cenfri

Public sector data sharing frameworks

Module 3

of a report submitted in relation to a public sector data cataloguing, classification and sharing project undertaken as part of the Rwanda Economy Digitalisation Programme

March 2023

In partnership with

mastercard foundation

GOVERNMENT OF RWANDA

Republic of Rwanda
Ministry of ICT & Innovation

Author(s)

Prof. Patrick McSharry

Claude K. Migisha

Cenfri South Africa

Tel. +27 21 913 9510
Email: info@cenfri.org

The Vineyards Office Estate
Farm 1, Block A
99 Jip de Jager Drive
Bellville, 7530
South Africa

PO Box 5966
Tygervalley, 7535
South Africa

Cenfri Rwanda

Tel. +250 788 312 132
Email: info@cenfri.org

www.cenfri.org

mastercard foundation

GOVERNMENT OF RWANDA

RISA Rwanda Information Society Authority

Table of contents

Introduction 1
Data sharing guidelines 2
The checklist 7
Requirements 8
Regulation 9
Technical 9
Funding 9
Recommendations 10
References 12
Appendix: Use-case template 13
Sample use-case 13

List of tables

Table 1: Examples of data sharing initiatives 4
Table 2: Checklist to be completed before initiating data sharing project 8

List of figures

Figure 1: The decision matrix 5

Introduction

Back in 2016 when Klaus Schwab, Founder and Executive Chairman of the World Economic Forum (WEF), coined the term “4th industrial revolution” Rwanda was one of those countries that bought into the idea. As a result, Rwanda embarked on a journey of developing regulatory frameworks that would ease the access, sharing and processing of data across government and with external stakeholders such as the private sector and its citizens.

Public sector data can be leveraged to increase access to services, make service delivery more efficient, inform policymaking, such as predicting tea productivity at the factory level (McSharry at al., 2016), and quantifying the economic cost of conflict (Mawejje & McSharry, 2021). Of course, it is paramount that studies of personal data and new applications that impact society are carefully evaluated and tested before embarking on a program to implement innovative technologies. AI ethics is an important part of the educational program at CMU-Africa, helping to safeguard the implementation of AI-enabled solutions (McSharry, 2023).

The Rwanda Economy Digitalisation Programme, implemented by Cenfri in partnership with the Rwandan Ministry of ICT and Innovation and the Mastercard Foundation, is an initiative that involves collaborating with stakeholders to leverage insights from data analysis to improve policymaking, catalyse innovation and, ultimately, improve livelihoods.

As part of this programme, Cenfri hired a team of consultants to work with the Government of Rwanda (GoR) – specifically the Rwanda Information Society Authority (RISA) and the Ministry of Education (MINEDUC), the Ministry of Agriculture (MINAGRI) and the Ministry of Finance and Economic Planning (MINECOFIN) in three domains:

- Data cataloguing (Module 1)
- Data classification (Module 2)
- Data sharing (Module 3)

This module covers some of the guidelines for **data sharing**. Data cataloguing and data classification are covered in previous modules.

While this consultancy assignment responded to the context in Rwanda, the needs of the government institutions with whom we worked, and specific programme objectives, this overview is likely to be of use to any public sector entity that is starting out on its data-for-decision-making journey. The datasets, examples and use cases referenced in this document are for indicative purposes and should be substituted with whatever is appropriate in the context in which these data cataloguing principles are being applied.

Data sharing guidelines

Data has value only when it is consumed. Data locked away and unutilized represents an opportunity cost. The value of data can increase exponentially if it is shared and combined with other sources of data. By sharing data, it is possible for innovation to take place, and this can offer considerable advantages for governments and both the public and private sectors. In addition, the citizens of a country with an active data-sharing policy stand to benefit from improved services, greater efficiency, and the promotion of social goods. The open data movement has already led to governments worldwide sharing their data. Much work still remains to be done, with many datasets remaining locked away, undiscovered, or difficult to access.

