

**Technology Forecasting in the National Research and Education Network Technology
Domain using Context Sensitive Data Fusion**

L. Staphorst^{1,2}, L. Pretorius³ and M.W. Pretorius⁴

¹South African National Research Network (SANReN),
Council for Scientific and Industrial Research, South Africa,
Postal Address: PO Box 395, Pretoria, 0001, South Africa

²Department of Engineering and Technology Management,
University of Pretoria, South Africa

Postal Address: Department of Engineering and Technology Management, University of Pretoria,
Lynnwood Road, Hatfield, Pretoria, 0001, South Africa

Email: leon.staphorst@gmail.com

Tel: +27 82 857 1135

(Corresponding Author)

³Department of Engineering and Technology Management,
University of Pretoria, South Africa

Postal Address: Department of Engineering and Technology Management, University of Pretoria,
Lynnwood Road, Hatfield, Pretoria, 0001, South Africa

Tel: +27 12 420 4148

Email: leon.pretorius@up.ac.za

⁴Department of Engineering and Technology Management,
University of Pretoria, South Africa

Postal Address: Department of Engineering and Technology Management, University of Pretoria,
Lynnwood Road, Hatfield, Pretoria, 0001, South Africa

Tel: +27 12 420 4606

Email: tinus.pretorius@up.ac.za

ABSTRACT

Using inductive reasoning this paper develops a framework for the Structural Equation Modeling based context sensitive Data Fusion of technology indicators in order to produce Technology Forecasting output metrics. Data Fusion is a formal framework that defines tools, as well as the application of these tools, for the unification of data originating from diverse sources. Context sensitive Data Fusion techniques refine the generated knowledge using the characteristics of exogenous context related variables, which in the proposed framework entails non-technology related metrics. Structural Equation Modeling, which is a statistical technique capable of evaluating complex hierarchical dependencies between latent and observed constructs, has been shown to be effective in implementing context sensitive Data Fusion. For illustrative purposes an example model instantiation of the proposed framework is constructed for the case of the National Research and Education Network technology domain using knowledge gained through action research in the South African National Research Network, hypotheses from peer-reviewed literature and insights from the Trans-European Research and Education Network Association's annual compendiums for National Research and Education Network infrastructure and services trends. This example model instantiation hypothesizes that a National Research and Education Network's infrastructure and advanced services capabilities are positively related to one another, as well as to the contextual influence it experiences through government control. Also, positive relationships are hypothesized between a National Research and Education Network's infrastructure and advanced services capabilities and its usage, which is defined as the technology forecasting output metric of interest for this example. Data from the 2011 Trans-European Research and Education Network Association compendium is used in the Partial Least Square regression analysis of the example model instantiation, which confirms all hypothesized relationships, except the postulation that a National Research and Education Network's infrastructure and advanced services capabilities are positively related. This latter finding is explained

by observing the prevalence of technology leapfrogging in the National Research and Education Network global community.

Key words: Technology Intelligence, Technology Indicators, Technology Forecasting, Data Fusion, Structural Equation Modeling, National Research and Education Network

1. Introduction

Technological advancement continues at an astounding rate, seemingly following exponential growth models such as Moore's [1], Nielsen's Law [2] and Metcalfe's Law [3]. Driven not only by the invention, innovation and diffusion of new technologies, but also by the move to the paradigms of globalization and open innovation [4], this has created highly competitive global markets for technology based products and services [5]. Hence, the survival, growth and profitability of firms that play in these markets depend highly on their ability to monitor current, as well as predict future technological changes in order to create a solid and sustainable technological base that can withstand, or adapt to rapidly changing market requirements [5]. Moreover, firms need to effectively and efficiently manage technological changes both internally and externally if they are to create sustainable competitive advantages in rapidly high-tech markets [6]. Technology Intelligence (TI), which is a core process within the discipline of technology management, involves the process of capturing technology related data, converting this data into information by determining relational connections and refining information to produce knowledge that can guide strategic decision makers during strategic planning [6][7]. Technology indicators, such as technology maturity and degree of innovation, are those measureable sources of technology related data that allow for the direct characterization and evaluation of technologies over their whole life cycle [7]. Scrutinizing the information that has been distilled from a set of technology indicators in a forward-looking approach, commonly referred to as Future-oriented Technology Analysis (FTA), can potentially provide decision makers with Technology Forecasting (TF) knowledge, amongst others [8].

Buchroithner [9] and Wald [10] define Data Fusion (DF), which was developed in the military domain for the generation of quality tactical knowledge through the multi-layered processing of sensor data [11], as "... a formal framework in which are expressed means and tools for the alliance of data originating from different sources. It aims at obtaining information of greater quality; the exact definition of 'greater quality' will depend upon the application." Within the discipline of DF, context can be viewed as synonymous with a situation, which in turn is defined as a set of relational connections (i.e. instantiated relations) [12]. Context can be used in each level of the DF process in order to refine data alignment and association, as well as during situation state estimation [12]. Recently, context sensitive DF techniques have been explored which effectively refine the generated knowledge at each level of processing based on the characteristics of exogenous context-related variables [12].

Regression analysis constitutes a family of statistical techniques geared at modeling and analyzing the relationship between dependent and independent variables from empirical data [13]. Moreover, regression analysis attempts to explain the variations in dependent variables as functions (commonly referred to regression functions) of variations in independent variables [13]. With this knowledge it is then possible to perform prediction and forecasting of the values that dependent variable will assume for specific independent variable values [13]. Classic regression techniques (such as multiple regression, discriminant analysis, logistic regression and analysis of variance) can be classified as first generation techniques, since these techniques explicitly assume independence between multiple dependent variables [13]. This, unfortunately, limits the ability of such techniques to comprehensively model complex interrelationships, such as the interplay between two or more output variables in a TF model. More specifically, classic first generation regression techniques are not able to model the potential mediating or moderating effect that output variables could have on one another. To overcome this limitation, Jöreskog [14] proposed covariance based Structural Equation Modeling (SEM) as a second-generation technique, which allows for the simultaneous modeling of relationships among multiple dependent and independent constructs. A further inherent limitation of first generation

regression techniques is their explicit assumption that all dependent and independent variables are directly observable [13]. This assumption implies that all variables' values can be directly obtained from real-world sampling experiments [13]. As such, any variables that cannot be directly observed need to be considered unobservable and have to be excluded from first generation regression models [13]. However, such unobservable variables, commonly referred to as latent constructs, are supported by SEM. Steinberg postulated that SEM is ideally suited to implement context sensitive DF [12][15]. Not only does SEM support the complex structural models used in situation state estimation (as is required in TF), it also allows for non-linear and non-Gaussian factors and cyclical dependencies amongst model variables that can be either latent or directly observable [12].

According to Sohn and Moon [16] most TF techniques rarely take into account the structural relationships amongst technology indicators and TF output metrics. SEM, however, provides an advantage over these limited TF techniques by allowing for the modeling of complex hierarchical relationships between technology indicators and TF outputs metrics. Sohn and Moon [16] have shown that SEM, which can be viewed as a generalization of factor and path analysis methods such as Bayesian Networks [12], can successfully implementing TF of the Technology Commercialization Success Index (TCSI) TF output metric.

An NREN is a specialized broadband network connectivity and service provider that explicitly caters for the needs of the research and education communities of a country [18]. In some instances, NRENs also service the needs of other public sector entities, such as hospitals, municipalities and libraries. Typically, one NREN is present per country (for example SANReN [19] in South Africa and the Joint Academic Network (JANET) in the United Kingdom), although separate NREN entities could potentially exist to service distinct in-country research and education sectors or geographic areas, for example the Energy Sciences Network (ESnet) and Kansas Research and Education Network (KanREN) in the United States [18]. NREN's are built primarily on fiber optic cabling infrastructure and provide researchers, educators and students with unparalleled connectivity speeds and advanced

services at a fraction of the price of commercial network providers [18]. These networks are currently experiencing rapid technology driven changes, resulting in evolving business models, innovative infrastructure solutions and service offerings, as well as increased international collaboration [20][18].

The objectives of this paper are twofold: Firstly, the paper builds on the work of Steinberg [12][15], as well as Sohn and Moon [16], by proposing a framework for the SEM based DF of technology indicators in order to produce TF output metrics. The proposed framework is an evolved and improved version of the framework first proposed in [17]. Secondly, application of the proposed framework is illustrated by through the use of a model instantiation example in the NREN technology domain. The proposed NREN example model instantiation was constructed using insights gained through action research in the South African National Research Network (SANReN) [19], insights from Trans-European Research and Education Network Association's (TERENA) NREN compendium for 2012 on global NREN infrastructure and services trends [18], as well as hypotheses and postulations in peer-reviewed literature. It is important to view the example NREN model instantiation presented in this paper as a mere illustrative example of the use of the proposed framework, not as a definitive platform for TF in the NREN domain.

The paper is structured as follows: Firstly, a theory review is presented on the use of SEM for context sensitive DF and the use of SEM in TF, as well as taxonomy of technology indicators and forecasting output metrics. An evolved version of the framework for SEM based DF for TF proposed by Staphorst, Pretorius and Pretorius in [17] is then developed through inductive reasoning. The example NREN model instantiation of the framework is then presented, including the definition of a number of research propositions relevant to this example model instantiation. This is followed by a quantitative evaluation of the example model instantiation using cross-sectional data extracted from TERENA's NREN compendiums for 2011 [20], including an evaluation of the research propositions defined for the example model instantiation. Lastly, the paper presents an evaluation of the reliability and the validity of the example NREN model instantiation.

2. Theory and Framework Development

Steinberg postulated that SEM is one potential statistical tool that lends itself naturally to implement DF, with the added benefit that it allows for the inclusion of context sensitivity during the solving of DF inferencing problems [12][15] [16]. Sohn and Moon showed in [16] that SEM can be used as a regression technique to evaluate a multi-layered hierarchal model through progressive aggregations and refinements of input technology indicator data in order to produce a reliable statistical estimate of a TF output metric [17]. In the following subsections it will be shown that, through inductive reasoning, Soon and Moon's [16] use of SEM for TF and Steinberg's use of SEM to implement context sensitive DF [12] can be combined to develop a framework for SEM based DF for TF. This was first proposed by Staphorst, Pretorius and Pretorius in [17].

