BUILDING BIODIVERSITY DATA INFRASTRUCTURE FOR SCIENCE AND DECISION-MAKING: INFORMATION NEEDS AND INFORMATION-SEEKING PATTERNS IN SOUTH AFRICA

BRENDA DALY

DLYBRE001

Supervised by: Emeritus Associate Prof Karin de Jager

Co-supervised by: Richard Higgs

A minor dissertation submitted in partial *fulfilment* of the requirements for the award of the degree of Master of Philosophy in Digital Curation

Faculty of the Humanities

University of Cape Town

2020

COMPULSORY DECLARATION

This work has not been previously submitted in whole, or in part, for the award of any degree. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works, of other people, has been attributed and has been cited and referenced.

Signature: Signed by candidate Date: 2020/10/26

Acknowledgements

I would like to thank my supervisors, Karin de Jager and Richard Higgs for the guidance and advice throughout the project and SANBI for the opportunity and funding.

To Ferozah Conrad for all her assistance and continuous encouragement.

Thank you to my colleagues Anisha Dayaram and Boyd Escott for the valuable input, and my best friend Linda Downsborough for being just that.

A huge thank you to Ashraf Conrad (my Number 50) for your extreme patience and willingness to help me with my words and numbers.

To my family and friends for your constant support.

To my folks as without them, I would not be where I am today.

Abstract

Biodiversity information is critical to inform science-based policy development as well as to support responsible and accountable land-use planning and decision-making practices. The uptake of available information for these uses is, however, not yet quantified or understood. Here, the extent to which the needs of biodiversity information end-users in South Africa are supported via existing information sources was investigated, at the science, practice and policy interface, using the South African National Biodiversity Institute (SANBI's) online conservation mapping service users as a case study.

A quantitative investigation of the information needs of end-users of biodiversity information was made, their information-seeking patterns analysed and the various uses of information by different user groups in South Africa investigated. This allowed for the implications of these needs and behaviour on system design and information provision to be formulated to better design the envisaged National Biodiversity Information System at SANBI. Based on a representative sample of end-users from policy, implementation and research backgrounds, a questionnaire was used and the responses were examined to determine which content was most useful, what barriers and enablers they face when trying to access biodiversity information, and what degree of interdisciplinary information is needed in addressing environmental problems. A sample of 778 (13%) respondents from a total of 5977 biodiversity information users was analysed from across the country.

The study found that the lack of appropriate or available information remains one of the three highest unmet needs of biodiversity information end-users. The absence of good prior knowledge of sources of biodiversity information and unreliable and inaccurate information are two additional factors that hinder respondents in finding biodiversity information and achieving their goals. The major implication of information deficiency identified by respondents related to uncertain and/or inaccurate outcomes resulting in ill-informed decision-making. A key outcome of the analysis of the survey results are a series of recommendations on how these issues might be addressed, and it is envisioned that these may be used to help guide the development of a National Biodiversity Information System. A broad range of recommendations have been proposed, principally that the interoperability of information from various adjacent and disparate fields of study be combined with biodiversity information as a means of addressing environmental problems.

Table of Contents

Figures

Tables

Abbreviations and Acronyms

ARC	Agricultural Research Council
BGIS	Biodiversity Geographic Information System
CBD	Convention on Biological Diversity
CBA	Critical Biodiversity Areas
CITES	Convention on International Trade in Endangered Species
COP	Conference of the Parties
CSIR	Council for Scientific and Industrial Research
DKIS	Department of Knowledge and Information Stewardship
ESA	Ecological Support Areas
EWT	Endangered Wildlife Trust
GBIF	Global Biodiversity Information Facility
NFEPA	National Freshwater Ecosystem Priority Areas
OBIS	Ocean Biogeographic Information System
SANBI	South African National Biodiversity Institute
SAWS	South African Weather Service
TOPS	Southern African Threatened and Protected Species
UCT	University of Cape Town
WCMC	World Conservation Monitoring Centre

Chapter 1: Introduction

This research emanates from a practical concern as to how biodiversity data and information can be presented to maximise its discoverability, usefulness and improved outcomes in biodiversity research, management and decision-making. In addition, enable better planning and policy development through the provision and packaging of information in a user-relevant context, to make it easier for end-users to build on existing knowledge. Identified solutions related to the concerns above are described in the literature; metadata that provides contextual details in information discovery is an example (Alford, 2009, Jones, Pryor & Whyte 2013:5). The primary concern, however, is the structuring and organisation of online data to serve biodiversity end-users better, as these data allow "researchers and practitioners to expedite the conversion of results into information products, thereby potentially improving conservation initiatives" (Endangered Wildlife Trust [EWT], 2011). This study investigated biodiversity end-users' information needs, how they undertake finding what they need (information-seeking patterns) and how they utilise the information. The results of this study will be used to ensure a user-friendly biodiversity information system. This data infrastructure intends to integrate and holistically manage all biodiversity information, support end-users of this information in achieving their goals, through informed decision making, and bring policy-relevant biodiversity information online to help institutions and individuals understand the impact of human activity on biodiversity. As defined by Thessen *et al.*, (2016:296) data infrastructure are the tools and services that support data-driven science.

1.1 Background

Researchers, practitioners and policy-makers need to find relevant information, assess the quality of the data and then use the information in a research or decision-making process. These tasks are often related to the use or management of living resources, and science-based policy development. Good policies also make efficient use of information throughout their lifespan to monitor progress and review the effectiveness of policy (World Conservation Monitoring Centre [WCMC], 1998:1). A major current obstacle to finding key biodiversity information required to support planning and policy formulation is often the result of ineffective data sharing (Tenopir *et al.,* 2011). This is due to a lack of skilled personnel and infrastructure (hardware, software, data formats, tools and protocols), and the benefits derived from sharing data are often greater for the end-user than the creator (Alford, 2009, Thessen & Patterson, 2011, Davis *et al.,* 2014).

Because biodiversity data and information might not immediately be at hand, there is a risk of biodiversity loss due to uninformed management decisions and the potential for inadequate policy development (Brownlie, Walmsle & Tarr 2006). Shanmughavel, (2007:368) writes that data is often spread across many databases, or the information has not been digitised. In addition, there is uncertainty about what information to use owing to the availability of many web-based platforms, and where to timeously find the information required. For these reasons, it is important to understand the information-seeking behaviour of biodiversity end users, to develop information systems that align with these patterns, thus addressing end users' needs (Agosto & Hughes-Hassell, 2005:141). Information-seeking behaviour is defined as "the purposive [purposeful] seeking for information as a consequence of a need to satisfy some goal" (Wilson, 2000:49) and this is explored in the theoretical discussion below.

The South African National Biodiversity Institute [SANBI] is a parastatal organisation legally mandated to manage biodiversity information, based on the *National Environmental Management: Biodiversity Act, No. 10 of 2004* (NEMBA, 2004). SANBI fulfils this obligation not only from information generated within SANBI but also from partner organisations (e.g. national-scale integration of herbaria and museum data). As such, SANBI has made significant investments in developing large biological information resources and assembling this information into publicly accessible databases. SANBI leads and coordinates research, and monitors and reports on the state of biodiversity in South Africa and their legal mandate includes proactively harnessing, organising, refining, synthesising and managing biodiversity information (NEMBA, 2004). The development of an integrated information system is being driven by meeting SANBI's mandate and the Institute's recognition that information-systems developed around end-users' needs are more likely to be used. This study aims to understand what the biodiversity information needs are of various user groups and to define, in the clearest possible terms, the problems that the new National Biodiversity Information System will be expected to address.

It is important to distinguish between the concepts of information and data and to define what is meant by information within this study. In their review of biologists' technical and sociological issues, Thessen and Patterson (2011) clarify that *data* is limited to unprocessed objective data (artefacts or observations), from which *information* is created through analysis, becoming meaningful and gaining context. *Knowledge* is how the information is internalised by an individual. An information need is defined by Nicholas (2000) as a need for specific information to complete a task or achieve a goal and

in this study, the word “information” can, therefore, refer to either data or information. This is examined in the theoretical discussion (section 2.4).

1.2 The rationale of the study

SANBI’s responsibility is to secure the future of South Africa’s biodiversity and it has organically developed a range of systems, tools and policies (data infrastructure) to respond to information needs identified by internal and external projects and programmes. However, management costs are escalating and, with the numerous websites (a symptom of multiple projects) through which SANBI makes data and information available, end-users are finding it difficult to navigate the maze of biodiversity information sources.

There is awareness by SANBI of the barriers to finding biodiversity information on the numerous SANBI websites: with 13 outward facing information sources, end-users are unaware of the data available and where to find information. Document and data silos mean that information is stored three different times in three different ways and possibly conflicting within the various systems. The specific intention of the survey within this study was to inform the construction of an integrated biodiversity information source and not necessarily to investigate whether there is a need for another information source. Whilst the current systems provide valuable data, information and tools, they do not capitalise on the potential knowledge and other synergies that can be generated through effective systems integration. A study of this nature will provide the data needed to assist in the development and redesign of a system for discovering relevant and coherent biodiversity information.

1.3 Research problem

Developing and maintaining useful information services takes considerable investment and it is, therefore, vital to understand end-user needs and information-seeking behaviour. Within the field of biodiversity information, there is a lack of understanding of the end-users and user groups’ information needs in South Africa. It is also important to understand what users do with the information and how this may be linked to science, practice and policy needs, and also what impact inaccessible information can have on biodiversity loss.

Needs for biodiversity information vary, and biodiversity is often not the only factor influencing, for example, sustainable development. Biodiversity information must often be integrated with social,

political and economic information for any analysis to be useful (Davis *et al.,* 2014:691). Biodiversity information is often available in different data types from diverse information sources and at various geographical scales, and often in a form that is not relevant to end-user needs. In some cases, metadata is used in discovering data and can facilitate cross-disciplinary connections and linkages. However, researchers often fail to compile comprehensive and clear metadata to retain data integrity or there is a lack of use by end-users (Tenopir *et al.,* 2011:3).

Previous studies (Shanmughavel, 2007, Tann, *et al.,* 2008, Avlonitis & Daly, 2014) show that good prior knowledge of where to find needed information can be a problem for a novice. An exploratory interview at a training session for technical local government officials (during the Eleventh session of the Conference of the Parties (COP 11), 2005 in Hyderabad), from the Directorate of Climate Change and Biodiversity, Western Cape Province (City of Cape Town) found that, even though readily accessible data exists in various data repositories such as the Biodiversity Geographic Information System (BGIS), Birds in Reserves Project (BIRP), Agricultural Research Council (ARC), staff often struggle to find the data necessary to carry out their jobs effectively, with data interpretation merely compounding this problem (Avlonitis & Daly, 2014).

The task of data integration and processing to generate and interpret the information from appropriate sources requires experience and expertise. Many decision-makers lack the time, and in some cases the background, to collate and analyse different data types and interpret the results. Furthermore, compiling disparate, heterogeneous and interdisciplinary data when addressing environmental problems is discouraging as it is time-consuming. The broad geographic scope and formats of this information and the difficulties with data integration, which may be technology or data-related, such as the differences in spatial representation and resolution exacerbate these challenges (Koesten *et al*., 2019). During these early exploratory interviews in Hyderabad, many of the participants were interested in the coordination of sub-national, centralised, streamlined, coordinated platforms where information would be easy to obtain. They also suggested that it would be useful to showcase and/or provide a set of case studies on how other local governments and other relevant end-users have used this information (Avlonitis & Daly, 2014).

SANBI has historically attempted to understand the needs of users through formal stakeholder consultation. Steenkamp and Smith (2003), following a national workshop to determine the needs of users of botanical information, hosted by SANBI in South Africa, found that the currency (whether the information is up-to-date) and completeness of information were important at the time of need. The

community felt it important to expand the collection to make it more relevant. Furthermore, it was felt that more physical and online points of access to the information were needed, as well as a library system providing updated literature. Although these needs had not been considered in the initial development of herbarium collections it was, however, at this point that the integration of biodiversity information was first considered by SANBI who since then has been working towards integration.

1.4 Research questions

The study aimed to ascertain the information needs and information-seeking patterns of user groups working with biodiversity data in South Africa and to assist and guide the design of an effective information service that could meet these needs. While SANBI is not the sole provider of national-level biodiversity information, SANBI's online conservation mapping service end-users served to inform this study as this platform also hosts other institutions information.

This research project was therefore guided by the following main research question:

To what extent are the needs of biodiversity information users in South Africa supported via existing information sources at the science, practice and policy interfaces?

To achieve the objectives of this study and address the broad research question above, the following sub-questions were defined:

1. Who is accessing SANBI's biodiversity information online?
2. What are the biodiversity information needs of SANBI's users?
3. What do users do with SANBI information once retrieved?
4. What are the information-seeking strategies of SANBI's online users?
5. What challenges do SANBI's online users encounter when trying to find information?
6. What information is needed from other disciplines when addressing environmental problems?
7. What was published in this field and on the concepts of information behaviour in South Africa?

1.5 Research design

The theoretical concepts underpinning this study were influenced by both Nicholas' qualitative framework (2000) which offer important concepts used to provide a range of perspectives related to

information needs (Afzal, 2017:119), and Wilson's general model of information-seeking behaviour (1997), which emphasises the information-seeking process and causative factors that influence this behaviour. Both models are discussed in section 1.6.

Potential end-users of biodiversity information (27 profiles) were identified following a review of the literature (Appendix 1) and were used to select a purposive sample. These users were sourced from those having previously requested data from, and registered on, the BGIS website (SANBI, 2018). BGIS is an online conservation mapping service hosted by SANBI and had 5977 registered users of which 778 responded to the survey. The online BGIS conservation mapping service aims to assist in biodiversity research, planning and decision-making by providing spatial biodiversity assessment and planning information. The website is generated and maintained by SANBI, whilst the data is sourced from SANBI and various environmental institutions (data partners). The wide variety of users of this system were deemed a suitable population for this study.

To determine the biodiversity information needs of different end-user groups and how people seek information, the survey was designed, using the online software SurveyMonkey (www.surveymonkey.com). The survey questions (Appendix 2) were adapted from the comparative studies of Tann, Kelly, and Flemons (2008), Jamali and Nicholas (2010) and Davis *et al.,* (2014). The survey consisted of both open-ended and closed questions and took approximately 20 minutes for respondents to complete. The link to the survey was sent via email and demographic questions were asked to determine who the end-users were. Consent was given if the questionnaire was completed. A reliable and valid survey was obtained by pre-testing the research instrument and using existing and already tested questions. Results of the survey were compared with the results from other studies and results were combined to provide solid recommendations. Table 1.1 provides the objectives achieved and the problem addressed by the research sub-questions and the data gathering methods used.

Research sub-questions	Objectives	Problem	Data gathering methods
1. Who is accessing SANBI's biodiversity information online?	To identify key user groups of biodiversity information in South Africa.	A lack of knowledge of who is accessing biodiversity information.	Literature review
2. What are the biodiversity information needs of SANBI's users?	To investigate different end-users' biodiversity information needs.	No clear understanding of end-user needs of biodiversity information.	Questionnaire
3. What do users do with SANBI information once retrieved?	To determine what information is required for what specific purpose (scenarios of use).	The context of relevance and purpose of biodiversity information is not understood and needs to be determined:	Questionnaire

Research sub-questions	Objectives	Problem	Data gathering methods
		how and for what reason is the end-user needing the information?	
4. What are the information-seeking strategies of SANBI's online users?	To determine the information-seeking patterns of end-users working with biodiversity data in South Africa.	A lack of understanding of how individuals navigate the complex landscape of biodiversity information and how they choose to satisfy their information needs.	Questionnaire
5. What challenges do SANBI's online users encounter when trying to find information?	To identify the limitations and constraints associated with not being able to find needed information.	A lack of understanding of factors influencing information-seeking strategies.	Questionnaire
6. What information is needed from other disciplines when addressing environmental problems?	To develop interdisciplinarity awareness to improve access to integrated biodiversity information for science, practice, and policy.	A poor understanding of how broad the cross-disciplinary needs are.	Questionnaire
7. What was published in this field and on the concepts of information behaviour in South Africa?	Gain insight into the findings in this field of study in South Africa.	Insufficient knowledge of similar studies completed in South Africa.	Literature review

Table 1.1: Research questions, objectives, problem statements and data gathering methods

Respondents were given 5 months between June and November 2018 to complete the survey. The survey was confidential with participants being asked to indicate their name or affiliation/organisation for the sole purpose of allowing the researcher to make follow-up telephone calls if clarity was needed. Data collection, coding and data analysis were done in MS Excel and data were analysed in the context of the purpose of each question.

1.6 Nicholas's and Wilson's models

This study adopted a quantitative approach and explores two conceptual frameworks, namely those of Nicholas (2000) and Wilson (1997). The first framework is a structure for evaluating information needs proposed by Line (1969) and adopted by Nicholas (2000) that focuses on investigating the context and content from which the user derives their need (Inskip *et al.,* 2008). The second framework provides valuable insights for this study on information-seeking behaviour (Wilson 1997) as it describes "how people need, seek, manage, give and use information in different contexts" (Savolainen, 2007). The primary purpose of these conceptual models was to provide tools to interrogate the research question, guide the development of the questionnaire, provide insight into the responses analysed and locate or frame the area of work.

Both Nicholas's and Wilson's models provide practical frameworks with which to explore: i) information needs (the information needed to effectively complete a task, solve a problem or pursue an interest), ii) information-seeking behaviour (related to how these needs are met), which happens regularly at times out of necessity (Case, 2006:18), and iii) the use of information (processing information found).

Although Nicholas's and Wilson's models have a common ground (discussed in section 2.5), there are conceptual differences. Nicholas centres on a needs assessment framework consisting of eleven characteristics that are described further in section 2.4. This structure considers the various characteristics of information needs, explains circumstances and is measurable, thereby providing a shared understanding and improving synergy with other studies. On the other hand, Wilson's framework presents the various stages in the information-seeking process and identifies information-seeking related activities (Wilson, 1994). These stages are essential in understanding the sequence of events from when an information need is triggered to a conscious search for information and influencing variables. Wilson provides a model of the information-seeking process and incorporates three theories that are used to explore various aspects in the process; these include stress-coping theory, risk-reward theory and self-efficacy theory, as explained in section 2.4.

The general model of information behaviour was founded by Wilson (1981) and has been used in a variety of disciplines (e.g. health, information science) providing a large body of work which has been criticised by Savolainen (2007) as being an umbrella concept that includes a particular topic with similar names such as information-seeking behaviour, information use, want and demand. The model, copyrighted in 1995, was progressively developed through expansion of Wilson's model of 1981. Although Wilson is still adapting and working on his model it best illustrates an iterative process from an information behaviour perspective, consisting of six stages: i) recognition of a knowledge gap (context), ii) activating mechanisms for seeking information (motivation), iii) causative factors influencing the behaviour of an end-user (intervening variables), iv) the reasons to pursue or forego the information-seeking process, v) the purpose, effect, or function of the information (information-seeking behaviour), and vi) the use of the information (Nussbaumer *et al.,* 2009). These stages were used to guide the researcher in this study in understanding the information-seeking behaviour of end-users of biodiversity and to facilitate the development of an information system.

1.7 Ethical implications

The survey was confidential rather than anonymous; this means that respondents were required to either indicate their name or organisation for follow-up. However, their names and organisations have not been revealed in the dissertation. Confidentiality was ensured by limiting access to identifiable information and securely storing data and documents. Ethics clearance procedures were followed as per the University of Cape Town (UCT) Humanities Guide and the Department of Knowledge and Information Stewardship (DKIS) procedures. The study received ethical approval from DKIS and research approval from SANBI. This research was conducted in compliance with the University of Cape Town's Policy on Research Ethics. The University of Cape Town Ethics Committee approved this study on the 19th of June 2018 (Ref No.: UCTLIS201806-07) via email as an online survey.

1.8 Limitations and delimitations of the study

A list of limitations and steps taken to ameliorate them is presented below in Table 1.2:

Limitations	**Measures taken to ameliorate**
The electronic survey that was conducted had a poor response rate (13%), raising concerns of representivity.	To mitigate nonresponse bias the survey was configured to be mobile responsive, an extended data collection period was allowed and multiple reminders were sent and an incentive was provided. Respondents were also reassured that the survey data would be kept confidential in case participants were less likely to participate if they thought their names would be disclosed.
The survey had an average completion rate of 57% as respondents did not answer every question. This resulted in certain questions having a lower sample size and level of reliability.	The first four questions of the survey were easy to answer questions (multiple-choice). The questions were substantially independent so where respondents did not provide an answer to a specific question, these were excluded from the stated frequencies in the analyses.

Limitations	Measures taken to ameliorate
A lack of detailed or explanatory responses (subject to miscommunication of the data) may affect the quality of the data.	Both qualitative and quantitative data type questions were included in the survey. The qualitative research (open-ended questions) provided descriptive information on the topic.
A disconnect between the respondents' understanding of the question and the researcher's interpretation of their responses due to different backgrounds and knowledge.	A pilot study was used to ensure a rapport was built with the respondents and definitions were included to clarify the meaning. Errors due to translation were minimised through comparative studies.
Qualitative type questions take longer to answer.	The total number of open-ended questions for the survey was limited to 7 of 26 questions so that completing the survey was not overly time-consuming.
Sets of answer choices may be imprecise.	This was mitigated by the provision of textual responses for unforeseen answers and an "other" option was included, to enable respondents to include additional answers.
The survey may be a burden on respondents.	This was reduced by explaining that the survey would be meaningful to the intended respondent and examples were given so the questions appeared less onerous.
The sample of respondents may potentially be non-representative of the population of eventual users of the system.	A literature review was used to clearly define the target end-user group.
The introduction of sources of bias that may reduce the accuracy of the results.	Biased words and phrases were avoided and definitions provided for the various terms in the questionnaire to ensure there was a full understanding of the topics by the respondents. A mix of rating scales and open-ended and closed survey questions were used in the type and design of the questionnaire to minimise bias. An introduction was provided for at the

Limitations	Measures taken to ameliorate
	start of the three sections in the questionnaire summarising the topic being considered. To combat researcher bias, several open-ended questions were asked. However, it is acknowledged that researcher bias may still have occurred during the analysis of open-ended questions.
Whilst many biodiversity professionals do speak English; if it were the respondents' second or third language, this might have compromised their written responses.	Mitigation measures included avoiding the use of idioms and jargon, providing clarification and defining what was meant by certain concepts.
The choice of conceptual frameworks was a limitation. The problem is that there is no clear systematic conceptualisation of the term 'information need' and the use of placeholders such as information seeking and use are applied as measurements which result in the imprecise use of terms within this field of research (Wilson, 1994, Afzal 2017).	Two conceptual frameworks, of Nicholas (2000) and Wilson (1997) were combined to address the two variables of 'information need' and 'information-seeking and use' that are represented in the sub-questions of this study. These issues are discussed further in section 2.4.

Table 1.2: Limitations of the study and the measures taken to ameliorate them

1.9 Dissertation structure

Although the chapters are interrelated, each focuses on a different aspect within the study area. Chapter 1 (Introduction) introduces the research project and objectives, provides a theoretical background and describes the research problem. This chapter also clearly outlines the limits of the study and describes the proposed research methodology.

Chapter 2 (Literature review) provides a synthesis of concepts and arguments from related studies, and in this way defines the diverseness of the literature and creates a conceptual framework so that the research question and methodology is better understood. The theoretical models for the dissertation are also discussed.

Chapter 3 (Methodology) provides an outline of the research methodology. This chapter ensures an understanding of what processes led to the results and conclusions that were detailed in chapters 4 and 5.

Chapter 4 (Results and discussion) describes research findings, presents the results of the survey and a short discussion. Research results were translated into a set of practical strategies that can be used in the development of a new biodiversity information system.

Chapter 5 (Conclusion and recommendations) summarises major findings and outlines the practical approaches and recommendations that may be used in the development of an information system with the intention that this research will find a direct and indirect application.

Chapter 2: Literature review

This chapter provides a critical overview of the research completed to-date on biodiversity information needs and information-seeking behaviour. In addition, the review also focusses on how other studies have defined who the end-users of biodiversity are, documented information needs of the international biodiversity community and describe intervening requirements needed in building an integrated system.

Biodiversity data deals with different kinds of heterogeneous data (species, taxonomy, ecological, genetic and geographic) from a variety of systems that are often not integrated and end-users have to alternate between information systems to collect relevant data (Torres *et al.,* 2006 Hoffmann *et al.,* 2014). Data infrastructure is a digital infrastructure that promotes data sharing and consumption. This spans storage and delivery mechanisms (networks, databases and technologies), resources (datasets, identifiers, licences, and registers), standards, policies, archives, data management, analysis and visualisation (Dodds & Wells, 2019). The challenge in building a data infrastructure service is not only about integrating heterogeneous data from various sources but also giving attention to technical and contextual requirements.

2.1 End-users of biodiversity information

Bodin and Prell (2011:4) in their book on the application of social network analysis in natural resource management, differentiate between researchers, practitioners and policy-makers. This, however, does not define the different types of users associated with biodiversity information usage. Appendix one is a listing of the end-users identified within these roles (27 profiles). This list was compiled following Swart *et al.*'s, (2017:15) recommendation that it is important to ascertain the various categories of users within these sectoral groupings in order to determine the various types of information required and the techniques (functionalities and services) in which these data should be presented online.

The Global Biodiversity Informatics Outlook, published by the Global Biodiversity Information Facility (Hobern *et al.,* 2012:8) distinguishes between seven groups of end-users of biodiversity information; i) policy-makers, ii) funders, iii) regional, national or international biodiversity organisations, iv) owners or custodians of biodiversity data, v) biodiversity researchers, vi) Information Technology professionals or biodiversity informatics specialists, and vii) members of the public. The Hobern *et al.,* (2012) categorisations are more closely aligned with the activities associated with biodiversity

information management which include data collection, data aggregation, analysis/science, policy advice, policy development, policy implementation and monitoring. Most tasks result in an output which is often the result of the outcome of the use of information and/or a result of the information-seeking process.

Beaman and Cellinese (2012) distinguish between four types of users: i) primary users (e.g. collection managers, curators) which may also be defined as the custodians (data stewards) of the data; those responsible for the curation of data; ii) downstream users (e.g. scientists, practitioners); iii) social network users which result from a crowdsourcing workforce (i.e. an online community of volunteers that assist with the transcription of labels and fields notes); and iv) citizen science communities (i.e. volunteers that collect biodiversity observations data). It is clear that they differentiate between the end-users of various forms of information. For example, Roetman and Daniels (2011) define citizen science as a research methodology that involves the contribution of citizens and this form of information is described as methodological by Nicholas (2000:53) under the information needs characteristic of 'nature' in his framework for evaluating information needs.

The end-users of biodiversity information listed above reflect the diverse functioning of individuals with a shared need for biodiversity information. This diversity has an impact on how the information should be managed. The data infrastructure to be built at SANBI will serve three biodiversity end-user groups: i) scientific community, ii) practitioners and iii) policy-makers. The first challenge in building such an information system is to accommodate the needs of all these user groups and associated end-users efficiently, creating value through the way that the information is organised and integrated. The second challenge is that it is not uncommon for end-users to have overlapping roles (section 4.1.5.2). For example, a researcher may also be a provincial practitioner whose work feeds into policy-making decisions.

2.2 Information needs of users working with biodiversity data in South Africa

Several surveys on biodiversity data need according to data types and biodiversity groups were completed internationally with global coverage (Faith *et al.,* 2013; Ariño, Chavan & Faith, 2013). Eight case studies were found from India, Australia, United Kingdom, Belize, Canada and New York State (Ariño *et al.,* 2013) and a country assessment from Ghana (Asase & Schwinger, 2018). These studies explored the needs of users working with biodiversity data using a Content Needs Assessment (CNA) survey which determined what data users may be using, what they would use if available and what primary biodiversity data is needed. Each study had a different focus and purpose, and was based on

a particular outcome. As described by Ariño *et al.,* (2013) a Content Needs Assessment in essence "examines the extent and adequacy" of biodiversity information currently available, against the desired objective or problem identified.

The World Conservation Monitoring Centre (WCMC) stated that each country responds differently to the task of generating, organising and providing information (WCMC, 1989:17). This is either due to their unique environments or environmental problems and therefore many international case studies may not be relevant in the South African context. Due to the diverse purpose of each CNA survey and considerable variation between countries a comparative approach and analysis was not feasible. As a result, it was not possible to determine the documented needs of the biodiversity community. However, similarities in these assessments that relate to this study include, the importance of linking new data types and sources (e.g. molecular, environmental, ecosystem services); knowing where to find primary biodiversity data, and the fact that their information is scattered in diverse formats and resolutions. The studies listed above show that research in determining data gaps on various topics and within different counties has received significant focus within the biodiversity community. It is due to these studies' significance and limited relevance, and that the results are not comparable, that this type of study was not considered.

Although there is limited literature on the information needs of biodiversity end-users in South Africa, several international studies have identified biodiversity information needs elsewhere. A study by Davis *et al.,* (2014) in the South-eastern United States, assessed the type of biodiversity information needed, tools needed to integrate information, attributes of preferred biodiversity information sources, barriers and facilitators to finding, accessing and using biodiversity information and characteristics of information sources that are valued. According to survey results, information is not always described in a way that allows for discovery, and the greatest challenge is users' ability to find different types of information (Davis *et al.,* 2014:698). A major barrier for respondents was the lack of time to find needed information with data management and metadata tools playing an important part in biodiversity work.