At present, many governments are engaging with the public and private sectors and actively experimenting with different ways of accelerating data sharing. A key part of this is promoting collaboration between the data gatekeepers and those that can deliver solutions to extract value from data. Data sharing is already helping to solve problems in many areas such as labour markets with a focus on the future of work, enhancing skills and creating jobs. The creation of smart cities depends on access to real-time data about mobility to improve transport systems and reduce congestion (Paul & McSharry, 2021). Information about electricity consumption can help to improve service delivery (Otieno & McSharry, 2019a; Otieno & McSharry, 2019b). By combining satellite imagery and mobile data, it is possible to predict multidimensional poverty (Njuguna & McSharry, 2016) and to identify which roads and routes to prioritise for investment based on mobility trends (Uwizera et al., 2022). Nonetheless, barriers to data sharing persist. Privacy concerns and several high-profile data breaches have shone a spotlight on data security and present substantial challenges to the safe sharing of data. Public sector organizations tend to be risk averse and for many, the privacy risks outweigh the unquantified potential benefits and saying no to sharing data is the easiest stance to take.

The potential rewards from data sharing are often difficult to measure. The best way to provide motivation is to look at the landscape of existing projects and the various areas where data sharing has already added value. The benefits range from improvements in efficiency to reduced costs to the ability to scale up services using automated systems. Data-sharing projects can be broadly categorized as enablers of novel applications, such as extracting actionable insights, optimising outcomes, improving decision-making, facilitating predictions, or unlocking innovation. The following matrix highlights some relevant case studies from Rwanda and other countries in these enabling areas:

Enabling	Country/Sector	Description
Insights	Netherlands / Smart Cities	AMdEX (Amsterdam Data Exchange) is a data exchange initiative by the Amsterdam Economic Board that aims to accelerate the transition to a fair data economy. It collects city data held by government agencies, companies, and others to provide broad access to data for researchers, businesses, governments, and individuals in a secure marketplace. Use cases include logistics, smart buildings, aircraft maintenance, and electric car charging (AMdEX, 2023).
	Rwanda / Public	Rwanda Revenue Authority (RRA) in partnership with the International Growth Centre, CMU-Africa, and GIZ considered the use of an AI-enabled approach to predict land and property prices. An accurate machine learning model was developed to predict these prices and now forms the basis of an equitable taxation system (Bachofer et al., 2019) in Rwanda.
	Rwanda / Transport	Rwanda Transport Development Agency (RTDA) in collaboration with CMU-Africa and The University of Rwanda shared and studied data from call detail records (CDRs) and satellite imagery to evaluate road conditions. A novel approach for infrastructure prioritization using deep learning and big data analytics was developed. The result was a ranking of roads which could be prioritized for construction given mobility trends (Uwizera et al., 2022).
Optimisation	South Korea / Transport	Seoul Owl Bus is a data-driven service. In Seoul, South Korea, the metro system shuts down from midnight to 5 am. The municipal government used its citizens' late-night calls and texts to plan routes for a new night bus service. The partnership between the government and Korean Telecom (KT corporation) was able to create a service that not only saved late-night commuters $1.2 million in taxi fares from 2012 to 2014, reducing car trips by 2.3 million annually (city buses emit 80 per cent less carbon monoxide than private cars); it was also beneficial to low-income communities, providing them with a viable solution to commute home to the outer boroughs after working night shifts in the city (Seoul Solution, 2023).
	Nigeria / Transport	Lagos State Bus Rapid Transit (BRT) in collaboration with CMU-Africa investigated urban travel demand in Lagos using individual commuter trip data from purchased tickets (Paul & McSharry, 2021). Simulations using fixed and dynamic bus scheduling demonstrate that the average waiting time could be reduced by as much as 80%. The load curves, insights and the approach developed will be useful for informing policymaking in Lagos and similar African cities facing the challenges of rapid urbanization.
	Sub-Saharan Africa / Health	Population-based HIV Impact Assessment (PHIA) data from sub-Saharan Africa allowed for a machine learning approach to identify predictors for HIV status, which can facilitate targeted screening interventions that improve health care. Amongst the eight most predictor features in both sexes were: age, relationship with family head, the highest level of education, highest grade at that school level, work for payment, avoiding pregnancy, age at the first experience of sex, and wealth quintile. It was predicted that 4·14% of males and 10.81% of females are at high risk of infection (Mutai et al, 2021).

