

---

**THE FACTORS AFFECTING A DATA  
HARMONISATION INNOVATION IN THE  
WESTERN CAPE, SOUTH AFRICA**

---

BEY-MARRIÉ SCHMIDT

THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN  
THE SCHOOL OF PUBLIC HEALTH AND FAMILY MEDICINE AT THE  
UNIVERSITY OF CAPE TOWN

APRIL 2019

SUPERVISORS: A/PROFESSOR CHRISTOPHER J. COLVIN  
AND DR NATALIE H. LEON

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

# PLAGARISM DECLARATION

I, Bey-Marrié Schmidt, confirm that the work presented in this thesis is my own, both in concept and execution. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:

Date: 02 April 2019

Signed by candidate

# INCLUSION OF PUBLICATIONS

I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publication and papers in my PhD thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publication:

## **Publication:**

- Bey-Marrié Schmidt, Christopher J. Colvin, Ameer Hohlfeld, Natalie Leon. Defining and conceptualising data harmonisation: a scoping review protocol. *Systematic Reviews*. 2018(7):226

## **Unpublished papers:**

- The history of health information system strengthening in the Western Cape Province, South Africa
- Definitions and concepts of data harmonisation: a scoping review
- The motivations and opportunities, design process and operationalisation of a data harmonisation initiative in South Africa: an ethnographic case study
- Institutional and conceptual dilemmas and key success factors associated with data harmonisation processes

Signature:

Date: 02 April 2019

Signed by candidate

# ABSTRACT

Lack of coordination and integration between routine electronic databases can limit effective data production and utilisation to support health management decision-making. There is currently a need to strengthen data support structures through the harmonisation of multiple databases across different types of health services and organisations. Data harmonisation (DH) is an innovative process of copying existing electronic data captured in various databases into a centralised data repository where the data is integrated and then transformed into useable formats for data users.

However, there is limited evidence about the wide range of factors (especially social factors) that impact on DH innovations, such as historical factors, stakeholder relationships and institutional terrain. This doctoral research aimed to identify and explore the factors affecting a DH initiative currently underway in the Western Cape Province of South Africa. The research was conducted using three methodological approaches, namely a historical analysis and synthesis, a scoping review and an ethnographic case study.

For the historical analysis, relevant articles were identified through literature searches and data were collected through document reviews and interviews with two key informants. Data were first organised chronologically according to key events that took place in the health information system (HIS). Text from websites, journal articles, internal documents, standard operating procedures and interview notes were then synthesised according to key themes related to HIS interventions.

For the scoping review, systematic literature searches were conducted to identify studies that met the eligibility criteria of the review. Two review authors (one being the doctoral student) screened titles, abstracts and full-texts and then sampled studies based on the range, variation and similarities or differences in definitions and concepts and intervention descriptions. Manual coding and the filter option in Excel were used to provide (a) numerical analysis of the characteristics of included studies; (b) narrative synthesis of the different DH definitions, components and processes, as well as intentions, suggestions and/or explanations of how DH may lead to improved health management decision-making.

For the ethnographic case study, data were collected using participant observation (including conversations, meeting attendance and telephone and email communication), document reviews and in-depth interviews. Participants included data clerks, facility managers, health information staff and managers, DH innovators, researchers, public health specialists and database managers. Raw data were collected in the form of meeting minutes, field notes, interview notes and document extracts. Data analysis was conducted using thematic data analysis. The doctoral student manually coded data by highlighting recurring themes and evidence, and by extracting prominent themes from the various sources of data. As a strategy for testing the validity of emerging themes, the doctoral student used triangulation of different data sources; including looking for consistencies or inconsistencies between data sources.

Five main findings emerged from the doctoral research. The first finding affirms that DH is a multi-faceted intervention. In the literature, it is defined and described using different terms for similar aims and activities (such as record linkage, data warehousing, health information exchange). Key characteristics emerging from a synthesis of DH studies include: a process of multiple steps to integrate electronic data; different types of databases, institutions and technical activities; integrating data involves using unique patient identifiers; and framing interventions or activities around a specific scope or purpose (such as geographic area, disease surveillance and treatment management). DH interventions contributed to three levels of health management decision-making, namely clinical support, operational and strategic management, and population-level disease surveillance.

The second finding relates to the concept of ‘cultivation’. Cultivation is an ongoing and iterative social process to deal with problems between people, institutions and technology as they engage with each other in the context of an emerging innovation. The third finding is about striking a balance between the role of champions in designing and piloting innovations and the role of institutions in operationalising innovations and incorporating them into the broader health system for acceptance amongst implementers and users and for sustainability in the future.

The fourth finding is about the motivations and opportunities that contributed to the emergence of a DH initiative in the Western Cape Province of South Africa. Opportunities for the new DH

initiative include well-developed individual electronic databases, a government-university collaboration, and the positive attitude of frontline health workers towards DH projects. The new initiative faced design and operational challenges such as difficulty to access data from different health authorities and the incompleteness of electronic data. However, new data access and transfer procedures and existing social relationships were important for dealing with the changes that occurred as DH projects were being operationalised.

The last finding highlights tensions that emerged between DH innovators and other health information technology (HIT) stakeholders because of institutional and conceptual differences (such as different approaches to data access and governance, differences in conceptualisations of the value of data, and misunderstandings about the purpose of formal data procedures). DH innovators were able to navigate conflicts emerging from institutional and conceptual differences because of their strong leadership and team setup, institutional positioning and stakeholder engagement activities, to become institutionalised within the health system.

These findings provide health system, information technology and research stakeholders with a broader understanding of the range of social factors that impact on DH innovations. This research promotes a more comprehensive approach in designing, implementing and evaluating DH innovations to limit poor outcomes of innovations and wasted resources.

# ACKNOWLEDGEMENTS

This research would not have been possible without financial support from the Division of Social and Behaviour Sciences at the University of Cape Town (UCT), through the US National Institute of Mental Health and the National Research Foundation of South Africa, and time provided by the South African Medical Research Council (SAMRC).

I would like to thank my supervisors, A/Prof Chris Colvin at the School of Public Health and Family Medicine at UCT and Dr Natalie Leon at the SAMRC for their time, generosity and guidance throughout my PhD journey.

This research is a product of willing City of Cape Town, Western Cape Government, Provincial Health Data Center, and Health Information System Program staff and managers. Thank you for your time and kindness.

Thank you to the entire iALARM team, all the researchers and coordinators, for providing me with intellectual and administrative support. Particular thanks to Myrna van Pinxteren and Eleanor Whyte who have been a source of encouragement. I am grateful to the South African Social Science and HIV (SASH) Programme for all the writing retreats and professional development opportunities. I am also thankful for moral and intellectual support from Tamara Kredo, Ameer Hohlfeld and Sara Cooper at Cochrane South Africa, SAMRC.

I am eternally grateful to my mother, Mpingana Schmidt, and my late father, Paul Schmidt, who always prioritised my education, especially because they themselves had no opportunity to attend tertiary education. I will never be able to thank them enough.

And finally, to Ntsikelelo Maduneni, the man who married me four months before I embarked on the PhD journey. Thank you for supporting my career, for often sacrificing quality time and for always believing that I can complete my PhD studies even if you weren't always sure what it entailed.

# TABLES OF CONTENT

## Contents

<b>PLAGARISM DECLARATION</b> .....	<b>2</b>
<b>INCLUSION OF PUBLICATIONS</b> .....	<b>3</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>7</b>
<b>TABLES OF CONTENT</b> .....	<b>8</b>
<b>LIST OF TABLES AND FIGURES</b> .....	<b>13</b>
<b>LIST OF ABBREVIATIONS</b> .....	<b>15</b>
<b>CHAPTER 1: INTRODUCTION</b> .....	<b>18</b>
1. Overview.....	18
2. Literature review: The relationship between data harmonisation, routine health information systems and health systems strengthening .....	21
2.1. The complex health system.....	21
2.2. The role of RHISs in health system strengthening.....	24
2.3. The role of DH as a HIT innovation to RHISs .....	29
2.3.1. Defining data harmonisation .....	29
2.3.2. How data harmonisation has the potential to enhance HIV care management .	33
2.3.3. Barriers and facilitators of data harmonisation innovations .....	36
3. Conceptual frameworks .....	39
3.1. Performance of Routine Information System Management (PRISM) framework	41
3.2. Complex Adaptive Systems approach .....	44
3.3. Interactive Sociotechnical Analysis approach .....	47
3.4. Summary of the methodological frameworks.....	49
4. Problem statement.....	49
5. Study objectives .....	50
6. Thesis structure.....	51
<b>CHAPTER 2: METHODOLOGY</b> .....	<b>53</b>
1. Country setting .....	53
2. Study setting.....	55
3. Methods .....	58
4. Study limitations .....	62
5. Methodological frameworks .....	63
5.1. Sociotechnical Evaluation.....	63

<b>CHAPTER 3: THE HISTORY OF HEALTH INFORMATION SYSTEM STRENGTHENING IN THE WESTERN CAPE PROVINCE, SOUTH AFRICA .....</b>	<b>67</b>
<b>1. Introduction .....</b>	<b>67</b>
<b>2. Methods.....</b>	<b>70</b>
2.1. Data collection and data analysis.....	70
2.2. The old South Africa .....	72
<b>3. Findings .....</b>	<b>74</b>
3.1. Key HIS interventions after apartheid.....	74
3.1.1. The emergence of a HIS initiative: Health Information System Program .....	75
3.1.2. The emergence of a HIS intervention for primary health care: minimum data set	77
3.1.3. Developing a HIS intervention from the bottom up: District Health Information Software.....	80
3.1.4. National indicator dataset (NIDS) .....	81
3.2. The barriers, facilitators and outcomes of HIS interventions.....	84
3.2.1. Barriers .....	84
3.2.2. Facilitators .....	89
<b>4. Discussion .....</b>	<b>91</b>
<b>CHAPTER 4: DEFINITIONS, COMPONENTS AND PROCESSES OF DATA HARMONISATION: A SCOPING REVIEW .....</b>	<b>96</b>
<b>1. Introduction .....</b>	<b>103</b>
<b>2. Methods .....</b>	<b>104</b>
2.1. Study objectives .....	105
2.2. Identifying relevant studies .....	105
2.2.1. Eligibility criteria.....	105
2.2.2. Search strategy.....	106
2.3. Selecting studies for inclusion .....	106
2.3.1. Screening records .....	106
2.3.2. Sampling of studies.....	107
2.4. Data extraction .....	108
2.5. Data synthesis: collating, summarising and reporting findings.....	108
2.6. Reflexivity .....	109
<b>3. Findings .....</b>	<b>110</b>
3.1. Results of the search.....	110
3.2. An overview of key characteristics of data harmonisation studies.....	112
3.3. Definitions, components and processes of data harmonisation.....	112

3.3.1. Alternative terms and definitions of data harmonisation .....	113
3.3.2. Components and processes of data harmonisation .....	116
3.4. The relationship between data harmonisation and health management decision-making .....	120
<b>4. Discussion .....</b>	<b>123</b>
4.1. Synthesis of findings .....	123
.....	125
.....	125
4.2. Study limitations .....	126
4.3. Implications for research and practice .....	126
4.4. Conclusion .....	127
<b>CHAPTER 5: THE MOTIVATIONS AND OPPORTUNITIES, DESIGN PROCESS AND OPERATIONALISATION OF A DATA HARMONISATION INITIATIVE IN SOUTH AFRICA: AN ETHNOGRAPHIC CASE STUDY .....</b>	<b>139</b>
<b>1. Introduction .....</b>	<b>139</b>
<b>2. Methods .....</b>	<b>144</b>
2.1. Setting: The Provincial Health Data Center .....	144
2.2. The iALARM project .....	146
2.3. The ethnographic study on DH innovations .....	149
2.4. Using the Sociotechnical Evaluation methodological framework to design our ethnographic study .....	150
2.5. Data collection methods .....	151
<b>3. Findings .....</b>	<b>153</b>
3.1. Overview .....	153
3.2. Motivations and opportunities leading to data harmonisation innovations in the Western Cape Province .....	153
3.3. Emerging design and operational challenges and responses .....	159
3.4. Balancing tensions emerging from innovating and institutionalising new data procedures .....	164
<b>4. Discussion .....</b>	<b>167</b>
<b>CHAPTER 6: INSTITUTIONAL AND CONCEPTUAL DILEMMAS AND KEY SUCCESS FACTORS ASSOCIATED WITH DATA HARMONISATION PROCESSES .....</b>	<b>172</b>
<b>1. Introduction .....</b>	<b>172</b>
<b>2. Methods .....</b>	<b>177</b>
2.1. The Provincial Health Data Center (PHDC) .....	177

2.2.	The i-ALARM study .....	179
2.3.	Data collection and analysis .....	180
3.	Findings .....	181
3.1.	Overview .....	181
3.2.	Institutional dilemmas of DH innovations .....	181
3.2.1.	Fragmentation of routine electronic data managed by different institutions... ..	182
3.2.2.	Tensions between centralised and decentralised approaches to data access and governance .....	184
3.2.3.	Lack of ongoing working relationships between City and Province health authorities .....	185
3.2.4.	Tensions that emerged as a result of the purpose and power of PHDC's data repository project .....	186
3.3.	Conceptual dilemmas of DH innovations.....	190
3.3.1.	Tensions between scientific and practical conceptualisations of the value of data .....	190
3.3.2.	Misunderstandings about the purpose of formal data practices .....	191
3.4.	Navigating institutional and conceptual dilemmas .....	193
4.	Discussion .....	195
<b>CHAPTER 7: DISCUSSION AND CONCLUSION .....</b>		<b>200</b>
1.	Summary of findings .....	200
2.	Positioning the research findings in the literature .....	204
2.1.	The barriers and facilitators of data harmonisation innovations.....	204
2.1.1.	Data harmonisation innovations in low- and middle-income settings versus high-income settings.....	206
2.1.2.	Different phases of data harmonisation innovations.....	209
2.1.3.	Wide versus narrow scope of data harmonisation innovations.....	211
2.1.4.	Section summary.....	212
2.2.	Strategies for navigating the barriers and facilitators of data harmonisation innovations .....	213
3.	Implications for research .....	217
4.	Recommendations for practice.....	217
5.	Strengths and limitations .....	219
6.	Conclusion .....	220
References.....		222
Appendices.....		230
Appendix 1. Search strategies .....		230
Appendix 2. Ethics approval for PhD project .....		231

**Appendix 3. Ethics approval for iALARM study ..... 233**  
**Appendix 4. Consent form..... 235**  
**Appendix 5. Interview guide ..... 238**

# LIST OF TABLES AND FIGURES

## **Chapter 1:**

Table 1. Barriers and facilitators of HIE

Figure 1. The WHO Health Systems Framework

Figure 2. The Health System Dynamics Framework

Figure 3. The components of a health information system

Figure 4. The steps for producing quality health information

Figure 5. PRISM framework

Figure 6. Complexities of HIT innovations in health care systems

Figure 7. Interactive Sociotechnical Analysis

## **Chapter 2:**

Table 1. Methods for primary research conducted in Chapters 3, 5 and 6

Figure 1. Interactions between technology and social processes in HIT innovations

Figure 2. Key components of the Sociotechnical Evaluation approach

## **Chapter 3:**

Table 1. HIS interventions, stakeholders and key events

Table 2. Barriers, facilitators and outcomes of HIS interventions

## **Chapter 4:**

Appendix 1. Search strategies

Table 1. Characteristics of included studies (n=181)

Table 2. Characteristics of sampled studies (n=61)

Table 3. Alternative terms and definitions of data harmonisation interventions

Table 4. Concepts of data harmonisation interventions and processes

Table 5. The relationship between DH interventions and health management decision-making

Figure 1. Extract of the Excel data extraction form

Figure 2. PRISMA diagram of eligible studies

Figure 3. A concept map of data harmonisation and its relationship to health management decision-making

### **Chapter 5:**

Table 1. Routine electronic databases relevant for building the iALARM cascade

Figure 1. Sociotechnical interactions

Figure 2. Data harmonisation at PHDC

Figure 3. The HIV treatment cascade, key indicators and relevant routine health information databases in the Western Cape

Figure 4. Key components of the Sociotechnical Evaluation approach

### **Chapter 6:**

Figure 1. Interactive Sociotechnical Analysis (ISTA) framework

### **Chapter 7:**

None

# LIST OF ABBREVIATIONS

AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral Therapy
ASOs	AIDS Service Organizations
CAS	Complex Adaptive Systems
CD4	Cluster of Differentiation 4
CDWH	Clinical Data Warehouses
CHCs	Community Health Centers
CHIC RHIO	Carolina HIV Information Cooperative Regional Health Information Organisation
DH	Data Harmonisation
DHA	Department of Home Affairs
DHIS	District-based Health Information Software
EHR	Electronic Health Record
EKAPA	Evaluation of the Khayelitsha AIDS Programme
EMR	Electronic Medical Record
FTP	File Transfer Protocol
HCP(s)	Health Care Provider(s)
HCT	HIV Counselling and Testing
HCWs	Health Care Worker(s)
HIE	Health Information Exchange
HIM	Health Information Management
HIS(s)	Health Information System(s)
HISP	Health Information System Program
HIT	Health Information Technology
HIV	Human Immunodeficiency Virus
HREC	Human Research Ethics Committee
iALARM	Using Information to Align Services and Link and Retain Men in the HIV Cascade
IDI	Infectious Disease Informatics
ISTA	Interactive Sociotechnical Analysis

IT	Information Technology
LaPHIE	Louisiana Public Health Information Exchange
LMICs	Low- and Middle-Income Countries
MDS	Minimum Data Set
MeSH	Medical Subject Headings
MPI	Master Patient Identifier
NCD(s)	Non-communicable disease(s)
NDoH	National Department of Health
NGO(s)	Non-governmental Organisation(s)
NHISSA	National Health Information Systems of South Africa
NHLS	National Health Laboratory Services
NIDS	National Indicator Data Set
NORAD	Norwegian Agency for Development Cooperation
NPO	Non-profit Organisation
NUFU	Norwegian University Council
PHC	Primary Health Care
PHCIS	Primary Health Care Information System
PHDC	Provincial Health Data Center
PREHMIS	Patient Registration and Health Management Information System
PRISM	Performance of Routine Information Systems Management
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RDP	Reconstruction and Development Programme
RHIS(s)	Routine Health Information System(s)
RMR	Routine Monthly Report
STE	Sociotechnical Evaluation
STI(s)	Sexually Transmitted Infection(s)
TB	Tuberculosis
UCT	University of Cape Town
UNAIDS	United Nations Programme on HIV/AIDS
USA	United States of America
UWC	University of Western Cape

WCDoh Western Cape Provincial Department of Health  
WHO World Health Organization

# CHAPTER 1: INTRODUCTION

---

The introduction chapter is divided into three sections. The first section provides an overview of the thesis topic. The second section is a literature review on the relationship between data harmonisation (DH), routine health information systems (RHISs) and health systems strengthening. And the last section is a synthesis of four conceptual frameworks that are useful in shaping the methods and interpreting the findings for the gaps in literature.

## 1. Overview

Data harmonisation (DH), or the process to harmonise data captured across different sources, has emerged amongst health authorities and managers, information technology (IT) stakeholders and researchers as an important intervention for RHISs. A RHIS is a system that collects, distributes and uses health information at regular intervals (such as monthly, quarterly or annually) through predictable mechanisms (such as reports or audits) to address specific decision-making needs. DH aims to address problems of standardisation, fragmentation, and lack of coordination and integration of data. Thus when successfully implemented, DH has the potential to improve data production and utilisation to support health management decision-making by integrating RHISs across various types of health facilities and organisations into single health information exchange (HIE) platform [133]. Health management decision-making is a proactive and interactive process that should demand and use the best available data (well-integrated, complete and accurate data) during health system improvements, as well as monitoring and evaluation [6, 9]. Thus, DH is not only an intervention to integrate electronic databases, but it is also an intervention which can influence and be influenced by the demand and use of health information for decision-making.

Unfortunately, DH has often been viewed as merely a technical solution, or a health information technology (HIT) innovation, to apparently simple technical problems. Views opposing this narrow focus on HIT innovations as merely technical are well-documented over the past two decades [42, 51]. HIT designers and implementers have cautioned against poor outcomes of innovations and the potential waste of resources when the impact of social factors is neglected

during development and implementation processes [42, 43]. Even though there is growing awareness that DH is shaped by both social and technical contexts, there is limited evidence on the social factors that impact on the development and implementation of DH innovations [8, 44, 51]. This PhD project aimed to address that gap in knowledge by exploring the social challenges, opportunities and processes of a DH innovation. I did so, primarily, by observing the operationalisation and institutionalisation of a DH initiative that is emerging in the Western Cape Province of South Africa.

There are two opportunities that made this PhD project possible. The first opportunity has to do with the new DH initiative that aimed to integrate multiple disparate databases capturing province-wide health data [112]. And the second opportunity had to do with a five-year HIV study that aimed to use routinely collected and newly harmonised health data to coordinate the work of health system and community-based service providers. A collaboration was established between the new DH initiative and the HIV study; the DH initiative aimed to test design processes for harmonising longitudinal data across a certain disease or cohort and the HIV study aimed to make use of this innovation to better understand the performance of HIV services in the province. This unique collaboration presented me with an opportunity to study DH innovations as they were unfolding in our local setting. The HIV study needed someone who could help facilitate the process of accessing data from the DH initiative. I took on a dual role as an iALARM researcher and doctoral student where I facilitated the process of accessing data for the HIV study which gave me access to participating in and observing some of the activities of the DH initiative for my PhD research. I was able to learn more about how the DH initiative came about, what its purpose was, who its key stakeholders were and what challenges it faced. My PhD project aimed to identify and explore how existing social relationships, organisational processes and technical systems interacted with the new DH innovation, given a particular historical context of health information system (HIS) interventions in South Africa. The sections of this chapter are organised as follows:

- a literature review of the relationship between the health system, RHIS and DH
- a synthesis of four conceptual frameworks on health information technology innovations (which guided the conceptualisation and methods of the PhD project)
- a summary of the study rationale

- an outline of the study objectives
- and an outline of the thesis structure.

## **2. Literature review: The relationship between data harmonisation, routine health information systems and health systems strengthening**

The purpose of the literature review is to provide a description of the relationship between DH, RHISs and health systems strengthening. This provides context for how DH innovations can strengthen RHISs for health systems performance, considering certain social challenges, opportunities and processes. The review also focuses on studies that describe the design, development and implementation of DH innovations for improving the quality and delivery of chronic care services (such as HIV), given that this PhD project is embedded in an HIV study.

The literature review is organised according to the following sub-sections:

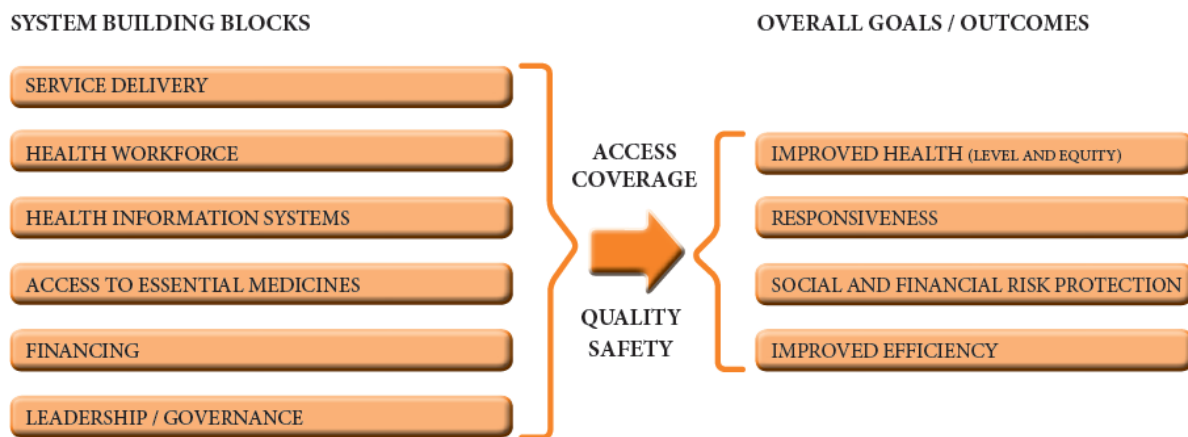
- the complex health system
- the role of RHISs in health systems strengthening
- DH as a HIT innovation to RHISs

### **2.1. The complex health system**

The health system is a complex system consisting of many components that influence each other and overall health system performance [1]. Ensuring that the health system performs at its best (that is, health system strengthening) is currently an important topic on the global public health agenda [2, 3]. Well-functioning RHISs are able to provide health systems with informational support for efficient and effective performance. DH innovations (which are interventions to harmonise data captured across different sources) have the potential to strengthen RHISs [4, 5]. Ultimately, health systems performance relies on various organisations, resources (including well-functioning RHISs) and people whose primary purpose is to improve health outcomes [1, 6]. It includes direct and indirect efforts to influence health determinants and health outcomes, through the delivery of preventive, promotive, curative and rehabilitative services. The health system needs to be responsive and financially efficient, which requires staff, money, supplies, transport, communication, and guidance [1].

The World Health Organisation (WHO) has identified six ‘building blocks’ that are central to the functioning of a health system; these are: service delivery, health workforce, health information system (HISs), access to essential medicines, financing, and leadership and governance [7]. According to the WHO Health Systems Framework, the building blocks presented in **Figure 1** can contribute to the strengthening of a health system in three ways.

**Figure 1.** The WHO Health Systems Framework

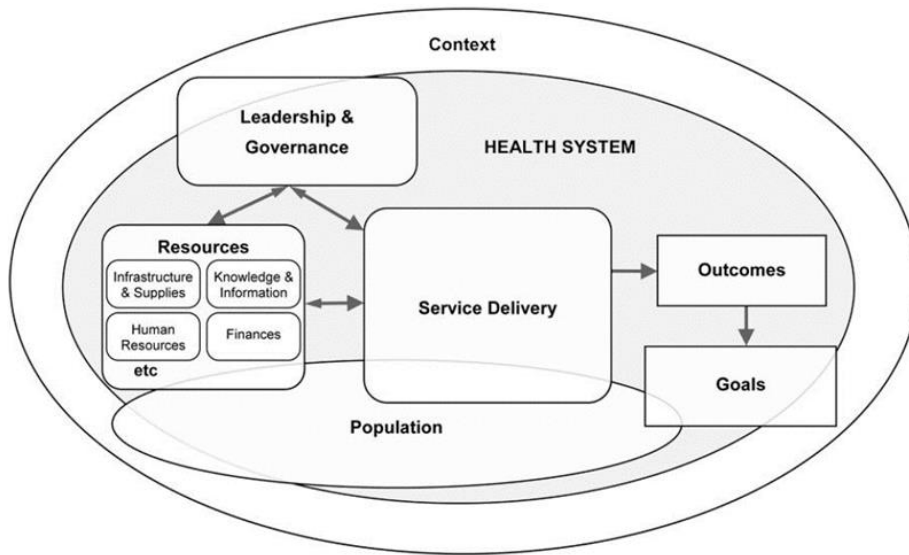


Firstly, leadership and governance and HISs can provide support to decision- and policy-making processes across all of the other building blocks; secondly, financing and health workforce are the key input components of a health system; and thirdly, access to essential medicines and service delivery reflects the immediate outputs of a health system, that is the availability and distribution of care. This framework is useful for defining the health system and dividing it into building blocks to monitor performance and optimise interventions. However, the links and interactions between these building blocks as well as the actions that influence people’s behaviours are not fully captured in the WHO Health System Framework [7].

The Health System Dynamics Framework builds onto existing frameworks, such as the WHO Health System Framework, by recognising the multifaceted nature of the health system, that is, it is comprised of complex interactions and equilibriums between different elements [1]. Instead of six building blocks, this framework consists of ten elements that focus on system interactions and

values that are central to the behaviours of people, which in turn influences choices and processes in a health system. **Figure 2** presents these elements and their dynamic interactions which are: goals and outcomes, values and principles, service delivery, population, context, leadership and governance, and the organisation of resources (finances, human resources, infrastructure and supplies, and knowledge and information) [1].

**Figure 2.** The Health System Dynamics Framework [1]



Essentially, the Health System Dynamics Framework is based on the notion that the health system is a complex adaptive social system; it is made up of many actors and organisations that interact with each other. These interactions or processes of communication and coordination are often not linear, and they are hard to predict, but they can result in actions that lead to temporary equilibriums [1]. Further, the health system is a social system, because it is an open system that is influenced by changes due to context or population, while it draws resources from its environment to respond to new needs. Like other social systems, the health system may anticipate new needs or changes and adapt and respond accordingly by taking advantage of available opportunities to improve interventions, health services and health outcomes [1, 8].

Leadership and governance are critical for health system responsiveness, that is engaging new challenges and optimising opportunities, so health managers and policy-makers rely on routinely

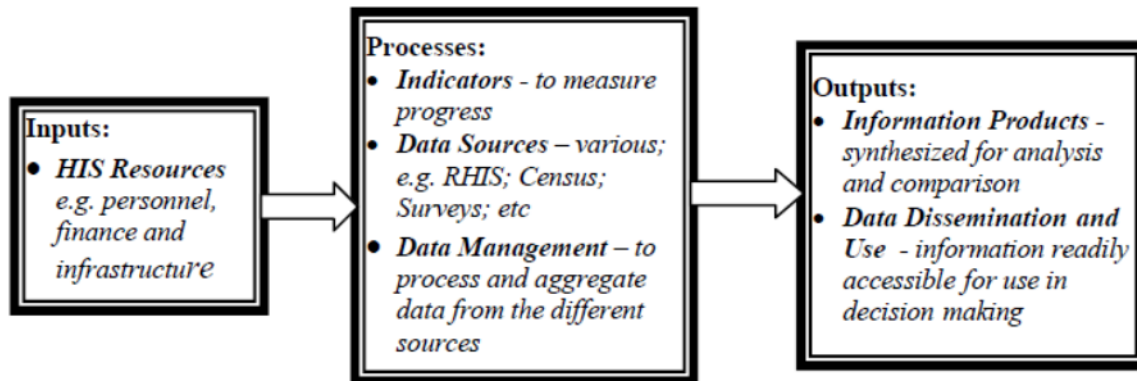
collected 'knowledge and information' captured in HISs for planning, monitoring and evaluation. **Figure 3** illustrates the inputs, processes and outputs required for a HIS to produce relevant, quality and timely information which can be used to support decision-making and thereby strengthen health system performance [9]. The relationship between a HIS and health system performance is explained in detail below.

## **2.2. The role of RHISs in health system strengthening**

For clinical and management decision-making to be effective, health system authorities and managers require high-quality, relevant and easily accessible routinely produced health information [6, 10]. A well-functioning health information system (HIS) is a system that collects, distributes and uses information, while a RHIS is able to specifically collect and distribute information for use at regular intervals through mechanisms designed to meet predictable information needs [10]. Examples of RHISs are surveillance systems (for identifying the incidence of death), patient health or medical records (to track the clinical care delivered to individuals), facility-based information management systems (used by managers to track the delivery of health services) and administrative and related support systems (for human resource, finance, drug supply and diagnostic services) [10].

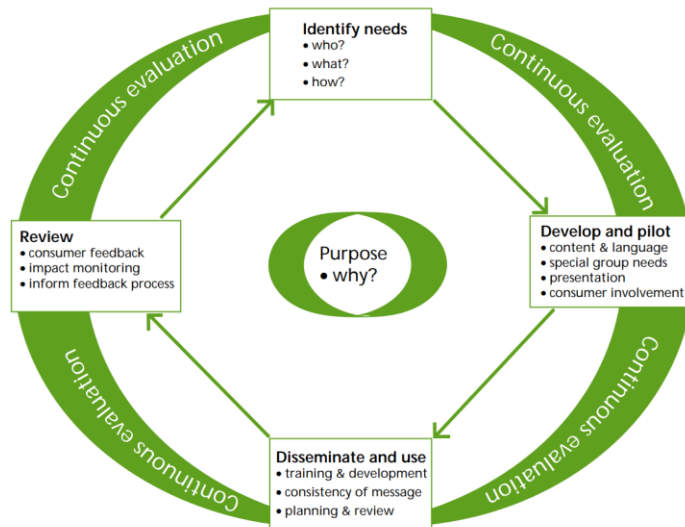
The relationship between a RHIS and health management decision-making can be illustrated using **Figure 3**.

**Figure 3.** The components of a health information system [9]



Health management decision-making should be a proactive and interactive process that demands and uses the best available data (ideally well-integrated, complete and accurate data) for service improvements and general monitoring and evaluation of the health system [6, 9]. As mentioned before, a HIS has three components: inputs, processes and outputs. The inputs of a HIS are the resources for capturing information into the system such as trained personnel and computers. Once information is inputted into a HIS it undergoes processes of producing aggregate data which is compiled from indicators such as key measurements and various data sources such as demographic, clinical, laboratory, pharmacy and administrative databases. This results in a HIS producing outputs or information products such as reports that can be used for analysis and decision-making. Since the components of a HIS are inter-linked, the path to producing quality health information for decision-making is complicated. This is because good quality health information for decision-making (by patients, clinicians, health managers and policy-makers) should clearly state its aims, be relevant, be accurate, be accessible, comprehensible and acceptable to its audience, help users to identify further sources of information and support, and help users to judge how reliable the information is. The steps for planning and producing quality health information are presented in **Figure 4**:

**Figure 4.** The steps for producing quality health information [134]



Even when the steps outlined in **Figure 4** are satisfied (that is, HISs consist of good quality health information), the HIS components (presented in **Figure 3**) still need to work together to encourage managers and decision-makers to demand and use information. The components of a HIS may function well separately, but their integration is what results in a well-functioning system. A well-functioning HIS brings together all relevant data sources and stakeholders to ensure that users of health information have access to reliable, authoritative, useable, understandable and comparative data [135].

Many developing countries lack well-functioning RHISs (where all components are integrated) that encourage demand for and effective use of health information for evidence-based decision-making [4]. Problems related to RHIS include the production of poor quality data that do not meet the needs of decision-makers; key variables and indicators for collecting, analysing and reporting health information vary across programmes so data cannot be harmonised into useable formats [11]; and fragmentation across levels of the health system can result in high burdens of workload for health workers, a consequence of duplication and excessive production of data [12]. Additionally, even when high-quality health information exist, it may be of low value for management decision-making if it is not transformed into useable formats and if feedback mechanisms are not established. Feedback reports received at lower levels are frequently delayed and information is obsolete for decision-making because facility managers face deadlines and time

constraints in their daily decision-making [4, 13] However, even where relevant health information is readily accessible, it cannot be assumed that this information will be used to inform health management decision-making [14].

The barriers for effective use of health information for management decision-making can be organised using the Performance of Routine Information Systems Management (PRISM) framework, which is one of the four conceptual frameworks that will be described in detail later in this section. PRISM identifies three levels that influence the performance of a RHIS, namely technical, organisational and behavioural determinants. One of the key objectives of a RHIS is data use. The use of data for health management includes the analysis, synthesis, interpretation and review of data as part of the decision-making process across levels and programmes of a health system [4]. Barriers to effective use of health information can occur at any stage of the process of collecting, producing and using data, and as described below, at any or all of the three levels identified by PRISM.

Firstly, the technical determinants of a RHIS are all the factors that relate to the specialised technology to develop, manage and improve RHIS processes and performance. These factors include the development of indicators, the types of information technology, and the development of software for data processing and analysis [13]. Technical barriers that result in weak RHISs are poor design of data sources, lack of technical interoperability among existing systems (the inability for data sources to exchange information), and absence of common metadata (standardised data definitions, data sources, frequency of reporting and levels of use) [15]. Technical interventions, such as DH, are widely recommended for addressing such technical barriers to the production and utilisation of information. However, there is a gap in the literature about the causal relationship between technical interventions and health management decision-making [16]. There is less theoretical literature explaining how and why the relationship between technical interventions and health management decision-making exists, that is, evidence on the casual mechanisms that result in technical interventions increasing health management decision-making.

Secondly, PRISM recognises that a RHIS and its users operate in an organisational context that is influenced by many social factors (a health system is itself a social system). RHIS users are

influenced by organisational rules, values and practices in terms of their work processes and performance [13]. For example, factors relating to organisational structure, resources, data procedures and support services promote a specific type of culture of information through the values and beliefs within the organisation for collecting, analysing and using information to achieve the organisation's goals [15]. There is evidence that the culture of information (especially in developing countries) is centred on the demand for massive data collection and reporting, even though there is under-utilisation of health information for decision-making at local levels [17]. There is a phenomenon called path dependency that explains that historical dynamics (that is rules, values and practices) of an organisation shape the processes and performance of a RHIS to such an extent that the RHIS is limited in evolving with organisational and technological changes [18]. Path dependency states that, in some instances, it becomes increasingly difficult to make changes to existing processes or paths for producing and using health information without incurring significant costs and restructuring.

Interventions for overcoming organisational barriers to RHISs are important for creating environments that encourage effective use of information for decision-making across all levels of the health system and for improving existing RHISs [10]. Organisational interventions to RHISs often go hand-in-hand with behavioural interventions, because they both address organisational and behavioural barriers to data quality and information use. This means, it is not only organisational rules, values and practices that influence the manner in which data is produced and used, but also the confidence, motivation and competence of health staff.

Lastly, RHIS users' demand, confidence, motivation and competence to perform RHIS tasks affects RHIS processes and performance directly [13]. This means that people's perceptions about the importance of a task and the level of difficulty to perform that task potentially determine whether that task will be performed or not [15]. Negative attitudes, for example, from a lack of knowledge of the usefulness of RHIS data, of those who work on RHIS tasks can hinder RHIS processes and performance. Other behavioural barriers to RHIS relate to narrow programmatic interests, inadequate training, and the lack of appropriate RHIS skills of health managers and providers. These barriers are influenced by both technical and organisational factors of RHISs in

that specialised technologies and initial designs of HISs, and institutionalised rules, values and practices can limit demand for health information [13, 15].

The next sub-section will provide insights into the interactive relationship between technical, organisation and behavioural components. DH has been identified as a technical intervention for strengthening RHISs, but it can influence and be influenced by social elements of a RHIS. This is because DH has two main aims: firstly, for two or more systems to exchange information (that includes produce, extract, merge, link and clean data); and secondly, for those systems to make information accessible and for data users to demand and use the information that has been exchanged. Both these aims relate to social elements of a RHIS, but for this PhD project I focused on the second aim which involves organisational structures and culture, individuals' perceptions and political motivations for demanding and using information.

The first aim can be viewed as mostly technical, however, the second aim includes social drivers such as organisational culture, individuals' perceptions and political agendas for using information. Below, I explore this multi-faceted nature of DH as a HIT innovation.

## **2.3. The role of DH as a HIT innovation to RHISs**

### **2.3.1. Defining data harmonisation**

There are a range of definitions and terms used for DH because of the dynamic technical and social contexts in which its design, development and implementation processes occur. There are alternative terms for DH which are used to describe its aims and activities such as HIE, electronic health records, data linkage and data warehousing. Below, is a review of four different descriptions of DH (where DH or a similar activity is sometimes referred to as a different term) that can usefully contribute to a working definition for the thesis.

The first definition to consider is from a systematic review assessing the barriers and facilitators to HIE in LMIC settings. The review defined HIE as follows:

*“Health information exchange (HIE) is the electronic mobilisation of clinical and administrative information within or across organisations in a region or community and, potentially, internationally between various systems according to locally and/or nationally recognised standards while maintaining the authenticity and accuracy of the information being exchanged, enabling stakeholders to make informed decisions to enhance healthcare quality of a patient and population. HIEs are multi-stakeholder organisations that oversee the business, operational and legal issues involved in the exchange of information.” [19].*

This definition highlights three important points; firstly, DH can take place across different settings related to databases, geographic areas and institutional levels or systems. Secondly, the data being exchanged should be real and accurate for stakeholders to make informed decisions that can enhance the quality of health services. Thirdly, harmonising data into a HIE platform involves multiple stakeholders who are concerned with the business, operational and legal aspects of such HIT projects.

The second definition is from a Chinese study that aimed to collect health data items from various HISs nationally and harmonise them into an electronic health record (a large centralised data repository) using a conceptual data model. The study defined the process of DH as follows:

*“EHR [Electronic Health Records] should be longitudinal and contain all the information about a person’s health... To set up EHR, which is patient-centric, firstly we, in the perspective of individuals’ healthcare, have to identify whether all the information necessary for EHR is available in existed systems, where the information is, and how the information is defined and formatted. Secondly, the heterogeneous information recorded by various systems should be made consistent or at least comparable with one another by reviewing, matching, redefining and standardizing each data item. For EHR will not be realized until the life-long health data of a person that are distributed in inhomogeneous information systems can be integrated, the process of identifying, reviewing, matching, redefining and standardizing information, which is defined as data harmonization...” [20].*

This study states that an EHR should contain individual-level data harmonised in a longitudinal manner. Therefore, to set up an EHR, DH designers should be well-acquainted with existing databases, in terms of where different data is located and in what format. The next step involves making the existing databases interoperable by reviewing, matching, redefining and standardising the data items. The implementation of an EHR relies on the successful transformation of health data for patient-centric care.

The third definition is from a recently conducted systematic review on the views of health care professionals to linkage of routinely collected health care data and to identify any potential barriers and facilitators to participation in a data linkage system. The review described data linkage as follows:

*“Linkage at individual patient level of routinely acquired health data between primary and secondary care could be important... Linkage of routine healthcare datasets by unique patient identifiers could provide an alternative or complementary approach to the identification of [a specific outcome]. It would permit following exposed individuals in real time and provide a denominator. Routine data linkage would also enable creation of exposure cohorts in order to monitor long-term outcomes and enable a more efficient screening for [the outcome] due to an ever-increasing data pool” [21].*

Data linkage was described as an important process for integrating individual patient-level data that is captured at the primary and secondary levels of care. This linkage can be done using unique patient identifiers to facilitate access to data for specific outcomes and to follow patients in real time. Data linkage can also make it possible to build disease-specific cohorts and monitor long-term outcomes (for example, for chronic diseases) or to monitor a certain group of individuals who are most at risk in relation to a specific outcome or dropping out of care.

The last definition is from a German study that aimed to develop a data warehousing approach that would enable researchers to access health data collected in hospital data repositories. The study provides two approaches to data warehousing, which are as follows:

*“CDWH [Clinical Data Warehouses] are generally built on one of two predominant architectural paradigms: either, data is directly extracted, transformed and loaded from applications systems and databases into a data mart, i.e. an integrated view over a defined subject, or it is stored in a centralized data repository from which data marts or views can be established. Both approaches rely on a process to extract data from sources, transform it appropriately and to load it to a target database schema (ETL process). The ETL process is necessary to gather data originating from diverse sources (databases, application systems, messages), to “clean it up”, make it fit a common information model and to harmonize terminologies.” [22].*

Data warehousing was described as consisting of various technical processes and activities for integrating patient-level data into a centralised data repository. One of the products that can be created using the centralised repository is a data mart (an application for viewing harmonised patient-level data). This application can help health staff access patient-level data to support clinical management decision-making.

These four descriptions were selected from the literature because of the different terms used for DH. All of them imply that DH is a technical intervention (there is electronic integration of databases and systems); it involves a process of moving data from one place to another; then transforming the data (a secondary data production process occurs) through various activities (such as reviewing, merging, linking and cleaning); and making the data accessible and useable for researchers and health staff (which is implied in the first and the third definitions).

The working definition of data harmonisation for this thesis is thus as follows:  
Data harmonisation is an innovative process of copying existing electronic data captured in various databases into a centralised data repository where the data is integrated and then transformed into useable formats for data users.

Given this working definition of DH and the explanation of what a HIS is presented earlier (as a system that collects, distributes and uses health information), the relationship between DH, HIT and HIS is clearer. DH can through the application of innovative, new and effective technologies

(that is, HIT) bring together data from different HISs or RHISs (which have the same functionality as HISs, except that the data is collected, distributed and used at regular and predictable intervals) into a HIE platform. HIE allows for harmonised and transformed data to be shared and used by specified data users.

What seems to be missing from these descriptions of DH [19-22] is a clear link between the technical activities and their intended outcomes. Even though DH involves the technical process of integrating databases, its key outcome is to increase the demand and use of health data for service improvements and overall health systems performance [9, 23]. The problem is that DH is often viewed as merely technical; there is a technical problem (that is, fragmentation of databases) and there is a technical solution (that is, integrating databases through computer software and applications). There is limited reference to: the social aspects that result in DH or shape its design and implementation (such as health staff's demand for efficient technologies); or the impact that DH may have on social aspects (such as changes to stakeholder relationships). Below, I provide examples of the intended outcomes of DH, specifically using the case of chronic illnesses such as HIV and non-communicable diseases. I then explore some of the barriers and facilitators that shape the design, development and implementation of DH innovation.

### **2.3.2. How data harmonisation has the potential to enhance HIV care management**

It is widely reported in the literature that DH has the potential to strengthen RHISs to provide informational support to the other building blocks of the health system, including health service delivery [21, 108, 133]. Although the effectiveness of DH on patient and health system outcomes is not well-studied (there are few or no experimental or longitudinal studies), there are descriptive studies (such as cross-sectional or qualitative studies) that highlight the potential benefits of DH.

Harmonised health services data that are often kept in RHISs are useful for planning, implementing, and monitoring and evaluating health service delivery [24, 25]. It can serve as an early warning system for impending health service emergencies, it helps health managers to assess the impact of interventions or tracks progress towards specific goals, it is useful for identifying health problems and priorities, and it can be used to inform health service policy and strategies

[24, 25]. In the context of HIV management, harmonised databases can facilitate access to and retrieval of clinical data to provide efficient, effective and equitable patient-centred care. Thus, DH is an important intervention for providing health staff with the necessary information for tracking, linking and retaining people who need to be enrolled in HIV testing, treatment and care services [26]. Harmonised databases are necessary for establishing surveillance systems for vulnerable groups such as those people at risk of dropping out of services [23]. The studies described below focus on DH innovations to support service delivery, specifically for chronic care management like HIV care.

Three American studies [26-28], one conducted in Louisiana, another in North Carolina and the last one across six sites, describe similar DH innovations that were implemented to enhance quality and continuity of HIV care and improve patient outcomes among HIV-positive individuals. In the first study, the Louisiana Public Health Information Exchange (LaPHIE) was developed to facilitate secure data exchange between a public health surveillance data platform and electronic medical record system [27]. LaPHIE alerts health service providers when HIV-positive individuals who have not received HIV care for more than 12 months are seen at any ambulatory or inpatient facility within the service delivery network. The study found that the LaPHIE significantly reduced otherwise missed opportunities to intervene with individuals not in HIV care who were accessing other health services [27].

In the second study, the Carolina HIV Information Cooperative Regional Health Information Organisation (CHIC RHIO) “was implemented to improve patient care and health outcomes by enhancing communication among geographically disconnected networks of HIV care providers in rural North Carolina” [28]. CHIC RHIO comprises one medical clinic and five AIDS Service Organizations (ASOs) serving clients in eight rural counties. The study reports that CHIC RHIO members felt that the intervention benefited them and their clients; for example, improvements were observed in data quality, clinic-AOS relationships, information exchange and perceived level of patient care [28].

The last study assessed the HIV services delivery using quality and coordination indicators from the electronic data systems of six sites and described the HIE interventions designed to enhance

service delivery [26]. The sites developed HIE interventions that aimed to integrate data from different types of data systems (such as surveillance, electronic health records, laboratory and billing). The intended outcome of these HIE interventions were improved linkage and retention, quality and efficiency of care and increased access to patient information. The study reported that HIE can play an integral role in bringing together the fragments of the health system to improve health outcomes for people with HIV as well as those at risk for HIV [26].

In terms of the South African context, one study assessing a DH innovation for non-communicable diseases (NCDs) is worth noting [29]. The South African study attempted to enumerate NCD patients in a rural district of KwaZulu-Natal using two innovative techniques, electronic data linkage and capture-recapture (CR). The researchers first attempted to construct a district register for NCDs, using as many data sources as possible, which they described as an electronic data linkage. Secondly, they tried to use a statistical method (called capture-recapture) to analyse the multiple patient lists for overlap and estimate the total NCD population (both counted and uncounted). The study reported that implementing an electronic data linkage system for NCDs (or other chronic illnesses) by name, age and diagnosis is feasible, but because of little overlap between data sources, capture-recapture calculations were not possible [29].

There are a few key differences and similarities amongst the studies. Although these are only four studies from the literature, there is widespread recognition that DH innovations are particularly useful for strengthening RHISs of long-term cohorts or chronic illnesses. The American studies reported that databases were sufficiently advanced for information exchange to take place while the South African study was more experimental in terms of testing an electronic data linkage technique using capture-recapture calculations. Important considerations for DH innovations can be drawn from each of the four studies. The first study (a system to alert health service providers) was concerned with the ethical and legal issues of sensitive HIV data and conflicts around data ownership. The second study (a system to improve communication between geographic networks) highlights the importance of understanding the terrain of health databases, establishing stakeholder relationships and collaborations, and resolving issues around funding (whether such innovations should be funded by the HIV programmes or IT budget). The third study (a system to integrate databases different for types of services at six different sites) reported that HIE interventions need

to be flexible and adaptable to different settings because there are usually variations in the quality of health care and databases across settings. And the last study (a system to integrate NCDs databases) pointed out a few contextual issues that impact on DH innovations, specifically in African contexts, such as difficulties in spelling patients' names, duplication of patient entries and migration of patients between health facilities.

In sum, these studies show that DH innovations are useful for improving health service delivery (especially for long-term or chronic illnesses) considering some context-specific challenges. There are a limited number of studies that report on DH innovations in South Africa. There is a need to report on emerging DH innovations in terms of the challenges, opportunities and processes that take place during design and development stages, and when fully implemented, to assess their effectiveness. Below, I review a few studies that report on the key features of DH innovations (or HIE) such as barriers, facilitators and activities.

### **2.3.3. Barriers and facilitators of data harmonisation innovations**

Although there is sparsity of evidence on the effectiveness of HIT innovations (DH or HIE) in LMICs, there are studies that report on the potential barriers and facilitators of DH and HIE [19]. It is useful to review those studies for a list of barriers and facilitators to provide insights into why interventions may or may not work in local settings. **Table 1** presents a list of barriers and facilitators adapted from two systematic reviews [19, 30]; one specifically about primary care practices (first point of entry into the health system) and another looking at LMICs.

**Table 1.** Barriers and facilitators of health information exchange (HIE) [19, 30]

<b>Barriers</b>	<b>Facilitators</b>
<b>Primary care practices</b>	
<ul style="list-style-type: none"> <li>• cost of establishing and maintaining links between EHRs and HIE networks</li> <li>• security and privacy issues given that HIE innovations use patient information</li> <li>• liability for data from outside sources (for example, laboratories)</li> <li>• misaligned incentives (who pays and who benefits)</li> <li>• provider reluctance to relinquish control of patient information to competing systems</li> <li>• lack of interoperability among databases and lack of IT training and support</li> </ul>	<ul style="list-style-type: none"> <li>• more efficient workflows (for example, improved access to clinical data and streamlined referral processes)</li> <li>• improved quality of care (fewer prescribing errors and fewer hospital readmissions)</li> <li>• eliminating costs of storing paper records and downsizing personnel</li> <li>• government incentives for use of HIT.</li> </ul>
<b>Low- and middle-income countries</b>	
<ul style="list-style-type: none"> <li>• lack of leadership and coordination and insecure environments due to corruption</li> <li>• lack of funding for sustainability</li> <li>• lack of equipment and power shortages</li> <li>• lack of training and absence of supervision</li> <li>• faulty, rigid and incomplete system and system design issues</li> <li>• unawareness, privacy concerns and resistance to new processes</li> <li>• lack of timely reporting, feedback and data analysis tools and poor data quality.</li> </ul>	<ul style="list-style-type: none"> <li>• political will and good administration</li> <li>• investments in cost-effective technology</li> <li>• office equipment (for example, phones and radios) and alternative power sources</li> <li>• training and hiring of HIT staff and offer incentives</li> <li>• simple and user-friendly technology and computerised systems</li> <li>• Assess user needs, perceived usefulness and willingness and cooperation</li> <li>• Regular feedback, supportive supervision, standardised data sets and synthesis and validation of data</li> </ul>

Both reviews state that barriers and facilitators are context-specific; therefore HIE innovations should be designed and implemented according to the health priorities of different countries and stakeholders [19, 30]. The barriers and facilitators listed for LMICs can be classified at the following levels: socio-political, financial, infrastructural, organisational, technical, individual and data management. Socio-political factors are related to data culture, organisational environment and politics. There seems to be a low culture of demanding and using health data, health stakeholders may perceive HIT innovations as threatening to their job security, and lack of leadership to coordinate HIE activities between the different levels of the health system [31, 32]. Financial constraints are widely reported as being one of the main barriers for the implementation of HIE innovations [19, 21, 30, 33]. This may be because of variations in the maturity of HISs, difficulties in coordinating funds between multiple stakeholders and the recurring costs associated with maintaining and upgrading innovations as technology evolves. Although, financial barriers of HIE innovations are widely reported in the literature, there is a gap in studies evaluating the cost of innovations versus efficiencies.

