A new trajectory for Spatial Data Infrastructure Evolution in the developing world

Prestige Tatenda Makanga
(2010)

Thesis Presented for the Degree of
Master of Science Engineering in Geomatics at
The UNIVERSITY OF CAPE TOWN

Supervisor: Dr Julian Smit

Declaration

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

2. I have used the Harvard convention for citation and referencing. Each contribution to, and quotation in, this research report from the work(s) of other people has been attributed, and has been cited and referenced.

3. This thesis is my own work.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Signature ______________________________

Date **13/05/2010**

Acknowledgements

I would like to thank the following people and organisations for the key role that they have played in this research:

1. Dr Julian Smit for the support and mentorship throughout my two years at UCT. You were there the whole time to make things happen (Funding, Networking, Conferences, academic guidance etc).
2. The African Centre for Cities through the Cape Urban Observatory for:
    a. Funding my Masters research
    b. Funding conference attendance (FOSS4G2008 (SA) and GSDI11 (Netherlands)
    c. Funding the research trip to the Stockholm Resilience Centre
3. The UCT International Academic Programs Office (IAPO) for funding and facilitating my semester of exchange to Simon Fraser University (SFU) Vancouver, Canada.
4. Prof Nadine Schuurman and the Health Geographics team at SFU for your critical input to my research and helping out with my write up.
5. The CUO team, Dr Ralph Hamann, Dr Yvonne Hansen, Justin Gichobi, Ratidzo Dhlembeu, Rebecca Tyler, Ruth Massey and Ayse Basbozkurt for your companionship and making the CUO a good working environment.
6. My family back in Zimbabwe for moral and financial support when I moved to South Africa and Fadzai my Fiancee for the moral support.

Abstract

Spatial Data is a key resource in the development of cities. There is a lot of socio-economic potential that is locked away in spatial data holdings and this potential is unlocked by making the datasets widely available for use. Spatial Data Infrastructures (SDIs) have served this primary purpose; to make data accessible through the use of web based technologies. However, SDIs have not had their anticipated impact at local levels of governance. They have traditionally served as platforms that facilitate access to raw spatial datasets. They have not fully facilitated for the use of these datasets and therefore have attracted minimal attention from decision makers and users.

This research suggests a new trajectory for SDI evolution; a trajectory that will allow them to evolve into more relevant platforms for confronting the urban crisis in developing nations and thereby ensuring that they have the societal impact that they are intended to. The research explores the characteristics of the mainstream efforts to counter urban crises in the developing world to determine how the new SDI should be re-conceptualised to more adequately assist in responding to the urban crisis. This leads to the incorporation of Evidence Based Practice (EBP) into SDI through the use of urban indicators and knowledge creation processes to reflect on the pressing societal issues. From the new SDI concept, an architectural design is implemented as a “proof of concept”. At the heart of this new concept is the SDIs ability to provide access to more than just raw spatial datasets but useful information products that are based on these data. This proves that EBP can be incorporated into SDI to make them more efficient in responding to the urban problems in developing nation and consequently more relevant Information Infrastructures for urban decision makers.

Glossary

CCT	City of Cape Town
CGC	Community Geomatics Centre
CSIR	Centre for Science and Industrial Research
CUO	Cape Urban Observatory
DBSA	Development Bank of South Africa
DFD	Data Flow Diagram
DLA	Department of Land Affairs
EBDM	Evidence Based Decision Making
EBP	Evidence Based Practice
ESRI	Environmental Systems Research Institute
EU	European Union
FDD	Feature Driven Development
FOSS	Free and Open Source Software
GI	Geographical Information
GIS	Geographical Information Systems
GrSD	Grabouw Sustainable Development Initiative
GSDI	Global Spatial Data Infrastructure
GUO	Global Urban Observatory
ICA	International Cartographic Association
ICT	Information and Communication Technology
IDP	Integrated Development Plan
INSPIRE	Infrastructure for Spatial Information in Europe
IS	Information System
ISO	International Standards Organisation
ISRF	Information Systems Research Framework
IT	Information Technology
KML	Keyhole Markup Language
KPAs	Key Priority Areas/Key Performance Areas
LUOs	Local Urban Observatories
MDGs	Millennium Development Goals
MER	Monitoring Evaluation and Reporting
NGO	Non Governmental Organization
NII	National Information Infrastructure
NSDI	National Spatial Data Infrastructure
NSIF	National Spatial Information Framework
ODP	Open Distributed Processing

OGC	Open Geospatial Consortium
RE	Requirements Engineering
RM-ODP	Reference Model for Open Distributed Processing
RRSU	Regional Remote Sensing Unit
RVU	Regional Vancouver Urban Observatory
SACN	South African Cities Network
SADC	Southern African Development Community
SASDI	South African Spatial Data Infrastructure
SDDF	Spatial Data Discovery Facility
SDF	Spatial Development Framework
SDI	Spatial Data Infrastructure
SE	Software Engineering
SES	Socio Economic Status
SHS	Sustainable Human Settlements
SMD	Spatial Metadata Discovery
SOA	Service Oriented Architecture
SSMIC	Sault Ste Marie Innovation Centre
Stats SA	Statistics South Africa
TWK	Theewaterskloof
UAP	Urban Atlas Portal
UCT	University of Cape Town
UML	Unified Modelling Language
UNECA	United Nations Economic Commission for Africa
UP	Unified Process
VGI	Volunteered Geographic Information
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service
XP	Extreme Programming

List of Tables

List of Figures

Chapter One: Introduction

1.0Introduction

Spatial Data Infrastructures (SDIs)have traditionally been defined as the totality of "technologies, policies, standards, human resources, and related activities necessary to acquire, process, distribute, use, maintain, and preserve spatial data throughout all levels of government, the private and non profit sectors, and academia (FGDC 1997). However different regions of the world are adopting more contextualised definitions because of different perceptions of SDIs and the role that they play in communities (Maguire and Longley, 2005). In some instances, SDI implementing agencies put more emphasis on policy that governs that sharing and use of spatial data and less on the technologies that facilitate the whole process. This results in the creation and recognition of policy documents first and later the establishment of the SDI enabling platforms through technology. Others put emphasis on the technology and ignore the necessity of cementing data sharing social networks and standardisation of procedures.

Despite these varying notions of the SDI concept around the world, the value of an SDI is demonstrated in its support to respond to perennial problems such as poverty, disasters, urbanisation, utilities and the environment (Nedovic-Budic et al., 2009, Williamson, 2003, Masser, 2005). In Nedovic-Budic et al. (2009) the authors refereed to the remark below by Craglia and Johnston (2005):

> Many of the challenges of contemporary society, such as protecting the environment, increased security, better transport, 'socially just' or 'sustainable' development and enhanced services to citizens, require decision makers to identify where need is greatest. To effectively target intervention, monitor outcomes, and assess impacts, access to Geographic Information (GI) is crucial. Ideally, it should be easy to identify who owns GI, whether it is fit for the purpose in hand, how it can be accessed and integrated with other information.

While the anticipated function for SDIs is to play a critical role in addressing the societal challenges at a local level of governance, they have not had the desired impact at this level (Nedovic-Budic et al., 2004).

1.1Background

SDIs have been evolving over the years. The first generation of SDIs, which were pioneered in the United States in the early 90s, resulted in the emergence of Clearinghouses which facilitated access to raw spatial datasets (Groot, 1997). The second generation resulted in the emergence of Geo-Portals, that facilitated access to services which were based on the data and not necessarily the raw datasets (Maguire and Longley, 2005). There is a big debate in SDI research on the possible characteristics of the next generation of SDIs. A number of scholars support the assertion that this next generation of SDIs will thrive on user-generated content in the form of Volunteered Geographic Information (VGI) (Budhathoki et al., 2008, Coleman et al., 2009, Goodchild, 2007, Fernández and Iglesias, 2009, Schuurman, 2009b).

1.1.1 A review of SDI developments in Africa

In Africa, formal SDIs have also been developing but at a seemingly slow pace (Lance and Bassolé, 2006). Like the rest of the world, in the few instances where they are being developed, SDI initiatives are mainly pioneered by government agencies.

A study on the status of SDI implementation on the African continent (Makanga and Smit, 2008) showed that SDI implementation is still in its infancy with most countries still trying to establish the basic frameworks upon which their SDIs will be implemented. There is very little political support for formal SDI initiatives on the continent (Clarke, 2009) and therefore, although there are a number of SDI coordinating bodies their SDI projects don't seem to take off the ground. When this study was conducted there were only 3 operational Clearinghouses on the continent; one more than when a similar assessment was done on a global scale in 2003 (Crompvoets and Bregt, 2003). The summary of the assessment is given in Table 1.1

Table 1.1: Overall status of SDI implementation in Africa

Indicator Category	Central Africa	North Africa	East Africa	Southern Africa	West Africa
NSDI Coordination	1 out of the 3 countries has a NSDI coordinating team.	3 out of 5 countries have a NSDI coordinating team.	3 out of 5 countries have a NSDI coordinating team.	6 out of 9 countries have a NSDI coordinating team.	6 out of 9 countries have a NSDI coordinating team.
Political Support	1 out of 3 countries has a reasonable level of political support	1 out of 3 countries has a reasonable level of political support	3 out of 5 countries has a reasonable level of political support	2 out of 9 countries has a reasonable level of political support	3 out of 9 countries has a reasonable level of political support
Funding	Gabon is the only country that has reasonable funding for NSDI	No country has reasonable funding for NSDI	Rwanda is the only country that has reasonable funding for NSDI	Swaziland is the only country that has reasonable funding for NSDI	Nigeria is the only country that has reasonable funding for NSDI
Stakeholder Participation	1 country has maximum stakeholder participation for the NSDI initiative	1 country has maximum stakeholder participation for the NSDI initiative	2 countries have maximum stakeholder participation for the NSDI initiative	1 country has maximum stakeholder participation for the NSDI initiative	1 country has maximum stakeholder participation for the NSDI initiative
Clearinghouses	1 NSDI Clearinghouse	No NSDI Clearinghouse	1 NSDI Clearinghouse	No NSDI Clearinghouse	1 NSDI Clearinghouse
Legal Component	1 country has at least a component of NSDI Legal framework implemented	2 countries have at least a component of NSDI Legal framework implemented	No country has at least a component of NSDI Legal framework implemented	1 country has at least a component of NSDI Legal framework implemented	2 countries have at least a component of NSDI Legal framework implemented

A score for each country included in the survey was generated which describes the extent to which the different countries have established the different components of their SDI according to the indicators column in Table 1.1 and the score results are shown in Figure 1.1.

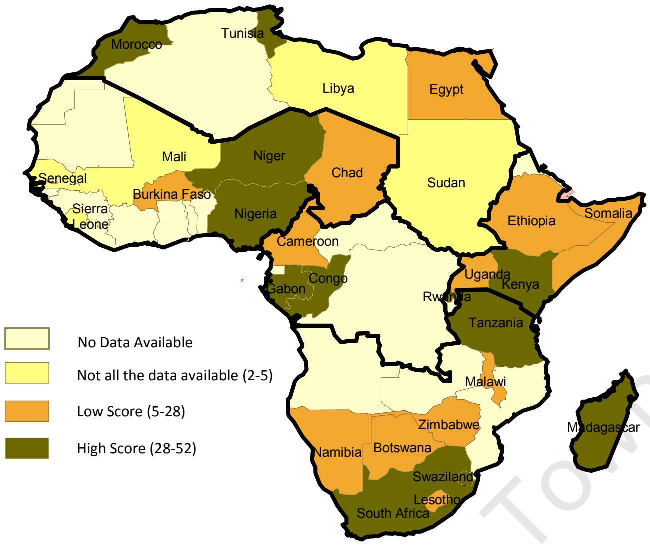


Figure 1.1: Map showing SDI implementation in Different African Countries

While this study showed that formal SDI activity on the continent in still generally in its infancy, there has been an emergence of a few informal SDI activities (Von Hagen, 2007) and these are mainly formed from networks of convenience among organisations that realise the need to share Geographic Information (GI) e.g. the Southern African Development Community's Regional Remote Sensing Unit (SADC RRSU). The challenge facing such informal SDI developments is in handing over these initiatives to governments to formally give them a national stature and priority (Von Hagen, 2007) therefore an access to funding streams from the national budget. All the same, the few SDI initiatives in existence are focusing on facilitating access to raw spatial datasets (through clearinghouses) and not really on the use of these datasets.

One of the main informal SDI activities is the Mapping Africa for Africa initiative that was commissioned in 2003 at a conference for the International Cartographic Institute (ICA) that was done in Durban South Africa. This initiative is premised by the fact that Geoinformation is critical for development and each African country need to have a basis set of fundamental datasets that will help in addressing the problems that they face. National Mapping Agencies, the academia, research institutes and NGOs have therefore had a key role to play in the determination of these fundamental datasets and also facilitating for their capture see (Mavima, 2004, Akingbade et al., 2004). It is envisaged that these fundamental datasets that are determined by the individual countries will play an important role in the process of improving the socio-economic status at community levels (Akingbade et al., 2004).

1.1.2 Responding to the Urban Problems in Africa

In order to start thinking of how SDIs in Africa will need to evolve or at the least emerge (in instances where there are no SDI activities); it is important to consider the nature of problems that an information infrastructure such as an SDI will actually aid to alleviate. In Africa, the main Local planning challenges revolve around problems inherent in human settlements with 62% of African urbanites living in Slums (UNHabitat, 2003). There is a global effort through the UN Habitat's Global Urban Observatory initiative to counter this trend in urbanisation (UNHabitat, 2009a).

The GUO has continually been helping local authorities to create Local Urban Observatories that serve to collect information that is relevant to understand the critical urban trends. This process is normally guided through the use of certain indicator sets that reflect on the pressing societal issues. The indicators are meant to inform policy and decision making as they expose the areas that require urgent attention. Indicators therefore become source objective evidence on the presence of pertinent urban trends. Evidence Based Practice (EBP) plays a critical role in responding to these challenges through the use of objective indicators that are tailor made for the communities where they are implemented (Holden, 2009).

This approach to countering urban challenges is widely accepted as an objective way of isolating the real problems that exist in society before responding to the symptoms. Local authorities normally think that they know the problem that is in hand when in fact they need to learn of the problem first and better understand how to respond (Pieterse, 2008). Indicators present an opportunity to explore and expose the problems by presenting reliable evidence on the problems that would have been determined by the community itself. In this regard, indicators are a very useful tool for Evidence Based Decision Making.

1.2Problem definition

While SDIs primary purpose is to facilitate data creation, sharing and ultimately use to help decision makers to be adequately informed as they response to societal problems, they generally do not seem to be developed in line with mainstream approaches to counter the urban problem that exists on the continent. Prominent SDI implementers on the continent (e.g. South Africa) are consumed in seemingly never ending efforts the create the enabling platform in the form laws and regulations while the data sharing and use, which constitute the main purpose of the SDI happen at a slow pace.

There is no evidence of any significant alignments between SDI initiatives and efforts to counter the urban crisis on the African continent.

The emphasis of MAFA is to create the necessary geospatial datasets that will help create better communities (Clarke, 2009) and then this data will be used as a tool towards sustainable development. The emphasis of the GUO is similar except that they do not put emphasis on Geospatial data but any datasets that help to report on the predetermined indicators. However, these two related initiatives can be done synergistically to achieve maximum results at the local governance level

The features that characterise most SDIs in Africa do not adequately help decision makers respond to the urban crisis. Whilst there is a huge potential in using spatial data as a resource in the move towards sustainable development on the African continent (UNECA, 2001) SDIs have not managed to reasonably facilitate its use to showcase urban trends and work 'upstream' in affecting decision making processes. They remain platforms that are used by people and organisations that need to use spatial data for their day-to-day business operation and not really as tools to communicate the problems on the ground to facilitate targeted and informed response strategies to the urban crisis. While the use of Geographic Information Systems as tools for Evidence Based Decision Making (EBDM) has been extensively explored in the past (Schuurman et al., 2008, Bedi et al., 2007), the possibility of strategically fusing EBDM into SDIs has not been fully explored, yet this is a meaningful characteristic if SDIs are to be any useful in changing the face of communities.

The features that characterise most SDIs have made them generally unfriendly to the ordinary user and therefore have attracted little attention from decision makers (van Oort et al., 2009). There is also very little acceptance among SDI implementing agencies (especially national implementers) of the need for a user driven approach to developing SDIs (Smit et al., 2009) whereas, this seems to be a new key trajectory in the evolution of other SDIs around the world.

1.3Research Objective and Questions

This research aims to explore and evaluate a new trajectory for SDI evolution, which incorporates mainstream urban problem solving strategies to make SDIs more relevant and useful in confronting the crisis of urbanization in Africa. It investigates the fusion of EBP with Spatial Data Infrastructures to make them more relevant to the problems faced at a local level, thereby achieving their intended purpose. The main assertion is that while access to raw spatial datasets is critical in an SDI, information products that are based on the data will be more useful in showcasing trends and presenting evidence therefore better informing decision making processes.

The key questions that this research seeks to answer are as follows:

1. **What key features should be incorporated into SDIs to make them more relevant in helping to address the urban crisis?**
2. **What SDI Architecture will meet the requirements for Sustainable Human Settlements initiative in developing Nations?**

1.4Research Hypothesis

The Hypothesis that guides this research is:

Incorporating Evidence Based Practice into Spatial Data Infrastructures using urban indicators will make them more useful in responding to the urban crisis.

1.5Scope of Research

This research seeks to reconceptualise the notion of SDI from being simply data access platforms to platforms for EBDM. Therefore, substantial effort goes into constructing a new conceptual framework for SDIs based on existing literature. The new concept is also presented using an Architectural design. The architectural design is implemented to demonstrate the feasibility of the newly constructed concept. The research however, does not go into too much detail to display the technical precision and robustness of the designed artefact as this does not help to answer any of the questions.

The Western Cape Province of South Africa is the testing ground and source for data used in this research, whereas the theoretical framework constructed and used to justify the proposed approach is taken from a body of literature on African Urbanisation in general. Therefore, it is anticipated that this new concept of SDI can be replicated and adapted in other developing nation contexts.

1.6Research Methods

The first question, which seeks to address issues on the characteristics of SDIs, is answered through a literature review. The first section of the review explores the problems birthed by urbanisation in Africa as well as the mainstream approaches to counter them. The second section reviews the evolution of SDI concept. The two sections are synthesized to construct a new perspective on SDIs that caters for the response mechanisms to the urban problems.

The second question concerning the architectural design of the SDI is answered through the inspiration of a philosophical framework based on the Design and Behavioural science paradigms in Information Systems research. The Reference Model for Open Distributed Processing (RM-ODP) is used as the methodological framework for conceptualising the SDI design from different viewpoints; while Software Processes are used to do the actual design (see Figure 1.2).

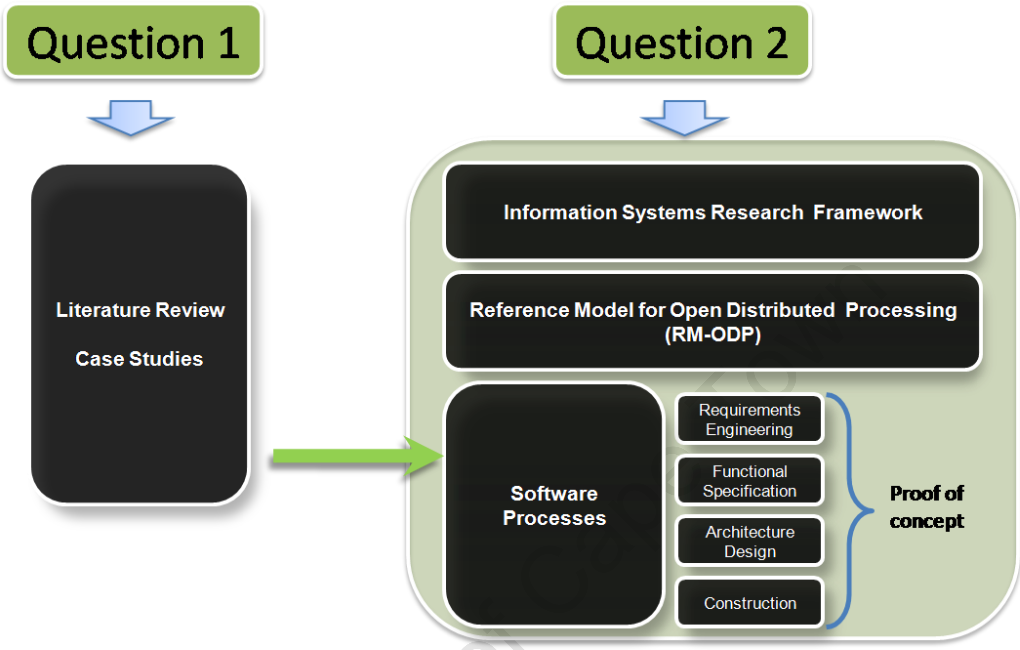


Figure 1.2: Research Methodology

1.7Thesis Structure

The rest of the thesis is structured in the following way:

Chapter 2: Response Mechanisms to the Urban Crisis

This chapter gives a concise review of the urban challenges facing developing nations coupled with a review of mainstream discourses to address these challenges.

Chapter 3: A new perspective on SDIs

An assessment of how the evolving SDI concept is relevant in confronting the urban crisis is done in this chapter. A new perspective for SDIs is proposed as is relevant to the urban challenges in developing nations.

Chapter 4: The Design Approach

This chapter describes the methods used to answer the two questions of this research. The use of these methods is backed up by a review of existing SDI design and software engineering approaches.

Chapter 5: Analysis Modelling

This chapter presents the first set of models that describe the main functional and non functional characteristics of the SDI. These models eventually form the basis of the architectural design in the following chapter.

Chapter 6: The Architectural Design

This chapter focuses on the discussion of the Architecture that meets the requirements modelled in the Analysis modelling phase of the project. The final architecture and the specific implementation technologies are discussed.

Chapter 7: Conclusions and Recommendations

This chapter summarises the key findings of the research and gives recommendations for successful implementation as well as areas that still require further research.

Chapter Two: Response Mechanisms to the Urban Crisis

2.0 Introduction

This chapter explores the main characteristics of urban settlements in the developing world as well as the mainstream response strategies to counter the problems that are experienced in these places. It serves as the basis for the discussions that are presented in later chapters. These characteristics are explored by drawing upon the work of the United Nations Commission for Human Settlements (UN-HABITAT) as well as some urban literature. Three cases are presented, two from the Western Cape Province of South Africa and one from Vancouver Canada. All three cases are part of the UN-HABITAT's Global Urban Observatory (GUO) initiative and are chosen from both the developing and developed world to draw some parallels and lessons. The last section of the chapter is dedicated to a discussion on four of the main features that characterise the efforts to respond to urban challenges.

2.1The Urban Crisis

Africa has the largest urban growth rate in the world. The continent hosts 20 percent of the world's slums and these have been increasing in number every year (UNHabitat, 2003). Most of the urban growth on the African continent will continue to take the form of slums and makeshift houses (Pieterse, 2009). The urban setting in Africa is bewildered by a myriad of problems that are a result of uncontrollable and unsustainable growth in the populations of cities. World over, there is a visible crisis for decent living conditions. This is however more pronounced in Sub-Saharan Africa, compared to the rest of the world, where close to 62% of the urban population lives in slums (UNHabitat, 2003). Figure 2.1 below shows the worldwide urban slums statistics.

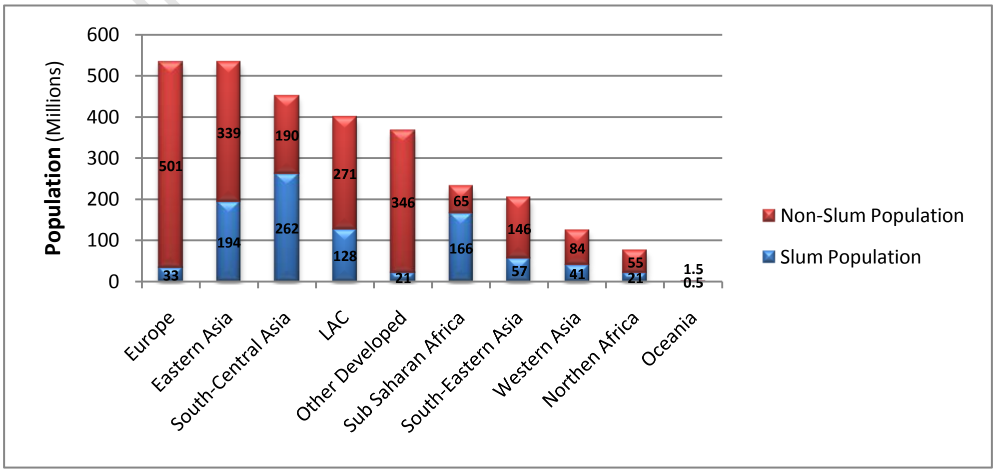


Figure 2.1: World Urban Slum Population (Source: UNHabitat, 2003)

The UN-Habitat (2003) defines a slum as "an area that combines, to various extents, the following characteristics: inadequate access to safe water, inadequate access to sanitation and other infrastructure, poor structural quality of housing, overcrowding and insecure residential status."

In developing countries, various methods are being employed by governments in an attempt to confront the threat of disaster that has emerged because of the growing urban population. Below are examples of how some African governments are reacting to the problem of slums (Turok and Parnell, 2009);

- Some appear reluctant to address the issue explicitly, having no obvious plan of action beyond piecemeal reactions to specific urban crises as they emerge
- Many believe that they cannot cope with burgeoning urbanisation and the way to limit the social dislocation and environmental effects is to reduce migration and promote rural development instead
- Providing a slight variant in seeking to divert growth pressures from existing cities into new towns elsewhere
- Some simply ignore the problem and watch the terror unfold

Seven of the eight United Nations Millennium Development Goals (MDGs) are meant to be achieved by 2015 (UNHabitat, 2009c). They respond to the world's main development challenges and they are drawn from the targets contained in the Millennium Declaration adopted during the United Nations summit in 2000. Target 4 of Goal 7 on the MDGs list is dedicated to achieving a significant improvement in the lives of at least 100 million slum dwellers by 2020. Considering the amount of work that has been done by the UN-Habitat so far, the UN and a number of scholars are sceptical as to whether the MDGs are going to be met in Africa. In fact, in 2005 Africa was the only continent not in line with meeting any of the targets of the MDGs (UNGA, 2005, IMF and WB, 2005, Easterly, 2009). All the same, assuming that the housing targets are met, only 10% of the total one billion slum dwellers would have gained access to housing (Riley et al., 2007). One would also question if building houses for the urban poor will adequately respond to the crisis of urbanization in developing nations. Is this a sustainable approach or does it simply address symptoms of a problem that is deeply rooted in inequality, segregation and poverty? A more radical approach to solving the urban problem is required for real progress on eradicating slums and the negative societal impact they have to be realised.

Informal settlements or slum dwellings must be perceived as a major characteristic of the African city (Pieterse, 2008). The premise behind this school of thought is that if 62% of the total urban population in Africa is living in slums, then it does not make sense to perceive the slum as an unwanted attachment to the formal city; it is the city! Governments have largely ignored the details of how informal settlements actually function, survive and continue to grow. Policy makers have generally not considered the delicate social networks and informal political structures that exist in these places. They tend to assume a too homogenous and uncomplicated view of the urban poor, whereas heterogeneity manifests itself in economic positioning, religious and ethnic identity, political inclination and many other issues (Pieterse, 2008). Intervention efforts are mainly in the form of bulldozing through these places in an attempt to integrate them into the formal city. Authorities assume that they know the problems that are inherent in slums when in fact they do not. Intervention should begin with trying to learn the actual problem that exists before prescribing solutions. Pieterse (2009) further notes that the African Urbanism is not adequately theorized. Knowledge creation is critical in order to understand the realm of the informal city. This will in turn allow for more objective and more informed intervention strategies. Learning city trends will potentially uncover underlying patterns and inform decision making and policy formulation towards the alleviation of urban problems and creation of sustainable human settlements.

2.2 Sustainability of Human Settlements

The Western Cape Province Human Settlements Reference Group in the Department of Local Government and Housing in South Africa defines Sustainable Human Settlements as: " Well-managed entities in which economic growth and social development are in balance with the carrying capacity of the natural systems on which they depend for their existence and result in sustainable development, wealth creation, poverty alleviation and equity" (HSRG, 2005). This localized definition is an adapted version of the definition of sustainability from a report given by the World Commission on Environment and Development (WCED, 1987). The definition appeared in a discussion report that was "aimed at redressing colonial and apartheid spatial planning and development through the delivery of socially, economically and spatially integrated housing delivery." It focuses on the aspired result of sustainability or what is expected as characteristic of the communities once they are transformed into sustainable settlements.

The National Science Foundation Workshop on Urban Sustainability (NSFWUS) proposed a definition of sustainability that has a slightly different focus to the one above. Below is an extraction from their report (NSFWUS, 2000).

> … We propose a definition of sustainability that focuses on sustaining lives and livelihoods rather than on the question of sustaining development. By "sustainable livelihoods," we refer to processes of social and ecological reproduction situated within diverse spatial contexts. We understand processes of social and ecological reproduction to be non-linear, indeterminate, contextually specific, and attainable through multiple pathways.

> Within the terms of this definition, sustainability: Entails necessarily flexible and ongoing processes rather than a fixed and certain outcome; Transcends the conventional dualisms of urban versus rural, local versus global, and economy versus environment; and supports the possibility of diversity, difference, and local contingency rather than the imposition of global homogeneity

This definition concurs with the viewpoint that human settlements are Complex and Adaptive Systems that are therefore not easily predictable. This means that it is very difficult to prescribe a solution but instead policy makers should cultivate appropriate crisis response mechanisms by responding to results derived from continuously learning and better knowing the problem. Like any other complex system, human settlements survive at the edge of chaos through constantly adjusting internal controls and variables to adapt to the environment that they exist in. Volatile human settlements like slums are never confined by the boundaries imposed by formal structures of policy and governance; instead these spaces are constantly working against any organised effort that threatens their existence. The NSFWUS definition suggests that sustainability of human settlements should involve the learning of trends over time and appreciating the behaviours of the different elements that make up human settlements. On this backdrop, it is highly unlikely that one can be able to prescribe uniform standards that measure sustainability in such volatile and uncertain environments. Instead, policy makers need to be agile and appropriately respond to both the short and long term threats that emerge from human settlements as they learn more from knowledge creation systems.

Knowledge is created through learning processes and requires continuous observation of trends which relate to specific themes of interest. Since learning is an open ended non linear process, there is a sense of uncertainty about what the outcome of the learning process will be. What one can be certain of though, is what he or she wants to learn at an abstract level. The results of the learning process can be objectively used as thematic evidence that can potentially interrogate existing policy or be used to create new intervention strategies and also create more relevant intervention policies. If evidence from the learning process is used to inform policy, energies will subsequently be channelled to where they are actually needed. A discussion on the learning process is presented later in this chapter.

2.3 UN-Habitat and the GUO network

The United Nations Human Settlements Program (UN-Habitat) is the United Nations agency for human settlements. It is mandated by the United Nations General Assembly (UNGA) to promote socially and environmentally sustainable towns and cities with the goal of providing adequate shelter for all (UNHabitat, 2009b). The UN-Habitat is playing a vital role at a global scale in fostering the creation of sustainable settlements. It is constantly working to quantify the world's population that lives in sub-standard housing conditions and proposing strategies to solve problems that result from the rapid urbanization of the globe. Once every two years, the UN-Habitat produces a report that presents varying statistics on urban settlements. This report has generally been a useful insight into the urban problems for governments and policy makers to consult for decision-making.

The UN-Habitat established the Global Urban Observatory (GUO) network in response to a need which called for ways to monitor global progress in implementing the Habitat Agenda and to monitor and evaluate global urban conditions and trends (UNHabitat, 2009a). The mandate of the GUO is to address the urgent need to improve the worldwide base of urban knowledge by helping Governments, local authorities and organizations of the civil society develop and apply policy-oriented urban indicators, statistics and other urban information. Currently 110 local and 24 national urban observatories are working in collaboration with the UN-Habitat to improve the collection, management, analysis and use of information in formulating more effective urban policies. The UN-Habitat is able to offer insight into urban issues at a global scale but does not have the authority to implement programs that address urban challenges at a local level. However Local Urban Observatories (LUOs) are better positioned to analyse local urban trends and provide evidence that can potentially influence policy and decision making.

2.3.1 Urban Indicators: Presenting the evidence

Urban indicators are a powerful way of presenting evidence that can potentially be used for decision making. They allow one to monitor whether or not, over time, policies have been successfully implemented. An indicator will often be a goal, a target, a threshold, or a benchmark against which one can assess change. Although indicators have traditionally been used to continuously monitor and/or evaluate progress of specific policy implementations, they can also be used to measure phenomenon that is not necessarily covered by policy to influence changes in the policy. In such a case, a secondary party will most likely be responsible for this and will be making an attempt to advice whoever is implementing policy to focus on other issues that may have been overlooked. Indicators are a powerful instrument for quantifying progress and presenting evidence relating to specific areas of interest. They can also act as a reliable basis for Evidence Based Decision Making.

Most of the literature on Evidence Based Decision Making (EBDM) is from the medical field. The phrase "Evidence Based Practice" was inspired by a specific approach in medicine called "Evidence Based Medicine" which was developed from the work of British Epidemiologist and medical researcher Archie Chochrane in the early 1970's (Stonor and Stutz, 2004). The Coalition for Evidence

Based Policy (CEBP) in the United States is a non-profit organisation whose mission is to increase government effectiveness through providing rigorous evidence as a means to making more informed decisions about policy formulation (CEBP, 2009). According to the CEBP, in the last 50 years evidence based policy has had significant impact in the field of medicine but in other areas of social policy like education, poverty reduction and crime and justice, government programs are implemented with little regard for evidence.

Evidence based practice is generally a four part process (Hamilton, 2003) which comprises of: understanding literature that backs up the subject of interest, creating hypotheses and doing the necessary research to address the hypotheses, sharing the results and finally having them reviewed. These processes result in the creation and use of accurate evidence. In the case of urban studies, indicators are incorporated into the second stage and are able to unambiguously present evidence. However, the major challenge facing the urban indicator movement internationally is how to successfully incorporate, the collection and reporting of indicators into decision making processes (Holden, 2009). This may be due to deliberate action by policy makers to ignore urban scholars who advocate for policy changes and yet they are not directly involved in the policy formulation and implementation. However, it makes sense to assume that if indicator systems are created and measured by independent parties they are likely to be objective and worth looking at. Independent analysts are most likely impartial and are therefore better positioned to facilitate the creation of unbiased information benchmarks than any other interested party. Indicator systems serve to drive home the message of the need for improvement in clear and specific domains (Holden, 2009) by providing evidence which is neutral.

There are three different types of indicators that serve different purposes: management indicators (on input, process, performance, output, etc.), policy indicators (outcome, impact, distance to target, etc.) and state indicators (objective, subjective, etc.) (Van Assch and Reynaert, 2009). Regardless of the type of indicator, all of them serve to present a measurement from a specified benchmark, which works to assess whether a program or policy implementation is working in a way that it is intended to or not. These measurements are made on data; data that may have existed before the design of an indicator system or data that is created specifically to be used with a set of indicators.

The next section is a review of the work at the City of Cape Town, The Cape Urban Observatory (CUO) and the Regional Vancouver Urban Observatory (RVU) to illustrate the use of urban indicators as tools for presenting evidence in urban research and how this is in turn used to affect policy formulation and revision. These initiatives are chosen to illustrate the use of indicators as tools for presenting evidence in different urban settings, that battle with different challenges. The two urban observatories are working directly with the GUO network, whilst the City of Cape Town is working indirectly through the South African Cities Network.

2.3.2 The City of Cape Town

The City of Cape Town (CCT) municipality's approach in responding to the human settlements crisis in the city has been chosen because Cape Town falls within the study area of this research and is the largest municipality in the Western Cape Province of South Africa. Cape Town like most African cities is battling daily with the challenges that result from urbanization. For example there has been a steady increase in the number of shacks in the city as shown in Figure 2.2.

