SACJ - volume 33 number 2 - December 2021

SACJ

DESUPER DEORSUM

South African Computer Journal

December 2021 - Volume 33 Number 2

South African Computer Journal

Volume 33 Number 2, December 2021
ISSN 1015-7999 (print)
ISSN 2313-7835 (online)

A publication of the South African Institute of Computer Scientists and Information Technologists
Open Access: https://sacj.cs.uct.ac.za

South African Computer Journal

NOTES FOR CONTRIBUTORS

The *South African Computer Journal* is an accredited specialist academic journal, publishing research articles, technical reports and communications in English in the Computer Science, Computer Systems and Information Systems domains. Its primary target is research of interest in Africa or published by African researchers, but all quality contributions are considered. All research articles submitted for publication are rigorously refereed by independent peer reviewers. The journal publishes original work that is of international stature. The editorial board comprises local and international scholars of high repute. The journal is published online using the *open access* model, making papers accessible in developing countries where funding to subscribe is scarce.

Submissions

Authors should submit papers for publication at the web site at http://sacj.cs.uct.ac.za/index.php/sacj/about/submissions. Please also check there for the latest version of the below guidelines.

Form of Manuscript

Manuscripts for *review* should be prepared according to the following guidelines, which summarize more detailed instructions on the web site.

SACJ has a double-blind reviewing policy. No author's name or affiliation should appear anywhere. Citing of previous work by the author or authors should be anonymised if appropriate. Acknowledgments and thanks should not be included in the draft for review. If you use Microsoft Word please make sure that your name and affiliation are not saved in the document properties.

- The article should start as follows:
  - the title (as brief as possible)
  - an abstract of less than 200 words
  - an appropriate keyword list
  - a list of ACM Categories (2012 scheme; as of 2016, we no longer use the 1998 scheme).
- Tables and figures should be numbered and titled, and referenced in the text by number.
- *SACJ* uses American Psychological Association 6th edition style as described here http://www.apastyle.org/. References should be listed at the end of the text in alphabetic order of the (first) author's surname and cited in the text.

If you use an appropriate BibTeX style, this will work automatically; do not spend a lot of time on reference citation minutiae since the production editor will take care of that sort of detail.

- if a DOI is available for any reference cited, include the DOI in the reference list (see DOI display guidelines here: http://www.crossref.org/02publishers/doi_display_guidelines.html).

SACJ is produced using the LaTeX document preparation system. Though we can also accept Microsoft Word submissions, delays in publication are more likely with the latter format. While we encourage submission in a format similar to the published paper, authors should not waste undue time on layout as this will be redone by the production editor.

Authors retain the right to republish their work, subject to any republished version including a reference to the *SACJ* version.

Publication Charges

A charge of R6000 will be levied on papers accepted for publication to cover costs of open access publication. Where the author's institution or research budget is unable to meet this charge, it may be waived upon request of the author and at the discretion of the editor-in-chief.

Proofs

Proofs of accepted papers will be sent to the corresponding author to ensure that typesetting is correct, and not for addition of new material or major amendments to the text. Corrected proofs should be returned to the production editor within three days.

Extended Conference Papers

Authors of conference papers are welcome to submit extended papers to *SACJ* for consideration on these terms:

- a covering letter accompanying submission should explain what is added to the paper to make it worth publishing as a journal paper
- the paper includes at least 30% new material
- a pointer is provided to the original paper or, if it is not freely available, upload it as supplementary material when submitting the extended paper to *SACJ*
- evidence is provided that republication in this form does not violate copyright of the conference publication

South African Computer Journal

Book Reviews, Letters and Communications

Book reviews are welcome, as are letters to the editor; both should carry the author's full name and affiliation, and should be limited to 500 words. Communications and Viewpoints up to two pages in length (or longer, by negotiation with the editor-in-chief) may also reflect minor research contributions, works-in-progress or ideas to stimulate debate.

This category of submission is accepted at the discretion of the editor-in-chief, not refereed and does not qualify as a research publication for South African government subsidy purposes. The major criteria for acceptance are that the item is coherently written and does not require significant editing, that it is timely and it is likely to be of interest to readers.

South African Computer Journal

Volume 33 Number 2, December 2021
ISSN 1015-7999 (print)
ISSN 2313-7835 (online)

A publication of the South African Institute of Computer Scientists and Information Technologists
Open Access: https://sacj.cs.uct.ac.za

CONTENTS

Editorial: Moving on

Philip Machanick sacj.editor@gmail.com
Department of Computer Science, Rhodes University, South Africa

Introduction

This is my last regular issue of *South African Computer Journal* before Katherine Malan takes over final editing of the July 2022 issue. I served two five-year terms as editor-in-chief and am pleased to hand the journal over as a going concern.

This issue marks two other changes. It is the last issue to be completed by production editor James Dibley and one of our Information Systems editors, Caroline Kene, will be stepping down.

In my previous editorial, I gave a retrospective of how *SACJ* has advanced since I took over; this time I want to reflect on how Covid-19 has made things difficult in academia, before listing contents of the current issue.

Covid-19 Retrospective

Perhaps it is premature to do a retrospective on Covid-19 as the Omicron variant is growing apace at time of writing. Nonetheless we have learnt from operating under a pandemic. I give some personal reflections that may or may not apply to others.

SACJ, despite all the disruptions of the Covid-19 pandemic, has brought out each issue on schedule. However, things have been slowed down: finding responsive reviewers has been more difficult than before and this issue is relatively thin.

In my own courses, 2020 was a particularly difficult year. One of my classes, a second-year computer architecture course, started right after South Africa's first lockdown. Much of the 4-week course was run with no certainty of access to computers and Internet. The course previously relied heavily on the practical component, mostly coding in MIPS assembly language, and I had to make major adjustments.

In 2021, I took a highly adaptive approach to the course. In retrospect, I invented a new approach to pedagogy that I now call *adaptive agile blended learning*. It draws on the core ideas of agile development and blended learning (not a novel idea (Tesar & Sieber, 2010)) but rather

Machanick, P. (2021). Editorial: Moving on [Editorial]. *South African Computer Journal 33*(2), vii–viii. 10.18489/sacj.v33i2.1042

SACJ is a publication of the South African Institute of Computer Scientists and Information Technologists. ISSN 1015-7999 (print) ISSN 2313-7835 (online).

than a fixed framework for applying these methods, I emphasize close attention to how a class is coping and rapid adjustments to the approach as problems arise – hence *adaptive*.

In a class survey, while there were some negatives, I had some of the most positive comments in any course survey I have run.

Another aspect of Covid-19 is countering disinformation. I have spent a lot of time on social media and writing for popular media – not usual academic activities but those with the knowledge to take part in public debate need to step up when they are needed. The extent of disinformation and anti-science contrarianism has not been seen since the height of AIDS denial.

In this issue

In this issue, we have three research papers:

- Mpangase *et al*: "nf-rnaSeqCount: A Nextflow pipeline for obtaining raw read counts from RNA-seq data"
- Muchenje and Botha: "Consumer-centric factors for the implementation of smart meters in South Africa"
- Pillay and Van der Merwe: "Big Data Driven Decision Making Model – a case of the South African Banking Sector"

We also provide a short communication which aims to clarify some best practices for preparing submissions to *SACJ*, with particular attention to preparing publication-ready reference lists.

References

Tesar, M., & Sieber, S. (2010). Managing blended learning scenarios by using agile e-learning development. *Proc. IADIS International Conference E-Learning*, *2*, 125–129.

SACJ 33(2) December 2021
Research Article

nf-rnaSeqCount: A Nextflow pipeline for obtaining raw read counts from RNA-seq data

Phelelani T. Mpangase[a, b], Jacqueline Frost[b, c], Mohammed Tikly[d], Michèle Ramsay[a, b], Scott Hazelhurst[a, e]

[a] Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg
[b] Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg
[c] Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York City
[d] Division of Rheumatology, School of Clinical Medicine, University of the Witwatersrand, Johannesburg
[e] School of Electrical & Information Engineering, University of the Witwatersrand, Johannesburg

ABSTRACT
The rate of raw sequence production through Next-Generation Sequencing (NGS) has been growing exponentially due to improved technology and reduced costs. This has enabled researchers to answer many biological questions through "multi-omics" data analyses. Even though such data promises new insights into how biological systems function and understanding disease mechanisms, computational analyses performed on such large datasets comes with its challenges and potential pitfalls. The aim of this study was to develop a robust portable and reproducible bioinformatic pipeline for the automation of RNA sequencing (RNA-seq) data analyses. Using Nextflow as a workflow management system and Singularity for application containerisation, the nf-rnaSeqCount pipeline was developed for mapping raw RNA-seq reads to a reference genome and quantifying abundance of identified genomic features for differential gene expression analyses. The pipeline provides a quick and efficient way to obtain a matrix of read counts that can be used with tools such as DESeq2 and edgeR for differential expression analysis. Robust and flexible bioinformatic and computational pipelines for RNA-seq data analysis, from QC to sequence alignment and comparative analyses, will reduce analysis time, and increase accuracy and reproducibility of findings to promote transcriptome research.

Keywords: bioinformatics, pipelines, workflows, nextflow, RNA-seq, singularity, container, reproducible

Categories: • **Applied computing ~ Bioinformatics**

Email:
Phelelani T. Mpangase Phelelani.Mpangase@wits.ac.za (CORRESPONDING),
Jacqueline Frost jacqueline.frost@mssm.edu,
Mohammed Tikly Mohammed.Tikly@wits.ac.za,
Michèle Ramsay Michele.Ramsay@wits.ac.za,
Scott Hazelhurst Scott.Hazelhurst@wits.ac.za

Article history:
Received: 26 May 2020
Accepted: 17 August 2021
Available online: 19 December 2021

Mpangase, P.T., Frost, J., Tikly, M., Ramsay, M. and Hazelhurst, S. (2021). nf-rnaSeqCount: A Nextflow pipeline for obtaining raw read counts from RNA-seq data. *South African Computer Journal 33*(2), 1–16. https://doi.org/10.18489/sacj.v33i2.830

SACJ is a publication of the South African Institute of Computer Scientists and Information Technologists. ISSN 1015-7999 (print) ISSN 2313-7835 (online).

1 INTRODUCTION

With the increase in the rate at which raw sequencing data is produced due to improved technologies and reduced cost of Next-Generation Sequencing (NGS), researchers in the field of bioinformatics and computational biology are able to perform "multi-omics" data analyses to answer many biological questions (Fan et al., 2014; Kluge & Friedel, 2018; Schulz et al., 2016). However, analysis of such large datasets comes with a number of challenges, especially when it comes to sharing data analysis methods with the scientific community and being able to reproduce consistent results using the same data across different computational platforms (Boettiger, 2015; Di Tommaso et al., 2017; Kurtzer et al., 2017). When performing computational analyses of NGS data, often different tools are required at each processing step of the analysis. For example, Haas et al. (2013) describe a procedure for assembling RNA sequencing (RNA-seq) data, quantifying expression levels for transcripts and identifying differentially expressed transcripts between samples. This protocol requires a number of applications in order to be executed successfully, including `Trinity`, `bowtie`, `samtools`, `R` and `NCBI-Blast+` (Haas et al., 2013).

To a bioinformaticist, computational biologist or someone who is familiar with the `Unix` environment, installing these applications and running this protocol described by Haas et al. (2013) would be a straight-forward procedure. However, to a novice, this would be a difficult task. Not being an administrator also significantly complicates installation of applications. Another challenge in performing such a procedure would be having to re-do the analysis, either multiple times whilst changing certain parameters, or performing the analysis using more than one dataset. In this case, simply executing the protocol commands on a command line interface (CLI) would not suffice. Custom scripts would have to be created in order to compile and order the multiple commands needed to execute the protocol procedure repeatedly or on multiple datasets (Piccolo & Frampton, 2016). Another option would be to implement "workflow management systems" to construct a pipeline (or "workflow") of the multiple analyses steps, handle input/output files between applications and also automate the analysis (Di Tommaso et al., 2017; Piccolo & Frampton, 2016; Schulz et al., 2016).

Another challenge that the scientific community faces in performing multi-step analysis that requires different pieces of software at each analysis step is software dependencies and libraries (Kurtzer et al., 2017). Many bioinformatics tools are built from sources, and thus there will be a complexity of dependencies and libraries between the softwares needed to perform the analyses (Schulz et al., 2016). In addition to software and library dependency, there is also a computational environment or an operating system (OS) dependency. Installation of different application softwares on different OSs requires different configuration steps, and some applications are only designed to be executed on a specific environment of a specific OS (Kurtzer et al., 2017; Piccolo & Frampton, 2016). A solution to software and OS dependency is to use virtual machines or software package managers (containers) (Boettiger, 2015; Kurtzer et al., 2017; Piccolo & Frampton, 2016; Schulz et al., 2016).

When big datasets are being analysed, personal desktop machines and laptops are not an option. In most cases, bioinformatic analyses will require a significant amount of computing power, memory and will sometimes need to be processed simultaneously (in parallel) in order to reduce the amount of time needed to perform each task (Di Tommaso et al., 2017; Kurtzer

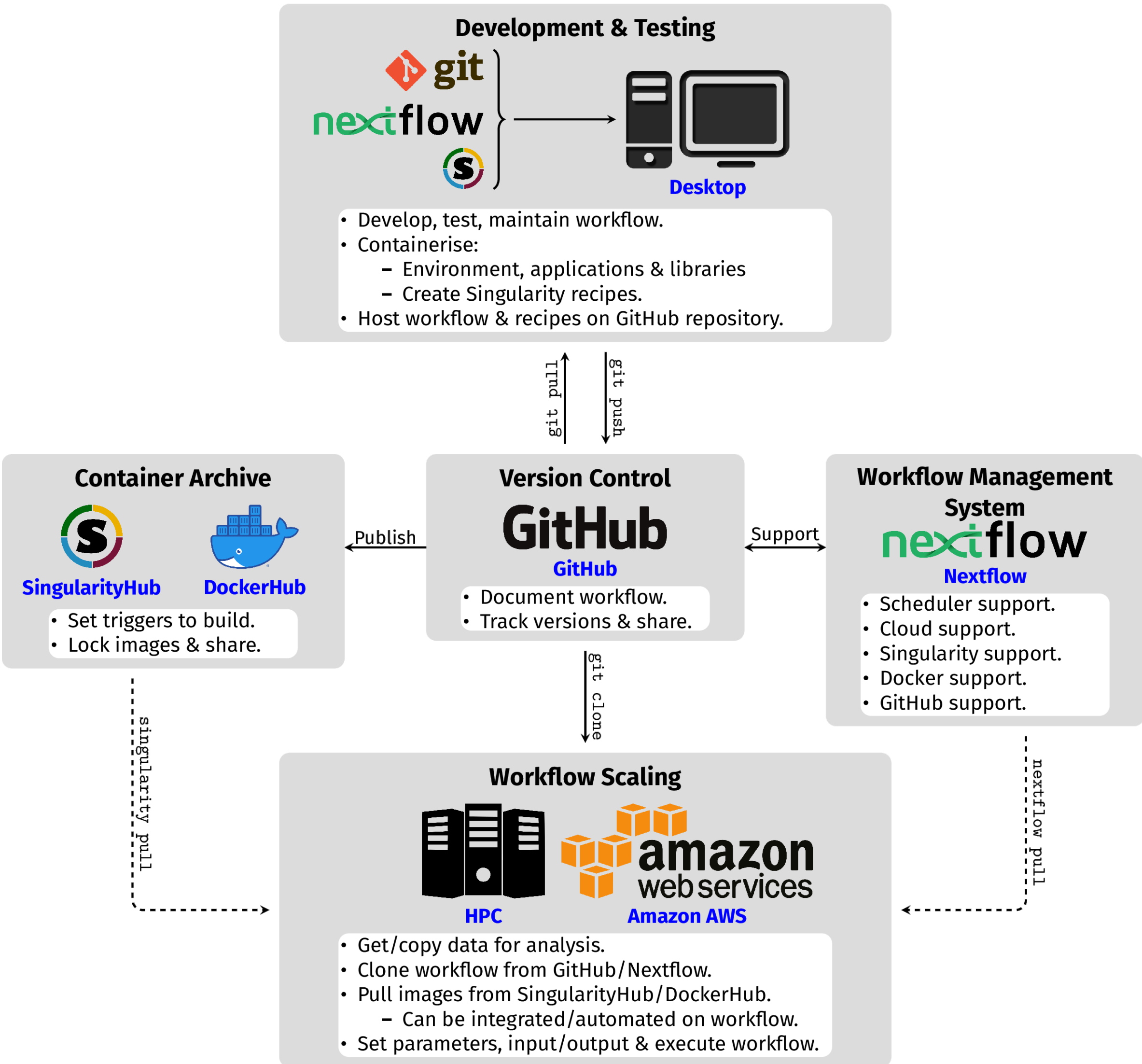

Figure 1: **Summary of resources and best practices for development, maintenance, sharing and publishing of reproducible and portable pipelines.** Development of reproducible pipelines start on individual desktop machines using `Nextflow` (Di Tommaso et al., 2017), `Singularity` (Kurtzer et al., 2017) and Git (https://git-scm.com/). A pipeline repository can be created on GitHub (https://github.com/) to track version changes. SingularityHub (https://singularity-hub.org/) or DockerHub (https://hub.docker.com/) can be used to create and archive containers triggered by a GitHub push. The pipeline can be cloned on HPC or cloud-services for analyses on a larger scale.

https://doi.org/10.18489/sacj.v33i2.830

et al., 2017). These analyses have to be performed on high-performance computing (HPC) clusters available in most research institutes or cloud-computing services which offer significantly high computing resources that can meet the requirements of intensive bioinformatic analyses (Kurtzer et al., 2017). This "scaling up" of bioinformatic analyses to cloud environments and HPC clusters is further enhanced by a combination of workflow management system and containerisation of software; making bioinformatic analyses pipelines "portable" across different computing platforms (Boettiger, 2015; Kurtzer et al., 2017; Piccolo & Frampton, 2016; Schulz et al., 2016). Figure 1 summarises the best practices and tools that researchers could apply to their research approach and reproducible pipeline development. This combination also overcomes the limitation of software installation on HPC clusters and cloud-services as sometimes the users do not have privileges to install softwares and their dependencies (Kurtzer et al., 2017). Furthermore, coupling the combination workflow management systems, software containers and HPC with proper documentation and storing code using version control systems (VCS) creates portable pipelines that can be shared amongst the scientific community and ensures reproducibility across different platforms (Di Tommaso et al., 2017; Kurtzer et al., 2017; Perkel, 2016; Piccolo & Frampton, 2016).

The availability of RNA-seq data from black South African patients affected with systemic sclerosis (SSc) and healthy individuals from the study by Frost et al. (2019) presented an opportunity to develop a robust computational pipeline in an effort to bridge the gap between repetitive (and most often complicated) bioinformatic analyses and the large datasets produced by NGS technologies. Human genome sequencing is still a relatively costly venture, mainly due to the genome size. However, since only a fraction of a genome is transcribed, the set of the transcribed RNA molecules (transcriptome) reflects the current state of the cell (or group of cells) in a given tissue and at a specific time. The analysis of the transcribed RNA molecules can often provide insights into the etiology and underlying pathological mechanism of a disease (Finotello & Di Camillo, 2015).

Although a useful approach, studying gene expression through the transcriptome alone has the limitation that it does not necessarily correlate directly with amount of protein present in a cell, and should therefore be interpreted with caution. Nonetheless, RNA-seq provides a quick and cost effective way of obtaining large amounts of transcriptome data, providing insights into the levels of gene expression. Such data allow us to identify transcribed genes, discover new disease-associated genes, measure transcript and gene expression abundance, study allelic information and identify splice variants for genes (Grabherr et al., 2011; Haas et al., 2013; Li et al., 2014; Trapnell et al., 2013; Trapnell et al., 2012). A typical RNA-seq analysis involves three major steps: (1) identifying a set of genes and/or transcripts from the hundreds of millions of short ($\sim$ 36-125 bases) RNA-seq reads, produced from the sequencing experiment, through mapping to a reference genome/transcriptome; (2) quantifying the abundance of the genes/transcript; and (3) performing differential expression analysis (Conesa et al., 2016). This paper presents `nf-rnaSeqCount`, a portable, reproducible and scalable `Nextflow` pipeline that addresses the first two steps of RNA-seq data analyses, i.e., (1) identification of genes from RNA-seq data and (2) quantifying their abundance.

2 PIPELINE IMPLEMENTATION AND WORKFLOW

The purpose of the pipeline is to make the task of producing raw read counts for performing differential expression analysis easier, especially for other researchers with little or no knowledge of bioinformatics. The pipeline also needs to be portable and reproducible in order to allow scaling to different computational platforms when large or small datasets are being analysed.

2.1 `nf-rnaSeqCount` Implementation

`Nextflow` (Di Tommaso et al., 2017) and `Singularity` (Kurtzer et al., 2017) were used to implement the pipeline into a portable and reproducible pipeline. `Nextflow` is a Groovy-based domain-specific language (DSL) specifically designed for bioinformaticists with strong programming knowledge to solve many of the challenges of the inability to reproduce data analysis. Some of these challenges are due to computational platform variations, software and database management, complexity of pipelines, intermediate file handling and lack of good practice (Di Tommaso et al., 2017). The advantages of `Nextflow` as a workflow management system are that it can handle input and output as channels between processes and reduce the need of having to create intermediate directories to store intermediate results. Variables can also be declared dynamically with no need to explicitly name files, and only the output that is required can be saved to files in each analysis step.

`Nextflow` also has a number of features that promote pipeline reproducibility and portability. It supports `Docker`[1] and `Singularity`[2], the two most used containerisation softwares in the bioinformatics community; it integrates/supports the popular VCS GitHub[3] for sharing of code, and version management; and it allows for scaling of computational pipelines on HPC and cloud systems by providing out of the box scheduler support for Sun Grid Engine (SGE), PBS/Torque, Platform Load Sharing Facility (LSF), Simple Linux Utility for Resource Management (SLURM), HTCondor and Amazon Web Services (AWS) (Di Tommaso et al., 2017).

To facilitate reproducibility and portability of the pipeline, `Docker` containers were created for each of the processes applications in the pipeline and hosted on DockerHub[4] to use with `Singularity` when executing the pipeline. `Singularity` is a lightweight platform for building and running containers that is gaining popularity in the bioinformatics community, especially for performing analysis on a large scale. It uses an image format that is supported across different versions of the C library and kernels, and gives users the ability to encapsulate their pipelines, all required applications, dependencies and base OS environment into a single file that can be locked, copied, shared and archived (Kurtzer et al., 2017). `Docker` containers hosted on DockerHub can be downloaded and converted into `Singularity` image format

[1] http://docker.io/
[2] https://www.sylabs.io/
[3] http://github.com/
[4] https://hub.docker.com/

(SIF). Such image files have standard Linux/UNIX file permission and cannot be modified (not even by the host OS), thus can be used with confidence that nothing within the image has changed. Each SIF contains the necessary software required by each process to run. This removes the need to install all the software tools used for these analyses.

2.2 `nf-rnaSeqCount` Workflow

The `nf-rnaSeqCount` pipeline depends on `Nextflow` and `Singularity` to run. These two softwares must be installed in order for the pipeline to be executed. The input for the `nf-rnaSeqCount` pipeline are `FASTQ` files (both paired- and single-ended) for the RNA-seq data to be analysed, a reference genome (in `FASTA` format) and an annotation file (in `GTF` format) for the reference genome. These can be passed as command-line arguments during pipeline execution or specified in a configuration file (e.g. `main.conf` in Figure 2) that can also be passed to the pipeline during execution. Other parameters of the pipeline can be specified in a similar way. Unlike most pipelines that are executed in one go, the `nf-rnaSeqCount` is a modular pipeline that can be executed in multiple stages (`main.nf` in Figure 2), allowing the users to interact with the results at each step of the pipeline.