Enabling	Country/Sector	Description
Predictive analytics	Rwanda / Development	Rwanda Utilities Regulatory Authority (RURA) shared mobile call detail records with CMU-Africa. Multidimensional poverty at the administrative sector level in Rwanda was accurately predicted (out-of-sample correlation of 88%) using a combination of features from satellite and mobile data and has potential to monitor development progress and target new interventions (Njuguna & McSharry, 2016).
	Mozambique, South Sudan, Ethiopia / Development	World Bank country economic reports rely on surveys and official statistics to inform policymakers. Big data sources (satellite, mobile) complement these traditional datasets and now offer the means to better predict economic activity at subnational level. Satellite imagery (nighttime lights, land cover, weather, and pollution) was used to estimate subnational economic output in Mozambique, Ethiopia, and South Sudan and to understand the drivers of economic growth (McSharry et al., 2019; World Bank, 2021; World Bank, 2022).
Decision-making	Global / Transport	Google Waze is a platform that provides real-time anonymised crowdsourced traffic data collected from participating drivers. In its Connected Cities programme, it shares its large amount of traffic data with government agencies, which can use this data to better inform policy or quickly deploy traffic assistance if needed. Some use cases of Waze data include reducing traffic; decreasing accident response times; and repairing infrastructure.
	Rwanda / Health	The Ministry of Health in collaboration with CMU-Africa, collected weight data using tablets at health clinics across Rwanda. This was shown to be more accurate than both paper-based records and the World Health Organization (WHO) growth curves and offers a means of identifying malnutrition in young children (Brown & McSharry, 2016).
	South Sudan / Economy	World Bank shared data with CMU-Africa to assess the economic impact of conflict. The study estimates the output loss in South Sudan because of the double shock of the protracted post-independence conflict and macroeconomic crisis. Using the synthetic control method for comparative studies, the analysis found that the cumulative loss in the growth rate of real per-capita gross domestic product (GDP) was 69.63% (or a yearly average of 15.65%) over the period 2012–2018 (Mawejje & McSharry, 2021).
Innovation	UK / Finance	Open Banking: Since 2018, a regulation from the Competition and Markets Authority mandates that UK-regulated banks allow authorised providers (such as licensed Startups offering budgeting apps, or other banks) direct access to customer account information and data at transaction level through APIs. Use cases include new tools that will positively impact on vulnerable communities' financial inclusion.
	Kenya, Tanzania / Energy	PowerGen Renewable Energy shared consumption data from microgrids in Kenya and Tanzania with CMU-Africa. This enabled a comprehensive study of the consumption patterns of households connected to microgrids and helped identify accurate demand forecasting models and a methodology for segmenting customers that helps providers increase the potential of solar power in the region (Otieno, Williams & McSharry, 2018a; 2018b).

Table 1: Examples of data sharing initiatives

NESTA (2020) recommends using a decision-matrix approach (see figure 1) to promote data sharing. This practical approach involves six components, and a canvas is used to frame the relevant questions for each one.

