PRT 840: IT THESIS

Final Report

On

Data Governance Framework for Smart City Initiative

Submitted By:

Nayan Maharjan (S3624877)

Pujan Katwal (S366540)

Samyog Chapagain(S372145)

Final report handed in partially fulfilling the requirements for MASTER OF DATA SCIENCE award eligibility.

Supervisor: Dr. Yakub Sebastian

Date: 27.10.2024

CHARLES DARWIN UNIVERSITY

COLLEGE OF ENGINEERING, INFORMATION

TECHNOLOGY AND ENVIRONMENT

DECLARATION

We are proud to say that this work, which is now being turned in as the final report for the Master of Data Science degree at Charles Darwin University, is the result of our own study. Any ideas or work that we borrowed from other researchers have been properly cited. This is our official statement that the work in this report has not yet been accepted for any degree and is not being submitted for any other degree at this time.

Name	Student No.	Signature	Date
Nayan Maharjan	S362487	N.M	27.10.2024
Pujan Katwal	S366540	P.K	27.10.2024
Samyog Chapagain	S372145	S.C	27.10.2024

DEDICATION

To begin, we'd like to present this thesis to our parents, who let us come to Australia to study. We made the right choice by choosing to go to Charles Darwin University and meet new people and grow as a group. Second, we'd like to dedicate this report to our families, bosses, friends, and peers, whom we love. Their help and encouragement push us to keep going and do good work.

ACKNOWLEDGEMENTS

Contents

List of Figures

Abstract

This thesis describes the creation of a data governance system tailored particularly for Darwin City, with the goal of facilitating its transformation into a smart city. This framework addresses the need for improved data management in urban settings by aligning with the Information Act 2002 and related Australian privacy regulations, providing robust data handling, security, and openness. The framework is organised around essential goals such as legal compliance, operational efficiency, data security, and encouraging an atmosphere of open innovation. While the framework has been fully conceptualised and is supplemented by specific Data Governance Guidelines in the appendix, it is still theoretical and requires actual application. This study introduces a revolutionary approach to urban data governance, providing a scalable model for other cities looking to exploit digital innovations for long-term urban growth. The ultimate objective is to establish a foundation that not only supports Darwin City's smart city projects but also acts as a model for comparable urban infrastructures across the world.

1. Introduction

This thesis describes the development of a complete data governance system designed specifically for Darwin City, with the goal of pushing its transformation into a smart city. This effort reacts to the growing reliance on digital data to improve urban management and public service delivery while adhering to strict regulatory requirements such as the Information Act 2002 and appropriate Australian privacy laws.

1.1 Project Context and Importance

Darwin City needs strong data governance to guarantee security, efficiency, and openness if it wants to lead in urban innovation. The suggested structure is meant to improve the operational effectiveness of the city, respect the privacy of its citizens, and create an atmosphere of free and honest invention. The Data Governance Guidelines, which are found in the appendix and offer organised rules and processes to direct the city's data management activities, are fundamental to this framework.

1.2 Objectives

The framework is structured around several key objectives:

- **Ensuring Legal and Regulatory Compliance**: Enhancing the city's data processes to prevent legal problems and foster public trust.
- **Implementing Best Practices**: Using well-known international data management standards to improve productivity and security.
- **Promoting Operational Efficiency and Transparency**: Simplifying data procedures helps to support public responsibility and better decision-making.
- **Enhancing Security Measures**: Building strong defences against data breaches and cyberattacks.
- **Fostering Continuous Improvement and Innovation**: Changing with the times and addressing developing issues with constant framework changes.

1.3 Scope

This study shows how strategic data handling may improve service delivery and assist city-wide smart projects, therefore offering a model for efficiently including strong data governance into urban administration. The thesis intends to provide insights on enhancing urban data systems that are safe, efficient, and flexible by theoretically assessing and suggesting implementations for this framework.