Tann, Kelly, and Flemons (2008) conducted a User Needs Analysis in southern Western Australia which reviewed the users and uses of biodiversity information. The analysis assisted the development team in setting priorities at the start of developing an information portal; this included determining expressed needs, key difficulties and identifying workflows. Tann *et al.,* (2008) identified the users of biodiversity, what biodiversity data was being used and sources and examples of tasks that were core

to their work-roles. Data was captured via survey responses, workshops, interviews and longitudinal studies.

Davis and Tenopir's (2012) assessed the information behaviour of natural resource managers from the Southern Appalachian region (United States), this study showed that this group represents a unique end-user group (20%, n=87 of 428 survey respondents) when compared with users, where biodiversity work-related tasks do not include natural resource management. This group had a higher need for summarised rather than raw data and a greater demand for resource restoration and management type information. They consulted a variety of information sources more specifically those provided by state environmental and wildlife resource agencies. The authors did not offer any thoughts on how information could be made more accessible for this particular end-user group.

White and Molina (2006) investigated the major challenges in managing biodiversity in the Pacific Northwest region of North America with a focus on forest and rangeland habitats. They found a lack of a common definition of biodiversity, of monitoring frameworks and standard legislative policies to direct biodiversity management. This was accompanied by difficulty in finding relevant information and social and economic impacts. The information needs and potential solutions identified in the study were specific to the study area which makes it difficult to extrapolate the finds to the South African context. In the context of this study, biodiversity is defined as "the diversity of genes, species and ecosystems on Earth, and the ecological and evolutionary processes that maintain this diversity" (SANBI, 2016).

Although the studies listed above pertain to biodiversity information users' information needs, the different studies are sourced from and refer to different countries, different environmental challenges, different guiding policies and legislation, and as such differ widely in resources associated with solving these environmental problems. Furthermore, information needs tend to evolve over time either with persistent advances in technology and the affordability of devices or storage space. In some instances as information becomes available, other needs for information and/or policy developments may be raised, such as the Aichi Targets (a strategy to reduce the loss of biodiversity) (Ariño, Chavan & Faith, 2013:3, Case, 2006:79).

Mostert and Ocholla (2005) investigated the information needs of a sub-population from the target population in this study. Mostert and Ocholla's (2005) study using questionnaires supplemented by observations, revealed the preferred information sources of South African parliamentarians using Wilson's general model of information behaviour. Results showed that parliamentary libraries which

should provide parliamentarians with necessary information are not well-utilised, online information sources are rapidly increasing and the need for oral sources are declining. The research was done by surveying all parliamentarians and parliamentary librarians in South Africa (23%, n=167 responses from 763 questionnaires distributed) on a national and provincial level. The authors found that 70% of parliamentarians conduct their own searches and prior knowledge of where to find information influences the selection of resources and the kind of information contained (Mostert & Ocholla, 2005:141). However, high-level diplomats (52%) rely on information intermediaries in acquiring information.

Gathering from the literature, three themes emerged when assessing biodiversity information needs. Firstly, the information needs that support biodiversity reporting and monitoring functions, which need to align to meet the national and international obligations of a country such as the twenty Aichi Targets 2020 of the Convention on Biological Diversity (CBD, 2013a); the 17 Sustainable Development Goals 2030 (169 targets and 244 indicators) which aim to reduce inequality, end poverty, protect the environment and promote justice and peace (Brooks *et al.,* 2015); and the Paris Climate Agreement under the United Nations Framework Convention on Climate Change (UNFCCC) to reduce the emission of gases contributing to global warming (UNFCCC, 2015). The best available data are a fundamental requirement to analysis, monitoring and reporting on the state of biodiversity. Brooks *et al.,* (2015) and Han *et al.,* (2014) highlighted the importance of data-driven problem solving or decision-making such as identifying and managing species and areas of importance (these include protected areas, key biodiversity areas, Red List of Species and Ecosystems or modelling the spread of invasive species) in tracking these targets and goals. These assessments are not possible with limited, inappropriate information or where there are quality concerns.

The second information need is biodiversity management (e.g. Environmental Impact Assessments), and spatial biodiversity assessment and planning (identifying and prioritising areas for biodiversity conservation (SANBI, 2016:10). Botts *et al.,* (2019) reported on the evolution of conservation planning in South Africa from theory to implementation and the key elements to this transition. The study showed various types of data used including information from other disciplines. The third is the scientific understanding of the data (research). It is therefore critical that biodiversity information is made available in a relevant and useful form to address these needs.

2.3 Information-seeking patterns of user working with biodiversity data in South Africa

A sub-question of this study aimed to determine the information-seeking strategies of SANBI's online users. How end-users undertake finding information was tested in the questionnaire by asking focused questions on the preferences of information sources, online sources consulted and the ease or difficulty of finding information (refer to Table 3.1). Understanding these patterns will better support the activity of information seeking and facilitate the design of the system. Nicholas (2000:47) and Wilson (1997:562) similarly describe information-seeking strategies. Table 2.1 compares the similarities of information-seeking strategies as described by Nicholas and Wilson.

Nicholas 2000	**Wilson 1997**
▪ Fact-finding (answer specific questions) ▪ Current awareness (keep up-to-date) ▪ Briefing/background (broadly familiar) ▪ Stimulus (provides ideas or heightens interest) ▪ Research (in-depth studies)	▪ Active search (actively seeking) ▪ Ongoing search (updates) ▪ Passive search (relevant information) ▪ Passive attention (non-intentional)

Table 2.1: Comparison of the information-seeking strategies of Nicholas and Wilson.

Transaction log analysis from various SANBI information services was considered at the start of the study. These are log files that contain the actual event logs and statistics for every information search query. Wilson's (1999:257) paper in which he reviews models in information behaviour research, differentiated between information-seeking and information searching. Information-seeking is concerned with general behaviour at the initiation or activation of information-seeking and searching is simply querying what is available on a particular topic or information system. Case and Given (2016:95) highlighted the fact that researchers rarely define information seeking and describe it as "a reaction to the recognition of an information need". As aptly described by Cole (2011:1220) "all information searches for one school assignment involve the same information need".

Cole's (2011) study analysed Taylor's (1968) four-level information need model and proposed a theory of information need for information retrieval by developing six propositions upon which the theory was based. These include that Taylor's four information need levels are in every search, user information needs do not evolve only the topics explored, the type of information used evolves, information systems are recognised as knowledge formation/acquisition systems, an additional social

framing input level is needed in Taylor's four-levels and the visceral (instinctual) level of Taylor's four-level information need leads to adaptive behaviour.

Information-seeking research has been criticised for being a subjective process, having poor generalisability and a lack of theory building, but authors such as Weiler (2005) and Fisher and Julien (2009) reason that information-seeking research can be applied practically to improve information delivery. Numerous methods, key concepts and models (theoretical frameworks) have been developed as outlined in Fisher and Julien's (2009) review of information behaviour research published from 2004-2008. Information-seeking researchers use the terms knowledge-seeking versus information-seeking interchangeably; Wilson (1981:4) suggests that this is due to inappropriate use of the terms related to the purpose of the research.

Information needs arise due to a knowledge gap hindering the user in completing a task. Various studies (Kujala, 2002, Mostert & Ocholla, 2005, Janse, 2006, Case, 2006:98, Rowley, & Hartley, 2008, Fisher & Julien, 2009) concluded that several elements modify and influence seeking strategies and the understanding of and interpretation of information. As described by Meyer (2016) these include the interaction between elements in the user's mental structures (cognitive – knowledge and skills of an individual), sensorimotor (motivation) and affective (anxiety or uncertainty) as well as the elements in the context in which an information need arises (e.g. tasks, a situation in action and dialogue).

Savolainen's (2012) study on conceptualising information need in context, identified three major contextual features of the information need. These included: i) situation of action (a situation or set of circumstances in which an information need develops that are effected by where an individual is at a specific time and place), ii) task performance (a work task or problem-solving which are also influenced by intervening factors (e.g. urgency of the problem or complexity of the task)) that can influence the information-seeking activity), and iii) dialogue (personal communication in which information needs are formalised during conversation).

A comprehensive literature review found no evidence of studies focussed on information-seeking patterns for end-users of biodiversity in South Africa. However, a comparative study in Australia associated with freshwater ecology and rivers investigated the knowledge-seeking strategies of natural resource professionals reflecting on the various models and processes used. Cullen *et al.,* (2001:4) defined knowledge-seeking as a process used to "actively acquire information from selective sources" and therefore suggested that information should be arranged from the viewpoint of the

information seeker (Cullen *et al.,* 2001:19). The authors explored traditional models of information seeking such as the library and repositories and described other strategies such as extension, knowledge broking and joint problem-solving models. Results showed that these users rely on professional networks to find the necessary information.

Janse (2006) assessed the seeking and use of scientific information by decision-makers in European forest policy development. His survey showed that personal communication was an important information source. Online information sources were another means of seeking information although it was also considered a barrier due to the quantity of information and complexity of the websites. Sinclair, Mozzotti and Graham (2003) were concerned with how the usefulness of information determines its consideration by end-users in a decision making process. This study in South Florida identified reasons for decision-makers seeking and using scientific information or not. The authors, Sinclair *et al.* (2003) examined land-use decision-makers' intentions to use threatened species information for planning and regulation. Results showed that although past behaviour and social norms play an important part, demonstrating the potential impact of their decisions on threatened species, had a greater effect.

Jamali and Nicholas (2010) evaluated the information-seeking behaviour of scientists from two different fields of study (physics and astronomy). The study explored the issue of scattered literature and interdisciplinarity (the extent to which researchers from a field of study rely on the literature of other disciplines) of both groups. The case study findings showed that the main method of keeping up-to-date was browsing and the methods used for finding articles were tracking references and searching in the subject database.

2.4 Theoretical framework: Information behaviour models

Two theoretical models and their relevance to this study were briefly introduced in Chapter 1. Nicholas's analytical model consists of a set of related variables (characteristics of information needs) that explain, describe or predict the phenomena of information needs, and Wilson's model describes and explains the sequence of events or actions by individuals in finding information of some kind. For this study, the information needs of specific categories of end-user are the primary focus in the investigation. Survey questions related to the respondents' information-seeking behaviour were also used to draw inferences about their information needs, as information-seeking is the observable evidence of information needs as described by Case (2006:81).

Both Nicholas (2000) and Wilson (1997) agree that information needs stem from the necessity to meet three basic human needs and that these are interrelated, highly personal and can vary amongst individuals with the same need; namely, physiological needs (includes the need for water, air, food and sleep), psychological needs (factors that drive individuals to achieve certain goals, for example, an attitude towards the environment e.g. saving water or recycling), and cognitive needs (mechanisms of learning or problem-solving).

Wilson (1999:250) described a model as the framework for thinking about a problem. The problem as highlighted by Wilson (2000:50), Cole (2011:1218), Meyer (2016) and Case and Given (2016:94) is that information needs are mental processes and only become observable through information activities such as seeking and use. Wilson's (1997) model was selected to explain the information activities respondents may perform when identifying their information needs. These theories of information seeking and use are tested through the questionnaire, refer to Table 3.1, which shows which research question was addressed by what questionnaire question. Although it is more productive to study observable behaviour then internal mental events this has inherent limitations such as the dependence on the provision of information which is discussed later in this section. Following Meyer's (2016) suggestion that the relationship between information needs, seeking and use is a cause and effect combination, Nicholas' (2000) model provided the categories and framework required to analyse and determine the information needs, and Wilson (1997) clarified the sequence of events and the intervening variables.

Nicholas used several concepts for measuring various aspects of information needs: ways of looking at or interpreting the information needs phenomenon. Nicholas's needs assessment framework based on Line (1969) for evaluating information needs, considers 11 major characteristics of information needs: subject, function, nature, intellectual level, viewpoint, quantity, quality/authority, date range/currency, speed of delivery, place, processing and packaging. Demographic factors affecting information needs considered by Nicholas, are, however, not considered in this study due to the limited scope on biodiversity information needs, these include job, country/culture, personality, information awareness, gender, age, time availability, access, resources/costs, and information overload. The needs across the user base were close or similar enough not to require analysis of these factors. However, the characteristics and work-roles (job) of end-users of biodiversity are discussed in section 4.1.

Meyer's (2016) study focused on establishing the fundamental components of information behaviour by analysing the contents of existing multifaceted models and proposed a model to guide novice

researchers. Following a careful comparison, the components that emerged from Meyer's study were similar to Nicholas's characteristics of information needs. For simplification purposes the following matches were made: the role of context relates to viewpoint and nature, the inner experience relates to the intellectual level, the role of information needs relates to function, the relevancy of information relates to the subject and the impact of technology relates to processing and packaging. Information activities relate to information-seeking behaviour as described by Wilson (1997) which are observable through information activities.

Nicholas discusses subject and functions in a continuum with the nature of the information, the intellectual level of an end-user (information complexity) and their viewpoints. Quantity, quality, currency and temporal extent, timeliness of delivery, spatial extent, processing and packaging are considered descriptive information. Whereas, Wilson locates information needs in a hierarchy as described by Nussbaumer *et al.,* 2009 namely: information behaviour, information-seeking behaviour, information search behaviour and information use behaviour. Nicholas's characteristics of information need and Wilson's core concepts are defined and elaborated on in the relevant sections of this dissertation. These key elements are used to make the approach of the research of this study clear and provide explanatory factors (Case, 2006:121). This study applies a combination of the two models to address various aspects in information behaviour, using Nicholas's framework of identifying, evaluating and comparing information needs and Wilson's model to illustrate the iterative process of how an information seeker resolves a need.

In linking the various stages in the information-seeking process, Wilson refers to three theories to define his general model of information behaviour (Figure 2.1). Stress/coping theory is used as an activating mechanism, an example is how finding information facilitates coping by reducing uncertainty in cancer patients in health studies (Wilson, 1999). In this study, task performance (the task or problem at hand) causes an individual to seek information. The rationale could be i) job satisfaction in which the respondent wants to do a good job or keep up-to-date (Nicholas 2000:24), ii) learning for the sake of knowing and gaining knowledge (Cullen *et al.,* 2001:10); or iii) a need to act (Case, 2006:88). Spencer (2011:9) suggested that understanding the mindset and motivations of visitors using a website will ensure the utilisation of the system.

The second theory used by Wilson is risk/reward theory which balances the risk perceived by an individual against the anticipated rewards of discovering useful information. An example is the consequences associated with not finding information that may result in ill-informed decision-making. This is closely related to the self-efficacy theory which considers the competence in one's

ability to complete a task and the difficulty of the task (Watters & Duffy, 2005:243). This then affects an individual's thoughts, behaviour and affective state which in turn affects the action taken and effort put in (Miwa, 2005:54, Case, 2006:87).

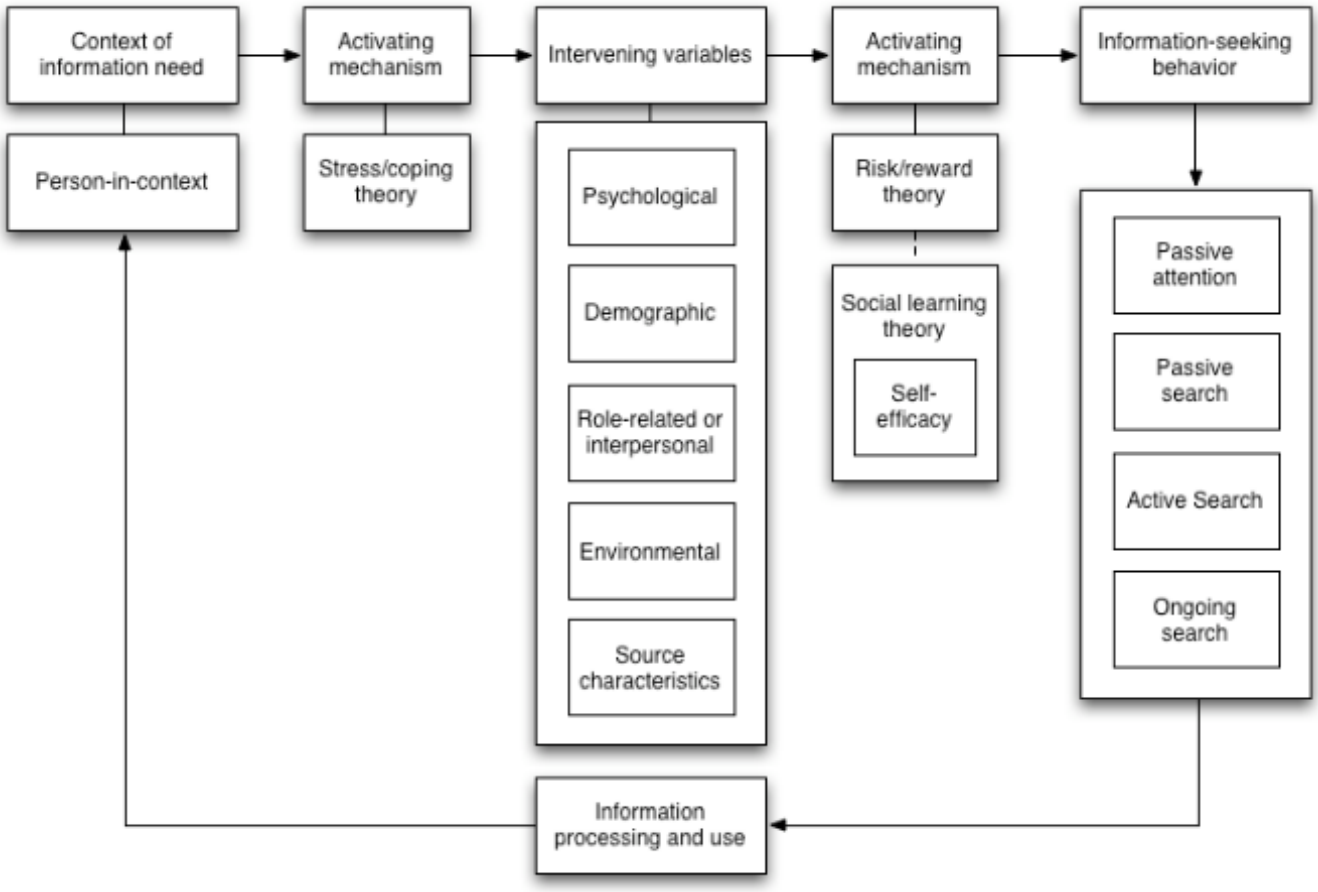


Figure 2.1: Wilson's general model of information behaviour (Wilson, 1997)

Nicholas's model was considered as it addressed the basic assumptions of information needs and provided a useful guide to identifying information needs. Wilson's model, having been adopted by many information-seeking studies, is the most reviewed model and applies to multiple contexts and knowledge domains compared with other models that are narrowly focussed on specific disciplines or tasks (Case, 2006:122). Wilson's model also considers the informal transfer of information between individuals, which is a frequently used or preferred information source in this study (further discussion in 4.2.2.1).

As briefly noted in 1.8, the lack of clear agreement on terms and associated concepts and the diverse subject matter and multiple groups such as journalists, engineers, medical scientists that have used various methods to study information behaviour, have resulted in fragmentation in this research field (Case, 2006:322). Wilson's focus on information processing and use relate to his suggestion that 'information need' is an unsubstantial concept in the information behaviour research process, and

that observable behaviour such as information-seeking and use has a more solid grounding (Wilson, 2005). Wilson is dismissive of Line's work and Nicholas's model and does not recognise either. However, each model has its strengths and weakness. Afzal's (2017:123) criticism of Nicholas is that the framework and the proposed data collection techniques are disconnected. Wilson's focus on the 'use or information-seeking behaviour data', while an important indicator, offers an incomplete view of need due to the dependency on the provision of information (Line, 1969, Nicholas & Martin, 1997, Nicholas & Herman, 2009). However, to gain a complete understanding of the need people have for information in this study, information needs, information-seeking behaviour and use are evaluated as suggested by Nicholas and Herman (2009:17). Both models lack a theoretical foundation underpinning the empirical research, which confused (Afzal 2017).

Other well-known theoretical frameworks that have been applied to information behaviour such as Dervin's sense-making theory (1976), Ellis's information-seeking model (1989), and Kuhlthau's information search process model (1991) are not suitable theoretical frameworks for this study. Their approaches only focused on the information search process and various associated aspects such as feelings, thoughts and actions (Wilson 1994, Ikoja-Odongo & Mostert 2006:149). Each of the information-seeking models listed above has a different intended application and as described by Case (2006:122) and Meyer (2016), many of the models do not consider the various variables associated with information behaviour research.

2.5 Conclusion

This detailed literature review provided a brief introduction to what has been researched by other authors and current research in this field. The literature informed the questions that were asked in the online survey. By combining Nicholas's and Wilson's models, an evaluation framework has been created to analyse the needs of end-users of biodiversity information and their information-seeking patterns and use, and also allows for addressing challenges related to the provision of biodiversity information. This cause and effect combination of information needs, seeking and use as described by Meyer, (2016) affirms the merge of these two models.

Chapter 3: Research design and methodology

This chapter outlines the choice of research design, and the methodologies, including the development and sampling processes that were used in collecting and processing the quantitative data for this study. Survey reliability and validity concerns were also addressed. As the purpose of the study was to obtain the views from the end-users of biodiversity information, a strategic non-random sampling technique which targeted representatives of a community of practice (Botts *et al.,* 2019:1244) that utilised SANBI's online conservation mapping service BGIS was used, with the expectation that each respondent would provide insightful opinions.

3.1 Choice of research design

Julien, Pecoskie and Reed (2011), in their review of articles on information behaviour following two previous studies, determined that survey methods (questionnaires and interviews) are still the most prevalent, with 58.1% for the period 1984-1998 and 44.7% for 1999-2008 (749 articles). Afzal (2017) and Ariño *et al.,* (2013) found that survey methods were commonly used and a successful method for measuring information needs, and provided various reviewed examples of studies where this approach was used.

Based on findings in the literature, and as it was the research method used by Davis *et al*., (2014) and Tann *et al*., (2008) a quantitative research method in the form of a survey questionnaire was found to be most appropriate for this study. The purpose of the questionnaire was to collect data through the gathering of individual opinions to gain a broad overview from a representative sample of the population (purposive sampling method). The survey questions were structured to achieve the objectives of the study, which investigated the information needs, purpose and information-seeking patterns of end-users working with biodiversity data in South Africa and the associated challenges hindering users from achieving their tasks.

A desktop study using related literature from a wide range of sources and documents that were not peer-viewed (technical reports) was completed to gain an understanding of who may potentially be accessing biodiversity information and their needs. These users were then categorised into key end-user groups (Appendix 1). Both of these methods are discussed in more detail in 3.2. A desktop study was done before the questionnaire was distributed, this guided the researcher in determining a

representative sample of potential respondents. The rationale for the data-gathering tools and research instruments is discussed below.

3.2 Methodology

A quantitative approach was used to quantify the behaviours and opinions of respondents. The quantifiable data was used to articulate the facts and reveal patterns (Fink, 2017). The type of quantitative research used (survey research) allowed for the comparison of 3 sub-groups and more accurate findings across a large number of respondents.

Assessing information needs through views expressed via a survey was widely used and highly acceptable in this type of study as seen from the literature. A survey design was chosen as it is a flexible method providing various privacy modes in self-administered surveys, it allows for anonymity (candid opinions) or confidentiality. In terms of the number of recipients in the sample size (5977 email recipients) an online survey was most suitable (Fink, 2017). The questionnaire consisted of mostly closed-ended questions (refer to section 3.2.3 for more detail) resulting in a survey design being more useful. Due to the length of the questionnaire and the rating scale type questions used a survey method was also most appropriate (Fink, 2017). The survey design also ensured that recipients were approached in a standard and consistent way.

3.2.1 Identifying the unit of analysis (end-user groups)

A review of the literature provided insight into the kind of end-users and their information needs and seeking behaviour. This provided a preliminary list of potential end-users and their information needs as reflected in these reports (Appendix 1). A total of 27 key end-users were identified and end-user groups of biodiversity information were categorised in this way. The SANBI Library Catalogue, online information sources, policies and management plans were searched for relevant articles and analysed in a systematic manner drawing out the required information.

Dawson *et al.,* (2004) described a user profile as a process of defining and grouping a user of a digital resource using different attributes. The complexity and huge amount of information contained in existing biodiversity systems made the task of describing and categorising users more difficult. Defining the target population through the review of the literature was important in determining to whom SANBI should and does serve biodiversity information. The list was compared to the registered users of the SANBI online conservation mapping service, and these users were identified as reflective of the potential users of biodiversity information.

3.2.2 Population and sampling

The online conservation mapping service BGIS, noted in Chapter 1, provided a platform for hosting information products used in monitoring and reporting on the state of biodiversity in South Africa. When downloading any information products from the website, compulsory end-user registration provides an email address and occupational sector. A sample population was drawn from the BGIS end-users.

All the registered users of the SANBI online conservation mapping service BGIS were emailed (N=5977) with a description of the project and survey aims and objectives and a request to access a link to an online survey. A total of 778 (13%) responses were collected from 5977 email recipients with an average completion rate of 57%, including 442 complete responses and 336 partial responses. Forty-six percent (355) of the 778 respondents self-identified on the questionnaire. The BGIS service collectively serves spatial biodiversity datasets from data partners to assist planners and decision-makers by providing information on inventories, biodiversity plans, spatial assessment and plans, land cover and ecosystem maps.

Benchmarking against comparative studies, both Tann, Kelly, and Flemons (2008) and Davis *et al*., (2014) sampled from individuals working directly with biodiversity data, including attendees from regional conferences, identified contacts, universities and environmental organisations. Davis had 457 responses collected from 8597 email recipients and Tann received 242 responses from 480 people contacted. Zhi (2014) as cited in Etikan, Musa and Alkassim (2016) stated that the sample size of purposive sampling is related to data saturation which means that increasing the number of responses will not provide anything new that has not already been provided by previous respondents and does not have any statistical implications.

3.2.3 Questionnaire

The quantitative research method used involved the completion of survey questionnaires by SANBI's online conservation mapping service users (BGIS). A pilot literature review identified existing survey questions with the idea of potentially broadening the knowledge of biodiversity information needs and seeking-patterns. Of the 26 survey questions, 18 were adapted from Davis *et al.,* (2014). Six additional questions (Q4 & Q12-16) were added from Tann, Kelly and Flemons (2008), one (Q25) from Jamali and Nicholas (2010), with the final question requesting the contact details of the person undertaking the survey. This facilitated the process as the questions had already been tested and

considered appropriate for use as suggested by Fink (2003:22). Despite Davis, Tann and Jamali not defining or evaluating any theories relevant to their research, their work allowed for comparisons with this study and ensured validity.

The questionnaire consisted of four sections, representing Wilson's hierarchy of information needs, information behaviour as firstly, a descriptor of the overall process, secondly, a goal-directed conscious effort to acquire information, thirdly, the physical or mental interaction with the information, and fourthly, processing and use looking towards the incorporation of the information. The questionnaire is reproduced in Appendix 2.

The disadvantages of a questionnaire are the dependence on the clarity of language, and if the respondent understands what is being asked (Fink, 2017:26). However, a questionnaire was found to be the most convenient due to the large geographical reach and its cost-effectiveness due to the size of the sample. Both qualitative and quantitative data were collected using the online questionnaire. The questionnaire contained four sections, each investigating different aspects: i) Respondents' primary work sectors and subject disciplines, ii) biodiversity information needs, iii) information-seeking patterns and factors influencing these patterns and iv) biodiversity information relevance and purpose.

Sub-questions	No. of questions	Information collected (what do the questions measure?)
1. Who is accessing SANBI's biodiversity information online?	3 (1-3)	1. Primary work sector and roles associated with biodiversity information; 2. Respondent's areas of work, to help in explaining the results of the survey (subject discipline); 3. Type of focus area sought in the different respondent's roles;
2. What are the biodiversity information needs of SANBI's users?	8 (4-11)	4. Biodiversity data currently used by respondents; 5. The necessity for raw primary data; 6. The necessity for synthesised/summarised secondary data; 7. The geographical scale needed; 8. Relevancy of biodiversity information to complete work; 9. Importance of data models; 10. Information gaps (data or information not found); 11. Types of challenges faced when information is not available;
3. What do users do with SANBI information once retrieved?	4 (12 -15)	12. Task scenarios (discover common tasks expected from the target user community); 13. Steps of the task describing the type of need; 14. Description of the type of data or information needed; 15. Source(s) consulted or might be used;
4. What are the information-seeking strategies of SANBI's online users?	5 (16-20)	16. Preferred information sources; 17. Online data/information sources consulted; 18. Ease and/or difficulty with finding information; 19. Types of challenges faced when looking for information needed; 20. Attitude statements to measure respondents experience in finding information (support, satisfaction and challenges);

5. What challenges do SANBI's online users encounter when trying to find information?	3 (21-23)	21. Ease of finding information tools; 22. Importance of tools when working with biodiversity information; 23. Importance of determining attributes of preferred information sources;
6. What information is needed from other disciplines when addressing environmental problems?	2 (24-25)	24. Types of data needed when addressing environmental problems; 25. Frequency of using interdisciplinary information.
7. User profile question	1 (26)	26. Contact details of the person taking the survey.

