Journal of the International Society for the Systems Sciences | 67th Meeting of the International Society for the Systems Sciences
J.M. Wilby, Editor.

SYSTEM THINKING MEETS DATA SCIENCE/ENGINEERING

Vusumuzi Malele
Unit for Data Science and Computing, .
School of Computer Science and Information Systems
Vanderbijlpark Campus, North-West University, South Africa
vusi.malele@nwu.ac.za

Abstract
With the advent of data 3.0 and analytics 3.0, system thinkers are in the position to provide a bigger picture in data science and data engineering. In the data life cycle, a system thinking approach emphasises data-driven decision-making. A System Thinker approaches problem-solving by viewing the problems as part of a wider, data-resourced and dynamic system, and a Data Practitioner supports the data life cycle by collecting, transforming, and analyzing data, and communicating results to inform and guide decision-making. This paper uses explanatory research and a pragmatic case study approach to look at the (i) What is the role of system thinking and data science/engineering skill in data-driven decision-making or organisation? (ii) Does the combination of system thinking and data science/engineering give rise to a new discipline? (iii) What are the skills needed in this new discipline? The research shows that the system thinking skills in the data life cycle are important. System thinkers meet data practitioners to provide a bigger picture of data-driven decision-making. The latter ascertains the position of a system thinker in any industrial revolution (i.e. industry 4.0. and industry 5.0). Furthermore, a Data-System Thinker is proposed as a new career field.

Keywords
System thinking, Data Science, Data Engineering, Data Science Thinking, Data System Thinking.

1 | Introduction

The fourth industrial revolution, gig economy, digital living, etc, changed the way life is perceived in terms of living, working and interaction. The interaction among people has moved to technologies that produce a vast amount of data. These days, every walk or jog, telephone call, and interaction with the environment and people produces data. These data, if well analysed to information, could create a better world for all. In this regard, the data could drive decision-making.

The bedrock of all data-driven decision-making is the availability of datasets and the analysis capabilities. The analysis of datasets is provided by different methodologies, among which are bibliometric analysis, mathematical and statistical modelling, artificial intelligence algorithms, machine learning, deep learning, etc. Data practitioners like data scientists and data engineers become responsible for the generation of such datasets, the analysis and contribution to decision-making. However, their contribution in most cases is channelled to a particular challenge or opportunity within life in general or an organization.

Exhibit 1. Analysis of System Thinking and Data Science based on Google Analytics (Source: Author).

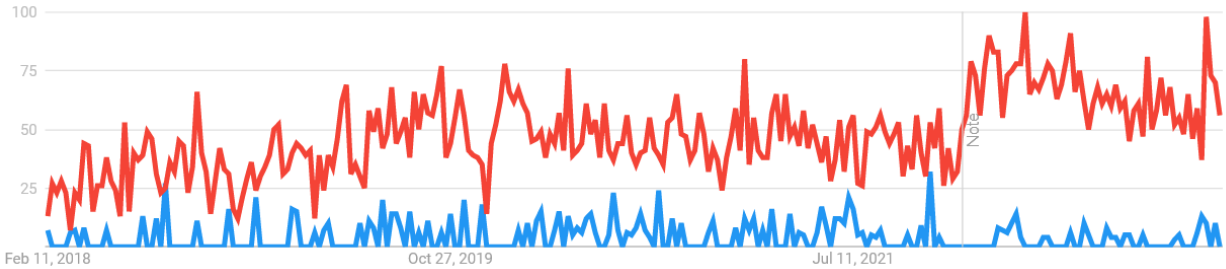


Exhibit 1, shows a quick Google Analytics that compares the popularity search between data science and data engineering (DSDE), and system thinking fields. Clearly, from Exhibit 1, most people seem to check out matters

of DSDE more than they check system thinking, leading to the fact that most challenges are addressed through data-driven decision-making, with a special focus on that challenge rather than the data contributing to the system or organization-wide decision-making.

The scope of providing decision-making that could positively the organization not just a challenged area but addressed the challenged area and effectively enhance others areas, is addressed by the involvement of a System thinker. System thinkers look at system parts with a connected whole perspective, with a curious and root cause seeking open-minded behaviour. They expand the range of options available for solving a problem. On the other hand, a Data Scientist/Engineer looks at a problem through a data glance.

Although, data science and data engineering (DSDE) is popular; however, some thinking capabilities are needed by DSDE practitioners for managing an overall life or organization-wide systems. The latter calls for the combination of system thinking and data science/engineering.