The different modules of the pipeline can be grouped into 4 steps, and each module (specified using the `–mode` argument) calls the required process in the workflow. The 4 workflow steps and their modules are as follows: (1) **Data Preparation**: `prep.Containers` and `prep.Indexes`; (2) **Quality Control**: `run.ReadQC` and `run.ReadTrimming`; (3) **Alignment and Quantification**: `run.ReadAlignment` and `run.ReadCounting`; and (4) **MultiQC**: `run.MultiQC`. The `nf-rnaSeqCount` pipeline can be obtained using the following command:

```
nextflow pull phelelani/nf-rnaSeqCount
```

The help menu for the pipeline can be accessed with the following command:

```
nextflow run nf-rnaSeqCount --help
```

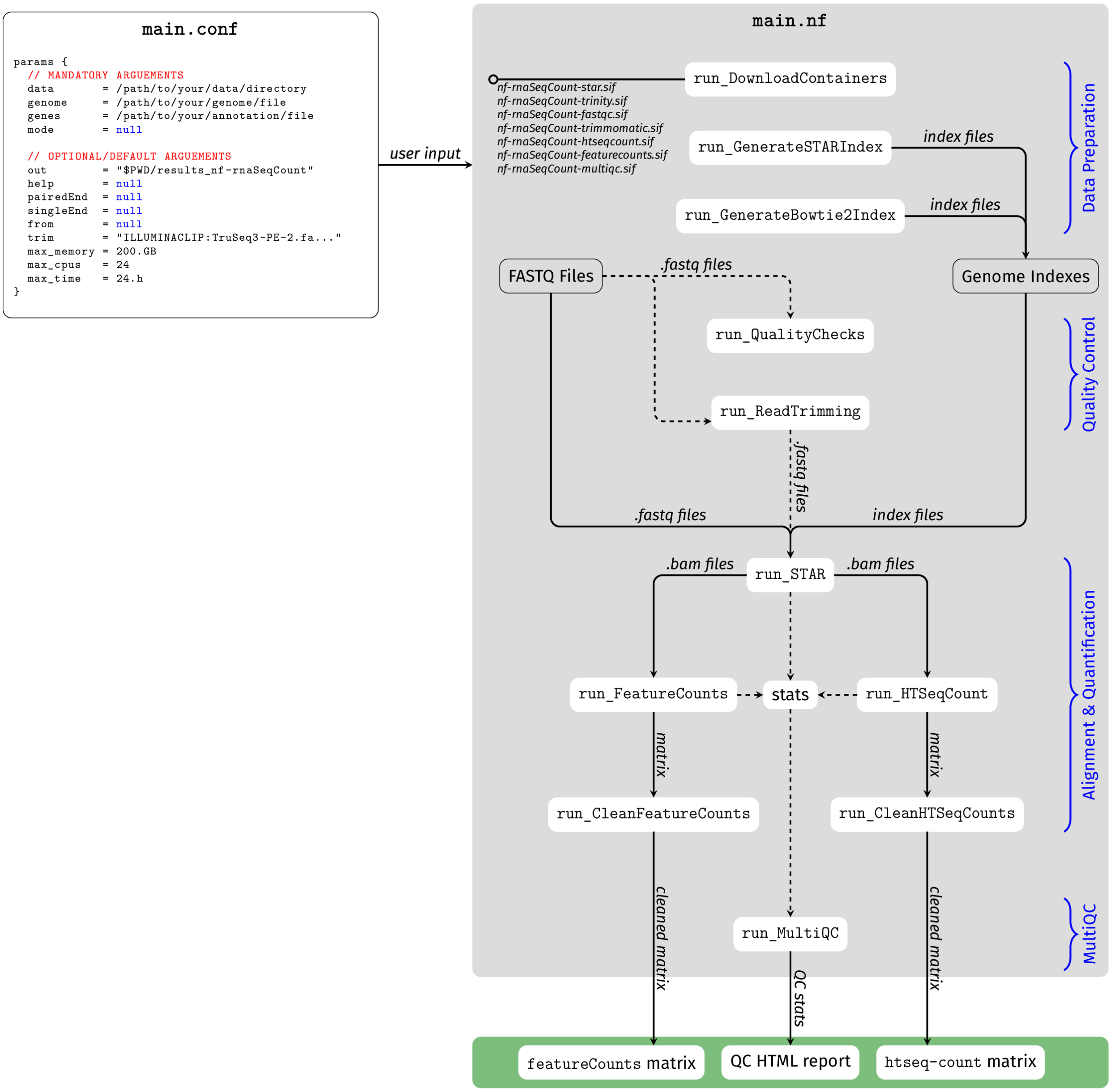

Figure 2: **Overall summary of the `nf-rnaSeqCount` pipeline.** The `nf-rnaSeqCount` pipeline works in 4 stages: (1) **Data Preparation:** for downloading `Singularity` containers and indexing the reference genome using `STAR` and `Bowtie`; (2) **Quality Control:** for assessing the quality of RNA-seq reads using `FastQC` and trimming low quality bases using `Trimmomatic`; (3) **Alignment & Quantification:** for aligning reads to the reference genome using `STAR` and quantifying abundance of identified genomic features using `featureCounts` and `htseq-count`; (4) **MultiQC:** for assessing the quality of the steps in the pipeline using `MultiQC`. The main output for the `nf-rnaSeqCount` pipeline are read count matrices produced by `featureCounts` and `htseq-count`, as well as a QC report from `MultiQC`.

2.2.1 Data Preparation

Data preparation is mandatory at the first step of the workflow. The first process in this step, `run_DownloadContainers`, downloads all the required workflow containers with the required software for executing the pipeline from DockerHub and converts them to `Singularity` format. This step is crucial as all processes in the pipeline depend on the applications that are packaged in these containers:

```
## Download Singularity containers
nextflow run nf-rnaSeqCount --mode getContainers
```

Once the `Singularity` containers have been downloaded, the `run_GenerateSTARIndex` and `run_GenerateBowtie2Index` processes will index the reference genome (for aligning the RNA-seq reads and quantifying the abundance of the identified genomic features) using `STAR` (Dobin et al., 2013) and `Bowtie` (Langmead et al., 2009), respectively. Before indexing, the location for the reference genome (FASTA), annotation file (GTF), input FASTQ files and output directory can be provided in a configuration file (as in Figure 2), and passed to the `Nextflow` command using the `-c` option. However, the files can also be passed as command-line arguments when executing the pipeline. The remainder of the pipeline assumes that all the required files are provided in a configuration file called `main.conf`. To index the reference genome using `STAR` and `Bowtie`, the following commands can be used:

```
## Generate STAR and Bowtie2 indexes
nextflow run nf-rnaSeqCount -c main.conf --mode prep.Indexes
```

2.2.2 Quality Control

The `nf-rnaSeqCount` pipeline incorporates `FastQC` (Andrews, 2010) and `Trimmomatic` (Bolger et al., 2014) for pre-processing of reads. This "quality control" (QC) step of the `nf-rnaSeqCount` workflow is optional; however, it is very useful to first assess the initial quality of the reads so that poor quality reads and contaminating "adapter" sequences can be removed. In this step, the `run_QualityChecks` process uses `FastQC` to assess the quality of the RNA-seq reads. The `run_ReadTrimming` process, which uses `Trimmomatic`, is then used to remove technical sequences and poor quality bases from the data. To assess the quality of the RNA-seq reads, the following command can be executed:

```
## Perform QC on reads
nextflow run nf-rnaSeqCount -c main.conf --mode run.ReadQC
```

Once the quality of the reads has been assessed, `Trimmomatic` can be used to remove low quality bases and adapter sequences. The `−trim` option can be used to pass `Trimmomatic` arguments to the pipeline:

```
## Trim low quality reads
nextflow run nf-rnaSeqCount -c main.conf --mode run.ReadTrimming --trim 'TRAILING:28 MINLEN:40'
```

https://doi.org/10.18489/sacj.v33i2.830

2.2.3 Alignment and Quantification

The alignment and quantification is the main step of the `nf-rnaSeqCount` workflow. In this step, the RNA-seq reads are aligned to the reference genome (indexed in the "Data Preparation" step) using STAR (Dobin et al., 2013) in the `run_ReadAlignment` process. The resulting BAM files are then used to quantify the abundance of the identified genes using both `featureCounts` (Liao et al., 2014) (`run_FeatureCounts` process) and `htseq-count` (Anders et al., 2015) (`run_HTSeqCount` process). To align the reads to the reference genome and quantify the abundance of the genomic features identified, the following commands can be used:

```
## Align reads to reference genome
nextflow run nf-rnaSeqCount -c main.conf --mode run.ReadAlignment

## Quantify the abundance of identified features
nextflow run nf-rnaSeqCount -c main.conf --mode run.ReadCounting
```

2.2.4 MultiQC

The MultiQC step is optional. In this step, the `run_MultiQC` process uses `MultiQC` (Ewels et al., 2016) to collect all the statistics from all the programs that were executed in the workflow, and give a summary of all statistics in an HTML file. To obtain the summary statistics of the workflow, the following command can be used:

```
## Obtain summary statistics from all tools
nextflow run nf-rnaSeqCount -c main.conf --mode MultiQC
```

2.3 `nf-rnaSeqCount` Pipeline Output

The output directory for the `nf-rnaSeqCount` pipeline contains a number of folders:

- n number of folders corresponding to each of the samples that were processed by the pipeline. These folders contain general statistics on mapping using STAR.
- ***featureCounts*** folder containing read counts matrix (`gene_counts_final.txt`) for `htseq-count`. This file can be used for differential expression analysis.
- ***htseqCounts*** folder containing read counts matrix (`gene_counts_final.txt`) for `featureCounts`. This file can be used for differential expression analysis.
- ***report_QC*** folder containing `MultiQC` QC reports in HTML format. This file can be used to assess the quality of read mapping and gene quantification.
- ***report_workflow*** folder containing pipeline execution reports. These files can be used to trace the execution of the pipeline and check other metadata in order to assign resources correctly to the processes.

2.4 Testing

The RNA-seq data of black South African patients with SSc were used to test and validate the usefulness of the `nf-rnaSeqCount` pipeline. This transcriptome data was from a study by Frost et al. (2019), conducted with ethics approval of the Human Research Ethics Committee (HREC [Medical]) of the University of the Witwatersrand (certificate number M120512). The `nf-rnaSeqCount` pipeline was initially developed and tested on the University of the Witwatersrand (Wits) Computing cluster using SLURM and PBS job schedulers. The pipeline also has been successfully tested on the University of Cape Town (UCT) eResearch HPC[5] and on Amazon's AWS[6] using RNA-seq data for this study. On the UCT's eResearch HPC, which has SLURM as the job scheduler, the same computing requirements as with the Wits Computing cluster were used.

For AWS execution of the pipeline, the `Nextflow` supported Amazon Machine Image (AMI), `ami-4b7daa32`, was used to deploy an Amazon Elastic Block Store (EBS) of 1000GB using the Elastic Compute Cloud (EC2), `m4.10xlarge`, with 40 virtual CPUs and 160 GB of memory. All AWS analyses were performed on the European (Ireland) region since the `Nextflow` AMI was only available for this region. The pipeline was executed using the standard computing environment (no job scheduler) on the EC2. Estimating the cost of running the analysis on the AWS (February 2019 pricing), the `m4.10xlarge` cost \$2.22 per hour[7]), the standard general purpose solid-state drive (SSD) costs \$0.11 per GB-month[8]. Given that the analysis took approximately 4 hours, the total approximated cost for running the `nf-rnaSeqCount` pipeline on the SSc data was:

$$\left(\frac{\$2.2}{hour} \times 4\ hours\right) + \left(\frac{\$0.11/GB}{month} \times 1000GB \times \frac{4\ hours}{740\ hours}\right) = \$9.39$$

2.5 Benchmarking

The `nf-rnaSeqCount` pipeline was further benchmarked for time, memory and CPU usage against the popular `Rsubread` package (Liao et al., 2019). Benchmarking of the `nf-rnaSeqCount` pipeline against the `Rsubread` package was to determine the speed at which both tools complete different tasks, computational resource requirements, scalability on large datasets, ability to perform tasks in parallel as well as usability. The `Rsubread` package was chosen as it is a comprehensive tool and performs similar RNA-seq analysis workflow (read alignment and read counting) as our `nf-rnaSeqCount` pipeline. The RNA-seq data used for benchmarking was downloaded from the Gene Expression Omnibus (GEO) with an accession GSE111073[9]. The data consisted of 21 RNA-seq breast cancer samples from walnut-consuming patients (10 samples) and control group (11 samples) (Hardman et al., 2019a,

[5] http://hpc.uct.ac.za/
[6] https://aws.amazon.com/
[7] https://aws.amazon.com/ec2/pricing/on-demand/
[8] https://aws.amazon.com/ebs/pricing/
[9] https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111073

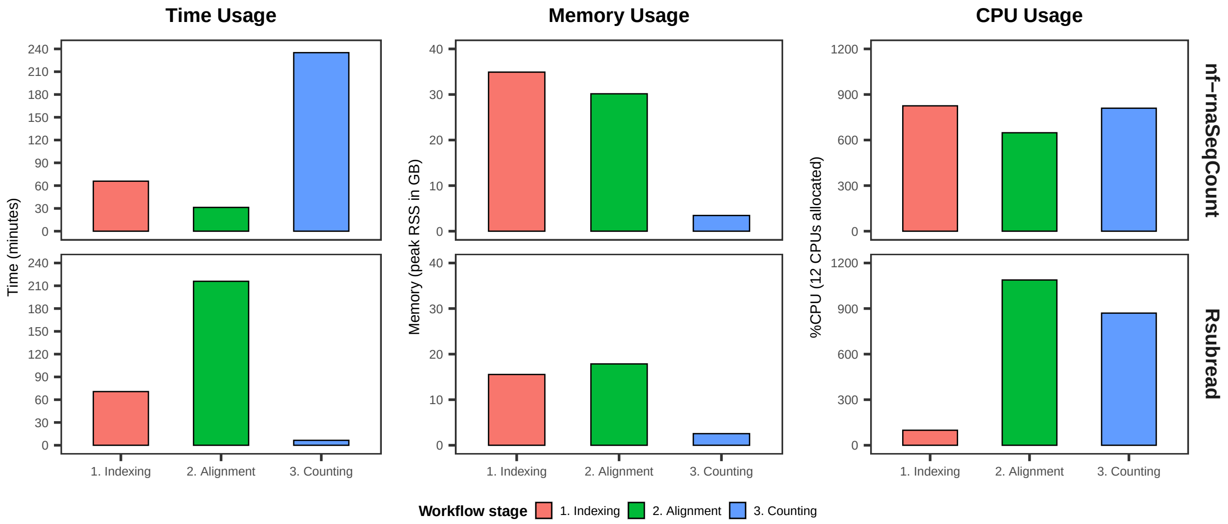

Figure 3: **`nf-rnaSeqCount` and `Rsubread` performance benchmarking.** The `nf-rnaSeqCount` pipeline (top row) was compared to the `Rsubread` package (bottom row) in terms of time (1st column), memory (2nd column) and CPU usage (3rd column) when performing the standard RNA-seq workflow, i.e., indexing (red), read alignment (green) and read counting (blue).

2019b). Benchmarking of the two tools was carried out on the Wits Computing cluster, with 48 GB of memory and 12 CPUs allocated to each task in the analysis workflow, i.e., reference genome indexing, read alignment and read counting.

Figure 3 summarises the benchmarking results between `nf-rnaSeqCount` and the `Rsubread` package in terms of their time, memory and CPU usage when performing genome indexing, read alignment and read counting. The `nf-rnaSeqCount` pipeline was able to distribute tasks across multiple nodes on the Wits Computing cluster, i.e., run jobs in parallel, in addition to multi-threading, whereas the `Rsubread` applications were multi-threaded only on a single node.

When it comes to reference genome indexing, both the `nf-rnaSeqCount` (using STAR and `Bowtie`) and `Rsubread` (using `index`) performed equally in terms of time usage, with each tool completing the task in 66 and 71 minutes, respectively. The `nf-rnaSeqCount` utilised more resources for indexing (35 GB of memory and 825% of allocated CPUs) compared to `Rsubread` (16 GB and 99% of allocated CPUs). For the alignment of the RNA-seq reads to the reference genome, `nf-rnaSeqCount` (using STAR) outperformed `Rsubread` (using `align`), with the alignment tasks completed 31 and 215 minutes, respectively. `nf-rnaSeqCount` completed the alignment tasks using 30 GB of memory and 648% of the allocated CPUs, whilst `Rsubread` completed the alignment tasks using 18 GB of memory and 1088% of the allocated CPUs.

Finally, for the read counting, the `nf-rnaSeqCount` (using `htseq-count` and `featureCounts`) was outperformed by `Rsubread` (using `featureCounts`), with the read

https://doi.org/10.18489/sacj.v33i2.830

counting tasks completed in 235 and 6 minutes, respectively. `nf-rnaSeqCount` completed the read counting tasks using 3 GB of memory and 810% of the allocated CPUs, whilst `Rsubread` completed the alignment tasks using 3 GB of memory and 870% of the allocated CPUs. The huge difference with time usage seen between `nf-rnaSeqCount` and the `Rsubread` when it comes to read counting can mainly attributed to `htseq-count`. Unlike other tools in both workflows, `htseq-count` cannot be multi-threaded, thus drastically reducing performance and increasing amount of time it takes to complete the read counting tasks for the `nf-rnaSeqCount` workflow.

3 DISCUSSION

The `nf-rnaSeqCount` pipeline has been successfully implemented in `Nextflow` and `Singularity`, and can be executed on any UNIX-based OS with `Nextflow` and `Singularity` installed. It is available on GitHub[10] and all the workflow containers with softwares required for running the pipeline are hosted on DockerHub[11]. In addition to running the pipeline locally, the `nf-rnaSeqCount` pipeline also supports the PBS and SLURM job schedulers on HPCs, and this information can be passed to the `-profile` option of `Nextflow` when executing the pipeline. Available options are `slurm` (for SLURM scheduler) and `pbs` (for PBS scheduler).

The benchmarking results reveal that the `nf-rnaSeqCount` pipeline compares almost as equally well as the popular `Rsubread` package in terms of runtime and resource usage. However, `nf-rnaSeqCount` has added advantages over `Rsubread`: (1) **Parallelisation:** in addition to applications being multi-threaded (with the exception of `htseq-count`) within the workflow, the implementation of `nf-rnaSeqCount` on `Nextflow` also allows processes to be run across multiple nodes on HPC, which drastically improves performance when working with large datasets; (2) **Installation:** there is no need for installation of packages and dependencies for the `nf-rnaSeqCount` pipeline, e.g., only `Nextflow` and `Singularity` are required; and (3) **Usability:** there is no need for writing tedious scripts for performing RNA-seq analysis with `nf-rnaSeqCount`, i.e., all inputs, outputs and parameters can be put into a `config` file which will be used to execute each step of the analysis.

The main requirements for a highly efficient pipeline include reproducibility (capability of the pipeline to reproduce the results under different computational resources), portability (capability of using the pipeline on different computational platforms) and scalability (being able to execute the pipeline on desktop machines, cloud or HPC environments). The `nf-rnaSeqCount` pipeline presented in this paper meets these requirements for an efficient pipeline. `nf-rnaSeqCount` is designed on `Nextflow` and all its application dependencies are packaged in `Singularity` containers. This makes it possible to run the pipeline on any machine, from desktop to HPC, with both `Nextflow` and `Singularity` installed. `Nextflow`

[10] https://github.com/phelelani/nf-rnaSeqCount
[11] https://hub.docker.com/r/phelelani/nf-rnaseqcount

supports a wide variety of job schedulers, and the nf-rnaSeqCount pipeline comes packaged with support for PBS and SLURM schedulers. Advanced users can add their own scheduler support using the nextflow.config file.

The pipeline also comes with detailed documentation on GitHub, for users interested in using this pipeline. The workflow containers hosted on DockerHub ensure that users do not have to go an extra step to install all the softwares required to execute the nf-rnaSeqCount workflow. The modularity of the pipeline not only allows users to interact with results from each step, but also to modify the parameters for the different applications used in the workflow.

4 CONCLUSION

The nf-rnaSeqCount pipeline presented here provides a quick and efficient way to obtain a matrix of read counts (matrix **N** of $n \times m$, where N_{ij} is the number of reads assigned to gene i in sequencing experiment j (Rapaport et al., 2013)) that can be used for differential expression and pathway analysis. The output from the nf-rnaSeqCount pipeline can be directly used with popular downstream differential expression analysis tools, such as DESeq2 (Love et al., 2014) and edgeR (Robinson et al., 2010), which take raw read counts as input. The nf-rnaSeqCount pipeline incorporates common tasks associated with RNA-seq data analyses, including QC, read trimming, read alignment and gene quantification. This pipeline largely eliminates the need for multiple scripts and tedious repetitive tasks associated with RNA-seq data analysis, especially when working with large RNA-seq datasets. Users wishing to use the nf-rnaSeqCount pipeline can do so by cloning the repository onto their computational platform (desktop, HPC or cloud) with UNIX-based OS. The availability of workflow containers on DockerHub for executing the nf-rnaSeqCount pipeline eliminates the need for manual installation of applications.

DECLARATIONS

Competing Interests

The authors declare that they have no competing interests.

Funding

This work is based on the research supported in part by the National Research Foundation of South Africa for the Thuthuka Funding Instrument grant (unique grant no. 99206) and Connective Tissue Diseases Research Fund, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg. SH is partially supported by the H3ABioNet grant from the NIH Common Fund Award / NHGRI Grant Number U24HG006941.

Authors' Contributions

P.T. Mpangase: conceptualisation, methodology, software, validation, formal analysis, investigation, data curation, writing (original draft), visualisation, project administration, funding acquisition.

J. Frost: resources, writing (review & editing).

M. Tikly: resources, writing (review & editing).

M. Ramsay: conceptualisation, methodology, writing (review & editing), supervision, funding acquisition.

S. Hazelhurst: conceptualisation, methodology, validation, writing (review & editing), supervision, funding acquisition.

ACKNOWLEDGEMENTS

The authors would like to thank J. Ponomarenko, S. Bonnin, L. Cozzuto, A. Vlasova and P. Di Tommaso from the Centre for Genomic Regulation (CRG) for hosting and assisting P.T. Mpangase; and T. Ngcungcu for kindly allowing us to use the RNA-seq data on albinism for testing the pipeline.