Table 3.1: Survey outline for investigating biodiversity end users' information needs, information-seeking behaviour and relevance and purpose of information

The survey instrument was pre-tested with a subsample of biodiversity end users and staff members from SANBI (due to resource constraints only 6 members participated). This pre-test group checked content validity, unclear items or questions and in some cases provided additional response categories. Pre-test respondents were not included in the sample group and the finalised survey was made available on SurveyMonkey between 19 June and 16 November 2018. Invitations to complete the survey were sent to 5977 of SANBI's online users via email. The reason for the long period granted to recipients to respond to the survey was due to the poor response rate and the time needed to recruit responses and send out reminders. Chapter four presents the outcomes of the survey with a short discussion relevant to each result and, where possible, critical comparison with similar studies.

3.2.4 Method of data analysis

Descriptive statistics and frequencies were used to summarise, analyse and describe measures of the sample. The data collection, coding and analysis of the data followed Babbie and Mouton's (2001:409) data analysis method. Through a process of discovery, important categories, patterns (emerging themes and topics) and relationships were identified from the survey data. The content received from respondents was translated (coded) from quantitative descriptions into numerical responses for mathematical calculation. Pre-established coding schemes were used to code question 10 (list of responses described by Davis *et al.,* 2014:694) and question 12 (major tasks described by Tann *et al.,* 2008:19) for comparison purposes. A structure for investigating information needs developed by Line (1969) and adopted by Nicholas (2000) was considered as a starting point for analysing and evaluating the data.

The survey had eight open-ended questions that enabled respondents to provide their own answers, resulting in non-numerical responses. These questions were coded before analysis to reduce the wide variety of responses to a set of attributes. All other open-ended questions were coded according to the different dimensions the responses reflected and derived from the research question's purpose. A coding sheet was developed, explaining and describing the variables. A Respondent ID was assigned to each respondent's record and all identifiable information was deleted to ensure data was treated confidentially. A unit of analysis for each question was determined, depending on the context and content.

Descriptive statistics was chosen as the method of data analysis as it best presents quantitative descriptions, and was used to summarise and describe single variables or the associations that connect more than one variable as suggested by Babbie and Mouton (2001:459). No predictions or inferences were made regarding the population parameters. The analysis tool provided by SurveyMonkey's online survey offered a very basic summary of results that were difficult to modify.

3.3 Conclusion

This chapter provided the data collection methods and data analysis used for the study. A purposive sampling design was most appropriate, following the determination of a list of potential end-users. As respondents were selected from previously registered users of SANBI's online conservation mapping service, the sampling method was also self-selected.

Chapter 4: Results and discussion

This chapter focuses on understanding the context in which the need for information may develop, and in this way defines the population being studied and the role of information in the end-users work-role or social setting. This is followed by the survey results and a brief discussion against Nicholas's structure for analysing information needs. The analysis is grouped according to the research questions as listed in Table 3.1 which shows which research question was addressed by what questionnaire question.

4.1 Identifying those who need the provided information

The information needs characteristic of 'intellectual level' and 'level of complexity' are defined by Nicholas (2000:54) as "the minimum knowledge and training the end-user requires" and "how abstract or compressed the information is" respectively. Biodiversity information consists of various interconnected levels of complexity ranging from species to communities and ecosystems, which link to genetics and landscapes. Understanding the information people may need, how they perform a task, and the way they think about concepts will help deliver an information product in a way that is more effective (Spencer, 2011:42).

Wilson's model (1997) proposed that information behaviour is influenced by contextual factors specific to the individual (refer to section 2.3) and the context in which the information is used, and he describes these factors as intervening variables. These elements influence and shape a user's perspective as they progress through the information-seeking process, and can include variables such as personal (psychological and demographic); role-related or interpersonal (researchers, practitioners and policy-makers); environmental (economic level); and information source characteristics (accessibility and credibility). These variables are defined as 'causative factors' which influence information-seeking behaviour. Wilson (1997) explains that these elements may also represent barriers to reaching information needs, although each of these may change over time through the course of the information-seeking process.

Similarly as discussed in section 2.4, Nicholas refers to 'demographic factors' that he also considers potential barriers to meeting information needs (Nicholas & Martin, 1997). Although these elements were not factored into the survey due to the scope of the project, it is with these in mind that the characteristics of the biodiversity end-users need to be considered.

4.1.1 Characteristics of end-users of biodiversity information and end-user groups

This section of the study explores the intended user community to identify primary user groups in order to become familiar with who is accessing biodiversity-related information. A summary of the characteristics of the end-users is then provided which in turn helps articulate why biodiversity information needs exist. A literature review investigated what is already known about potential users of biodiversity information and the information they may require (refer to section 3.2.1). Taking into account that the end-user forms a dynamic component in the system, Spencer (2011:22) explained the importance of understanding the target audience to support the ordering and grouping of website content in a way that makes sense. She also illustrated the importance of understanding the content (what information is available and how to deliver it to the user), and context (how people use information and relevance for the purpose and/or significance to the needs of users).

4.1.2 Context of information needs

The context of information needs is described by Wilson (1981) as the motives and purposes of seeking information and focuses on the individual that needs specific information (person-in-context). An individual's context and viewpoints determine their perception during information-seeking, which influences the choice of information sources and the meaning that is derived (Case, 2005:115). This study sought to establish who the primary users of SANBI's online conservation mapping service are in order to provide an understanding of the context of why and then how users find information. The intention of defining the categories of users was to reduce preconceived assumptions about who is accessing biodiversity information. The end-user groups listed below are described by taking account of the context from which the end-users derive their needs and by identifying tasks that may be completed by different user groups:

Researchers (e.g. scientists): their key goal or task is building or expanding knowledge through scientific inquiry and their output is often literature-based. These are studies undertaken by universities, museums, herbaria, zoos and others (Fornwall, 2000). They tend to use raw data for analytical work (Despot-Belmonte *et al.,* 2017:5).

Biodiversity practitioners: this user group utilises the information to generate products which they proceed to implement (e.g. conservationists, reserve managers, environmental analysts or Environmental Assessment Practitioners (EAPs)). This group of users applies their knowledge to practical tasks, which may include ecosystem-based management of resources through landscape

conservation planning and priority mapping. By integrating the best available information, they develop biodiversity planning products such as environmental management plans or provincial biodiversity plans. In this way, Critical Biodiversity Areas (CBA) and Ecological Support Areas (ESA) are identified and these plans provide land-use management recommendations and it is therefore vital that practitioners have access to spatial biodiversity assessment and planning information.

Policy-makers (e.g. national government officials): they are generally characterised by officials in relevant municipal, provincial and national departments. Their tasks include spatial and development planning (e.g. urban and regional Spatial Development Frameworks, Integrated Development Plans), environmental assessment authorisation and decision making. Processes such as policy-development and reporting are participatory, resulting in the need for authoritative, current, high-quality information (WCMC, 1998:8, Mostert & Ocholla, 2005). This group prefer products from digested processed raw data for policy-reporting and assessments (Despot-Belmonte *et al.,* 2017:5).

The section above provides the context in which the information needs and behaviour were investigated; this includes characteristics and user situations, and motives for seeking information. However, the reality in South Africa is that many end-users have multiple roles that overlap the end-user groups identified in this study, hence the information disseminated is more complex in this context. This diversity of science, policy and practice settings in which information is sought and applied frames the scope in which information needs to be provided.

The differences between these categories are significant and this adds to the challenge of investigating the needs associated with each role. The section that follows is a discussion of the survey findings and the reader should consider the parameters described above in the description of the sample population.

4.1.3 Primary work sector of the end-user groups

The survey needed to ascertain the user roles of respondents in order to customise the analysis and recommendations for the prospective new system. Three required questions at the start of the survey asked respondents about their primary work sector (professional position), subject discipline and biodiversity-related work. The results of the three questions were combined to provide various user profiles. Davis *et al.,* 2014 often used primary work sector as a unit of analysis, but based on the experience of this researcher, the role of the users does not always correlate with the primary work sector. It was considered that the role or function of an end-user would provide more insight than merely providing the work sector. Therefore a schema of key end-user groups or roles (researcher,

policy-maker and practitioner) was adopted and used as the unit of analysis in this study. Information needs arise out of the roles individuals fill or perform (e.g. parliamentarian, environmental assessment practitioners, and urban planners) and these work-roles must be understood.

4.1.4 Description of key end-user groups

Nicholas' characteristic of viewpoint relates to the subject orientation, approach or angle of the information requirements. He describes viewpoint as the various approaches to writing up or presenting information using categories such as a school of thought (e.g. climate change versus habitat loss as a dominant driver of species extinction), political orientation, discipline orientation and positive or negative approaches (Nicholas, 2000:56). Biodiversity end users in this study are groups of individuals who share similar sets of concerns and interests; but have varied views and requirements (Fornwall, 2000). The end-users are often from diverse backgrounds and different disciplines, although interconnected either by a common concern for the environment or its economic value (e.g. tourism, environmental consultant) (White and Molina, 2006:87). Although determining the viewpoint from which a piece of information was compiled is not always clear or possible, it is important to note that it might render the information unusable (Nicholas & Herman, 2009:62).

With an awareness of the biodiversity end-user context, it is important to understand the relevance of biodiversity information and consider to what extent these different end-user groups require different types of information. Potential biodiversity users of biodiversity information from the literature (Appendix 1) have been subsumed within the researcher, policy-maker and practitioner roles. Researchers include scientists, outdoor enthusiasts, graduate students and the general public (citizen scientists). The information needs of those engaged in research include checklists, species descriptions and distribution, identification keys and published research.

Policy-makers include researchers within governments or government agencies, and their information needs focus on the overall status of biodiversity, concentrating rather on whether decisions are conserving biodiversity or leading to its degradation and loss. Practitioners include ecological/environmental consultants, ecologists, agricultural scientists, conservation and spatial planning experts. Their focus is on summarising the facts and reporting on them, usually in a temporal sequence in order to determine trends. Their information needs include species distribution, lists of threatened species and ecosystems at a site-level evaluation.

4.1.5 Characteristics of respondents (Q1-3)

Using the broad definitions of each end-user group as described above, this section of the survey identifies which end-user groups have the greatest interest in biodiversity information. The results cover the primary work sectors and subject disciplines of the respondents and thus describe the sample population. Respondents were asked to indicate their primary work sector, subject discipline and the role that they play in their biodiversity-related work. A "prefer not to answer" option was provided to ensure respondents did not give valueless data in trying to move onto the next question or to avoid violating ethical principles (Fink, 2003).

4.1.5.1 Primary work sector and subject disciplines (Q1 and Q2)

The three end-user group schema (Appendix 1) had not been finalised when the survey was deployed. It became obvious that the three sub-groupings that emerged in the data were also reflected when cross-referenced in the literature (refer to section 2.1). While not deliberate at the outset, asking detailed questions allowed for the schema/framework to be applied and it also emerged as a useful unit of analysis in understanding the end-user group needs concerning respondent's roles. The primary work sector results (Question 1) were used to group respondents into researchers, policy-makers and practitioners. Where the work sector results were unclear, the subject discipline was used to identify the appropriate end-user group. Of the survey respondents, 60% (n=469) were researchers, 4% (n=29) were policy-makers and 36% (n=280) were practitioners (refer to section 4.1.4 for a description of key user groups).

Subject discipline	Researchers (n=469)	Policy-makers (n=29)	Practitioners (n=280)	Number of respondents (n=778)
Life Sciences	31%	17%	13%	24%
Physical Sciences	14%	7%	9%	12%
Environmental Sciences	80%	52%	70%	75%
Agriculture and natural resources	31%	38%	26%	30%
Professional / technical field	22%	45%	50%	33%
Other Environmental Law	0%	3%	0%	0%
Other Planning	1%	0%	2%	1%

* Multiple response question

Table 4.1: Primary work sectors and subject disciplines

Respondents were analysed in relation to their overall response and their key end-user group. Question 2 requested respondents to indicate their field of work, providing as many subject

disciplines as appropriate. As expected, 75% (n=584) of the respondents indicated environmental sciences as one of the broad disciplines in which they work. One-third (n=259) of all respondents selected a professional/technical field while 30% (n=230) selected the subject discipline of "agriculture and natural resources". Definitions of the subject disciplines were provided with the survey questions (Appendix 2).

Amongst researchers, 80% (n=374) of the respondents indicated environmental sciences, while an equal proportion of 31% (n=146) each specified life sciences and agriculture and natural resources as their field of work. Policy-makers have a very similar order in relation to their subject discipline with 52% (n=15) selecting environmental sciences, followed by 45% (n=13) selecting the professional/technical field and 38% (11) agriculture and natural resources. Of the respondents from the practitioner's user group, 70% (n=195) selected environmental sciences, 50% (n=141) selected the professional/technical field and 26% (n=73) agriculture and natural resources as their discipline.

In a similar study published in the United States in 2014 by Davis, comparing respondent characteristics, 59% (n=132) were academic, 22% (n=49) were government and 42% (n=42) were from other work sectors which include environmental non-profits, land trusts, and/or consulting firms. The majority of Davis's respondents listed life sciences (51%, n=114), or agricultural and natural resources (25%, n=55) as their primary subject discipline.

4.1.5.2 Primary role with respect to biodiversity information (Q3)

Respondents were asked to indicate the various biodiversity-related work they performed in their roles (Question 3). The question allowed respondents to tick as many activities as applicable from a list of biodiversity work. Additionally, a further role, Environmental Management, was added after analysing the frequency in the option "Other (please specify)".

Biodiversity-related work	Researchers (n=469)	Policy-makers (n=29)	Practitioners (n=280)	Number of respondents (n=778)
Research	90%	52%	16%	62%
Education	49%	17%	7%	33%
Information management	28%	59%	28%	29%
Land conservation	52%	45%	43%	48%
Marine conservation	12%	28%	7%	11%
Coastal conservation	15%	28%	13%	15%
Estuarine conservation	13%	24%	11%	13%
Inland aquatic conservation	23%	24%	20%	22%
Biodiversity related programme manager	16%	14%	10%	14%
Biodiversity related programme implementer	18%	21%	14%	17%
Elected, appointed and/or organisational decision-maker	7%	10%	6%	7%
Consultant	28%	10%	58%	38%
Engineer	3%	7%	6%	5%
Environmental Management*	1%	0%	5%	2%

* Multiple response question

Table 4.2: Roles with respect to biodiversity information

Researchers indicated research (90%) as one of the frequently mentioned activities and also cited land conservation and education in large proportions. Amongst policy-makers, the frequently indicated activity was information management (59%) with substantial proportions indicating research (52%) and then land conservation (45%). Consultancy (58%) work was the most frequently chosen activity for practitioners, followed by land conservation (43%) and information management (28%). Similar to Davis's study, respondents have numerous roles related to their biodiversity work and in both studies, the majority of respondents indicated research as their primary role (50.2%, n=111).

4.2 Survey method results

As the context and viewpoints of end-users of biodiversity were broadly defined above, the goal was to drill into the detail. The reason for the comparison between registered BGIS Users (sample population) and survey respondent's end-user group profiles was to show the relative representativity of the respondents against an existing well-utilised information source sample population, thereby illustrating potential biases, gaps and uncertainties in the survey results.

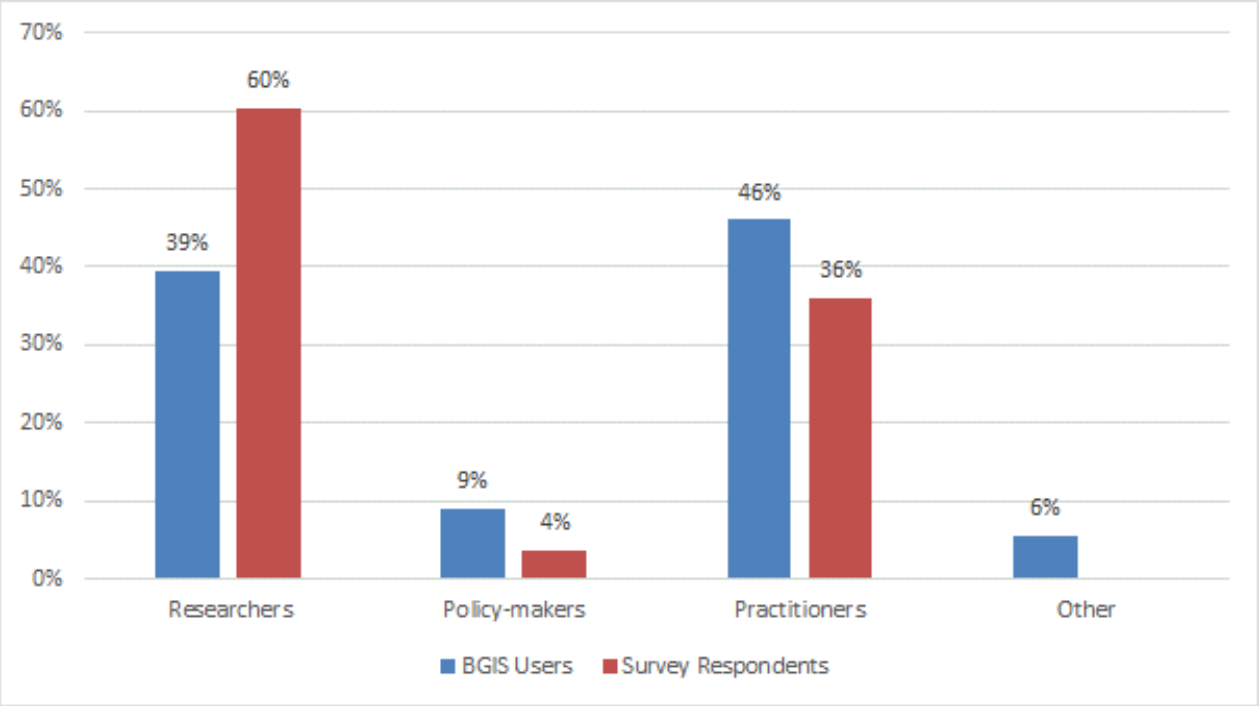


Figure 4.1: Comparison of registered BGIS Users (sample population) and survey respondent's end-user group profiles

In a comparison of the profile of user groups of registered BGIS Users and Survey Respondents, findings show that there were more researchers (60% of respondents versus 39% of BGIS Users) and a concomitant under-representation of practitioners (36% of respondents versus 46% of BGIS Users) and policy-makers (4% of respondents versus 9% of BGIS Users). About 6% of BGIS Users indicated 'other' in their occupational sector, although these users were included in this comparison, none of the survey respondents indicated "other" in the survey. The bias was removed by reporting according to the various end-user groups for the major questions in the survey.

4.2.1 Biodiversity information needs (Q4-10)

In order to design a system which ensures that information is useful and meets the information and decision support needs of different user groups, it was important to identify, assess and analyse the information needs of the users of SANBI's online conservation mapping service. This section explored the needs of the ultimate user, the user who needs access to the information to achieve their task or goal. These tasks often address questions related to the use and/or conservation of natural resources and tracking changes. To do this it is necessary to understand what information they need (Hardisty and Roberts, 2013).

Referencing Table 3.1, seven questions (Q4 – Q10) and two questions on interdisciplinarity (Q24 and Q25) were used to investigate the “information needs” of users. This section of the survey looked at what type of content users find most useful in their different roles and identifies information gaps and challenges experienced with not being able to find the information needed.

4.2.1.1 Data categories sought in the different respondent's roles (Q4)

The ‘nature of information’ is an important concept considered by Nicholas (2000:53) when investigating information needs. According to Nicholas (2000:53), there are four forms in which information is classified: i) conceptual/theoretical, ii) historical, iii) descriptive, and iv) statistical or methodological. Whilst these forms are utilised by the end-users in a manner in which the information was intended, information seekers are often unaware of the nature of the information type being sought (Nicholas & Herman, 2009:51).

Nicholas suggests that practitioners tend to use descriptive, methodological and statistical forms whilst researchers tend to gravitate towards theoretical and historical forms of information (Nicholas, 2000:54). This section discusses the information needs of respondents in relation to various forms, data categories and types.

A list of raw and synthesised data categories was provided to determine what types of content people currently find most useful. Survey respondents were asked to choose up to three data categories that they currently use. The distinguishing criteria between raw data and synthesised information types are that raw data has not been processed (uninterpreted) and therefore there is no potential for analytical errors (Thessen & Patterson, 2011:18, Nicholas & Herman, 2009:54).

	Data Category	Researchers (n=469)	Policy-makers (n=29)	Practitioners (n=280)	Number of respondents (n=778)
Raw primary data	Animal and plant descriptions	44%	21%	34%	**40%**
	Species checklists	41%	17%	33%	**37%**
	Maps and distribution data	64%	69%	63%	**64%**
	Flora and fauna surveys	39%	24%	39%	**38%**
	Citizen Science	24%	17%	16%	**21%**
	Images and videos	24%	21%	16%	**21%**
	Identification keys	32%	21%	22%	**28%**
	Ecological data	58%	52%	61%	**59%**
	Invasive species data	39%	45%	35%	**38%**
	Gene sequences	7%	0%	1%	**5%**
Summarised	Threatened species data	45%	38%	49%	**46%**
	Forecasted trends, possible scenarios, speculations, predictions	22%	17%	13%	**18%**
	Planning/prioritisation outputs	41%	55%	56%	**47%**

* Multiple response question

Table 4.3: Data categories currently used by survey respondents

Thirteen data categories (adapted from Tann *et al.,* 2008) were provided for respondents to indicate current usage. These categories included research, policy-related and planning types of information. As can be seen in Table 4.3, approximately two-thirds (64% or 498 respondents) of all respondents selected maps and distribution data, 59% (n=459) ecological data and 47% (n=367) and 46% (n=360) selected planning/prioritisation outputs and threatened species data respectively.

Amongst researchers, maps and distribution data (64%), ecological data (58%) and threatened species data (45%) were cited most frequently. Practitioners, who make up 36% of the total respondents, cited maps and distribution data (63%) and ecological data (61%) most frequently. The 29 policy-makers cited maps and distribution data (69%) and planning/prioritisation outputs (55%) most frequently. Gene sequencing data was the least cited data category overall (5%) and the least cited amongst each of the roles. Similar to Tann *et al.*'s, (2008) report, spatial datasets (e.g. maps, species distributions, geographical ranges and vegetation distributions) are the most used data categories among biodiversity end-users (72%, n=457).

4.2.1.2 The necessity for biodiversity information types (Q5-Q9)

Biodiversity datasets across broad temporal, geographic and taxonomic scales are required to answer biodiversity questions. These heterogeneous and highly distributed datasets create a challenge with data integration which Hoffmann *et al.,* (2014:60) suggested can be addressed by analytical methods

and visualisation tools. The aim of the section below was to identify the types of information needed and their importance to biodiversity-related work and how this specifically should be targeted in the new biodiversity information-system. In the next five survey questions, the respondents were requested to consider the relevance of biodiversity information, the importance in the type of data needed to do biodiversity-related work (raw primary biodiversity data, synthesised/summarised secondary and data models) and geographical scale of biodiversity information needed.

4.2.1.2.1 **Importance in the type of data needed to do biodiversity-related work (Qs 5-6)**

Nicholas (2000:84) describes processing as the "different ways the same ideas and data can be represented". This is closely associated with packaging which he describes as the "external presentation or physical form of the information" (Nicholas, 2000:87). Information can be presented in numerous ways, such as a collection of raw data, or as a concise summary (Line, 1969:11). As described in section 1.6, raw data does not convey information (which has meaning and context) and therefore it is necessary to process (with associated measures of uncertainty) the data into information (secondary data) making it more relevant and accessible.

Based on the concept of processing, the questionnaire differentiates between the need for i) raw primary data (which are data that have not been subjected to analysis), ii) synthesised/summarised secondary data, and iii) modelled data. Survey questions 5 and 6 asked respondents on separate sliding scales how much of the two different types of data (raw and synthesised/summarised secondary or data models) are needed to do their work. Synthesised/summarised and modelled data were combined as both are processed data where there has been analysis, recombination, or modification of raw data (Thessen & Patterson, 2011:18). Respondents chose a position on a numeric sliding scale for both questions. The scale ranged from 1 to 5, where 1 = none of these data types is needed, 3 = half and 5 = all, with the instrument generating a percentage value.

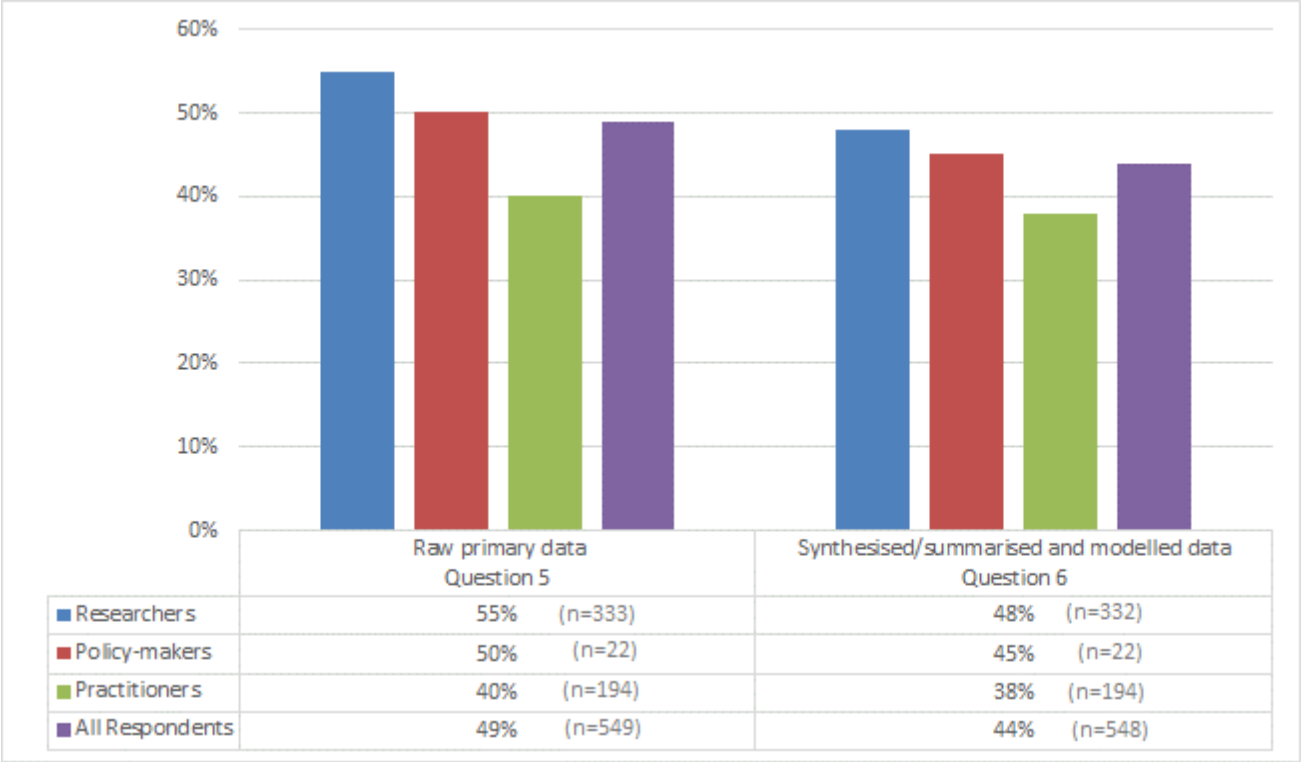


Figure 4.2: Percentage of respondents reporting whether raw primary biodiversity data versus synthesised/summarised and modelled data are necessary to complete more than 50% to do their work

From Figure 4.2, 49% (n=549) of respondents who provided a position on the sliding scale said that at least half of their work is based on raw primary biodiversity data. In comparison, of the 548 respondents that provided a position, 44% said that at least half of their work is based on synthesised/summarised secondary biodiversity data. Similarly, across all end-user groups, the proportion requiring raw data is higher than synthesised/summarised data, although in higher proportions for researchers (55% vs 48%), than for policy-makers (50% vs 45%), and practitioners (40% vs 38%).

