UFS UV BUSINESS SCHOOL SAKESKOOL BE WORTH MORE WEES MEER WERD

A COMPETENCY MODEL FOR DATA SCIENTISTS IN GRAIN SA

YOLANDI KRUGER

Field study submitted to the UFS Business School

in the Faculty of Economic and Management Sciences in partial fulfilment of

the requirements for the degree of

MAGISTER

in

BUSINESS ADMINISTRATION

at the

University of the Free State

Supervisor: Prof. M. Kotzé

Co-supervisor: Mr J.F. de Villiers

2015

DECLARATION

I declare that the field study hereby handed in for the qualification Masters in Business Administration at the UFS Business School at the University of the Free State is my own independent work and that I have not previously submitted the same work, either as a whole or in part, for a qualification at another university or at another faculty at this university.

I also hereby cede copyright of this work to the University of the Free State.

Name: Yolandi Kruger

Date: November 2015

ACKNOWLEDGEMENTS

I would like to acknowledge and express my most sincere gratitude towards the following persons for their role in my life as well as their support in my journey to obtaining my MBA degree. The Lord who guides, protects, supports and enables me; Prof. Tina Kotzé, my supervisor, for her outstanding guidance, support and mentorship throughout the whole process; Mr Jannie de Villiers of Grain SA, my co-supervisor, who came up with the idea of applying data science to South African agriculture, and for his leadership; Monsanto, especially Kobus Steenekamp and Magda du Toit, for sponsoring the visit to Monsanto and Climate Corporation in the USA; Grain SA for sponsoring the remainder of my visit to the USA and enabling more extensive data collection; and my family, especially my parents, for their support, words of encouragement, understanding and motivation.

I would like to dedicate this field study to my grandmother, Martie Kotzé, who passed away before she could see the end results. She set the example, believed in the power of lifelong education as well as my ability to obtain this degree. She motivated and prayed for successful completion of this chapter of my life. This one is for you!

ABSTRACT

With the current global population growth and the consequential increasing demand for food, agricultural productivity needs to increase. Grain SA is a leading role player in the agricultural industry and needs to serve the grain producers in South Africa effectively. Data science is a fairly new concept and is described as the management of large data sets from disparate sources to show results which assist in informed decision making. It is believed that the application of data science principles in agriculture may deliver many benefits, including increased productivity and profitability.

Since data science is a new discipline that has not yet been implemented in Grain SA, it would need to be introduced to farmers and the agribusiness as a whole and the implementation thereof would need to be monitored. To capitalise on "big data", Grain SA would be required to recruit and appoint a data scientist with the necessary skills and expertise to manage and distribute large data sets. The aim of this study is to conceptualise a competency model for data scientists in Grain SA.

The adopted approach for the research was qualitative. Since the field of data science in agriculture is fairly new and information on the topic is very limited, the use of an exploratory study method was most suitable. The researcher conducted face-to-face interviews with 20 participants from nine organisations. The participants included individuals who are data scientists or work closely with data scientists. The interviews were conducted in the USA because the nation plays a leading role in agricultural innovation and offers a rich source of information for researchers in the field.

The current role of data science in agriculture was explored by means of a literature review and an empirical study. The study describes the core competencies of a data scientist in agriculture and, based on this information, articulates the role of a data scientist in Grain SA.

The core competencies of an effective data science coordinator in Grain SA are conceptualised and the development of the new competency model for data science coordinators in Grain SA is discussed in detail.

Keywords: *competencies, competency model, competency-based approach, data science, data scientist, Grain SA, agriculture, data science coordinator.*

TABLE OF CONTENTS

DECLARATION **2**

ACKNOWLEDGEMENTS **3**

ABSTRACT **4**

LIST OF TABLES **8**

LIST OF FIGURES **8**

ACRONYMS AND ABBREVIATIONS **9**

1. CHAPTER 1 – INTRODUCTION AND PROBLEM STATEMENT **10**

1.1. INTRODUCTION 10
1.2. PROBLEM STATEMENT 13
1.2.1. RESEARCH QUESTIONS 13
1.2.2. RESEARCH OBJECTIVES 14
1.2.3. STRUCTURE OF STUDY 14

2. CHAPTER 2 – LITERATURE REVIEW **16**

2.1. INTRODUCTION 16
2.2. THE DEVELOPMENT OF A COMPETENCY MODEL 16
2.2.1. INTRODUCTION 16
2.2.2. DEFINITIONS OF KEY TERMS 16
2.2.3. A COMPETENCY-BASED APPROACH 20
2.2.4. PROCESS FOR THE DEVELOPMENT OF A COMPETENCY MODEL 21
2.3. DATA SCIENCE IN THE AGRICULTURAL SECTOR 29
2.3.1. INTRODUCTION 29
2.3.2. OVERVIEW OF THE GRAIN INDUSTRY 30
2.3.3. DEFINITION OF DATA SCIENCE 32
2.3.4. THE ROLE OF DATA SCIENCE IN AGRICULTURE 33
2.3.5. THE ROLE AND COMPETENCIES OF A DATA SCIENTIST 37
2.4. SUMMARY 40

3. CHAPTER 3 – RESEARCH DESIGN AND METHODOLOGY **41**

3.1. INTRODUCTION 41
3.2. CONCEPTUAL FRAMEWORK 41
3.2.1. RESEARCH DESIGN 42
3.2.2. RESEARCH STRATEGY 43

3.2.3. SAMPLING STRATEGY 43
3.2.4. DATA COLLECTION STRATEGY 45
3.2.5. DATA ANALYSIS STRATEGY 46
3.2.5.1. CATEGORISING STRATEGIES 46
3.2.5.2. CONNECTING STRATEGIES 46
3.2.5.3. MEMOS AND DISPLAY 46
3.3. RELIABILITY AND VALIDITY OF THE STUDY 47
3.3.1. CREDIBILITY 47
3.3.2. TRANSFERABILITY 47
3.3.3. DEPENDABILITY 48
3.3.4. CONFORMABILITY 48
3.4. ETHICAL CONSIDERATIONS 48
3.5. TIME FRAME AND SETTING 49
3.6. SUMMARY 49

4. CHAPTER 4 – RESULTS AND DISCUSSION 50

4.1. INTRODUCTION 50
4.2. RESULTS 50
4.2.1. THE CURRENT ROLE OF DATA SCIENCE IN AGRICULTURE 50
4.2.2. THE CORE COMPETENCIES OF A DATA SCIENTIST IN AGRICULTURE 56
4.2.3. THE ROLE OF A DATA SCIENTIST IN GRAIN SA 68
4.2.4. THE COMPETENCIES OF A DATA SCIENCE COORDINATOR IN GRAIN SA 73
4.2.5. A PROPOSED COMPETENCY MODEL FOR A DATA SCIENCE COORDINATOR IN GRAIN SA 76
4.3. SUMMARY 81

5. CHAPTER 5 – CONCLUSIONS AND RECOMMENDATIONS 83

5.1. INTRODUCTION 83
5.2. CONCLUSION 83
5.3. RECOMMENDATIONS 85
5.4. LIMITATIONS 85

REFERENCE LIST 87

ADDENDUM: BRIEFING SHEET FOR INTERVIEWS 95

LIST OF TABLES

Table 2.1 Specific competency modelling stages and steps (Brits, 2012).............. 28

Table 2.2 Production of different crops in South Africa (CEC, 2015) 31

Table 3.1 Sample participants... 44

Table 4.1 Identified categories for the current role of data science in agriculture .. 51

Table 4.2 Core competencies of a data scientist in agriculture in terms of knowledge, skills and attributes... 58

Table 4.3 Different types of data scientists, as described by the participants 69

Table 4.4 Different types of data scientists – combined categories........................ 70

Table 4.5 A proposed competency model for data science coordinators 77

LIST OF FIGURES

Figure 2.1 A competency as a cluster of related knowledge, skills and attributes (Brits, 2012)……………………………………………………………17

Figure 2.2 Competency terminology and how it is connected – an example (Brits, 2012)………………………………………………………………………20

Figure 2.3 Proposed process for developing a competency model with specific stages (Brits, 2012)…………………………………………………… 22

Figure 2.4 Advances in crop production (Monsanto, 2014b)…………………….. 35

Figure 2.5 Sources and application of data – an example from Tom Farms (Tom, 2014)……………………………………………………………………. 36

Figure 4.1 A proposed process for developing a competency model with specific stages and steps (Brits, 2012)…………………………………….... 56

Figure 4.2 An example of competency terminology and connected subcategories (Brits, 2012)……………………………………………………………. 65

Figure 4.3 A graphical representation of the core competencies for a data scientist in agriculture in terms of competency clusters, domains and individual competencies..67

Figure 4.4 A graphical representation of the core competencies for a data science coordinator in Grain SA in terms of competency clusters, domains and individual competencies..75

ACRONYMS AND ABBREVIATIONS

BFAP	Bureau for Food and Agricultural Policy
CEC	Crop Estimates Committee
CIPD	Chartered Institute of Personnel and Development
CTSI	Clinical and Translational Science Institute
FAO	Food and Agriculture Organization
GDP	Gross Domestic Product
GPS	Global Positioning System
HRSG	Human Resource Systems Group
IFAD	International Fund for Agricultural Development
OECD	Organisation for Economic Co-operation and Development
RFID	Radio-Frequency Identification
SA	South Africa
TMA	Talent Motivation Assessment
UFS	University of the Free State
WFP	World Food Programme
WSU	Wayne State University

1. CHAPTER 1 – INTRODUCTION AND PROBLEM STATEMENT

1.1. INTRODUCTION

As the world's population grows, the demand for food increases and more pressure is placed on natural resources. By 2050, agricultural productivity needs to increase by 70% despite the limited natural resources available to feed the growing world population (Syngenta, 2014). The agricultural sector is one of the main contributing industries to the Gross Domestic Product (GDP) of many countries through value adding to raw materials (Goldblatt, 2012) and it plays a crucial role in ensuring the food security of all countries. Recent forecasts indicate the world population will grow from the current seven billion to approximately 9.1 billion people in 2050 (Food and Agriculture Organization (FAO), 2003; Syngenta, 2014). Investment in agriculture is required for "promoting agricultural growth, reducing poverty and hunger, and promoting environmental sustainability" (Biodiversity et al., 2012, p. 13). Authorities recognise that agriculture holds the key to food security and an increase in food production is needed to feed a growing population. Therefore, agriculture is receiving greater attention all over the world.

According to Syngenta (2014), the global demand for grain specifically has increased by almost 90% since 1980 and will continue increasing at an average rate of 1.4% per year, with a current annual demand of almost 2.3 billion tons. To meet this demand, farmers worldwide need to produce around 1.4% more grain annually (Syngenta, 2014). This situation is applicable in South Africa where grain plays an important role in ensuring food security for the nation. South Africa produced 14.3 million tons of maize during the 2014/15 production season, of which 10.24 million tons were consumed locally (Grain SA, 2015a).

One of the organisations concerned with the well-being of the grain industry in South Africa is Grain SA. The organisation's mission is to provide strategic commodity services to South African grain producers to support sustainability, continuous production and a food-secure country (Grain SA, 2015b). Grain SA is a commodity organisation owned by its members – the grain producers of South Africa. It is involved in all matters affecting the profitability and sustainability of grain production. Due to its competence and leading role in the industry, the organisation is acknowledged, locally and internationally, as the grain producers' only and official voice (Grain SA, 2015b).

Grain SA has been an important role player in research and development in the grain industry (Grain SA, 2013). Previously, farmers did not have access to the necessary market information to ensure effective decision making regarding their farms and the organisation identified this gap in the grain industry. In order to close this gap, Grain SA started to compile and distribute a daily market report to all its members. Subsequently, other agribusinesses followed Grain SA's approach and now make sufficient and timely information available to more industry players. Grain SA has gathered large quantities of data over the past decade by means of research including market research, new technology, industry reports, conservation agriculture and information on inputs necessary for the production of grain. The organisation is dedicated to investigating any events and factors that might affect agriculture and specifically the grain industry.

For Grain SA to meet its goal of providing continuous strategic support to farmers in the 21st century, the organisation needs to stay up to date with relevant technological innovations. Therefore, De Villiers (2013, p. 15) points out that "Grain SA's challenge is to continuously think creatively and innovatively about how the issue of the grain producer and agriculture can be promoted at various levels." A new opportunity identified by Grain SA is the optimal use of all relevant available research data by means of the application of data science (De Villiers, 2013). Grain SA's agricultural economists and scientists have access to large amounts of data that is not always tapped into for use in effective decision making at all levels. Applying the principles and tools of data science will allow grain producers to access more consumable, accurate and timely information that can be used as a basis for making informed choices.

Data science is a fairly new concept to the business world but is not new to statisticians and analysts. Data science concepts have originated from a combination of different, existing disciplines (Stanton, Palmer, Blake & Allard, 2012; Loukides, 2012; Harris, Murphy & Vaisman, 2013). Data science is the management of large data sets from disparate sources to generate specific results which assist in informed decision making (Umachandran, 2013). It has been implemented very successfully in other industries and sectors such as healthcare, retail, manufacturing and the public sectors as well as by Google (Loukides, 2012; Manyika et al., 2011).

Ellis (1999, in Shiya, 2011) argues that decision making forms the basis of management in farming. Farmers rely on both external and internal sources for effective decision making (Shiya, 2011). Therefore, the application of data science in agriculture will enable farmers to analyse different variables, such as weather challenges, soil health, weed management, insect and disease management, simultaneously and enable them to make quick, informed decisions for better farming outcomes and improved productivity (Monsanto, 2014a). If data science is applied in modern production, producers would have the potential to use natural resources more efficiently, increase potential crop yields and evaluate past, present and future management decisions based on the analysis of the data they have generated in the field (Monsanto, 2014a).

Since data science is a new discipline that has not yet been implemented within Grain SA, it must be introduced to farmers and the agricultural sector as a whole and its implementation and efficacy should be monitored. To capitalise on big data, Grain SA would be required to recruit and appoint a data scientist with the necessary skills and expertise to manage and distribute the information. According to Davenport and Patil (2012), one true challenge for any organisation is to identify and attract talented personnel to the organisation and ensure they are productive. In order for Grain SA to do so, the role and required competencies of an effective data scientist within the agricultural sector need to be clarified.

Competencies are general descriptions of the underlying knowledge, skills, abilities and other characteristics needed by people to ensure worthy performance in their occupation (Coetzee & Schreuder, 2013). In other words, competencies are the set of behaviours instrumental in the delivery of desired organisational results or outcomes (Brits, 2012). Various authors (Bartram, 2012; Brits, 2012; Campion et al., 2011) collectively refer to related sets of knowledge, skills and abilities as a competency model and describe it as a selection of competencies required by a specific occupational group – in this case for a data scientist in Grain SA. The development of a competency model allows an organisation to “identify the behaviours that drive successful performance and enables the organisation to deliver their technical expertise effectively" (Chartered Institute of Personnel and Development (CIPD), 2014, p. 2). From the literature it is evident that a competency model forms the basis for the recruitment and evaluation of potential candidates for specific positions (Campion et al., 2011).

1.2. PROBLEM STATEMENT

In light of the above, the growing need for the use of data as a scientific tool to improve the effectiveness of decision making within the agricultural sector becomes evident. However, there is no model currently available that captures the required role and competencies of a data scientist in Grain SA or for any other organisation in the grain industry in South Africa. Although the role and competencies of data scientists in other industries have been discussed to some extent in the literature (Davenport & Patil, 2012; Harris et al., 2013; Loukides, 2012; Sanders, 2013; Stanton et al., 2012), an extensive search revealed no literature or empirical research regarding the role and competencies of a data scientist within the grain sector. For Grain SA to use data science as a vehicle to add value in terms of the sustainability, profitability and continuous production of grain in South Africa, a competency model, including the required role and competencies of an effective data scientist within the grain sector, would need to be developed. This will help avoid a situation in which unqualified individuals are appointed, training is ineffective or qualified people are poorly managed which will ultimately result in Grain SA missing its primary goal of providing strategic support to grain producers.

1.2.1. RESEARCH QUESTIONS

In order to operationalise the research, the following research questions have been formulated:

1. What is the current role of data science in agriculture?
2. What are the core competencies for a data scientist in agriculture?
3. What is the role of a data scientist in Grain SA?
4. What would be included in a competency model for data scientists in Grain SA?

1.2.2. RESEARCH OBJECTIVES

On the basis of the research questions, the following objectives have been formulated:

1. Explore the current role of data science in agriculture.
2. Describe the core competencies of a data scientist in agriculture.
3. Formulate the role of a data scientist in Grain SA.
4. Conceptualise the competencies of a data scientist in Grain SA for the development of a competency model.

1.2.3. STRUCTURE OF STUDY

The purpose of **chapter 1** is to provide an introduction and background to the study. It starts by giving the reader a brief overview of the agricultural sector, its contribution to food security and the challenge farmers face of increasing production with limited resources to feed a growing population.

The role of Grain SA is explained in terms of the grain industry of South Africa and how the organisation supports sustainability, profitability and continuous production through the services they render to grain farmers. A need to appoint a data scientist in Grain SA has been identified as well as the lack of a competency model to guide the human resource process.

The world of data science and data scientists is briefly considered and the link between data science, competencies, competency models and Grain SA is explained. Background to the problem statement is given and the research questions as well as the research objectives of the study are listed.

Chapter 2 presents a brief review of the literature relating to competencies and competency modelling. Definitions of key terms, the role of competency modelling and the process for the development of a competency model are discussed. The second part of the chapter discusses the use of data science in the agricultural sector and gives an overview of the grain industry, the definition of data science, the role of data science in agriculture and the role and competencies of a data scientist. The selected research design and methodology is based on the literature review.

Chapter 3 explains the research design and methodology used to investigate the problem statement and research questions. It lists the research objectives and discusses the conceptual framework in terms of the research design, research strategy, sampling strategy, data collection strategy and data analysis strategy. The chapter further describes how the reliability and validity of the study is tested with regard to credibility, transferability, dependability and conformability. Ethical considerations and the demarcation of the field of study conclude the chapter.

Chapter 4 presents the research findings and answers the research questions. The current role of data science in agriculture is explained through the eyes of the research participants. Firstly, the core competencies necessary for an effective data scientist in agriculture are identified and discussed in terms of knowledge, skills and attributes. Secondly, the competencies are examined according to competency domains, domain clusters and individual competencies. The remaining part of the chapter focusses on the formulation of a role for data scientists in Grain SA as well as the competencies of a data science coordinator in Grain SA and the development of an overarching competency model.

Chapter 5 is the final chapter and consists of the conclusion, limitations and recommendations of the study.

2. CHAPTER 2 – LITERATURE REVIEW

2.1. INTRODUCTION

The previous chapter gave a brief introduction to the agricultural sector and its challenges, the role Grain SA plays in the grain industry of South Africa, the use of data science and the need to develop a competency model for a data scientist in Grain SA. The purpose of this chapter is to discuss definitions of key terms relating to competencies and competency models and to explain why a competency-based approach is followed. The process utilised in the development of a competency model is outlined. The final two sections of chapter 2 examine the role of data science and data scientists in agriculture.