References

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. [(01 September 2018, last accessed)]. Retrieved September 1, 2018, from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Boettiger, C. (2015). An introduction to Docker for reproducible research. *ACM SIGOPS Operating Systems Review*, *49*(1), 71–79. https://doi.org/10.1145/2723872.2723882

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*(1), 13. https://doi.org/10.1186/s13059-016-0881-8

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, *35*(4), 316–319. https://doi.org/10.1038/nbt.3820

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fan, J., Han, F., & Liu, H. (2014). Challenges of Big Data analysis. *National Science Review*, *1*(2), 293–314. https://doi.org/10.1093/nsr/nwt032

Finotello, F., & Di Camillo, B. (2015). Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. *Briefings in Functional Genomics*, *14*(2), 130–142. https://doi.org/10.1093/bfgp/elu035

Frost, J., Estivill, X., Ramsay, M., & Tikly, M. (2019). Dysregulation of the Wnt signaling pathway in South African patients with diffuse systemic sclerosis. *Clinical Rheumatology*, *38*(3), 933–938. https://doi.org/10.1007/s10067-018-4298-5

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hardman, W. E., Primerano, D. A., Legenza, M. T., Morgan, J., Fan, J., & Denvir, J. (2019a). Dietary walnut altered gene expressions related to tumor growth, survival, and metastasis in breast cancer patients: a pilot clinical trial. *Nutrition Research*, *66*, 82–94. https://doi.org/10.1016/j.nutres.2019.03.004

Hardman, W. E., Primerano, D. A., Legenza, M. T., Morgan, J., Fan, J., & Denvir, J. (2019b). mRNA expression data in breast cancers before and after consumption of walnut by women. *Data in Brief*, *25*, 104050. https://doi.org/10.1016/j.dib.2019.104050

Kluge, M., & Friedel, C. C. (2018). Watchdog – a workflow management system for the distributed analysis of large-scale experimental data. *BMC Bioinformatics*, *19*(1), 97. https://doi.org/10.1186/s12859-018-2107-4

Kurtzer, G. M., Sochat, V., & Bauer, M. W. (2017). Singularity: Scientific containers for mobility of compute (A. Gursoy, Ed.). *PLOS ONE*, *12*(5), e0177459. https://doi.org/10.1371/journal.pone.0177459

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Li, B., Fillmore, N., Bai, Y., Collins, M., Thomson, J. A., Stewart, R., & Dewey, C. N. (2014). Evaluation of de novo transcriptome assemblies from RNA-Seq data. *Genome Biology*, *15*(12), 553. https://doi.org/10.1186/s13059-014-0553-5

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Liao, Y., Smyth, G. K., & Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*, *47*(8), e47–e47. https://doi.org/10.1093/nar/gkz114

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Perkel, J. (2016). Democratic databases: science on GitHub. *Nature*, *538*(7623), 127–128. https://doi.org/10.1038/538127a

Piccolo, S. R., & Frampton, M. B. (2016). Tools and techniques for computational reproducibility. *GigaScience*, *5*(1), 30. https://doi.org/10.1186/s13742-016-0135-4

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C. E., Socci, N. D., & Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology*, *14*(9), R95. https://doi.org/10.1186/gb-2013-14-9-r95

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Schulz, W., Durant, T., Siddon, A., & Torres, R. (2016). Use of application containers and workflows for genomic data analysis. *Journal of Pathology Informatics*, *7*(1), 53. https://doi.org/10.4103/2153-3539.197197

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, *31*(1), 46–53. https://doi.org/10.1038/nbt.2450

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*(3), 562–578. https://doi.org/10.1038/nprot.2012.016

SACJ 33(2) December 2021
Research Article

Consumer-centric factors for the implementation of smart meters in South Africa

Tonderai Muchenje, Reinhardt A. Botha

School of Information Communication Technology, Nelson Mandela University, Port Elizabeth, South Africa

ABSTRACT

Smart meter implementation is still in its infancy in many African countries, including South Africa. This is evident from the fact that most research studies are either Eurocentric or American-centric. Hence, this research aimed to identify consumer-centric factors for planning considerations in implementation of smart meters in South Africa. We used various behavioural theoretical models found in literature to identify potential factors relevant to this study. Based on quantitatively gathered data ($n = 705$), a structural equation model (SEM) was used to evaluate the identified factors. This study found that only ten consumer-centric factors were significant to smart meter consumers. These factors include behavioural intention, attitude, trust in technology, social norms, facilitating conditions, perceived usefulness, perceived ease of use, privacy risk, monetary cost, and perceived value. In conclusion, the study shows that not all factors suggested within the European and American context are relevant for smart meter implementation within the South African context. Hence, results of this study hold some practical implications in assisting utility companies in identifying consumer-centric factors that are relevant to the South African population. Finally, consumer-centric factors can be used by policy makers and energy regulators as baseline factors for future pervasive technology acceptance studies.

Keywords: Smart meter, consumer-centric, perceived usefulness, perceived value, behavioural intention, structural equation model, trust in technology, privacy risk, facilitating conditions, social norms

Categories: • **Human-centered computing ∼ Ubiquitous and mobile computing theory, concepts and paradigms**

Email:
Tonderai Muchenje muchenjet@tut.ac.za (CORRESPONDING),
Reinhardt A. Botha ReinhardtA.Botha@mandela.ac.za

Article history:
Received: 4 November 2020
Accepted: 25 November 2021
Available online: 19 December 2021

1 INTRODUCTION

The global demand for electric energy consumption. both for consumer and commercial use, has continued to increase rapidly owing to population and economic growth (Organisation for Economic Co-operation and Development, 2012). A sustainable future energy strategy is required to create innovative ways and systems that are relevant to transform the way

Muchenje, T. and Botha, R. A. (2021). Consumer-centric factors for the implementation of smart meters in South Africa. *South African Computer Journal 33*(2), 17–54. https://doi.org/10.18489/sacj.v33i2.909

SACJ is a publication of the South African Institute of Computer Scientists and Information Technologists. ISSN 1015-7999 (print) ISSN 2313-7835 (online).

electrical energy is produced, transmitted, distributed, and consumed. Otherwise, the access to electrical energy in the future will only be for the privileged, but unaffordable to the poor.

Currently, many countries worldwide are being forced to rethink their electricity infrastructure investment and energy generation, transmission, supply, and management in general (Eberhard et al., 2017), as a matter of urgency. Developed countries, such as the United States of America, and those in Europe, have taken initiatives to replace their ageing traditional electricity infrastructure with electricity grids commonly referred to as smart grids (J. Kranz et al., 2010; J. Kranz et al., 2015; J. J. Kranz & Picot, 2011). It has been found that many parts of the traditional electricity grids in various countries are decades old, wearing out and failing to contain the increasing electricity demand, monitoring and control needed to support economic growth prospects (Bazilian et al., 2013; Manyika et al., 2013). Therefore, utility companies across the world are investing billions of dollars to upgrade and modernise electricity grids with smarter technologies, aiming to improve electricity usage efficiency, reliability, privacy, and security in the digital age (Camarinha-Matos, 2016; Ma et al., 2018). Although the implementation of smart meters, as part of the smart grid system, offers many benefits to both consumers and service providers (utility companies), it has been faced with wide resistance in many countries across the world.

Resistance to smart meter implementation is a reality in South Africa too. From consideration of the challenges mentioned above, several gaps can be identified that should be researched from an African customer perspective, because factors impacting smart meter adoption in Africa are potentially different from those in both America and Europe (J. Kranz et al., 2010; J. Kranz et al., 2015; J. J. Kranz & Picot, 2011). Therefore, this research study was delineated to focus only on the South African context.

In South Africa, there is a need to understand and explore planning considerations for smart meter implementation and it was this need that became the major motivation for conducting this study. In relation to this, research is needed to support the development of a consumer-focused predictive model and planning considerations in smart meter implementation of planning projects, with specific reference to technology acceptance and use. Therefore, the need to investigate the consumer-centric factors that facilitate high acceptance of smart meter implementation for planning within the South African context can be considered vital research.

The paper first provides a background to the study on smart meters, discussing smart meter adoption in the rest of the world, highlighting the role of customer-centric factors. Subsequently, a justification for using the technology acceptance model as the foundational research model for this research is discussed. This is followed by a discussion of the supplementary models and development of the research model, as well as the theoretical underpinnings that the study. The discussion of the research methods is followed by hypothesis development, model testing and the presentation of the empirical results. The research paper concludes with a discussion of the theoretical implications.

2 BACKGROUND

Energy is becoming the oxygen and lifeblood of the mass industrialized world and is a critical resource for fostering growth in emerging economies (World Economic Forum, 2012). The global demand for and consumption of energy, including electricity, is growing faster than its generation and supply capacity. To generate electricity at a capacity that can meet both residential and industrial demands, while keeping pace with economic growth and climate change requires striking a balance that is difficult to maintain (Wunderlich et al., 2012). Unfortunately, for many decades, electricity infrastructure has not changed much. Most components of the traditional electricity grid used in various countries are decades old and, consequently, are wearing out and failing to meet the increasing demand to supply electricity as well as control and monitor its consumption (Barrett et al., 2013; Manyika et al., 2013). Therefore, utility companies across the world are investing billions of dollars into upgrading and modernizing their electricity grid infrastructure with smarter technologies to improve its efficiency, reliability, privacy, and security, and launching them into the digital age (Bargh & McKenna, 2004).

Incorporating smart meters (Wang et al., 2019) into the traditional electricity grid has become inevitable for countries that are pursuing sustainable development and energy efficiency (Zamrudi et al., 2019). Smart meters provide monitoring and control capabilities to the smart grid (Dileep, 2019). Consumers can remotely manage their energy consumption, while utility companies can manage and control the electricity supply, billing, and the monitoring of this finite resource (Zamrudi et al., 2019). The building of smart grids is a driving force behind energy efficiency and economic growth (Han et al., 2017).

2.1 What are smart meters?

As part of the modernization of the electricity infrastructure, smart meters are currently being deployed in various countries, including South Africa. Smart grids utilizing smart meter technology will modernize existing grids with bidirectional communication and pervasive communication capabilities for the smart generation, distribution, management, and consumption of electricity (Guo et al., 2015). Atkins (2014) defines a smart meter as

> an electric meter that measures energy consumption data over specified intervals, has two-way communications capability, stores metering data in registers, supports a variety of tariffs (e.g., time of use, inclined block, maximum demand, free basic electricity) which can be remotely updated, can switch attached loads on command and interfaces to data concentrators.

Smart meter technology simplifies the meter reading process for energy utilities and enables new services, flexible tariffs, and demand response programs in the context of the smart grid. As the electricity grid becomes "smarter", it gains many new data collection, communication, and information sharing capabilities related to energy usage, and the related technologies in turn introduce new challenges that were not associated with the traditional system.

2.2 Smart meter adoption in European countries and the USA

Although smart meters have benefits for both utility providers and consumers, there is still some resistance to the adoption of this technology. This section identifies consumer-centric factors that have impacted smart meter adoption, with specific reference to the USA and Europe as pioneers.

2.2.1 The USA

Jay et al. (2019) found that people generally support smart meter installation in the United States of America. This study included participants from 17 states with high smart meter installation rates. The results showed that privacy concerns, experience with privacy violations, and optimism in regard to accepting new technologies have an impact on consumer support for smart meter installations. They further suggested that communication to the public regarding privacy and economic benefits need to be clear to enhance adoption. In addition, Jay et al. (2019) also pointed out that social norms and technology readiness factors have some impact on the level of support for smart meter installations. Based on the Ithaca, New York study by Bugden and Stedman (2018), familiarity and climate change risk perceptions are the two factors that have the greatest impact on the acceptance of smart meters. S. Zhou (2021) discovered that dynamic pricing had a small impact on smart meters' penetration compared to supportive policies that facilitate the implementation and use of smart meters.

2.2.2 European countries

In their study of 23 European countries, Faruqui et al. (2010) found that the quantification of environmental benefits, transparency, and financial rewards based on low flat tariffs were among the factors that were proposed to boost smart meter acceptance. Based on a Dutch use case, Cuijpers and Koops (2013) found that privacy was a major driver in the acceptance of smart meters. They discovered that electric energy consumption intrudes on the privacy of personal life; therefore, they proposed a careful balance between smart metering and privacy protection.

S. Zhou and Brown (2017) studied four European countries, namely Finland, Sweden, the Netherlands, and Germany. The success stories of rolling out smart meters in Finland and Sweden were mainly based on regulatory mandates, positive financial regulations for the distributed system operators (DSOs), and policies to enhance social acceptance. From their perspective, the proper functioning of DSOs is vital to smart meter acceptance, as they deal with all operations, including the installation of smart meter devices, accurate billing and reading, marketing, data security and protection, and the authorization of third-party access to customer data. In the Netherlands, rollout resistance was caused by privacy and data security concerns leading to low social acceptance. In Germany, privacy was a major concern that led to low acceptance.

McKenna et al. (2012) propose addressing consumer privacy concerns early on to avoid a delay in smart meter deployment in their United Kingdom research study. In addition, Rausser et al. (2018) also suggest that economic benefit through a reduction in the cost of smart meters and other related costs could also increase smart meter adoption.

In summary, the consumer-centric factors identified as enhancing smart meter acceptance in the USA were privacy concerns, experience with privacy violations, optimism in regard to accepting new technologies, familiarity, smart meter supportive policies and social norms, and climate change risk perceptions. The quantification of environmental benefits, transparency, financial rewards, regulatory mandates, positive financial regulations for the DSOs, privacy and data security concerns, economic benefits, and policies that enhance social acceptance were identified as such factors in European countries. Although privacy concerns, economic or financial rewards, and environmental benefits or climate risk perceptions were found to be common consumer-centric factors in both Europe and the USA, some country-specific factors were also identified. Familiarity, technology readiness, optimism in regard to accepting new technology and social norms and were specific to the USA, while regulatory mandates, financial regulations of the DSOs, and policies for social acceptance were found to have aided in smart meter implementations in Europe.

Since the level of smart meter installation in South Africa is in its infancy, most municipalities are still battling to get smart meters implemented and functioning to realize their benefits. This is evident in reported cases of incorrect billing and poor customer support. Therefore, this study needs to identify consumer-centric factors that can enhance smart meter acceptance in South Africa, as some of the factors outlined in the USA and Europe may not directly fit South Africa's geo-economic and political landscape.

3 THEORETICAL FRAMEWORK: THE TECHNOLOGY ACCEPTANCE MODEL

As the decision to accept and use a new technology depends on uncertain benefits and uncertain costs, understanding the factors affecting choice is vital for both technology innovators and policy makers in relation to future technology developments (World Economic Forum, 2016). Theories that predict how a user comes to accept and use a specific technology have been dealt with extensively in past research, as is evident in Table 1. These theories suggest a number of constructs that influence a user's decisions about how and when they will use a new technology. TAM was chosen based on its focus on user acceptance and use (Table 1) and an extension of the TRA and TPB (Miltgen et al., 2013a). As for the UTAUT, it focuses on organisational technology acceptance (Venkatesh et al., 2012).

Table 1: Technology and user acceptance theories

Model	Constructs	Definitions	Source
Privacy calculus theory	Perceived ease of use Perceived usefulness Relevant social groups Institutional privacy assurance Perceived privacy risks	The Privacy Calculus Theory argues that a consumer's ability to take risks (disclosure of personal information) is influenced by the consumer's perception of benefits against risks (the calculus).	Morosan and DeFranco (2015) Sun et al. (2015) James et al. (2015) Keith et al. (2013) Dinev et al. (2006)
Theory of Reasoned Action	Attitude Subject norms	The Theory of Reasoned Action suggests that a person's behaviour is determined by a person's intention to perform the behaviour and that this intention is, in turn, a function of a person's attitude toward the behaviour and a person's subjective norm.	Fishbein and Ajzen (1975) Vallerand et al. (1992) Rehman et al. (2003)
Theory of Planned Behaviour	Attitude Subject norms Perceived behavioural control	An individual's behaviour is driven by behaviour intentions, where behaviour intentions are a function of three determinants: an individual's attitude toward behaviour, subjective norms and perceived behavioural control. The concept was proposed by Ajzen in 1991 to improve on the predictive power of the theory of reasoned action by including perceived behavioural control.	Rehman et al. (2003) Cheon et al. (2012) L. E. Davis et al. (2002) Bamberg et al. (2003) Notani (1998) Lynne et al. (1995) Ajzen (1985)
Technology Acceptance Model	Perceived usefulness External variables Perceived ease of use Attitude towards behavioural intention	The Technology Acceptance Model (TAM) is an information systems theory that models how users come to accept and use a technology.	Ajzen (1991) Sánchez-Prieto et al. (2015) Miltgen et al. (2013a)
Diffusion of Innovation Theory	Relative advantage Compatibility Trialability Observability Complexity	Diffusion research centres on the conditions which increase or decrease the likelihood that members of a given social system will adopt a new idea, product or practice.	F. D. Davis et al. (1989) Srivastava and Moreland (2012)
Unified Theory of Acceptance and Use of Technology	Performance expectancy Effort expectancy: The ease of use of the technologies Social factors Facilitating conditions Attitude Behavioural intentions	The Unified Theory of Acceptance and Use of Technology model aims to explain user intentions to use an information system and subsequent usage behaviour.	Miltgen et al. (2013b) Thomas et al. (2013) Martins et al. (2014) Venkatesh et al. (2012)

Based on the literature review, TAM has been useful in studying the intent to accept new technologies in a variety of contexts in Table 2. TAM is useful in different situations to explain information system acceptance and use. TAM has also been shown to be reliable and valid in many studies (Lai, 2017); thus, it provides a useful starting point. Therefore, TAM was used as a foundational framework for this study, contributing extensively to the research model. Though the TAM is a widely used model for studies about technology acceptance and use, it has some weaknesses in that it overlooks certain individual factors that could influence the choice to accept or reject a technology. Such individual factors can either provide additional variables to the TAM or provide an integrative view of the variables needed to explain or predict technology acceptance (S. Chen et al., 2011). Micheni et al. (2013) posit that, although the TAM provides valuable insights that focus mainly on the determinants of intention, it does not predict how perceptions are formed and how they can be manipulated to enhance user acceptance and increase technology usage. In this study, the motives that drive the acceptance of smart meter technology, in reality, may introduce additional significant constructs such as trust (Gefen et al., 2003; Wu & Chen, 2005), privacy concerns (Keith et al., 2013; T. Zhou, 2011), price value, and facilitating conditions (Venkatesh et al., 2012). These additional constructs cannot be explained or explicitly dealt with in the TAM or individually in other acceptance models.

As has been suggested above, TAM could not adequately cover all the consumer-centric factors that are relevant for smart meters; therefore, an extended TAM was developed for smart meters for this study. The next section provides an extended TAM for smart meters.

4 EXTENDED TECHNOLOGY MODEL FOR SMART METERS

Through an extensive review of the literature, TAM was identified as the seminal model for modelling behavioural intention to accept and use smart meters. In applying the TAM to this study, behavioural intention was defined as a consumer's indication of his/her readiness to accept and use smart meter technology (Miltgen et al., 2013a). The TAM suggests that the intention towards a behaviour can be predicted from perceived ease of use and perceived usefulness that is governed by a consumer's attitude towards using a technology, such as a smart meter, as is the case in this study.

4.0.1 Attitude

Attitude is expected to have a direct positive relationship with behavioural intentions to accept smart meter technology. According to Fishbein and Ajzen (1975), attitude is defined as the degree to which the performance of the behaviour is positively or negatively valued by an individual. In support, Miltgen et al. (2013a) suggest that the attitude towards performing a behaviour is jointly influenced by perceived usefulness and perceived ease of use. This joint relationship confirmed by Miltgen et al. (2013a), is widely accepted and verified by several past studies. As hypothesised by the TAM (Miltgen et al., 2013a), if a consumer has a positive

Table 2: Past research studies that have used the Technology Acceptance Model

Application domain	Research study context	Source
Banking	Customer acceptance of internet banking	Maduku (2013)
	Adoption of internet banking	Lee (2009)
	Adoption of Mobile Money Services	Micheni et al. (2013)
	Mobile banking	Gu et al. (2009)
Government services	Electronic toll collection service	C. Chen et al. (2007)
	Hospital information system acceptance	Lu and Gustafson (1994)
	End users' reactions to health information technology	Holden and Karsh (2010)
Commerce	Online trading	Gefen et al. (2003)
	Consumer acceptance of online auctions	Holden and Karsh (2010)
	E-commerce acceptance	Roca et al. (2009) and Stern et al. (2008)
	User acceptance of world-wide web	Gefen and Straub (2000)
Education	Examining faculty use of Online Learning Systems	Moon and Kim (2001)
	Understanding academics' behavioural intention to use Learning Management Systems	Fathema et al. (2015)
	E-learning attitudes	Alharbi and Drew (2014)
Privacy and security	Biometric applications	Holden and Karsh (2010)
	User acceptance of radio-frequency identification (RFID)	Park (2009)
Geography and environmental service	Consumer acceptance of location-based services in the retail environment	T. Zhou (2011)
	Investigating the impact of privacy concerns on user adoption of location-based services	Muller-Seitz et al. (2009)
General application	User acceptance of interface agents in daily work	Uitz and Koitz (2013)
	Understanding of self-service technologies	Serenko et al. (2007)
	Perception about the use of electronic mail	S. Chen et al. (2009)

attitude towards using smart meter technology, he or she is more likely to have an intention to use a smart meter. Therefore, the following relationship was tested:

> H1: Consumers with a positive attitude towards smart meter technology will have a positive behavioural intention to use smart meter technology.

Though the TAM provided the base on which the structural theory of the proposed research model was established for this study, it has certain weaknesses in overlooking other individual constructs that could influence the choice to accept or reject a technology (S. Chen et al., 2011). Consequently, other individual factors were included to provide the interrogative view needed to explain or predict technology acceptance as it relates to this study about smart meter technology implementation.

https://doi.org/10.18489/sacj.v33i2.909

4.0.2 Perceived usefulness

Perceived usefulness is expected to influence both attitude and behavioural intention positively to accept and use smart meter technology. Perceived usefulness is defined as the prospective user's subjective probability that using a specific application will increase his or her job performance within an organisational context (Miltgen et al., 2013a). In the context of this study, the intention to accept smart meter technology is expected to be linked to benefits that can be derived from using it. Based on previous studies, perceived usefulness has been found to be a strong determining factor in predicting behavioural intention (Miltgen et al., 2013a; Tan et al., 2012; Venkatesh et al., 2012). Similar to the TAM hypothesis, perceived usefulness is most likely to encourage a consumer to derive more benefits from managing electricity using a smart meter. Therefore, the following hypothesis was tested:

> H2: Consumers with higher perceived usefulness will have a positive behavioural intention to use smart meter technology.

Apart from perceived usefulness influencing behavioural intentions, Miltgen et al. (2013a) note that there is empirical evidence that shows perceived usefulness to have an influence on attitude towards using a technology. Taking the findings of Miltgen et al. (2013a) into consideration, this study posits that those benefits derived from using a smart meter will impact positively on consumer perception of how favourable smart meter technologies are. Therefore, the following hypothesis was tested:

> H3: Consumers with higher perceived usefulness will have a positive attitude towards the use of smart meter technology.

4.0.3 Ease of use

In this study, perceived ease of use is expected to have a positive influence on both perceived usefulness and attitude. Perceived ease of use is defined as the degree to which a user perceives that the effort required to use a particular technology will be minimal (Miltgen et al., 2013a; Ten Kate et al., 2010). The structural model of this study predicted that the ease-of-use construct has a direct relationship with both perceived ease of use and attitude. This notion is based on prior studies (Miltgen et al., 2013a; Venkatesh et al., 2012). For the purposes of this study, it is postulated that if a consumer believes that using a smart meter is effortless, they are more likely to exhaust all the benefits that can be derived from using it and, when realising the maximum benefits, they will have a positive attitude towards using the system. Considering the findings of prior studies, if a consumer finds it easy and effortless to use a smart meter, he or she is more likely to use it and have a positive attitude towards smart meter technology (Ma et al., 2018). Therefore, the following hypothesis is tested:

> H4: Consumers with higher ease of use will have a significantly positive relationship with perceived usefulness towards the use of smart meter technology.

H5: Consumers with higher ease of use will be significantly positive towards the attitude to use smart meter technology.

4.1 Technology Acceptance Model integrated constructs within smart meters

Since the TAM was found to be inadequate–and did not provide all the relevant factors that might influence customer behaviour–supplementary factors that might have relevance in predicting the behavioural intention to use smart meters were added to the model developed for this study. Sections 4.1.1 to 4.1.4 consider these supplementary factors and the proposed hypotheses that were investigated in relation to these factors.