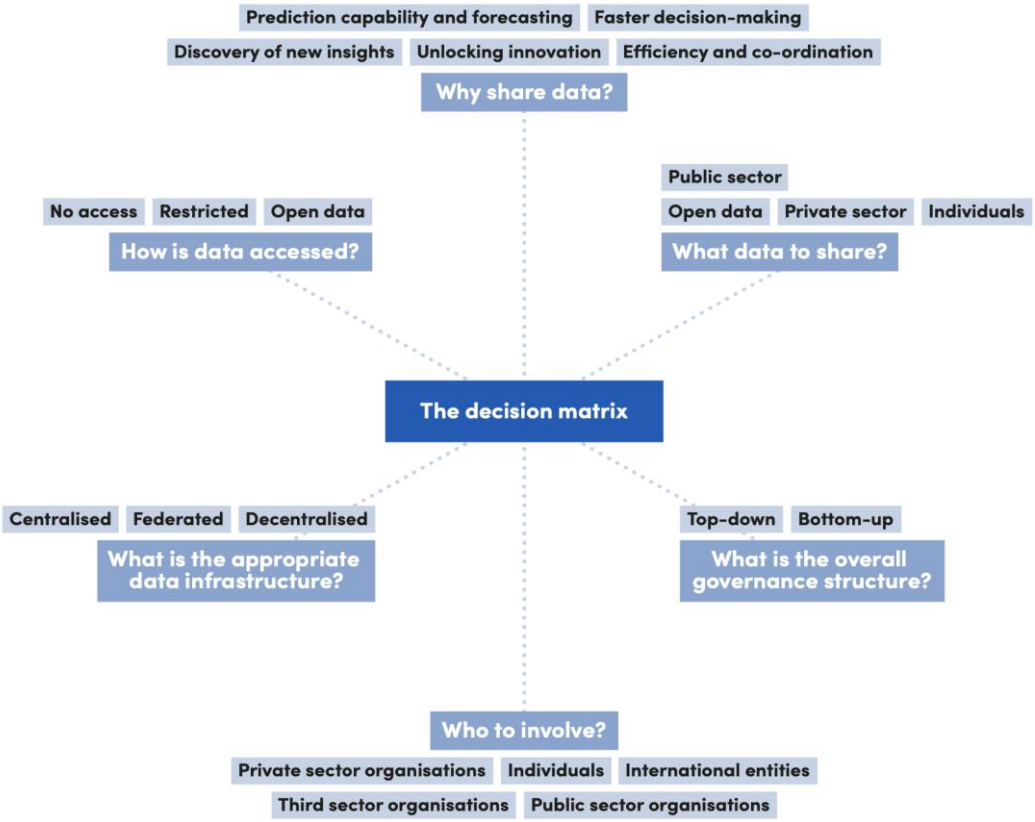

Figure 1: The decision matrix

Why share data?		
Questions	**Answers**	**Examples**
• What type of problem are you trying to solve? • What other things could you do instead of sharing data? • Are there examples of similar projects that could help identify benefits and make the case for the partnership to be formed? • Is the output of the initiative going to be a one-off analysis or does it require continuous data sharing? • Is the data you want to share going to be used for strategic reasons or operational? • Are you starting with a specific problem in mind or with the data? • How is the project being communicated?	Discovery of new insights Unlocking innovation Faster decision-making Increased prediction capability and forecasting Optimised process efficiency and coordination	AMdEX

What data to share?		
Questions	**Answers**	**Examples**
• What data would you need, and how much of it is already available? • How are you going to incorporate data if it becomes available in the future? • What are the data gaps, and can you mitigate the effect of inequality in data availability? • Where do these datasets fit on the spectrum of closed–open? • How do we understand public norms around this type of data (e.g., mandatory portability)? • What form does each of the datasets take? • How mature is the data? • How shareable and easy to link is it? • What is the level of anonymisation of the data?	Open data Public sector Private sector	Data Catalogue

Who to involve?		
Questions	**Answers**	**Examples**
• Who is initiating the partnership? • What are the incentives for stakeholders to take part in the partnership? • Assess what the value distribution • is in the partnership (i.e., is there • equity of value among all stakeholders? Who will not benefit?) • Where is the funding coming from? • Who holds data that relates to this use case? • Is there anyone else besides data providers who are needed to make this project work? For example, • do you have expertise in terms of data science? • When data subjects' involvement is required, how are they represented in the decisions? • Is there an option for opting in or out?	Public sector organizations Private sector International organizations Individuals	ODI Mapping Data Ecosystems NESTA Partnership Toolkit

How is data accessed?		
Questions	**Answers**	**Examples**
• Will data be accessible? • If so, what is the access model to the data? • Is there a need to have multiple access models? • If restricted, what kind of restriction does it require and why?	No access Open access Restricted access	Data Classification Matrix

What is the overall governance structure?		
Questions	**Answers**	**Examples**
• How will each party report progress? • How will decisions be made and in which forum? • How will conflicts between the parties be resolved? • Would anyone external be brought in? • Who is accountable for what? • How is risk being managed and, if needed, mitigated?	Top-down Trusted intermediaries Bottom-up	Royal Society and British Academy, Data management and use ODI Lessons from pilots

What is the appropriate data infrastructure??		
Questions	**Answers**	**Examples**
• What is the structure that best fits the purpose and why? • Will the data need curation? If so, who is responsible for it? • Futureproofing: what measures have been considered if circumstances change?	Centralised Federated Decentralised	Data.gov.uk is the central platform that provides storage and open access to a wide variety of government data sources for the UK.

Having answered all the questions in the decision matrix, the next step is to ensure that the foundations are in place for initiating data sharing. Two elements, namely a checklist and requirements, are proposed as foundational prerequisites to any data-sharing efforts.

The checklist

The purpose of the checklist is to ensure that key items have been adequately addressed before initiating data sharing. Otherwise, it is likely that issues will arise that will hinder the project. The three boxes to check are

1) senior buy-in
2) incentivization
3) equitable contributions

Table 2 provides a summary of the considerations and important questions to ask, the individuals responsible for signing off and the outputs required to demonstrate that the boxes have been checked.