DSDE is interdisciplinary. In this paper, interdisciplinary means involving more than one branch of knowledge (Online Oxford Dictionary). System thinking could be seen as an interdisciplinary field in nature. In this regard, the amalgamation of system thinking and DSDE lead to a multidisciplinary field. In this paper, multidisciplinary means combining or involving several academic disciplines or professional specializations in an approach to a problem (Online Oxford Dictionary).

An amalgamated system thinking and DSDE approach to data-driven decision-making could bring a better life for all. The purpose of this paper is to contribute to the fact that to provide a bigger picture of data-driven decision-making, system thinkers need DSDE practitioners and DSDE need system thinkers. The latter could lead to a new field that could be coined as Data System Thinking (DaSysT).

Despite this introduction, this paper is subdivided into four other sections, DaSysT, methodology, findings and discussion, and then conclusion.

2 | Data Science/Engineering and System Thinking approach

In the System thinking approach a sum is greater than its parts, and looks at how the whole is connected rather than separate parts. For example, in an organisational helicopter view illustrated in Exhibit 2, a system thinker will identify interconnections among the projects and people needed by each of the subprogrammes for all managerial and functionality of the organization. While the DSDE practitioner will concentrate on the data emanating from each of the departments making them practitioners working more in silos; the system thinker apply the skill of emergence. Emergence is the opposite of silos because it is where a larger idea or outcome is born from smaller parts (https://www.phoenix.edu/blog/what-is-systems-thinking.html). In this regard, the system thinker combines existing things to create something new, i.e. two departments can be combined to produce a new output.

Exhibit 2. System Thinking approach (Source: www.ensembletechnology.com)

In Exhibit 2, the organizational design and the organizational development can be combined to yield organizational human capacity and resources. In this regard, a system thinker will always keep in touch with all departments for proper organizational feedback and enhancement. The system thinker sees a bigger picture and the inter-relationships between seemingly unrelated areas. While a DSDE practitioner sees a solution within the relationships. Therefore, the

combination of system thing and DSDE emerges if an independent system or organizational-wide data-driven decision-making is needed. The latter introduces Data System Thinking.

The work by Peters (2014) illustrates the role of system thinking in the field of global health. Data-driven decisions together with a large body of theories, methods, and tools were identified as useful elements for decision-making in global health. In global health, Peters (2014) concludes that the latter elements could intervene and assist in improving people's health.

Kenett, Swarz, & Zonnenshain (2019) illustrate that in the past, system engineers used intuition and/or qualitative assessments to make decisions. These days, system thinkers in systems engineering are creating and using data-driven engineering requirements, engineering calculations, testing, modelling and simulations, to make decisive decisions (Kenett, Swarz, & Zonnenshain, 2019). In this regard, system thinkers provide the organization with data-driven decisions.

A workshop that was organized by the National Academies of Sciences, Engineering, and Medicine's Forum on Regenerative Medicine used a system thinking approach to tackle the regenerative medicine issues (Beachy, et al., 2021) to resolve the role of data science in regenerative medicine should be strengthened, as well as the importance of data science training and data literacy for the current and future regenerative medicine workforce (Plant, et al., 2022). According to Viljoen, et al., (2022), regenerative medicine focuses on harnessing the body's ability to heal itself by restoring the functionality of damaged tissues and organs. Different fields are role players within the regenerative medicine space, in the work of Beachy, et al., (2021), Plant, et al., (2022) and Viljoen, et al., (2022) system thinking and data science played a role in providing empirical influence on establishing regenerative medicine.

Langen, Falk & Mansouri (2022) illustrated that society produces information that leads to big data. Suitable tools and methods were needed to harvest and analyse such big data about products in the industry. In this regard, they proposed system thinking as one tool that can analyze the current state of the industry. Unfortunately, Langen, Falk & Mansouri (2022) neglected some data analytics simulation or modelling tools. For example, machine learning could assist them to predict and model the future of the products that are produced by the relevant industry.

All the studies mentioned about did not see a need for a discipline that could graduate individuals with a combination of system thinking and data science/engineering. The studies indicate a need for a combination of system thinking and data science/engineering discipline. There is less evidence in the literature for the latter move; hence this paper begins to create and contribute to that debate. The studies above show that the future demands multidisciplinary disciplines that will lead practitioners and experts to work in harmony with other disciplines and with a bigger picture of organisational outcomes. This move allows organization people to move away from outputs to outcomes.

3 | Methods

To strengthen and corroborate the contribution made in Section 2, a case studies method is used in this section. In this regard, this case study answers the following three important research questions that were used to foster this paper.