4.1.1 Perceived value

Perceived value is expected to have a positive and direct influence on behavioural intention to accept smart meter technology. Perceived value can be defined as a consumer's overall subjective evaluation of the utility of a product or service mainly based on the trade-off between perceived benefits (utility) and perceived sacrifices or cost (Hazen et al., 2015; Zeithaml, 1988). As found in other studies, individuals are more likely to adopt a new technology if the perceived benefits outweigh the cost of acquiring or using the system (Sun et al., 2015; van Ittersum et al., 2006; Xiong, 2013). Perceived value can be expressed as an equation:

$$\text{Perceived value} = \frac{\text{Benefits (functional benefits + emotional benefits)}}{\text{Costs (monetary costs + time costs + energy costs + privacy risks)}} \quad (1)$$

In agreement, Xiong (2013) suggest that consumers tend to value mobile banking more if the benefits of using it are greater than the monetary cost and privacy risk to be incurred. Therefore, if consumers believe that using smart meter technology is valuable, they are more likely, in turn, to have positive intentions towards accepting and using it. The following hypotheses related to perceived value were tested:

H6: Customers with higher monetary cost will have a negative effect on the perceived value of using smart meter technology.

H7: Consumers with higher privacy/perceived risk will have a negative influence on the perceived value of smart meter technology.

H8: Consumers with higher perceived value will have a positive behavioural intention to use smart meter technology.

H9: Monetary cost positively affects the perceived usefulness of using smart meter technology.

H10: Perceived usefulness will have a positive effect on the perceived value of using smart meter technology.

4.1.2 Facilitating conditions

Facilitating conditions are expected to have a positive influence on both ease of use and behavioural intentions to accept and use smart meters. According to Ghalandari (2012) and Venkatesh et al. (2012), facilitating conditions can be defined as the extent to which consumers perceive that the resources and support required to perform a behaviour are available. In the context of this study, resources and support could be in the form of awareness campaigns, educational workshops and demonstration sessions on how to use the smart meters in order to enhance effortless use of the system. As suggested in some studies, facilitating conditions can influence intention to accept a technology both directly and indirectly through attitude (Ajzen, 1991; Ghalandari, 2012; Moon & Kim, 2001; Venkatesh et al., 2012). In contrast, other studies suggest that facilitating conditions are found to influence ease of use as compared to their influence on attitude (Moon & Kim, 2001; Venkatesh et al., 2012). Hence, if a consumer believes that facilitating conditions make it easy to operate a smart meter without much effort, he or she is likely to have the intention to accept smart meter technology. Therefore, the following two hypotheses were tested:

> H11: Consumers with high levels of facilitating conditions will have a significant relationship with ease of use.

> H12: Consumers with high levels of facilitating conditions would have a positive behavioural intention to use smart meter technology.

4.1.3 Trust in technology

Trust in technology is another of the additional independent constructs incorporated into the model used in this study. Trust in technology is expected to have a positive influence on both attitude and behavioural intention to accept and use smart meter technology. Prior research has shown that trust has been a critical factor in predicting the intention to accept and use technology (Dinev & Hart, 2006; Gefen et al., 2003; Joubert & van Belle, 2013; T. Zhou, 2011). In general, Mayer et al. (1995) define trust as the willingness of a user to be vulnerable to the actions of another party based on the expectation that the other party will perform a particular action important to the truster, irrespective of the ability to monitor or control that other party. Based on the findings of other studies, trust becomes a key driver for intention and use of online systems owing to its relevance in dealing with uncertainty and risk vulnerability transactions (Doney & Cannon, 1997; Gefen et al., 2003; Thatcher et al., n.d.). McKnight et al. (2002) further suggest that institutionally based trust (referred to as trust in technology) also has an impact on attitude toward accepting and using a particular technology. Hence, if a consumer perceives that the smart meter environment, including supporting structures and regulations, makes the environment feel safe, he or she is likely to have a positive attitude towards the system, which in turn will lead to an intention to accept and use smart meter technology and a belief and trust in the technology. Therefore, the following three hypotheses were tested:

https://doi.org/10.18489/sacj.v33i2.909

H13: Consumers with higher trust in technology will have a positive perceived usefulness towards the use of smart meter technology.

H14: Consumers with higher trust in technology will have a positive attitude toward the use of smart meter technology.

H15: Consumers with higher trust in smart meter technology will have a significant positive relationship with behavioural intention to use smart meter technology.

4.1.4 Social norms

Social norms are another independent construct added to the modified TAM, which is expected to relate positively to behavioural intention to accept smart meter technology. The social norms construct is described as social pressure placed on the consumer to perform or not to perform a behaviour (Cuijpers & Koops, 2013; Fishbein & Ajzen, 1975). While most studies concur, that social norms have an influence on behavioural intentions (Beldad & Hegner, 2018), some TPB studies emphasise that the assertion holds when technology acceptance is mandatory. In this study, the influence of friends, family, colleagues, and other social groups have a positive impact on the consumer's intention to accept smart meter technology. Hence, the following hypothesis was tested:

H16: Consumers with higher social norms will have a positive behavioural intention to use smart meter technology.

After the specification of the research model structural theory, all the constructs used were classified into two groups, namely exogenous constructs and endogenous constructs. This was done as part of the process of developing the structural model. The exogenous and endogenous constructs are both represented in the proposed research model below.

4.2 Proposed research model

In this study, a path diagram was developed of the proposed structural model to represent the structural theory. Facilitating conditions, social norms, monetary cost and privacy risk were classified as exogenous constructs. As suggested by Hair et al. (2010), exogenous constructs are independent constructs that are used to predict other constructs in the structural model; hence, they are not hypotheses that can be tested on them. Therefore, as per the path diagram in Figure 1, they are not single arrows that are supposed to enter them; rather, the arrows that came from them enter endogenous constructs. Hair et al. (2010), Hair et al. (2017) refer to endogenous constructs as the constructs in the structural model that are determined by other constructs (exogenous constructs). They further emphasise that endogenous constructs are usually the result of the hypotheses to be tested and verified. Therefore, in order to test the proposed structural relationships in the proposed structural model, structural equation modelling was used to test all 16 hypotheses. Figure 1 shows all the hypotheses that were tested in the structural model.

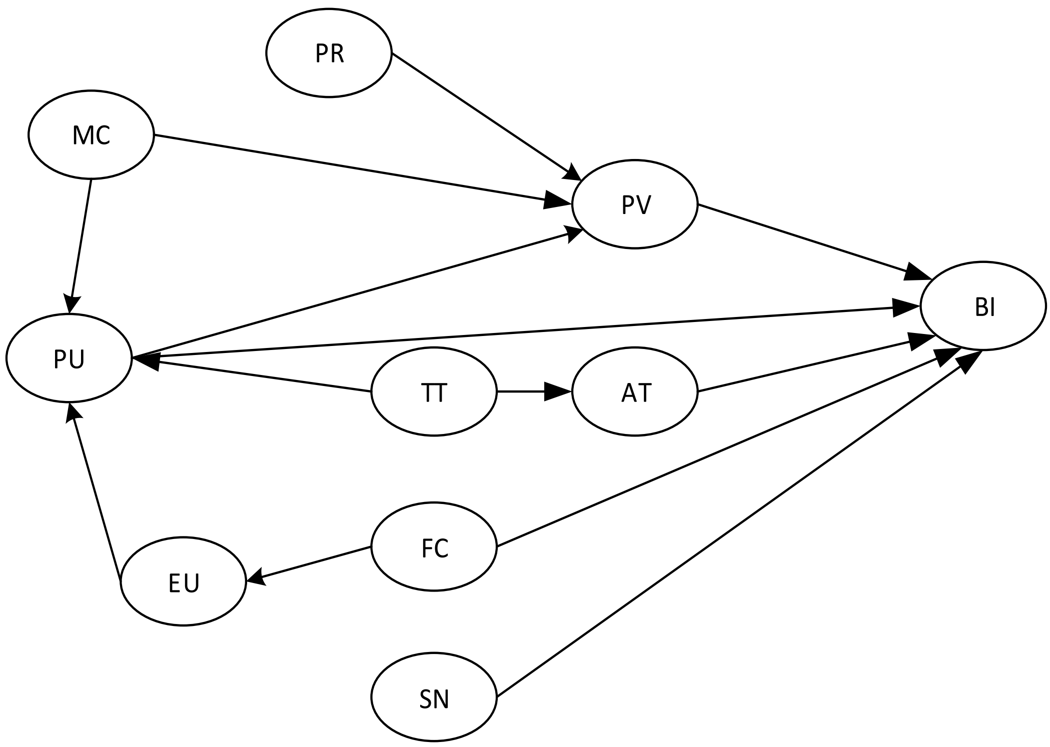

Figure 1: Proposed research structural model
BI: Behavioural intention, AT: Attitude, PV: Perceived value, PR: Privacy risk, MC: Monetary cost, PU: Perceived usefulness, EU: Ease of use, TT: Trust in technology, FC: Facilitating conditions, SN: Social norms

5 DATA COLLECTION METHODOLOGY

The quantitative data for this research study were collected in two ways. Online and manual data collection methods were employed. The online data collections mainly targeted those we could not have physical access to in various municipalities and towns, while the manual data collection was focused on those where we had easy physical access, like residents in Pretoria and Johannesburg. The two-research assistant chosen were trained and inducted in terms of the research data collection procedures in order to enhance data collection consistency. Both manual and online data collection was only targeted at those within South African borders.

The survey of smart and non-smart meter consumers was chosen to test the predicted constructs and their relationships within the model. Since smart meter technology implementation is still in its infancy in South Africa (Smart Energy International, 2019), the questionnaire included a section with an infographic of the smart meter display, and a brief description and illustration of smart meter technology to the participants in order to establish a common understanding among all respondents.

6 PARTICIPANTS AND TARGETED SAMPLE

The survey sample consisted of residential consumers within South Africa. Consumers were targeted in municipalities that have started implementing smart meters, such as City of Tshwane, City of Johannesburg, City of Cape Town, and City of Ekurhuleni. As such, both smart meter and non-smart meter users were included. Participants were approached based on being employed and thus likely to be responsible or co-responsible for energy decisions in their household. South Africa was chosen to be a test bed for the investigation in order to identify the consumer-centric factors that might be applicable in an African context since most previous research studies were either Eurocentric or American-centric (Smart Energy International, 2019).

Overall, 768 respondents completed the questionnaire. Thereafter, the gathered questionnaires were screened and those not completely filled were eliminated from the final analysis using the list-wise missing data handling. After removing 63 questionnaires from the 768, the result in response rate was 88.1%. The final sample size was 705 respondents from the target South African electricity population at the time of study of 31 million, which was considered acceptable, according to Hair et al. (2010) and Krejcie and Morgan (1970). The ages of survey participants ranged from 18–50 and above, with 54.15% males and 45.84% females. For more details, the characteristics of the targeted sample demographics are presented in Table 3.

7 MEASUREMENT DEVELOPMENT

Although TAM is widely used for researching technology acceptance and use, its applicability was found to be inadequate in explaining customer-centric factors in smart metering. The questionnaire was developed to measure the constructs in the research model. The constructs in the research model were trust in technology, perceived usefulness, perceived ease of use, perceived privacy risk, monetary cost, perceived value, attitude, facilitating conditions, social norms, and behavioural intention.

The behavioural intention, attitude, perceived usefulness, and perceived ease of use were measured using an adapted model from Belanche et al. (2012), Kaushik et al. (2015), Venkatesh et al. (2012), and Xu et al. (2011), while some were self-developed in the context of smart metering. The items to measure the facilitating conditions construct were adapted from Martins et al. (2014) and Venkatesh et al. (2012). The perceived value items were adapted from Agarwal et al. (2007) and Kim et al. (2009). The perceived privacy risk was adapted from Krasnova and Veltri (2010), Taneja et al. (2014), and Venkatesh et al. (2012), and the items for trust in technology were modified from Belanche et al. (2012) and Thatcher et al. (n.d.). Within some constructs, items were self-developed and reworded in the smart meter context of the study. All the constructs were measured using a 7-point Likert scale, with one (1) meaning 'strongly agree' and seven (7) meaning 'strongly disagree'. For more detail on all the constructs and their measurement items used in the research model, see the derivation of construct items in Appendix A and the final constructs in Appendix B.

https://doi.org/10.18489/sacj.v33i2.909

Table 3: Descriptive statistics

Gender	Frequency	Percentage
Male	358	54.16
Female	303	45.84
Missing data	44	24
Age		
18–25	100	15.27
26–35	195	29.77
36–45	181	27.63
46–50	96	14.66
> 50	83	12.67
Missing data	50	7.00
Education level		
No schooling	5	0.76
Has some schooling	21	3.20
Matriculated	84	12.80
Certificate	76	11.76
Diploma	164	25 00
Undergraduate (with Honours)	203	30.95
Masters and Doctorates	103	15.70
Missing data	49	6.95
Average annual income		
Less than R150 000	263	41.55
R150 000–R299 000	143	22.59
R300 000–R449 000	92	14.33
R450 000–R599 000	57	9.00
R600 000–R749 000	34	5.73
Above R900 000	23	3.63
Missing data	103	10.21
Smart users descriptive		
Current users	287	44.50
Non-users	358	55.50
Missing data	60	8.50
Smart meter service provider		
Tshwane Metropolitan	363	55.42
Eskom	89	13.59
Others	61	9.31
Johannesburg Metropolitan	49	7.48
Private companies	41	6.26
Ekurhuleni Metropolitan	26	3.97
Non-Metropolitan	10	1.53
Cape Town Metropolitan	9	1.37
eThekwini Metropolitan	7	1.07
Missing data	50	7.09
Total	705	100%

In addition to the construct measurements for the research model, smart user demographics information was also collected, mainly to gather information about the target sample population. Before the final data collection, three-stage pre-testing (15, 73 and 55 participants) of the questionnaire was conducted to address misunderstandings regarding some statements, words, ambiguity, flow, and the overall layout of the questionnaire. Some adjustments were made to the construct and construct items, which did not achieve recommended levels for items internal consistency, construct reliability, and validity (Hair et al., 2010).

Table 4: Recommended goodness-of-fit indices for the research model (Hair et al., 2010; Hair et al., 2017)

Name of category		Index name	Acceptable level
Chi Square (X^2)	X^2	Discrepancy Chi Square	n/a
Degrees of freedom	Df	Degrees of Freedom	n/a
Absolute fit	RMSEA	Root mean squared error of approximation	n/a
	GFI	Goodness of fit	> 0.9
	90% CI of RMSEA	90% confidence interval of root mean squared error of approximation	$0.03- > 0.08$
	RMR	Root mean residual	< 0.05
	SRMR	Standardised root mean residual	< 0.05
	Normed X^2	Normed Chi Square	$< 2.0- < 5.0$
Incremental fit	NFI	Normed fit index	> 0.90
	NNFI	Non-normed fit index	> 0.90
	CFI	Comparative fit index	> 0.90
	TLI	Tucker-Lewis index	> 0.95
Parsimonious fit indices	PNFI	Parsimonious normed fit index	> 0.95
	AGFI	Adjusted goodness-of-fit index	> 0.95

https://doi.org/10.18489/sacj.v33i2.909

Table 5: Goodness of fit for the proposed smart meter measurement model validity (Hair et al., 2010; Hair et al., 2017)

Name of category	Index	Recommended level	Measurement model values	Comments
Chi Square (X^2)	X^2	n/a	2658.66	
Degree of freedom	Df	n/a	944	
Absolute fit	RMSEA	< 0.06 or 0.07	< 0.05	Achieved level
	90% confidence interval RMSEA	$0.03 - 0.08$	$0.04 - 0.05$	Achieved level
	SRMR	< 0.05	0.02	Achieved level
	Normed X^2	< 2.0 and < 5.0	2.8	Achieved level
Incremental fit	CFI	> 0.90	> 0.97	Achieved level
	TLI	> 0.95	> 0.96	Achieved level

Table 6: Construct correlation matrix
EU: Ease of use, FC: Facilitating conditions, PU: Perceived usefulness, PR: Privacy risk, PV: Perceived value, TT: Trust in technology, SN: Social norms, AT: Attitude, BI: Behavioural intention.
Discriminant validity: AVE values $>$ Squared correlations.
Note: Values below the diagonal are correlation estimates among constructs, the diagonal element is constructed variance, and values above the diagonal are squared correlations.

Construct	EU	FC	PU	MC	PR	PV	TT	SN	AT	BI
EU	1.000									
FC	0.731	1.000								
PU	0.675	0.660	1.000							
MC	0.025	0.028	0.043	1.000						
PR	0.141	0.143	0.184	0.200	1.000					
PV	0.505	0.466	0.549	0.015	0.208	1.000				
TT	0.550	0.563	0.575	0.017	0.179	0.746	1.000			
SN	0.387	0.390	0.433	0.017	0.190	0.501	0.620	1.000		
AT	0.530	0.509	0.588	0.013	0.160	0.683	0.769	0.620	1.000	
BI	0.556	0.524	0.598	0.007	0.136	0.705	0.778	0.615	0.901	1.000

https://doi.org/10.18489/sacj.v33i2.909

Table 7: Summary of the standardised factor loadings, reliability and average extracted variance
Note: If the factor loading $>= 0.7$, AVE cut-off values > 0.5 and Cronbach alpha $>= 0.7$ then the constructs show convergent validity.

Construct	Factor loading	Cronbach alpha	Average extracted variance
Ease of use	0.904 0.944 0.904 0.848	0.944	0.811
Facilitating conditions	0.904 0.944 0.904 0.848	0.961	0.864
Perceived usefulness	0.904 0.944 0.904 0.848	0.972	0.896
Monetary cost	0.904 0.944 0.904 0.848	0.947	0.540
Privacy risk	0.904 0.944 0.904 0.848	0.949	0.549
Perceived value	0.904 0.944 0.904 0.848	0.945	0.543
Trust in technology	0.904 0.944 0.904 0.848	0.975	0.865
Social norms	0.904 0.944 0.904 0.848	0.971	0.849
Attitude	0.904 0.944 0.904 0.848	0.977	0.540
Behavioral intentions	0.904 0.944 0.904 0.848	0.981	0.600

https://doi.org/10.18489/sacj.v33i2.909

Table 8: Construct reliability: internal consistency and average extracted variance
If the AVE cut=off values > 0.5 and Cronbach alpha $>= 0.7$ the construct has achieved construct reliability.

Construct	Average item test correlation	Cronbach alpha	Average extracted variance
Ease of use (EU)	0.926	0.944	0.811
Facilitating conditions (FC)	0.945	0.961	0.864
Perceived usefulness (PU)	0.962	0.972	0.896
Monetary cost (MC)	0.930	0.947	0.54
Privacy risk (PR)	0.931	0.949	0.549
Perceived value (PV)	0.926	0.945	0.543
Trust in technology (TT)	0.943	0.975	0.865
Social norms (SN)	0.934	0.971	0.849
Attitude (AT)	0.967	0.977	0.608
Behavioural intentions (BI)	0.956	0.981	0.600

8 HYPOTHESIS TESTING OF THE STRUCTURAL MODEL

The assessment of the structural model validity was not sufficient to confirm the structural relationships between constructs; therefore, individual parameter estimates of the proposed model were measured to establish whether the parameter estimates were significant or not. The SEM results in Table 9 were processed using STATA version 13; the standardised estimates were assessed based on the coefficient β value and the p-value. According to Hair et al. (2010), a significant parameter estimates entails that the t-value must be greater than 1.96 and the p-value $<= 0.05$. Hair et al. (2010) also emphasise that a significant parameter estimate value must be > 0 for positive relationships and < 0 for negative relationships whereas the p-value must be < 0.01 in both instances.

The sections below will discuss in detail how the proposed hypotheses were either accepted or rejected. For ease of presentation of the results, the discussion is organised into categories of hypotheses with similar exogenous constructs.

8.1 Facilitating conditions → Ease of use

The structural relationship between facilitating conditions and ease of use of smart meters was examined. The results indicate that there was a positive significant relationship between facilitating conditions and ease of use ($\beta = 0.867, t = 78, p < 0.001$). Thus, H11 was confirmed and accepted. This means that participants agree and confirm that facilitating conditions play an important role in assisting consumers to be able to use smart meters easily. These results concur with previous research (Venkatesh et al., 2012) in that, as individuals are provided with information about how to use smart meter technology, it further enhances their cognitive level. This means that when facilitating conditions increase, the ease of use also increases.

Table 9: Structural parameter estimates for the smart meter model
Significance level: $p < 0.05$*, $p < 0.01$**, $p < 0.000$***, $p > 0.1$ Rejected

Hypothesis	Paths	Std. Err	Coefficient (β)	t-value	Hypothesis
Facilitating conditions→Ease of use	FC→EU	0.110	0.867	78.30	Supported***
Ease of use→Perceived usefulness	EU→PU	0.030	0.566	18.83	Supported***
Monetary cost→Perceived usefulness	MC→PU	0.021	0.080	3.75	Supported***
Trust in technology→Perceived usefulness	TT→PV	0.031	0.362	11.57	Supported***
Perceived usefulness→Perceived value	PU→PV	0.022	0.687	30.02	Supported***
Monetary cost→Perceived value	MC→PV	0.029	-0.126	-4.26	Supported***
Privacy risk→Perceived value	PR→PV	0.032	0.2276	7.07	Supported**
Ease of use→Attitude	EU→AT	0.379	0.055	1.47	Rejected
Perceived usefulness→Attitude	PU→AT	0.039	0.202	5.11	Supported***
Trust in technology→Attitude	TT→AT	0.028	0.689	24.30	Supported***
Perceived usefulness→Behavioural intention	PU→BI	0.028	0.0380	1.34	Rejected
Perceived value→Behavioural intention	PV→BI	0.027	0.111	4.01	Supported***
Attitude→Behavioural intention	AT→BI	0.031	0.699	22.38	Supported***
Facilitating conditions→Behavioural intention	FC→BI	0.025	0.020	0.80	Rejected
Trust in technology→Behavioural intention	TT→BI	0.037	0.103	2.76	Supported*
Social norm→Behavioural intention	SN→BI	0.023	0.058	2.49	Supported

8.2 Perceived usefulness, privacy risk and monetary cost → Perceived value

The structural relationship between perceived usefulness and perceived value of smart meters was also examined. The results indicated that perceived value was positively and significantly related to perceived usefulness ($\beta = 0.688, t = 30.02, p < 0.001$), thus confirming hypothesis H10. Benefits that come from the use of smart meter technology can make people value the use of smart meters. The relationship between privacy risk and perceived value of smart meters was also examined and it was found that privacy risk was significantly associated with perceived value ($\beta = 0.223, t = 7.07, p < 0.001$). Though these results confirmed hypothesis H7, the strength of the relationship was weak ($\beta = 0.223$). These results may suggest that, if there are many perceived risks towards the use of smart meters, this will tend to impact negatively on the use of smart meter technology. Furthermore, the relationship between monetary cost and perceived value of smart meters was also examined and found to be significant ($\beta = -0.127, t = -4.26, p < 0.001$). Hypothesis H6 was confirmed, even though the relationship between the two constructs was negative and weak with a coefficient ($\beta = -0.127$).

The results indicate that if there is high monetary cost towards the use of smart meters, people tend to become negative towards the use of smart meter technology. In summary, the results showed that perceived usefulness ($\beta = 0.688$) was the best predictor for perceived values in comparison to privacy risk ($\beta = 0.223$) and monetary cost ($\beta = -0.127$), respectively.

https://doi.org/10.18489/sacj.v33i2.909

8.3 Ease of use, monetary cost and trust in technology $\rightarrow$ Perceived usefulness

The structural relationship between ease of use and perceived usefulness of smart meters was found to be significant ($\beta = 0.567, t = 18.83, p < 0.0001$), thus verifying hypothesis H4. The structural relationship between monetary cost and perceived usefulness of smart meters was found to be positive and significant ($\beta = 0.080, t = 3.75, p < 0.0001$), thus verifying hypothesis H9. The results indicated that participants believed that they derived more benefits from the use of smart meters if the cost towards the use of smart meters was less. Furthermore, trust in technology and perceived usefulness were also examined and it was found that trust in technology and perceived usefulness of smart meters was positive and significant ($\beta = 0.362, t = 11.57, p < 0.0001$), thus verifying hypothesis H13. Participants believe that if technology is trustworthy, dependable and reliable it will, in turn, improve the perceived benefit of using smart meters.