2.1 Structural Equation Modeling for Context Sensitive Data Fusion

Within SEM theory distinction is made between exogenous and endogenous latent constructs, with the former being variables that are not explained by the internal interrelationships embodied by the model and therefore always act as independent variables [13]. Due to its generality, SEM terminology does not refer to regression analysis' dependent and independent variables, but rather only to exogenous constructs, which are independent variables that are not functions of any relationship in the model, as well as endogenous constructs, which are either dependent or independent variables that are explained by the relationships with other dependent and/or independent variables present in the model.

With reference to the indicators measured as proxies to represent latent constructs, such latent constructs can be classified as follows [13]: A latent construct with reflective indicators is one in which all measured indicator proxies, also commonly referred to as factors, are expected to have high correlations to the latent construct, as well as other potential reflective indicators, Therefore it will have the ability to represent the variance in the unobserved variable sufficiently. In contrast, latent constructs with formative indicators are those that are represented by a weighted combination of

indicators that are not highly correlated to either the latent construct itself, or the other formative indicators included in the weighted combination. The formative indicators of a latent construct can therefore be seen as representing different dimensions of this construct.

Although Jöreskog in 1973 [14] originally proposed that the parameters of a SEM model be estimated using covariance based techniques, of which the LISREL program that was developed by Jöreskog in 1975 is arguably the most popular, variance based techniques, also commonly referred to as component based techniques, have also gained popularity [13]. Partial Least Squares (PLS), which was first introduced by Wold [21] as Non-linear Iterative Partial Least Squares (NIPALS), is one such variance based technique [13]. While covariance based techniques attempt to minimize the difference between the sample covariance values and those predicted by the regression model, which is equivalent to estimating the model parameters such that the covariance matrix of the observed measurements is reproduced, PLS regression, which is also sometimes referred to Projections to Latent Structures, focuses on maximizing the variance of the dependent variables explained by the independent variables [13].

It is common practice in SEM to represent models using path diagrams that depict the exogenous and endogenous constructs, the path coefficients of interconnections between these constructs, reflective and formative indicators, as well as the loadings of these indicators on constructs [17]. Graphically any SEM path diagram can be condensed into layered groupings of constructs, as is shown in Fig. 1 [17]. While the groupings and layering of these groupings can be done based on any arbitrary set of criteria, this paper's proposed use of SEM to implement context sensitive DF for TF warrants distinct groupings of exogenous and endogenous constructs, representing context and technology related constructs, respectively. Furthermore, it is proposed that the layering of the groupings is performed in such a manner that context and technology related constructs of a similar nature or complexity occupy the same DF layer.

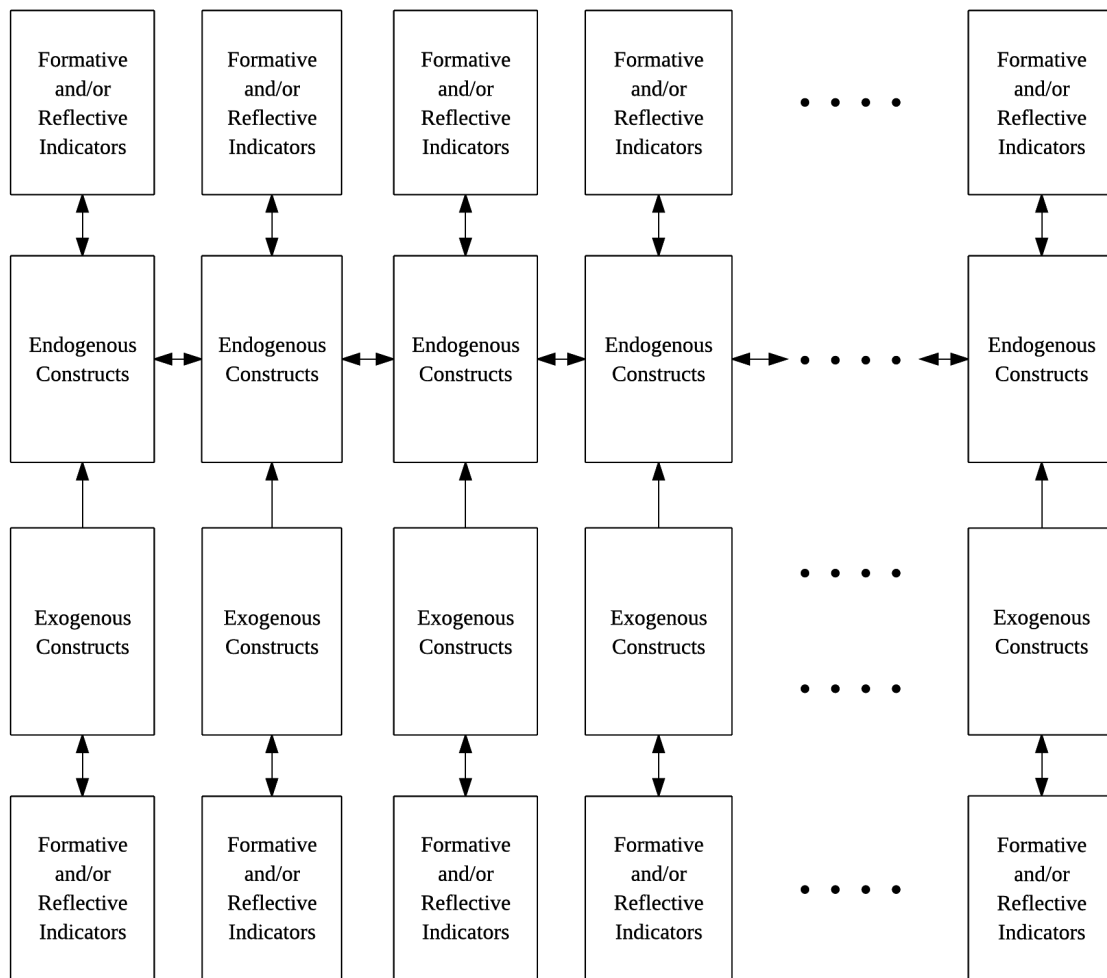


Fig. 1: Grouping and Layering of Constructs in a SEM Path Diagram

Recall that DF is essentially a framework for the multi-layered refinement of estimates of problem variables from multiple measurements, either directly or indirectly observable [12]. The Joint Directors of Laboratories' Data Fusion Group (JDL/DFG) recognized that, in a military environment, DF entails the progressive aggregation and refinement of sensor data in order to produce quality tactical knowledge [12]. In an attempt to standardize the structure of the multi-layered DF process across all possible military applications and implementations, the JDL/DFG defined the following six levels of processing [12]: Level 0 involves signal/feature/subject assessment, level 1 involves object assessment, level 2 involves situation assessment, level 3 involves impact assessment, level 4 involves process refinement and level 5 involves user refinement.

While these JDL/DFG DF level definitions are not appropriate for the use of DF in TF, the concept of multi-layered progressive aggregation and refinement of measurement indicator data is core to this paper's proposed framework. Hence, Fig. 2 depicts a more generalized context sensitive DF framework that supports N levels of aggregation, with N an integer that represents the user of the framework's required number of levels of aggregation and refinement. Applicable context related information can be incorporated into the aggregation and refinement process at any of the N levels of DF processing.

By noting that SEM is capable of the simultaneous modeling of relationships among multiple dependent and independent constructs, Steinberg [12] postulated that SEM is one potential statistical tool that lends itself naturally to implement DF. Moreover, based on the following argumentation Steinberg [12] showed that SEM allows for the inclusion of context sensitivity during the solving of DF inferencing problems: Firstly, Steinberg [12] defined a situation, or a context, as a set of relationships, where a relationship can be viewed as a specific instantiated relation. In general, context is used in DF inferencing problems in order to refine ambiguous estimates, explain available data and constraint processing during data acquisition, cueing or fusion [12]. Next, Steinberg harmonized DF and SEM terminology by noting that DF problem variables are in fact SEM endogenous constructs, context variables can be viewed as SEM exogenous constructs and classic DF sensor measurements are the reflective and formative indicators present in SEM.