Checklist	Considerations	Who? What?	Outputs
Institution senior buy-in	Are you engaging with people senior enough to make decisions and unlock issues when they arise? What signals have you been given that there is senior buy-in, on both sides?	Minister; Permanent Secretary; Chief Digital Officer	Internal Memo, Memorandum of Understanding (MOU) signed
Incentivization	For a data sharing partnership to work, all parties must benefit. You might have to help your potential partner to understand how they will benefit from the partnership. Benefits can be derived from the following incentives: (1) legal; (2) economic (increasing revenue or decreasing costs); (3) non-economic (benefits in kind, social good, reciprocity).	Regulators, Public Sector bodies, private sector, civil society	Contractual agreements signed
Equitable contributions	All partners should make an equitable contribution to the partnership. Making an equitable contribution does not mean making an equal contribution. Examples of contributions include money, time, resources, expertise, connections, or data. What is required for a partnership to succeed is for all stakeholders to be clear and happy that the contributions brought into the partnership are fair and valuable to all parties.	Money Time Resources Expertise Connections Data	Partnership agreement, Memorandum of Understanding (MOU) signed

Table 2: Checklist to be completed before initiating data sharing project

Requirements

The requirements for initiating data-sharing can be categorized into three areas:

1) Regulation
2) Technical
3) Funding

It is recommended to seek adequate advice from both legal and technical experts when designing the project. The following is a suggested list of tasks to undertake for each requirement:

Regulation

- Identify relevant legal and regulatory requirements. These can be defined at the institutional level, national level or for specific public bodies.
- Ensure that any data-sharing initiative complies with these requirements.
- Communicate to relevant authorities that requirements are satisfied.
- Undertake an annual review of evolving legal requirements.
- Ensure effective communication takes place with relevant stakeholders.

Technical

- Carry out a periodic evaluation of data maturity at various public sector bodies.
- Undertake annual technological infrastructure audit.
- Assess the data quality.
- Require interoperable systems.
- Establish the minimum acceptable frequency of data sharing.
- Promote the creation of application programming interfaces (APIs).
- Design the data-sharing architecture.
- Decide whether architecture will be a commercial pre-built solution or tailor-made.

Funding

The type of funding structure will reflect the specific type of partnership that is involved in undertaking data-sharing. The considerations are as follows:

- Investment in one lead or coordinating organization such as a government.
- Equal funding: several partners share the funding costs equally.
- Tiered funding: tiers are organized based on the level of input or benefit-in-kind.
- External funding: a grant is utilized for a specific project or pilot.
- Commercial funding: income generated from selling data is used to fund the project.

To conclude, the above (i.e., the data-sharing decision matrix, the checklist, and the requirements) could form the basis of a data governance strategy for public sector entities and ultimately unlock data sharing across governments and beyond.

Recommendations

From working with institutions to come up with data catalogues, to providing data classification approaches, it has been possible to better understand the barriers to sharing data within the government. This experience has motivated the proposal of data sharing guidelines that could facilitate the processes that are required to unlock data and enable data-driven decision-making and innovations. The following are the key recommendations:

- Data legislation in Rwanda is rapidly growing and more developments are expected as the country continues to position digital technologies as an important engine for socioeconomic development. However, as highlighted in previous sections, a lot still needs to be done to unlock the value of data for various actors. The starting point should be to enact a data sharing law that would see more data being shared across the public sector or with the private sector, innovators, academia, and civil society. Understandably, such a law would need to align with existing ones (i.e., data protection law, access to information law). The proposed data classification and sharing guidelines may form the basis for the data sharing law.
- After passing the aforementioned law, some kind of a public sector data governance framework that is adopted across all public sector entities will be required. This is currently missing. Once put in place, such a framework would, on the one hand, ensure all emerging approaches are integrated with existing laws (i.e., data protection), but also, on the other hand, align with best practices and ongoing initiatives (i.e., Government service bus, etc.) across government. In doing so, there is a greater chance that public sector bodies will move away from duplicating efforts and working in silos.
- The proposed data cataloguing efforts proved useful for the pilot institutions. It's recommended that, with the coordination support of RISA, all government institutions may adopt the practice and have it as a mandatory internal practice to uphold data management best practices. A more robust, web-based version of the data catalogue, which is ideally uniform across government, will need to be developed per the industry standards.
- The recent appointment of Chief Digital Officers within key public sector entities is commendable. To continue fostering the "data culture", these CDOs need to be equipped in order to act as data champions or data stewards within their respective institutions. As such, they would be responsible for ensuring the quality of the solution which is fit for purpose and captures the organisation's data assets.
- A whole-of-society approach is recommended to fully reap the dividends offered by the data economy. For a robust data ecosystem in Rwanda, it's critical that both the data supply and demand sides collaborate. These two sides need to come together and discuss data-related matters (i.e., access, privacy, re-use, etc.) and co-create relevant use-cases. In order to motivate the creation of use-cases, a framework was developed (see Appendices). A national working group that convenes on a regular basis bringing together various actors (i.e., public sector, innovators, private sector, academia, civil society, etc.) to discuss data-related

matters is needed in Rwanda. Data and statistics differ, hence the need to complementarily add to a few attempts to convene stakeholders done by the NISR.