1.2 Structure of Report
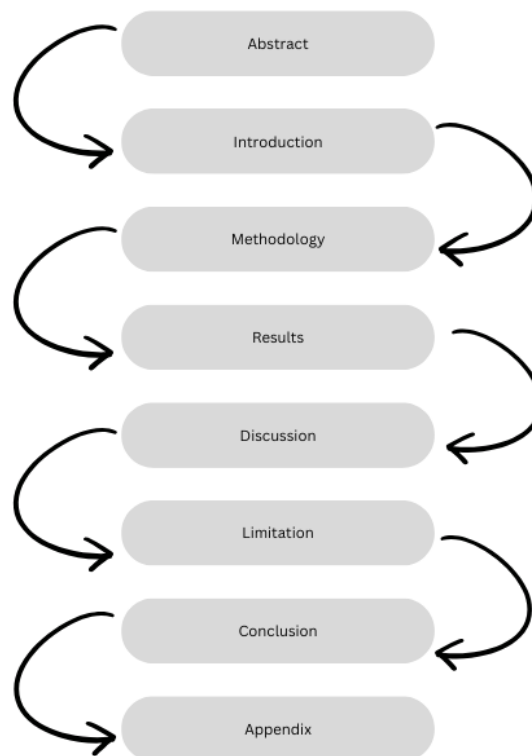
Figure 1: Structure of Report

2. Methodology

This thesis's methodology is meant to meet the important need for a data workflow that handles the sharing of information well within Darwin City's smart city framework. As the city's data-driven projects move forward, it becomes clear that it needs a system that can wisely decide whether data should be shared with the public, with third parties, among workers, or thrown away. The main goal of this study is to create a workflow that not only sorts and processes data quickly but also follows strict rules for making decisions to make sure that data is handled safely and correctly. This method uses a structured development process that includes feedback from important stakeholders. Its goal is to create a workflow that makes data sharing more open and accountable, which supports Darwin City's commitment to open government and public trust.

2.1 Development of the Data Workflow

2.1.1 Phase 1: Initial Diagram

Figure 2: Phase 1 data workflow

2.1.1.1 Workflow Functionality

Strategically, the first workflow diagram was meant to control and simplify data handling from acquisition through delivery, either inside or outside of the company. The following describes the working flow:

- **Data Acquisition**: The first step in the process is gathering of data, which means getting raw data from different sources in the city's infrastructure. This step is very important because it makes sure that all the data needed for smart city services and city management is gathered quickly and correctly.
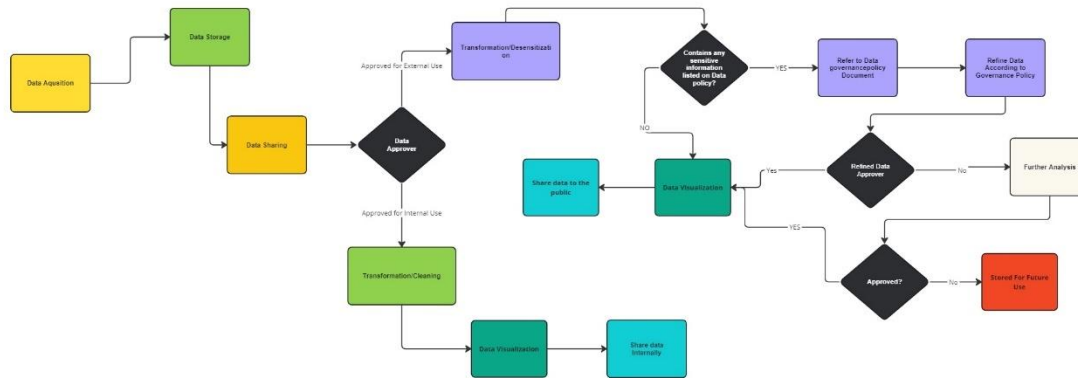- **Data Storage**: The information is kept in a safe method once it has been gathered. This system is made to handle a lot of data while keeping it safe and only letting authorised users view it.
- **Data Approval**: Most of the work is done by the Data Approver. This job is very important because it decides what to do with the data next based on what it was meant to be used for. The Data Approver looks at the data and decides if it can be used internally or externally, or if it needs to be processed further:
  - **For External Use**: Data meant for external use goes through a desensitising and transformation procedure to make sure it doesn't include private or security-compromising sensitive data. It is then checked to guarantee sensitive info is absent. Should clearance, the material moves to the Data Visualisation phase, ready for public sharing. Should sensitive material be discovered, it is returned into the process for data governance policy-based refining before it can be reviewed for approval.
  - **For Internal Use**; Data approved for internal use is directed towards Transformation/Cleaning to ensure it is in a usable state for internal stakeholders. Following this, it moves to the Data Visualization stage, where it is formatted for internal reports and analytics.
- **Data Visualization**: At this level, data is visualised to support knowledge and application either for internal decision-making or public distribution. This stage is vital since it turns unprocessable data into a format the public and stakeholders can readily understand.
- **Decision Nodes for Data Refinement**: If private information is found in the data, especially after it has been approved for external use for the first time, it may be sent back to be fixed in line with governance policies. This repeated process makes sure that all private and regulatory rules are followed.
- **Final Approval and Distribution**: Once the data has passed all checks and improvements (if necessary), it is either kept for future use, shared internally, or released for public consumption.