2.2. THE DEVELOPMENT OF A COMPETENCY MODEL

2.2.1. INTRODUCTION

Competencies are increasingly implemented in the lives of individuals, employees, career practitioners, team leaders and organisational managers and leaders (Brits, 2012). Due to their growing importance, it is necessary to understand what competencies mean, as well as to have a vocabulary and framework for conceptualising and discussing this important Human Resources concept. For this purpose, a short discussion on the definitions of competencies, competency models and competency modelling follows.

2.2.2. DEFINITIONS OF KEY TERMS

2.2.2.1. COMPETENCIES

The study and use of competencies is not a recent phenomenon but dates back as early as the 1950s when Flanagan (1954) introduced key methodologies. Later, White (1959) identified and explained the human trait he called “competence”. However, McClelland (1973) “is credited with coining the term competency” (Scheer, Cochran, Harder & Place, 2011, p. 65) in the context of his studies on the factors that lead to superior performance.

Roach (in Sephton & Kemp, 2014, p. 43) defines a competency as a "state of having the knowledge, judgement, skill, energy, experience and motivation required to respond adequately to the demands of one's professional responsibilities." Competencies are most commonly defined as knowledge, skills and attributes required to adequately respond to the demands of professional responsibilities (Akkermans, Brenninkmeijer, Huibers & Blonk, 2012; Bell, Lee & Yeung, 2006; Brits, 2012; Campion et al., 2011; Lubbe, 2010). **Attributes** include the personal characteristics, traits, motives and values or ways of thinking that affect an individual's behaviour. **Knowledge** is the factual information that a person knows and that is needed for a specific job (Akkermans et al., 2012). A **skill** is an ability that has been acquired by training and education (Brits, 2012; Croucamp, 2013) enabling a person to consistently perform a complex task accurately, effectively and efficiently (Murphy, 2010). A skill is thus the demonstration of a particular talent (Mirabile, 1997).

Figure 2.1 below illustrates the conception of a competency as a cluster of related knowledge, skills and attributes.

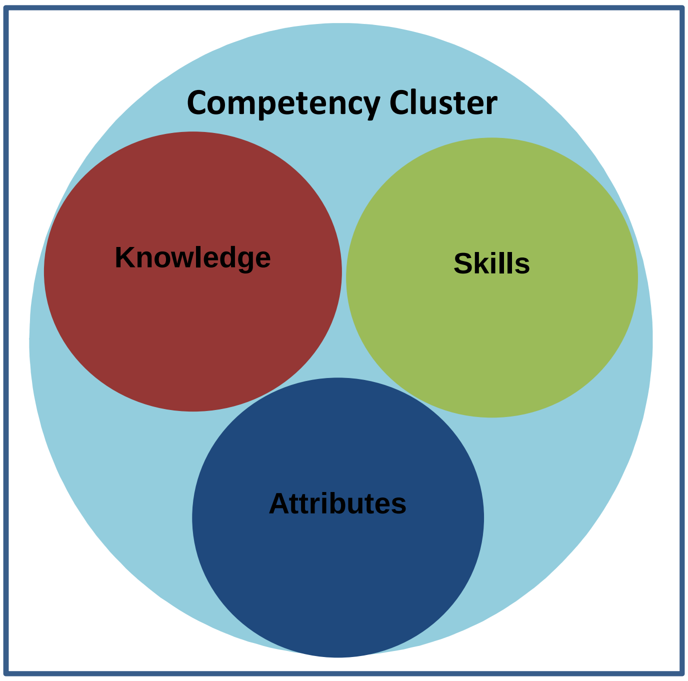

Figure 2.1: A competency as a cluster of related knowledge, skills and attributes (Brits, 2012)

The relationship between competencies and work outputs, but not activities, seems to be mentioned often in the literature. Campion et al. (2011) and Lubbe (2010) suggest that knowledge, skills and characteristics should correlate with job performance and should be measurable against well-accepted standards. Dubois (2007) asserts that competencies can be defined as personal characteristics which drive superior job performance, while Edgar and Lockwood (2011) describe core competencies as the capabilities held by people. When applied correctly in the organisation, competencies will make a critical contribution to the competitiveness of the firm (Edgar & Lockwood, 2011). Also, Bell et al. (2006, p. 410) state "competencies involve the knowledge required to achieve a given outcome, the skills to implement that knowledge, and the personality characteristics required to motivate the implementation of the knowledge and skills to achieving a desired outcome."

There seems to be a general agreement that competencies are of a more comprehensive or even strategic nature than job tasks (Brits, 2012). This is supported by Buford and Linder (2002, p. 3) who describe a competency as "a validated decision tool of activities that described key knowledge, skills, and abilities for performing those activities." From these arguments it is apparent that the link between learned behaviour (knowledge and skills) and the achievement of success in the workplace is paramount. It implies that there should be a transfer of knowledge to the work place to ensure "operational competence" (Brits, 2012). Therefore, for the purpose of this study, competencies are defined as composites of knowledge, skills and attributes that lead to the worthy performance of occupational groups in an organisation (Brits, 2012).

2.2.2.2. COMPETENCY MODEL AND COMPETENCY MODELLING

Most researchers agree on the definition of a competency model and refer to it as a collection or cluster of required competencies for effective performance in a particular job, job family or functional area (Brits, 2012; Campion et al., 2011; Lubbe, 2010). Mirabile (1997, p. 75) describes a competency model as "the output from analyses that differentiate high performers from average and low performers."

A competency model is also defined as a selection of competencies required by a specific occupational group. This approach is supported by Parry (1996, p. 52) who argues that “a list of competencies is more useful to job holders and their managers if similar competencies are grouped under broad headings”. According to Brits (2012), the development of a competency model could help to link the competencies of individuals with those of the organisation. The identification of core competencies related to occupational groups in an organisation is therefore a logical step preceding the identification of job competencies for individual job holders. This would ensure the competency model serves as the foundation for personnel management processes which in turn supports the business strategy of the organisation. Therefore, competency modelling is an efficacious way to identify behavioural expectations that should form part of strategic personnel management frameworks and processes (Brits, 2012).

Lubbe (2010) refers to a competency model as “an integrated cluster of competencies required in the selection process and competency modelling denotes the process followed to design and develop a competency model.” Stevens (2012) states that competency modelling focuses on the future roles that align with a strategic plan and defines maximum performance in those roles through worker attributes. These attributes range from characteristics common to those in a particular role to attributes of an occupational group (Stevens, 2012). A competency model is also seen as a decision-making tool which describes the key capabilities for performing a specific job (McLagan, 1997). Yet, managers and employees should not only understand the model and its relationship to the job, but also be able to apply it (Brits, 2012).

According to Garrett (n.d.), a competency model consists of the collection of success factors necessary for achieving results in a specific job (work role) in a particular organisation. For the purposes of this study, the definition of a competency model by Brits (2012) is applicable. This author defines a competency model as a cluster of competency domains and its associated competencies required by a specific occupational group. A competency domain is regarded as a collective name for a group of similar competencies, while a competency cluster is seen as a collective name for a set of competency domains (Brits, 2012). Figure 2.2 illustrates how competencies, competency clusters and competency domains are connected.

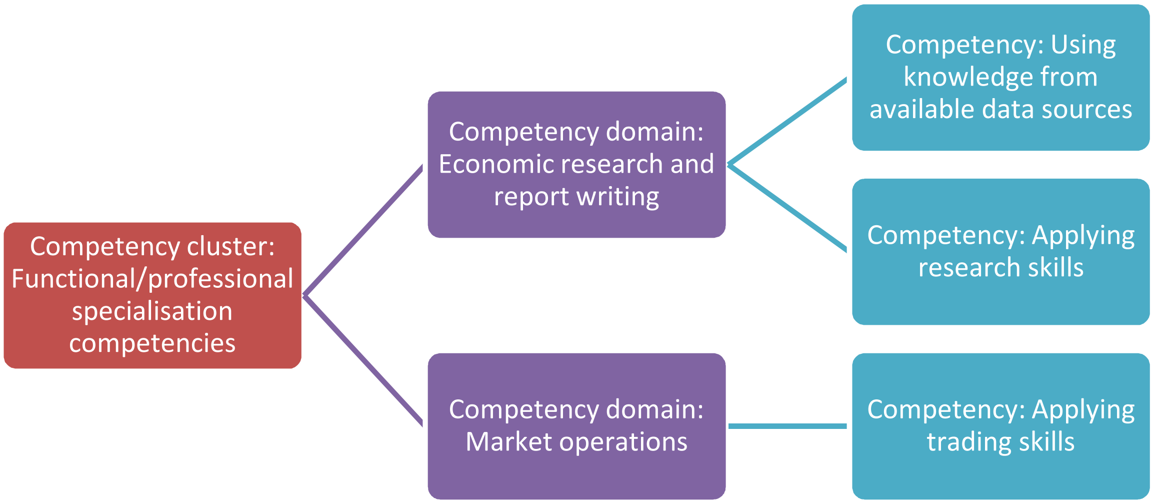

Figure 2.2: Competency terminology and how it is connected – an example (Brits, 2012)

In terms of the benefits of competency models, Brown (2006) asserts that the resources devoted to developing and implementing competency models have the potential to significantly impact employee performance. Other advantages of developing a competency model include increased productivity, effective training, reduced costs, reduced staff turnover and better organisational performance resulting in a competitive advantage (Robinson et al., in Lubbe, 2010). Furthermore, when used effectively, it can provide the organisation with a flexible and dynamic base that increases competitive advantage (Soderquist, Papalexandris, Loannou & Prastacos, 2010). Although the process of identifying the competencies may sometimes be cumbersome, the benefits to the organisation could be great. It may be necessary to customise the competency model to suit the unique circumstances of different organisations and industries. Leaders of organisations should realise that the organisation's effectiveness can only be improved by competency models when they are aligned with, and linked to, the organisation's culture, values and expectations (Brits, 2012).

2.2.3. A COMPETENCY-BASED APPROACH

A competency-based approach is an alternate to the job analysis approach. The latter is inductive because it starts with the required job task to arrive at conclusions about the important parts of the job while the former approach is more deductive, starting with the outcomes and then creating tasks (Campion et al., 2011). The use of competencies shifted the focus from the job, its tasks, roles and responsibilities towards the person, their capabilities and behaviours (Croucamp, 2013; Soderquist et al., 2010). According to Soderquist et al. (2010), the main focus currently is on the competencies that are possessed by superior performers who execute a range of activities successfully.

Sanchez and Levine (2009) reckon the main compelling reason for the adoption of a competency-based approach would be to assist the organisation to create a competitive advantage through increased performance. A competency-based approach can be an alternative or addition to the well-established and well-known models for job analysis and personnel specifications required for understanding an organisation's human capital requirements (Bell et al., 2006).

2.2.4. PROCESS FOR THE DEVELOPMENT OF A COMPETENCY MODEL

The development of a competency model is not a new process and various models can be found in the literature (Akkermans et al., 2012; Bartram, 2012; Brits, 2012; Burnett, 2011; Campion et al., 2011; Croucamp, 2013; Edgar & Lockwood, 2011; Lubbe, 2010; Sanchez & Levine, 2009; Soderquist et al., 2010). For the purpose of the study, the competency model development process described, implemented and validated by Brits (2012) will be discussed, adapted and applied.

Brits (2012) proposed an eight-stage process for developing a competency model with detailed steps for completing each stage. Although all the stages will be discussed, only some of the stages will be adopted in the research methodology, according to their relevance to the research questions and objectives of this study.

2.2.4.1. STAGES OF THE DEVELOPMENT OF A COMPETENCY MODEL

A competency model can be developed by following a logical step-by-step framework that is easily communicated to the stakeholders and assists the developer to track the process. The eight stages (Brits, 2012) can be summarised as shown in Figure 2.3 below. Each stage is discussed in this section.

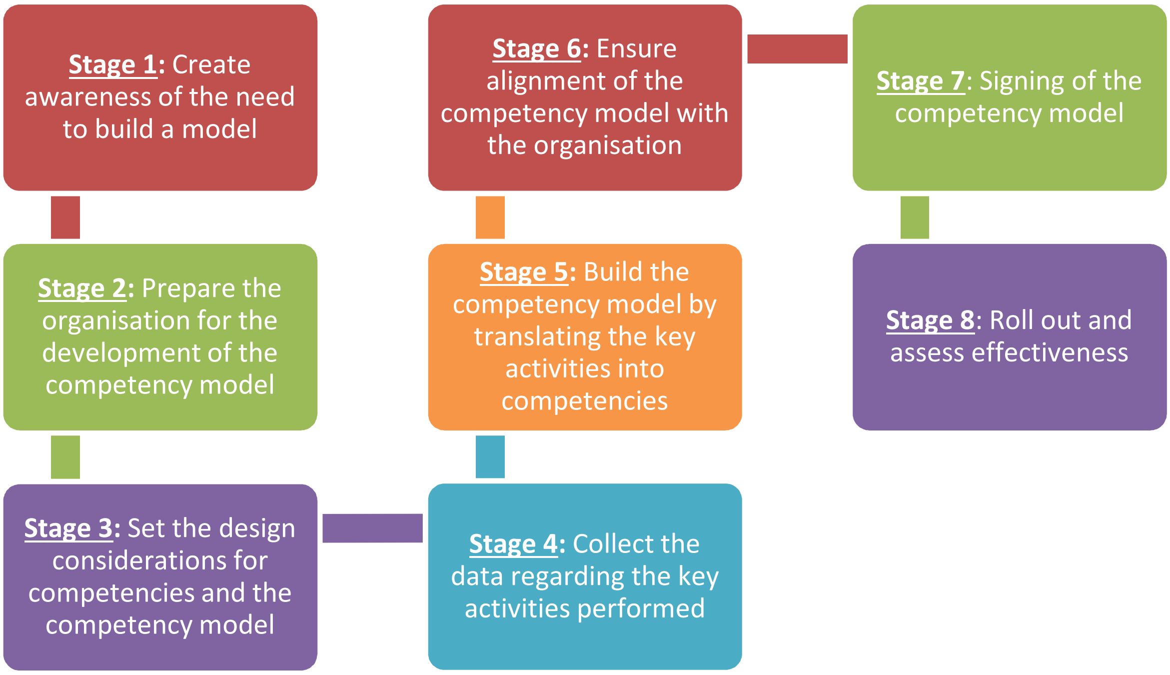

Figure 2.3: Proposed process for developing a competency model with specific stages (Brits, 2012)

Stage 1: Create awareness of the need to build a competency model

- Define and communicate the purpose and objective of the competency model exercise throughout the organisation in order to set the scene for the process that will follow (Brits, 2012).
- Determine the organisational strategy and objectives to inform the competency modelling exercise. Successful competency models identify competencies that are aligned to the organisational strategy and foster competitive advantage (Campion et al., 2011). A competency model should also show all the stakeholders how the model is aligned with the organisation's lifecycle and business strategy (Brits, 2012).
- Campion et al. (2011, p. 231) reason that the "business objective linkage of competency models is critical to the interest and commitment of senior management" and they suggest that the development of the model should start with a definition of the organisation's objectives and goals. This brings into perspective the reason why this is one of the first steps in competency modelling.

Stage 2: Prepare the organisation for the development of the competency model

- Communication with all stakeholders in the organisation is vital to the successful development and implementation of a competency framework. The benefits of the model to the personnel management processes, as well as the productivity and competitiveness of the organisation, should be communicated clearly to management. The employees should be prepared for the process that will follow (Doerflein, 2007; Martone, 2003) so that they fully understand what is expected of them.
- Formulate a change navigation plan. The change navigation plan will ensure that all employees understand the reasoning behind the development of the model (Lievens, Sanchez & De Corte, 2004). Organise an information session with senior management to co-plan the process to be followed and schedule regular review and feedback sessions to inform senior management of the progress made (Brits, 2012).

Stage 3: Set the design considerations for competencies and the competency model

- The design considerations should be documented and distributed to the stakeholders to show how the process is aligned with the personnel management process as well as the organisation's strategy (Brits, 2012). This ensures that a systematic process is followed and includes considerations such as the relevancy, level of analysis, similarity of competencies and the possible need to cluster competencies together (Brits, 2012).
- The design considerations include the establishment of the relevancy of competencies to be included and the required level of analysis and similarity among the competencies (Parry, 1996).
- Once the design considerations are set, prioritise the competencies. Brits (2012) and Mirabile (1997) suggest making use of the 80:20 principle when prioritising competencies. The focus should be on analysing 20% of the

competencies that would have an influence on 80% of the outcomes achieved by the occupational group.

- Following identification and prioritising, competencies should be validated for relevancy and accuracy. Brits (2012) suggests this should be done by subject-matter experts who are familiar with the required activities and its necessary competencies.

Stage 4: Collect the data regarding the key activities performed

- After the formulation of the design considerations, data collection regarding the key performance activities can begin. It is useful to follow a step-by-step process to ensure collection of the correct data (Brits, 2012) and care should be taken to select the correct process.
- The first step in data collection would be to answer the question of why the function exists and what the scope of the function is (Brits, 2012). The facilitator should understand the role of the function in order to ask the right questions and steer participants in the right direction. All the activities associated with the function should be listed, participants asked about the responsibilities of those doing the job and how they interact with other occupational groups across the organisation (Brits, 2012). The collection of data forms a considerable part of the development process and has a major influence on the outcome of the model. It is therefore important that the relevant people participate in the interviews, focus groups or complete surveys to add to the legitimacy of the obtained data (Brits, 2012).
- Verify the frequency of key activities performed and ensure their relevancy. The relevancy of activities can be determined by ranking the activities associated with the occupational group competencies from the most to the least important or relevant using the frequency as a yardstick. Only competencies that correlate with the job activities of the occupational group should be included (Brits, 2012). Kennett (2008) agrees that if competencies which are not relevant to the function are included it will encourage or reward behaviour that does not increase output and can contribute to confusion and a lack of role clarity.

- Measure the difficulty level of key activities and determine the standards required to perform key activities. Determining the level of difficulty of key activities could help establish the extent to which certain tasks need to be analysed (Aziz, 2005).
- Describe the level of commitment required to perform the key activities and consider stability versus transience of the key activities performed (Brits, 2012).

Stage 5: Build the competency model by translating the key activities into competencies

- In order to build a competency model, it is not necessary to develop it from scratch if there are models readily available that can be adjusted to meet the needs of the organisation. If no models are appropriate, the following steps should be completed to translate key activities into competencies (Brits, 2012).
- Cluster key activities into knowledge, skills and personal attributes. Prioritise the knowledge, skills and personal attributes based on key activities. It is the role of the facilitator to prioritise the knowledge, skills and personal attributes to ensure only the most important are included in the model (Brits, 2012).
- Each occupational group within an organisation will have different knowledge, skills and personal attribute requirements and should be clustered accordingly. There can be overlapping competencies, although the facilitator should guide participants to cluster the competencies that are applicable to, and makes sense in, their specific jobs (Dreachslin, 1999, in Brits, 2012). Truesdell (2001, p. 50) supports this view and states that "a bundle or group of skills, associated with a particular occupational group, is necessary to make up a core competency rather than a single skill."
- Develop a taxonomy for knowledge acquisition and application at the different levels of complexity. It should be noted that the knowledge acquisition and application tasks at different levels of complexity can be divided in two parts as explained by Bloom, Englehart, Hill and Krathwohl (1956, cited in Brits, 2012). Knowledge (part 1) is the behaviours and situations that require

remembering of ideas, material or phenomena. Intellectual abilities and skills (part 2) include comprehension, analysis, evaluation, synthesis and application and points to complex cognitions like critical and reflective thinking and problem solving (Brits, 2012).