There is lack of infrastructure in most LMICs such as inconsistent internet access and shortages of electricity are still problems in some countries [34, 35]. As a result, there may be difficulties in continuously using the innovations and data exchanges between systems may not be up to date. Such issues have the potential to affect the quality of health services and the quality of data, which may lead health staff to abandon the use of HIE innovations [36]. There are studies that report on the infrastructure challenges that are specific to poor settings, but there is a gap in the literature when it comes to studies that report on strategies for dealing with such challenges. It is important to consider the real-world experience of health staff when developing strategies to deal with infrastructure challenges.

Organisational barriers have to do with lack of training and effective communication and management amongst different stakeholders. Health institutions need to make sure that their staff are continuously updating their skills and that clear communication channels are established between different types of services (such as clinical, laboratory and pharmacy) for HIE innovations to be effective. Other barriers to HIE may be related to technical factors. Incomplete, faulty, rigid, fragmented and limited functionality of electronic health systems are all key technical barriers to

HIE [35, 37]. HIE innovations are usually developed and implemented in response to a technical problem such as poor-quality data, lack of interoperability between system and duplication of data capturing.

The next barrier has to do with handling individual-level data and inadequate English language skills amongst health staff. The legal and ethical aspects of implementing HIE innovations are well-reported in the literature [27, 38], these are especially a concern when dealing with sensitive data (such as health data). Language issues can negatively affect the work of health staff in two ways; one, data errors may occur if health staff incorrectly enter patients' names, and two, data items may not be well-defined according to the data needs of health staff. More research is need to assess patient and providers' perceptions around trust and confidentiality [39], as well as strategies for overcoming language issues [29]. The last barrier listed here, data management, is about the usefulness and usability of data. Lack of timely reporting of health data and feedback mechanisms between levels of the health system are key barriers to data utilisation [6, 14, 40]. HIE innovations should be based on research of users' needs for information, in terms of types of data, frequency of use and accessible formats.

Identifying various factors that impact on HIE innovations revealed important gaps in the literature. There is a need for studies: comparing the costs of HIE innovations with their effectiveness in different settings and contexts, reporting on strategies for dealing with infrastructure challenges in the context of technology innovations, outlining flexible and adaptable HIE processes, and assessing patient-provider relationships and data users' needs in the context of HIE innovations. In the next section, I review conceptual and methodological frameworks that are useful for conducting such sociotechnical studies.

### **3. Conceptual frameworks**

Thus far, I have described the relationship between DH, RHISs and health systems strengthening, reviewed definitions of DH and formulated a working definition for the thesis and provided examples of the usefulness of DH innovations for chronic care management. In this section, I explore three conceptual frameworks; PRISM (introduced above), the Complex Adaptive Systems

and the Interactive Sociotechnical Analysis frameworks. I also provide a summary of a methodological framework called the Sociotechnical Evaluation approach. The methodological framework is explained in detail in Chapter 2, however, it is important to briefly describe its relationship to the conceptual frameworks here.

Given the gaps emerging from the literature review, I use the four frameworks (in combination) to conceptualise the PhD project and guide its methods. For example, the frameworks can help me frame the thesis in terms of the broad field of HIT, they can provide the language to talk about HIT innovations; they can guide the interpretation of research findings in relation to other studies and contexts; and they can provide a comprehensive platform for thinking about what enables and hinders HIT innovations in our local setting. Below, I briefly explain how the conceptual frameworks fit together before describing each of them in detail.

As mentioned before, DH is an innovative process of copying existing electronic data captured in various databases (such as RHISs) into a single exchange platform (such as a centralised data repository) Globally, there is political drive to implement DH innovations to strengthen RHISs in hope that RHISs can provide informational support for decision-making to improve quality, safety and efficiency of health services [41]. In parallel, there is interest in the evaluation of such innovations so that their impact, in terms of processes and outcomes, is known especially in relation to unintended consequences of technology and the unpredictable patterns of social interactions [42]. In terms of RHIS strengthening, evaluators are encouraged to specifically focus on how DH can enhance data production, accessibility and utilisation by influencing the determinants of a RHIS (that are technical, organisational and behavioural) [1, 8, 43]. The determinants of a RHIS are well-described in the Performance of Routine Information System Management (PRISM) framework [13]. However, the pathway between data production, accessibility and utilisation is complicated and requires that the determinants of a RHIS work well together.

It is important to review the PRISM framework, which describes the determinants of a RHIS, because DH (which is central to this thesis) is trying to fix RHISs and PRISM may identify social and technical factors, dynamics and contexts that impact on the successful development and

implementation of RHIS interventions. DH usually takes place at several health system levels (such as facility or management levels) and ranges from technical to social levels (such as computer infrastructure to human relationships) where the interplay between these levels can be complex. This complexity is well-described using the Complex Adaptive Systems and Interactive Sociotechnical Systems conceptual frameworks, as well as the Sociotechnical Evaluation approach that takes on the same sociotechnical perspective as the first two but primarily focuses on the methods for studying the complexities of HIT innovations [42-44]. These selected frameworks all provide a sociotechnical perspective, where it is assumed that human factors (such as confidence, motivation and competency) and organisational factors (such as rules, values and practices) and HIT factors (such as electronic databases) are interrelated parts of one system, each shaping the other.

### **3.1. Performance of Routine Information System Management (PRISM) framework**

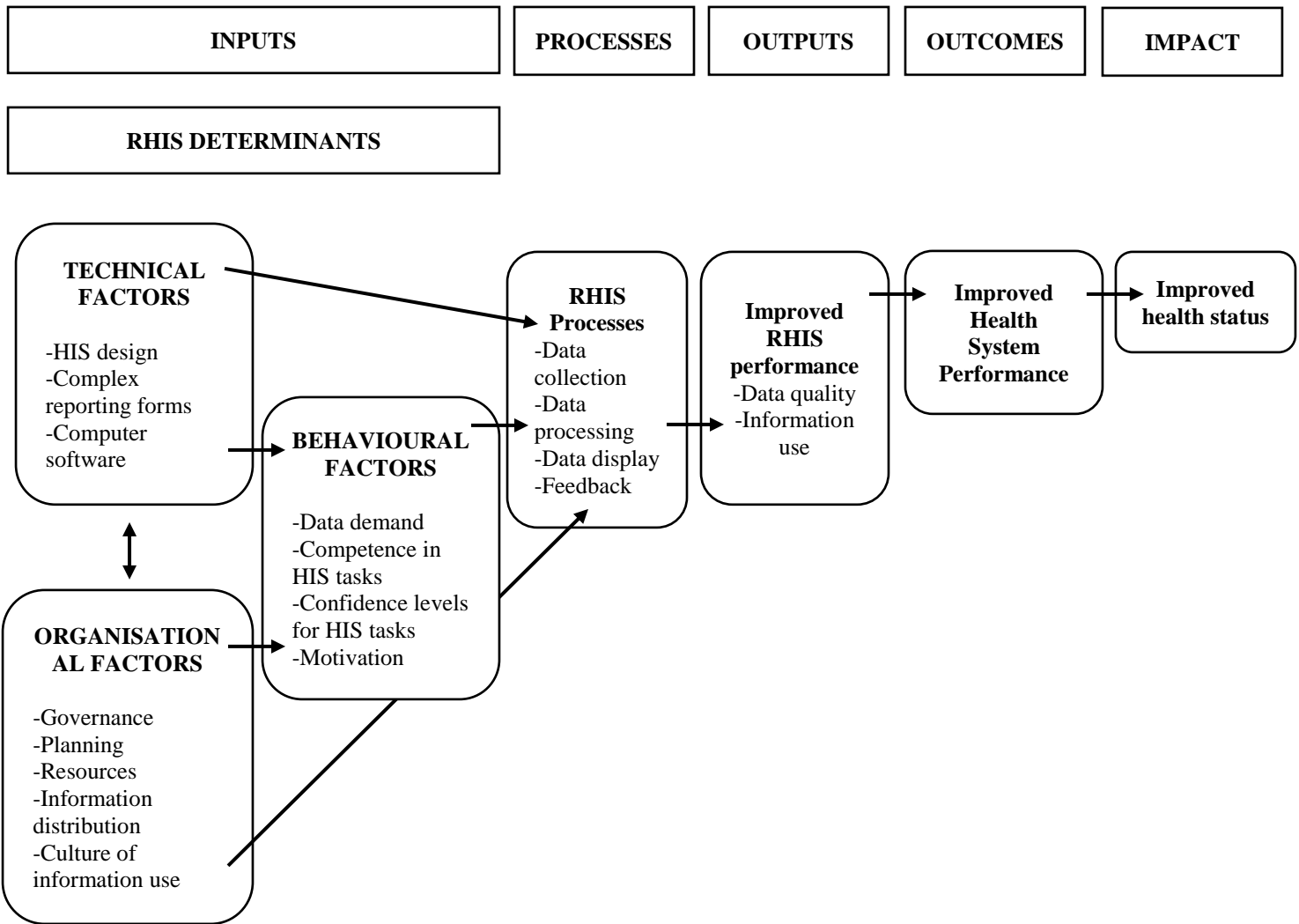
The PRISM framework is a response to known weaknesses of RHIS in developing countries (4). Examples of RHIS weaknesses are: inconsistencies between key variables and indicators for collecting, analysing and reporting health information across programmes [13]; the production of poor quality data that cannot easily be exchanged [15]; and programmatic fragmentation across levels of the health system which can result in the duplication and excessive production of data [12, 45]. PRISM was developed as an innovative approach to appraise RHISs, and to guide the design, implementation and evaluation of interventions that aimed to strengthen RHISs. PRISM is therefore concerned with RHIS performance which relates to how the technical, organisational and behavioural determinants of a RHIS work together [13]. DH is not explicitly mentioned in the PRISM framework, but as described earlier, it has the potential to address the problems of RHISs through coordination, linkage and integration of existing databases.

Electronic RHISs were introduced in developing countries (as early as the 1990s) to support health system decision-making [4, 12]. Earlier studies assumed that if health system managers provided technical resources and introduced organisational rules related to RHISs, then the RHISs would be easily used and sustained. The influence of international donors on RHISs was also assumed to

be a positive contributor towards the development of RHISs; donors helped governments build disease-specific databases and epidemiological surveillance systems. However, it is only by the early 21<sup>st</sup> century that increasing evidence showed that RHISs were not producing the intended results, because of the multiple vertical databases that had been created and the outsourcing of the management of databases [13]. Additionally, technical managers ('data people') and health system managers ('action people') did not understand each other's roles and responsibilities in terms of managing databases [12, 17]. These problems emerged because RHIS development were primarily based on technical approaches for data analysis, reporting and using new information and communication technology (ICT). There was little recognition of the effects of contextual issues, behavioural factors or organisational processes. When attention was eventually given to these social factors there was a need to put them in a coherent framework to understand their effects on RHIS processes and performance [13, 15]. The PRISM framework was therefore developed, which outlines the three determinants (technical, organisational and behavioural) that impact on RHIS performance and their relationship to each other [13].

According to PRISM, a RHIS is composed of inputs, processes and outputs (or performance) which in turn affect health system performance and consequently lead to better health outcomes (see **Figure 5**). RHIS performance is defined as 'improved data quality and continuous use of information', where improved data quality has to do with the relevance, completeness, timeliness and accuracy of data, while continuous use of information depends on the power that an individual has to make decisions and the importance given to other considerations despite availability of information [13]. PRISM promotes the continuous improvement of RHISs, which can be done by analysing the three determinants and identifying appropriate interventions to address determinants that negatively influence RHIS performance. It also promotes a more integrated (or 'holistic') approach to RHIS development; it moves beyond technical RHIS processes and performance outcomes and adds a new layer of individual and contextual determinants.

**Figure 5.** PRISM (Performance of Routine Information System Management) framework [13]



PRISM is useful for identifying the three determinants of RHISs and explaining the relationship between RHIS performance and the health system [13], but it has some limitations. Firstly, PRISM promotes the idea that everyone (within the HIT environment) is responsible for achieving RHIS objectives and performance; that means people are expected to monitor their own performance and design tools for evaluating it. This idea goes against the RHIS status quo where there is a divide between ‘data people’ who assess RHIS performance and ‘action people’ who assess health system performance. Secondly, to use PRISM one requires additional skills to develop performance improvement tools and skills in communication and advocacy in order to promote its usefulness and applicability. Thirdly, PRISM suggests four tools for assessing the overall level of RHIS

performance, interactions and overlap between existing databases, RHIS management practices (governance, planning, training, use of tools and resources) and the role of behavioural factors (motivation, confidences, competence, problem-solving skills and data demand). These tools are very comprehensive and time consuming and there is a misconception that they should be applied at the same time to get a good understanding of a RHIS. However, only those tools that are appropriate for a specific purpose should be used; for example, mapping should only be used when the objective is to study the interaction and overlap of existing databases and to strengthen integration of data across multiple services. Although, PRISM recognises that both social (organisational and behavioural) and technical determinants shape RHIS, another limitation is that it appears to simplify the complexity of the relationship between the RHIS determinants by presenting the relationship as linear [13].

For the thesis, PRISM was useful for identifying and naming the different components of a RHIS (that is inputs, processes, outputs, outcomes and impact) and for describing the three determinants. It also provided a broad frame for explaining DH as a potential intervention for RHISs even though it is not specifically explaining DH processes.

### **3.2. Complex Adaptive Systems approach**

PRISM introduced three determinants that impact on RHIS performance and processes, but it provided no insight into the complex systems that RHIS interventions interact with or are embedded in. In this section, I provide a summary of the Complex Adaptive Systems (CAS) approach which is a useful framework for explaining the complex systems that HISs and HIT innovations are a part of. CAS compares HIT innovations (such as DH) with ‘the occasional pleasant surprise and unpleasant frustration of raising children’, which speaks to the complex health systems that HIT innovations are typically embedded in [43]. In the case of raising children, it is presumed that parents intuitively know what is likely to work and not work, so similarly it is often assumed that enough is known about complex systems. The CAS approach acknowledges that a HIT innovation initiated within a health system is more complex than assumed; where the HIT innovation becomes part of a complex adaptive system. A complex adaptive system is comprised of individual agents (human and non-human) that have the freedom to act in ways that

are not always totally predictable and whose actions are interconnected in such a way that one agent's actions could change the context for others. Research into complex systems has revealed core properties that are described in CAS to advance understandings of HIT innovations.

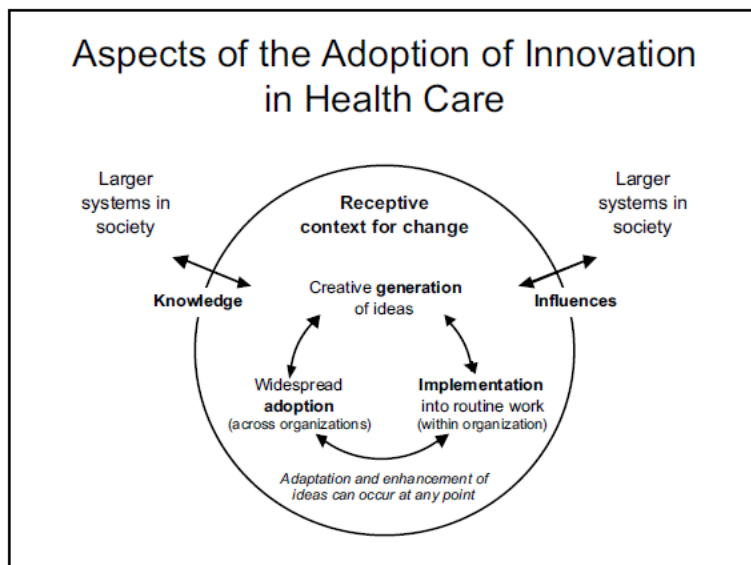
The first property states that relationships between agents are central to understanding the system. This is because interactions amongst parts of a complex system can produce valuable, new, and unpredictable outcomes that would likely not occur if any of the parts were acting alone. The second property has to do with the interconnected nature of structures, processes, and patterns of a system. For example, if a medical administration system was digitised, it may be ignored until real-world changes are made in the process of ordering and dispensing medications. It is also important to remember that individual agents have freedom, patterns of relationships, beliefs, traditions, power and conflict which may influence how they interact with new structures and processes. Thirdly the CAS approach states that actions are based on internalised simple rules and mental models, that is, agents respond to their environment using internalised rules that lead to certain actions. The fourth property is about 'attractor patterns'; underlying attractor patterns in a system may explain complex behaviours (attractor patterns are patterns that can help us understand agents' support or movement towards certain aspects within a system). For example, some innovations may be taken up quite naturally by a group of health professionals because they support professionals' autonomy or enhance their professional image amongst patients.

The fifth property that is outlined in the CAS approach has to do with the idea that the elements of a complex system can change themselves. The ability of a system to adapt is necessary for innovative change in one part of the system and resilience in another part. The sixth property states that an evolving and progressive system needs structures, processes, and patterns that support experimenting with new ways of doing things and providing feedback on the effectiveness of these new ways. The CAS approach states that the seventh property is inherent nonlinearity which means that small changes within a complex adaptive system can have large effects while large changes may only have small effects. And lastly, systems are embedded within other systems and co-evolve. The boundaries of a complex system are somewhat arbitrary; for example, "a medical group (nurses and doctors) is a complex system, which is embedded within a regional health care system,

which is embedded within a national health care system, which is embedded within a political system, and so on” [43].

In addition to this list of core properties for understanding HIT innovations, the CAS approach also describes three interrelated processes of generation (or development), implementation and adoption of innovative ideas (see **Figure 6**) [43]. “The generation process involves creative thinking that leads to the birth and initial pilot testing of an innovative clinical, business, or service delivery process idea” [43]. Implementation refers to the processes and challenges that come with actioning an innovative concept and incorporating it into practice within an existing health care system. The final process involves actions that hinder or enhance the adoption of a new HIT innovation and its practices across a section of the system or organisation, and eventually throughout. Although these processes are thought about from the start, further development occurs while implementation is underway as a response to unintended outcomes that gradually unfold. ‘Receptive context’ describes the extent to which groups or organisations accept change and are open to new ideas, so groups or organisations with a high receptive context are ‘ripe’ for change and they quickly adjust or adopt to new innovations [43].

**Figure 6.** Complexities of HIT innovations in health care systems [43]



For the thesis, the CAS approach supplemented the PRISM framework; it provided more insight into the context in which RHIS interventions (HIT innovations) occur. A strength of the CAS approach is that it provides details about the sociotechnical elements of HIT innovations in complex systems; it is useful for comparing the listed CAS properties with the findings of the PhD project. The three processes of generation, implementation and adoption of HIT innovations were useful to guide my thinking about the stage at which our local DH innovations were. It widened my understanding that the generation, implementation and adoption of HIT innovations take place as non-linear and iterative processes.

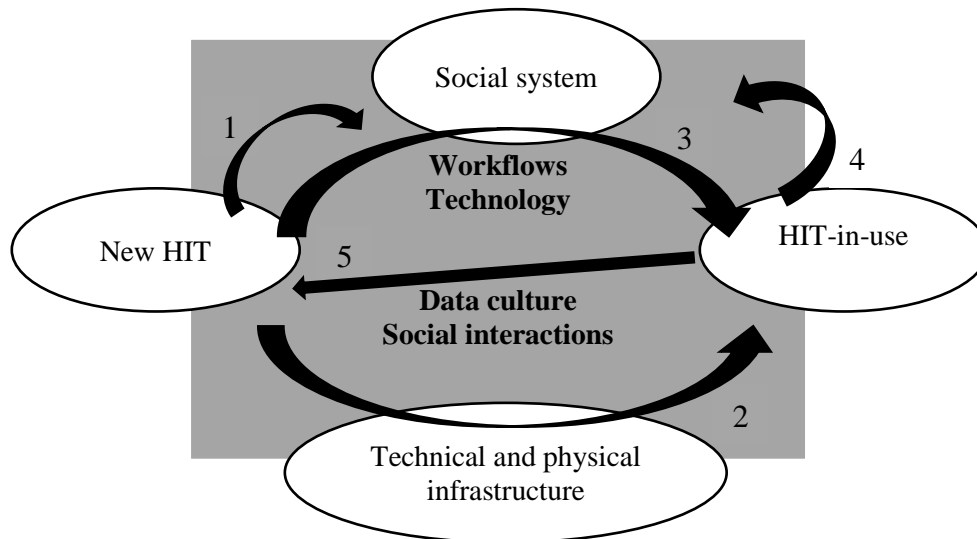
### **3.3. Interactive Sociotechnical Analysis approach**

PRISM and CAS were useful for explaining the determinants of RHISs and identifying the characteristics of HIT innovations in complex systems. In this section, I describe the Interactive Sociotechnical Analysis (ISTA) approach which specifically focuses on the interactions between social processes and technology [42]. The framework is derived from combining different perspectives on social and technical interactions (or sociotechnical interactions). Sociotechnical interactions are types of relationships between social systems, technical and physical infrastructure, existing HIT (which are in use) and new HIT (which are being developed and implemented). These interactions are understood to form recursive processes where feedback loops of newly introduced HIT innovations result in recurring changes to a social system [42].

Firstly, ISTA includes traditional sociotechnical approaches which merely focused on the benefits of appropriate design of technologies. Traditional approaches view HIT innovations as stable with time and consistent across contexts. This included some understanding on how technology affects social systems, but it regarded social systems as uninfluential which means that social systems were viewed to have no impact on technology and its uses. A second perspective is that HIT innovation users (such as clinicians and health service managers) help select, reinterpret, modify and sometimes create technologies. This way people adopt and use HIT innovations as they deem appropriate, by altering them and transforming the relationship between technology and their contexts. Then lastly, there is also the understanding that HIT innovations are embedded within organisations and broader social contexts, so similar technologies can be applied and used in

different contexts, that is between people, between people and equipment, and between sets of equipment. Five types of sociotechnical interactions of HIT innovations are derived from these different perspectives (see **Figure 7**) [42].

**Figure 7.** Interactive Sociotechnical Analysis framework [42]



The five types of sociotechnical interactions are summarised here. Firstly, the ISTA approach recognises that new HIT innovations have the potential to change prior workflows, communication and relationships, because changes emerging from interactions between existing HIT and new HIT innovations may lead to unintended consequences. Secondly, there may be a poor fit between new HIT innovations and existing technical and physical infrastructure, so this lack of interoperability may pose challenges (such as poor decisions, delays, data loss, errors and unnecessary testing) to new HIT innovations. Thirdly, HIT users may reinterpret the purpose of new HIT innovations which may lead to different uses and practices from those intended by HIT designers, in order to deal with ‘unforeseen’ workflows and stakeholder relationships. Fourthly, the development and implementation of new HIT innovations requires recursive processes that determine their usability. Lastly, social systems (such as people and organisational structures) may drive the process of recursive change in that HIT users’ reactions and their local adaptations of new innovations may move away from original conceptualisations, so HIT designers may be forced to reconfigure some HIT features.

Like the CAS approach, the ISTA approach recognises that there are different types of complex interactions between social and technical processes; this makes the framework a useful tool for making meaning of the thesis findings. ISTA focuses on a limited number of sociotechnical elements and types which makes it easy to compare the sociotechnical interactions described in the ISTA approach with those that I observed while DH innovations were being developed and implemented in our local setting.

### **3.4. Summary of the methodological frameworks**

The PRISM, CAS and ISTA provided insights into the elements of RHISs and how HIT innovations need to operate in the complex settings that the RHISs they are trying to fix are embedded in. The Sociotechnical Evaluation (STE) approach is also based on the premise that HIT innovations are developed and implemented in complex systems; but what makes it different is that it provides very clear methodological steps for evaluating HIT innovations. The sociotechnical perspective (which is described in ISTA approach above) is about investigating how HIT innovations change social processes (for example, the way health care is delivered by integrating clinical, laboratory and pharmacy databases) and how HIT innovations change over time because of new technologies and user demands (for example, ongoing improvements to usability) [8, 42]. The STE approach takes it a step further by guiding researchers to identify potentially transferable lessons emerging from development and implementation processes of new HIT innovations while considering the increasing availability of technical capacity and diverse settings and contexts. In Chapter 2, I detail the methods outlined in the STE framework. These methods were used in the conceptualisation of this PhD project, the data collection and analysis, and reporting of the findings.

## **4. Problem statement**

Lack of coordination and integration between routine electronic databases can limit effective data production and utilisation to support health management decision-making. There is currently a need to strengthen data support structures through the harmonisation of multiple databases across different types of health services and organisations. Health managers and policy-makers need

harmonised data for disease surveillance and chronic care management, and resource allocation especially in LMICs. DH entails identifying, reviewing, matching, redefining and standardising data captured in various electronic health records (or databases) [20]. The harmonised data is ideally kept in a single HIE platform where authorised data users can easily access it and use it to make timely clinical and operational decisions.

DH is often defined as a technical intervention to a HIS (a HIS aims to effectively produce health data and promote its use) in the literature. However, in practice, DH is usually developed and implemented in complex ‘messy’ settings where social systems exist alongside technology and sometimes influence DH processes and outcomes in unintended and unexpected ways. The neglect of social factors (including elements, activities, processes and outcomes) in studies and designs of HIT innovations can lead to poor results of DH and result in wasted resources used during development and implementation stages. There is limited evidence on the social factors of DH; for example, the challenges and opportunities related to relationships between stakeholders, institutional structures, the design and conceptualisation of DH innovations, and buy-in and motivation to use DH products. This study therefore aims to examine the social factors of DH innovations in the Western Cape Province, specifically in relation to a DH initiative that is currently underway. The new DH initiative was started by the Western Cape Provincial Department of Health (officially in 2015) to integrate patient-level data captured in multiple, unconnected electronic databases across different types of health services (such as clinics and hospitals) and government levels. The study objectives are listed below.

## **5. Study objectives**

The specific objectives of the PhD project are as follows:

- to understand how historical factors shaped the development and implementation of HIS interventions in South Africa (chapter 3);
- to conduct a scoping review on the definitions and conceptualisations of data harmonisation and to explore the relationship between data harmonisation and health management decision-making (chapter 4);

- to explore the challenges, opportunities, and factors) affecting the development and implementation of data harmonisation innovations in South Africa (chapters 5 and 6).

## **6. Thesis structure**

This PhD project takes four approaches to the above-mentioned objectives. Firstly, a historical analysis and synthesis of the development and implementation processes of HIS interventions in South Africa. Secondly, a scoping review of the literature to define and conceptualise DH and explain the relationship between DH and health management decision-making. Thirdly, a qualitative assessment of using a new DH initiative in our local setting to access province-wide harmonised data. Lastly, a case study of institutionalising and operationalising a new DH initiative in a province of South Africa. In terms of the write-up of this PhD project, the thesis combines both monograph and publication formats; the findings chapters (3, 4, 5 and 6) are written as papers for publications.

After this introductory chapter is the methods chapter in which I identify the setting and context in which this research took place. I then provide a description of a methodological framework that guided the scope, data collection and data analysis approaches of this PhD project. At the end of Chapter 2, I reflect on some of the opportunities and challenges that I experienced in conducting this research.

Chapter 3 focuses on the key events and barriers and facilitators of the development and implementation of HIS interventions in South Africa. This chapter provides an understanding of how historical factors can impact on new interventions and I draw out important lessons for the planning and monitoring of future interventions. This chapter is different to the others in that I take a retrospective approach in understanding HIT innovations.

In Chapter 4, I outline the methods for conducting a scoping review to define and conceptualise DH in a published protocol. I also present the review in the same chapter, in the review where I identify alternative terms for DH, I provide a synthesis of definitions, components and processes of DH, and I identify three levels of health management decision-making that are supported by

DH efforts. This chapter is important for identifying some of the characteristics of DH and showing that DH can contribute to different levels of health management decision-making. It is also important for illustrating that there is a lack of evidence on what DH interventions and activities entail in diverse settings and contexts, especially in LMICs.

Chapter 5 takes a unique methodological approach in that I used the new DH initiative to produce data for another study that I was involved in and then reflect on the process and experience of engaging with the new DH initiative. This chapter (using an ethnographic approach to study DH innovations) provides an understanding of the motivations and opportunities, design process and operationalisation of DH innovations in South Africa.

The last findings chapter, Chapter 6, presents an ethnographic account of the institutional and conceptual dilemmas that affected the implementation of the new DH initiative as well as the key success factors in navigating these dilemmas. There are cross-cutting themes between this chapter and Chapter 3 and 5, for example the role of context, design processes, institutions, leadership and stakeholder resistance or acceptance on HIS or DH innovations.

In the discussion chapter, Chapter 7, I provide a summary of the key findings and discuss the findings in relation to existing literature. I then identify important lessons for practice and research and point out the main strengths and limitations of the thesis.

## CHAPTER 2: METHODOLOGY

---

This second chapter provides a description of the setting, a summary of the methods used in each of the four chapters, and further details about the methodological framework used to guide the thesis. Given that the findings chapters are written as standalone papers, more details for their methods are provided in the individual chapters.

### 1. Country setting

South Africa has an estimated population of 55 million people [46]. Health services are delivered by public and private organisations. The public health services are divided between national, provincial and local levels of authority. The majority of the population relies on public health services, usually government-run clinics and hospitals. The public health services are delivered at the primary (clinics and community-based care), secondary and tertiary levels (hospitals). Primary, secondary and tertiary health services are managed by local and provincial governments, and each of the nine provinces in South Africa are responsible for their own health workforce and delivery of health services. The National Department of Health is mainly responsible for policy development and distributing finances among the provinces from the national budget [46].

The Western Cape Province has a well-resourced health system compared to other provinces in the country [47]. At the time of independence (1994), the Western Cape had the second-highest income per capita and the highest human development index which translated into better health indicators. This economic advantage as well as academic and research opportunities for health professionals (there are three universities in the province) spurred innovative projects and provided a fertile environment for testing new innovations. Additionally, the Western Cape Province had the highest health expenditure per capita across provinces as well as the highest availability relative to population of clinics and public doctors and nurses, and the second-best availability in terms of acute public hospital beds [47].

The Western Cape is comprised of six districts: the Cape Winelands, Central Karoo, Cape Town Metropole, Eden, Overberg and West Coast districts. In the Cape Metropolitan district (which is the largest urban district), primary health care is delivered by both the local municipal health authority, known as the City of Cape Town (referred to as City health authority) and the Western Cape Provincial Department of Health (referred to as the Province health authority). The City health authority provides mostly preventive mother, child and reproductive health services (such as immunisations, family planning, HIV testing and STI treatment) and some curative care, while the Province health authority provides a full PHC package of care, including curative care services. There is overlap between City and Province PHC services, for example, with both providing STI, HIV and TB care. The Province health authority is responsible for managing hospital care services and the National Department of Health (NDoH) co-ordinates country-wide laboratory services, via the National Health Laboratory Services (NHLS).

Some of the health programmes and services, especially for epidemics such as HIV and TB, were established as collaborations between government and non-governmental organisations (NGOs) with the help of international donor funders. Separate databases were developed for the disease- or programme-specific services. This resulted in new databases for specialised services (for treatment initiation) and existing routine (or administrative) databases (for headcount data) operating in parallel to each other. The programme-specific databases existed side by side with the routine databases because the routine databases were considered incapable of and insufficient for monitoring and evaluating disease-specific programmes. Even though the disease-specific databases (often set up by donors) had the potential to provide informational support for decision-making, there was fragmentation and redundancy in data collection, with multiple data flow processes and reports.

Currently, various demographic, clinical, laboratory, pharmacy and mortality data is dispersed across multiple databases. For example, demographic, clinical and pharmacy data is collected within provinces, while mortality data (from the national death register) and laboratory data (National Health Laboratory Services) are collected centrally at the national level. Fragmentation between these databases is a problem for disease surveillance, tracking individual patient outcomes

and performance in health services, and broader health system planning and monitoring and evaluation [6].

## **2. Study setting**

This PhD project is embedded in a study titled ‘Using Information to Align Services and Link and Retain Men in the HIV Cascade’ (or iALARM). The iALARM study is focused on helping health system and community service providers use routinely collected information more efficiently to improve local HIV services. One of the ways in which the iALARM study is doing this is through a collaboration with a new data harmonisation (DH) initiative called the Provincial Health Data Center (PHDC). PHDC is able to help the iALARM study team access newly harmonised data which iALARM can use to facilitate communication (in the form of a task team) between health system and community service providers. This PhD project was conceptualised at a time when iALARM was trying to build a single dataset that contains longitudinal HIV-related data using the new innovative harmonisation approaches at PHDC. The iALARM researchers were trying to assess HIV-positive individuals’ outcomes and performance in treatment and care services. This iALARM-PHDC project provided me with an opportunity to learn more about the challenges and opportunities for conducting DH, especially in terms of the social aspects that may impact on them (such as people, relationship dynamics and existing data practices).

PHDC recently emerged in the Western Cape Province to optimise data production and data utilisation of demographic, clinical (including laboratory and pharmacy data), mortality and administrative data (including financial and human resources) captured in disease- or programme-specific and routine administrative databases. The DH innovations being introduced at PHDC are novel in their approach because they make use of individual patient public health services data captured electronically within City, Province and National databases; as opposed to previous HIS interventions that primarily focused on aggregate data. Individual-level data refers to an individual patient’s demographic, clinical, laboratory and pharmacy data which can be linked using a unique identifier (such as a shared patient record or patient folder number). Aggregate-level data is high-level data that are combined into summary reports for assessing overall service performance and cost efficiencies.

PHDC aimed to address the problem of routine electronic data relevant for individual patients' clinical management and overall service improvements being captured in multiple, disparate, large-scale databases across various types of health services, levels and organisations. One of the DH projects that PHDC worked on was a HIE platform in the form of a data repository where all health-related individual-level data could be stored centrally. PHDC was housed at the Western Cape provincial department of health and collaborated with a research unit at a local university. This led to PHDC working with various research and government (health services) stakeholders who produce data (for example, they own, manage, govern or pay for it) and/or use data (for example, they analyse, interpret and make decisions based on it).

Another DH project that PHDC worked on was an electronic single-viewer patient record useful for individual patient and chronic care management, as well as for supporting epidemiological analyses and routine operational reporting. The electronic single-viewer patient record was a computer application, in the form of an information dashboard, that frontline health staff (such as nurses and doctors) could use to access clinical data linked using a unique identifier, called a Clinicom number. Clinicom is a relatively new electronic system (about a decade old) that is used to allocate an individual patient with a unique identifier that can be used universally at health facilities within the province to admit, discharge and transfer, as well as to schedule appointments and maintain records. The ability to harmonise individual-level data through, for example Clinicom, could help clinicians easily and timeously track patients as they move through different types of health services and geographic areas.

One way in which the data repository could enhance epidemiological analysis is through the development of chronic care cohorts or cascades, for example, for TB, HIV/AIDS and diabetes. Patients in chronic care can be defined as part of a specific cohort or cascade on the basis of treatment initiation, laboratory test results and subsequent clinical visits; and linkages between databases makes it possible to determine the impact of co-morbidities on clinical and service management. In terms of routine operational reporting, there are already clinical and administrative (for costs or human resources) systems that generate summary reports from aggregate-level data, but the data repository could eliminate the problem of data being duplicated in various databases across different units in the same hospital. Each unit could only capture data

in one appropriate system, linked to or copied into the data repository, without having to also duplicate the data in the Clinicom register for universal access.

These PHDC's initiative were spurred by researchers and health services practitioners needing linked and merged individual-level data. The founder of PHDC was the principal investigator on various research projects assessing patient outcomes and long-term effectiveness of disease or health programmes (such as HIV and antenatal services) as well as being a health services manager within the Province health authority. Some PHDC staff were also involved in research activities at a local university, and also participated in the development and implementation of primary-level HIV and TB databases in the Western Cape Province, which made them knowledgeable about research, the health services and HISs. Although proof of concepts for conducting DH innovations were underway a few years earlier, PHDC was officially started in 2015 and the team was given a few months to set up within the Province health authority, engage with the research directorate about data access applications, identify key stakeholders, and maintain collaborations within the health services and the local university.

The aspirations of PHDC are to strengthen the routine HIS, to develop capacity and pilot technical procedures for harmonising public health services data and evaluate outcomes of DH innovations. The province-wide data repository which centrally collects individual-level data from the different health-related databases is one innovation for strengthening the technical infrastructure of the province. PHDC hopes that by starting DH innovations, it will motivate the health services to consider upgrading computer programmes, so that data requests can be processed easily and timeously as well as improve the overall quality of data practices. Additionally, taking a comprehensive approach is not only about DH innovations involving various databases and stakeholders, but it is about PHDC continuously learning about the HIS context, uncovering new challenges as practical work is underway and being flexible in making changes as new opportunities emerge.

### 3. Methods

As detailed in the previous chapter, the aim of this PhD project was to explore the social factors shaping DH innovations in South Africa, in a context where a new DH initiative is being implemented. The research was approved by the University of Cape Town Human Research Ethics Committee (HREC ref: 320/2015 and 738/2018); the approval letters are **Appendix 2** and **Appendix 3**. The overall study design is a multi-method qualitative study; including three methodological approaches, namely a historical analysis and synthesis (Chapter 3), a scoping review (Chapter 4) and an ethnographic approach (Chapters 5 and 6). The methods for each of the four findings chapters (Chapters 3-6) are outlined in detail in each of the chapters; however, below I provide below a summary of these methods, and specifically a summary of the primary research methods (for Chapters 3, 5, and 6) in Table 1.

Objective 1 of the PhD project is to examine the historical events that shaped the development and implementation processes of HIS interventions in post-apartheid South Africa. I wanted to know how HISs were reformed after apartheid, what challenges and opportunities HIS developers faced, and what the important lessons were for emergent DH innovations. In simple terms, I wanted to understand where HIS interventions in South Africa originate from, how they are similar or different to new or emerging HIS interventions, and how HIS intervention development and implementation processes should look like the future. Chapter 3 thus provides a chronological synthesis of key historical events and identifies key barriers, facilitators and outcomes of HIS interventions in South Africa. Data related to the history of HIS interventions were collected from different sources, such as journal articles, institutional reports or documents and by interviewing two key informants who worked for Health Information System Program (HISP) when HIS interventions were being developed and implemented for the new government. I analysed the data chronologically and according to key events that took place, which led me to an understanding of the key barriers, facilitators and outcomes of HIS interventions.

Objective 2 is to define and conceptualise DH innovations and to try and provide an explanation of the relationship between DH and health management decision-making. While conceptualising this PhD project, I realised that there was no uniform term, definition or description for the process

of integrating fragmented data. I therefore wanted to search the literature to identify and synthesise the different terms and make meaning of DH. Studies eligible into the scoping review were identified through systematic literature searches. A colleague and I screened titles, abstracts and full-texts and then sampled studies based on the range, variation and similarities or differences in definitions and concepts and intervention descriptions. I used manual coding and the filter option in Excel to provide (a) a numerical analysis of the characteristics of included studies; (b) a narrative synthesis of the different DH definitions, components and processes, as well as intentions, suggestions and/or explanations of how DH may lead to improved health management decision-making.

Chapter 4 consists of two parts. The first part is the published protocol of the scoping review which outlines the methods for conducting the review which can be briefly summarised as: identifying the research question, identifying relevant studies, selecting studies for inclusion, and data extraction and data synthesis. The second part is a report of what I found in conducting the review. To arrive at the findings, data were collected from studies that are included in the scoping review through a systematic literature search and eligibility assessment at the title and abstract stage and at the full-text stage. Data from the included studies are analysed both numerically and qualitatively to arrive at three groups of findings: a summary of the characteristics of included studies; alternative terms, definitions and concepts of DH interventions; and an exploration of the relationship between DH and health management decision-making.

The last objective, Objective 3, involves an investigation of the challenges, opportunities and factors affecting the development and implementation of DH innovations in South Africa. Through iALARM stakeholder mapping and networking activities, I was able to identify health system and DH stakeholders who are key informants for my research. My key informants are DH innovators at PHDC (founder and staff), City and Province data stakeholders (data clerks, health information officers and database managers), City and Province health services staff and managers (nurses, facility and programme managers, public health specialists, sub-district managers) and researchers (iALARM researchers and other researchers at a local university). I sought verbal or written consent from all key informants after talking them through the consent form (**Appendix 4**). I collected data from February 2016 to September 2017 through iALARM-PHDC meetings, PHDC-

stakeholder meetings, iALARM-stakeholder meetings, participant observation and interviews. Interviews were guided by the questions outlined in **Appendix 5**. Raw data from meetings, interviews and field notes are coded and organised using thematic data analysis.

There are no immediate or direct benefits to the informants who were enrolled in this research. However, the research findings can contribute to the ongoing development and implementation processes of DH innovations in South Africa. A potential risk to DH innovators was them unintentionally sharing confidential information during the process of trying to explain something else related to PHDC. I reassured DH innovators that they could withdraw any information during the course of data collection, and that specific details of sensitive social or political issues and confidential information about technical processes (such as codes and algorithms) were not going to be included in the write-up.

**Table 1.** Methods for primary research conducted in Chapters 3, 5 and 6

Summary of primary research methods (chapters 3, 5 and 6):		Additional details
Participants and recruitment	<ul style="list-style-type: none"> <li>• <b>HISP managers</b> were purposefully selected because of their long-term (since 1994) involvement in the development and implementation of HIS interventions</li> </ul>	n=2
	<ul style="list-style-type: none"> <li>• <b>PHDC founder and staff</b> (also referred to as DH innovators) were purposefully selected because of their efforts to harmonise individual-level data across demographic, clinical, laboratory, pharmacy and administrative databases</li> </ul>	n=6
	<ul style="list-style-type: none"> <li>• <b>Health services stakeholders</b> (such as nurses and doctors; facility and programme managers, public health specialists, sub-district managers) working closely with PHDC or iALARM were selected because they are potential implementers and users of DH innovations</li> </ul>	n~20 <i>Inconsistent numbers at meetings</i>
	<ul style="list-style-type: none"> <li>• <b>Health information stakeholders</b> (such as data clerks, health information officers and database managers) were selected because they collected data and compiled reports, or maintained database features and controlled access to databases</li> </ul>	n=5
		n=3

	<ul style="list-style-type: none"> <li>• <b>Researchers</b> working on the iALARM study (specifically those dealing with PHDC) were selected because they are potential users of public health services data and DH products.</li> </ul>	
Informed consent	<ul style="list-style-type: none"> <li>• <b>Written consent</b> was obtained for all informants who were interviewed individually (e.g. HISP and PHDC staff)</li> <li>• <b>Verbal consent</b> was obtained for all other participants (e.g. group observations and meetings)</li> </ul>	Copies of written consent forms were stored with the iALARM coordinator and verbal consent was noted in the field notes
Data collection methods	<ul style="list-style-type: none"> <li>• <b>Document reviews</b> of journal articles, reports, standard operating procedures and information on institutional websites</li> <li>• <b>In-depth interviews</b> with key participants, such as innovation designers, implementers and users, to better understand context, challenges, opportunities, motivations and processes of HIS or DH innovations</li> <li>• <b>Participant observation</b> of key participants (specifically DH innovators) included conversations, meeting attendance (PHDC, iALARM and health services), and telephone and email communication <ul style="list-style-type: none"> <li>○ Meeting attendance: iALARM-PHDC meetings (n~5), PHDC-stakeholder meetings (n~10), iALARM-stakeholder meetings (n~15)</li> <li>○ Telephone and email communication: updates and resolving challenges related to accessing data for the iALARM study</li> </ul> </li> </ul>	<p>Topics extracted: context, key events, key actors, challenges, opportunities, motivations and processes of HIS interventions or DH innovations</p> <p>Interview guide: Appendix 5 and emerging questions Interviews lasted approximately two hours each No recordings Detailed interview notes</p> <p>Detailed field notes Meeting minutes: taken by an allocated minute-taker; minutes captured discussion and action points</p> <p>Data collection for Chapters 3, 5 and 6 was conducted at the workplace of all participants, except interviews with the two HISP managers took place at coffee shops</p>
Data collection periods	<ul style="list-style-type: none"> <li>• Chapter 3 <ul style="list-style-type: none"> <li>○ Literature searches: <b>mid-2015 to mid-2017</b></li> <li>○ Interviews: <b>between October 2016 and February 2017</b></li> </ul> </li> <li>• Chapter 5: <b>from February 2016 to December 2016</b></li> </ul>	<p>Interviews lasted approximately two hours each</p> <p>Meetings (one-two hours each) n=24, observed</p>

	<ul style="list-style-type: none"> <li>Chapter 6: <b>from August 2016 to September 2017</b></li> </ul>	<p>participants before and after meetings</p> <p>Meetings (two-three hours each) n=22, observed participants before and after meetings</p> <p>Data collection periods for Chapters 3, 5 and 6 overlapped</p>
Data analysis	<ul style="list-style-type: none"> <li>Chapter 3: <b>inductive analysis</b> (coding and synthesising according to key themes)</li> <li>Chapter 5 and 6: <b>iterative thematic analysis</b> (themes were identified intermittently while data collection was still ongoing)</li> </ul>	<p>Triangulation: compare and contrast findings emerging between documents and between documents and interviews</p> <p>Triangulation was used to test the validity of emerging themes by looking for consistencies or inconsistencies between field notes, interview notes and meeting minutes</p>

#### 4. Study limitations

The study limitations applicable to the individual findings chapters are described in detail in their Discussion sections. The overall limitations of this PhD project can be summarised as five points.

Firstly, there is scarcity of literature that defines DH innovations. The lack of clarity of what DH means can result in me making judgements about what I think DH means, especially during the early stages of the research.

Secondly, the literature on HIS and DH innovations is limited to a specific time and place. This can result in me making assumptions about whether certain aspects of DH described in the literature are applicable to the DH innovations in our local setting or not.

Thirdly, the thesis findings are emerging in a context where DH innovations are still being operationalised and are therefore continuously changing to adapt to emerging technical and social aspects. This may impact the timeliness and relevance of the findings, especially in relation to different phases of DH innovations.

Fourthly, it is difficult to include multiple informants in the research, as DH innovations at PHDC are not yet well-known or understood within the health services and HIT environment.

Lastly, there is the dilemma of me being both an iALARM researcher (and therefore a PHDC stakeholder) and a doctoral student. iALARM researchers and PHDC staff are key informants of my PhD project. Researching emerging DH innovations means being able to keep track of key databases and stakeholders. On the one hand, I am in the unique position where I can enter specific spaces (such as PHDC) because of my affiliation to iALARM and I can identify potential informants for the PhD project while coordinating research activities and networking for iALARM. On the other hand, I have to however be aware of what role I take on in specific interactions with people. I need to strike a balance between being a representative of the iALARM project and being a doctoral student. This distinction is important for ethical reasons; for example, iALARM stakeholders may interact with me because of their own interests or responsibilities on the iALARM project, but they may choose to interact differently when I am wearing my doctoral student 'hat'. Additionally, having knowledge about certain aspects of PHDC and establishing relationships with PHDC staff can potentially influence the way I interpret research findings and lead to a bias towards PHDC.

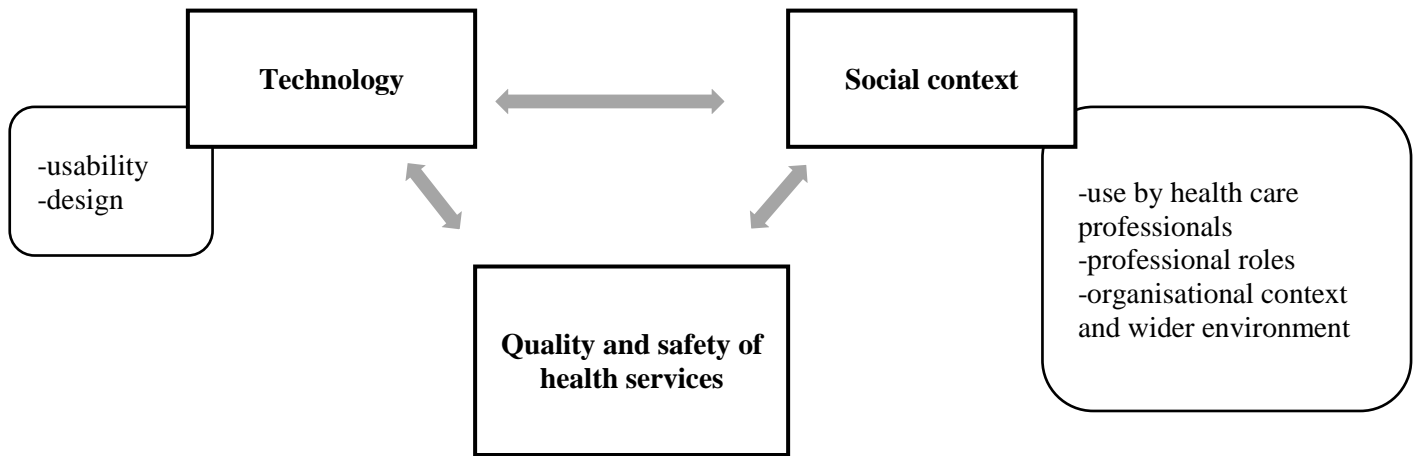
## **5. Methodological frameworks**

### **5.1. Sociotechnical Evaluation**

The Sociotechnical Evaluation (STE) methodological framework is based on the sociotechnical perspective described by the Interactive Sociotechnical Analysis (ISTA) and Complex Adaptive Systems (CAS) frameworks (see Chapter 1). The sociotechnical perspective recognises the complex interactions between health information technology and social processes. Hence, the STE

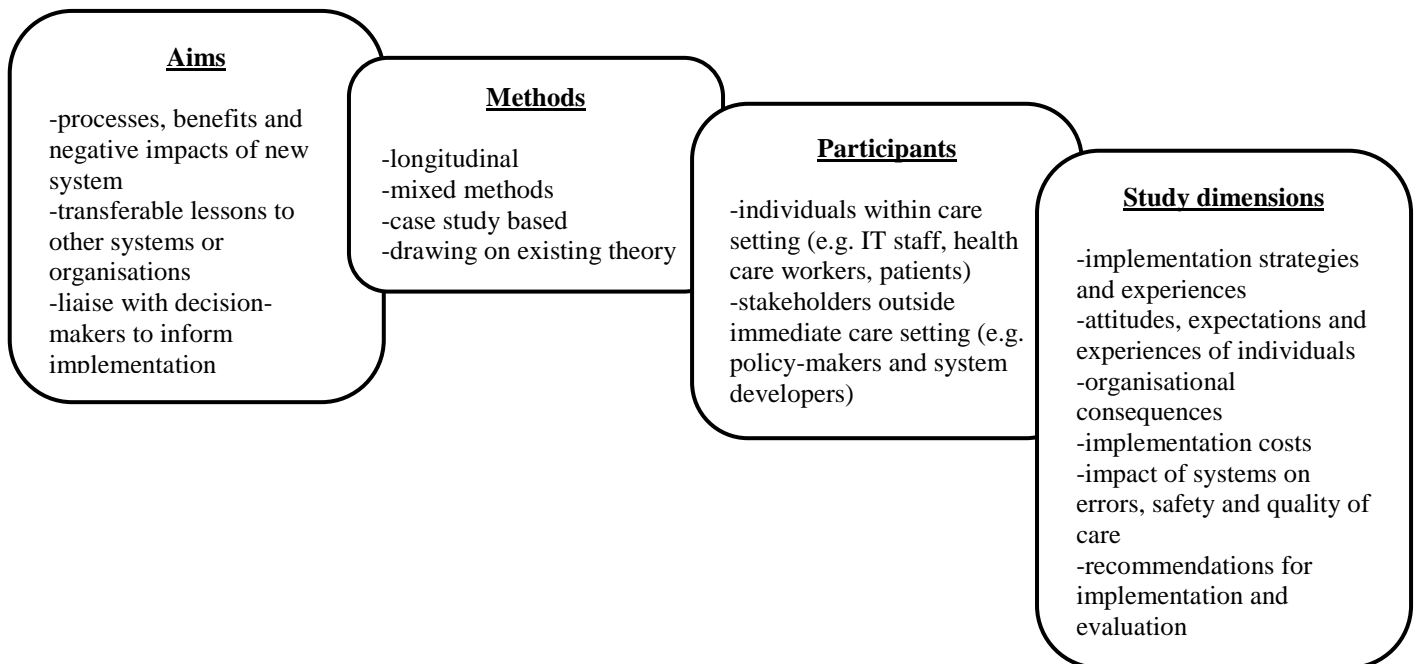
framework provides guidance for identifying and exploring the changes and outcomes of both social and technical aspects of HIT innovations. A limitation of other methodological or evaluation frameworks is that they solely focus on the impact of technical outcomes, so they end up having limited recommendations and/or generalisability beyond a specific clinical setting. The STE framework is useful for the PhD project for two reasons. Firstly, it provides an explanation of the interactions between technology and social processes (which is illustrated in **Figure 1**) as a prelude to the methodological aspects to considering in doing research in this area [44]. And secondly, it provides a summary of the key components to consider in examining HIT innovations which are presented in **Figure 2** [44].

**Figure 1.** Interactions between technology and social processes in HIT innovations [44]



**Figure 1** explains the relationship between three dimensions: technology, social context and quality and safety of health care services, while **Figure 2** presents guidelines on aims, methods, participants and specific dimensions that should be considered in sociotechnical research related to HIT innovations.