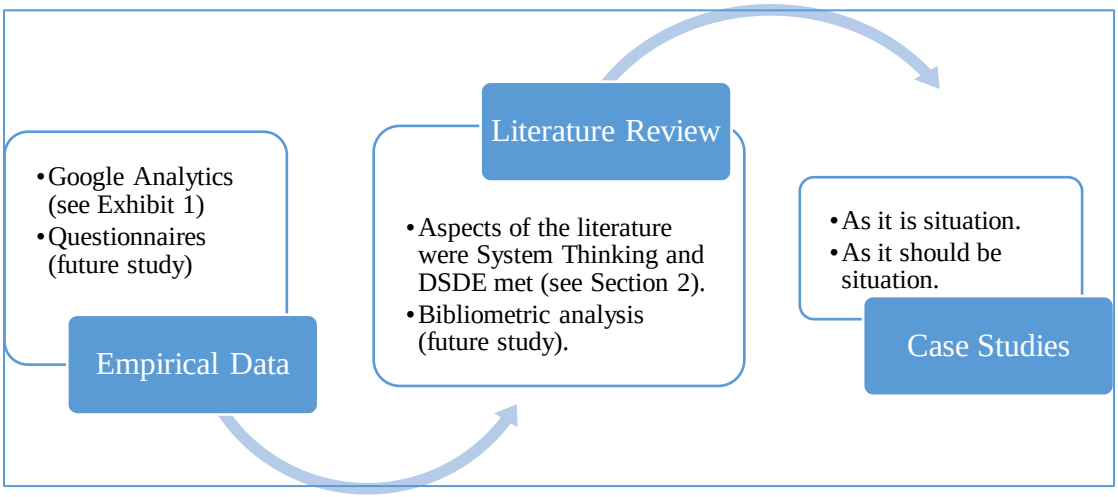
(i) What is the role of system thinking and data science/engineering skills in data-driven decision-making organisations? The answer to this question is provided in Section 2 above. The Section identified the importance of system thinking and data science/engineering-driven decision-making. It highlighted how different authors manage to lift the role of either system thinking or data science/engineering role in data-driven decisions or the combination of the two, to address the challenges that could be faced by any organization.

(ii) Does the combination of system thinking and data science/engineering give rise to a new discipline? The answer to this question begins to emerge in Section 1 whereby the importance of the dataset was discussed as the bedrock of all data-driven decision-making which enables analysis capabilities. The data scientists/engineers become responsible for the generation of datasets, the analysis and contribution to decision-making. However, their contribution in most cases is channelled to a particular challenge or opportunity within life in general or an organization. System thinkers used such datasets to create knowledge that addresses organization-wide challenges.

(iii) What are the skills needed in this new discipline? This Section will be answering this question by looking at the different skills highlighted in Exhibit 3.

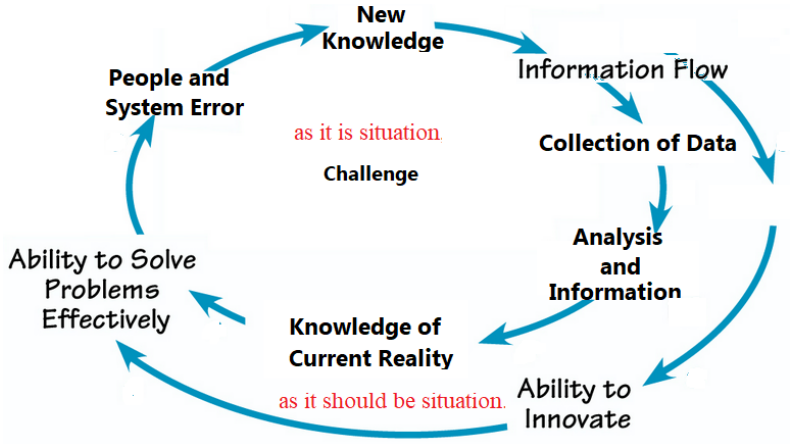
A pragmatic research philosophy was adopted to answer three questions raised in the study. Pragmatism is more of a practice and empirical point-of-view than mainly relying on abstract thinking. In this regard, pragmatism looks at the situation in two different scenarios, "as it is the situation" and "as it should be a situation". Pragmatism relies on mixed methods that lean on providing the idea or solution to the identified challenge or point of discussion. In this regard, the research design illustrated in Exhibit 3 was adopted for this study. It uses case study methods to visualize and validate the amalgamation of system thinking and data science leading to DaSysT.

Exhibit 3. Research Design (Source: Author).



The case study method uses the "As it is situation" which is the Real World scenario and "As it should be situation" which is the System Thinking about Real World scenario. Both scenarios were conducted with the help of the Loops diagram (Exhibit 4). The loops indicate that in the "As it is the situation" current data, information and knowledge about the challenge are analysed and used for information flow to decision-making; while that information is used to enable innovation and effective solutions which are provided in the "As it should be a situation".