The Strategic Development Information and Geographical Information Systems (SDI & GIS) department at the CCT supplies the city with information and knowledge that is necessary to address the development needs of Cape Town. Their main functions are:

- providing information, research, surveys and policy support for the City's strategic and urban development processes and projects, such as the Integrated Development Plan (IDP)
- providing and co-ordinating spatial information for the City
- developing the City information and knowledge strategy, and improving the accessibility of information
- planning and managing a Knowledge Resource Centre within the City to encourage knowledge sharing and facilitate easy access to information

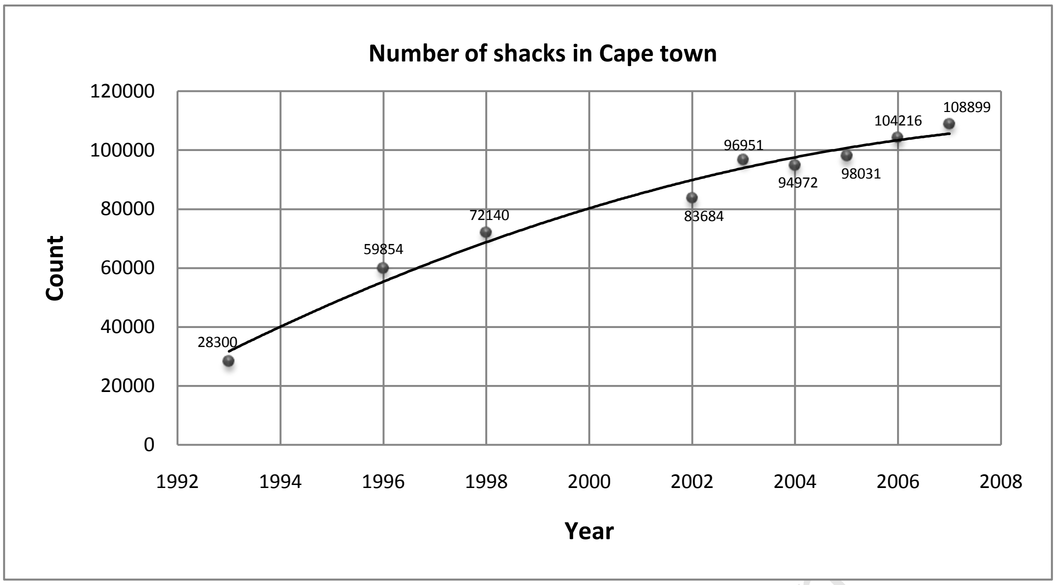


Figure 2.2 Numbers of shacks in Cape Town (from 1993-2007). Source (CCT, 2009)

To streamline their efforts of analysing and responding to the City's urban problems, the SDI & GIS department is using a set of nine indicator groups, which are further broken down into 29 more specific indicators (Table 2.1). The department is continuously creating new data and getting data from the statistical authority in South Africa (StatsSA) to report on the different indicators and inform them on the effectiveness of their policy implementation and also reveal patterns that were otherwise unnoticed. Two composite and more generic indicators have also been generated from this comprehensive list: the Socio Economic Status (SES) index and the Service index.

The SES index combines indicators on education levels, unemployment rate, poverty and unskilled labour whilst The Service index combines indicators on informal dwellings, access to electricity, access to a flush or chemical toilet, access to clean drinking water and weekly refuse removal. All the indicators are assumed to have equal weighting therefore the two indices are simply an arithmetic mean of the constituent indicators.

The SDI &GIS department publishes reports about the different indicators on their website, in some instances showing the status quo of the different themes under scrutiny and in some instances trends over time. Spatial data plays a critical role in presenting the findings of the departments' research. There is a map in PDF format that can be downloaded for almost every indicator, revealing the spatial patterns of the different phenomenon under study.

Table 2.1: Indicators for the City of Cape Town (CCT, 2009)

Demographics		**Year**
Total area	2,461 km2	2007
Population (total)	3,4 million	2007
Number of households	904,000 (estimated)	2007
Total lenghth of coastline	294 km	2006
Economy		
GDP	R130 billion	2007
% Unemployment	16.9	2007
Education		
Literacy-% of people aged 20+ with grade 5 education or less	8.4	2007
Highest level of education for people aged 20+	Below Matric: 58.3%	2007
	Matric: 23.6%	
	Post Matric: 16.8%	
Number of Libraries	108	2008
Health		
The prevelance of HIV as per Antenatal Survey	18,2%	2006
The incidences of TB in Cape Town per year (cases and deaths)	Cases: 26,754	2005
	Deaths: 2,122	
Living Conditions		
Number of informal dwellings sericed by the city	108,899	2007
% of households living below the poverty line (<R1,600 per month)	38,8%	2005
% of households with no access to safe drinking water on site	6,7%	2007
% of households with no access to adequate sannitation	5,8%	2007
% of households with no access to electricity for lighting	2,8%	2007
% of households with no access to at least weekly refuse removal	4,4%	2007
Number of Parks and Reserves	144	2006
Indigent population	201,867 households	2008
Safety and Security		
Number of murders reported per 100,000 population per year	60	2008
Number of reported rape cases per 100,000 population per year	73 (Apr to Dec 2007)	2007
Number of drug related crimes per 100,000 population per year	830	2008
Number of commercial and/or industrial crimes reported per 100,000 per year	464	2008
Number of Police stations	62	2006
Number of Fire stations	29	2006
Transport		
Public and Private transport into the Cape Town CBD in a day	Buses: 4%	2004/5
	Minibus taxis: 11%	
	Metered taxis: 1%	
	Rail: 17%	
	Cars: 67%	
Environment		
Number of days having particulate matter (PM10) exceedances	City centre: 7	2007
	Goodwood: 25	
	Khayelitsha: 86	
Indigenous Plant Species	2,621	2006
Tourism		
Number of International and domestic **tourists** (Western Cape)	International: 1,763,631	2007
	Domestic: 5,5 million	2006

Since the CCT is the local authority directly responsible for creating and implementing policy in an attempt to turn Cape Town into a sustainable city, it is tempting to question the objectivity of their indicator system. This is not to rob the City of the credit it deserves for visible efforts to counter the urban crisis in Cape Town, with the City's' Mayor being voted the World's mayor of the year in 2008 for her efforts to make Cape Town a liveable city (WorldMayor, 2009). This is an indication that there is significant effort being made to achieve sustainability of livelihoods in the City. However, Independent analysts are still in a better position to create unbiased, more objective and trustworthy indicators than the policy makers themselves (Holden, 2009).

2.3.3 The Cape Urban Observatory (CUO)

The CUO's overarching purpose is to contribute to the development of sustainable human settlements in the greater Cape Town area. This will be achieved by facilitating evidence-based decision-making, and improved collaboration and learning by providing a public platform for the storage, dissemination and analysis of timely and reliable geo-spatial information, and analysis on themes relevant to integrated development planning (CUO, 2009). The CUO was launched in July 2008 and at the time of writing, it has served to network people and organizations that have data that is relevant for understanding urban trends in the Western Cape Province of South Africa. This has resulted in the creation of a forum, which consists of Government departments, and research organizations that have been capturing data on different themes over time. The Forum provides overarching strategic direction and guidance for the observatory. Regular meetings ensure that the Forum is closely involved in the development of the CUO. The City of Cape Town and its 6 adjacent municipalities are represented on the forum.

The CUO has adopted a 'three-way convergence' model to determine the set of indicators that are useful to understand the trends associated with different themes within human settlements in the Western Cape Province of South Africa (see figure 2.3). This model was developed by the CUO as part of a process to create a Monitoring, Evaluation and Reporting Framework for a Sustainable Development Initiative for Grabouw, a small town in the Western Cape Province. This approach has resulted in the identification of a list of benchmark indicators that are currently being refined through consulting the different stakeholders represented by the CUO Forum (see Appendix 1). The approach facilitates the creation of relevant indicators by considering; mainstream indicators that are being used around the world; local needs through community participation, and an analysis of the available data from existing and possible sources.

Although there are high levels of collaboration with local authorities and other relevant institutes, it is anticipated that the CUO will play the lead role in speaking out about the urban trends through an evidence based public platform. At the time of writing, the CUO is collecting data from different municipalities to host them on a central server at the University of Cape Town (UCT) and is also documenting metadata for these data. These datasets will be accessible on the internet as public domain information and will also be used by the Observatory to perform different analysis, including the calculation of indicators and comparison of different indicators over time.

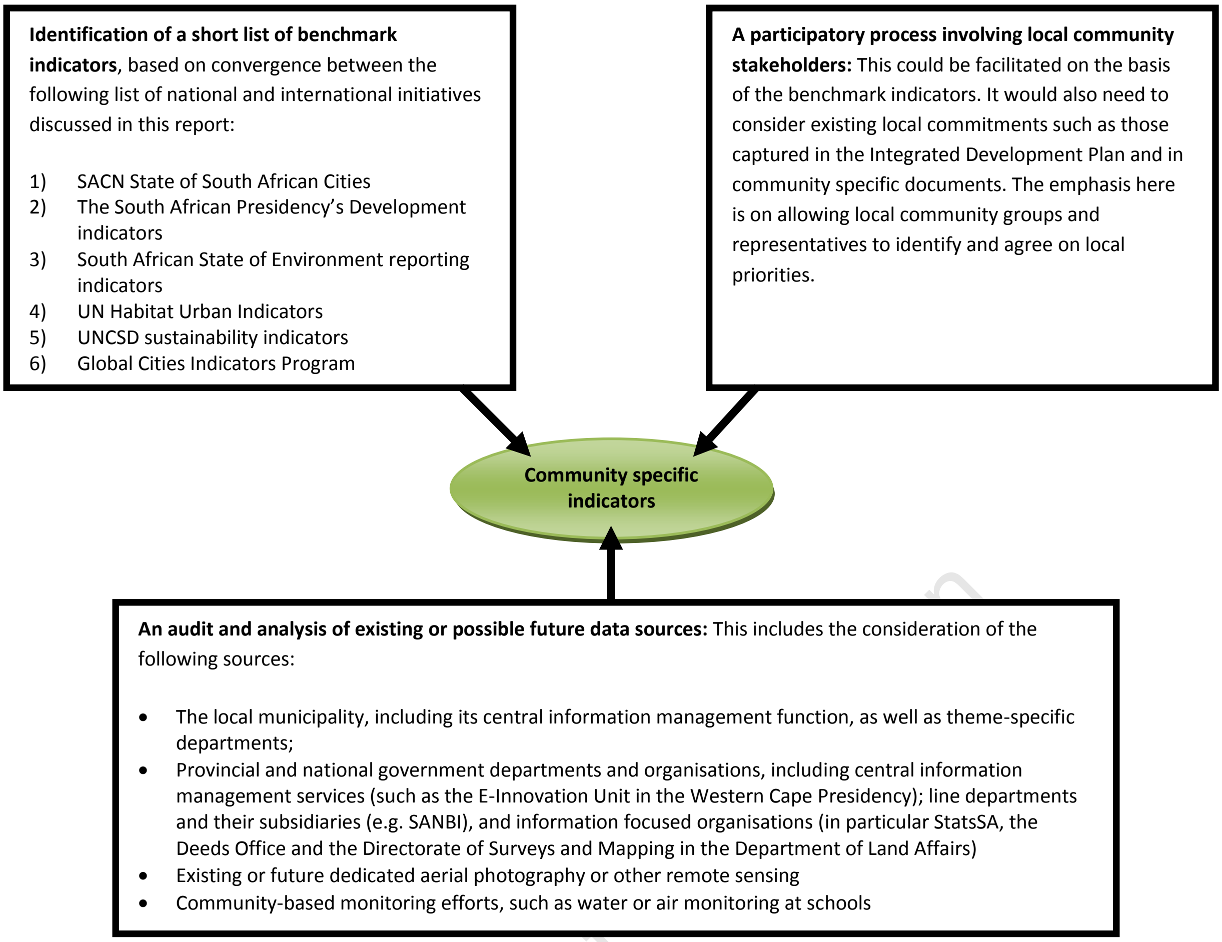


Figure 2.3: 'Three-way convergence' approach for determining indicators (EEU, 2009)

2.3.4 The Regional Vancouver Urban Observatory (RVU)

The Regional Vancouver Urban Observatory (RVU) was established in 2004 as part of the UN-Habitat GUO network. It is the first Local Urban Observatory from the developing world (RVU, 2006b). The vision of the RVU is to achieve a Vancouver region in which public decisions are informed by information that is relevant and in which there is accountability for dynamic targets aimed towards a global sustainable future (RVU, 2009). At Inception, the first priority of the Observatory was to convene a region wide process of dialog with representatives from citizens, business, academia and civil society. The purpose of this was to facilitate a participatory approach to learning about sustainability indicators locally and from global experiences, in order to eventually create relevant indicators for the region. This resulted in the set of eight indicator categories listed below which represent the key issues that the region is concerned about: Mobility, Poverty, Economy, Governance, Community, Environment, Food Systems and Arts and Culture. Each category has 3 indicators making the total number of indicators 24 (Appendix 2).

According to its inaugural report (RVU, 2006a), the whole RVU process is summarised in the three phases below:

i. Engaging academic researchers to investigate the state of knowledge about the use of indicators

ii. Involving the public in a participatory process to define the focus areas and also decide on the indicators

iii. Aligning the work in (i) and (ii) above to the objectives of the UN-HABITAT Millennium Development Goals

It's interesting to note that unlike the CUO, the RVU seems not to have valued the need for a pre-audit of the data that would be available from the different sources to decide on the indicators. It probably did not foresee problems with access to data as much as the CUO did, considering the environment that it is working in. One would naturally be inclined to worry about data when trying to get information about informal settlements where the majority of African urbanites live, and which are characterised by rapid changes daily, than in developed nations which are relatively well built, well documented and less volatile.

However not all the data that was required by the RVU was in fact available. For example, the Counting on Vancouver (2006a) report remarks that "Results of an opinion poll conducted in 2005 showed transportation to be the 'number one issue' for Greater Vancouver residents but existing data does not go deep enough to reveal many of the key issues and trends related to mobility and they do not do justice to the integral linkages of mobility trends to data in health, housing, urban planning, employment, income, social behaviour, and politics." There are other examples in the document that show unavailability of adequate data to answer some of the questions posed by the indicators. The above mentioned facts are useful evidence to advocate for the capture of these datasets by the relevant authorities, as it is a testimony of the fact that the responsible authorities may not be catering for some of the community needs in their policies. This may result in a bottom up creation of relevant data that can potentially result in better and more informed decision making for the benefit of citizens.

2.4 Summary: Key features for urban crisis response

This chapter has presented the main characteristics of the problems that are evident in human settlements. A number of key features have been identified as crucial to ensuring appropriate response to the challenges posed by human settlements. These can be summarised in three points; Indicators, Participation and Ongoing research/ Continuous Learning.

2.4.1 Indicators

Indicator systems are a key feature of unambiguously presenting evidence on specific trends of interest in the city. This is evident in the three cases that were presented where in all three instances indicators were critical elements for assessing progress with regards to policy implementation and also identifying trends overtime. Indicators are being used as sources of objective evidence to expose the realities that exist in cities. It is important that any effort to counter the challenges that are faced by urban areas should be directed by evidence on the ground and urban indicators are a useful tool in that regard.

2.4.2 Participation

There should be active participation of the communities in responding to urban problems. This can be at two levels; the first is in creating relevant data that can be analysed to reveal hidden or unknown but useful patterns and the second is in getting to know the main issues that trouble communities so as to focus energy and prioritise these issues. Communities can participate through community representatives that they appoint (EEU, 2009) or technologies can be used to foster individual participation to create data ubiquitously (Goodchild, 2007).

2.4.3 Ongoing Research/Continuous Learning

The process of learning is a three part cycle consisting of Romance, Precision and Generalization (Whitehead, 1916). Romance is a stage where the learner explores the wealth of possibilities on a new subject. In this period the learner should become fascinated with the broad significance of the idea and actively pursue the more precise investigations that follow (Denton et al., 2002). Precision involves the learner focusing on refining their knowledge about the subject through rigorous research, intensive and relevant data collection and analysis. In the Generalization stage the learner is guided to discover the worth of their learning efforts and appreciate the realised patterns, meaning and general applications. These patterns uncover new questions prompting the learner to pursue these related topics and the cycle goes on.

Considering that ongoing research is a critical component of EBDM (Hamilton, 2003), agents that are responsible for responding to the urban crisis should be in a position to accept that there is a lot that needs to be learnt and knowledge that needs to be created in order to make more informed decisions regarding human settlements. Therefore knowledge creation and continuous research should be strategically embedded in the approaches to counter the problems in human settlements.

2.5 The role of Geo-Information

Geo-Information or spatial data has a long history of being used as a communication tool for a wide range of phenomena that occur on the surface of the earth. Geo-Information is also at the heart of spatial analysis using Geographic Information Systems, which has a lot of applications in urban systems. Maps are able to tell stories and display trends that are not immediately apparent in tabular or other forms of data to better inform policymaking. According to Bedi et al.(2007) maps are "more than just pretty pictures" but are in fact useful tools for presenting poverty related data to inform the formulation of policy.

In the learning cycle, Geo-Information has traditionally been critical in generalizing understandings of urban problems and presenting evidence in ways that can effectively reveal an urgent need for responsive action; informed and targeted action (Figure 2.4). It reveals patterns in situations that exhibit a difficulty in understanding the causes and effects of phenomenon. It facilitates the immediate identification of spatial patterns and therefore stimulates spatial thinking. Spatial thinking comprises the knowledge, skills and habits of mind, to use concepts of space, tools of representation and reasoning processes to structure, solve and express solutions to a problem (Bednarz and Bednarz, 2008).

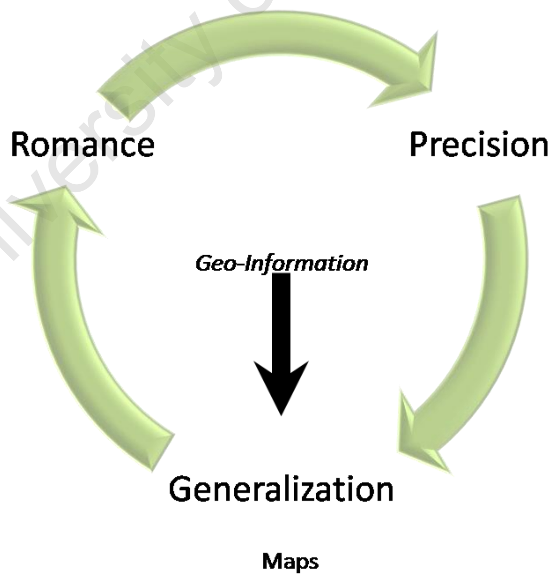


Figure 2.4 The role of Geo-Information in learning urban trends. Constructed from Whitehead's (1916) Cyclical learning process

There is a huge potential for the use of Geo-Information to provide evidence on trends in cities. Numerous organisations are simultaneously and continuously creating Geo-Information on multiple themes in cities, from government departments to private companies, research institutes to Non-Governmental Organisations. Without proper coordination, collaboration and well defined data sharing mechanisms, this data ends up being stored away after single use and remains undiscovered until it is eventually forgotten. Substantial financial investments go into creating data and other potential users can save money by reusing already existing data instead of duplicating data creation effort. Geo-information collected for specific purposes also contains stories and evidence that remains hidden if not used. There is a lot of economic potential in spatial data, but this potential is only realised by making the data widely available (UNECA, 2001). This is the reason for Spatial Data Infrastructures; to facilitate access to disparate spatial datasets to reduce duplication of data creation efforts and encourage the wider use of these datasets. However, this classical definition of SDIs has been evolving over time. The next chapter proposes a new trajectory for this evolution to ensure that SDIs are more useful in meeting the needs of decision makers by helping them better understand and respond more appropriately to the urban challenges that communities face in the developing world.

Chapter 3: A new perspective on SDIs

3.0 Introduction

This chapter advances an argument for the use of SDIs as platforms for evidence-based decision making in efforts to develop sustainable human settlements in developing nations. It begins with a discussion of how SDIs have been evolving from the time that the concept was first conceived. While the landmark features of the evolution apparently appear to be largely technological, the role of policy, institutional arrangements, and general purpose for an SDI have not changed much from when the concept started. These elements therefore do not come to the fore in the discussion. The section on the SDI evolution is followed by a discussion on the changing role of the user of spatial data from just being a mere recipient of data to being an active participant in the data creation process. The principles of Wikinomics are also explored in order to understand some recent innovations in data sharing and collaboration.

To get an idea of the current SDI models some examples are given and their characteristics described. The concluding section of the chapter draws from the main discussion in this chapter to determine a new trajectory for the evolution of SDIs to be more useful infrastructures in responding to the main challenges faced by urban settlements. This is done by incorporating some of the features described in Chapter 2 which enable them to be used as platforms for Evidence Based Decision Making.

3.1 The Evolution of the Spatial Data Infrastructure concept

SDI features have been evolving from the time that the concept was first conceived. The first generation focused on data discovery, the second generation on service delivery (Maguire and Longley, 2005) and the emerging third will potentially thrive on user generated content (Budhathoki et al., 2008).

3.1.1 First Generation of SDIs

Initial efforts to develop Spatial Data Infrastructures (SDI) started in the late 1970's when many national mapping agencies in Europe began to recognize the need to justify the large public investments they had received (Groot, 1997). To achieve this, the mapping agencies needed to improve access to and encourage a broader use of spatial information in their custody; as a result there was a need to come up with new strategies that would facilitate access to these data resources.

However, more visible efforts to establish SDIs were seen in 1992 when the Clinton administration assumed power in the USA with one of its key objectives involving the creation of a National Information Infrastructure (NII) (Dutton and Peltu, 1996). The main purpose for this initiative was to broaden the activity of the US government in promoting new ICT products and services to address major social and economic objectives. This resulted in the NII pioneering the 1st generation of SDIs which were mainly data driven, providing public access to National Spatial Data holdings in disparate locations through the internet. Metadata clearinghouses emerged as the main feature supporting this function for this generation of SDIs.

Metadata is normally defined as data about data (GSDI, 2004). Geospatial metadata are structured facts that describe spatial information, or information services (ANZLIC, 2009). According to the GSDI cookbook (2004) there are three key levels of metadata these are stated below as well as the questions that they try to answer:

> **Discovery metadata** - What data sets hold the sort of data I am interested in? This enables organizations to know and publicize what data holdings they have.
>
> **Exploration metadata** - Do the identified data sets contain sufficient information to enable a sensible analysis to be made for my purposes? This is documentation to be provided with the data to ensure that others use the data correctly and wisely.
>
> **Exploitation metadata** – What is the process of obtaining and using the data that is required? This helps end users and provider organizations to effectively store, reuse, maintain and archive their data holdings.

Documenting spatial metadata helps organizations that own spatial data to manage it more efficiently. Normally if people who created the spatial data for an organization leave, they leave with the value of the spatial data if it's not documented. This can result in unnecessary duplication of effort by recreating already existing datasets.

Metadata records have the potential to be repositories for qualitative information about both quantitative and qualitative spatial and non-spatial attributes (Schuurman, 2009a). This is through how metadata provides contextual information that helps the user to evaluate the quality of information that they can extract from the data. Metadata Standards describe the mandatory elements that should be captured to make them meaningful and useful to whoever will use them. One benefit of standards is that they have been developed through a consultative process with other experts (GSDI, 2004) and therefore they are normally exhaustive of the need of the community of spatial data users. Standards allow for uniformity in structure of metadata record, which in turn increases interoperability, standardizes and makes easier the process of discovering the metadata.

A spatial data clearinghouse can be defined as an electronic facility for searching, viewing, transferring, ordering, advertising, and disseminating spatial data from numerous sources via the Internet (Crompvoets and Bregt, 2003). Metadata standards as well as interoperability standards like the Catalogue Services for the Web (CSW) (OGC, 2007) have facilitated collaborative efforts for the creation of spatial data clearinghouses. These are still key components of any SDI and are the main gateway to spatial information in disparate data holdings.

3.1.2 The Second Generation of SDIs

Geoportals were the landmark feature of the second generation of SDIs (Maguire and Longley, 2005). Beyond facilitating access to spatial data on the internet, Geoportals have managed to allow for use of this Geoinformation on the web through aggregating map based content from a variety of sources. They also allow for users to consume GIS services online and facilitate viewing of maps through web map viewers (Goodchild et al., 2007). Unlike the first generation of SDIs which were mainly data driven, the second generation were more process driven (Budhathoki et al., 2008). These changes reflect some of the early visions of SDIs which were centred on the use of information rather than mere discovery. The former US Vice President Al Gore in reference to the Landsat Satellite reflects this in a statement:

> "...The Landsat program, designed to help us understand the Global environment, is a good example. The Landsat satellite is capable of taking a complete photograph of the entire planet every two weeks, and it has been collecting data for more than 20 years. In spite of the great need for that information, the vast majority of those images have never fired a single neuron in a single human brain. Instead they are stored in electronic silos of data. We used to have an agricultural policy where we stored grain in Midwestern silos and let it rot while millions of people starved to death. Now we have an insatiable hunger for knowledge, yet a great deal of data remains unused (Gore, 1998)."

The real use of data is what was actually needed to achieve any visible form of impact in societies. By the turn of the 21st century, SDIs had started evolving to more service oriented architectures. Changes in web computing catapulted this at the same time that Geoportals emerged as tools to facilitate both discovery and use of spatial data on the internet. In this evolution of SDIs, open standards had a huge role to play in enabling spatial data to be used on the internet. Geoportals are an essential part of most SDIs now, whether local, national or regional (e.g. http://www.inspire-geoportal.eu/, http://gos2.geodata.gov/wps/portal/gos). They provide the capabilities to query metadata records for relevant data and services, and then link directly to the on-line content services themselves.

In most cases, governments have been directly responsible for SDI implementation programs in different countries around the world. The efforts to establish SDIs has resulted in NSDI policies and strategies at the national level of governance; prescribing standards, procedures and legislation that guide implementation at all levels of governance from national to local. However NSDI strategies do not meet the needs for local planning, where real societal impact is realized (Nedovic-Budic et al., 2004). The heterogeneity that characterizes cultures, politics and landscapes in different locations within the same country is an immediate indication that different strategies would be required to intervene into these diverse spaces. Governments mainly determine the general national goals and policies but how these are achieved in the different places is normally left to the local authorities.

In an investigation of the relationships between municipal, provincial and national government SDI implementers in South Africa, Smit et al (2009) demonstrated that Local government appears to be more progressive in accepting a user driven process than the national government and that the latter is inherently entrenched in the establishment of the overarching framework within which SDI activities are to take place. This bottom up and user driven approach to generating and using spatial data is the typical inclination of the emerging third generation of SDIs, which will potentially thrive on user-generated content (Goodchild, 2007, Schuurman, 2009b, Elwood, 2008, Budhathoki et al., 2008).

3.1.3 Third generation of SDIs

The evolution of the web is influencing new perceptions of SDIs because this is the technological platform upon which they function. The web has evolved from the first generation (Web 1.0) where communication was mainly one directional with web content mainly being authored by publishers for consumption by the public. It was a read or write web (Yihong, 2007). Web 2.0 broadly refers to a new generation of Internet services and technology. It allows for participation by the user community through contribution of information onto the web. While there is an ongoing debate on the characteristics of Web 2.0 (Anderson, 2007), the participatory Web as opposed to 'the Web as information source' is frequently attached to descriptions of Web 2.0, of which user-created content and collaboration are the hallmarks (Schuurman et al., 2008).

Cloud computing and Wikitecture, which are key features of Web 2.0 are revolutionizing the way we present, share and analyse Geographic Information (Hudson and Crooks, 2008). Cloud computing refers to a style of computing where massively scalable IT-related capabilities are provided as a service using internet technologies to a wide variety of clients (Gartner, 2008). Wikitecture is the architecture behind websites that function through voluntary content generation by the user community. These two principles have facilitated voluntary participation by citizens in the creation of Geographic information on the web in the form of Volunteered Geographic Information.

Applications like Google Earth and Wikimapia function on this principle of Volunteered Geographic Information (VGI) which was defined by Goodchild (2007) as the harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals. GIScience (more specifically SDI) research is sparking a growing interest in the VGI phenomenon. There is huge potential that VGI will aid in the creation of more relevant datasets as part of existing SDIs (Budhathoki et al., 2008, Fernández and Iglesias, 2009). Although there are still substantial research efforts needed to determine the societal impact of such an approach to SDI implementations (Elwood, 2009a), immediately apparent is the huge potential that this approach has in creating more timely, relevant and useful spatial datasets at the local level, therefore serving as a resource to augment, update or complete existing formal spatial datasets (Goodchild, 2007).

The Geoweb is facilitating wider use of Geo-Information and Geo-Information services on the internet. The Geoweb is an integrated and discoverable collection of geographically related web services and data that spans multiple jurisdictions and geographical regions (Lake and Farley, 2007). This definition ties in closely with the definition of SDI according to the GSDI cookbook (2004) except for the fact that the GSDI definition of SDI mentions the relevance of policy in facilitating the sharing of spatial data. Policy was probably key in the GSDI definition because the concept emerged from the US government Information infrastructure initiative (Groot, 1997) which advocated for prescriptive data sharing mechanisms among government departments and this was engraved in policy. To understand the characteristics of the Geoweb, the definition given above can be broken into four parts:

1. a collection: indicates harnessing the value of information resources from diverse sources
2. integrated: indicates the fusion of these information resources to form a well woven data infrastructure
3. discoverable: gives the impression that the external environment can look into the infrastructure and explore the value stored in it
4. geographically related web services and data: the contents of the Geoweb have a strong location component and geography is an organizing principle for the data and services

SDIs are evolving to this definition of the Geoweb (Budhathoki et al., 2008). While literature on the Geoweb and Geo-Information echoes their use in analysis of trends in urban settings and presenting useful evidence on different spatial phenomenon, little has been said about how they have thrived on using geography as an organizing principle for information.

An organising principle is the logic by which work is coordinated and information is gathered, disseminated and processed (McEvily et al., 2003). It helps one to simplify and get a handle on a particularly complicated domain; for example, most modern cities are based on the organizing principle of the grid plan in order to better manage transportation and addressing.

Geography is the organising principle for Google (Jones, 2007) which draws ideas from philosophy, psychology and sociology to appreciate the importance of delivering meaningful information in a spatial context. Jones remarks;

> "The questions; what, where, when, why and who constitute the rod by which we divide the important from the unimportant in every aspect of life. A rich sense of where requires providing an interactive, exploration oriented user modality presenting information visually in its natural spatial context and supporting search and search refinements in the same context – In essence we have learned that context is everything. We aim to help users to rapidly select important information from less relevant information and we offer information analysis so that the user can discover implications contextually."

Context is a key characteristic that makes data and information make sense (Schuurman, 2009a). Geography as an organizing principle gives context and meaning to data and makes it more useful in understanding trends or patterns. Through the Geoweb, there is potential to aggregate content from many different and seemingly unrelated sources to abstract meaningful information through sorting the content using geography.

The emergence of the Geoweb is resulting in the redefinition of what spatial information really is. Traditional definitions of spatial data implied GIS data that could be used in a GIS and for mapping. The ESRI Shapefile (shp) and Autodesk's drawing (dwg) file are common examples of the format in which spatial data was found. The first generation of SDIs was meant to facilitate discovery of such datasets and services that are based on this data. However, the Geoweb is resulting in the redefinition of spatial data to any form of information that has a location reference. So this has resulted in information and data like georeferenced news feeds, photographs, web pages, videos and any other piece of information that has a location reference being termed Geo-Information. As a result of web standards there is more Geo-Information in different forms that exists electronically and can be integrated and used on a web platform. This in turn provides a broader information base for many forms of analysis that users might be interested in.

The Geoweb appears to be a more informal platform for data integration than the formal SDIs that are pioneered mainly by government and are used to search for data in formal data silos. However the principles that fuel the Geoweb have potential to be fused with traditional SDIs to create a hybrid bottom up and Top Down approach. The bottom up creation of Geo-Information is catalyzed by ubiquitous creation of spatial data by ordinary users of the web. Therefore, there is an emerging need to understand the changing role of the user in information generation; the role of both data user and producer played by organizations and individuals who have traditionally simply been recipients of information and services.

3.2 The User/Producer Convergence

The initial idea of SDIs focused on making spatial data available to the public. The structures that made this possible (Clearinghouses and Geoportals) are still not really relevant to people who don't require access to the raw geospatial datasets (Elwood, 2009a, van Oort et al., 2009). Geoweb applications like Google earth, Wikimapia, virtual earth, etc. are becoming popular with ordinary amateur (GIS) users mainly because these platforms are easy to use and also facilitate the voluntarily input of information by users. The active participation of ordinary citizens in data creation has to some extent made such public platforms more relevant to the ordinary user than the formal models for sharing data through the traditional concept of an SDI. The Geoweb thrives on the principles of mass collaboration and ubiquitous content creation through voluntary participation of specialized and amateur citizens

Wikinomics is a concept that also thrives on these principles (Tapscott and Williams, 2006). Organisations all over the world are adopting this approach and tapping into the intellectual capacity of the rest of the world, to build capacity and inform decision making in new and more efficient ways. Mass collaboration and crowd sourcing helps organizations to adjust to the needs of the community that they intend to reach out to by having them on board to contribute to the business operations. The user or consumer of a service or product has an active part in the production process and this is what is termed the **User/Producer convergence** in this research. The next section draws some lessons from four key principles of wikinomics, which echo the key principles of the VGI phenomenon and the Geoweb.

3.2.1 Openness

Traditionally, organisations have closed their internal mechanisms, daily business operations and intellectual property from the outside world. This has meant that end-users have been mere recipients of products and services. Opening up to the public however results in getting external insight on how to improve processes to meet the requirement of the end-user. Of course, this should be done in a cautious way, otherwise it would be "suicidal" to reveal top secrets to the wrong people, but the key is to have the end user's opinion on how to serve them better. An example of this is the Free and Open Source Software (FOSS) movement, which unlike proprietary software where the source code is hidden from the user, FOSS software allows everyone access to the source code and they can modify it to suit their needs. Therefore, software evolves faster and becomes more relevant to the users with significantly less effort from the originator.

This also applies to VGI and the Geoweb. Public platforms that are open for editing and contributions from the community of end-users result in the creation of datasets that would not have otherwise, been found at the mapping authorities or other organisations that create and manage data. In some cases, participants get access to the raw data, which is good enough motivation to participate (e.g. www.openstreetmap.org), but this would be a motivation for the specialized users. Even in instances where users don't have access to the raw data, but instead services based on their data (e.g. Google maps); they are still willing to volunteer information because they can immediately give their data through these open platforms and immediately see the value of their contributions .

3.2.2 Peering

Hierarchies and organizational structures dominate most organizations. This has resulted in rigid communication channels, protocols and operational procedures being at the core of how organizations function. Although structure is an important framework, its ability to adjust internal controls in response to fickle environmental factors is even more important (Dooley, 1997, Nogueira et al., 2000). Businesses operate in a dynamic environment and therefore they should naturally be agile to adjust to the uncertain nature of the environment they operate in. Structure should be placed in context with the operational environment at a specific epoch but should change and adapt to the environment as it changes (Ludwig Von, 1972). Peering is one radical way of breaking the barriers that are imposed by rigid structures. It involves deliberately targeting key resource people or organizations and collaborating with them for increased efficiency. An example is how Linus Torvalds of Linux collaborated with peers through the internet to make significant changes to the Linux operating system when it was launched in 1991. Linux is now perhaps the single largest collaborative project in the planet's history (Rivlin, 2003). Peering means tapping into the knowledge and experience of your peers for increased efficiency.

A 2008 study that was done to assess the status of National SDIs in Africa showed that formal NSDI implementation on the continent is still in its infancy (Makanga and Smit, 2008). This is mainly because most governments in Africa do not see the implementation of SDI as a priority as compared to other pressing issues (such as food, shelter, health etc). This is slowly changing, though the pace is just too slow considering the potential for such an initiative to fuel economic development in these nations (UNECA 2001). On another front there seem to be a lot of informal SDI activities that are fuelled by organizations that appreciate the value in sharing geospatial data (Lance and Bassolé, 2006, Lance, 2004). Examples are the SDI East Africa initiative (http://www.ungiwg.org/sdi-ea/?q=about) and the Southern Africa Regional Remote Sensing Unit (http://www.sadc.int/geonetwork). Both organizations are cataloguing data in East and Southern Africa but do not have the authority to officially sanction the NSDI activities in the areas that they operate in.

Although this may paint a poor picture of the development of formal SDI's on the African content, it certainly does not depict a disaster. SDIs on the African continent are being built from the bottom up using self-organized voluntary networks. This could mean that African data sharing mechanisms exhibit some important characteristics that are not available in other formal SDIs; meeting the needs of the user at the local level and having SDI activity more as a bricolage from the local level up, with organizations that need data corresponding in informal peer to peer social or professional networks. The challenge though is to transfer ownership of these informal SDI initiatives to government agencies (Von Hagen, 2007).

Open Geospatial Standards are able to facilitate this peering process in SDIs. These standards are used to create networks for data custodians to share data with the public or amongst themselves. However it should be noted that most SDIs have deliberately excluded informal or non government data custodians from these networks (DLA, 2003). Increased peering with informal data custodians can potentially increase the quantity of useful data circulating amongst those that need to use it.