2.1.1.2 Evaluation and Feedback

The first workflow was seen by stakeholders as unnecessarily complicated, especially regarding the excessive looping in data approval and sensitivity checks that can cause inefficiencies. To enable better data flow and lower possible mistakes, they advised minimising the links between phases and streamlining of these procedures for a more linear workflow. Furthermore, there was a great demand for more thorough views of every stage to increase workflow transparency and clear roles and duties strengthening management and knowledge.

Moreover, stakeholders advised including a formal feedback system to enable ongoing development depending on consistent input. This would allow dynamic changes to the workflow to better satisfy changing needs and handle unanticipated difficulties, hence improving the general responsiveness and functioning of the data management system inside the framework of the smart city.

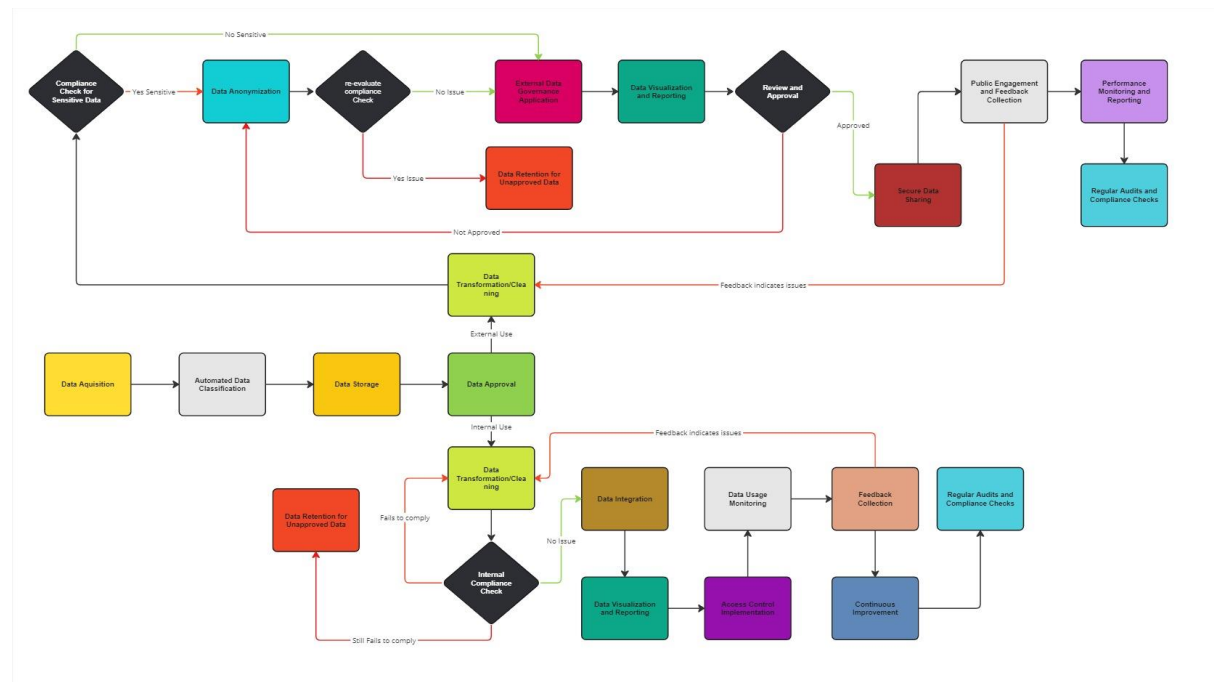## 2.1.2 Phase 2: Enhancement and Refinement