Exhibit 4. Loop diagram adopted in this study (Source: Author).



4 | Findings and Discussion

4.1.1 | What is the role of system thinking and data science/engineering skills in a data-driven decision-making system?

In this paper to answer this question the experience of establishing a National Data Policy Observatory (NPDO) by the government of the Republic of South Africa will be reported using two perspectives: (i) as it is the situation, and (ii)
as it should be situation.

(i) As it is the situation

At the beginning of the COVID-19 pandemic declaration, the South African government and the different expect did not know how to manage the situation. Then during the advent of COVID-19, the South African government realised that there was a lack of publicly available national health data observatory that would provide data to anyone who might need it for research or decision-making process. Some of the government Ministers didn't even know how their ministries were affected by COVID-19 or how they could contribute to addressing the challenges brought by COVID-19. The Director Generals of the Department of Science and Innovation, the Department of Health, and the Department of Monitoring and Evaluation met and agreed to publicly make the data of those affected by COVID-19 in terms of

hospitalization, out-of-hospital and recovery available. These statistics were publicly made available following the direction of the World Health Organization (WHO).

These Departments fostered a Health Data Observatory (HDO). This observatory gave data scientists/engineers a chance to use data analytic tools to analyse the data to provide health-related decision-making. The process was conducted solely to provide COVID-19 statistics that led to several country, productivity and services lockdowns. If the system thinking approach was not adopted the situation could have been the same, collecting and analysing data to decide on lockdowns without thinking about what were the effects and impact of such a decision on livelihoods in terms of human living and business.

(ii) As it should be situation

When the system thinking approach was adopted, the COVID-19 situation and the country's healthcare system were analysed to realise what affects lockdowns, what are its effects and who are vulnerable individuals or businesses. This led to the establishment of the National Policy Data Observatory (NPDO) which is a government-led, national policy data observatory, governed through a secretariat led by the Department of Science and Innovation, Department of Planning, Monitoring and Evaluation, Statistics South Africa, the South African Revenue Services (SARS) and the Council for Scientific and Industrial Research (CSIR). The system thinking approach allowed the NPDO to comprise an electronic data platform hosted by the CSIR (see https://dataobservatory.csir.co.za/).

The platform is made up of the HDO and socio-economic data (drawn from different research sectors such as universities, science councils, civil organizations, industry, think tanks, etc) that is analysed by data scientists and displayed on the dashboard. Then the system thinkers use the data to provide advice for different systems. For example, they advised the government through the NATJOINTS, and inter-Ministerial Committee (IMC) to speed up the unemployment grant system because lockdowns were beginning to affect several families, Secondly, they advised the government to support through rebates, the South African business sector through the lockdowns. At the time of writing this paper, the analysed data was used to provide advice on vaccine rollout, South African sector indicators, and logistics industry systems. Exhibit 4 below highlights the NPDO welcoming webpage and the NPDO's different main dashboard screens.

Exhibit 4a. The NPDO main webpage (Source: www.csir.co.za).