- List the competency domains and group the knowledge, skills and personal attributes under each domain. A competency domain is a collective name for a group of similar competencies (Brits, 2012).
- Lastly, subject-matter experts need to validate the competency domains and associated competencies (Brits, 2012).

Stage 6: Ensure alignment of the competency model with the organisation

- Once the model has been built, it should be assessed for alignment with the organisation. As part of stage 3 – setting the design considerations – it was mentioned that the model should be aligned with the organisation's business strategy. Now the model should be reviewed to verify that it is aligned with the organisation's strategy (Brits, 2012). The process forms part of vertical alignment.
- Following the vertical alignment, the competency model should be aligned with the personnel management framework and processes. This is horizontal alignment. Horizontal alignment of the competency model with the personnel management process is necessary in order for it to serve as an integrative framework for personnel management in the organisation (Brits, 2012).

Stage 7: Signing of the competency model

- Before implementation of the competency model, it must be signed off by the senior management of the organisation, preferably the head of the organisation (Brown, 2006) to demonstrate that top management is committed to the modelling process.

Stage 8: Roll out and assess effectiveness

- The final stage includes preparing the organisation for implementation of the competency model, implementation, monitoring and assessing the effectiveness of the competency model, taking corrective action and evaluating the success of the corrective action. The true value of the competency model will only be known after it has been successfully implemented and the added value is measured. Continuous assessment will keep the model up to date and maintain buy-in from stakeholders (Brits, 2012).

Brits (2012) designed the eight-stage process to develop and validate a competency model for various occupational groups within the organisation. For the purpose of this study, the applicable steps from stages 3, 4 and 5 will be applied. Table 2.1 below outlines the step-by-step process.

The next section is an introduction to data science in the agricultural sector. The purpose of the study is to conceptualise the competencies of a data scientist in Grain SA for the development of a competency model. In order to understand the specific industry, an overview of the grain industry will be discussed in the second part of this chapter. The role of data science in agricultural will be explored and the role and competencies of data scientists underlined.

Table 2.1: Specific competency modelling stages and steps (Brits, 2012)

Stage 1: Create awareness of the need to build a competency model

- Step 1.1: Define the purpose and objectives of the competency model excercise and communicate this within the organisation
- Step 1.2: Determine the organisational strategy and objectives to inform the competency modelling excercise

Stage 2: Prepare the organisation for the development of the competency model

- Step 2.1: Obtain buy-in from key interest groups and change agents to act as sponsors of the change process for building and implementing the competency model
- Step 2.2: Formulate the change navigation plan
- Step 2.3: Organise information session with senior management to co-plan the process to be followed
- Step 2.4: Schedule regular review and feedback sessions to inform senior management of progress made

Stage 3: Set the design considerations for competencies and the competency model

- Step 3.1: Establish the relevancy of competencies to be included in the competency model
- Step 3.2: Decide on the level of analyis required
- Step 3.3: Verify the similarity of the competencies
- Step 3.4: Prioritise the competencies
- Step 3.5: Validate the competencies
- Step 3.6: Define the levels of work applicable to the organisation

Stage 4: Collect the data related to the key activities performed

- Step 4.1: Ascertain the purpose of the function
- Step 4.2: Verify the frequency of key activities performed
- Step 4.3: Ensure the relevancy of activities
- Step 4.4: Measure the difficulty level of key activities
- Step 4.5: Determine the standards required to perform key activities
- Step 4.6: Describe the level of commitment required to perform key activities
- Step 4.7: Consider stability versus transcience of key activities performed

Stage 5: Build the competency model by translating the key activities into competencies

- Step 5.1: Cluster key activities into knowledge, skills and personal attributes
- Step 5.2: Prioritise the knowledge, skills and personal attributes based on key activities
- Step 5.3: Cluster knowledge, skills and personal attributes as these apply to the various occupational groups
- Step 5.4: Develop a taxonomy of knowledge acquisition and application at the different levels of complexity
- Step 5.5: List the competency domains and group the knowledge, skills and personal attributes under each domain
- Step 5.6: Validate the competency domains and associated competencies

Stage 6: Ensure alignment of the competency model with the organisation

- Step 6.1: Aligning the competency model with the business strategy, focus areas and philosophy – vertical alignment
- Step 6.2: Aligning the competency model with the personnel management framework and processes – horizontal alignment

Stage 7: Signing off on the competency model

- Step 7.1: Signing off on the competency model

Stage 8: Roll out and assess effectiveness

- Step 8.1: Preparing the organisation for implementation of the competency model
- Step 8.2: Implementation
- Step 8.3: Monitoring and assessing the effectiveness of the competency model by measuring the value added by the model against key indicators
- Step 8.4: Taking corrective action to improve the effectiveness of the competency model
- Step 8.5: Evaluating the success of the corrective action

2.3. DATA SCIENCE IN THE AGRICULTURAL SECTOR

2.3.1. INTRODUCTION

Agricultural activities occupy roughly 38% of the earth's terrestrial surface (Foley et al., 2011) and supplies food for direct human consumption as well as producing feed (for livestock), fuel (for transportation and energy, including household kitchen fires), fibre (for clothing) and, increasingly, agricultural biomass used to produce a host of industrial chemical and material products (Alston & Pardey, 2014). Globally, 62% of crop production is allocated to food, 35% to animal feed and 3% to bioenergy (Foley et al., 2011).

South African farming activities include crop production, cattle ranching and sheep farming as well as various other industries (Goldblatt, 2012). About 12% of the country's soil is suitable for rain-fed crops, of which only 3% is considered truly fertile. The greatest part of South Africa's surface (69%) is suitable for grazing, resulting in livestock farming being the largest sector (Goldblatt, 2012). Agricultural activity is the main food provider and "principle source of income and employment in rural areas" (FAO, International Fund for Agricultural Development (IFAD) & World Food Programme (WFP), 2014, p. 13).

Worldwide, agriculture receives greater attention as authorities recognise that it holds the key to food security and increase in food production in order to feed a growing population. To achieve this goal, farmers need to produce around 1.4% more grain annually (Syngenta, 2014). In order to meet these challenges, the agricultural sector would need to improve productivity, ensure economies of scale and proper investment in agriculture and focus on the links between land, people and technology (Biodiversity et al., 2012; Syngenta, 2014).

During the past 50 years, the expansion of agricultural production was mainly through the increase of output per land unit against a slow growing land base or, in other words, increased productivity (Alston & Pardey, 2014). Foley et al. (2011, p. 339) discovered that there is a large variation of crop yields across the globe, even in places with similar growth conditions, which they call 'yield gaps'. They define yield gaps as "the difference between crop yields observed at any given location and the crop's potential yield at the same location given current agricultural practices and technologies" (Foley et al., 2011). These yield gaps are often caused by poor management and limit productivity. Many of the yield gaps can be closed effectively with better deployment of existing crop varieties and improved management (Foley et al., 2011).

2.3.2. OVERVIEW OF THE GRAIN INDUSTRY

The grain industry consists of eight major crop groupings grown globally. These crop groupings represent the majority of food production and include corn, cereals, soybean, vegetables, rice, diverse field crops, sugar cane and a number of smaller more diverse crops referred to as specialty crops (Syngenta, 2014). Due to the fact that the grain industry contributes to both biofuel production, as well as food production directly and indirectly (as a supplier of feed to livestock and poultry) (Biodiversity et al., 2012; Syngenta, 2014), it plays a central role in the agricultural sector.

Syngenta (2014) reported that the demand for grain has increased almost 90% since 1980 and that each year 2.4 billion tons of grain is consumed annually through food, fuel and feed. The four main contributing crops include soybean and maize (feed), and rice and wheat (food). In South Africa, half of the maize produced is used for animal feed, of which 70% is for poultry (Goldblatt, 2012). Furthermore, any significant rise in the demand for meat results in a similar rise in the demand for grain because one kilogram beef requires seven kilograms grain to produce, one kilogram pork requires four kilograms of grain and a kilogram of poultry requires two kilograms of grain (Syngenta, 2014). It is thus evident that agriculture is mainly demand driven and the grain industry specifically will continue to play a vital role in the global economy. The challenge will be to meet the growing demand by means of increased production.

In South Africa, depending on the region, production can be divided into summer and winter rainfall crops. Summer crops, as captured by the Crop Estimates Committee (CEC), include white and yellow maize, sunflower, soybeans, groundnuts, sorghum and dry beans, while winter crops include wheat, malting barley and canola (CEC, 2015). An overview of the production of the different crops in South Africa is given in Table 2.2. It displays the hectares planted and crop estimates for all the major crops in South Africa for the 2013/14 and 2014/15 production seasons. From this table it is evident that maize is by far the most produced crop followed by wheat, soybeans and sunflower seed.

Table 2.2: Production of different crops in South Africa (CEC, 2015)

CROP	2014 AREA PLANTED (HA)	2014 CROP ESTIMATE (TONS)	2015 AREA PLANTED (HA)	2015 5TH CROP ESTIMATE (TONS)
White maize	1 551 200	7 710 000	1 448 050	4 649 800
Yellow maize	1 137 000	6 540 000	1 204 800	5 105 500
Sunflower seed	598 950	832 000	576 000	612 400
Soybeans	502 900	948 000	687 300	1 008 100
Groundnuts	52 125	74 500	58 000	62 855
Sorghum	78 850	265 000	70 500	114 700
Dry beans	55 820	82 130	64 000	73 390
Wheat	476 570	1 775 534	Still to be released	
Malting barley	85 125	310 360	Still to be released	
Canola	95 000	123 500	Still to be released	

Note: Estimates are for the calendar year, e.g. production season 2013/14 is 2014 and production season 2014/15 is 2015.

South African producers experienced a very good 2013/14 growing season with favourable weather resulting in the production of a bumper crop and high average yields (Chaura et al., 2015). During the 2014/15 growing season, a severe drought hit certain areas and this led to the declining crop estimates (Chaura et al., 2015). This illustrates the effect weather conditions can have on the outcomes of produced crops.

The following section focuses on the definition of data science as well as the role data science plays in the agricultural sector.

2.3.3. DEFINITION OF DATA SCIENCE

Data science is a relatively new concept and is used across different industries, agriculture being one of them. According to Loukides (2012), data science involves more than merely finding data but also finding out what to do with the data once acquired. Data application acquires its value from the data itself and creates more data as a result. Therefore, it involves more than just the application of data; it enables the creation of data products.

Sanders (2013) states that data science revolves around asking the right questions. People engage in data science whenever they interactively and iteratively search for deep, hidden patterns. It is inherently a collaborative and creative field, where the successful professional can work with database administrators, businesspeople and others with overlapping skill sets to complete data projects in innovative ways (Harris et al., 2013). Stanton et al. (2012, p. 98) summarise it as follows:

> Think of a hundred project folders full of paper forms, photographs, sketches, formulas, and handwritten notes or a hundred thousand PDF files containing reports with tables, graphics, and narratives: lots of data but little for a statistician to work with in these scenarios. In addition, data science covers the entire information lifecycle and requires a combination of technical and interpersonal skills necessary to understand existing information behaviours that surround data generation, access and reuse.

Umachandran (2013) describes data science as a technology that manages large volumes of data from disparate sources and constructs a system with built-in intelligence. This system can think, find and correlate various pieces of information before generating a result. For the purpose of this study, Umachandran's (2013) definition will be used and built upon. In the next section, the role of data science will be explored with specific reference to agriculture.

2.3.4. THE ROLE OF DATA SCIENCE IN AGRICULTURE

Spielman and Birner (2008) argue that technological change is not only an improvement, but essential to reduce poverty, foster development and stimulate economic growth. Farmers who want to be successful are advised to react quickly to changes in markets, learn about new technology and start using data to improve their efficiency and decision making (Tom, 2014). The role of data science in agricultural decision making is becoming more evident and the information that farmers require for decision-making purposes can be obtained from external and internal sources. External sources refer mainly to other farmers, agricultural journals and state and private institutions, while the internal sources refer to the farmer's own record system or farm management information system (Shiya, 2011). Data science could be applied to improve both internal and external decision making. The application of it in modern production agriculture can improve crop yields and natural resources can be utilised more efficiently by enabling farmers to evaluate all farm management decisions (past, current and future) through agronomic data analysis from data generated in the field (Monsanto, 2014b).

The Climate Corporation is a team of software engineers and data scientists who make use of three technology competency areas, namely, hyper local weather monitoring, simulation of seasonal weather and agronomic modelling on the field level. These competencies form the basis for crop insurance, weather insurance and software tools which assist farmers in protecting and improving their operations (Crosbie & Friedberg, 2013). The software tools provide farmers with continuous current data (up-to-the-minute) for field monitoring, yield forecasting, crop insights and what they call 'decision support' recommendations (Crosbie & Friedberg, 2013). Within the next five to ten years there will be a significant shift in the flow of information and manner in which it is used with the predominance of data services being a leading indicator (Batchelor, Scott, Manfre, Lopez & Edwards, 2014). "The future of agriculture is Big Data" (Vogt, 2013, p. 2) and continuous decision making will play an important role in the modern farming framework (Bureau for Food and Agricultural Policy (BFAP), 2014).

Monsanto is a leading company in the agricultural sector that supplies agricultural inputs. Flowing from their primary interest in agriculture, they are one of the first organisations in the agricultural sector to identify and act on the growing trend of data science use in agriculture and have positioned themselves to act innovatively and quickly based on significant trends. Figure 2.4 below portrays the noteworthy advances in crop production since the early 1900s when farmers ploughed their lands using animal traction (Monsanto, 2014b). The invention of tractors and machinery replaced the use of animals in land preparation which increased productivity immensely. This was followed by hybrid advancements, crop protection products, conservation tillage, plant biotechnology, GPS advancements, molecular breeding and agricultural biologicals (Monsanto, 2014b). All of these advancements have changed crop production radically and have also guided the industry towards increased productivity.

Moreover, Monsanto and the Climate Corporation believe that the use of data science is the advancement next in line to change agriculture as it is currently known (Crosbie & Friedberg, 2013). Friedberg iterates on the importance of data science in agriculture as it helps farmers unlock yield potential as well as managing their risks (Crosbie & Friedberg, 2013). The application of data science in agriculture will enable farmers to analyse weather challenges, soil health, weed management, insect and disease management simultaneously and enable them to make quick, informed farming decisions for better outcomes and improved productivity (Monsanto, 2014a).

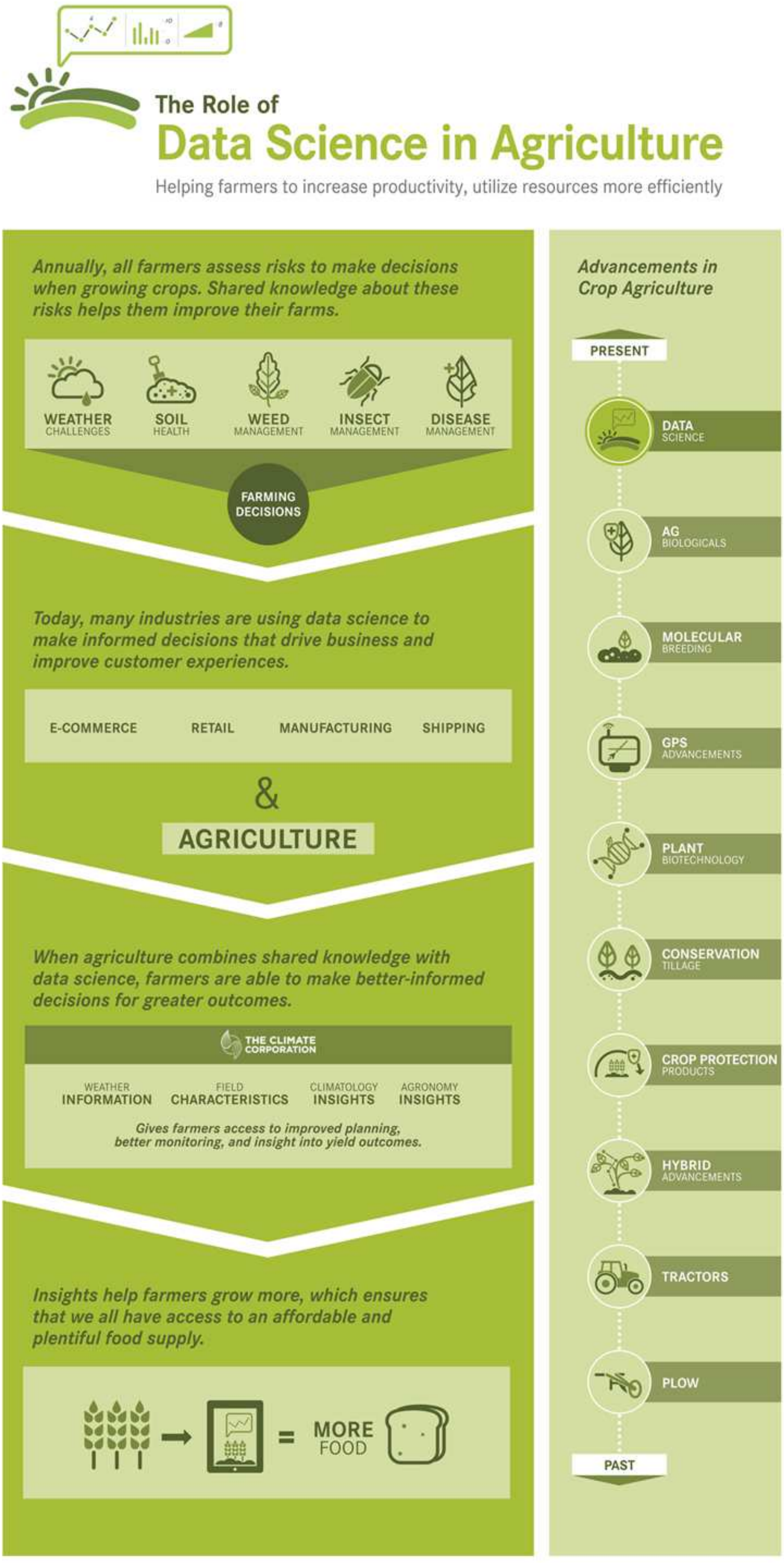

Figure 2.4: Advances in crop production (Monsanto, 2014b)

Figure 2.5 below (Tom, 2014) is a graphical representation of the sources of data and application of data science in a farming operation. Data about fertilizer application, tillage, planting, spraying grain plants, irrigation, fertilizer trucks, chemical trucks, grain trucks, fertilizer facilities, warehouses (seed and chemical) and commodity markets is collected through soil and plant testing, field equipment, satellites, planes or drones, personnel, soil sensors, weather stations, commodity accounts, bank accounts and vendor accounts, field scouts, utility companies and radio-frequency identification (RFID) tags and bar codes fitted to resources.