**Figure 2.** Key components of the Sociotechnical Evaluation approach [44]



These three dimensions have the potential to impact on the key elements, activities and outcomes of DH innovations, given that they are interrelated and can be influenced by or have influence over other factors. For example, the state of technology impacts on the design of HIT innovations which in turn impacts of the level of accessibility and usability of HIT innovations, while being implemented in a specific social context that is influenced by organisational factors, the roles of health care professionals (HCPs), the demand of HIT innovations by HCPs, and the wider environment [44]. Optimally, these interactions should have positive outcomes on the quality and safety of health care services but given the diverse and non-linear way in which interactions occur, this is not always the case. What makes most HIT innovations complex is that technology and social processes are continuously evolving over time, for example, as a response to newly arising organisational needs or shifting boundaries of existing HIT systems.

The next chapters (3 to 6) are findings chapters. Theses chapters are written as standalone papers, each with their own introduction, methods, findings and discussion sections.

# **CHAPTER 3: THE HISTORY OF HEALTH INFORMATION SYSTEM STRENGTHENING IN THE WESTERN CAPE PROVINCE, SOUTH AFRICA**

---

## **1. Introduction**

The history of a health information system (HIS) is important for explaining the contextual challenges and opportunities that have emerged in relation to data processes and products over time. A HIS refers to “any system that captures, stores, manages or transmits information related to the health of individuals or the activities of organisations that work within the health sector” [10]. There have been many efforts over the past decades, especially in low- and middle-income countries (LMICs), to develop and implement HIS interventions to strengthen the health system [4, 48]. A HIS is one of the six essential and interrelated building blocks of the health system and has the potential to affect change in the other building blocks [1]. However, in developing and implementing HIS interventions much of the focus has been on the technical components of HISs, such as available technologies and computer infrastructure. This narrow definition of HIS interventions as merely addressing technical problems neglects the influence of social factors (such as organisational, behavioural, political and historical aspects) on the development and implementation processes [43, 49].

Examples of technical problems are poor designs of existing HISs [15], lack of technical interoperability among existing HISs [9, 45], and absence of common metadata (such as key indicators and data definitions) [15]. At the surface level these problems appear to be technical, but they may be caused by social factors, such as poor governance over financial resources, hierarchical organisational structures, and low staff motivations and competencies to perform HIS tasks [42, 50]. Applying a narrow technical approach in the development and implementation of HIS interventions may lead to poor outcomes and wasted resources, because of the unintended and sometimes unexpected consequences that may emerge from sociotechnical interactions [42]. Sociotechnical interactions can be defined as the elements, activities and outcomes that come into

existence because of complex, iterative relationships between social processes (such as people and institutions) and technology (such as HIS interventions) [42, 44, 50].

A HIS is therefore best optimised when various social and technical factors work jointly to advance its development and implementation processes. The benefits of a well-functioning HIS have to do with it being able to produce reliable and timely information on individual- and population-level health outcomes and overall health system performance [134, 135]. A HIS is embedded in a health system which means it is influenced by some of the technical and physical infrastructure and social systems of the health system. For example, the leadership of the health system may demand health information for guiding the activities of the other health system building blocks. This means that a HIS provides informational support to decision-makers at all levels of the health system for planning, monitoring and evaluation. It can further enable evidence-based decision-making for health policies and programmes, especially for optimal allocation of resources [10, 51]. Thus, a few approaches (like the Sociotechnical Evaluation and the Interactive Sociotechnical Analysis) have outlined the role of sociotechnical interactions in HIS intervention processes [42, 44]. These approaches should encourage HIS designers, implementers and users to consider the multiple and complex social and technical factors that shape the development and implementation processes of HIS interventions.

Sociotechnical evaluations (STE) can provide insights into potentially transferrable lessons, especially in contexts where complex iterative relationships between technology and social processes are present [44]. But, individual studies of HIS interventions are usually unable to capture all complexities and relationships of sociotechnical factors (such as technical, historical, political, organisational and behavioural factors) at once [44]. Most STEs are undertaken prospectively to map out emerging or ongoing interactions between technology and social processes, as well as to determine how newly implemented HIS interventions were received and used [44, 51, 52]. There is lack of historical analysis and synthesis of sociotechnical factors of HIS interventions. Retrospective evaluations are beneficial in that they allow HIS stakeholders, health system managers and social scientists to reflect on and draw lessons from HIS interventions that have already been developed and implemented. The rationale for this paper is based on two limitations of once-off formative evaluations; that is, research that was done over a short period of

time at the time when the HIS intervention was being developed or implemented. One, they generally cannot capture the multiple and complex sociotechnical factors of HIS interventions, especially for those factors that only become evident once interventions are complete. And two, there is a need for retrospective evaluations of HIS development and implementation processes that can inform future planning and monitoring of HIS interventions [53].

Various HIS interventions were introduced as part of the Reconstruction and Development Programme (RDP) in South Africa since 1994 [54]. One of the goals of RDP was to restructure the health sector by establishing several decentralised health districts within each of the nine provinces of South Africa [11, 54]. Decentralising the health system involved the highest level of government (that is the national government) transferring authority and responsibility of some health system functions to the lower levels of government (that is provincial and local governments). Strategic management teams specialising in different aspects of the health system were established to operationalise this goal across all provinces. The strategic management team leading the development and implementation of HIS interventions in the Western Cape Province was determined to restructure the health system by creating district and provincial management information systems. The strategic management team, in preparation for these HIS interventions, consulted IT experts in two universities in the Western Cape Province in 1995. IT experts were asked to provide guidance in the development and piloting of the HIS interventions, which were seen as the main vehicle through which health districts in the Western Cape could become decentralised. The pilots were rolled out in different health facilities across the six districts of the province (namely the Cape Winelands, Central Karoo, Cape Town Metropole, Eden, Overberg and West Coast).

Over two decades later, HIS intervention processes are still ongoing the Western Cape Province and there is one particularly important intervention that aims to harmonise multiple disparate health-related databases across different types of health services and organisations. This new province-wide HIS intervention can benefit from an understanding of the historical factors that shaped previous HIS development and implementation processes in the province. For example, placing the development and implementation processes of the current HIS and its interventions into historical context will provide a widened understanding of the emerging barriers and

facilitators of HIS processes and products. Current designers, implementers and users of HIS interventions may be able to draw lessons from elements, activities and patterns of history that impacted on previous HIS development, implementation and strengthening efforts [45, 54]. By having an in-depth look at the social factors that influenced past HIS, we can identify the barriers and facilitators of HIS interventions as they were being implemented, as well as highlight the key events that led to the current state.

## **2. Methods**

### **2.1. Data collection and data analysis**

This paper aims to identify and interpret the key historical events and barriers and facilitators that impacted on the development and implementation of HIS interventions in the Western Cape Province of South Africa- from 1994 onwards. This research contributed towards the doctoral studies of the lead researcher and the research was approved by the University of Cape Town Human Research Ethics Committee (HREC ref: 320/2015 and 738/2018).

Relevant articles on the history of HIS in South Africa were identified by the lead researcher (BS) through literature searches. Literature searches were conducted in Google, Google Scholar and relevant organisational websites (such as the Health Information Systems Program, the Health Systems Trust and the South African Department of Health websites). Several searches were conducted intermittently from mid-2015 (from the start of the PhD project) up to mid-2017, until the first draft of this paper was prepared. Examples of the keywords that were used (in various combinations) are history, politics, HIS, health information management and health system reform. The literature searches aimed to identify websites, journal articles, reports or other types of documents that reported on the development and implementation of HIS interventions in South Africa from 1994 onwards (when South African became independent). About 20 records were identified through the literature searches most of which were journal articles. BS then conducted document reviews; the process of collecting data by reviewing existing literature and documents.

She also held multiple interviews with two key informants who were instrumental in the development and implementation of HIS interventions in South Africa. These informants were specifically selected into the study because of their involvement in the development and implementation of HIS interventions since the beginning (from 1994 onwards). The interviews were meant to bring out key historical and contextual information related to the literature and document reviews, which were the main body of data. The interviews were semi-structured; BS asked informants open-ended questions to probe informants to speak on issues that they deemed important in the development and implementation of HIS interventions. BS also asked informants to verify, expand on, or correct details emerging from the document reviews. While only two well-placed informants were interviewed, they were (because of their longevity in the field) able to help contextualise the information and provide clarity on what was emerging from the literature and document reviews. The lead researcher took detailed notes during the interviews (the interviews were not recorded) and asked the informants to recommend and provide her with documents that were relevant for the research (such as, standard operating procedures or reports). Data collection was ongoing, although the interviews were conducted between October 2016 and February 2017.

Data were analysed manually using an inductive analysis approach. The data were first organised chronologically according to key events that took place in the HIS environment post-apartheid. Text from websites, journal articles, internal documents, standard operating procedures and interview notes were then coded and synthesised according to key themes (barriers, facilitators and outcomes) of HIS interventions. Data collected through the document reviews and interviews were triangulated; extracts from the documents were compared and contrasted with interview notes. Triangulation was used to enrich, refute, confirm or explain differences and similarities in findings emerging between documents and between documents and interviews. It was also a useful method for identifying new or clarifying questions for follow-up interviews with the two informants. Data were analysed qualitatively by categorising key aspects of HIS interventions into three themes: barriers, facilitators and outcomes.

## 2.2. The old South Africa

There are two key contextual aspects that influenced health system reform including the development and implementation of HISs post-apartheid. Firstly, apartheid policies (the word ‘apartheid’ means separation in Afrikaans) and the organisation of the apartheid health system led to fragmentation between different levels of the health system and health services (such as clinics and hospitals). The apartheid government had a strong focus on the hospital sector which primarily served the White minority living in urban areas. Other racial groups relied on PHC services, but these were underdeveloped. Secondly, the new South African government introduced a district-based approach to reform the health system through decentralisation. This approach aimed to shift control, planning and allocation of resources from a central office (national) to peripheral offices (districts and provinces) that are semi-autonomous even though they are bound to the national government by a common policy frame. We describe these two contextual aspects (apartheid policies and the district-based approach) below.

The first contextual aspect that impacted on health system reform has to do with apartheid policies that caused fragmentation within the health system and across different levels of the health services. As a start, the apartheid government forcefully implemented political, economic and land policies that structured societies according to race hierarchies [55]. These policies had many consequences, including racial discrimination, introduction of the migrant labour system, the destruction of family life, vast income inequalities and extreme violence [55]. Health policy was specifically used as an instrument of the government in achieving apartheid goals [56, 57]. For example, there were parallel health services and authorities for different racial groups (White, Asian, Coloured and African) which were allocated different operational budgets. Additionally, there were differences in the quality of health services for Africans (or Black people) between rural and urban areas.

Then, the apartheid government created ‘Bantustans’ (or homelands) outside of South Africa and reallocated people to them based on race and ethnicity [55]. Bantustans’ were created to act separately from each other as quasi-independent states, but they were still controlled by the apartheid government. This resulted in 14 health departments; the central department, three White,

Asian and Coloured departments, and ten departments within the Bantustans. The apartheid government expected each Bantustan to create its own health system; however, lack of resources and infrastructure led to poorly organised health services within the Bantustans (where the majority of the Black people lived). As the implementation of apartheid policies progressed, the government seized control over the missionary hospitals that were responsible for providing health services to the people living in the Bantustans. Over time, there were fewer clinicians who remained in the Bantustans; they directed their focus on curative care (patients with life threatening conditions) which made it difficult for the Bantustans to implement resource intensive PHC services. By the end of apartheid, the health system was fragmented into various health departments with a strong focus on the hospital sector, while PHC services were underdeveloped [55].

The second contextual aspect that impacted on health system reform was the district-based approach which the new government introduced to overcome fragmentation of the health system and health services. Many countries have adopted health system decentralisation; that is, to move health administrative functions, responsibilities and resources from national to provincial to district level [58]. Decentralisation “is seen as a mechanism to achieve the following: greater equity and efficiency; greater involvement of and responsiveness to communities; the reduction in the size of the bureaucracy far removed from the communities being served; and greater coordination between social sectors” [59]. In South Africa, decentralisation occurred in the form of a district-based approach for two reasons; one, developing PHC services required local management, and two, health system integration needed to start within smaller units (for one team to manage local clinics, the district hospital and network of dispensaries).

The district-based approach for PHC services faced two main challenges. One, there were power struggles between old and new institutional structures, such as old managers with a curative background were resistant towards new public health interventions with a strong focus on health promotion and disease prevention [11, 54]. And two, there were no comprehensive data standards for collecting, analysing, and reporting health data which led to each district (and each vertical programme within each district) having its own data standards with little or no coordination with others [45].

The Western Cape Province identified the need to develop a district HIS as a mechanism for supporting the district-based approach for PHC services early on (in 1995). The district HIS was defined as a system for gathering, analysing, presenting and using aggregated routine information related largely to health service delivery at district level [60]. It was based on four principles; it promotes the use of data at the level at which it is collected, the data indicators that are collected are useful to measure performance and targets, data collection of all health services at all levels of the health system, and there is an overall district HIS to reduce duplication and parallel information systems.

The district-based approach was spurred on by the sudden urgency of epidemics in South Africa [55]. At the time, the government started disease-specific prevention programmes with the intention to alleviate the high-burden of HIV and TB with the support of donor agents and non-governmental organisations (NGOs) [15]. A decade later (in 2004), the national ART programme was launched, one of the largest and most intensely-funded health care programmes in South Africa and the largest ART programme in the world. HISs were developed alongside these programmes for assessing whether HIV and TB programmes running in parallel were achieving the service providers' and funders' targets [61, 62]. This means that programme-specific HISs were not linked with each other or other HISs.

### **3. Findings**

The findings are organised as two sections. In the first section we identify key events related to the development and implementation of HIS interventions. We look at four specific interventions, that is, a HIS initiative, a software programme, and a provincial and national data set. In the second section we identify key barriers and facilitators related to the four HIS interventions.

#### **3.1. Key HIS interventions after apartheid**

The South African health system was previously organised according to apartheid government policies, including access to health services and the purpose and function of HISs. The development and implementation of HISs became one of the initiatives through which the new

government could reform the health system. Below, we describe four different but related HIS interventions: the Health Information System Program, the minimum data set, the District Health Information Software and the national indicator dataset.

### **3.1.1. The emergence of a HIS initiative: Health Information System Program**

The emergence of the Health Information Systems Pilot Project, or HISP(P) in the Western Cape Province after South Africa became independent in 1994 is one example that stands out of a HIS initiative to reform the health system [11, 54]. HISP originally emerged as a collaboration between the strategic management team responsible for HISs in the province and two local universities who had conducted research on the state of the HISs in Atlantis and Mitchell's Plain (two townships in the Western Cape). Jointly, the strategic management team and the researchers conceptualised a pilot project to develop a district health and management information systems, first using the Western Cape Province as a pilot site and then later expanding into other provinces across South Africa and more recently (from the mid-2000s) to other LMICs. The key actors of HISP were the School of Public Health at the University of the Western Cape (UWC), the Department of Community Health at the University of Cape Town, the Norwegian Computing Centre affiliated to the Institute of Informatics at the University of Oslo, and the Western Cape Department of Health and the National Department of Health (NDoH). The HISP team consisted of a district facilitator and a project coordinator in each district, university staff, activists (from the health sector and NGOs), and two Norwegian researchers (the founders of HISP who were well-connected to government and research stakeholders) [11, 54].

The proposal to design and pilot district HIS interventions received funding from the Norwegian Agency for Development Cooperation (NORAD) for two to three years from 1996. The funding from NORAD ended in 1998. The national government endorsed the roll-out of HISP projects in the following year (early 1999) which then led to an additional 3 years of funding (1999 to 2001) from the Norwegian University Council (NUFU), UNAIDS (through the EQUITY project) and various provinces and councils. Funds were targeted towards national roll-out of software, HIS related research and capacity building (such as institutional infrastructure and training). From

2002, HISP did not receive further funding from international or local institutions and was therefore restructured to do consultancy work as a way of raising funds for its operations.

Key events related to the establishment of HISP have involved the implementation of pilot projects in the Western Cape Province (between 1996 and 1998) and later the implementation of software, training and data standardisation processes nationally (between 1999-2001); as summarised in **Table 1**. The HISP pilot project was started in three areas in Cape Town in 1996; the health sub-districts were Khayelitsha, Mitchells Plain, and Blaauwberg (and the South Peninsula was added three years later). The pilots were started at the lower levels of the health system so that health staff at those levels could take ownership of the HIS interventions and HISP managers could manage any issues arising from the interventions at that small-scale level and make the necessary revisions for future implementations. The first phase of HISP aimed to identify information needs and support interim district management teams. This led to the development of an ‘essential’ data set (the later version is called a ‘minimum’ data set) where data requirements for vertical and donor-funded programmes are integrated into a set of routinely reported data and standards for PHC data [40]. Details about the development of the essential or minimum data set are discussed later. A District Health Information Software (DHIS) application was then developed to support the implementation and use of the essential data set.

The essential data set was implemented in all local government health facilities (Cape Metropole district) in 1997 after several months of negotiations between HISP and local HIS managers. In parallel, the first version of DHIS was developed; it was implemented in 1998 and used to capture and analyse monthly data across management levels in the Western Cape. HISP hosted an Open Day conference in the same year which was seen as the end of the first phase. The conference was attended by representatives from the national government who were impressed with HISP; this led to HISP becoming involved in the development of a new national essential data set (which was formally implemented in 1999) and other provinces (such as the Eastern Cape) deciding to use the DHIS.

The second phase of HISP started with the NDoH officially endorsing HISP as the national collaborator in developing and implementing HIS-related interventions in 1999. Several pilot

projects were then started in other provinces, such as KwaZulu-Natal, Mpumalanga, Northern Cape and the North West. The Joint United Nations Programme on HIV/AIDS (UNAIDS) was the primary funder of the national roll-out of HISP; the national roll-out continued in all provinces in South Africa before DHIS was implemented in other developing countries as well. Even though HISP was endorsed by the NDoH and DHIS was integrated into the government's operations, HISP was not absorbed into the government structure and it did not secure long-term funding beyond 2002. It is at this point that HISP became a not-for-profit (NPO) company in 2003 (HISP-SA) and focused primarily on software development and technical support and consulting. Later on (from 2008 onwards), HISP entered agreements with the NDoH to support emerging HIS interventions.

### **3.1.2. The emergence of a HIS intervention for primary health care: minimum data set**

Once HISP was well-established as a HIS initiative amongst different government and research stakeholders, the team began to develop and pilot different types of HIS interventions, firstly to address known information needs of the PHC services. It emerged from the first phase of HISP (when information needs were assessed) that an essential data set and standards for PHC data were required. The old data sets were biased towards work-related elements (such as staff performance and labour costs) and the data were of poor quality. The process of standardising the PHC data set occurred in three stages. The first stage was for developing and piloting the new minimum data set (MDS) which had additional features to the essential data set. In addition to the data elements provided in the essential data set, it also provided definitions and standards for collecting, analysing and reporting health data and information useful for decision-making, particularly for vertical programmes at the PHC level. The MDS became a key element of the district-based HIS approach especially because each district needed to develop its own HISs. By mid-1997, the new MDS (referred to as the Routine Monthly Report) was implemented in the Western Cape Province with a reduced number of data elements (from around 300 down to 40) to decrease workloads and costs associated to data processing (such as labour or resource costs).

The second stage was about expanding the MDS into other provinces, taking into consideration the key experiences of the MDS process conducted in the Cape Metropole. Key lessons related to

the role of a mediator (a HISP manager) who continued to facilitate communication between different stakeholders (such as district and provincial managers) who were working on the same MDS process; the need for a group of local and provincial managers who jointly took ownership of the new MDS; and data sets already established in other districts and provinces can be used by colleagues who are in the process of identifying their data needs [54]. After the Western Cape Province, MDS development processes took place in the Eastern Cape and Mpumalanga provinces in late 1997 and were finalised in the following year.

A national dataset for PHC was being developed by the National Health Information Systems of South Africa (NHISSA) committee at the same time that HIPS was helping provinces to establish their own minimum data sets. The national data set (referred to as the national MDS) was meant to collate key data elements related to PHC centrally, but provincial stakeholders were inconsistent with submitting data to the national MDS (they did not prioritise the national data collection process). Provincial stakeholders felt like the national MDS resembled the former centralised HIS structure in that national stakeholders were taking a top-down approach in its implementation.

By mid-1998, national stakeholders realised that introducing the national MDS was not a useful approach for several reasons. Firstly, provincial stakeholders were preoccupied with developing and implementing their own MDS processes. They were being helped by HISP which somewhat gave them the independence to establish their own HISs without interference. Secondly, provincial stakeholders were concerned that the introduction of the national MDS would create additional tasks for them and shift the focus of data processes back to national government.

Thus far we have described the first two stages that involve developing and piloting the MDS and implementing it across provinces. These stages revealed the importance of the Western Cape Province as an innovator and pilot site because HISP was successfully established as a HIS initiative and the MDS became a widely recognised HIS intervention for PHC. Additionally, the national government was impressed with HISP developing and implementing the provincial MDS, but later became concerned with HISP only developing provincial MDS processes without integrating them centrally at the national level. The national government's idea to introduce a national MDS to integrate and align provincial MDS centrally was not well-received by the

provincial stakeholders. As mentioned, provincial stakeholders were busy with establishing their own MDS and they were unhappy with the top-down approach of the national MDS. However, concerns about fragmented and misaligned provincial MDS (which national stakeholders were trying to address by establishing a national MDS) were confirmed in a national survey of HISs conducted by HISP [11, 54].

The third stage involved HISP conducting a national survey of HISs in mid-1998. The survey revealed that there was misalignment between the MDS at different levels and provinces. The survey aimed to assess existing MDS processes and knowledge about the type of data being collected within provinces. It was supplemented by a comparison of data collection forms, reports and tables across provinces. Although the template for the MDS was the same across provinces, some provinces modified the MDS (they added or removed data elements) to suit local needs and there were also differences in how long it took for provinces to implement the MDS because of limited resources (such as computers) and level of staff competencies. The survey data echoed concerned stakeholders' perceptions that the HIS was extremely 'messy' and fragmented and there was a need for coordination and standardisation of data processes.

Three key findings have emerged from the three stages of the MDS development process. Firstly, MDS pilots in the Western Cape Province shaped the development and improvements of the other provincial MDS. Secondly, fragmentation and variation between the provincial MDSs was an unintended consequence of the MDS rollout. Thirdly, the MDS processes and the survey revealed that the national stakeholders needed to rethink the national MDS, in terms of the purpose for collecting data at the national-level and data flow between provincial and national levels. The survey revealed that there were tensions between national and provincial stakeholders regarding the national MDS, which led them to realise that they needed to reimagine their relationship to each other and reorganise their roles in the health system.

A fourth finding that cuts across the three stages is related to the tension between standardising but decentralising HIS interventions, and national government wanting to collate some MDS processes centrally. On the one hand, HISP proposed a bottom-up approach where stakeholders in the different provinces could assess their own data needs and develop or modify HIS interventions

accordingly. On the other hand, HISP also introduced HIS interventions such as the MDS to standardise and integrate data processes across provinces, so that provinces can at a minimum use the same MDS template. This caused tension between standardising HIS interventions across provinces (HIPS aimed to build a minimum level of IT infrastructure across provinces) and allowing for flexible HIS interventions. Another source of tension involved national stakeholders wanting provinces to submit data to a central MDS. HISP was focused on developing local data collection and reporting processes, rather than data collection being about lower level staff reporting back to higher level managers. There was a push between decentralising data collection processes and having some records of data collection at the national level for assessing and setting targets such as health services performance.

### **3.1.3. Developing a HIS intervention from the bottom up: District Health Information Software**

Following on from MDS processes and the national survey, HISP developed a free software package between 1998 and 1999 which could be used throughout South Africa. The software was developed by HISP after local stakeholders were unable to develop their own data capturing and data processing applications. HISP developed the software to address fragmentation issues between the provincial and national MDS and deal with information flow inefficiencies caused by hierarchies between those people collecting data at the lower levels of the health system and those people using data to make high-level decisions about the health system. HISP aimed to overcome hierarchies across levels of the health system by shifting control over software and HIS infrastructure, and therefore moving decision-making and accountability, from the central to the local levels. This was to empower various data producers and users, such as local management, health workers and the community and promote a horizontal flow of information and knowledge. Developing the software was however not a straightforward process and it involved the two stages described below.

The first attempt to develop the software, referred to as the District Health Information Software (DHIS), was initiated in mid-1997 (but only took place in 1998 and 1999) when the HISP team invited lower level health staff to work with them to develop the software. As a starting point,

HISP sent the health staff who were going to be involved in developing the software to standard Microsoft Office courses. The assumption was that the health staff (specifically data clerks) would be able to develop simple applications for capturing and processing RMR data once they completed the courses. However, they did not develop any useful applications, because they did not have the support of their managers to develop such applications in the context of competing workloads and they did not feel like they had the technical skills to do so. This led to a meeting in late 1997 where IT developers from local private companies were asked to develop an electronic data capturing and processing platform. Following that HISP hired a skilled IT developer from a local private company and one of the specialists on the HIPS team started working fully with HISP and together they developed the first version of the DHIS that health staff started to use for their core tasks.

This second stage involved the two full-time HISP developers piloting the DHIS in a few Western Cape districts in March 1998 and revising it as required. Revisions of the software took place frequently (on a weekly basis) particularly because it became popular amongst local stakeholders as it facilitated the implementation of the MDS, it helped cement the district-based HISs and it empowered staff and managers in the health services to collate their own data and assess their performance targets locally (decentralisation of health districts). In mid-1998 when an advanced version of DHIS was available, it was then adopted throughout the Western Cape Province to capture and process RMR data, including specific disease and programme data.

#### **3.1.4. National indicator dataset (NIDS)**

The national government decided to standardise data processes across provinces by having a national data set (similar to the provincial MDS, but at the national-level). The need for a national data set arose when HISP started to expand beyond the Western Cape and Eastern Cape provinces. The purpose of establishing a MDS in each province was to empower lower level staff and allow for context-specific data collection; however, there was uniformity in the way HISP implemented software and data processes. HISP software and data processes were meant to, from the start, become the national standard. The national data set (which was conceptualised in mid-2000) therefore became an important feature in consolidating the district and provincial information

systems. The aim was for the provinces to keep their data sets, but the national data would contain all the minimum data elements that provinces must collect and report on. This national data set was revised numerous times, often because additional data elements were added as existing programme data sets expanded and also to accommodate vertical programmes (such as HIV/AIDS programmes). By 2005, the national dataset was expanded and renamed as a national indicator dataset (NIDS). The national data set had the same data elements and standards captured in the provincial MDS, but users of the NIDS focused more on the use of indicators that could be analysed and used to measure health targets for decision-making.

The NIDS accessed data collected from health services and vertical programmes through DHIS; vertical programmes often had different type of data sets sometimes based on donor-funders requirements or data sets that were modified for specific sites or geographic areas. Although data collection happens using different data sets across health facilities and vertical programmes, the same key data elements and indicators are collected for the NIDS. One of the key features of the NIDS is that it was revised regularly to incorporate emerging data needs.

**Table 1.** HIS interventions, stakeholders and key events

	<b>Purpose</b>	<b>Stakeholders</b>	<b>Key events</b>
<b>HISP</b>	To develop and implement efficient HISs that would lead to the decentralisation of and equal access to health services.	University of Cape Town, University of Oslo, Western Cape Department of Health, National Department of Health	<p>First phase: identify information needs and support interim district management teams</p> <p>Second phase: HISP endorsed by NDoH as national collaborator for HIS interventions</p> <p>Current phase: HISP reorganised as a not-for-profit company</p>
<b>MDS</b>	To develop an essential data set and standards for the PHC data that were required.	HISP, Western Cape Department of Health, National Department of Health	<p>First stage: MDS was developed and piloted in the Western Cape Province</p> <p>Second stage: MDS was implemented in other provinces. Implementation was guided by key experiences from the Western Cape Province.</p> <p>Third stage: national survey was conducted to assess existing MDS processes and the types of data collected</p>
<b>DHIS</b>	To support data production and data utilisation, particularly at the lower levels of the health system	HISP, Western Cape Department of Health	<p>First approach: PHC staff were sent to Microsoft Office courses in hope that they would develop applications for RMR data</p> <p>Second approach: Skilled IT specialists developed a PHC software application, which resulted in the development of an application for quarterly tuberculosis data</p>
<b>NIDS</b>	Like the MDS, the NIDS aimed to collect and aggregate PHC data from all the provinces to monitor and evaluate national performance	HISP, National Department of Health	<p>First stage: national essential dataset was continuously revised to capture additional data elements that were added to accommodate new health programmes</p> <p>Second stage: NIDS was established with an increased focus on the use of indicators to measure health targets.</p>

## 3.2. The barriers, facilitators and outcomes of HIS interventions

As described above, HIS interventions focused on identifying information needs, establishing data flows between provincial and national levels, standardising data processes using technical infrastructure and building a national data set (later renamed as NIDS) for aggregating data across provinces. There are several factors that enabled and hindered the processes, activities and outcomes of these HIS interventions when they were initiated. Below, we identify key barriers, facilitators and outcomes of the HIS interventions, and later we draw out key lessons for strengthening existing HIS interventions.

### 3.2.1. Barriers

There is a lack of evidence on the barriers facing HIS interventions [63]. This analysis of the barriers of HISs is useful for identifying important lessons for future interventions, especially in the context of limited resources. We identified several barriers related to the development and implementation processes of HISP, MDS, DHIS and NIDS which are listed in **Table 2**. The key challenges had to do with stakeholders disagreeing about data processes and software, the institutionalisation and sustainability of new interventions, and misconceptions about staff competencies and motivations and about existing workflows, cultures and practices.

As mentioned before, HISP emerged out of a need to develop and pilot district HISs to support the newly decentralised government administrative structures. When HISP first started it invited all relevant local stakeholders to participate in the process of defining, designing, developing and testing the district HIS. The aim was to standardise data collection and reporting (through data processes and software) across all health facilities and types of services within the Western Cape Province. HISP managers felt that HISs needed to be in line with a set of data standards to communicate with other HISs at the same management level or above. They followed a ‘pyramid’ model where HISs were based on a hierarchy of data standards. This meant that each level of the health system (district, provincial and national) could define or modify its own data sets and indicators while at the same time adhering to the minimum data standards across the hierarchy.

Trying to get multiple stakeholders to agree to the pyramid model was challenging, especially when it came to developing the provincial MDS.

The development of the provincial MDS faced three main challenges. Firstly, there were conflicts between frontline staff (who collected the data) and health information officers (who aggregated data sets) about the types of data that should be collected. Frontline staff were concerned with clinical data for patient management, while health information officers were concerned with producing reports for high-level managers who used the reports to assess health system targets and performance. Frontline staff were frustrated with submitting data to the health information officers on a routine basis (monthly and quarterly) without receiving any feedback. They also felt that the data submission requirements were developed without their input many years earlier and data tools were not revised to accommodate new health priorities. Health information officers were concerned that the MDS would impact on their ability to meet reporting requirements; the data needs of clinicians were not their primary focus and they were merely following instructions from high-level managers.

Secondly, within HISP circles it appeared that some vertical programme managers made excessive demands for detailed datasets that contained hundreds of data elements because they saw themselves as independent stakeholders. Each vertical programme focused on a specific disease or aspect of health (for immunisation, women's health, TB and HIV/AIDS and STIs) so managers, in an effort to ensure that all angles of service delivery were reported on and monitored, required large amounts of data. They were concerned that the MDS would not meet their data needs and it would only provide them with data necessary for measuring broad targets set out by provincial and national managers. However, the programme managers' approach to data collection was opposed by the rest of the stakeholders (such as frontline staff and data managers) because it caused duplication between programmes, it added to the workload of frontline staff who captured the data, and the data quality was poorer across a large number of data elements.

Thirdly, data managers (who developed or managed databases) saw themselves as the ultimate figures of authority when it comes to data. They were reluctant to accept the new MDS because they feared that converting to the new system would create gaps in the data collection processes

or replace the data processes they had previously put in place. Data managers were sceptical about whether the MDS would capture all the data that high-level managers required as its primary focus was to support lower level stakeholders in decision-making. They were also concerned that the current hierarchical approach to data processing and reporting would be eliminated by the MDS, which would cause them to lose power. Additionally, discussions about data elements for the MDS reignited long-standing feuds between two groups of data managers. There were managers who worked for the apartheid government and were involved in the development of databases then; previous databases were heavily focused on demographics and staff performance. There were other managers (some worked for the apartheid government) who were in support of new databases to reduce the heavy focus on reporting certain aspects such as demographics. Managers also disagreed about how services should be provided, managed or assessed through data collection; for example, some wanted disease-specific programmes (for TB) to have their own MDS.

Due to similar challenges as those listed above, HIS interventions (as seen with the MDS) were generally delayed, but the national government regarded HISP as successful in starting the HIS development process and HISP was therefore recommended for national roll-out. Despite this national recognition and HISP driving data processes and systems across South Africa, it was not absorbed into the government structure and it struggled to secure funding after the pilot phase. HISP managers felt that a possible reason for this was because of the ambiguous nature of HISP processes and interventions. The HISP founders felt that on the one hand HISP was seen a neutral party (a bottom-up grassroots movement) that could promote the district-based approach of the new government by developing district HISs. But on the other hand, because the organisation was made up of academics and IT specialists and it was associated with foreign funders, the HISP founders thought that the national government did not see HISP fitting in with its long-term political agendas. The issues of HISP not being institutionalised within government and relying on short-term funding was a concern for its sustainability which sometimes led HISP managers to pause HIS activities to complete other income-generating tasks abroad.

HISP was creative in drawing on human resources (a non-monetary resource) within the government structure to minimise the cost of hiring IT specialists; for example, when HISP managers trained data clerks to develop simple applications using Microsoft Office.

Simultaneously, HISP managers were able to alleviate some of the concerns that frontline staff had about the new data collection processes by involving data clerks in the development of the DHIS application. In addition to stakeholder issues and lack of funding, another challenge that HISP managers faced has to do with their own assumptions that data clerks were the most suitable group to train on developing simple applications for capturing and processing RMR data. However, they were faced with several problems when they tried to empower data clerks by providing them with training on data handling processes and helped them to develop a simple data application.

One of the reasons why HISP chose to focus on data clerks was because they were likely the least skilled at the lower level of the health system but were in charge of a valuable task (that is data capturing). The perception that data clerks were unskilled was widespread amongst lower level stakeholders, so they were concerned when HISP decided to train data clerks and also it clashed with the levels of hierarchy when it comes to data collection. In some districts, data collection was split between data clerks (lower level staff) and information officers (higher level staff). Data clerks were responsible for capturing and submitting data routinely to health information officers who would validate and analyse the data before submitting reports to district or provincial information managers. Health information managers reportedly felt threatened that HISP was training data clerks to perform tasks across the data handling process. They questioned what their role would be once data clerks were able to validate, analyse and produce reports independently.

Another issue related to HISP wanting to empower data clerks. HISP managers felt that data clerks were not interested in learning about the data handling process because they did not perceive themselves as the end users of the data that they produce. Data clerks reportedly felt that they would continue to be data producers without receiving meaningful feedback from their seniors. They were also concerned that data capturing applications (one of the aspects of the training) would cause them to lose their jobs because the applications were aimed at streamlining and decentralising data capturing, analysis and use at the facility level. A further issue had to do with facility and district managers (who were usually nurses and doctors); managers reportedly felt side-lined and may have had doubts about the appropriateness (in terms of competence and motivations) of using data clerks. This delayed the training (because they did not immediately

approve for data clerks to attend) and when data clerks returned to the facilities they were expected to continue with their usual work tasks and did not have the time, capacity or support to develop any data applications. Instead of data clerks developing their own data capturing applications, HISP had to later hire an IT specialist from a private company to develop such an application.

Beyond the barriers faced by provincial stakeholders, there were two important barriers at the national level related to NIDS. The barrier had to do with differences between provincial databases and resistance of provincial stakeholders towards a top-down data integration approach. Firstly, there was still fragmentation between data sets, and old data collection tools continued to be used in parallel to the MDS which led to duplication of data processes. Stakeholders were heavily focused on numerator data (headcounts) rather than data indicators with denominators (comparable measures). Some local authorities collected large amounts of data and the data could not easily be analysed within and across provinces because of multiple definitions, for example, for age categories. These issues of large data sets, numerators versus indicators, and different definitions of data elements were some of the barriers to the NIDS.

Secondly, provincial stakeholders were opposed to the top-down approach of the NIDS. The development of the NIDS was limited to NHISSA committee members so provincial stakeholders opposed the NIDS when it was time for implementation at the local levels. Provincial stakeholders appeared to be opposed to interventions that resembled the centralised structure of the apartheid government; they were sceptical whether the data standards developed at the national-level considered local data needs and processes.

In sum, the key barriers identified here are in many ways expected. The process of developing and implementing any type of intervention usually involves multiple stakeholders who have different interests and priorities. Lack of acceptance of the HIS developers into a formalised structure not only results in resistance from some stakeholders but it also creates uncertainties around funding and sustainability of an intervention. HIS developers' misjudgements about data staff's technical competencies and motivations and existing workflows, culture and practices led to poor planning and additional resources to rectify HIS processes.

### 3.2.2. Facilitators

Despite the various challenges, HIS interventions were in many instances successfully developed and implemented across provinces in South Africa. Our analysis of the facilitators of HISs is useful for identifying the key strategies that future HIS developers should consider in their planning of interventions. The role of change agents (the HISP founders), the evaluation of the state of existing HISs (the national survey conducted by HISP) and the horizontal sharing of experiences and replication of data sets were important facilitators for the development and implementation of HIS interventions. The facilitators of HISP, MDS, DHIS and NIDS are also listed in **Table 2**.

**Table 2.** Barriers, facilitators and outcomes of HIS interventions

Barriers	Facilitators	Lessons
<ul style="list-style-type: none"> <li>• multiple stakeholders with conflicting interests and priorities</li> <li>• not being institutionalised within government</li> <li>• uncertainty about sustainability due to lack of funding</li> <li>• misleading assumptions that health staff will develop their own data management tools using technology</li> </ul>	<ul style="list-style-type: none"> <li>• collaboration between government, activists and universities</li> <li>• change agents who are developers, implementers and users of HIS interventions</li> <li>• conducting research on the state of the HIS and identifying actions</li> <li>• drawing on the experiences of pilots to revise interventions</li> </ul>	<ul style="list-style-type: none"> <li>• “Cultivation” can be used to transform social systems while new HIS interventions are being introduced</li> <li>• Balancing the role of champions against the need to institutionalise HIS interventions</li> </ul>

One of the first tasks of HISP was to bring together different stakeholders across the hierarchy of the HIS. HISP itself was made up of University staff, activists from the health sector and NGOs and two Norwegian researchers (they also founded HISP). It was further able to gain the interest of provincial stakeholders who were involved in data production and health service delivery, as well as national stakeholders who were involved in designing data policies and making decisions to uniformly improve health services across the country. The founders of HISP played a key role in bringing these stakeholders together. When the MDS was being developed, one of the founders served as the moderator between different stakeholders; in that he tried to help frontline staff and data managers resolve their conflicts about provincial data collection processes. Such an individual

can be described as a change agent, because he was respected by the disagreeing parties and he knew the issues well enough to kick-start a transformation process. Other change agents were a select few individuals (such as facility or district managers) who took ownership of the new HIS interventions and championed them among other stakeholders.

Identifying change agents was not only important for navigating conflicts between stakeholders, but it was also necessary for promoting HIS interventions especially amongst usually disempowered HIS actors. For example, the data clerks who participated in the HISP training were able to promote the uptake of HIS interventions because they felt included in the process of developing the DHIS application. The HISP founders identified their approach to empower lower level health staff and thereby transform old social and technical systems as ‘cultivation’. Cultivation is based on the sociotechnical perspective (outlined in the Interactive Sociotechnical Analysis framework) that HISs and HIS interventions are embedded in social processes and systems. In the discussion section, we explain how HISP applied a cultivation strategy to deal with conflicts between different stakeholders and encouraged lower level staff to participate in the development and implementation of HIS interventions.

Another facilitator of HIS interventions was the national survey that was conducted by the HISP team to assess the state of the HIS especially at the PHC level. The benefits of the survey was that it used different methods to assess the MDS used at all levels of the health system, the results and recommendations were quickly discussed and adopted in the NHISSA committee, as well as incorporated into the work of HISP and training courses at UWC, and the collation and analysis of data sets also guided the development of the DHIS application. Overall, the survey triggered important actions and developments because it revealed that there was still poor coordination and standardisation across provinces. One action that emerged out of the survey was the establishment of a new national MDS (which later became the NIDS).

HISP managers relied on the wide range of experiences from pilot sites to replicate data sets for the rest of the provinces. The HISP managers were open to engage various stakeholders and experiences but they also knew when to make decisions and initiative action which was an important facilitator for HIS interventions. After some time of provincial stakeholders not being

able to come to an agreement about the data elements for the MDS in the Western Cape, the HISP managers decided on how to proceed with the development to prevent discouraged stakeholders from abandoning the process entirely. This led to the development of the first MDS version which was later revised to include new needs, and then it was finalised for other provinces who could also modify it according to their needs.

The key facilitators of HIS interventions identified through the examples of HISP, MDS, DHIS and NIDS have to do with the role of change agents as mediators of tense discussions and champions for new interventions. Other facilitators, such as conducting research on existing data processes and adapting future developments according to past successes, highlight important lessons.

## **4. Discussion**

This paper provides a historical analysis of the development and implementation of HIS interventions, the years after South Africa became independent. We described how HIS processes emerged over time and the impact of both social and technical factors on future events. We identified and interpreted key events and barriers and facilitators of different HIS interventions taking place across time. By chronologically tracking the key events (the establishment of HISP, MDS, DHIS and NIDS) we were able to better understand the step by step process of developing HIS interventions and the iterative process (of testing and trying) of implementing them. The key barriers and facilitators were related to stakeholder engagement, resistant actors or change agents, institutional and financial factors, and the role of pilots in developing interventions. Below, we summarise the key findings.

HISP engaging with multiple stakeholders was important for identifying data needs, but it was sometimes a barrier to advancing HIS interventions quickly because of conflicting interests and priorities amongst stakeholders. The position of HISP as a bottom-up grassroots movement (their focus was on empowering lower level staff by involving them in data processes) made it possible to collaborate with different stakeholders (such as government, activists and researchers) because they were generally perceived as neutral in debates about data needs. HISP was well-positioned to

provide all stakeholders across hierarchical levels with an equal opportunity to participate in the development of HIS interventions. However, HISP did not become institutionalised into the government structure because it did not fit into the political agendas of national stakeholders and HISP was too involved with university, non-governmental and overseas stakeholders and other projects.

On the up side, HISP initially received funding for pilot projects from external donors which was useful for getting HIS reform off the ground. Pilots projects were used as a testing ground and HISP was therefore able to identify aspects that needed improvement and revise HIS interventions for large-scale rollout. The challenge came when the funding period for the pilot projects ended. It was difficult for HISP to continue implementing HIS interventions across the country without much financial support from the government. Lack of funding threatened HISP's sustainability and to overcome this, implementation processes of HIS interventions were intermittently stopped so that the HISP team could work on paid projects to generate income to fund its operations. When HISP tried to minimise the problem of funding by training data clerks on data handling processes and helping them develop a data capturing application, they faced resistance from health information officers and facility and district managers. These stakeholders reportedly felt that data clerks were incompetent and unmotivated, and it seemed that they were concerned about their job security because newly trained data clerks could replace them, or data applications would take over their data analyse and reporting tasks.

Our findings are generally consistent with other studies that have examined the key barriers and/or facilitators of HIS interventions [49, 64, 65]. We will draw on three studies to illustrate this. The first study is by Waterson et al. and it focused the introduction and evolution of Health Information Technology (HIT) innovations in the United Kingdom (UK) National Health Service (NHS) [49]. The authors state that there were tensions between local and national strategies for implementing HIT because different stakeholders had different data needs and priorities. Although this study was conducted in a different setting to ours (high income setting), we have a similar finding related to tensions between different stakeholders, that is, stakeholders at different levels of the local health systems and stakeholders at provincial or national levels.

The second study by Kostadinovski et al. is focused on the institutional and financial factors that impact on HISs and HIS interventions [64]. The authors emphasise the importance of incorporating HIS interventions into well-established organisational structures (such as government) in order to tap into existing human and financial resources. This was also a finding in our study; for example, HISP aligned itself with local and national government and HISP managers tried to involve health staff as a form of additional help in developing and implementing HIS interventions.

The last study by Barnett et al. is focused on champions and how they experienced the development and implementation of HIT innovations [65]. The authors identified four factors that champions felt facilitated or obstructed HIT innovations; namely, the role of evidence (to prove effectiveness), the function of interorganisational partnerships, the influence of human resources and the impact of contextual factors. In our study, we also found that pilots were important to prove that HIS interventions were necessary and useful; HISP was based on a collaboration between government, NGOs and universities, HIS developers sought buy-in from local and national stakeholders right from the start; and disparities in staff competencies and IT infrastructure across provinces impacted on the time and efforts required to implement HIS interventions.

We identified two key findings in addition to those documented in the literature. The first finding involves a process of ‘cultivation’. Cultivation is a slow and incremental process to transform a social system [54]. In this case, this involved HISP bringing together various stakeholders from different hierarchical levels to jointly build HIS infrastructure. One point about cultivation is that both people (stakeholders) and technology are part of a social system. People interact with each other and people interact with or about technology to cause certain challenges and opportunities for the HIS intervention that is being developed. For example, debate between stakeholders about data needs and priorities was a vehicle for cultivating a culture where people working across different levels of the health system (data clerks, health information officer, facility managers and district managers) can interact with one another. The cultivation strategy was for HISP managers to focus the debate on data issues or HIS interventions in order to empower lower level health staff (who collect the data). This strategy created a space for different people to talk about the different PHC and data management contexts that they work in and how emerging HIS interventions will impact their work.

Another point about cultivation is that there is no clear-cut division between what HIS developers envisioned and how users will ‘translate’ (or operationalise) the HIS intervention [11]. A cultivation strategy can lead to the design and implementation of a HIS intervention being changed numerous times. On the one hand, HIS developers can inscribe their desired user-behaviours into the intervention they develop. For example, the DHIS (software) development process initially started with HISP managers training data clerks to make their own data management applications. HISP managers were inscribing a culture of empowerment of lower level staff in data management processes. On the other hand, users can react differently or not at all to what HIS developers expect and diminish the control that HIS developers had to change user-behaviours. In the case of HISP, data clerks did not use the computer skills they learned to develop data management applications as HISP managers intended, so HISP managers had to come up with another plan (they hired external IT developers but kept data clerks involved). Cultivation as a strategy to transform social systems (such as to encourage open discussion and empower lower level staff) can therefore lead to an effective, iterative development and implementation process.

The second finding is about balancing the role of change agents (or champions) against the need to institutionalise HIS interventions. The case of HISP, the HISP managers can be described as champions because they played an important role in designing, developing and implementing innovative HIS interventions. Pilots for HIS interventions in South Africa initially started in the Western Cape Province where the HISP managers were based. National government later endorsed HISP as their main collaborator to implement data management software and processes. Lack of funding and institutionalisation of HISP within the government structure posed as a key challenge to HIS interventions moving from pilots to practice. This highlights a tension, where on the one hand, HIS interventions need to be led by champions, especially earlier on when local buy-in is still required and when interventions still need to be tested and revised [66]. On the other hand, the role of the institution (government) becomes increasingly important as the HIS interventions become fully developed and are ready for large-scale rollout. In the case of HISP, it was important for the national government to endorse HISP as the national standard (this was a key moment in HISP continuing its operations). But because HISP was not a government structure, the implementation of HIS interventions was delayed in some provinces as HISP managers needed to balance national rollout activities with generating independent income.

To summarise, the two key findings that emerged from this historical analysis relate to the concept of cultivation as a strategy to engage and empower different stakeholders; and the need to balance the short-term role of champions to get innovations off the ground against the long-term role of institutions to provide funding and sustainability for innovations. A possible limitation of this historical analysis was that we could not explore the issues of HIS interventions from the viewpoints of the multiple stakeholders involved. There were only two interviewees (it would have been difficult to identify other stakeholders) and there was little written in the literature on HIS interventions in the South Africa. Combining evidence from the interviews and the literature did however help us triangulate the findings.

In conclusion, we found similar barriers and facilitators of HIS interventions as other studies. These related to stakeholder engagement, the role of champions, institutional and financial factors, and the role of pilots in advancing the development of HIS interventions. Two additional aspects that we found in our study are related to the strategy (cultivation) that HIS developers used to manage stakeholder discussions, and the role of champions at inception of HIS interventions and the need for institutionalisation as HIS interventions progress. A key lesson is the importance of iterative development, flexibility in implementation and ongoing engagement with key stakeholders.

## **Contributions**

The PhD student (BS) conducted data collection and data analysis. Her PhD supervisors (CC and NL) provided guidance throughout the research process and provided critical feedback of drafts.

# **CHAPTER 4: DEFINITIONS, COMPONENTS AND PROCESSES OF DATA HARMONISATION: A SCOPING REVIEW**

---


Chapter 4 consists of two parts. The first part is a copy of the published protocol of the scoping review on definitions and concepts of data harmonisation. The second part is the write-up of the scoping review.

PROTOCOL

Open Access



# Defining and conceptualising data harmonisation: a scoping review protocol

Bey-Marrié Schmidt<sup>1,2\*</sup> , Christopher J. Colvin<sup>3,4,5</sup>, Ameer Hohlfeld<sup>2</sup> and Natalie Leon<sup>4,6</sup>

## Abstract

**Background:** Data harmonisation is an important intervention to strengthen health systems functioning. It has the potential to enhance the production, accessibility and utilisation of routine health information for clinical and service management decision-making. It is important to understand the range of definitions and concepts of data harmonisation, as well as how its various social and technical components and processes are thought to lead to better health management decision-making. However, there is lack of agreement in the literature, and in practice, on definitions and conceptualisations of data harmonisation, making it difficult for health system decision-makers and researchers to design, implement, evaluate and compare data harmonisation interventions. This scoping review aims to synthesise (1) definitions and conceptualisations of data harmonisation as well as (2) explanations in the literature of the causal relationships between data harmonisation and health management decision-making.

**Methods:** This review follows recommended methodological stages for scoping studies. We will identify relevant studies (peer-reviewed and grey literature) from 2000 onwards, in English only, and with no methodological restriction, in various electronic databases, such as CINAHL, MEDLINE via PubMed and Global Health. Two reviewers will independently screen records for potential inclusion for the abstract and full-text screening stages. One reviewer will do the data extraction, analysis and synthesis, with built-in reliability checks from the rest of the team. We will use a combination of sampling techniques, including two types of 'purposeful sampling', a methodological approach that is particularly suitable for a scoping review with our objectives. We will provide (a) a numerical synthesis of characteristics of the included studies and (b) a narrative synthesis of definitions and explanations in the literature of the relationship between data harmonisation and health management decision-making.

**Discussion:** We list potential limitations of this scoping review. To our knowledge, this scoping review will be the first to synthesise definitions and conceptualisations of data harmonisation in the literature as well as the underlying explanations in the literature of the causal links between data harmonisation and health management decision-making.

**Keywords:** Data harmonisation, Data linkage, Health information exchange, Routine health information system, Scoping review, Health management decision-making

## Background

An effective health system relies on a routine health information system (RHIS) that provides the informational support needed by health managers to identify gaps in service delivery and to inform planning, implementation and monitoring of interventions [1]. However, many countries, especially in low-and middle-income settings, do not have

well-functioning routine health information systems (RHISs) to monitor and evaluate their work [2, 3]. This limits countries' ability to improve the effectiveness, efficiency, quality and equity of their health services.

Given the increasing availability of large electronic databases of routine health information, health authorities and managers, information technology (IT) stakeholders and researchers have identified data harmonisation as an important intervention for strengthening routine health information systems (RHISs) [4, 5]. There is often a lack of coordination and integration of large electronic databases; this is typically due to inconsistencies between key variables and indicators for collecting, analysing and

\* Correspondence: [schbey001@myuct.ac.za](mailto:schbey001@myuct.ac.za)

<sup>1</sup>School of Public Health and Family Medicine, Faculty of Health Sciences, University of Cape Town, Falmouth Rd, Observatory, Cape Town 7925, South Africa

<sup>2</sup>Cochrane South Africa, South African Medical Research Council, Cape Town, South Africa

Full list of author information is available at the end of the article



reporting health information across programmes [8]; the production of poor quality data that cannot easily be exchanged [6]; and programmatic fragmentation across levels of the health system which can result in the duplication and excessive production of data [9]. Data harmonisation has the potential to address all these problems, through coordination, linkage and integration of existing large-scale databases [6–8].