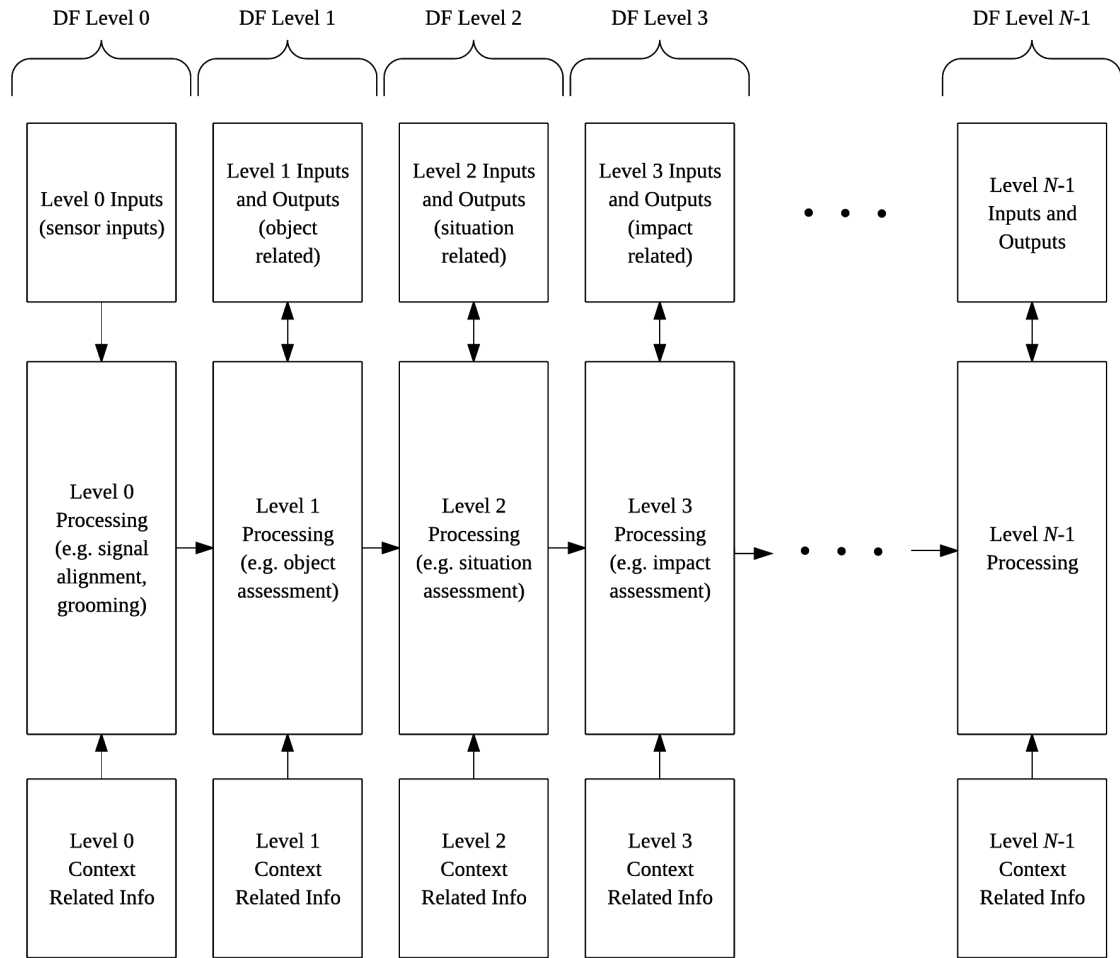


Fig. 2: Generalized Framework for Context Sensitive DF

2.2 Technology Indicators and Forecasting Output Metrics

According to Porter and Cunningham [22] technology indicators employ empirical information to estimate technology characteristics that affect technological advance and successful commercialization. Watts and Porter [23] state that technology indicators are empirical measures stemming from models of technological innovation and progression, such as the S-curve. Nyberg and Palmgren [4] expands on these definitions by describing technological indicators as those indices or statistical data that allow for the direct characterization of characteristics of technology throughout their life cycles in order to allow decision makers to take strategic actions. Such indicators can in general be divided into three major categories based on their intended function: input indicators, byput

indicators and output indicators [4][24]. Grupp [24] states that input indicators are variables related to drivers of technological progress, byput indicators are variables that are related to sub-phenomena of the technological progress and output indicators are variables related to the qualitative, quantitative or value-rated progress in process or product development [4].

As depicted in Fig. 3 input, byput and output technology indicators can be grouped based on similar characteristics or complexity. Furthermore, complex relationships can exist between input, byput and output technology indicators. For the purposes of this paper it is assumed that input technology indicators drive technology processes at Level 0, byput technology indicators indicate sub-phenomena between Level $x-1$ and Level x for $x = 1, 2, 3, \dots, N-1$, and output technology indicators indicate product or process related progress at Level x for $x = 1, 2, 3, \dots, N-1$.

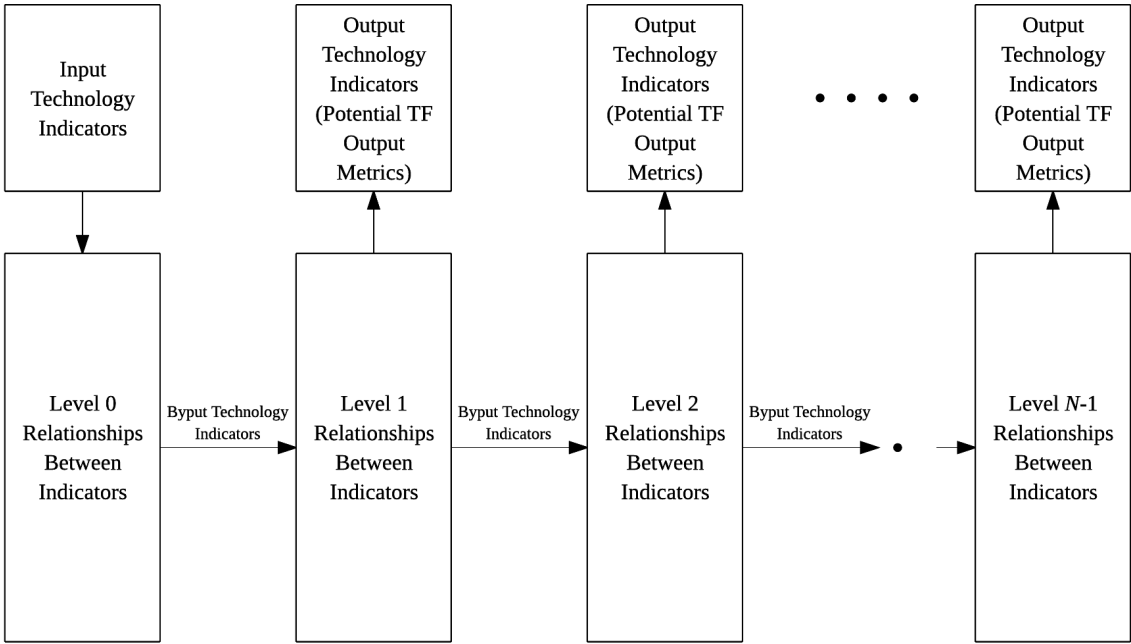


Fig. 3: Relationships between Input, Byput and Output Technology Indicators

A wide variety of sources exist that can be used to harvest technology indicators, ranging from patent databases and scientific publications [22], through to the rumor mill and financial market indicators [4]. In monitoring and mining these potential sources of technical indicators, bibliometrics have

emerged as one of the most popular set of quantitative techniques [4]. Bibliometrics uses counts of citations, publications or patents to produce indicators of technological progress in a specific technology domain [4].

Various frameworks have been proposed for the systematic categorization of technology indicators. In [4] Nyberg and Palmgren presents a succinct summary of the frameworks proposed by Watts and Porter [23], Grupp [24] and Chang [26], which is repeated here: The Watts and Porter [23] framework consists of the following three categories:

- *Technology Life Cycle Status Indicators*: Based primarily on the S-curve, these metrics determine the level of progress of a technological development along its life cycle, as well as the growth rate of the technology [4].
- *Innovation Context Receptivity Indicators*: These indicators gauge the sufficiency of supporting technology, as well as the development of standards and regulations surrounding the technology under investigation [4].
- *Market Prospect Indicators*: The potential commercial payoffs of the technology are considered by this type of indicator. Of specific importance with these indicators are factors such as technology application areas, intellectual property and market competitiveness [4].

Although Grupp was the instigator of the general function based classification of technology indicators into input, byput and output indicators [24], he originally referred to these three types of indicators based on the stage in the technology's life cycle at which the measurement was performed:

- *Resource Indicators*: This input indicator type measures the various possible expenditures on research, development and innovation [4].

- *Research and Development (R&D) Results Indicators*: This is the output indicator type, which measures qualitative, quantitative, or value rated advances in production processes or products [4].
- *Progress Indicators*: Indicators of this type, for example the technometric indicator [24] that measures the number of features or product specifications, are byput metrics that measures sub-phenomena of the technological progress [4].

The Technology Indicator Ontology (TIO) proposed by Chang [26] divides technology indicators into the following two broad groupings, each with a number of sub-groups:

- *Technology Development Indicators*: This broad grouping includes measures that track the development, change, progress and trend of a technology from a technological perspective [4].
- *Market Development Indicators*: This broad grouping includes all indicators related to the market development and potential application areas of the technology, including sales, investment and industrial applications [4].

The proposed framework for SEM based DF for TF allows for the use of any of the above stated types of technology indicators as latent or formative indicators for endogenous and exogenous constructs in the model. More specifically, input technology indicators are used with exogenous constructs. Conversely, bypass and output indicators are used for endogenous constructs. The TF output metrics, which will be used by decision makers to drive strategic action, consist of output metrics related to endogenous constructs in the SEM model. External environment related indicators contributing to exogenous constructs that realize context sensitivity in the DF process could also include technology indicators. For example, Sohn and Moon [16] used the Technology Commercialization Success Index (TCSI) metric, which is an example of a market prospect indicator in the framework proposed by Watts and Porter [23], as the primary TF output metric for their SEM model.

2.3 Proposed Framework for SEM Based DF for TF

Sohn and Moon showed in [16] that SEM can be used as an effective regression technique to evaluate a multi-layered hierarchal model through progressive aggregations and refinements of input technology indicator data in order to produce a reliable statistical estimate of the TCSI TF output metric. By extending Soon and Moon's [16] use of SEM for TF through Steinberg's use of SEM to implement context sensitive DF [12], this paper proposes the framework depicted in Fig. 4 for SEM based context sensitive DF for TF. The framework was developed through inductive reasoning by overlaying Fig. 3's relationship framework for technology indicators, as defined by Grupp [24], on the general context sensitive DF framework of Fig. 2 and applying the SEM construct grouping and layering framework of Fig. 1.

In this framework multi-layered aggregation and refinement of technology and context related information is accomplished by the processing performed at DF Levels 0 through $N-1$, where N is user selected. The number of levels N will be determined not only by the complexity of the technology domain under consideration, but also by time and cost constraints. Also, potential diminishing returns resulting from additional levels of aggregation and refinement will also be determining factors in defining N .

Input technology indicators [4][24] and context related indicators [12] are used as inputs to technology related endogenous constructs and context related exogenous constructs, respectively. Note that the use of bi-directional interconnections between indicators and constructs, as well as between multiple constructs, is based on the SEM path diagram conventions defined in [25]. This illustrates that positive or negative correlation can exist between constructs, as well as the fact that indicators can be either reflective or formative in nature.