In summary, the results showed that ease of use ($\beta = 0.567$) was the best predictor of perceived usefulness in comparison with trust in technology ($\beta = 0.362$) and monetary cost ($\beta = 0.080$). The participants think that the ease of use of smart meters does have an influence on their perceived usefulness of smart meters. The results also concur with the findings of previous studies (Miltgen et al., 2013a; Tan et al., 2012).

8.4 Perceived usefulness and trust in technology and ease of use $\rightarrow$ Attitude

The relationship between perceived usefulness and attitude towards smart meters was found to be positive and significant ($\beta = 0.202, t = 18.83, p < 0.0001$). Consequently, hypothesis H3 was verified and confirmed. These results suggest that, as more benefits are derived from smart meter perceived usefulness, people will eventually change their attitude towards smart meters in managing electricity usage. The relationship between trust in technology and attitude was also examined and was found to be significant ($\beta = 0.690, t = 24.30, p < 0.0001$), thus hypothesis H14 was verified and accepted. The structural relationship between ease of use and attitude towards smart meters was examined and found not to be significant ($\beta = 0.055, t = 1.47, p > 0.142$), thus hypothesis H5 was verified and rejected. This might suggest that participants think that ease of use of smart meters does not have much influence on their attitude toward using smart meters.

In summary, of the three constructs that influence attitude towards smart meters, only trust in technology and perceived usefulness were found to be significant, with trust in technology being the best construct to predict the consumer's attitude towards the use of smart meters. Ease of use was rejected as a construct that influenced attitude.

8.5 Perceived value, attitude, social norms, trust in technology and perceived usefulness → Behavioural intention

The relationship between attitude and behavioural intention to accept smart meters was significant ($\beta = 0.70, t = 22.38, p < 0.001$), thus hypothesis H1 was verified and accepted. The results suggest that positive attitude towards smart meter technology can be influenced by trust in technology and perceived usefulness which, in turn, have a significant influence on the behavioural intention to use and accept smart meter technology. The relationship between perceived usefulness and behavioural intention to accept smart meter technology was found not to be significant ($\beta = 0.038, t = 1.34, p > 0.01$), thus hypothesis H2 was verified and rejected. The relationship between perceived value and behavioural intention towards use and acceptance of smart meters was found to be significant ($\beta = 0.11, t = 4.07, p < 0.001$), thus hypothesis H8 was verified and accepted. The more people see value in the use of smart meter technology, the more consumers will tend to accept and use smart meter technology.

The relationship between facilitating conditions and behavioural intention was found to be not significant ($\beta = 0.020, t = 0.80, p > 0.1$), thus hypothesis H12 was verified and rejected. The relationship between trust in technology and behavioural intention towards accepting smart meter technology was found to be positive and significant ($\beta = 0.103, t = 2.76, p < 0.05$), thus hypothesis H15 was verified and accepted. These results suggest that consumers are willing to accept smart meter technology when it is trustworthy, dependable and reliable. The relationship between social norms and behavioural intention to accept smart meters was found to be significant ($\beta = 0.058, t = 2.49, p < 0.01$), thus hypothesis H16 was verified and accepted.

Based on statistical analyses of both the measurement and structural models, documented in Table 5 and Table 10, respectively, this section summarises the structural model relationships that were evaluated as significant for this research study. The significant factors and the relationships between these are illustrated in Figure 2 below.

Table 10: Goodness-of-fit measure for the structural model (Hair et al., 2010; Hair et al., 2017)

Name of category	Index	Proposed structural model
Chi Square (ΔX^2)	ΔX^2	3199.66
Degree of freedom	Df	963
Probability	P	0.05
Absolute fit	RMSEA	0.057
	90% confidence interval RMSEA	0.055–0.06
	SRMR	0.06
	Normed X^2	3.32
Incremental fit	CFI	0.955
	TLI	0.952

https://doi.org/10.18489/sacj.v33i2.909

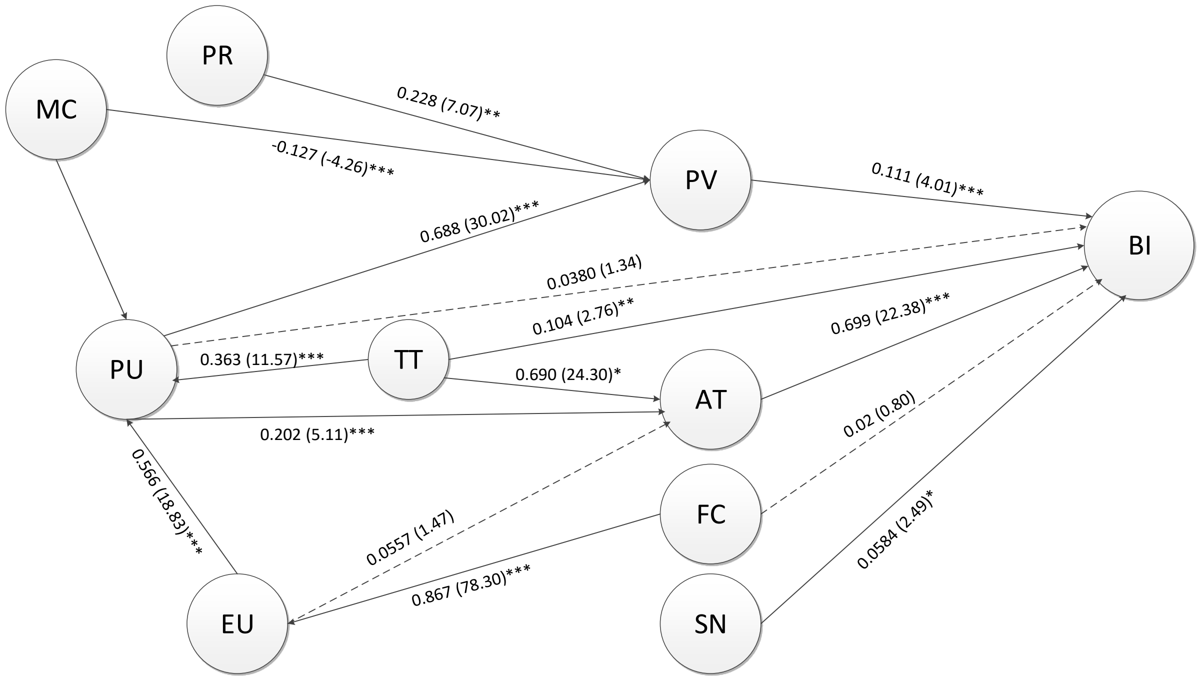

Figure 2: Final research model: proposed structural relationships
Note Numbers in brackets are t-values. The numbers outside the brackets are standardised path coefficients. Dotted arrow lines indicate that the hypothesis is not a significant path ($p > 0.1$) *. The solid arrow lines indicate that the hypothesis is significant ($p < 0.05$) *, ($p < 0.01$) ** and ($p < 0.001$) ***

9 DISCUSSION

The smart meter adoption and implementation within South Africa is still in its infancy (Smart Energy International, 2019). Hence, the identification of consumer-centric factors is important for the implementation of smart meters in South Africa. In order to better understand the consumer-centric factors involved in the implementation of smart meters, an extensive investigation was conducted into the factors that affect smart meter consumers. A survey was conducted of 768 participants, from both smart meter and non-smart meter users within South Africa. Then, the research identified a myriad of consumer-centric factors that affect the planning for smart meter implementation in South Africa.

From the smart technology acceptance perspective (Ponce-Jara et al., 2017), there has not been a great deal of research in the area; hence, the use of prior work did not assist much in the identification of factors that are important to smart meter technology. The results above show that 13 factors were found to be most important in the implementation of smart meters in South Africa (Table 9). From the findings, utility companies or smart meter service providers must consider these consumer-centric factors for better implementation of smart meters in South Africa.

As depicted in Table 9 and Figure 2, facilitating conditions were found to be significant when considering ease of use, while perceived ease of use appeared to be important for perceived usefulness from the perspective of the smart meter electricity consumer. These results were in line with other studies (Moon & Kim, 2001; Teo, 2009; Venkatesh et al., 2012). However, facilitating conditions did not seem to have any direct influence on attitude towards the use and acceptance of smart meter technology. This means that facilitating conditions must be considered in the implementation of smart meters as they can assist the smart meter users in the operation of smart meter technology and, in the process, allow the smart meter users to derive benefits from using smart meters.

Trust in technology, monetary cost and ease of use were found to be significant when considering the perceived usefulness of smart meter technology (Figure 2). This means that, in order to make the smart meter users benefit from using a smart meter, the smart meter needs to be trustworthy, affordable and easy to operate. Otherwise, if the trust in technology, monetary cost and perceived ease of use are not addressed properly, this might have an adverse impact on smart users to even consider using them. In the same light, perceived usefulness, and trust in technology (Orlando & Vandevelde, 2021) were found to be relevant in enhancing the attitude of potential smart meter consumers towards acceptance and use of smart meter technology.

While Beldad and Hegner (2018) suggest that social norms have impact on ease of use, and although this study found social norms to be significant, Teo (2009) found them to be insignificant as modern consumers are more independent and are not influenced by other people when making important decisions.

This research was conducted with an objective to generalise the findings to other pervasive computing technologies (smart technologies) in future. Evaluating the applicability of the research proposed model to other similar pervasive technologies like smart cities opens another door for future research. As the research was conducted only in South Africa as a country, conducting the same research in multiple countries within Africa can assist in evaluation of the extent to which the factors affecting smart meter technology acceptance and use can be generalised within the African context. Finally, as much as smart meters may bring better consumer management of electricity consumption, the comments from the data collection phase suggest that most participants were not aware of smart meters; hence, there is a need for more awareness campaigns and education to reduce smart meter implementation resistance.

10 THEORETICAL IMPLICATIONS

This research study has made several contributions to behavioural studies in a number of ways. Most significant is the contribution to knowledge about the implementation of pervasive computing technologies, such as smart meter acceptance and use in developing countries. This section discusses the theoretical research contributions that can emanate from the research study.

From the theoretical perspective, this research contributes most importantly to the pervasive technology domain with a rich empirical study that assists technology innovators, utility companies, and policy makers to understand the consumer-centric factors that might affect technology acceptance and use. Since most of the behavioural models and theories were developed a long time ago (such as the TRA, TPB and TAM (Miltgen et al., 2013a)), their application and use in identifying factors that influence smart technology acceptance and use might not be adequate or relevant. The rapid change in technology and user technology interactions has impacted significantly on the way technology is viewed. Hence, the proposed theoretical model developed in this study presents new empirical knowledge that can be refined further in technology acceptance and use studies in the future.

In addition, this study contributes to the model development literature and design. Various theoretical models were identified to assist in developing a competing model for this study. Based on an extensive literature review, the TAM was identified as a fundamental research framework. However, since smart meter technology is a new technological advancement, the TAM was found to be inadequate in helping to understand and identify factors that affect user behaviour to accept and use pervasive technology (Gefen et al., 2003; Wu & Chen, 2005; T. Zhou, 2011). Therefore, this study contributed to the model development literature by integrating other relevant variables to supplement the TAM inadequacy. In view of smart meter technology, the factors that were found relevant and incorporated into the TAM for the study model development were: privacy risk (PCT), facilitating conditions and monetary cost (UTAUT), social norms (TPB), trust, and perceived value. Consequently, the proposed model output in this study confirms a new contribution to the body of knowledge (Table 2).

Another theoretical contribution that emanated from this study is an enrichment of the African literature about smart meter technology acceptance modelling. From the literature review, it is evident that most smart grid and smart meter roadmaps and studies on technology adoptions acceptance, implementations and post-implementations have been conducted either in the United States of America or in European countries (Rausser et al., 2018; Teo, 2009). Considering that these are developed countries, which have better electricity grid infrastructure, better electricity policies, high-level technical skills, a high sense of environmental consciousness, and high privacy appetite, it must be acknowledged that the situation is different in developing countries like South Africa. It is, therefore, meaningful that the results of this study, which focuses on a South African situation, can contribute or can be used as a reference model in understanding the African smart meter consumer's perspective.

Finally, the results from this research can be adapted by other African countries for the planning of smart meter implementation as opposed to using American or European views, which are far from the African setting.

11 CONCLUSION

This study has improved the understanding of consumer-centric factors that must be considered in the implementation and acceptance of smart meters in South Africa. Since there is no specific technology acceptance model and theory available to explain in the behavioural intention to accept smart meter technology in general, the research relied on various models postulated. The following models and theories were used in this research: the Technology Acceptance Model, Privacy Calculus Theory, and the Unified Theory of Acceptance and Use of Technology; hence, leading to the complex research model depicted in Figure 1 and Figure 2.

The analysis of the research model allowed important interrelationships amongst consumer-centric factors identified and examined in the implementation of smart meters in South Africa. The findings showed that 10 factors, namely perceived ease of use, facilitation conditions, monetary cost, perceived risk, perceived value, trust in technology, social conditions, attitude, and behavioural intentions were identified to be important to smart meter consumers. In the process, it was revealed that perceived value, attitude, trust in technology, and social norms are important factors that directly impact the implementation of smart meters, whilst facilitation conditions, perceived ease of use, perceived usefulness, trust in technology, monetary cost, and privacy risk have indirect impacts on either attitude or perceived value, respectively (Figure 2). These findings were found to concur with the literature review: attitude (Miltgen et al., 2013b), perceived usefulness (Miltgen et al., 2013a), price value (Xiong, 2013), social norms (Beldad & Hegner, 2018), privacy risk (Xu et al., 2011), trust in technology (Belanche et al., 2012), facilitating conditions (Teo, 2009). Therefore, utility companies, technological innovators and designers should make sure that the smart meters are beneficial and valuable without compromising their easy in usability. On the other hand, policy makers, technology designers and vendors must ensure that smart meters are secure, trustworthy and reliable in providing electricity services to its smart meter consumers, with adequate support.

References

Agarwal, N., Wang, Z., & Xu, Y. (2007). Factors affecting 3G adoption: An empirical study. *PACIS 2007 Proceedings*, 256–277. http://www.pacis-net.org/file/2007/1205.pdf

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckman (Eds.), *Action-control: From cognition to behavior*. Springer.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Alharbi, S., & Drew, S. (2014). Using the technology acceptance model in understanding academics' behavioural intention to use learning management systems. *International Journal of Advanced Computer Science and Applications*, *5*(1), 143–155.

Atkins, P. (2014). Smart meters and smart-enabled prepaid meters project [Department of Economic Development and Tourism, Western Cape Government].

Bamberg, S., Ajzen, I., & Schmidt, P. (2003). Choice of travel mode in the theory of planned behavior: The roles of past behavior, habit, and reasoned action. *Basic and Applied Social Psychology*, *25*(10), 175–187. https://doi.org/10.1207/S15324834BASP2503_01

Bargh, J. A., & McKenna, K. Y. A. (2004). The internal and social life. *Annual Review of Psychology*, *55*(1), 573–590. https://doi.org/10.1146/annurev.psych.55.090902.141922

Barrett, J., Peters, G., Wiedmann, T. O., Scott, K., Lenzen, M., Roelich, K., Quéré, L., & C. (2013). Consumption-based GHG emission accounting: A UK case study. *Climate Policy*, *13*, 451–570. https://doi.org/10.1080/14693062.2013.788858

Bazilian, M., Welsch, M., Divan, D., Elzinga, D., Strbac, G., Howells, M., Jones, L., Keane, A., Gielen, D., Balijepalli, V. S. K. M., Brew-Hammond, A., & Yumkella, K. (2013). Smart and just grids for sub-Saharan Africa: Exploring options. *Renewable and Sustainable Energy Reviews*, *20*, 336–352. https://doi.org/10.1016/j.rser.2012.11.004

Belanche, D., Casado, L. V., & Flavián, C. (2012). Integrating trust and personal values into the Technology Acceptance Model: the case of e-government services adoption. *Cuadernos de Economía y Dirección de la Empresa*, *15*(4), 192–204. https://doi.org/10.1016/j.cede.2012.04.004

Beldad, A., & Hegner, S. (2018). Expanding the technology acceptance model with the inclusion of trust, social influence, and health valuation to determine the predictors of German users' willingness to continue using a fitness app: A structural equation modeling approach. *International Journal of Human–Computer Interaction*, *34*(9), 882–893. https://doi.org/10.1080/10447318.2017.1403220

Bugden, D., & Stedman, R. (2018). A synthetic view of acceptance and engagement with smart meters in the United States. *Energy Research and Social Science*, *47*, 137–145. https://doi.org/10.1016/j.erss.2018.08.025

Camarinha-Matos, L. (2016). Collaborative smart grids – A survey on trends. *Renewable and Ssustainable Energy Reviews*, *65*, 283–294. https://doi.org/10.1016/j.rser.2016.06.093

Chen, C., Fan, Y., & Farn, C. (2007). Predicting electronic toll collection service adoption: An integration of the technology acceptance model and the theory of planned behavior. *Transportation Research Part C: Emerging Technologies*, *15*(5), 300–311. https://doi.org/10.1016/j.trc.2007.04.004

Chen, S., Chen, H., & Chen, M. (2009). Determinants of satisfaction and continuance intention towards self-service technologies. *Industrial Management and Data Systems*, *109*(9), 1248–1263. https://doi.org/10.1108/02635570911002306

Chen, S., Li, S., & Li, C. (2011). Recent related research in technology acceptance model: A literature review. *Australian Journal of Business and Management Research*, *1*(9), 124–127.

Cheon, J., Lee, S., Crooks, S. M., & Song, J. (2012). An investigation of mobile learning readiness in higher education based on the theory of planned behavior. *Computers and Education*, *59*(3), 1054–1064. https://doi.org/10.1016/j.compedu.2012.04.015

Cuijpers, C., & Koops, B. (2013). Smart metering and privacy in Europe: Lessons from the Dutch case. In G. S., L. R., de Hert P., & P. Y (Eds.), *European data protection: Coming of age*. https://doi.org/10.1007/978-94-007-5170-5_1

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*(8).

Davis, L. E., Ajzen, I., Saunders, J., & Williams, T. (2002). The decision of African American students to complete high school: An application of the theory of planned behavior.

Journal of Educational Psychology, *94*(4), 810–819. https://doi.org/10.1037/0022-0663.94.4.810

Dileep, G. (2019). A survey on smart grid technologies and applications. *Renewable Energy*, *146*, 2589–2625. https://doi.org/10.1016/j.renene.2019.08.092

Dinev, T., Bellotto, M., Hart, P., Russo, V., & Colautti, C. (2006). Privacy calculus model in e-commerce—A study of Italy and the United States. *European Journal of Information Systems*, *15*(4), 389–402. https://doi.org/10.1057/palgrave.ejis.3000590

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, *17*(1), 61–80. https://doi.org/10.1287/isre.1060.0080

Doney, P. M., & Cannon, J. P. (1997). An examination of the nature of trust in buyer-seller relationships. *Journal of Marketing*, *61*(2), 35–51.

Eberhard, A., Gratwick, K., Morella, E., & Antmann, P. (2017). Independent power projects in sub-Saharan Africa: Investment trends and policy lessons. *Energy Policy*, *108*, 390–424. https://doi.org/10.1016/j.enpol.2017.05.023

Faruqui, A., Harris, D., & Hledik, R. (2010). Unlocking the €53 billion savings from smart meters in the EU: How increasing the adoption of dynamic tariffs could make or break the EU's smart grid investment. *Energy Policy*, *38*(10), 6222–6231. https://doi.org/10.1016/j.enpol.2010.06.010

Fathema, N., Shannon, D., & Ross, M. (2015). Expanding the technology acceptance model (TAM) to examine faculty ease of learning management systems (LMSs) in higher education institutions. *MERLOT Journal of online learning and teaching*, *11*(2), 210–232.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Addison-Wesley.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Inexperience and experience with online stores: The importance of TAM and trust. *IEEE Transactions on Engineering Management*, *50*(3), 307–321. https://doi.org/10.1109/TEM.2003.817277

Gefen, D., & Straub, D. W. (2000). The relative importance of perceived ease of use in IS adoption: A study of e-commerce adoption. *Journal of the Association for Information Systems*, *1*(8), 1–30.

Ghalandari, K. (2012). The effect of performance expectancy, effort expectancy, social influence and facilitating conditions on acceptance of e-banking services in Iran: The moderating role of age and gender. *Middle-East Journal of Scientific Research*, *12*(6), 801–807. https://doi.org/10.5829/idosi.mejsr.2012.12.6.2536

Gu, J. C., Lee, S. C., & Suh, Y. H. (2009). Determinants of behavioral intention to mobile banking. *Expert Systems with Applications*, *36*(9), 11605–11616. https://doi.org/10.1016/j.eswa.2009.03.024

Guo, C., Bond, C., & Narayanan, A. (2015). The adoption of new smart-grid technologies: Incentives, outcomes, and opportunities [Last accessed 10 Dec 2021]. https://doi.org/10.1088/1757-899X/199/1/012062

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th). Prentice Hall.

Hair, J. F., Matthews, L. M., Matthews, R. L., & Sarstedt, M. (2017). Pls-SEM or CB-SEM: Updated guidelines on which method to use. *International Journal of Multivariate Data Analysis*, *1*(2). https://doi.org/10.1504/ijmda.2017.10008574

Han, L., Chen, W., Zhuang, B., & Shen, H. (2017). A review on development practice of smart grid technology in China. *IOP Conference Series: Materials Science and Engineering, Volume 199*. https://doi.org/10.1088/1757-899X/199/1/012062

Hazen, B. T., Overstreet, R. E., & Wang, Y. (2015). Predicting public bicycle adoption using the technology acceptance model. *Sustainability*, *7*(14558–14573). https://doi.org/10.3390/su71114558

Holden, R. J., & Karsh, B. (2010). The technology acceptance model: Its past and its future in health care. *Journal of Biomedical Informatics*, *43*(1), 159–172. https://doi.org/10.1016/j.jbi.2009.07.002

James, T. L., Warkentin, M., & Collignon, S. E. (2015). A dual privacy decision model for online social networks. *Information and Management*, *52*(8), 893–908.

Jay, D., Hmielowski, B., D., A., Harvey, G., & Joo, J. (2019). The social dimensions of smart meters in the United States: Demographics. *Energy Research and Social Science*, *55*, 189–197. https://doi.org/10.1016/j.erss.2019.05.003

Joubert, J., & van Belle, J. (2013). The role of trust and risk in mobile commerce adoption within South Africa. *International Journal of Business, Humanities and Technology*, *3*(2), 27–38. http://www.ijbhtnet.com/journals/Vol%5C_3%5C_No%5C_2%5C_February%5C_2013/3.pdf

Kaushik, A. K., Agrawal, A. K., & Rahman, Z. (2015). Tourist behaviour towards self-service hotel technology adoption: Trust and subjective norm as key antecedents. *Tourism Management Perspectives*, *16*, 278–289. https://doi.org/10.1016/j.tmp.2015.09.002

Keith, M. J., Thompson, S. C., Hale, J., Benjamin, P., & Greer, C. (2013). Information disclosure on mobile device: Re-examining privacy calculus with actual user behavior. *Journal of Human Computer Studies*, *71*(12), 1163–1173. https://doi.org/10.1016/j.ijhcs.2013.08.016

Kim, Y. J., Chun, J., & Song, J. (2009). Investigating the role of attitude in technology acceptance from an attitude strength perspective. *International Journal of Information Management*, *29*(1), 67–77. https://doi.org/10.1016/j.ijinfomgt.2008.01.011

Kranz, J., Gallenkamp, J., & Picot, A. (2010). Exploring the role of control–smart meter acceptance of residential consumers. *AMCIS 2010 Proceedings, Paper 315*. https://aisel.aisnet.org/amcis2010/315

Kranz, J., Kolbe, L. M., Koo, C., & Boudreau, M. C. (2015). Smart energy: Where do we stand and where should we go? *Electronic markets*, *25*(1), 7–16. https://doi.org/10.1007/s12525-015-0180-3

Kranz, J. J., & Picot, A. (2011). Toward an end-to-end smart grid: Overcoming bottlenecks to facilitate competition and innovation in smart grids. *Communications of the National Regulatory Research Institute*, *138*.