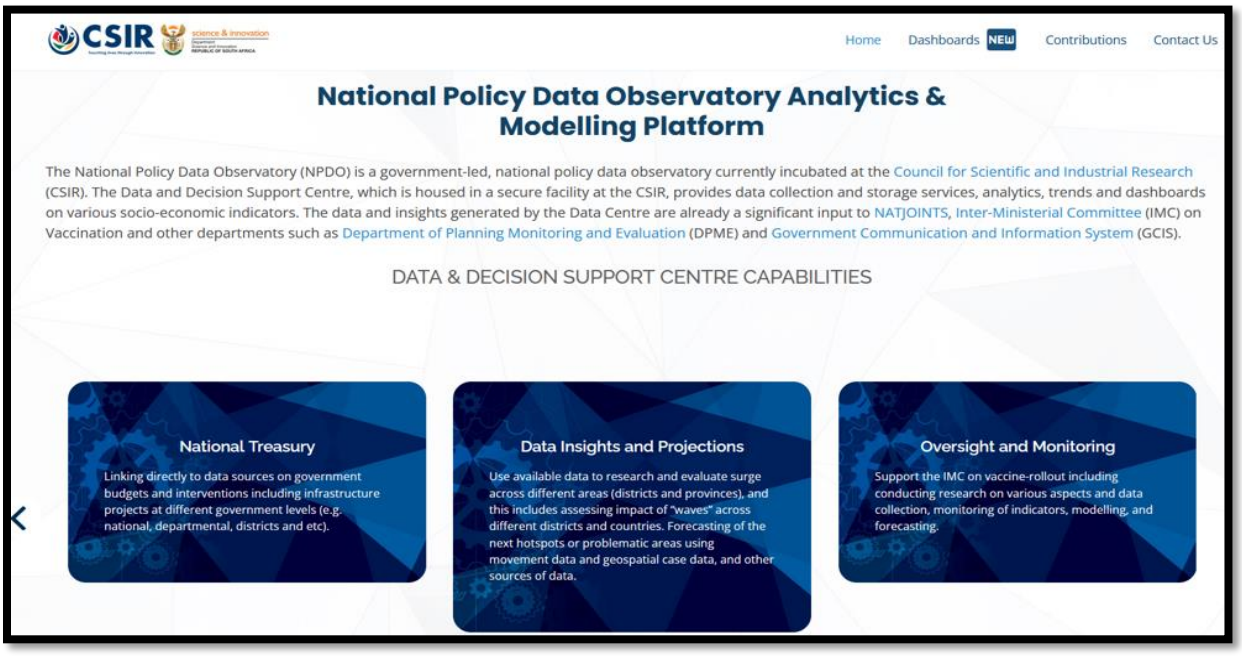
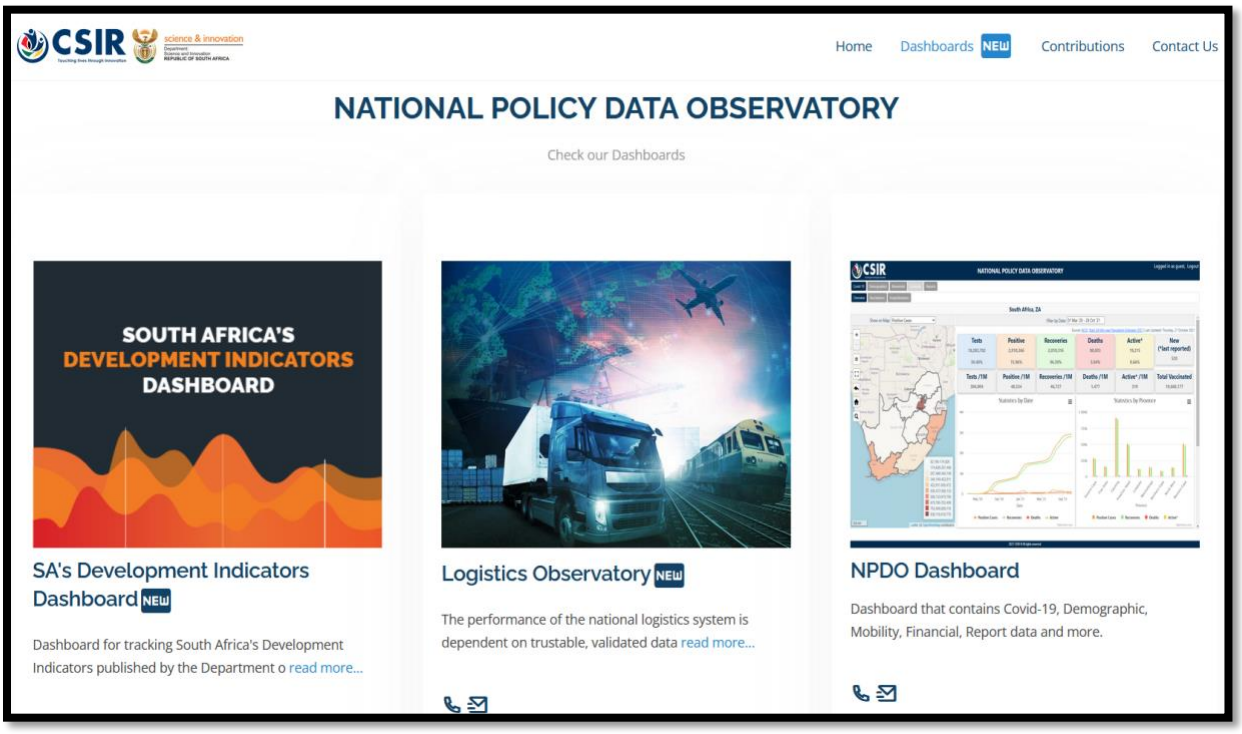


Exhibit 4b. NPDO main dashboard (Source: www.csir.co.za).



The involvement of system thinkers within this data science/engineering environment gave data scientists/engineers direction on which data is relevant for decision-making. The collection of the dataset was stored and analysed to yield trends and predictions of various socio-economic indicators that affected South African citizens. The amalgamation of system thinking and data science/engineering led to relevant indicators to be collected. Some of these indicators are poverty and food security indicators, social living (i.e. gender-based violence, social traders), unemployment, etc. The above case illustrated that when system thinking meets data science/engineering a more holistic decision and solutions towards a problem is primarily taken.