Harmonised data sets also have the potential to improve informational support for health management decision-making and, in turn, support health systems strengthening [9, 10]. However, data harmonisation interventions may take on different parts of the problem of fragmented systems, use different definitions and may have different intended outcomes with regard to improving routine health information systems. This makes comparison and assessment of its usefulness for improving health systems functioning difficult to assess. A second challenge is that sometimes even when quality and timely health information are available, limited access to and use by management for planning, monitoring and evaluation and quality improvement is still a problem [6, 7, 9]. Data harmonisation has the potential to provide timely, relevant and accessible informational support for health management decision-making [12, 13], but we need to better understand how data harmonisation might actually work to improve decision-making. In this review, we are interested to learn more about the scope of data harmonisation definitions and activities as well as how those working in this field understand its effect on management decision-making. Our assumption is that harmonised routine health information may increase access to and use of relevant routine health information which could improve management decision-making and tasks of monitoring, evaluation, planning and ongoing quality improvement. We define effective management decision-making as the proactive and interactive process that demands and uses the best available data (well-integrated, complete and accurate data) during programme development as well as monitoring and evaluation [9].

#### Why it is important to do this scoping review

There is growing recognition that the successful implementation of data harmonisation interventions occurs in multiple technical and social (i.e. organisational and behavioural) contexts. This multi-faceted nature of data harmonisation has resulted in a range of different terms being used for interventions with similar aims and activities [11]. For example, terms such as *data integration* [12], *data linkage* [13] and *health information exchange* [10] are all used to describe data harmonisation-type activities, and it is not always clear the extent to which these efforts are similar in practice, scope and relevance. While the use of multiple terms is not a problem in

itself, lack of clarity on what constitutes ‘data harmonisation’ makes it difficult to compare studies and synthesise evidence on impact.

Lack of understanding of the underlying causal mechanisms between the data harmonisation activities and the intended outcomes for health management decision-making also makes it difficult to compare interventions and to evaluate the impact and implications for health systems strengthening. Having a clearer idea of the range of definitions and concepts used, the various components and activities included in data harmonisation interventions and the proposed underlying causal mechanisms being tested can help inform researchers and health system decision-makers on the design, implementation and evaluation of different data harmonisation interventions [14].

Since 2012, there have been three systematic reviews on data harmonisation and related activities, indicating a growing interest in the topic. The reviews were concerned with the integration of health information found in multiple databases across multiple organisations, for the purposes of clinical and service improvements, and for research analyses. One review focused on the determinants of RHIS performance and its role in improving health systems functioning and performance at the local level [9]. Another focused on views of health care professionals on data sharing or data linkage of clinical data for research purposes [8], while the third focused on barriers and facilitators of health information exchange (HIE) in LMICs [12]. Consistent with what was found in primary studies of data harmonisation processes, these reviews used a variety of terms to explain the integration and exchange of health information [15]. Data harmonisation was defined both narrowly and broadly depending on its objectives; in one review, *data linkage* was used solely to describe the technical stages of combining multiple databases [8], while in another, *health information exchange* was used to describe similar as well as broader processes involving multiple stakeholders to mobilise information across various systems, organisations and geographical areas [16]. It is important to identify and synthesise these variations in terminology in a systematic way, to reflect both the range of activities, but also to identify the commonalities, and build an understanding of how data harmonisation interventions are thought to work to support the different needs of implementers and/or users of harmonised data.

#### Methods

This scoping review will follow the methodological stages for scoping studies proposed by Arksey and O'Malley [15] who recommend a process that is “not linear but iterative, requiring researchers to engage with each stage in a reflexive way” in order to achieve both

'in-depth and broad' results. The steps involved are identifying the research question, identifying relevant studies, selecting studies for inclusion, data extraction and data synthesis.

### Study question and objectives

This scoping review aims to appraise the characteristics of studies on data harmonisation and the definitions used for data harmonisation activities and to develop an understanding of the intended effect of data harmonisation interventions on management decision-making. The objectives are:

1. To identify and synthesise the characteristics of studies of data harmonisation;
2. To identify and synthesise the various definitions and concepts used to describe data harmonisation interventions, and
3. To develop a conceptual understanding of explanations in the literature of the causal relationship between data harmonisation interventions and health management decision-making.

In order to inform our understanding of the causal mechanisms (including the role of key contextual socio-technical dynamics) (objective 3), we will draw on information extracted for objectives 1 and 2 and, in addition, extract data on the descriptions of the components, processes, contexts and intended causal pathways of data harmonisation interventions. Such a synthesis has the potential to broaden and clarify the knowledge base of researchers and health management about the range of and variation in data harmonisation interventions, and the intended relationship between the components (individually or in combination) and management decision-making.

### Identifying relevant studies

#### Eligibility criteria

Peer-reviewed research studies (no methodological restrictions) and grey literature on data harmonisation in health-related information databases are eligible if they provide (a) a definition and/or a conceptualisation of data harmonisation (and/or related terms) and/or (b) a description of a data harmonisation intervention (in terms of components and processes and causal mechanisms) and/or (c) contribute to an explanation of the causal relationship between data harmonisation and health management decision-making (for example, through improved quality and accessibility of harmonised information for management and or the utilisation of harmonised health information for management decision-making). Studies concerned with various

technical aspects of data harmonisation, such as changes in key variables and indicators, software and hardware infrastructure for data generation, and in reporting and feedback procedures, are also eligible, provided it is considered part of a data harmonisation intervention.

### Search strategy

The search will identify all relevant studies from the year 2000 onwards (01 January 2000 to 31 July 2018). This is around the time that large-scale digitisation of routine information started to be implemented (especially in LMICs), and when policy-makers and researchers became interested in harmonisation of large digital databases [2, 3, 9, 16]. The following electronic databases will be searched for eligible studies:

- CINAHL, EbscoHOST
- MEDLINE via PubMed
- Global Health
- Science Citation Index and Social Sciences Citation Index, ISI Web of Science
- Relevant websites, such as the World Health Organization (WHO) and MEASURE Evaluation websites

Search terms will include a distillation of keywords and Medical Subject Headings (MeSH) terms related to data harmonisation (concept A) and health information system (concept B). We have developed a preliminary search strategy using relevant keywords and MeSH terms (see Additional file 1). To ensure that we do not miss potential studies, we will apply an iterative approach using known studies that meet the inclusion criteria identified during preparation of the protocol. Studies known to meet the inclusion criteria will be searched for among "hits" (search records) and used to identify new keywords and MeSH terms not already included in the search strategy. Once the search strategy has been finalised using the PubMed database, we will tailor it to each database and report on the adaptations. Searches will be limited to English as we do not have the resources required for reviewing non-English literature. There will be no geographic restrictions.

In addition to the electronic searches, review authors will (a) search the reference lists of all included studies and key references (for example, relevant systematic reviews) and (b) contact authors of included studies and/or experts in the field for additional references.

### Selecting studies for inclusion

#### Screening records

The initial search from different sources will be conducted to identify a database of records (title and abstracts) of relevant studies. The search results will be

collated in the Endnote reference management programme and duplicates removed [17]. The final search database will then be uploaded into Covidence, an electronic programme designed for managing the screening process in systematic reviews (<https://www.covidence.org>). Two reviewers (BS and AH) will then independently screen the records to evaluate their eligibility for full-text review. The full texts of those studies identified as potentially relevant will be retrieved and read by the two reviewers to make a final decision about inclusion. During this full-text review stage, where necessary, study authors will be contacted for further information. At both the abstract and full-text screening stages, conflicts will be resolved by the two reviewers (BS and AH) first attempting to reach a consensus view; failing which, a third reviewer (NL) will be the final arbitrator. The study selection process will be summarised using a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram.

### **Sampling**

We will use a combination of sampling techniques, including two types of 'purposeful sampling', a methodological approach that is particularly suitable for the focus of our scoping review [18]. These sampling techniques are intended to address both the breadth (for example, exploring the characteristics of studies on data harmonisation) and depth (for example, definitions, concepts, components, processes and explanations of casual mechanisms of data harmonisation) [18, 19] of the review.

For objective 1, we will not apply any sampling strategy, in order to ensure we capture the characteristics of the widest range of studies on the topic. For objectives 2 and 3, we will apply both maximum variation sampling (to identify both variation and similarities in definition and concepts and intervention descriptions) as well as theoretical sampling (where we will sample in relation to emerging theoretical insights and questions to provide a sufficiently 'rich' synthesis of descriptions of underlying causal mechanisms) [19]. The theoretical sampling will be iterative as we will start with synthesising emerging insights and may then loop back and look for more studies.

### **Data extraction or 'charting the data'**

Once the list of papers to be included is finalised, the data extraction and sorting process (also referred to as 'charting of data' in Arksey and O'Malley) is the next step [15]. Data extraction of all the included studies will be conducted by one reviewer (BS), using the data extraction framework presented in Fig. 1. The extraction framework will be used to collect, sift and sort data that can address the three objectives of the review. This will be a mixture of general 'demographic' information about

the study (such as country, level of the health care system) and specific information about the data harmonisation intervention (such as definitions, types of routine information systems, components, outcomes) and suggested causal mechanisms for the effect of data harmonisation on management decision-making. The framework will be piloted on the first few studies and revised where necessary. One other reviewer (AH) will independently conduct data extraction for a random sample of 10% of the included studies to increase reliability.

The process of data extraction and sorting will be done in Excel, using the data items in the data extraction framework (Fig. 1) to fill in information for each of the items in the framework. This will also allow for comparison of key items across studies and allow for synthesis within and across data items (for example, comparing definitions across studies, or comparing within one study, the definition and the description of the intervention components and processes).

As this scoping review aims to identify various characteristics, definitions and causal mechanisms of data harmonisation, we will not conduct any risk of bias or quality assessment of included studies. This approach is consistent with scoping reviews of similar aims and methodological frameworks for conducting scoping reviews [15, 20, 21].

### **Data synthesis or 'collating, summarising and reporting the findings'**

One review author (BS) will conduct data analysis, using manual coding and data synthesis methods on the extracted data from included studies. Another reviewer (NL) will review the data analysis work on an ongoing basis as an additional quality check.

This review will combine quantitative and qualitative syntheses to provide an overview of our findings. First, we will present an overview of all the included studies using a numerical analysis of the key characteristics of the studies [15]. The numerical synthesis will include following categories: income level of the country, the level of the health care system targeted in the intervention (for example primary health care, hospital-level, community-based health care), the particular type of routine health information systems involved (for example, clinical care, finance, human resources or drug supply information systems), the governance/management level targeted in the intervention (for example facility, district, regional or national levels) and types of patient population or disease programme (for example non-communicable disease or adult reproductive health).

The second synthesis approach will be a qualitative narrative synthesis [21] of data harmonisation definitions and of the conceptual models for understanding of how

**Data extraction items: Defining and conceptualising data harmonisation****Study and author details**

Title: .....  
 Author: .....  
 Publication date: .....  
 Publication type: .....  
 Study design: .....  
 Country of study: .....

**Detailed characteristics of studies**

Country income classification   
 Level of the health care system   
 Types of routine health information systems   
 Governance/management level targeted in the intervention   
 Types of patient population or disease programmes targeted

**Study content**

Definition of data harmonisation   
 Alternative term (s) used for similar activities   
 Description of the data harmonisation intervention (for example, components and processes)   
 Description of causal mechanisms, conceptual framework, or logic model of data harmonisation effects on health management decision-making

**Comments****Fig. 1** Data extraction framework

data harmonisation is meant to improve health management decision-making. We will collate and summarise definitions of data harmonisation and related concepts describing data harmonisation activities by looking for the key components across definitions and for key variations. We will code and synthesise the extracted data to identify the key issues that emerge regarding components and processes of data harmonisation interventions, the expected outcomes and impacts, and the factors influencing data harmonisation effects on management decision-making (including the steps of production, access and/or utilisation of health information).

To summarise, the numerical and narrative synthesis will result in three sets of findings: (a) an overview of key characteristics of data harmonisation studies, (b) the definitions and conceptualisations of data harmonisation, and (c) a narrative synthesis of the relationship between data harmonisation and health management decision-making.

Finally, we will ensure that the reporting of our findings is aligned with the PRISMA 2015 statement presented in Additional file 2.

**Ethics and dissemination** This is a scoping review of completed studies, so no ethical approval is required. The results will be disseminated through peer-reviewed publications and conference presentations as well as shared with local and national stakeholders engaged in data harmonisation projects.

**Discussion**

To our knowledge, this scoping review will be the first to synthesise definitions and conceptualisations of data harmonisation in the literature, as well as the underlying explanations in the literature of the causal links between data harmonisation and health management decision-making. Given time and financial constraints, we will only search for English studies published after 2000; potentially relevant studies may be missed. Applying purposeful sampling techniques will assist with addressing both breadth and depth of explanation in this scoping review, but it may also result in missing potentially useful content [18, 19]. This scoping review will be of interest to designers, implementers and users of data harmonisation interventions; it will broaden understandings of the range and complexity of studies, definitions, systems, organisations and stakeholders involved in such interventions and of the intended causal pathways for improving health management decision-making.

**Additional files**

**Additional file 1:** Search strategy developed in PubMed database. (DOCX 14 kb)

**Additional file 2:** PRISMA-P 2015 Checklist. (DOCX 32 kb)

**Abbreviations**

HIE: Health information exchange; LMICs: Low- and middle-income countries; MeSH: Medical Subject Headings; PRISMA: Preferred Reporting Items for

Systematic Reviews and Meta-Analyses; RHIS: Routine health information system; WHO: World Health Organization

#### Acknowledgements

We would like to thank Ms. Gill Morgan, University of Cape Town, who assisted with developing the search strategy.

#### Funding

Time to write this paper was supported by the US National Institute of Mental Health (grant number 1R01MH106600) and the South African Medical Research Council (SAMRC). The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the US National Institutes of Health or the SAMRC.

#### Availability of data and materials

Not applicable.

#### Authors' contributions

BS drafted the protocol together with NL and CC. AH contributed to the development of search strategy. All authors reviewed and approved the final manuscript before final submission for peer review.

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Author details

<sup>1</sup>School of Public Health and Family Medicine, Faculty of Health Sciences, University of Cape Town, Falmouth Rd, Observatory, Cape Town 7925, South Africa. <sup>2</sup>Cochrane South Africa, South African Medical Research Council, Cape Town, South Africa. <sup>3</sup>Division of Social and Behavioural Sciences, School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa. <sup>4</sup>Department of Epidemiology, School of Public Health, Brown University, Providence, USA. <sup>5</sup>Department of Public Health Sciences, University of Virginia, Charlottesville, USA. <sup>6</sup>Health Systems Research Unit, South African Medical Research Council, Cape Town, South Africa.

Received: 4 July 2018 Accepted: 21 November 2018

Published online: 06 December 2018

#### References

- World Health Organization. Everybody's business-strengthening health systems to improve health outcomes: WHO's framework for action. Geneva: WHO; 2007.
- Lippeveld T, editor. Routine health information systems: the glue of a unified health system. Keynote address at the Workshop on Issues and Innovation in Routine Health Information in Developing Countries, Potomac, March; 2001.
- World Health Organization. Country health information systems assessments: overview and lessons learnt. Geneva: WHO; 2012.
- Boule A. The use of routine health data sources in the Western Cape to address health service questions (presentation). Health Impact Assessment Directorate, Western Cape Department of Health. 2014.
- Mehta U, Heekes A, Kalk E, Boule A. Assessing the value of Western Cape Provincial Government health administrative data and electronic pharmacy records in ascertaining medicine use during pregnancy. *S Afr Med J*. 2018; 108(5):439–43.
- Mutale W, Chintu N, Amoroso C, Awoonor-Williams K, Phillips J, Baynes C, et al. Improving health information systems for decision making across five sub-Saharan African countries: implementation strategies from the African Health Initiative. *BMC Health Serv Res*. 2013;13:59.
- Karuri J, Waiganjo P, Daniel O, Many A. DHIS2: the tool to improve health data demand and use in Kenya. *J Health Inform Dev Countr*. 2014;8:38–60.
- Heywood A, Boone D. Guidelines for data management standards in routine health information systems. *Measure Evaluation*. 2015.
- Nutley T, Reynolds H. Improving the use of health data for health system strengthening. *Glob Health Action*. 2013;6(1):20001.
- Akhlaq A, McKinstry B, Muhammad KB, Sheikh A. Barriers and facilitators to health information exchange in low- and middle-income country settings: a systematic review. *Health Policy Plan*. 2016;31(9):1310–25.
- Fichtinger A, Rix J, Schäffler U, et al. Data harmonisation put into practice by the HUMBOLDT Project. *Int J Spat Data Infrastruct Res*. 2011;6:234–60.
- Lippeveld T. Routine health facility and community information systems: creating an information use culture. *Glob Health Sci Pract*. 2017;5(3):338–40.
- Hopf YM, Bond C, Francis J, et al. Views of healthcare professionals to linkage of routinely collected healthcare data: a systematic literature review. *J Am Med Inform Assoc*. 2014;21:e6–10.
- Cresswell KM, Sheikh A. Undertaking sociotechnical evaluations of health information technologies. *J Innov Health Inform*. 2014;21:78–83.
- Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. 2005;8(1):19–32.
- Nutley T, Gnassou I, Traore M, et al. Moving data off the shelf and into action: an intervention to improve data-informed decision making in Cote d'Ivoire. *Glob Health Action*. 2014;7:25035.
- Bramer WM, Giustini D, de Jonge GB, et al. De-duplication of database search results for systematic reviews in EndNote. *J Med Libr Assoc*. 2016; 104(3):240.
- Benoot C, Hannes K, Bilsen J. The use of purposeful sampling in a qualitative evidence synthesis: a worked example on sexual adjustment to a cancer trajectory. *BMC Med Res Methodol*. 2016;16:21.
- Suri H. Purposeful sampling in qualitative research synthesis. *Qual Res J*. 2011;11(2):63–75.
- Tricco AC, Lillie E, Zarin W, et al. A scoping review on the conduct and reporting of scoping reviews. *BMC Med Res Methodol*. 2016;16:15.
- Popay J, Roberts H, Sowden A, et al. Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme. 2006.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



# **Definitions, components and processes of data harmonisation: a scoping review**

## **1. Introduction**

Health information technology (HIT) innovations involve the design, development and implementation of electronic and interoperable systems for the health sector. They have the potential to strengthen aspects of the health system; for example, through RHISs that can produce accurate, complete and up-to-date information and support decisions, actions and changes across all components and levels of the health system [6, 12]. This means that information captured by well-functioning RHISs can help health managers identify gaps in health service delivery as well as inform their planning, implementation and monitoring of interventions [48, 67]. Data harmonisation (DH) is one type of a HIT innovation that can help RHISs function at their best; it involves coordination, linkage and integration of existing large-scale health-related databases [20]. This is especially useful in low-and-middle income (LMIC) settings where problems related to the inconsistencies between key variables and indicators for collecting, analysing and reporting health information across programmes are more common[15]. Other challenges to effective RHIS functioning include the production of poor quality data that cannot easily be exchanged and programmatic fragmentation across levels of the health system which can result in the duplication and excessive production of data [9].

Lack of standardised data production processes, fragmentation of databases, and errors and duplication in data production are only some of the challenges of RHISs and these may appear technical in nature at first glance [12, 42]. Introducing new data forms, setting up warning systems to detect potential errors and developing algorithms for integrating different databases are possible technical solutions to such apparently technical challenges. However, there is a need to consider the influence of social factors as well, such as people's competencies in dealing with new data production processes, institutional values about data utilisation, and existing relationships between data producers and decision-makers [42, 44, 51]. DH interventions for RHISs may not be used effectively if data production and utilisation processes are viewed as merely technical. Given that RHISs are embedded in complex health systems and DH interventions are trying to improve RHIS

functions, they too are influenced by the broader setting in which dynamic and complex social and technical factors interact [1, 8, 43]. There is thus growing recognition that the development and implementation of DH interventions occurs in multiple technical and social contexts, where different types DH interventions and activities tackle different parts of the problem of fragmented RHISs. Data harmonisation may therefore have different definitions and intended outcomes. There are a wide range of terms being used for interventions with similar aims and activities [68]. For example, terms such as record linkage, data warehousing, data sharing and health information exchange are all used to describe data harmonisation-type activities [19, 69-71], and it is not always clear the extent to which these efforts are similar in practice, scope and relevance. The use of multiple terms may not be a problem in itself, but there is no common understanding of its components and processes and this lack of clarity on what constitutes ‘data harmonisation’ makes it difficult to compare and assess the usefulness of DH interventions in strengthening RHISs.

To take it a step further, even when DH interventions have been implemented for RHISs and RHISs are able to provide timely, relevant and accessible informational support, it is still unclear whether this has a positive effect on health management decision-making. Lack of understanding of the relationship between DH activities and health management decision-making makes it difficult to evaluate the impact and implications for health systems strengthening [44]. The scope of this review is to therefore understand the range of definitions for DH, alternative terms for DH and DH interventions, and to identify explanations of the relationship between DH and health management decision-making. Synthesising variations and communalities in terminology, identifying the key components of DH interventions and identifying explanations of how DH can support management decision-making can all enhance the potential usefulness of DH interventions for researchers and implementers.

## **2. Methods**

This scoping review was conducted according to the methods outlined by Arksey and O’Malley [72]. They recommend a process that is “not linear but, requiring researchers to engage with each stage in a reflexive way” to achieve both ‘in-depth and broad’ results. This review followed the standard steps for systematic reviews: identifying the research question, identifying relevant

studies, selecting studies for inclusion, data extraction and data synthesis. These are detailed in our published study protocol included as part of this PhD thesis [73].

## **2.1. Study objectives**

This scoping review aims to appraise the characteristics of studies on data harmonisation and the definitions used for data harmonisation activities and to develop an understanding of the intended effect of data harmonisation interventions on management decision-making. The objectives are:

1. To identify and synthesise the characteristics of studies of data harmonisation;
2. To identify and synthesise the various definitions and concepts used to describe data harmonisation interventions, and
3. To develop a conceptual understanding of explanations in the literature of the causal relationship between data harmonisation interventions and health management decision-making.

We took a stepped approach in addressing these objectives. All included studies were used to address Objective 1. To address Objective 2, while extracting the characteristics of all the included studies, we sampled studies that were using alternative terms for DH interventions and used those to identify, synthesise and compare similarities and differences in definitions. While executing Objective 1 and 2, we identified a smaller number of studies that contributed to Objective 3.

## **2.2. Identifying relevant studies**

### **2.2.1. Eligibility criteria**

Peer-reviewed studies and grey literature were considered eligible for inclusion into the scoping review if they provided a definition, description or concept (in the form of a model, framework or process) of a DH intervention. Additionally, studies were eligible if they provided an explanation of the causal relationship between DH and health management decision-making (such as through improved quality and accessibility of harmonised information for management and/or the utilisation of harmonised health information for management decision-making). Any studies

concerned with different technical activities of DH (such as linking, merging, cleaning and transferring) were considered.

### **2.2.2. Search strategy**

A systematic literature search was conducted in PubMed, CINAHL and Web of Science for eligible studies from 1 January 2000 to 30 September 2018. We limited our search to the year 2000 as this is around the time that work on HIT innovations (such as health information exchange) began in high-income countries (predominantly in the United States of America) [74] and when researchers and health system managers in LMICs became interested in the integration of large digital databases [12]. In **Appendix 1** we present the search strategies of the three databases. Based on preliminary searches we anticipated that these databases would yield the highest results. The search strategies include a combination of keywords and Medical Subject Headings (MeSH) terms related to data harmonisation (concept A) and health information system (concept B). There were no geographic restrictions, but we only searched for English studies.

## **2.3. Selecting studies for inclusion**

### **2.3.1. Screening records**

The first reviewer (BS) conducted all the searches with the help of a librarian and collated the records in the EndNote reference management programme where duplicates were removed. Two reviewers (BS and AH) then independently screened the records (titles and abstracts) to assess eligibility for full-text review. BS and AH resolved conflicts that emerged at this stage by talking through the inclusion criteria and arriving at a joint decision.

The full-texts of potentially eligible studies were retrieved and assessed by the two reviewers (BS and AH). Final inclusion into the review was based on whether at a minimum the study had a definition or description of a DH intervention or referred to its relationship with health management decision-making. The first reviewer read all full-texts and the second reviewer only read a sample (roughly a third) of the full-texts to verify the first reviewer's decision about

inclusion. BS and AH disagreed about the inclusion of four studies, but after discussing, they decided to exclude the studies.

After finalising screening, the two reviewers then mapped out the characteristics of included studies in an Excel spreadsheet. They recorded the name of the first author, the date, the type of study (primary, review, conceptual, commentary), the term used for the intervention they described (DH or alternative), the country in which the study was taking place, the level of the health care system (frontline, management, research), and ticked whether there was a conceptual model, framework, diagram or process description of DH and health management decision-making. This detailed mapping of study characteristics was useful for informing sampling options for Objectives 2 and 3.

### **2.3.2. Sampling of studies**

A scoping review aims to map the literature on a particular topic rather than to provide an exhaustive explanation of a particular phenomenon of interest [72, 75]. Thus, the number of included studies is expected to be high in scoping reviews. To manage the high numbers for a scoping review such as this one (where the aim was to provide definitions and concepts) it was necessary to make use of a qualitative sampling approach. A qualitative sampling approach for this review aimed for variation and depth rather than an exhaustive sample; because reviewing too large a number of studies can impair the quality of the analysis and synthesis [76]. We used two types of purposive sampling techniques called maximum variation sampling and theoretical sampling [77]. These techniques were used to identify both the range, variation and similarities or differences in definitions and concepts and intervention descriptions (as per Objective 2) and to provide a rich synthesis of explanations of causal relationships between DH and health management decision-making (as per Objective 3). For Objective 1, we did not apply a sampling strategy which means that all the included studies that at a minimum provided a definition or description of a DH intervention contributed data towards Objective 1.

## 2.4. Data extraction

The first reviewer (BS) extracted data for Objective 1 from all the included studies (n = 181). A second reviewer (AH) independently extracted data from 81 (45%) of included studies to verify data extraction done by the first author. We used an Excel spreadsheet for data extraction as presented in **Figure 1**. AH and BS extracted a few studies first before clarifying the items in the spreadsheet. Once data extraction was complete, the reviewers were easily able to filter according to the individual items extracted to synthesise and compare studies. Given the objectives of the scoping review, we did not extract any information relevant for conducting risk of bias or quality assessment. Not conducting risk of bias or quality assessment is consistent with scoping reviews of similar aims and methodological approaches [72, 75, 78].

**Figure 1.** Extract of the Excel data extraction form

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Authors	Study title	Date	Include, exclude, not sure	Status of fulltext	Definition or description (y/n)	DH term used in the study	Conceptual framework (e.g. process or diagram)	Purpose of HIE/Aim of study	Name of country	Level of health system (e.g. clinic, hospital, management)	Type of study (primary, commentary, conceptual)	Should we sample this study? (y/n)
1	Callan	Making health information exchange work	2014	include	found - online	y	HIE	no	analysis of current state of HIE organisations	USA	management		no
32	Callan	Tracking HIE's ever evolving operational models	2014	include	found - online	y	HIE	no	analysis of current state of HIE organisations	USA	management		no
33	Campion	Push and pull: physician usage of and satisfaction with health information exchange	2012	include	found	y	HIE	no	push and pull of HIE and user satisfaction	USA	hospital		no
34	Campion	Health information exchange system usage patterns in three communities: practice sites, users, patients, and data	2013	include	found	y	HIE		implementation of HIE across diverse care settings and patient populations	USA	communities		no
35	Campion	Patient encounters and care transitions in one community supported by automated query-based health information exchange	2013	include	found	y	HIE	no	extent to which HIE support patient encounters and care transitions	USA	clinics		no
36													

## 2.5. Data synthesis: collating, summarising and reporting findings

The first reviewer (BS) conducted data analysis using manual coding and the filter option in Excel. Another reviewer (NL) reviewed the data analysis work on an ongoing basis as an additional quality check. For Objective 1, we conducted a numerical analysis to provide an overview of the characteristics of all the included studies. For Objective 2, we conducted a qualitative analysis to provide a narrative synthesis of the different DH definitions and concepts, and to identify different

components or activities that are considered part of the DH processes. For Objective 3, we reviewed data related to intentions, suggestions and or explanations of how DH may lead to improved health management decision-making. We extracted and analysed data relevant to objective 2 and 3 at the same time. We first created a list of all the different terms used to describe DH interventions and then compared definitions across alternative terms by looking for similarities or differences in the definitions or descriptions of DH interventions. We then coded key components, processes and outcomes of DH interventions and the factors reported as important in the relationship between DH and health management decision-making.

The findings are structured according to three themes matching the three study objectives: an overview of the key characteristics of included studies, alternative terms and definitions of DH, and a narrative synthesis of the relationship between DH and health management decision-making.

## **2.6. Reflexivity**

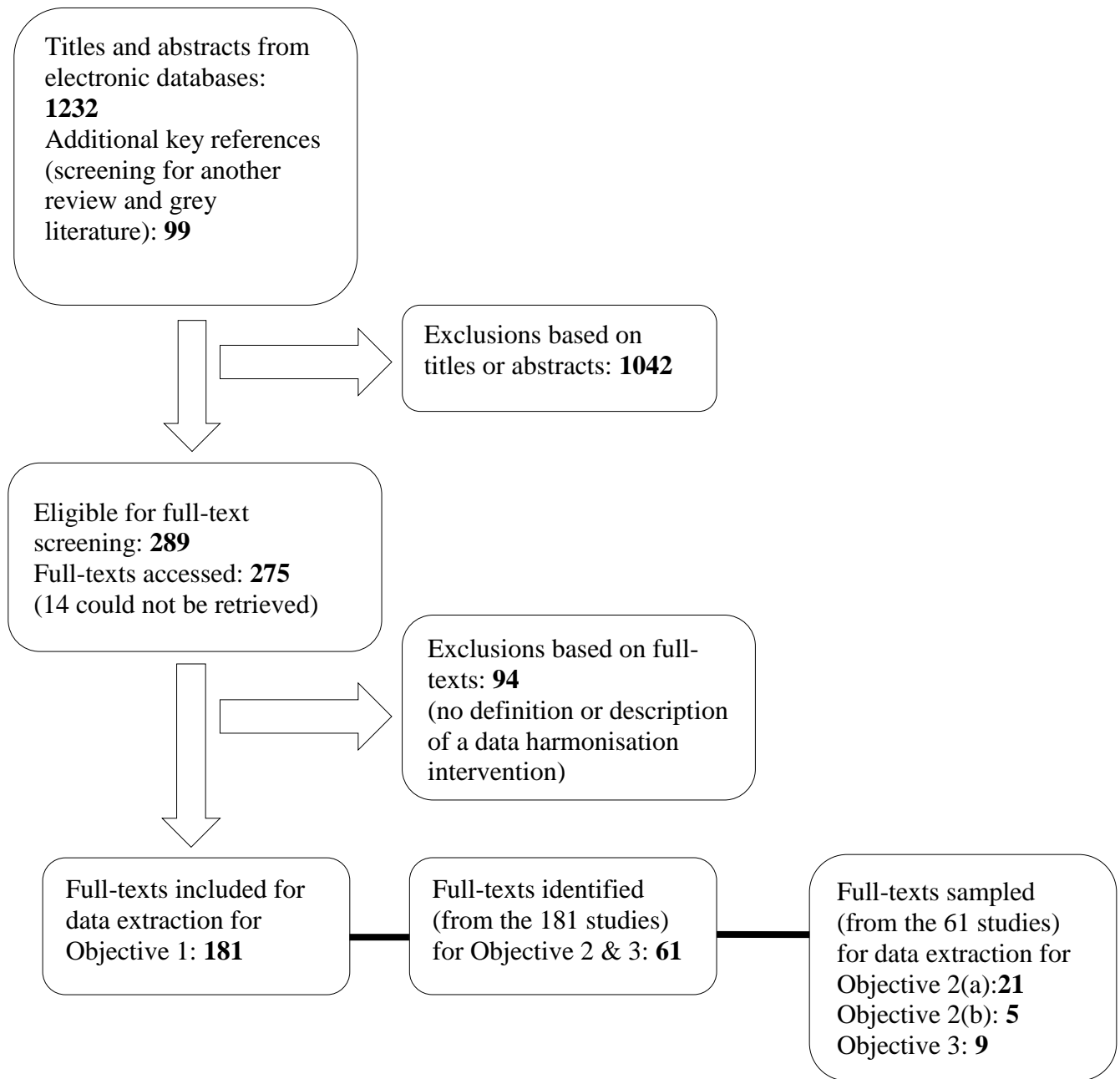
Throughout the review, the authors were aware of their own positions and reflected on how these could influence the study design, search strategy, inclusion decisions, data extraction, analysis, and synthesis, and interpretation of the findings [76]. The review authors are trained in anthropology, epidemiology and health systems research. The first author was involved in participant observation of an innovative DH project in the Western Cape Department of Health in South Africa as part of her doctoral research where she grappled with questions that informed the objectives of this review. Three of the authors (BS, AH and NL) were involved in a Cochrane systematic review on RHIS interventions when this scoping review was conceptualised, so they were familiar with some of the HIS literature and had some appreciation for the conceptual and methodological complexities of studying the field of health information management. This experience informed the way the first author developed the search strategy. She used an iterative approach to narrow down the search as much as possible because of her prior knowledge that it was difficult to balance sensitivity and specificity when developing a search strategy for HIS literature that is often multi-disciplinary in nature.

## 3. Findings

### 3.1. Results of the search

**Figure 2** shows a PRISMA diagram of the search results. We screened a total of 1331 (1232 and 99) titles and abstracts identified from searching three electronic databases, grey literature and while screening for a Cochrane systematic review assessing the effectiveness of RHIS interventions on health systems management [79]. Almost a quarter (289 of 1331) were deemed potentially eligible for full-text screening. We managed to access full-texts for 275 studies and of those 181 were included into the scoping review for Objective 1. We excluded 94 full-text articles because they did not meet the minimum criteria; that is, provide a definition or description of a DH intervention or activity. We sampled 61 studies from the 181 for Objective 2 and 3. We arrived at 61 studies by including all reviews (systematic or literature reviews) and all studies (irrespective of the type of study) that also had a process description, conceptual framework or theory of a DH intervention (in addition to the minimum criteria for Objective 1).

**Figure 2.** PRISMA diagram of eligible studies



### **3.2. An overview of key characteristics of data harmonisation studies**

A total of 181 studies which are presented in **Table 1** were included into this scoping review for Objective 1. Given the high number of included studies, we decided to only map the following key characteristics of those studies: first author, date, type of study, intervention term (DH or alternative), country and level of the health care system. The majority of included studies (126 of 181) are primary studies assessing various aspects of developing and implementing DH interventions (quantitative studies n=86) or patient, providers or stakeholders' perspectives (qualitative studies n = 34) or a combination of both (mixed methods studies n = 6). We found a total of 42 study protocols, conceptual papers and commentaries and 13 reviews (systematic and literature reviews).

Of the 181 included studies, 9 were not country specific (these were global reviews), 151 were from the USA and the rest were from other countries (specifically Australia, Brazil, Canada, China, Finland, Germany, Israel, Japan, Jordan, Korea, Malaysia, Netherlands, South Africa and South Korea). In terms of the level of the health care system, 128 studies were on a DH intervention or activity that was concerned with the frontline level (health service providers), 48 studies were concerned with health system factors or policy-related activities at the managerial level, and only 5 studies focused on DH interventions specifically for research purposes. The majority of studies (92%) used the term health information exchange (HIE), while the remaining studies (8%) used a variety of terms to describe various DH interventions and activities, specifically, record linkage, data mining, data linkage, data warehousing, data sharing and data harmonisation.

### **3.3. Definitions, components and processes of data harmonisation**

We first discuss the alternative terms and definitions of DH and then we summarise key components and processes of DH using studies sampled from the 61 ('rich') studies identified for Objective 2 and 3. **Table 2** presents the 61 studies; that is, the type of study design, the intervention terms, the country, the level of the health care system and the purpose of the study. These studies were concerned with the challenges and opportunities of DH, the barriers and facilitators of DH,

the various factors affecting DH (such as technical and financial factors), the outcomes of DH (such as patient safety and quality of care), and privacy and security issues of patient information.

### **3.3.1. Alternative terms and definitions of data harmonisation**

To describe alternative terms and definitions of DH (Objective 2(a)), we sampled 21 studies from the from the 61 studies identified for Objective 2 and 3. The alternative terms and definitions are summarised in **Table 3**. During the process of sampling, we realised that the majority of studies (53 of 61) used the term ‘health information exchange’. We therefore only used 13 of the 53 studies that used the term HIE to compile the definition for it, presented in the table. These 13 studies were chosen to represent the term HIE because they were reviews and we assumed that reviews provided synthesised definitions of interventions. In addition to the 13 studies, 8 other studies were sampled (hence 21 studies in total) because they provided different terms to describe DH interventions or activities besides HIE.

When looking across the table, there is overlap between some of the terms and definitions. For example, data linkage and record linkage both focus on ‘linkage’ as a core activity in combining different databases using a unique patient identifier. HIE is described as a key outcome of data interoperability, while data sharing is described as a key outcome of HIE. Definitions for data harmonisation, record linkage and data warehousing explicitly state that these interventions involve a process of having to integrate different or ‘homogeneous’ databases or information systems. Below, I report on the various definitions of DH interventions and activities.

Data harmonisation is considered a multi-step process with a range of activities (such as identifying, reviewing, matching, redefining and standardising information). Data harmonisation interventions rely on interoperability between databases and systems which means copying standardised patient-level data into a separate repository. Data linkage and record linkage are a type activity of a broader intervention (data harmonisation) that focus on mechanisms (such as unique patient identifiers) for integrating large datasets. Data warehousing is concerned with extracting, transforming and loading large datasets using information technology (IT) platforms, application systems and data marts. Data sharing and HIE are two outcomes of DH interventions. They are about accessing data found across different platforms (such as clinical and financial

systems) and being able to share patient outcomes across a trajectory. These various definitions indicate that the aim of DH is to improve patient outcomes, coordination of health services, quality of care and efficiency and facilitate public health interventions.

In reviewing the definitions, we were able to identify nine characteristics of DH presented below. No single study included all of these characteristics and there are no specific factors such as study design, country or level of the health care system associated with the definitions.

- Any type of DH intervention or activity is a *process of multiple steps* involving both technical and social processes.
- There are at least *two or more databases* involved in any DH intervention or activity.
- A data harmonisation intervention or activity involves *electronic data* (no reference is made to data found in paper-based sources).
- The goal of a DH intervention or activity is to *integrate, harmonise and bring together* different electronic databases into useable formats.
- Data harmonisation occurs when there is an increasing availability of electronic data that can be pooled together using *unique patient identifiers*.
- *Different types of data* can be linked and shared such as individual patient clinical, pharmacy and laboratory data, health care utilisation and cost data, and personnel-related data.
- Electronic data required for DH processes can be found within and across *different departments and institutions at facility, district, regional and national levels*.
- A data harmonisation process consists of *different types of technical activities* such as identifying, reviewing, matching, defining, redefining, standardising, merging, linking, merging and formatting data.
- DH interventions or activities are defined according to a *specific scope and purpose* such as disease surveillance, monitoring of long-term outcomes, screening for adverse events, geographic area, secondary data use and data marts or dashboards.

**Table 3. Alternative terms and definitions of data harmonisation interventions**

Where multiple studies used a similar definition, the review authors synthesised the data from similar definition into the composite definition for each term, as presented in this table.

Liu, 2010 [80]	<p><b>Data harmonisation</b> is the process of integrating life-long health data of a person that are distributed in inhomogeneous information systems through identifying, reviewing, matching, redefining and standardising information. This process involves two steps. Firstly, identifying whether all the information necessary for a single electronic platform is available in existing systems, where the information is, and how the information is defined and formatted. And secondly, to make the heterogeneous information recorded by various systems consistent or at least comparable with one another by reviewing, matching, redefining and standardising each data item.</p>
Boyd, 2014 [69]	<p><b>Record linkage</b> is the process of bringing together data relating to the same individual from within and between different datasets. When a unique person-based identifier exists, linkage can be achieved by simply merging datasets on the identifier. However, when a person-based identifier does not exist, then some other form of data matching or record linkage is required for integrating data.</p>
Gill, 2001 [81] Hopf, 2014 [21]	<p><b>Data linkage</b> can be used to construct a register for a specific geographic area and disease (for example, a district non-communicable disease register). Linkage of routine datasets by unique patient identifiers can provide an opportunity for identifying adverse drug reactions and tracking exposed individuals in real time. Routine data linkage can also enable the creation of exposure cohorts to monitor long-term outcomes and enable a more efficient screening for adverse drug reactions due to an ever-increasing data pool.</p>
Haarbrandt, 2016 [82]	<p><b>Data warehousing</b> is the process of establishing specialised databases by integrating information systems (the authors specifically referred to hospital information systems) to facilitate secondary use of data. Clinical data warehouses are generally built on one of two predominant architectural paradigms: either, data is directly extracted, transformed and loaded from applications systems and databases into a data mart (an integrated view over a defined subject), or it is stored in a centralised data repository from which data marts can be established. Both approaches rely on a process to extract data from sources, transform it appropriately and to load it (or copy it) to a specific database.</p>

<p>Hu, 2007 [71] Jones, 2012 [83]</p>	<p><b>Data sharing</b> is based on the need for a more robust method for defining and sharing expected and actual patient outcomes. It must leverage existing informatics tools since a great deal of patient-specific information is already available in medical record systems and billing and administrative systems. One type of data sharing system is an infectious disease informatics (IDI) system. An IDI system should encompass sophisticated algorithms for the automatic detection of emerging disease patterns and the identification of probable threats or events. It should also have advanced computational models that overlay health data for spatial–temporal analysis to support public health professionals’ analysis tasks.</p>
<p>Elysee 2017 [84]</p>	<p><b>Data interoperability</b> is one of two functionalities of an advanced electronic health record. The first function is health information exchange, which is the ability to electronically share patient-level information among unaffiliated providers across organisational boundaries. The second function is interoperability, which is the ability to produce standardised patient-level health information that can be integrated into unaffiliated health care providers’ electronic health records.</p>
<p>Akhlaq, 2016 [19] Dixon, 2010 [85] Esmailzadeh, 2016 [86] Esmailzadeh, 2017 [87] Fontaine, 2010 [88] Hopf, 2014 [21] Kash, 2017 [89] Mastebroek, 2014 [90] Parker 2016 [91] Rahurkar, 2015 [92] Rudin, 2014 [93] Sadoughi, 2018 [94] Vest, 2012 [95]</p>	<p><b>Health information exchange (HIE)</b> is a type of health information technology (HIT) intervention. It involves the electronic mobilisation of clinical and administrative data or information within or across data repositories or organisations in a community or region, between various systems as per recognised standards. This is to ensure that the HIE maintains the authenticity and accuracy of the information being exchanged, thereby enabling stakeholders to make informed decisions to enhance healthcare quality and delivery of patients and populations. Sharing clinical data can potentially improve patient safety, care coordination, quality of care and efficiency, facilitate public health efforts and reduce mortality and healthcare costs. Lastly, HIE involves multi-stakeholder organisations that oversee the business, operational and legal issues involved in the exchange of information.</p>

### 3.3.2. Components and processes of data harmonisation

To synthesise key components and processes of DH interventions (Objective 2(b)) we sampled 5 studies from the 61 studies identified for Objective 2 and 3. Three of the 5 studies (Hu 2007, Boyd 2014 and Elysee 2017) overlap with the 21 studies sampled to describe the alternative terms and definitions of DH interventions and activities above. The 5 studies, presented in Table 4, were

selected from the 61 studies because they provided more detail about the intervention and had a visual representation of DH components or processes.

The first study listed in the **Table 4** (Ji 2017), provides a diagrammatic illustration of what appears to be a fairly comprehensive conceptual model of a DH intervention, illustrating different types of data, different levels of the health care system (e.g. clinics and hospitals), the multiple processes of exchanging data, the multiple directions in exchange of data and the key role of the unique patient identifier in enabling the DH process [96]. In the next model, Boyd et al. [69] and Santos et al. [97] both lay out the technical processes involved in the linkage process of different databases, but Santos et al. specifically focuses on linking data required for individual patient clinical management into a central repository. Lastly, Elysee et al. [84] and Hu et al. [71] both describe a DH intervention for a specific purpose, that is medication reconciliation and disease outbreak surveillance respectively.

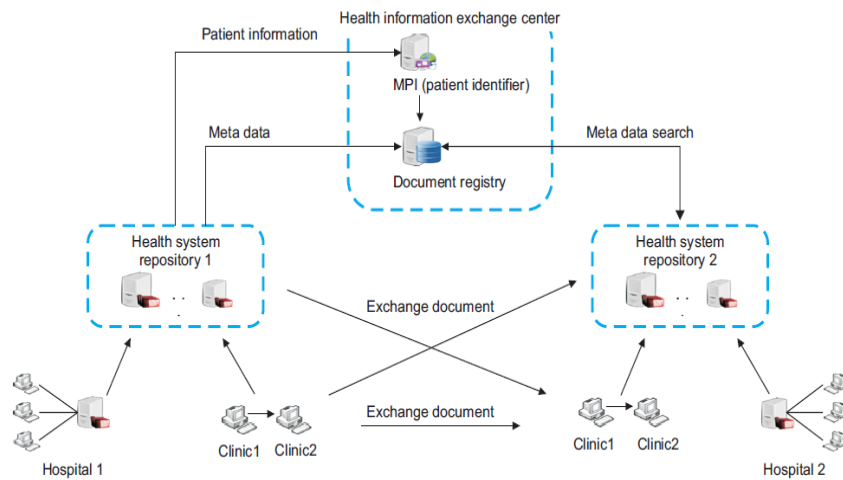
These conceptual models of DH interventions and activities highlight that there are various steps involved in the integration of databases and in the transformation of data into useable formats. Integrating databases means bringing together data of the same individual from within and between different electronic databases, through various activities involving identifying, reviewing, matching, redefining and standardising data [69, 80]. Once data is harmonised, it can be categorised by various criteria of interest, such as geographic area or disease or patient population, and transformed into different formats such as graphs, tables or dashboards to make it easier for users to access and use the information [82]. There may be different ways that the data is harmonised; in some studies DH is described as a linear and one-directional process, while other studies described it as an iterative and multi-directional process.

**Table 4. Concepts of data harmonisation interventions and processes**

The table presents the different conceptual models of data harmonisation and the review authors provide a summary of how key components and processes were described by the authors of these models.

**A comprehensive data harmonisation process**

Ji (2017) [96] present a health information exchange (HIE) architecture for data exchange between hospitals and clinics. The HIE center consists of a master patient identifier (MPI) server and a document registry which is interchangeable between all institutions. Such a HIE center is only possible when there are already various data repositories in existence because its aim is to manage metadata, large groups of patients and multiple health organisations.



**The multiple steps of a technical linkage process**

Boyd (2014) [69] identify four steps in the linkage process, that is, file verification, data cleaning, linkage and grouping. File verification is to check that all data items needed for record linkage are correctly organised in a file. Data cleaning involves a predetermined cleaning strategy so that data items are standardised. The process for linking data items usually involves complicated linkage maps where linkage needs and linkage scenarios are identified. And lastly, once satisfied with the linkage, the linkage outputs can be grouped into categories that are user-friendly.

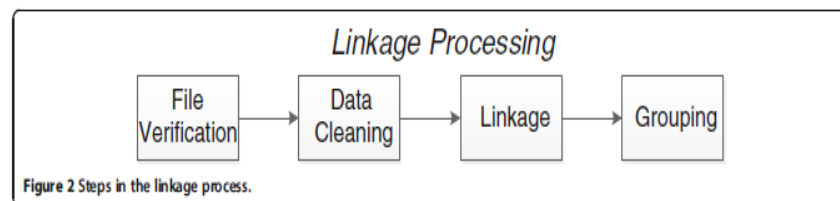
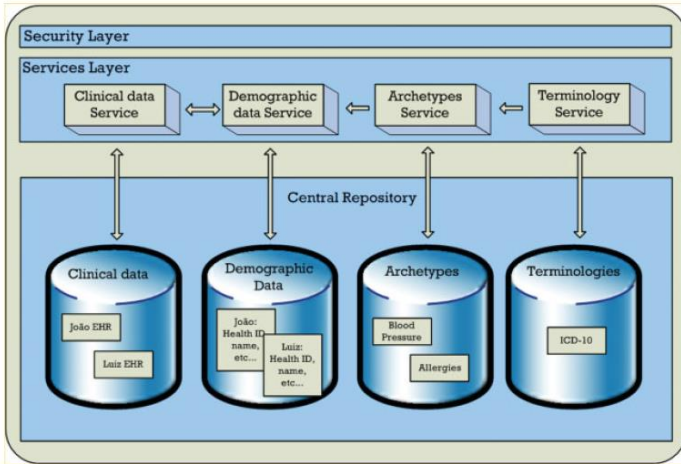


Figure 2 Steps in the linkage process.

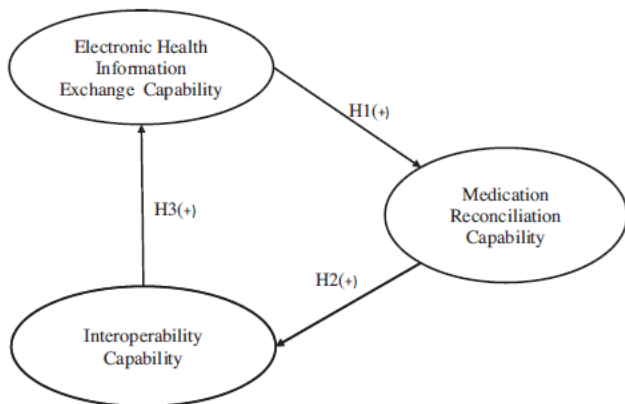
**Different types of data involved in a technical linkage process**

Santos (2017) [97] provide a proof of concept for integrating clinical and demographic data, archetypes and terminologies related to maternal and neonatal data into a central repository. Clinical and demographic data can be sent and retrieved from the central repository. Archetypes represent clinical concepts; they make it possible to establish rules for data sharing and define vocabulary and terminologies.



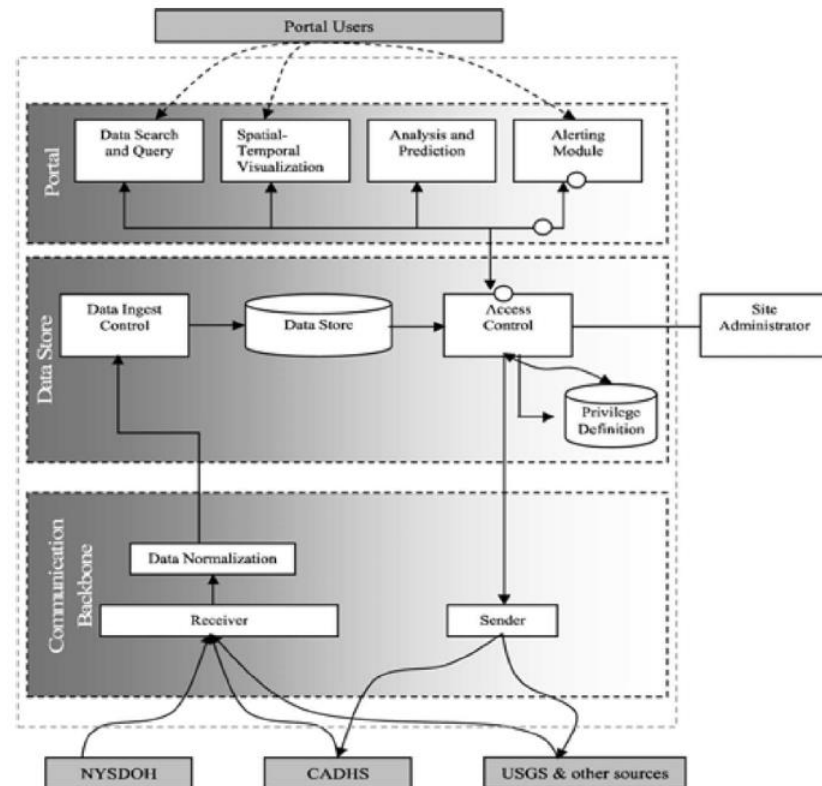
**A data harmonisation intervention specifically focused on medication reconciliation, health information exchange and interoperability**

Eylsee (2017) [84] describe the importance of using health information exchanges (HIEs) in resolving medication administration discrepancies and improving patient safety. They state that there is a positive relationship between medication reconciliation, HIE and interoperability. The positive relationship is based on the notion that as more electronic data becomes available, clinicians will not be able to process high volumes of data from different places on their own. Hospitals will therefore seek to increase interoperability between electronic databases of different hospitals. This will enable hospitals to implement the use of electronic medical reconciliation which will improve timeliness, accuracy and completeness of information sharing.



### A data harmonisation intervention specifically focused on disease outbreak surveillance

Hu (2007) [71] outline the different functions of an infectious disease HIE. The functions include infectious disease data search and query, spatial-temporal visualisation, outbreak detection and analysis and automatic alert generation. The authors state that the challenge and complexity of designing such a HIE extends beyond heterogenous databases. There are multiple layers that should be considered in the design of a HIE, such as the communication backbone, the data store, the portal and portal users.