Figure 3: Phase 2 improved data flow diagram

2.1.2.1 Improvements and Enhancements

Phase 2 brought numerous important changes meant to solve the complexity and inefficiencies found in the first workflow:

- **Streamlined Data Flow**: To cut the number of loops and boost process efficiency, the data flow was greatly simplified. Simplifying the routes between data collecting, storage, and approval phases helped to speed decisions and lower the possibility of bottlenecks.
- **Enhanced Data Security and Compliance**: To protect data and make sure that rules were followed, security steps and compliance checks were made stronger. To better handle sensitive information, automated data classification and improved data anonymization methods were put in place. This made sure that all data met strict privacy standards before it was visualised or shared with the outside world.
- **Improved Decision Points**: Particularly regarding data fit for internal or external use, the decision-making procedures were improved. Establishing clear criteria and checkpoints helped to enable more accurate assessments, therefore lowering the possibility of mistakes and guaranteeing that only correctly handled data moved forward to next phases.

- **Robust Feedback Integration**: A formal feedback system was added to the process so that stakeholders' ideas and thoughts could be constantly collected. This made sure that the process could be changed on the fly to meet changing needs and quickly fix any problems that were found.
- **Automated and Advanced Processing Techniques**: Advanced data processing and cleaning methods helped guarantee that data meant for external sharing was handled with increased accuracy, hence further matching the workflow with external data governance criteria.

These changes made in Phase 2 made the general functionality, security, and efficiency of the data workflow much better. They also laid the groundwork for further growth and improvement in later phases.

2.1.2.2 Evaluation and Feedback

During Phase 2, the changes that were made to the data flows were mostly judged by reviews from stakeholders, who mostly looked at how the streamlined data processes and better security measures worked. Stakeholders gave useful feedback, pointing out that the process was becoming more efficient and meeting regulatory standards. But they did point out that to better handle the city's complex data interactions, process visualisation and documentation needed to be clearer and more standardised.

Stakeholders suggested utilising the Business Process Model and Notation (BPMN) for Phase 3 to improve clarity and standardisation across the process. They also advised that the workflow might be more structured, with separate pools for managing internal data, third-party data sharing, and open data. This feedback was critical in designing the subsequent phase, resulting in considerable redesign efforts focused at merging these separate data management pools and implementing BPMN to better align the workflow with the City of Darwin's strategic data governance objectives.