4.1.2 | Does the combination of system thinking and data science/engineering warrant a new discipline?

To answer this question, the practical work that was conducted as a system engineer at the kraft and paper company, SAPPI Kraft Pty Ltd, Ngodwana Mill, will reported using two perspectives: (i) as it is situation, and (ii) as it should be situation.

(i) As it is the situation

Ngodwana Mill has been in operation since 1966. The mill produces paper pulp for its consumption, as well as newsprint, and kraft linerboard. In 2004, the author was provided with the role of being a Mill Systems Engineer reporting to the company's Group Specialist Systems Engineer. One of the biggest challenges that SAPPI Mills had, was a problem of a bad smell odour that was due to the use of sulphide on wood chips from the tree bulk. Exhibit 5 illustrates the company's odour exhaled from the chimneys and towers. This odour affected neighbouring citizens who were staying closer to the company's premises. It led to a lot of environmental issues and the company embarked on solving this challenge. As a system engineer, we begin to collect and develop different datasets from all company chimneys and points that exhaust smoke. Different sensors and industrial information technology tools such as PLC and SCADA were used to achieve the latter task. The collected data was captured by data professionals and analysis was conducted by a mill systems engineer.

Exhibit 5. SAPPI Kraft Ngodwana Mill (Source: www.sappi.co.za).

(ii) As it should be situation

Using the dataset collected through different methods mainly sensors that were placed within the company premises, three questions were important for the company's executive: (i) What does the company do to reduce the odour? (ii) Does the odour pose occupational health and safety hazards to employees, and people living and working in the neighborhood? And (iii) Were there any experts invited to do the consultant work?

In an environment such as SAPPI Kraft, the pollution rate is frequently increasing leading to the creation of chemical pollutants that affect the environment. The pollution could be produced by most of the factors illustrated in Exhibit 3. All these factors produce data. Each section could have its own DSDE practitioners, however, DaSysT practitioners will have a helicopter view and understanding of the entire process of the energy generation plant. They use data to determine the number of pesticides neurotoxic, acids, spills, etc that affect the stream needed to generate water for the organisation. They will use machine learning skills to predict the amount of pollutants that could disturb the organization and system thinking to provide a strategic direction that could influence and direct the organization.

The application of the system thinking approach in the SAPPI Kraft situation led to the use of a collected dataset and analysis that could lead to what information could be learned from such a dataset. The dataset was created from the collection of data from three consecutive years. As illustrated by Exhibit 4, the collected dataset led to the knowledge of the then-current reality which after debugging the people and system errors that might have existed led to new knowledge that created the ability to innovate and provided an effective solution which was agreed upon as the company's decision. Exhibit 6 provides a table that summarises the findings from the dataset.

Exhibit 6. The SAPPI Kraft Ngodwana Mill odour prediction for decision making.

Intervention	Description and impact on sulphide odour	Predicted implementation difficulty	Predicted effect	Directly connected to Odour
What does the company do to reduce the odour?	The company introduced new systems for collecting data and a new team for analysis. The team was led by the Mill System's Engineer.	⊕	✓✓	R40 million
Does the odour pose an occupational health and safety hazard to employees, and people living and working in the neighborhood?	An electronic portal was used to schedule the collection and storage of employees' well-being data obtained through workshops and consultation. The same portal was used to collect data from workshops conducted for the public living near the Company's premises.	⊕	✓✓✓	90 Employees 220 public

Were there any experts invited to do the consultant work?	A standardized structured triage tool was developed and provided by the experts. It is used to perform a triangulation technique to validate the collected and analyzed data. The tool also collected a new dataset.	⊕⊕	✓	22 cases from 15 units of the Company's premises were used.

Predicted by the workshop participants; ⊕ easy, ⊕⊕ rather easy, ⊕⊕⊕ very difficult, ✓ moderately high effect, ✓✓ rather high effect, ✓✓✓ high effect;

Most employees' well-being data obtained through workshops and consultations predicted no hazard. About 60% (54) of 90 Employees and 73% (161) of 220 non-employees responded that the use of the electronic portal for datasets has been highly effective and easy to use. Of which the trend shows that they will continue to use the portal. The dataset revealed decisions that were organized for public consumption as illustrated by Exhibit 7.

Exhibit 7. The outcome of the systematic analysis of the SAPPI Kraft Ngodwana Mill odour dataset (Source: www.sappi.co.za).

4.1.3 | What are the skills needed in this new discipline?