To gain insight into the functioning of this framework, consider the aggregation and refinement that occur in progressing from DF Level 0 to DF Level 1: Regression analysis outputs generated for the

technology related exogenous constructs at DF Level 0 contribute formatively or reflectively to technology related endogenous constructs at DF Level 1. Regression analysis outputs for the context related exogenous constructs of DF Level 0 contribute to context related exogenous and technology related endogenous constructs at DF Level 1. The regression analysis results produced at DF Level 1 for context related exogenous constructs can also contribute to technology related endogenous constructs at this same level. Technology indicators for the technology related constructs at DF Level 1 could potentially be selected as the TF output metrics, or could simply be byput technology [4][24] indicators if additional DF levels are required for further aggregation and refinement. The aggregation and refinement achieved by moving from DF Level $x-1$ to DF Level x , for $x = 1, 2, 3, \dots, N-1$, follows a similar interconnection structure as the progression from DF Level 0 to DF Level 1, with the exception that now constructs at DF Level $x-1$ contribute to constructs at DF Level x .

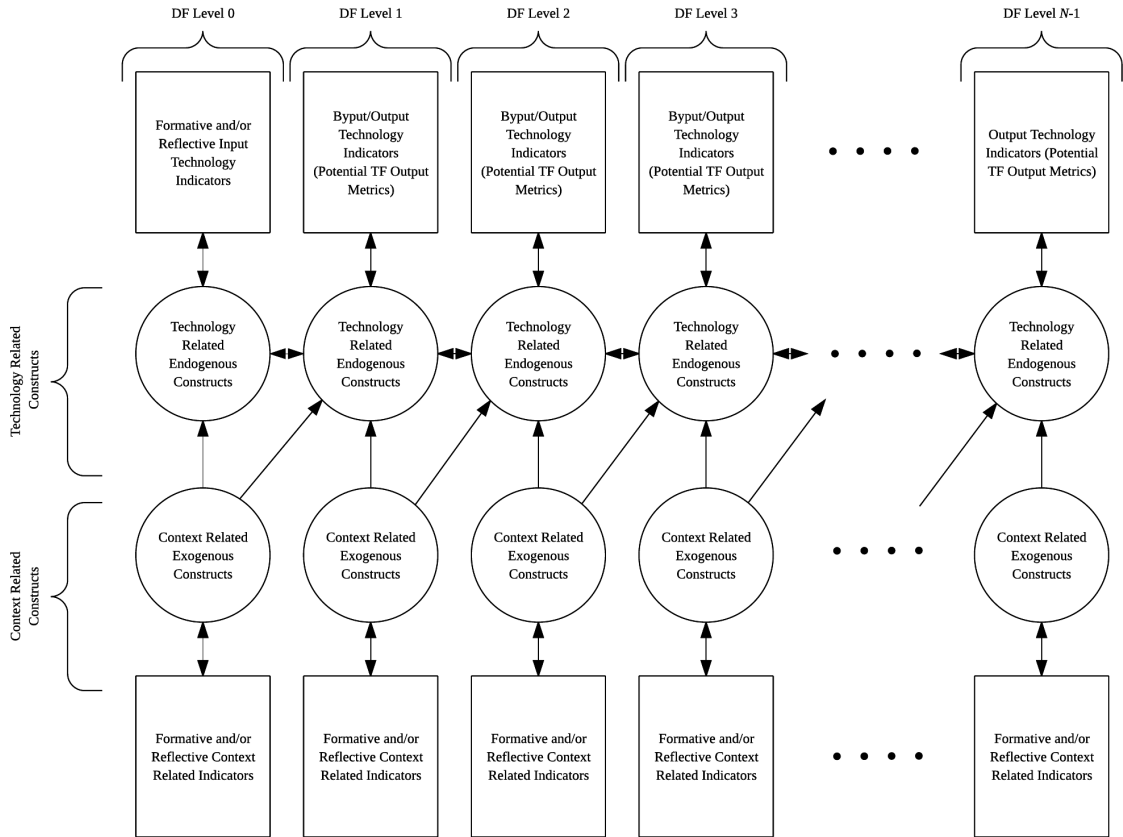


Fig. 4: Proposed Framework for SEM Based DF for TF

2.4 Construction and Utilization of a Model Instantiation of the Framework

Constructing a model instantiation of Fig. 4's proposed framework for SEM based DF for TF (referred to as SEM model building in the remainder of this study), involves firstly defining an appropriate set of technology related endogenous constructs and context related exogenous constructs, with relevant technology and context related measurement indicators, respectively, which reflects the fundamental characteristics of the technology domain under investigation. This is then followed by defining a set of hypothesized relationships between these constructs, emanating from theory, action research, colloquial knowledge or speculation. The last stage in the SEM model building exercise involves using empirical data, captured for the each of the technology and context related measurement indicators, in a PLS regression analysis [37] in order to determine the significance (including indicator loadings) of these measurement indicators, as well as the significance (including the path coefficients) of the defined set of hypothesized relationships.

Exploratory research efforts that make use of SEM frequently terminate at this final stage of SEM model building, as the research objectives for such studies typically involve testing the significance of the hypothesized relationships in the model instantiation [37]. In the case of the proposed framework for SEM based DF for TF, such hypothesis testing could prove especially useful in determining the impact of changing contextual factors, such as technology policy decisions, on the technology related endogenous constructs defined for the technology domain under investigation. As such, the proposed framework effectively provides one with a capability to forecast the relational influence between technology related endogenous constructs and context related exogenous constructs.

A model instantiation of the proposed framework, constructed using the SEM model building process described above, can also be used to forecast TF output metrics through a process best referred to as SEM post-processing [37]. This process involves populating the structural equations defined by the SEM model instantiation with the known context and input/output technology indicator data from a single metric measurement snapshot and solving these equations to determine the unknown TF output metrics by calculating their associated output technology indicator values.

3. Constructing an Example NREN Model Instantiation of the Framework

NREN's are frequently used as incubators for the development of new networking technologies and services [18]. Hence, NREN's contribute significantly to the creation of new Internet based business ventures, innovative business models and game changers in the way society works and plays. For

example, Facebook and Google have their roots within the NREN environments of Harvard University and Stanford University, respectively.

TERENA, which now forms part of GÉANT, was an was a not-for-profit association of European NRENs with the objective to provide a platform for NREN's to collaborate and openly share knowledge on networking technologies, services and infrastructure. TERENA performs an extensive yearly survey amongst the global NREN community in order to determine current technology and services trends. The results and interpretation of these surveys are then openly published as part of TERENA's NREN compendium series.

The NREN model instantiation example detailed in the following subsections, constructed using the model building process described in Section 2.4, was created using insights captured in TERENA's NREN compendium for 2012 [18], knowledge gained through action research [28] performed by the authors during their involvement with the management and operations of SANReN [19], as well as hypotheses presented and tested in peer-reviewed literature. It is an improved version of the example NREN model instantiation originally proposed and analyzed in [27]. Note that the use of this example NREN model instantiation to performed TF in this study, presented as part of the analysis results discussed in Section 4, was limited to forecasting the relational influence between technology related endogenous constructs and context related exogenous constructs. Forecasting of TF output metrics using SEM post-processing of the example NREN model instantiation will be considered during future research.

3.1 Overview of the Example NREN Model Instantiation

Fig. 5 presents an example NREN technology domain model instantiation of Section 2's framework employing the SEM path diagram conventions defined in [25]. This example model instantiation employs $N=3$ DF levels. Level 0, Level 1 and Level 2 focus on NREN Connectivity (i.e. the NREN provided infrastructure to deliver advanced services), NREN Services (i.e. the portfolio of advanced

services provided to users in order to make use of the NREN provided infrastructure) and NREN Utilization (i.e. a measure use of the NREN provided through the advanced services available to users), respectively.

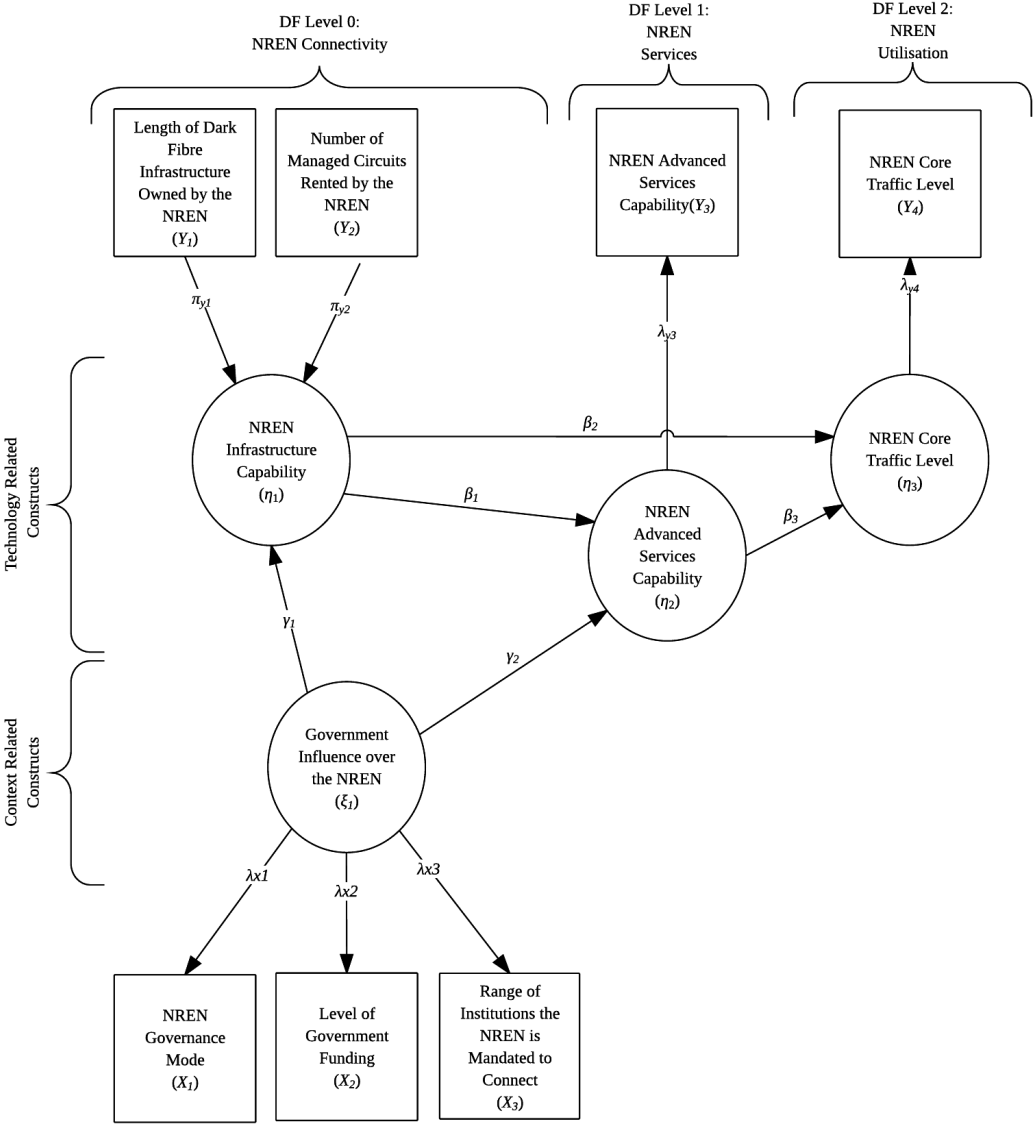


Fig. 5: Example Model Instantiation for the NREN Technology Domain

In essence the NREN Connectivity level is an aggregation of Layer 1 (Physical) through to Layer 6 (Presentation layer) in the 7-layered Open Systems Interconnection (OSI) model [29], while the NREN Services level represents Layer 7 (Application layer). The 7-layered OSI model has been