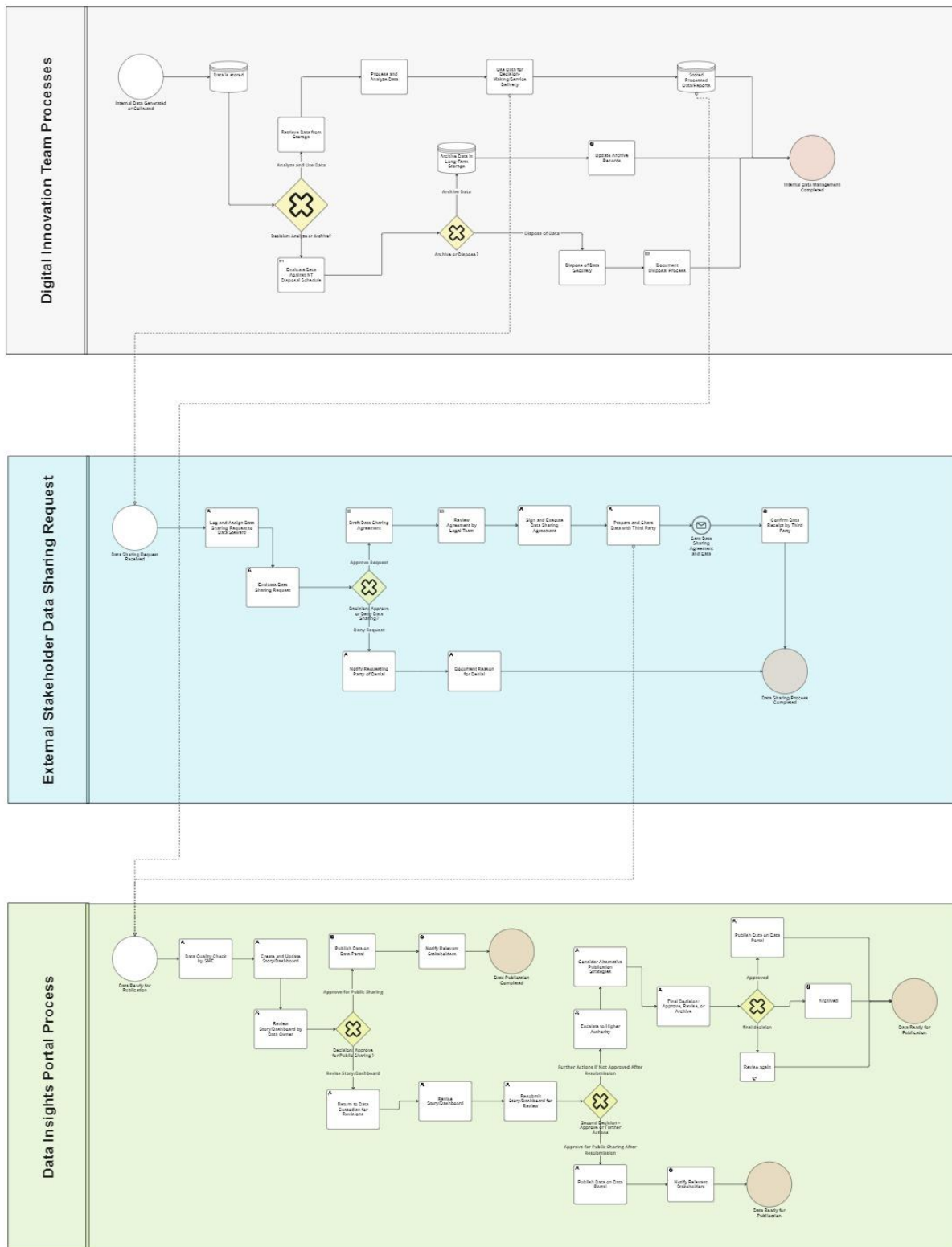## 2.1.3 Phase 3: Integration of BPMN and Structured Pools

Figure 4: Phase 3 BPMN model

2.1.3.1 Overview of Workflow Organization

In Phase 3, the data management processes are restructured in a detailed way using the Business Process Model and Notation (BPMN). The workflow is split into three groups that are each designed to handle different types of data. These groups are the Digital Innovation Team Processes, the External Stakeholder Data Sharing Request, and the Data Insights Portal Process. This separation makes operations clearer and makes sure that each stream of data dealing is done in the most efficient, secure, and compliant way possible.

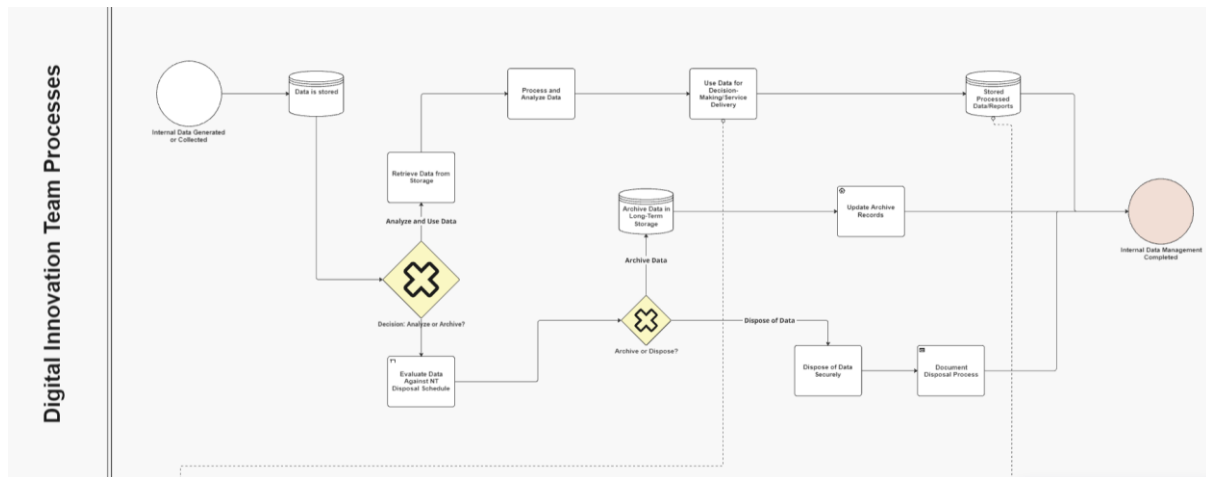**Digital Innovation Team Processes (Pool 1):**