Krasnova, H., & Veltri, N. F. (2010). Privacy calculus on social networking sites: Explorative evidence from Germany and USA. *43rd Hawaii International Conference on System Sciences, 1–10.*, 1–10.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, *30*(3), 607–610.

Lai, P. (2017). The literature review of technology adoption models and theories for the novelty technology. *Journal of Information Systems and Technology Management*, *14*(1), 21–38. https://doi.org/10.4301/S1807-17752017000100002

Lee, M. (2009). Factors influencing the adoption of Internet banking: An integration of TAM and TPB with perceived risk and perceived benefit. *Electronic Commerce Research and Applications*, *8*(10), 130–141. https://doi.org/10.1016/j.elerap.2008.11.006

Lu, H., & Gustafson, H. (1994). An empirical study of perceived usefulness and perceived ease of use on computerized support system use over time. *International Journal of Information Management*, *14*(5), 317–329.

Lynne, G. D., Casey, C. F., Hodges, A., & Rahmani, M. (1995). Conservation technology adoption decisions and the theory of planned behavior. *Journal of Economic Psychology*, *16*.

Ma, Z., Asmussen, A., & Jørgensen, B. N. (2018). Industrial consumers' smart grid adoption: Influential factors and participation phases. *Energies*, *11*(1), 182.

Maduku, D. K. (2013). Predicting retail-banking customers' attitude towards Internet banking services in South Africa. *South African Business Review*, *17*(3), 76–100.

Manyika, J., Chui, M., Bughin, J., Dobbs, R., Bisson, P., & Marrs, A. (2013). Disruptive technologies: Advances that will transform life, business, and the global economy [McKinsey Global Institute]. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/disruptive-technologies

Martins, C., Oliveira, T., & Popovič, A. (2014). Understanding the Internet banking adoption: A unified theory of acceptance and use of technology and perceived risk application. *International Journal of Information Management*, *34*, 1–13. https://doi.org/10.1016/j.ijinfomgt.2013.06.002

Mayer, R. C., Davis, J. H., & Schoorman, F. C. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734.

McKenna, E., Richardson, I., & Thomson, M. (2012). Smart meter data: Balancing consumer privacy concerns with legitimate applications. *Energy Policy*, *41*, 807–814. https://doi.org/10.1016/j.enpol.2011.11.049

McKnight, D., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334–359.

Micheni, E. M., Lule, I., & Muketha, G. M. (2013). Transaction costs and facilitating conditions as indicators of the adoption of mobile money services in Kenya. *International Journal of Advanced Trends in Computer Science and Engineering*, *2*(5), 9–15.

Miltgen, L. C., Popovič, A., & Oliveira, T. (2013a). Determinants of end-user acceptance of biometrics: Integrating the "Big 3" of technology acceptance with privacy context. *Decision Support Systems*, *56*(1), 103–114. https://doi.org/10.1016/j.dss.2013.05.010

Miltgen, L. C., Popovič, A., & Oliveira, T. (2013b). Determinants of end-user acceptance of biometrics: Integrating the "Big 3" of technology acceptance with privacy context. *Decision Support Systems*, *56*(1), 103–114. https://doi.org/10.1016/j.dss.2013.05.010

Moon, J., & Kim, Y. (2001). Extending the TAM for a World-Wide-Web context. *Information and Management*, *38*(4), 217–230.

Morosan, C., & DeFranco, A. (2015). It's about time: Revisiting UTAUT2 to examine consumers' intentions to use NFC mobile payments in hotels. *International Journal of Hospitality Management*, *53*(10), 17–29. https://doi.org/10.1016/j.ijhm.2015.11.003

Muller-Seitz, G., Dautzenberg, K., Creusen, U., & Stromereder, C. (2009). Customer acceptance of RFID technology: Evidence from the German electronic retail sector. *Journal of Retailing and Consumer Services*, *16*(1), 31–39.

Notani, A. (1998). Moderators of perceived behavioral control's predictiveness in the theory of planned behavior: A meta-analysis. *Journal of Consumer Psychology*, *7*(3), 247–271.

Organisation for Economic Co-operation and Development. (2012). Green growth studies: Energy [Last accessed 10 Dec 2021]. https://www.oecd.org/greengrowth/greening-energy/49157219.pdf

Orlando, D., & Vandevelde, W. (2021). Smart meters' roll out, solutions in favour of a trust enhancing law in the EU. *Journal of Law, Technology and Trust*, *2*(1). https://doi.org/10.19164/jltt.v2i1.1071

Park, S. Y. (2009). An analysis of the technology acceptance model in understanding university students' behavioral intention to use e-learning. *Educational Technology and Society*, *12*(3), 150–162. https://doi.org/10.1007/s00340-009-3513-0

Ponce-Jara, M. A., Ruiz, E., Gil, R., Sancristóbal, E., Pérez-Molina, C., & Castro, M. (2017). Smart grid: Assessment of the past and present in developed and developing countries. *Energy Strategy Reviews*, *18*, 38–52. https://doi.org/10.1016/j.esr.2017.09.011

Rausser, G., Strielkowski, W., & Štreimikienė, D. (2018). Smart meters and household electricity consumption: A case study in Ireland. *Energy and Environment*, *29*(1), 131–146. https://doi.org/10.1177/0958305X17741385

Rehman, T., Yates, C. M., McKemey, K., Garforth, C., Cooke, R., Tranter, R. B., Park, J., & Dorward, P. (2003). Modelling the uptake of new technologies on dairy farms in South West England using the Theory of Reasoned Action and mathematical programming. *Agricultural Economics Society conference, Seale Hayne, England.*

Roca, J. C., Garcia, J. J., & de la Vega, J. J. (2009). The importance of perceived trust, security and privacy in online trading systems. *Information Management and Computer Security*, *17*(2), 96–113. https://doi.org/10.1108/09685220910963983

Sánchez-Prieto, J., Olmos-Migueláñez, S., & García-Peñalvo, F. J. (2015). Informal tools in formal contexts: Development of a model to assess the acceptance of mobile technologies among teachers. *Computers in Human Behaviour*, *55*, 519–528. https://doi.org/10.1016/j.chb.2015.07.002

Serenko, A., Bontis, N., & Detlor, B. (2007). End-user adoption of animated interface agents in everyday work applications. *Behaviour and Information Technology*, *26*(2), 119–132. https://doi.org/10.1080/01449290500260538

Smart Energy International. (2019). The factors behind low smart meter penetration in Africa [Last accessed 10 Dec 2021]. https://www.smart-energy.com/industry-sectors/policy-regulation/the-factors-behind-low-smart-meter-penetration-africa

Srivastava, J., & Moreland, J. J. (2012). Diffusion of innovations: Communication evolution and influences. *The Communication Review*, *15*(40), 294–312. https://doi.org/10.1080/10714421.2012.728420

Stern, B., Royne, M. B., Stafford, F. T., & Bienstock, C. C. B. (2008). Consumer acceptance of online auctions: An extension and revision of the TAM. *Psychology and Marketing*, *25*(10), 619–636. https://doi.org/10.1002/mar.20228

Sun, Y., Wang, N., Shen, X., & Zhang, J. (2015). Computers in human behavior location information disclosure in location-based social network service: Privacy calculus, benefit structure, and gender differences. *Computers in Human Behavior*, *52*, 278–292. https://doi.org/10.1016/j.chb.2015.06.006

Tan, X., Qin, L., Kim, Y., & Hsu, J. (2012). Impact of privacy concern in social networking web sites. *Internet Research*, *22*(2), 211–233. https://doi.org/10.1108/10662241211214575

Taneja, A., Vitrano, J., & Gengo, N. J. (2014). Rationality-based beliefs affecting individual's attitude and intention to use privacy controls on Facebook: An empirical investigation. *Computers in Human Behavior*, *38*, 159–173. https://doi.org/10.1016/j.chb.2014.05.027

Ten Kate, S., Haverkamp, S., Mahmood, F., & Feldberg, F. (2010). Social network influences on technology acceptance: A matter of tie strength, centrality and density. *23rd Bled eConference*, 18–32.

Teo, T. (2009). The impact of subjective norm and facilitating conditions on pre-service teachers' attitude toward computer use: A structural equation modeling of an extended technology acceptance model. *Journal of Educational Computing Research*, *40*(1), 89–109. https://doi.org/10.2190/EC.40.1.d

Thatcher, J., McKnight, D., Baker, E., Arsal, R. E., & Roberts, N. H. (n.d.). The role of trust in postadoption IT exploration: An empirical examination of knowledge management systems. *IEEE Transactions on Engineering Management*, *58*, 56–70.

Thomas, T. D., Singh, L., & Gaffar, K. (2013). The Utility of the UTAUT model in explaining mobile learning adoption in higher education in Guyana. *International Journal of Education and Development Using Information and Communication Technology*, *9*(3), 71–85.

Uitz, I., & Koitz, R. (2013). Consumer acceptance of location based services in the retail environment. *International Journal of Advanced Computer Science and Applications*, *4*(12).

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Brière, N. M., Senécal, C., & Vallières, E. F. (1992). The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement*, *52*, 1003–1019.

van Ittersum, K., Rogers, W. A., Caine, K. E., O'Brien, M. A., & Parsons, L. J. (2006). *Understanding technology acceptance: Phase 1 – Literature Review and Qualitative Model Development* [Technology Report HFATR-0602]. Georgia Institute of Technology, School of Psychology, Human Factors and Aging Laboratory. http://hdl.handle.net/1853/40580

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157–178.

Wang, Y., Chen, Q., Hong, T., & Kang, C. (2019). Review of smart meter data analytics: Applications, methodologies, and challenge. *IEEE Transactions on Smart Grid*, *10*(3), 3125–3148. https://doi.org/10.1109/TSG.2018.2818167

Willis, T. (2009). An evaluation of the technology acceptance model as a means of understanding online social networking behavior.

World Economic Forum. (2012). Energy for economic growth. Hydrocarbon Processing [Last accessed 10 Dec 2021]. https://www3.weforum.org/docs/WEF_EN_EnergyEconomicGrowth_IndustryAgenda_2012.pdf

World Economic Forum. (2016). The future of electricity in fast-growing economies attracting investment to provide affordable, accessible and sustainable power [Last accessed 10 Dec 2021]. http://www3.weforum.org/docs/WEF_Future_of_Electricity_2016.pdf

Wu, I. L., & Chen, J. L. (2005). An extension of trust and TAM model with TPB in the initial adoption of on-line tax: An empirical study. *International Journal of Human-Computer Studies*, *62*(6), 784–808.

Wunderlich, P., Veit, D., & Sarker, S. (2012). Adoption of information systems in the electricity sector: The issue of smart metering. *18th Americas Conference on Information Systems 2012, AMCIS 2012*, 3669–3678.

Xiong, S. (2013). Adoption of mobile banking model based on perceived value and trust. *2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering*, 632–635.

Xu, H., Robert, X., Carroll, J. M., Beth, M., Luo, X., Carroll, J. M., & Rosson, M. B. (2011). The personalization privacy paradox: An exploratory study of decision-making process for location-aware marketing. *Decision Support Systems*, *51*(1), 42–52. https://doi.org/10.1016/j.dss.2010.11.017

Zamrudi, Z., Karim, S., Faridha, M., Maharani, D., & Kuraesin, A. (2019). Smart meter adoption: The role of consumer experience in using smart device. *Journal of Physics: Conference Series*, *1175*. https://doi.org/10.1088/1742-6596/1175/1/012038

Zeithaml, A., V. (1988). Consumer perceptions of price, quality and value: A means-end model and synthesis of evidence. *Journal of Marketing*, *52*(3), 2–22. https://doi.org/10.1177/002224298805200302

Zhou, S. (2021). The effect of smart meter penetration on dynamic electricity pricing: Evidence from the United States. *The Electricity Journal*, *34*(3). https://doi.org/10.1016/j.tej.2021.106919

Zhou, S., & Brown, M. A. (2017). Smart meter deployment in Europe: A comparative case study on the impacts of national policy schemes. *Journal of Cleaner Production*, *144*, 22–32. https://doi.org/10.1016/j.jclepro.2016.12.031

Zhou, T. (2011). The impact of privacy concern on user adoption of location-based services. *Industrial Management & Data Systems*, *111*(2), 212–226. https://doi.org/doi:10.1108/02635571111115146

A APPENDIX: DERIVATION OF CONSTRUCT ITEMS

Table 11: Derivation of construct items

Monetary cost (MC)	MC1	Smart meter technology is expensive to a consumer like me. *	Agarwal et al. (2007)
	MC2	Smart meter technology makes me pay more than the old manual system. *	
	MC3	Smart meter technology causes me to incur a higher cost than the old manual system. *	
	MC4	Smart meter technology will make me pay more money unnecessarily. *	Self-developed
Perceived value (PV)	PV1	Smart meter technology provides good value.	Venkatesh et al. (2012)
	PV2	Smart meter technology is worthwhile considering.	Agarwal et al. (2007)
	PV3	Smart meter technology provides me more benefits than disadvantages.	
	PV4	I appreciate what smart meter technology can do for me.	
Attitude (AT) **	AT1	I think using smart meters is a good idea***	Belanche et al. (2012)
	AT2	I think using smart meters is a wise idea***	
	AT3	I think using smart meters would be a pleasant experience. ***	
	AT4	Generally, I like the idea of using smart meters. ***	
Facilitating conditions (FC)	FC1	I will easily access information about the use of smart meter technology. *	McKenna et al. (2012)
	FC2	I will easily get instructions on smart meter use in my home language. *	Venkatesh et al. (2012)
	FC3	I will easily get guidelines on how use smart meters. *	
	FC4	I will easily get guidelines on how use smart meters. *	
	FC5	I can easily get support when I have difficult using smart meter technology. *	Micheni et al. (2013)
Social norms (SN)***	SN1	I will be happy to have a smart meter installed in my home. ***	Taneja et al. (2014)
	SN2	I will volunteer to have a smart meter installed in my home. ***	
	SN3	I am comfortable with having a smart meter installed in my home. ***	
	SN4	I am positive about a city-wide roll-out of smart meters. ***	Venkatesh et al. (2012)
	SN5	I support the installation of smart meters in the city. ***	
	SN6	I will be happy to have a smart meter installed in my home. ***	
Personal consciousness (PC) ***	PC1	I will support smart meter technology use because my family supports it. **	Taneja et al. (2014)
	PC2	I will support smart meter technology use because my friends support it. **	
	PC3	I will support smart meter technology use because my colleagues support it. **	
	PC4	I will support smart meter technology use because people important to me say it helps save the environment. **	Venkatesh et al. (2012) and Willis (2009)

https://doi.org/10.18489/sacj.v33i2.909

	PC5	I will support smart meter technology use because people important to me think it is the right thing to do. **	
	PC6	I will support smart meter technology use because I believe it is the right thing to do. **	
Community consciousness (CC)**	CC1	I will support the use of smart meter technology because my community thinks it is good to manage and distribute electricity. **	Willis (2009)
	CC2	I will support the use of smart meter technology because my community thinks it prevents electricity theft. **	
	CC3	I will support the use of smart meter technology if my political affiliate party supports it. **	
	CC4	I will support the use of smart meter technology if my community thinks it saves electricity. **	
	CC5 **	Overall, I will support the use of smart meter technology if my community thinks it saves the environment. **	
Behavioural intention to use smart meters (BI)	BI1	I will be happy to have smart meter installed at my home.	Xu et al. (2011)
	BI2	I am favourable towards having a smart meter installed in my home.	
	BI3	I will volunteer to have a smart meter installed in my home.	Venkatesh et al. (2012)
	BI4	I am comfortable to have a smart meter installed in my home.	Kaushik et al. (2015)
	BI5	I plan to have smart meter installed in my home**	
	BI6	I support the installation of smart meters in the city.	
	BI7	I am positive about a city-wide roll-out of smart meters.	

B APPENDIX: FINAL CONSTRUCT ITEMS

Table 12: Final construct items

Trust in technology (TT)	TT1	Smart meter technology is trustworthy.
	TT2	Smart meter technology is dependable.
	TT3	Smart meter technology is credible in managing electricity demand & supply.
	TT4	Smart meter technology has a good reputation in the electricity industry.
	TT5	Smart meter technology improves reliable electricity supply.
	TT6	Smart meter technology records electricity billing information accurately.
Perceived usefulness (PU)	PU1	Smart meter makes it easier for me to monitor & adjust my electricity usage.
	PU2	Smart meter makes it easier to manage electricity usage.
	PU3	Smart meter makes it easy for me to get timely billing information.
	PU4	Smart meter makes it easier for me to use electricity efficiently.
Perceived ease of use (EU)	EU1	I will find it easy to use a smart meter.
	EU2	I will find it easy to learn how to operate the smart meter.
	EU3	I will find it easy to get the smart meter to do what I want it to do.
	EU4	It will not require any mental effort to use the smart meter.
Privacy risk (PR)	PR1	I think smart meter technology makes it easier for my personal data to be misused for market research and advertising without my knowledge.
	PR2	I think smart meter technology allows easier access to my personal data without my knowledge.
	PR3	I think smart meter technology makes me vulnerable to criminals.
	PR4	I think smart meter technology put my privacy at risk.
Monetary cost (MC)	MC1	Smart meter technology will make me pay more money unnecessarily.
	MC2	Smart meter technology will make me pay more than the old manual system.
	MC3	Smart meter technology will cause me to incur a higher cost than the old manual system.
	MC4	Smart meter technology will be expensive to a consumer like me.
Perceived value (PV)	PV1	Smart meter technology provides good value.
	PV2	Smart meter technology is worthwhile considering.
	PV3	Smart meter technology provides me more benefits than disadvantages.
	PV4	I appreciate what smart meter technology can do for me.
Attitude (AT)	AT1	I think using smart meters is a good idea.
	AT2	I think using smart meters is a wise idea.
	AT3	I think using smart meters would be a pleasant experience.
	AT4	Generally, I like the idea of using smart meters.
Facilitating conditions (FC)	FC1	Gaining access to information about the use of smart meters will be easy.
	FC2	Obtaining instructions for smart meter use will be easy.
	FC3	Obtaining guidelines on how to use smart meters will be easy.
	FC4	I can easily get support when I experience difficulties using smart meters.
	FC5	Gaining access to information about the use of smart meters will be easy.
Social norms (SN)	SN1	I will be happy to have a smart meter installed in my home.
	SN2	I will volunteer to have a smart meter installed in my home.
	SN3	I am comfortable with having a smart meter installed in my home.
	SN4	I am positive about a city-wide roll-out of smart meters.
	SN5	I support the installation of smart meters in the city.
	SN6	I will be happy to have a smart meter installed in my home.
	BI1	I will be happy to have smart meter installed in my home.

	BI2	I intent to have a smart meter installed in my home.
Behavioral intention (BI)	BI3	I will volunteer to have a smart meter installed at my home.
	BI4	I am comfortable to have a smart meter installed in my home.
	BI5	I am positive about a city-wide roll-out of smart meters.
	BI6	I support the installation of smart meters in the city.

SACJ 33(2) December 2021
Research Article

Big Data Driven Decision Making Model: A case of the South African banking sector

Komla Pillay, Alta van der Merwe
Department of Informatics, University of Pretoria, South Africa

ABSTRACT
The quest to develop a Big Data Driven Decision Making framework to support the incorporation of big data analytics into the decision-making process resulted in the development of a decision making model. The study was conducted within the banking sector of South Africa, with participants from three leading South African banking institutions. The conducted research followed the design science research process of awareness, suggestion, development, evaluation and conclusion.

This study developed a theoretical Big Data Driven Decision Making model which illustrates the decision-making process in banking using big data. The study further determined the organizational supports that need to be in place to support the big data analytics decision-making process.

Keywords: big data, big data analytics, innovation, decision-making, banking sector, design science, theoretical model, organizational support

Categories: • **Information systems** ∼ **Data analytics**

Email:
Komla Pillay komla.pillay@up.ac.za (CORRESPONDING),
Alta van der Merwe alta.vdm@up.ac.za

Article history:
Received: 6 December 2020
Accepted: 27 September 2021
Available online: 19 December 2021

1 INTRODUCTION

Fast paced technology advancements have reshaped organizations over the years and have challenged these organizations to keep up with the changes. One such technology trend that has emerged is the use of big data analytics with the potential to enhance knowledge within an organization (Ishwarappa & Anuradha, 2015). Decision-making regarding big data analytics within organizations ultimately influences innovation within organizations. Many organizations are unsure of how to proceed with this innovation in order to embrace all of the possibilities. An investigation into existing big data analytics business models only yielded the Gartner's Business Analytics Framework. This framework defines

> the people, processes and platforms that need to be integrated and aligned to take a more strategic approach to business intelligence, analytics and performance. (Chandler et al., 2011)

Pillay, K. and van der Merwe, A. (2021). Big Data Driven Decision Making Model: A case of the South African banking sector . *South African Computer Journal 33*(2), 55–71. https://doi.org/10.18489/sacj.v33i2.928

SACJ is a publication of the South African Institute of Computer Scientists and Information Technologists. ISSN 1015-7999 (print) ISSN 2313-7835 (online).

The framework does not identify organizational supports and guidelines that need to be in place for the incorporation of big data analytics into managerial decision-making.

Big data is a collection of large, complex sets of data which includes data from social media and real-time system records (Ishwarappa & Anuradha, 2015). The big data analytics era has introduced new methods for extracting and analyzing information in order to gain insights into the organizations, customers and business processes (Xu et al., 2016). Big data was used to highlight the challenges that the visualization of data would mean in the computing world. It was then defined for its volume, velocity, and variety of information which meant that sophisticated software had to be designed in order to handle the data. At the beginning of 2009, big data analytics became revolutionary and signaled new opportunities for business intelligence. Researchers began to predict that data collected will shift from structured to unstructured (Wang et al., 2018).

Big data analytics can help organizations make better and faster decisions. There is an opportunity to supplement the current managers' intuitive and experienced based decisions with data driven insights. There is a need for managers to better incorporate big data analytics into the decision-making process. In order for managers to better make use of big data analytics, they must first understand what the decision-making process looks like with big data analytics. They also need to understand what organizational supports need to be in place to support the big data analytics decision-making process.

1.1 Big Data Analytics in Banking

Decision makers within organizations rely on the quality, relevance and validity of the information when making key organizational decisions. Data storage technology and data gathering techniques have progressively improved. This innovative environment has created a new era for various sectors especially the banking sector. Banks have the opportunity to improve their customer information, products, risk measurement and market expectations. Banking institutions have embraced analytics but turning the analytical insights into business outcomes has been a challenge. Big data analytics success requires business adoption and change management. In 2018, only 7% of banks in the EMEA (Europe, Middle East & Africa) region had attained complete incorporation of important analytics use cases. Only 15% of banks in the EMEA region think that the management relies on analytics to make decisions. There is 20% of staff in EMEA banks that believe their management can be convinced by big data analytical insights that go against their original belief. Half of the management in the surveyed banks in the EMEA region, responded positively to the value that analytics offers their institution with only 25% effectively communicating how operational capacity can be improved (Naveira et al., 2018).

Many banking and other organizations are focusing on the development of tools that enable the storing and processing of these large volumes of structured and unstructured data. Organizations thereafter invest time and financial resources to enable their staff to use these tools. There is a gap in the research that there is not much attention placed on managers who

https://doi.org/10.18489/sacj.v33i2.928

use these progressive tools that generate smarter information. There is no process or system to monitor the impact of this on the organizations' decision-making cycle. Furthermore, there are no models, guidelines or frameworks to assist managers in incorporating big data analytics to support their decision-making.