3.2.3 Sharing

Sharing is also a non-conventional way of doing business. It is a "give and take", and when organizations share information appropriately, they can yield more from their synergy than they would do from not sharing. Wikipedia is a good example of the power of sharing information. By 2006, Wikipedia had ten times more information than Encyclopaedia Britannica its closest rival but still as accurate (Milne et al., 2006). Clearly, sharing software in the FOSS community has continually distributed development efforts. This property is inherent in VGI activities as well. Data is freely created and shared for most non-commercial VGI projects e.g. open street map.

In South Africa, private firms and Non-Governmental organizations are not recognized as data custodians (DLA, 2003). Therefore they are not mandated to fully operate within the confines of the SDI Act. For example, they are not obliged to document metadata for their datasets and or contribute to the formal data stores when they create new data. They normally keep data in their local databases and yet the data is potentially useful in addressing some of the key challenges facing the nation. The principle of data sharing as is inherent in the VGI phenomenon, should also be the heart of the evolving concept of SDI to facilitate maximum participation from all the people and organizations who have data, so that they are able to not only draw from national mapping agencies but also give back value added products.

3.2.4 Acting globally

The web is turning the world into a global village and this is an opportunity to learn from global trends. User generated content presents opportunities for users to have seamless access to information that will facilitate social and other forms of learning in ways that were not imaginable before their advent.

One example of an initiative that promotes a global experience to affect local decision-making strategies is the Urban Atlas Portal (UAP) (www.urbanatlasportal.org) which was launched by the Stockholm Resilience Centre in October 2009. Starting with twelve nodes (Figure 2.5), the UAP will serve as a platform where urban researchers will voluntarily upload information on various urban themes presenting their perspectives, focus and work at each of the participating urban sites. The UAP is in the form of a Geoweb application that serves as a learning platform where valuable urban research efforts will be published on a voluntary basis. The information presented by the portal ranges from map data to narratives in the form of text and graphics.

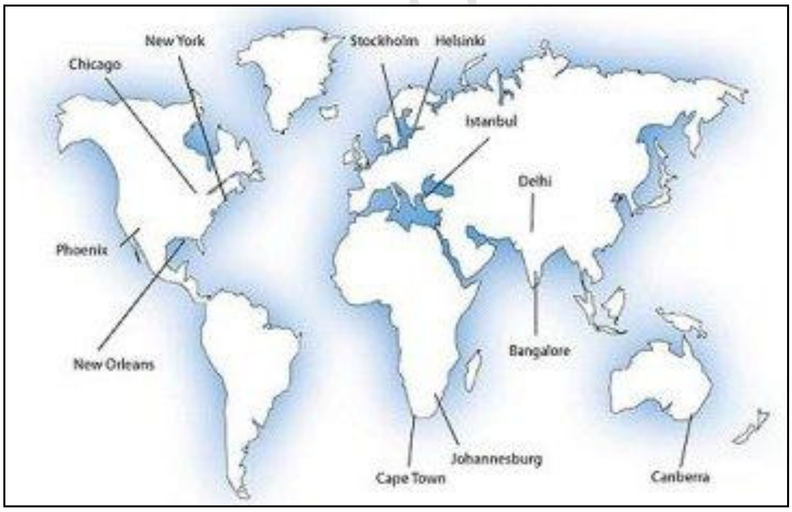


Figure 3.1: Nodes participating in the Urban Atlas Portal (SRC, 2009). Used with permission from the Stockholm Resilience Centre

The Regional Vancouver Urban Observatory mentioned earlier on in this chapter is also another initiative that draws lessons from Global experiences. This allows sustainability to not only be contextualized to a local setting but also be compared to the rest of the world.

The parallels between the VGI concept and Wikinomics presented in this section can powerfully revolutionize society by enabling access to large volumes of useful information that is contained in the minds of people and organizations. Careful thought however needs to go into coming up with appropriate strategies to capture this information in ways that will make it useful for informing various decision making processes.

The motivations for volunteering in Free and Open Source Software (FOSS) and Wikipedia have been well researched (Krishnamurthy, 2002, Raymond, 2001, Kittur and Kraut, 2008), but these may not be the same for the VGI phenomenon, considering that these are more formal and authoritative sources. Whereas some contribute VGI out of altruism, professional interest, intellectual stimulation and as a social responsibility, some do it out of mischief by replacing legitimate entries with nonsensical or overtly offensive content (Coleman et al., 2009). This results in inaccurate and misleading information which cannot be used for any meaningful analysis.

There is a need to assess the impact that different data providers will have on the quality of information gathered using VGI if it is to be used by decision makers. Will the information be representative of the status quo of the different areas of focus in societies? How useful will the information be in responding to the challenges faced in the different environments that it is generated? Will this approach not simply cement the social injustices and inequality profiles that affect society? These questions are critical and help assess the applicability of VGI in different situations.

Although participation by the citizens in unsustainable settlements is definitely critical and in line with mainstream urban discourses that advocate for participatory democracy as a means to addressing these problems, the people living in these areas may not access the internet in ways that enable them to participate appropriately. Therefore, having public internet based platforms of participation will most likely mean participation will come from the elite who are living in sustainable communities and not the target communities. Such a result has the potential to further alienate the already excluded demographic group of society (Crutcher and Zook, 2009).

In settings such as informal settlements, community representatives play a crucial role in ensuring that the voices of the people are heard and that government attempts to meet their needs. An example is the role played by stakeholder forums in sustainable development initiatives in South Africa (Hamann et al., 2008). Stakeholder forums consist of different representatives of people groups that have an interest in a particular community. They can be ward leaders, business representatives or political representatives. The people that they represent normally elect them. This is a platform to be heard and may not necessarily result in the creation of useful data but it can inform policy makers of the main issues that need urgent attention in communities and therefore allow for more informed responses. Community data can be found from organizations that are doing work in these communities; for example, an NGO that has a network of clinics in an informal settlement can provide data on different health themes for the community.

3.3 Examples of SDI activity

Table 3.1 below lists some examples of Geoweb platforms, relevant to this research, that are currently available on the internet. Six examples have been chosen, the first three that have features prominent in most SDIs and the last three are an example of the use of indicators in web mapping applications.

Table 3.1: Examples of selected SDI and related activity. The letters represent the following V=VGI, M=Metadata, D=Data Download, G=Map Viewer, I=Indicators, S= Web Services

Name	URL	V	M	D	G	I	S	Description of services
NSIF South Africa	www.nsif.dla.gov.za							The National Clearinghouse for South Africa which is a repository for metadata records. Users can search through distributed catalogues but are not able to download data
INSPIRE	http://www.inspire-geoportal.eu/							An inventory of datasets available in the EU specifically for chosen themes that present different environmental phenomenon. Map viewer with ability to add map services
US NSDI	www.fdgc.gov/nsdi							The US Clearinghouse with a map viewer and access to map services for consumption on other platforms and in multiple formats
Sault Ste Marie GIS portal	http://portal.ssmic.com/							Thematic data and GIS services for social and public safety, economic and education, culture tourism and recreation and environmental health.
Water and the Rural Poor: Interventions for improving livelihoods in sub-Saharan Africa (FAO)	http://www.fao.org/nr/water/art/2008/flash/ruralmaps/gallery1.html							A flash interface displaying maps reporting on different issues about water and sanitation in rural areas of Africa. Although Spatial data is used for reporting it is not interactive.
Demographic Health Surveys	http://macroint.mapsherpa.com/hivmapper/							Interactive mapping and display of HIV/AIDS indicators on a mapping interface
CIESIN new web portal for global-scale data	http://sedac.ciesin.columbia.edu/mapviewer/							An indicator based mapping interface that allows user to add their own web map services to the application

Like most SDIs the ones presented in the table focus mainly on delivering data and Geodata services through the web. These also try to categorize the data into meaningful themes for example the 5 themes of the Sault Ste Marie GIS portal. The Community Geomatics Centre (CGC), which developed this site, has won numerous awards for being the most comprehensive Local SDI in North America. The CGC is a good example of the positive results of peering and collaboration of organizations that own data and is discussed in chapter 5. Their thematic approach also applies to the INSPIRE Geoportal which serves environmental data to the EU countries. However, the portal only displays selected core environmental datasets. Through the portal, users have access to other datasets that they can search for in the catalogue and add to the map viewer.

The indicator based initiatives (bottom three in the table) show the power of maps in presenting evidence and trends using pre-determined indicators. However, they are static to the content that is available and not flexible enough to allow for user generated queries and analysis of the data. The site however still serves its purpose because its objective is to present indicators on available data. SDIs would become more useful in responding to the urban challenges if the concept is fused with the capability to present indicators as is done by the last three sites in the table.

The next section of this chapter draws from the findings from Chapter 2 and 3 to re-conceptualize SDI platforms and incorporate mechanisms that allow for presentation of evidence and continuous learning of urban trends. This provides opportunities to fully explore urban spaces and respond to the emergent needs of societies as they occur through provision and use of timely and relevant information based on collaborative content generation efforts.

3.4SDIs as platforms for evidence based decision making

Five key areas have been identified that help to construct a new perspective on SDIs to transform them from merely being public platforms for accessing spatial data and services to being platforms that serve information products which inform local decision making based on the evidence produced from the data. These areas need to be addressed in order for the conventional SDI to evolve into useful platforms for evidence-based decision making towards achieving sustainability of human settlements.

3.4.1 Content generation for the SDI

SDIs are evolving into even more open networks of networks. VGI has a critical role as a means of generating relevant user content than what is available in formal data stores. While this is a valid assertion, VGI can potentially reproduce the inequalities that already exist in societies as some disadvantaged groups will not participate in the creation of data. Therefore, VGI will conceal as much as it is able to reveal. However at an organizational level, peering and mass collaboration are imperative. Organizations that have data should be able to volunteer their information to the SDI in order to ensure collaboration and maximise potential use of data.

3.4.2 Community participation

Ordinary citizens and organizations can participate in different ways in the creation of relevant information and its use. Within the community of participants, there are those that have the capacity to volunteer information and those that do not have access to the internet in a way that will allow for meaningful participation. In the case of the latter, participation can be through stakeholder forums and community representatives. This will allow for decision makers to hear different viewpoints about the problems that exist in the communities that they are trying to learn about. In cases where the community has capacity to freely volunteer information, the principles of the Geoweb can be exploited to capture relevant information as VGI. This allows for SDIs to be built by both formal and informal datasets; the hybrid Top-Down/Bottom-Up approach (Fernández and Iglesias, 2009)

Indicators that are developed through a consultative and community participatory process and used as the backbone for information presented and stored in an SDI will make them more relevant in addressing the pressing societal challenges

3.4.3 Redefining Geo-Information

SDIs serve to facilitate access to and the use of spatial data. The form in which spatial data is found has been changing over the years and SDIs should also evolve to cater for these changes and consequently be more useful information gateways. Spatial data refers to any form of information that has a location reference and this includes (but is not limited to) Geo-Tagged content on the web including photos, news feeds, web pages, videos etc. All these new Geo-Information resources carry potential to provide a contextualised understanding of what goes on in specific places. SDIs should evolve into platforms that provide access to a wide range of such Geo-Information resources and help to better respond to the problems in urban settlements.

3.4.4 Indicators: Presenting the evidence

Indicators are a key feature in attempts to counter the urban problems in the world as was presented in Chapter 2. They are important tools for understanding the extent to which specific problems have affected communities and are a direct measure of the work that needs to be done to correct them. Proper use of indicators results in the generation of new knowledge, which will allow for appropriate response mechanisms to be formulated. If SDIs are to be useful Information Infrastructures in the attempt to find lasting solutions to urban problems, they need to facilitate access to the needed and essential information. The Spatial Data that they serve should display city trends, and indicators allow for this. Decision makers will be able to better understand the real problems that exist in communities and therefore channel mitigation and recovery efforts appropriately

3.4.5 New Organizing Principles: Context, Geography and Themes

The conceptual framework for SDIs being proposed in this chapter will handle more diverse data than what has been handled through the traditional SDI concept. This is true for both data formats and the quantity of data. The process of discovering relevant data and services may be more complicated than before and therefore will need to be better organised. Three organising principles have been identified which will make it easier to sieve out useful information whenever it's requested from the SDI.

3.4.5.1 Context based Metadata

Data is made in different ways and for different purposes; therefore there is a need to define the context within which the data was made so that it is not misused. The context for the data can be effectively described in the metadata records for the data and services. Most spatial metadata describes the technical characteristic of the data and miss the context-based and tacit information about semantic attributes (Elwood, 2009b, Schuurman, 2009a). Information like sampling methodologies, collection rationale, policy constraints etc, should be captured to give meaning to spatial data (Schuurman and Leszczynski, 2006).

3.4.5.2 Geography

Geography is also a good organising principle for data. This is being cemented by the evolving characteristics of Geo-Information, which enables all sorts of electronic information to be categorized and contextualised by location.

3.4.5.3 Themes

Sorting data thematically helps the user to find whatever they are looking for faster than they otherwise would. This is the approach that has been taken by the SSMIC and the INSPIRE initiatives that have been presented in this chapter.

Figure 3.2 is a depiction of the fusion of these five features to create a new conceptualisation of SDIs. The new concept facilitates a learning process and an evidence based approach toward decision-making.

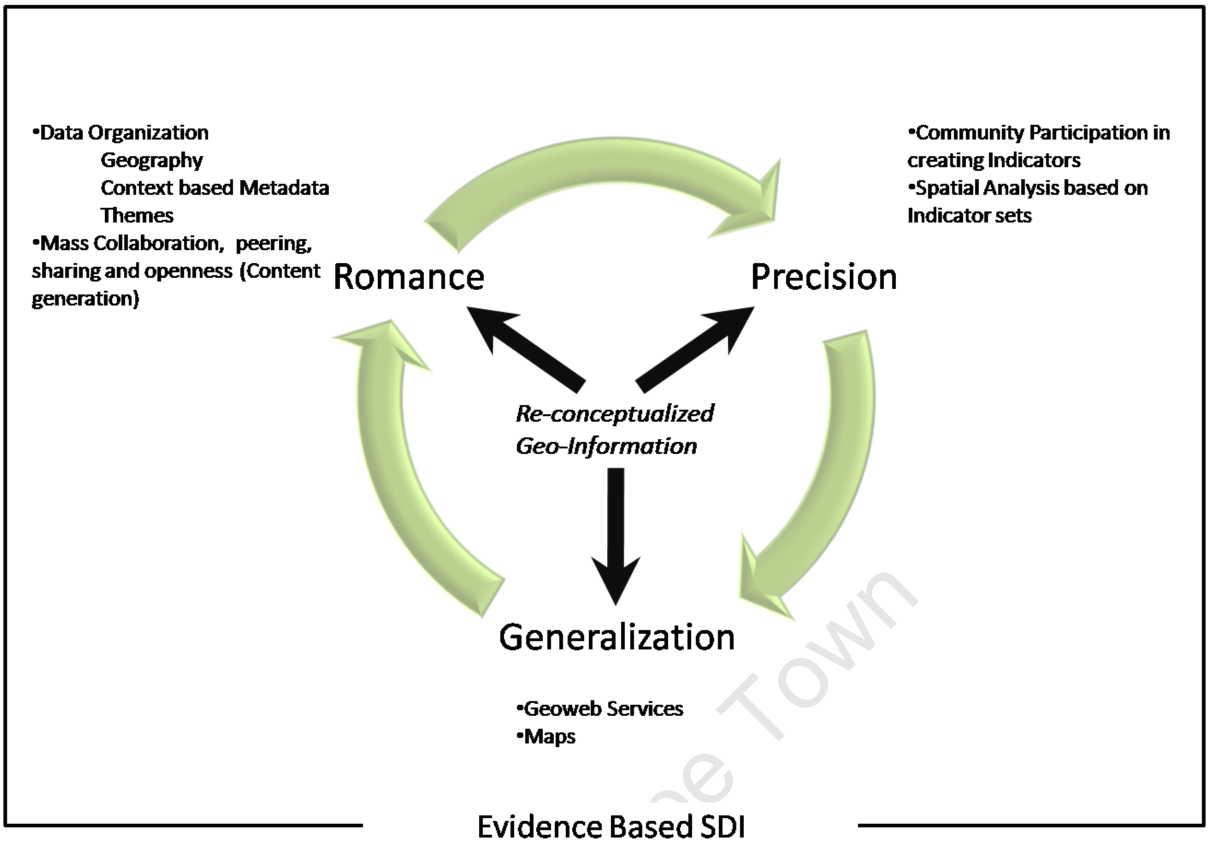


Figure 3.2: Reconceptualising SDI as a platform for continuous learning and Evidence Based Decision making

Having described the new concept of SDIs, the rest of the thesis focuses on the design processes for coming up with an architectural description of an SDI that will typically possess these essential characteristics. The next chapter describes the methods that are used to design and test the SDI architecture.

Chapter 4: The Design Approach

4.0 Introduction

The purpose of this chapter is to present a detailed description of the methodology that was used to design and prove the concept of an SDI that takes into account the factors described in Chapter 3 to address urbanisation challenges. Information Infrastructures (including SDIs) are an assembly of Information and Communication Technology (ICT) components converted to useful shared ICT services by a human ICT infrastructure of knowledge, skills, architecture and experience (Weill et al., 1999). Research on Information Infrastructures is essentially located within the Information Systems (IS) discipline. On these grounds, this chapter draws from IS literature to present principles that govern the design of relevant and useful artefacts in IS research. These principles are then applied in the SDI design process relevant to this research.

The first section focuses on presenting an Information Systems Research Framework that is constructed from two main paradigms that characterize IS research; behavioural Science and Design Science (Philosophical framework). These paradigms provide the philosophical grounding for the design process that is presented in the second section and is based on the Reference Model for Open Distributed Processing (RM-ODP) (Methodological framework). Some lessons are then borrowed from the Software Engineering discipline to illustrate how the different phases of the RM-ODP are implemented to design and evaluate the final artefact (Implementing methodology) (See Figure 4.1).

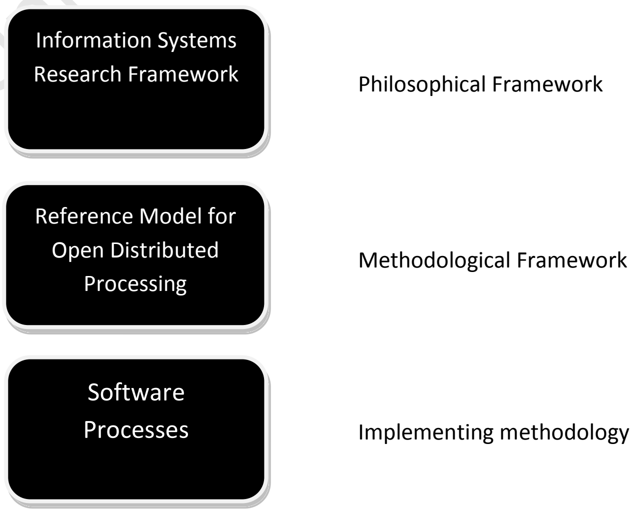


Figure 4.1: An overall description of the Design Approach

4.1 The Information Systems Research framework (ISRF)

When Studying SDIs, it is imperative to consider the underlying philosophies of knowledge, otherwise such studies will be greatly constrained to the technical and administrative organizational dimensions (Georgiadou et al., 2009). Solving a research problem in IS requires that the problem space be thoroughly defined and an appropriate IT solution be tailor made for the problem. Consequently, the main debates in IS research revolve around relevance and utility of IT artefacts (Agarwal and Lucas, 2005, Hirschheim and Klein, 2003). These debates are in line with the two main paradigms that characterize IS research; Behavioural Science and Design Science (Hevner et al., 2004, Carlsson, 2006). The behavioural science paradigm forms the mainstream IS research and should be complemented by Design Science (Carlsson, 2006).These paradigms also apply to SDIs as they fall within the broad discipline of IS. While SDIs have mainly had a technical focus and have mainly been pioneered by agents from the Geomatics and GIS background, some key lessons can be learnt from IS research (Georgiadou et al., 2009).

Behavioural Science has its roots in the natural sciences (Simon, 1996) and it seeks to develop theories and explain or predict organizational or human phenomena surrounding the analysis, design, implementation, management and use of an information System (Hevner et al., 2004). It helps to define the problem that requires being solved. This paradigm is useful to ensure that the design artefacts that are produced are relevant for their intended purpose. The theories, knowledge and other outputs of the behavioural science paradigm in turn affect or determine the design characteristics of the IT artefact. These design characteristics are implemented and tested through the inspiration of design science research philosophies.

Design science has its roots in the engineering sciences (Simon, 1996). It seeks to extend the boundaries of human/organizational capabilities by creating new artefacts. Design based research methods are normally employed by learning scientists in their inquiries because this methodological framework considers the subject of study to be a complex system involving emergent properties that arise from interactions of more variables than initially known by the researchers (Brown, 1992). The rich phenomena that emerge from the interactions of people, organizations and technology need to be qualitatively assessed to yield an understanding of the phenomena and therefore generate understanding adequate for theory generation and problem solving.

To get a better understanding of design science as an Information Systems research paradigm it is important to understand design both as a process and a product (Walls et al., 1992). The design process is a sequence of activities that produces innovative products (Hevner et al., 2004). Design as a product or artefact refers to the output of a design process. The evaluation of a design product results in refinement of the design process and eventually the product. This is done iteratively until the optimal design is created (Markus et al., 2002). March and Smith (1995) describe two design processes and four design artefacts as:

Design Processes

- Build: Artefacts are constructed to solve specific problems
- Evaluate: They are evaluated to see if they serve the purpose they were created for

Design Artefacts

- Constructs: Provide the language in which the problem is specified and communicated (Schon, 1983)
- Models: Tools that aid problem and solution understanding. They use constructs to represent the real world situation
- Methods: these are the processes that provide guidance on how to solve the problem
- Instantiations: These show that constructs, models and methods can be implemented in a real world solution

Behavioural Science and Design Science are complementary but distinct (March and Smith, 1995). An entire bias towards the design science paradigm may result in technically precise IS artefacts that do not meet the users' need and are therefore not really useful. The same danger exists for a bias towards behavioural science which can result in an over emphasis on understanding the problem, defining the context and generating theories at the expense of developing systems that work. There is a need to strike a balance between design and behavioural sciences to result in the creation of appropriate artefacts.

Figure 4.2 presents the conceptual framework for understanding, executing and evaluating Information Systems research by combing the behavioural and design science paradigms. The Environment defines the problem space where the phenomena of interest reside. Behavioural science addresses research through development and justification of theories that explain or predict phenomena from the problem space. Design science addresses the research through building appropriate artefacts designed to meet the required need. There is a continuous assessment of the design and refinement of the system requirements until the optimum design is established. These iterations are due to the fact that the design requirements of an Information System are normally not adequately known at the inception stage of a project. Real life projects are characterized by continuously changing business and user requirements and an ever evolving view of the required Information System architecture (Larman, 2003).

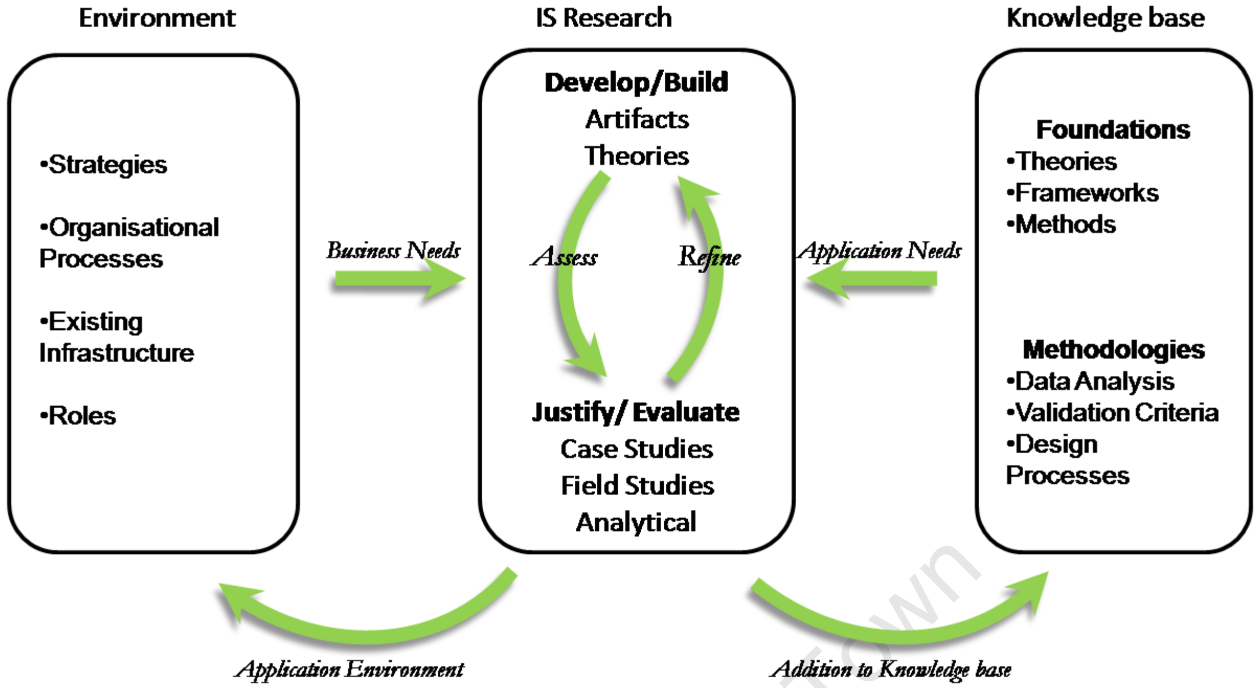


Figure 4.2: The IS research framework (Adapted from (Hevner et al., 2004))

The knowledge base is composed of Foundations and Methodologies. These provide the necessary tools through which the IS research is accomplished. Knowledge from previous research efforts is used to provide a foundation for theories, frameworks, instruments, constructs, models, methods and instantiations in the develop/build phase. Methodologies are the guidelines used in the Justify/evaluate phase. Rigor is achieved by appropriately applying the foundations and methodologies. The output of IS research are new foundations and methodologies as well as new artefacts to influence the real world.

4.2 The Design Process

The ISRF presents a philosophical underpinning for developing and evaluating Information Systems artefacts (Carlsson, 2006, Dobson, 2001). It does not specify the detailed components of a design process. In fact, IS design processes are fuelled by the philosophies behind the framework. The **design process** defines methods and procedures for creating systems/software designs. So, design science does not define a methodology but a conceptual framework upon which different methodologies can be constructed. Environmental factors (e.g. type of project, available technology and the perceived IT solution) result in different methodologies being applied for different projects.

However, design science defines an appropriate philosophical foundation and conceptual framework for these different methodologies.

4.2.1 Reference Model for Open Distributed Processing

Open distributed processing (ODP) is a term that describes systems that support heterogeneous distributed processing both within and between organisations through the use of a common interaction model (Raymond and Armstrong, 1995). SDIs function on the principles of ODP (Fernández and Fernández 2008). The Open Geospatial Consortium (OGC) is constantly developing SDI standards and technical specifications that fuel access to and use of distributed geospatial resources through the internet. OGC is a non-profit, international, voluntary consensus standards organization that is leading the development of standards for geospatial and location based services (OGC, 2009). OGC standards are at the core of SDIs today and they form the common interaction models that facilitate discovery and use of Geospatial Information on the web. These standards are constantly being created and modified to suit the ever-changing needs for Geo-Computing on the web.

The Reference Model for Open Distributed Processing (RM-ODP) defines the methodological framework that is applied in the design process of this research. RM-ODP helps to provide a coordinating framework for the standardization of ODP by creating an architecture which supports distribution, interworking, interoperability and portability (Raymond and Armstrong, 1995). It provides a "big picture" that helps organize the components of a system in an appropriate manner to achieve the intended result. The RM-ODP provides techniques for abstracting from reality in order to make sense of a complex system and design an appropriate solution to a problem.

The RM-ODP prescribes a framework using five different but interrelated viewpoints explained below and shown in Figure 4.3 from which to abstract or view ODP systems (Farooqui et al., 1995).

Enterprise viewpoint: This viewpoint describes the purpose, scope and policies that are associated with the IS. It describes the relationship between a system and its environment, its role and associated policies.

Information viewpoint: This describes the semantics of information and information processing incorporated into the IS.

Computational viewpoint: A functional decomposition of the system into a set of services that interact through interfaces.

Engineering viewpoint: Contains the mechanisms and functions required to support distributed interaction between services and Data within the system. Its chief concerns are communication, computing systems and software processes.

Technology viewpoint: Contains the specific technologies chosen for implementation

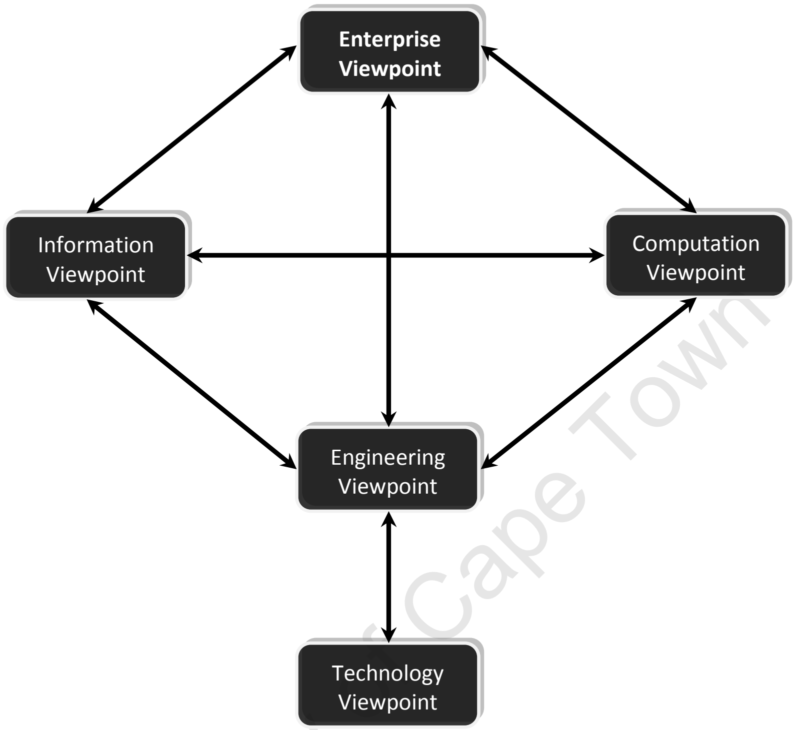


Figure 4.3: The Reference Model for Open Distributed Processing (Adapted from Farooqui et al., 1995)

The RM-ODP has been used before to model SDIs (Fernández and Fernández 2008, Cooper et al., 2007, Hjelmager et al., 2008), from identifying and defining their purpose (Enterprise Viewpoint), determining functionality (Information and component viewpoints), designing an appropriate architecture (Engineering viewpoint) to implementing the design using appropriate technologies (technology viewpoint). The International Cartographic Association (ICA) is working on defining formal models and technical characteristics of SDIs using this reference framework (Hjelmager et al., 2008, Cooper et al., 2007). Work has been done to define the Enterprise, Information and Computation viewpoints but the last two viewpoints are yet to be defined. It appears that the ICA is coming up with standardised components and procedures for implementing SDIs, an approach that may not be appropriate considering the volatile nature of SDIs themselves and the diverse environments in which they may be implemented. However, the framework still works as a guiding principle for developing distributed Information infrastructures.

4.3 Software Engineering Processes

While the RM-ODP provides a useful and guiding methodological framework for implementing SDIs, Software Engineering (SE) principles provide the tools that are necessary to develop working IS artefacts (Nunamaker and Chen, 1990). Software Engineering is the establishment and use of sound engineering principles in order to obtain software systems that are reliable and relevant to the needs of the user. Over the years, many models have been developed to make systems development more efficient. Traditional/Prescriptive models prescribed systematic ways for creating software systems (Wang and King, 2000). Recent developments in the Software Engineering discipline however have a bias towards user oriented approaches (Aroyo and Dicheva, 2000). Therefore, there has been a paradigm shift, where Software Engineers are now valuing efforts to understand the user needs over rushing to complete the product development through a set of rigid guidelines. This has seen the rise of new "agile" systems development approaches, where the main focus is adapting to the client's needs to create more relevant products without adopting a prescriptive approach. This however, does not totally nullify the importance of the previously used systematic approach to software development.

4.3.1 Prescriptive Models

Prescriptive models were introduced into Software and Systems Engineering practices to bring order to the way that software systems are designed and built. According to (Pressman, 2009), the general framework for all prescriptive models is summed up in the following steps:

a. Communication (Project initiation and requirements gathering)
b. Planning (Estimating costs, project planning and task scheduling)
c. Modelling (Systems Analysis and design)
d. Construction (Code generation, component integration and systems testing)
e. Deployment (Delivery support and feedback)

Most prescriptive models are built around this framework. Two examples of prescriptive models are given below.

4.3.1.1 Waterfall Model

There are times when the requirements of a problem are clearly defined and the work flows from communication to deployment in a relatively linear fashion. The waterfall model suggests a systematic sequential approach to software development from the requirement gathering through to deployment.

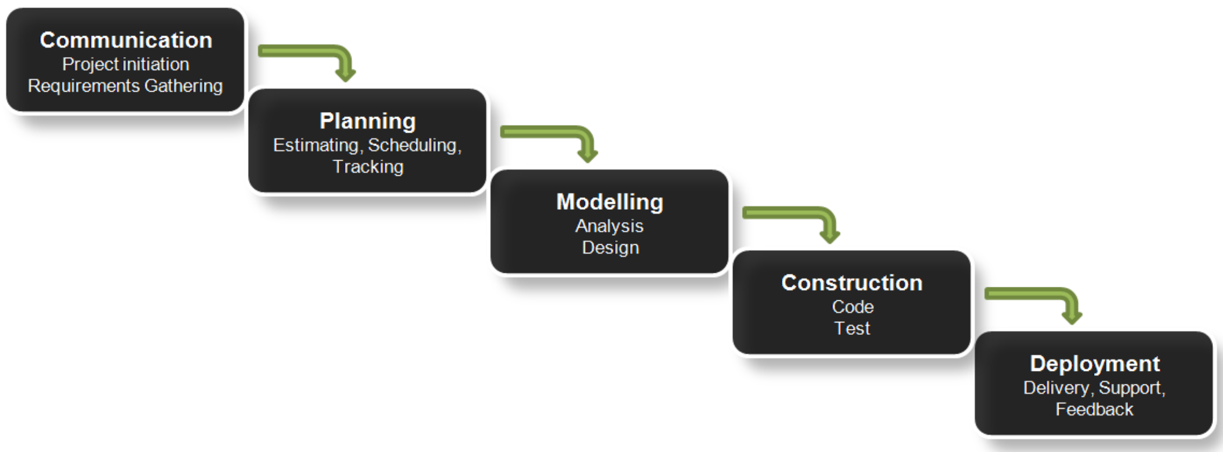


Figure 4.4: The waterfall model (adapted from (Agarwal et al., 2009))

The main problem with the Waterfall model is that real projects rarely follow the sequential flow of the waterfall model and it is difficult to handle changes in the requirement once the project is at an advanced stage (Agarwal et al., 2009). It is also often difficult for the client to explicitly state all requirements at the beginning of the project.

4.3.1.2 Incremental Process Models

These are similar to the Waterfall model in that most of the requirements should be well defined at project initiation. However, the system is built incrementally with bits and pieces of the working system being delivered at different phases within the project.

Versioning is a key characteristic of Incremental Models where the later version is an upgrade of the previous version. As the system requirements become clear, a better working version of the system is given to the client for assessment (Forouzan and Mosharraf, 2007). The main disadvantage with this approach is that it is tempting to regard a reasonably working increment as the final product whereas further rigor and research could result in new perceptions for the system and therefore new functionality.