Figure 5: Phase 3 first pool Digital Innovation Team Processes

- **Data Generation and Collection:** This pool starts with the internal generation or gathering of data, which is then safely stored.
- **Decision Making on Data Use:** The purpose of evaluating data is to decide whether it will be used for research or storage in the future.
  - **Analysis Route:** If the data is designated for analysis, it is retrieved, processed, and used to make decisions or provide services. The results are then saved as processed data.
  - **Archiving Route:** If data is not required for immediate use, it is compared to the NT Disposal Schedule to determine if it should be archived or securely disposed of, ensuring that all actions meet regulatory standards.

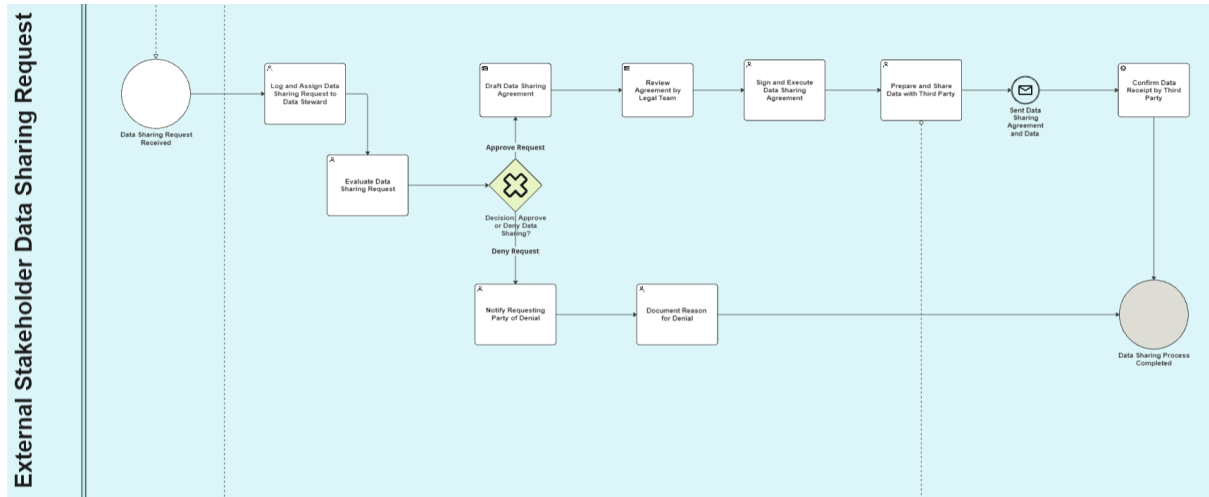**External Stakeholder Data Sharing Request (Pool 2):**