### **3.4. The relationship between data harmonisation and health management decision-making**

We sampled 9 studies from the 61 studies (identified for Objective 2 and 3) that provide an explanation of the relationship between DH and health management decision-making. We present extracts of explanations of the relationship in **Table 5**. These 9 studies were the only ones out of the 61 studies that referred to the relationship between DH and health management decision-making. Of the 9, the study by Eylsee et al. (listed forth in the table) provided the most detail; the study states that there is a positive relationship between increased electronic data, clinicians being

unable to deal with high volumes of data and the need for interoperability between electronic databases at different hospitals. The study also states that interoperability between electronic databases should improve timeliness, accuracy and completeness of information sharing.

The other studies (Ji, Boyd, Santos and Hu) did not explicitly explain the relationship between DH interventions and health management decision-making, but they broadly stated the benefits and concerns of DH interventions to health management decision-making, including clinical decision-making [69, 71, 96, 97]. For example, most of the studies stated that DH interventions make it possible for health providers to use data over time and across organisations to support clinical management decision-making. But, some studies stated that DH interventions were sometimes unable to deal with incompleteness and poor quality of data and inconsistencies in definitions and codes of data items.

From the 9 studies, we identified three types of health management decision-making that DH contributed to. These are as follows:

- Clinical decision-making for individual patient clinical management or clinical support and quality improvement tools
- Operational and strategic decision-making for health system managers and policy-makers
- Population-level decision-making for disease surveillance and outbreak management

The first level involves frontline clinicians being able to access their patients' medical information and treatment data and timelines (datasets of longitudinal clinically relevant individual-level data) through DH interventions. In these situations, DH can make it easier for frontline clinicians to develop tools for reminding them about patients' performance in treatment and care services as well as help them improve the quality of health care services. At the operational and strategic decision-making level, DH interventions have the potential to support high-level health managers make decisions involving a wide network of stakeholders such as consumers, patients and professionals. Lastly, disease surveillance and outbreak management decision-making rely on harmonised data to plan, monitor and evaluate population-level interventions.

**Table 5.** The relationship between DH interventions and health management decision-making  
*The review authors directly quoted text from the primary studies where a description of the link between data harmonisation and health management decision-making was provided.*

Cimino, 2014 [98]	“Data completeness: A promise of HIEs is to use consolidated information over time and across providers to improve <b>medical decision-making for the patient</b> . When presenting a medical timeline for a patient, how does a provider know whether the HIE presentation of history is missing information? The consequences to patients can be devastating.”
Downs, 2010 [99]	“...community-based approach to establish a common pathway based on common data standards to facilitate the incorporation of interoperable, clinically useful genetic or genomic information and analytical tools into EHRs to <b>support clinical decision-making for the clinician and consumer</b> .”
Grossman, 2008 [100]	“...the exchanges going beyond core clinical data exchange activities that give physicians access to data at the point of care to offering <b>physicians clinical decision support, reminders and other quality improvement tools aimed at individual patients</b> .”
Kuperman, 2013 [101]	“Ideally, a physician would have access to complete, accurate and timely patient data to <b>support optimal decision making</b> . Health information exchange capabilities will reduce the extent of data fragmentation but will not eliminate it entirely.”
Politi, 2014 [102]	“In this scenario, an HIE system is likely to have a significant impact on <b>clinical decision making</b> if information is readily accessible; the need for rapid decisions might render the scrutiny of an HIE system impractical.”
Vest, 2010 [103]	“The anticipated benefits of more data to inform <b>physician decision making</b> , sparing patients of needless tests, helping organization identify inappropriately managed patients, and improving the health of the public will only be achieved by HIE that does not exclude providers in an area, limit what data elements are available, or restrict exchange to specific subpopulations.”

Shapiro, 2006 [104]	“The goal of a nationwide health information network would be to deliver information to individuals– consumers, patients, and professionals –when and where they need it, so they can use this information to <b>make informed decisions about health and health care.</b> ”
Vest, 2015 [105]	“Improved access to more comprehensive information may <b>support decision-making</b> , inform providers of additional medications or allergies, and help avoid repeated or duplicate testing.”
Zaiden, 2015 [106]	“Combined with data mining and statistical analysis tools, these repositories of health information can greatly advance medical knowledge, healthcare quality, and <b>good strategic management.</b> ”

## 4. Discussion

### 4.1. Synthesis of findings

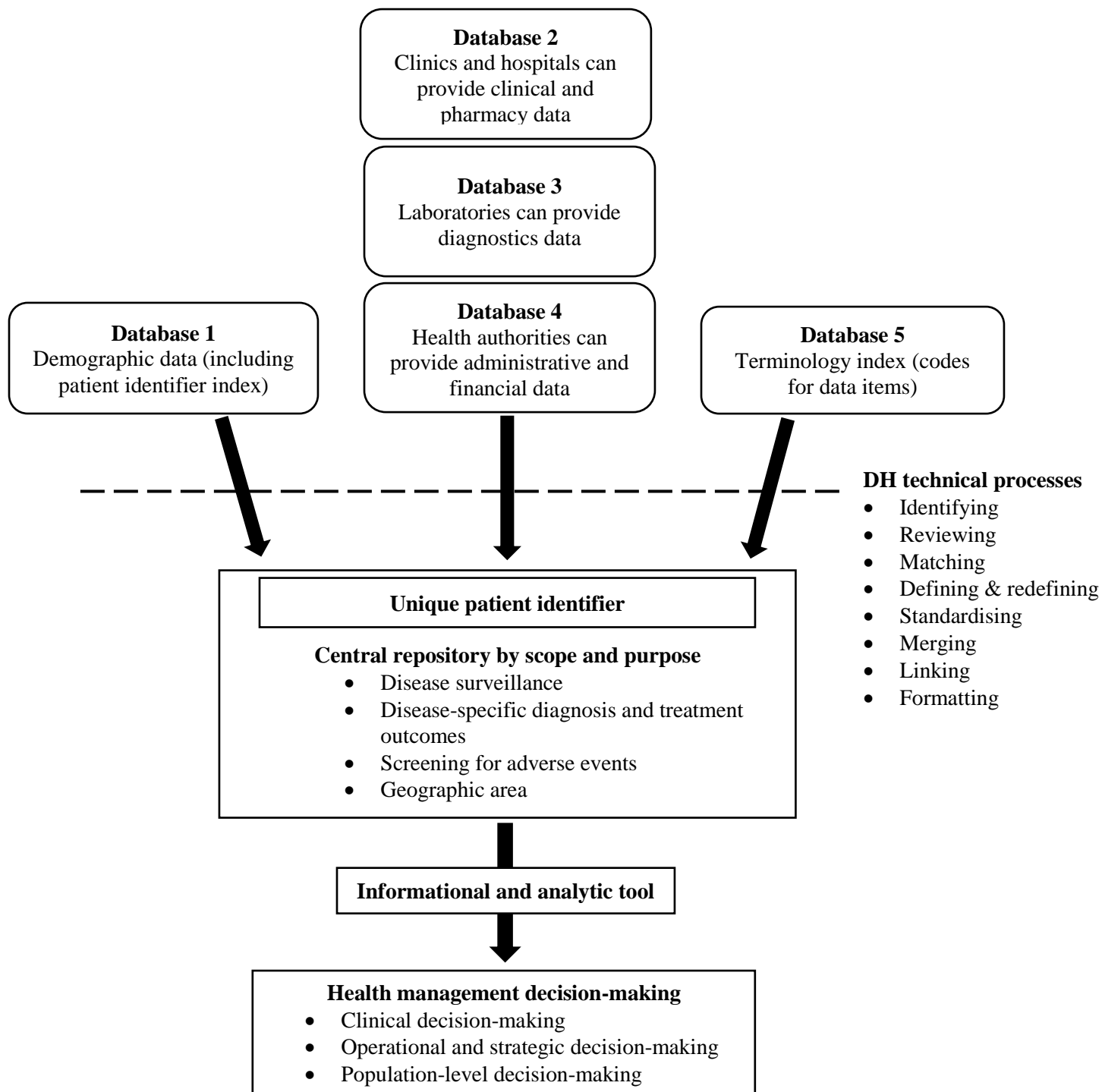
This review aimed to provide an overview of the key characteristics of DH studies, provide alternative terms, definitions, components and processes of DH interventions and provide explanations of the relationship between DH and health management decision-making in the literature. Of the 181 studies that at a minimum provided a definition or description of a DH intervention or activity, 86 were primary quantitative studies, 151 were studies conducted in the United States of America (USA), 128 were aimed at improving frontline level health services, and 164 used the term ‘health information exchange’ for the intervention or activity. The majority of the studies that used the term HIE were conducted in the USA. In addition to the term data harmonisation (and HIE), we identified five other terms; those are record linkage, data linkage, data warehousing, data sharing, and data interoperability. Terms about linkage and sharing appear to be describing activities of DH interventions, while terms like data harmonisation, data warehousing and HIE seem to describe a more comprehensive approach to DH interventions involving both data production and data utilisation aspects. The term data interoperability refers to the intended function of a DH intervention (to enable to exchange data) but is focused on technical processes that occur between databases.

There were no specific factors (such as study type, country or level of the health care system) associated with studies that used the alternative terms of DH expect for HIE. Even though different studies used different terms, there was consensus amongst them that DH should be a useful intervention or activity to improve patient and health system outcomes.

We identified nine characteristics of DH interventions and activities. Using these nine characteristics, DH can be summarised as a process that aims to integrate two or more electronic databases, it involves different types of data captured within and across various institutions at different health care system levels, and varying activities are required to pool together data using unique patient identifiers for the purpose of providing information support for health management decision-making. The review identified three types of health management decision-making that DH contributed to: (a) clinical decision-making for individual patient management, clinical support and quality improvement tools; (b) operational and strategic decision-making for health system managers and policy-makers; and (c) population-level decision-making for disease surveillance and outbreak management.

Drawing on the definitions and the conceptual models of DH identified in this review, we developed a concept map (see **Figure 3**) which presents one possibility for explaining how different aspects of DH interventions and activities work together to support health management decision-making. The concept map consists of different types of databases (1 to 5) containing different types of data such as demographic, clinical, pharmacy, laboratory, administrative and financial, and terminology data. A technical process involving different types of activities (such as matching, merging and linking) takes place to integrate the different types of data using a unique patient identifier. The central repository, where the data is harmonised, is defined according to specific criteria such as a geographic area or disease outcomes. The data kept in the repository should be accessible to data users, who can then use this harmonised data as an information and analytic tool to support health management decision-making for clinical, operational, strategic, and or population-level decision-making.

**Figure 3.** A concept map of data harmonisation and its relationship to health management decision-making



## **4.2. Study limitations**

There are two main differences between the published protocol and this scoping review. Firstly, in the protocol we listed the Global Health database as one of the electronic databases that we would search, but this was not feasible as we realised late that none of the reviewers had permissions to access that database and gaining access was not affordable. We did however manage to search at least three electronic databases, as is the convention in reviews [115]. Secondly, in the protocol we stated that we would conduct full-text screening in duplicate, but due to the large volume of studies included for full-text screening, this was not feasible. The first reviewer (BS) assessed all full-texts and then the second reviewer (AH) verified the decisions of the first reviewer in a third of the included studies, which allowed for additional quality checks.

There are two main limitations of the review. Firstly, we restricted our literature search to English. We did not have the resources required for reviewing non-English studies. The majority of studies identified were from the English-speaking parts of the USA, but it is possible that studies from other non-English high-income countries, with extensive electronic health systems (such as France) may have been missed. Secondly, there is a possibility that because of sampling, we may have missed relevant studies for Objectives 2 and 3, though applying sampling aimed to identify variety, comprehensiveness and meaningfulness of the definitions and explanations.

## **4.3. Implications for research and practice**

There is a need to understand what DH interventions and activities are comprised of in diverse settings and contexts, especially in low- and middle-income countries. There were fewer studies from LMICs, which may be due to a lower prevalence of electronic health information systems in those settings. Nevertheless, DH interventions hold promise for improving the informational support in LMICs also, and more studies in this context could usefully expand the evidence base. The review highlights the importance of providing detailed descriptions of DH interventions, to allow for better comparisons and to improve the transferability of study results. Additionally, many resources are spent on the technical development of DH projects, with the implicit assumption that this will provide the informational and analytic support for health management decision-making, but this assumption is seldom tested in the research. There is a need for qualitative research on the

health system factors of implementing DH and for formative work to inform design of DH interventions. Finally, primary research and evidence synthesis of the experiences of key stakeholders involved (implementers and users of harmonised data) would improve our understanding of the causal mechanisms between data harmonisation and health systems strengthening.

#### **4.4. Conclusion**

The review aimed to widen our understanding about what DH interventions entail and how they can contribute to health management decision-making. The review revealed that most studies of DH interventions and activities were conducted in high-income settings and used the term ‘health information exchange’. We also identified nine characteristics of DH related to the process, technical activities, types of data, mechanisms for integrating data and scope and purpose of the interventions. The review showed that DH interventions contributed to three types of health management decision-making, that is, clinical, operational and strategic, and population-level surveillance decision-making. We provided a concept map of the components of DH and make recommendations for future research.

#### **Contributions**

The PhD student (BS) conducted all the searches and with the help of AH (co-author) screened records and conducted data extraction. BS analysed the data with help from NL. NL, CC and AH provided critical feedback on the drafts.

**Table 1.** Characteristics of included studies (n=181)

<b>Study name</b>	<b>Date</b>	<b>Type of study</b>	<b>Intervention term</b>	<b>Country</b>	<b>Level of the health care system</b>
<b>Commentary</b>					
Burris	2017	Commentary	HIE	USA	Frontline: hospitals
Figge	2010	Commentary	HIE	USA	Management
McIlwain	2009	Commentary	HIE	USA	Management
Murphy	2010	Commentary	HIE	USA	Management
Overhage	2007	Commentary	HIE	USA	Management
Rudin	2010	Commentary	HIE	USA	Frontline: workers
<b>Conceptual</b>					
Boyd	2014	Conceptual	RL	Australia	Research
Carr	2013	Conceptual	HIE	USA	Frontline: hospitals
Cimino	2014	Conceptual	HIE	USA	Management
Deas	2012	Conceptual	HIE	USA	Management
Del Fiol	2015	Conceptual	HIE	USA	Frontline: prisons, hospitals
Dimitropoulos	2009	Conceptual	HIE	USA	Management
Downs	2010	Conceptual	HIE	USA	Management
Feldman	2017	Conceptual	HIE	USA	Management
Frisse	2010	Conceptual	HIE	USA	Frontline: patients, workers
Frisse	2008	Conceptual	HIE	USA	Frontline: organisations
Frohlich	2007	Conceptual	HIE	USA	Management
Godlove	2015	Conceptual	HIE	USA	Frontline: patients
Greene	2016	Conceptual	HIE	USA	Management
Grossman	2008	Conceptual	HIE	USA	Management
Haarbrandt	2016	Conceptual	DW	USA	Management
Hu	2007	Conceptual	DS	USA	Management
Jones	2012	Conceptual	DS	USA	Management
Kuperman	2013	Conceptual	HIE	USA	Management
Langabeer	2016	Conceptual	HIE	USA	Management

Liu	2011	Conceptual	HIE	China	Management
McDonald	2009	Conceptual	HIE	USA	Management
McMurray	2015	Conceptual	HIE	USA	Management
Miller	2014	Conceptual	HIE	USA	Frontline: hospitals
Nelson	2016	Conceptual	HIE	USA	Frontline: prisons, hospitals
Politi	2014	Conceptual	HIE	n/a	Management
Ranade-Kharkar	2014	Conceptual	HIE	USA	Management
Shapiro	2016	Conceptual	HIE	USA	Frontline: workers, organisations
Shapiro	2006	Conceptual	HIE	USA	Management
Thorn	2013	Conceptual	HIE	USA	Frontline: health care workers
Thorn	2014	Conceptual	HIE	USA	Frontline: health care workers
Vest	2010	Conceptual	HIE	USA	Management
Williams	2012	Conceptual	HIE	USA	Management
Yaraghi	2014	Conceptual	HIE	USA	Management
Zafar	2007	Conceptual	HIE	USA	Management
Zaidan	2015	Conceptual	HIE	Malaysia	Management

**Primary studies**

Abramson	2012	Primary, quantitative	EHR, HIE	USA	Frontline, hospitals
Adjerid	2011	Primary, quantitative	HIE	USA	Management, states
Adler-Milstein	2011	Primary, quantitative	HIE	USA	Frontline: organisations
Adler-Milstein	2013	Primary, quantitative	HIE	USA	Management, organisations
Adler-Milstein	2016	Primary, quantitative	HIE	USA	Management
Alexander	2016	Primary, qualitative	HIE	USA	Frontline, health care workers
Alexander	2015	Primary, qualitative	HIE	USA	Frontline, health care workers
Ancker	2012	Primary, quantitative	HIE	USA	Frontline, consumers
Bahous	2016	Primary, quantitative	HIE	Israel	Frontline, hospital
Bailey	2013	Primary, quantitative	HIE	USA	Frontline: hospital
Ben-Assuli	2013	Primary, quantitative	HIE	USA	Frontline: hospitals
Boockvar	2017	Primary, quantitative	HIE	USA	Frontline: hospital
Butler	2014	Primary, qualitative	HIE	USA	Frontline: prisons, communities
Campion	2012	Primary, quantitative	HIE	USA	Frontline: health care workers
Campion	2013	Primary, quantitative	HIE	USA	Frontline: communities

Campion	2013	Primary, quantitative	HIE	USA	Frontline: clinics, hospitals
Campion	2014	Primary, quantitative	DE	USA	Frontline: organisations
Carr	2014	Primary, quantitative	HIE	USA	Frontline: hospitals
Carr	2016	Primary, quantitative	HIE	USA	Frontline: hospitals
Cochran	2015	Primary, qualitative	HIE	USA	Frontline: clinics, communities
Cross	2016	Primary, qualitative	HIE	USA	Management, organisations
Dalan	2010	Primary, qualitative	DM	USA	Research
Dimitropoulos	2011	Primary, quantitative	HIE	USA	Frontline: consumers
Dixon	2013	Primary, quantitative	HIE	USA	Frontline: hospitals
Dixon	2011	Primary, quantitative	HIE	USA	Frontline: laboratories
Downing	2017	Primary, quantitative	HIE	USA	Management: policy
Dullabh	2013	Primary, qualitative	HIE	USA	Management: organisations
Elysee	2017	Primary, quantitative	HIE, IO	USA	Frontline: hospitals
Foldy	2007	Primary, quantitative	HIE	USA	Management: organisations
Fontaine	2010	Primary, qualitative	HIE	USA	Frontline: primary health care
French	2016	Primary, quantitative	HIE	USA	Management: organisations
Fricton	2008	Primary, quantitative	HIE	USA	Frontline: patients, workers
Frisse	2012	Primary, quantitative	HIE	USA	Frontline: organisations
Furukawa	2013	Primary, quantitative	HIE	USA	Frontline: hospitals
Furukawa	2014	Primary, quantitative	HIE	USA	Frontline: health care workers
Gadd	2011	Primary, quantitative	HIE	USA	Frontline: health care workers
Garg	2014	Primary, quantitative	HIE	USA	Frontline: hospitals
Gill	2001	Primary, quantitative	DL	South Africa	Frontline: patients, disease
Grinspan	2013	Primary, quantitative	HIE	USA	Frontline: patients
Grinspan	2014	Primary, quantitative	HIE	USA	Frontline: health care workers
Grinspan	2015	Primary, quantitative	HIE	USA	Frontline: patients
Hassol	2014	Primary, quantitative	HIE	USA	Frontline: health care workers
Herwehe	2012	Primary, quantitative	HIE	USA	Frontline: health care workers
Hincapie	2011	Primary, qualitative	HIE	USA	Frontline: health care workers
Holman	2008	Primary, quantitative	DL	USA	Frontline: organisations, research
Hypponen	2014	Primary, quantitative	HIE	Finland	Frontline: health care workers
Ji	2017	Primary, quantitative	HIE	Korea	Frontline: hospitals

Johnson	2011	Primary, mixed	HIE	USA	Frontline: hospitals
Jung	2015	Primary, quantitative	HIE	USA	Frontline: hospitals
Kaelber	2013	Primary, quantitative	HIE	USA	Frontline: hospitals
Kierkegaard	2014	Primary, qualitative	HIE	USA	Frontline: health care workers
Kierkegaard	2014	Primary, qualitative	HIE	USA	Management
Kim	2012	Primary, qualitative	HIE	Korea	Management
Knaup	2006	Primary, quantitative	DS	Germany	Frontline: hospitals
Kralewski	2012	Primary, qualitative	CIE	USA	Frontline: organisations, workers
Laborde	2011	Primary, quantitative	HIE	USA	Frontline: hospitals
Lee	2012	Primary, quantitative	HIE	South Korea	Frontline: health care workers
Li	2011	Primary, quantitative	DE	Japan & China	Frontline: organisations
Liu	2010	Primary, qualitative	DH	China	Management
Lobach	2007	Primary, quantitative	HIE	USA	Management
Maenpaa	2011	Primary, quantitative	HIE	Finland	Frontline: hospital
Maiorana	2012	Primary, mixed	HIE	USA	Frontline: workers, disease
Martinez	2015	Primary, quantitative	HIE	USA	Frontline: hospitals
Massoudi	2016	Primary, qualitative	HIE	USA	Frontline: organisations
Mastebroek	2017	Primary, qualitative	HIE	Netherlands	Frontline: patients
Mastebroek	2017	Primary, qualitative	HIE	Netherlands	Frontline: patients
Mastebroek	2016	Primary, quantitative	HIE	Netherlands	Frontline: health care workers
Matsumoto	2017	Primary, qualitative	HIE	USA	Frontline: workers, hospitals
Medford-Davis	2017	Primary, quantitative	HIE	USA	Frontline: patients, hospitals
Mello	2018	Primary, qualitative	HIE	USA	Management: policies
Merrill	2013	Primary, quantitative	HIE	USA	Frontline: managers
Messer	2012	Primary, mixed	HIE	USA	Frontline: clinics, organisations
Miller	2012	Primary, qualitative	HIE	USA	Frontline: consumers, organisations
Miller	2017	Primary, quantitative	HIE	USA	Frontline: disease, workers
Moore	2012	Primary, quantitative	HIE	USA	Frontline: workers, hospitals
Motulsky	2018	Primary, quantitative	HIE	Canada	Frontline: workers
Myers	2012	Primary, qualitative	HIE	USA	Frontline: disease, workers

Obeidat	2014	Primary, quantitative	IE	Jordan	Frontline: hospitals
O'Donnell	2011	Primary, quantitative	HIE	USA	Frontline: workers
Onyile	2013	Primary, quantitative	HIE	USA	Frontline: patients
Opoku-Agyeman	2016	Primary, quantitative	HIE	USA	Frontline: hospitals
Overhage	2017	Primary, quantitative	HIE	USA	Management
Ozkaynak	2013	Primary, qualitative	HIE	USA	Frontline: hospitals, workers
Park	2015	Primary, quantitative	HIE	South Korea	Frontline: clinics, hospitals
Park	2013	Primary, quantitative	HIE	South Korea	Frontline: clinics, hospitals
Patel	2011	Primary, quantitative	HIE	USA	Frontline: clinics, hospitals
Politi	2015	Primary, quantitative	HIE	Israel	Frontline: hospital
Ramos	2016	Primary, mixed	HIE	USA	Frontline: patients
Ramos	2014	Primary, qualitative	HIE	USA	Frontline: patients
Reis	2016	Primary, quantitative	HDE	USA	Frontline: hospital
Richardson	2014	Primary, qualitative	HIE	USA	Frontline: organisations, workers
Ross	2010	Primary, qualitative	HIE	USA	Frontline: clinics
Ross	2013	Primary, quantitative	HIE	USA	Frontline: workers, clinics, hospitals
Rudin	2009	Primary, qualitative	HIE	USA	Frontline: health care workers
Rundall	2016	Primary, qualitative	HIE	USA	Management: policy makers, leaders
Saef	2014	Primary, quantitative	HIE	USA	Frontline: hospitals
Santos	2017	Primary, quantitative	HIE	Brazil	Frontline: clinics, hospitals
Shade	2012	Primary, quantitative	HIE	USA	Frontline: clinics, hospitals
Shank	2012	Primary, quantitative	HIE	USA	Frontline: health care workers
Shapiro	2013	Primary, quantitative	HIE	USA	Frontline: hospitals
Shapiro	2007	Primary, quantitative	HIE	USA	Frontline: health care workers
Sicotte	2010	Primary, qualitative	HIE	Canada	Frontline: workers, hospitals
Sprivulis	2007	Primary, quantitative	HIE	Australia	Frontline: workers, organisations
Squire	2002	Primary, mixed	HIE	USA	Frontline: health care workers
Sridhar	2012	Primary, quantitative	HIE	USA	Frontline: hospital

Thornewill	2011	Primary, mixed	HIE	USA	Frontline: consumers, organisations
Unertl	2012	Primary, qualitative	HIE	USA	Frontline: clinics, hospitals
Vest	2010	Primary, qualitative	HIE	USA	Frontline: hospitals
Vest	2017	Primary, qualitative	HIE	USA	Frontline: consumers, organisations
Vest	2015	Primary, qualitative	HIE	USA	Frontline: consumers, organisations
Vest	2013	Primary, qualitative	HIE	USA	Management: policy makers
Vest	2009	Primary, quantitative	HIE		Frontline: workers, patients
Vest	2017	Primary, quantitative	HIE	USA	Frontline: consumers, organisations
Vest	2011	Primary, quantitative	HIE	USA	Frontline: patients, hospitals
Vest	2014	Primary, quantitative	HIE	USA	Frontline: patients, hospitals
Vest	2014	Primary, quantitative	HIE	USA	Frontline: hospitals
Vest	2015	Primary, quantitative	HIE	USA	Frontline: hospitals
Vreeman	2008	Primary, quantitative	HIE	USA	Frontline: laboratory, radiology
Wen	2010	Primary, quantitative	HIE	USA	Frontline: patients
Winden	2014	Primary, quantitative	HIE	USA	Frontline: clinical care
Wright	2010	Primary, quantitative	HIE	USA	Frontline: health care workers
Yeager	2014	Primary, qualitative	HIE	USA	Frontline: consumers
Yeaman	2015	Primary, quantitative	HIE	USA	Frontline: hospital
Zech	2015	Primary, quantitative	HIE	USA	Frontline: patients, organisations
Zech	2016	Primary, quantitative	HIE	USA	Frontline: patients, organisations
Zhu	2010	Primary quantitative	HIE	USA	Research
<b>Study protocol</b>					
Dixon	2013	Protocol, mixed	HIE	USA	Frontline: organisations
<b>Reviews</b>					
Esmailzadeh	2016	Review	HIE	n/a	Management: policy
Esmailzadeh	2017	Review	HIE	n/a	Frontline: patients
Fontaine	2010	Review	HIE	n/a	Frontline: primary health care
Hopf	2014	Review	DL	n/a	Frontline: health care workers

Kash	2017	Review	HIE	n/a	Frontline: hospitals
Mastebroek	2014	Review	HIE	USA	Frontline: disease, workers
Parker	2016	Review	HIE	USA	Research
Rahurkar	2015	Review	HIE	n/a	Frontline: hospital
Rudin	2014	Review	HIE	USA	Frontline: clinical care
Sadoughi	2018	Review	HIE	n/a	Management
Vest	2012	Review	HIE	n/a	Management
Dixon	2010	Review	HIE	USA	Research
Akhlaq	2016	Review	HIE	LMICs	Management, countries

**Table 2.** Characteristics of sampled studies (n=61)

<b>Study name</b>	<b>Date</b>	<b>Type of study</b>	<b>Intervention term</b>	<b>Country</b>	<b>Level of the health care system</b>	<b>Purpose of the study</b>
Akhlaq	2016	Review, qualitative	HIE	LMICs	Management, countries	Barriers and facilitators of HIE
Boyd Boyd	2014	Conceptual	RL	Australia	Research	Functions of record linkage
Burris	2017	Commentary	HIE	USA	Frontline: hospitals	Benefits of HIE
Campion	2012	Primary, quantitative	HIE	USA	Frontline: health care workers	Push and pull of HIE
Cimino	2014	Conceptual	HIE	USA	Management	Debates around consumer-mediated HIE
Dalan	2010	Conceptual	DM	USA	Management	Possibilities for clinical data mining and research
Dimitropoulos	2009	Conceptual	HIE	USA	Management	Privacy and security of interoperable HIE
Dixon	2010	Review, framework	HIE	USA	Research	Costs, effort and value of HIE
Downing	2017	Primary, quantitative	HIE	USA	Management: policy	Relationship between HIE and organisational HIE policy decisions
Downs	2010	Conceptual	HIE	USA	Management	Improving laboratory services through HIE
Dullabh	2013	Primary, qualitative	HIE	USA	Management: organisations	Experience of HIE implementation
Elysee Elysee	2017	Primary, quantitative	HIE, IO	USA	Frontline: hospitals	Relationship between HIE, interoperability and medication reconciliation
Esmailzadeh	2016	Review	HIE	n/a	Management: policy	HIE assimilation and patterns for policy
Esmailzadeh	2017	Review	HIE	n/a	Frontline: patients	Patients' perceptions of HIE
Fontaine	2010	Review	HIE	n/a	Frontline: primary health care	HIE for primary health care practices
Fontaine	2010	Primary, qualitative	HIE	USA	Frontline: primary health care	Barriers and facilitators of HIE in primary care practices
Frisse	2010	Conceptual	HIE	USA	Frontline: patients, workers	Impact of HIE on patient-provider relationships
Gadd	2011	Primary, quantitative	HIE	USA	Frontline: health care workers	Users' perspectives on the usability of a regional HIE

Gill	2001	Primary, quantitative	DL	South Africa	Frontline: patients, disease	Linkage of non-communicable diseases data
Greene	2016	Conceptual	HIE	USA	Management	Technical and financial aspects of HIE
Grossman	2008	Conceptual	HIE	USA	Management	Barriers to stakeholder participation in HIE
Haarbrandt	2016	Conceptual	DW	USA	Management	Approaches for a clinical data warehouse
Herwehe	2012	Primary, quantitative	HIE	USA	Frontline: health care workers	Implementation of an electronic medical record and HIE
Hincapie	2011	Primary, qualitative	HIE	USA	Frontline: health care workers	Physicians' opinions of HIE
Hopf	2014	Review	DL	n/a	Frontline: health care workers	Healthcare professionals' views of linking routinely collected data
Hu	2007	Conceptual	DS	USA	Management	Challenges in implementing an infectious disease information sharing and analysis system
Hypponen	2014	Primary, quantitative	HIE	Finland	Frontline: health care workers	User experiences with different regional HIE
Ji	2017	Primary, quantitative	HIE	Korea	Frontline: hospitals	Technology and policy changes for HIE
Jones	2012	Conceptual	DS	USA	Management	An overview of electronic data sharing
Kash	2017	Review	HIE	n/a	Frontline: hospitals	Hospital readmission reduction and the role of HIE
Kierkegaard	2014	Primary, qualitative	HIE	USA	Frontline: health care workers	Applications of HIE information to public health practice
Kierkegaard	2014	Primary, qualitative	HIE	USA	Management	Health practitioners' needs and HIE
Kuperman	2013	Conceptual	HIE	USA	Management	Potential unintended consequences of HIE
Liu	2010	Primary, qualitative	DH	China	Management	Defining data elements for HIE
Maiorana	2012	Primary, mixed	HIE	USA	Frontline: workers, disease	Trust, confidentiality and acceptability of sharing HIV data for HIE
Massoudi	2016	Primary, qualitative	HIE	USA	Frontline: organisations	HIE for clinical quality measures
Mastebroek	2014	Review	HIE	USA	Frontline: disease, workers	HIE in general care practice for people with disabilities

Mastebroek	2016	Primary, quantitative	HIE	Netherlands	Frontline: health care workers	Priority setting and feasibility of HIE
Mastebroek	2017	Primary, qualitative	HIE	Netherlands	Frontline: patients	Experiences of people with intellectual disabilities in HIE
Matsumoto	2017	Primary, qualitative	HIE	USA	Frontline: workers, hospitals	HIE in managing hospital services
Parker	2016	Review	HIE	USA	Research	The use of HIE in supporting clinical research
Politi	2014	Conceptual	HIE	n/a	Management	Use patterns of HIE
Rahurkar	2015	Review	HIE	n/a	Frontline: hospital	Impact of HIE on cost, use and quality of care
Ramos	2016	Primary, mixed	HIE	USA	Frontline: patients	HIE consent process in an HIV clinic
Ranade-Kharkar	2014	Conceptual	HIE	USA	Management	Improving data quality integrity through HIE
Ross	2010	Primary, qualitative	HIE	USA	Frontline: clinics	Motivators, barriers, and potential facilitators of adoption of HIE
Rudin	2014	Review	HIE	USA	Frontline: clinical care	Use and effect of HIE on clinical care
Vest	2016	Primary, qualitative	HIE	USA	Management: policy makers, leaders	Information-sharing needs and HIE
Sadoughi	2018	Review	HIE	n/a	Management	Quality and cost-effectiveness, and the rates of HIE adoption and participation
Santos	2017	Primary, quantitative	HIE	Brazil	Frontline: clinics, hospitals	HIE for continuity of maternal and neonatal care
Shade	2012	Primary, quantitative	HIE	USA	Frontline: clinics, hospitals	HIE for quality and continuity of HIV care
Shapiro	2016	Conceptual	HIE	USA	Frontline: workers, organisations	HIE in emergency medicine
Shapiro	2006	Conceptual	HIE	USA	Management	Approaches to patient HIE and their impact on emergency medicine
Vest	2012	Review	HIE	n/a	Management	National and international approaches of health information exchange
Vest	2015	Primary, qualitative	HIE	USA	Frontline: consumers, organisations	HIE to change cost and utilisation outcomes

Vest	2010	Conceptual	HIE	USA	Management	Challenges and strategies for HIE
Williams	2012	Conceptual	HIE	USA	Management	Strategies to advance HIE
Yaraghi	2014	Conceptual	HIE	USA	Management	Professional and geographical network effects on HIE growth
Yeager	2014	Primary, qualitative	HIE	USA	Frontline: consumers	Factors related to HIE participation and use
Zaidan	2015	Conceptual	HIE	Malaysia	Management	Security framework for nationwide HIE
Zhu	2010	Primary quantitative	HIE	USA	Research	Facilitating clinical research through HIE

# **CHAPTER 5: THE MOTIVATIONS AND OPPORTUNITIES, DESIGN PROCESS AND OPERATIONALISATION OF A DATA HARMONISATION INITIATIVE IN SOUTH AFRICA: AN ETHNOGRAPHIC CASE STUDY**

---

## **1. Introduction**

Effective health management decision-making relies on high-quality, relevant and easily accessible information that can be found in a routine health information system (RHIS) [6, 14]. A RHIS is any system that collects, distributes and enables use of health information, at regular intervals through routine mechanisms, to address predictable health information needs [10]. Its purpose is to provide informational support for health management decision-making, through effective data production and data utilisation mechanisms [10, 12].

The function of a RHIS is to produce quality routine health information and support effective use of routine health information for decision-making [4, 12]. Problems related to routine health information systems (RHISs) include variations across different databases of key variables and indicators for collecting, analysing and reporting health information across programmes that mean that data cannot be integrated into useable formats [11, 21]; fragmentation and duplication across levels of the health system, which result in high workloads for health care workers; and the production of poor quality data that do not meet the needs of decision-makers [9]. Even when there is adequate production of health information in useable formats, poor feedback mechanisms and outdated information are a further concern. Feedback reports are frequently delayed and information can quickly become obsolete as health managers and policy-makers face deadlines and time constraints in decision-making [4, 12, 107]. Health information that is outdated, even if high-quality, is of low value for management decision-making [12].

Data harmonisation (DH) is an intervention that can provide timely, relevant and accessible informational support for management decision-making [4, 9, 15]. DH is an innovative process of copying existing electronic data captured in various databases into a centralised data repository where the data is integrated and then transformed into useable formats for data users. [20]. It aims to improve the quality and utilisation of health information for monitoring and evaluation, and for decision-making [68, 108].

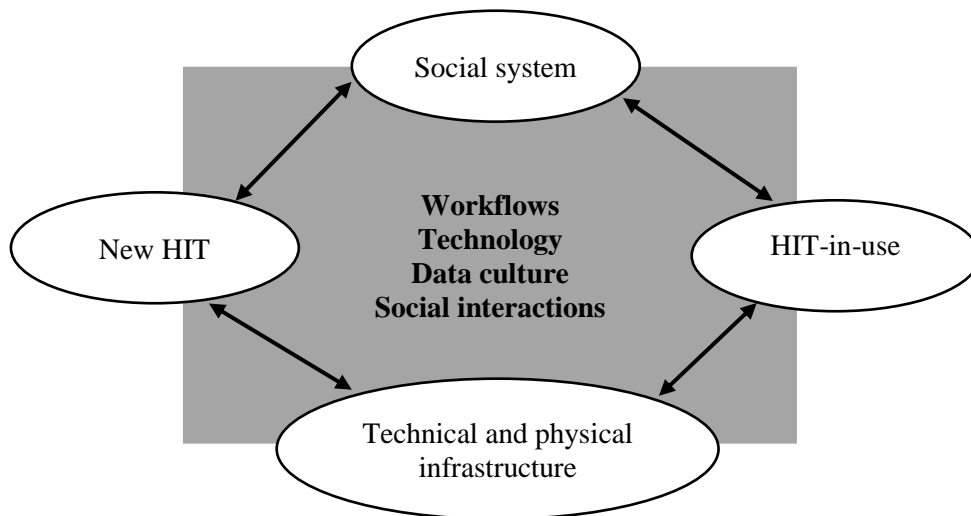
The operation and impact of a RHIS is shaped by three sets of determinants—technical, organisational, and behavioural—that work together [13]. On the one hand, DH has the potential to strengthen RHISs by addressing problems of poor data quality, fragmentation and duplication. This is possible through a technical process whereby heterogeneous data sources (with clinical, laboratory, pharmacy, mortality and administrative data) are integrated into a single information exchange platform [68]. Such a platform can facilitate easy and timely access to and retrieval of health data needed to provide informational support for management decision-making. On the other hand, developing and implementing DH innovations can be a time-consuming and expensive exercise that involves complex technical and operational processes to identify all relevant data sources and to set up the necessary technical infrastructure.

DH innovations are often viewed as merely technical solutions to seemingly simple technical problems of RHISs [68]. Technical interventions are generally aimed at improving the design and technical aspects of a RHIS [13]. They address specific technical problems through the development of new indicators, new data collection forms and procedural manuals; and updated computer software and hardware [13]. Reducing DH innovations to merely technical interventions (a technicist approach), however, assumes that fragmentation between RHISs can easily be solved by introducing standardised data indicators and by coding software that enables RHISs to exchange data amongst each other [42, 44]. This narrow focus on the technical requirements of DH neglects the fact that RHISs are embedded in a complex social system (the health system), where various health information technology (HIT) actors (those people who develop, implement and/or use HIT) across different levels of the health system interact with each other, in sometimes unpredictable ways [43]. DH innovations will be impacted by the same complexities and social factors that affect RHISs [13]. Neglecting the impact of social factors (including historical, political, organisational,

and behavioral factors) on DH processes may lead to poor outcomes of the innovation and result in wasted resources [109, 110].

A more comprehensive approach acknowledges that health information technology (HIT) innovations (including DH innovations) are influenced by interactions between new technologies and existing social and technical systems [42, 49]. The Interactive Sociotechnical Analysis (ISTA) approach was developed to counter the notion that HIT innovations are merely technical and that all unintended consequences of HIT processes are caused by technical factors or unpredictable social factors. Interactions between new HIT and existing social and technical systems can in fact be explained and predicted using the ISTA approach [42], which assumes that although technical problems exist, the functioning and ultimate impact of HIT innovations are shaped by a wide array of ‘sociotechnical interactions’. Sociotechnical interactions are interplays between new HIT innovations and existing social and technical systems of health services, which usually include workflows, culture (established social practices for producing and using data), social interactions, and technologies (see **Figure 1**).

**Figure 1.** Sociotechnical interactions (adapted from Harrison 2007) [42].



**Figure 1** identifies various sociotechnical elements that are interconnected and whose actions may lead to changes in workflows, data culture and social interactions in response to new technologies [42].

This paper reports on the sociotechnical factors affecting a DH initiative started in our local setting. The ISTA is a useful approach for analysing any development processes and dynamics associated with the new DH initiative. Firstly, it promotes a holistic view, considering various sociotechnical factors and their interactions in the development and implementation of HIT innovations. Secondly, it recognises that new HIT innovations have the potential to change prior workflows, communication and relationships, and that the changes emerging from interactions between existing HIT and new HIT innovations may lead to unintended consequences. Lastly, HIT users may, through recursive processes, re-interpret the purpose of new HIT innovations. Recursive processes are feedback loops that alter newly introduced HIT innovations and that may promote changes in the social system [42]. HIT users' re-interpretations of HIT innovations may lead to different uses and practices than those intended by HIT designers, which HIT users introduce to deal with unforeseen workflows and stakeholder relationships.

Our study used the ISTA approach to investigate DH innovations currently underway in the Western Cape Province of South Africa. We took a unique methodological approach in this study by using the DH innovation to produce data for another study we were conducting and then reflecting on the process and experience of engaging with the DH innovation. This provided an opportunity for not only accessing the data we needed for the other study, but also for learning about the processes and dynamics of DH innovations.

The Provincial Health Data Center (PHDC) (a new DH initiative) was started by the Western Cape Provincial Department of Health (WCDoH) in 2015 to integrate individual-level data captured in multiple, unconnected electronic databases across different types of health services and organisations. PHDC has been led by a government-university partnership called the Provincial Health Data Center (PHDC) [111-113]. Individual-level data refers to an individual patient's demographic, clinical, laboratory and pharmacy data which can be linked using a unique identifier (such as a shared patient record or patient folder number). The DH initiative is focused on collating

and integrating individual-level data into one shared platform as a way of providing informational support to researchers, health managers and policy-makers. This is meant to enable epidemiological and health service analyses to determine programme effectiveness and support disease surveillance, clinical and service management, and for financial (cost effectiveness) reporting [111].

The key goal of the DH initiative has been to create a province-wide electronic data repository. The data repository captures electronic disease-specific cascade data, at the individual level, over time and across different health care service levels (i.e. primary to tertiary health care levels). Electronic, harmonised cascades of individual-level data then enable health care workers to track individual patients as they move between the municipal and provincial primary health care (PHC) services; and between PHC and hospital care. The data repository captures different types of health services data related to health care utilisation, treatment initiated and received, health status and treatment outcomes, and morbidity and mortality.

Our engagement with this DH initiative as part of a separate epidemiological project presented us with a unique opportunity to learn more and reflect on the sociotechnical factors that impact on DH innovations. In this study, we investigated the opportunities, processes and potential of a DH innovation process as it was unfolding. We identified the contextual facets that enabled the DH innovation to take off, the responses made to design challenges and other barriers, and the attempts at institutionalising the new harmonisation processes.

The ISTA conceptual framework provides us with a lens through which to reflect on and analyse DH processes. This framework makes it possible for us to compare our research with other studies done in different contexts since it provides a common language to talk about HIT innovations (such as DH) and in turn enhance the potential for international comparisons of research findings. More specifically, our reflections on the sociotechnical facets of DH innovations may provide a more comprehensive and dynamic picture of what enables DH innovations in our local setting in terms of the context, aspirations and design processes, and operational challenges. The findings may be of interest to those involved in DH activities such as developers, implementers and users. It may lead to better understandings of the factors that influence DH innovation efforts and this

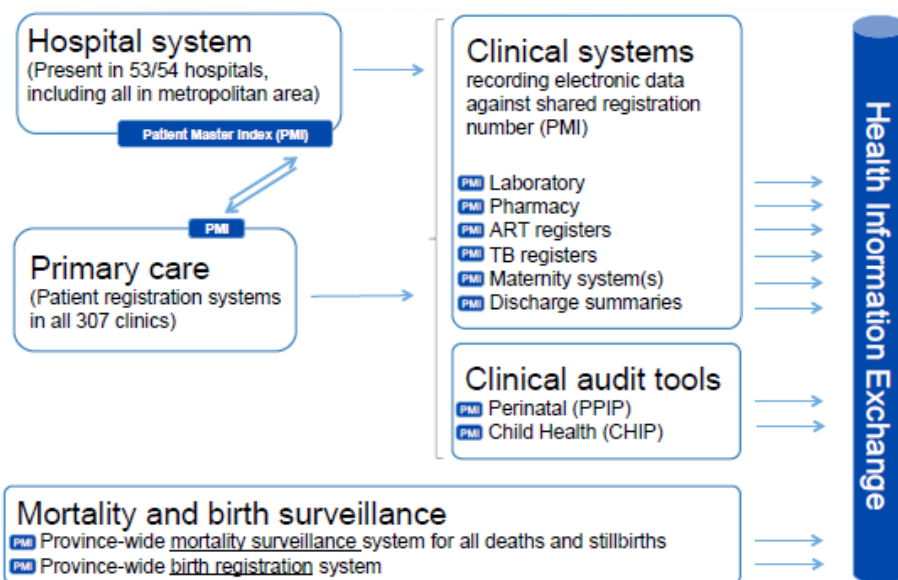
could in turn inform decision-makers and implementers about how to plan, monitor and optimise DH innovation projects. Additionally, this research may be of interest to individuals who are interested in the social complexities of HIT innovations and help deepen their understandings of the complex layers of systems, organisations and stakeholders involved [49].

## 2. Methods

### 2.1. Setting: The Provincial Health Data Center

PHDC is the unit within the WCDoH that has established a central data repository of harmonised individual-level data. The leadership and staff of PHDC are professionals who work in the interface between government and university epidemiology. PHDC has recently started to integrate demographic, clinical (including laboratory and pharmacy data), financial and human resources data captured electronically in routine health-related databases through DH innovations [111, 112, 114]. **Figure 2** illustrates the complexity of data harmonisation performed at PHDC.

**Figure 2.** Data harmonisation (health information exchange) at PHDC [114]



There is currently fragmentation between multiple large-scale databases that contain routine data relevant for individual patient clinical management. Routine electronic databases are spread across different types of public sector health services, that is, primary, secondary and tertiary health care services. PHC services are the first point of entry for preventive services (such as immunisation and family planning) and some curative services (such as TB and HIV treatment); and secondary and tertiary services provide hospital and specialised care. Public sector health services are free of charge and they serve the majority of South Africans (80% of the population and mostly no or low-income earners) [115].

The Western Cape Province is one of nine provinces in South Africa. Each province is divided into several health and administrative districts. The Western Cape has five health districts, the Cape Metropolitan district being the most urbanised and it serves the largest with a population of approximately 4.5 million. Health services are delivered by public and private organisations, with public sector health care serving the largest, mostly low socio-economic communities. Health care services are delivered at PHC level (clinics and community-based care); and secondary and tertiary hospital level care. In the Cape Metropolitan district, PHC is delivered by both the local municipal health authority (which we refer to as the City health authority) and the WCDoH (which we refer to as the Province health authority). The City health authority provides mostly preventive mother, child and reproductive health services (such as immunisation, family planning, HIV testing, STI treatment) and some curative care, while the Province health authority provides a full PHC package of care, including curative care services. There is overlap between City and Province PHC services, for example, with both providing STI, HIV and TB care, which sometimes leads to duplication of services, and even of databases. The WCDoH is responsible for managing hospital care services and the National Department of Health (NDoH) coordinates country-wide laboratory services, via the National Health Laboratory Services (NHLS).

Integrating the routine databases of the City and Province PHC health services has been very difficult because there has not been a consistent and unique patient identifier across health authorities. There is also not a routine single unique patient identifier that allows for tracking of patient care across the primary and hospital health care levels. This has been a problem for long-term clinical management because health care workers cannot track patients as they move between

different health levels and facilities; and decision- and policy-makers cannot efficiently plan and monitor service management [23, 116, 117].

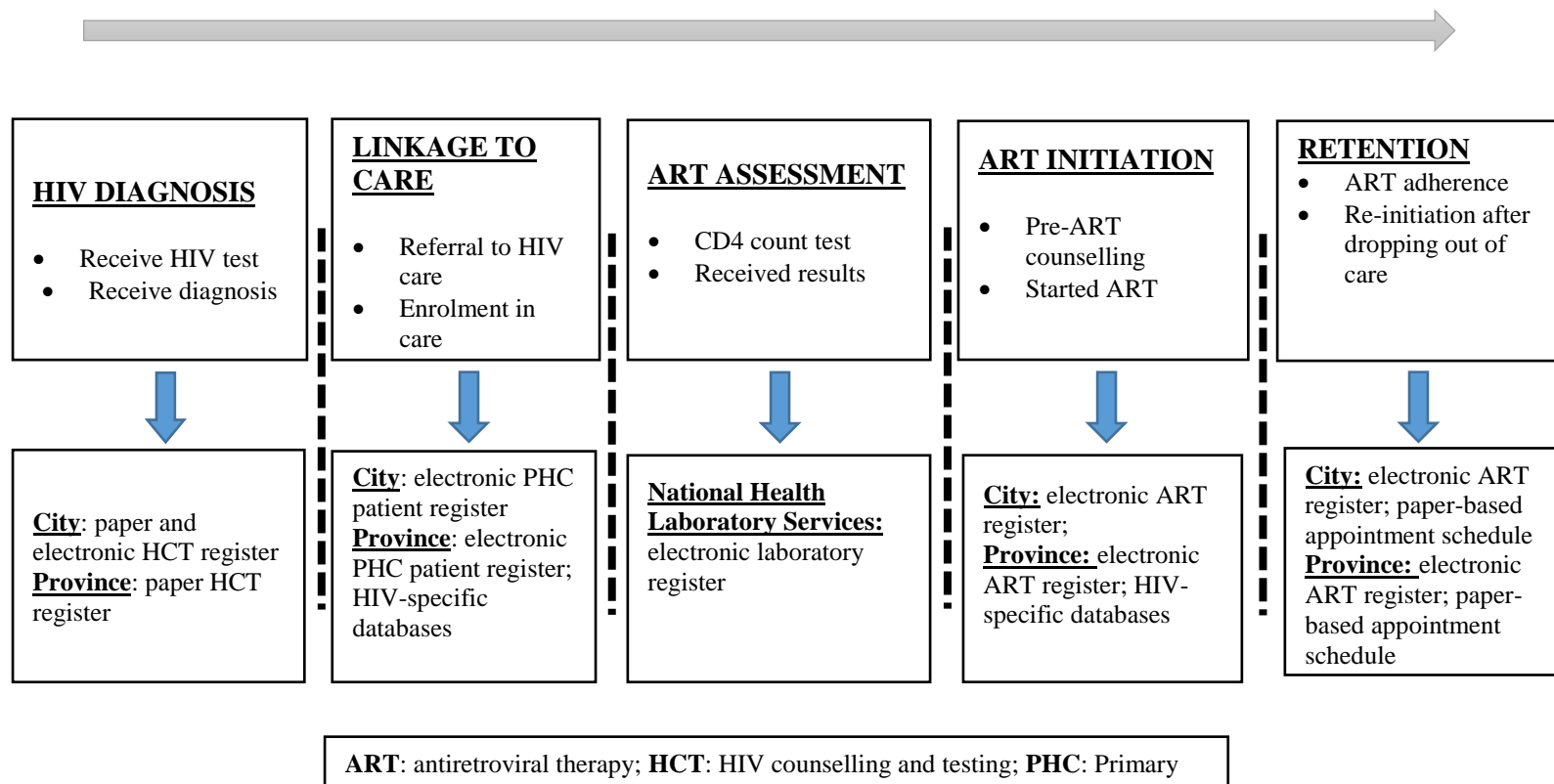
The linkage of various routine databases, across the two PHC health service delivery authorities and the across PHC and hospital level care services, became feasible when Clinicom, a new electronic patient register, was launched in 2006. Clinicom is used mainly by clinic reception or data clerks to allocate individual patients with unique identifiers (Clinicom numbers) that can be used universally at City and Province health facilities to admit, discharge and transfer, as well as to schedule appointments and maintain records. Clinicom had the potential for linking records kept in City and Province databases, enabling an opportunity for DH efforts to emerge. At the time of this study, DH innovators were in the process of building an application that synthesises patient information using Clinicom as well as generate disease-specific cohorts with patient-level data [111]. The application and the electronic cohorts would provide frontline workers and health service managers with the information they needed to track the performance of long-term patients across different services.

## **2.2. The iALARM project**

The researcher was first introduced to the PHDC via the iALARM research project that needed harmonised cohort data from the PHDC for a retrospective study on retention in HIV care. “Using Information to Align Services and Link and Retain Men in the HIV Cascade”, abbreviated as iALARM, investigates ways to enhance linkage to HIV care for men, using routine health information to support collaborative efforts between health facility and community-based care organisations. The retrospective cohort study quantitatively assessed the performance of an existing cascade of HIV services (a continuum of the stages of accessing care for HIV testing, treatment, and retention in care). The HIV cascade data would contribute to the development and piloting of a health information management (HIM) intervention aimed at promoting collaboration between health system and community stakeholders to enhance linkage to HIV services for men. The data collection process relied on iALARM researchers having access to health services data across a wide range of different services, levels and databases.