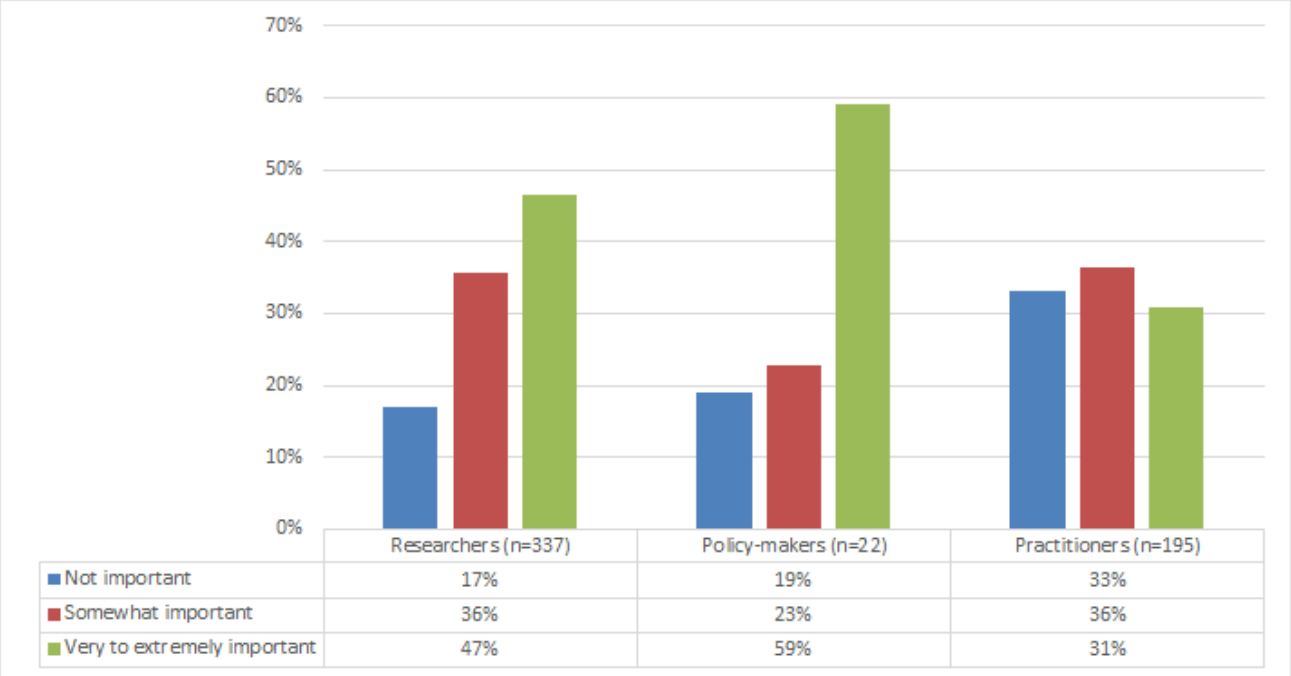
A comparison of results with Davis *et al.,* (2014:694) is difficult in this case as the authors asked three separate questions. However, results showed that both raw and summarised information is important or needed in nearly equal proportions (67.3%, n=217 vs 65.1%, n=218) and data models are slightly less important (52.1%, n=219).

The results of scientific research are often not used effectively in environmental planning and decision-making (Sinclair, Mozzotti & Graham, 2003). Houtkamp *et al.,* (2016) suggests that this is due to research projects not generally being designed to provide specific answers to environmental

questions and therefore do not match decision-makers' information needs, therefore the nearly equal importance of raw primary data for policy-makers and researchers.

4.2.1.2.2 **Importance of data models to biodiversity-related work (Q9)**

The study investigated the importance of data models, defined in the questionnaire as a "process of mathematically linking biodiversity variables" according to Ferrier, Jetz and Scharleman (2017), in relation to biodiversity-related work. The respondents were asked to indicate the relative importance on a Likert scale, from not at all important, not so important, somewhat important, very important and extremely important.



	Researchers (n=337)	Policy-makers (n=22)	Practitioners (n=195)
Not important	17%	19%	33%
Somewhat important	36%	23%	36%
Very to extremely important	47%	59%	31%

Figure 4.3: Percentage of respondents reporting the relative importance of data models to biodiversity-related work

Note: 'Not at all important' and 'not so important' were combined to create 'not important' in Figure 4.3 and 'very important' and 'extremely important' were combined to form 'very to extremely important'.

According to the total respondents (n=554), the importance of data models is 'very to extremely important' (42%) to their biodiversity-related work. All three patterns of responses are very different for each end-user group. The highest incidence of 'very to extremely important' was recorded amongst policy-makers (59%). For researchers, the proportion is slightly lower (47%), with the relative importance of data models being the lowest for practitioners (31%). This is expected as typically

practitioners use raw data to generate their own models. For respondents of this survey, the relative importance of data models to biodiversity-related work is slightly less important 42% (n=554) than 52.1% (n=219) of overall responses found in Davis *et al.,* (2014:694).

4.2.1.2.3 **Geographical scale (Q7)**

The concept of "place" or more specifically "place of publication" is defined by Nicholas as the place or country of origin of the information. The concept of place according to Nicholas (2000:80) is dependent on three aspects, i) subject, ii) language proficiency, and iii) whether the users are practitioners or academics.

Biodiversity information can be required and/or provided at a range of spatial resolutions (Pocock, 2018:17). It was therefore critical to understand the preferred spatial scale needed by respondents, as this makes the data more relevant and useful. More importantly, each geographical scale has inherent limitations to their use and the exchange of data at or between spatial scales is restricted (Walters & Scholes, 2017:20). Also, biodiversity databases usually have different objectives and data structures at various geographical scales to match the often multi-scaled decision-making process as confirmed by Huang *et al.,* (2012). Respondents were asked to rank the geographical scale of biodiversity data in order of importance. Four geographical scale options were provided, local scale refers to information at the scale of habitats and catchments. Province is information at a provincial scale or within South Africa for national or international for global. Of the respondents, 554 answered the question.

Geographic Scale	Total Score	Average Rank
Local	1123	2.027
Province	1238	2.235
National (SA)	1405	2.536
Global	1740	3.141

Table 4.4: Respondents need for biodiversity data at four different geographic scales

The responses were converted from a ranking position to a point score and multiplied the point score by the number of respondents selecting a ranking position for the geographic scale. This resulted in a total score which was then divided by the total number of responses to provide an average rank. The most frequently top-ranked option was Local (n=525) and the most frequently bottom-ranked option (n=480) was Global.

When compared to the study by Davis *et al.,* (2014:695), on the other hand, results showed that state-level information is the geographical scale most frequently used. This can be seen as the equivalent to a province in South Africa.

4.2.1.2.4 **The relevance of biodiversity information (Q8)**

Nicholas (2000:39) explains that the facet 'subject' is the most noticeable characteristic of an information need (often used as keywords) as it relates to whether the request for information consists of one or a number of subject terms (subject-oriented data collections). The subject description considers three aspects: i) the number of subjects involved; ii) the depth to which these subjects are pursued; iii) the problems associated with the subject being too specific or vague (Nicholas 2000:40). Line (1969) describes this characteristic as the "…subject range of potentially useful information" and therefore Question 8 is discussed in conjunction with Question 24 and 25, as these three survey questions looked at other subject disciplines that need to be integrated with biodiversity information when addressing environmental problems and how often information from other disciplines are sought.

Respondents were asked to choose a position on a sliding scale to state how much of their needed information relates specifically to biodiversity, on a scale that ranged from "none of the information needed" to "at least half of the information" to "all" information needed to do the work.

Figure 4.4: Proportion of user groups where >50% of the information needed to do work relates to biodiversity

Figure 4.4 shows that for 59% of 555 respondents, at least half the information they need to do their work relates specifically to biodiversity and when compared to Davis's (2000:693) it was nearly two-thirds of the respondents (63.2%, n=267). For researchers, the proportion is 60%, for policy-makers 55% and practitioners 58% that indicated that at least half the information they need to do their work relates specifically to biodiversity. The results are very similar for all three end-user groups indicating that these groups are heavily invested in biodiversity information. This more than justifies the investments made in biodiversity-related processes and infrastructure to provide long-term sharing and hosting of biodiversity information for scientists, decision/policy-makers and biodiversity practitioners.

4.2.1.2.5 Interdisciplinary awareness to facilitate data integration (Q24 and Q25)

Although a major part of the information needed in the life sciences discipline is scientific, ecological, geographical and genetic studies and/or inquiries, it often entails interactions with other fields of study. An integrated understanding of other discipline data types and/or information in addressing environmental problems is therefore vital as access to information in various disciplines may be a barrier to getting the job done. Two survey questions were used to determine what adjacent (e.g. climatology, geology, hydrology) and disparate (e.g. economics, sociology) fields of study data/information are currently being combined with biodiversity information and how often this information is sought.

By default, taxonomy and systematics disciplines have a close association due to the use of species names in the classification of biodiversity data. There is, however, a need to understand the significance and impact of the integration of information across disciplines and in this way determine what data integration services are needed or possible. These qualitative comments of what other types of information are needed when addressing environmental problems were categorised into subject disciplines in order to generate frequencies in relation to other types of information needed. Table 4.5 provides a list of information needed for other discipline type information and examples of what respondents use the information for.

Fields of study	Use case descriptions	Frequency	Percentage
Climatology	Climate change model impacts. Climate variation data (historical temperature, rainfall, evapotranspiration, radiation). Predict behavioural changes in the target species. Climate data for a specific local region. To determine the effects of climate on the flora and water systems in South Africa.	145	19%
Socio-political and economic	To determine the living conditions of people in areas surrounding those being studied, as this helps to determine the social issues that	119	15%

Fields of study	Use case descriptions	Frequency	Percentage
	could arise among stakeholders. To indicate the reliance of people on natural resources and demand for ecosystem services.		
Soil Science	Soil types and patterns to assess factors contributing to land degradation and rehabilitation. Need the digital data for Soils of South Africa by Martin Fey.	101	13%
Urban Planning	Planning and design of land-use and land development in the built environment; includes land use which is how the land is being used (e.g. agriculture, mining, urban, etc.); land cover showing physical land types (e.g. open water, forest, etc.); infrastructure (e.g. human, electrical grid, etc.), cadastre. Integrated Development Plans (IDPs); Spatial Development Framework (SDF); State-Owned Enterprises (SoE) Reports.	89	11%
Ecosystem Ecology	NFEPA data – conservation of wetlands to give effect to the Overberg Wetland strategy and action plan.	83	11%
Hydrology	Hydrological data to determine the effects in the change of hydrology on wetland habitats. Hydrological modelling. The synergy between the Department of Water and Sanitation (DWS), Department of Environmental Affairs (DEA) and SANBI datasets is needed.	64	8%
Anthropology	The projected vulnerability of sub-quaternary and local catchments to future water stress, fire and alien invasion. Human-induced environmental issues.	59	8%
Water Resources	Groundwater depth and direction datasets for mitigation and tracking underground pollution threats. Ground and surface water, aquifer. Water use datasets – how people in the area use the water and how important a specific river is to the general public including water-use laws.	54	7%
Geology	Geological data (maps, metadata, descriptions of the geological types of South Africa).	44	6%
Topography	Climate, soils, topography, hydrology, fire (other disturbance) data to understand the role of these factors in determining the distribution of plants.	36	5%
Oceanography	Coastal management lines and flood lines – assess vulnerability in terms of coastal processes. Data on harvesting types and levels in coastal areas. Data on marine tourism activities.	24	3%
Agricultural	Another agriculture census required (the last one was undertaken in 2007). This information helps to understand changes in major land-use practices in South Africa (e.g. area cultivated to different crops; livestock numbers, farm sizes, etc.).	21	3%
Conservation	Delineation of protected areas and sensitive areas to stop development and encroaching of the high water flood line. Access to CBA, ESA, species use (medicinal plant use) and trade data.	21	3%

* Multiple response question

Table 4.5: Information needed from various fields of study and use cases

The list spans a broad sector of different disciplines and often a use case draws from two or more 'fields of study', therefore, the use of the term 'interdisciplinary'. Climatology (19%) was found to be the most frequently requested discipline information type, followed by socio-political and economic (15%); soil science (13%); urban planning (11%); and ecosystem ecology (11%). Eight other fields of

study were mentioned that were cited by fewer than 10% by the respondents. SANBI at present serves a limited number of the discipline type data listed in the table above.

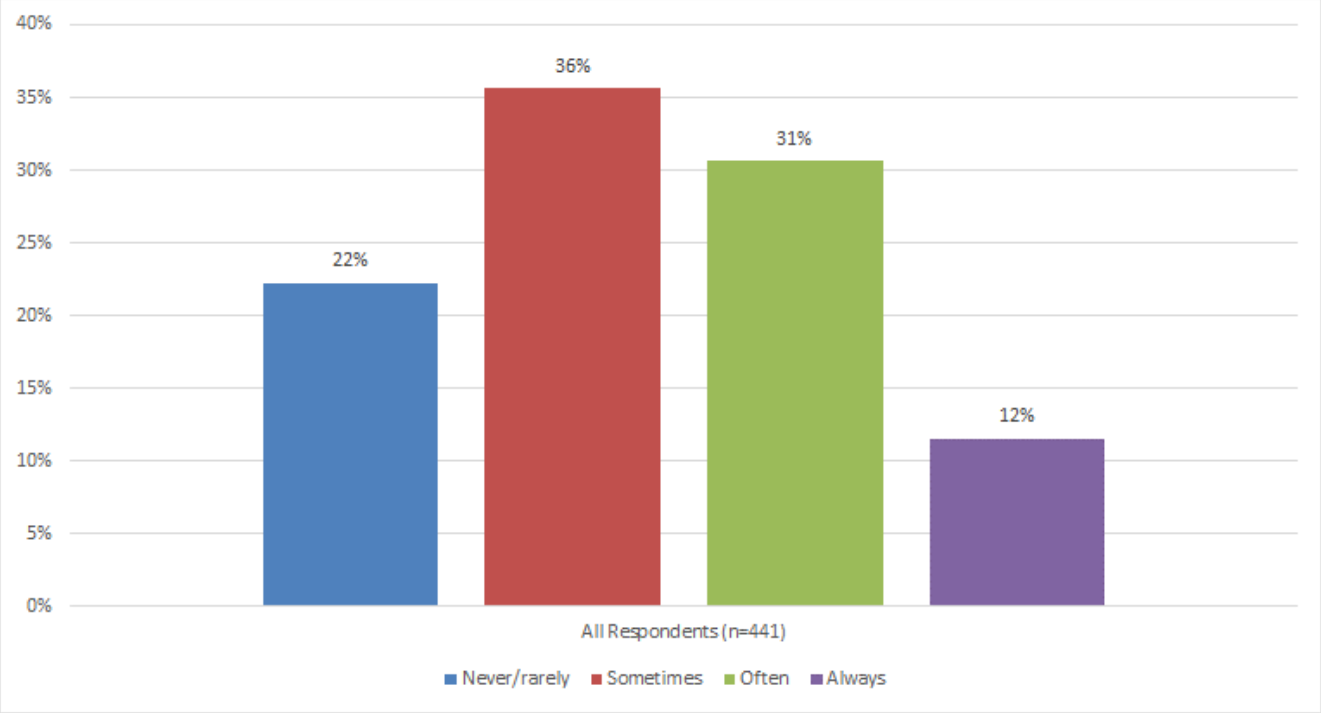
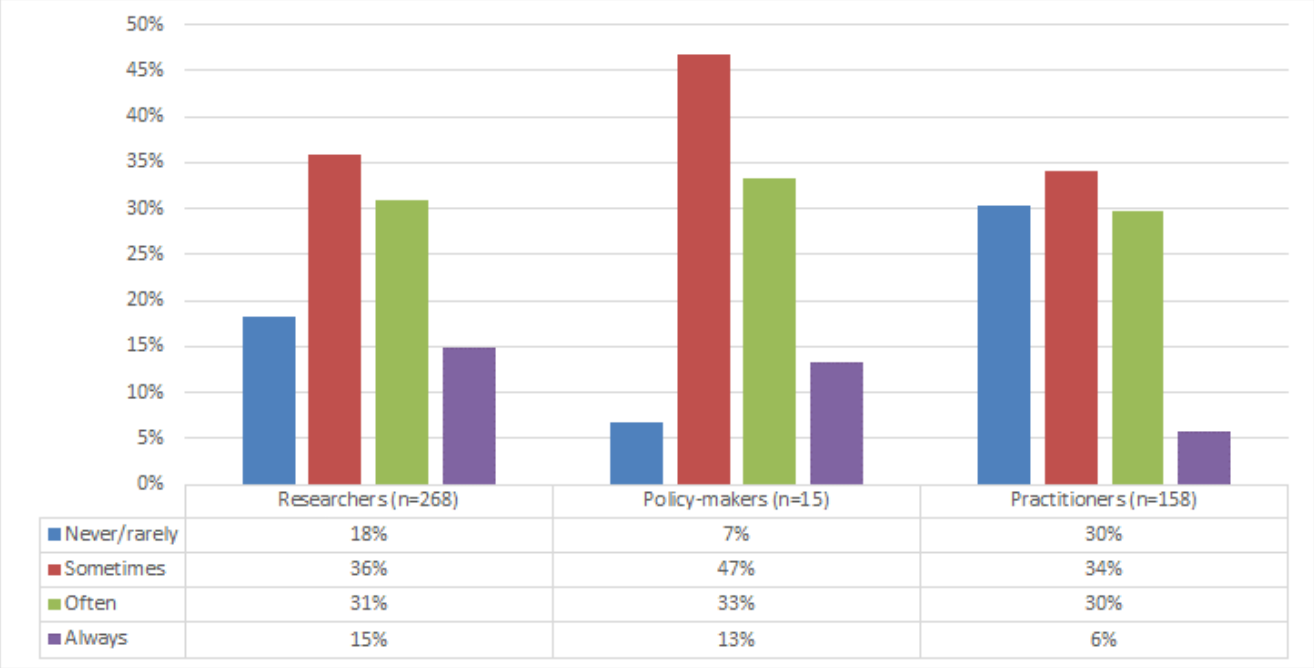


Figure 4.5: Frequency of seeking and utilisation of information from other disciplines for total responses

Respondents were then asked to indicate how often they need to look for and use the results of research by people in other disciplines as suggested by Jamali and Nicholas (2010). In this way, it was possible to estimate how interdisciplinary the work or research of end-users of biodiversity really is. Of the overall respondents, 22% (n=441) specified that they never or rarely rely on information from other fields of study, however, the response patterns from the different end-user groups differed as shown in Figure 4.6.



	Researchers (n=268)	Policy-makers (n=15)	Practitioners (n=158)
Never/rarely	18%	7%	30%
Sometimes	36%	47%	34%
Often	31%	33%	30%
Always	15%	13%	6%

Figure 4.6: Frequency of seeking and utilisation of interdisciplinary information across the end-user groups

Figure 4.6 shows how often end-user groups seek and use information from other fields of research. Results show that policy-makers are more reliant on interdisciplinary information with 47% reporting they 'sometimes' and 33% 'often' use information from other fields. Practitioners indicated that they are less reliant with 30% 'never/rarely' using interdisciplinary information. Researchers show the highest dependence on information from other fields of study, with 15% of researchers always reliant on this type of information.

4.2.1.3 Identified current content gaps (Q10)

Speed of delivery/supply is defined by both Line (1969) and Nicholas (2000) as having the necessary information available at the time of need or knowing which topics are current. It is understood that information priorities are often ranked differently depending on the interests of the user (e.g. loss of indigenous knowledge) or the value and interests attached to the natural resources (e.g. the water crisis in the Western Cape in 2016). These priorities also change over time and it is often difficult to predict the environmental concern for a particular period or predict the length of time a data type will stay topical (WCMC, 1998:3).

Respondents were asked what biodiversity data or information they have not been able to find. Although current systems provide access to biodiversity data, the intention was that this survey question would highlight the inadequacies of existing biodiversity data, including difficulty in access

to these data. By identifying the current needs that exist, which press individuals towards information-seeking behaviour, there will be a better understanding of the use and the ability to design a more effective information system (Wilson, 1981). The following gaps were identified which biodiversity information end-users face when they trying to find information:

Content gap	Frequency	Percentage
Environmental change data	72	22%
Occurrence / distribution data	69	21%
Ecosystem data	52	16%
Specific taxa / species	44	13%
Environmental data	36	11%
Planning / prioritisation outputs	28	8%
Increased access to survey collections	10	3%
Geographic locations	9	3%
Literature	5	2%
Increased access to specimen collections	3	1%
Data associated with published papers	2	1%
Increased access to museum collections	1	0%
Total	**331**	**100%**

Table 4.6: Views of 331 respondents on data or information they have not been able to find

Three hundred and thirty-one responses were provided for the open-ended question soliciting what data or information respondents are unable to find. More than a fifth of the respondents reported an inability to find environmental change data (22%, for example, habitat loss, landscape and climate changes, invasive data and land cover) and occurrence/distribution data defined as the existence of an organism at a particular place at a particular time (21%, for invasive and invertebrates data). Ecosystem data which is landscape-level units comprised of communities and their ecological drivers (16%, such as wetlands and riverine) were also cited as types of data or information that respondents are unable to find.

The list of responses described by Davis *et al.,* (2014:694) that were used to categorise the data in Question 10 included: information concerning specific taxa, species, or geographic locations, occurrence and distribution information, and increased access to historical literature, specimen databases, and museum collections. Although there are similar themes listed in Table 4.6 the content gap is different between the two studies.

4.2.2 Biodiversity information relevance and purpose (Q12-Q15)

Much of SANBI's information system development or adoption in the past has been prototyping, being one of several methods to determine user requirements (Vissers *et al.,* 2017:50). Scenarios of use is another technique and provided a derived understanding of the expectations that respondents currently have of the system. Four survey questions (Q12 - Q15) were used to explore a task analysis of how and for what purpose biodiversity information is currently needed.

4.2.2.1 Common tasks or patterns of use by different user groups (Q12)

Respondents were asked to give an example of a task, process or application where biodiversity data was used to achieve an outcome that was core to their work or study. Respondents were requested to answer the questions in their own words. This type of questioning approach was selected because the user may not be able to communicate their system requirements precisely. They were, however, able to explain their goals and steps in the tasks to achieving their goal (Kujala, 2002:34).

Scenarios of use/tasks were seen as functionality that could be built as one or more components of the system. A predefined list of themes was adopted from Tann *et al.,* (2008) of major tasks and topics of interest. The table below provides a list of the themes, their descriptions, an example of the task given by a respondent and the percentage of cases in which these types of tasks were provided.

Use case	Description	Example of task	Count	Percentage of cases
Site assessment	Determining how appropriate an area is for conservation or development	Assessment of biodiversity value and ecological significance of land portions during environmental impact assessments; this information is used to guide site layouts as well as contributing to recommendations regarding whether or not environmental authorisations should be granted.	168	37%
Distribution analysis	Analysing distribution for a species (occurrence data)	Mapping species ranges to investigate patterns, niche, breadth and geographic range size.	77	17%
Habitat management planning	Managing habitats for conservation (life cycle, distribution and other data)	▪ Compiling management plans for Forest Nature Reserves. Assisting in compiling the Principles, Criteria and Indicators manual used to assess the performance in the management in State Forests. ▪ Developing CBA Maps and Protected Area Expansion Strategies.	42	9%
Invasive species analysis	A special case of 'Distribution analysis' to model the occurrence	Invasive tendencies of Nile Tilapia to be captured in a Biodiversity Risk Assessment for permitting purposes.	40	9%

Use case	Description	Example of task	Count	Percentage of cases
	and spread of invasive species			
Population/habitat monitoring	Tracking the population of an organism or habitat type	Estimating population changes over time in a specified location in relation to local climate and landscape.	34	8%
Threatened species categorisation	Assigning or discovering the categorisation of the conservation status for an organism	Quantifiable data on the occurrence of threatened species within an area to determine if critical habitat thresholds are triggered.	20	4%
Vegetation modelling	Modelling the composition and distribution of vegetation communities	Using relevé (plot) data to model the potential distribution of ecological groups.	15	3%
Identification	Identifying and determining the name for an organism	The scientific naming of plants that are being utilised by rural communities.	12	3%
Site/region checklist	Generate a list of species known to occur at a site or in a region (includes 'Pre-impact checklist')	Compiling a list of plants that are endemic to an area for purposes of indigenous gardening in that particular area and then cross-referencing with what is available from commercial nurseries (advisory to the neighbourhood).	11	2%
Site selection	Identification of priority sites for a given species (closely related to 'Site assessment')	Predicting habitat suitability for a critically endangered cycad.	10	2%
Genetic analysis	A special case of 'Distribution analysis' to map genetic variation across the landscape	Mapping genetic diversity across the landscape and over time.	7	2%
Population/habitat modelling	Modelling the dynamics of the populations of an organism or habitat	Climate change modelling for plant species	5	1%
Pre-impact checklist	A special case of 'Site/region checklist' to document species lists before modification or to infer likely original species list after the damage has occurred	Compiling of expected species lists for areas where development is due to take place.	4	1%
Disease outbreak analysis	Exploring epidemiology of diseases (occurrence data, biology, molecular)	Lack of types of terrapin pathology studied in South Africa creates a problem whereby low identification of pathogens occurs.	4	1%
		Total	**449**	

Table 4.7: Descriptions of the use cases according to Tann *et al.,* (2008) and examples of the tasks provided by respondents

An analysis of the survey responses of the major tasks using biodiversity information was completed using 14 of the 28 categories of major use cases identified by Tann *et al.,* (2008:19). Question 13

(steps of the task), Question 14 (the types of data or information needed) and Question 15 (sources and platforms to find the information to complete the task) were not included in this analysis. The responses from these three questions helped inform 'use cases' and formed the kernel of SANBI's User Requirements Specification documentation.

Site assessment was the leading task and therefore the ability to determine what occurs in a particular area, being able to extract data from literature and query various spatial plans and historic site assessment reports were key to ensuring that these information needs are met (Tann *et al.,* 2008:32).

Tann *et al.,* (2008:28), on the other hand, found that "distribution analysis was the dominant task of their study and this meant that the ability to retrieve information spatially was essential, varying over time, scale and various forms of content".

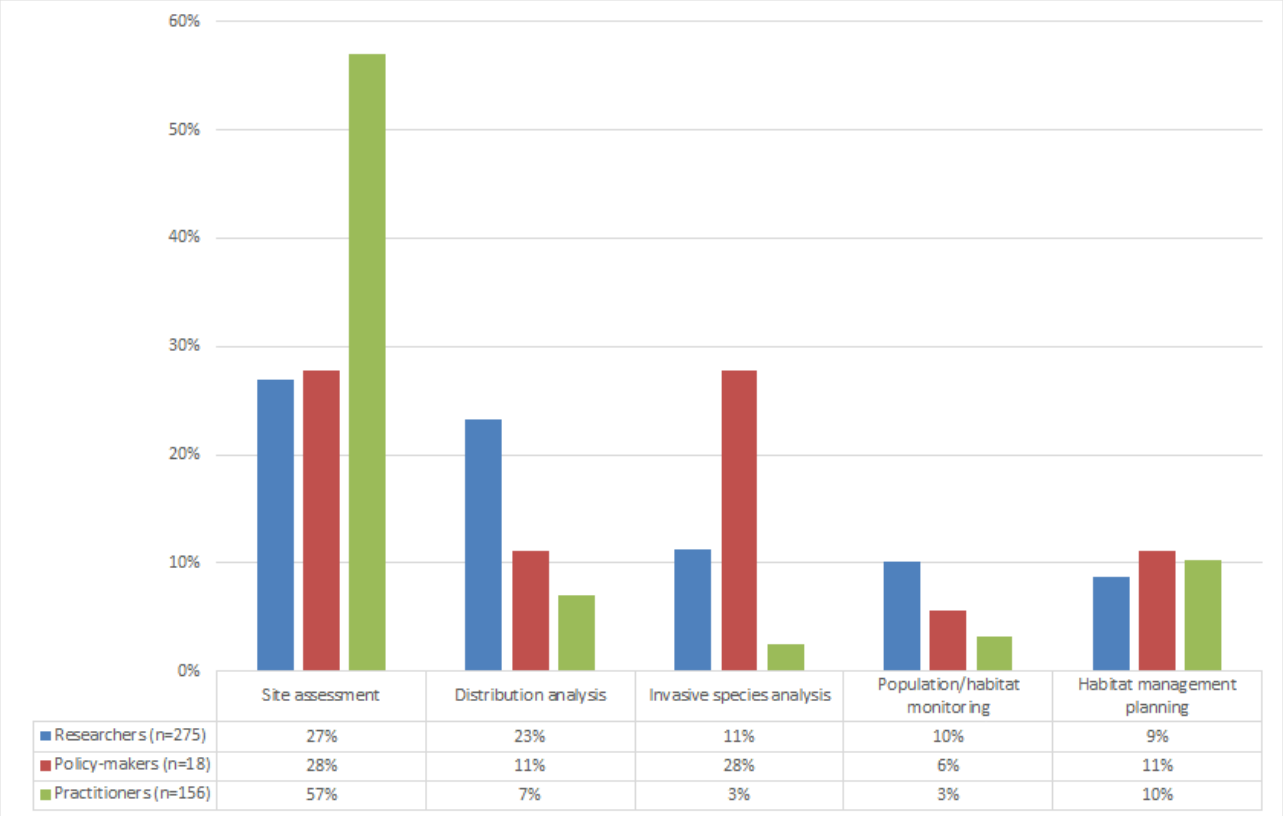


Figure 4.7: Survey responses showing tasks of importance by end-user groups

When the major tasks were analysed by end-user group, 57% of practitioners noted site assessment marginally more than the overall therefore placing it as the leading task. The predominance of the site assessment is thus less marked for researchers (27%) and policy-makers (28%). As mentioned above invasive species analysis is a special case of 'Distribution analysis' which is a need to study the geographical range of species and in this case, is a major task for policy-makers (28%). The negative impact of invasive species on the economy and ecosystems in South Africa currently as described in

the Status of Biological Invasions and their management in South Africa (van Wilgen & Wilson, 2018) is top of the policy-makers agenda.

4.2.3 Information-seeking of end-users of biodiversity information (Q16-Q18 & Q20)

The function concept is defined by Nicholas as the main function for which the information is gathered (Nicholas, 2000:46) and is closely associated with Wilson's four information-seeking behaviour modes (Wilson 1997:562) which describe for what purpose a particular information-seeking act is undertaken. Whether it is actively seeking or gathering information versus keeping up-to-date or non-intentionally finding information as described in Table 2.1.

As information needs are intended to motivate search behaviour, this section of the study examines the information behaviour from the respondent's perspective, as interpreted by the researcher. Four questions in the survey were used to investigate the information-seeking patterns of respondents (questions Q16-Q18 & Q20). This section of the questionnaire was concerned with understanding how respondents seek information, so recognising their preferred information sources (Q16), gaining a sense of the number of online biodiversity resources being used (Q17), and the ease of finding needed information (Q18) and identifying barriers and enablers (Q20).