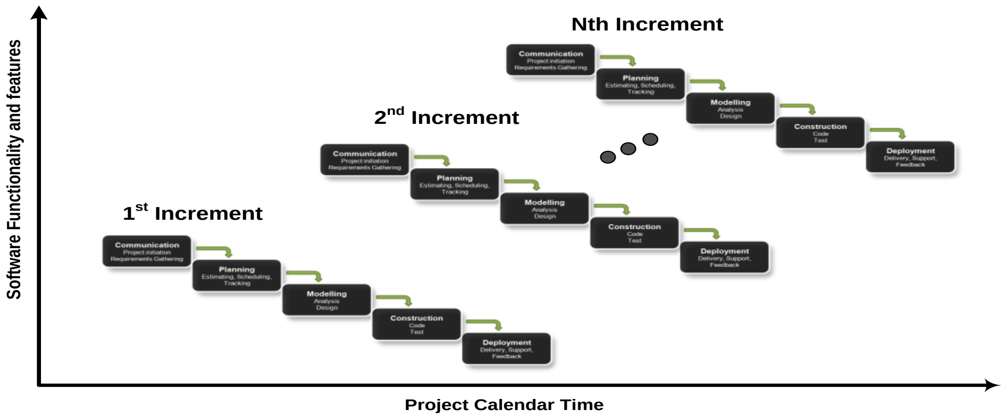


Figure 4.5: The Incremental process model

Other Prescriptive models include Rapid Application Development (RAD) (Martin, 1991), Prototyping (Forouzan and Mosharraf, 2007) and the Spiral Model (Tsui and Karam, 2006)

4.3.2 Agile Models

Agile Software engineering combines a philosophy and a set of development guidelines. The philosophy encourages user satisfaction, early incremental delivery and minimal software engineering products (Cockburn, 2002). The development guidelines stress delivery of a working system over Systems Analysis and Design (although those activities are not discouraged). Unlike prescriptive models, that follow the "plan your work and then work your plan" approach, Agile Software Engineering is driven by the "plan as you go" approach. To achieve this, there is active participation of stakeholders from project inception right through to delivery (Cockburn, 2002). This ensures that Systems Developers, who do not necessarily know the business case to the required levels, work on delivering a more relevant product. The Agility characteristics of this approach help the software development team to adapt to changes in the requirements at any stage of the project.

Some of the assumptions that characterise the Agile approach (Turk et al., 2004) are listed below:

1. It is difficult to adequately address all Business and User requirements at inception. It is equally difficult to predict changes in the requirements over a period of time.
2. Analysis, design and construction are not as predictable (from a planning point of view) as we would want them to be.

3. Design and construction are interleaved and one cannot really determine the amount of design work required unless there is a level of construction and testing.

Examples of Agile development approaches include, Extreme Programming (XP), Feature Driven Development (FDD) and SCRUM (Boehm and Turne, 2003)

4.3.3 The Unified Process

The Unified Process (UP) is an attempt to draw the best features and characteristics of conventional software process models but characterise them in a way that implements many of the best principles of Agile development (Pressman, 2009). In order to develop relevant and user oriented IS artefacts, the User Needs require to be modelled as accurately and thoroughly as possible. UP recognises the importance of stakeholder communication and streamlined methods of defining the customer's view of the system (Use Cases) (Kroll and Kruchten, 2003). It recognises the importance of architecture and helps the architect to focus on the right goals such as ease of use, reliance to future changes and component reuse (Ambler and Constantine, 2000). It provides an approach that is incremental and iterative and therefore has an evolutionary feel.

4.3.3.1 Phases in UP

There are four phases in the USDP and these are described below (Kroll and Kruchten, 2003):

Inception

This phase encompasses planning activities. Business and User requirements are assessed at this stage. The system is modelled at a high level of abstraction to show the different components of the system and how they will be knitted together. At this point, the required resources are identified and a risk assessment is done. Another product of the inception phase is a high-level use case model, which will most likely be 10-20 % of the final Use Case model. This will show the different actors in the system and how they will interact with it.

Elaboration

Elaboration involves communicating with the client and modelling different components of the system. At this stage the Use Cases from the preceding phase are elaborated and the architectural representation is refined to four different viewpoints of the system; Analysis Model, Design Model, Implementation Model, Deployment Model (Pressman, 2009).

In addition, the project plan is reviewed at this stage to ensure that deadlines and risks are reasonable.

Construction

The Architectural Model is used as reference for developing or acquiring the necessary software components to build the system. The components are built such that each function as depicted in the Use Case is achieved.

Transition

This encompasses the final phases of construction and testing of the different components of the system. Testing is done in units (i.e. the individual modules) and also as a whole.

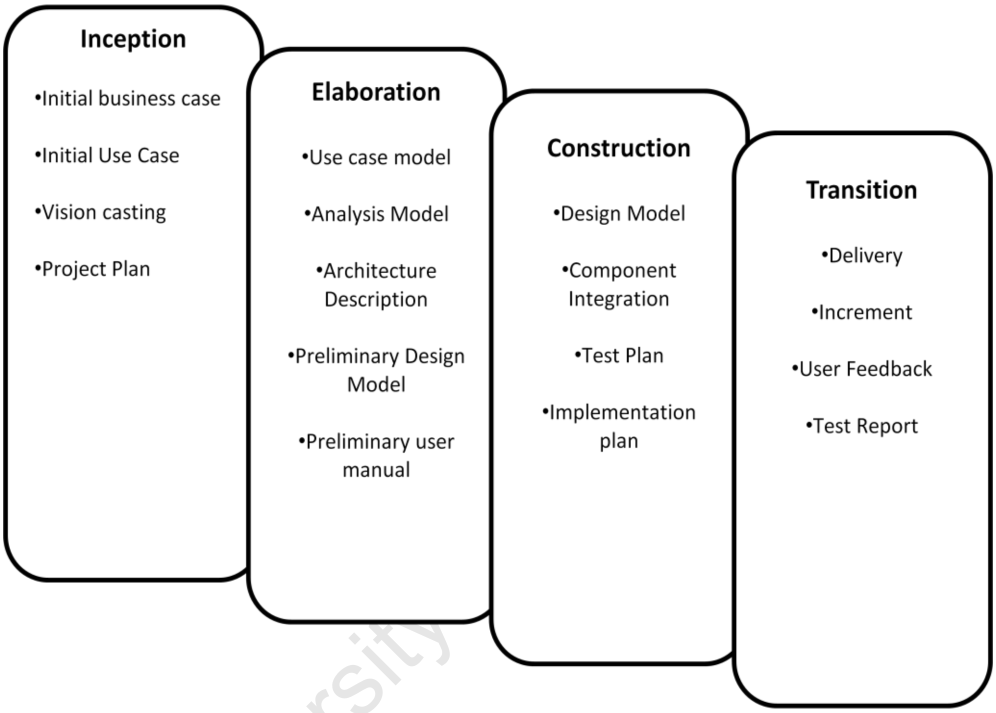


Figure 4.6: The Unified Process (adapted from Pressman 2006)

The UP strikes a balance between structure and being adaptive to change. Communication with stakeholders is crucial at all points of the project. This is to ensure that there is maximum user satisfaction upon delivering the final solution. This method is adopted for this research because at inception, the specifications of the system are not clearly defined and are only known through a thorough requirements engineering exercise.

4.4 The Hybrid Model

The overall methodology for the design process of the SDI is show in the hybrid model in Figure 4.7. This approach is grounded on the philosophies of the Design Science and Behavioural Science paradigms.

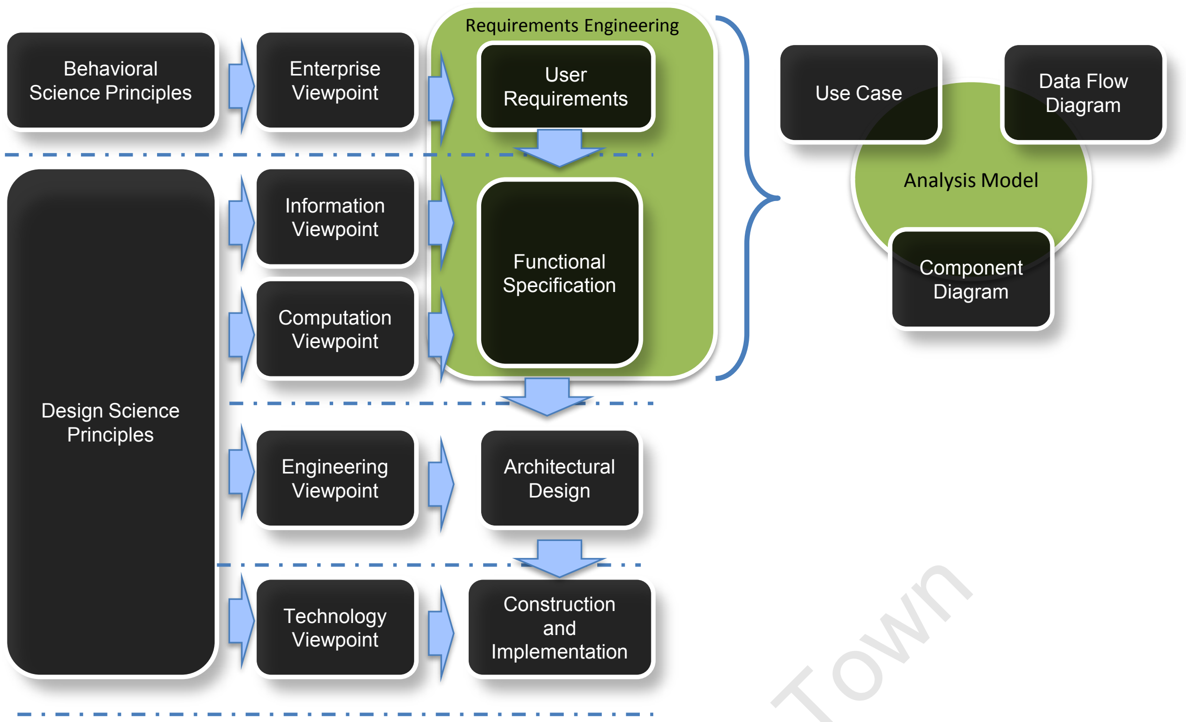


Figure 4.7: The Hybrid model showing the overall methodology with concepts borrowed from ISRF, RM-ODP and Software Engineering

The RM-ODP is then built upon the principles of this ISRF with the Enterprise viewpoint being aligned with the principles of Behavioural Science and the other four viewpoints with the Design Science paradigm. The main components and procedures that make up the UP are used to implement the methodology through well-defined and established SE processes. As part of implementing the methodology, a Requirements Engineering exercise (explained in the next section) is carried out to determine the user and functional requirements for the SDI. The exercise results in three models which make up the Analysis model and these are in the form of Use Case, Data Flow and Component Diagrams. The Analysis model forms the basis for the Architectural Design, which is developed and tested using relevant technologies.

While Hjelmager et al. (2008) proposed the use of class diagrams to model the Information viewpoint, this research uses the Data Flow diagram to model the SDI from this viewpoint. The reason for this is that data is seen as a critical element within the SDI and all the information processes and semantics should be modelled around how information will be stored and how it should flow in and out of the system, and not the relationships between the classes and objects that make up the system.

As shown in Figure 4.7, the Enterprise Viewpoint is modelled through a User Requirements Analysis, which results in a Use Case diagram. Both the Information and Computation Viewpoints are modelled through a Functional specification exercise resulting in a Data Flow diagram and a Component Diagram respectively.

4.5 Requirements Engineering

The hardest part of building a software system is deciding what to build. No part of the work so cripples the resulting system if done wrong and no other part is more difficult to rectify later (Brooks, 1987). The usefulness of an artefact is a measure of its successful implementation. Therefore understanding the requirements of the system to be built is critical to ensure relevance and utility and to avoid unnecessary wastage of time and resources. Requirements Engineering (RE) helps to better understand what is to be accomplished through the resultant system. It provides the appropriate mechanisms for understanding the problem, analysing need, and assessing feasibility, negotiating a reasonable solution, specifying the solution unambiguously, validating specifications and managing the requirements as they were transformed to an operational system (Futrell et al., 2002). Pressman (2009) describes four RE processes shown in Table 4.1. These are executed in an iterative manner as the requirements evolve during the life span of the project.

Table 4.1: Steps in Requirements Engineering (adapted from(Pressman, 2009))

RE Step	Purpose and description
Inception and Elicitation	Enables Analyst to get an overview of the project and the context of the problem to be solved. This is done through Interviewing stakeholders/users to assess their need
Elaboration	An Analysis Modelling task that is driven by creating and refining User scenarios that describe how the different actors in the system interact with the system. The Analyst is able to appreciate the normal, expected and exciting requirements
Negotiation	Negotiation is the process of coming up with reasonable requirements for the project. The stated requirements are ranked in terms of importance and they are modified, combined or eliminated to get a reasonable set of requirements.
Specification	Used to specify the requirements of the system. It can take different forms; in some instances a set of models, in other instances, a written document depicting different user scenarios and in other a combination of both. The Specification is the final work product for the Requirements Engineer and serves as a foundation for other Systems Engineering work.

In this study, there are two main sources of information for the RE and these are in the form of Case Studies and the Literature review. Both form the basis for the design artefacts that make up the Analysis Model as shown in Figure 4.8.

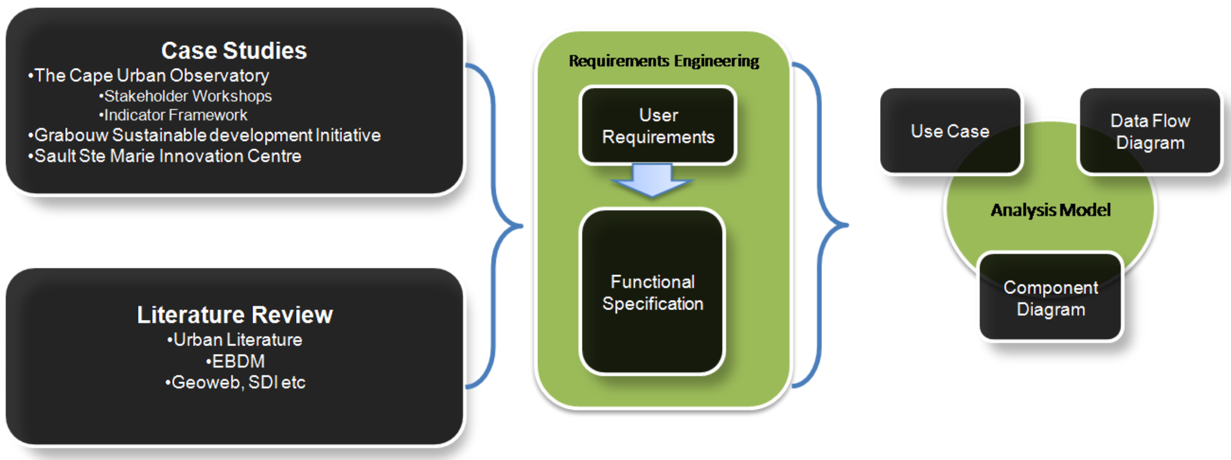


Figure 4.8: Sources of information for and output from the Requirements Engineering exercise

4.5.1 User Requirements (Use Case Modelling)

4.5.1.1 Case of the Cape Urban Observatory

During 2008/2009 the Cape Urban Observatory (CUO) held two Stakeholder workshops. The Inception Workshop was held in November 2008 and served to introduce the CUO initiative to a broad range of Stakeholders. A discussion on the main human settlements problems that require intervention and the appropriate response strategies was also started at this workshop. The Second Workshop was held in August 2009 and was aimed at notifying stakeholders of the progress that the CUO had made on different fronts and mapping a way forward for the rest of the year.

These two workshops are a source for knowledge on how to respond to the challenges facing the different municipalities in the Western Cape Province of South Africa. The List of Indicators that were developed by the CUO (discussed in Chapter 2) is used as benchmark indicators to determine the information and information processing requirements for the system.

4.5.1.2 Case of the Grabouw Sustainable Development Initiative (GrSDI)

The GrSDI is an Initiative being pioneered by the Development Bank of South Africa (DBSA) and the Theewaterskloof (TWK) Municipality in Grabouw, a small town located about 120km East of Cape Town in the Western Cape Province. The Initiative is one of six pilot projects that are being done in different provinces in South Africa in an attempt to develop models for sustainable development that can be adopted in other towns in South Africa (Hamann et al., 2008). There are three main groups of participants in this initiative:

a. The DBSA: The funders and Project managers for the initiative

b. TWK Municipality: The Local Authority responsible for implementing the different programs of the Initiative

c. Stakeholder Forum: Composed of Community representatives, the business community, municipal officials and officials from the DBSA. The forum is meant to be a discussion forum representative of the Grabouw community to discuss how to implement the program (Hamann et al., 2008).

Thirteen interviews were done with different representatives from the three groups of participants listed above. The interview questions were open ended and focused more on understanding the initiative, its implementation plan and progress instead of specific SDI questions (see Appendix 3). The reason for this was to allow for an open approach that would lead to a better understanding of the actual problems that exist and the current approaches to mitigating them. This is a better way of understanding the real problems that exist and also an objective way of thinking of an appropriate solution to the problem.

4.5.1.3 Case of the Sault Ste Marie Innovation Centre

The Community Geomatics Centre (CGC) at the Sault Ste Marie Innovation Centre (SSMIC) in Ontario Canada has been chosen as a third case study. They use GIS as a pivotal component in most of their projects. The CGC offers services to government and other organisations in the following areas (SSMIC, 2009) :

a. Spatial Data Capture and Manipulation

b. Information Systems / Information Technology / GIS Strategic Planning, Consulting, Development and Integration

c. Health and Social Services

The CGC's approach to creation and use of data as well as their data sharing mechanisms can potentially be a role model for similar initiatives around the world on how to use public domain information for public good. The number of awards (listed below) that they have won is a good testimony in support of this assertion:

- URISA Ontario Best Public Sector GIS Award (2009)
- ESRI Health GIS Conference – Communication Award (2008)
- URISA Leadership in the Field of GIS (2006)
- URISA Best Municipal GIS Award (2006)
- ESRI Canada Award of Excellence (2006)
- Award for Commitment to the Community – SSM Police (2005)

- URISA – Silver Award – GIS Leadership in Ontario (2005)
- URISA Best Municipal GIS Award (2003)
- Designated Most Comprehensive Municipal GIS Dataset in Canada by ESRI Canada (2003)
- ESRI Canada Business Partner Award (2002)

An Interview was done with one of the managers at the CGC. This was to complement the knowledge that had already been gathered about their initiatives through their website. This interview allowed for a better appreciation of the power of mapping and different forms of visualization in presenting evidence as well as the extent to which mass collaboration in data creation and sharing can amplify the possibility of more informed decision making.

4.5.1.4 Literature Review

The literature review that was presented in Chapters 2 and 3 of this thesis complemented the Case studies where a new perspective on SDIs was presented. Chapter 2 provided insight into the problems that are currently being faced by African cities and the mainstream theoretical frameworks behind current intervention efforts, their strengths, weaknesses and opportunities. The new model for SDI that was constructed in Chapter 3 defined the theoretical framework that shaped the requirements of the system and inspired the new SDI design that would more efficiently help in responding to current urban challenges.

This two-pronged approach fed into the Requirements Engineering process (Figure4.8). The four-part cycle of Inception, Elaboration, Negotiation and Specification was applied to continuously model the requirements in an evolutionary fashion throughout the life span of this research.

4.5.2 Functional Specification

A functional specification is the documentation that describes the requested behaviour of a system. The documentation typically describes what is needed by the system user as well as requested properties of inputs and outputs. Two models were created to describe the two viewpoints that make up this part of the design cycle: a Data Flow Diagram for the Information Viewpoint and a Component Diagram to describe the Computation viewpoint.

4.5.2.1 Information Viewpoint: Data Flow Diagrams

The Information viewpoint focuses on the Information content for the system. The Modelling activity involves identifying data elements of the system, manipulations that should be performed on the data and the information flows within the system.

The concept of a Data Flow Diagram (DFD) is not mentioned in the ISO RM-ODP; however, it provides some useful data modelling concepts that are used in this research. DFDs allow the Systems

Analyst to model the Information as well as the functional domain of the system. A DFD takes an Input/Process/Output view of the system. The initial phase of modelling the data flows involves creating a highly abstract view of the processes that model the flow of information through the system and the main processes that are done on the data before output of a desired result from the system. This is called level zero modelling (Liu and Tang, 1991). The level zero DFD is further broken down to illustrate more detailed processes. As the DFDs become more detailed it is possible to perform an implicit functional decomposition of the system which is critical for the design engineering process (Wynn, 2004).

A DFD has three components: a rectangle denotes an Input or Output, Transformations on the data or processes are denoted using a circle or bubble and labelled arrows represent data objects and their direction of flow (See figure 4 below):

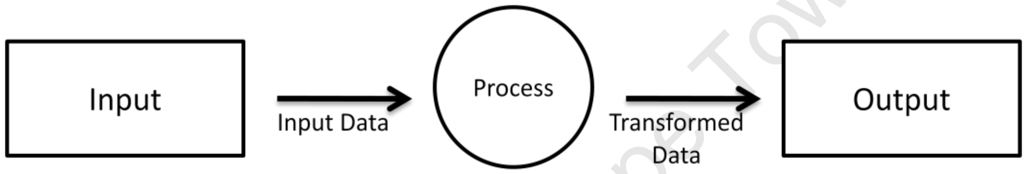


Figure 4.9: Elements of a Data Flow Diagram

The Use Cases described in the previous section are used as input for the DFD modelling process. A grammatical parse is done on the Use Case to identify processes and data elements for the DFD. Nouns point to the data elements and verbs point to processes.

4.5.2.2 Computation Viewpoint: Component Diagram

The computational viewpoint deals with the logical partitioning of the distributed system independent of any specific distributed environment on which they run (Farooqui et al., 1995). In this study, an understanding of the service components that make up SDIs was derived from the ICA work on modelling SDIs that has been presented earlier in this chapter (Cooper et al., 2007). The UML Component Diagram is used to model the components of the SDI and how they interface with each other. A component is a modular, deployable and replaceable part of the system that encapsulates implementation and exposes a set of interfaces. More specifically for Information infrastructures, components are responsible for functions that support the processing required in a problem domain.

A component diagram is made up of the following features:

- Component (rectangle with small symbol in upper right corner);

- Provided Interface (connector with circlet);
- Required Interface (connector with arc); and
- Dependence (dashed arrow).

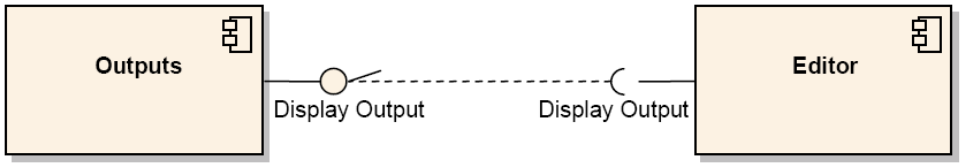


Figure 4.10: Elements of a component Diagram (OMG, 2009)

4.6 Architectural Design

Design is important because it allows the Systems Analyst as well as the stakeholders of the system to access the design before it is implemented. The Design phase requires brainstorming on alternative architectures for the system. The goal of design is to produce models that exhibit firmness, commodity and delight (Kapor, 1991). The design process first involves the acquisition of a number of alternatives which form the raw material for the design in the form of components, component solutions, and knowledge contained in catalogues, textbooks and the mind (Belady, 1981). From these alternatives, the Analyst picks up relevant elements that meet the requirements from the Analysis modelling exercise.

The purpose of the Design Engineering exercise is to come up with a technical description of the Architectural design. The Architectural design represents the proposed SDI solution as is constructed from the results of Analysis modelling. The System Architecture is a comprehensive framework that describes the structure of the system, its components and how they are designed to fit together (Bass et al., 2003). The main products of the design engineering process are a series of diagrams that depict the architecture of the system from different viewpoints.

4.7 Construction

This phase requires the Analyst to identify suitable technologies that allow the design to be implemented successfully. The final technologies are chosen based on the following factors:

a) Features had to exist within the software packages of choice that allowed for the implementation of the design features;
b) The software had to be easy to use and require minimal configuration therefore resulting in a short turnaround time for implementation and testing;

c) The software needed to be available. As a result some components are implemented using Free and Open Source software and some using technologies that had already been bought by the University of Cape Town (UCT) where the research was conducted.

The Architecture design is evaluated for fitness of purpose. Black box tests are done against the requirements specification to ensure that the system delivers the intended services. It should be noted that this research aims is to demonstrate how SDIs should evolve to be useful platforms in responding to the urban challenges in developing nations. Therefore, the main aim of the evaluation is to demonstrate a working model and not necessarily efficiency of internal software processes at a micro level. White box tests that seek to verify the technical precision of the designed artefact were therefore not performed.

Chapter 5: Analysis Modelling

5.0 Introduction

This chapter presents the first set of models that were created as part of the design process for the SDI. These models depict the requirements from the first three viewpoints of the RM-ODP that was explained in the previous chapter. A Use Case diagram was created from the Requirements Engineering (RE) process and this modelled the Enterprise viewpoint of the system. The User case shows the different actors involved in the SDI and how they interact with the system.

The second model is a Data Flow Diagram (DFD) which models the information viewpoint of the SDI. The DFD is created through performing a grammatical parse on the Use Case diagram to identify Information object and processes. The DFD allows for the functional and information domains of the system to be modelled from an Input – Process – Output perspective.

The Component diagram is the last model that is presented in this chapter. It shows the functional decomposition of the SDI and how the individual components that make it up should interact to ensure that the SDI serves its purpose. This model also shows how the system should be modularized.

These models together make up the Analysis model, which became the basis for the Architecture design presented in Chapter 6.

5.1 Requirements Engineering

The Process of gathering and defining the requirements of the system is the most important stage in the development lifecycle. The initial step involves defining the actors in the system and how they are anticipated to interact with the system through a Use case model (Pressman, 2009). This phase of the design determines how other components of the system are structured to meet the requirements portrayed in the Use Case. The three case studies that are mentioned in Chapter 4 are used as the main source of information that was fed into the use case.

5.1.1 The case of the Cape Urban Observatory

Information on the requirements of the SDI is created from two workshops that were organised by the CUO. The first is an Inception workshop which aimed at creating a shared vision for the CUO among the different stakeholders. It was attended by more than 40 participants, including representatives from the City of Cape Town and surrounding municipalities, various departments of the Provincial Government of the Western Cape, the Gauteng Provincial Government, the CSIR, the

Development Bank of Southern Africa (DBSA) and the South African Cities Network (SACN) (Smit, 2008). These organisations each have a role to play in alleviating the urban problems in the Cape Province in different but sometimes complementary ways.

The workshop resulted in a discussion that led to identifying the key problems that the CUO may be able to alleviate, and how they can alleviate these. These have been summarised below:

5.1.1.1 Key Problems (Smit, 2008)

- Data in the different organisations is not known
- Data in the different organisations is not easily available
- There are a lot of competing Geospatial dataset platforms
- Data is not publicly available as it is normally created for clients that want to retain confidentiality
- Data is easily misinterpreted and misused because it is not well documented
- There is uneven GIS capacity within the different municipalities (e.g. while there are large volumes of data at the City of Cape Town, other adjacent municipalities barely have any data).

As is evident from the inception workshop, most of the problems that the workshop participants are facing revolve around data and the CUO is seen as an opportunity to address these problems. A number of functions and roles are defined for the CUO in order for it to be useful in meeting the needs of the stakeholders that attended, these are listed below:

5.1.1.2 Key functions for the CUO

- Storage of Data
- Dissemination of Data to users
- Facilitating collaboration and data sharing amongst organisations
- Facilitating continuous learning on a broad range of urban themes
- Analysis of urban trends
- Facilitate creation of timely and relevant data
- A platform for evidence based decision making

In an attempt to help solve the above problems and achieve the goals stated above the following propositions were made for the CUO:

1. There was need to create a user friendly web based portal that facilitated access to distributed databases of many organisations that host useful data

2. Interoperability emerged as a major issue to achieve the above functions. Data should be accessed in various formats and used on various software platforms
3. Tools should be created to easily use GIS data for those that do not have the financial capacity to purchase expensive GIS software
4. The CUO should facilitate the creation of metadata
5. The CUO should play a leading role in the analysis of urban trends on a wide range of themes in the region
6. The CUO should facilitate the creation of new datasets for indicators that will be used to report on various themes in the region
7. The CUO should create an important link between monitoring, decision-making and analysis (The CUO is well placed to influence decision making because the municipal authorities (who are responsible for local decisions) are some of the stakeholders in the CUO).

A second workshop was held in August 2009 and served to get stakeholders together and showcase the work that had been done by the CUO since inception. Three stakeholder groups were identified as having a critical role in the overall function of the CUO. These are:

- Data Custodians
- Data Users
- Administrator (CUO)

At a high level of abstraction Figure 5.1 below is a Use Case diagram that describes the roles of the three stakeholder groups from the CUO's core mandate defined at the inception and follow up workshops.

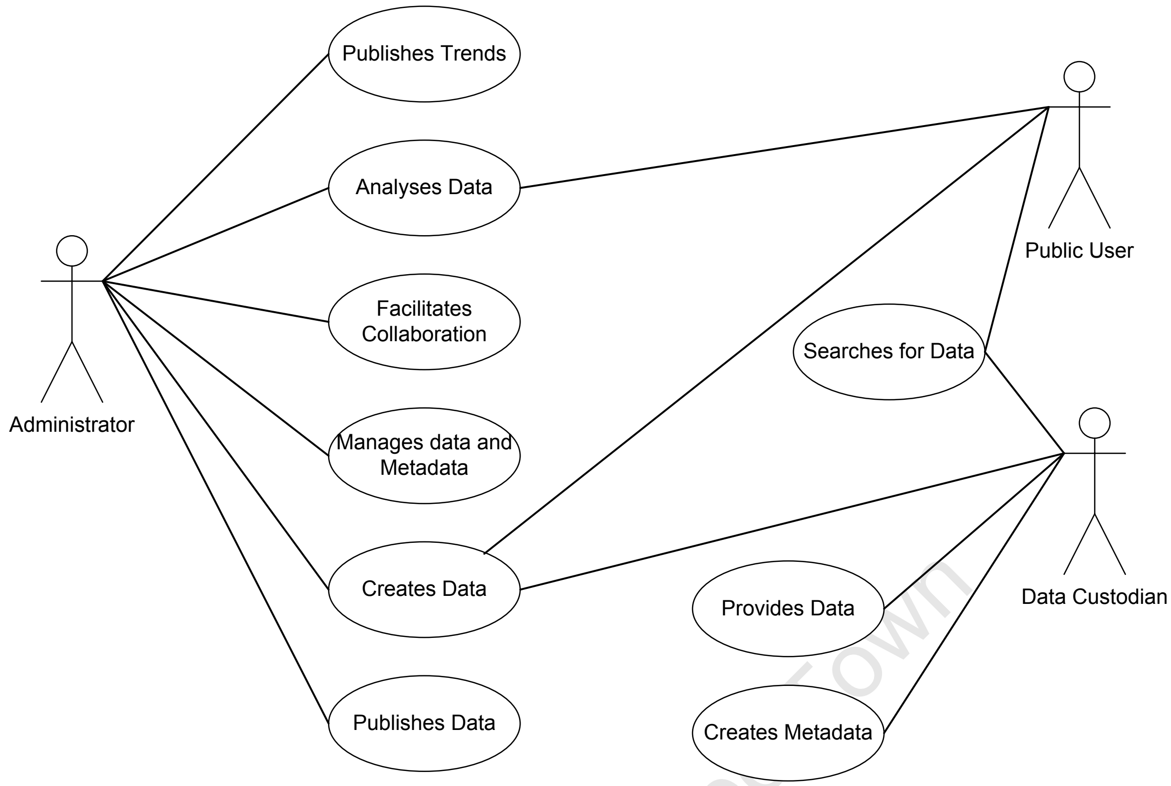


Figure 5.1: Initial Use Case Diagram derived from CUO workshops

5.1.2 Lessons from Sault Ste Marie

An interview was held with the Sault Ste Marie Innovation Centre (SSMIC) Community Geomatics Centre (CGC). It was an open ended interview that aimed at understanding the Centre's successful approach to Community participation and collaboration in data creation and use. The interview helped to understand the business model for the CGC, which is focused on centralised services (data management, GIS Analysis) allows for:

Cutting on operational Costs

The CGC is the GIS and related services consultant for the organizations that it serves, therefore these organisations do not necessarily have to invest heavily in annual software licences and building human capacity at the scale that they would if they did all the work. They pay for the services offered by the CGC as and when needed.

Better management of information resources

The organisations served by the CGC remain custodians and owners of the data although it is stored at the SSMIC where it is managed, maintained and used. The daily workflows of the participating organisations are closely tied in to the operational structures of the CGC such that updates in data at each organisation are reflected in the CGC database at a monthly interval.

Use of Public Data for Public Good

Besides offering services to the data custodians themselves, the CGC uses the data for general public good. Data from one organisation can be used in another organisation's project. This is done in a framework that is built on trust which plays a key role in data sharing initiatives at a local level (Harvey, 2003). Although the details of how this is done by the CGC cannot be released in this thesis to protect their privacy, the CGC has structures that allow for maximum use of data resources but still protecting privacy and strategic information for the organisation. This synergistic approach has displayed the power of collaboration and sharing of information resources to confront the different problems that are faced by the SSMIC.

These key points as well as other important ones are highlighted below in the extracts from the interview.

1. How are policy makers embracing the evidence that you are producing?

"We have managed to get the attention of policy makers because we are a neutral and independent body. We demonstrate the value of shared data through our data products and challenge them to collaborate. A lot of the times organisations and leaders, conceptually or theoretically get it, but it is in seeing evidence and actually connecting with examples that really drive the message home that real value is found. The key is to getting decision makers in front of these examples and have them connect to a path of exploring how to make change in their societies"

This extract shows the importance of presenting results to the important people in an appropriate way. This principle should be inherently visible in SDIs in order to deliver useful information products to decision makers and also facilitate the identification of critical issues in a way that can define new policy trajectories.

2. Do you get to keep the data or your clients/participants keep their data?

"We keep the data in a secure data warehouse. Our model is a three pillar approach that is built entirely on trust. The three pillars are:

   a. *Data Management: We have demonstrated safety standards for data and organisations trust us not to misuse their data. Even within the CGC, only the people who are working directly on the data have got access to it.*
   b. *Project management: Because of our standard for managing the data, we are trusted for projects as well.*
   c. *Values: We are interested in maintaining trust and our values reflect that."*

3. Do you build capacity within organisations or they constantly need to consult you for your services as and when they have need?

"We are the Municipal GIS team so they consult us..."

Collins highlighted that sustainability is important but the synergies that they have created with the participating organisations allow them to achieve much and make more visible impact in the community.

It makes sense for the organisations to be empowered to respond to the challenges that they face especially in instances where no one organisation has the capacity to efficiently respond to the challenges. Although centralising processes introduces an independent voice that will most likely be impartial and more objective than any interested party, it is important to still build capacity within municipalities and relevant organisations so that there is no creation of a dependency syndrome.

Below are some other important extracts from the interview:

"There is value in visualising data. Data exists to inform us but a lot of the times it is not effective in sending the message across because it is in tabular form. I do not understand what an excel spreadsheet is trying to say, I don't see the trends and patterns. You can tell me with words that you have done all your analysis and modelling and give me a 5 paged report that summarises it but for some reason, when I see it I believe it, because I believe it I feel empowered to do something about it. Visualizing is critical to decision making."

"Beyond the importance of visualising data, we need to have access to data and present it in a way that will enable decision making at the highest level."

In summary the main lessons that are learnt from the CGC concerning the use of data to respond to urban challenges are

5.1.2.1 Importance of access to data

Data is a critical tool to facilitate informed responses to societal problems. There is therefore a need for unhindered access to relevant data that is hosted by different organisations.

5.1.2.2 Importance of participation

It is more efficient for organisations that have data to form synergies and share data and knowledge than to work individually in response to urban challenges.

5.1.2.3 The Power of GIS and visualization

GIS plays a key role in providing analysis functions that allow for the data to be converted into meaningful information products. It also allows for patterns that are embedded in data to be uncovered and this has a huge impact on the way that evidence is presented and the potential of its use.

5.1.2.4 Centralising services

Centralising services reduces duplication of effort in data creation. It also allows for more efficient and cost effective workflows since GIS may not be the core business of an organisation but an essential service that is not necessarily needed all the time. Outsourcing the GIS services would therefore be more sensible economically.

The setup at CGC is characterised by three main actors as well; Data Custodians, Administrators (the CGC) and individual users who consume the information resources produced from the data.

5.1.3 The Case of the Grabouw Sustainable Development Initiative (GrSDI)

The GrSDI is a pilot project that aims to develop and test a Sustainable Development model that can be replicated in other parts of South Africa. This project is funded by the Development Bank of South Africa (DBSA) which also has an active role in implementing the program (Hamann et al., 2008).

11 Interviews were conducted in Grabouw to get an understanding of the problems in this area and the response plans being tailored by the DBSA and other stakeholders in the initiative. These included Municipal Officials, the Business community, the farmers' community and the local residents of the town. Four main issues are gathered from the interviews and these are described below.

5.1.3.1 The Need for a Monitoring, Evaluation and Reporting (MER) Framework

The GSDI requires an MER to monitor the progress of the initiative. There is a clear indication from the interviews especially from the municipal authorities that since the GSDI commenced, there has been some work that has been done, but a lot of it is not quantified and therefore it is difficult to tell whether the initiative is achieving its goals. Below is an extract from one interview with a municipal official that echoes what has been mentioned here.

"There are goals, objectives and KPAs (key performance areas) and targets for performance as well as time and expenditure that must be communicated."