The concept of the HIV cascade provided iALARM researchers with a useful model for identifying the points at which HIV-related data (i.e. demographic, clinical, laboratory, pharmacy and mortality data) were needed, and in which databases, across City and Province health authorities. **Figure 3** shows the steps of the cascade along with the relevant indicators and databases required to assess progress at that cascade step.

**Figure 3.** The HIV treatment cascade, key indicators and relevant routine health information databases in the Western Cape



iALARM researchers needed to extract routine data that could show the state of HIV services, in terms of coverage, retention and dropout rates from HIV testing to long-term retention in antiretroviral treatment (ART) services. iALARM researchers planned to extract a cohort (cascade profile) of adult men and women who first tested HIV positive in the health sub-district between 1 January 2012 and 31 December 2013. We needed to identify a cohort of patients who had been enrolled into HIV care in one sub-district, over a two-year period (2012-2013), and to track their health care utilisation data as well as their health status indicators over the subsequent three-year period, based on available, retrospective routine data. Follow-up data for clinical factors (HIV diagnosis, ART eligibility assessment, ART initiation, ART retention), and time-to-event factors (time between HIV cascade steps) were to be extracted up to 31 December 2015 and this would provide a retrospective, longitudinal assessment of HIV service performance across the cascade.

Extracting the electronic data for such a comprehensive retrospective assessment of the HIV cascade was not practical prior to this, given the fragmented nature of the routine databases at the time, as well as the lack of a single unique patient identifier. However, with the DH project of the PHDC, developing an electronic HIV cascade became a possibility. This also became an opportunity to test the progress of the DH efforts, in effect, using the iALARM data request as a practice run to harmonise data for a specific research purpose, and in the process, to use it as a demonstration of the functionality and potential of the data harmonising project.

### **2.3. The ethnographic study on DH innovations**

This ethnographic study began when iALARM researchers asked DH innovators at PHDC to assist them with building an electronic HIV cascade necessary for the retrospective cohort study, which we refer to as the iALARM cascade. The request for building the iALARM cascade happened at a time when DH processes, such as building the data repository or extracting datasets for users, were still new and emerging. Thus, the relationship between iALARM researchers and PHDC staff was mutually beneficial; iALARM researchers wanted PHDC staff to assist them with designing and collecting data for the electronic HIV cascade, while DH innovators wanted to test technical processes that they were working on. Given the newness of DH processes and

ongoing work to test and improve these processes, it was necessary for PHDC staff to first design the process for building the iALARM cascade before engaging in any technical or practical work. Collaborative efforts to design and build the iALARM cascade involved continuous communication, for example, through email and telephonic communication and joint meetings. Although iALARM researchers primarily engaged with PHDC staff, they learned during the design process that PHDC did not yet have a complete data repository. There were negotiations underway for PHDC to access all available routine data from the City health authority and other government departments to build a comprehensive data repository. PHDC did not have access to all the data required for the cascade so iALARM researchers additionally engaged with sub-district health information staff for this ‘missing’ data.

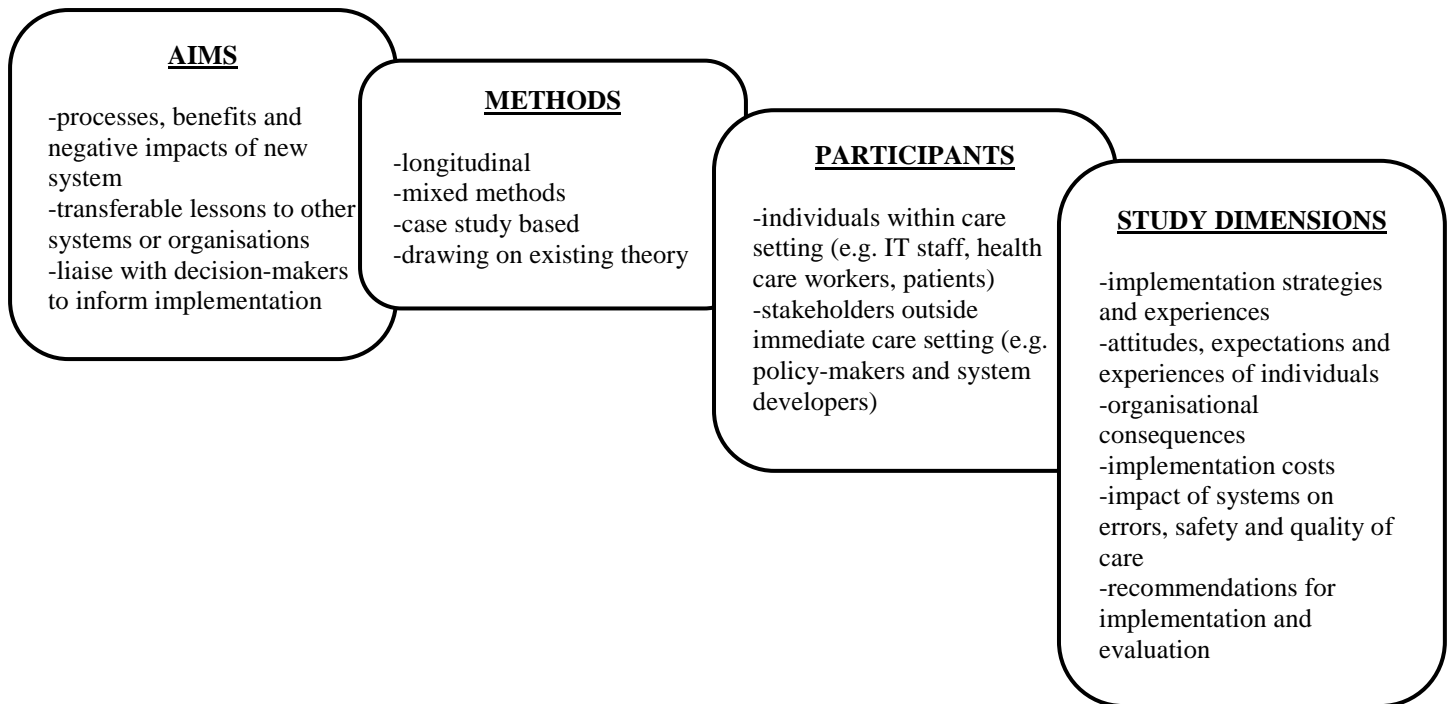
Regular communication between iALARM and PHDC entailed the lead researcher (BS) working closely with some of the PHDC staff to collect data for the cascade. Exposure to these DH processes presented a unique opportunity for the lead researcher to explore and reflect on the sociotechnical factors of DH innovations as development and implementation processes were taking place in the local setting. Thus, the design process of the iALARM cascade, which set out the steps for data collection, became a ‘lens’ through which to learn more about and reflect on DH innovations.

#### **2.4. Using the Sociotechnical Evaluation methodological framework to design our ethnographic study**

The Sociotechnical Evaluation (STE) framework is a methodological framework for evaluating the development and implementation of HIT interventions. It provides a summary of key components to consider in examining HIT innovations such as the DH innovation process studied here [44]. **Figure 4** presents the key components, i.e. aims, methods, participants and study dimensions, identified by the STE approach. The aim of our study was to learn about the context, processes and dynamics of DH innovations, which we did using ethnographic research methods. We used the iALARM cascade process as an entry point for collecting data for this study, since we were already engaging with DH innovators and health information staff to assist us with the iALARM cascade. We draw out key lessons about the context, and processes for the design and

operationalisation of the DH innovation. Drawing on the STE approach may allow for lessons learnt to be compared with those identified in other studies, where the STE or similar approaches are used in studying DH innovation. The lessons we share about DH innovations may advance strategies for planning, monitoring and optimising DH innovation projects.

**Figure 4.** Key components of the Sociotechnical Evaluation approach (adapted from Cresswell, 2014) [44]



## 2.5. Data collection methods

Using an ethnographic study approach, the lead researcher used qualitative methods to examine the context, sociotechnical processes, challenges, and opportunities of DH innovations in progress, as they were being developed by PHDC. Data were collected alongside the design process of the iALARM cascade. This research contributed towards the doctoral studies of the lead researcher (BS) and the research was approved by the University of Cape Town Human Research Ethics Committee (HREC ref: 320/2015 and 738/2018).

Conducting an ethnographic study allowed for the use of various qualitative research methods, including participant observation, document reviews and interviews [118]. The primary method

of data collection was participant observation, which involved (a) attending meetings such as iALARM, PHDC and iALARM-PHDC meetings and health services data review meetings; and telephonic and email communication and conversing with various data stakeholders such as data clerks, facility managers, health information staff and managers, and DH innovators while they were performing their tasks. This enabled BS to participate in and observe the work activities of data stakeholders, to learn about existing databases and emerging DH innovation processes and dynamics. Data were collected in the form of meeting minutes and field notes taken during meetings or conversations. Data were also collected through interviews with facility managers, data clerks and health information staff; and by reviewing meeting notes and internal documents. The lead researcher took detailed notes during interviews and highlighted key ideas in meeting notes and internal documents. Data collection took place from February 2016 to December 2016; BS conducted informal interviews and spent time understanding the work tasks of ten participants (four health information staff, two DH innovators, three frontline health workers, one health services manager). Cumulatively, BS attended twenty-four meetings; these included iALARM, iALARM and PHDC, and PHDC meetings for planning and designing the iALARM cascade; and sub-district data review meetings. The unique positioning of BS as a participant in the iALARM-PHDC project and as an independent researcher brought its own strengths and limitations to the study which are outlined in the Discussion section.

The data were analysed using an iterative process which “involves moving back and forth between concrete bits of data and abstract concepts, between inductive and deductive reasoning, between description and interpretation” [119]. While data collection was ongoing BS and her PhD supervisors intermittently had discussions to identify broad topics that were prominent, BS then constructed categories (such as challenges and opportunities) for these broad topics. The lead researcher was able to clarify or expand on these broad topics by asking clarifying questions at meetings or by setting up follow-up interviews. Data analysis was conducted using thematic data analysis; BS manually coded data by highlighting recurring themes and evidence therefore, and by extracting prominent themes from the various sources of data. As strategy for testing the validity of emerging themes, we used triangulation of different data sources; including looking for consistencies or inconsistencies between data sources [120].

## **3. Findings**

### **3.1. Overview**

The findings are organised into three sub-sections. The first sub-section is about the motivations and opportunities that contributed to the emergence of DH innovations in the Western Cape Province. The second one is about emerging design and operational challenges and responses of the new DH initiative (PHDC). The last sub-section is about PHDC needing to balance tensions emerging from innovating and institutionalising data procedures.

### **3.2. Motivations and opportunities leading to data harmonisation innovations in the Western Cape Province**

PHDC emerged in response to both the challenges posed by having multiple routine electronic databases managed by the two different health authorities (City and Province), as well as the opportunities this presented for harmonising data.

Within the clinical area, information is fragmented across disease-specific electronic databases (for example, for TB and HIV), whilst there are also gaps in information. For example, there are no electronic monitoring systems for longitudinal follow-up for chronic disease like diabetes and cancer which rely on harmonised individual-level data. Historically, routine electronic data were primarily used to generate aggregate-level data; aggregate-level data is high-level data that are combined to assess overall service provision, performance and cost efficiency. The key challenge to integrating these databases is that they do not share a single unique patient identifier, making it extremely challenging to extract harmonised individual-level data. PHDC directly addressed these challenges by developing the technical tools and procedures to link the various electronic databases in ways that allowed for a more integrated, comprehensive and accessible RHIS, capable of producing meaningful and useful data for decision-making and planning.

To illustrate this challenge of fragmentation, we describe how it applied to our efforts to build a profile of HIV service delivery (HIV care cascade). For our iALARM research project, we needed

to access retrospective cohort data of all HIV-positive patients who utilised health services over a three-year period since their first enrollment into HIV services in 2012 and 2013. One of the key challenges for collecting the data that they required to build the iALARM cascade had to do with the fragmentation of routine electronic databases across City and Province health authorities, and the fragmentation and overlap within provincial health department databases. In Table 1, we show the multiple routine electronic databases that provide relevant information to complete the iALARM cascade. The table provides details on the name of the database, the health authority where the database is used, the scope and indicators used in the database and the relevant data for the iALARM cascade. As we show, the databases have different and sometimes overlapping purposes, which also mirrors the different and sometimes overlapping service delivery patterns between the City and Province health authorities. These databases jointly contained most of the data required for the iALARM cascade but given that the databases were not linked in any way and split across health authorities, it would have been nearly impossible to access the relevant data in any meaningful way. PHDC staff were able to link and integrate some of these databases, and they took on the task of accessing relevant data for the iALARM cascade through their emerging PHDC data repository. Below we identify two key opportunities, both technical and structural, that facilitated DH innovations such as the PHDC data repository.

**Table 1.** Routine electronic databases relevant for building the iALARM cascade

<b>Domain name and authority</b>	<b>Scope and indicators</b>	<b>Relevant data for the iALARM cascade</b>
<b>Patient Registration and Health Management Information System (PREHMIS)</b>  <b>City of Cape Town Health Department</b>	PREHMIS is used for patient administration and routine information collection in city primary health care facilities.  PREHMIS stores basic patient information and the services provided to patients; it is not meant to replace patient folders.	<ul style="list-style-type: none"> <li>• Patient demographics</li> <li>• HIV counselling and testing data</li> <li>• Record of clinical visits</li> </ul>
<b>Primary Health Care Information System (PHCIS)</b>  <b>Western Cape Department of Health</b>	PHCIS is used for patient administration and routine information collection in provincial community health centers (CHCs).  PHCIS stores patient demographics, basic facility information, services provided to patients, record of clinical visits, appointment scheduling; and admissions, discharge, and transfers	<ul style="list-style-type: none"> <li>• Patient demographics</li> <li>• Records of clinical visits</li> </ul>
<b>Evaluation of the Khayelitsha AIDS Programme (EKAPA)</b>  <b>Western Cape Department of Health</b>	EKAPA is the provincial centralised online solution for HIV and TB monitoring. It is being merged to the Primary Health Care Information System (PHCIS), and it is further being developed into a multi-disease monitoring platform and Electronic Medical Record (EMR) system. EKAPA stores TB- and HIV-related clinical visits	<ul style="list-style-type: none"> <li>• Patient demographics</li> <li>• HIV diagnosis</li> <li>• Co-morbidities (such as TB)</li> <li>• TB/HIV clinical visits</li> </ul>

<b>Interlinked Electronic Registers (Tier.Net)</b>  <b>Western Cape Department of Health</b>	<p>Tier.Net is a standalone electronic register for monitoring HIV/AIDS services (particularly ART) in the Western Cape Province. It functions as a stepping stone between paper-based registers and a full electronic medical record (EMR) software (or online solution), called EKAPA. Tier.Net is in use in most provincial primary health care providing ARVs. It is further being developed to add TB and maternal and child health functionalities.</p> <p>HIV-related data, including pharmacy and laboratory services and available mortality data</p>	<ul style="list-style-type: none"> <li>• Patient demographics</li> <li>• HIV diagnosis</li> <li>• ART-related services (eligibility assessment, treatment initiation, retention)</li> </ul>
<b>DISA and TRAC: National Health Laboratory Services (NHLS)</b>  <b>[DISA and TRAC are not abbreviations]</b>	<p>DISA and TRAC store laboratory data, including laboratory tests performed and results for clinics, CHCs and hospitals.</p>	<ul style="list-style-type: none"> <li>• Patient demographics</li> <li>• Identifying health facility requesting testing</li> <li>• CD4 count test information</li> <li>• Drug resistance or toxicity information</li> <li>• Co-morbidities (such as TB)</li> </ul>
<b>National Death Index: Department of Home Affairs (DHA)</b>	<p>Births and deaths are registered at the national DHA; records are immediately accessible electronically</p>	<ul style="list-style-type: none"> <li>• Mortality data</li> </ul>

The first opportunity to operationalise PHDC related to the maturity of the RHIS in the Western Cape; for example, a patient identification system called Clinicom was introduced to link some electronic databases. Clinicom works by allocating individual patients with unique identifiers (Clinicom numbers) that can be used universally at health care facilities, from primary to hospital care to admit, discharge and transfer, as well as to schedule appointments and maintain records.

PHDC recognised that the Clinicom system could potentially be used as a first step to link various clinical, laboratory, pharmacy and mortality data captured across multiple routine electronic databases managed by both City and Province health authorities.

However, the Clinicom system was challenging in itself, in that its implementation across the health authorities was uneven and it created new problems related to duplication and errors in patient records. The transition from using a range of identifiers (such as databases- or clinic-specific identifiers) to using Clinicom numbers happened at different times across health authorities. This resulted in staff at some health facilities having to check if a patient had been assigned a new Clinicom number using the routine database interface. If not, a new Clinicom number was then created using the Clinicom system and then manually linked with the old patient identifier in an attempt to harmonise a patient's information. This process resulted in additional workload for staff and it increased chances for data errors and duplications of patient details, because of the use of both the routine database and Clinicom in creating the Clinicom number.

Clinicom was eventually directly linked to individual City and Province routine databases and the linkage of an old identifier with a new Clinicom number happened automatically. Although Clinicom initially seemed like an opportunity for PHDC, the delayed transition to and automatisisation of Clinicom initially caused technical challenges for PHDC to operationalise interoperability between databases.

The second opportunity to operationalise DH innovations involved PHDC's institutional structure and positioning as a government-research collaboration. Firstly, the PHDC team was known amongst health authorities; that enabled closer ties of accountability with key stakeholders and it gave them the authority to engage with City and Province health authorities directly. Secondly, the leader of PHDC had expertise in the RHIS; he was involved in the development of EKAPA and Tier.Net (described in **Table 1**) which begun about two decades prior. Thirdly, PHDC staff could test their capacity to harmonise individual-level data by piloting DH processes on research projects before attempting province-wide innovations. This allowed PHDC to gradually progress from being a small-scale DH initiative (that was only providing informational support to research projects) to becoming fully institutionalised within the health services and across the health

authorities. The setup of PHDC was helpful in operationalising and institutionalising DH processes. This is because DH processes required direct engagement with health authorities, permissions to access their databases, familiarity with and expertise in the management of the various databases and need for negotiating difficult issues of governance around databases.

The positive attitude of frontline health staff towards emergent DH innovations was another aspect that showed progress of DH innovations, particularly at the lower levels of the health system. Frontline health staff were one of PHDC's stakeholders; they could provide PHDC with feedback on the usability of DH products in clinical settings and identify areas of improvement. They were optimistic that harmonised data could support them in daily clinical decision-making; for example, in that they would be able to track patients across different health services and geographic areas. Frontline health staff's optimism towards DH innovations also indicated that they would be unlikely to oppose PHDC once DH innovations were rolled out.

i-ALARM researchers spoke to frontline health staff about the iALARM cascade, and staff said that the electronic HIV cascade would be useful for their work for several reasons. Firstly, they would use the electronic HIV cascade to assess trends of cascade outcomes and events over time, for example, by age categories, gender and health care facilities. Frontline health staff explained that data captured in an electronic cascade could then easily be translated into other formats (such as diagrams or tables) to point out gaps in HIV services and guide collaborations with community-based organisations. Secondly, they imagined the possibility of a routinised HIV cascade report which would enable them to evaluate service performance, motivate for additional resources (such as clinicians and community health workers), as well as initiate patient-level tracking interventions. For sub-district health managers, test cases such as the iALARM cascade, were important because they could contribute towards the development of large-scale provincial surveillance interventions.

The Clinicom system and the institutional positioning of PHDC as a government-research consortium presented two important opportunities for designing and operationalising DH innovations locally. Frontline health staff's positive attitudes towards DH innovations was a motivating factor for PHDC given that frontline staff were the intended users of DH products. On

the down side, PHDC faced challenges in the process of designing and operationalising DH innovations.

### **3.3. Emerging design and operational challenges and responses**

Challenges of DH innovations at PHDC became more obvious when we began the process of building the iALARM cascade. Below, we present two examples of challenges related to the iALARM cascade design process. The first example is about the difficulty of accessing HIV-cascade-related data captured in the City database (that is PREHMIS) at the sub-district level, and the second example is about the challenges associated with incompleteness of electronic HIV testing data.

The first example describes the chronological steps that we and PHDC followed to build the iALARM cascade. Using this example, we are able to illustrate two points; firstly, that initial ideas about the feasibility of DH innovations may change once practical work is underway and secondly changes to DH innovations occur through iterative processes and interactions between DH designers, users and other stakeholders.

The first step in building the iALARM cascade involved us approaching PHDC to assist us with extracting, linking and merging a subset of clinical, laboratory, pharmacy and mortality data into a single dataset. At the time of our request, PHDC was only beginning to operationalise DH innovations so technical processes were still being tested and continuously improved. But because PHDC staff were better positioned to access the data that we required, they helped us lay out the steps for collecting the iALARM cascade data. PHDC staff planned that we would access routine HIV services data from both City and Province databases. However, PHDC staff could at that point only access routine data for patients who had accessed HIV-related services at Province facilities. The data that PHDC received from the City was extracted from PREHMIS retrospectively (on a quarterly basis) which meant there was a gap in the data in terms of the follow-up period of the iALARM cascade. There were ongoing and promising discussions between PHDC and City stakeholders to link a ‘live’ version of PREHMIS to the data repository. But because they had not reached an agreement yet, PHDC staff advised iALARM researchers as

a workaround to access the City data that they required with the help of City health information staff in the sub-district.

Once this gap in the process was identified, the second step involved us approaching City health information staff working in the health sub-district of interest to assist us with accessing any further PREHMIS data that would complete the iALARM cascade. Health information staff in the sub-district were responsible for aggregating health services data (from City clinics) captured in PREHMIS into routine monthly reports for analysing health service performance. They could search for an individual patient using a patient folder number or personal details (such as name, surname and date of birth) through the PREHMIS interface. This enabled them to see patient registration information (such as demographic information) and some health care utilisation information (such as last visit to the clinic and associated conditions).

iALARM researchers first approached health information staff in the sub-district to extract a longitudinal dataset of our enrolled patients who accessed City clinics. Health information staff in the sub-district were willing to assist iALARM researchers, but they realised when they began to think through the necessary steps, that it was not feasible to do so using their PREHMIS interface. They could only search for one patient at a time. They were also unable to search for individual visit or outcome data collected outside that specific sub-district, even if they had the patient identifier. Even though health information staff were able to extract large datasets of aggregated data, they were unable to input a list of patient identifiers, filter for specific data and time variables, and extract a longitudinal dataset for a specific cohort (such as the cascade profile) with multiple outcome variables across all City facilities. Once sub-district health information staff realised this limitation, they referred iALARM researchers to health information managers at the district level, who could perform a wider range of functions in PREHMIS.

Before we began this third step of approaching the health information managers, we again asked PHDC staff for their assistance in identifying the right people who could assist us. Neither PHDC nor iALARM researchers had realised that health information staff would be limited in their access to PREHMIS data. As a next step, the director of PHDC advised us to approach the main City health information managers for help. PHDC had been in negotiations with these health

information managers about routine access to PREHMIS data for their data repository. They also suggested that iALARM researchers ask the City to work directly with PHDC instead of through iALARM staff. Our original plan had been to collect the cascade data from the City (sub-district level) and PHDC separately and then merge these into a single analysis dataset. In this new approach, PHDC staff would work with City health information managers directly to extract and transfer data; PHDC would perform the linkage of datasets and give i-ALARM a de-identified version.

One of the reasons why PHDC offered to work directly with City health information managers was so that iALARM researchers would not have to access named data (of the cascade profile). PHDC staff explained that one of the functions of the Centre was to minimise any ethical risks to data users and patients associated with researchers accessing health services data. They wanted to uphold the confidential and sensitive nature of health services data, particularly in the case of the iALARM data request where they were dealing with HIV-related data. The intended analyses of the retrospective cohort study did not require iALARM researchers to obtain named data. PHDC staff de-identified all relevant data (including that obtained from PREHMIS) by removing patient identifying information and assigning each patient with a unique study identifier before giving iALARM the data. PHDC staff saw direct communication with City health information managers as an improvisation for protecting health services data, while also solving a practical data access problem for iALARM.

This new arrangement was more, however, than just a way to protect against the release of identifiable information. PHDC also saw the data access request by iALARM researchers as an opportunity to further negotiations with the City health authorities for routine access to PREHMIS. iALARM's data access request happened at a time when DH innovators at PHDC were already negotiating with health information managers for PREHMIS data, and these negotiations continued in parallel to the data collection activities of the iALARM study. PHDC hoped that our request would demonstrate the kinds of opportunities a routinely updated data repository would have for health research. Shortly after iALARM researchers discussed the new way forward with PHDC staff and began coordinating with the City, PHDC staff informed

iALARM researchers that the City and Province health authorities had agreed for PREHMIS data to be linked to the PHDC data repository.

The fourth step for collecting data for the iALARM cascade involved PHDC staff extracting health care utilisation data and outcome indicators for the cascade profile directly without the complicated step of us going back and forth between PHDC and the City health authorities. The change did not appear to happen because of iALARM's request, though it may have added a last-minute push to the process. This illustrates that there may initially have been gaps between aspirations of DH innovations and practical reality; to close these gaps, innovations are often therefore rapidly changing in the early phases of implementation. Additionally, technical systems challenges may be quickly addressed when there is effective collaboration amongst stakeholders.

The second example illustrates a similar gap between the aspirations of DH innovations, in terms of how they are conceptualised and designed, and the practical realities of operationalising them. It offers an additional lesson, however, that DH innovators and users needed to be flexible, realistic and practical in adapting new approaches to deal with gaps between the aspirations and practical realities of DH. Below is an example of how an unexpected challenge in building the iALARM cascade resulted in PHDC and iALARM researchers needing to adjust their initial plans, in terms of the scope of the electronic cascade.

Before iALARM researchers started data collection for the iALARM study, they consulted PHDC staff about the possibilities for building an electronic HIV cascade (including HIV testing data) using retrospective cohort data. Their conversations were not detailed at this stage, as iALARM researchers were trying to map out the services offered at PHDC and the types of data that were available electronically. The director of PHDC, who was also involved in the iALARM study, was optimistic that building the iALARM cascade was possible because of the various types of routine electronic data across the HIV cascade that they had linked to the data repository. However, once iALARM researchers and PHDC staff started to more concretely operationalise the steps for building the cascade, PHDC staff became uncertain about the availability of electronic HIV counselling and testing (HCT) data. PHDC staff knew that City and Province health authorities had started the process of transitioning HCT data from paper-based registers to

an electronic database, but they were uncertain whether this transition was complete amongst different health facilities across the province. Conversations about the availability of electronic HCT data also led iALARM researchers to the realisation that the relevant HCT data were not only held by the public sector, but also by non-governmental organisations (NGOs), private general practitioners and workplace wellness programmes. PHDC staff and iALARM researchers had not thought about this in their earlier conversations; their focus was on the data that was available at PHDC linked to the data repository.

To deal with the gap between aspirations of DH innovations and practical realities, DH innovators adopted a flexible approach in building the iALARM cascade. They were sceptical about the extent to which there was complete electronic HCT data in the data repository, but they still planned to start off the data extraction process by searching for all adults who tested HIV positive in a health sub-district in Cape Town in 2012 and 2013. PHDC staff knew that the data repository was growing (i.e. more data were being linked and archived) so they wanted to test whether they could extract a complete subset of the cascade. PHDC staff, however, saw that there was too little HIV testing data to extract (even for a sample analysis) meaning that most provincial health facilities had not yet transitioned to the electronic system.

PHDC staff and i-ALARM researchers again engaged in conversation to think through steps for redesigning the cascade; they decided to start the search with all adults who were first registered in HIV-related services in the health sub-district by having a first CD4 count done. This was the earliest step in the cascade for which routine data were available in the public sector; HCT data were also unlikely to be found in the private sector. DH innovators proceeded to extract datasets for the iALARM cascade, excluding HCT data, which slightly changed the cohort from those who tested positive in the sub-district at public sector health facilities, to those who initiated ART eligibility screening (CD4 testing) at public sector health facilities. DH innovators had to improvise by building the iALARM cascade starting with a later step (CD4 count measurements) for which there was complete electronic data.

Difficulty accessing the City database and the incompleteness of electronic HIV testing data were two examples of challenges that PHDC faced in building the iALARM cascade. PHDC tried to

address these challenges through collaboration with the City health authority and flexibility to change DH products. Below, we share examples related to similar challenges of balancing tensions emerging from operationalising and institutionalising new data procedures.

### **3.4. Balancing tensions emerging from innovating and institutionalising new data procedures**

As illustrated above, DH innovations at PHDC were in their early stages of development so potential opportunities and challenges were still emerging. This meant that gaps between aspirations of DH and practical realities were continuously being discovered and needing to be addressed. Additionally, PHDC was still trying to figure out how to standardise and routinise DH processes, that is, how to operationalise tasks related to DH processes so that it could eventually become part of the standard functioning of the RHIS rather than a separate standalone innovation. Using two examples, one data access and transfer procedures and the other about the role of social relationships in dealing with these changes, we share key lessons about the operationalisation and institutionalisation of new data procedures.

The first example revealed that the need for new practices to deal with user demands and manage workloads may emerge from pilots of innovations. DH innovators at PHDC realised that they needed to standardise data request processes since they started working on DH projects such as the iALARM cascade. In the case of the iALARM cascade, we only started to have more detailed conversations about the steps for building the electronic HIV cascade once we had obtained ethical approval from the university and were granted permission to access the required data through the Province health authority. PHDC did not require iALARM to formally engage with an internal process for requesting data (such as completing an application form) at that time, as the formal application process was still being developed and institutionalised. The director of PHDC had in principle agreed for DH innovators to help us with accessing the data that we required. However, PHDC later asked us to engage in the formal data access application process even though the process for building the iALARM cascade was already being operationalised. PHDC explained that institutionalising the internal application process was necessary for record-keeping and resource management. The data access application form that we were asked to

complete covered the following areas: details of the data request, the outcome of the application and the terms of agreement for data.

The example revealed that once DH innovators started engaging with real-life data requests they realised that they needed to plan for how data requests would be managed in a systematic manner in the future. Establishing formal data request processes was important for both the purpose of administration and workload management, and to have transparent guidelines (that would eventually become institutionalised practice) for data users. Even though we had already begun working with PHDC, we were later required to submit a formal application. PHDC was trying to operationalise processes for building the iALARM cascade while also trying to institutionalise administrative procedures such as the data access application. By developing and implementing both operational and administrative procedures, PHDC was trying to demonstrate that it could manage data requests and fulfil institutional requirements at the same time. Formal administrative processes were important for PHDC to gain legitimacy and become fully institutionalised as a DH initiative that could handle data users' demands. Given that early DH processes started off as research projects, it was necessary for PHDC to establish formal procedures that would demonstrate its transition into the health services. Becoming fully institutionalised within the Province was important for the sustainability of PHDC, in terms of accessing health services data and providing informational support to health services stakeholders.

In dealing with early DH projects, PHDC also identified the need to document and operationalise systematic processes for data transfer (as they did for data requests). When PHDC staff began standardising data transfer processes, they asked the data users that they were working with at the time (including iALARM researchers) to provide feedback on a draft of a data transfer format that they hoped to implement. The data transfer document included instructions about the format in which datasets would be transferred to users, instructions to users on how to check for duplications, and tables listing variables related to patient demographics, mortality, births, and pharmacy and laboratory services. The draft also included a basic flowchart identifying key stages of data extraction and data transfer which DH innovators plan to perform once a data access request was approved. For research requests, the steps outlined in the flowchart were as follows: the researcher provides the DH innovator with the characteristics of patients, the DH innovator

generates study identifiers, the researcher validates the patient cohort, the DH innovator extracts the required data according to the study cohort, and as a last step, the DH innovator sends the required data to the researcher.

This example of PHDC introducing new data practices emphasises the lesson that DH innovations require more than just technical processes of data production. DH innovations also involve administrative processes, and institutional politics, of data access and data transfer. Data access and data transfer processes may be tested and operationalised in parallel to the technical implementation of DH innovations. We observed that PHDC was working on technical processes to build the iALARM cascade at the same time as they were trying to institutionalise the data access application and data transfer processes.

The second example has to do with DH innovators and data users balancing tensions emerging in the new data practices through informal social relationships. Up to now, we have pointed out that DH innovators realised the need to standardise and routinise data access applications and data transfer processes which would eventually contribute to the institutionalisation of PHDC. While these formal processes were underway, PHDC staff relied on informal and social relationships with us to iron-out any emerging challenges arising from their formalising processes while implementation was ongoing. The newness of administrative processes at PHDC led to iALARM researchers and PHDC staff frequently interacting about the application form and about the steps for transferring the iALARM cascade dataset. We needed to ask questions about the application form, for example, how to capture details about the iALARM cascade taking into consideration known challenges for accessing HCT and PREHMIS data. We also wanted to ask questions about the data transfer process, for example, the format in which we would receive the data and how we should define the different variables in the iALARM cascade dataset when conducting data analysis. This required frequent but different interaction with PHDC to when we were conceptualising the iALARM cascade. It was beneficial for us and PHDC to interact in more flexible ways while we were completing the data access application form and during the data transfer process, most often through telephonic or email chats.

To overcome communication barriers (that could potentially delay the submission of the data access application form and the data transfer process) the lead researcher (BS) and the DH innovator working on the iALARM cascade began communicating through flexible mediums, such as SMS and email conversations. The iALARM researcher began to interact with another DH innovator who was working on establishing administrative procedures at PHDC. The iALARM researcher also attended weekly meetings at PHDC which made it easier for her and the DH innovators to talk through challenges of and next steps for the iALARM cascade. PHDC did not yet provide data users with written explanations of the datasets that they provided, in terms of how the data were merged and the definitions of indicators. Even though we were asked to provide feedback on the data transfer format; the format was not specific for the iALARM dataset as PHDC was still working on the practical aspects. The iALARM researcher and the DH innovator interacted informally, through electronic and face-to-face conversation, to discuss matters related to the datasets.

This example is about the social relationship between an iALARM researcher and two DH innovators. An important lesson emerged; namely that social relationships are critical for supporting the evolution and consolidation of DH processes and products. Social interactions between DH innovators and data users can increase the usability of innovation products and outputs because innovators can explain the intended use of products, while also gaining insights into the challenges that users face. Users are able to obtain help with definitions and interpretations of data from innovators.

## **4. Discussion**

In this study we studied the opportunities, processes and potential of a DH initiative as it was unfolding. We identified the motivations and opportunities that enabled the DH innovation to take off, the responses to design challenges and the attempts at institutionalising new harmonisation processes. Key lessons can be drawn from the findings. Firstly, DH innovations can in some cases (such as PHDC) be strengthened through government-research collaborations. Secondly, DH innovations (especially during their initial stages of piloting) benefit from the support of relevant stakeholders such as frontline health staff who are potential implementers and/or users of DH

processes and products. Thirdly, existing social dynamics (ethical and institutional) related to work tasks, scope and hierarchical levels as well as the safeguarding of confidential and sensitive health services data used in DH innovations can have an impact on its successful implementation. Fourthly, DH innovators often come across new needs or barriers while DH projects are already underway which they need to respond in order to continue with implementation of the innovation. Fifthly, DH innovations are continuously changing due to new technologies and unintended sociotechnical interactions, so social relationships are important for navigating emerging challenges and closing gaps between design, administrative and technical processes. The overall message from the findings is that DH innovations are influenced by multiple and dynamic factors, including new technologies, existing relationships, different stakeholders, available infrastructure and the politics of organisations.

Our findings are consistent with other studies that have shown that HIT innovations (including DH innovations) are best optimised when sociotechnical factors and interactions are well-considered in the development and implementation processes [49, 118]. A study recently conducted in England reported on the design of HIT innovation to enhance an existing patient administration system (PAS) [118]. The PAS innovation aimed to integrate HISs that record patient data including in-patient admissions and discharges, outpatient appointments, and operation theatre activities. The authors participated in the innovation and conducted research on it at the same time. They showed how adopting a sociotechnical approach to the PAS innovation enabled them to identify all data stakeholders that needed to be involved. They were also able to identify data users who were initially neglected during the system development. In this study, the sociotechnical approach was applied as a subsequent ‘remedy’ to the technology-driven and top-down approaches taken during the system’s development process [118]. Like in the PAS innovation study, we were also able to learn more about the context, aspirations and operational facets of DH innovations while participating in a DH exercise. This also put us as researchers and data users in a position to offer feedback and a broader sociotechnical perspective to the DH innovators on the development and implementation processes.

A conceptual paper written by the United States (US) Department of Health and Human Services

also used the sociotechnical approach and identified ten key considerations for the successful optimisation of any large-scale HIT innovation [51]. The key lessons from that paper are aligned with ours: (a) HIT innovations face operational challenges related to differences in existing technical systems and organisational contexts; (b) HIT innovations involve data integration across organisational boundaries and stakeholder groups (such as HIT designers, researchers, clinicians and patients); and even newly implemented HIT innovations are challenged by new user needs and should be approached as a work-in-progress [51].

Another study (also conducted in England) described three case studies across different stages of a HIT innovation [49]. The first case study was on the design and conceptualisation stage and how health staff perceived the move from paper-based to electronic data systems; the second one was on the implementation stage and health staff were asked about the advantages and disadvantages of sharing information across a network; and the third one was on the use, adaptation and evolution stage and health staff were asked about the flow of information and working across organisational boundaries. Two of their key lessons are similar to ours; namely, that HIT designers need to capitalise on the optimism of data users to pilot and simplify new HIT processes and products for everyday use. And secondly, HIT designers need to be able to improvise and use a set of ‘workarounds’ in order to overcome unintended problems of working with HIT. In our study, PHDC staff worked with City health managers directly as a workaround when iALARM researchers could not access City data from health information staff at the sub-district level. Another workaround was the use of social relationships to deal with any technical or operational challenges emerging in relation to the practical use of datasets.

As mentioned earlier, the ISTA conceptual and STE methodological frameworks are useful for studying and analysing emerging HIT innovations in a holistic manner [42, 44]. These frameworks emphasise the importance of considering both social and technical factors and interactions of HIT innovations which include staff workflows, communication pathways, stakeholder relationships, existing HIT systems and the new HIT innovation. Our study has demonstrated the interplay between some of these factors. For example, the lack of integration between databases and limited communication between City and Province health authorities delayed the process for building the iALARM cascade; working on a joint technical project (the

iALARM cascade) was a catalyst for data exchange processes between City and Province health authorities. PHDC worked on technical processes in parallel to administrative processes for data access and data transfer which potentially changed the workflows of staff; and new social relationships were formed between iALARM researchers and PHDC staff to manage emerging queries about the data.

While our findings are broadly consistent with lessons from other studies and key principles identified in sociotechnical frameworks, there are remaining gaps in knowledge of HIT innovations [118]. Firstly, this study took place at a time when DH innovations were newly emerging in that development and piloting of DH innovations were still ongoing. Thus, further research is necessary for assessing the scale-up, institutionalisation, and sustainability of such HIT innovations, especially in contexts with limited resources where HIT innovations are weighed up against other urgent health priorities, and where there is still division between institutions such as local, regional and national governments. Secondly, given the newness of HIT innovations in general, more research from different contexts and looking at different types of HIT innovations is required. DH (the focus of this study) is only one type of HIT innovation and even though we participated in and observed a DH process, we were only able to report back on a few activities in one DH process. HIT stakeholders, health system managers, policy makers and researchers can be guided by conceptual or theoretical and practical evidence on the activities and processes of different types of HIT innovations. This evidence would guide policy development and ensure that implementation processes are well-planned to maximise the use of resources and obtain positive outcomes. Lastly, there is a need for researchers to get involved in the field of sociotechnical systems and HIT innovations specifically using methodologies such as participant observation that allow them to experience some of the processes of HIT innovations while also observing associated opportunities and challenges that different HIT stakeholders face [42, 49]. Their findings can help health system managers and policy-makers better understand the intentions of HIT innovations, the complex range of activities involved in developing and implementing HIT innovations and the social, technical and cost implications.

To our best knowledge, this is the first study in South Africa that examined the contextual, aspirational and operational facets of DH innovations. A limitation of this study has to do with

the continuously changing state of DH innovation processes; we only collected data over a period of approximately one year and the findings are of the initial phase of DH innovations (that is early development and implementation stages). However, this study was based on an actual data request submitted to PHDC (hence the study timeline) at a time when DH innovators were trying to operationalise DH processes and products in their local setting. The data request imitated a DH exercise requiring DH innovators to harmonise specific data from various clinical, laboratory, pharmacy and mortality databases into a single analysis dataset for a research project. The DH exercise presented a unique opportunity for the researchers to identify and reflect on sociotechnical factors of DH innovations, alongside the data collection process of the research project. Participant observation made it possible for lead researcher to participate in and observe the work activities of DH innovators and other HIT stakeholders. Additionally, data were collected from interviews and informal conversations which allowed for triangulation (verification of research findings) using different sources of data.

In conclusion, we identified potential opportunities and motivations that enabled DH innovations to become implemented. We also highlighted two key strategies (that is stakeholder collaboration and social relationships) for responding to design challenges and institutionalising new harmonisation processes. This study on the social aspects of DH innovations (and not only the technical ones) can guide future innovators' plans on developing and implementing DH projects for research and health services practice. Health system managers can also understand and manage the complex layers of new and existing social and technical systems, organisations and stakeholders involved in the development and implementation of DH innovations.

## **Contributions**

The PhD student (BS) conducted data collection and data analysis. Her PhD supervisors (CC and NL) provided guidance throughout the research process and provided critical feedback on the drafts.

# **CHAPTER 6: INSTITUTIONAL AND CONCEPTUAL DILEMMAS AND KEY SUCCESS FACTORS ASSOCIATED WITH DATA HARMONISATION PROCESSES**

---

## **1. Introduction**

Health information technology (HIT) innovations are being introduced into health care settings globally to improve individual health outcomes, health service delivery and overall health system performance [51]. HIT innovations involve technology-based interventions to improve the design, development, implementation, use and maintenance of information systems within health care settings [121]. When HIT innovations are optimised, health information systems (HISs) can be more effective at data production and data utilisation. In terms of health service delivery, well-functioning HISs can synthesise and integrate health information into useable formats (such as routine monthly reports) that provide informational support for real-time clinical and service management decision-making and they can reduce unnecessary testing of health service interventions [52]. The benefits of HIT innovations further lie in their ability to spur positive changes across other components of the health system, such as health care financing, health workforce, and medical supplies [1, 52]. This is because leadership and governance structures of health systems, comprised of health managers and policy-makers, rely on health information made accessible to them through the successful implementation of HIT innovations. They can use health information for planning, monitoring and evaluation of all health system components [1].

Overall, HIT innovations have the potential to improve several data aspects such as data quality, data synthesis, data usability, and timely data access. In light of these benefits, there have been numerous attempts to implement different types of HIT innovations across diverse health care settings [51, 52]. In parallel there has been growing recognition that the benefits associated with HIT innovations can only be realised when HIT innovations are fully optimised through successful implementation [51]. Successful implementation of HIT innovations involves joint

optimisation of social and technical elements of the health system, such as people, technologies and their use, the physical environment, and organisational structures [122]. As such, HIT innovations are developed and implemented within complex health systems where various social and technical elements sometimes act independently, in unpredictable ways, and whose actions are interconnected and may influence each other [43].

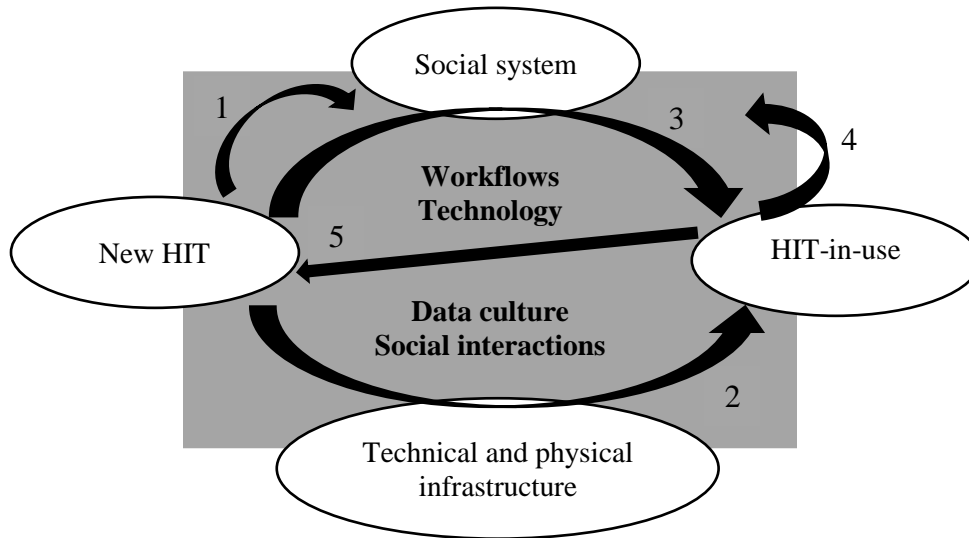
In many cases, however, HIT innovations are developed and implemented in a ‘technicist’ manner, with a primary focus on addressing technical aspects of the innovation [42, 68]. The technicist approach relies on the assumption that HIT innovations or technical interventions are sufficient to solve what appear to be technical problems related to data quality and accuracy, and even data accessibility and utilisation. This approach neglects the fact that HIT innovations are embedded in complex health systems, which consist of social systems where various health system actors (such as HIT designers and health service providers) across different levels of the health system (such as facility, district and region) interact with each other, in often unpredictable ways [13, 43]. As a result, it ignores the impact of social factors (including historical, political, organisational, and behavioral factors) on HIT innovation processes. This may lead to failures of HIT innovations, failures often reflected in poor outcomes of HIT innovations and wasted resources [42, 43].

There is growing evidence that HIT innovations should be developed and implemented in such a way that they can address and mitigate the impact of unintended consequences from interactions between social and technical factors within the HIT environment [8, 44]. However, there are a limited number of studies [44, 51] that have adopted a comprehensive approach to sociotechnical factors in their evaluations of HIT innovations. There is a need to conduct qualitative evaluations or case studies, taking a social perspective, on the factors that impact HIT innovations in real-world settings [44, 50]. It may not be feasible to study various types of HIT innovations across diverse health care settings at the same time; it may be costly, implementation and development processes may be at different stages or at different time points; and important details may be missed in the context of a large study [37]. Therefore studying one type of HIT innovation that is already in progress provides a good opportunity for learning about the real-world challenges and solutions.

One example of a HIT innovation that has recently received increasing attention is data harmonisation (DH). DH aims to address the problem of lack of interoperability between various databases across the health system, viewed as solely a technical problem requiring a technical intervention [20, 68]. DH innovations are meant to enable one or more systems to exchange information, to enable systems to use information in outputs or products, and to enable information users to easily access and retrieve a wide range of data. In the context of disease management, such innovations should enable information users to extract individual patient information and longitudinal treatment trajectories.

The potential benefits of DH are related to data synthesis, integration, organisation, and availability. However, it cannot be assumed that there will not be social challenges while technical processes of DH innovations are being developed and implemented. Interactions between technical factors and social processes may lead to unintended consequences [42] which may, for example, be reflected in a sudden increase in governance over existing technical infrastructure for use in new HIT innovations. In addition to poor technical fit between existing technical infrastructure and new HIT innovations, DH processes may involve ongoing negotiations between DH innovators and other HIT actors. DH innovations, like the HISs they are trying to strengthen, are embedded in complex health systems, which are affected by a range of social and technical factors. **Figure 1** illustrates that social and technical (or sociotechnical) factors and interactions are made up of social systems, technical and physical infrastructure, existing HIT (which are in use) and new HIT (which are being developed and implemented). These factors are all brought together in the Interactive Sociotechnical Analysis (ISTA) approach [42].

**Figure 1.** Interactive Sociotechnical Analysis framework [42]



The ISTA framework was derived from different perspectives on social and technical interactions that emerge out of HIT innovations. It specifies important relationships between new HIT innovations, workflows, health system actors and organisations. The framework emphasises the recursive and iterative nature of these relationships and their potential for producing unintended consequences [42]. It covers five sociotechnical dynamics that are of importance to this study. Firstly, the ISTA approach recognises that new HIT innovations have the potential to change prior workflows, communication and relationships, because changes emerging from interactions between existing HIT and new HIT innovations may lead to unintended consequences. Secondly, there may be a poor fit between new HIT innovations and existing technical and physical infrastructure, so this lack of interoperability may pose challenges (such as poor decisions, delays, data loss, errors and unnecessary testing) to new HIT innovations. Thirdly, HIT users may reinterpret the purpose of new HIT innovations, which may lead to different understandings and data practices from those intended by HIT designers, in order to deal with ‘unforeseen’ outcomes, workflows and stakeholder relationships. Fourthly, the development and implementation of new HIT innovations require recursive processes that determine their usability. Lastly, social systems (such as people and organisational structures) may drive the process of recursive change in that HIT users’ reactions and their local adaptations of new innovations may move away from original conceptualisations, so HIT designers may be forced to reconfigure some HIT features [42].

The ISTA approach is useful because it incorporates the key social and technical elements that impact on the development and implementation processes of DH innovations (as listed above). The technical elements relate to existing technical and physical infrastructures, including already established HISs, that facilitate or hinder the development and implementation processes of DH innovations. The social elements relate to HIT stakeholders who already manage and/or use existing systems through specific workflows within their organisational context. Beyond identifying these elements, the framework is also useful for explaining the relationships (and activities) between various elements of DH innovations. One of the key arguments of the framework is that interactions are repeated until changes gradually occur in pathways to data production and organisational cultures of data use. In this model therefore, DH innovations would be developed and implemented through iterative processes between DH innovators and various HIT stakeholders.

There are currently DH innovations underway in the Western Cape Province of South Africa. A DH initiative called the Provincial Health Data Center (PHDC) has started to align, synthesise and integrate various health-related databases across different types of health services and organisations. Its aim is to provide informational support to researchers, health managers and policy-makers relevant for epidemiological analyses, disease surveillance and clinical and service management. The new DH initiative presented a unique opportunity for us to learn more about the sociotechnical factors of DH innovations. The aim of this research was to identify and examine (a) the institutional and conceptual dilemmas that affected the operationalisation of the DH initiative; and (b) identify the key success factors in the DH initiative navigating these dilemmas, in terms of its leadership, institutional positioning, and operational and relationship-building activities.

## **2. Methods**

### **2.1. The Provincial Health Data Center (PHDC)**

The setting for this research was the Provincial Health Data Center (PHDC) which was officially started in 2015, through a collaboration between a provincial health department and a university research unit. PHDC is an emerging DH initiative which aims to provide informational support to researchers, health staff and managers and policy-makers (who we identify as data users) for health management decision-making.

The need for DH innovations emerged from research projects and health programmes that the founding director of PHDC was involved in. Data users such as researchers and health service practitioners required harmonised individual-level data for epidemiological analysis, disease surveillance, clinical and service management, and operational reporting. Harmonised, individual-level data refers to the integration of an individual patient's demographic, clinical, laboratory and pharmacy data across unconnected databases. Routine and electronically harmonised individual-level data were not previously available because of the absence of a common patient identifier that could be used at different health facilities across the province.

The introduction of a system called Clinicom made it possible for individual-level data to be linked. Clinicom electronically allocates individual patients with unique identifiers that can be used universally at health facilities within the Western Cape Province to admit, discharge and transfer, as well as to schedule appointments and maintain records. PHDC is making use of new technical opportunities such as Clinicom to gradually curate all available individual-level data into a province-wide data repository to enhance data extraction, data linkage and data transfer.

PHDC is made up of a team of individuals (PHDC staff or DH innovators) with expertise in public health research, clinical practice and information technology (IT) who are working on DH innovations such as the data repository. The aims of PHDC are to integrate different types of data derived from various routine electronic databases in the province; to develop an electronic single-viewer that enables health care workers (HCWs) to view patient-level information (such as

demographics, clinical, history and treatment initiation); and to provide informational support for epidemiological analyses and routine operational reporting.

PHDC is still an emerging initiative and it is not yet fully institutionalised (where DH processes are standardised and routinised) into the provincial government structure but has gradually grown to provide informational support to a wide variety of research and health services stakeholders. PHDC is a unique initiative in South Africa; harmonised individual- or patient-level public health services data were not previously available, so frontline health workers could not easily and timeously track patients enrolled in different types of services (such as diagnosis, treatment and care) over a long period of time (such as TB and HIV patients on treatment, or pregnant women in antenatal care) across different health facilities in the province. Additionally, PHDC is able to provide researchers, health managers and policy-makers with the necessary informational support to conduct epidemiological analysis of the effectiveness of health programmes on individual and collective patient outcomes or to assess the performance of costs and service delivery in relation to available human and financial resources [111].

PHDC is positioned between a government department and a research unit. Its key stakeholders (and those relevant to this research) can be grouped into five categories: (a) HIT actors (for example, database managers) who were involved in the design, development, implementation and/or management of databases, such as updating database features and controlling access; (b) HIT implementers and/or users who were involved in providing or managing health services (for example, clinicians, public health specialists and health managers); (c) business analysts who used routine data related to finances and human resources to evaluate health service performance; (d) provincial and national policy-makers who required synthesised evidence to make decisions in hopes of improving health services; and (e) researchers who requested access to public health services data for study analysis, so as to provide recommendations to improve the effectiveness of health programmes.