unofficially extended through the addition of Layers 8 to 10, representing Human-Computer Interaction (HCI) related aspects [30]. NREN Utilization is one possible representation of these HCI related layers.

At Level 0 of the example NREN model instantiation, which focuses on infrastructure related technology metrics, a single technology related endogenous construct, namely *NREN Infrastructure Capability* (η_1), is defined. The purpose of this construct is to model the extent to which the NREN has invested in dark fibre infrastructure and managed circuits [18][20][31]. Dark fibre is defined fibre infrastructure that is either owned outright by the NREN, or where the NREN has secured a long-term Indefeasible Right of Use (IRU) for the use of fibre [18][20]. Managed circuits are fibre infrastructure owned by another party and leased by the NREN [18][20]. Based on [31], in the example NREN model instantiation it is postulated that the *NREN Infrastructure Capability* (η_1) construct will be related to two formative input technology indicators (i.e. both indicators jointly represent the construct) that measure the length of available dark fibre infrastructure (denoted as *Length of Dark Fiber Infrastructure Owned by the NREN* (Y_1) with indicator loading π_{y1}) and the number of rented managed circuits [18][20] (denoted as *Number of Managed Circuits Rented by the NREN* (Y_2) with indicator loading π_{y2}), respectively.

Also defined at Level 0 is a single context related exogenous construct entitled *Government Influence over the NREN* (ζ_1). This construct has three reflective indicators (i.e. each indicator is capable of individually representing the construct) that measure the NREN governance mode (denoted as *NREN Governance Mode* (X_1) with indicator loading λ_{x1}), level of government funding provided to the NREN (denoted as *Level of Government Funding* (X_2) with indicator loading λ_{x2}) and the range of institutions the NREN is mandated to connect (denoted as *Range of Institutions the NREN is Mandated to Connect* (X_3) with indicator loading λ_{x2}), respectively. NREN governance mode can range from full government driven governance through to no government driven governance [18][25]. The range of institutions that the NREN is mandated to connect can vary from only type of institutions, such as universities, to

a suite of various types of institutions, such as research organizations, universities, schools, etc. [18]. A positive relation between *Government Influence over the NREN* (ξ_1) and *NREN Infrastructure Capacity* (η_1) is postulated and represented by the path coefficient γ_1 . This relation was derived from the notion that government intervention is required at various points in the NREN value chain, such as infrastructure funding, policy definition, regulation, etc. in order to ensure that an NREN successful matures in terms of the connectivity and advanced services that it provides [32][33].

It is important to note that additional context related measurement indicators and constructs from the political, economic, sociological, legal and environmental domains can be added to a model instantiation such as this example NREN model instantiation in order to potential improve the model's ability to forecast output technology metrics. However, given that this example NREN model instantiation was tested using the data available from the 2011 TERENA NREN compendium, the context related measurement indicators were limited to those associated with the *Government Influence over the NREN* (ξ_1) construct.

Level 1 of the example NREN model instantiation, which focuses on services related technology metrics, defines a single exogenous technology related construct entitled *NREN Advanced Services Capability* (η_2). This construct embodies the NREN's capability to provide a suite of advanced NREN services [18][20], such as Authentication and Authorization Infrastructure (AAI) services, provisioning of Identity Federation Services, hosting of Identity Federation Services and inter-federating with other NRENs [18]. The construct has a single reflective byput technology metric as reflective indicator, measuring the size of the portfolio of advanced services offered and hosted by the NREN (denoted as *NREN Advanced Services Capability* (Y_3) with indicator loading λ_{y_3}). A postulated positive relationship between *NREN Infrastructure Capacity* (η_1) and *NREN Advanced Services Capability* (η_2) is represented by the path coefficient β_1 . This relationship emanates from the postulation in [33] that an NREN requires an advanced infrastructure capability in order to be able to deliver a portfolio of advanced services.

While no exogenous context related construct is defined for data fusion Level 1, it is postulated that the Level 0's *Government Influence over the NREN* (ξ_1) is positively related to the NREN's ability to deliver advanced services [18]. This relationship is captured in the SEM model of Fig. 5 by means of path coefficients γ_2 . The postulated relationship is based on the reasoning in [32][33] that government intervention is required at various points in the NREN value chain in order to ensure that an NREN successful matures in terms of the advanced services portfolio that it provides.

Level 2 in the example model instantiation focuses on the utilization of the NREN, which is frequently used as a proxy to measure the impact that an NREN creates in its beneficiary communities [18][19], as well as the Return of Investment (ROI) of the funders of the NREN [34]. A single context related exogenous construct, entitled *NREN Core Traffic Level* (η_3), which represents the bandwidth usage in the core network of the NREN, is used to represent the utilization of the NREN [18][20]. This construct is directly measure by means of the reflective measurement indicator *NREN Core Traffic Level* (Y_4), with indicator loading λ_{y4} . This measurement indicator is also the technology forecasting output metric for the example NREN model instantiation.

Postulated positive relationships between *NREN Infrastructure Capability* (η_1) and *NREN Core Traffic Level* (Y_4), as well as between *NREN Advanced Services Capability* (η_2) and *NREN Core Traffic Level* (Y_4), are represented by path coefficients β_2 and β_3 , respectively. The positive relationship between infrastructure capability and network utilization (i.e. core network traffic level) is supported in [31]. In [32][33] it is postulated that the maturity of the advanced service portfolio is a driver in the utilization of broadband networks, thereby justifying the positive relationship between the advanced services capability and network utilization.

3.2 Research Propositions for the Example NREN Model Instantiation

The postulated relationships between constructs in Fig. 5's example NREN model instantiation give rise to the set of research propositions below, which are evaluated in Section 4.3. These research propositions' association with the various paths defined in the example NREN model instantiation is detailed in Fig. 5, as well as Table 3.

- **Research Proposition H1:** The NREN's infrastructure capability is positively related to the level of government influence over the NREN. This hypothesis stems from notion that government influence is required in order to ensure that an NREN is successful in maturing its infrastructure capability [32][33].
- **Research Proposition H2:** The advanced services capability of the NREN is positively related to the NREN's infrastructure capability. This hypothesis is supported by the postulation in [33] that an NREN requires an advanced infrastructure capability in order to be able to deliver a portfolio of advanced services.
- **Research Proposition H3:** The advanced services capability of the NREN is positively related to the level of government influence over the NREN. This hypothesis stems from notion that government influence is required in order to ensure that an NREN is successful in maturing its advanced services portfolio [32][33].
- **Research Proposition H4:** The level of core network traffic in the NREN is positively related to the infrastructure capability of the NREN, as postulated in [31].
- **Research Proposition H5:** The level of core network traffic in the NREN is positively related to the advanced services capability of the NREN, as postulated in [32][33].

4. Analysis Results for the Example NREN Model Instantiation

Secondary data from TERENA's NREN compendiums for 2011 [20] was used to determine Fig. 5's indicator loadings and path coefficients through PLS regression analysis. Table 1 below summarizes the composition of the NREN model instantiation indicator data using the secondary data extracted from the 2011 TERENA NREN compendium [20]. A total of 61 NRENs responded to TERENA's

survey to collect data for this. The original survey distributed by TERENA to NRENs is available from [20].

Table 1: Technology and Context Related Indicator Data Composition

| <i>Technology or Context Related Indicator</i> | <i>Indicator Composition</i> |
|---|---|
| <i>NREN Governance Mode (X_1)</i> | <p>Extracted from the online profiles of the respondent NRENs of the 2011 compendium [20] using following the scaling:</p> <ul style="list-style-type: none"> • The NREN is a government agency or part of a ministry = 3 • Government appoints at least half of the NREN's governing body = 2 • Indirect relationship between the NREN and government = 1 • No formal relationship between the NREN and government = 0 |
| <i>Level of Government Funding (X_2)</i> | Level of government funding (as a percentage of total funding) received by respondent NRENs, as summarized in Graphs 6.4.2 and 6.4.3 in [20] |
| <i>Range of Institutions the NREN is Mandated to Connect (X_3)</i> | Sum of the institution types supported by respondent NRENs, as shown in in Table 2.2.1 of [20] |
| <i>Length of Dark Fiber Infrastructure Owned by the NREN (Y_1)</i> | Total length of dark fiber [in kilometers] owned by respondent NRENs as summarized in Table 3.6.3 of [20] |
| <i>Number of Managed Circuits Rented by the NREN (Y_2)</i> | Total number of managed circuits rented by respondent NRENs as summarized in Table 3.3.2 of [20] |
| <i>NREN Advanced Services Capability (Y_3)</i> | <p>Total number of positive answers to the following questions in Table 5.3.1.1 in [20]:</p> <ul style="list-style-type: none"> • Does the NREN provide of Authentication and Authorization Infrastructure (AAI) services? |

| | |
|--|---|
| | <ul style="list-style-type: none"> • Does the NREN provide Identity Federation services? • Does the NREN operate the Identity Federation services? • Does the NREN's Identity Federation services inter-federate with those provided by other NRENs? |
| <i>NREN Core Traffic Level (Y₄)</i> | Annual level (measured in terabytes per year) of traffic sent on to the backbone networks of respondent NRENs, as measured by T ₁ +T ₄ in Graphs 4.2.1 and 4.2.2 in [20] |

In this study the SmartPLS [36] freeware software package was employed to realize the example NREN model instantiation of Fig. 5 and calculate all loadings and path coefficients through PLS regression. SmartPLS was configured to normalize all indicator data, as a variety of scaling approaches and ranges was used by TERENA in collecting the original data. SmartPLS was also used to evaluate the reliability and validity test criteria defined in [17] with the results discussed in Section 4.2. Note that only 28 NRENs provided all of the survey inputs in order to calculate the indicator inputs according to Table 1. Hence missing data was flagged and SmartPLS configured to use a mean replacement algorithm to compensate for this [36].