4.2.3.1 Preferred information sources (Q16)

Line (1969:12) and Nicholas (2000:84) emphasise packaging as the various physical forms in which information is stored or presented. Question 16 determined which information sources respondents preferred. Information sources are not only physical objects such as books or maps as defined by Line (1969), but also digital formats such as catalogues.

Six answer options (listed in Table 4.8) were provided for respondents to rank their preferred information sources. To determine the relationship between the information sources and work sectors and therefore reflecting on Tann *et al.'s,* (2008:17) work, the list of options was adapted and expanded to include existing management plans and personal contacts. In addition, respondents were requested to rank their preferences.

Information sources	Researchers (n=279)	Policy-makers (n=15)	Practitioners (n=163)	Average
Online data/information sources	45%	33%	56%	49%
Literature (journal publications)	23%	7%	13%	19%
Personal contacts	13%	7%	4%	9%
Existing management plans	8%	33%	10%	9%
Local expert opinions or subject specialists	8%	13%	11%	9%
Online library catalogues	4%	7%	6%	5%

Table 4.8: Analysis of survey responses of the top-ranked information sources

Across the end-user groups (n=457) nearly half (49%, n=222) indicated the online data/information sources as their first ranked option as the preferred information sources. One-fifth of the respondents (n=88) listed literature (journal publications) as their top preferred information sources, while another 9% each tended to go to personal contacts (n=43), existing management plans (n=42) and local expert opinions or subject specialists (n=41) as their number one preferred information sources. Only 5% (n=21) indicated online library catalogues as their preferred information sources.

This is to some extent consistent with Zipf's (1949) Principle of Least Effort (Case, 2005), the human tendency towards ease of access to information. With library catalogues, the burden is on the user to acquire the item from the library and so end-users are reluctant to use library resources and prefer the convenience of online information (speed of delivery).

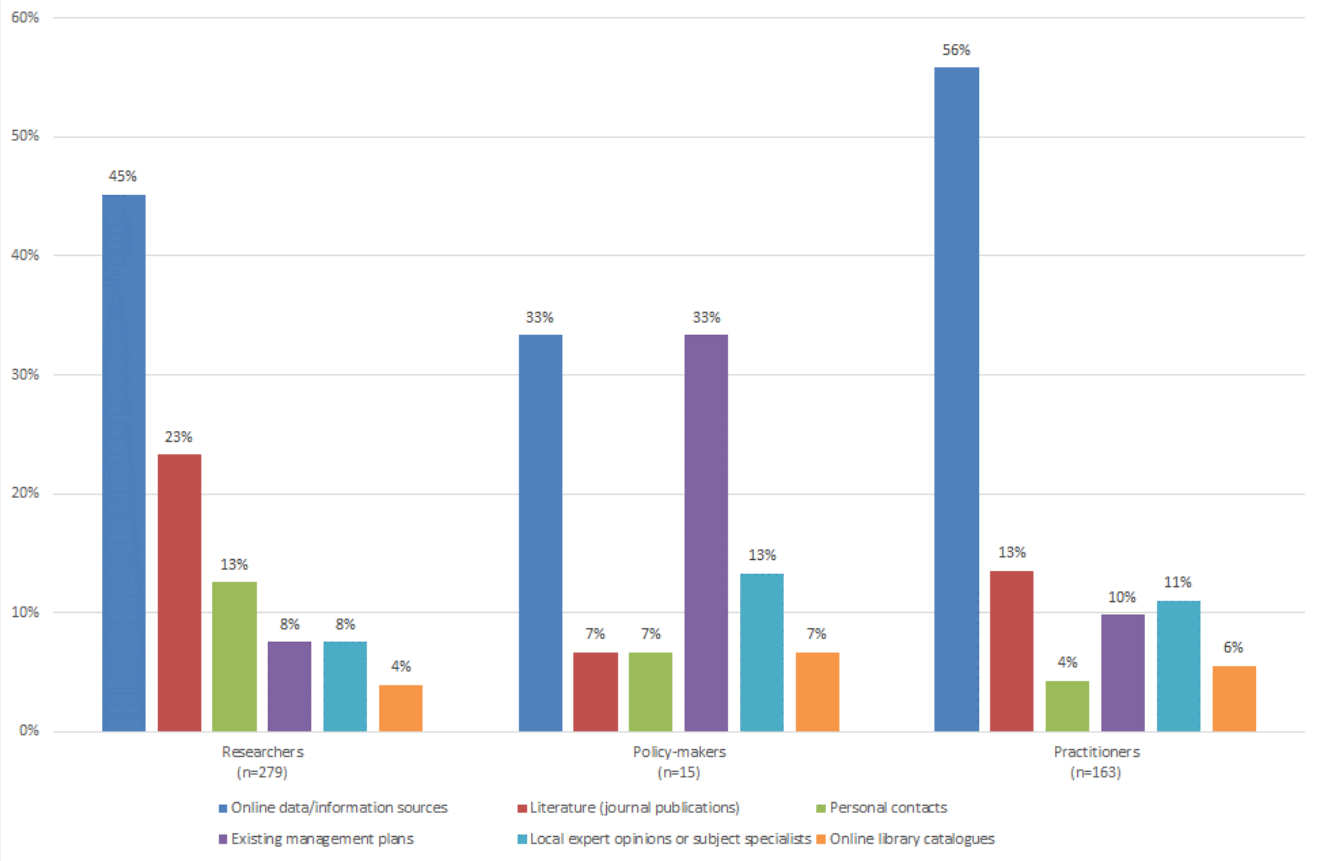


Figure 4.8: Survey responses showing top-ranked information sources by end-user group

Amongst researchers (n=279), a slightly smaller proportion (45% versus 49% in Table 4.8) ranked online data/information sources as their top choice. However, 23% (compared to 19% overall) ranked literature (journal publications) as their top preferred information source. Journal articles are generally read to keep up-to-date (current awareness function) with a certain field of study or research. Published literature is more difficult to access by the other user groups unless open access articles are available. The preference of online data/information sources and literature (journal publications) may be due to these information sources providing well organised and structured data (data arranged in tabular form as defined by Thessen & Patterson, 2011:30), thereby making them more accessible.

Policy-makers (n=15) ranked online data/information sources and existing management plans equally (33%), as their top choices for preferred information sources. Of the policy-makers, 13% consulted local expert opinions or subject specialists, the highest amongst the user groups, as their second-ranked information sources. This may be due to the time pressures where an individual chooses the advice of a specialist, often an easily accessible source in lieu of reading lengthy scientific documents or the specialist is the only source of information. However, Mostert and Ocholla (2005:143) found that seeking advice is less used by parliamentarians than print and electronic sources stating that it is

time-consuming. The majority (56%) of practitioner respondents (n=163) listed online data/information sources as their number one preferred information sources.

These end-user search strategies to find information, match findings from Tann *et al.'s* (2008) showing that respondents use more than one source. However, Tann's study showed that literature (field guides, fact sheets, books, journals and grey literature) were preferred, followed by databases (in-house, external) and a person or agency.

4.2.3.2 Online data/information sources consulted (Q17)

The purpose of Question 17 was to gain a sense of the number of online biodiversity resources being used and explore how current online biodiversity resources support respondents in their biodiversity-related work. The different websites referenced in the survey were tabulated and their frequency determined.

Respondents were asked which websites they primarily use when searching for biodiversity information and in this way, the researcher determined where a user group goes to fulfil a need and which currently-available websites are being selected to solve their information problems. The analysis looked at the various types of websites consulted and the average number of websites consulted by end-user groups. Three categories of websites were identified, i) SANBI websites (the most popular included BGIS, Plants of southern Africa (POSA), Red List of South African Plants, Biodiversity Advisor, Citizen Science (iNaturalist / iSpot) and PlantZAfrica), ii) other South African websites such as national government websites (Environmental Geographical Information Systems (E-GIS), Department of Water and Sanitation (DWS), Chief Surveyor-General, South African Weather Service (SAWS), agricultural-related websites, such as Agricultural Research Council (ARC), Elsenberg Department of Agriculture (CapeFarmMapper), Agricultural Geo-Referenced Information System (AGIS), and iii) international websites such as journal publishing houses, Research Gate, Ocean Biogeographic Information System (OBIS), FishBase, United States Geological Survey (USGS), International Union for Conservation of Nature (IUCN).

Categories of websites	Researchers (n=258)	Policy-makers (n=13)	Practitioners (n=148)	All Respondents (n=419)
SANBI Websites	78%	85%	87%	82%
International Websites	52%	62%	30%	44%
South African Websites	36%	54%	42%	39%

* Multiple response question

Table 4.9: Frequency of websites primarily used by end-user groups

Table 4.9, shows the comparison between the usage of SANBI versus other South African and international websites among the three end-user groups. As expected, being a case study of SANBI, end-users most often accessed online biodiversity resources on SANBI websites (82%). This was followed by international websites (44%) and other South African websites (39%). Although many respondents admitted to starting their search via Google, the challenge voiced by respondents from this study was that "hours are wasted sifting through garbage" when searches are conducted via Google.

Although the South African national government websites are regularly consulted, international literature-based websites such as journal publishing houses (including Science Direct, JSTOR, Web of Science), Google Scholar and Research Gate are frequently consulted, therefore foreign websites are ranked higher than South African ones. It was clear that multiple, often disconnected, online biodiversity resources (websites) were consulted such as online biodiversity databases, online mapping services, and aggregated content.

SANBI websites were frequently consulted and the preferred information source, showing that current SANBI websites are useful. This indicates an understanding of user needs and expectations to a certain extent or there may be no other viable South African alternative. The results have provided a catalogue of preferred information sources and a sense of the number of online biodiversity resources being used by information seekers. With further analysis, the differences and preferences between the various website types can be determined.

	Researchers (n=258)	**Policy-makers (n=13)**	**Practitioners (n=148)**	**Average (n=419)**
Average number of websites consulted	2.64	2.69	2.38	2.55
Minimum number of websites consulted	1	1	1	1
Maximum number of websites consulted	19	10	5	19
Sum of the number of websites consulted	682	352	35	1069

Table 4.10: Average number, range of minimum and maximum of websites consulted by end-user groups

A total of 419 individuals responded to the open-ended question 17, and Table 4.10 shows that the average number of online biodiversity resources consulted is 2.55. While Davis *et al.,* (2014:695) provided a vetted list of 17 information sources, one-fourth of respondents (24.8%, n=64) indicated that they regularly consult ''other'' sources. Davis *et al.,* (2014:695) determined the number of information sources regularly consulted by the majority of respondents (60.8%, n=143) was between

two to five information sources. Although the range for Davis is wider compared to this study which all cluster from 2.38 to 2.69, approximately a third of respondents only provided one response which automatically brought down the average as shown in Table 4.11.

Number of responses	Researchers	Policy-makers	Practitioners	Average
1	32.56%	30.77%	33.78%	32.94%
2	24.03%	23.08%	31.76%	26.73%
3	19.77%	0.00%	14.19%	17.18%
4	9.30%	38.46%	12.16%	11.22%
5	7.75%	7.69%	4.05%	6.44%
6	3.10%	0.00%	2.70%	2.86%
7	2.33%	0.00%	0.00%	1.43%
8	0.78%	0.00%	0.00%	0.48%
9	0.00%	0.00%	0.68%	0.24%
10	0.00%	0.00%	0.68%	0.24%
19	0.39%	0.00%	0.00%	0.24%
Total	**100.00%**	**100.00%**	**100.00%**	**100.00%**

Table 4.11: Shows the frequency of responses for each number of websites

Table 4.10 shows that practitioners are consulting less online biodiversity resources on average (2.38) than researchers (2.64) and policy-makers (2.69). Davis *et al.'s,* (2014:696) research showed that academics (57%, n=73) are significantly more likely to consult fewer information sources than respondents overall (46.3%, n=99). Government (41.3%, n=19) and other respondents (42.5%, n=17) are significantly more likely to consult more (>6) than respondents overall (24.8%, n=53).

Results show that biodiversity end users depend on various online data/information sources and tend to stick with what they know or have the time to look for (time availability). A respondent expressed frustration and disappointment with finding it too difficult to retrieve information in a useful format for current needs, from SANBI and government websites.

4.2.3.3 Finding needed biodiversity information (Q18 and Q20)

The quantity of information needed, which is described by Nicholas (2000:63) as the "information appetite" of users, is often dependent on the main purpose (nature of the need), the sources of information available, time pressures, and the diligence and degree of thoroughness (Case, 2006:34) with which the search is pursued. This information characteristic can also be a cause of constraint if there is time pressure, no inclination or need to go through volumes of information, there is an inundation of information or the need for meeting deadlines versus the completeness of the information have to be balanced. Case (2005) explains that information can reduce uncertainty in a

decision-making task; however, this can also become a barrier if there is too little information or having too much (overload).

4.2.3.3.1 Ease of finding needed biodiversity information (Q18)

Respondents were asked about the relative ease of finding needed biodiversity information for the three data types: raw, synthesised/summarised and data models, on a scale of easy or extremely easy, neither easy nor difficult to difficult or very difficult.

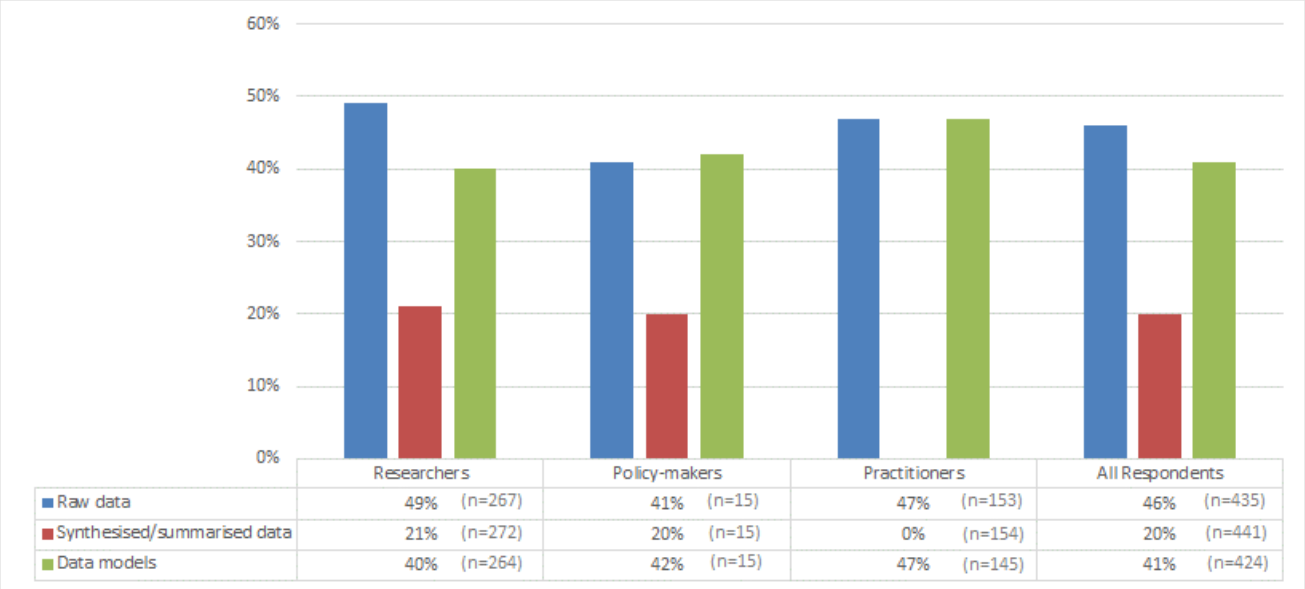


Figure 4.9: Survey responses showing the difficulty in finding needed biodiversity information

The proportion of responses indicating difficulty in finding needed biodiversity information is listed in Figure 4.9 for the three data types. Overall only 20% of respondents indicated that synthesised/summarised data was difficult or very difficult to find. This ranges from no practitioners reporting that synthesised/summarised data is hard to find, as compared to approximately 20% of researchers and policy-makers finding this to be the case. Reports that data models were difficult or very difficult to find, came from 41% of respondents. All three user groups reported raw data was the most difficult to find (46%).

Although the results above provide an overall perspective, many of the respondents commented that *it depends on the species or community being researched* and an example provided, indicated that insect data were more difficult to find at the correct or relevant geographical scale.

4.2.3.3.2 Barriers and enablers to finding biodiversity information (Q20)

Respondents were asked to rate their level of agreement with eight statements about finding biodiversity information. Response choices ranged from strongly disagree, disagree, neither agree nor disagree and agree to strongly agree. The statements provided in Question 20 were summarised to simple variables statements.

Barriers and enablers to finding biodiversity information	Variable
My organisation provides training to help me find the biodiversity information I need to do my work	Training
I already have the biodiversity information I need to do my work	Already have
Readily accessible data exists in the various data repositories	Readily accessible
I feel besieged with information, and totally overloaded	Overloaded
Knowing where to find the biodiversity information needed is a challenge	Challenge to find
The tools for finding biodiversity information work well	Tools work well
Good prior knowledge of where to find biodiversity data is needed	Prior knowledge needed
It will be useful to showcase and/or provide a set of case studies on how other institutions have used biodiversity data	Showcase/case studies

Table 4.12: Description of the variables representing the eight statements

In relation to barriers, Davis (2014:698) highlighted the issues of an organisation providing training to find biodiversity information, knowing where to find information and already having the information as key aspects. Davis (2014:697) reported that the greatest proportion of responses (61.3%, n=133) for any barrier or facilitator measured, disagree that their organisation provides training to help them find the biodiversity information they need to do their work, with 54.8% (n=120) reported knowing where to find information was a challenge. In addition, 54.8% (n=121) reported already having the information they need to do their work.

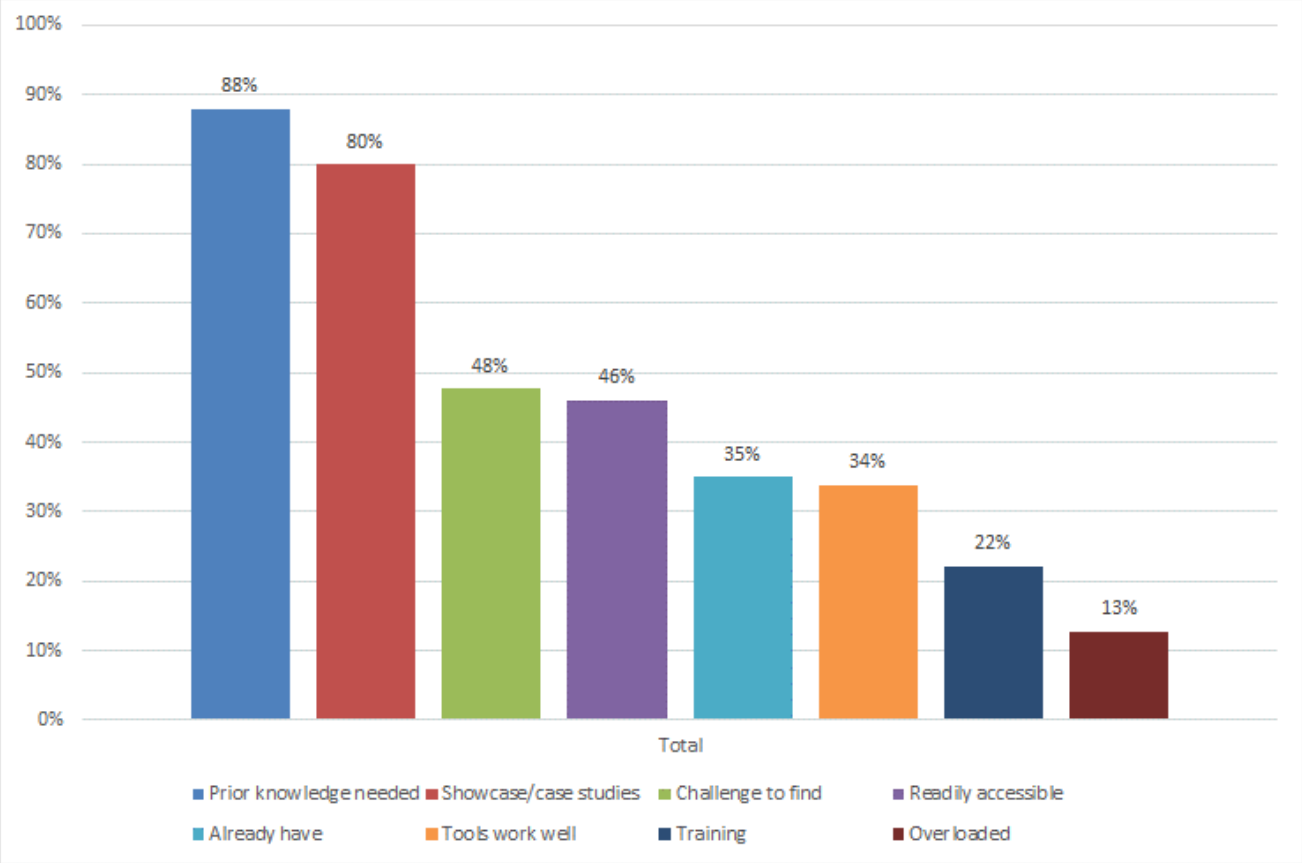
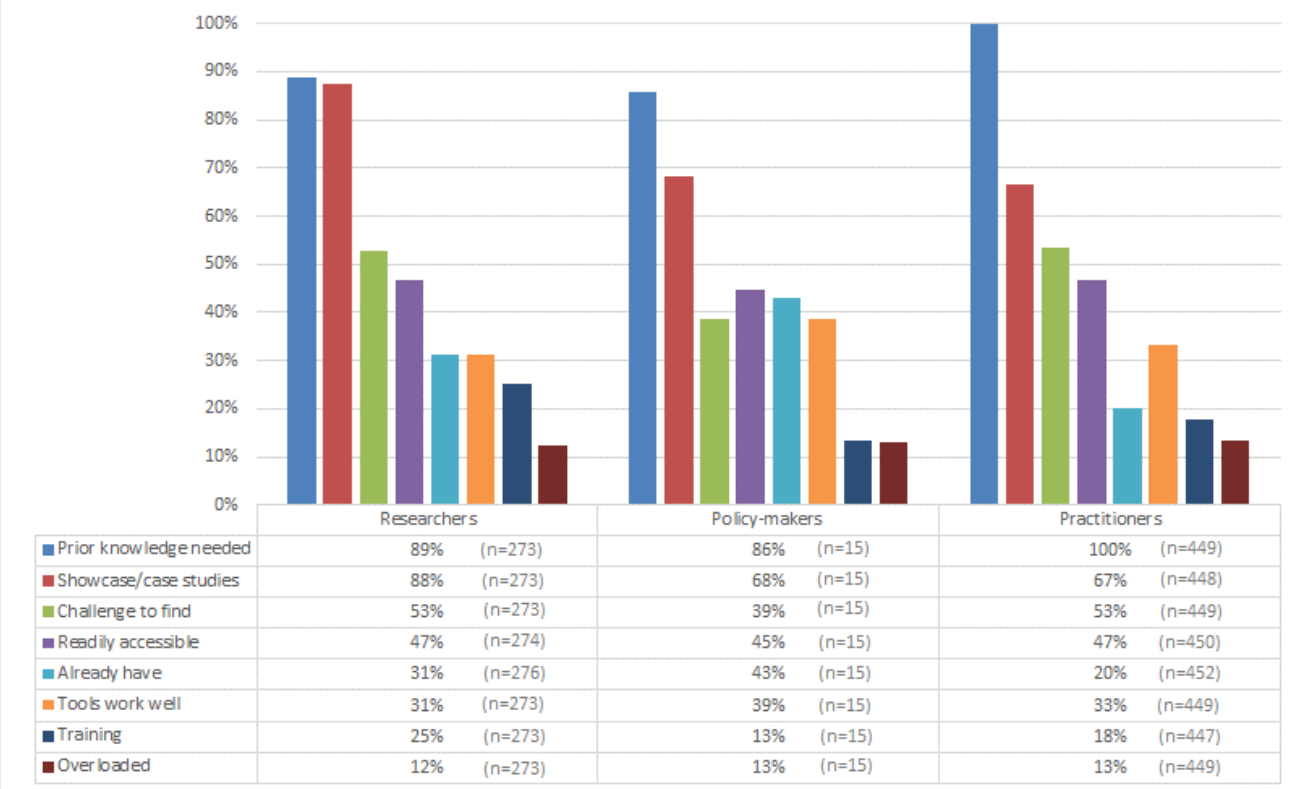


Figure 4.10: Level of agreement (%) with barriers and enablers to finding biodiversity information for total responses

Note: 'Strongly disagree' and 'disagree' were combined to create 'overall disagree' in the question above and 'agree' and 'strongly agree' were combined to form 'overall agree', with 'neither agree nor disagree' remaining as the neutral point. Figure 4.10 summarises the overall agreement proportion to the eight statements provided.

Of the total responses (n=452), 88% of the respondents stated that good prior knowledge is needed on where to find information, followed by providing a set of case studies can be useful and an enabler to understanding how other institutions use biodiversity data (80%). Nearly half of the respondents agreed that they encounter difficulties in where to find needed biodiversity information (48%), and 46% agreed that readily accessible data exists in the various data repositories. Other respondents reported that they already have the necessary biodiversity information to do their work (35%), that the tools work well (34%), that their organisation provides training (22%) and that they felt besieged with information and totally overloaded (13%).



	Researchers		Policy-makers		Practitioners	
Prior knowledge needed	89%	(n=273)	86%	(n=15)	100%	(n=449)
Showcase/case studies	88%	(n=273)	68%	(n=15)	67%	(n=448)
Challenge to find	53%	(n=273)	39%	(n=15)	53%	(n=449)
Readily accessible	47%	(n=274)	45%	(n=15)	47%	(n=450)
Already have	31%	(n=276)	43%	(n=15)	20%	(n=452)
Tools work well	31%	(n=273)	39%	(n=15)	33%	(n=449)
Training	25%	(n=273)	13%	(n=15)	18%	(n=447)
Overloaded	12%	(n=273)	13%	(n=15)	13%	(n=449)

Figure 4.11: Level of agreement (%) with barriers and enablers to finding biodiversity information by end-user groups

Results by end-user groups show that practitioners agree 100% with the statement that good prior knowledge of where to find biodiversity data is needed. Researchers agree almost equally that prior knowledge (89%) and case studies (88%) are enablers to finding biodiversity information. Nearly 50% of each of the three groups agree that readily accessible data exists in various data repositories: researchers, 47%; policy-makers, 45%; and practitioners, 47%. While 22% overall thought that they get the necessary training from their organisation to find biodiversity information, only 13% of policy-makers reported getting the training they needed.

4.2.4 Obstacles encountered and constraints faced (Q11 & 19)

The value of information depends upon the ability to access, analyse, interpret, and communicate results and none of this is possible if the data are not worth using or determined to be irrelevant (Nicholas, 2000:28). Part of the study included determining the problems or difficulties experienced by end-users, as often information needs are not met due to these difficulties. The overall goal of improving the usefulness of data infrastructure services is to provide information in an accessible and easy to understand package. Both Wilson's and Nicholas's models try to identify the elements that represent barriers in accessing biodiversity information. Wilson focuses on factors affecting the

seeking behaviour (time, geography, culture) and Nicholas on the content (quality, quantity). Question 11 and 19 discuss the challenges respondents face when they are not able to find the information needed and the importance or relevance of these challenges are explored.

4.2.4.1 Limitations and constraints experienced in accessing biodiversity information (Q19)

A clearer idea of the challenges faced by end-users when looking for biodiversity information was needed so the situation could be improved where possible. Respondents were asked to complete the following sentence "When looking for the biodiversity information I need to do my work, I am limited by…" and select as many as appropriate from six answer choices.

Survey respondents were asked to tick as many of the answer options as appropriate. Overall, the lack of appropriate information was found to be the greatest barrier (61%) when respondents reported searching for biodiversity information (Table 4.13). However, policy-makers battle with knowing where to look, in equal proportion with lack of appropriate information (53%) and therefore in higher proportions than for the researchers and practitioners.

Response categories	Researchers (n=275)	Policy-makers (n=15)	Practitioners (n=159)	Average (n=449)
Lack of appropriate information	65%	53%	55%	61%
Lack of available information	59%	33%	50%	55%
Lack of time to seek or synthesise relevant information	48%	27%	43%	45%
Knowing where to look	40%	53%	36%	39%
Not knowing how to find what I need	29%	13%	23%	27%
Poor searching skills	9%	7%	5%	8%
Lack of fine scale data	2%	0%	2%	2%
Lack of standardised data formats	1%	0%	3%	2%
Lack of metadata	2%	0%	1%	1%
Lack of financial resources	0%	7%	1%	1%
No limits experienced	0%	0%	3%	1%

* Multiple response question

Table 4.13: Limiting barriers encountered when searching for information

Challenges raised by respondents with regards to financial resources included a lack of access to the internet and resources such as scientific journals. The government's current lack of financial resources limits almost everything: travelling, fieldwork, accessing the internet, replacing out-dated equipment.