## 2.2. The i-ALARM study

Our research is situated within a larger study— *Using Information to Align Services and Link and Retain Men in the HIV Cascade*—or ‘iALARM’, which seeks to ‘raise the alarm’ about poor linkage and retention of men in HIV-related services. The iALARM study aimed to use routinely collected health information to more effectively coordinate the work of health system and community stakeholders. This was done by establishing a task team consisting of health system and community stakeholders who participated in the development and piloting of a health information management (HIM) intervention [123]. This HIM intervention incorporated findings from a retrospective cohort study that was conducted to quantitatively assess patient outcomes as they engaged with HIV testing, treatment and care services. The retrospective cohort study relied on the integration of electronic HIV-related data found in multiple clinical, laboratory, pharmacy, and mortality databases across the Western Cape Province of South Africa.

In 2015, i-ALARM researchers (including the lead researcher, BS) began actively engaging with health information staff and managers across the province, for example, data clerks, health information staff and managers and DH innovators. They primarily did so for the data collection process of the retrospective cohort study and to identify individuals to include in the task team of the HIM intervention. The director of PHDC (who was also a co-investigator on the i-ALARM study) was identified during the conceptualisation of the study as a potential vehicle for gathering the cascade data. Collaborative work between i-ALARM researchers and DH innovators to collect relevant data for the retrospective cohort study (that is to build an electronic HIV cascade) also provided i-ALARM researchers with a unique opportunity to learn more about DH innovations. This continuous engagement presented BS with an opportunity to conduct a parallel qualitative study to examine the challenges and strategies associated with operationalising and routinising DH processes at PHDC. There is a separate paper (Chapter 5) that reflects on our experiences of trying to gather data for building the electronic HIV cascade and that paper is written from a more experiential and participatory point of view. However, this paper specifically looks at PHDC in the broader context and through a series of additional interviews and observations.

### **2.3. Data collection and analysis**

This research contributed to the doctoral studies of BS and the research was approved by the University of Cape Town Human Research Ethics Committee (HREC ref: 320/2015 and 738/2015). This study aimed to identify and examine institutional and conceptual dilemmas and design, leadership and negotiation strategies associated with DH processes (operational activities and social relationships). BS conducted data collection using ethnographic research methods, including participant observation and interviews. Ethnography allows researchers to study (observe and/or interact) with participants in their real-life environment using methods such as participant observation and face-to-face interviewing [124, 125]. Participant observation was the primary method of data collection and was conducted between August 2016 and September 2017. BS attended and participated in meetings between PHDC staff, and research and health services stakeholders; work-in-progress PHDC staff meetings; and iALARM task team meetings (a total of twenty-two meetings, two to three hours each). She also conducted eleven in-depth interviews (approximately two hours each) with PHDC staff, researchers and public health specialists working with PHDC and national database managers. Additionally, BS participated and observed in the data collection process of the i-ALARM retrospective cohort study. Data collection for the retrospective cohort study required a DH process, specifically for building an electronic HIV cascade. The lead researcher conducted a qualitative study alongside that DH process, so some of the data collected then (between February 2016 to December 2016) also contributed to the findings of this paper. She was both a participant in the iALARM-PHDC project as well as an independent researcher, which had its own strengths and limitations for the study; more will be explained in the Discussion section.

Raw data (collected through participant observation and interviews) were in the form of meeting notes, field notes and 'diary-type' recordings of the researcher reflecting on specific aspects of the research. Data were analysed using an iterative process of intermittently identifying broad topics emerging from the data while data collection was still underway; authors identified new questions or areas of clarification from these broad topics. Thematic data analysis was applied which entailed BS manually highlighting (coding) recurring and prominent themes. The authors

conceptualised three main thematic areas as findings; through continuous discussions they ensured that emerging themes and stories were coherent.

### **3. Findings**

#### **3.1. Overview**

The findings are organised according to three broad themes. The first two relate to the institutional and conceptual dilemmas that were revealed in interactions between PHDC and other HIT stakeholders. PHDC staff engaged with various HIT stakeholders for the purpose of integrating local databases into the province-wide data repository as well as to provide data users with access to harmonised data. The third relates to the key success factors of PHDC navigating institutional and conceptual dilemmas; its success was underpinned by its leadership, its institutional positioning and the activities it engaged in to institutionalise itself (that is, to become absorbed into the health authorities).

#### **3.2. Institutional dilemmas of DH innovations**

This section will explore the institutional dilemmas that PHDC faced while trying to operationalise the data repository and while in the process of becoming fully institutionalised as a government data support structure. For the purposes of this paper, institutional dilemmas are the challenges and tensions that emerged from sociotechnical interactions between existing institutional data stakeholders, systems and processes and new DH innovations being developed and implemented in our local setting. Below, we will describe four institutional challenges. Firstly, fragmentation of databases between levels of the health services and the City and Province health authorities posed a challenge to PHDC. Secondly, tensions emerged between PHDC and other HIT stakeholders because of centralised versus decentralised approaches to data access and governance. Thirdly, lack of technical interoperability posed a challenge for secure and efficient data exchange between PHDC and the health authorities. Finally, different understanding about the purpose and power of PHDC led to tensions between PHDC and other HIT stakeholders which posed a challenge for PHDC to become more accepted within the HIT environment.

### **3.2.1. Fragmentation of routine electronic data managed by different institutions**

One of the key challenges that PHDC faced while trying to implement DH innovations, such as the PHDC data repository, related to the fragmentation of routine electronic data. The routine data that PHDC needed for building the data repository was dispersed across the health system in two ways. One, fragmentation occurred between levels of the health services and health authorities. And two, fragmentation resulted from previous approaches of databases being developed according to the characteristics and needs of specific health programmes; and development processes were influenced by donor funding. These aspects were reflected in the challenges that PHDC faced in implementing DH innovations.

Fragmentation between levels of the health services and health authorities was an expected dilemma since DH innovations at PHDC were specifically trying to harmonise databases at those different levels. PHDC knew that the databases were fragmented, but the added layer of complexity was that the databases were embedded within different institutions that were not always working together efficiently or effectively. The setup of data systems between City and Province in Western Cape Province was distinct. One of the challenges that PHDC staff faced involved them needing to go to different institutions to access the routine data that they required to build the PHDC data repository. It was an added challenge for PHDC that the City and Province health authorities offered the same services in the same area in parallel to each other, often without sufficient coordination, which sometimes led to duplications. Lack of coordination between the health services spilt over to lack of integration between databases and PHDC needed to engage with different health services to access individual databases for inclusion into the province-wide data repository. Another challenge that PHDC faced could be considered as built-in resistance of the different institutions. The two health authorities, laboratory services and other government departments operated separately from each other and were not used to initiatives such as PHDC that collaborated across institutional boundaries.

Additionally, the organisation of databases in the local setting were influenced by the needs of specific disease programmes and the agendas of donor funders. There were no uniform standards (beyond those required for minimum data sets built earlier on) for building databases so what data

people collected and how the data were collected varied between and within provinces. Many databases were developed to feed into the assessment of provincial and national targets, and thus may not have included all the information that might be useful programmatically. The National Department of Health (NDoH) generally laid out health policies and programmes for the provinces to implement without providing standardised guidance. There was flexibility in how the provinces could implement the health policies and programmes as well as in how provinces could develop suitable databases that would provide the necessary data for monitoring and evaluating programmes and services.

Databases for many vertical health programmes, particularly for HIV and TB, were initially set up by local health authorities in collaboration with foreign donor organisations rather than by National government. Databases were developed to support specific individual programmes and the content of the databases was driven by donor reporting requirements. It was also more difficult than usual for other health system stakeholders to access the data collected in those databases because donors strictly controlled access or because donor and health system IT people could not easily coordinate data exchange (they were not in the same place and data collection elements differed). Local health authorities sometimes hired IT partners from overseas (associated with donor funders) or from private companies to develop the programme-specific databases.

In addition to the database challenges posed by the lack of national data standards and the influence of donor funders, local database managers have sometimes also struggled to make significant changes to databases without the input of the original developers. Given that most programme databases were developed with donor funding and involved external IT partners, local database managers were unable to modify database features without requiring additional financial and technical resources to make the desired changes to the databases. For PHDC, the challenge was that they did not have control over the technical features of the databases. PHDC staff had to learn more about the individual databases and make do with the data that they could access and use; this led to DH processes being implemented gradually and in an iterative manner rather than in a consistent or once-off manner.

In sum, DH processes were started to address the problem of fragmentation between databases due to divisions between levels of the health services and the authorities. Parallel databases and built-in resistance for institutions to work together were experienced as challenges for building the PHDC data repository, this led to PHDC spending time and resources working across the different levels and institutions, developing relationships with stakeholders and iteratively operationalising DH processes. Additionally, the influence of donors on the development of databases and reliance on external IT partners also caused problems for PHDC to access and modify database features for copying relevant data into the repository.

### **3.2.2. Tensions between centralised and decentralised approaches to data access and governance**

Tensions also emerged between PHDC and other HIT stakeholders because of differences in approaches to data access and governance. PHDC's approach was to have a centralised data repository for the province. However, this approach was in contrast to the existing approach where data access and governance processes were decentralised. Differences in approaches to data access and governance were reflective of the different kinds of relationships that PHDC and HIT stakeholders had to databases.

On the one hand, the general sense at PHDC was that there was a need for centralising access to and governance (management and practices) over routine data. PHDC imagined that it would be able to do so by harmonising data from different sources in some way (although not exclusively) in the form of a data exchange platform such as the data repository. It would manage the data repository, help data users access the data that they required, as well as develop standardised data extraction and data transfer processes. PHDC was dealing with data access and governance as a custodian; its aim was to centrally safeguard data on behalf of the health services and make it accessible to data users. On the other hand, from the view point of PHDC, advocating for changing the status quo around ownership and management of databases was going to be difficult as the other HIT stakeholders seemed satisfied with existing processes for managing databases. Even though National government paid for the development and implementation of databases, there were no standardised requirements for producing and reporting data. HIT stakeholders

viewed themselves as the owners of databases; each database manager put in place systems to produce data and govern access to it.

An example of the differences in approaches to data access and governance can be seen when PHDC and the City authority were discussing the terms for linking a “live” version of the City database into the new PHDC data repository. Discussions on how to go about linking an up-to-date City database to the data repository often centred on the ownership of databases; the City health authority viewed itself as the owner of the City database and controlled access to the database. PHDC’s plans posed a challenge to the existing data ownership and governance system because the data repository aimed to combine all individual-level data (including City-specific data) and centralise it for the purposes of DH. This meant that PHDC would obtain copies of all electronic individual-level data and it would be archived and integrated into the data repository. The City applied a more decentralised approach, however, where City managers aimed to control access to the data; Province managers could not easily access City data without lengthy approval processes or data exchange negotiations.

Earlier we described how fragmentation between health services and health authorities led to the need for DH, and how the implementation of programme or disease-specific databases and donor funding impacted on DH activities at PHDC. Here the challenges that PHDC faced were the result of contrasting approaches to data access and governance. The next institutional challenge that we describe below flows from the previous ones—the lack of ongoing working relationships between City and Province health authorities.

### **3.2.3. Lack of ongoing working relationships between City and Province health authorities**

Lack of ongoing relationships and technical interoperability for secure and efficient data exchange between PHDC and HIT stakeholders was a third institutional challenge that impacted on the implementation of DH innovations. Difficulties in the discussions between PHDC and HIT stakeholders to access data in a routine manner reflected a lack of ongoing working relationships between different health authorities. Although there was general agreement about working

together, there were no established procedures for doing so. For example, there was no technical platform through which PHDC and City health authorities could share individual-level data. They had to jointly identify how they could exchange data in an efficient manner, while safeguarding the confidentiality of named health services data. PHDC staff felt that lack of technical interoperability was the surface-level challenge to operationalising the data repository; the real challenge was the lack of ongoing and formalised working relationships between the authorities, which resulted in lack of institutional collaboration. The challenge of figuring out how to share and exchange data became a platform for working out relational issues.

For example, while PHDC staff and City managers were still figuring out the terms for data exchange, they in one instance had to use a USB flash drive and in another, a makeshift File Transfer Protocol (FTP) setup as a workaround to the issue of data exchange. In the first instance, a City representative delivered a USB flash drive to PHDC which a PHDC staff member was subsequently going to return with extracts of Province data that City database managers had requested. In the second instance, a FTP was set up as a temporary mechanism to transfer data files between the City health authority and PHDC using a computer network. The USB and FTP workarounds not only reflected the lack of routine technical interoperability between the two parties, but also hesitations about the PHDC data repository. It was the general perception amongst PHDC circles that by providing PHDC with retrospective data first, the City health authority was testing how the data were going to be used in DH activities before they made any commitments to engage in more routine and sustainable exchanges. This test resulted in PHDC and City staff having to use inefficient ways of exchanging data as a technical solution. The use of temporary mechanisms was also an opportunity for City and Province health authorities to slowly build trust and navigate existing institutional boundaries.

#### **3.2.4. Tensions that emerged as a result of the purpose and power of PHDC's data repository project**

The last institutional challenge has to do with PHDC trying to gain acceptance within the HIT environment. Tensions emerged because some HIT stakeholders did not see the purpose of PHDC as a DH initiative and data support structure, and some perceived it as threatening to their existing

roles, data processes and autonomy. PHDC faced resistance (to varying degrees) from stakeholders at different levels of the health services, including local management (City and Province database managers), national management (National database managers), and district hospital managers. Resistance towards PHDC not only delayed the technical implementation of DH innovations, but also pushed PHDC staff to devise strategies midway implementation to increase PHDC's relevance and to ensure its acceptance (described in a later section).

There were parallel plans amongst local database managers to strengthen data production and data utilisation across health programmes and departments in the Western Cape; similarly, the National government planned to develop DH innovations for aggregate-level and disease-specific databases. PHDC was advanced in its plans to perform DH processes compared to these other initiatives; its power lay in its ability and readiness to start implementing DH processes. There were two issues related to PHDC. One, PHDC had the ability to efficiently organise itself into a DH initiative which was perceived as threatening by both local and national stakeholders who were also planning to develop DH innovations. And two, some stakeholders were hesitant about the purpose of PHDC because they could not understand the innovative element that PHDC was bringing when compared to their own plans to conduct DH. These were some of the challenges that PHDC faced while trying to complete the data repository project.

From the perspective of PHDC staff, the process for accessing local databases was generally more complex and drawn out than they had anticipated. Local database managers were initially uncertain about how to respond to PHDC requesting to access their databases for linkage to the data repository. They were dealing with a new and unusual health information initiative that was requesting to have direct technical access on a large scale for the purpose of harmonising individual-level data across various databases. Database managers had several concerns. Firstly, database managers were concerned about whether PHDC had the appropriate institutional and legal permissions, especially since PHDC occupied an unusual space between the health authorities and the university. They were also concerned about whether PHDC had the technical capacity to effectively execute such a complex task, and the safeguards to ensure the data remained protected. Finally, the database managers were concerned about who would gain access to this newly centralised data repository and how access would be controlled.

Tensions related to the purpose and power of PHDC extended beyond the local management level. PHDC had to consider national level issues as well, so the director of PHDC tried to engage with National HIT managers to keep up to date with their plans as well as to share the lessons that were emerging from operationalising DH innovations at the provincial level. Even though National managers did not regularly interact with PHDC (for example, by being part of the PHDC meetings), they were aware of the emergence of PHDC. Like with managers within the province, there was a general sense of confusion amongst National managers about the role of PHDC in terms of what unique aspect it was bringing given that DH innovations were already planned at national level. Within PHDC circles, there were views that National managers were maybe resistant towards PHDC because they lacked clarity on the differences between National and PHDC's operational mandates and agendas. PHDC's unique focus was on harmonising individual-level data and facilitating data access for research and health service stakeholders in the province. The perception was that were focused on producing data useful for assessing national performance targets; hence they did not consider the DH activities at PHDC to be of importance to them and they felt that PHDC was concerned with research projects or small-scale projects exclusive to the Western Cape.

At both the Province and National management levels, the focus was on ensuring that PHDC adhered to existing institutional processes for obtaining data access and safeguarding confidential health services data, and that PHDC did not interfere with high-level DH innovations. At the lower level of the health system, concerns were about PHDC reinventing the wheel. For PHDC, district hospital managers were an important stakeholder group. However, some district hospital managers were under the impression that PHDC was replicating data processes. For example, when PHDC was trying to link hospital data to the data repository, district hospital managers thought that PHDC was merely storing copies of their data. They did not fully understand that the data repository was being built so that individual-level data could be harmonised into useful formats that they could benefit from such as the electronic single viewer (described earlier).

In addition, district hospital managers sometimes thought that the PHDC data repository was replicating their data collection processes. They were also initially resistant towards the data repository because of PHDC's approach to centralised data management, that is by having copies

of all the data in one place for exchange and use by different stakeholders. PHDC stakeholders and staff who interacted with district hospital managers got the sense that district hospital managers were resistant to their data being accessible centrally. District hospital managers were concerned about wasting time and resources that they had invested in setting up workflows to collect the programme and hospital data that they needed to assess their performance against local district health targets.

There are important lessons that can be drawn from the challenges that PHDC faced when it tried to gain acceptance amongst other HIT stakeholders. Firstly, resistance towards an emerging HIT innovation may be for different reasons; for example, at the management level concerns were about legal permission for data access, safeguards for protecting health services data, and provincial DH projects not interfering with high-level national plans. At the lower level, concerns were around PHDC replicating and replacing existing systems and workflows that district hospital managers had invested time and money into developing. And secondly, resistance against a HIT innovation may reflect something about its purpose and power that is unclear or is perceived to be threatening to existing stakeholders, systems and processes.

To summarise, this section described four institutional challenges that PHDC faced when trying to operationalise DH projects (such as the data repository) and becoming fully institutionalised as a data support structure based within the Province health authority. The start of the institutional challenges begins with the fragmentation between routine electronic data caused by the lack of standardised approaches and influence of donor funders on data production and reporting requirements. Other institutional challenges have to do with tensions between centralised and decentralised approaches to data access and governance; lack of ongoing relationships and technical interoperability for secure and efficient data exchange between PHDC and the health authorities; and tensions that emerged in response to the differing views of the purpose and power of PHDC's data repository project.

### **3.3. Conceptual dilemmas of DH innovations**

The previous section was about the tensions that emerged when PHDC interacted with various HIT stakeholders in the context of existing social and technical systems and processes of different institutions. PHDC also faced conceptual dilemmas while trying to complete the data repository and in the process of becoming fully institutionalised as a province-wide DH initiative. This section will explore two kinds of conceptual dilemmas that emerged. One, there was tension between a scientific versus pragmatic approach to data; and two, there was resistance towards formal procedures for data requests and data transfers because of misconceptions that these procedures were mechanisms through which PHDC tried to monitor other HIT stakeholders' data practices.

#### **3.3.1. Tensions between scientific and practical conceptualisations of the value of data**

Tensions emerged between PHDC and data users because of their different conceptualisations of the value of data. PHDC was focused on providing high quality data, while health service or programme managers (referred to as frontline clinician managers) were merely looking to access 'good enough' data. For frontline clinician managers, good enough data were any routine data that was immediately available to support decision-making.

Contrasting perceptions about the value of data (in terms of data quality and data usability) are common especially when a new innovation is being implemented [53, 54]. This was evident when disagreements between PHDC staff and frontline clinician managers emerged during the early pilots of DH projects. For PHDC, the pilots were important not only to support the work of frontline clinician managers, but also to show other HIT stakeholders its value. Firstly, PHDC was setting up new data systems and practices. PHDC needed to be cautious in operationalising DH processes and products because it was being monitored by resistant stakeholders. For example, some stakeholders were sceptical about PHDC handling public health services data; PHDC used pilot projects with frontline clinician managers to show that ethical data practices were in place (such as anonymising data for research purposes). Secondly, PHDC wanted to market its service to HIT stakeholders by demonstrating its ability to conduct secondary data

production, for example, cleaning datasets and adequately linking them, transforming data into different formats (such as reports, tables and applications) and password protecting the data before transferring it.

Frontline clinician managers (were interested in the pilot projects at PHDC because they saw it as an opportunity to access newly harmonised data (from a central place), which they assumed would help them address urgent clinical and operational needs. They did not anticipate that PHDC would base the value of data on high scientific standards of accuracy and completeness and engage in data transfer processes to safeguard the data. There was a general perception that DH processes at PHDC were complex (many stakeholders did not understand the purpose of DH innovations), so the focus on data quality at PHDC was seen as an added layer of complication.

Thus far we have described the tensions that emerged between PHDC and data users because of scientific versus practical conceptualisations of the value of data; these tensions were related to the iterative nature of implementing DH processes and the lengthy timelines and multiple steps involved in completing data requests. A key lesson has to do with how DH innovators and data users may conceptualise the value of data differently, which in turn shapes their expectations of DH processes and products. PHDC valued data according to its quality (accuracy and completeness) while data users valued data according to its usability (feasibility and practicality). Hence, PHDC staff anticipated that it would take time to complete the multiple steps involved in preparing good data, while data users expected a short turnaround time to access useable data, especially because they knew it was already stored in the PHDC data repository. Below, we describe the perceptions that some HIT stakeholders had of the formal data request and data transfer processes (referred to above) that PHDC introduced.

### **3.3.2. Misunderstandings about the purpose of formal data practices**

DH innovations are first conceptualised, then developed into operational steps, and finally implemented through standardised procedures for conducting DH processes and generating DH products. Tensions may emerge between DH innovators and HIT stakeholders when they each have different conceptualisations of the use or purpose of formal data procedures. For example,

PHDC introduced formal procedures for data requests and data transfer processes while in the process of piloting the data repository. PHDC viewed formal procedures as a mechanism to institutionalise itself; the aim was to show that there was structure and formality associated with its data practices. Some HIT stakeholders, particularly those who were also based within the Province health authority, disagreed with PHDC's formal approach and generally questioned the purpose of the formal procedures. When PHDC started, formal data request and data transfer procedures were not yet established. Initial engagements between PHDC staff and health managers were informal; health managers would talk to PHDC staff about the data they needed and PHDC staff would share the data with them. PHDC was flexible about data requests at the time, but in order to become institutionalised and legitimate, it needed to establish formal procedures (such as introduce data request forms).

Many people outside of PHDC became concerned and felt threatened by PHDC; firstly, because they were used to more flexible systems than what PHDC was introducing and so the formal procedures were seen as a hassle. Some HIT stakeholders felt that PHDC was enforcing higher (scientific) standards for handling data through formal procedures without considering that health managers (who are data users) had previously used routine data not transformed at PHDC to make health management decisions. The second complaint was that these formal procedures might be used to hold HIT stakeholders accountable for certain data practices. The perception amongst HIT stakeholders was that PHDC would monitor how they shared data with each other (including with PHDC) and compare that against the new formal data transfer guidelines. PHDC would then be able to reference the guidelines (as a way of holding them accountable and correcting them) if HIT stakeholders reverted to informal procedures. Unlike PHDC, the HIT stakeholders who were opposed to the formal procedures were already institutionalised, but PHDC (as a new initiative) was using the formal procedures to demonstrate that it was responsibly performing a needed service.

To summarise, this section described two conceptual dilemmas; one, tensions emerged because of differences in scientific versus practical approaches to data, and two, formal procedures implemented at PHDC were perceived as a threat by health managers who used to interact with PHDC staff through informal channels. An important lesson is that DH innovations are not only

affected by institutional factors (such as lack of ongoing relationships between institutions), but also by people's thoughts about why data is important and how it should be used. Conflicts may emerge when HIT designers (including DH innovators) are not in agreement with other HIT stakeholders (such as data users) about the state of available data and how it can be transformed for use (for example, in terms of the timeline and processes required).

### **3.4. Navigating institutional and conceptual dilemmas**

Even though there were quite a number of factors challenging PHDC, there were other factors that helped PHDC function effectively and overcome these challenges. This section will provide a summary of the key success factors of PHDC in navigating the institutional and conceptual challenges described above. These factors have to do with the emergence of a DH champion (and then mini champions), PHDC being located between a government and university, and PHDC establishing formal procedures and engaging with the different stakeholders.

The first aspect that helped PHDC navigate the institutional and conceptual challenges that it faced has to do with its leadership. PHDC's leadership was flexible, multi-skilled and distributed across different tasks. There was a DH 'champion' and a team of 'mini-champions' who had suitable traits to navigate both expected and unexpected challenges. Firstly, the person who initiated DH innovations and founded PHDC played a key role as the champion. His role was to guide the design, development and implementation of DH innovations that responded to health services and research stakeholders' needs for harmonised data. The idea to conduct large-scale DH processes was born out of the PHDC's director's experiences as a clinician, researcher and public health expert needing harmonised data to do his work. He was knowledgeable about local databases; for example, he was involved in the development and implementation of a tiered paper-electronic system and TB and HIV-specific databases. Additionally, the director of PHDC relied on relationships that he had established with technical staff and managers of local databases and other stakeholders. Even though it was not a deliberate strategy, his unique position as an employee of the Province health authority and the university was an added advantage that he drew on to mobilise financial and operational resources from both institutions.

Secondly, ‘mini-champions’ emerged in the form of PHDC staff who were well-skilled to work on different technical tasks and to engage with various HIT stakeholders. PHDC staff were responsible for operationalising emerging ideas related to DH processes and products; one of their primary tasks was to build the data repository. PHDC staff were identified as mini-champions because they had the technical skills to implement DH innovations, and one way in which they did this was to become specialists in different tasks. They continuously shifted between technical tasks (such as programme coding) and setting up administrative procedures (such as writing up forms, capturing data requests and documenting outputs). In addition to their skills, PHDC staff also had personal interests and motivations that contributed to the progression of PHDC; they enjoyed working on a wide range of tasks and they saw it as an opportunity to merge their new career interests with their academic background and previous work experience. They felt a serious responsibility towards patient data, so they aimed to use patient data correctly to improve patient outcomes and health service delivery. PHDC staff were generally passionate about working in a consistently changing environment where the director expected them to deal with emerging challenges by adapting existing skills or learning new ones, and they felt that they received positive recognition.

The second aspect has to do with PHDC being located between a government and university. The director of PHDC’s dual employment was the initial reason for where PHDC was located, but it became an important strategy for PHDC to remain there. On the one hand, setting up PHDC within the Province health authority provided it with a sense of legitimacy. This way HIT stakeholders who were uncertain about PHDC or resistant to it could more frequently interact with it about its purpose. On the other hand, PHDC could continue working with HIT stakeholders at the local university to pilot and scale-up DH projects. Related to this aspect is the issue of PHDC needing to become institutionalised, that is, it needed to devise ways of engaging with the different stakeholders and establish formal data procedures.

The lead of PHDC was aware of the social and political nature of data ownership, access and use, and he knew that it was going to be a challenge to move from conducting small-scale DH processes (primarily conducted in the context of research projects) to supporting larger health services projects. PHDC therefore introduced formal data procedures at the same time that pilot

projects were underway. As mentioned before, PHDC needed to show that it had a proper governance structure to guide various administrative and technical processes, even though pilot projects were still underway. At the beginning, PHDC staff communicated directly with data users who were involved in pilot projects about the data that they required. However, some HIT stakeholders felt that PHDC was merely a research project and it lacked novelty or capacity to deal with high-level health services data problems. Standardising and routinising data procedures was one way for PHDC to demonstrate that it had moved from a pilot or an innovation to an institutionalised service. Formal procedures were introduced so that PHDC could track the number of new, ongoing and completed data requests and to show how it was dealing with different types of data requests. PHDC was also able to develop new systems to demonstrate that it was prioritising internal requests from health services stakeholders over external requests from university researchers. Formal procedures were therefore to help PHDC become institutionalised as well as to ensure accountability with stakeholders.

## **4. Discussion**

The case of PHDC was useful for examining the institutional and conceptual dilemmas that affect the operationalisation of DH projects and for identifying the key aspects related to leadership, institutional positioning, and operational and social activities for navigating such dilemmas. The Western Cape Province presented a unique set of challenges for PHDC to harmonise data across different levels of institutions and build relationships and operationalise DH processes. Tensions between centralised and decentralised approaches to data access and governance and lack of ongoing relationships and technical interoperability between PHDC and the health authorities caused a delay to the PHDC data repository project.

To overcome some of these challenges, PHDC invested more time and resources than anticipated in devising temporary solutions and mechanisms to navigate institutional boundaries and build trust amongst HIT stakeholders and data users. Another tension was about PHDC's purpose to build the data repository; this was perceived as threatening to existing roles, data processes and autonomy. Two conceptual dilemmas emerged while DH projects were underway. One, differences between scientific and practical understandings of the value of data resulted in

conflicting expectations of DH products and processes. Two, formal data procedures implemented at PHDC were perceived as threatening to data users who previously interacted with PHDC staff through informal channels. Overall, these dilemmas revealed an important lesson about DH innovations; they are not only affected by institutional factors but also by people's conceptualisation of the value of data.

We only found a few conceptual frameworks and primary studies that highlight similar social challenges of HIT innovations as the institutional and conceptual dilemmas described above [37, 42-44, 122, 126]. From our knowledge, this is the first study evaluating the social challenges of a DH initiative in a LMIC setting. Additionally, there is a scarcity of literature on key factors for navigating social challenges of HIT innovations. Three important lessons can be learned from the case of PHDC about the key success factors of HIT innovations. One, for HIT innovations to be successful, they need to align to real-world needs, they should be driven by a few individuals (champions) who conceptualised innovative ideas about how to solve data-related issues, while taking into consideration available technical infrastructure and making use of their networks and expertise. Two, to move from conceptualising a HIT innovation to implementing it, the team of HIT innovators (the leadership) needs to be comprised of individuals who are like-minded, multi-skilled and flexible enough to adapt their skills with the transition of HIT innovations. Three, there is a transitional phase from piloting to institutionalising a HIT innovation; introducing formal procedures can help clarify the HIT innovation's purpose and legitimise it.

A recent conceptual paper on the successful optimisation of large-scale HIT innovations [51] is useful for highlighting three key lessons. The first lesson is applicable to our findings on the role of leadership in navigating challenges of DH innovations. The authors state that for any HIT innovation to be successful it needs ongoing commitment from a few individuals to ensure that there are sufficient resources for implementation activities. The leadership team needs to have a good mix of clinical, informatics, analytics, and organisational skills and it needs to be able to use these skills in different ways [51]. In the case of PHDC, the leadership was able to draw on the government and university's financial and human resources. PHDC staff were also skilled enough to work on a wide range of tasks, from technical tasks (such as programme coding) to

setting up administrative procedures (such as writing up forms, capturing data requests and documenting outputs).

The second lesson on the successful optimisation of HIT innovations has to do with innovators engaging multiple stakeholders, for stakeholders to come to an agreement about what data will be collected, how it will be used and who will have access to it. This study states that agreement about the goals of the HIT innovation is important to ensure that data activities of stakeholders are aligned [51]. PHDC faced several institutional and conceptual dilemmas that were revealed when engaging the different City and Province stakeholders, especially when trying to introduce the province-wide data repository. One of the ways in which PHDC tried to navigate through those challenges was to keep its institutional position between the government and university; it could use research projects to pilot DH innovations and increase user demand, while introducing formal data procedures as a form of institutionalising itself in the government structure. Pilot projects and formal data procedures were also used to clarify questions around the data repository, how data were safeguarded, how it was used and by whom it was used.

The third lesson reported in the conceptual paper that resonates with our findings is about the evolving nature of HIT innovations; the study states that HIT innovation activities are best conceptualised as an ongoing process of improvement [51]. This means that HIT improvements need to be aligned with technical changes (for example, emerging innovations and refinement of existing functionalities) and social dynamics (for example, changes in stakeholder or user relationships). It is also important to pay attention to both formal or planned and informal or unplanned activities because these often identify or address existing technical and social issues associated with new procedures and systems. PHDC had to be open to unexpected discoveries about the databases that needed to be linked to the data repository; databases were historically developed for specific vertical programmes and to meet donor funders' agendas. This meant that PHDC staff had to familiarise themselves with individual databases; DH processes were therefore implemented gradually and in an iterative manner rather than in a consistent or once-off manner.

This study had two key strengths. Firstly, we were able to extend our research on DH innovations beyond the iALARM HIV cascade (the reason for initial contact with PHDC) and learn more

about the social dilemmas and key success factors associated with DH processes. Secondly, the lead researcher was in the unique position of being an iALARM representative at PHDC stakeholder meetings and also conducting qualitative research on DH innovations. This dual role allowed her to access otherwise restricted spaces within PHDC circles and identify potential research informants amongst PHDC stakeholders.

One potential limitation had to do with the lead researcher's close engagement with PHDC; she had to navigate the dual role of being a stakeholder and being a researcher and uphold the confidentiality and trust built amongst PHDC stakeholders. For data collection, for example, this meant that BS had to make a clear distinction between discussions or events that were for public consumption (that could be reported here) and those that were confidential matters between PHDC and iALARM or other stakeholders. Interviews with key informants (such as PHDC staff) were therefore conducted towards the end of participant observation to reassure trust and clarify what could be consumed for research purposes.

In conclusion, we described a diverse range of institutional and conceptual challenges related fragmented databases, different approaches to data access and governance, lack of institutional relationships, perceived power of innovations, scientific versus practical conceptualisations and perceived intentions of formal data procedures. We described three key factors for navigating the institutional and conceptual challenges which have to do with leadership, institutional positioning, and operational and relationship-building activities. These research findings may be of interest to individuals who are involved in DH innovations, for example, developers, implementers and users of such innovations. They may also be of interest to individuals (such as health system managers and social scientists) who are interested in strategies for introducing and institutionalising HIT innovations into complex organisational structures and existing technical infrastructure and social systems.

## **Contributions**

The PhD student (BS) conducted data collection and data analysis. Her PhD supervisors (CC and NL) provided guidance throughout the research process and provided critical feedback on the drafts.

# CHAPTER 7: DISCUSSION AND CONCLUSION

---

This PhD project examined various factors affecting data harmonisation (DH) (a type of health information technology (HIT) innovation) for strengthening health information systems (HISs). The thesis consists of four standalone findings chapters; one scoping review, one historical analysis and two ethnographic papers. Data were collected using literature searches, participant observation, interviews and document reviews. Jointly the findings chapters are about (a) the historical factors (events, barriers and facilitators) that impacted on health information system (HIS) interventions in the years after South Africa became independent; (b) definitions and concepts of data harmonisation in the literature and explanations of the relationship between DH and health management decision-making; and (c) a wide range of social factors (such as context, design, institutions, perceptions) associated with introducing new DH projects and institutionalising a province-wide DH initiative. In this final chapter of the thesis, I summarise the key findings of the thesis, which I then discuss in the context of existing literature. I draw out key lessons from the findings and provide recommendations for practice and lastly identify the gaps for future research.

## 1. Summary of findings

In chapter 3, I synthesised key historical factors such as events, barriers and facilitators, that emerged and influenced the development and implementation of the new HIS and HIS interventions after South Africa became independent. The key events that took place were the development and implementation of the Health Information System Program (HISP), the Minimum Dataset (MDS), the District Health Information Software (DHIS) and the National Indicator Dataset (NIDS). HISP was initially a pilot project to identify information needs and support new district management teams. HISP then developed datasets and software such as provincial MDS and DHIS. The National Department of Health later endorsed HISP as the main collaborator in developing and implementing datasets and software across all provinces in South Africa. The national government eventually decided to develop a national dataset (called NIDS) to collect standardised data items.

One barrier that unfolded as HISP was emerging was related to conflicts between HISP managers (HIS intervention developers) and HIS stakeholders (information technology (IT) and health services practitioners). They disagreed about how data processes should be operationalised, that is, in terms of which data elements should be prioritised and the general approach to data collection and reporting. Lack of funding to continue national rollout of HIS interventions and HISP not becoming institutionalised within the government structure was another barrier to HIS interventions. As a result, HISP managers intermittently stopped HIS intervention processes to work on other income-generating projects overseas. One way that HISP managers tried to overcome this challenge was by training health staff how to develop their own data management tools using basic computer skills. But, health staff did not develop data management tools after the training because of lack of time, capacity and support which ended up delaying new HIS processes.

On the upside, however, positioning HISP as a government-university collaboration made it possible for different stakeholders to participate in the initial development of HIS interventions led by HISP managers. HISP managers were regarded as neutral parties in helping to resolve disagreements between stakeholders who disagreed about the design of HIS interventions. They were regarded as change agents (or champions) amongst the HIS stakeholders because they introduced innovative HIS products and processes; also because they conducted research to assess the state of the HIS and to identify user needs; and they used earlier HIS intervention pilots as a testing ground for new interventions and to revise their design processes.

Chapter 4 is a scoping review on the definitions and concepts of DH and its relationship to health management decision-making. The review reports on three key findings. Firstly, the review included 181 studies and the majority of included studies were: primary studies; studies conducted in high-income settings and studies describing a DH intervention for frontline health management; and studies using the term 'health information exchange'. Secondly, a subset of included studies identified the following nine characteristics of data harmonisation: involves a technical and social process, involves at least two or more databases; involves different types of data; makes use of electronic data; aims to convert data to useable formats; integrates data within and across different departments and institutions; consists of different types of activities; makes

use of unique patient identifiers; and addresses a specific scope and purpose. The review also identified six alternative terms to describe a DH intervention or activity: record linkage, data linkage, data warehousing, data sharing, data interoperability and health information exchange (HIE).

Additionally, the review identified five studies that included a visual representation of DH components or processes. The models of DH illustrated three aspects of DH: one, a comprehensive approach to DH involves different types of data, different levels of the health care system (clinics and hospitals), multiple activities for exchanging data and an index for patient identifiers; two, the technical process of DH involves linking different electronic data into a central repository; and three, a DH intervention generally has a specific purpose such as medical reconciliation (updated list of patients' medicines) and disease outbreak surveillance. Finally, the review synthesised evidence on the relationship between DH and health management decision-making. The review identified nine studies that describe three levels of the relationship. The first level that DH contributes to is the clinical decision-making level, which involves individual patient clinical management or clinical support and quality improvement tools; the second level is the operational and strategic decision-making level which involves support for health system managers and policy-makers; and the third level is about population-level decision-making for disease surveillance and outbreak management.

The focus of chapter 5 is on the motivations and opportunities that contributed to the emergence of a DH and the design and operational challenges and responses of the new DH initiative. There were two opportunities that made it possible to operationalise DH innovations at the Provincial Health Data Center (PHDC). Firstly, the routine health information system (RHIS) in the Western Cape Province progressed to a point where a new patient identification system (Clinicom) could be introduced to link the key electronic databases in province. Secondly, the institutional structure and positioning of PHDC as a government-university collaboration made it possible for PHDC to establish closer ties of accountability with key stakeholders and it gave them the authority to engage with the health authorities directly. The positive attitude of frontline health staff also contributed towards the emergence of DH innovations.

Like other technology innovations, DH innovations in the Western Cape Province also faced design and operational challenges. Two examples of such challenges were revealed when iALARM and PHDC worked together to build an electronic HIV cascade. In that process they struggled to access the data they required in the City database and they also realised that electronic HIV testing data were incomplete. At the same time, PHDC had to balance tensions emerging from operationalising and institutionalising new data procedures. Two solutions involved PHDC introducing data access and transfer procedures and building and using social relationships to deal with these changes.

In chapter 6, I present three areas of findings: institutional dilemmas of DH innovations; conceptual dilemmas of DH innovations; and the key success factors in navigating institutional and conceptual dilemmas. PHDC faced four types of institutional challenges when it tried to operationalise a province-wide data repository and become fully institutionalised as a data support structure within government. The institutional challenges related to fragmentation between routine electronic data; tensions between centralised and decentralised approaches to data access and governance; lack of ongoing working relationships between the City and Province health authorities for secure and efficient data exchange; and stakeholders' response to the purpose and power of PHDC's data repository project. Additionally, PHDC faced two types of conceptual challenges related to tensions that emerged because of differences in scientific versus practical approaches to data, and formal procedures implemented at PHDC were perceived as threatening by stakeholders who previously interacted with PHDC staff through informal channels.

There are three success factors that helped PHDC navigate these institutional and conceptual challenges. The success factors relate to the emergence of a DH champion and mini champions; the positioning of PHDC as a government-university collaboration; and activities to formalise data procedures and engage different stakeholders. Later in this chapter, I extract key lessons emerging from the four chapters and provide recommendations in line with those lessons for researchers, health managers and policy-makers involved in the development, implementation and evaluation of HIT innovations.

## **2. Positioning the research findings in the literature**

### **2.1. The barriers and facilitators of data harmonisation innovations**

The thesis findings are well-aligned with the sociotechnical perspective identified and explained in the three conceptual frameworks described in introduction chapter (Chapter 1). The conceptual frameworks are the Performance of Routine Information System Management (PRISM), the Complex Adaptive Systems (CAS) and the Interactive Sociotechnical Systems (ISTA) frameworks [13, 42, 43]. The PRISM framework identified both the technical and social (organisational and behavioural) determinants of a RHIS [13]. The CAS framework explained the social complexities of health systems which HISs are embedded in, which means DH innovations are also impacted by this complexity [8, 43]. The ISTA framework focused on the unintended consequences of HIT innovations that occur when social processes and technology interact with each other [42, 44].

The barriers and facilitators of HIS and DH innovations identified in Chapter 3, 5 and 6 related to stakeholder engagement, funding and institutionalisation, the role of frontline or lower level health staff, government-university collaborations and the role of champions, overlap with findings in the literature. Two systematic reviews identify similar factors of HIE as those described in the thesis. The first systematic review assessed the barriers and facilitators to the implementation and adoption of HIE in low- and middle-income countries (LMICs) [19]. This review identified seven broad categories of factors that influence HIE; namely socio-political, financial, infrastructure, organisational, technical, individual and data management factors. The review particularly focused on the influence of these factors on implementation and future use.

The second systematic review that identified similar factors of HIE as those described in the thesis, assessed stakeholder participation (or engagement) in HIE, and the benefits, barriers, and overall value of HIE to primary care practices [30]. This review reported workflow, costs, and leadership and strategic planning as factors that impact HIE in primary care settings. The historical analysis and ethnographic chapters (Chapters 3, 5 and 6) provide insights into how the factors mentioned in the two systematic reviews played out in the South African context. In South Africa, the HIS strengthening efforts and DH projects were also primarily focused on the primary

health care (PHC) level. The majority of studies included in the scoping review (Chapter 4) also report on DH interventions or activities that aimed to improve data production, accessibility and usability of frontline level (or PHC level) health service providers.

In addition to the two systematic reviews mentioned above, two primary studies conducted in the United States of America (USA) are of relevance. The first study was published about a decade ago and the authors asked the question ‘can barriers to stakeholder participation [in the implementation of HIE] be overcome?’ [127]. This study listed fragmentation between databases and institutions, stakeholder engagement and buy-in, institutional collaborations and exchanges, and clinic and hospital staff’s perceptions as factors that can impact on the sustainability of HIE innovations. The second study aimed to determine the functions of HIE and the potential motivators, barriers, and facilitators of adoption of HIE innovations [128]. The authors report that user engagement with HIE tools (in terms of their desired functions), financial implications (costs and gains), and trust between HIE stakeholders were major contributing factors to the adoption of HIE. Like the systematic reviews, these studies also reported on similar factors of HIE as those reported in the thesis findings related to users’ perceptions, stakeholder relationships and finances as those described in the thesis.

There are three main differences between the four studies described above and the thesis findings. Firstly, the thesis findings are based on a low- and middle-income setting in South Africa as opposed to the high-income settings reported in the four studies (and the literature in general). Secondly, the PhD project reported primarily on the design and piloting of a DH initiative as it was emerging, while the literature generally seemed to report on innovations that were already implemented at regional levels. Thirdly, DH interventions or activities in the literature focused on a specific scope or purpose, while the DH initiative emerging in our local setting seems to have a wide scope. Below, I will describe the three differences between the literature and the thesis findings in detail.

### **2.1.1. Data harmonisation innovations in low- and middle-income settings versus high-income settings**

There is generally limited evidence on the factors affecting DH innovations across all income settings [8, 42, 44]. Of the little evidence available, it is observed that similar factors are identified as impacting DH innovations across low- and middle- income settings (such as those described in the thesis) and high-income settings (what is mostly found in the literature). However, there are two importance differences related to (a) the availability and focus of evidence and (b) the specific factors that impact DH innovations, in low- and middle- versus high-income settings.

Firstly, the majority of evidence on DH innovations is from high-income settings with a focus on post-implementation aspects, as described in the scoping review (Chapter 4). Studies from high-income settings generally claim that HIE innovations can improve quality, efficiency, and cost of health services [30, 85, 129, 130]. There is likely experiential evidence that supports this claim for high-income settings. For example, the second systematic review and the two primary studies conducted in the USA (mentioned above) are all based on HIE innovations in high-income settings and are focused on issues related to the scale-up and sustainability of HIE innovations [30, 127, 128]. Contrary, evidence from low- and middle- income settings, as described in the first systematic review and the thesis findings, is primarily focused on design, piloting and implementation issues. Even though, the PHDC case study (Chapter 6) refers to issues of institutionalising and sustaining DH innovations, less of the focus is on these later issues possibly because implementation of DH innovations are still ongoing and not yet near completion.

As mentioned before, the literature from high-income settings are primarily focused on HIE innovations that have already been implemented and it seems like recent problems (from 2000 onwards) are related to ensuring that lower level stakeholders (or primary care practices) join or continue to participate in regional HIE innovations [30, 129, 131]. There may be less evidence available on the design and pilot aspects of HIE innovations in high-income settings, possibly because HIE innovations are widely accepted and endorsed by local governments so research on earlier phases or proof of concepts are likely not required. Additionally, institutional and financial challenges associated with the start of innovations may be less intense and starting an innovation

is widely accepted, so understanding how innovations can be maintained and sustained is more critical.

Secondly, there are differences in how certain factors of DH innovations (social, institutional and technical) play out in low- and middle- income settings compared to high-income settings. Three key differences are: the state of IT infrastructure, stakeholder collaborations, and funding and institutionalisation of innovations. The historical analysis and ethnographic chapters provide a different perspective of these factors compared to the literature from high-income settings.

IT infrastructure challenges are more common in low- and middle- income settings compared to high-income settings. Although, the state of RHISs has improved in low- and middle- income settings over the past two decades (for example, as reported in Chapter 5), they still faces several challenges such as poor IT infrastructure and lack of personnel training [12, 15, 108]. This means that DH innovations in low- and middle- income settings are implemented in contexts where RHISs are still a work-in-progress, which may be different from RHISs in high-income settings. The four studies conducted in the USA (mentioned above) report on well-developed electronic medical records (EMRs) across health facilities and organisations. This advanced state of RHISs in the USA made it possible for innovators to implement HIE innovations with minimal IT infrastructure issues. In the case of HISP, the innovations were in themselves about establishing electronic platforms for capturing and processing data. While PHDC tried to harmonise electronic data captured across the RHIS, it also had to deal with incomplete electronic databases and establish new data management procedures. The difference in the state of RHISs in low- and middle- versus high income settings is important for explaining why innovations in low- and middle- income settings (like South Africa) face more difficulty from the start.

Another challenge that seemed to be more persistent in low- and middle- income settings than high-income settings was related to stakeholder engagement and collaborations, especially because of lack of existing or ongoing relationships between different people and institutions. In South Africa (a middle- income setting), various local and national stakeholders were involved in the early development and implementation processes of HIS interventions, and similarly various City and Province health authorities are involved in current DH innovations. Both HISP

and PHDC were positioned as government and university collaborations which meant they worked with HIS (database), health services and research stakeholders in designing, piloting and implementing the innovations. HIS and health services stakeholders were divided between City, Province and National health authorities. HISP and PHDC involved a wide range of stakeholders to ensure good buy-in of innovations, to establish a line of accountability, and to combine financial and personnel resources amongst institutions. However, there were lack of existing or ongoing relationships between stakeholders and conflicts between stakeholders emerged because of long-standing feuds (not directly related to HISP or PHDC); they had different data needs and priorities, and they perceived HIS or DH innovations threatening to their roles. Bringing different health authorities and HIT stakeholders together is an ongoing challenge in low- and middle-income settings, possibly because of past institutional divides and long-standing conflicts (see Chapter 3) and competition for limited resources.

Lastly, lack of funding for DH innovations and delays to institutionalise them was an ongoing challenge in low- and middle- income settings versus in high-income settings. Chapters 3 and 6 (on the HISP and PHDC case studies) both report on lack of funding and institutionalisation as challenges to fully implementing HIS and DH innovations. HISP initially acquired funding from two local universities and an international institution to design and pilot HIS interventions in a few provinces. Lack of funding was a challenge for HISP especially when HIS interventions were endorsed for national roll-out. This struggle for funding was linked to HISP not being institutionalised within government. Even though HISP was commissioned to roll-out HIS interventions on behalf of government, it was not funded by government or considered as part of government. PHDC faced similar challenges related to funding and institutionalisation. Design and pilots of DH innovations were initially funded through research grants, and some PHDC staff were co-employed by a local university and government. But, the need for PHDC to become institutionalised became increasingly necessary to access databases from different government departments in a routine manner, to be able to make use of IT infrastructure within government and to fund personnel salaries. Lack of funding and institutionalisation of HIS or DH innovations in low- and middle- income settings are ongoing challenges across phases of innovations unlike in high-income settings.

Given that high-income settings may be faced with less challenges than low- and middle- income settings in terms of IT infrastructure, stakeholder collaborations and funding and institutionalisation of DH innovations, they may generally be able to implement innovations without extensive delays and they are able to make better use of the innovations. This means that HIE innovations in high-income settings can be planned for, developed and implemented, and monitored and evaluated in contexts where supportive factors such as advanced RHISs, existing collaborations and financial capacities are present. It is likely that due to existing social, technical and financial resources, HIE innovations in high-income settings can be associated with: improved workflows (less time is spent on searching for data and referral processes are streamlined); cost savings (the costs of storing paper records are eliminated); and increased revenue for stakeholders (there are government incentives for using HIT and pay-for-performance incentives) [30]. Raising revenue through the use of HIE innovations seems distinct to high-income settings where governments give financial incentives to health facilities and organisations to remain within HIE networks and to use other HIT-related innovations [30].

### **2.1.2. Different phases of data harmonisation innovations**

The phases of HIT innovations such as DH include conceptualisation, design, development, piloting, implementation, institutionalisation, and monitoring and evaluation [8, 51, 132]. The PhD project looked at the early phases of DH innovations (conceptualisation, design, development and piloting) while the four studies above looked at implementation, sustainability and use issues. There are two important aspects to consider regarding the factors that can affect DH innovations across phases.

Firstly, different aspects of the broad categories of factors (such as social, technical, financial) are applicable in different ways at different phases of DH innovations. The ethnographic chapters (5 and 6) described the broad categories of factors that impacted on DH innovations in the Western Cape, but also how different aspects of those broad categories of factors featured across the different phases of DH projects at PHDC. For example, in our study, stakeholder engagement emerged as an important social factor of DH. These included issues such as differences amongst stakeholders regarding perceived purpose and who had what power to deliver on the expectation

of the PHDC project; and not having structured, formal and continuous working relationships between City and Province health authorities.

Earlier on, PHDC used small-scale projects such as the iALARM cascade to engage City and Province stakeholders about the usefulness of DH projects. In the case of the iALARM cascade, PHDC hoped that working collaboratively with City managers on the data collection process would ease City managers' fears that PHDC was trying to replace them, and it was an opportunity for them to identify mechanisms for exchanging data. Later on, PHDC engaged City and Province stakeholders for a different purpose, namely to access the data required to complete the data repository on a regular basis. PHDC involving stakeholders in the earlier phases of DH innovations was a mechanism for obtaining stakeholder buy-in, which made it possible for innovators to later engage stakeholders about data access. This illustrates the point that social interactions (with stakeholders) occur in different ways at different phases of DH innovations for different reasons.

Secondly, the degree of influence of the different factors of DH innovations may vary across different phases of DH innovations. The ethnographic chapters provided insights into the different factors and challenges of starting DH projects and moving small-scale DH projects from pilot testing to operationalisation. Chapter 5 showed that motivations and opportunities related to the RHIS environment, PHDC's institutional structure and data users' attitudes, and design factors are important in starting a DH project. Chapter 6 showed that institutional and conceptual challenges (such as different approaches to data access and governance, differences in conceptualising the value of data, misunderstandings about the purpose of formal data procedures and others) are more obvious once DH innovations start to move from the pilot phase to becoming institutionalised.

The historical analysis and case study of PHDC (Chapters 3 and 6) were distinct in that they described financial factors as important to the later phases of HIS and DH innovations (that is, moving from piloting to operationalising innovations). The degree of influence of financial factors varies across different phases of innovations. For example, less money is required for developing a proof of concept of an innovation, but more money is required when starting to

scale-up and institutionalise an innovation. This means that financial factors are more of a concern at later phases than earlier phases of innovations because new IT infrastructure and additional personnel salaries are required for sustaining the innovations.

In addition to the two points mentioned above, the phases of DH innovations may overlap or occur concurrently. For example, in the case of PHDC, some aspects of the earlier phases (design and piloting) were also seen in later phases (implementation and institutionalisation) of the data repository project. While PHDC was still trying to figure out how to access data from the City database regularly, PHDC was at the same time also trying to institutionalise new data procedures (such as application forms and reference documents) for dealing with data users' requests.