4.1 Measurement Indicator and Path Coefficient Results

The reporting of the PLS regression results for the example NREN model instantiation, presented in the following subsections, was based on the reporting standard defined by Vinzi, Chin, Henseler and Wang [37]. According to this reporting standard, the PLS regression results for the measurement portion of the SEM path diagram, consisting of the loadings for all of the measurement indicators in the model, are reported first, followed by the PLS regression results for the structural portion of the SEM path diagram, consisting of the path coefficients for all interrelationships between constructs.

4.1.1 Measurement Portion SEM Regression Results

The indicator loadings for the measurement portion of the example NREN model instantiation, determined using SmartPLS [36], are listed in Table 2. Although these loadings were not used directly

in order to evaluate the research propositions stated in Section 3.2, a detailed investigation thereof was crucial in order to determine those reflective indicators that did not comply with the minimum Indicator Reliability level of 0.4 (see Section 4.2.1).

Table 2: Measurement Portion Indicator Loading Results

| <i>Constructs</i> | <i>Type</i> | <i>Measurement Indicators</i> | <i>Loadings</i> |
|--|-------------|---|------------------------|
| <i>Government Influence over the NREN (ξ_1)</i> | Reflective | <i>NREN Governance Mode (X_1)</i> | $\lambda_{x1} = 0.892$ |
| | Reflective | <i>Level of Government Funding (X_2)</i> | $\lambda_{x2} = 0.805$ |
| | Reflective | <i>Range of Institutions the NREN is Mandated to Connect (X_3)</i> | $\lambda_{x3} = 0.854$ |
| <i>NREN Infrastructure Capability (η_1)</i> | Formative | <i>Length of Dark Fiber Infrastructure Owned by the NREN (Y_1)</i> | $\pi_{y1} = 0.473$ |
| | Formative | <i>Number of Managed Circuits Rented by the NREN (Y_2)</i> | $\pi_{y2} = 0.737$ |
| <i>NREN Advanced Services Capability (η_2)</i> | Reflective | <i>NREN Advanced Services Capability (Y_3)</i> | $\lambda_{y3} = 1.0$ |
| <i>NREN Core Traffic Level (η_3)</i> | Reflective | <i>NREN Core Traffic Level (Y_5)</i> | $\lambda_{y4} = 1.0$ |

4.1.2 Structural Portion SEM Regression Results

The path coefficients for the structural portion of the example NREN model instantiation, which were determined using SmartPLS [36], are listed in Table 3. Significance testing for these path coefficients, based on asymptotic *t*-statistics, is presented in Section 4.2.2. These path coefficients and their associated significance test results were used in Section 4.3 to evaluate the research propositions listed in Section 3.2.

Table 3: Structural Portion Path Coefficient Results

| <i>Research Proposition: SEM Path</i> | <i>Coeff.</i> |
|--|--------------------|
| H1: <i>Government Influence over the NREN (ξ_1) → NREN Infrastructure Capability (η_1)</i> | $\gamma_1 = 0.599$ |
| H2: <i>NREN Infrastructure Capability (η_1) → NREN Advanced Services Capability (η_2)</i> | $\beta_1 = 0.016$ |
| H3: <i>Government Influence over the NREN (ξ_1) → NREN Advanced Services Capability (η_2)</i> | $\gamma_2 = 0.855$ |
| H4: <i>NREN Infrastructure Capability (η_1) → NREN Core Traffic Level (η_3)</i> | $\beta_2 = 0.289$ |
| H5: <i>NREN Advanced Services Capability (η_2) → NREN Core Traffic Level (η_3)</i> | $\beta_3 = 0.187$ |

4.2 Reliability and Validity Analysis for the Example NREN Model Instantiation

Similar to the reporting standard for SEM indicator loading and path coefficient results Vinzi, Chin, Henseler and Wang [37] suggest that the reporting of reliability and validity test results first considers the measurement portion, which include Indicator Reliability, Construct Reliability and Convergent Validity [17]. This is then followed by the structural portion, which includes Coefficients of Determination, Path Coefficient Significance and Predictive Validity [17]. The reasoning behind this approach is that a lack in confidence in the accuracy and representivity of the measurement indicators in a model instantiation negates the need to test the reliability and validity of the structural portion [37].

4.2.1 Measurement Portion Reliability and Validity Results

This subsection details the reliability and validity test results, determined using SmartPLS [36], for the measurement portion of the SEM for the example NREN model instantiation, based on the metrics defined in [37] and detailed in [25]. Table 4 presents the Indicator Reliability judgment [25], Construct Reliability [25] and Convergent Validity [25] test results.

Table 4: Indicator Reliability, Construct Reliability and Convergent Validity Test Results

| <i>Constructs</i> | <i>Measurement Indicators</i> | <i>Indicator Reliability Judgment</i> | <i>Construct Reliability</i> | | <i>Convergent Validity</i> |
|--|---|---------------------------------------|--|------------------------------|----------------------------|
| | | | <i>Cronbach's Alpha</i> | <i>Composite Reliability</i> | |
| <i>Government Influence over the NREN (ζ_1)</i> | <i>NREN Governance Mode (X_1)</i> | Included | 0.810 | $\rho_{\zeta,1} = 0.887$ | $AVE_{\zeta,1} = 0.724$ |
| | <i>Level of Government Funding (X_2)</i> | Included | | | |
| | <i>Range of Institutions the NREN is Mandated to Connect (X_3)</i> | Included | | | |
| <i>NREN Infrastructure Capability (η_1)</i> | <i>Length of Dark Fiber Infrastructure Owned by the NREN (Y_1)</i> | Included | Tests not applicable: This construct has formative indicators [25] | | |
| | <i>Number of Managed Circuits Rented by the NREN (Y_2)</i> | Included | | | |
| <i>NREN Advanced Services Capability (η_2)</i> | <i>NREN Advanced Services Capability (Y_3)</i> | Included | Tests not applicable: This construct is directly observable [25] | | |
| <i>NREN Core Traffic Level (η_3)</i> | <i>NREN Core Traffic Level (Y_4)</i> | Included | Tests not applicable: This construct is directly observable [25] | | |

The Indicator Reliability test [25], which gives an indication of the level of variance in the measurement indicator that can be explained by its associated latent construct [25], revealed that none

of the reflective indicators exhibited a loadings less than 0.4 during a first-run PLS regression SEM analysis. As a result, no reflective indicators had to be removed and all subsequent SEM analyses could be performed on the model instantiation as is. All formative indicators were retained regardless of their loadings, since the concept of Indicator Reliability is not applicable to formative indicators. This is because these indicators can exhibit low correlation with their associated latent constructs, but still contribute significantly to their overall variance [37].

Construct Reliability [25], which considers whether the set of reflective indicators associated with a latent construct jointly measure it adequately [37], employs both the classic Cronbach's Alpha metric [37] and the more contemporary Composite Reliability measure [38]. This study's final judgment on the adequacy of a set of reflective indicators to measure their related latent construct was based on the requirement that the Composite Reliability measure needs to exceed a minimum level of 0.6 [37]. As is clear from Table 4 the only latent construct with reflective indicators present in the model was *Government Influence over the NREN (ξ_1)*, which complied with this requirement for Composite Reliability [38].

Convergent Validity considers the correlation between responses obtained by maximally different methods of measuring the same construct [37]. It is determined through the Average Variance Extracted (AVE) metric [25][38], which measures the variance of each latent construct's reflective indicators (as captured by the construct itself) relative to the total measured variance. Measured against the study's elected threshold value of 0.5 for this metric [25], it can be concluded from Table 4's results that the reflective indicators of the only latent construct *Government Influence over the NREN (ξ_1)* exhibited a sufficient AVE level, indicating that for this construct the majority of the total variance measured was due to indicator variance and not due to measurement error.

Discriminant Validity for the measurement portion of a SEM model considers the level of dissimilarity in the measurements obtained by the measurement tool for different constructs [37]. A necessary

condition to achieve Discriminant Validity requires that the shared variance between a latent construct and its indicators (determined by taking the square root of its AVE) exceeds the shared variance between this latent construct and any other latent constructs. Since the example NREN model instantiation only had one latent construct with reflective indicators, namely *Government Influence over the NREN* (ξ_1), this test was unnecessary.

4.2.2 Structural Portion Reliability and Validity Results

The results for the reliability and validity tests for the structural portion of example NREN model instantiation, based on the metrics defined in [37] and detailed in [25], are presented in this subsection. Table 5 details the Path Coefficient significance test results, while Table 6 considers the Coefficients of Determination and Predictive Validity test results. These results were obtained using SmartPLS [36].