When compared to Davis *et al.,* (2014), the limiting factors were cited less frequently than in this study. The lack of appropriate information was cited 36.7% across all work sectors by Davis *et al.,* (2014), compared to 61% across the role or function of an end-user in South Africa. A possible

explanation for this difference between results may be due to the rich variety of biodiversity in South Africa which is still inadequately researched and documented, so the required information may not be available (Steenkamp & Smith 2003:305). Further studies are needed to separate the lack of actual availability from a lack of competence of the end-user in finding the appropriate information.

Davis *et al.,* (2014) identified various barriers that prevent end-users from finding preferred and appropriate information: i) lack time associated with trawling the numerous information sources, with downloading and analysing the data or locating and reading the scientific literature; ii) the information is often difficult to interpret; iii) often out-of-date; iv) incomparable boundaries; v) lack of data integration and gaps in the scientific effort; vi) search skills are inadequate; vii) the complexity of biodiversity information contained within diverse subject area; and viii) jargon or acronyms associated with scientific information. These barriers are similar to those found in 4.2.4.2 below.

4.2.4.2 Causes and consequences when unable to find information needed (Q11)

To gain an overall understanding of respondents' in South Africa's perceptions of the obstacles they face when looking for information to do their work, respondents were asked to list the challenges they encounter when they were not able to find the information they needed. The intention was to gain insight into the challenges users have and in this way to understand what they need.

The key concepts cited by respondents were mapped and collated into themes. In many instances, the respondent conflated the challenge and consequence. Responses were therefore categorised into the presenting issue (challenge) and what is likely to happen if no action is taken (consequences). This study identified the following challenges that respondents face when not able to find the information they need:

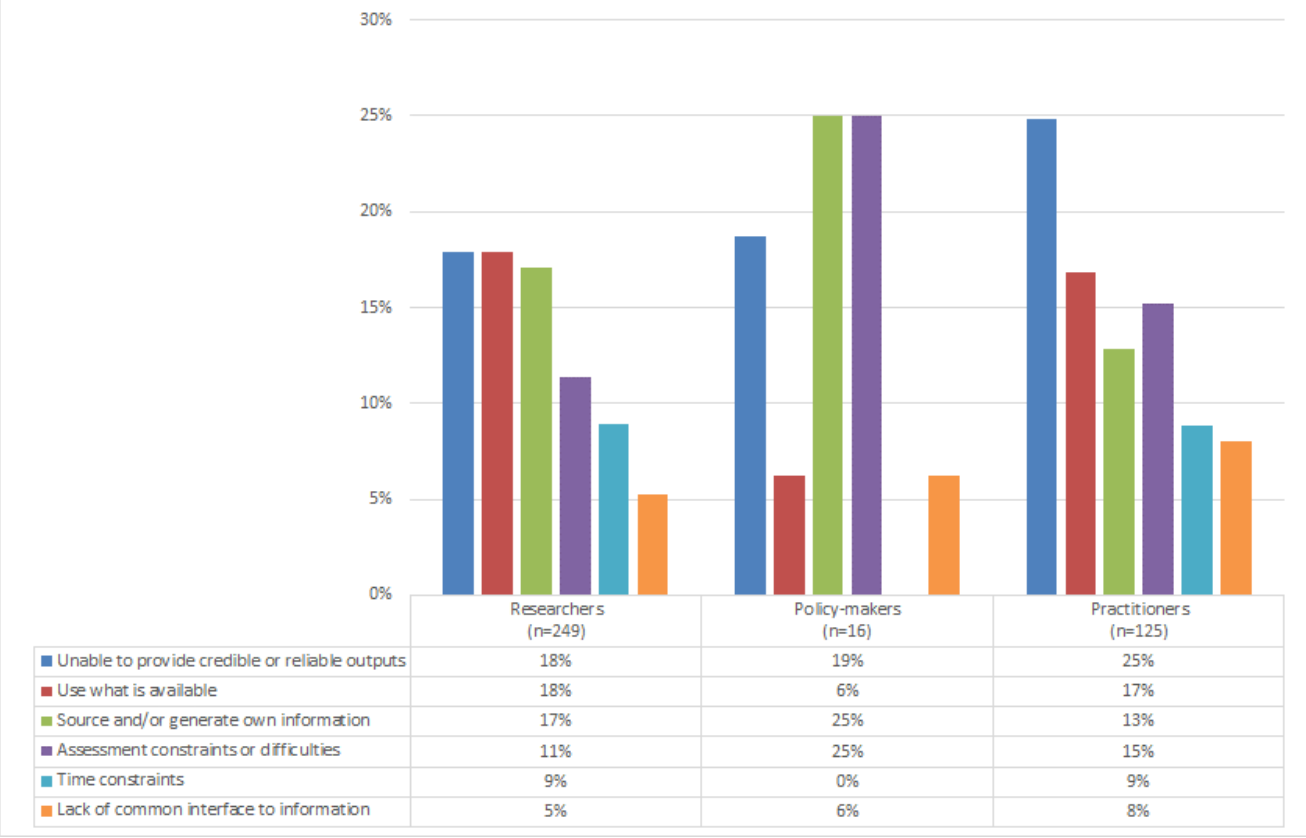
Challenges	Description	Example of task	Count	Percentage of cases
1. Unable to provide credible or reliable outputs	There is difficulty in completing an analysis, inaccurate maps and reduced accuracy when modelling.	Uncertainty in reporting, we are therefore forced to follow the precautionary principle due to deficient data, which in turn may hamper or constrain the development. We have to fill the void with expert opinion.	78	20%
2. Use what is available	If not able to find the necessary information, use what is available.	How to extrapolate sensibly to cover information gaps and surmise when there is data deficiency due to under-reporting.	66	17%
3. Source and/or generate their own information	Consultants need to be hired or costly field surveys are initiated.	The information has to be created internally by the company, due to lack of funds this is not always possible.	62	16%

Challenges	Description	Example of task	Count	Percentage of cases
4. Assessment constraints or difficulties	There is difficulty in completing a site assessment due to historic, baseline, environment data or current conditions for an area not being available.	When the information required is not available then the confidence level of the assessment being conducted is affected. Without adequate information, preparing an accurate assessment is challenging.	51	13%
5. Time constraints involved in collecting data or using alternative sources	The time needed to locate and/or analyse information.	One needs to contact researchers in SANBI directly to obtain the data or inputs, which is time-consuming and relies on personal contacts. It also makes it hard to make the case, if one does access the relevant information, that this is what should be used by all specialists and Environmental Assessment Practitioners (EAPs) who may not go to the same lengths to obtain the most recent information.	33	9%
6. Lack of common interface to information	There are various information sources to query and finding the appropriate or authoritative source is difficult.	The contact person through whom to direct enquiries. Biodiversity data appears to be decentralised making it difficult to access (in several places, managed by different entities).	24	6%
7. Planning constraints or difficulties	Unable to conduct or formulate informed management plans.	Inadequate descriptions of project areas limited ability for developing an understanding of relationships and interpretation of patterns or prediction hinders planning of management and operations of natural resource management work.	24	6%
8. Lack of information sharing culture	Existing data is not accessible and permission to access the data or cost is problematic.	I've found it hard to find people to ask for some of this information and to get responses back from them. For example, I was interested in looking at the original application for a river transfer scheme to see who assessed the potential admixture of different lineages of indigenous fish, but when I contacted the relevant conservation authority they forwarded this to the "right people" who just never got back to me.	23	6%
9. Failure in what duty requires	Unable to do the job or respond to queries or advise clients	Unable to complete that particular project.	15	4%
10. Disparate nature of the information	Heterogeneous information, difficult format to access, incorrect spatial scale (including both data resolution and spatial extent).	Conflicting information and sometimes navigating multiple web pages on the SANBI site and BGIS that don't link to each other e.g. the VEGMAP webpages. Structure of the web pages is also sometimes not that intuitive on BGIS. It is difficult to find all datasets easily with the current search function e.g. I would like to	11	3%

Challenges	Description	Example of task	Count	Percentage of cases
		search for all provincial vegetation maps. On the old BGIS we could search by theme or data was arranged by theme but with the new BGIS we need to search by name, and you have to know the names of the datasets which don't always say 'provincial'. This is similar for government datasets and city datasets. It would be nice if portals existed from BGIS to these sites as well.		

Table 4.14: Descriptions of the challenges when not able to find information and examples provided

The lower response rate to this question (n=387) is due to many of the responses being consequences and not challenges and therefore records were marked as null and not included in the analysis. An analysis of the survey responses showed that being unable to provide credible or reliable outputs was a major challenge when unable to find information needed (20%, n=78). This was followed by having to use what is available (17%, n=66).



	Researchers (n=249)	Policy-makers (n=16)	Practitioners (n=125)
Unable to provide credible or reliable outputs	18%	19%	25%
Use what is available	18%	6%	17%
Source and/or generate own information	17%	25%	13%
Assessment constraints or difficulties	11%	25%	15%
Time constraints	9%	0%	9%
Lack of common interface to information	5%	6%	8%

Figure 4.12: Top six challenges respondents face when not able to find needed information by end-user group

The challenges experienced by all three end-user groups were very different, indicating that each group experienced unique difficulties and therefore their needs were also different. The greatest challenges for researchers were that they were unable to provide credible or reliable outputs, in equal proportion with using what is available (18%), followed by the sourcing and/or generating their own data which increases costs (17%). Policy-makers placed having to source and/generate their own information and assessment constraints or difficulties equally (25%), followed by the difficulty of being able to provide credible or reliable outputs (19%). Practitioners battled with concerns about credible and reliable outputs (25%) in higher proportions than researchers and policy-makers.

4.2.5 Factors influencing information-seeking (Q21 - Q23)

Case (2006:94) explores a list of needs in actively seeking information, these include the perceived value of information such as relevance, pertinence and salience (information is perceived as relevant) and how contextual factors (as discussed in section 4.1.2) or a user's internal or mental structures modelled by an individual's background and education or environment (set of circumstances) can determine one's perception during seeking information. Three questions in the survey were used to investigate the significance of biodiversity information tools and factors that influence the selection of information sources.

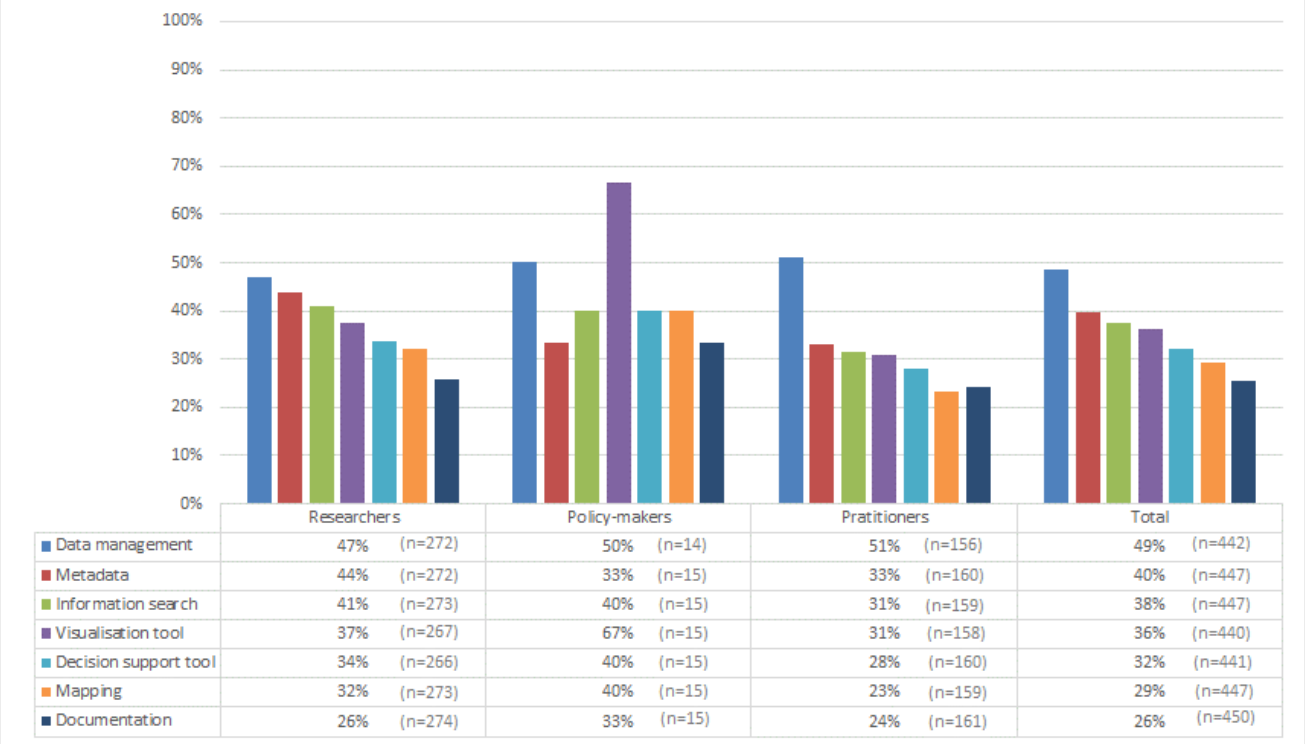
4.2.5.1 Significance of biodiversity information tools in conducting biodiversity work

As mentioned in section 4.2.1.2.1, packaging is the various means in which information may be stored (primary or secondary information) or presented (Line, 1969:12, Nicholas, 2000:84). Information tools are another way of presenting information and help end-users discover, manage, analyse, integrate, visualise and describe data/information (Chapman 2005, Davis *et al.,* 2014, Corrêa *et al.,* 2018). The ease of finding and the importance of biodiversity information tools can inhibit or encourage information use.

4.2.5.1.1 Ease of finding biodiversity information tools (Q21)

Davis identified seven biodiversity information tools and respondents were provided with a definition (Kagan, 2006, Schuh, 2012 & Houtkamp *et al.,* 2016) and asked to rate the ease of finding these tools. The responses to 'very easy' and 'easy' were combined to create 'overall easy' in the question above and 'difficult' and 'very difficult' were combined to form 'overall difficulty', with 'neither easy nor

difficult' remaining as the neutral point. Figure 4.13 summarises the proportion of overall difficulty in finding biodiversity information tools.



	Researchers	Policy-makers	Pratitioners	Total
Data management	47% (n=272)	50% (n=14)	51% (n=156)	49% (n=442)
Metadata	44% (n=272)	33% (n=15)	33% (n=160)	40% (n=447)
Information search	41% (n=273)	40% (n=15)	31% (n=159)	38% (n=447)
Visualisation tool	37% (n=267)	67% (n=15)	31% (n=158)	36% (n=440)
Decision support tool	34% (n=266)	40% (n=15)	28% (n=160)	32% (n=441)
Mapping	32% (n=273)	40% (n=15)	23% (n=159)	29% (n=447)
Documentation	26% (n=274)	33% (n=15)	24% (n=161)	26% (n=450)

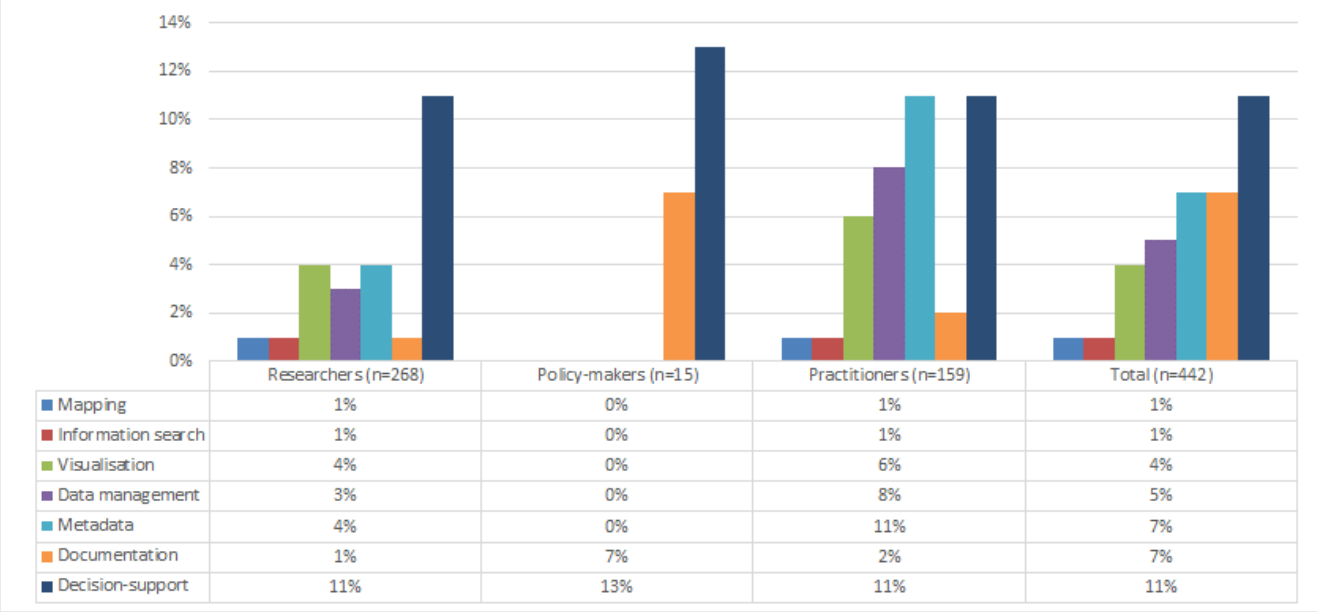
Figure 4.13: Proportion of overall difficulty of finding needed biodiversity information

Overall, 49% of the respondents experienced difficulty in finding data management tools with a further 40% reporting difficulty in obtaining metadata tools. Information search tools were difficult to find for 38% of respondents, while only 26% of respondents reported difficulty in obtaining documentation tools. The pattern of responses is similar for the researcher and practitioner groups, however, results show that policy-makers particularly battled to find visualisation tools (67%) and then metadata or documentation tools (both 33%).

Davis et al (2014:695) reported that "less than one-third of respondents (29.6%, n=83) said it was easy or extremely easy to find the biodiversity information tools they needed".

4.2.5.1.2 Importance of biodiversity information tools (Q22)

Respondents were asked to rate the importance of the seven biodiversity information tools on a Likert scale, from not at all important, not so important, somewhat important, very important to extremely important. 'Not at all important' and 'not so important' were combined to create 'not important' in the figure below.



	Researchers (n=268)	Policy-makers (n=15)	Practitioners (n=159)	Total (n=442)
Mapping	1%	0%	1%	1%
Information search	1%	0%	1%	1%
Visualisation	4%	0%	6%	4%
Data management	3%	0%	8%	5%
Metadata	4%	0%	11%	7%
Documentation	1%	7%	2%	7%
Decision-support	11%	13%	11%	11%

Figure 4.14: Biodiversity information tools that were ranked "not important"

Respondents from each user group rated the relative importance of biodiversity information tools, as per definitions described in Appendix 2. The least important biodiversity information tools, using the proportions reporting "not important", ranked across all the respondents is as follows: decision-support tools (11%), documentation and metadata (7% each respectively) and data management (5%). Consequently, survey results show that mapping and information search (1% respectively), are therefore relatively more important to biodiversity end-users, followed by visualisation (4%).

These results are confirmed by Davis (2014:695) who found that decision-support tools were also the biodiversity information tools least likely to be used by respondents (24.3%, n=73). The percentage of respondents who do not use the other tools measured is as follows: metadata tools (19.2%), documentation tools (9.7%), visualisation tools (8.8%), data management tools (6.3%), mapping tools (4.4%), and information search tools (2.0%).

4.2.5.2 Information source attributes (Q23)

Elements influencing respondents' preferences for specific information sources were addressed in Question 23. Nicholas only draws attention to two (quality/authority and date range/currency) of the ten information attributes identified in this study. Although Nicholas describes quality/authority as highly subjective, it is defined as a major information characteristic as it denotes completeness, accuracy, validity and consistency. Assessing quality is aided by the understanding of whether or not the producer of the information is perceived as a competent authority and how this perception can change over time (Nicholas, 2000:67). The date/currency of information is another influencing factor

that relates to the currency of the information (how up-to-date) and how far back in time the information is required (Nicholas, 2000:73).

Respondents were asked to rate the importance of ten information source attributes and the definitions for each attribute were provided in the questionnaire according to Davis *et al.,* (2014), Tann *et al.,* (2008), Houtkamp *et al.,* (2016) and Weiler (2005). Once respondents had rated the importance of each source attribute on a scale, numerical values were assigned to the relative importance with 1 denoting 'not at all important', 2 'not so important', 3 'somewhat important', 4 'very important' and 5 'extremely important'. This resulted in a total score which was then divided by the total number of responses to provide an average rank for each of the information source attributes.

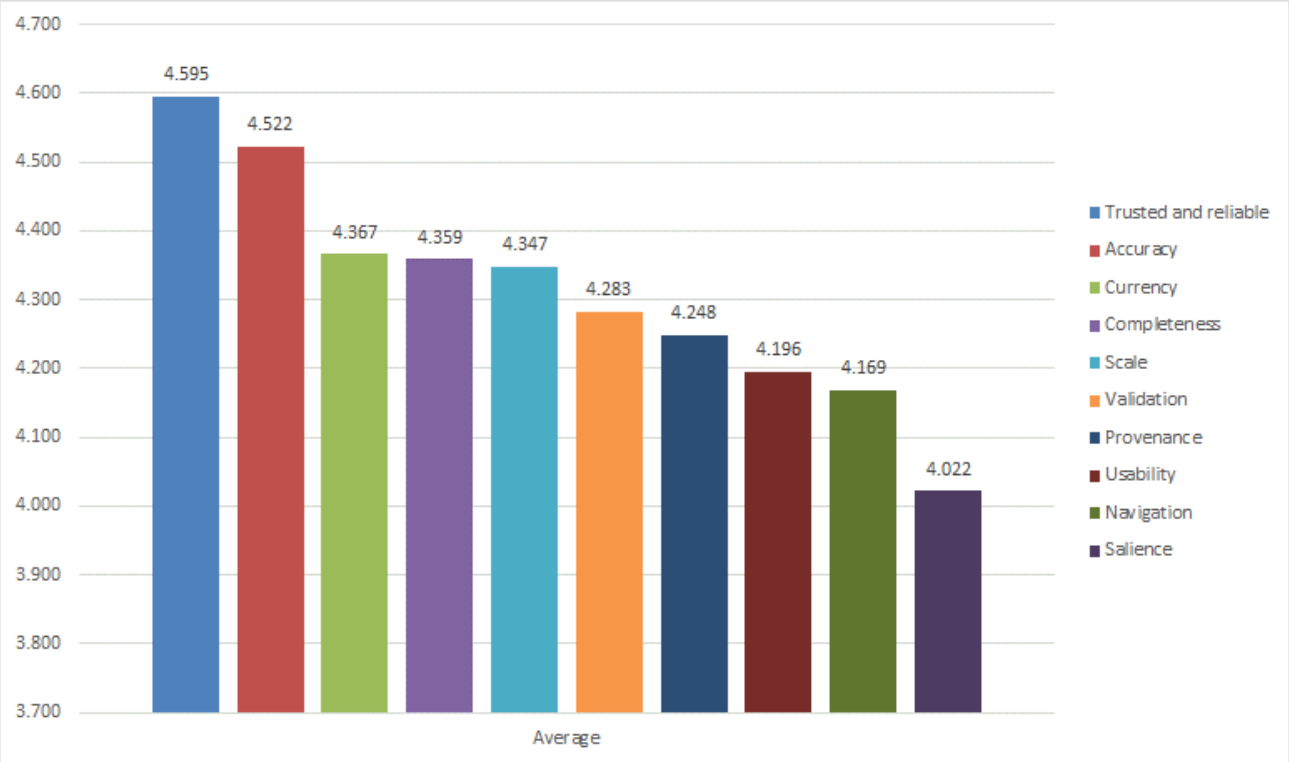
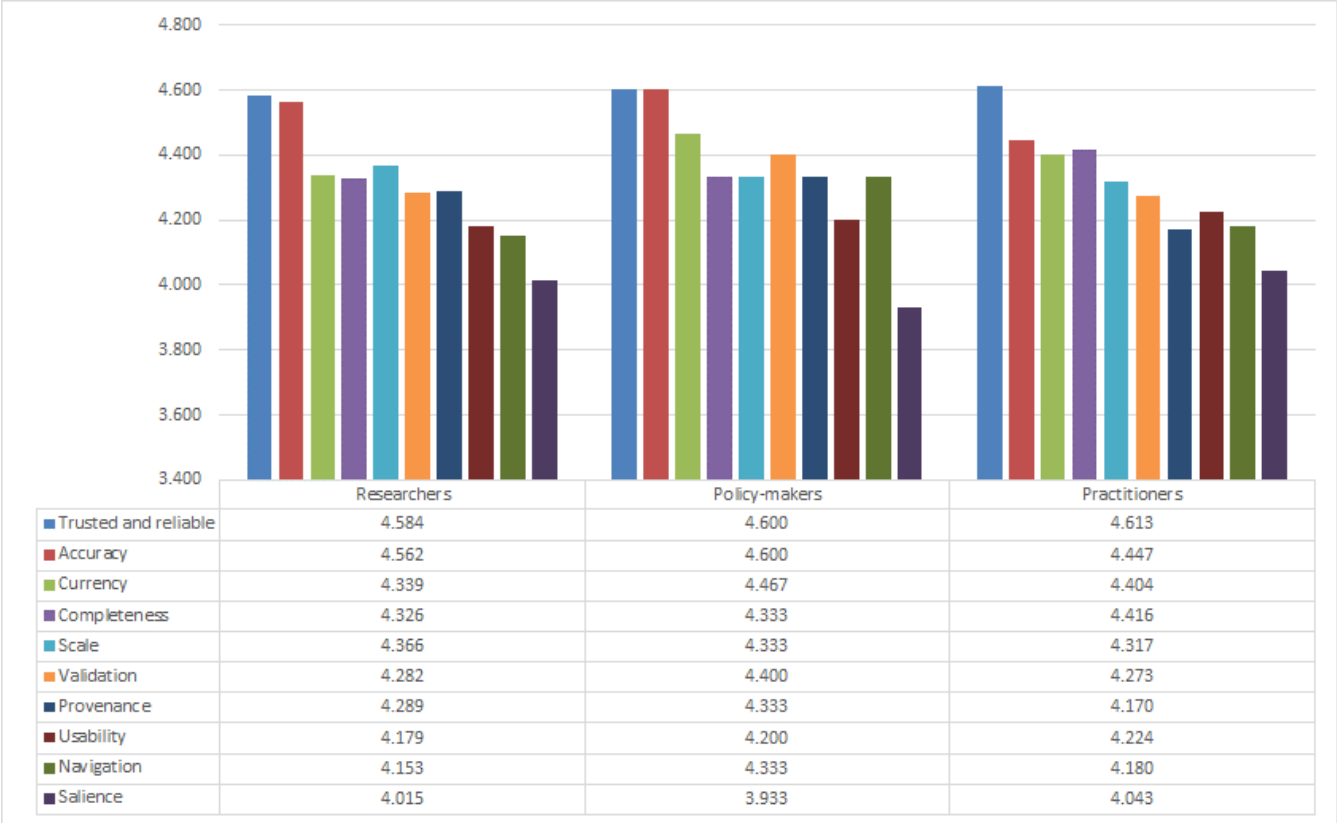


Figure 4.15: Ranking the importance of source attributes

Using the overall average score, respondents ranked trusted and reliable sources of information as the most important information source attribute with an average of 4.595. This was followed by the accuracy of the information (4.522 average score); while the salience of the source was ranked lowest in importance (4.022 average score).



	Researchers	Policy-makers	Practitioners
Trusted and reliable	4.584	4.600	4.613
Accuracy	4.562	4.600	4.447
Currency	4.339	4.467	4.404
Completeness	4.326	4.333	4.416
Scale	4.366	4.333	4.317
Validation	4.282	4.400	4.273
Provenance	4.289	4.333	4.170
Usability	4.179	4.200	4.224
Navigation	4.153	4.333	4.180
Salience	4.015	3.933	4.043

Figure 4.16: Ranking the importance of source attributes across the end-user groups

Trusted and reliable sources were ranked as the most important of the ten information source attributes across all three end-user groups. While accuracy was rated second most important by end-user groups, scale ranked highly amongst researchers (4.366 average score), currency (4.467) and validation (4.400) being important to policy-makers and completeness (4.416) and currency (4.404) of importance to practitioners.

Davis's (2014:696) research shows that approximately 80–95% of the respondents rated all seven attributes as important or essential, with no significant difference found between work sectors (academic, government and others). Accuracy, validation and salience were not included in Davis' list of source attributes.

4.3 Conclusion

This chapter defined the type of population studied and the context in which they work. A set of questions assessed end-users' current usage trends, which forms and types of biodiversity information were most useful, information needs that existed, and problems experienced when not able to find information. The findings from the survey also identified major tasks of importance,

information sources and influencing factors as well as areas of importance related to data models, information tools and source attributes.

The theoretical models proposed for this study provide a range of inter-related aspects to evaluate information needs and information-seeking patterns. Chapter 5 reflects on how the results presented in Chapter 4 provide a clearer understanding of the research questions and recommendations to address information needs presented. The results in the chapter to follow were analysed and understood within the perspective of the chosen theoretical frameworks and the research questions that were originally posed.

Chapter 5: Conclusion and recommendations

The main Research Question for this dissertation (section 1.4) asked to what extent the needs of biodiversity information users in South Africa were supported via existing information sources at the science, practice and policy interface. In this final chapter, findings from the research question and consequent sub-questions, are summarised, discussed and interpreted. Detailed answers to the research questions are provided, which in turn and where appropriate, produced recommendations for implementation.