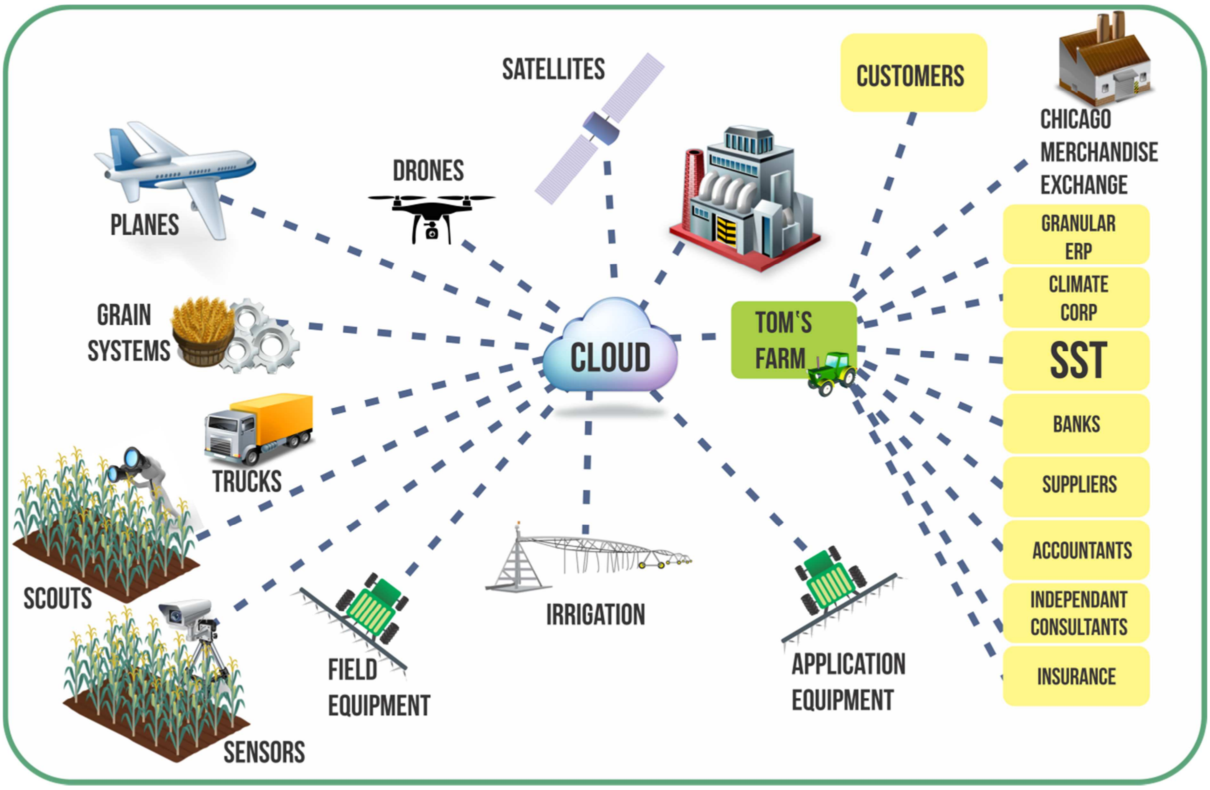

Figure 2.5: Sources and application of data – an example from Tom Farms (Tom, 2014)

Now that the role of the discipline (data science) is clarified, the role and competencies of the person who fulfils the role of a data scientist will be investigated.

2.3.5. THE ROLE AND COMPETENCIES OF A DATA SCIENTIST

Currently, there is little consensus about the role of a data scientist in an organisation. However, an organisation that intends to capitalise on big data will need to hire data scientists and the true challenge will be to identify individuals with the right talent, attract them to the specific enterprise and ensure they are productive (Davenport & Patil, 2012). Stanton et al. (2012, p. 108) argue that the role of the data scientist is to “emphasize the value for information users and decision makers that can come about through application and innovative use of existing technology to organize, analyse, and curate data.” Data scientists make new discoveries as they explore data, (Davenport & Patil, 2012) they communicate these discoveries and make recommendations and suggestions for new directions in business. They are said to combine entrepreneurship and patience (Loukides, 2012).

A data scientist is an information professional with the knowledge and skills to conduct sophisticated, systematic data analysis (Hosack, Power & Sagers, 2014; Power, 2015; Stanton et al., 2012). They contribute to the “collection, cleaning transformation, analysis, visualization, and curation of large, heterogeneous data sets” (Stanton et al., 2012, p. 97). According to Steven Hillion (2011, p. 1), data scientists are “analytically minded, statistically and mathematically sophisticated data engineers who can infer insights into business and other complex systems out of large quantities of data.” Companies acknowledge the need for staff with advanced analytical and problem-solving skills (Hosack et al., 2014). These staff members need to be analytical thinkers who can manage an abundance of data on a daily basis and provide solutions to inefficiencies within the organisation.

Rauser (2011) calls the German astronomer Tobias Mayer the first data scientist. Mayer had pure mathematical and applied engineering skills that he used very effectively and Rauser (2011) believes these skills can help organisations face the “big data” challenges of the 21st century.

Harris et al. (2013) identify four types of data scientists in their attempt to analyse and describe data science in depth. They conclude that the differentiation is not based on data scientists’ breadth of knowledge but rather their depth of knowledge in a specific area in relation to others and their preferred methods of addressing data science problems (Harris et al., 2013). The four types of data scientists include data businesspeople, creatives, developers and researchers. A short description of each is provided in the sections to follow.

2.3.5.1. DATA BUSINESS PEOPLE

Data businesspeople are firstly concerned with their organisation and how data projects can be turned into profit (Harris et al., 2013). They see themselves as entrepreneurs and leaders with technical data science skills (Harris et al., 2013).

2.3.5.2. DATA CREATIVES

Data creatives extract and integrate data, perform advanced analyses, create visualisations, conduct interpretations and build tools to make the analysis scalable and applicable to other users (Harris et al., 2013). Data creatives usually have academic experience with undergraduate degrees in the fields of economics or statistics. Harris et al. (2013) perceive this group as representative of the broadest view of data scientists.

2.3.5.3. DATA DEVELOPERS

Data developers usually have a computer science or computer engineering degree and are concerned with the technicality of data, that is, "how to get it, store it, and learn from it" (Harris et al., 2013, p. 15). Their daily activities can include coding and machine learning.

2.3.5.4. DATA RESEARCHERS

Data researchers have a sound academic research background in statistics, physical or social science (Harris et al., 2013). Organisations value academic training to aid in understanding the complexity of processes through the use of data analyses and have started using researchers to address these data problems (Harris et al., 2013).

The skills for analysing data have a high correlation with the expected expertise in many scientific disciplines. The complexity of large data from non-traditional sources also requires expertise in data retrieval and analysis, statistical analysis, hypothesis testing, interpretation and presentation of data, report writing and storytelling (Hosack et al., 2014).

Loukides (2012) goes further when he underlines the traits of a data scientist. They have a willingness to build data products incrementally, iterate over a solution and the ability to explore (Loukides, 2012). Data scientists attend to all aspects of a problem from data collection to conditioning, to the drawing of conclusions and thinking “outside the box”. They come up with new perspectives of a problem and are able to work with broadly defined problems (Loukides, 2012). Data scientists know they face technical limitations but do not allow it to deter them from finding solutions. They are comfortable in the digital realm and have the ability to structure large quantities of formless data (Davenport & Patil, 2012).

Given the nascent nature of data science, data scientists often have to fashion their own tools or conduct research and thus need to be more than skilled technicians (Hosack et al., 2014). They need interpersonal and communication skills, they must understand how to present data so that other people can reuse it and learn about new technologies as they are developed (Stanton et al., 2012).

According to Rauser (2011), curiosity is the final plank in the framework for a data scientist because, when a new dimension surfaces, the inquisitive person is excited to become an expert on that topic and only an intensely curious person will enjoy acquiring new domains of knowledge and mastery of new areas of expertise continuously.

Data scientists are also in great demand. The US alone has a shortage of 140 000 to 190 000 people equipped with deep analytical skills and 1.5 million analysts and managers to analyse large amounts of data and make decisions based on the findings (Manyika et al., 2011).

To summarise, data science is a new movement in science, an epic wave that is gathering and starting to crest. To catch it, you need people who can “surf” this wave (Davenport & Patil, 2012). However, after an extensive search, it was concluded that no literature or empirical research could be found on the development of a competency model for data scientists in the agricultural or grain industry. This gap in current knowledge thus needs to be filled.

2.4. SUMMARY

This chapter provided a background to all the elements included in the research topic to allow the reader a clear understanding and perspective of the study. The first part of the chapter dealt with the development of a competency model and definitions of the key terms followed by a discussion on why a competency-based approach is preferred. The process for the development of a competency model was explained.

The second part of the chapter provided an introduction to data science in the agricultural sector, followed by a brief overview of the grain industry. Data science was defined and discussed in terms of its role in agriculture. The contributions of various researchers on the use of data science and the suggested role and competencies of data scientists have been outlined and compared. The chapter concludes with a brief summary of the types of data scientist.

Chapter 3 will take the reader through the research design and methodology applied during the study by explaining the logic behind the applied methods.

3. CHAPTER 3 – RESEARCH DESIGN AND METHODOLOGY

3.1. INTRODUCTION

The goal of this chapter is to provide the reader with a map of the researcher's journey in planning and conducting the study while gaining insights along the way. It focusses on the goal of the study, the conceptual framework that was applied, how the study was tested for reliability and validity, the ethics considered and the demarcation of the field of study. Some of the initial questions that need to be answered are why the study should be conducted, what the researcher wanted to understand through the study and why the results are relevant. In an attempt to answer the research questions set out in chapter 1 (section 1.2.1), specific research objectives were formulated (section 1.2.2). The research objectives are repeated here below.

1. To explore the current role of data science in agriculture.
2. To describe the core competencies of a data scientist in agriculture.
3. To formulate the role of a data scientist in Grain SA.
4. To conceptualise the competencies of a data scientist in Grain SA for the development of a competency model.

3.2. CONCEPTUAL FRAMEWORK

According to Maxwell (2009, p. 222), a conceptual framework is "the system of concepts, assumptions, expectations, beliefs, and theories that supports and informs your research." It should explain what the researcher thinks is happening and why it is happening. The researcher's thinking forms an integral part of the reasoning behind the study and it is thus important to examine the foundation of the researcher's thinking. This can be done by explaining the research paradigm in which the researcher situated her work. Research paradigms refer to a set of philosophical assumptions about ontology (how we see reality) and epistemology (how we can understand it) (Maxwell, 2009; Schurink & Schurink, 2012). The conceptual framework will be discussed in terms of the research design, research strategy, sampling strategy, data collection and data analysis strategy.

3.2.1. RESEARCH DESIGN

The adopted approach for the research was qualitative. Qualitative research is exploratory, is used when the researcher is not sure which variables are important and should be examined and is useful when a new topic is addressed with a certain group of people (Creswell, 2003). Exploratory research relies heavily on qualitative approaches to data gathering such as interviews, focus groups, case studies, projective methods or informal discussions (Sekaran & Bougie, 2013).

The field of data science is fairly new and there is limited information on the topic. Furthermore, no examples of competency models for data scientists in the agricultural or grain industry are available. Therefore, the use of an exploratory study was most suitable (Sekaran & Bougie, 2013). Some facts are known about the topic but more information is needed on the competencies of a data scientist in order to develop a model for a data scientist according to the needs of Grain SA.

The paradigm in which the researcher situated the study is objectivism which is based on the belief that an "external reality could be studied objectively" (Schurink & Schurink, 2012, p. 20). Bryman (2007) and Schwandt (2007) argue that qualitative researchers who believe the real world should be discovered through a systematic, interactive, methodological approach and that knowledge arises from interpretation and observation could also be regarded as objectivists. A belief of ontology is that the "life-world of subjects could be discovered in an objective manner" (Schurink & Schurink, 2012, p. 14) with an epistemological stance that interpretation arises from the observation of the researcher. Acknowledgement of the researcher's stance in relation to the study is important as it enhances credibility (Patton, 2002). The research was conducted in a natural setting and the researcher was interested in the participants' point of view. Tracy (2010) marks qualitative methodological research of a high quality if it has a worthy topic, rich rigor, sincerity, credibility, resonance, meaningful coherence and makes a significant contribution and is ethical.

3.2.2. RESEARCH STRATEGY

Research strategies provide specific direction for all the procedures in a research design (Creswell, 2003). This is determined through the researcher's choice of approach and methodology (Schurink & Schurink, 2012). The strategy followed during the course of the study is of phenomenological research. The study describes the meaning of experiences of a phenomenon, topic or concept for individuals (Creswell, 2003), in this case, the competencies of a data scientist. The researcher identified the "essence" of the role and competencies of a data scientist, as described by the participants of the study. Schurink and Schurink (2012) suggest that researchers using this strategy should mainly use participant observation and/or interviewing as methods of data collection.

3.2.3. SAMPLING STRATEGY

In South Africa and at the time of the study no data scientists have been employed in agriculture and therefore no South African participants could be included in the study. The USA is one country where various agricultural innovations originate and the same is true for the use of data science in agriculture. The researcher identified this country as a setting where the individuals to be studied would most likely be situated and would be a rich source of information. When exploring how different individuals behave, the unit of analysis is the individual (Lubbe, 2010).

A limited number of people in a small category have the information that is sought for the purposes of this study. Judgement sampling involves the choice of participants who are in the best position to provide the required information (Sekaran & Bougie, 2013). The researcher thus conducted non-probability judgement sampling and interviewed 20 participants from nine different organisations who have a profound knowledge of data science and its role in agriculture. The study employed a small sample in order to gain an in-depth understanding and thereby work towards theory building (Collins, 2010). This approach facilitated open-ended investigations which allowed the researcher to uncover unanticipated findings for further exploration (Leko, 2014).

The researcher's primary intent was to develop themes from the data (Creswell, 2003). The participants include individuals who work directly with data scientists in the same teams or are data scientists themselves. Two of the participants are end users of data science products and five employ data scientists and oversee activities performed by them. Sampling continued until theoretical saturation had been reached. Theoretical saturation is reached when no new information arises in repeated cases (Sekaran & Bougie, 2013). Table 3.1 below lists the participant's number as no names are used, their relationship to data scientists, as well as their location.

Table 3.1: Sample participants

Participants	Relationship to Data Scientists	Location
Participant 1	Works with data scientists in different teams.	St. Louis, Missouri
Participant 2	Provides training on products delivered by data scientists and feedback to data scientists on their products/findings.	St Louis, Missouri
Participant 3	Works with farmers and agronomists and tests products developed by data scientists. Sends feedback to data scientists on their products/findings.	St Louis, Missouri
Participant 4	Works in a sales team and markets products developed by data scientists.	St Louis, Missouri
Participant 5	Works with data scientists and has a special focus on measurements.	St Louis, Missouri
Participant 6	Works with data scientists and has a special focus on agronomy. Sends feedback to data scientists on their products/findings.	St Louis, Missouri
Participant 7	Works with data scientist/s in the same team.	St Louis, Missouri
Participant 8	Works with data scientist/s in the same team.	St Louis, Missouri
Participant 9	Works with data scientist/s in the same team.	St Louis, Missouri
Participant 10	End user of data products for farms.	Yorkville, Illinois
Participant 11	End user of data products for farms.	Champaign, Illinois
Participant 12	Markets products developed by data scientists and has an interest in data science in general.	Tremont, Illinois
Participant 13	Works with data scientist/s in the same teams.	Tremont, Illinois
Participant 14	Employs and oversees all activities done by data scientists.	San Carlos, California
Participant 15	Employs data scientists and works with them in different teams.	St Louis, Missouri

Participant 16	Data scientist.	Chicago, Illinois
Participant 17	Works with data scientist/s in the same team.	Chicago, Illinois
Participant 18	Employs data scientists and works with them in different teams.	Moline, Illinois
Participant 19	Employs data scientists and works with them in different teams.	San Francisco, California
Participant 20	Employs data scientists and works with them in different teams.	San Carlos, California

3.2.4. DATA COLLECTION STRATEGY

Data collection methodology is the theory of “how inquiry should proceed” (Schurink & Schurink, 2012, p. 31). The interview is one of the oldest data collection methods, is the most commonly used and researchers agree that it is a powerful way to understand people (Al-Yateem, 2012). The researcher applied a qualitative research method and made use of semi-structured face-to-face interviews. A list of semi-structured open-ended questions relating to the different objectives was prepared by the researcher (Sekaran & Bougie, 2013). (See the Addendum for the list of questions sent to all participants).

Al-Yateem (2012) suggests that data obtained from interviews be recorded, transcribed and inspected for evident themes. Factors that have an effect on the quality of data include the interviewer, the participants and the format of the questions. The gathered data should be rich, accurate and very close to reflecting the realities of the phenomena and situation to ensure the conclusions are accurate (Al-Yateem, 2012).

Participants were requested to provide information on the role and competencies of data scientists. They were informed of the interviews beforehand by means of a briefing letter (see the Addendum) to ensure they were well prepared and that they understood the goal of the interview. All interviews were recorded with an audio recorder and transcribed afterwards. The participants were asked for their consent before the conversations were recorded.

3.2.5. DATA ANALYSIS STRATEGY

The data analysis strategy used consists of three main subtypes: categorising strategies (thematic analysis and coding), connecting strategies (case studies and narrative analysis) and memos and display (Maxwell, 2009). These strategies will be discussed briefly in this section.

3.2.5.1. CATEGORISING STRATEGIES

Data is rearranged into different categories by means of coding. Coding includes the “fracture” (Strauss, 1987, p. 29) of data – rearranging it into different categories that “facilitate comparison between things in the same category and between categories” (Maxwell, 2009, p. 237). Baptiste (2001) states further that categorising of data includes the tagging of data followed by the grouping of tagged data. Tagging involves the selection of portions of the body of material that satisfies the researcher’s curiosity while supporting the purpose of the study (Baptiste, 2001). After completion of tagging the material, the researcher places data with similar characteristics in the same group or category. Once the data has been categorised, the researcher applies connecting strategies (Baptiste, 2001).

3.2.5.2. CONNECTING STRATEGIES

Connecting strategies involve an attempt to understand the data in context by making use of different methods to identify relationships between the different elements in the text (Maxwell, 2009). Connections can be made when the researcher positions an integrated set of associations and relationships between and among various concepts – relationships that did not exist prior to the research, were unknown, obscure or undocumented (Baptiste, 2001).

3.2.5.3. MEMOS AND DISPLAY

The third concept, memos and display, requires functions that are not directly linked to data analysis but help the researcher facilitate critical thinking about data and its relationships as well as stimulating and capturing the researcher’s ideas about the data (Maxwell, 2009).