The purpose of this study was to develop a model that can be used to assist managers in incorporating big data analytics into decision-making in the banking sector. The development of the model entailed establishing a set of criteria for the selection of an existing decision-making model that supports the incorporation of big data analytics, establishing the content of a decision-making model that includes big data analytics and developing an organizational supports model that can be used in conjunction with a decision-making model to promote the facilitation of big data analytics for decision-making in banking.

2 THEORETICAL FOUNDATION

The fundamental theories and models used for the development of the Big Data Driven Decision-Making models include Capgemini's Big Data Maturity model, Simon's decision-making model and the data pyramid.

2.1 Capgemini's Big Data Maturity Model

The level at which an institution has adopted big data analytics can be described using the Capgemini's Big Data Maturity model – outlined in Table 1. This model defines three levels of big data maturity: beginner, proficient and expert. The model rates each level against four categories: culture, capabilities and operating model, data and technology. Organizations within the banking sector can move to a higher level of maturity by possibly hiring and training the right analytics staff with continuous upskill options to stay abreast of new developments in big data analytics tools and techniques. Further to this, decision-making staff within the organization should also be trained on the use of analytics so their reliance on specialized analytics staff is diminished thereby enhancing the decision-making process (Capgemini, 2014).

https://doi.org/10.18489/sacj.v33i2.928

Table 1: Roadmap to building analytics maturity (Capgemini, 2014)

	Level of maturity		
	Beginner	Proficient	Expert
Culture	Preliminary analytics strategy, but little buy-in from leadership	Analytics used to understand issues, develop data-based options across the business	Full executive sponsorship of analytics
Capabilities and operating model	Pockets of reporting and analysis capability	Well-defined recruitment process to attract analytics talent	Analytics Centre of Excellence to promote best practices
	Dispersed talent	Budget for analytics training	Strategic partnerships for supplementary analytics skills
Data	No defined data infrastructure	Data available for existing and potential customers	Internal, external and social media data is merged to build an integrated and structured dataset
	Conflicting, informal and dispersed data	Most data are still unstructured and internal	
Technology	Poor data governance	Use of some statistical and forecasting tools	Established, robust master data management framework for structured and unstructured data sets
	Basic data reporting using mainly spreadsheet-based tools	Coherent procedures for data management	
Sample applications of customer data analytics	Mass/random targeting of customers to increase product profitability using basic product eligibility criteria	Basic profiling of customer base with customized analysis on drivers of purchase of each product individually	Analysing customer behaviour across channels to predict interest areas; developing personalized products and services

https://doi.org/10.18489/sacj.v33i2.928

2.2 Simon's Decision-Making Model

Herbert Simon's (1979) decision-making model as depicted in Figure 1, forms the basis of many decision-making models and has proven to stand the test of time. Simon's model depicts the decision-making process as either a linear or iterative flow of steps. At any point in the process, the decision maker may choose to revisit a previous step for further refinement. The phases of Simon's decision-making model are - Intelligence: the problem or opportunity is identified and relevant information is thereafter gathered. This is a time consuming, critical stage because the decision emanates from the information. The decision can only be as good as the information gathered; Design: several possible solutions for the problem or approaches to capitalize on an opportunity are developed. This phase includes intense research into the different available options; Choice: this phase evaluates the alternatives outlined in the design phase. The end result is a chosen option for implementation; and Implementation: the chosen option is implemented. If the chosen option is unsuccessful, then the decision-making process starts again at Intelligence or alternatively, the Design and Choice phases are revisited (Simon, 1979)

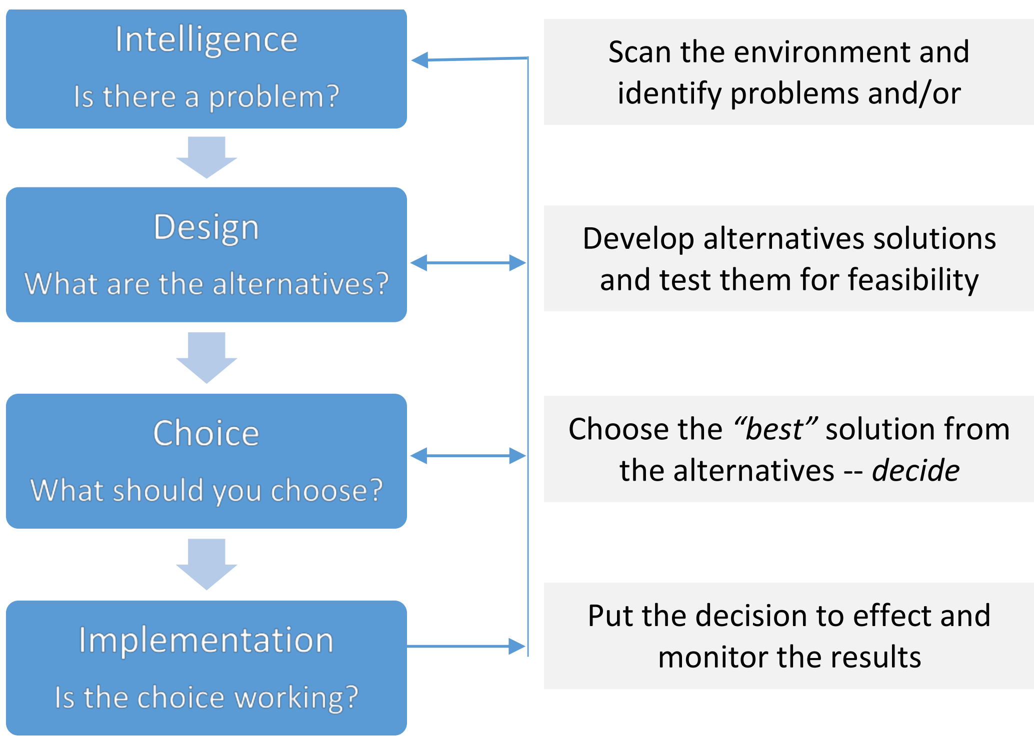

Figure 1: Simon's Decision-Making Model

https://doi.org/10.18489/sacj.v33i2.928

2.3 Data Pyramid / Wisdom Hierarchy

The wisdom hierarchy as depicted in Figure 2, starts at data, progressing to information, then knowledge and finally wisdom. Data is seen as the initial building block which can be categorized and corrected to become information. Likewise, information can be compared or connected to become knowledge and lastly wiser decisions can thereafter be made (Rowley, 2007). The importance of the generation of suitable knowledge being input into the decision-making process is known. The theoretical BDDDM model shows how the steps involved in decision-making process correspond to the wisdom hierarchy. It references the wisdom hierarchy from data to information to knowledge to wisdom. The wisdom level is able to be achieved due to big data analytics enabling action oriented automatic decisions and the ability to measure efficiency using tracking analytics.

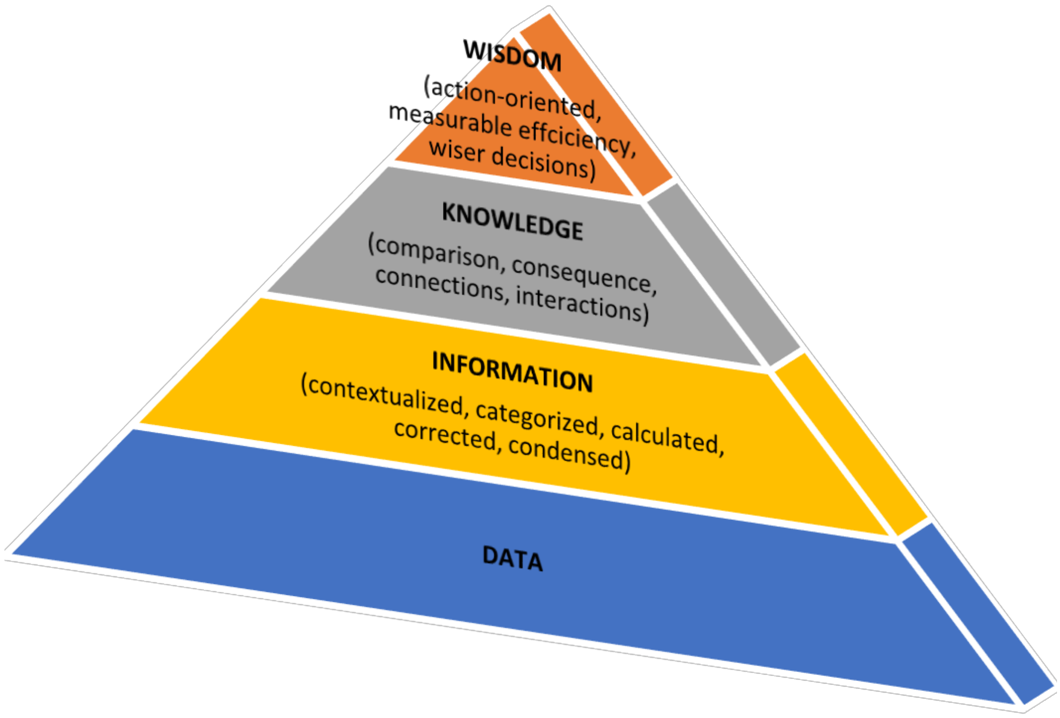

Figure 2: Wisdom Hierarchy (Rowley, 2007)

3 METHODOLOGY

The design science research methodology as outlined by Vaishnavi and Kuechler (2011) was followed for this research. A convenience sample was used for the study that included three of the five major banking institutions in South Africa. One interview and two questionnaires were administered for this research. In total six participants, two from each sampled institution were interviewed and surveyed. The participants were the Chief Information Officers (Chief Information Technology Officers) and another senior member of staff. A letter was sent to the Chief Information Officers of the selected banks, requesting their participation together with the participation of another senior member of their team. The senior staff members that participated were selected by the Chief Information Officers themselves.

The Big Data Driven Decision Making (BDDDM) framework was developed via three iterations with the findings from one of the iterations adding value to the next development iteration. The design research strategy as implemented by this research project is outlined below:

Cycle 1 was to be aware of, suggest and develop the theoretical model for big data driven decision-making for banking. This was achieved through a literature review and insight gained from interviews. The theoretical model provided the foundation for the next core development cycle and represented the development of the first construct for the Big Data-Driven Decision-Making (BDDDM) framework. The theoretical DDDM model was used as a base to construct the theoretical BDDDM model.

Cycle 2 was to be aware of, suggest and develop the organizational supports for big data driven decision-making for banking. This was achieved through a literature review and analysis of the questionnaire and interview data. The organizational supports model provided the input for the next core development cycle and represented the development of the second construct for the BDDDM framework.

Cycle 3 was to be aware of, suggest and develop the management guideline structure for the use of big data in decision-making. This cycle is not included in this paper.

The development phase was followed by the testing of the presented theoretical BDDDM model and guidelines by means of a proof of concept. The assessment of the practicality or usability, reliability, and efficiency was conducted by administering a test questionnaire to the participants from the three cases.

This paper will only detail the findings of cycle 1 and cycle 2, which resulted in the theoretical and organizational supports model for Big Data Driven Decision Making Model.

4 DATA COLLECTION

The data collection techniques used to establish a set of criteria for the selection of a base decision-making model (for cycle 1) that supports the incorporation of big data analytics was conducted by surveying the literature for managerial and data-driven decision-making models. The sample of the decision-making models was managerial or data driven due to the nature of the research being undertaken which is the development of managerial guidelines in a big data environment.

https://doi.org/10.18489/sacj.v33i2.928

The data collection techniques used to establish the content of a decision-making model and the necessary organizational supports included literature review, structured questionnaire and semi-structured interview.

Purposive non-probability research sampling was used. Purposive sampling is based on using one's judgment to select cases that will enable the research objectives to be met. It is often used when working with small samples. This research project selected three banking institutions where participants were administered questionnaires and interviewed. The three chosen banking institutions are on the top five biggest banks in South Africa list (Smith, 2017). South Africa has five big banking institutions whose rating is based on the number of South African banking customers. The research population for this research study is the top five South African banking institutions. The reason that this was chosen as the research population – the more established institutions are likely to be market leaders and have the necessary infrastructure to adopt the latest market trends like big data. The sample size for this research project was three institutions of a sample population of five. The number of respondents from each sample institution was small due to the nature of the information that was gathered together with the high-profile positions of the personnel that were interviewed. Sampled personnel included executive and senior information officers from the chosen institutions.

5 DATA ANALYSIS

5.1 Cycle 1 data analysis

The data analysis to determine a suitable model to adapt for big data analytics decision-making was carried out by comparing the various decision-making models according to the criteria that was determined by a Systematic Literature Review (SLR). The SLR established selection criteria were: focus on quality, incorporates decision support system technology, focus on meaningful insights, staff data literacy in an organizational environment, incorporates evaluation, incorporates the core phases of decision-making (intelligence, Design, Choice and Implementation), and focus on big data. The Data Driven Decision Making (DDDM) model was the most suitable decision-making model for adaptation. Figure 3 depicts the DDDM model. Although this data driven decision-making model originated in the education sector, the translation of this model to incorporate big data is seamless as it best fits the identified criteria. The key criteria is illustrated by the blocks with italic font—this was added on the original DDDM model in order to demonstrate the suitability of the model.

The theoretical Big Data Driven Decision Making (BDDDM) Model was created using exploratory data analysis. The process of adapting the theoretical DDDM model to the theoretical BDDDM model followed four steps:

1. illustrate how the phases of Simon's decision-making model maps to the different steps of big data analytics,

2. re-map the levels of the wisdom hierarchy to the steps for the analysis of big data,

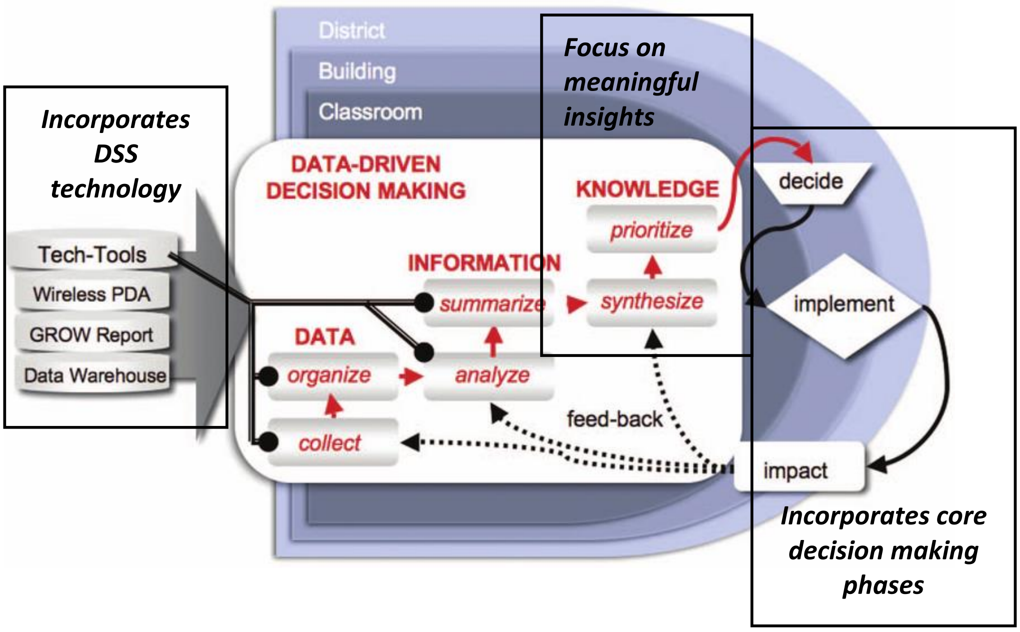

Figure 3: Theoretical Model for Data-Driven Decision Making (Goldenburg, 2017; Mandinach et al., 2006)

3. map the different levels of decision-making in the banking sector, and
4. explain the role of technology tools from the perspective of big data analytics.

5.2 Cycle 2 data analysis

The BDDDM Organizational Supports was also created using exploratory data analysis. This model was formed by identifying the organizational supports necessary for big data and thereafter identifying the corresponding steps for the organizational supports.

6 RESEARCH FINDINGS

6.1 Cycle 1 Research Findings

The shortlisted decision-making models were assessed against the criteria to determine that the Data Driven Decision Making (DDDM) model was the most suitable for adaptation. The Data Driven Decision Making Model met six of the seven established criteria. This model resonates the best with the proposed research due to its focus on the role of data in the decision-making process.

The objective for developing the theoretical BDDDM model was to demonstrate how decision-making works when big data analytics is incorporated into the decision-making process. Incorporating Simon's Decision-Making model resulted in the addition of the intelligence, design, choice and implementation phases. The decision-making phases outlined in Simon's decision-making model was incorporated into the newly created BDDDM model. The reason for the inclusion is that Simon's model is seen as a standard for decision-making models as many decision-making models use this model as their base. It also serves to assist the decision-maker to understand which phase of the decision-making cycle the different big data analysis steps are positioned.

Thereafter the identified DDDM model shows how the steps involved in the decision-making process correspond to the wisdom hierarchy. It references the wisdom hierarchy from data to information to knowledge. It does not reference the wisdom hierarchy up to the wisdom stage. The big data analytics use in artificial intelligence allows for the wisdom level of the wisdom hierarchy to be referenced in the created BDDDM model. The wisdom level is able to be achieved due to big data analytics enabling action oriented automatic decisions and the ability to measure efficiency using tracking analytics.

The DDDM model illustrated the steps involved in the processing of data as it is being converted to knowledge. This set of steps were collect, organize, analyses, summarize, synthesize and prioritize. The reason why these steps were changed in the newly created BDDDM model was due to the steps involved in the processing of big data differing. The steps involved in the analysis of big data are recording, acquisition, cleaning, extraction, integration, aggregation and interpretation. The phases of Simon's decision-making model are intelligence, design, choice and implementation. The steps for the analysis of big data, the phases of Simon's decision-making model and the different levels of the wisdom hierarchy all come together in the theoretical BDDDM model.

- Due to big data analytics enabling artificial intelligence, the decision can be made automatically within the system. Alternatively, the decision can be recommended and approved outside the system. This is the reason why choice from Simon's decision-making model can be both in and out of the system.

- The intelligence phase of Simon's decision-making model corresponds with data and information on the data pyramid and acquisition, recording, cleaning and extraction phases of big data analysis.

- The different alternatives generated for each decision makes up the design phase of Simon's decision-making model and this corresponds to knowledge on the wisdom hierarchy. The corresponding big data analysis steps for design phase are aggregation and integration.

- At the wisdom level, the corresponding big data analytics step is interpretation and this can also be corresponded to the choice phase of Simon's decision-making model.

Once a decision is taken and implemented, the results are fed back into the system to the relevant phase or step. The implementation and impact part of the BDDDM model is the same as that of DDDM except for the additional feedback loop to the Wisdom band of the data pyramid.

The next step of creating a theoretical BDDDM model looked at the decision-making levels. The created theoretical BDDDM model shows the different possible levels of decision-making in banking being branch level, regional level and head office. The interviewees revealed that decisions are taken at all three of these levels with decisions effected by big data only being made at regional and head office level. The final step for the creation of the theoretical BDDDM model involves technology tools. Technology tools are demonstrated as an arrow entering into the BDDDM model as the tools facilitate the data entering into the banking system. Technology tools assists big data decision-making as it brings in the raw data into the system. Some technology tools have pre-processing functionality built in which already starts the big data analysis process. Big data mining is not possible without the use of technology tools. Data and technology are linked together with the data being entrenched into the technology and is therefore regarded jointly. The type of data differs according to the type of technology and there also exists different user groups of the organization who will use different technology tools.

The theoretical BDDDM model is depicted in Figure 4 which illustrates the different steps involved in the analysis of big data, how each step correlates to the wisdom hierarchy and thereafter maps these two parts to Simon's Decision-Making Model. The theoretical BDDDM model was evaluated by the research participants against three objectives: the model clearly demonstrates how decision-making will look if big data analytics were incorporated; the model clearly illustrates how Simon's decision-making model (i.e. intelligence, design, choice and implementation) maps to the different steps of big data analytics; and the models clearly demonstrates the different levels of decision-making that big data analytics should support in the banking environment. Respondents conclusively agreed that all the outlined objectives were successfully achieved. The theoretical BDDDM model outlined what big data driven decision-making will look like by illustrating the big data analysis steps in relation to the elements of data, information, knowledge and wisdom—this is the output from cycle 1.

6.2 Cycle 2 Research Findings

Cycle 2 of the study determined the organizational supports that need to be in place to support the big data analytics decision-making process. This is a two-part process of identifying the organizational supports and identifying the corresponding steps for the organizational supports.

The organizational supports necessary for big data analytics decision-making in banking were determined from the interview response and mapping the response to the literature. The first discussion point of the semi-structured interview required the respondents to outline the organizational supports that are necessary for the effective use of customer data to improve customer service (sample use case used for the purpose of the interview) in banking. The

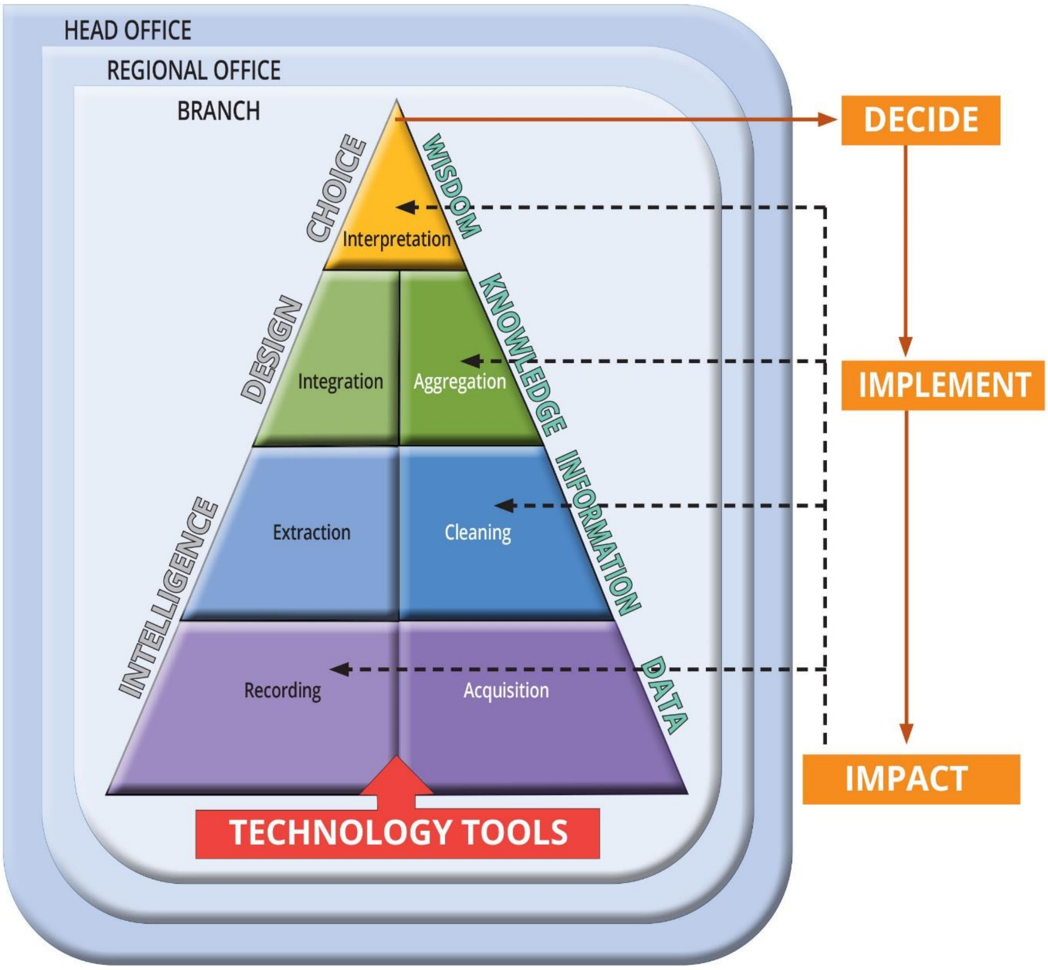

Figure 4: Theoretical Big Data Driven Decision Making Model

results were data-driven strategies, big data analytics organization structure, organizational understanding of big data processes, market leading tools, focus on technology and talent needed to support big data analytics culture. The six different organizational supports were categorized into Capgemini's Big Data Maturity organizational support categories of data, infrastructure, capabilities and operating model, and culture. The reason for the categorization was to determine if the respondents identified categories that fit into the organizational support categories identified by Capgemini Big Data Maturity Model (Capgemini, 2014). The Capgemini Big Data Maturity model is a model developed for the banking sector that helps determine an organization's level of big data maturity. This model also categorizes organizational supports for the banking sector. The feedback received from the respondents can be grouped into the established categories as in Table 2. All responses were able to be logically grouped into one of the existing categories so this confirms that the organizational supports found in big data literature can be used for the BDDDM Organizational Supports model.