### **2.1.3. Wide versus narrow scope of data harmonisation innovations**

As described in the scoping review (Chapter 4) and most commonly described in the literature, DH innovations (interventions or activities) are usually centered on a specific (narrow) scope or purpose such as geographic area, disease surveillance, treatment management, or producing a data mart (or dashboard). However, DH innovations at PHDC were centered on a broader (wide) scope or purpose, that is, a province-wide data repository capturing different data across different government departments for different diseases and conditions. Below, I discuss this contrast between wide and narrow scope of DH innovations further.

Firstly, studies on DH innovations in Chapter 4 (the scoping review) generally described a narrow focus of DH innovations. The scoping review identified that DH innovations can be defined according to the level of the health care system, the level of health management decision-making, geographic area, disease or condition. Secondly, the focus of PHDC seemed to be primarily on one large project (the province-wide data repository) which is different to the studies of DH innovations described in Chapter 4 (the majority of those studies were from high-income countries). While PHDC initially focused on harmonising data related to specific conditions or health services (HIV, TB and maternal care), the data repository gradually expanded to capture data for all diseases and conditions and across all types of health services in the province. The wide scope of PHDC represented both strengths and weaknesses to the implementation of DH

projects. An important strength of the wide scope was that PHDC could involve a wide range of stakeholders (health services and research and City and Province). As discussed in Chapters 5 and 6, drawing on different stakeholders was a key tool for PHDC to gain acceptance, to identify key individuals, to gain access to databases, and to get financial and personnel resources. The inclusive approach of PHDC made it possible to push for the institutionalisation of data repository project as it was a province-wide innovation that would benefit different stakeholders across all levels of the health care system and support health management decision-making at all levels.

A potential weakness of PHDC having a wide scope was linked to the multiple, lengthy and time-consuming processes of collecting large volumes of data across different stakeholders, especially amongst those who were resistant. PHDC had to engage researchers and City and Province stakeholders separately for different purposes such as for piloting small projects and accessing the data they required for the data repository. Additionally, because of its wide scope, PHDC was affected by various issues that affected the RHIS such as the unequal transition of health facilities and types of health services to electronic databases. These issues not directly related to PHDC nevertheless impacted on its progress, for example, there was a delay in the process of completing the data repository because DH innovators needed to find workarounds for gaps in the data.

#### **2.1.4. Section summary**

This section described similarities and differences of the thesis findings and the literature. The characteristics of DH innovations are wide-ranging in terms of processes, activities, databases, institutions and people. However, from the literature and as shown in the scoping review (Chapter 4) it seems like more attention is paid to the technical factors and processes of DH innovations because there is limited reference to social factors such as context, stakeholders, institutions and perceptions [44, 50]. The alternative terms (exchange, linkage, interoperability) used to describe DH innovations or activities do not seem to offer a different perspective either. This is contrary to the experiences described in the historical analysis and ethnographic chapters of this thesis (see Chapters 3, 5 and 6) where social factors play a big role across settings of DH innovations. Models of DH innovations from the literature do not seem to capture the influence of factors (or aspects of those factors) and the level of influence of factors across different, overlapping or concurrent

phases of DH innovations. It is worth noting that ethnographic findings in the thesis showed the diversity and influence of various factors on DH innovations.

In reviewing the factors that impact on DH innovations across different phases in low- and middle-income settings and high-income settings, we also identified key tools and strategies that innovators used in dealing with the challenges associated with low- and middle- income settings. Below, I describe four strategies for navigating barriers and facilitators of DH innovations.

## **2.2. Strategies for navigating the barriers and facilitators of data harmonisation innovations**

Another aspect worth noting relates to the ways in which DH innovators (or HIS intervention developers) in the South African context dealt with the unintended challenges of DH processes. The idea that HIT innovations, in general, are prone to unintended challenges is well-described in the literature, but there is limited evidence explaining how unintended challenges can be dealt with [37, 42]. One such study is by Cresswell et al. and the study states that ignoring the unintended challenges of HIT innovations can have major consequences, and it proposes factors to consider in dealing with the unintended challenges [51]. The authors list leadership, strategy, vision and continuous cycles of improvements as key tools in navigating challenges. For example, HIT innovations should be developed and implemented in the context of a strong but flexible leadership; HIT designers should identify and apply key lessons learned from organisations that have developed and implemented HIT innovations before; there should be clarity around what data processes entail; and HIT processes and products should be continuously monitored and improved based on shared lessons and experiences. Similarly, I describe (in Chapters 3 and 5) several success factors or strategies for dealing with the unintended challenges of DH innovations. Below, I will highlight four key success factors and strategies for dealing with the unintended challenges of DH innovations as identified in this thesis. The four key success factors and strategies are related to the concept of cultivation, the role of champions, government-university (or research) collaborations, and formalising data procedures.

Firstly, cultivation is one strategy for navigating social challenges emerging during the development and implementation of DH innovations. Even though it is unclear where the term cultivation originates from, HISP managers used it to describe the process of slowly transforming a social system through new or unusual activities centered on a HIT innovation [11, 54]. Cultivation is an important stakeholder engagement strategy for innovation designers, implementers and users to jointly discuss different approaches to developing and implementing an innovation, which then leads to the transformation of stakeholder relationships. Local stakeholders, especially lower level staff, are empowered to participate in the development of an innovation through mediated discussions with high level managers, and innovation processes are flexible and slow enough to allow for learning, training and user feedback. Additionally, cultivation is about inscribing social aspects such as data practices and organisational values into the innovation, but it also considers that new interpretations or translations of the innovation may occur as users interact with the innovation [54].

One weakness of cultivation is that there is no clear division between innovators' design ideas and new design ideas that have emerged as the implementers and/or users of the innovation have interacted with it. This not only means that development and implementation of an innovation is continuous, but it also means that social systems and processes are being transformed as a response to the 'inscribed' data practices and organisational values of the innovation. It is difficult to identify a clear point of transformation of social systems and processes, just as it is difficult to sometimes predict the unintended challenges of innovations. A strength of cultivation is getting people across different hierarchies to engage with each other and the new innovation, it disrupts usual practices where a top-down approach is generally applied [54].

For example, in the case of HISP and PHDC cultivation was the process of bringing together stakeholders from different health authorities or institutions (such as City and Province) and levels of the health system (such as data clerks and information managers) to debate about different data approaches for developing and implementing the new innovations and through mediation collectively agreeing on a way forward. An agenda to transform existing relationships and data processes was, for example, inscribed in the DHIS (software) development process. HISP managers specifically provided computer training to data clerks to empower them in developing

their own data capturing applications. Transforming the role of data clerks (from data producers to innovators) was one way of moving from a top-down to a bottom-up approach to data production and utilisation.

Secondly, the role of a champion is another important strategy for navigating social challenges emerging during the development and implementation of DH innovations. A champion can be defined as someone who takes ownership of an innovative idea and promotes it, firstly in an informal manner within an organisation and then formally using a proof of concept motivates for the operationalisation of the idea [66]. It is a positive that champions are able to drive new innovations in terms of designing, developing and implementing them, because of existing connections to key stakeholders and they can establish new collaborations between different sectors [65, 66]. In addition to the champion, who is the leader of an innovation, ‘mini-champions’ also start to emerge as the champion develops a team to work on the innovation [65, 66]. Mini champions, like new data procedures, are a type of institutional capacity for scaling-up innovations.

The HISP and PHDC case studies demonstrated the important role of champions (especially at the start of innovations) while also highlighting the need for innovations to become institutionalised for long-term sustainability. In the case of HISP, champions initiated HIS interventions and facilitated discussions between different stakeholders and identified the data needs of future users. The PHDC champion also engaged with different stakeholders (City and Province health authorities) to obtain buy-in and to access the data that they required to complete the province-wide data repository. Additionally, the founder of PHDC established a team of DH innovators who designed and piloted different DH projects. HISP and PHDC were both successful in introducing new innovations (such as the DHIS and province-wide data repository), but both initiatives struggled to become institutionalised within the government structure.

Thirdly, the institutional positioning of an innovation is important in overcoming some of the social challenges that emerge during the development and implementation processes [11, 51]. In our case, it was ideal that the DH initiative was a government-university collaboration. Collaborations may be useful in contexts where socio-political factors are highly influential and

where resources are limited. Chapter 3 and 6 described tensions between provincial and national stakeholders and City and Province health authorities respectively as HISP and PHDC introduced new innovations. Although disagreements between different stakeholders were about data processes and reporting requirements at the surface level, there were underlying political issues between levels of the health care system. Political issues emerged from data processes implemented during apartheid and from the district-based to decentralise the health care system and the HIS. While a district-based HIS and a province-wide data repository were important for strengthening the health system, there was lack of technical skills and financial resources and flexible testing grounds for new innovations. HISP and PHDC were both established as government-university collaborations. The university provided the technical skills and resources required for piloting new innovations, while government provided access to public health services data, health services stakeholders (data users) and prospects for scaling-up and institutionalising the new innovation.

Lastly, balancing tensions emerging from formalising data procedures while maintaining informal (social) relationships is an important strategy for dealing with challenges that emerge during the development and implementation of innovations. The thesis findings reveal that institutionalising new innovations can be challenging. PHDC introduced formal data procedures to overcome this challenge; for example, application forms helped PHDC keep track of data requests and to show data users as well as high level managers that their activities were standardised and PHDC was transitioning into the government structure. Given the complexities of DH innovations (DH processes and products are continuously needing to be revised), DH innovators at PHDC used social relationships with data users to navigate the gaps in the formal data procedures. From the PhD project it emerged that stakeholder engagement is not only important for identifying data needs and obtaining buy-in for data processes, but also for grappling with practical issues of data products.

### **3. Implications for research**

There is a need for further research on what DH interventions and activities entail in diverse settings and contexts especially in LMICs. The scoping review highlighted the need for a better understanding of the components and processes of DH to allow for comparisons and to improve transferability of study results. Additionally, much resources are spent on the technical processes of DH innovations with the assumption that technical processes will lead to data utilisation for health management decision-making. But, the relationship between DH and health management decision-making is barely tested or theorised in the research.

There is limited ethnographic research on DH innovations documenting key social factors and complexities of DH innovations. There is an opportunity for further research on innovators, implementers and users' perceptions and experiences of DH processes, especially on what challenges they envisage and potential strategies for overcoming them. It may be useful to engage in long-term research (participant observation), from design to full implementation of a DH innovation to provide clarity on the spectrum of activities and short- and long-term outcomes. Interdisciplinary research may inform the development, implementation and evaluation of DH innovations from different perspectives, including perspectives of health information technology development, health information management, health systems development and social science.

### **4. Recommendations for practice**

This PhD project showed how DH innovations in our local setting in the Western Cape Province were influenced by various factors, such as prior HIS processes, institutional terrains and different design conceptualisations. Above, I identified key success factors and strategies from the literature and from this study on how HIT innovators dealt with the unintended challenges that emerged as innovations were being operationalised. Below, are more specific recommendations for developing and implementing a DH initiative in a low- and middle- income setting.

Data harmonisation innovations are faced with a wide range of social and technical factors, and tangible (people, institutions and technology) and non-tangible (design concepts and perceptions)

factors. Firstly, for a DH project such as the province-wide data repository to be successfully implemented, DH innovators should in advance ask key stakeholders (such as database managers and data users) what they foresee as barriers to implementation and how to potentially overcome the barriers. DH innovators should also conduct a situational analysis to identify the influence of key social and technical factors and plan for dealing with emerging challenges.

Secondly, DH innovators should consider key lessons of HIT innovations previously implemented in the same setting, such as the introduction of the district-based HIS in South Africa in earlier years. As lessons from the past can be used to guide the design of a flexible and adaptive DH project. For example, PHDC's experiences highlight the importance of balancing the need to formalise engagement with frontline implementers and data users, while still remaining flexible to engage with them during the earlier phases of trying to routinise procedures.

Thirdly, those involved in the DH project should clearly define what the purpose of the project is, which stakeholders are involved and what data is required. There may be value in defining DH according to factors identified in a situational analysis and having a common understanding about the innovation can be a way of clarifying the scope of discussions and managing conflicts amongst stakeholders.

Lastly, DH innovators should decide upfront who the primary users of their innovations are (in terms of the level of the health system) and how DH innovations are likely to impact decision-making processes. It is important for innovators to design DH products that are perceived as adding value to the workflows of the intended users. DH innovators should also consider different approaches for operationalising DH innovations across levels of the health system.

## **5. Strengths and limitations**

This PhD project used a multi-method qualitative approach to examine a new DH initiative in a middle- income setting in South Africa. Firstly, the study design in itself was a strength in that the research was conducted using qualitative mixed methods; I used a chronological analysis approach, scoped the literature systematically and participated in and observed ongoing DH activities. These methods provide the PhD project with credibility because of the long-term observations (mostly during 2016 and 2017); more than one data collection method was used (in chapters 3, 5 and 6 where primary research was conducted) which made triangulation possible; and I intermittently discussed data collection and data analysis with my PhD supervisors. The data collected was dependable because similar content emerged from different data collection notes (interviews, meetings and observations); and the findings could be presented in different formats as summaries, examples and interpretations and in relation to the literature thus making parts of it transferable to similar settings.

Secondly, the findings captured a good timeframe of events and factors ranging from 1994 onwards. From the historical analysis, which provides insights into earlier experiences of HIS interventions in South Africa; to the scoping review which focused on studies of electronic HISs; and finally, to the current state of HISs, specifically in terms of emerging DH innovations. Thirdly, the findings contribute new knowledge to three research gaps by providing (a) a common definition for DH innovations while highlighting the need to better incorporate social elements in its conceptualisations; (b) an early understanding of the challenges and opportunities (particularly the social ones) of a DH innovation as it was emerging in a real-world setting; and (c) evidence on how DH innovations and potentially other HIT innovations are initiated, developed and implemented in low- and middle- income settings.

There are three main limitations of the PhD project that need to be taken into consideration when interpreting the findings. Firstly, the historical analysis and scoping review (Chapters 3 and 4) relied on existing literature. There were a limited number of studies on the early developments of HISs and HIS interventions in South Africa and DH innovations in low- and middle- income settings. The historical analysis was also limited by the number of key informants. Only two key

informants, both working primarily as researchers and consultant, were interviewed which means the perspective of government or other HISP stakeholders was not captured through interviews.

Secondly, I had the dual role of being a doctoral researcher and an iALARM representative (therefore a PHDC stakeholder). I continuously reflected on what role (researcher or stakeholder) I was assuming at a particular time when engaging with PHDC to ensure that certain information that I was privy to as a stakeholder remained confidential and separate from the PhD project. There is a possibility that my knowledge of certain circumstances and events related to PHDC may have unintentionally impacted on the analysis and interpretation of the data. Lastly, the continuously changing nature of DH innovations may have impacted on the primary research findings. The findings only are based on specific moment in time in the development and implementation process. These findings are of the early phases of designing and piloting a DH innovation in South Africa.

## **6. Conclusion**

Data harmonisation is a complex and dynamic HIT innovation to strengthen HIS and health system performance. The PhD project found that existing HIT infrastructure, technical components and processes; the role of innovation champions; institutional and financial aspects; design concepts and pilot projects; stakeholder perceptions; and formalised data procedures; all determine the success of a DH initiative. The interaction between these various social and technical factors result in unintended challenges, but the willingness of innovators to use emerging opportunities to deal with the unintended challenges is critical to the completion of DH projects. Four approaches for dealing with HIT innovation challenges in South Africa (in earlier years and currently) relate to cultivating a space for different stakeholders to engage about their conflicting data needs and priorities, the emergence of a champion who can mediate conflicts, establishing a government-university collaboration, and formalising data procedures as a way of gradually becoming institutionalised within government for long-term sustainability.

There are still gaps in the literature on the spectrum of activities and outcomes of DH innovations as well as the relationship between DH and health management decision-making across levels of

the health system. However, this research is the basis for a set of key recommendations for future practice. Conducting a situational analysis of influential factors, drawing out lessons from the past, defining the scope and purpose of an innovation, and identifying intended users and outcomes, are all critical considerations in planning, implementing and monitoring DH innovations. Overall, this research is providing health system, information technology and research stakeholders with a broader understanding of the range of factors that affect DH innovations. This can lead to a more comprehensive approach in designing, implementing and evaluating DH innovations to limit poor outcomes of innovations and wasted resources.

## References

1. Olmen, JV et al. The Health System Dynamics Framework: The introduction of an analytical model for health system analysis and its application to two case-studies. *Health, Culture and Society*. 2012. 2(1):p.1
2. Kutzin, J and Sparkes, SP. Health systems strengthening, universal health coverage, health security and resilience. *Bulletin of the World Health Organization*. 2016. 94(1):p.2
3. Kieny, MP et al. Strengthening health systems for universal health coverage and sustainable development. *Bulletin of the World Health Organization*. 2017. 95(7):p.537
4. Mutale, W et al. Improving health information systems for decision making across five sub-Saharan African countries: implementation strategies from the African Health Initiative. *BMC Health Services Research*. 2013. 13(Suppl 2):p.S9
5. Prodinger, B et al. Harmonizing routinely collected health information for strengthening quality management in health systems: requirements and practice. *Journal of Health Services Research and Policy*. 2016. 21(4): p.223
6. Nutley, T and Reynolds, H. Improving the use of health data for health system strengthening. *Global Health Action*. 2013. 6(1):p.20001
7. World Health Organization. *Monitoring the building blocks of health systems: a handbook of indicators and their measurement strategies*. World Health Organization Geneva. 2012
8. Sittig, DF and Singh, H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. *BMJ Quality and Safety*. 2010. 19(Suppl 3):p.i68
9. Karuri, J et al. DHIS2: The tool to improve health data demand and use in Kenya. *Journal of Health Informatics in Developing Countries*. 2014. 8(1):p.38
10. Hotchkiss, DR et al. How can routine health information systems improve health systems functioning in low-and middle-income countries? Assessing the evidence base, in *Health Information Technology in the International Context*. Emerald Group Publishing Limited. 2012
11. Braa, J and Hedberg, C. The Struggle for District-Based Health Information Systems in South Africa. *The Information Society*. 2002. 18(2): p.113
12. Lippeveld, T. Routine health information systems: the glue of a unified health system. In *Keynote address at the Workshop on Issues and Innovation in Routine Health Information in Developing Countries*. Potomac. 2001
13. Aqil, A. et al. PRISM framework: a paradigm shift for designing, strengthening and evaluating routine health information systems. *Health Policy Plan*. 2009. 24(3): p.217
14. Nutley, T et al. Moving data off the shelf and into action: an intervention to improve data-informed decision making in Cote d'Ivoire. *Glob Health Action*. 2014. 7: p.25035
15. Heywood, A. and Boone, D. *Guidelines for Data Management Standards in Routine Health Information Systems*. Measure Evaluation. 2015
16. Hendriks, PH. Knowledge management, health information technology and nurses' work engagement. *Health Care Management Review*. 2016. 41(3): p.256
17. Hurter, T. *The culture of data use in the management structures of a rural health district in the Western Cape Province*. University of Cape Town. 2015
18. Aanestad, M et al. *Strategies for development and integration of health information systems: coping with historicity and heterogeneity*. 2005

19. Akhlaq, A et al. Barriers and facilitators to health information exchange in low- and middle-income country settings: a systematic review. *Health Policy Plan*. 2016. 31(9): p. 1310
20. Liu, D et al. Harmonization of health data at national level: a pilot study in China. *International Journal of Medical Informatics*, 2010. 79(6): p.450
21. Hopf, YM et al. Views of healthcare professionals to linkage of routinely collected healthcare data: a systematic literature review. *Journal of the American Medical Informatics Association*. 2014. 21(e1): p.e6
22. Haarbrandt, B et al. Automated population of an i2b2 clinical data warehouse from an openEHR-based data repository. *Journal of Biomedical Informatics*. 2016. 63(2016): p. 277
23. Foreit, K et al. Data demand and information use in the health sector: conceptual framework. *MEASURE Evaluation*. 2006
24. Benson, FG. Analysing the national notifiable diseases surveillance system in South Africa. *WIREDSpace*. 2018
25. Baker, MG and Fidler, DP. Global public health surveillance under new international health regulations. *Emerging Infectious Diseases*. 2006. 12(7): p.1058
26. Shade, SB et al. Health information exchange interventions can enhance quality and continuity of HIV care. *International Journal of Medical Informatics*. 2012. 81(10): p.e1
27. Herwehe, J et al. Implementation of an innovative, integrated electronic medical record (EMR) and public health information exchange for HIV/AIDS. *Journal of the American Medical Informatics Association*. 2011. 19(3): p. 448
28. Messer, LC et al. The development of a health information exchange to enhance care and improve patient outcomes among HIV+ individuals in rural North Carolina. *International Journal of Medical Informatics*. 2012. 81(10): p. e46
29. Gill, GV et al. Enumeration of non-communicable disease in rural South Africa by electronic data linkage and capture–recapture techniques. *Tropical Medicine and International Health*. 2001. 6(6): p. 435
30. Fontaine, P et al. Systematic review of health information exchange in primary care practices. *The Journal of the American Board of Family Medicine*. 2010. 23(5): p. 655
31. Braa, J et al. Developing health information systems in developing countries: the flexible standards strategy. *Mis Quarterly*. 2007. 31(2): p. 381
32. Ishijima, H et al. Challenges and opportunities for effective adoption of HRH information systems in developing countries: national rollout of HRHIS and TIIS in Tanzania. *Human Resources for Health*. 2015. 13: p. 48
33. Eden, KB et al. Barriers and facilitators to exchanging health information: a systematic review. *International Journal of Medical Informatics*. 2016. 88 (2016): p. 44
34. Lewis, D et al. Understanding the role of technology in health information systems. *Pacific Health Dialog*, 2012. 18(1): p. 144
35. Braa, JE et al. Networks of action: sustainable health information systems across developing countries. *MIS quarterly*. 2004: 28(3): p. 337
36. Chang, LW et al. Perceptions and acceptability of mHealth interventions for improving patient care at a community-based HIV/AIDS clinic in Uganda: a mixed methods study. *AIDS Care*. 2013. 25(7): p. 874
37. Vest, JR and Gamm, LD. Health information exchange: persistent challenges and new strategies. *Journal of the American Medical Informatics Association*. 2010. 17(3): p. 288
38. Mello, MM et al. Legal Barriers to the Growth of Health Information Exchange—Boulders or Pebbles? *The Milbank Quarterly*. 2018. 96(1): p. 110

39. Maiorana, A et al. Trust, confidentiality, and the acceptability of sharing HIV-related patient data: lessons learned from a mixed methods study about Health Information Exchanges. *Implementation Science*. 2012. 7(1): p. 34.
40. Garrib, A et al. An evaluation of the district health information system in rural South Africa. *SAMJ: South African Medical Journal*. 2008. 98(7): p. 549
41. Sæbø, JI et al. Comparing strategies to integrate health information systems following a data warehouse approach in four countries. *Information Technology for Development*, 2011. 17(1): p. 42
42. Harrison, MI et al. Unintended consequences of information technologies in health care—an interactive sociotechnical analysis. *Journal of the American Medical Informatics Association*. 2007. 14(5): p. 542
43. Plsek, P. Complexity and the adoption of innovation in health care. *Accelerating Quality Improvement in Health Care: Strategies to Accelerate the Diffusion of Evidence-Based Innovations*. National Institute for Healthcare Management Foundation and National Committee for Quality in Health Care. 2003
44. Cresswell, KM and Sheikh, A. Undertaking sociotechnical evaluations of health information technologies. *Journal of Innovation in Health Informatics*. 2014. 21(2): p. 78
45. Jacucci, EV et al. Standardization of health information systems in South Africa: The challenge of local sustainability. *Information Technology for Development*. 2010. 12(3): p. 225
46. Mahlathi, P. Minimum data sets for human resources for health and the surgical workforce in South Africa's health system. *African Institute of Health and Leadership Development*. 2015
47. Gilson, L et al. Development of the health system in the Western Cape: experiences since 1994. *South African Health Review*. 2017. 2017(1): p. 59
48. World Health Organization, Country health information systems assessments: overview and lessons learnt. World Health Organization Geneva. 2012
49. Waterson, P. Health information technology and sociotechnical systems: A progress report on recent developments within the UK National Health Service (NHS). *Applied Ergonomics*. 2014. 45(2): p. 150
50. Berg, M. Patient care information systems and health care work: a sociotechnical approach. *International Journal of Medical Informatics*. 1999. 55(2): p. 87
51. Cresswell, KM et al. Ten key considerations for the successful optimization of large-scale health information technology. *Journal of the American Medical Informatics Association*. 2016. 24(1): p. 182
52. Jha, AK et al. The use of health information technology in seven nations. *International journal of medical informatics*. 2008. 77(12): p. 848
53. Cho, S et al. Contextual dynamics during health information systems implementation: an event-based actor-network approach. *European Journal of Information Systems*. 2008. 17(6): p. 614
54. Braa, J. and Hedberg, C. Developing district-based health care information systems: the South African experience. *Proceedings of IRIS*. 2000
55. Coovadia, H et al. The health and health system of South Africa: historical roots of current public health challenges. *The Lancet*. 2009. 374(9692): p. 817
56. Horwitz, S. *Health and Health Care under Apartheid*. University of the Witwatersrand 2009.

57. Price, M. Health care as an instrument of apartheid policy in South Africa. *Health Policy and Planning*. 1986. 1(2): p. 158
58. Hendricks, JD et al. App challenged: Are midwives prepared? *Australian Nursing and Midwifery Journal*. 2016. 23(7): p. 32.
59. Pillay, YD et al. The district health system in South Africa: progress made and next steps. National Department of Health. 2001.
60. Kumalo, F. Health management information systems: Core health issues. *South African health review*. 2006. 2006(1): p. 65
61. Megan Osler, AB. Three interlinked electronic registers (Tier.Net) Project: a working paper. 2010.
62. Cornell, MB. The evolution and effectiveness of the South African antiretroviral therapy program. PhD Thesis. 2014
63. Serrano, KJ et al. Willingness to exchange health information via mobile devices: findings from a population-based survey. *Annals of Family Medicine*. 2016. 14(1): p. 34
64. Kostadinovski, A. and Atanasoski, D. Determinants of a healthcare information system. Jubilee International Scientific Conference. 2013
65. Barnett, J. et al. Understanding innovators' experiences of barriers and facilitators in implementation and diffusion of healthcare service innovations: a qualitative study. *BMC Health Services Research*. 2011. 11(1): p. 342
66. Hendy, J. and Barlow, J. The role of the organizational champion in achieving health system change. *Social Science & Medicine*. 2012. 74(3): p. 348
67. World Health Organization. Everybody's business-strengthening health systems to improve health outcomes. WHO's framework for action. 2007
68. Fichtinger, A et al. Data Harmonisation Put into Practice by the HUMBOLDT Project. *International Journal of Spatial Data Infrastructures Research*. 2011. 6: p. 234
69. Boyd, JH et al. Technical challenges of providing record linkage services for research. *BMC Medical Informatics and Decision Making*. 2014. 14 (1): p. 23
70. Haarbrandt, BT et al. Automated Transformation of openEHR Data Instances to OWL. *Studies in Health Technology and Informatics*. 2016. 223 (2016): p. 63
71. Hu, PJ et al. System for infectious disease information sharing and analysis: design and evaluation. *IEEE Transactions on Information Technology in Biomedicine*. 2007. 11(4): p. 483
72. Arksey, H. and O'Malley, L. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*. 2005. 8(1): p. 19
73. Schmidt, B-M. et al., Defining and conceptualising data harmonisation: a scoping review protocol. *Systematic reviews*. 2018. 7(1): p. 226
74. Cimino, JJ et al. Consumer-mediated health information exchanges: The 2012 ACMI debate. *Journal of Biomedical Informatics*. 2014. 48 (2014): p. 5
75. Tricco, AC et al. A scoping review on the conduct and reporting of scoping reviews. *BMC Medical Research Methodology*. 2016. 16(1): p. 15
76. Cochrane Effective Practice and Organisation of Care (EPOC) Group. EPOC Qualitative Evidence Syntheses: protocol template. 2018.
77. Suri, H. Purposeful sampling in qualitative research synthesis. *Qualitative Research Journal*. 2011. 11(2): p. 63
78. Popay, J et al. Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme. 2006.

79. Leon, N et al. Routine Health Information System (RHIS) interventions to improve health systems management. *Cochrane Database of Systematic Reviews*. 2015. 12 (CD012012): p.1
80. Liu, D et al. Harmonization of health data at national level: a pilot study in China. *International Journal of Medical Informatics*. 2010. 79(6): p. 450
81. Gill, GV et al. Enumeration of non-communicable disease in rural South Africa by electronic data linkage and capture-recapture techniques. *Tropical Medicine and International Health*. 2001. 6(6): p. 435
82. Haarbrandt, BE et al. Automated population of an i2b2 clinical data warehouse from an openEHR-based data repository. *Journal of Biomedical Informatics*. 2016. 63(2016): p. 277
83. Jones, RB et al. Overview of electronic data sharing: why, how, and impact. *Current Oncology Reports*. 2012. 14(6): p. 486
84. Elysee, GJ et al. An observational study of the relationship between meaningful use-based electronic health information exchange, interoperability, and medication reconciliation capabilities. *Medicine (Baltimore)*. 2017. 96(41): p. e8274
85. Dixon, BE et al. A Framework for evaluating the costs, effort, and value of nationwide health information exchange. *Journal of the American Medical Informatics Association*. 2010. 17(3): p. 295
86. Esmailzadeh, P and Sambasivan, M. Health Information Exchange (HIE): A literature review, assimilation pattern and a proposed classification for a new policy approach. *Journal of Biomedical Informatics*. 2016. 64 (2016): p. 74
87. Esmailzadeh, P and Sambasivan, M. Patients' support for health information exchange: a literature review and classification of key factors. *Implementation Science*. 2017. 17(1): p. 33
88. Fontaine, P et al. Systematic review of health information exchange in primary care practices. 2010. *Journal of the American Board of Family Medicine*. 23(5): p. 655
89. Kash, BA et al. Review of successful hospital readmission reduction strategies and the role of health information exchange. *International Journal of Medical Informatics*. 2017. 104(2017): p. 97
90. Mastebroek, M et al. Health information exchange in general practice care for people with intellectual disabilities-a qualitative review of the literature. *Research In Developmental Disabilities*. 2014. 35(9): p. 1978
91. Parker, C., M. Weiner, and M. Reeves, Health information exchanges-unfulfilled promise as a data source for clinical research. *International Journal of Medical Informatics*. 2016. 87 (2016): p. 1
92. Rahurkar, S et al. Despite the spread of health information exchange, there is little evidence of its impact on cost, use, and quality of care. *Health Affiliation*. 2015. 34(3): p. 477
93. Rudin, RS et al. Usage and effect of health information exchange: a systematic review. *Annals of Internal Medicine*. 2014. 161(11): p. 803
94. Sadoughi, F et al. The impact of health information exchange on healthcare quality and cost-effectiveness: a systematic literature review. *Computer Methods and Programs in Biomedicine*. 2018. 161(2018): p. 209
95. Vest, JR et al. Health information exchange: national and international approaches. *Advances in Health Care Management*. 2012. 12 (2012): p. 3
96. Ji, H et al. Technology and Policy Challenges in the Adoption and Operation of Health Information Exchange Systems. *Advances in Health Care Management*. 2017. 23(4): p. 314

97. Santos, MR et al. Health information exchange for continuity of maternal and neonatal care supporting: a proof-of-concept based on ISO Standard. *Applied Clinical Informatics*. 2017. 8(4): p. 1082
98. Cimino, JJ et al. Consumer-mediated health information exchanges: the 2012 ACMI debate. *Journal of Biomedical Informatics*. 2014. 48 (2014): p. 5
99. Downs, SM et al. Improving newborn screening laboratory test ordering and result reporting using health information exchange. *Journal of the American Medical Informatics Association*. 2010. 17(1): p. 13
100. Grossman, JM et al. Creating sustainable local health information exchanges: can barriers to stakeholder participation be overcome? *Res Brief*. 2008(2): p. 1
101. Kuperman, GJ and McGowan, JJ. Potential unintended consequences of health information exchange. *Journal of General Internal Medicine*. 2013. 28(12): p. 1663
102. Politi, L et al. Use patterns of health information exchange through a multidimensional lens: conceptual framework and empirical validation. *Journal of Biomedical Informatics*. 2014. 52 (2014): p. 212
103. Vest, JR et al. More than just a question of technology: factors related to hospitals' adoption and implementation of health information exchange. *International Journal of Medical Informatics*. 2010. 79(12): p. 797
104. Shapiro, JS et al. Approaches to patient health information exchange and their impact on emergency medicine. *Annals of Emergency Medicine*. 2006. 48(4): p. 426
105. Vest, JR and Abramson, E. Organizational uses of health information exchange to change cost and utilization outcomes: a typology from a multi-site qualitative analysis. *AMIA Annual Symposium Proceedings*. 2015
106. Zaidan, B et al. A security framework for nationwide health information exchange based on telehealth strategy. *Journal of Medical Systems*. 2015. 39(5): p. 51
107. Nicol, E et al. Perceptions about data-informed decisions: an assessment of information-use in high HIV-prevalence settings in South Africa. *BMC Health Services Research*. 2017. 17(2): p. 765
108. Lippeveld, T. Routine health facility and community information systems: creating an information use culture. *Global Health: Science and Practice*. 2017
109. Peute, LW et al. Anatomy of a failure: a sociotechnical evaluation of a laboratory physician order entry system implementation. *International Journal of Medical Informatics*, 2010. 79(4): p. e58
110. Karsh, B-T et al. Health information technology: fallacies and sober realities. *Journal of the American Medical Informatics Association*. 2010. 17(6): p. 617
111. Boulle, A. Strategic approach to patient-level health data harmonisation and integration. *Western Cape Department of Health*. 2013
112. Boulle, A. The use of routine health data sources in the Western Cape to address health service questions (presentation). *Western Cape Department of Health*, 2014
113. Mehta, U et al. Assessing the value of Western Cape Provincial Government health administrative data and electronic pharmacy records in ascertaining medicine use during pregnancy. *South African Medical Journal*. 2018. 108(5): p. 439
114. Boulle, A., The Western Cape Provincial Health Data Center (presentation). *Health Impact Assessment Directorate, Western Cape Department of Health*. 2017
115. SouthAfrica.info, Health care in South Africa. 2019.  
<http://wwwsouthafricainfo/about/health/health.htm#Vs8imvs9IaI>

116. Hallett, TB and Eaton, JW. A side door into care cascade for HIV-infected patients? *JAIDS Journal of Acquired Immune Deficiency Syndromes*. 2013. 63(2): p. S228
117. Boulle, A. Antiretroviral therapy and early mortality in South Africa. *Bulletin of the World Health Organization*. 2008. 86(9): p. 678
118. Adaba, GB and Kebebew, Y. Improving a health information system for real-time data entries: An action research project using socio-technical systems theory. *Informatics for Health and Social Care*. 2018. 43(2): p. 159
119. Merriam, SB. *Qualitative Research and Case Study Applications in Education. Case Study Research in Education*. 1998
120. Carter, N et al. The use of triangulation in qualitative research. *Oncology Nursing Forum*. 2014
121. Hersh, W Health care information technology: progress and barriers. *Jama*. 2004. 292(18): p. 2273
122. Patel, VL and Kannampallil, T. Human factors and health information technology: current challenges and future directions. *Yearbook of Medical Informatics*. 2014. 9(1): p. 58
123. Colvin, CJ et al. Health information as a catalyst for community health system engagement. *South African Health Review*. 2018. 2018(1): p. 135
124. Green, J and Thorogood, N. *Qualitative methods for health research*. Sage. 2018
125. Liamputtong, P and Ezzy, D. *Qualitative research methods*. Oxford University Press Melbourne. 2005
126. Cresswell, KM et al. Actor-Network Theory and its role in understanding the implementation of information technology developments in healthcare. *BMC Medical Informatics and Decision Making*, 2010. 10(1): p. 67
127. Grossman, JM et al. Creating sustainable local health information exchanges: can barriers to stakeholder participation be overcome? *Center for Studying Health System Change*. 2008
128. Ross, SE et al., Health information exchange in small-to-medium sized family medicine practices: motivators, barriers, and potential facilitators of adoption. *International Journal of Medical Informatics*. 2010. 79(2): p. 123
129. Fontaine, P et al. Health information exchange: participation by Minnesota primary care practices. *Archives of Internal Medicine*. 2010. 170(7): p. 622
130. Hyppönen, H et al. User experiences with different regional health information exchange systems in Finland. *International Journal of Medical Informatics*. 2014. 83(1): p.
131. Haarbrandt, B et al. *IHE XDS. Informatics, Management and Technology in Healthcare*. 2013
132. Cresswell, K. and Sheikh, A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. *International Journal of Medical Informatics*. 2013. 82(5): p. e73
133. Gautham, M et al. District decision-making for health in low-income settings: a qualitative study in Uttar Pradesh, India, on engaging the private health sector in sharing health-related data. *Health Policy and Planning*. 2016. 31(2): p ii55
134. Public Health Strategy Division, Welsh Assembly Government. *Framework for best practice: the production and use of health information for the public*. [http://www.wales.nhs.uk/sites3/Documents/420/framework\\_bestpractice\\_e1.pdf](http://www.wales.nhs.uk/sites3/Documents/420/framework_bestpractice_e1.pdf)

135. World Health Organization. Health information systems. 2008.  
[https://www.who.int/healthinfo/statistics/toolkit\\_hss/EN\\_PDF\\_Toolkit\\_HSS\\_InformationSystems.pdf](https://www.who.int/healthinfo/statistics/toolkit_hss/EN_PDF_Toolkit_HSS_InformationSystems.pdf)

# Appendices

## Appendix 1. Search strategies

### PUBMED

data harmonisation [tiab] OR data harmonization [tiab] OR data linkage [tiab] OR data sharing [tiab] OR health information exchange [tiab] OR "Health Information Exchange" [Mesh]

AND

electronic medical record [tiab] OR health information management [tiab] OR hospital information management system [tiab] OR district health information system [tiab] OR district health management information system [tiab] OR healthcare information system [tiab] OR "Electronic Health Records"[Mesh] OR "Health Records, Personal"[Mesh] OR "Health Information Systems"[Mesh] OR "Medical Records Systems, Computerized"[Mesh])

AND

Filters - English AND Publication Date: 2000/01/01 – 2018/09/30

### CINAHL




TI data harmonization OR TI data linkage OR TI data sharing OR MH health information exchange OR TI health information exchange AND MH health information systems

Limiters - Published Date: 20000101-20181231; Narrow by Language: English

### WEB OF SCIENCE

(TS=(data harmonization AND health information system)) AND LANGUAGE: (English)  
Indexes=SCI-EXPANDED Timespan=2000-2018

## Appendix 2. Ethics approval for PhD project

 <b>UNIVERSITY OF CAPE TOWN</b> <small>UNIBESITHI YASIKAPA - UNIVERSITEIT VAN KAAPSTAD</small>	 <b>HUMAN RESEARCH ETHICS COMMITTEE</b> 13 NOV 2018 <b>HEALTH SCIENCES FACULTY</b>	 <b>FACULTY OF HEALTH SCIENCES</b> Human Research Ethics Committee
<b>FHS016: Annual Progress Report / Renewal</b>		

<b>HREC office use only (FWA00001637; IRB00001938)</b> <b>This serves as notification of annual approval, including any documentation described below.</b>			
<input checked="" type="checkbox"/> Approved	Annual progress report	Approved until/next renewal date	30/11/2019
<input type="checkbox"/> Not approved	See attached comments		
Signature Chairperson of the HREC		Date Signed	15/11/2018

Comments to PI from the HREC <i>Thank you for the deviation document</i>
---

### Principal Investigator to complete the following:

#### 1. Protocol information

Date (when submitting this form)	23/10/2018		
HREC REF Number	738/2016	Current Ethics Approval was granted until	28/02/2018
Protocol title	Harmonising routinely collected health information to improve the performance of the HIV cascade for men in the Western Cape, South Africa		
Protocol number (if applicable)			
Are there any sub-studies linked to this study?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No		
If yes, could you please provide the HREC Ref's for all sub-studies? <b>Note:</b> A separate FHS016 must be submitted for each sub-study.			
Principal Investigator	A/Prof Christopher J. Colvin		
Department / Office Internal Mail Address	Division of Social and Behavioural Sciences, School of Public Health and Family Medicine		

1.1 Does this protocol receive US Federal funding?	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No
1.2 If the study receives US Federal Funding, does the annual report require full committee approval?	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> No
1.3 Has sponsorship of this study changed? If yes, please attach a revised summary of the budget.	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> No



**2. List of documentation for approval**

--

**3. Protocol status (tick ✓)**

<input type="checkbox"/>	Open to enrolment
<input checked="" type="checkbox"/>	Closed to enrolment (tick ✓)
<input checked="" type="checkbox"/>	Research-related activities are ongoing
<input type="checkbox"/>	Research-related activities are complete, long-term follow-up only
<input type="checkbox"/>	Research-related activities are complete, data analysis only
<input type="checkbox"/>	Main study is complete but sub-study research-related activities are ongoing
<input type="checkbox"/>	Study is closed → Please submit a Study Closure Form (FHS010)

**4. Enrolment**

Number of participants enrolled to date	19
Number of participants enrolled, since last HREC Progress report (continuing review)	n/a
Additional number of participants still required	1

**5. Refusals**

Total number of refusals (participants invited to join the study, but refused to take part)	0
---	---

**6. Cumulative summary of participants**

Total number of participants who provided consent	19
Number of participants determined to be ineligible (i.e. after screening)	0
Number of participants currently active on the study	19
Number of participants completed study (without events leading to withdrawal)	19
Number of participants withdrawn at participants' request (i.e. changed their mind)	0
Number of participants withdrawn by PI due to toxicity or adverse events	0
Number of participants withdrawn by PI for other reasons (e.g. pregnancy, poor compliance)	0
Number of participants lost to follow-up. Please comment below on reasons for loss of follow-up.	0
Number of participants no longer taking part for reasons not listed above. Please provide reasons below:	0

### Appendix 3. Ethics approval for iALARM study



UNIVERSITY OF CAPE TOWN  
UNIVERSITEIT VAN KAAPSTAD

FACULTY OF HEALTH SCIENCES  
Human Research Ethics Committee



#### FHS016: Annual Progress Report / Renewal

<b>HREC office use only (FWA00001637; IRB00001938)</b>			
<b>This serves as notification of annual approval, including any documentation described below.</b>			
<input checked="" type="checkbox"/> Approved	Annual progress report	Approved until/next renewal date	30/06/17
<input type="checkbox"/> Not approved	See attached comments		
Signature Chairperson of the HREC		Date Signed	21/6/2016

Comments to PI from the HREC	HUMAN RESEARCH ETHICS COMMITTEE  21 JUN 2016  HEALTH SCIENCES FACULTY UNIVERSITY OF CAPE TOWN

**Principal Investigator to complete the following:**

**1. Protocol information**

Date (when submitting this form)	20 June 2016		
HREC REF Number	320/15	Current Ethics Approval was granted until	30 June 2016
Protocol title	Assessing men's performance in the HIV cascade in Klipfontein sub-district: A retrospective cohort study (sub-study linked to 802/2014)		
Protocol number (if applicable)			
Are there any sub-studies linked to this study?	X Yes No		
If yes, could you please provide the HREC Ref's for all sub-studies? <b>Note:</b> A separate FHS016 must be submitted for each sub-study.	802/2014		
Principal Investigator	A/Prof Christopher Colvin		
Department / Office Internal Mail Address	School of Public Health and Family Medicine		

1.1 Does this protocol receive US Federal funding?	<input checked="" type="checkbox"/> XX Yes	<input type="checkbox"/> No
1.2 If the study receives US Federal Funding, does the annual report require full committee approval?	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> XX No
1.3 Has sponsorship of this study changed? If yes, please attach a revised summary of the budget.	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> XX No

HUMAN RESEARCH  
ETHICS COMMITTEE



UNIVERSITY OF CAPE TOWN  
UNIVERSITEIT VAN KAPSTAD

20 JUN 2016 FACULTY OF HEALTH SCIENCES  
Human Research Ethics Committee



HEALTH SCIENCES FACULTY  
UNIVERSITY OF CAPE TOWN  
**Form FHS006: Protocol Amendment**

HREC office use only (FWA00001637; IRB00001938)

<input checked="" type="checkbox"/> Approved	<input checked="" type="checkbox"/> Type of review: Expedited	<input type="checkbox"/> Full committee
--	---	---

This serves as notification that all changes and documentation described below are approved.

Signature Chairperson of the HREC		Date	21/6/2016
-----------------------------------	--	------	-----------

**Note:** All major amendments must include a local PI Synopsis justifying the changes for the amendment. Please note that incomplete amendment submissions will not be reviewed.

Comments from the HREC to the Principal Investigator

**Note:** The approval of this protocol amendment does not grant annual approval. Please complete the FHS016 / FHS017 form for annual approval at least one month before study expiration.

**Principal Investigator to complete the following:**

**1. Protocol information**

Date (when submitting this form)	14 June 2016	
HREC REF Number	320/2015	
Protocol title	ASSESSING MEN'S PERFORMANCE IN THE HIV CASCADE IN KLIPFONTEIN SUB-DISTRICT: A RETROSPECTIVE COHORT STUDY	
Protocol number (if applicable)		
Principal Investigator	A/Prof Christopher J Colvin	
Department / Office Internal Mail Address	Falmouth 3.46, School of Public Health and Family Medicine	
1.1 Is this a major or a minor amendment? (see FHS006hlp) Major (tick box) Minor (tick box)	<input checked="" type="checkbox"/> XX Major	<input type="checkbox"/> Minor
1.2 Does this protocol receive US Federal funding?	<input checked="" type="checkbox"/> XX Yes	<input type="checkbox"/> No
1.3 If the amendment is a major amendment and receives US Federal Funding, does the amendment require full committee approval?	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> XX No

## Appendix 4. Consent form

### Consent to participate in research

Harmonising routinely collected health information to improve the performance of the HIV cascade for men in the Western Cape, South Africa

#### Who is doing this study and why?

You are being asked to participate in the PhD research project of Ms Bey-Marrié Schmidt from the School of Public Health and Family Medicine at the University of Cape Town. This PhD research project is situated within a larger project – *Using Information to Align Services and Link and Retain Men in the HIV Cascade*—or ‘i-ALARM, which seeks to ‘raise the alarm’ about the under-recognised but urgent need to link and retain men in HIV-cascade-related services.

The PhD research project aims to understand the technical and perceived processes, promises and challenges of harmonising routinely collected health information to improve the performance of the HIV cascade particularly for men in the Western Cape, South Africa. While the researcher is interested in the harmonization and management of health information, she will ask a wide range of questions and speak to a wide range of people working with health information across all levels of the health system at both city and provincial-levels.

If you agree to participate in this project, you will be asked to speak to Ms Bey-Marrié Schmidt.

#### What will you do in this study?

If you volunteer to participate in this study, the researcher will ask you to answer some questions and talk about your perceptions of and experiences with health information systems, as well as to participate in some of your daily tasks. If you feel uncomfortable about answering any of the research questions, feel free not to participate or not to answer a particular question.

Some of these interviews will be informal and will take place in places where you are already spending time working. These conversations will likely be short and informal. Some of the interviews, though, will be more formal and in-depth. These could last between 1 and 1.5 hours. Please tell the researcher if you have any time limits or if you need to leave at any time. Nothing will happen if you do not wish to participate or if you decide to withdraw from the study before its conclusion.

The researcher may ask you for permission to record some of these interviews. You will have the choice to have the interview recorded or not. If you not to have it recorded, the researcher will simply take notes about your conversation.

#### Are there any risks in this research?

The researcher may ask you to speak to uncomfortable or difficult issues, such as challenges that you experience in performing your daily tasks. If at any time you do not want to answer a particular question, please tell the researcher and you will not be asked to answer. You are free to not answer any question or speak about any subject that you do not want to. If you feel upset during or after the interview, please tell the researcher.

#### Are there any benefits for participating in this research?

There are no direct benefits to you for participating in this study. There will be no payment for participating in this study, for example. There may be some long-term, indirect benefits to you if the project is able to provide recommendations to strengthen monitoring and evaluation, and support service improvements in terms of the performance of the HIV cascade.

#### Will my name be shared with anyone?

The researcher will not share your name with anyone and when she writes about the research, she will not use your name. All the information from this project will be kept by the researcher in a safe place. The researcher may share some of the information with the research team, but no one outside the research team will have access to your information. Extracts from your interviews may be published in research reports but any direct information that could identify who you are will be removed.

#### Who is part of the research team?

The researcher you will speak to is Ms Bey-Marrié Schmidt from the School of Public Health and Family Medicine at the University of Cape Town. Her supervisors for her PhD research are Dr Christopher J Colvin who is also from the School of Public Health and Family Medicine at the University of Cape Town and Dr Natalie Leon from the South African Medical Research Council. If you have any questions or concerns about the research, please feel free to contact:

Ms Bey-Marrié Schmidt  
Cell: 079 070 3631 (during office hours)  
Email: [schbey001@myuct.ac.za](mailto:schbey001@myuct.ac.za)

or

Dr. Christopher J. Colvin  
Tel: 021 406-6706 (during office hours) or 084-684-7202 (anytime)  
E-mail: [cj.colvin@uct.ac.za](mailto:cj.colvin@uct.ac.za)

#### What are my rights as a research participant?

You may withdraw your consent (permission) to participate in this study at any time and stop participating without any penalty. When you participate in this study, you are not giving up any legal claims, rights or remedies that you may have. If you have questions about your rights as a

research participant, contact the Human Research Ethics Committee (HREC) at the Faculty of Health Sciences at the University of Cape Town at 021 406 6338.

Signature of the research participant

The information above was described to me by \_\_\_\_\_. I was given the opportunity to ask questions and these questions were answered to *my* satisfaction.

*I hereby consent voluntarily to participate in this study.* I have been given a copy of this form.

\_\_\_\_\_  
Name of Participant

\_\_\_\_\_  
Signature of Participant

\_\_\_\_\_  
Date

Signature of the researcher

I declare that I explained the information given in this document to \_\_\_\_\_ [*name of the participant*]. [*He/she*] was encouraged and given ample time to ask me any questions.

\_\_\_\_\_  
Signature of Researcher

\_\_\_\_\_  
Date

## Appendix 5. Interview guide

### Guide for formal interviews

Harmonising routinely collected health information to improve the performance of the HIV cascade for men in the Western Cape, South Africa

#### Introductory comments

Introduce the study and the purpose of the interview

Explain the purpose of the consent form and its contents: risks and benefits, confidentiality, voluntary participation, use of results

Explain note-taking and/or recording

Ask if the participant has any questions and address them

#### Section 1: Role in health information system or management

What is your role? Prompt: Tell me more about your day to day work tasks. How does your role engage with the processes for collecting, analysing, reporting, using or managing health data?

How long have you been in this role?

Are you employed by the Western Cape Provincial or City of Cape Town Health Department?

What health data meetings (if any) do you attend and what is their purpose?

What policy or process changes have occurred (for example in the last 5 years) that relate to your role in health information system or management?

How have these changes affected you?

#### Section 2: HIV-cascade-related health information systems

What is your understanding of the HIV cascade?

Which health information systems do you know about that capture data related to the HIV cascade? Prompt: Are there different health information systems (both paper and electronic) that capture HIV testing, ART eligibility assessment, ART initiation, migration, death etc.?

Which health information systems (both paper and electronic) do you work with?

How do the health information systems you work with differ? Prompt: What is the pathway of health information from patient consultation to management decision-making (collection, use, flow and management)?

#### Section 3: Processes, challenges and opportunities of health information systems

Are there guidelines, policies, or other documents that describe your role and how you should collect/use/analyse/manage health data?

What happens in reality? Prompt: What are the differences between what the guidelines/policies/documents state and what happens in practice due to certain constraints?

What aspects do you like about the current health information systems or management processes? Prompt: What do you like about your role?  
What technical/professional/personal challenges do you face that impact on your role? Prompt:  
What do you think are the weaknesses of the health information systems you work with? Or what do you not like about the health information systems you work with?  
Why do you think these health information system or management problems have not been addressed?

#### Section 4: Developing a health information system intervention

What do you think are viable solutions to these health information system or management problems? Prompt: Tell me about interventions that you have (or would) developed or implemented as solutions?

What other health information system or management interventions do you know about?  
Prompt: Experience from previous role or another province etc.

What do you know about data harmonisation? Prompt: Do you think it is technically possible for health data to be harmonised between city and provincial facilities, laboratories and death registries?

Who do you know (colleagues) that is working on data harmonisation or other health information system or management interventions?

What aspects of your role would become easier if health data were harmonised? For example, if individual, cohort or aggregate data were timeously and easily available at all stages of the HIV cascade: testing, ART eligibility assessment, ART initiation, migration, death etc.

#### Concluding comments:

Reassure participant of confidentiality and anonymity

Ask if the participant has any questions and address them

Discuss future engagements

Thank participant