Table 5: Path Coefficient Significance Test Results

| <i>Research Proposition: SEM Path</i> | <i>Asymptotic t-Statistic</i> | <i>Calculated p-Value</i> | <i>Significance Judgment for $\alpha = 0.10$</i> |
|--|--|--------------------------------------|--|
| H1: <i>Government Influence over the NREN</i> (ξ_1) → <i>NREN Infrastructure Capability</i> (η_1) | 3.952 | < 0.001 | Yes |
| H2: <i>NREN Infrastructure Capability</i> (η_1) → <i>NREN Advanced Services Capability</i> (η_2) | 0.124 | 0.901 | No |
| H3: <i>Government Influence over the NREN</i> (ξ_1) → <i>NREN Advanced Services Capability</i> (η_2) | 4.840 | < 0.001 | Yes |
| H4: <i>NREN Infrastructure Capability</i> (η_1) → <i>NREN Core Traffic Level</i> (η_3) | 2.015 | 0.044 | Yes |
| H5: <i>NREN Advanced Services Capability</i> (η_1) → <i>NREN Core Traffic Level</i> (η_3) | 1.662 | 0.097 | Yes |

As with covariance based multiple regression techniques, the quality of the structural portion of a model instantiation can be investigated by means of a bootstrapping procedure [37] in order to determine the significance levels of the path coefficients. The significance of path coefficients (also sometimes referred to as Goodness-of-Fit) can be tested via asymptotic t -statistics. From Table 5's Path Coefficient Significance test results, obtained using SmartPLS's bootstrapping function [36] configured to generate 500 sets of subsamples from the 61 cases in the original sample, the only path that exhibited a p -value larger than the maximum acceptable significance level of $\alpha = 0.10$ was *NREN Infrastructure Capability* (η_1) \rightarrow *NREN Advanced Services Capability* (η_2). Hence, this path was deemed insignificant.

Table 6: Coefficient of Determination and Predictive Validity Test Results

| <i>Technology or Context Related Indicator</i> | <i>Coefficients of Determination (R^2)</i> | <i>Predictive Validity (Q^2)</i> | |
|--|---|---|--|
| | | <i>Cross-validated Communalities (H^2)</i> | <i>Cross-validated Redundancies (F^2)</i> |
| <i>Government Influence of the NREN (ζ_1)</i> | Test not applicable: Exogenous variable | 0.440 | Test not applicable: Exogenous variable |
| <i>NREN Infrastructure Capability (η_1)</i> | 0.359 | 0.054 | 0.191 |
| <i>NREN Advanced Services Capability (η_2)</i> | 0.749 | 1.0 | 0.727 |
| <i>NREN Core Traffic Level (η_3)</i> | 0.175 | 1.0 | 0.142 |

The Coefficients of Determination (R^2) reflect the share of an endogenous construct's variance explained by related endogenous or exogenous constructs [37]. The test results given in Table 6

revealed that the interrelationships between the endogenous latent constructs and their related constructs produced explained variances exceeding the minimum level of 0.1 [25]. Also, interrelationships with the *NREN Advanced Services Capability* (η_2) endogenous latent construct were deemed to be strong, since the R^2 for this construct exceeded 0.7 [25]. Interrelationships with the *NREN Core Traffic Level* (η_3) endogenous latent construct were viewed as weak, since the R^2 for this construct was lower than 0.3 [25].

In order to determine the Predictive Validity of a model instantiation the Stone-Geisser (referred to as Q^2) non-parametric test is performed [25][37] based on a blindfolding procedure [39]. The model instantiation is considered to have Predictive Validity if $Q^2 > 0$ [37]. The Stone-Geisser test criterion can take on two distinct forms, depending on the type of prediction that is investigated: The first form, which is geared at determining the Predictive Validity of the measurement portion (although usually calculated during the structural portion's validity evaluation), is referred to as the Cross-validated Commuality [37] and is denoted by H^2 . Cross-validated Commuality measures the ability of the model instantiation to predict the observable endogenous constructs from their own latent construct scores [37]. The second form, which evaluates the Predictive Validity of the structural portion, is referred to as Cross-validated Redundancy [37]. This metric, denoted by F^2 , measures the model instantiation's ability to predict the observable endogenous constructs using latent constructs that predict the block of data in question [37].

A review of the Predictive Validity test results for the *NREN Core Traffic Level* (η_3) construct, directly observable via the output forecasting technology metric of interest *NREN Core Traffic Level* (Y_4), revealed that both the Cross-validated Commuality (H^2) and the Cross-validated Redundancy (F^2) tested positively. Hence, the both the example NREN model instantiation's measurement indicators and the defined structural relationships are well suited to forecasting the NREN's core network traffic level.

4.3 Evaluation and Discussion of the Framework Outputs for the Example NREN Model

Instantiation

Using the calculated path coefficients in Table 3 and the path coefficient significance test results in Table 5, the research propositions defined for the example NREN model instantiation in Section 3.2 were evaluated as follows:

- **Research Proposition H1:** The path coefficient of $\gamma_1 = 0.599$ supports the direction of the proposed relationship between *Government Influence over the NREN* (ξ_1) and *Infrastructure Capability* (η_1) *Government Influence of the NREN* (ξ_1). Furthermore, the path coefficient was judged to be significant at the maximum allowed significance level of $\alpha = 0.10$. Hence, this hypothesized relationship was not rejected and supports the notion in [32][33] that government influence is required in order to ensure that an NREN is successful in maturing its infrastructure capability. Moreover, the positive influence that the government has over the infrastructure capabilities of an NREN was to be expected, since most NRENs are government interventions geared at supporting a country's research and education communities by enhancing the available research and education infrastructure and services [32][33].
- **Research Proposition H2:** This hypothesized relationship between *NREN Infrastructure Capability* (η_1) and *NREN Advanced Services Capability* (η_2) was rejected. While the path coefficient $\beta_1 = 0.016$ supported the direction of the proposed relationship, the path coefficient was judged to be not significant at the maximum allowed significance level of $\alpha = 0.10$. This finding is counter to the assertions in [33] that an NREN requires an advanced infrastructure capability in order to mature its portfolio of advanced services. This can be explained by noting that, in the telecommunications industry providers in developing countries frequently leapfrog their more developed counterparts by offering advanced services, even though their infrastructure capability might still be nascent [35]. In the case of the NREN community, European NRENs frequently support fledgling NRENs in Africa and Asia to rapidly deploy

advanced services through development programs driven by the GÉANT Association [18][20].

- **Research Proposition H3:** The path coefficient $\gamma_2 = 0.855$ supported the direction of the hypothesized relationship between *Government Influence over the NREN* (ξ_1) and *NREN Advanced Services Capability* (η_2). Furthermore, the path coefficient was deemed significant. Hence, this research proposition was not rejected and supports the notion in [32][33] that government influence is required in order to ensure that an NREN is successful in maturing its advanced services portfolio. As with Research Proposition 1, this finding was to be expected, since most NRENs are government interventions geared at supporting a country's research and education communities by enhancing the available research and education infrastructure and services [32][33].
- **Research Proposition H4:** The postulated relationship between *NREN Infrastructure Capability* (η_1) and *NREN Core Traffic Level* (η_3) was not rejected since the path coefficient $\beta_2 = 0.289$ was judged to be significant at the maximum allowed significance level of $\alpha = 0.10$. Therefore, the postulated relationship in [31] that an NREN's infrastructure capability is positively related to its usage is supported. This finding correlates with the notion that enhancing NREN infrastructure will lead to improved usage [18][33].
- **Research Proposition H5:** This research proposition was not rejected, since the path coefficient $\beta_3 = 0.187$ support the direction of the hypothesized relationship between *NREN Advanced Services Capability* (η_2) and *NREN Core Traffic Level* (η_3). Also, this path was deemed significant at the maximum allowed significance level of $\alpha = 0.10$, thereby supporting the notion in [32][33] that an NREN's advanced services capability is positively related to the usage of the NREN. This finding correlates with the notion that, by providing the beneficiaries of an NREN with a portfolio of advanced services to fully utilize the infrastructure available, the usage of an NREN will increase [18][33].

In summary, the data available from the 2011 TERENA NREN compendium supported all of the postulated relationships in the example NREN model instantiation, with the exception of the postulation that an NREN's advanced services capability is related to its infrastructure capability. This latter result emanated from technology leapfrogging that is prevalent in the NREN community. Furthermore, both the selected input and byproduct selected metrics, as well as the context related metrics, contributed successfully to the forecasting of the output technology metric measuring the usage of the NREN. The same was also true of the structural relationships defined throughout the three DF levels of the example NREN model instantiation.

5. Conclusions

Applying inductive reasoning to the work of Sohn and Moon [16], as well as Steinberg [12][15], the paper derived a framework for SEM based DF for TF. Unlike most TF approaches, the proposed framework not only caters for complex and hierarchical structural relationships between technology indicators and TF output metrics, but also allows for non-linear and non-Gaussian factors and cyclical dependencies amongst model variables, which can be either latent or directly observable [16].

An example model instantiation of the proposed framework was presented for the NREN technology domain. This example model instantiation was created using knowledge gained through action research in SANReN [28], data captured by TERENA in its yearly NREN compendium series [18][20], as well as hypotheses presented in peer-reviewed literature. The example model instantiation, which consisted of 3 DF levels, suggested that an NREN's infrastructure capability (defined at DF level 0) and advanced services capability (defined at DF level 1) were both positively related to the government influence over the NREN, which was regarded as a contextual influence. Furthermore, the model postulated that an NREN's infrastructure capability was positively related to its advanced services capability and that both the infrastructure and services capabilities were positively related to the usage of the NREN (defined at DF level 2), which was chosen as the technology forecasting output metric of the example model instantiation.

Data from the 2011 TERENA compendium was also used to perform a PLS regression in order to determine the path coefficients and indicator loadings in the example NREN model instantiation. From the PLS regression results obtained for the example NREN model instantiation it can be concluded that all of the selected technology indicators were able to adequately measure their respective technology related model constructs. Furthermore, the TERENA data substantiated all hypothesized relationships in the example NREN model instantiation, except the postulation that an NREN's infrastructure capability is positively related to its advanced services capability. This last finding can be explained by the high prevalence of technology leapfrogging [35] in the global NREN community [18], since developing NRENs are rapidly implementing extensive advanced services portfolios (with the assistance of their more advanced NREN peers), while their infrastructure capability is still somewhat nascent. Lastly, a reliability and validity evaluation of the example NREN model instantiation using the 2011 TERENA compendium data highlighted that both the measurement and the structural portions of the model were capable of contributing adequately to the forecasting of the technology forecasting output metric that measured the usage of an NREN.