There have been so many publications on future skills and future careers. In 2018, Professor Longbing Cao, a professor of data science, who holds two doctorates, PhD in Pattern Recognition and Intelligent Systems (from the Chinese Academy of Sciences, China) and PhD in Computing Science (from the University of Technology Sydney, Australia) published a book titled: "Data Science Thinking: The Next Scientific, Technological and Economic Revolution". The book explores how data science transforms existing science, technology, industry, economy, profession and education. In the book, Prof Cao looks at (i) a comprehensive study of data science, technology and economy perspectives, (ii) presents rich and deep thinking and insights into data-driven research, innovation, industrialization, and opportunities, (iii) addresses the needs of decision-makers who are responsible for managing the new realm of data science, and (iv) investigates the mindset and skillset of data scientists to help define the projected course of the data revolution (Cao, 2018).

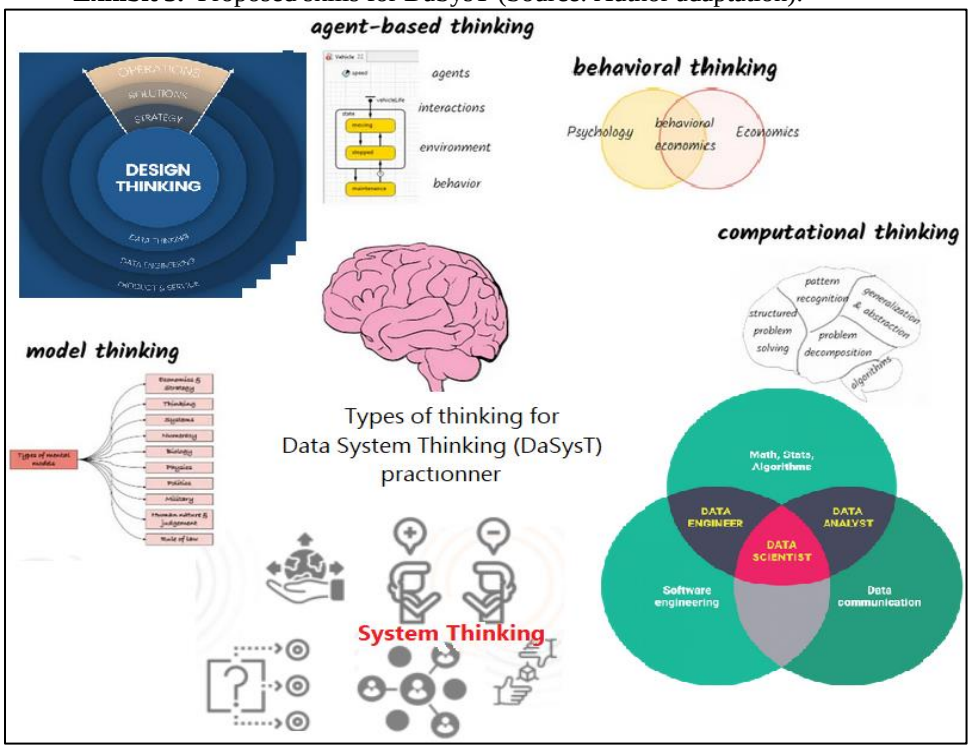
Data Science Thinking (DST) refers to methodologies, processes, structure, traits and habits of the mind in handling data problems and systems (Cao, 2018). Although, Cao (2018) began to deliberate on data science skills, unfortunately, no discussion of the latter beginning to argue for establishing a field or discipline of Data Science Thinking (DST).

Notably, Prof Cao, holds two PhDs, meaning there was a need for him to be an expert in both areas (Pattern Recognition and Computer Science). In the past, during Prof Cao's university days as a student, the two fields could have not been closely related, unlike today. Arguably, scholastic debates of the past were so entrenched in philosophical approaches that created a separate way of solving problems and managing life. This is still entrenched in how some scholars researchers or industrialists relate to interdisciplinary or multidisciplinary matters. The latter is one of the challenges that creates a bridge between university and industry since trained students become one-sided and fail to integrate technical, social and business understanding.

The aforementioned departmentalism continues and affects how individuals begin to lack a balanced way of solving problems emanating from the spheres of societal influence (Cunningham & Rogers, 2007). The seven spheres of societal influence are (i) religion, (ii) family, (iii) education, (iv) government, (v) media, (vi) celebration (arts, entertainment and sports), and (vii) economics (Business, Science, and Technology).

The seven spheres of societal influence point to viewing the world as a system that has different sub-systems. System thinkers are equipped to view and manage every problem as a system. Notably, all these spheres produce data, which means the role of DSDE cannot be ignored, it is significant and helps in making informed decisions.

Exhibit 3. Proposed skills for DaSysT (Source: Author adaptation).