The researcher followed the above approach to analyse the data, starting with categorising the data, followed by connecting data and then memo and display strategies. Data display gives direction on how to present the data while data coding assists the researcher in developing ideas while drawing preliminary conclusions simultaneously (Sekaran & Bougie, 2013).

After application of the conceptual framework, the reliability and validity of the study should be tested (Golafshani, 2003). The process and contents of the reliability and validity of the study follow in the next section.

3.3. RELIABILITY AND VALIDITY OF THE STUDY

Reliability and validity are treated separately in quantitative studies which is not the case in qualitative studies (Golafshani, 2003). In qualitative studies, the reliability and validity are expressed in terms of credibility, transferability, dependability and conformability (Lincoln & Guba, 1985). The reliability and validity of the study will therefore be discussed according to these four requirements.

3.3.1. CREDIBILITY

Credibility refers to the trustworthiness of the research findings and can be achieved through thick description, triangulation and multivocality (Tracy, 2010). Thick description was obtained as the researcher illustrated findings in depth with abundant concrete detail. Enough detail is provided so that readers may come to their own conclusions. Triangulation assumes that if two or more sources converge on the same conclusion, credibility is augmented (Denzin, 1978). The researcher thus made use of a wide range of informants to increase the credibility of the study.

3.3.2. TRANSFERABILITY

Transferability is achieved when readers intuitively transfer the research to their own situation and feel it overlaps with their own behaviour (Tracy, 2010). To promote transferability, the researcher aimed to report on the research findings in such a way that the interpretations allow new understandings and perspectives on the phenomenon, even when applied to other contexts.

3.3.3. DEPENDABILITY

Dependability of the study contributes to reliability (Shenton, 2004). In other words, if the study is repeated in the same context using the same methods and participants, the results would be similar (Shenton, 2004). In order to address the dependability of the study, the processes within the study are reported in detail to allow a future researcher to repeat the work.

3.3.4. CONFORMABILITY

Conformability refers to the quality of the results with regard to the researcher's objectivity (Shenton, 2004). It is concerned with the fact that the results should represent the experiences and ideas of the participants rather than characteristics or preferences of the researcher (Shenton, 2004). The researcher thus attempts to describe the research design in adequate detail. The process followed to display the results is made clear and the researcher indicates where the results were explored, described, formulated or conceptualised.

3.4. ETHICAL CONSIDERATIONS

Ethics in research refers to "a code of conduct of behaviour while conducting research" (Sekaran & Bougie, 2013, p. 208). Proper permission was obtained from Grain SA to conduct the study on their behalf. Participation in the study was voluntarily, conducted in good faith and information was dealt with in a confidential manner. The individuals who participated are treated as autonomous agents (Clinical and Translational Science Institute (CTSI, 2014). The researcher aimed to provide complete and honest feedback to Grain SA to assist them in making decisions based on the new information that was gathered. The researcher made efforts to maximise benefits and minimise possible harms for all parts of the research process (CTSI, 2014). Participants were selected because they can make a valuable contribution and not for the sake of obtaining general information.

3.5. TIME FRAME AND SETTING

The competency model was developed by the researcher for Grain SA's Grain Economy division under the Human Resources training and development umbrella during the course of 2015. Data collection occurred during April over a period of three weeks when the researcher visited the US. The competency model was finalised towards the end of 2015.

3.6. SUMMARY

The discussion on the research design and methodology, as captured in this chapter, aimed to describe the step-by-step process the researcher followed in her attempt to answer the research objectives. The research objectives were reiterated, underlining the importance of aligning the research design with the research objectives. The framework addressed the research design, strategy, sampling strategy, data collection and analysis strategy.

The reliability and validity of the study and ethical considerations were also addressed. The last part of the chapter paid attention to the timeframe and setting of the study and explains when, where and under which department the study was conducted.

The following chapter reveals the results as it obtained through the empirical research. The results are discussed under headings corresponding to the research objectives.

4. CHAPTER 4 – RESULTS AND DISCUSSION

4.1. INTRODUCTION

This chapter gives a detailed description of the results obtained from the interviews with the different participants. The results are discussed according to the different research objectives, starting with the current role of data science in agriculture. Thereafter, the core competencies of a data scientist in agriculture and the role of a data scientist in Grain SA are examined. Finally, the competencies of a data scientist in Grain SA are conceptualised with a view to developing a competency model.

4.2. RESULTS

4.2.1. THE CURRENT ROLE OF DATA SCIENCE IN AGRICULTURE

The role of data science in agriculture has evolved rapidly over recent years but is not properly defined yet. This role was explored through the empirical data collected. Table 4.1 below displays the categories as identified by the researcher from the interviews, a description of each category and the frequency it was mentioned by the interviewees. The identified categories are discussed below the table.

Table 4.1: Identified categories for the current role of data science in agriculture

IDENTIFIED CATEGORY	DESCRIPTION	FREQUENCY MENTIONED
Decision making	• Leverage data to make decisions. • Farmers have more information at their fingertips to make better decisions. • Data science in agriculture is important in driving the broader decisions. • Helps farmers make better decisions most of the time. • Farmers can make decisions on the fly. Quick decisions they have not been able to make before.	12
Data analysis	• Data science helps analyse data.	7
Form insights out of information	• What insights can I create from the data?	5
Problem solving by answering complex questions	• Problem solving before producers know the problem exists.	4
Increased profitability	• Data science can help turn information from the farm operation into profitable insights.	4
Visualisation of data	• How to visualise the data that has been analysed in a way that is easy to read and is understandable.	3
Creating value from data	• Creating something useful with the data. • Farmers "are starting to see value in their data, just like value in their crop" (Participant 1).	3
Data management	• Farmers generate large amounts of information they cannot manage and analyse without a computer system.	2
Cleaning up data	• Data scientists detect and correct or remove corrupt or inaccurate records	2
Data as a commodity	• A commodity that farmers have not been able to leverage in the past.	1

4.2.1.1. DECISION MAKING

Decision making as part of the role of data science was mentioned the most often by participants. Twelve out of the 20 participants indicated that data science drives broader, quicker and more efficient decision making. Farmers are perceived to be the data customer and make use of decision support tools, as developed by data scientists. Participant 10 said data science is “Essentially helping you make decisions for the future year.”

Participant 2 indicated that, for the past 20 years, large amounts of data have been collected by means of superficial analysis. He mentioned that there are companies that provide software programs a farmer or retailer can use for data analysis to provide cursory insights. However, there has been very little deep statistical analysis and almost no use of information with regards to weather and how it impacts decisions. Participant 2 also stated, “If you are around agriculture at all, you understand that weather is probably one of the largest impacts that we have on a crop. The fact that we have never really calculated that into the equation is mind boggling.” For this participant the power of data science lies in using all historical weather information collected over the last 30 to 50 years and making inferences on how that weather impacts crops. Furthermore, the participant feels that data science can be used to predict with relative certainty what is expected to happen throughout the next season.

Participant 2 explained that data science can help producers have more information at their fingertips to make better decisions. Most agricultural producers have 30 to 40 chances to generate a pay check in their entire career according to Participant 2. Data science can be used to ensure a better probability of a good outcome for each one of those chances.

This view is supported by another participant, a producer, who mentioned that if they take large amounts of data, for example yield maps that are taken by GPS, the yield is known for every acre or even down to a tenth of an acre. With the application of data products built by data scientists, the producer will understand their field potential very well. For instance, the weather is predicted and the planted hybrid and amount of nitrogen applied is known. The yield map serves as a verification of what farmers already expected before they started harvesting.

4.2.1.2. DATA ANALYSIS

Data analysis was also mentioned a significant number of times (seven) by the participants. One of the participants (Participant 13) noted that “farmers today have a lot of information about their fields, about their yields, about their machinery but it is very hard for them to analyse that information and be able to decide what decision can they make that will impact profitability for their farm.”

Participants view data analysis as the step following data collection. Large volumes of data are already being collected and data science supports the analysis of the data. Participant 19 explained the steps as follows: “Analyse the data, separate the signal from the noise, so to speak, and get actual recommendations.”

4.2.1.3. INSIGHTS FROM INFORMATION

Five of the participants feel the role of data science involves forming insights from the information. According to one participant, “True data science is statistics [used] to form insights out of the information.” Participant 4 stated that, once the information has been collected, managed and analysed, the insights that are created and what happens with those insights are the concern of data scientists.

4.2.1.4. PROBLEM SOLVING THROUGH ANSWERING COMPLEX QUESTIONS

Participants referred to the role of data science in agriculture as helping solve producers’ problems by answering complex questions. Participant 16 said data science involves solving producer’s problems before they may even know that the problem exists. Data science is thus about “Having models in place or being able to quickly create a model to solve someone’s problems” (Participant 16).

4.2.1.5. INCREASED PROFITABILITY

In applying data science, producers can increase their profitability. Four of the participants mentioned this advantage. According to participant 18, the role of data science in agriculture involves:

> understanding how we take a big pool of data and start to create simple products that improve growers' operations efficiencies. Allow a grower to utilise, leverage and incorporate it in their business processes without a lot of complexity and it is kind of natural in their flow to increase productivity, yield, top line revenue or optimise cost.

4.2.1.6. VISUALISATION OF DATA

Visualising analysed data results in information that is easy to read and understandable. This category was mentioned by three participants in answering the question on the role of data science in agriculture. "Everything is becoming so visual that the people really want interactive visual data" (Participant 1).

4.2.1.7. CREATING VALUE FROM DATA

As part of its role in agriculture, data science must create something useful such as a product. This is the view of three of the participants. Participant 1 said, "They are starting to see value in their data, just like value in their crop."

4.2.1.8. DATA MANAGEMENT

Data management was mentioned by two of the participants. From the interviews it is evident that the need arose to manage large volumes of data collected on farms. Participant 1 explained that "precision agriculture", a concept which became popular in the US in the 1970s and 1980s, did not become what the industry envisaged at the time because the computer systems needed to manage the data in order to answer complex questions did not yet exist.

4.2.1.9. CLEANING DATA

Two of the participants indicated that data cleaning should be regarded as part of the role of data science in agriculture. “Data scientists are trying to make sure the data is as clean as possible and accurate” (Participant 3). Participant 19 indicated that only in the last few years has the industry become truly interested in the role of data scientists in cleaning data and gathering it in a way that allows it to be mined, analysed and for recommendations to be derived from it.

4.2.1.10. DATA AS A COMMODITY

One participant referred to producing data commodities as part of the role of data science in agriculture.

> People are starting to realise that data can be seen as a commodity – one which farmers have not had the privilege or ability to leverage in the past – but now they can with the application of data science. Data science delivers data products supporting decision making on all levels (Participant 1).

4.2.1.11. SUMMARY OF THE ROLE OF DATA SCIENCE IN AGRICULTURE

From the above it is evident that participants regard the role of data science as follows:

- Facilitating effective decision making and problem solving in agriculture by means of computerised data management.
- Cleaning data, data analysis and the visualisation of analysed data in order to derive profitable insights.
- Therefore, data science creates value from data by turning information from the farm operation into decisions that lead to increased profitability.

Now that the current role of data science in agriculture has been explored, the core competencies of a data scientist in the sector can be described.

4.2.2. THE CORE COMPETENCIES OF A DATA SCIENTIST IN AGRICULTURE

In the process of identifying the core competencies of a data scientist in agriculture, the competencies are first identified in terms of knowledge, skills and personal attributes. Later in the chapter the same competencies are classified according to competency domains, clusters and individual competencies.

4.2.2.1. CORE COMPETENCIES IN TERMS OF KNOWLEDGE, SKILLS AND ATTRIBUTES

As part of the literature review, the process for the development of a competency model (as described, implemented and validated by Brits (2012)) was discussed, adapted and applied. The process involves eight stages (Brits, 2012) and the researcher applied certain steps from selected stages that are relevant to this study. Figure 4.1 below summarises the stages described by Brits (2012), some of which will be used to formulate and describe the core competencies of a data scientist in agriculture.

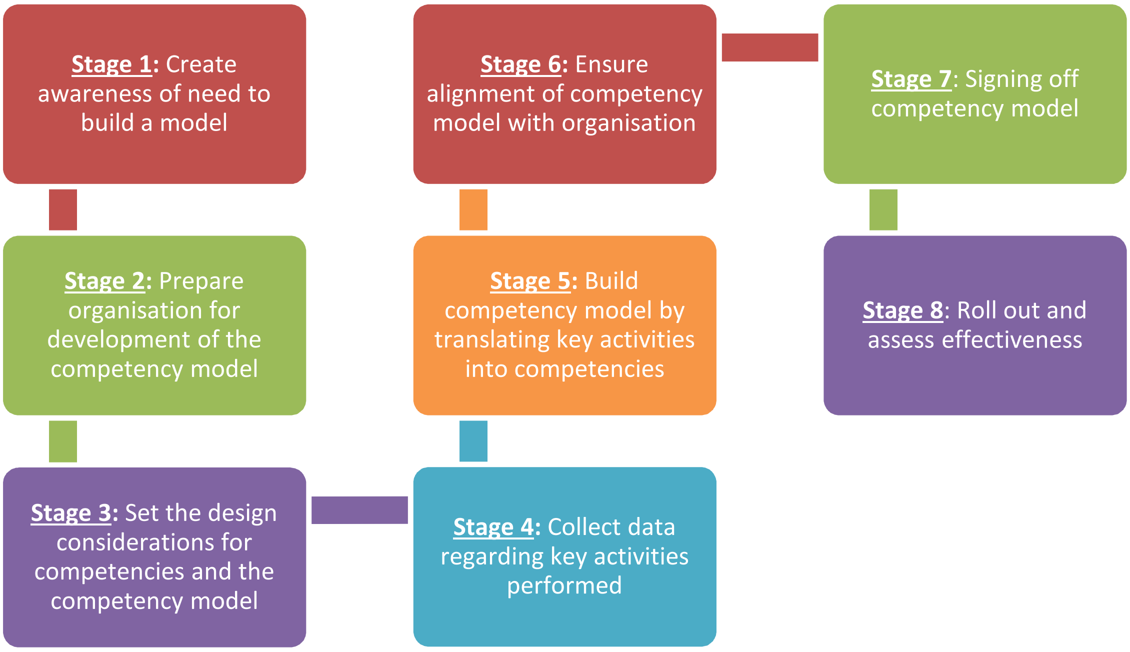

Figure 4.1: A proposed process for developing a competency model with specific stages and steps (Brits, 2012)

In order to determine the core competencies need by an effective data scientist in agriculture, the applicable steps from stages 3, 4 and 5 were applied. The knowledge, skills and personal attributes were prioritised to ensure only the most important competencies are included in the model. Participants were asked what would differentiate an excellent data scientist from a mediocre one. Their feedback was used to identify competencies that would play a role in the efficiency of the data scientist. Furthermore, the competencies that were mentioned by most of the participants guided the prioritising process.

Table 4.2 presents the categorised core competencies of a data scientist in agriculture. The competencies are listed in terms of knowledge, skills and attributes. The table includes the identified category, a description of the category and the frequency it was mentioned by the participants. The list is sorted according to frequency mentioned which underlines the priority assigned to the specific competencies. A discussion of the knowledge, skills and attributes follows after the table.

Table 4.2: Core competencies of a data scientist in agriculture in terms of knowledge, skills and attributes

IDENTIFIED CATEGORY	DESCRIPTION	FREQUENCY MENTIONED
KNOWLEDGE		
Domain knowledge and experience	• Knowledge of a specific domain including farming, plant breeding, agronomy, soil science or some aspect of these. • Real farming experience is also highly valuable.	9
Postgraduate qualification	• Masters or PhD level qualification in physics, computer science, data science, statistics, other related sciences or agronomy.	7
Agronomic science	• Understands interaction between soils and crop physiology, how to apply insights from this and make recommendations.	4
Mathematics	• Knowledge on how to apply a mathematical strategy to solve a practical problem.	4
Data collection	• Knowledge of data collection equipment and format of data collected by the device, particularly how to upload the data from the device into a cloud or other means to access it. • Geo-spatial information system knowledge.	4
Software engineering systems	• Bachelors to PhD in computer science or mathematics. • Expertise in computer languages like R or Python.	2
Statistics	• Sound knowledge of statistics.	2
Data analysis and modelling	• Knowledge on the relationship between data and outcomes. • Knowledge of how to rewrite complex problems in a mathematical model.	2

Field-based research science	• Knowledge of how to conduct the field research component of data science and manage the interface with the user of the platform. • Knowledge of how to conduct field research to measure statistical differences in terms of performance.	1
Equipment expert	• Understands all the tools and measuring devices used to collect accurate data.	1
Machine learning	• Knowledge of machine learning algorithms and how to use each one for different problems.	1
IDENTIFIED CATEGORY	**DESCRIPTION**	**FREQUENCY MENTIONED**
SKILLS		
Communication skills	• Able to effectively transfer thoughts and express ideas verbally in individual or group situations.	12
Data analysis	• Statistical skills. • Able to apply analytics on available data in order to gain insights from it.	7
Software engineering systems	• Computer programming skills. • Data processing skills. • Coding skills including experience in translating models into code. • Data ingestion skills • Able to develop the right interface to allow third group analysis.	6

Achieving results and satisfying customer expectations.	• The ability to take direct action in order to attain or exceed objectives. • Develops alternatives when certain actions have not led to a desired result. • Uses knowledge and expertise on data science and applies it to practical situations. • Actively seeks the best way to achieve goals.	5
Quantitative skills	• Quantitative mind-set, ability to code and use different statistical software. • Skills for manipulating data.	3
Data modelling	• Able to conduct data modelling data and extract value out of it. • Able to link different processes, place correct weighting on each one and connect these with the help of the agronomist so that the output is accurate.	3
Teamwork	• Cooperative and works well in a team to find solutions which generally benefit all involved parties.	2
Visualisation of data	• Able to present data visually in a way that is easy to see, read and understand.	1
Writing skills	• Able to transfer thoughts and ideas onto paper and create reports.	1
Multitasking	• Ability to work on multiple projects at the same time	1
Field-based research skills	• Ability to understand trials and trial results.	1

IDENTIFIED CATEGORY	**DESCRIPTION**	**FREQUENCY MENTIONED**
	ATTRIBUTES	
Creativity	• Ability to come up with original and innovative ideas and solutions and to adopt points of view outside the usual parameters.	11
Attention to detail	• Thoroughness in accomplishing a task with concern for all areas involved, no matter how small. • Monitors and checks work or information and plans and organises time and resources efficiently.	7
Collaboration	• Fosters cooperation and teamwork while participating in a group and working toward solutions which benefit all involved parties.	6
Continuous learning	• Demonstrates eagerness to acquire necessary technical knowledge, skills and judgement to accomplish a task effectively. • Desire to acquire knowledge and skills necessary to perform job more effectively. • Ability to absorb new information readily and put it into practice effectively. • Recognises mistakes and attempts to correct or prevent them.	5
Accuracy	• Consistently delivers work of a high quality which is precise and meets standards, procedures, rules, regulations and expectations. • Recognises and separates quality data from insignificant data.	3

Receptivity	• Open to dialogue and a good listener. • Absorbs and understands important verbal and non-verbal information and asks further questions when necessary. • Ability to understand their audience.	3
Problem solving	• Generates creative approaches to problems and opportunities.	3
Passionate	• Engaging and enthusiastic.	2
Questions traditional methods	• Considers non-traditional farm practices.	1
Vision	• Ability to see beyond daily tasks and explore ideas for the future. Regards the facts from a distance and sees them in a broader context or over a longer term. • Consider the big picture and understand all of the different factors a farmer has to consider in producing a crop and how they are connected. • Connect variables with outcomes.	1
Investigative	• Asks the right questions on technical and functional matters.	1
Tenacious	• Shows persistence in finding a solution despite hurdles as most problems are very complex.	1
Agile and flexible	• Able to use variety of devices from different manufacturers.	1

Knowledge

Domain knowledge is the third-most competency mentioned overall. Participants felt strongly that data scientists need knowledge on farming, plant breeding, agronomy and science or some aspects of these disciplines. This knowledge could also be obtained through personal experience in the domain, for example, if someone grew up on a farm.