5.1.3.2 An Independent Organisation/Person to do the MER

"Maybe we need someone from outside to sit us down and explore what we are trying to achieve..."

Just like the CGC in the previous section, the stakeholders of the GSDI have expressed the need for an independent and objective individual or organisation to help with monitoring the progress of the initiative.

5.1.3.3 Communication

Communication emerged as one of the most pressing challenges for this initiative. Whilst some of the community representatives remarked that, *"Nothing has been achieved that people can see and delivery is a big problem... Excitement was built up and then nothing happened"* one of the municipal authorities was *"happy with the progress of the infrastructure project".* This is an indication that there actually has been some work done but the local communities have not noticed it.

Some other remarks that are gathered from the interviews that illustrate the definite need for more transparent communication are listed below:

"The forum should have a communication medium/strategy which can continuously break down some of these complicated communications."

"The initial vision of the SDI was never caught by most of the people involved... Information needs to be communicated more effectively to all stakeholders involved."

"The community however doesn't know what is happening with these proposals."

5.1.3.4 Community participation

The last important issue that is identified from the interviews is that there is a need for the community to participate. So far, this has been done mainly through community representatives who are actually not well represented in the Stakeholder forum. This has meant that their voice has not been adequately heard. Instead, the funders as well as the business community have dominated the forum meetings. This has resulted in the community not having a deep sense of ownership of the initiative as highlighted in the extract below:

"There was a perception within the community that the DBSA which was a role player from the beginning through sourcing capacity and providing skilled professionals to start the projects was the key driver and therefore the community had no ownership of the developments."

5.2 Final Use Case

From the requirements assessment above and the new SDI model that was presented in Chapter 3, a composite Use Case diagram was created to illustrate the perceived uses of the SDI and how the different actors would interact with it. The three Actors that were described are consistent in all the cases presented above and these are 1) Data Custodian, 2) Administrator and 3) the Individual User. In this case, the data custodian represents the individuals or organisations that have data that is potentially useful in responding to the urban challenges. In some cases, these organisations would have been responsible for creating these data sets. The Administrator represents the authority that is responsible for coordinating most of the SDI activity including data management, system administration, metadata management etc. The individual user is anyone who can access the system for any reason from research purposes to getting general information on selected themes.

Five main use cases are identified

- Manage Clearinghouse
- Author GIS Services
- Manage various reporting material
- Manage Spatial and Non Spatial Database
- Publish Services onto the web

These use cases are a step towards a more technical description of what is to be achieved by the system and how the different actors interact with the SDI. Below is a description of these Use Cases and how the different actors fulfil their roles in the system. A composite Use Case Diagram is shown in Figure 5.2.

5.2.1 Manage Clearinghouse

As was shown in chapter 3, metadata is required for data to make sense and not be misused. Context based metadata gives a better meaning to data and facilitates proper use of datasets. A clearinghouse facilitates access to spatial data and services by searching a collection of metadata resources. These can be stored locally on a central server or can be stored in distributed sources. The Administrator is responsible for managing the clearinghouse. There are many technical and operational issues when managing the clearinghouse and these are left for the next chapter, which presents the Architecture of the SDI. Within the Manage Clearinghouse use case, three use cases were developed and these are explained below.

The “extend” Labels on the Use Case diagram indicate an extend dependency. It is a generalization relationship where an extending use case continues the behaviour of a base use case.

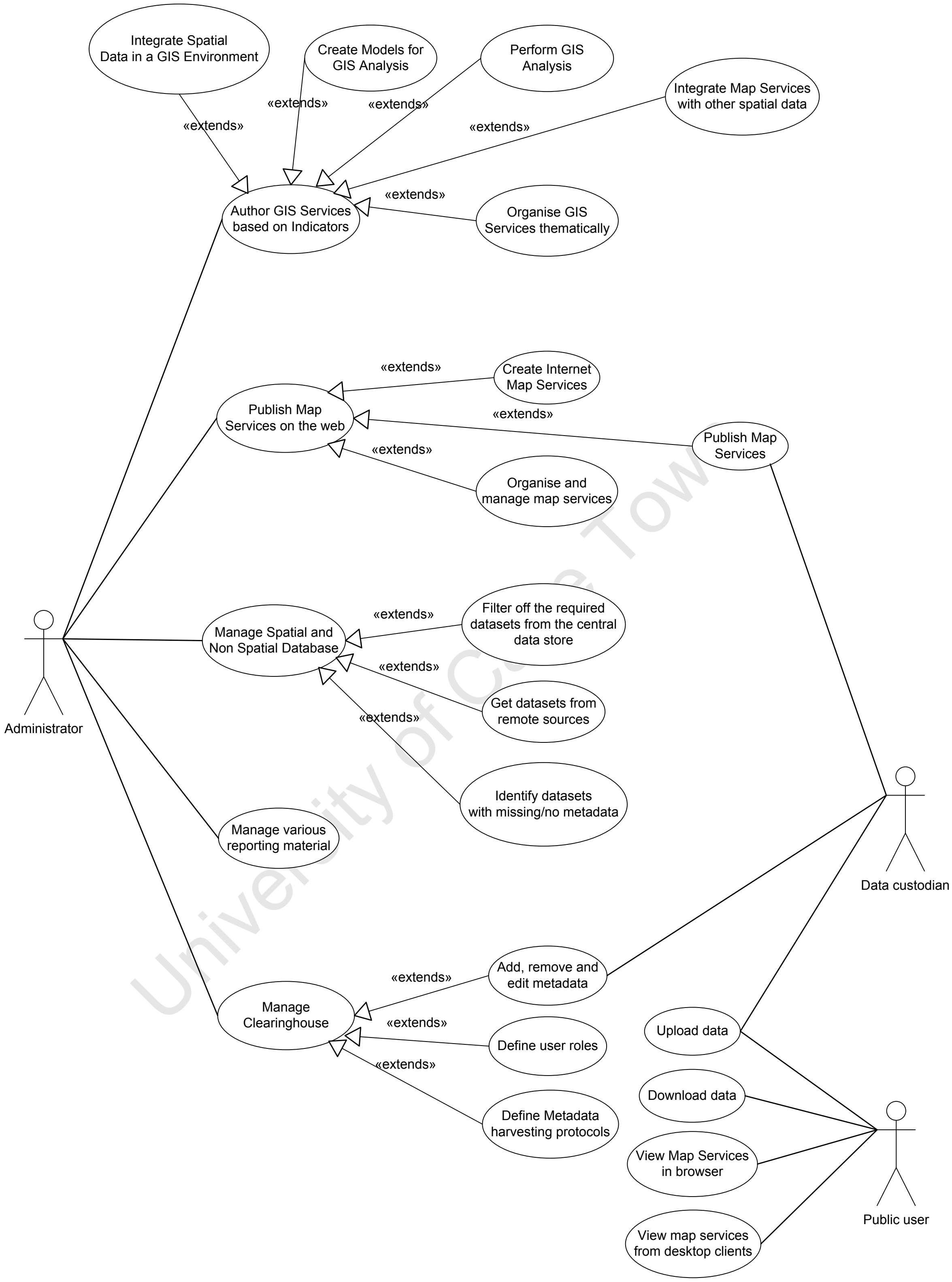


Figure 5.2: The Proposed Use Case Diagram for the New SDI Perspective

5.2.1.1 Define User roles

The Administrator is responsible for defining user roles. This means that different levels of access are required for different users and user roles. The Administrator should also be able to add other users as administrators to the system.

5.2.1.2 Add, Remove and Edit Metadata

Both the Administrator and the Data Custodian should be able to add, remove or edit metadata records to the system. The data custodians know their data best and the conditions under which it was created and the purpose it was created for, therefore they are in a position to generate metadata records for datasets. The Administrator is responsible for authoring and publishing GIS services and therefore needs to create metadata records for this.

5.2.1.3 Define metadata harvesting strategies

The Administrator is responsible for defining harvesting protocols to get metadata resources from the data custodians. This process involves automatically gathering metadata that is already associated with a resource, and which has been produced via automatic or manual means (DCMI, 2007). The harvesting procedures can be based on agreement between the Administrator and the Data Custodians.

5.2.2 Author GIS Services based on Indicators

GIS is a critical tool in presenting evidence on diverse urban themes (See chapter 3 and the RE section). Geography as an organising principle helps to group data that is contextually related through location. GIS provides the necessary tools to create, integrate and analyse information to present evidence blended into a location context to better inform and facilitate targeted decision making. The Author GIS services Use Case comprises of five Use Cases that are more specific:

5.2.2.1 Integrate Spatial Data in a GIS Environment

The Administrator needs to integrate data resources from different sources into a GIS environment and author specific services that can be discovered on the internet. This allows analysis to take place using desktop GIS software.

5.2.2.2 Create models for GIS Analysis

The Administrator should also be able to create models based on specific themes and define standard GIS processes for analysis. An indicator framework can be the basis for understanding the GIS analysis needs and therefore, the models and processes that are required to be run on the data to produce the desired result.

5.2.2.3 Perform GIS Analysis

Analysis is critical in order to uncover lessons and trends from data. The Administrator should be able to perform GIS analysis on the data to draw meaningful and accurate results that showcase evidence useful for decision making.

5.2.2.4 Integrate Map Services with other forms of data

In instances where data may not be available in its raw form because of many reasons spanning from privacy to size of data, it will be necessary for the Administrator to integrate Map services on the internet with other datasets to perform analysis

5.2.2.5 Organise GIS Services Thematically

As was noted in the literature review, it is essential to organise data and services thematically. This will make data more meaningful and a lot easier to access and assess for fitness of purpose.

5.2.3 Manage various reporting material

The definition of geographic information is changing. The Geoweb is able to facilitate the integration of a wide range of datasets in a GIS environment. The administrator should be able to manage all sorts of information that is relevant to showcase and report on urban issues. These can be Videos, Photographs, documents etc. These should be coupled with other geographic information in a meaningful way to allow for as much learning as possible within a desired scope of work.

The Users and data custodians should also be able to contribute this information in suitable ways. This will allow the SDI to host all sorts of information that is relevant in understanding the societies that are represented by this information. The three organising principles discussed in chapter 3 (themes, context and geography) are the principles by which these potentially large volumes of data are sorted.

5.2.4 Manage Spatial and Non Spatial Datasets

5.2.4.1 Filter off required datasets from central database

Considering that the SDI will potentially host excessively large volumes of data, it becomes necessary to have a strategy of siphoning out relevant datasets as a Data Management principle. Therefore inherent in the workflows of the SDI should be mechanisms to do this. An example is the CUO, which is in the process of getting copies of all the datasets that the different municipalities and organisations within its area of focus have. It would be a daunting task to search for datasets, know which data has metadata and which data is useful in specific instances from the data warehouse by manually navigating through the database. The administrator should have access to more automated ways of doing this.

5.2.4.2 Get Datasets from remote sources

In instances where datasets are too large to be transmitted electronically there is a need to define the mechanisms of data transportation from remote sources and easily integrating this data into the central data store.

5.2.4.3 Identify Datasets with missing or no metadata

It is important that most of the data that is hosted by the SDI either centrally or remotely has metadata. This determines how useful the data will be for different applications. It's therefore important to keep a record and track of data that has or does not have metadata records. If there is a way that the metadata can be created then a decision to do so will have to be made. It is possible to assume that since there is a lot of data in the database then most of it is useful whereas the usefulness of data is revealed through the metadata attached to it.

5.2.5 Publish Services onto the web

Communication is of paramount importance for the success of such an SDI design since it is meant to be a platform for showcasing evidence on a range of urban themes. Data and Products of data analysis should therefore be made publicly available to communicate the intended message to different people. Inherent in the SDI architecture should be the ability to maintain privacy in instances where potentially sensitive information can be published or in instances where the originator of the data has reservations about the data being publicly available. Lessons can be drawn from the CGC, which in some instances does analysis at very low levels of data abstraction and at high very high scales to draw meaningful results at a community level but still maintaining the privacy that is due to citizens.

5.2.5.1 Create and publish Internet Map Services

The results of GIS Analysis should be published onto the web. Web mapping standards provide the necessary tools to facilitate access to map datasets on the internet.

5.2.5.2 Organise and manage map services

These web map services should be organised logically and thematically to make it easy to find and understand them.

The individual user and the data custodian do not have as much a role compared to the administrator. It should however be understood that the SDI is not an end in itself but it is a means to an end. Very little is actually known in terms of the actual impact that such an SDI could potentially have on the face of human settlements in developing nations. It mainly facilitates access to information and information products leaving the decision making process to the responsible authorities. However, through public participation, collaboration and sharing of data, such a platform, will slowly recreate human settlements in a virtual environment. Individual users, research organisations and data custodians use this data in many ways that cannot be imagined at the design stage. Unlike the CGC that has access to many data resources but limits operations and data usage to itself and its operations, this model of SDIs will allow for publication of data resources from many organisations and individuals and at the same time facilitate ubiquitous use of this data. However, it still maintains focus on specific issues that would have been defined as pertinent by the community, scholars, research initiatives and organisations through a participatory and consultative framework.

5.3 The Information Viewpoint

The Data Flow Diagramming technique is applied to model the SDI from the Information viewpoint. Figures 5.3 and 5.4 represent the Level zero and Level one data flow diagrams.

5.3.1 Level Zero DFD

The Level Zero DFD (Figure 5.3) shows an abstract view of the data that flows through the system, the inherent processes and operations on the data and the output from the system.

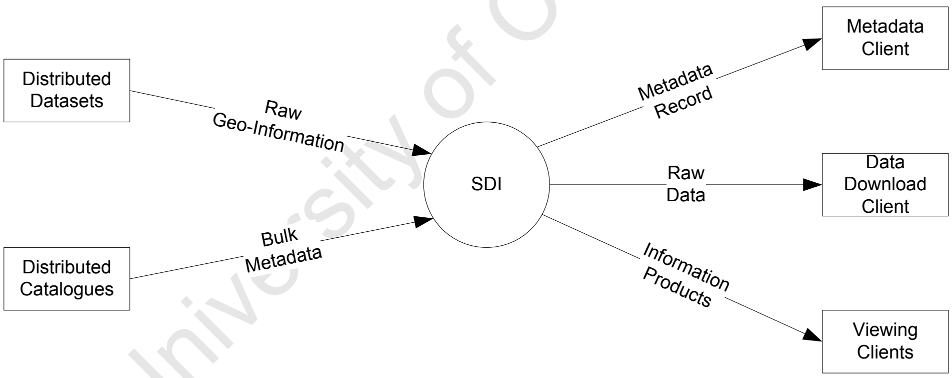


Figure 5.3: Level zero DFD

5.3.1.1 Inputs

There are broadly two inputs that feed into the SDI; these are raw Geo-Information resources from **Distributed Datasets** and bulk metadata that is contained in **Distributed Catalogues**.

The distributed datasets are the datasets that different data custodians have. These datasets are a critical resource for the SDI and it is essential to get access to them in order to use them to author GIS services and to make them publicly available where possible. These datasets can either be stored on a central server or used from the custodians' server remotely through a Service Oriented Architectures (SOA).

The distributed catalogues are a network of metadata catalogues that host the essential metadata to allow data to be discovered and used appropriately. This metadata is harvested from the catalogues using standard protocols and is stored in the central clearinghouse.

5.3.1.2 Outputs

Three types of outputs were identified as **Metadata, Raw Spatial data** and **Information products** based on the data. Metadata should be viewable in at least the standard web browser and in an easy to read format. Users should also be able to download spatial data. This is to allow for a wide range of users to access data for increased research based on the data, which will potentially yield meaningful output to facilitate decision making. The viewing client must allow for data and information products to be viewed on the web. There are a number of options for the viewing clients that can consume the web mapping services published through the SDI and these are described and explained in Chapter 6.

5.3.2 Level One DFD

To understand the information semantics and processes in detail, and consequently the functional requirements for the SDI, it is necessary to break down the single SDI process of the level zero DFD into more specific processes. These are shown in Figure 5.4. This in turn is useful for determining the functional decomposition of the whole system. In the following description of the DFD, the processes performed on the data are highlighted in bold.

The distributed datasets can either be stored on a central server **(Spatial Data Storage)** or used directly from the data custodian using web mapping services. Central storage tends to be more efficient since the access to data is not hindered by the constraints that are introduced by trafficking of the data through the web. However, it may not be applicable in all cases, as some data custodians may need to preserve the privacy contained in their data and therefore may not be too willing to give the data away. In such instances, it might be appropriate to access the data through information services published from the data over the web. An example of this approach is the Google Application Programmable Interface (API), which allows users to access the world maps and embed then in web mapping applications without necessarily having to access the raw datasets.

Whatever method of access, the data will in most instances need to be integrated with data from other sources. This process of **data integration** may require some operations to be done on the data to allow for the data to be interoperable. This may require some coordinate transformations, format conversions, editing the attributes etc. Data integration becomes the starting point of creating meaningful datasets for later use. There are however some limitations to the manipulations that can be done on data that is available as a service, even though this information is normally served in ready to use formats. For example, the OCG Web Map Service (WMS), is served in the standard WGS 84 co-ordinate reference system, which most GIS packages are able to read and in some instances perform “on the fly” coordinate transformations to integrate with other datasets. However one may not alter the table structures for the attributes of the data which would have been used to author the WMS without access to the raw dataset.

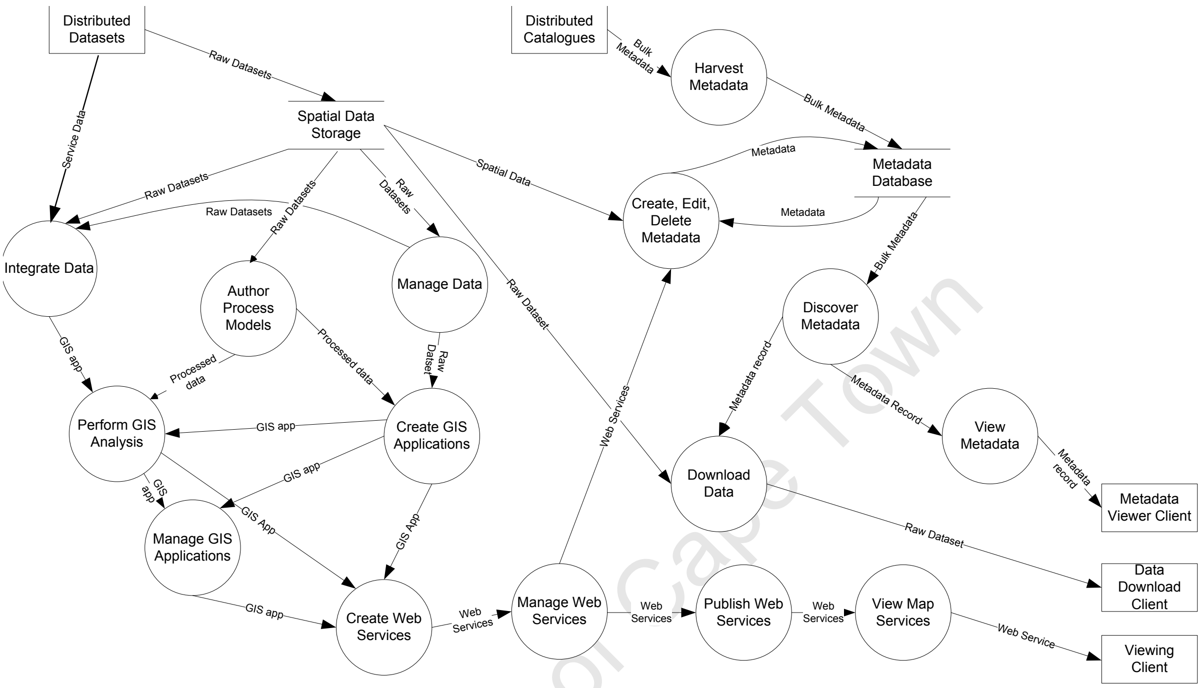


Figure 5.4: Level 1 DFD

There is a need to manage the data that is stored in the Central Server. **Managing data** improves operations and allows for the monitoring of the data that is stored locally, as well as a better understanding of its properties and potential use. It may involve functions such as “crawling over” various datasets in a directory structure, to determine the number of datasets registered on a specific coordinate system, number of datasets with or without metadata, finding datasets that have particular fields in their attributes etc. Managing Data also feeds into the data integration process because once there is a clear understanding of the available datasets then these can be more efficiently used together for different purposes.

The datasets are used as input when **authoring process models** for analysis. In some instances, it is necessary to create and keep models that allow for routine operations on the data. The advantage of process models is that when the data elements are changed the model will generate a new result rather than making use of static datasets.

The data that is managed in the SDI is also used in the **create GIS applications**. In this case, datasets are carefully chosen from the database depending on the intended purpose of the application. In the applications environment, the data can be analysed using the pre-authored models. In some instances, it will be essential to create applications that simply display the layers of available datasets with no major processing. It is important to author different applications if they are meant to display data on different themes. This allows related data to be bundled together and this can make the use and discovery of data potentially easier. As a result, there will also be a need to **manage GIS applications** as well as the information products that result from the GIS analysis.

Since it is important to communicate the results of the GIS analysis and data themes available through a public platform, the **GIS applications should be published** onto the internet to be consumed by a wide range of clients. The desktop GIS applications are used to **create web services** using standard protocols. This can result in services that are as many as the GIS applications and therefore the **services will need to be managed** in order to achieve efficiency in effectively communicating to the intended audience. Part of the management process involves creating services metadata, which is documentation of the different services, their capabilities and other properties. Services metadata is created and stored in the **Metadata database**. Standard protocols are used to expose the metadata to external applications that allow the data to be discovered and described on a public platform.

Metadata for spatial data is also created and stored in the same metadata database. The administrator has to define the user roles and rights of access to the metadata database for the different users. Some may edit data; others can create new metadata whilst others are able to simply view the data. However, for metadata that exists on different servers that belong to the data custodian, the administrator will need to define harvesting protocols that allow the system to "crawl over" **distributed Catalogues** to harvest metadata at predefined times. Users can automatically update these metadata records on the central metadata database to make them discoverable.

The discovered metadata provides direct links to **downloadable data** stored in the spatial data store. These are downloaded through an appropriate interface. Once data has been discovered it should also **be viewable** for the appropriate clients and used for GIS analysis.

5.4The Computational Viewpoint

The Elements of the SDI Computational viewpoint according to Cooper et al (2007) are adopted in the computational viewpoint of the proposed SDI. This is to get an understanding of the components that make up the SDI and how they interact. It is chosen to adopt this model because it exhibits all the characteristics that are presented in the DFD presented above. Below is a description and illustration of the service components that are presented in the component model.

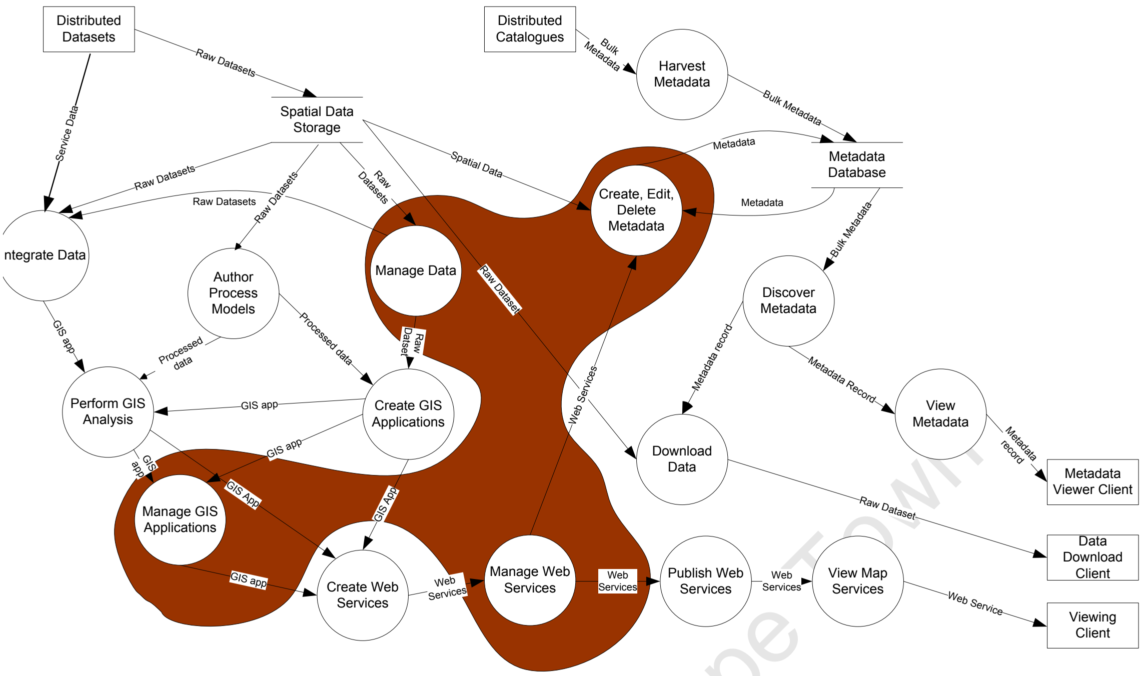


Figure 5.5: Management Services

5.4.1 Management Service

The Management Services component monitors the overall functionality of the SDI. It exposes functionality that focuses on managing the SDI. Some of the functionalities that are contained in Management services are: Managing GIS applications, Managing Web Services, Managing Data and Managing Metadata (Figure 5.5)

5.4.2 Registry Services

The main purpose of the component Registry Services is to register other services, publish them, and later to search through them. Highlighted in Figure 5.6 are the following features. The Harvest and Create, Edit and Delete metadata functions allow for metadata to be registered in the database. The discover metadata function describes a search process through the database. Create and publish web services are also part of this component.

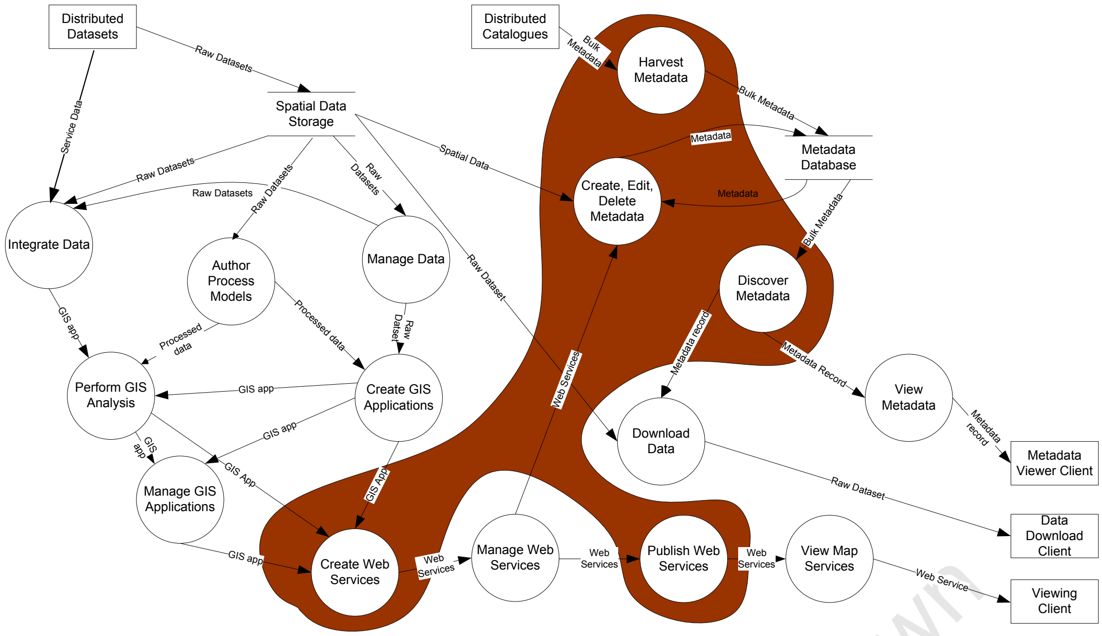


Figure 5.6: Registry Services

5.4.3 Data Services

The Data Services component deals with data sets shared and registered on the Internet. For example, Data Services provide access to collections of data in repositories and databases as well as integration of the datasets to create GIS applications (Figure 5.7).

5.4.4 Portrayal Service

The Portrayal Services component deals with displaying the results of application services. It is a critical component of the SDI as it allows for the data and services that are offered to be visualised and made sense of appropriately for decision-making. The functions that make up this component are shown in Figure 5.8.

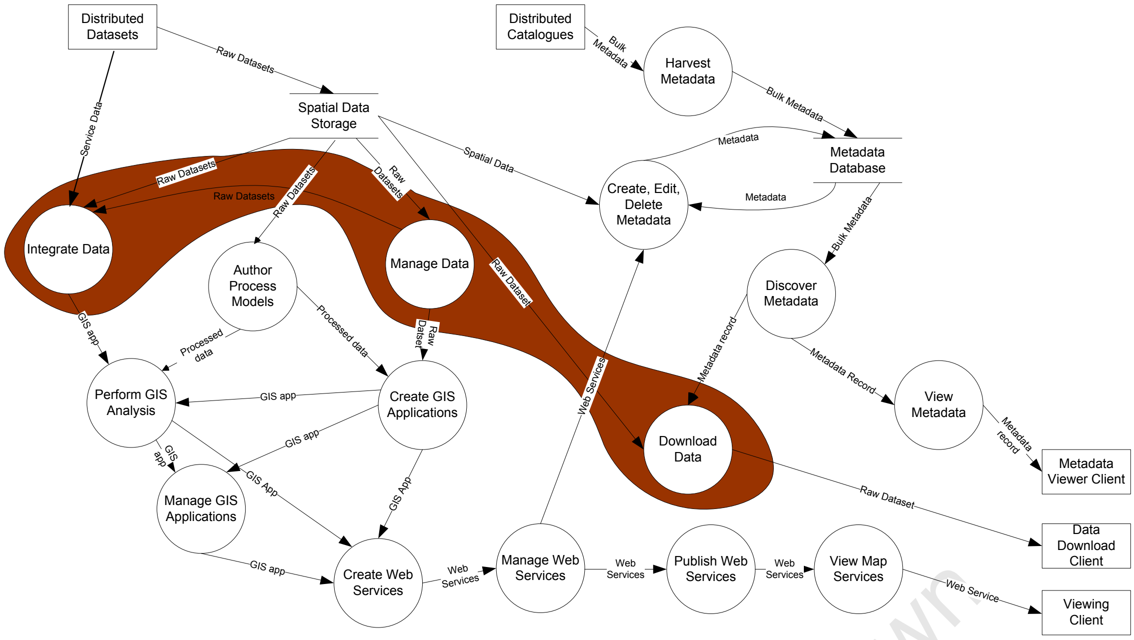


Figure 5.7: Data Service

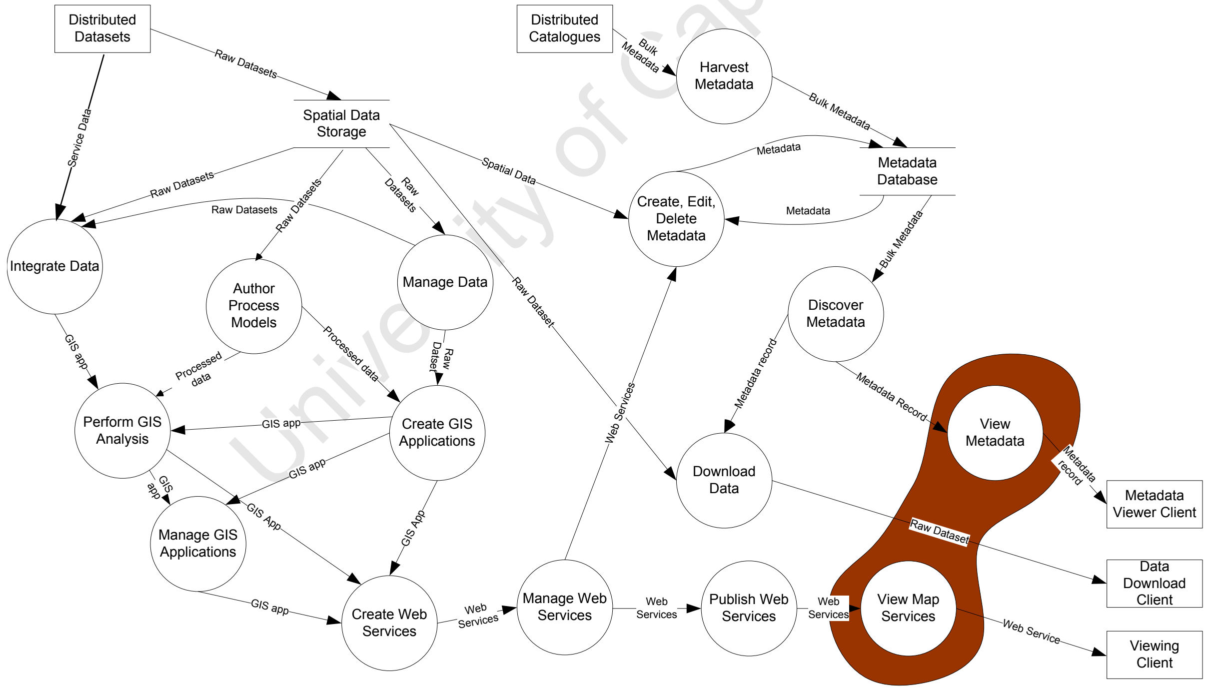


Figure 5.8: Portrayal Services

5.4.5 Processing Services

The Processing Services component depicted by Figure 5.9 contains the services for data processing. These different processes run on the data elements to produce a result that is displayed on a mapping interface.

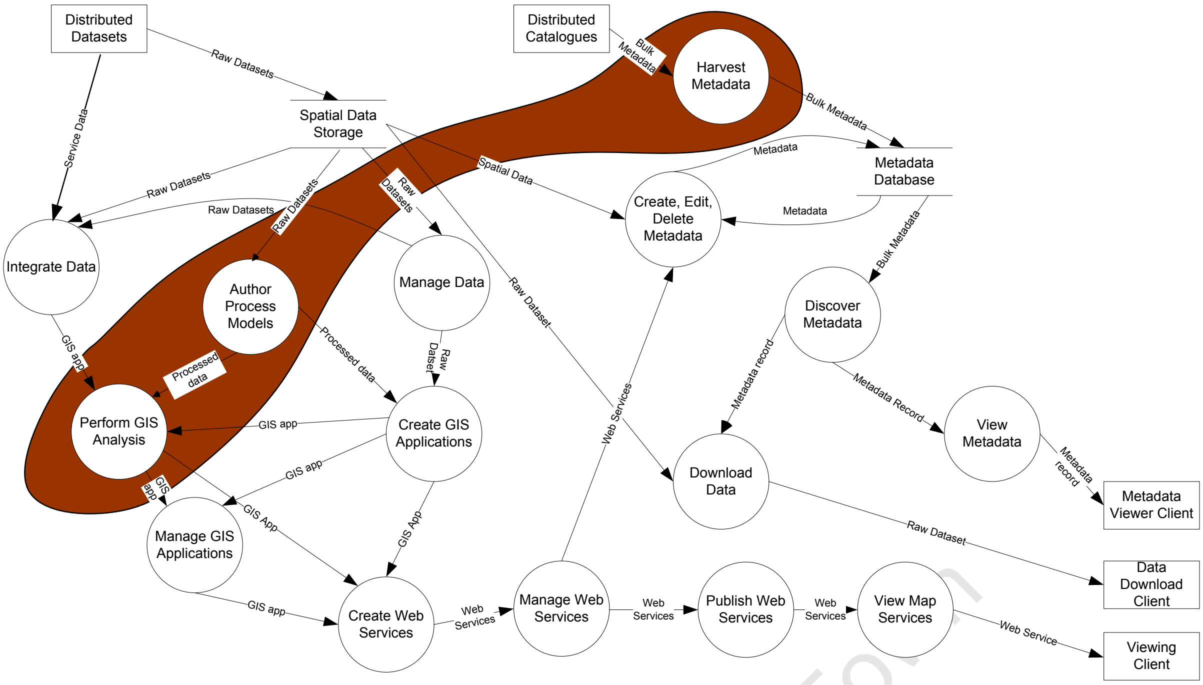


Figure 5.9: Processing Services

The computational viewpoint of SDIs as was modelled by the International Cartographic Association, blends in very well with the requirements of the SDI that was proposed here. This is because the component diagram that was presented by the ICA is an abstract model. The composition of the individual components may vary depending on the needed functionality of the SDI. For example, although the first generation of SDIs had processing services in the form of the protocols that are executed to render a map online or a set of metadata records to be retrieved; they did not have the set of service capabilities that allowed for intensive data analysis.