Table 2: Organisational Support Grouping according to Capgemini Categories

Organisational Support Grouping	Respondent grouping
Data	1 – Data-driven strategies
Infrastructure	4 – Market-leading tools 5 – Focus on the technology
Capabilities and operating model	2 – Big data analytics organisational structure
Culture	3 – Organisational understanding of big data processes 6 – Talent needed to support big data analytics culture

The identified organizational supports were further streamlined by combining Data and Infrastructure. The reason for the merge is that you cannot separate the data from the technology with regard to big data analytics. The data and technology are inextricably linked as the choice of technology determines the format and structure of the data. The final list of organizational supports is Data and Technology, Capabilities and Operating Model, and Culture.

Part 2 of the organizational supports model is identifying the corresponding steps for the organizational supports. In order to identify the corresponding steps that talk to the organizational supports, the steps in the analysis of big data were mapped to the relevant organizational support. The three chronological and inter-reliant steps are outlined in Figure 5. It is also important to note that the Data and Technology organizational supports have been combined at the process level. The reason for this is the interconnectedness of the two supports for the purposes of gathering and storing the data. Figure 5 outlines the process with the identified organizational support, its component parts and its' relation to the steps in the analysis of big data.

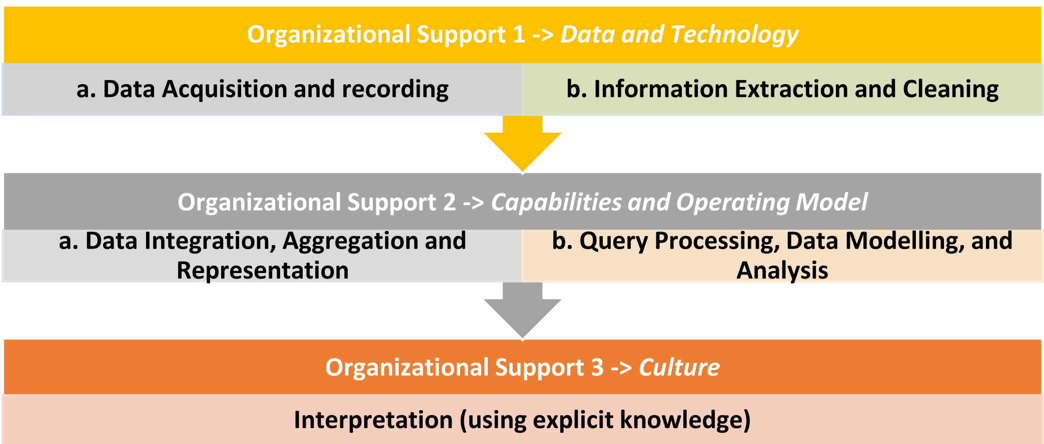

Figure 5: Steps for the BDDDM Organizational Supports Model

The mapping process was done using the literature. The interviewees were also asked during the semi-structured interviews about how these processes work on the ground. The relevant responses are listed below according to process and are discussed in relation to the literature.

Organizational Support 1. From the interviews, it was established that the banking institutions make use of third-party organizations to collect some of their data. Banking institutions do not have total control over the data collection methods. The institutions do gather some of their own data. It was also determined that customer data containers are decentralized with customer records for transactions, risk profiling and credit history, and are managed by various units within the organization. South African Banks analyses mostly structured voluminous data. The interview with key stakeholders revealed that banks do not have entrenched mechanisms in place to deal with unstructured data. Important to note at this juncture that dealing with unstructured data is a key component of working with big data.

Organizational Support 2. The interviewees from all the case studied institutions shared the same sentiment that the necessary big data analytics skills with respect to available big data analytics technology within their organizations are a challenge. The interviewees went as far as citing that the industry does not have sufficient or adequate big data analytics capacity. The interviewees also stated the close working partnerships that have been developed with external entities that assist banking institutions with their data analytics capacity. This process is responsible for conducting the analysis in order to create the explicit knowledge that is vital to managerial decision-making within the organization. This process requires the identified decision maker's knowledge requirements, use of data and the relationship with the identified data variables.

Organizational Support 3. This process requires the explicit knowledge output from Organization Support process 2, to be used by the decision makers to inform or justify their strategic and operational decisions. This is the last process of the BDDDM Organizational Supports model for banking.

The identified organizational supports together with the associated steps are demonstrated in Figure 6, BDDDM Organizational Supports model for banking—this is the output for cycle 2.

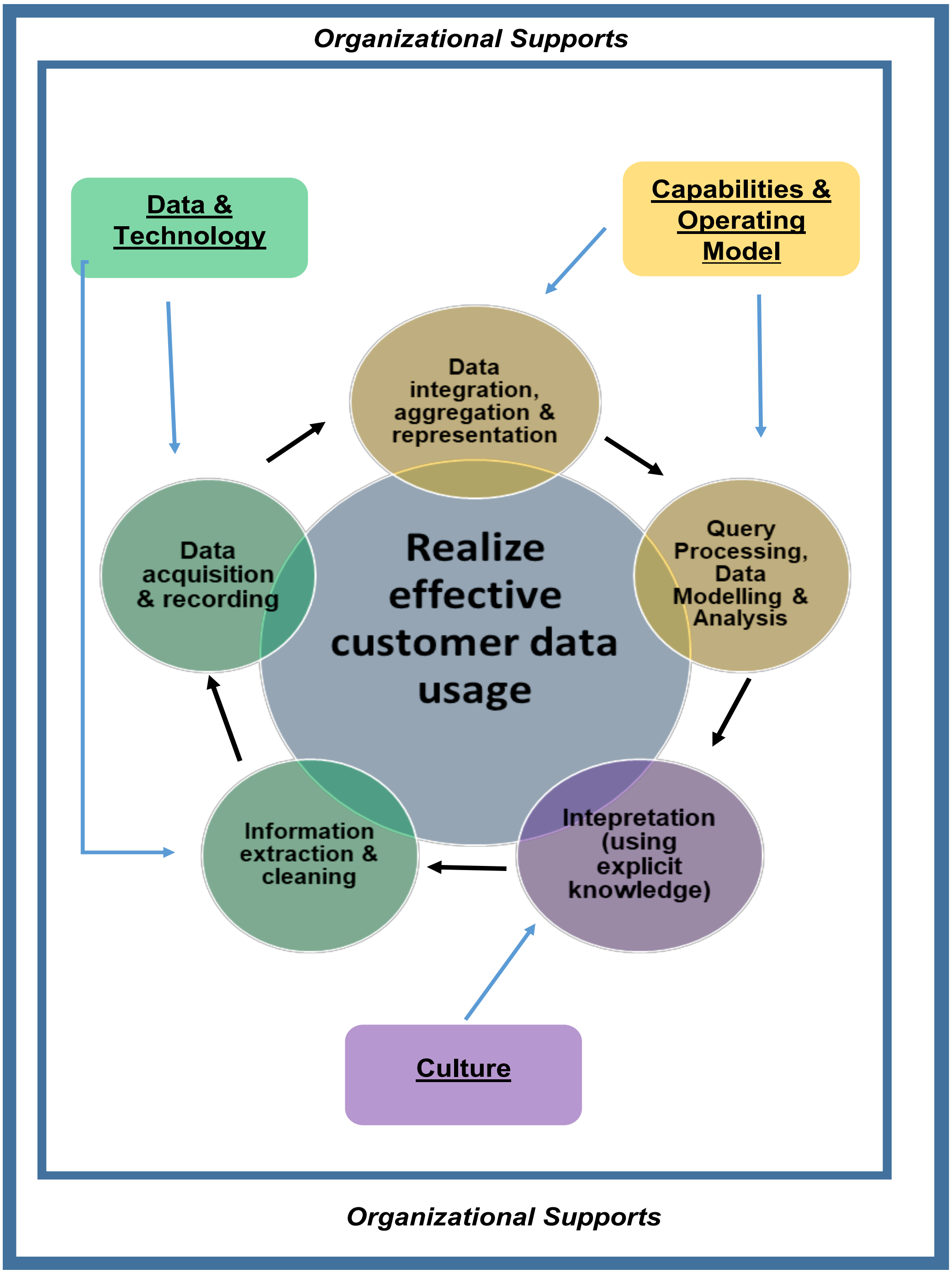

Figure 6: BDDDM Organizational Supports Model for Banking

6.3 How do the created constructs add value to the banking industry?

The theoretical model is a good tool to use for the dissemination of information regarding what is involved in the big data decision-making process. Typical banking institutions have a Design and Development unit (as was the case with all 3 cases for this study) that is a nucleus of the decision-making process. This unit will use the BDDDM model to disseminate how decision-making in a modern financial institution that is dependent on technology works. This unit will bring the decision-making stakeholders up to speed with the process of technological big data decision-making. The foundational understanding of how big data decision-making works ensures that all stakeholders have the same view of the structure of decision-making. This knowledge can promote a common understanding thereby facilitating more streamlined functioning of the operational teams. The theoretical BDDDM model can be used to demonstrate to non-technical banking staff (some managers) the internal workings of the big data decision-making process. It can be used to determine bottlenecks in the decision-making process.

The organizational supports as identified in the BDDDM organizational supports model need to be in place to ensure effective decision-making. Typically in banking institutions, the Chief Information Officer (CIO) will delegate the responsibility for each organizational support to the relevant unit. For example, the responsibility for the data and technology will lie with the Information Security unit, the Capabilities and Operating support will be in the System Development unit, and the Culture support will be with the CIO. The derived organizational supports have to be constantly checked for validity. Sometimes the structure of the organizational supports may remain the same as in the name, but the composition of the structure may change due to technological advances or organizational culture change. So each relevant unit will be awarded a check list to update every quarter to ensure the right actions are being performed on or with the organizational support. The BDDDM organizational supports model will assist the responsible banking unit to periodically assess the composition of the organizational support to establish if any amendments are necessary.

7 CONCLUSION AND LIMITATIONS

The theoretical and organizational supports model together illustrate the Big Data Driven Decision Making Model. Both models are linked by the big data processing steps. The theoretical model demonstrates how the big data decision making process works and the organizational supports model demonstrates what supports need to be in place within the organization to enable an efficient big data driven decision making process.

The following limitation must be taken into consideration when evaluating the creation of the theoretical and organizational supports Big Data Driven Decision Making Models:

- This study did not include the specific types of technology used in the banking sector for the processing of the big data. These technologies have a lot of processing ability which could have some impact on what the decision-making model would look like based on the steps involved in big data processing.

References

Capgemini. (2014). Big data alchemy: How can banks maximize the value of their customer data. [Last accessed 10 Dec 2021]. https://www.capgemini.com/resources/big-data-customer-analytics-in-banks/

Chandler, N., Hostman, B., Rayner, N., & Herschel, G. (2011). Gartner's business analytics framework. https://www.gartner.com/imagesrv/summits/docs/na/business-intelligence/gartners_business_analytics__219420.pdf

Goldenburg, B. (2017). The components of effective, data-driven decision-making [Last accessed 10 Dec 2021]. https://ismguide.com/data-driven-decision-making/

Ishwarappa, J., & Anuradha, J. (2015). A brief introduction on big data 5Vs characteristics and Hadoop technology. *Procedia Computer Science*, *48*, 319–324. https://doi.org/10.1016/j.procs.2015.04.188

Mandinach, E., Honey, M., & Light, D. (2006). *A theoretical framework for data-driven decision-making*. American Educational Research Association.

Naveira, C. F., Jacob, I., Rifai, K., Simon, P., & Windhagen, E. (2018). Smarter analytics for banks [Last accessed 10 Dec 2021]. https://www.mckinsey.com/industries/financial-services/our-insights/smarter-analytics-for-banks

Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information and Communication Science*, *33*(2), 163–180.

Simon, H. (1979). Rational decision making in business organizations. *The American Economic Review*, *69*(4), 493–513.

Smith, A. (2017). Top 5 biggest banks in South Africa [Last accessed 10 Dec 2021]. https://buzzsouthafrica.com/top-5-biggest-banks-in-south-africa/

Vaishnavi, V., & Kuechler, W. (2011). Extending prior research with design science research: Two patterns for DSRIS project generation. *Lecture Notes in Computer Science*, *6629*, 166–175.

Wang, Y., Kung, L., & Byrd, T. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organisations. *Technological Forecasting and Social Change*, *126*(3–13).

Xu, Z., Frankwick, G., & Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research*, *69*, 1562–1566.

SACJ 33(2) December 2021
Communication

Citation and referencing guidelines

James Dibley sacj.production.editor@gmail.com, Philip Machanick sacj.editor@gmail.com
Department of Computer Science, Rhodes University, South Africa

1 INTRODUCTION

This short article provides some guidelines and worked examples of reliable citation and referencing practice for contributing authors and reviewers to/for the *South African Computer Journal*. This is motivated by several observations:

1. that citation data obtained through search engines and reference managers are often unreviewed, misformatted, or incomplete;
2. while *SACJ* is prepared for publication using the LaTeX toolchain, a great number of accepted contributions to the journal are prepared using other software, and in these cases, *SACJ* production staff must manually transfer the citation data;
3. better guidelines for contributors and reviewers can support more rigorous review, resulting in a faster and more effective publication pipeline for *SACJ* overall.

As such, this article aims to provide enough detail:

- to prepare an article for submission, and
- to establish that a submitted article is ready to be accepted for publication.

1.1 Background

For historical and interdisciplinary reasons, *SACJ* published articles employ a citation and referencing style that closely tracks the American Psychological Association style guide, 6th edition. We are greatly assisted in this by Philip Kime's biblatex-apa6 package[1].

Submissions to *SACJ* are *not* required to use the APA 6th citation style, and we recommend that contributors do not attempt to do so.

However, *SACJ* does recommend that supplied reference sets include all of the data required by the APA 6th referencing style. Section 3 details this.

Dibley, J. and Machanick, P. (2021). Citation and referencing guidelines [Communication]. *South African Computer Journal 33*(2), 72–78. https://doi.org/10.18489/sacj.v33i2.1048

SACJ is a publication of the South African Institute of Computer Scientists and Information Technologists. ISSN 1015-7999 (print) ISSN 2313-7835 (online).

[1] https://www.ctan.org/pkg/biblatex-apa6

2 SUBMISSION GUIDELINES

2.1 Citation style

Authors using LaTeX, Microsoft Word or alternative word processing software should prepare their submissions using an IEEE-style numeric citation style, e.g.:

> As Pade observes, "in introductory courses on quantum mechanics, the practice of formal skills often takes priority (this is subsumed under the slogan 'shut up and calculate')." [1, p. xvii]
>
> [1] Pade, J. (2018). *Quantum mechanics for pedestrians* (2nd). Springer. https://doi.org/10.1007/978-3-030-00464-4

This citation style enables the most direct workflow for reviewers to evaluate a submission, as well as the most direct workflow for the production editor to prepare a pre-publication proof.

2.2 Referencing

Please make best efforts to provide complete, up-to-date and correctly-styled reference entries at the point of submission. Submissions with inaccurate or incomplete referencing may result in significant delays to preparation of the final proof.

2.2.1 Abbreviation

Please do not abbreviate the titles of journals or conference proceedings. Where abbreviations have been used in downloaded reference entries, these should be corrected.

2.2.2 Recommendations for contributors using Microsoft Word

SACJ recommends that authors preparing submissions in Microsoft Word make use of one of several freely-available reference manager plugins[2] that are able to export reference sets as BibTeX files. This simple step can significantly reduce time and risk of transcription error during the preparation of the final proofs.

The list of references at the end of this document may be used as a guide for manual preparation of reference entries, while the BibTeX code examples itemise the attributes required.

Numeric keys should be used for inline citations and references should use numeric keys ([1], [2-3], [4, 5, 6], etc.), rather than author names.

2.2.3 Recommendations for contributors using LaTeX

The *SACJ* production editor also cautions that BibTeX reference entries obtained through Google Scholar (e.g., through the 'cite search result' link) are captured from existing documents

[2] e.g. https://www.zotero.org/download/

rather than automatically generated from article metadata, and as such may be poorly-styled, incomplete, or inaccurate.

Authors may consult the BibTeX entries provided in this document to check the style and information requirements.

3 EXAMPLE REFERENCE ENTRIES

3.1 Article

Aaaa and Bbbb (2047) is provided by the following BibTeX code:

```
@article{1,
  author    = {Aaaa, P. Q. and Bbbb, R. S. },
                  % individual author names can be given in any format ('Donald
                     Knuth' or 'D.E. Knuth' or 'Donald E. Knuth' or 'Knuth,
                     Donald E.' are all fine, and will be processed and
                     formatted the same).
                  % give an author's name in a single consistent form across all
                      reference entries;  but giving 'Donald Knuth' for one
                     entry and 'D.E. Knuth' in another will fool the processor
                     into differentiating them
                  % for lists of authors, use 'and' (e.g. {Erich Gamma and
                     Richard Helm and Ralph Johnson and John Vlissides} -- long
                     lists will be abridged in line with APA requirements)
                  % for works with corporate/institutional authorship, brace the
                      entire name (e.g. {{American Mathematical Society}})
  year      = {2047},
  title     = {The quick brown fox jumps over the lazy dog: {An} overview of
     keyboard test procedures in the short twentieth century},
                  % capitalise the first word of the title, subtitle, any proper
                      names, and capitalise entire acronyms ('SACJ')
  journal   = {Historical Perspectives On Quality Assurance},
                  % capitalise every word in a journal title
  volume    = {17},
  number    = { 3},
  pages     = { 210--211}, % endash ('--') for page or date ranges
  publisher = {Springer},
                  % just the publisher's name, not their location
  doi       = {https://doi.org/001.0001/001.1.0001},
                  % if doi is available, omit url
                  % give DOI in form  -> https://doi.org/...
                  %           NOT   doi:...
                  %           NOT   DOI:...
                  %       AND NOT   https://dx.doi.org/
  url       = { },
  note      = { }   % if url is given, record 'Last accessed XX Mon Year' here
}
```

3.2 Conference paper

Cccc (2017) is provided by the following BibTeX code:

```
@inproceedings{2,
  author    = {T. Cccc},
  title     = {Think like a computer, scientist: {Debugging} {Prolog}},
  year      = {2017},
  booktitle = {2017 {IET} 4th {International Conference on Logic Programming}, {
     Toronto, Canada}, {22--24 May}},
                  % full title of conference proceedings, not abbreviated
                  % booktitles are given in sentence case, but name of
                     organisation, title of conference, location, month are all
                     proper names
                  % endash ('--') for page or date ranges
  pages     = {437--586},
                  % endash ('--') for page or date ranges
  organization = {IET},
  doi       = {https://doi.org/002.0002/002.2.0002}
}
```

3.3 Book

Dddd (2008) is provided by the following BibTeX code:

```
@book{3,
  author  = {Ursula Dddd},
  edition = {3rd},
  publisher = {Megadodo}, % do not provide location
  title   = {Modern developments in modelling development},
              % Book titles are given in sentence case, unlike journal titles
  year    = {2008},
  doi     = {https://doi.org/003.0003/003.3.0003} % if available
  % do not provide ISBN
}
```

3.4 Book chapter

Eeee (1998) is provided by the following BibTeX code:

```
@incollection{ 4,
  author    = {Victor Eeee},
  booktitle = {Mastering on-target debugging for wireless nanodevices},
  title     = {Dude, where's my {UART}?},
  year      = {1998},
  editor    = {Robert Ffff},
  url       = { },          % optional
  pages     = {157--184},   % endash ('--') for page or date ranges
  note      = { },
  publisher = {Zarniwoop-Verlag}
}
```

3.5 Theses

Gggg (1888) is provided by the following BibTeX code:

```
@mastersthesis{ 5,
  author = {Xavier Gggg},
  title = {Procedural power solutions: {A} multidisciplinary intervention},
                 % double-brace first character of subtitle
  year = {1888},
  school = {Xyzzy Institute of Technology},
  address = {Vancouver}
}
```

Hhhh (1999) is provided by the following BibTeX code:

```
@phdthesis{ 6,
  author = {Wilhelmina Hhhh},
  title = {Quantitative studies in quantum quintessence},
  year = {1999},
  school = {Qwerty College},
  address = {Wellington}
}
```

3.6 Technical report

Iiii (2000) is provided by the following BibTeX code:

```
@techreport{ 7,
  author     = {Eileen Iiii},
  title      = {Palatino, {Helvetica}, and {Gill Sans}},
                % double-brace technology names (in this case, font names)
  year       = {2000},
  type       = {Technical report},
  institution = {Font Institute},
  doi        = { }, % if available
  url        = {https://en.wikipedia.org/wiki/Palatino },
  note       = {Last accessed 15 Jun 17 }
                % if url is given, record 'Last accessed XX Mon Year' here
}
```

3.7 Webpages and online resources

South African Institute for Computer Scientists and Information Technologists (1998) is provided by the following BibTex code:

```
@misc{ 8,
  author     = {{South African Institute for Computer Scientists and Information
     Technologists }},
                % double-brace names of institutions / research groups /
                   commercial entities
  title      = {{SACJ} announcements},
                % double-brace acronyms
  year       = {1998},
  url        = { https://sacj.cs.uct.ac.za/index.php/sacj/announcement},
  note       = {Last accessed 10 Sep 99}
                % if url is given, record 'Last accessed XX Mon Year' here
}
```

References

Aaaa, P. Q., & Bbbb, R. S. (2047). The quick brown fox jumps over the lazy dog: An overview of keyboard test procedures in the short twentieth century. *Historical Perspectives On Quality Assurance*, *17*(3), 210–211. https://doi.org/001.0001/001.1.0001

Cccc, T. (2017). Think like a computer, scientist: Debugging Prolog, In *2017 IET 4th International Conference on Logic Programming, Toronto, Canada, 22–24 May*. IET. https://doi.org/002.0002/002.2.0002

Dddd, U. (2008). *Modern developments in modelling development* (3rd). Megadodo. https://doi.org/003.0003/003.3.0003

Eeee, V. (1998). Dude, where's my UART? (R. Ffff, Ed.). In R. Ffff (Ed.), *Mastering on-target debugging for wireless nanodevices*. Zarniwoop-Verlag.

Gggg, X. (1888). *Procedural power solutions: A multidisciplinary intervention* (Master's thesis). Xyzzy Institute of Technology. Vancouver.

Hhhh, W. (1999). *Quantitative studies in quantum quintessence* (Doctoral dissertation). Qwerty College. Wellington.

Iiii, E. (2000). *Palatino, Helvetica, and Gill Sans* (Technical report) [Last accessed 15 Jun 17]. Font Institute. Last accessed 15 Jun 17. https://en.wikipedia.org/wiki/Palatino

South African Institute for Computer Scientists and Information Technologists. (1998). SACJ announcements [Last accessed 10 Sep 99]. https://sacj.cs.uct.ac.za/index.php/sacj/announcement