Thereafter, the importance of knowledge at a *post-graduate level* in *physics, computer science, actual data science, statistics or other sciences* was mentioned. Knowledge of *agronomy* was then cited including an understanding of the interactions between soil and crop physiology. How to apply insights and make recommendations was also referred to.

Furthermore, knowledge of the *application of mathematical strategies* to problem solving, *data collection* (including from geo-spatial information systems) and knowledge of *software engineering systems* (including computer languages like R or Python and computer science) are valuable. Some participants included knowledge related to *statistics, data analysis and modelling, field research, equipment and tools used to collect data* as well as an understanding of *machine learning algorithms* and how to use each one for different problems.

Skills

Most participants regard *communication skills* as a core competency. Data scientists should be able to communicate clearly and articulately when speaking with an individual or before a group, ensuring that others fully comprehend the intended message. A good candidate is cooperative and works collaboratively in a team.

Secondly, data scientists need to have *data analysis* skills such as the ability to apply analytics on available data in order to gain insights from it, including *statistical skills*. For data scientists to properly analyse data, they need to use *software engineering systems*, ingest the data and develop the right interface. They should have adequate *computer programming* and *data processing skills*.

Other skills that were mentioned include *achieving results and satisfying customer expectations* – the ability to take action in order to attain or exceed objectives, *quantitative skills* related to manipulating data and *data modelling* to extract value.

The skills mentioned by a single participant are *visualisation of data, writing skills, multitasking* and *field-based research* referring to the ability to understand trials and trial results.

Personal Attributes

Creativity seems to be an important attribute and was mentioned by 11 of the participants. It refers to an individual's ability to generate original ideas. *Attention to detail* was mentioned by seven participants and refers to thoroughness in accomplishing a task and concern for all areas involved, no matter how small. Six participants cited an attribute involving someone that fosters cooperation and teamwork while participating in a group and working toward solutions which benefit all involved parties.

Following these top three attributes, someone that demonstrates eagerness to acquire necessary technical knowledge, skills and judgement to accomplish a task effectively was mentioned by five participants. The person has a desire to acquire knowledge and skills necessary to perform a task more effectively, recognises mistakes and attempts to correct or prevent them.

Attributes pointed out by three participants are *accuracy*, which requires the scientist to consistently deliver high quality accurate work that meets expected standards, being a *good listener* in order to understand their audience and *problem solving* or generating creative approaches to address both problems and opportunities.

Two of the participants regard being *passionate* and engaging as a required attribute, while the least frequently mentioned (once each) attributes include *questioning traditional methods,* having *vision, asking relevant questions* on technical and functional matters and being *tenacious, agile and flexible.* In the following section, the core competencies of a data scientist in agriculture are classified in terms of competency domains, clusters and individual competencies.

4.2.2.2. CORE COMPETENCIES IN TERMS OF COMPETENCY DOMAINS, CLUSTERS AND COMPETENCIES

According to the definition by Brits (2012), a competency model consists of a cluster of competency domains and its associated competencies. A competency cluster is a collective name for a set of domains while a domain is a collective name for a group of similar competencies as illustrated in Figure 4.2 below. The figure includes the cluster examples of functional and professional specialisations in economic research and market operations with a few associated competencies.

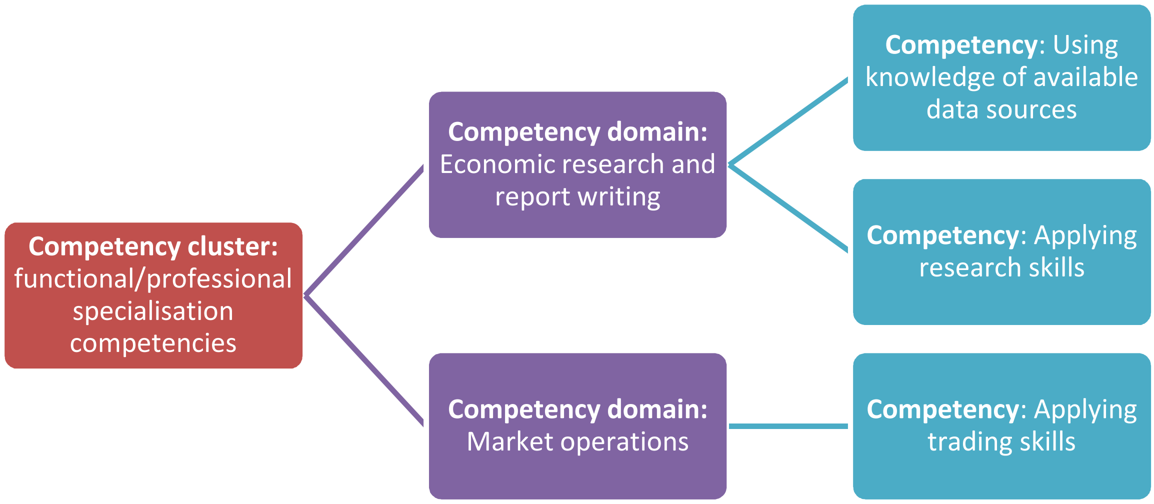

Figure 4.2: An example of competency terminology and connected subcategories (Brits, 2012)

Competencies are usually assigned a descriptive label or title and a definition (Campion et al., 2011). How the competency manifests on the job in behavioural terms is also described (Campion et al., 2011). The researcher made use of a competency library to add structure to the existing competencies as listed by the participants. A competency library refers to “a list of competencies from which to select when developing a competency model” (Campion et al., 2011, p. 245). At the same time, the knowledge, skills and personal attributes were clustered as they apply to the various occupational groups.

The Greater London Authority (2014) identifies clusters that outline the essential behaviour required for effective performance as leadership, delivering results and working well with others and in the organisational context. While creating links between competencies, the researcher discovered that the clusters published by the Greater London Authority (2014) could be used for the purpose of the study. Competencies were clustered for one occupation group, namely, the data scientist.

After listing the competencies, they were grouped into competency domains. A competency domain is a collective name for a group of similar competencies (Brits, 2012). The competency domains used are adapted from various competency libraries (Harvard University, 2015; HRSG, 2015; TMA, 2012; WSU, 2015) and include:

- Business domain expertise and knowledge
- Critical thinking, analysis, problem solving and business skills
- Programming and database skills
- Interpersonal and communication skills
- Analysis and presentation skills
- Data modelling, warehouse and unstructured data skills

After completing all the necessary steps, the core competencies needed by a data scientist in agriculture were identified, prioritised, labelled and organised according to competency domains and clusters. Figure 4.3 displays the core competencies of a data scientist classified in terms of competency clusters, domains and competencies.

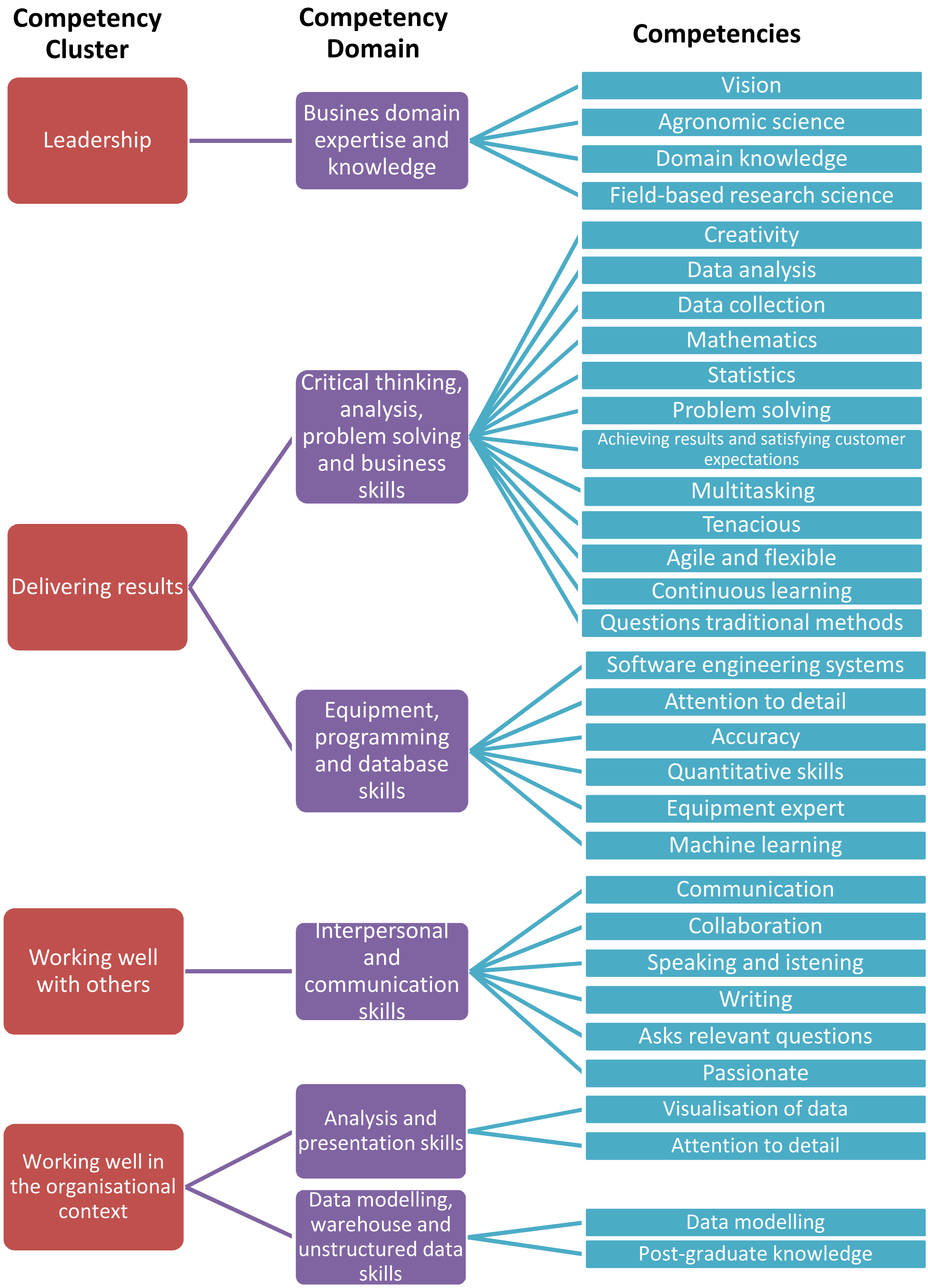

Figure 4.3: A graphical representation of the core competencies for a data scientist in agriculture in terms of competency clusters, domains and individual competencies

Now that the core competencies have been described, the role of a data scientist in Grain SA can be formulated.

4.2.3. THE ROLE OF A DATA SCIENTIST IN GRAIN SA

The current role of data science in agriculture has been explored and the core competencies of an agricultural data scientist identified. In order to conceptualise the competencies of a data scientist in Grain SA for the development of a competency model, the role of a data scientist in Grain SA needs to be formulated first. Nationally, no information is available on the role of a data scientist in agriculture. Therefore, data from the empirical research results obtained from the participants outside South Africa, as well as information from the literature, was used to formulate the role of a data scientist in Grain SA.

4.2.3.1. TYPES OF DATA SCIENTISTS

As previously discussed in section 2.3.5, Harris et al. (2013) divide data scientists into four types, namely, data businesspeople, data creatives, data developers and data researchers. During the interview process, two participants felt strongly that the role of a data scientist should be well defined and understood by the data scientist for them to be successful within the organisation. Both these participants referred to different types of data scientists and their explanations support the typology of Harris et al. (2013), as will be shown in this section.

Participant 5 stated:

> If you go to San Francisco, they will say data scientists will be somebody who takes the data collected and compiled by someone else, put the data into a database and create insights. That’s a data scientist, but that’s only one part of this big chain.

Participant 6 explained:

> I think it’s important to define them because, in the absence of defining the different types of data scientists, I don’t think you

can be successful. It’s going to be very difficult to find one person that has all of the expertise.

Participants 5 and 6 then went on to describe different types of data scientists as they see it. Table 4.3 below summarises the four types of data scientists, as identified by these two participants.

Table 4.3: Different types of data scientists, as described by the participants

Participant 5	Participant 6
Type 1: Data Science Collector	**Type 1: Equipment Expert**
One who collects data with precision and accuracy using the appropriate equipment and uploads various meta-data into the cloud.	An individual who understands all the tools and measuring devices being to collect accurate data.
Type 2: Software Systems Engineer	**Type 2: Field-based Research Scientist**
Someone who receives the meta-data from the cloud and decides how to store it and develop an interface to access the data so that the next user can analyse it.	A scientist who conducts field research and manages the interface with the platform user. Field research includes measuring statistical differences in terms of performance.
Type 3: Data Analyst	**Type 3: Modeller**
Analyses data, conducts data modelling and derives insights from the data.	Analyses data, conducts data modelling and derives insights from the data.
Type 4: Data Science Applier/Agronomist	**Type 4: Agronomic Scientist**
Applies the insights received from the data analyst and converts it into actual recommendations which are conveyed to producers.	Understands the interaction between soil and crop physiology, how to apply insights and make useful recommendations.

From Table 4.3 it can be seen that there is duplication in the different types of data scientists as explained by the participants. The researcher found that these types can be categorised according to the typology of Harris et al. (2013). Table 4.4 below displays the new categories as derived and adapted from the literature and the empirical results of the current study.

Table 4.4: Different types of data scientists – combined categories

Combined Categories	Participant 5	Participant 6
Data Science Coordinator	**Type 1 & 4**	**Type 1 & 4**
The data science coordinator is concerned with their organisation and to what effect data projects can be initiated and managed. They see themselves as entrepreneurs and leaders with technical data science skills as well as a profound knowledge of the domain they work in. They know all the role players – who can contribute, who needs support and who will utilise the data products as well as the relationships between these role players.		
Data Analyst	**Type 3**	**Type 3**
Data analysts extract and integrate data, perform advanced analysis, create visualisations, conduct interpretations and build tools to make the analysis scalable and applicable to other users. Data analysts usually have academic experience with undergraduate degrees in the field of economics or statistics. They can rapidly transform data into something of value. Harris et al. (2013) describe this group as representative of the broadest view of data scientists.		
Data Developer	**Type 1 & 2**	**Type 1**
Data developers usually have a computer science or computer engineering degree and are concerned with the technicality of data, that is, how to obtain, store and learn from it. Their daily activities include coding and machine learning. They should be knowledgeable of the different types of equipment used to collect and measure data with precision and accuracy.		
Data Researcher	**Type 4**	**Type 2**
Data researchers have a sound academic research background in the agricultural sciences or a related domain. They apply insights received from data analysts, converting these into useful recommendations that are relevant to the specific domain.		

4.2.3.2. PROPOSED ROLE FOR A DATA SCIENTIST IN GRAIN SA

Grain SA identified an opportunity to make optimal use of the organisation's available research data by means of the application of data science. In terms of the role that Grain SA plays in the grain industry in South Africa, as well as the context within which the organisation operates, specific aspects need to be taken into consideration when formulating the role of a data scientist in Grain SA. These are discussed individually in this section.

Building relationships

The opportunities that data science brings to agriculture in the future are numerous. Helping grain producers understand the positive role that data science can play and building relationships with individuals and companies in order to work together will be crucial. Grain producers' goal is to take care of the land and earn a living from it. Producers require assistance with the elements of data collection, data analysis, visualisation and interpretation of data because they are time consuming activities and producers do not necessarily have all the expertise required. The data scientist should also know all the role players – who can contribute, who needs support and who will utilise the data products as well as how the role players and outputs work together.

Building trust

Any new discipline or innovation comes with uncertainties. The same is true for the use of data science in agriculture. Grain SA should start building trust between data science and grain producers. Farmers will likely need to feel comfortable sharing more information than they ever have before. Grain SA is in the right position to take up the role of facilitator due to the organisation's objectivity and the existing trust relationship between Grain SA and its members.

Facilitating policy development

A third aspect that should be considered when formulating the role of a data scientist in Grain SA would be focused on data privacy and data protection. Grain SA can be a driver or facilitator in setting appropriate standards or policy for the use of producer data that is beneficial for the whole industry. It will help assure producers that it is safe to share their data and will also support the building of trust between role players.

It became evident that organisations need to clearly define and understand the role the position entails before appointing a data scientist. The specific role of the individual indicates what type of data scientist needs to be sourced. Each type of data science position requires a slightly different skill set. In light of the identified aspects, it is recommended that Grain SA's starting point would be to appoint a data science coordinator. The role of the data science coordinator will include:

- Building relationships with producers as well as other role players in agriculture with regards to data science;
- Knowing all the role players and their interrelationships;
- Building trust among producers so that they feel comfortable sharing their data; and
- Facilitating policy development that is beneficial for the whole industry and attends to issues such as data privacy and protection.

Summary of the role of a data scientist in Grain SA

The role of a data science coordinator in Grain SA is to build relationships with the organisation's producers and other companies entering the data science arena. The individual is concerned with initiating and managing data projects for Grain SA. The data science coordinator should have knowledge of the different role players, build trust among producers and facilitate policy development that is beneficial for the grain industry.

Now that it is clear what role data science coordinators will play in Grain SA, their competencies can be conceptualised for the development of a competency model.

4.2.4. THE COMPETENCIES OF A DATA SCIENCE COORDINATOR IN GRAIN SA

The core competencies that were identified and described for a data scientist in agriculture (see section 4.2.2) are used as a baseline and adapted to fit the role of the data science coordinator in Grain SA. The competencies are first identified in terms of knowledge, skills and attributes. Later on the same competencies are classified according to competency domains, clusters and individual competencies.

4.2.4.1. CORE COMPETENCIES IN TERMS OF KNOWLEDGE, SKILLS AND ATTRIBUTES

Knowledge

The data science coordinator in Grain SA should have a profound knowledge of the domain they work in. This includes knowledge of farming, plant breeding, agronomy and the other sciences or some aspects of it. Valuable knowledge is also gained through personal experience in the domain. Knowledge at a post-graduate level of physics, data science, statistics, science or agronomy is required. Such knowledge may be complimented by an understanding of applied mathematical strategies for problem solving.