If an individual masters the art, science and engineering of managing these spheres, well, then that would lead to a balanced living. Arguably, the latter would be one of the fields needed for the future, studying and managing multidisciplinary fields. This is a gap that a discipline such as DaSysT will address. Exhibit 3, illustrates the type of skills that the discipline of DaSysT will provide to its candidates.

5 | Contribution

This paper intended to show that the amalgamation of two interdisciplinary approaches leads to a multidisciplinary approach could lead to fostering new avenues and challenge the current status core.

In this paper, system thinking meeting data science/engineering was discussed and their combination led to the proposal of the Data System Thinking discipline. This paper uses Cunningham and Rogers (2007), and Cao (2018) as a necessary benchmark to contribute to DaSyT as the new discipline that needs to be harnessed for a balanced living.

The proposal is at its earlier stages and this paper is one of many that seek to source the scientific community's thinking around the idea of DaSyT. To strengthen the proposal different scenarios have been provided, some of which could be seen from other perspectives; however, their debate contributes significantly to the idea of DaSyT mindset and skillset.

6 | Conclusion

One does not need to have two PhDs to be an expert in the field of DaSysT but needs training from undergraduate to postgraduate in DaSysT disciplinary thinking. In this regard, a standard curriculum needs to be developed. The DaSysT covers a wider scope beyond scientific, technological and economic revolution but also covers humanity, political, and social aspects. In this regard, the DaSysT should include concepts such as data analytic thinking, cognitive thinking, statistical thinking, mathematical thinking, creative thinking, critical thinking, and project problem-solving skills. A DaSysT will be a techno-pol-socio-economic individual.

Is there a need for a DaSyT degree and/or curriculum? To participate in the future mindset and skillset it seems like among other study disciplines, there will be a need to train students in DaSyT. Can there be a Bachelor of DaSyT degree? A deep study on curriculum development among fields that relate to developing the DaSyT will be needed. In this regard, the future intent of this study will be to explore and strengthen the idea of establishing DaSyT as a new multidisciplinary degree or discipline. Furthermore, a critical systems heuristics approach towards professionalising DaSyT will be conducted.

7 | References

Cao, L. (2018). *Data Science Thinking*. In: Data Science Thinking. Data Analytics. Springer, Cham. https://doi.org/10.1007/978-3-319-95092-1_3

Beachy, S.H., Nicholson, A., Teferra, L., Hackmann, M., & Addie, S. (2021). Applying Systems Thinking to Regenerative Medicine. *Proceedings of a Workshop. National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Health Sciences Policy; Forum on Regenerative Medicine*. Washington (DC): National Academies Press (US). ; 2021 Mar 26. ISBN-13: 978-0-309-15182-5ISBN-10: 0-309-15182-1

Goutam, M.A., Ramesh, W.U., Eladl, M.A., El-Sherbiny, M, Elsherbini, D.M.A., Sukumar, A., Kannampuzha, S., Ravichandran, M., Renu, K., Vellingiri, B., Kandasamy, S., & Valsala G.A. (2022). Mixed Contaminants: Occurrence, Interactions, Toxicity, Detection, and Remediation. *Molecules. 16;27(8):2577*. doi: 10.3390/molecules27082577. PMID: 35458775; PMCID: PMC9029723.

Kenett, R.S., Swarz, R. S., & Zonnenshain, A. (2019). *Systems Engineering, Data Analytics, and Systems Thinking*. Chapter 1 in Systems Engineering in the Fourth Industrial Revolution. https://doi.org/10.1002/9781119513957.ch1

Langen, T., Falk, K., & Mansouri, M. (2022). A Systems Thinking Approach to Data-Driven Product Development. *International Design Conference – Design 2022*. https://doi.org/10.1017/pds.2022.194

Loren Cunningham, L., & Rogers, J. (2007). *The Book That Transforms Nations: The Power of the Bible to Change Any Country*. YWAM Publishing.

Peters, D.H. (2014). The application of systems thinking in health: why use systems thinking? *Health Res Policy Sys 12, 51* https://doi.org/10.1186/1478-4505-12-51

Plant, A.L., Piscopo, N., Saha, K., Zylberberg, C., Roy, K., Tsokas, K. Schumm, S.N., & Sarah H. Beachy, S.H. (2022). Implementing systems thinking and data science in the training of the regenerative medicine

workforce. *NPJ Regen Med*. *24;7(1):76*. doi: 10.1038/s41536-022-00271-2. PMID: 36566283; PMCID: PMC9790008.

Viljoen I.M., Hendricks, C.L., Malherbe, H.L., & Pepper, M.S. (2022). Regenerative medicines: A new regulatory paradigm for South Africa. *Biochimie*. *196:123-130*. doi: 10.1016/j.biochi.2022.02.010. Epub 2022 Mar 3. PMID: 35248613.