Readers should note that the above recommendations are applicable in the Rwandan context and may not be equally applicable elsewhere.

For information on data cataloguing and data classification refer to modules 1 and 2.

For more information on this project, contact ***Marcellin Nyirishyaka.***

References

Mawejje, J. & McSharry, P.E. (2021). The economic cost of conflict: Evidence from South Sudan. Rev Dev Econ. 2021; 00: 1– 22.

McSharry, P.E., Swartz, T. & Spray, J. (2016). Index Based Insurance for Rwandan Tea. International Growth Centre.

McSharry, P.E. (2023). Promoting AI Ethics Through Awareness and Case Studies. In "AI Ethics in Higher Education: Insights from Africa and Beyond". Editors: C. Corrigan, S. Asakipaan, J. Kponyo & C. Lütge, Springer Briefs in Ethics, Springer Publishing, Berlin, Germany.

Mutai, C.K., McSharry, P.E., I. Ngaruye, Musabanganji, E. (2021). Use of machine learning techniques to identify HIV predictors for screening in sub-Saharan Africa. BMC medical research methodology 21 (1), 1-11.

NESTA (2020). Data sharing toolkit. National Endowment for Science, Technology, and the Arts. https://www.nesta.org.uk/toolkit/data-sharing-toolkit

Njuguna, C. & McSharry P.E. (2016). Constructing spatiotemporal poverty indices from big data. Journal of Business Research 70: 318-327.

Paul, O. & McSharry, P.E. (2021). Public transportation demand analysis: a case study of metropolitan Lagos (May 25, 2021). SSRN: http://dx.doi.org/10.2139/ssrn.3852747

Uwizera, D., Ruranga, C. & McSharry, P.E. (2022). Deep learning inter-city road conditions in East Africa for infrastructure prioritization using satellite imagery and mobile data. Research Journal of the South African Institute of Electrical Engineers (SAIEE) 114(1):14-24.

World Bank (2021). Country Economic Memorandum for Mozambique. World Bank, Washington, US. World Bank (2022). Ethiopia's great transition: A Country Economic Memorandum. World Bank, Washington, US.

Otieno, F. & McSharry, P.E. (2018a). Forecasting energy demand for microgrids over multiple horizons. IEEE PES&IAS PowerAfrica Conference, Cape Town, South Africa.

Otieno, B. & McSharry, P.E. (2018b). Customer segregation of an East African Microgrid. IEEE PES&IAS PowerAfrica Conference, Cape Town, South Africa.

AMdEX (2023). Amsterdam Data Exchange. https://amdex.eu

SeolSolution (2023). https://seoulsolution.kr/en/content/night-bus-called-owl-bus-route-design-using-big-data.

Appendix: Use-case template

Use-cases help organizations to make critical decisions with clarity, accuracy, and speed.

Each use-case describes how the user leverages data and analytics to derive insights to answer tangible questions for decision making. Building the use-cases is the first step in formulating a data analytics strategy. The following template provides questions that can help to build use-cases:

Sample use-case

Item	Description	Questions
1	Strategic goal	What is the strategic goal that this data use-case will support?
2	Objective	What is the objective of this data use-case? List some questions that this data will help you answer.
3	Success	How will success be measured? List relevant key performance indices (KPIs)
4	Ownership	Who is the owner of this data use-case?
5	Users	Who are the users that will benefit from generated insights?
6	Requirements	What data, tools, personnel, and funding will be required?
7	Governance	What data governance, security, privacy, access, and ownership challenges are there?
8	Analytics	How will the data be turned into insights? List the proposed analytics approaches.
9	Technology	What are the technology challenges and requirements? List data collection, storage, processing, sharing considerations.
10	Capacity	What are the challenges for skills, capabilities, and resources? Who will deliver and fund?
11	Implementation	What are the implementation and change management challenges and requirements?