The data science coordinator in Grain SA would need to be somebody who has knowledge of the business sector or competitive intelligence, understands all the relevant tools and keeps up to date with new technology as it becomes available. The coordinator would also need to be able to work with the algorithms data science companies supply.

Skills

Communication skills are regarded as a core competency for a data science coordinator in order to build relationships and trust among the different role players. The data science coordinator could outsource certain tasks or projects if they are the only data scientist in the organisation and do not have the in-depth skills or capacity required for a certain task. Programming and data base skills are thus not included in the competency list as they can be outsourced. The remaining skills are relevant for inclusion in the competency model as the data science coordinator in Grain SA must have them.

Personal Attributes

The attributes listed and discussed for a data scientist in agriculture can be applied for a data science coordinator in Grain SA as well. The attributes include, amongst other attributes, creativity, attention to detail and someone that fosters cooperation and teamwork. Furthermore it should be someone that demonstrates an eagerness to acquire the necessary knowledge, skills and judgement to accomplish a task effectively.

In the next section, the same competencies are classified in terms of competency domains, clusters and individual competencies.

4.2.4.2. CORE COMPETENCIES IN TERMS OF COMPETENCY DOMAINS, CLUSTERS AND INDIVIDUAL COMPETENCIES

All the competencies for a data scientist in agriculture (see section 4.2.2) are applied for the data science coordinator in Grain SA except for the programming and data base skills domain and its connected competencies. Figure 4.4 includes the competencies needed for the development of a competency model for a data science coordinator in Grain SA.

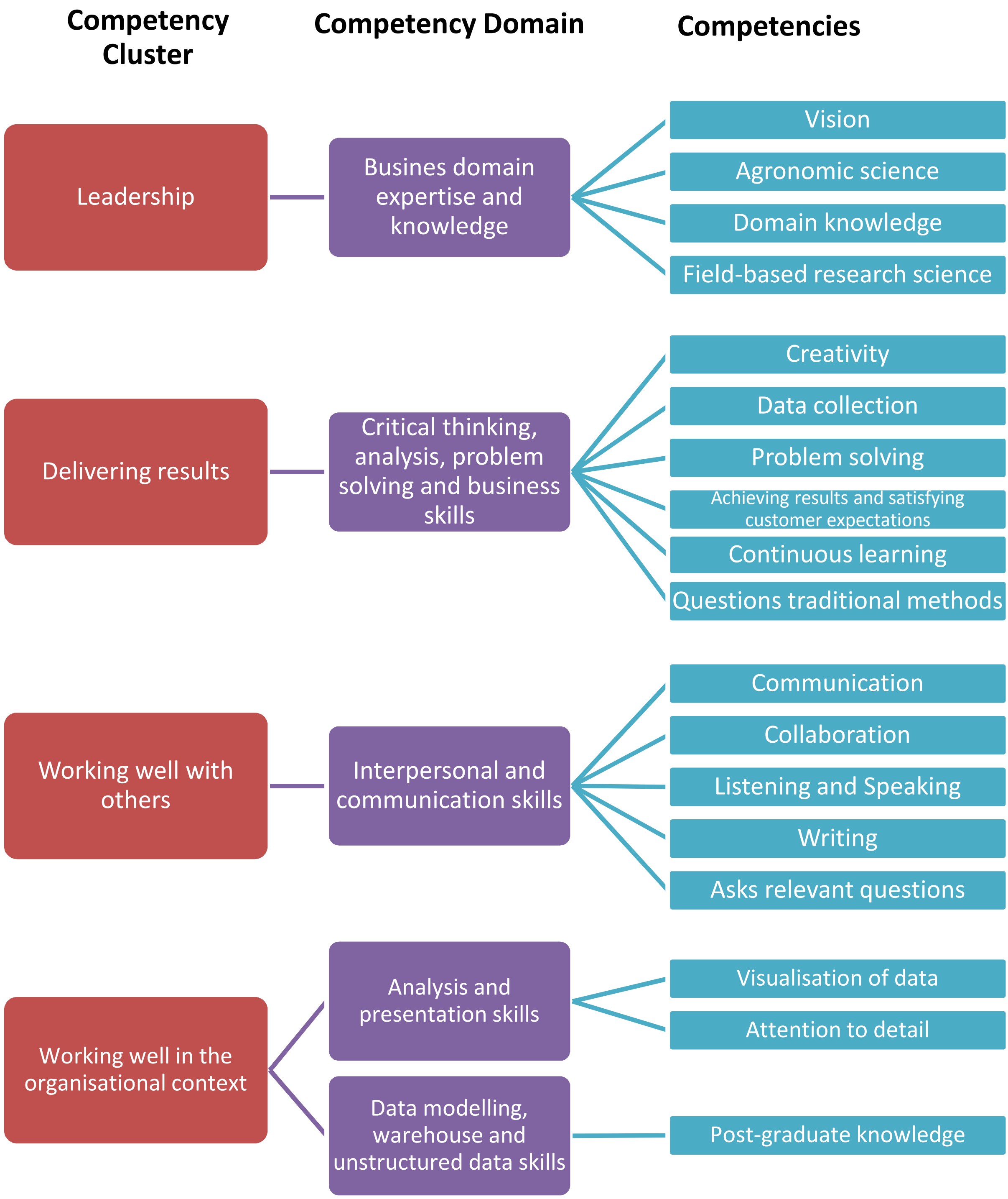

Figure 4.4: A graphical representation of the core competencies for a data science coordinator in Grain SA in terms of competency clusters, domains and individual competencies

4.2.5. A PROPOSED COMPETENCY MODEL FOR A DATA SCIENCE COORDINATOR IN GRAIN SA

The final step is the development of the competency model. Brits (2012) regards competency modelling as a way to identify expected behaviours that will enable the effective implementation of strategic processes. Although behavioural indicators do not form part of the eight-stage process discussed by Brits (2012), competencies are usually described in behavioural terms (Campion et al., 2011).

The behavioural terms were identified through the use of competency libraries and the empirical results of this study. The researcher made use of the same competency libraries used for the clusters and domains to add behavioural indicators (Harvard University, 2015; HRSG, 2015; TMA, 2012; WSU, 2015). The conceptualised competencies are included in a proposed competency model for a data science coordinator in Grain SA.

Table 4.5 represents the proposed competency model for a data science coordinator in Grain SA. The competencies are listed according to their domains and clusters. Each competency has a title and definition followed by a breakdown of the behavioural indicators that contribute to the specific competency.

Table 4.5: A proposed competency model for data science coordinators in Grain SA

Competency Cluster: Leadership			
Competency Domains	**Competencies**	**Definition**	**Behavioural Indicators**
Business domain expertise and knowledge	Vision	Ability to see beyond daily tasks and explore ideas for the future. Regards the facts from a distance and sees them in a broader context or over a longer term	Able to plan ahead despite limited information.
			Has an idea of the direction in which the discipline will develop based on social developments.
			Open to unusual and daring ideas to implement in the discipline.
			Takes time to think about the discipline.
			Understands the consequences of developments and translates these appropriately for their position or discipline.
			Communicates essentials and does not become distracted by details.
			Recognises innovative ideas proposed by the team and knows how to connect them.
			Combines various social trends and developments into an integrated vision of the future.
			Creates and considers possibilities others may ignore or deem impossible.
			Recognises national and international trends early on and manages the consequences for the organisation.
			Sees chances and opportunities for the organisation before others do and acts accordingly.
	Agronomic Science	Understands the interactions between soil and crop physiology and how to formulate insights into useful recommendations	Recognises trends in theory and practice in one's own technical area and effectively prepares for anticipated changes.
			Agronomic knowledge and experience and/or fundamental understanding of agriculture in general.
	Domain knowledge and technical expertise	Applies and improves their extensive or in-depth specialised knowledge, skills, and judgment to accomplish a result or to accomplish one's job effectively.	Understands technical aspects of the occupation, continuously builds knowledge and stays up to date with technical and procedural aspects.
			Makes oneself available to help solve technical or procedural problems or issues.
			Designs ways to apply new developments to improve organisational performance or customer service.
			Applies technical and procedural knowledge to correctly address a problem in a timely manner.
	Field-based research science	Conducting and managing the field research component of data science	Conducts field research to measure statistical differences in terms of performance.
			Reads and understands trial reports. Able to link the purpose and outcome of the trial and interpret the observations.

Competency Cluster: Delivering results			
Critical thinking, analysis, problem solving and business skills	Creativity	The ability to come up with original and innovative ideas and solutions and adopt points of view outside the usual parameters	Generates original ideas.
			Considers new ideas that currently seem impossible to others.
			Generates unconventional solutions and ideas.
			Able to connect concepts and views from different disciplines.
			Speaks in terms of possibilities rather than problems.
			Experiments with new methods and opportunities.
			Seeks better alternatives.
			Recognises connections between seemingly unconnected aspects.
	Data collection	How data is collected from different devices and in different formats	Knowledge of the different data collection equipment.
			Able to upload data from the different devices to an interface from which it can be accessed again.
			Understands geo-spatial information systems.
	Problem solving	Generates creative approaches to addressing problems and opportunities. Makes sound decisions after reviewing all relevant information	Considers all areas involved when accomplishing a task.
			Shows concern for all aspects of the job, no matter how small.
			Accurately checks processes and tasks.
			Monitors trends and problems over the long term.
	Achieving results and satisfying customer expectations	The ability to take direct action in order to attain or exceed objectives	Attains or exceeds set goals.
			Comes up with alternatives when certain actions have not lead to a desired result.
			Defines objectives in terms of tangible results (measurable results within a deadline).
			Indicates how objectives will be attained in terms of tangible actions (who, what, when etc.).
			Takes corrective action when objectives are in danger of not being attained.
			Tenacious, actively seeking alternative possibilities when confronted with adversity.
			Actively seeks the best way to achieve goals and considers options carefully.
			Redirects processes regularly to focus on objectives.
			Sets high standards and defines challenging yet feasible objectives.
			Considers challenging objectives that have an impact on the performance of other individuals or departments.

Competency Cluster: Working well with others			
Interpersonal and communication skills	Communication: oral and written	Effectively transfers thoughts and expresses ideas verbally or in writing in individual or group situations	Presents oneself clearly and articulately when speaking with individuals or before a group. Ensures others fully comprehend the intended message.
			Checks for understanding of the communication by asking open-ended questions that draw out the listener's views.
			Effectively uses appropriate literature or visual aids during product/service demonstrations or when giving presentations.
			Sources material for presentations in advance and organises presentations logically.
			Pays attention and repeats message back to speaker in a way that demonstrates understanding.
	Collaboration	Fosters cooperation and teamwork while participating in a group. Works toward solutions which generally benefit all involved parties	Demonstrates respect for the opinions of others.
			Identifies and pushes for solutions all parties can benefit from.
			Helps and supports fellow employees in their work to contribute to organisational success.
			Keeps others informed and up to date.
			Shares information and own expertise with others to enable them to accomplish group goals.
	Receptivity	The ability to show one absorbs and understands important verbal and non-verbal information and to ask further questions when necessary	Demonstrates they are listening through body language and eye contact.
			Does not interrupt but allows the other person to finish speaking.
			Paraphrases the other person's story.
			Asks questions until everything is answered.
			Briefly summarises the other person's point of view.
			Checks whether their summary is a correct representation.
			Is able to read non-verbal communication cues.
			Listens to content and retrieves information from the non-verbal behaviour at the same time.
			Listens to what is said but also pays attention to what is not said.
	Writing	Able to transfer thoughts and ideas onto paper and write reports	Writes to convey a certain message in an understandable manner.
	Investigative	Asks relevant questions	Asks numerous questions on technical and functional matters.

Competency Cluster: Working well in the organisational context			
Analysis and presentation skills	Visualisation of data	Able to present data visually and in a way that is easy to see, read and understand	Able to present solutions and recommendations in an easy-to-read visual manner.
	Attention to detail	Thoroughness in accomplishing a task through concern for all the areas involved, no matter how small. Monitors work or information and plans and organises time and resources efficiently	Double checks the accuracy of information and products to provide consistently high quality work.
			Provides information on a timely basis and in a usable form to others who need it.
			Carefully monitors the details and quality of one's own and others' work.
			Ensures tasks are done correctly, thoroughly or precisely.
			Completes all work to relevant procedures and standards.
	Lifelong learner	Demonstrates eagerness to acquire necessary technical knowledge, skills and judgment to achieve a result or to serve a customer's needs effectively. Desire and drive to acquire knowledge and skills necessary to perform tasks more effectively	Keeps up to date with current research and technology and identifies and pursues areas for development and training that will enhance job performance.
			Takes responsibility for one's own development.
			Maintains fluency in appropriate work applications, software or tools.
			Reviews, selects and disseminates information regarding key technologies, best practices and tools to others in the group.
			Continually looks for ways to expand job capabilities.
	Questions traditional methods	Does not accept or rely on traditional methods alone	Questions traditional methods and considers innovative products and approaches.
Data modelling, warehouse and unstructured data skills	Post-graduate level knowledge	Masters or PhD level competency in physics, computer science, data science, statistics and/or mathematics	Knowledge of the relationship between data and outcomes.
			Transforms a real world problem into a mathematical problem. Interprets mathematical objects or information in relation to the situation presented.
			Understands machine learning algorithms and how to use each one for different problems.

4.3. SUMMARY

The role data science plays in agriculture is receiving more attention as companies in the sector begin investing in data science tools and expertise. Data science includes the use of large volumes of data from disparate sources to help solve producers' problems. Data scientists turn information into profitable insights, delivering data products that support decision making on all levels.

At the beginning of the study, the importance of agriculture to support food security in any country was established. Grain SA's presence and role in the grain industry was described and a problem was identified. Grain SA will likely need to appoint data scientists in order to stay relevant as a commodity organisation but the organisation does not have a competency model or guidelines to point them in the right direction when sourcing personnel.

The second chapter outlined the use of a competency-based approach and why it was decided to develop a competency model for data scientists in Grain SA. It also explained a step-by-step process to develop the competency model. Literature on the agricultural sector, data science in general as well as data science in agriculture was then reviewed. The literature suggests that technological change is not only an improvement but necessary to reduce poverty, foster development and stimulate economic growth (Spielman & Birner, 2008). Data science can be applied in agriculture to improve crop yields and assist farmers to utilise natural resources more efficiently.

Through the empirical research, the researcher came to the conclusion that much of the feedback and insights from the role players who participated in the study is aligned with the theory. This helped the researcher acquire an in-depth understanding of the role and competencies of data scientists in agriculture specifically. It also assisted in the process of conceptualising a competency model for data science coordinators in Grain SA.

The literature and empirical results show that there are different types of data scientists and the role of the data scientist should not be narrowly defined. Harris et al. (2013) suggest that the differentiation should not be based on the individual's breadth of knowledge but rather on their depth of knowledge in a specific area in relation to others and their preferred methods of addressing data science problems.

While addressing the research question on the role of a data scientist in Grain SA, it was revealed that Grain SA will benefit from appointing a data science coordinator. The coordinator should be concerned with the organisation and to what effect data projects can be initiated and managed. They need to know the various industry agents and their interrelationships.

In the last section of chapter 4 a competency model was conceptualised for a data science coordinator in Grain SA. Chapter 5, the final chapter, contains the conclusion, limitations and recommendations of the study.

5. CHAPTER 5 – CONCLUSIONS AND RECOMMENDATIONS

5.1. INTRODUCTION

Pressure on food security as a result of population growth is a current reality as is the limit on natural resources. Agriculture is a major player in the world economy and the only industry that can feed growing nations and ensure food security on all levels. The use of data science is supportive of the goal to increase productivity, profitability and efficiency in the use of limited natural resources to produce food. A data scientist in Grain SA needs to have a specific set of competencies and the appropriate combination of knowledge, skills and attributes to confront these challenges. The future of agriculture depends on the effective use of "big data" and continuous decision making as part of the modern farming framework. This final chapter includes the conclusion, limitations and recommendations made to Grain SA stemming from the research of the study.

5.2. CONCLUSION

The use of data science in agriculture is becoming more evident as producers and agribusinesses recognise the value derived from it as a decision-making tool. Data scientists support producers by supplying more timely information in a simple format to help them make quick and informed decisions.

It is clear from the results of this study that there is no universal role for data scientists in all organisations; it depends on the context and objectives of the organisation. It is advised that organisations that intend to appoint data scientists should start by clarifying the role and competencies they require based on the different role definitions. They should clearly define the context in which the data scientist is required to function before recruiting. This would be beneficial to the job holder, the organisation as well as the other role players involved.

At first, the role of data scientists in Grain SA was unclear due to the fact that the creation and maintenance of data science products is a complex process that requires multiple expertise on various levels. Both the literature and the results of the empirical research indicate that researchers and practitioners need to make a distinction between the various roles that data scientists may play within certain contexts and different organisations. These different roles should be defined clearly according to the specific contributions made by data scientists in each of these roles. The researcher combined the descriptions gained from both the literature and the empirical data to classify the role of data scientists into four types, namely, data science coordinators, data analysts, data developers and data researchers.

Due to the specific role that Grain SA plays in the grain industry in South Africa, as well as the context within which the organisation operates, the researcher needed to identify the most appropriate role for a data scientist appointed in Grain SA, based on these various types of data scientists that were identified during this research. It was concluded that the role of a data science coordinator would be most appropriate within the current Grain SA context. Therefore, the role of a data science coordinator in Grain SA should include the following:

- Building relationships with producers as well as other role players in agriculture with regard to data science;
- Knowing of all the role players and how they are connected;
- Building trust among producers so that they are willing to engage in data sharing; and
- Facilitating policy making that is beneficial for the agricultural industry and addresses the issues of data privacy and protection among others.

A data science coordinator in Grain SA should have the competencies identified in the study.

A secondary contribution of the study is to indicate the important role data scientists can play in increasing production and food security for South Africa and the world. Data scientists should be driven by this challenge and need to work with organisations such as Grain SA to do so. In order to use resources effectively and efficiently, including data and technology, capable human resources are needed. This study provided a road map for individuals, organisations and other role players in agriculture as to how this can be done.

5.3. LIMITATIONS

The use of data science in agriculture is relatively new and little research has been done within the discipline. After an exhaustive search, the researcher had to work with limited literature with regard to the use of data science in agriculture. No literature could be found on data science in South African agriculture.

Although the empirical section of the current study focused on data scientists in the USA as it is a country where various agricultural innovations originates from and perceived as a rich source of information. However, it is not the only source from which data can be collected. The study did not include data on the use of data science in agriculture from other countries.

The competency model is based on both an extensive literature review as well as the information gathered through the empirical study; it has not been validated by subject-matter experts.

5.4. RECOMMENDATIONS

The results from the interviews show that data science in agriculture is a growing trend of which the importance should not be underestimated. Based on the role and competencies identified by means of this research, the following recommendations can be made:

Firstly it is recommended that Grain SA appoint a data science coordinator.