Finding a suitable theoretical framework was particularly challenging, due to the absence of any theory used or reference made in any of the related literature including the studies from which the survey instrument was adapted. Nicholas's (2000) model of the information needs analysis framework recognised the characteristics of information needs and thereby provided the variables that were tested and also facilitated discussions on concepts and their rationale. Wilson's (1997) general model of information behaviour offered systematic answers to questions with practical applications, and this laid the groundwork for future analysis of the dynamic nature of changing needs. Although both models provided direction and contributed in different ways to the understanding of the information behaviour phenomenon, they were limited, as was also found by Afzal (2017), by the lack of clarity in conceptualising and measuring information needs, terminological concerns, and no clear agreement on appropriate methodologies.

5.1 What are the characteristics of the information needs of end-user groups?

In addition to the main research question, seven ensuing sub-questions were posed. Sub-question one explored who the users were that accessed SANBI's online biodiversity information. Responses to this sub-question defined and described the population and investigated the primary roles associated with biodiversity information to understand who is using biodiversity data in South Africa. The desktop study distinguished between 27 unique types of users interacting with biodiversity data/information, with these types being grouped into three broad categories (researchers, policy-makers and practitioners) called 'end-user groups'. The study found that research was the primary role in the various biodiversity-related work performed and the majority of respondents indicated environmental sciences as one of the broad disciplines in which they work, followed by agriculture and natural resources. This is important in the prioritisation of end-user needs in an effort to establish a system that supports these aspects.

Peacock and Brownbill (2007) explained that the way we think about end-users is the way we shape a webspace and this is often an unexplored avenue of assumptions. End-users of biodiversity information are not only actively engaged in performing technical tasks in the field (practitioners) but also include researchers and decision-makers (refer to section 4.1.2). This diverse end-user base of biodiversity information created a challenge in identifying a suite of requirements that applied to all. Whilst researchers are interested in expanding a body of knowledge by generating new scientific information, policy-makers manage biodiversity through the development and better delivery of policy, and biodiversity practitioners use biodiversity information to inform the practice of managing biodiversity.

Keeping in mind that science means researchers and practitioners and that decision-makers are policy-makers the results from the analysis showed that the information needs and information-seeking behaviour of these three sub-groups are distinct due to the different patterns of responses to key questions. Researchers differed from practitioners and policy-makers; with the former reporting a higher raw data requirement (primary data) than secondary information (Q5). Raw data, however, was perceived as the most difficult to find by all respondents (Q18). The frequency of seeking and using interdisciplinary information was highest among researchers (Q25) and as described by Jamali and Nicholas (2010:233), these disparate subject areas may rely on different information types and sources.

It is important to acknowledge the dependence of policy-makers on concise information which is evident in the importance of data models (Q9), planning/prioritisation outputs (Q4), and their top-ranked information source being that of existing management plans (Q16) when looking for biodiversity information. Results showed that the briefing function, as described by Nicholas (2000:51), is vital to policy-makers as their need for background summaries, synthesis information and advice are greater than for the other two end-user groups (Nicholas & Herman, 2009:44).

Accessing information through online data/information sources were the practitioners' preferred information source (Q16). However, they consulted less online biodiversity resources on average (Q17) and were less reliant on interdisciplinary information (Q25). White & Molina (2006:12) suggested that this could be due to the nature of the information sought by practitioners which are constantly shifting due to the dynamic nature of ecosystems (natural and unnatural) and the challenges associated with changing landscapes (e.g. fire, land use, grazing and climate change).

The distinct needs highlighted above show that the "context of information needs" has a significant impact on requirements, information-seeking behaviour and use. It is important that the integrated information system meets these differentiated needs and considers how information is offered to the three end-user groups. Exploring which of their information needs will have the most impact and how these interests can be balanced between the groups, are essential. This relates to Nicholas's processing and packaging characteristic of providing information in a way that speaks to each end-user group. Nicholas and Herman (2009:57) however warn of people's varied ability to comprehend information and the various problems associated with the different levels of information that are required to satisfy the uninformed, partially informed and informed end-user (intellectual level).

Recommendations to consider in planning a new system:

- Provide adequate guidance to end-users who are unfamiliar with the biodiversity discipline and information structure (e.g. intuitive navigation, grouping content into concepts).
- Tailor the different data/information views to meet the needs of different types of end-user groups and ensure they apply to their specific needs e.g. concise formats that are reliable, accurate, and up-to-date, with some measures of validation for policy-makers (Vissers *et al.,* 2017).
- Provide a platform where all three sub-groups can regularly consult information to guide decisions related to biodiversity; this will ensure for example, that parliamentarians use authoritative and verified information.
- Store, maintain and index unpublished material such as existing management plans for ease of access.
- Create a space for SANBI's science and policy advice divisions to provide proactive advisory mechanisms on summaries of emerging and topical biodiversity information.

5.2 What do respondents need from the information system?

Sub-questions two, three and six were linked and focused on aspects of information needs. Sub-question two explored what the biodiversity information needs of users were. Sub-question three identified and examined major uses of retrieved data, and sub-question six considered the demand for information from other disciplines (refer to Table 3.1).

Ariño *et al.,* (2013:3) suggested that assessing future information needs should be based on current usage trends and survey question four investigated current use and revealed that of the thirteen data

categories respondents selected, they indicated that maps and distribution data and ecological data were the two most demanded data categories. These results corresponded with question 12. The task analysis included four questions (Q12-Q15) and provided insight into how users go about the work-role. Results show that of the 14 major tasks identified, site assessments (current and historical inspection of a site) and distribution analyses (determining the range of species) were the main tasks; hence the need for distribution and ecological data. Integrating data to complete these tasks can be time-consuming in trying to resolve heterogeneity issues as described by Torres *et al.,* (2006) and an integrated system should facilitate this process. Knowing what common tasks have to be completed, helped in determining the current information needs of respondents as proposed by Tann *et al.,* (2008). Nicholas and Martin (1997) however, cautioned that these types of investigations are dependent on what information is available and only monitors what currently exist. This, however, provided an improved comprehension of the tasks respondents try to perform with available information.

Responses from Q5 and Q6 indicated that raw primary data were used by all three end-user groups in their investigations to generate interpreted information, however, these results did not directly convey the information in a format that was most useful for decision-making, and provision needs to be made for this. Findings affirmed the importance of data models (Q9), showing that across the user groups, data models were important to policy-makers and slightly less important for researchers and practitioners. The key criteria identified by respondents that influenced information-seeking, and which may result in information not translating into effective use, was the importance of geographical scale. Results showed that biodiversity data is most frequently used at a local geographical scale (Q7), and therefore providing information at a finer scale will ensure the relevancy and usefulness of the data. These findings respond to sub-question two and three.

Of the numerous roles and responsibilities end-users are required to undertake, each has its own subject requirements ranging in detail and extent (Nicholas & Herman, 2009:29). Although results in Q8 showed that at least half the information respondents need to do their work, specifically relates to biodiversity, there is still a high degree of interdisciplinary information required. Questions 24 and 25 investigated the reliance on information from other adjacent fields of science. Results showed that researchers, policy-makers and practitioners drew from a variety of disciplines, and the implication of this broad scope was one challenge, and the other was in finding authoritative information. Although all three end-user groups indicated they were reliant on interdisciplinary information to a certain extent, the frequency of seeking interdisciplinary information is higher for researchers and

policy-makers which can potentially result in different information-seeking patterns compared to practitioners (Jamali & Nicholas, 2010:234). Nicholas and Herman (2009:45) suggested in an attempt to satisfy the interdisciplinary information needs of respondents, that a briefing function type search is implemented to provide a broad understanding of the numerous subject areas. These findings answer sub-question six.

As listed above several forces influence the biodiversity end-users' information needs and seeking patterns such as disciplinary and geographical boundaries across which information needs to be integrated. The integration of information across subject areas remains a challenge for respondents due to potentially limited knowledge outside their field of expertise, the various data formats, and the need for interchangeable data standards and structures (Thessen & Patterson, 2011). Findings derived from these questions indicate that the variations in subject requirements and the current lack of biodiversity information integration with key disciplinary information will remain a barrier and can result in failure to achieve tasks.

Question 10 identified information gaps that exist which compelled individuals towards information-seeking. The top three content gaps included:

(i) environmental change data - where long-term research data is currently unavailable, difficult to find or problematic to compare;
(ii) occurrence/distribution - relates to the data deficiencies in coverage of certain taxa (e.g. invertebrates);
(iii) ecosystem data - although terrestrial (National Vegetation Map 2006), wetland and river (National Freshwater Ecosystem Priority Areas (NFEPA) 2011) information is available, it is only recently that estuarine, marine and sub-Antarctic marine ecosystem maps became available following the 2018 National Biodiversity Assessment.

Recommendations to consider in planning a new system:

- SANBI needs to consider its ability to effectively manage and disseminate the broad range of data categories identified (Q4) and focus on the most demanded data categories.
- Through further analyses of Q13-Q15, investigate how data or information should be presented so that the major tasks identified are easily completed. Ascertain how tasks can be broken down into more manageable steps so the system can be built to present these data more constructively.
- Accept and store site assessment data and reports as suggested by Tann *et al.,* (2008).

- Combine geospatial and biological data for species distribution analysis, also aggregate information from many biodiversity institutions that hold data, and make available and accessible a richer store of data for reuse.
- Work with data partners to advance interoperability and data integration, developing tools and models to combine heterogeneous interdisciplinary information.
- Focus on aggregating or mobilising biodiversity information that is local in scope.
- The integration of a broad mix of disciplinary information is needed. Further investigation is needed into firstly data standards to accommodate different database schemas, and secondly institutional requirements and policies regarding data sharing. Allowing for the linking of datasets from various fields of study should reduce the time and effort respondents currently spend on searching for authoritative interdisciplinary information.
- Determine important sources of environmental change information that end-users need, and how best to represent and measure these changes.
- Encourage the publishing of environmental change data through institutional repositories and integrate where possible.
- Deficient species-specific data should be crowd-sourced through citizen science projects.
- In the future, iterative ecosystem information should be updated, versioned and served.
- Using the results from Q10 and Q24, conduct a Data Gap Analysis as suggested by Chavan, Sood and Ariño (2010) to map data that is currently inaccessible with information needs identified in this study.

5.3 What is important to end-users of biodiversity information?

Sub-question four enquired about the information-seeking strategies of SANBI's online users and was aimed at understanding how end-users search for information. Question 16 investigated how and through what form users currently access or fulfil their information needs. The findings concluded that online data/information sources were the most popular channels employed to gain access to information. Literature (journal publications) were considered to be trusted and reliable sources and therefore also favoured.

Responses to Q17 indicated that SANBI websites are the more popular online/information sources regularly consulted, showing to some extent that SANBI websites are meeting current information needs. Due to the vast number of online/information sources being consulted, it is hardly surprising that results in Question 19 showed that confusion frequently ensued about where to look for the

information required timeously, especially for policy-makers (53% n=15), due to the existence of many platforms.

This follows on the question of the difficulty in finding biodiversity information (Q18) and responses showed that this varied based on the type of information sought, with raw data and data models being the most difficult to find. Respondents' attitudes were that good prior knowledge of where to find information was essential and that it would be useful to showcase and/or provide a set of case studies on how other institutions have used biodiversity data.

Recommendations to consider in planning a new system:

- The accessibility to peer-reviewed research outputs are highly valued, therefore it is vital to enable access to research outputs via SANBI's institutional repository without breaking copyright laws; this includes open access to data, code and papers.
- Ensure raw data from research projects are made accessible.
- Provide case studies that contain a particular problem or decision that needs to be made.

5.4 What pressures do respondents face?

The fifth sub-question of the study enquired about the challenges that hindered SANBI's online users in solving problems or achieving their goals. It was intended to understand the current challenges experienced within the biodiversity community and to mitigate these limitations where possible. Although survey question 11 and 19 addressed two different aspects, they showed that if SANBI was going to provide a useful information service, then it would be required to address a host of challenges experienced by respondents.

Responses to Question 11 explored the challenges users faced when not able to find information and introduced ten challenges. The top six of these enumerate the most serious adverse consequences for users. Lacking sufficient information, respondents were challenged as follows:

Challenges	In consequence
i. Unable to provide credible or reliable outputs	Little confidence in the accuracy of results. Filling the void with expert opinion perpetuates and maintains information gaps.
ii. Making use of what is available	Extrapolation or abstraction of data compromises or biases the information; suboptimal use of data with reduced effectiveness. Monitoring is not possible and the verification of information also poses a challenge.
iii. Having to source or generate their own data	Duplication of effort and additional costs (consultancy-based service contracts).
iv. Assessment constraints or difficulties	Inability to recommend management actions, and less effective planning, resulting in ill-informed decision making.

Challenges	In consequence
v. Time constraints to seek or synthesise relevant information	Incapable of meeting deadlines and the quality of work is compromised. Time is an external challenge that SANBI has the least control over, however it is important to understand that end-users may resist pursuing a need due to a lack of time (Nicholas & Herman, 2009:19).
vi. Lack of common interface to information	Leads to users struggling to find relevant information or contact person. Although the internet has radically increased the likelihood of finding information, the disparate information systems create other barriers as discussed above.

Table 5.1: Challenges and consequences faced by respondents when not able to find needed information

The challenges listed in table 5.1 point to respondents "not wanting to base a decision or a course of action on unstable foundations" when not able to find needed information (Nicholas & Herman, 2009:72). This corresponds with results from Q23 in which respondents are concerned about the authority and accuracy of information.

The lack of appropriate information was identified as a significant barrier for Q19. Nicholas and Herman (2009:31) suggest that the root cause of being unable to find appropriate information relates to ineffective subject descriptions. Locating information on a subject is difficult when topic descriptions contain jargon and acronyms make it difficult to search. Appropriate language needs to be employed to address the particular needs of newcomers to the subject area. Although Nicholas and Herman (2009:31) suggested the use of a controlled vocabulary, he warns that there is still the issue of the end-users having to guess the indexer's choice. While certain technology can index an unlimited number of words in a document, this type of functionality inevitably produces irrelevant information that users still need to sort, however, it does not require the end-users to have prior knowledge of the system (Nicholas & Herman, 2009:36).

Recommendation to consider in planning a new system:

- Develop and adopt a hybrid system of controlled vocabularies and natural language terms to enable a shared and common understanding and improved quality of query results. The requirements of the Biodiversity Information Standards (TDWG) Vocabulary Maintenance Specification and Standards Documentation Specification need to be followed.

5.5 What are the respondent's values, interests and concerns?

Answers to sub-question five of the study inquired about challenges related to information tools and investigated the preferred key features of information sources, the ease of finding information tools and their importance. Davis *et al.,* (2014) showed that information search and data management tools are necessary to integrate and facilitate access to information. Survey respondents were requested to rate the ease of finding (Q21) and the importance of these tools (Q22). Key findings revealed that respondents struggle to find all seven biodiversity information tools to some extent, with at least half of the respondents unable to find data management tools. Considering the effort to create and maintain metadata (the second most difficult information tool to find) these catalogues must be made more accessible. All respondents indicated that mapping tools and information search tools are important, while decision-support tools are the least important and therefore the least likely to be used (Q22). With ever-shrinking resources, SANBI must prioritise and build on those tools in greatest demand.

Research shows that good information has numerous qualities (Faith *et al.,* 2013). Question 23 identified what qualities (source attributes) were important to respondents and how they influenced the preference of information sources. Ranked highest on the list of needs affecting the use of information, were trusted and reliable sources. Ariño, Chavan and Otegui (2016) suggested measures of data trust and reliability that can be implemented. Ariño, Chavan and Faith (2013) stated the lack of quality checks results in less reliable data. Based on the literature it is no surprise that accuracy, currency and completeness were returned as consecutive rankings. Providing measures of reliability and maintaining accuracy and currency will necessitate time and effort from SANBI, but are important in developing authoritative information.

Recommendations to consider in planning a new system:

- Improve the accessibility and visibility of existing data management and metadata information tools.
- Improve and enhance the interactive map functionality on BGIS, focusing on the integration of geographic biodiversity information that allows users to query and retrieve data on the fly.
- Provide measures of reliability, accuracy (e.g. quality checks to indicate checked latitude/longitude values, citation analysis), currency and degree of completeness to facilitate trust.
- Develop annotation and feedback mechanisms of possible errors and data quality.

5.6 Recommendations for further research

The following questions were identified during the study and using statistical analysis, further research can be concluded on exploring the following:

- Are the findings from the target population representative of all SANBI's online conservation mapping service end-users?
- Are there significant differences between the rankings of biodiversity needs, the answer options of preferred information sources and the user's roles?
- How might the work sectors of respondents influence their selection of websites as information sources?

Research into the difficulty of separating the actual lack of availability of information from the lack of competence of the end-user in finding the appropriate information would provide insight into what a lack of availability means for the user (refer to section 4.2.4.1). Tracking user behaviour using system audits to gain a sense of where users are leveraging the system or not, could determine the end-users failure points and frustration (e.g. keyword search that if done differently could provide the necessary results?).

5.7 Conclusion

The purpose of this study was to gain an understanding of the information needs and information-seeking patterns of end-users of biodiversity information using SANBI's online conservation mapping service. Results showed that SANBI is only to a certain extent fulfilling the needs of biodiversity information users in South Africa and several challenges and priorities highlighted in Chapter 5 need to be addressed.

Using a survey, the research aimed at developing a set of recommendations for effectively supporting the information needs of biodiversity information end-users in South Africa. The intent behind the research produced outcomes and recommendations that are particularly focussed on facilitating the building biodiversity data infrastructure for science and decision-making. The analysis produced 25 recommendations or primary design requirements. Certain requirements, like developing a controlled vocabulary had already been documented in the literature. Others, such as a proactive advisory service for decision-makers, would be expected.

The results of this study explain how SANBI's online conservation mapping service users seek and navigate the landscape of biodiversity information, and recommendations were made accordingly. As the nature of needs and information-seeking behaviour of future users, or the potential changes in technology in shaping upcoming systems cannot be anticipated, we must get it right now so that future users have something to work with (Maceviciute *et al.,* 2010:11). The ongoing evolution of any biodiversity information system is based on continually changing information requirements and it is therefore vital that similar investigations are conducted frequently.

Going forward, it should be ensured that technologies or data already mobilised do not dictate the biodiversity questions to be addressed. As Peterson *et al.,* (2010) suggested, it is important that SANBI develop a conceptual framework to guide scientific questions and in this way leverage what is currently available and actively seek information that is still needed.

With the complex social-ecological issues the world is facing, it is vital to provide scientists, policy-makers and practitioners with the necessary information and tools to be better equipped at shaping and applying innovative solutions to environmental issues. By the consolidation and integration of information, it is believed that the gaps and shortcomings in the scientific workflows, and the relevant needs of the community, will be addressed. The monitoring and evaluation of end-users' information needs and use of information are vital to ensure that SANBI can anticipate how needs change and adapt accordingly (Nicholas & Herman, 2009:16).

References

Alford, J. 2009. Metadata challenges faced by producers and users of spatial data in South Africa. Dissertation Thesis. University of KwaZulu-Natal. Available: https://researchspace.ukzn.ac.za/xmlui/handle/10413/659 [2019, June 12].

Afzal, W. 2017. Conceptualisation and measurement of information needs: a literature review. *Journal of the Australian Library and Information Association*, 66(2):116–138. DOI:10.1080/24750158.2017.1306165.

Agosto, D.E. & Hughes-Hassell, S. 2005. People, places, and questions: an investigation of the everyday life information-seeking behaviours of urban young adults. *Library and Information Science Research*, 27(2):141-163. DOI:10.1016/j.lisr.2005.01.002.

Ariño, A.H., Chavan, V. & Faith, D.P. 2013. Assessment of user needs of primary biodiversity data: analysis, concerns, and challenges. *Biodiversity Informatics*, 8(2):59-93. DOI:10.17161/bi.v8i2.4094.

Ariño, A.H., Chavan, V., Macklin, J.A., Ghosh-Harihar, M., Mathur, V., Gaiji, S., & Flemons, P. 2013. *GBIF best practice guide for content needs assessment of stakeholder communities*, Ver. 1.0. Global Biodiversity Information Facility. Available: http://www.gbif.org/resources/3024.

Ariño, A. H., Chavan, V.S. & Otegui, J. 2016. *Best practice guide for data gap analysis for biodiversity stakeholders*. Global Biodiversity Information Facility. Available: http://www.gbif.org/resource/82566.

Asase, A. & Schwinger, G.O. 2018. Assessment of biodiversity data holdings and user data needs for Ghana. *Biodiversity Informatics*, 13:27–37. DOI:10.17161/bi.v13i0.7108.

Avlonitis, G. & Daly, B.G. 2014. Local Governments for Sustainability (ICLEI) / South African National Biodiversity Institute (SANBI), Global Biodiversity Information Facility (GBIF) mentorship project mobilising biodiversity data from local governments: an overview of challenges and opportunities. (Unpublished).

Babbie, E.R. & Mouton, J. 2001. *The practice of social research*. Oxford University Press.

Beaman, R.S. & Cellinese, N. 2012. Mass digitization of scientific collections: new opportunities to transform the use of biological specimens and underwrite biodiversity science. *ZooKeys*, 209:7–17. DOI:10.3897/zookeys.209.3313.

Bodin, Ö. & Prell, C. (Eds.). 2011. *Social networks and natural resource management: uncovering the social fabric of environmental governance*. Cambridge: Cambridge University Press.

Botts, E.A., Pence, G., Holness, S., Sink, K., Skowno, A., Driver, A., Harris, L.R., Desmet, P., Escott, B., Lötter, M., Nel, J., Smith, T., Daniels, F., Sinclair, S., Stewart, W. & Manuel, J. 2019. Practical actions for applied systematic conservation planning. *Conservation Biology*, 33(6):1235-1246. DOI:10.1111/cobi.13321.

Brooks, T.M., Butchart, S.H.M., Cox, N.A., Heath, M., Hilton-Taylor, C., Hoffmann, M., Kingston, N., Rodríguez, J.P., Stuart, S.N., & Smart, J. 2015. Harnessing biodiversity and conservation knowledge products to track the Aichi targets and sustainable development goals. *Biodiversity*, 16(2-3):157-174. DOI:10.1080/14888386.2015.1075903.

Brownlie, S., Walmsle, B. & Tarr, P. 2006. *Guidance document on biodiversity, impact assessment and decision making in southern Africa*. Southern African Institute for Environmental Assessment. Available: https://www.environment.gov.za/sites/default/files/docs/guidancedoconbiodiversity.pdf.

Case, D.O. 2005. Principles of lease effort. In *Theories of information behaviour.* K.E. Fisher, S. Erdelez & L. McKechnie, Eds. Medford, N.J.: Published for the American Society for Information Science and Technology by Information Today.

Case, D.O. 2006. *Looking for information: a survey of research on information seeking, needs, and behavior (Library and Information Science)*. 2nd ed. Edited by B. R. Boyce. Emerald.

Case, D.O. & Given, L.M. 2016. *Looking for information: a survey on the information seeking, needs and behaviour needs and behaviour*. 4th ed. Bingley, UK: Emerald.

Chapman, A.D. 2005. *Uses of primary species-occurrence data*, Ver. 1.0. Global Biodiversity Information Facility. Available: http://www.gbif.org/document/80545.

Cole, C. 2011. A theory of information need for information retrieval that connects information to knowledge. *Journal of the American Society for Information Science and Technology*, 62(7): 1216-1231. DOI:10.1002/asi.21541

Convention on Biological Diversity [CBD]. 2013a. Aichi Biodiversity Targets. Available: https://www.cbd.int/sp/targets/.

Convention on Biological Diversity. 2013b. Quick guide to the Aichi Biodiversity Targets, invasive alien species prevented and controlled. Available: https://www.cbd.int/doc/strategic-plan/targets/compilation-quick-guide-en.pdf.

Corrêa, P.L.P., de Moraes Batista, A.F, da Silva, D.L., Rodrigues, R.S., Frame, M., Morandini, M., Stanzani, S. & Correa, F. 2018. Evaluation of biodiversity data portals based on requirement analysis. *Ecological Informatics*, 48(May):215–225. DOI:10.1016/j.ecoinf.2018.09.008.

Cullen, P., Cottingham, P., Doolan, J., Edgar, B., Ellis, C., Fisher, M., Flett, D., Johnson, D., Sealie, L., Stocklmayer, S., Vanclay, F. & Whittington, J. 2001. *Knowledge seeking strategies of natural resource professionals*. (Technical report 2/2001). Cooperative Research Centre for Freshwater Ecology. ISBN:1876810661.

Davis, M.L.E.S. & Tenopir, C. 2012. Biodiversity information for natural resource management: Information needs and practices in Southern Appalachia. In *34th Annual College of Communication and Information Research Symposium*. Knoxville Tennessee, pp. 1–5.

Davis, M.L.E.S., Tenopir, C., Allard, S. & Frame, M.T. 2014. Facilitating access to biodiversity information: a survey of users' needs and practices. *Environmental Management,* 53(3)690-701. DOI:10.1007/s00267-014-0229-7

Dawson, D., Grant, A., Miller, P. & Perkins, J. 2004. User evaluation: sharing expertise to build shared values. In *Proceedings of the 8th Annual Museums and the Web Bibliography*. 31 March–3 April 2004, Arlington, Virginia / Washington DC, USA. 1–8. Available: https://www.museumsandtheweb.com/mw2004/papers/dawson/dawson.html.

Despot-Belmonte, K., Neßhöver, C., Saarenmaa, H., Regan, E., Meyer, C., Martins, E., Groom, Q., Hoffmann, A., Caine, A., Bowles-Newark, N., Bae, H., Canhos, D.A.L., Stenzel, S., Bowler, D., Schneider, A., Weatherdon, L.V. & Martin, C.S. 2017. Biodiversity data provision and decision-

making-addressing the challenges. *Research Ideas and Outcomes*, 3:e12165. DOI:10.3897/rio.3.e12165.

Dodds, L & Wells, P. 2019. Data infrastructure. In *The State of Open Data: Histories and Horizons*. T. Davies, S. Walker, M. Rubinstein & F. Perini, Eds. Cape Town and Ottawa: African Minds and International Development Research Centre.

Endangered Wildlife Trust [EWT]. 2011. Biodiversity data. Available: https://www.ewt.org.za/resources/resources-biodiversity-data/ [2020, February 2].

Etikan, I., Musa, S.A. & Alkassim, R.S. 2016. Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1):1-4. DOI:10.11648/j.ajtas.20160501.11.

Faith, D.P., Collen, B., Ariño, A.H., Koleff, P., Guinotte, J., Kerr, J. & Chavan, V.S. 2013. Bridging biodiversity data gaps: recommendations to meet users' data needs. *Biodiversity Informatics*, 8(2):41–58. DOI:10.17161/bi.v8i2.4126.

Ferrier, S. Jetz, W. & Scharleman, J. 2017. Biodiversity modelling as part of an observation system. In *The GEO Handbook on Biodiversity Observation Networks*. M. Walters & R. Scholes, Eds. Springer Nature, pp. 239–257. DOI: 10.1007/978-3-319-27288-7.

Fink, A. 2003. *How to manage, analyze and interpret survey data*. 2nd ed. SAGE Publications.

Fink, A. 2017. *How to conduct surveys: a step-by-step guide*. 6th ed. SAGE Publications.

Fisher, K.E., & Julien, H. 2009. Information behavior. *Annual Review of Information Science and Technology, 43*:1-73. DOI:10.1002/aris.2009.1440430114.

Fornwall, M. 2000. Planning for OBIS: examining relationships with existing national and international biodiversity information systems. *Oceanography*, 13(3):31-38. DOI:10.5670/oceanog.2000.06.

Groom, Q.J., Weatherdon, L.V. & Geijzendorffer, I.R. 2016. Is citizen science an open science in the case of biodiversity observations? *Journal of Applied Ecology*, 54(2):612-617. DOI:10.1111/1365-2664.12767.

Han, X., Smyth, R.L., Young, B.E., Brooks, T.M., de Lozada, A.S, Bubb, P., Butchart, S.H.M., Larsen, F.W., Hamilton, H., Hansen, M.C. & Turner, W.R. 2014. A biodiversity indicators dashboard: Addressing challenges to monitoring progress towards the Aichi biodiversity targets using disaggregated global data. *PLOS One*, 9(11). DOI:10.1371/journal.pone.0112046.

Hardisty, A., Roberts, D. & The Biodiversity Informatics Community. 2013. A decadal view of biodiversity informatics: challenges and priorities. *BMC Ecology*, 13(16):1-23. DOI:10.1186/1472-6785-13-16.

Hobern, D., Apostolico, A., Arnaud, E., Bello, J.C., Canhos, D., Dubois, G., Field, D., Alonso Garcia, E., Hardisty, A., Harrison, J., Heidorn, B., Krishtalka, L., Mata, E., Page, R.D.M., Parr, C., Price, J. & Willoughby, S. 2012. *Global biodiversity informatics outlook: delivering biodiversity knowledge in the information age*. Copenhagen: Global Biodiversity Information Facility. DOI: 10.15468/6jxa-yb44.