The next chapter builds upon the models that have been presented in this chapter and focuses on the Architectural Design and implementation of the SDI using relevant technologies. It demonstrates the incorporation of indicators as tools that build the principles of evidence based practice into the mainstream SDI, and illustrates how this can potentially foster learning of urban trends and aid in better decision making in dealing with urban challenges.

Chapter 6: The Architectural Design

6.0 Introduction

This chapter presents the last two viewpoints of the Reference Model for Open Distributed Processing (RM-ODP) for the SDI. These are the Engineering and Technology viewpoints. The Engineering viewpoint describes the Design structure of the SDI and the Technology viewpoint describes the specific technologies that are used for implementation. The elements in the Analysis model that were presented in Chapter 5 serve as input into the design process. Four sub-systems that make up the SDI are generated from the Data Flow Diagram and they are each described in detail. Unlike in the last chapters where the SDI viewpoints were presented in sequence, both the design and the technologies used to demonstrate how it can be implemented are discussed concurrently. A diagram that depicts the summary of the final Architecture design in presented towards the end of the chapter.

6.1 The Subsystems for the SDI

Four subsystems as shown in figure 6.1 have been identified as the main elements that constitute the SDI Architecture. These are a) The Desktop GIS, b) Internet Map Services, c) Data Discovery and d) Client domains. The subsystems are modelled from the DFD that was presented in the Analysis modelling chapter.

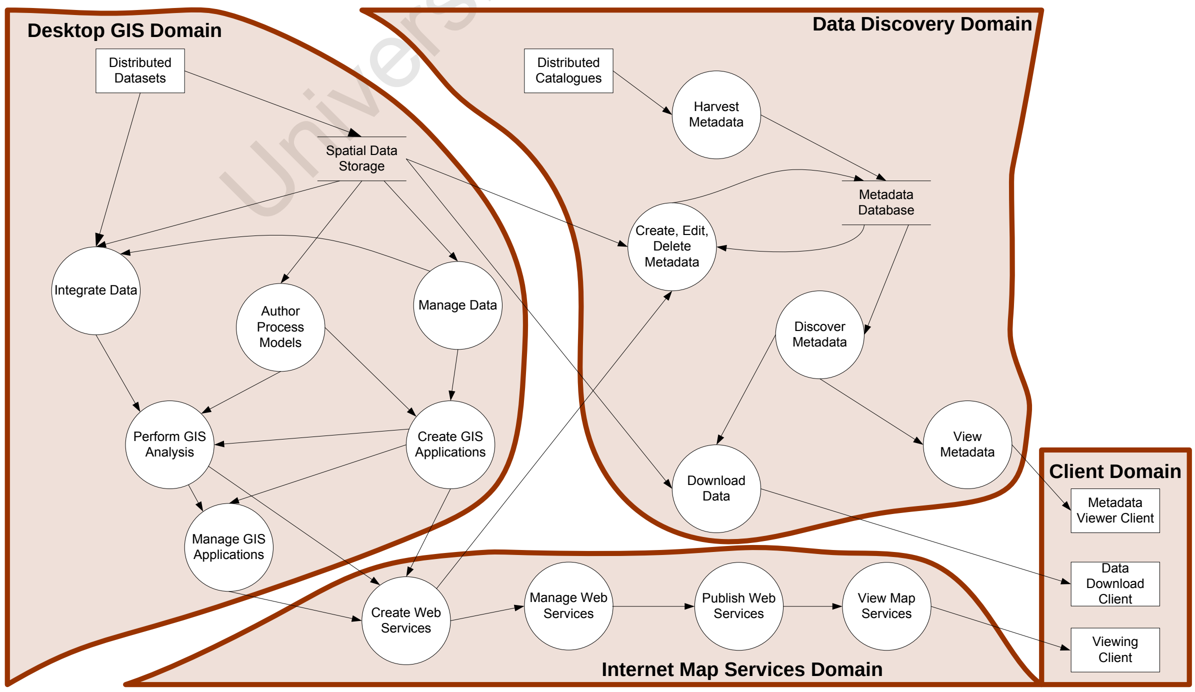


Figure 6.1: Subsystems making up the SDI

The main characteristics of these subsystems as well as the different technologies that are used to implement them are shown in Table 6.1.

Table 6.1: The technologies used to implement the SDI Subsystems

Category	Technology		Application/Role
Desktop GIS	ArcGIS Desktop		Integrating data from different sources and creating GIS applications. Defining processes using modelling tools to perform various forms of analyses.
	Arc Catalog		Managing Datasets stored locally as well as web mapping services.
	ArcGIS Geodatabase		The platform for storage of spatial and non spatial data
Internet Map Services	ArcGIS Server		Authoring, managing and publishing map services based on the data and applications from the desktop GIS domain. These services are mainly vector based.
	ArcGIS Image Server		Creating and publishing services that are based on raster datasets.
	OGC Standards		These define the standard protocols that are used to publish GIS data onto the web.
Data Discovery	Geonetwork Open Source		Managing the harvesting and discovery processes for metadata. It is also used to provide links to these data and online resources. It also serves as a content management system providing access to a wide range of information types other than GIS based ones (e.g. Photographs, documents etc).
	OGC Standards		Standards implemented to facilitate the discovery of Spatial Data and services on the web.
	Postgre SQL		The database that is used to store the metadata and other information that facilitates the data discovery process.
Client Domain	Thin Clients	Browsers	These consume light weight applications that have most of the data processing done on a central server.
		Mobile Clients	Mobile clients access the SDI through relevant thin client applications (like the browser)
	Fat Clients	Desktop GIS	Integrating services and data to consume more process intensive GIS based applications.
		Globe Clients	Consuming more process intensive GIS based applications in a 3D environment.

6.2 Desktop GIS Domain

This domain of the SDI design contains the necessary functions that allow for data to be integrated and GIS applications to be authored. ArcGIS Desktop 9.3, a leading proprietary GIS software package is chosen as the platform for implementing this part of the SDI due to the following reasons:

a) It is one of the most mature GIS software in the industry and offers a wide range of easy to use analysis function tools

b) ArcGIS desktop 9.3 is part of a bigger family of the ArcGIS 9.3 suite of softwares that are used in the Internet Map Services and Client domains of the SDI. Using the same suite of GIS softwares for the subsequent domains makes their implementation and workflows better manageable.

c) The software was readily available to the researcher through the University of Cape Town where the research was conducted

d) The software is relatively easier to use compared to other GIS packages on the market

While the decision to use the ArcGIS Suite was arrived at after the above considerations, in instances where the software is not accessible due to financial constraints, Free and Open Source Software (FOSS) are a more viable option. There are also a wide range of both Open Source and Proprietary GIS software options that can be considered when implementing the SDI.

Figure 6.2 below is a summary of how the different parts that make up the Desktop GIS domain. There are two main parts, the application layer where the GIS applications are authored and various forms of GIS tools are used to manipulate data and the data layer that defines how data is stored in a local Geodatabase and how it can be accessed from remote sources to be used in GIS applications.

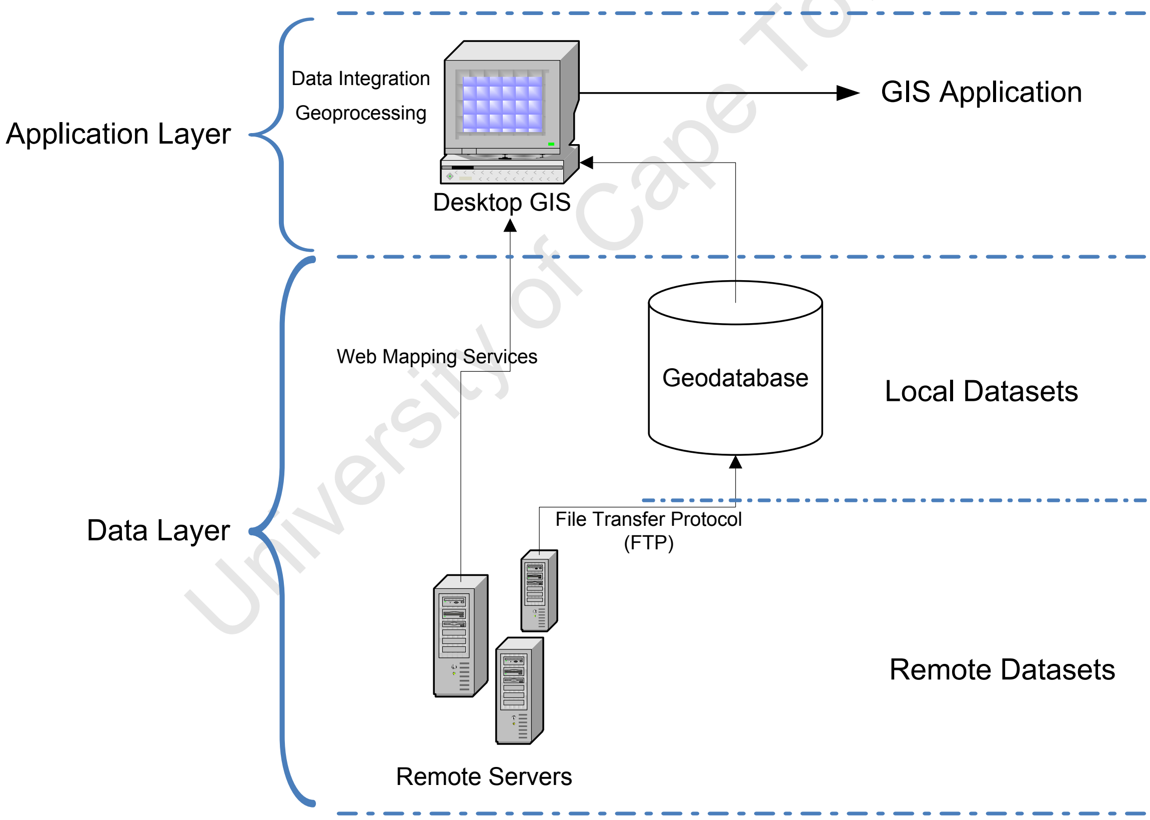


Figure 6.2: The desktop GIS domain for the SDI

6.2.1 The Application Layer

Within the application layer there are two main functions; Data Integration and authoring GIS applications and performing GIS analysis.

6.2.1.1 Data Integration

Authoring GIS applications requires integration of spatial datasets from a variety of sources. The administrator either integrates data from a local drive or from data that is stored remotely and is made available through web mapping standards. In the latter, the administrator has limited capabilities to manipulate the data. However, some datasets remain useful even when one may not be able to perform analysis using them, particularly when they are required to form the base layer for the other datasets. A common example of such an instance is the use of images services to form the backdrop of a mapping application (see Appendix 4).

The process of data integration is a critical and necessary component in this new SDI concept because to derive maximum value from the mapping and data capture activities that are being done by different organisations, a strategy must be formulated that allows for these datasets to be used together. This process involves, format conversions, coordinate transformations and restructuring and merging attribute tables among other issues to make the dataset usable in one application for analysis purposes. Some of these processes are done automatically in ArcGIS 9.3 whilst some require the intervention of an operator.

6.2.1.2 Authoring applications and performing GIS analysis

Applications should be authored strategically to present information on pressing societal matters using indicators generated through a community participatory process. Indicators therefore become the guiding rod for authoring relevant applications that are made publicly available through the SDI to serve as tools for presenting urban trends.

Analysis of data is performed through GIS analysis functions pre-packaged in GIS Software. This is how meaningful information products are produced from the raw datasets that typically characterise the output of most SDIs. The results of analysis are more useful in presenting thematic evidences than simply availing raw datasets. Process models are used to define routine tasks that can be performed on data to yield a desired result. Figure 6.3 is an example of such a model, constructed using ArcGIS 9.3 Model Builder.

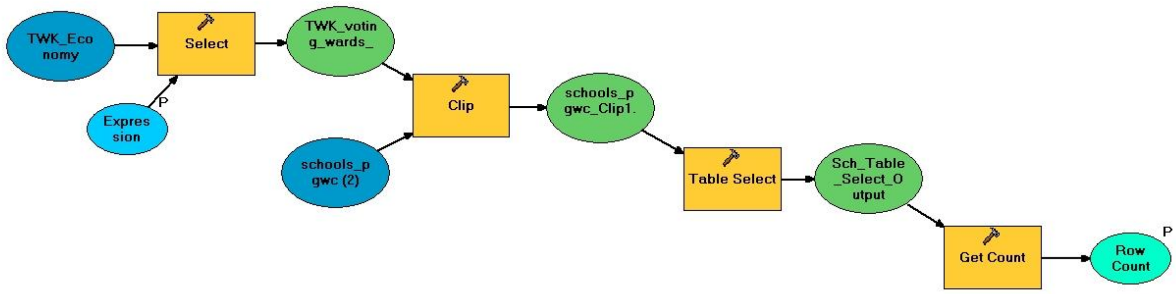


Figure 6.3: Model for calculating the number of schools that have access to piped water in a specific ward. Created using ArcGIS 9.3 Model builder

The model above shows a sequence of model steps that are executed to find out the number of schools within a chosen ward in the Theewaterskloof (TWK) municipal region that have piped water facilities. The ward boundaries for the TWK are defined in the 'TWK_Economy' dataset. Upon running the model, the user is prompted to select the ward that they require to find out information about. The schools that are in the selected ward are extracted from the schools layer ('schools_pgwc') through the clip command. The 'schools_pgwc' layer contains a field which contains water access attributes. There are five variables in this field namely; Borehole, Dam, Piped, Tanks and Other. A select query is run on the clipped schools layer to extract only the schools that have piped water in the ward and then a 'Get Count' command is executed to count the number of records in the new dataset and consequently the number of schools that have access to piped water.

Models result in well-defined workflows to combine data layers and run analysis functions on the data to yield a desired result. Their advantage is that they automate processes and therefore more easily expose GIS functionality to the user. The administrator creates the models whose results reflect the specifications of an indicator set.

In some instances, data processing will not be required; for example, users may need to view the properties of a dataset through the metadata records and then simply view the dataset online and download it. The datasets are still integrated in the ArcGIS environment and then exposed to the user community through web mapping standards (section 6.3)

6.2.2 The Data Layer

This layer contains the data and mechanism of acquiring and using data to author the GIS applications. It has two parts; the remote datasets that are hosted on different machines by data custodians and data producers and the local datasets that are stored in a geodatabase on the central server.

6.2.2.1 Remote Datasets

These are the datasets that are stored remotely by the data originators or custodians. They are made available to the GIS desktop environment in two ways:

a) Through web mapping services as was discussed in section 6.2.1.1
b) By uploads onto the central server through a File Transfer Protocol (FTP). This method will allow for user to gain access to the server through the internet and upload data. Each user is assigned a user name and password by the administrator to ensure that only authorised people end up uploading data. A separate section in the file structures can be left for unanimous users to also upload their data. The FTP site can be configured to set different access privileges to different users. Once this is done, the uploaded datasets may be integrated into the Geodatabase structures to be used for analysis in a GIS environment.

6.2.2.2 The Geodatabase

ArcGIS provides two main options for storing potentially large datasets. These are the File Geodatabase and an ArcSDE Geodatabase. A third option, the personal Geodatabase is not a suitable one for large datasets. Table 6.2 below gives a concise comparison of the two formats.

Table 6.2: A comparison of the File and ArcSDE Geodatabases (Source: (ESRI, 2010))

Characteristic	ArcSDE Geodatabase	File Geodatabase
Description	A collection of various types of GIS datasets held as tables in a relational database.The recommended format for data stored in a relational database	A collection of various types of GIS datasets held in a file system folder The recommended format for data stored and managed in a file system.
Number of Users	Multiuser Many readers and many writers	Many readers or one writer per feature dataset, feature class, or table.
Storage Formats	Oracle Microsoft SQL Server IBM DB2 IBM Informix PostgreSQL	Each dataset is a separate file on disk A file geodatabase is a file folder that holds its dataset files.
Size Limits	Up to DBMS limits	One TB for each dataset. 256 TB for extremely large image datasets. Each feature class can scale up to hundreds of millions of vector features per dataset.
Versioning Support	Fully supported across all DBMSs; includes cross-database replication, updates using check-out and check-in, and historical archiving	Only supported as a geodatabase for clients who post updates using check-out and check-in and as a client to which updates can be sent using one-way replication.

Characteristic	ArcSDE Geodatabase	File Geodatabase
Security and Permissions	Provided by DBMS	Operating file system security
Database Administration tools	Full DBMS functions for backup, recovery, replication, SQL support, security, and so on	File system management

A file Geodatabase structure was chosen over the ArcSDE Geodatabase for this research due to a number of reasons that are explained below:

1. Data from remote sources is uploaded into folder structures on the central server through the FTP. Defining the workflow for integrating this data into the file geodatabase structures automatically is a lot easier than doing the same for an ArcSDE Geodatabase.
2. The environment within which the clearinghouse was created (see section 6.4) does not support data downloads from an ArcSDE Geodatabase instance
3. The format is recommended by the software manufacturers for data stored and managed in a file system.
4. ArcSDE's main advantage is that it allows multiple users to edit the data at the same time. However, that is not necessary for an SDI considering that there is minimal editing of data; this is done by the data providers.
5. The size limits for a file Geodatabase favour storage of large amounts of data

Storing data in a file Geodatabase allows for compression and compaction which reduces the storage requirements and tidies up the storage of record by reordering them and clearing free spaces respectively, unlike storing the data directly on the hard drive using the operating systems' file structures (Childs, 2009).

6.3 The Internet Map Services Domain

The Internet Mapping Services domain can be summarized by three operations as shown in Figure6.4 below.

1. Creating Internet map services from the GIS applications
2. Creating Web mapping applications based on the services and the data they represent
3. Managing services

This is an important part of the SDI as in enables data and information that is contained on servers to be availed to interested and relevant people through the internet. It is in this domain that useful information products that are based on the data used to author the applications in the desktop domain are made publicly available to showcase specific urban trends and patterns.

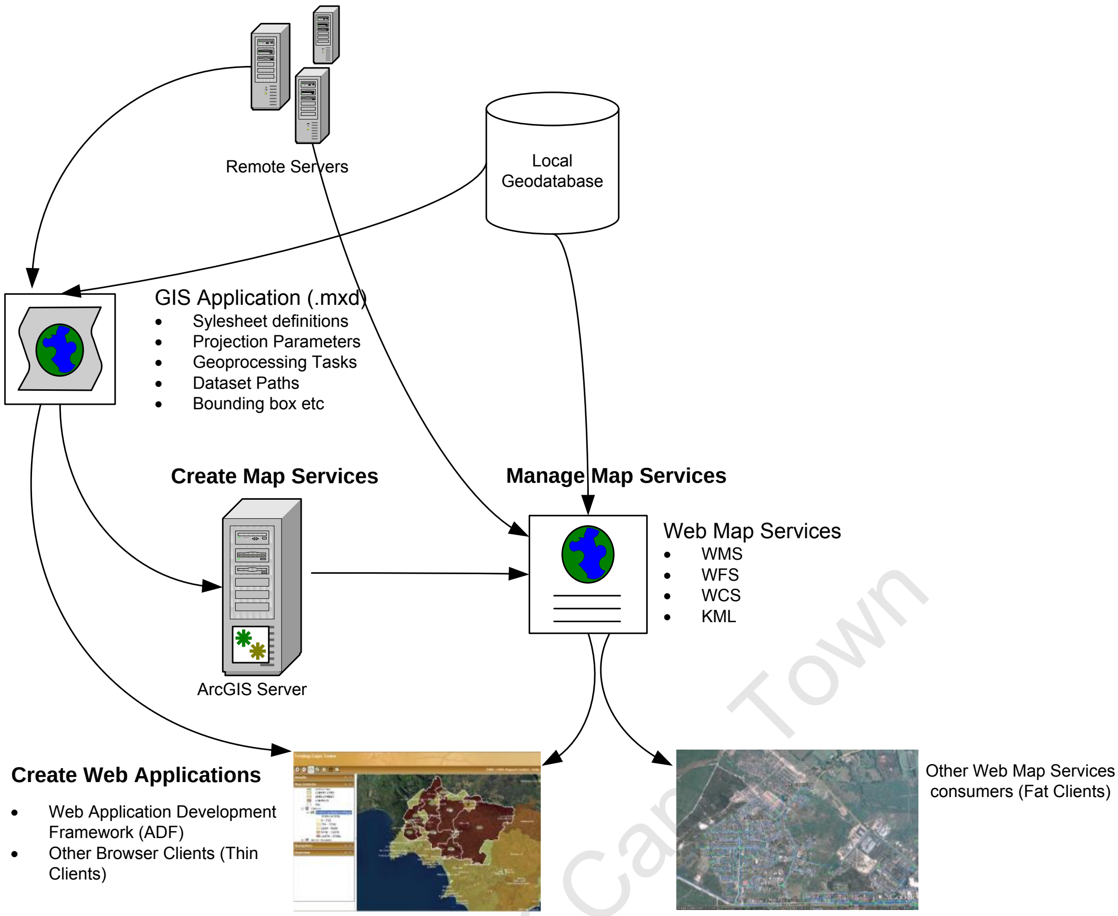


Figure 6.4: The Internet Map Services Domain

6.3.1 Create Internet Map Services

ArcGIS server comes with standard tools for converting the desktop GIS applications into mapping services. This process makes the data publicly available using open standards. When a GIS desktop application is authored, the properties of the application are stored in an mxd file. These properties include stylesheet definitions, projection parameters, the geoprocessing tasks, paths to the datasets used in the application, the bounding box etc. All these parameters are used as input into the services that are created in ArcGIS server. When the map service is created, it establishes direct connection with the datasets that were used in the application and configures them applying the parameters from the mxd file.

The most common standards in the web-mapping arena are shown in Table 6.3. While the first standard in the table helps to make metadata records available on line, the other four facilitate for the discovery and use of map-based content.

Table 6.3: OGC standards used to implement the SDI (OGC, 2009)

Name of Standard	Description
Catalogue Service for the Web (CSW)	Specifies the interfaces, bindings, and a framework for defining application profiles required to publish and access digital catalogues of metadata for geospatial data, services, and related resource information.
KML	An International standard language for expressing geographic annotation and visualization on existing or future web-based online and mobile maps (2d) and earth browsers (3d)
Web Coverage Service (WCS)	Defines a standard interface and operations that enables interoperable access to content such as satellite images, digital aerial photos, digital elevation data, and other phenomena represented by values at each measurement point.
Web Feature Service (WFS)	Allows a client to retrieve and update geospatial data encoded in Geography Markup Language (GML) from multiple Web Feature Services
Web Map Service (WMS)	Provides a simple HTTP interface for requesting geo-registered map images from one or more distributed geospatial databases

KML, WFS and WMS are used to render vector data while the WCS is used to render raster based services. The main function for the WCS is to expose raster analysis capabilities over the web. An Image service more efficiently serves raster data especially aerial photographs for providing some background data. Image processing and image distribution are traditionally regarded as two separate stages in image exploitation. An Image service however combines the two processes and makes them faster. The image service definition is used to define an image service including the data, processing, rendering, metadata, extent, and properties. The original sets of images are compressed before being sent through to a client application over the web. This compression reduces the size of transmitted imagery but puts an additional load on the service provider to compress the data first. However, efficiency is increased significantly. For example, the Aerial Images for the whole of Cape Town are close to 8 GB in size but the Image service definition (see Appendix 5) that was created from the same set of images is close to 3.5 GB. Besides the fact that the images are significantly compressed, tiling makes the service render the images faster because only the tile that corresponds to the area that the user is viewing at a specific scale is transmitted through the internet and not the whole image set in cases where no tilling would have been applied. Not compressing the data results in better quality images but requires significantly more bandwidth to be transmitted over the web.

WFS serves data in vector format over the web and will be very useful in instances where the users will be required to edit the data. WMS however work well in instances where data simply requires to be viewed through a mapping interface or queries done on the data that send back a map or tabular result to the user. WMS however results in poor picture quality when used in globe applications like Google earth and ArcGIS explorer. This is due to the fact that the WMS standard is not made to factor in the height variable in data and whenever a request is sent to return an output from a WMS, it is sent through either as a planar .jpg or .png image file. Therefore, superimposing the WMS image on a globe results in some distortions where some areas are above the surface of the globe whilst some are below. Some areas will also have lesser resolution than others will.

KML is an eXtensible Markup Language (XML) based standard that is supported by the Open Geospatial Consortium (OGC) and corrects for these errors. It is excellently consumed in 3D Based clients. However, since Google owns the KML data format and therefore maintains it, Google Earth tends to be a much better client for KML network links than ArcGIS explorer is. However, Google Earth has limited GIS functionality and primarily works well to simply view data. ArcGIS Explorer is able to consume tasks and model processes (e.g. Figure 6.3) that are hosted on an ArcGIS server platform, making it more suitable for performing different spatial analyses from a remote station through the internet.

The Map Services described above are authored in the ArcGIS server environment and the metadata for these services are stored in the metadata database. This allows end users to search for these services through the clearinghouse (described in section 6.4) and consume them with various client applications.

6.3.2 Manage Services

The Administrator also has access to tools that for managing services. Managing services involves processes like; deleting out of date services, switching service on or off at certain times, trouble shooting errors in the services (e.g. deletion of source data which would cause a service to stop working) and managing the storage of services.

Services are packaged in such a way that data that is related thematically will be preferably bundled up into one service. This means that if a user makes a request for environmental data for example, all that data is available through one service. An appropriate folder structure should also be defined to store the services on the central server where they will be made publicly available. This makes it easier for the administrator to know how to navigate to specific locations to add, edit or delete services.

Managing services also allows the administrator to start or stop them. This is useful in the event that there is need to make repairs to services or replace them with new data sources.

6.3.3 Creating web applications

Creating web GIS applications is also the Administrator's role. It involves packaging services for the end users to have access to functionality that will enable them to use the data and services. The ArcGIS Server Application Development Framework was used to develop these browser based GIS applications (See Appendix 6). The administrator specifies the functionality what they would want to expose to the end user of the application. In the process of authoring the application, as is the case when authoring a desktop application, the administrator may also integrate services from different sources to communicate whatever message they intend to through the application. The data analysis and processes that are contained in the models can also be used to give users access to tools to perform analysis on the data.

6.4 Data Discovery Domain

The Data Discovery Domain focuses on exposing data and services using metadata. This domain is made up of three parts namely; the Application Layer, services layer and the database layer (figure 6.5). The application layer defines the sets of functions and tools that the user will interact with to search through metadata records. The services layer contains the necessary functions and standards based toolsets and procedures that process requests from the application layer to return a result. The Database layer contains a Relational Database Managemement System (RDBMS) that stores the metadata records.

Geonetwork Opensource is standards based Free and Open Source catalogue application to manage spatially referenced resources through the web. This application was chosen to implement the application and services layers of the data discovery domain. It was chosen because many of the open standards have been built up into Geonetwork and it supports many metadata standards and therefore can harvest metadata from many interfaces. Being an open source application it is widely supported by a large community of users and developers worldwide.

The PostgreSQL Relational Database management system was used for the storage of the metadata and other information that is useful in enabling access to spatial datasets or services.

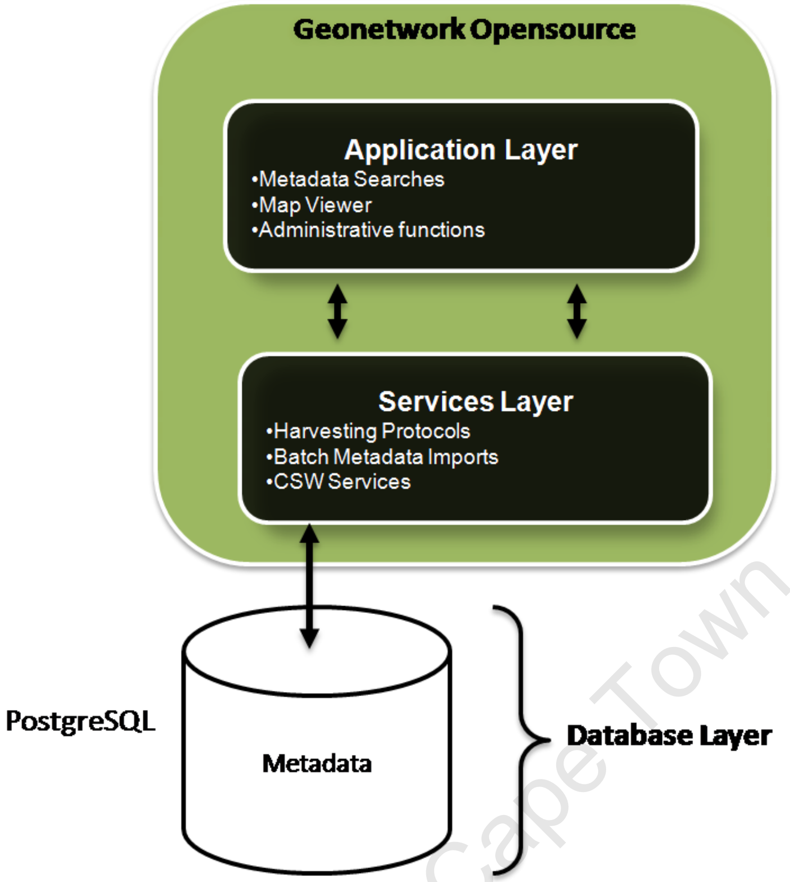


Figure 6.5: The data discovery domain

6.4.1 The Application Layer

The application layer contains a set of interfaces that allow for users to interact with the metadata database through the services provided in the service layer. Below is a list of some of the functions that the user has access to:

Metadata searches: This function provides the users with an interface that allows them to search for metadata record in the database through the use of key words. This input is converted into an SQL query that searches through the database and returns a result that is displayed on the same interface (see appendix 7). The user can then read through each metadata record to see the data that corresponds to their search criteria.

MapViewer: The Map viewer enables the user to consume web map services that are stored both on the central server or any other remote server. This user can use this platform to preview a map service before using it in their application (see appendix 8).

Administrative functions: The application has an interface that allows registered users to log in and access administrative functions. The administrative functions include:

- Creating new metadata
- Importing pre-authored metadata
- User management (changing passwords and changing user information)
- Harvesting management (defining harvesting settings e.g. from which catalogue and at what time is the harvesting being done)

6.4.2 Services Layer

6.4.2.1 Metadata Harvesting and Batch Metadata imports

Harvesting is the process of extracting metadata from remote information resources and adding them to a metadata catalogue (iRODS, 2010). Traditional models for SDIs have been based on distributed searches where the clearinghouse would redirect metadata searches to appropriate servers whenever there was a request. These processes are hindered whenever 1) there is a slow internet connection and 2) when the remote server with the metadata catalogue ceased to function. Through Metadata harvesting, all metadata from nodes registered with the central catalogue are periodically transferred from the owner to the central server to allow for faster and more reliable metadata searches.

The clearinghouse can harvest metadata from a range of standard interfaces namely:

- Geonetwork v2.1 Remote Node
- Geonetwork v2.0 Remote Node
- Web DAV Server
- Catalog Services for the Web (CSW) v2.0
- OGC Web Services (WMS, WFS, WPS etc)
- OAI Protocol for metadata harvesting
- ArcSDE Server
- Local file system

It is the role of the administrator to specify the harvesting protocols so that the central metadata database has up to date metadata records. This is achieved through the cooperation of those that host datasets. The harvesting workflows can be set to ‘crawl’ over remote metadata catalogues at specified times to look for and copy new or modified metadata records. This process also helps to distribute work on creating metadata considering that the administrator would not have any reasonable capacity to generate meaningful metadata records for data that he would not have created. Having the data custodians creating their own metadata as they create the data allows for the context within which the data is created to be captured. This reduces the risk of data being misused.

6.4.3 Database Layer

A PostgreSQL relational database is used as the central repository for information managed through the clearinghouse. While PostgreSQL can be configured to store spatial data, the file structures are used in this research due to reasons explained in section 6.2 of this chapter. The Figure 6.6 presents and overview of the database schema in the form of an Entity Relationship diagram. Entity Relationship modelling is a database modelling method used to produce a conceptual schema of the system and its requirements. The database has 20 tables that are interrelated. This schema shows the primary and foreign keys for each table. For a more comprehensive schema that shows all the attributes see Appendix 9.

The tables that store the data for the clearinghouse contain information on the following broad themes:

Categories: These tables stored the information of the type of data that is stored. The administrator can add and remove categories. Examples of Categories are (Interactive Maps, videos, documents, raw datasets, etc). This capability also makes the clearinghouse a content management system for the different datasets that it is a gateway to.

Languages: Although English is the only language that is used in this project, the database structure will allow for the use of other languages. The use of multiple languages may facilitate better community participation in the creation of data.

Metadata: These tables store the metadata records. Each metadata element is stored in a separate field

Operations: These tables store information that is used to manage the operations that are done on the data elements for the SDI. This includes information that makes some data downloadable and some just viewable in a browser. This security feature ensures that data does not get into the wrong hands and that the data providers' interested are safeguarded whilst maximum benefit is derived from the data.

Region: These tables store information about the Geographical extents within which users can search for data. The tables include that name of the geographical extent (e.g. a district or city) as well as the four coordinates that define the bounding box.

User Groups: The information on the roles that different users have is stored in these tables. Some examples of the user profiles are administrator, content reviewer, editor and registered user. The Administrator may add new roles and define the operations according to the domain of functions specified in the Operations tables.

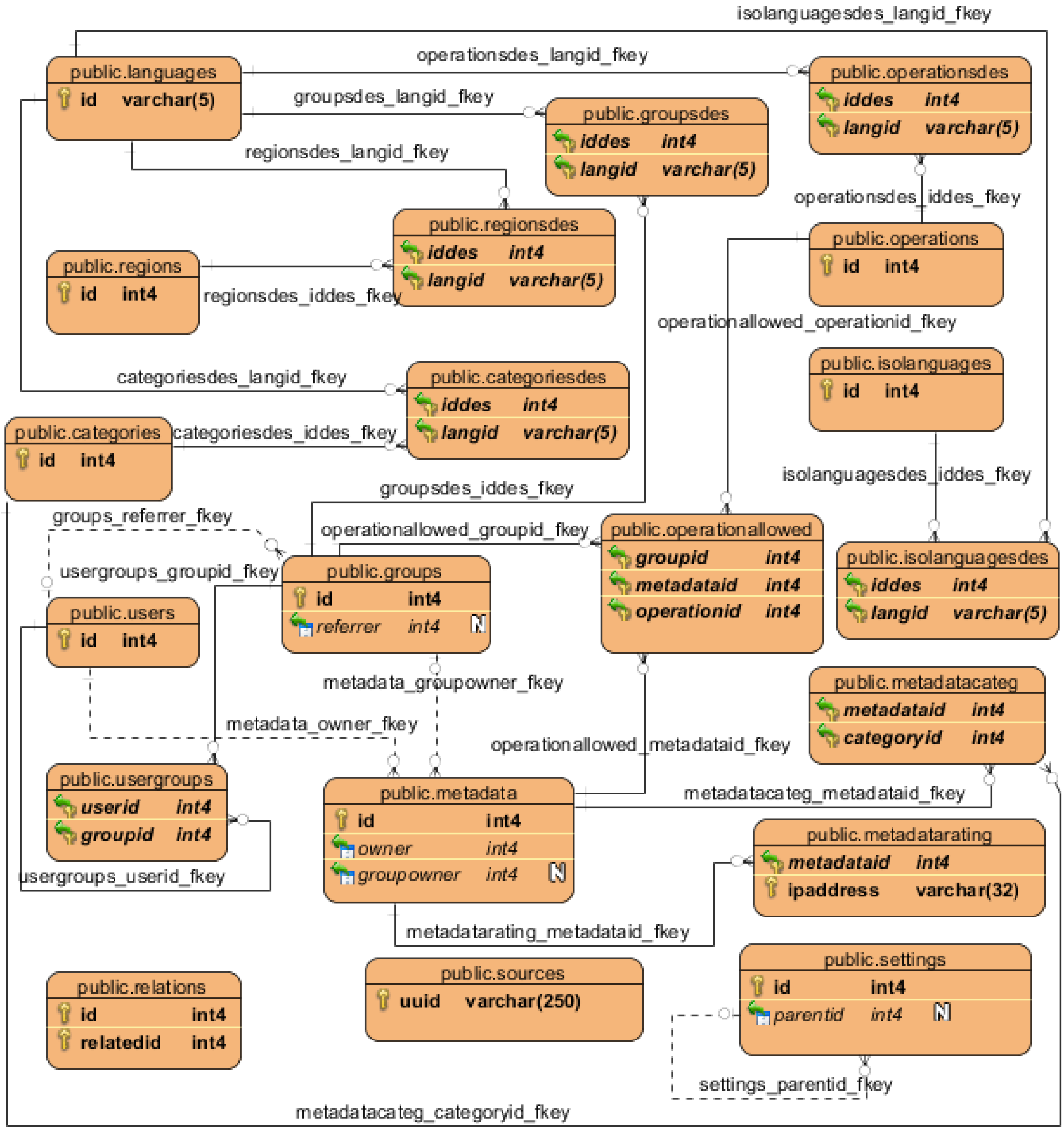


Figure 6.6: Metadata Database Entity Relationship Diagram

As was discussed in chapter 3, Geography, Themes and Context based metadata are powerful organising principles for large amounts of spatial data. The amount of information that an SDI can potentially host may mean that it might be difficult to find the datasets and services that one might be looking for. Therefore, the three organising principles are catered for in the following ways:

Geography: When users search for data, they have an option to search for it by region. This helps to streamline the search operations and increases the chances of returning more relevant search results based on geography. Geography therefore becomes a critical organising principle in an attempt to sieve out relevant information from the huge pool of data resources.