Secondly, the data science coordinator needs to have a solid foundation in domain knowledge, in this case, grain production as well as critical thinking and analysis, problem solving and business skills. Communication and collaboration are some of the competencies that differentiate mediocre data scientists from excellent ones. An individual who is not afraid of setting challenging goals and pursuing new avenues to achieve these goals is best suited to the role. The role is new within Grain SA and a first for South Africa’s grain industry. The learning curve will thus be steep and needs to move at a fast pace.

Grain SA operates objectively and has earned producers' trust. Therefore, the third recommendation is that the organisation plays a leading role in South Africa in the use of data science and data scientists and thereby sets an example for other role players. By doing that Grain SA can enable farmers to become more efficient in managing their operations which will translate into increased production and profitability and thereby contribute to food security.

In terms of the limitations of the empirical research, it is also recommended that future studies be conducted in other organisations in South Africa and other countries that are employing data scientists and that the competency model be validated and tested by interviewing experts and possibly HR managers to add value to this research.

REFERENCE LIST

Akkermans, J., Brenninkmeijer, V., Huibers, M., & Blonk, R.W.B. (2012). Competencies for the contemporary career: Development and preliminary validation of the Career Competencies Questionnaire. *Journal of Career Development*, 40(3), 245-267.

Alston, J.M., & Pardey, P.G. (2014). Agriculture in the Global Economy. *Journal of Economic Perspectives,* 28(1), 121-146.

Al-Yateem, N. (2012). The effect of interview recording on quality of data obtained. A methodological reflection. *Nurse Researcher,* 19, 31-35.

Aziz, Z. (2005). Developing a robust competency framework for the Malaysian banking industry. Industry Competency Framework Study Focus Group Discussion, Kuala Lumpur, 6 December.

Baptiste, I. (2001).Qualitative Data Analysis: Common Phases, Strategic Differences 2. QDA: Four Phases, Numerous Strategies. *Forum Qualitative Sozialforschung/ Forum: Qualitative Social Research*, 2(3), 26.

Bartram, D. (2012). The SHL Universal Competency Framework. SHL White Paper 2012.

Batchelor, S., Scott, N., Manfre, C., Lopez, A.V. & Edwards, D. (2014). *Is there a role for mobiles to support sustainable agriculture in Africa?* Amsterdam: Atlantis Press.

Bell, B.S., Lee, S. & Yeung, S.K. (2006). HRM: Implications for the Professional. *Human Resource Management*, 45(3), 295-308.

BFAP. (2014) *BFAP Baseline Agricultural Outlook 2014-2023.* Pretoria.

Biodiversity, CGIAR Consortium, FAO, IFAD, IFPRI, IICA, OECD, UNCTAD, WFP, World Bank, WTO. (2012) *Sustainable agricultural productivity growth and bridging the gap for small-family farms. Interagency report to the Mexican G20 Presidency.* Mexico.

Brits, D.W. (2012). *Building and validating a competency model delivered by a corporate university.* University of Johannesburg.

Brown, T. (2006). Stop Competency Blunders. *The American Society for Training and Development (ASTD) Magazine,* January, 1-3.

Bryman, A. & Bell, E. (2007). *Business Research Methods* (2nd edn). New York: Oxford University Press.

Buford, J.A. & Lindner, J.R. (2002). *Human resource management in local government: Concepts and applications for students and practitioners*. Cincinnati, OH: South-Western.

Burnett, E. (2011). Development of the competency framework: Our journey. *Journal of Infection Prevention*, 12(2), 50-52.

Campion, M.A., Fink, A.A, Ruggeberg, B.J., Carr, L., Phillips, G.M. & Odman, R.B. (2011). Doing competencies well: Best practices in competency modelling. *Personnel Psychology*, 64, 225-262.

CEC. (2015). *Preliminary area planted estimate for summer crops 2015.* Pretoria.

Chaura, B., Gachoud, A., Kafera, G., Kara, G., Manja, J., Mdladla, P., Ndambukia, A., Rammala, V., Sinnathamby D. & Zeelie, H. (2015). Food and Nutrition Security Working Group Paper. Retrieved 14 September 2015 from https://www.humanitarianresponse.info/fr/system/files/documents/files/fnswg_update_issue2_june2015.pdf.

CIPD. (2014). Competence and Competency Frameworks. Retrieved 19 October 2014 from http://www.cipd.co.uk/hr-resources/factsheets/competence-comptency-frameworks.aspx.

Coetzee, M. & Schreuder, D. (2013). *Personnel Psychology: An Applied Perspective.* Cape Town: Oxford University Press.

Collins, H. (2010). *Creative Research: The Theory and Practice of Research for the Creative Industries.* Worthing, West Sussex, UK: AVA Publishing.

Creswell, J.W. (2003). *Research Design Qualitative, Quantitative, and Mixed Methods.* Lincoln, TX*:* Sage.

Crosbie, T. & Friedberg, D. (2013). New info tools for agriculture. *Agri Marketing*, October, 46.

Croucamp, A. (2013). *Developing a Competency model for Head of Departments at Tertiary Education Institutions in South Africa.* University of Cape Town.

CTSI. (2014). Ethical Considerations in Research. Clinical Translational Research Fellowship Program Curriculum Core. Retrieved 20 September 2015 from https://accelerate.ucsf.edu/files/CTRFP_Ethical_Considerations.pdf.

Davenport, T.H. & Patil, D.J. (2012). Data scientist: The sexiest job of the 21st Century. *Harvard Business Review*, 19(10), 70-76.

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods*. New York: McGraw-Hill.

Doerflein, S. (2007). Measure the effectiveness of your competency model. Retrieved 10 February 2015 from http://www.slideshare.net/sdoerflein/leadership-competency-modeling-best-practices-presentation.

Dubois, D.D. & Rothwell, W.J. (2004). *Competency–Based Human Resource Management*. Palo Alto, CA: Davies Black Publishing.

Edgar, W.B. & Lockwood, C.A. (2011). Understanding, finding and applying core competencies: A framework, guide, and description for corporate managers and research professional. *Academy of Strategic Management Journal*, 10(2), 61-82*.*

FAO, IFAD & WFP. (2014). *The State of Food insecurity in the World 2014. Strengthening the enabling environment for food security and nutrition.* Rome: FAO.

FAO. (2003). *World Agriculture: Towards 2015/2030. An FAO Perspective.* London: Earthscan.

Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327-358.

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M. & Zaks, D.P.M. (2011). Solutions for a cultivated planet. *Research Analysis,* 478, 337-342.

Garrett, S. (n.d.). Competency Mapping. What is it and how it can be done by individuals. Retrieved 23 January 2015 from http://www.citehr.com/1283-skill-mapping-competency-mapping-career-planning.html.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *Test*, 8(4), 597-606.

Goldblatt, D.A. (2012). *Agriculture: Facts & Trends South Africa.* Cape Town: WWF.

Grain SA. (2013). *Inputs research and development.* Pretoria.

Grain SA. (2015a). *Supply and Demand for Maize in South Africa.* Pretoria.

Grain SA. (2015b). *Graangids 2015.* Pretoria: Infoworks.

Greater London Authority. (2014). Competency Framework. Guide for Managers and staff. Retrieved 19 October 2014 from https://www.london.gov.uk/sites/default/files/Competency-framework-guidelines.pdf.

Harris, H.D., Murphy, S.P. & Vaisman, M. (2013). *Analysing the Analysers – An Introspective Survey of Data Scientists and their Work.* Sebastopol, CA: O'Reilly Media Incorporated.

Harvard University. (2015). Harvard University Competency Dictionary. Retrieved 1 May 2015 from https://www.campusservices.harvard.edu/system/files/documents/1865/harvard_competency_dictionary_complete.pdf.

Hillion, S. (2011). What is a data scientist? Retrieved 25 August 2014 from http://www.forbes.com/sites/danwoods/2011/10/11/emc-greenplums-steven-hillion-on-what-is-a-data-scientist.

Hosack, B., Power, D. & Sagers, G. (2014). Designing Data Science Graduate Programs: A Case for Applied Doctorates in IS. MWAIS 2014 Proceedings, Paper 16. AIS Electronic Library retrieved 10 February 2015 from http://aisel.aisnet.org/mwais2014/16.

HRSG. (2015). Accelerate competency-based management in your organization with competency dictionaries. Retrieved 1 May 2015 from http://www.hrsg.ca/wp-content/uploads/2014/08/HRSG-Dictionary-Overview-List.pdf.

Kennett, P. (2008). Pitfalls to avoid when implementing Competency Frameworks. Retrieved 8 February 2015 from http://www.articlesbase.com/business-articles/pitfalls-to-avoid-when-implementing-competency-frameworkds-312804.html.

Leko, M.M. (2014). The value of Qualitative Methods in Social Validity Research. *Remedial and Special Education,* 35(5), 275-286.

Lievens, F., Sanchez, J.I. & De Corte, W. (2004). Easing the inferential leap in competency modelling: The effects of task-related information and subject matter expertise. *Personnel Psychology*, 57, 881-904.

Loukides, M. (2012). *What is Data Science? The future belongs to the companies and people that turn into products.* Strata. Sebastopol, CA: O'Reilly Media Incorporated.

Lubbe, L. (2012). *A competency model for security officers: A qualitative design.* Pretoria: University of South Africa.

Manyika, J., Chuis, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A.H. (2011). *Big data: The next frontier for innovation, competition, and productivity.* Mumbai: McKinsey Global Institute.

Martone, D. (2003). A guide to Developing a Competency-Based Performance-Management System. *Employment Relations Today*, 30, 23-32.

Maxwell, J.A. (2009). Designing a Qualitative Study. *The Sage Handbook of Applied Social Research Methods,* pp. 214-253.

McClelland, D.C. (1973). Testing for Competence rather than for “Intelligence”. *American Psychologist,* 14.

McLagan, P.A. (1997). Competencies: The next generation. *Training & Development*, 51(5), 40-47.

Mirabile, R. (1997). Everything you wanted to know about Competency Modeling. *Training & Development*, 51(8), 73-77.

Monsanto (2014a). The role of Data Science in Agriculture. Retrieved 30 September 2014 from http://news.monsanto.com/image/infographic/role-data-science-agriculture.

Monsanto (2014b). Launch of Open Ag Data Alliance Supports Farmer Privacy and Control of Their Data, Enables Adoption of Data Science-Powered Agriculture. *Business Wire.*

Murphy, K.R. (2010). “Individual differences that influence performance and effectiveness. What should we assess?” In Scott, J.C & Reynolds D.H. *Handbook of Workplace Assessment*. San Francisco, CA: Jossey-Bass.

Parry, S.B. (1996). The quest for competencies. *Training.* Accessed on 8 February 2015 from http://0-eds.b.ebscohost.com.wagtail.ufs.ac.za/eds/detail/detail?vid=25 &sid=0629e63f-073f-4557-94b5-66f8f355c4ba%40sessionmgr115&hid=108&bdata =JnNpdGU9ZWRzLWxpdmU%3d#db=edsgbe&AN=edsgcl.18756435.

Patton, M.Q. (2002). *Qualitative research and evaluation methods* (3rd edn). Thousand Oaks, CA: Sage.

Power, D. (2015). What universities offer masters degrees in analytics and data science? Retrieved 22 January 2015 from http://dssresources.com/faq/index.php?action=artikel&cat=id=250&artlang=en.

Rauser, J. (2011). What is a Data Scientist? Retrieved 10 October 2014 from http://www.forbes.com/sites/danwoods/2011/10/07/amazons-john-rauser-on-what-is-a-data-scientist.

Sanchez, J.I. & Levine, E.L. (2009). What is (or should be) the difference between competency modelling and traditional job analysis? *Human Resource Management Review,* 19(2), 53-63.

Sanders, M. (2013) Data Scientist Core Skills. Retrieved 12 October 2014 from http://www.datasciencecentral.com/profiles/blogs/data-scientist-core-skills?xg_source=activity.

Scheer, S., Cochran, G., Harder, A. & Place, N. (2011). Competency Modeling in Extension Education: Integrating an Academic Extension Education Model with an Extension Human Resource Management Model. *Journal of Agricultural Education*, 52(3), 64-74.

Schurink, W. & Schurink, E. (2012). Workshop in Qualitative Research (course notes). Presented at the University of the Free State, Bloemfontein, 29-31 May 2012.

Schwandt, T.A. (2007). *The Sage Dictionary of Qualitative Inquiry*. Thousand Oaks, CA: Sage. pp. 298-299.

Sekaran, U. & Bougie, R. (2013). *Research Methods for Business: A Skill-Building Approach* (6th edn). West Sussex, UK: John Wiley & Sons.

Sephton, M., Kemp, K. (2014). A competency framework for inflammatory bowel disease nurses. *Nursing Standard (Royal College of Nursing (Great Britain),* 27, 41-5.

Shenton, A. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22, 63-75.

Shiya, MA. (2011). *Implementation of the South African excellence model (entry level) to measure and improve performance of SMEs in agriculture and related businesses.* Bloemfontein: Central University of Technology.

Soderquist, K.E., Papalexandris, A., Loannou, G. & Prastacos, G. (2010). From task-based to competency-based: A typology and process supporting a critical HRM transitions. *Personnel Reviews*, 39(3), 325-346.

Spielman, D.J. & Birner, R. (2008). *How Innovative is your agriculture? Using Innovation Indicators and Benchmarks to strengthen National Innovation Systems.* Washington.

Stanton, J., Palmer, C.L., Blake, C. & Allard, S. (2012). Interdisciplinary Data Science Education. *Special Issues in Data Management, 1110,* 97-113.

Stevens, G.W. (2013). A critical review of the science and practice of Competency modelling. *Human Resource Development Review*, 12(1) 86-107.

Strauss, A. (1987). *Qualitative Analysis for Social Scientists*. Cambridge, UK: Cambridge University Press.

Syngenta (2014). *Syngenta: Our Industry 2014.* Basel, Switzerland.

TMA. (2012). TMA Competency Library. Retrieved 1 May 2015 from http://www.tmamethode.nl/downloads/tmaCompetencymodel1-120eng.pdf.

Tom, K. (2014). The Green Data Revolution. Frankfurt, Germany.

Tracy, S.J. (2010). Qualitative Quality: Eight “Big-Tent” criteria for excellent qualitative research. *Qualitative Inquiry*, 16(10), 837-851.

Truesdell, J. (2001). Focusing on core competencies. *Supply House Time,* 50-53.

Umachandran, K. (2013). *Data Science for Agriculture.*

Vogt, W. (2013). Looking at Big Data one plant at a time. Retrieved 28 September 2014 from http://farmindustrynews.com/blog/looking-big-data-one-plant-time.

WSU. (2015). Performance Planning and Development Program. Competency Library. Detroit: Wayne State University.

White, R.W. (1959). Motivation reconsidered: The concepts of competence. *Psychological Review*, 66, 297-333.

ADDENDUM: BRIEFING SHEET FOR INTERVIEWS

Grain SA: Bloemfontein
Suite 3, Private Bag X11, Brandhof, 9324
7 Collins Street, Arboretum
Bloemfontein
Telephone: 08600 47246, Fax: 051 430 7574

1 April 2015

To whom it may concern

DEVELOPMENT OF A COMPETENCY MODEL FOR DATA SCIENTISTS IN GRAIN SA

Data science is a fairly new concept in the business world and has been implemented successfully in various sectors/industries including agriculture. Decision making forms the basis of management in farming and farmers rely on both external and internal sources for effective management. Therefore, the application of data science in agriculture will enable farmers to analyse different variables simultaneously and help them to make quick, informed farming decisions.

Grain SA is a commodity organization owned by its members, the grain producers of South Africa and is concerned with the well-being of the grain industry. Their mission is to provide commodity strategic services to the producers to support sustainability, continuous production and a food-secure country. For Grain SA to stay true to its mission to provide continuous strategic support to farmers in the 21st century, the organization needs to stay technological relevant in changing times and a new opportunity that has been identified by Grain SA is the optimal use of all relevant available research data by means of the application of data science.

Since data science is a new discipline that has not yet been implemented in Grain SA, it would not only need to be introduced to farmers and the agribusiness as a whole, but the implementation thereof would need to be monitored. To be able to capitalise on big data, Grain SA would be required to recruit and appoint a data scientist with the necessary skills and expertise to manage and distribute big data.

In order for Grain SA to identify and attract the necessary talent (data scientist), the role and competencies related to an effective data scientist within the agricultural sector need to be clarified.

Competencies are general descriptions of the underlying knowledge, skills, abilities and other characteristics needed in people to ensure worthy performance on the job. In other words, the set of behaviours instrumental in the delivery of desired organisational results or outcomes. Various authors refer to a competency model and describe it as a selection of competencies required by a specific occupational group – in this case that of a data scientist in Grain SA. The development of a competency model allows an organisation to identify the behaviours that drive successful performance and enables the organisation to deliver their technical expertise effectively. From the literature it is evident that a competency model forms the basis for the recruitment and evaluation of potential candidates for specific positions.

In order to operationalize the research, the following research questions have been formulated:

- What is the current role of data scientists in agriculture internationally?
- What is the appropriate role for data scientists in Grain SA?
- What are the core competencies that are necessary for an effective data scientist in Grain SA?
- What would a competency model for data scientists in Grain SA entails?

The researcher will interview participants with a profound knowledge of data science as well as its role in agriculture. Sampling will continue until theoretical saturation has been reached. Theoretical saturation is reached when now new information arises in repeated cases. Some of the questions that will be asked in the interviews are for the purpose of the field study and of an academic nature while other questions are to be practical. The following questions will be used as a guideline for the interviews although it may differ amongst the various interest groups.

POSSIBLE RESEARCH QUESTIONS

1. According to you, what is the role of a data science in agriculture currently?
2. What would the role/position of a data scientist need to deliver in order to add value to you as a stakeholder?
3. According to you, what knowledge, skills, attributes does a data scientist need to be effective in his/her role?
    3.1. Knowledge: factual information that a person knows and that is needed for a specific job
    3.2. Skills: an ability that has been acquired by training and education that enables a person to consistently perform a complex task efficiently.
    3.3. Attributes: include personal characteristics, traits, motives and values or ways of thinking that affect an individual's behaviour.
4. What would a competency model for a data scientist entail?
5. What would differentiate an excellent data scientist from a good/mediocre one?
6. Please tell me about your thoughts on training on the subject? Who should do it? Who should be trained?
7. What type of education would a typical data scientist need to undergo? How would you train a data scientist?
8. What type of experience would you require from a data scientist?
9. In which areas should a data scientist be trained in?
    9.1. IT Skills, econometric skills, statistical/mathematical/agricultural skills?

PRACTICAL QUESTIONS

1. To what extent are you making use of data science in your farming operations?
2. Since when/how long have you been making use of precision agriculture? What is the timeline of your existing data?
3. Who is the driver/user of technology in your business?
4. Please tell me about your experience with data science software
5. What makes their service good?
6. Are you part of a commodity organization like Grain SA? What role do they play in the industry with regards to data science?
7. What is your expectation of them (commodity organization)? How can they support farmers in the use of data science? What role should they play/are they playing?
8. How do you see the future of data science/data scientists?
9. Any closing remarks?

Any other insights or comments are welcome at any time. Thank you for your interest and participation in the study.

Kind regards

Landi Kruger