Future research activities that will be undertaken include an evaluation of the strengths and weaknesses of the proposed framework for SEM based DF for TF, comparing the proposed framework to various popular TF techniques currently receiving attention from the technology management research community, as well as improving the model instantiation for the NREN technology domain. Strengths that will be explored include the ability of the proposed framework for SEM based DF for TF to utilize contextual information in order to improve its forecasting capability. Weaknesses that will be evaluated include the potential negative impact that poorly defined structural configurations in SEM models have on their resultant Goodness-of-Fit [37].

Improving the model for the NREN technology domain will entail a two-phase process, with the first phase involving a qualitative study [39] that will attempt to identify improved endogenous and

exogenous model constructs, technology indicators, as well as interactions between the various indicators and constructs. The unit of analysis [39] for this envisioned qualitative phase will be a single NREN, while the population will be all NRENS in existence worldwide at the time of the study. Data collection will be accomplished through the Delphi method [29] using a panel of experts comprising the senior technical managers at leading NRENS from the global community. Analysis of the qualitative data collected through various rounds of engagements with the panel of experts will start with narrative inquiry by means of a process of theme extraction [25]. This will then be followed by performing a frequency analysis on the extracted themes in order to produce a final set of importance ranked indicators, constructs and interconnections from which the improved NREN model instantiation will be constructed [25]. Testing the reliability and validity of the collected qualitative data will be accomplished by means of theory triangulation [25], as well as data triangulation [25] using as baseline published technology indicators from secondary data sources, such as TERENA's NREN compendium series.

The second phase will be quantitative in nature and will aim to determine, using PLS regression analysis, the indicator loadings and path coefficients of the NREN model constructed during the qualitative first phase. As with the qualitative phase the population will be all NRENS in existence at that point in time, with the unit of analysis being a single NREN [39]. While the data available from the TERENA NREN compendium series will be used as far as possible to populated technology and context related measurement indicators, any additional qualitative data required will be obtained using an online survey consisting of close-ended questions with Likert scaling [39], targeted a sample of senior technical managers at the NRENS in the population, selected through a process of convenience sampling [39].

Acknowledgements

The Council for Scientific and Industrial Research, as well as the University of Pretoria supported this study through funding contributions. Dr. Pieter de Villiers at the University of Pretoria is recognized

for his contribution towards the development the proposed framework for SEM based DF for TF by assisting the authors in the effort apply SEM as a context sensitive DF analysis technique.

References

- [1] C.A. Mack, Fifty Years of Moore's Law, *IEEE Transactions on Semiconductor Manufacturing* 24 (2) (2011) 202-207.
- [2] J. Nielsen, Nielsen's Law of Internet Bandwidth, <http://www.nngroup.com/articles/niensens-law-of-internet-bandwidth/> (accessed 08/01/2013), April 1988.
- [3] B. Metcalfe, Metcalfe's Law: A Network Becomes More Valuable as it Reaches More Users, *Infoworld* 17 (40) (1995) 53-54.
- [4] A. Nyberg, S. Palmgren, Using Indicators for Technology Monitoring: Steps Toward a Proposed Framework, Chalmers University of Technology, Master of Science Thesis in the Master Degree Programme: Business Design, 2011.
- [5] A.L. Porter, Tech Mining to Drive Open Innovation, *Proceedings of the 1st International Conference on Technology Innovation, Risk Management and Supply Chain Management* (2007) 1-13.
- [6] E. Lichtenthaler, Technological Change and the Technology Intelligence Process: A Case Study, *Journal of Engineering and Technology Management* 21 (2004) 331-348.
- [7] H. Chang, J. Gausemeier, S. Ihmels, C. Wenzelmann, Innovative Technology Management System with Bibliometrics in the Context of Technology Intelligence, *Trends in Intelligent Systems and Computer Engineering: Lecture Notes in Electrical Engineering* 6 (2008) 349-361.
- [8] A.L. Porter, QTIP: Quick Technology Intelligence Processes, *Technological Forecasting and Social Change* 72 (2005) 1070–1081.
- [9] M. Buchroithner, Geodata Interrelations: Inventory and Structuring Attempt of Taxonomic Diversity, *Proceedings of the 2nd Conference on Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images* (1998) 11–15.

- [10] L. Wald, An Overview of Concepts in Fusion of Earth Data, Proceedings of the EARSeL Symposium 1997, Future Trends in Remote Sensing (1997) 385–390.
- [11] L. Wald, Some Terms of Reference in Data Fusion, IEEE Transactions on Geoscience and Remote Sensing 37(3) (May 1999).
- [12] A.N. Steinberg, Context-Sensitive Data Fusion using Structural Equation Modeling, Proceedings of the 12th IEEE International Conference on Information Fusion, July 2009.
- [13] M. Haenlein, A.M. Kaplan, A Beginner's Guide to Partial Least Squares Analysis, Understanding Statistics 3 (4) (2004) 283-297.
- [14] K.G. Jöreskog, A General Method for Estimating a Linear Structural Equation System, in Structural Equation Models in the Social Sciences, eds.: A.S. Goldberger and O. D. Duncan, New York: Seminar, 1973.
- [15] A.N. Steinberg, G. Rogova, Situation and Context in Data Fusion and Natural Language Understanding, Proceedings of the 11th IEEE International Conference on Information, 2008.
- [16] S.Y. Sohn, T.H. Moon, Structural Equation Model for Predicting Technology Commercialization Success Index (TCSI), Technological Forecasting and Social Change 70 (9) (2003) 885-899.
- [17] L. Staphorst, L. Pretorius, M.W. Pretorius, Structural Equation Modeling Based Data Fusion for Technology Forecasting: A Generic Framework, Proceedings of PICMET2013, July 2013.
- [18] TERENA COMPENDIUM of National Research and Education Networks in Europe: 2012 Edition, <https://www.terena.org/publications/files/TERENA%20Compendium%202012.pdf> (accessed 30/12/2013), 2012.
- [19] The South African National Research Network (SANReN), <http://www.sanren.ac.za/> (accessed 27/12/2013), 2013.
- [20] TERENA COMPENDIUM of National Research and Education Networks in Europe: 2011 Edition, <https://www.terena.org/publications/files/TERENA%20Compendium%202011.pdf> (accessed 30/12/2013), 2011.

- [21] H. Wold, Path Models with Latent Variables: The NIPALS Approach, in *Quantitative Sociology: International Perspectives on Mathematical and Statistical Modeling*, eds.: H.M. Blalock, A. Aganbegian, F.M. Borodkin, R. Boudon and V. Capecchi (eds.), New York: Academic Press (1975) 307–357.
- [22] A.L. Porter, S.W. Cunningham, *Tech Mining: Exploiting New Technologies for Competitive Advantage*, Wiley-Interscience 29 (2004).
- [23] R.J. Watts, A.L. Porter, *Innovation Forecasting, Technological Forecasting and Social Change* 56 (1) (1997) 25-47.
- [24] H. Grupp, *Foundations of the Economics of Innovation: Theory, Measurement and Practice*. Cheltenham: Edward Elgar, 1998.
- [25] L. Staphorst, *The Impact of Intellectual Property Rights from Publicly Financed Research and Development on Governance Mode Decisions for Research Alliances*, Gordon Institute of Business Science, Master of Business Administration Thesis, 2010.
- [26] H. Chang, *A Methodology for the Identification of Technology Indicators*, HNI-Verlagsschriftenreihe, 233 (2007).
- [27] L. Staphorst, L. Pretorius, M.W. Pretorius, *Structural Equation Modeling Based Data Fusion for Technology Forecasting: A National Research and Education Network Example*, Proceedings of PICMET2014, July 2014.
- [28] B. Gustavsen, *Action Research, Practical Challenges and the Formation of Theory*. *Action Research* 6 (4) (2008) 421-437.
- [29] H. Zimmerman, *OSI Reference Model - The ISO Model of Architecture for Open Systems Interconnection*, *IEEE Transactions on Communications* 28 (4) (1980) 425-432.
- [30] B. Bauer, A.S. Patrick, *A Human Factors Extension to the Seven-Layer OSI Reference Model*, <http://www.andrewpatrick.ca/OSI/10layer.html> (accessed 18/01/2016), 2004.
- [31] B. Savory, *Build vs. Buy: The NREN Network Connectivity Dilemma: About Building Research & Education Fibre Networks*, Proceedings and Report of the 5th UbuntuNet Alliance Annual Conference (2012) 160-173.

- [32] R. Janz, A. Kutanov, The Case of NRENs in Central Asia, *Central Asian Journal of Global Health* 1 (1) (2012).
- [33] D. Greaves, An NREN Capability Maturity Model, http://www.ubuntunet.net/sites/default/files/NREN_Capability_Maturity.pdf (accessed 18/01/2016), 2009.
- [34] M. Bech, Calculating the Return on Investment of an NREN, In *TERENA Networking Conference 2011*, 2011.
- [35] V.W. Mbarika, T.A. Byrd, J.E. Raymond, P. McMullen, Investments in Telecommunications Infrastructure are not the Panacea for Least Developed Countries Leapfrogging Growth of Teledensity, *International Journal on Media Management* 2 (3/4) (2000) 133–142.
- [36] C.M. Ringle, S. Wende, S. Will, SmartPLS 2.0 (M3) beta, <http://www.smartpls.de> (accessed 24/01/2014), 2005.
- [37] V.E. Vinzi, W.W. Chin, J. Henseler, H. Wang, *Handbook of Partial Least Squares: Concepts, Methods and Applications*. Berlin, Heidelberg: Springer, 2010.
- [38] C. Fornell, D.F. Larcker, Evaluating Structural Equation Models with Unobservable Variables and Measurement Errors, *Journal of Marketing Research* 18 (1981) 39–50.
- [39] W. Zikmund, *Business Research Methods*, 7th Edition. Ohio: South-Western Cengage Learning, 2003.