Themes: The themes are captured in key words during the process of creating metadata such that when metadata searches are done, metadata records that fall within the same theme are bundled up together.

Context: The context for the data is captured in the Abstract and Purpose of the metadata form. Describing the context within which the data was created will ensure that data is not misused. Figure 6.7 shows a snapshot from the form that is used to capture metadata using the ISO 19139 template on the Geonetwork based clearinghouse.

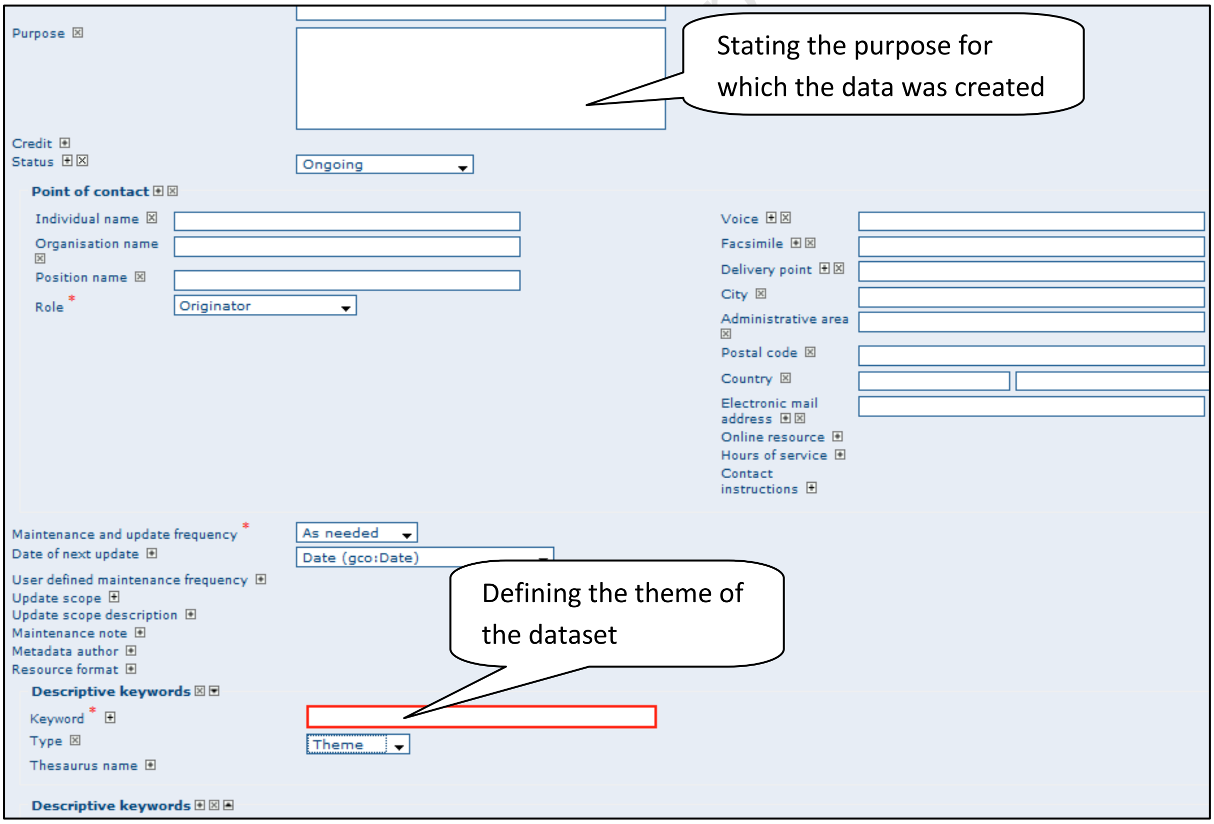


Figure 6.7: Defining the context and thematic scope for the data

6.5 The Client Domain

In Client/Server environments like the one that is presented in this research, there are broadly two types of computing architectures; thin and fat client architectures. The thin client architecture is where by all processing for the data is done on a central server whereas a client application simply requests for data and gets a result. The traditional web browser is an example of such a client where normally the client sends a request and gets a result. However with the advent of plug-ins (e.g. flash, Silverlight etc.) some server applications are now requiring some processing to be done on the client side. This reduces computing pressure on the server and therefore increases overall processing speed. However, such browsers still remain 'light weight' fat clients compared to desktop applications that use an access to some content on the web. An example is GIS software that consumes OGC based services and combines them with other datasets to perform GIS analysis.

6.5.1 Fat and Thin Clients

Examples of fat clients that consume web services and data exposed by the SDI are Globe clients. Most globes (e.g. NASA Worldwind and Google Earth) consume map services that are OGC compliant. This allows for services from the SDI to be integrated with local datasets and earth imagery for different purposes (see Appendix 10)

ArcGIS Explorer is a better client for ArcGIS anchored services than Google Earth is. Using ArcGIS Explorer, the end user can use more GIS related functions to perform analysis based on the services and data published through ArcGIS Server. The end user is also able to consume services from process models authored and published through the ArcGIS server. ArcGIS explorer can also be configured to become a CSW client and allow the user to search through CSW metadata catalogues and use interactive map resources associated with the metadata within the same platform

Other Fat Clients include desktop GIS software packages. These can consume map services that are authored through ArcGIS server to perform analysis or view data. These services may in turn be used by other organisations to integrate them with their local datasets and author maps that can also be published onto the web to be viewed by 3rd or 4th party clients.

The browser application (Appendix 6) is an example of a thin client application. In this instance, the user will simply need a web browser to consume the services that the application provided. The user will be limited to pre-structured functions that would have been embedded in the application. However, such applications serve the present the indicators perfectly well and do not need a GIS expert to use them. Decision makers will therefore be well equipped if they possess the simple skills that allow them to surf the web and will be able to see the different patterns that the applications will be trying to communicate.

The web applications are also organised according to the different indicator sets that they fall in. This makes the discovery process for the applications easier. To illustrate this, 8 indicator sets that were created by the CUO (see appendix 1) were used to create a web interface that provides access to web based maps that are indicator based (Figure 6.8). Clicking on an indicator category provides access to a webpage that links to the online mapping resources that are consumed by the user in the appropriate format that they would want from the given options. Users can typically consume these applications based on Geography. Although the web based maps may be accessed through the clearinghouse, a dedicated page that provides direct thematic access to them makes it easier to find relevant map services and clearly presents information on the key and pressing challenges that communities face. Such an approach to presenting data makes it easier for the non-GIS expert (but maybe key decision maker) to easily gain access to the macro trends that are presented through these indicator based maps.

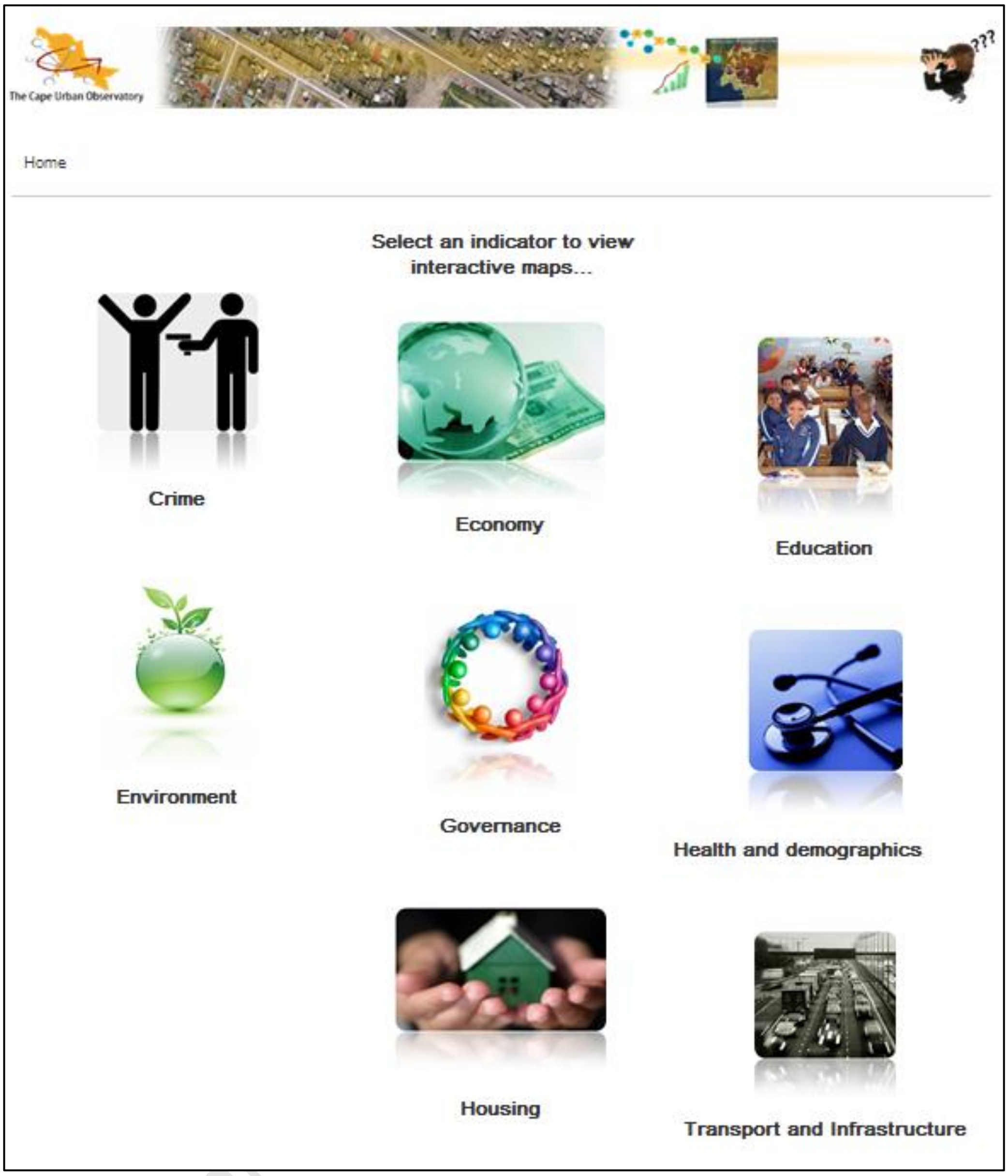


Figure 6.8: Interface providing access to indicator based map applications

6.5.2 Mobile Clients

OGC map services may also be consumed in GIS software that runs on mobile devices to be integrated with other location-based datasets like GPS waypoint, tracks, local data etc. This facility allows for individuals who need mapping resources from the field to access the SDI and consume GIS datasets as they work. Examples of mobile GIS software are ArcPad and gvSIG mobile. Mobile browser applications can be used to access the web mapping applications that are indicator based as in Figure 6.8.

6.6 Summary of Architecture

The architectural design and the specific technologies for implementation that have been described in this chapter are summarised in Figure 6.9 below. The four categories that were highlighted in the DFD in figure 6.1 are shown as well as what constitutes them. This rest of this section focuses on describing how the architectural design features echo the conceptual framework of a new perspective of SDIs that was described in chapter 3.

6.6.1 Data Organisation

The three organising principles that were described in Chapter three are reflected in the architecture in the following ways:

Geography

- Contents of the SDI can be searched for using regions that define certain geographic extents.
- The Web applications that showcase urban trends using indicators are also sorted according to location in the indicator browser interface

Context Based Metadata

- The context within which data was created is captured in the abstract and purpose sections of the metadata records. Data custodians or originators are best placed to know and therefore document this information.

Themes

- The thematic grouping within which a dataset fall is captured in the keywords section of the metadata record. This allows users to streamline searches to specific themes of interest.
- The browser interface (Figure 6.8) categorises the web-based map applications according to the indicator classes within which they fall.
- Related data is packaged in the same map service.

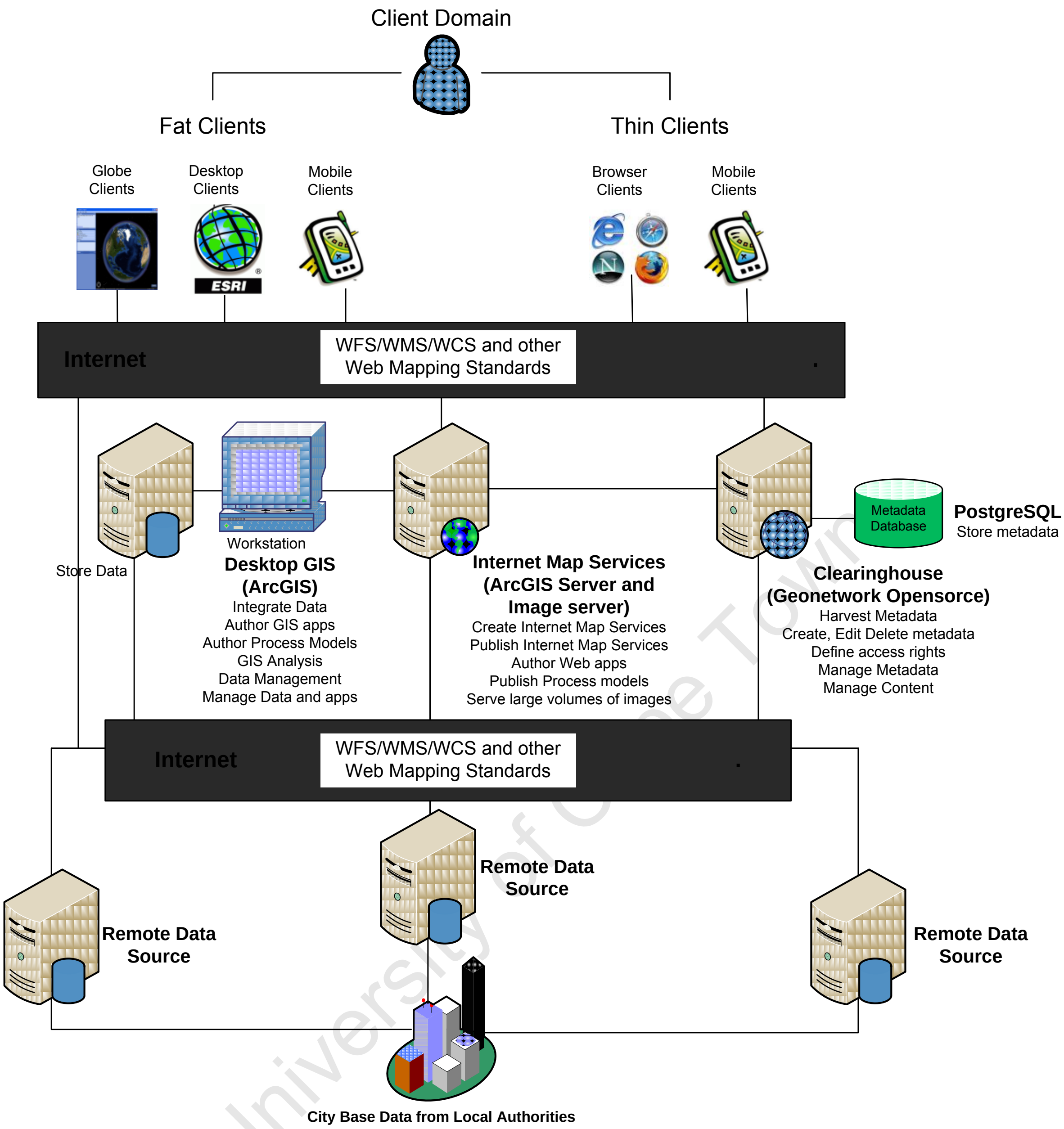


Figure 6.9: The proposed Architectural design

6.6.2 Principles of Wikinomics

The key principles of Wikinomics that were described in Chapter 3 are embedded in the architectural design in the following ways:

Openness

- The SDI is implemented using open standards and is publicly available on the web for use by virtually any interested party. This facilitates public access to information resources that are published through this platform and consequently can result in a bottom up interrogation of policy implementations.

- The open nature of the SDI will allow for limitless contributions from people/organisations that may have any meaningful data.

Peering

- The SDI facilitates discovery of datasets and data services. This in turn allows for better collaborative efforts in tackling urban problem as it turns to a platform where decision makers get to know of the work that is being done by others in the same domain of work.

Sharing

- There is a reduction in the duplication of data creation efforts, as users will get to better appreciate the wealth of data that other users have and therefore download these over the internet.

Acting Globally

- Having key urban trends available on the internet as public domain information allows for critic from a Global audience on urban problems and therefore allows for a broad range of ideas on how to adequately and properly respond to the urban crisis.

6.6.3 Community participation

While this architecture does not allow communities to participate in the generation of local datasets as VGI, they have an opportunity to be heard using community based indicators. Their input in generation relevant indicators means that the map based information products, which are a direct output from the SDI, communicate the major and pressing issues that communities face. Consequently, efforts to respond to the urban crisis are channelled to where they are required the most.

6.6.4 Spatial Analysis based on Indicators

Data Analysis is done in the GIS environment to display different spatially aware trends in the urban environment. The indicators play a key role in defining the type of analysis that should be done on the data. Process models like the one shown in figure 6.3 help for data to be analysed and report on the key issues contained in the indicators sets.

6.6.5 Map applications and Geoweb services

Maps are a powerful tool for generalizing urban trends. Having the map based web applications as part of the architecture facilitates the learning of trends the immediate use of data online to support decision-making processes. The ability to discover map services on the internet facilitates for the immediate use of data that may not be available for download for example private organisations that have value added data may avail it for use as a service and not as the raw dataset for business reasons.

The next chapter summarises the key findings from this research. It also highlights on the topics that future research efforts should address and suggests a few recommendations for successful implementation of the new SDI concept.

Chapter 7: Conclusion and Recommendations

7.0 Key Research findings

This research has succeeded in demonstrating the possibility of a new trajectory in the evolution of SDIs, which incorporates Evidence Based Practice (EBP). The features within EBP allow SDIs to be more useful platforms in responding to the urban crisis. This research has demonstrated that the incorporation of the following features into SDIs makes them more useful in responding to the urban crisis in developing nations:

a) **Community Participation**
As was shown in Chapter 2, community participation is crucial in responding to the challenges that societies face. The SDI concept constructed in this research allows community's voice to be heard through their contributions in the creation of urban indicators. This allows decision makers to focus on the pressing societal problems by responding to the key issues that would have been determined by the affected community groups.

b) **Urban Indicators**
Urban indicators are a critical tool in presenting evidence on different themes within the urban setup. The use of indicators in authoring relevant GIS services is demonstrated in this research in section 6.2.1 under Authoring applications and performing GIS analysis. These Indicator based Geoweb applications communicate effectively whether policy implementers are addressing the challenges that communities face. This approach results in a theme based focus for SDIs rather than the mainstream approach to developing them, which has largely focused on delivering access to all the available datasets without any inclination to the important ones that help to address the main societal problems.

c) **Access to Information products vs. access to raw data**
An important development in the SDI concept presented in this research is in its capability to serves as a gateway to information products based on spatial data rather that serving the raw spatial datasets only. This is shown through how this new SDI concept facilitates access to GIS services and applications as was explained in Chapter 6. This feature facilitates for the real use of data to reveal the important patterns that are hidden in it and therefore be used in decision-making applications. This key feature should facilitate for an increase in the use

of SDIs by decision makers and other ordinary users, as they are now a source for highlighting trends that both the decision makers and the community are interested in.

From an Architectural perspective, the following key concepts have been incorporated into the conceptual framework for SDIs:

d) **Redefined Geo-Information**
SDIs have traditionally been platforms that facilitate access to raw geospatial datasets that are used mainly in GIS applications. The definition of Geo-Information is changing to include any other form of information that has a location reference. The SDI concept that has been developed in this research as shown in section 3.4 allows for storage of diverse forms of Geo-Information through the content management capabilities of Geonetwork (Chapter 6). This allows for the use of various kinds of information resources in understanding the urban space e.g. photographs, videos, documents etc.

e) **Three organising principles**
Considering that the growth of the information repository that the SDI is a gateway to will be amplified in this new concept because of the new data types that can be classified as "Geoinformation", data will need to be organised in a way that will ensure that searches will return meaningful results. Data and information resources are therefore organised according to Geography and thematic scope (See sections 3.4 and 6.4.3). Context based metadata also gives meaning to data and therefore facilitates for its appropriate use.

These research findings show how Evidence Based Practice can be incorporated into SDIs and therefore cement the assertion made by the hypothesis in chapter 1 that "Incorporating Evidence Based Practice into Spatial Data Infrastructures using urban indicators will make them more useful in responding to the urban crisis."

7.1 Other Research findings

Besides the contributions made to the conceptual framework of SDIs, the following issues have been raised through this research and can form the basis for future SDI research and implementation strategies.

a) **Design and behavioural science**
The Information Systems Research Framework, which looks at the Behavioural and Design Science paradigms, has been used as the philosophical framework to ensure the relevance and utility of the SDI design. This framework may also be used in other SDI related research

and implementations. SDI implementers in developing nations may particularly use this approach to clearly define the purpose and scope of SDIs. Defining scope and purpose for the SDI is situated in the behavioural science domain of IS research and can be useful to more effectively communicate the reason for SDIs to attract attention from government as SDIs are not a high priority for most governments in developing nations (Makanga and Smit, 2008).

b) Use of the Data Flow Diagram (DFD)

The use of the DFD to model the information viewpoint of the SDI has been demonstrated in this research. It allowed for clearly visualising the information flows and consequently the functional decomposition for the SDI. While the Class and Use case diagram have been used to model SDIs from this viewpoint (Hjelmager et al., 2008), the DFD allows for easier conceptualisation of the information flows and processes within an SDI. The thesis has also demonstrated the applicability of other Software Processes in modelling SDIs

The research has managed to show that the above-mentioned features can be accommodated in SDIs to correct the main problems (below) that were highlighted in chapter 1:

- Lack mechanisms that facilitate real use of data in decision making processes
- Lack mechanisms that facilitate learning of trends across different themes in the Urban environment
- Lack mechanisms to facilitate community participation in responding to the problems that they face

7.2 Future research

While community participation plays a critical role in governing the urban space, the SDI concept developed in this research does not fully allow communities to participate in responding to the challenges that they face. Although the model constructed allows communities to be heard through their participation in the creation of relevant urban indicators, they have been excluded from the data creation process. This is due to the assumption that the target communities do not have the capacity to generate their own data as Volunteered Geographic Information. There is a potential role for Public Participation Geographic Information Systems (PPGIS) and Participatory Geographic Information Systems (PGIS) in the data creation processes, which have not been fully explored. Future research efforts may look into the possible role of these disciplines in facilitating a more community driven approach to the creation of data, considering that meaningful datasets on informal settlements may not be available from formal sources that participate actively in the SDI.

There is a need for increased participation of the Non government/private sectors in contributing data elements into the SDI. Future research efforts may be channelled towards understanding the business implications to such organisation for participating in SDI initiatives. While it may appear suicidal for a company that uses its value added spatial data as a competitive edge in business, to give it away freely, SDIs may be presenting an opportunity to advertise data through relevant information services that are discoverable through the clearinghouse. Therefore, it may turn out to be a win/win scenario for both companies and sustainable development initiatives.

References

AGARWAL, B. B., TAYAL, S. P. & GUPTA, M. (2009) *Software Engineering & Testing: An Introduction*, Jones and Bartlett Publishers

AGARWAL, R. & LUCAS, H. C. (2005) The information systems identity crisis: focusing on high-visability and high-impact research. *MIS Quarterly,* 29**,** 381-398.

AKINGBADE, A., SOUMAH, M., KUFONIYI, O. & MBAYE, M. (2004) EVALUATION OF FUNDAMENTAL GEOSPATIAL DATA SETS IN WEST AFRICA IN AFRICA, E. (Ed.) *Fundamental Geospatial Data Sets for Africa* Pretoria, EIS Africa.

AMBLER, S. W. & CONSTANTINE, L. L. (2000) *The unified process elaboration phase: best practices in implementing the UP*, Publishers group west.

ANDERSON, P. (2007) What is Web 2.0? Ideas, technologies and implications for education. *JISC Technology & Standards Watch.* JISC.

ANZLIC. (2009) Spatial Metadata. Australian, New ZeaLand Information Council (http://www.anzlic.org.au/infrastructure_metadata.html: ***Accessed 29/03/2010***)

AROYO, D. L. M. & DICHEVA, D. (2000) Domain and user knowledge in a web-based courseware engineering course, knowlegde-based software engineering. *In T. Hruska, M. Hashimoto (Eds.) Joint Conference knowledge-Based Software Engineering 2000.*

BASS, L., CLEMENTS, P. & KAZMAN, R. (2003) *Software Architecture in Practice*, Addison-Wesley.

BEDI, T., COUDOUEL, A. & SIMLER, K. (2007) *More than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions* Washington DC, The World Bank.

BEDNARZ, R. S. & BEDNARZ, S. W. (2008) The Importance of Spatial Thinking in an Uncertain World. *Geospatial Technologies and Homeland Security.*

BELADY, L. (1981) Foreword. IN PETERS, L. J. (Ed.) *Software Design: Methods and Techniques.* Yourdon Press.

BOEHM, B. W. & TURNE, R. (2003) *Balancing agility and discipline: a guide for the perplexed*, Pearson Education.

BROWN, A. L. (1992) Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *Journal of the Learning Sciences,* Vol 2.

BUDHATHOKI, N., BRUCE, B. & NEDOVIC-BUDIC, Z. (2008) Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal,* 72**,** 149-160.

CARLSSON, S. (2006) Towards an Information Systems Design Research Framework: A Critical Realist Perspective *DESRIST 2006. .* Claremont, CA.

CCT. (2009) City statistics and population census City of Cape Town (http://www.capetown.gov.za/en/stats/Pages/Indicators.aspx ***Accessed 29/03/2010***)

CEBP. (2009) Increasing Government Effectiveness Through Rigorous Evidence About "What works". Coalition for Evidence-Based Policy (http://coalition4evidence.org/wordpress/: ***Accessed 29/03/2010***)

CHILDS, C. (2009) The Top Nine Reasons to Use a File Geodatabase A scalable and speedy choice for single users or small groups *ESRI Education Services.*

CLARKE, D. (2009) Mapping Africa for Africa: Lack of political will, financial capability and skilled manpower - African nations are battling out all odds to unite geodetically. *The Geospatial Magazine.* Cape Town, GIS Development.

COCKBURN, A. (2002) *Agile software development*.

COLEMAN, D., YOLA, G. & LABONTE, J. (2009) Volunteered Geographic Information: The Nature and Motivation of Produsers. *International Journal of Spatial Data Infrastructures Research,* 4**,** 332 - 358.

COOPER, A., MOELLERING, H., DELGADO, T., DÜREN, U., HJELMAGER, J., HUET, M., RAPANT, P., RAJABIFARD, A., LAURENT, D., IWANIAK, A., ABAD, P. & MARTYNENKO, A. (2007) An initial model of the computation viewpoint for a spatial data infrastructure. *Cartography for everyone and for you, 23rd International Cartographic Conference.* Moscow, Russia.

CRAGLIA, M. & JOHNSTON, A. (2005) Assessing the Impacts of Spatial Data Infrastructures: Methods and Gaps. *7th AGILE Conference on Geographic Information Science.* Heraklion, Greece.

CROMPVOETS, J. & BREGT, A. (2003) World Status of National Spatial Data Clearinghouses. *Journal of the Urban and Regional Information Systems Association,* 15**,** 43-50.

CRUTCHER, M. & ZOOK, M. (2009) Placemarks and waterlines: Racialized cyberscapes in post-Katrina Google Earth. *Geoforum,* 40**,** 523-534.

CUO (2009) Cape Urban Observatory Report – July 2008 to July 2009. Cape Town, University of Cape Town.

DCMI. (2007) Dublin Core Metadata Initiative Tools Glossary. (http://dublincore.org/groups/tools/glossary.shtml: ***Accessed 29/03/2010***)

DENTON, L. F., DORAN, M. V. & MCKINNEY, D. (2002) Integrated Use of Bloom and Maslow for Instructional Success in Technical and Scientific Fields *Proceedings of the 2002 American Society for Engineering Education Annual Conference & Exposition* Montréal, Quebec, Canada American Society for Engineering Education

DLA (2003) Spatial Data Infrastructure Act, 2003. *54.* South Africa, Government of South Africa, Government Gazette

DOBSON, P. J. (2001) The Philosophy of Critical Realism—An Opportunity for Information Systems Research. *Information Systems Frontiers,* 3**,** 199-210.

DOOLEY, K. J. (1997) A Complex Adaptive Systems Model of Organization Change. *Nonlinear Dynamics, Psychology, and Life Sciences,* 1**,** 69-97.

DUTTON, W. H. & PELTU, M. (1996) *Information and communication technologies: visions and realities*, Oxford University Press.

EASTERLY, W. (2009) How the Millennium Development Goals are Unfair to Africa. *World Development,* 37**,** 26-35.

EEU (2009) An Integrated Monitoring, Learning and Reporting Framework for the DBSA Sustainable Communities Programme Pilot Projects. Cape Town, University of Cape Town.

ELWOOD, S. (2008) Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice. *GeoJournal,* 72**,** 133-135.

ELWOOD, S. (2009a) Geographic information science: emerging research on the societal implications of the geospatial web. *Progress in Human Geography***,** 0309132509340711.

ELWOOD, S. (2009b) Geographic Information Science: new geovisualization technologies -- emerging questions and linkages with GIScience research. *Progress in Human Geography,* 33**,** 256-263.

ESRI. (2010) ArcGIS Desktop Help 9.3, including 9.3.1 ESRI (http://support.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=welcome: ***Accessed 29/03/2010***)

FAROOQUI, K., LOGRIPPO, L. & DE MEER, J. (1995) The ISO Reference Model for Open Distributed Processing: an introduction. *Computer Networks and ISDN Systems,* 27**,** 1215-1229.

FERNÁNDEZ, D. & IGLESIAS, R. C. (2009) Spatial Data Infrastructure: From national to local level. *Global Spatial Data Infrastructure (GSDI) 11.* Rotterdam, Netherlands, GSDI.

FERNÁNDEZ, T. D. & FERNÁNDEZ , J. L. C. (2008) Towards Semantic Spatial Data Infrastructures: A framework for sustainable development. IN GSDI (Ed.) *GSDI10* St. Augustine, Trinidad, GSDI.

FOROUZAN, B. A. & MOSHARRAF, F. (2007) *Foundations of Computer Science*, Gaynor, Redvers-Mutton.

FUTRELL, R. T., SHAFER, D. F. & SHAFER, L. (2002) *Quality software project management*, Prentice Hall.

GARTNER. (2008) Gartner Says Cloud Computing Will Be As Influential As E-business. (http://www.gartner.com/it/page.jsp?id=707508: ***Accessed 29/03/2010***)

GEORGIADOU, Y., HARVEY, F. & MISCIONE, G. (2009) A Bigger Picture: Information Systems and Spatial Data Infrastructure Research Perspectives IN GSDI (Ed.) *GSDI11.* Rotterdam, Netherlands, GSDI.

GOODCHILD, M. (2007) Citizens as sensors: the world of volunteered geography. *GeoJournal,* 69**,** 211-221.

GOODCHILD, M. F., FU, P. & RICH, P. (2007) Sharing Geographic Information: An Assessment of the Geospatial One-Stop. *Annals of the Association of American Geographers,* 97**,** 250-266.

GORE, A. (1998) The Digital Earth: Understanding our planet in the 21st Century. *California Science Center, Los Angeles, California.*

GROOT, R. (1997) Spatial data infrastructure (SDI) for sustainable land management. *TC Journal* 3/4**,** 287 - 294.

GSDI (2004) *Developing Spatial Data Infrastructures: The GSDI Cookbook*, The Global Spatial Data infrastructure Association.

HAMANN, R., BOULOGNE, F., MARA, I. & CHANDAKA, S. (2008) A National Model for Innovation and Sustainable Development: The Grabouw Sustainable Development Initiative. *Sustainable Cities 2008.* Johannesburg, South African Cities Network.

HAMILTON, K. (2003) The four levels of Evidence Based Practice. *The American Institute of Architects*.

HARVEY, F. (2003) Developing geographic information infrastructures for local government: the role of trust. *The Canadian Geographer,* 47**,** 28(9).

HEVNER, A. R., MARCH, S. T., JINSOO, P. & RAM, S. (2004) Design Science in Information Systems Research. *MIS Quarterly.* MIS Quarterly & The Society for Information Management.

HIRSCHHEIM, R. & KLEIN, H. K. (2003) Crisis in the IS field? A critical reflection on the state of

the discipline. *Journal of the Association for Information Systems,* 4**,** 237-293. .

HJELMAGER, J., MOELLERING , H., COOPER, A., DELGADO, T., RAJABIFARD, A., RAPANT, P., DANKO, D., HUET, M., LAURENT, D., AALDERS, H., IWANIAKV, A., ABAD, P., DUREN, U. & MARTYNENKO, A. (2008) An Initial Formal Model for Spatial Data Infrastructures. *International Journal of Geographical InformationScience,* 22**,** 1295-1309.

HOLDEN, M. (2009) Community Interests and Indicator System Success. *Social Indicators Research,* 92**,** 429-448.

HSRG (2005) Breaking new ground: A comprehensive plan for the development of Sustainable Human Settlements. IN GROUP, H. S. R. (Ed.). Cape town.

HUDSON, A. & CROOKS, A. (2008) The Renaissance of Geographic Information: Neogeography, Gaming and Second Life. *Working Papers Series.* London, University College London.

IMF & WB (2005) Global Monitoring Report. IN BANK, I. M. F. A. T. W. (Ed.).

IRODS. (2010) Glossary. (https://www.irods.org/index.php/iRODS_Glossary ***Accessed 29/03/2010***)

JONES, M. T. (2007) Google's Geospatial Organizing Principle. *IEEE Computer Graphics and Applications,* 27**,** 8-13.

KAPOR, M. (1991) A software design manifesto. *Dr. Dobb's J.,* 16**,** 62-67.

KITTUR, A. & KRAUT, R. E. (2008) Harnessing the wisdom of crowds in wikipedia: quality through coordination. *Proceedings of the ACM 2008 conference on Computer supported cooperative work.* San Diego, CA, USA, ACM.

KRISHNAMURTHY, S. (2002) Cave or Community?: An Empirical Examination of 100 Mature Open Source Projects. *SSRN eLibrary*.

KROLL, P. & KRUCHTEN, P. (2003) *The rational unified process made easy: a practitioner's guide to the RUP*, Pearson Education.

LAKE, R. & FARLEY, J. (2007) Infrastructure for the Geospatial Web. IN SCHARL, A. & TOCHTERMANN, K. (Eds.) *The Geospatial Web: How Geobrowsers, Social Software and the Weeb 2.0 are Shaping the Network Society.* Springer.

LANCE, K. (2004) Spatial Data Infrastructure in Africa. *Nigerian Journal of Surveying and Geoinformatics*.

LANCE, K. & BASSOLÉ, A. (2006) SDI and national information and communication infrastructure (NICI) integration in Africa. *Information Technology for Development,* 12**,** 333-338.

LARMAN, C. (2003) *Agile and Iterative Development: A Manager's Guide*, Pearson Education.

LIU, T. & TANG, C. (1991) Semantic specification and verification of data flow diagrams. *Journal of Computer Science and Technology,* 6**,** 21-31.

LUDWIG VON, B. (1972) The History and Status of General Systems Theory. *The Academy of Management Journal,* 15**,** 407-426.

MAGUIRE, D. J. & LONGLEY, P. A. (2005) The emergence of geoportals and their role in spatial data infrastructures. *Computers, Environment and Urban Systems,* 29**,** 3-14.

MAKANGA, P. T. & SMIT, J. (2008) A review of Spatial Data Infrastructure implementation in Africa. *International conference on Free and Open Source Software for geospatial.* Cape Town, South Africa.

MARCH, S. T. & SMITH, G. F. (1995) Design and natural science research on information technology. *Decision Support Systems,* 15**,** 251-266.