Hoffmann, A., Penner, J., Vohland, K., Cramer, W., Doubleday, R., Henle, K., Kõljalg, U., Kühn, I., Kunin, W.E., Negro, J.J., Penev, L., Rodríguez, C., Saarenmaa, H., Schmeller, D.S., Stoev, P., Sutherland, W.J., Ó Tuama, É., Wetzel, F.T. & Häuser, C.L. 2014. Improved access to integrated biodiversity data for science, practice, and policy - the European Biodiversity Observation Network (EU BON). *Nature Conservation*, 6:49–65. DOI:10.3897/natureconservation.6.6498

Houtkamp, J., La Rivière, I., de Groot, H., Janssen, S. & de Jong, A. 2016. From research data to web-based policy tools: user-centered design techniques in the development of the AgMIP Impacts Explorer. In *Proceedings of the 8th International Congress on Environmental Modelling and Software*. Sauvage, S., Sánchez-Pérez, J. & Rizzoli, A. Eds. July 2016, Toulouse, France. Paper 140. Available: http://scholarsarchive.byu.edu/iemssconference/2016/Stream-D/140.

Huang, X., Hawkins, B., Lei, F., Miller, G.L., Favret, C., Zhang, R. & Qiao, G. 2012. Willing or unwilling to share primary biodiversity data: results and implications of an international survey. *Conservation Letters*, 5(5):399–406. DOI:10.1111/j.1755-263X.2012.00259.x.

Ikoja-Odongo, R. & Mostert, R. 2006. Information seeking behaviour: a conceptual framework. *South African Journal of Libraries and Information Science*, 72(3):145-158. DOI:10.7553/72-3-1112.

Inskip, C., Butterworth, R. & MacFarlane, A. 2008. A study of the information needs of the users of a folk music library and the implications for the design of a digital library system. *Information*

Processing and Management, 44(2):647-662. Available: https://openaccess.city.ac.uk/id/eprint/1713

IUCN Species Survival Commission (SSC). 2009. Guidelines for appropriate uses of Red List Data. Gland, CH: IUCN SSC. Available: https://portals.iucn.org/library/sites/library/files/documents/Rep-2009-007.pdf (Unpublished).

Jamali, H.R. & Nicholas, D. 2010. Interdisciplinarity and the information-seeking behavior of scientists. *Information Processing & Management*, 46(2):233–243. DOI:10.1016/j.ipm.2009.12.010.

Janse, G. 2006. Information search behaviour of European forest policy decision-makers. *Forest Policy and Economics*, 8(6):579–592. DOI:10.1016/j.forpol.2004.10.001.

Jones, S., Pryor, G. & Whyte, A. 2013. *How to develop research data management services - a guide for HEIs. DCC How-to Guides.* Edinburgh: Digital Curation Centre. Available: http://www.dcc.ac.uk/resources/how-guides [2019, October 17].

Julien, H., Pecoskie, J.J.L. & Reed, K. 2011. Trends in information behavior research, 1999-2008: a content analysis. *Library and Information Science Research*, 33(1):19–24. DOI:10.1016/j.lisr.2010.07.014.

Kagan, J.S. 2006. Biodiversity informatics: challenges and opportunities for applying biodiversity information to management and conservation. *Northwestern Naturalist*, 87(1):80-85. DOI:10.1898/1051-1733(2006)87[80:BICAOF]2.0.CO;2.

Koesten, L. Kacprzak, E. Tennison, J. & Simperl, E. 2019. Collaborative practices with structured data: do tools support what users need? In *Proceedings of the CHI Conference on Human Factors in Computing Systems, May 4–9, 2019, Glasgow, Scotland UK. ACM, New York, NY, USA.* DOI:10.1145/3290605.3300330.

Kujala, S. 2002. *User Studies: A practical approach to user involvement for gathering user needs and requirements*. Ph.D. Thesis. Helsinki University of Technology. Acta Polytechnica Scandinavica Mathematics and Computing Series No. 116, Espoo 2002, 132 pp. Published by the Finnish Academies of Technology.

Line, M.B. 1969. Information requirements in the social sciences: some preliminary considerations. *Journal of Librarianship and Information Science*, 1(1):1-19. DOI:10.1177/096100066900100101.

Maceviciute, E. & Wilson, T.D. 2010. Information behaviour research and information systems development: the SHAMAN project, an example of collaboration. *Information Research*, 15(4) paper 445. Available: http://informationr.net/ir/15-4/paper445.html.

Meyer, H. 2016. Untangling the building blocks: a generic model to explain information behaviour to novice researchers. In *Proceedings of ISIC, the Information Behaviour Conference, Zadar, Croatia, 20-23 September, 2016: Part 1*. *Information Research, 21*(4), paper isic1602. Available: http://InformationR.net/ir/21-4/isic/isic1602.html

Miwa, M. 2005. Bandura's social cognition. In *Theories of information behaviour.* Fisher, K.E., Erdelez, S. & McKechnie, L., Eds. Medford, N.J.: Published for the American Society for Information Science and Technology by Information Today.

Mostert, B.J. & Ocholla, D.N. 2005. Information needs and information seeking behavior of parliamentarians in South Africa. *South African Journal of Library and Information Science*, 71(2):136-150. DOI:10.7553/71-2-616.

National Environmental Management: Biodiversity Act, No.10 of 2004. 2004. *Government gazette*. 467(700). 7 June. Government notice no. 26436. Cape Town: Government Printer.

Nicholas, D. & Martin, H. 1997. Assessing information needs: a case study of journalists. *ASLIB Proceedings*, 49(2):43–52. DOI:10.1108/eb051447.

Nicholas, D. 2000. *Assessing information needs: tools, techniques and concepts for the internet age*. 2nd ed. London: ASLIB.

Nicholas, D. & Herman, E. 2009. *Assessing information needs in the age of the digital consumer*. 3rd ed. London: Routledge. DOI:10.4324/9780203855799.

Nussbaumer, P., Slembek, I., Lueg, C., Mogicato, R. & Schwabe, G. 2009. Understanding information seeking behaviour in financial advisory. In *Proceedings of the International Symposium of Information Science (ISI)*. 1-3 April 2009, Constance, Germany. DOI:10.5167/uzh-25264.

Peacock, D. & Brownbill, J. 2007. Audiences, visitors, users: reconceptualising users of museum on-line content and services. In *Proceedings of the 11th Annual Museums and the Web Bibliography*. 11-14 April 2007, San Francisco, California. Available: https://www.museumsandtheweb.com/mw2007/papers/peacock/peacock.html.

Peterson, A.T., Knapp, S., Guralnick, R., Soberón, J. & Holder, M.T. 2010. The big questions for biodiversity informatics. *Systematics and Biodiversity*, 8(2)159-168. DOI:10.1080/14772001003739369.

Pocock, M.J.O. (2018) An assessment of the biodiversity information needs of the UK's environmental public bodies. *JNCC Report No. 618*, JNCC, Peterborough, ISSN 0963- 8091.

Roetman, P.E.J & Daniels, C.B. 2011. The benefits of citizen science in research, education and community engagement. In *Creating Sustainable Communities in a Changing World*. Roetman, P.E.J & Daniels, C.B., Eds. Crawford House Publishing, Adelaide, Australia.

Rowley, J. & Hartley, R. 2008. *Organizing knowledge: an introduction to managing access to information*. 4th ed. Burlington: Ashgate.

South African National Biodiversity Institute [SANBI] – http://www.sanbi.org

South African National Biodiversity Institute [SANBI]. 2016. *Lexicon of biodiversity planning in South Africa*. Beta Version, June 2016. South African National Biodiversity Institute, Pretoria. 72 pp.

South African National Biodiversity Institute [SANBI]. 2018. BGIS: biodiversity GIS. SANBI, Pretoria. Available from http://bgis.sanbi.org/ (accessed June 2018).

Savolainen, R. 2007. Information behavior and information practice: reviewing the "umbrella concepts" of information-seeking studies. *The Library Quarterly*, 77(2):109-132. Available: https://www.jstor.org/stable/10.1086/517840.

Savolainen, R. 2012. Conceptualizing information need in context. *Information Research* 17(4) paper 534. Available: http://www.informationr.net/ir/17-4/paper534.html#.VorJ1hV96hc.

Schuh, R.T. 2012. Integrating specimen databases and revisionary systematics. In *No specimen left behind: mass digitization of natural history collections*. Blagoderov, V., Smith, V.S. Eds. ZooKeys 209: 255-267. DOI:10.3897/zookeys.209.3288.

Sinclair, J., Mazzotti, F. & Graham, J. 2003. Motives to seek threatened and endangered species information for land-use decisions. *Science Communication*, 25(1):39–55. DOI:10.1177/1075547003255293.

Shanmughavel, P. 2007. An overview on biodiversity information in databases. *Bioinformation*, 1(9):367-369. DOI:10.6026/97320630001367.

Spencer, D. 2011. *A practical guide to information architecture*. 2nd ed. Edited by B. Harper. UX Mastery.

Steenkamp, Y. & Smith, G.F. 2002. Addressing the needs of the users of botanical information: proceedings of a national workshop for stakeholders and end-users of botanical information and herbaria. In *Southern African Botanical Network Report No. 15 SABONET*. Pretoria.

Steenkamp, Y. & Smith, G.F. 2003. Needs of users of botanical information in South Africa: outcomes of a national workshop for the stakeholders and end-users of botanical information and herbaria. *Taxon*, 52(2):303-306. Available: http://www.jstor.org/stable/3647401.

St. Jean, B., Rieh, S.Y., Yakel, E. & Markey, K. 2011. Unheard voices: institutional repository end-users. *College & Research Libraries*, 72(1):21-42.

Swart, R.J., de Bruin, K., Dhenain, S., Dubois, G., Groot, A. & von der Forst, E. 2017. Developing climate information portals with users: promises and pitfalls. *Climate Services*, 6(2017):12-22. DOI:10.1016/j.cliser.2017.06.008.

Tann, J., Kelly, L. & Flemons, P. 2008. Atlas of Living Australia user needs analysis. Canberra: Atlas of Living Australia. Available: http://www.ala.org.au/wp-content/uploads/2010/07/ALAUserNeedsAnalysisReportExtract.pdf. (Unpublished).

Taylor, R.S. 1968. Question-negotiation and information seeking in libraries. *College & Research Libraries*, 29(3), 178–194.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Manoff, E.M. & Frame, M. 2011. Data sharing by scientists: practices and perceptions. *PLoS ONE* 6(6):e21101. DOI:10.1371/journal.pone.0021101.

Thessen, A.E. & Patterson, D.J. 2011. Data issues in the life sciences. In *e-Infrastructures for data publishing in biodiversity science*. Smith, V. & Penev, L. Eds. ZooKeys. 150:15-51. DOI:10.3897/Zookeys.150.1766.

Thessen, A.E., Fertig, B., Jarvis, J.C. & Rhodes, A.C. 2016. Data Infrastructures for estuarine and coastal ecological syntheses. *Estuaries and Coasts,* 39, 295–310. DOI:10.1007/s12237-015-0045-1

Torres, R.D.S., Medeiros, C.B., Gonçalves, M.A. & Fox, E.A. 2006. A digital library framework for biodiversity information systems. *International Journal on Digital Libraries*, 6(1):3–17. DOI:10.1007/s00799-005-0124-1.

United Nations Framework Convention on Climate Change [UNFCCC]. 2015. Paris Agreement. Available: https://unfccc.int/sites/default/files/english_paris_agreement.pdf.

van Wilgen, B.W. & Wilson, J.R. (eds.) 2018. The status of biological invasions and their management in South Africa in 2017. South African National Biodiversity Institute, Kirstenbosch and DST-NRF Centre of Excellence for Invasion Biology, Stellenbosch.

Vissers, J., Van den Bosch, F., Bogaerts, A., Cocquyt, C., Degreef, J., Diagre, D., De Haan, M., De Smedt, S., Engledow, H., Ertz, D., Fabri, R., Godefroid, S., Hanquart, N., Mergen, P., Ronse, A., Stévart, T., Stoffelen, P., Vanderhoeven, S. & Groom, Q.J. 2017. Scientific user requirements for a herbarium data portal. *PhytoKeys*, 78:37–57. DOI:10.3897/phytokeys.78.10936.

Watters, C. & Duffy, J. 2005. Motivational factors for interface design. In *Theories of information behaviour.* Fisher, K.E., Erdelez, S. & McKechnie, L., Eds. Medford, N.J.: Published for the American Society for Information Science and Technology by Information Today.

Walters, M. & Scholes, R.J. 2017. *The GEO handbook on biodiversity observation networks*. Springer Nature. DOI:10.1007/978-3-319-27288-7.

Weiler, A. 2005. Information-seeking behavior in generation Y students: motivation, critical thinking, and learning theory. *The Journal of Academic Librarianship*, 31(1):46-53. DOI:10.1016/j.acalib.2004.09.009.

White, R. & Molina, R. 2006. The Pacific Northwest Research Station Biodiversity Initiative: scoping out the challenges in managing for biodiversity. *Northwestern Naturalist*. Society for Northwestern Vertebrate Biology, 87(1):10-17. Available: https://www.fs.usda.gov/treesearch/pubs/27246.

Wilson, T.D. 1981. On user studies and information needs. *Journal of Documentation*, 37(10):3-15. DOI:10.1108/eb026702

Wilson, T.D. 1994. Information needs and uses: 50 years of progress? In *Fifty years of information progress: a Journal of Documentation Review*. B.C. Vickery Ed. pp. 15- 51 London: ASLIB

Wilson, T.D. 1997. Information behaviour: an interdisciplinary perspective. *Information Processing & Management*, 33(4):551–572. DOI:10.1016/S0306-4573(97)00028-9.

Wilson, T.D. 1999. Models in information behaviour research. *Journal of Documentation*, 55(3), 249–270. DOI:10.1108/EUM0000000007145

Wilson, T.D. 2000. Human information behavior. *Information Science Research*, 3(2):49-55.

Wilson, T.D. 2005. Evolution in information behaviour modeling: Wilson's model. In: Fisher, K.E., Erdelez, S. & McKechnie, L. (Eds.) *Theories of information behavior*. Information Today.

World Conservation Monitoring Centre [WCMC]. 1998. *WCMC Handbooks on Biodiversity Information Management. Volume 1: Information and Policy*. Reynolds, J.H. (Series Editor). Commonwealth Secretariat, London. ix + 29pp.

Appendix 1: List of potential end-users of biodiversity information

The table below provides a list of potential end-users and their specific information need.

End-user	Information needs	Role	Reference
1. Ecological/environmental consultants, ecologist	▪ Species composition changes with distribution in a region. ▪ Conservation status, environmental preferences and tolerance of various species. ▪ Description and distribution of the vegetation types and threat status. ▪ Dominant structural elements and environmental preferences (e.g. geology, climate).	Practitioners	Steenkamp & Smith (2002:5)
2. Scientist (e.g. herbarium and museum curator, botanists, zoologists, horticulturists)	▪ Checklist of species with valid scientific names and attribute data. ▪ Species distribution models. ▪ Access to literature. ▪ Biogeography to answer fundamental questions about why certain species occur where they do, and why certain types of species occur together. ▪ Solutions to problems that arise when alien species become invasive. ▪ The condition of biodiversity and the factors that affect it.	Researchers	Steenkamp & Smith (2002:5) St. Jean *et al.,* (2011:31) CBD, 2013b IUCN, 2009

End-user	Information needs	Role	Reference
	▪ Images that reveal information on e.g. size, colour, habitat and feeding method of a species. ▪ Verification of field observations. ▪ Where to conduct species-specific survey work, ecological studies and research based on identified data gaps (e.g. Data Deficient species).		
3. Vegetation surveyors	▪ Species that grow together under certain environmental conditions (ecosystem information). ▪ Where surveys have been conducted and where further research is needed.	Practitioners	Steenkamp & Smith (2002:5)
4. Rehabilitation specialists	▪ Plants used for erosion control, which will grow on toxic or nutrient-poor soil. ▪ How to protect endangered species in areas earmarked for mining/development.	Practitioners	Steenkamp & Smith (2002:5)
5. Farmers and animal feed companies	▪ Palatability of grasses. ▪ Identification keys for weeds and toxic plants. ▪ Species composition and if anything endangered occurs on the property. ▪ Invasive species checklist.	Practitioners	Steenkamp & Smith (2002:5)
6. Landscape architects/gardeners	▪ Mature plant forms of a species. ▪ Species that can be transplanted, have attractive foliage, flowers	Practitioners	Steenkamp & Smith (2002:5)

End-user	Information needs	Role	Reference
	and/or fruit and what, if any, are poisonous.		
7. Interior designers	Species that are tolerant of drought and low light conditions.	Practitioners	Steenkamp & Smith (2002:5)
8. Nursery owners	Preferences and tolerances of species that naturally occur in an area.	Practitioners	Steenkamp & Smith (2002:5)
9. Outdoor enthusiasts	▪ Identification keys and guides to local flora and fauna. ▪ Plants and parts that are edible. ▪ Endangered and invasive species.	Researchers	Steenkamp & Smith (2002:5)
10. The general public including school children and teachers.	▪ Species descriptions for school projects (including invasive alien species). ▪ Images of biodiversity. ▪ Where a plant can grow and indigenous plants that have medicinal use, etc.	Researchers	Steenkamp & Smith (2002:5)
11. Citizen Scientist, enthusiastic amateurs, educators and retired professionals.	▪ Checklist of species with valid scientific names. ▪ Access to an up-to-date evidence-based list of sensitive species to mask these species online (species vulnerable to collecting, poaching or over-exploitation). ▪ Checklist of invasive species. ▪ Characteristics of environmental change.	Researchers	CBD, 2013b Groom, Weatherdon & Geijzendorffer 2016
12. Red List Scientists	Species distribution as part of the red list assessment process.	Researchers	Steenkamp & Smith (2002:5)
13. Reserve and National Parks Managers	▪ Species distribution in national and provincial parks. ▪ Species checklist for these areas.	Practitioners	

End-user	Information needs	Role	Reference
14. Graduate students	▪ Open access to published research or unpublished works (e.g. conference papers, reports, newsletters, etc.). ▪ Access to raw data underlying the findings presented in published articles or research project.	Researchers	St. Jean *et al.,* (2011:31)
15. Government (setting national targets)	▪ Checklist of invasive species with valid scientific names. ▪ Invasive alien species control or eradication programmes underway and ecosystems affected and how? ▪ Species are threatened in the development of national and regional threatened species lists and national biodiversity strategies and action plans?	Policy-makers	CBD, 2013b IUCN, 2009
16. Law enforcement officers	▪ Southern African Threatened and Protected Species (TOPS) and CITES-listed species, listed on one of the Appendices of the Convention on International Trade in Endangered Species (CITES). ▪ Access to species identifications keys to make critical law enforcement decisions such as whether to confiscate a species or to issue the relevant permit.	Practitioners	

End-user	Information needs	Role	Reference
17. Provincial conservation authorities	Species that are listed on TOPS and CITES listed species related to the issuing of permitting.	Practitioners	
18. Environmental Impact Assessment (EIA) Practitioner	▪ List of threatened species and ecosystems at a site-level. ▪ Data to compile risk assessment reports for impact assessment and permitting purposes.	Practitioners	IUCN, 2009
19. Conservation agencies	▪ Where to set conservation priorities for areas and species (e.g. Important Bird Areas, Critical Biodiversity Areas, biodiversity hotspots, etc.). ▪ Access to a current evidence-based list of sensitive species.	Practitioners	IUCN, 2009
20. Monitoring and evaluation practitioners	The overall status of biodiversity, varying status between regions and countries, the rate at which biodiversity is being lost, where loss is most prolific and the main drivers of this decline.	Practitioners	IUCN, 2009
21. Biodiversity planning practitioners	Access to existing biodiversity planning products as well as biodiversity information that is used during the production of biodiversity planning products.	Practitioners	
22. Decision-makers	If decisions are conserving biodiversity or leading to degradation and loss.	Policy-makers	
23. Custom, Border and Airport Officials	Tools to identify illegally traded and transported endangered species, and	Practitioners	

End-user	Information needs	Role	Reference
	the names of these species that are in use, and their synonyms.		
24. Environmental Management Inspectors (EMIs)	Identification keys to better regulate and monitor the trade in TOPS and CITES-listed species.	Practitioners	
25. Government agencies	To know the state of biodiversity in South Africa for the formulation of policies and legislation and action plans.	Policy-makers	
26. Collectors	Access to information to assess applications for permits in terms of Chapter 7 of the National Environmental Management: Biodiversity Act, 2004.	Practitioners	
27. National, provincial and municipal officials	Development of Spatial Development Frameworks (SDF's).	Practitioners	

Appendix 2: Survey questions

Introduction

I appreciate you taking the time to respond to this survey and would like to thank you in advance for your valuable input. Participation in this study is voluntary and all data will be treated confidentially.

You are welcome to contact the project team at the South African National Biodiversity Institute by e-mail with any further questions or comments on B.Daly@sanbi.org.za.

Time: I know your time is extremely valuable. This survey has 25 questions and should take approximately 15 minutes to complete.

Primary work sectors and subject disciplines

This section will identify the primary roles associated with biodiversity information.

1. *What is your primary work sector (choose the most correct response)?

- ◯ Academic / educational institution
- ◯ Non-Governmental Organisation (NGO)
- ◯ Community Based Organisation (CBO)
- ◯ State institution (e.g. SANBI, ARC, SAWS, CSIR, etc.)
- ◯ National government
- ◯ Provincial government
- ◯ Local government
- ◯ Research institution (incl. students)
- ◯ Prefer not to answer

Other (please specify)

2. *Subject discipline – field of work (tick as many as appropriate)

- ☐ Life sciences (study of living organisms e.g. bacteria)
- ☐ Physical sciences (study of natural non-living objects e.g. soil)
- ☐ Environmental sciences (study of interactions between the physical, biological and information sciences e.g. ecology)
- ☐ Agriculture and natural resources (e.g. forestry, fisheries, water resources, etc.)
- ☐ Professional/technical field
- ☐ Prefer not to answer

Other (please specify)

3. *Biodiversity-related work (tick as many as appropriate)?

- [ ] Research
- [ ] Education
- [ ] Information management
- [ ] Land conservation
- [ ] Marine conservation
- [ ] Coastal conservation
- [ ] Estuarine conservation
- [ ] Inland aquatic conservation
- [ ] Biodiversity related programme manager
- [ ] Biodiversity related programme implementer
- [ ] Elected, appointed and/or organisational decision maker
- [ ] Consultant
- [ ] Engineer
- [ ] Prefer not to answer

Other (please specify)

Biodiversity information needs

This section of the survey assesses existing biodiversity information to determine if it meets key needs and identifies information gaps.

4. What biodiversity data do you currently use? (Tick at least 3 items)

- [ ] Animal and plant descriptions
- [ ] Species checklists
- [ ] Maps and distribution data
- [ ] Flora and fauna surveys (site assessments)
- [ ] Threatened species data
- [ ] Gene sequences
- [ ] Invasive species data
- [ ] Citizen Science (images and observations)
- [ ] Images and videos
- [ ] Identification keys (e.g. apps, field guides, drawings, images)
- [ ] Ecological data (e.g. vegetation map)
- [ ] Forecasted trends, possible scenarios, speculations, predictions
- [ ] Planning / prioritisation outputs (e.g. Critical Biodiversity Areas (CBAs), Ecological Support Areas (ESAs), Protected Areas Expansion, etc.)

Other (please specify)

5. **How much of the biodiversity-related information needed to do your work consists of raw primary biodiversity data (e.g. taxonomic descriptions, location, photographs, ecological plot inventories, etc.)?**

1 = None 3 = Half 5 = All

6. **How much information consisting of synthesised/summarised secondary biodiversity data or data models of primary data (e.g. niche or habitat model) is needed to do your work?**

1 = None 3 = Half 5 - All

7. **At what geographic scale do you need biodiversity data? Please rank them in order of importance to you.**

Global

South Africa

Provincial

Local

8. **How much of the information needed to do your work relates specifically to biodiversity?**

None of the information needed to do the work relates specifically to biodiversity

At least half of the information

All

9. **What is the importance of data models (i.e. mathematically linking biodiversity variables) to your biodiversity-related work?**

- Not at all important
- Not so important
- Somewhat important
- Very important
- Extremely important

10. What data or information have you not been able to find?

11. What challenges are you facing when <u>not</u> able to find the information needed?

12. Can you give an example of a task, process or application where you use biodiversity data to achieve an outcome that is core to your work or study? Examples: mapping genetic variation across the landscape or analysing the distribution for an invasive species.

13. Please write down each step of the tasks.

14. List the types of data or information needed.

15. What sources and platforms will you find these data and information?

Information-seeking behaviours and preferred information source

The next set of questions determines whether sufficient biodiversity information exists and the ease and/or difficulty in finding biodiversity information. Determine information-seeking strategies, identify the attributes of preferred information sources visited and the importance of tools when working with biodiversity information.

16. When looking for biodiversity data and/or information where do you tend to go?

- [ ] Literature (journalpublications)
- [ ] Existing managementplans
- [ ] Online library catalogue
- [ ] Local expert opinion or subject specialist
- [ ] Online data / information sources
- [ ] Personal contacts

17. Which websites do you primarily use when searching for biodiversity data and/or information (e.g. Biodiversity Advisor, etc.)? Information sources regularly consulted.

18. Ease of finding needed biodiversity information

	Easy or extremely easy	Neither easy nor difficult	Difficult or very difficult
Raw data	○	○	○
Synthesised/summarised data	○	○	○
Data models	○	○	○

Other (please specify)

19. In completing the sentence, ‘‘When looking for the biodiversity information I need to do my work, I am limited by…,” (tick as many as appropriate).

- [ ] Lack of available information
- [ ] Lack of appropriate information
- [ ] Lack of time to seek or synthesise relevant information
- [ ] Knowing where to look
- [ ] Not knowing how to find what I need
- [ ] Poor searching skills

Other (please specify)

20. Please rate your level of agreement with eight statements about finding biodiversity information:

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
My organisation provides training to help me find the biodiversity information I need to do my work	○	○	○	○	○
I already have the biodiversity information I need to do my work	○	○	○	○	○
Readily accessible data exists in the various data repositories	○	○	○	○	○
I feel besieged with information, and totally overloaded	○	○	○	○	○
Knowing where to find the biodiversity information needed is a challenge	○	○	○	○	○
The tools for finding biodiversity information work well	○	○	○	○	○
Good prior knowledge of where to find biodiversity data is needed	○	○	○	○	○
It will be useful to showcase and/or provide a set of case studies on how other institutions have used biodiversity data	○	○	○	○	○

Other (please specify)

21. Ease of finding the following biodiversity information tools

	Very easy	Easy	Neither easy nor difficult	Difficult	Very difficult
Decision support tool - information resource designed to support conservation decisions e.g. Land Use Decision Support (LUDS).	○	○	○	○	○
Metadata – structured information that describes the content, quality, condition and features of digital objects e.g. datasets, images, publications, etc.	○	○	○	○	○
Documentation – information resource of documents and/or literature	○	○	○	○	○
Visualisation tool - a visual representation of abstract data to better understand the data	○	○	○	○	○
Data management e.g. R analytical environment, niche models	○	○	○	○	○
Mapping - export lat-long data to permit the easy visualisation of those data and the creation of maps	○	○	○	○	○
Information search – link directly to content, and allow for refined secondary searches by geography or subject matter	○	○	○	○	○

22. Importance of the following biodiversity information tools

	Not at all important	Not so important	Somewhat important	Very important	Extremely important
Decision support	○	○	○	○	○
Metadata	○	○	○	○	○
Documentation	○	○	○	○	○
Visualisation	○	○	○	○	○
Data management	○	○	○	○	○
Mapping	○	○	○	○	○
Information search	○	○	○	○	○

23. Please rate the importance of ten information source attributes

	Not at all important	Not so important	Somewhat important	Very important	Extremely important
Currency: the information is up-to-date	○	○	○	○	○
Completeness: the information is comprehensive – not just a portion of what is potentially available	○	○	○	○	○
Accuracy: an understanding of data accuracy – particularly in relation to geography and taxonomy	○	○	○	○	○
Scale: the information is provided at the geographic or temporal scale needed	○	○	○	○	○
Usability: the information is easy to use	○	○	○	○	○
Trusted and reliable source: the information is authoritative	○	○	○	○	○
Provenance: the information is well documented	○	○	○	○	○
Navigation: the information is easy to find	○	○	○	○	○
Validation: there is some measure of validation of data (a process to ensure data have undergone data cleansing to ensure data quality)	○	○	○	○	○
Salience: the information provided is perceived as relevant	○	○	○	○	○

Other (please specify)

Policy and management

This final section aims to determine the relationship between biodiversity information that you require and policy and decision-making processes.

24. **What other types of data or information do you need to find when addressing environmental problems (e.g. social, political, economic, physical features such as climate, soils, hydrology, etc.)?**
    Example: Sea-level rise dataset - to determine the effects of sea-level rise on coastal breeding habitats

25. **How often do you need to look for and use the results of research by people in other disciplines (e.g. chemistry etc.)?**

- ◯ Always
- ◯ Often
- ◯ Sometimes
- ◯ Rarely
- ◯ Never

Other (please specify)

26. Please provide your name and contact information so that I can contact you if I have any questions. If you provide this information, I will share a summary of the survey results with you directly. This information is OPTIONAL and CONFIDENTIAL.

Name	
Company	
Address	
Address 2	
City/Town	
State/Province	
ZIP/Postal Code	
Country	
Email Address	
Phone Number	