MARKUS, M. L., MAJCHRZAK, A. & GASSER, L. (2002) A Design Theory for Systems That Support Emergent Knowledge Processes. Management Information Systems Research Center, University of Minnesota.

MARTIN, J. (1991) *Rapid application development*.

MASSER, I. (2005) GIS Worlds: Creating Spatial Data Infrastructures. *ESRI Press.* Redlands.

MAVIMA, R. (2004) DETERMINING THE FUNDAMENTAL GEOSPATIAL DATA SETS FOR SADC COUNTRIES IN AFRICA, E. (Ed.) *Fundamental Geospatial Data Sets for Africa* EIS Africa.

MCEVILY, B., PERRONE, V. & ZAHEER, A. (2003) Trust as an Organizing Principle. *Organization Science,* 14**,** 91-103.

MILNE, D., MEDELYAN, O. & WITTEN, I. H. (2006) Mining Domain-Specific Thesauri from Wikipedia: A Case Study. *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence.* IEEE Computer Society.

NEDOVIC-BUDIC, Z., FEENEY, M.-E. F., RAJABIFARD, A. & WILLIAMSON, I. (2004) Are SDIs serving the needs of local planning? Case study of Victoria, Australia and Illinois, USA. *Computers, Environment and Urban Systems,* 28**,** 329-351.

NEDOVIC-BUDIC, Z., PINTO, J. K. & BUDHATHOKI, N. R. (2009) SDI effectiveness from a user perspective. IN CROMPVOETS, J., RAJABIFARD, A., VAN LOENEN, B. & DELGADO, F. T. (Eds.) *A Multiview Framework to access Spatial Data Infrastructures.* Melbourne, Wageningen, Space for Geo-Information, Centre for SDI and Land Administration.

NOGUEIRA, J. C., JONES, C. & LUQI (2000) Surfing the Edge of Chaos: Applications to Software Engineering. *Department of Defence Command and Control Research and Technology Symposium.* Montery, CA, DOD.

NSFWUS, N. S. F. W. O. U. S. (2000) Towards a comprehensive Geographical perspective on Urban Sustainability: Final Report of the 1998 National Science Foundation Workshop on Urban Sustainability.

NUNAMAKER, J. F., JR. & CHEN, M. (1990) Systems development in information systems research. *System Sciences, 1990., Proceedings of the Twenty-Third Annual Hawaii International Conference on.*

OGC (2007) OpenGIS Catalogue Services Specification IN NEBERT, D., WHITESIDE, A. & VRETANOS, P. (Eds.). Open Geospatial Consortium.

OGC. (2009) Open Geospatial Consortium homepage. (http://www.opengeospatial.org/: ***Accessed 29/03/2010***)

OMG. (2009) Object Management Group: Unified Modeling Language (UML), Version 2.0. (http://www.omg.org/technology/documents/formal/uml.htm: ***Accessed 29/03/2010***)

PIETERSE, E. (2008) *City Futures: Confronting the crisis of urban development,* Cape Town, South Africa, UCT Press.

PIETERSE, E. (2009) African Cities: Grasping the unknowable. *African centre for Cities.* Cape Town, University of Cape town.

PRESSMAN, R. (2009) *Software Engineering: A Practitioner's Approach*, McGraw-Hill

RAYMOND, E. S. (2001) *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, O'Reilly \& Associates, Inc.

RAYMOND, K. & ARMSTRONG, L. (1995) Open distributed processing: experiences with distributed environments. *International Federation for Information Processing*.

RILEY, L., KO, A., UNGER, A. & REIS, M. (2007) Slum health: Diseases of neglected populations. *BMC International Health and Human Rights,* 7**,** 2.

RIVLIN, G. (2003) Leader of the Free World: How Linus Torvalds became benevolent dictator of Planet Linux, the biggest collaborative project in history. *Wired.*

RVU (2006a) Counting on Vancouver: Our view of the region. *Inaugural Report of the Regional Vancouver Urban Observatory.* Vancouver, Urban Studies, Simon Fraser University.

RVU (2006b) Expanding our view: Sustainability indicator results of the RVU study group process. Vancouver, Urban Studies, Simon Fraser University.

RVU. (2009) Regional Vancouver Urban Observatory. (http://www.rvu.ca/: ***Accessed 29/03/2010***)

SCHON, D. A. (1983) *The Reflective Practitioner: How Professionals Think in Action,* New York, Basic Books.

SCHUURMAN, N. (2009a) Metadata as a site for imbuing GIS with qualitative information IN COPE, M. & ELWOOD, S. (Eds.) *Qualitative GIS: A Mixed methods approach.* SAGE.

SCHUURMAN, N. (2009b) The new Brave New World: geography, GIS, and the emergence of ubiquitous mapping and data. *Environment and Planning D: Society and Space,* 27**,** 571-572.

SCHUURMAN, N., LEIGHT, M. & BERUBE, M. (2008) A Web-based graphical user interface for evidence-based decision making for health care allocations in rural areas. *International Journal of Health Geographics,* 7**,** 49.

SCHUURMAN, N. & LESZCZYNSKI, A. (2006) Ontology-Based Metadata. *Transactions in GIS.* Blackwell Publishing Limited.

SIMON, H. A. (1996) *The Sciences of the Artificial* The MIT Press.

SMIT, J., MAKANGA, P., LANCE, K., DE VRIES, W. & (2009) Exploring relationships between municipal and provincial government SDI implementers in South Africa *Global Spatial Data Infrastructure (GSDI) 11 World Conference.* Rotterdam Netherlands.

SMIT, W. (2008) Cape Urban Observatory Inception Workshop. Cape Town, University of Cape Town.

SSMIC. (2009) Sault Ste Marie Innovation Centre Awards and Achievements. (http://www.ssmic.com/index.cfm?fuseaction=content&PageID=1048&PageCategory=43: ***Accessed 29/03/2010***)

STONOR, T. & STUTZ, C. (2004) Towards Evidence Based Urban Design. Space Syntax.

TAPSCOTT, D. & WILLIAMS, A. (2006) *Wikinomics: How Mass Collaboration Changes Everything*, Portfolio Hardcover.

TSUI, F. F. & KARAM, O. (2006) *Essentials of software engineering*, Jones and Bartlett Publishers.

TURK, D., FRANCE, R. & RUMPE, B. (2004) Assumptions Underlying Agile Software Development Processes. *Journal of Database Management* 1-38.

TUROK, I. & PARNELL, S. (2009) Reshaping Cities, Rebuilding Nations: The Role of National Urban Policies. *Urban Forum,* 20**,** 157-174.

UNECA (2001) The Future Orientation of Geoinformation Activities in Africa: A Position Paper. IN DIVISION, D. I. S. (Ed.) *Second Meeting of the Committee on Development Information.* Addis Ababa, Ethiopia, United Nations Economic Commission for Africa.

UNGA, U. N. G. A. (2005) Resolution adopted by the General Assembly: 2005 World Summit Outcome. New York.

UNHABITAT (2003) The Challenge of Slums: Global report on human settlements.

UNHABITAT. (2009a) Global urban Observatory. United Nations Human Settlements Programme (http://ww2.unhabitat.org/programmes/guo/: ***Accessed 29/03/2010***)

UNHABITAT. (2009b) UN habitat for a better urban future. United Nations Human Settlements Programme (http://www.unhabitat.org/categories.asp?catid=1: ***Accessed 29/03/2010***)

UNHABITAT. (2009c) UN to Review Progress on the Millennium Development Goals at High-level Meeting in September 2010. United Nations Human Settlements Programme (http://www.un.org/millenniumgoals/sept_2010_more.shtml: ***Accessed 29/03/2010***)

VAN ASSCH, J. & REYNAERT, H. (2009) Harnessing Knowledge through Urban Indicators: Measuring quality of life in Flemish Cities. *City Futures in a Globalising World.* Madrid, Spain.

VAN OORT, P. A. J., KUYPER, M. C., BREGT, A. K. & CROMPVOETS, J. (2009) Geoportals: An Internet Marketing Perspective. *Data Science Journal,* 8**,** 162-181.

VON HAGEN, C. (2007) Towards a Spatial Data Infrastructure for Somalia using Open Source Standards. *Journal of Spatial Science,* 52**,** 157-169.

WALLS, J. G., WIDMEYER, G. R. & EL SAWY, O. A. (1992) Building an Information System Design Theory for Vigilant EIS. *INFORMATION SYSTEMS RESEARCH,* 3**,** 36-59.

WANG, Y. & KING, G. (2000) *Software engineering processes: principles and applications*, CRC Press.

WCED (1987) Our Common future. IN DEVELOPMENT, W. C. O. E. A. (Ed.). Oxford, England.

WEILL, P., BROADBENT, M., WILLCOCKS, L. & LESTER, S. (1999) Four views of IT infrastructures. In Beyond the IT Productivity Paradox. *Wiley Series in INFORMATION SYSTEMS RESEARCH***,** 335-362.

WHITEHEAD, A. N. (1916) The Organisation of Thought. *Proceedings of the Aristotelian Society,* 17**,** 58-76.

WILLIAMSON, I. (2003) SDIs setting the scene. IN WILLIAMSON, I., RAJABIFARD, A. & FEENEY, E. (Eds.) *Developing Spatial Data Infrastructures: From Concept to reality.* CRC Press.

WORLDMAYOR. (2009) World Mayor, caring for amazing cities (http://www.worldmayor.com/ ***Accessed 29/03/2010***)

WYNN, A. (2004) What Is Architecture? An Explanation for IT Professionals. *Workshop on ODP for Enterprise Computing (WODPEC 2004)* Monterey, California.

YIHONG, D. (2007) A simple picture of Web evolution. *Thinking Space.*

Appendices

Appendix 1

Indicator	Likely data source	Notes
1) Population 2) Birth rate 3) Death rate 4) Under 5 mortality 5) HIV positive rate 6) TB rate 7) Life expectancy 8) Number of households	Municipal data / StatsSA Health department / local clinics Health department / local clinics Health department / local clinics Health department / local clinics Health department / local clinics Health department / local clinics Municipal data or if necessary inferred from aerial photos	Indicators 1-8 are basic demographic and health indicators that provide an important baseline. They are common in most indicator frameworks. Furthermore, including indicators 2 and 3 allows for a better understanding of migration patterns, which is a vital issue in places like Grabouw.
9) Geographic Value Added per capita 10) Total property tax revenue per capita 11) Unemployment rate 12) Household gini co-efficient 13) Number of registered businesses	Uncertain Municipal data Labour Force Survey Uncertain CIPRO	Indicators 9 to 13 are generally considered in terms of economic conditions. The calculation of GVA at the local community level may be problematic, as is household gini co-efficient.
14) % households without formal shelter 15) % households without water on site (i.e. in dwelling or yard) 16) % households with VIP toilets or less 17) % households not using electricity for lighting 18) % households without weekly refuse collection	Municipal data or if necessary inferred from aerial photos Municipal data Municipal data Municipal data Municipal data	Indicators 14 to 18 are about access to housing and municipal services – again this is common to most indicator models.

19) Water quality in streams 20) Air quality 21) GHG emissions 22) Annual tonnes of waste disposed to landfills per capita 23) % solid waste recycled 24) Total energy use 25) Consumption of electricity per capita 26) Proportion of open space / natural areas to total land area 27) Proportion of agricultural land to total land area	Municipal data plus possibly community-based monitoring Municipal data plus possibly community-based monitoring Uncertain Municipal data Municipal data Municipal data Municipal data Municipal data, aerial photography Municipal data, Dept of Agriculture, aerial photography	Indicators 19 to 27 are broadly about environmental quality and resource use.
28) Crime rate 29) Homicides	Police Police	Crime is a key concern for most South African communities. Indicator 27 is a general, inclusive measure, while indicator 28 focuses on homicide, which is commonly a focus for international comparison.
30) Pedestrian or cycle traffic accidents 31) Average travel time to work opportunities	Police Uncertain	Indicators 30 and 31 are broadly about transportation. Furthermore, indicator 31 also provides a measure of the 'compactness' of the settlement
32) School leaving certificate pass rate 33) Matriculants with mathematics passes	Dept of Education Dept of Education	Indicators 32 and 33 are focused on education
		Indicators 34 and 35 are about governance and democratic

34) % registered voter turnout 35) Local ward committees – published minutes of ward committee meetings on TWK website (Yes / No)	Independent Electoral Commission Municipal data	participation
Possible composite indicators?		
36) Human Development Index 37) *Ecological footprint*	This is a combination of life expectancy (covered by 7), education indicators (not covered by 32 and 33 above), and GDP (possibly covered by 9) This is a broad measure of natural resource use – uncertain whether this would be covered by indicators 19 to 27	Indicators 36 and 37 are twp possible composite indicators to be used. The HDI is proposed because this is explicitly called for in the DBSA Programme outline. Furthermore, the City Development Index is probably inappropriate at the local community level. However, both the HDI and the Ecological Footprint entail a number of methodological requirements, which are not covered by the indicators proposed above. A decision will hence need to be made whether to aim for such composite indicators.

Appendix 2

The regional Vancouver Urban Observatory indicators (RVU, 2006a)

	INDICATOR	TREND
Sustainable Mobility	Percent of children who walk or cycle to school	Negative
	Percent of household income spent on transportation within the region	No Change
	Level of agreement with the statement: "I live in a neighbourhood in which I can walk to work and to meet my personal needs."	?
Overcoming Poverty	Availability of emergency services (food, beds, detox) as a proportion of demonstrated need for these services.	Negative
	Percent of households in the region consistently able to meet their basic needs.	Negative
	Quality of media coverage of poverty as a regional sustainability issue.	?
Economic Developmnt	Local Index for a Vital Economy (LIVE).	?
	Number of land use bylaws passed by municipalities that contravene the vision and principles outlined in the Livable Region Strategic Plan (LRSP).	?
	Efficient resource use in local municipalities (oil equivalent per capita).	Positive
Governance	Percent of Vancouver region residents who feel they have opportunity to voice thoughts on major community decisions.	?
	The success of a sample of attempts by municipalities to reach diverse groups of the public in strategic work toward sustainability.	?
	Percent of Vancouver residents who are aware of the Ecological Footprint and understand their contribution to it.	?
Building Community	The number and location of "third spaces" around the region.	?
	The number of institutions, organizations and businesses which engage with the public on a regular basis.	?
	The number of public consultations which achieved "true dialogue."	?
Natural Environmt	Total regional waste produced per capita.	Positive
	Percent of citizens who participate in environmental stewardship activities.	?
	Percent of development on greenfield vs. brownfield land.	?
Food Systems	The gap between the percent of income spent by each of 4 income groups needed to purchase a "healthy" food basket.	Negative
	Ratio of all land available for growing food to the potentially productive land in both urban and rural areas.	?
	Ratio of food items produced and consumed within the region those imported and consumed within the region for selected foods.	?
Arts and Culture	Quantity and quality of opportunities for cultural activity, as represented by an annually updated cultural events matrix.	?
	Percent of individuals who feel that they have adequate access, freedom and time for cultural and artistic activity.	?
	Ratio of dollars spent promoting multicultural awareness and artistic work to the dollars these activities contribute to the region.	?

Appendix 3

Grabouw Fieldwork Questionnaire April 2009. These interviews were done in TWK with

2) Municipal Officials
3) Grabouw Town Council
4) Stakeholders in the GrSDI

Aim: To understand the GrSDI.
NB: SDI stands for Grabouw Sustainable Development Initiative

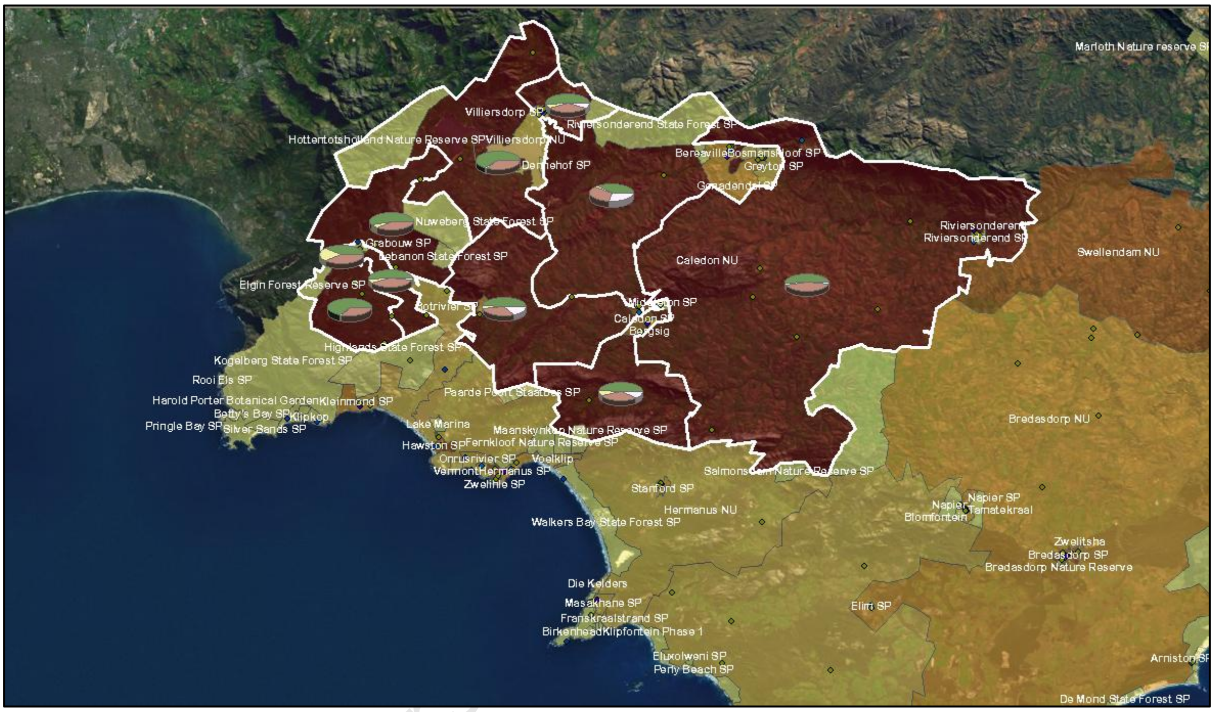
1. What happened in 2008? (main events and activities, goals, impressions)
2. Expectation, hopes for 2008 and to what extent were they met?
3. How were you involved /what role did you play?
4. What have been the strengths and weaknesses of the SDI in 2008?
5. Is there still a compelling/shared vision for the SDI?
6. Was there effective communication of what the SDI was doing/trying to do?
7. What were the key governance and management structures in the implementation process?
8. How successful has the forum been in the SDI (what has the role of the forum in implementation of the SDI)?
9. How well is progress being monitored and reported on/communicated?
10. What are the key issues to be monitored/ reported on? (show the list of generic indicators we have prepared)
11. Who should monitor or how should monitoring take place in the future?

- Thank the interviewee for their time

Appendix 4

Integrating data from diverse sources

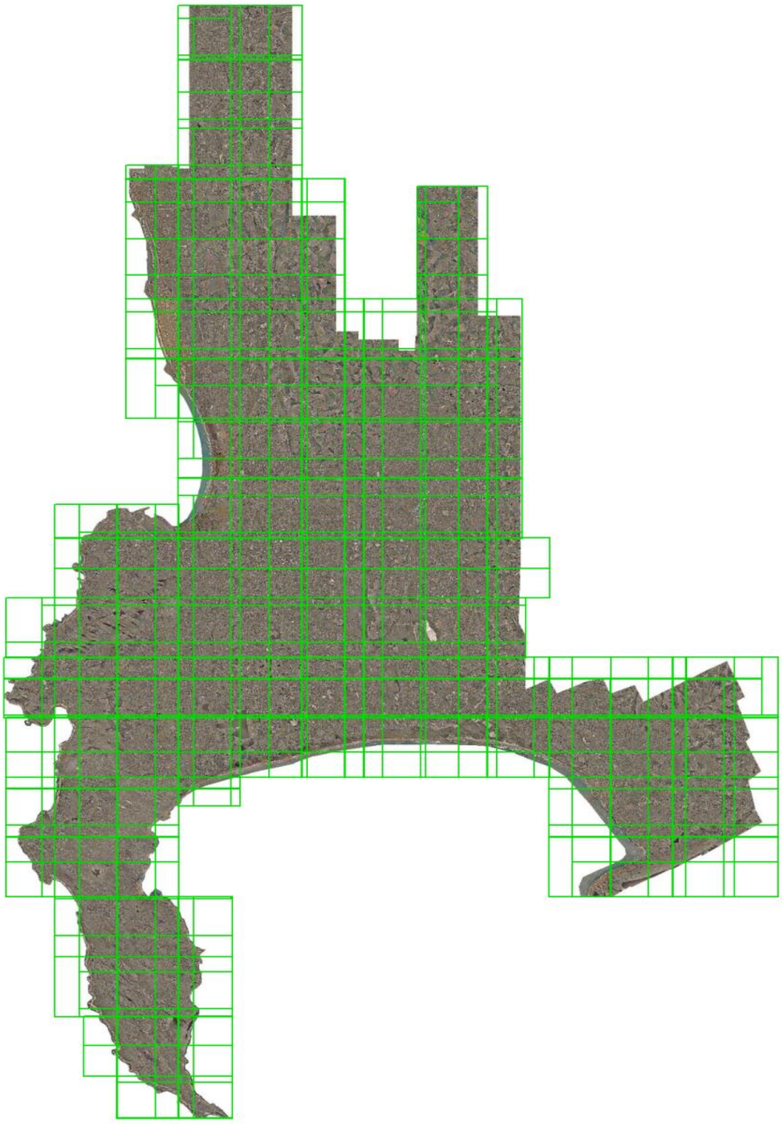
In this example data is integrated from different sources in an ArcGIS environment. The background image data is from an ESRI image service that provides access high resolution image data for most cities through the internet while the other datasets from the Theewaterskloof municipality are stored in a local file Geodatabase. These could have been transferred from data producers through a file transfer protocol (FTP).



Appendix 5

Image Service Definition for Cape Town

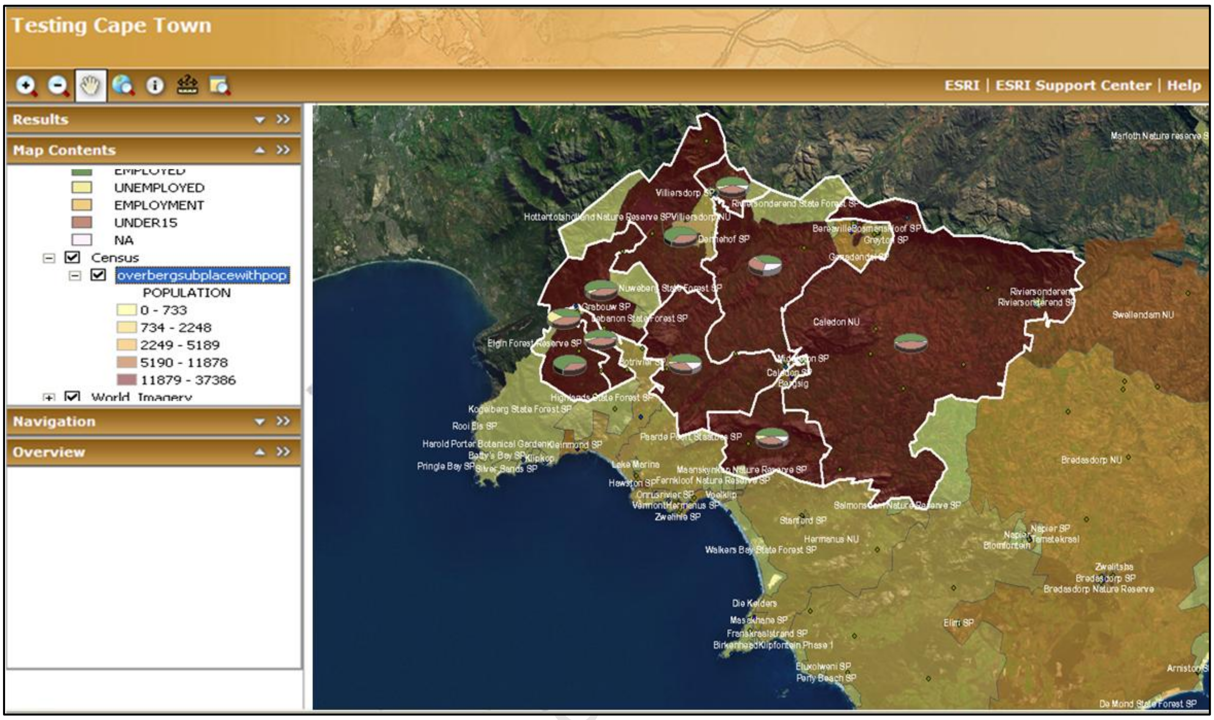
The image below is a picture of the image service definition for the set of images coering the whole of Cape Town, as well as the frames that define the boundaries of the different tiles that make up the service. The Images that were initially about 8GB were compressed to 3.5 GB (more than 50 % compression). Tiling resulted in faster transmission of the image set over the internet because when a client makes a request to view a certain portion of the image set, they are served with the specific tiles that they need to view and not the whole image as when no compression and tiling is applied.



Appendix 6

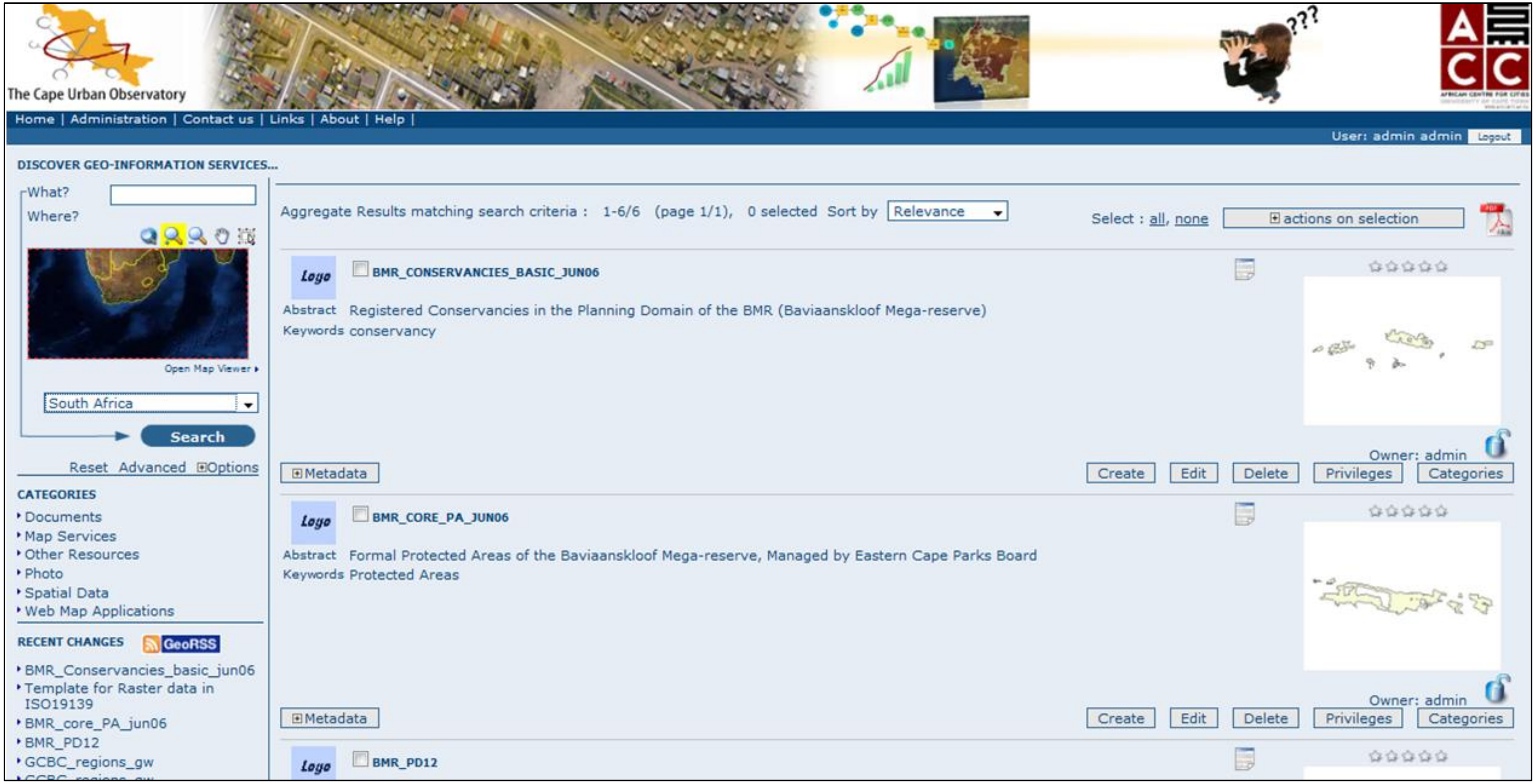
WebGIS application

Below is an example of a GIS application that is being consumed in a browser environment. This was created using the ArcGIS server Application Development Framework (ADF).



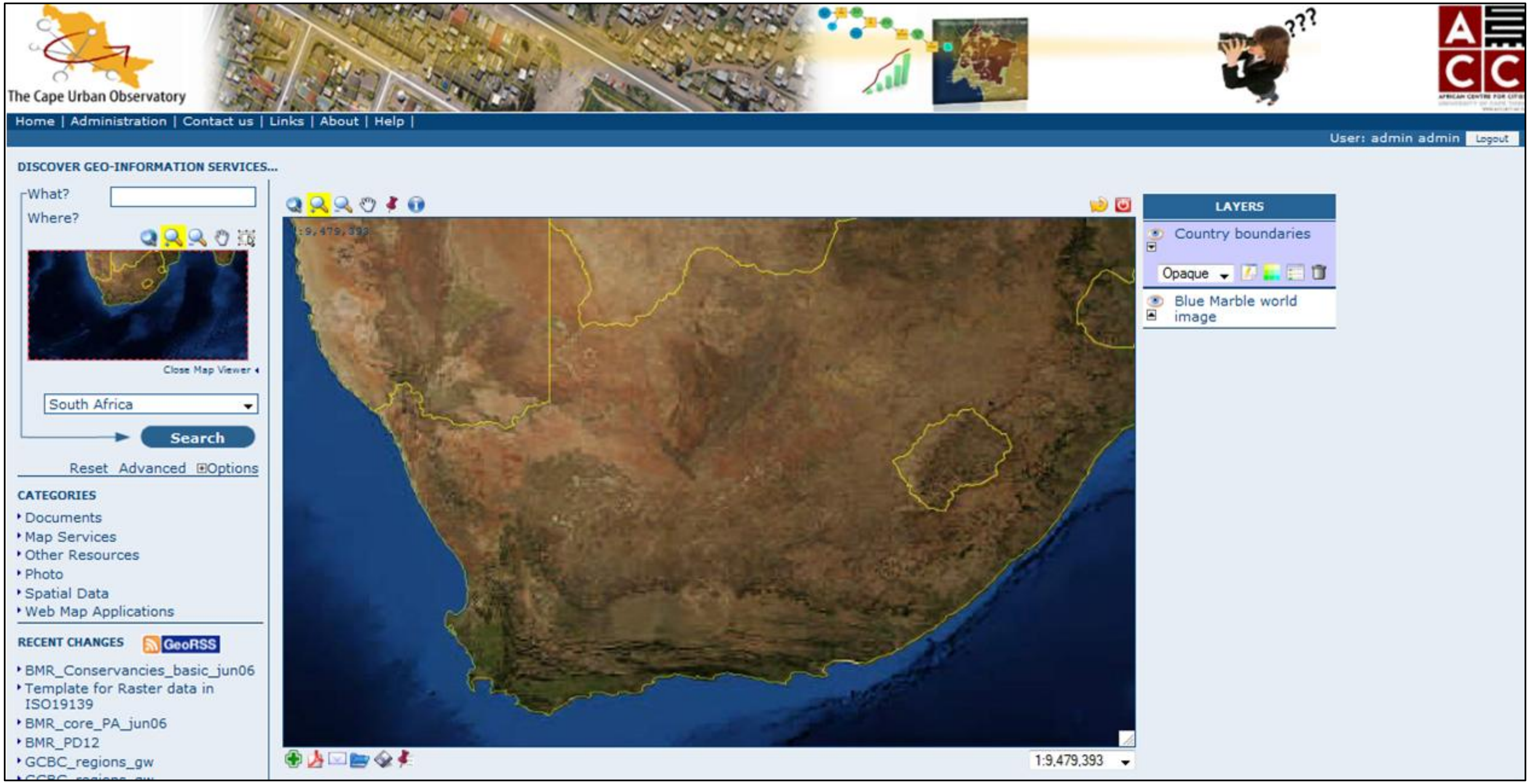
Appendix 7

Interface for searching through metadata records

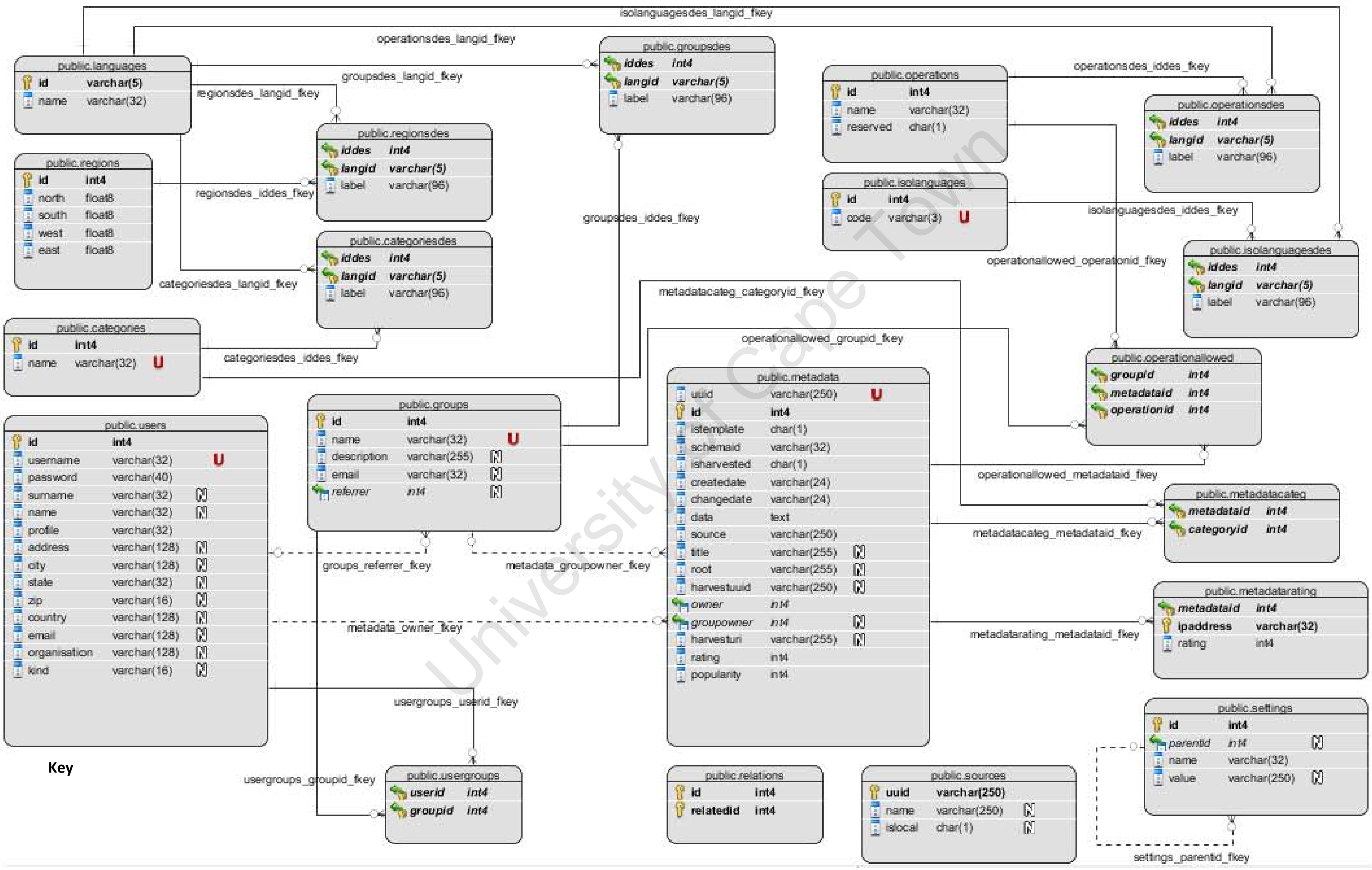


Appendix 8

The Map viewer interface for getting previews of map services



Appendix 9: Detailed ER Diagram for Metadata Database



Appendix 10

Consuming map services in Google Earth through a KML network link. This figure shows water pipes data for Grabouw.