Figure 6: Phase 3 Second Pool External Stakeholder Data Sharing Request

- **Handling Data Requests:** This pool is activated when a third-party data sharing request is received.
- **Evaluation and Agreement Process:** The request is then assigned to a Data Steward, who is responsible for evaluating it after it has been logged. Following the approval of the proposal, the next step in the process involves the creation of a data sharing agreement, which is then evaluated by the legal team, signed, and completed.
- **Data Preparation and Sharing:** Once everything is agreed upon, the data is prepared and sent to the third party, making sure that every step is carefully carried out, from confirming receipt to keeping records of the process.
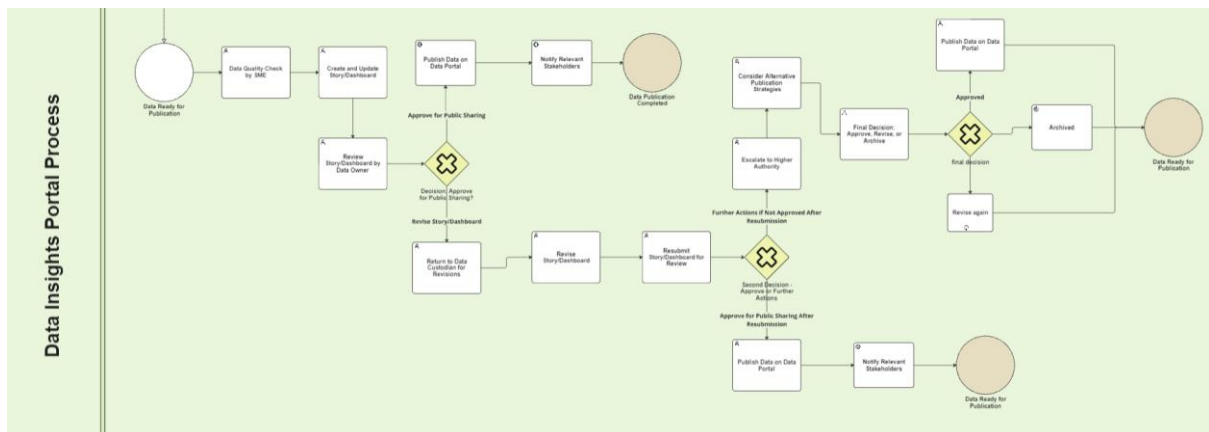
Data Insights Portal Process (Pool 3):

Figure 7: Phase 3 Third Pool Data Insights Portal Process

- **Preparation for Publication:** Data from Pool 1 that is processed and saved or data from Pool 2 that is shared with the outside world can be sent to publication.
- **Quality Checks and Reviews:** Prior to publication, data is subjected to a rigors quality check before being utilised to produce or update stories and dashboards. The data owner reviews them to ensure accuracy and appropriateness.

- **Publication Decision:** If accepted, data is posted to the data portal. If not, it is either returned for modifications or escalated for additional review, depending on the extent of the concerns discovered.

This clear but detailed explanation of Phase 3 shows not only how the BPMN integration has improved operations but also how the workflow fits in with the City of Darwin's strategic goals for data governance, which include making sure that managing city data is open, accountable, and flexible.

2.1.3.2 Evaluation and Feedback

Stakeholders gave important feedback during Phase 3 implementation, which was used to improve the BPMN-enhanced data process. Feedback for Pool 1 stressed the need for more freedom in making decisions about data handling. In particular, stakeholders suggested making the decision point for analysing or storing data better. They said that data that wasn't going to be analysed right away should instead be looked at regularly to see if it could be useful in the future. There was also a call to make a loopback from the "Evaluate Data Against NT Disposal Schedule" to the analysis process. This would let saved data that might still be useful be used again. Concerns were also made about how data is stored and retained. It was suggested that the decision-making process be changed so that it is clearer when to keep aggregated data and when to get rid of granular data. This would improve data privacy and storage efficiency.

For Pool 2, stakeholders wanted to change the name of the starting point to better reflect what it does and get rid of a few steps, such as assigning requests to a Data Steward and reviewing data sharing agreements by lawyers, which were seen as extra steps that added too much complexity and didn't add much value. In Pool 3, it was suggested that the way new or heavily updated stories and dashboards are reviewed should be changed so that Executive Managers are directly involved. This would make sure that important public communications are always supervised at a high level. Stakeholders also gave advice on how to divide the decision-making process into different parts for public sharing. They suggested separate paths for elected officials, staff, and the public to better shape the data sharing. They also stressed how important it was to add strong processes for archiving and securely destroying data, making sure that data control standards were followed.

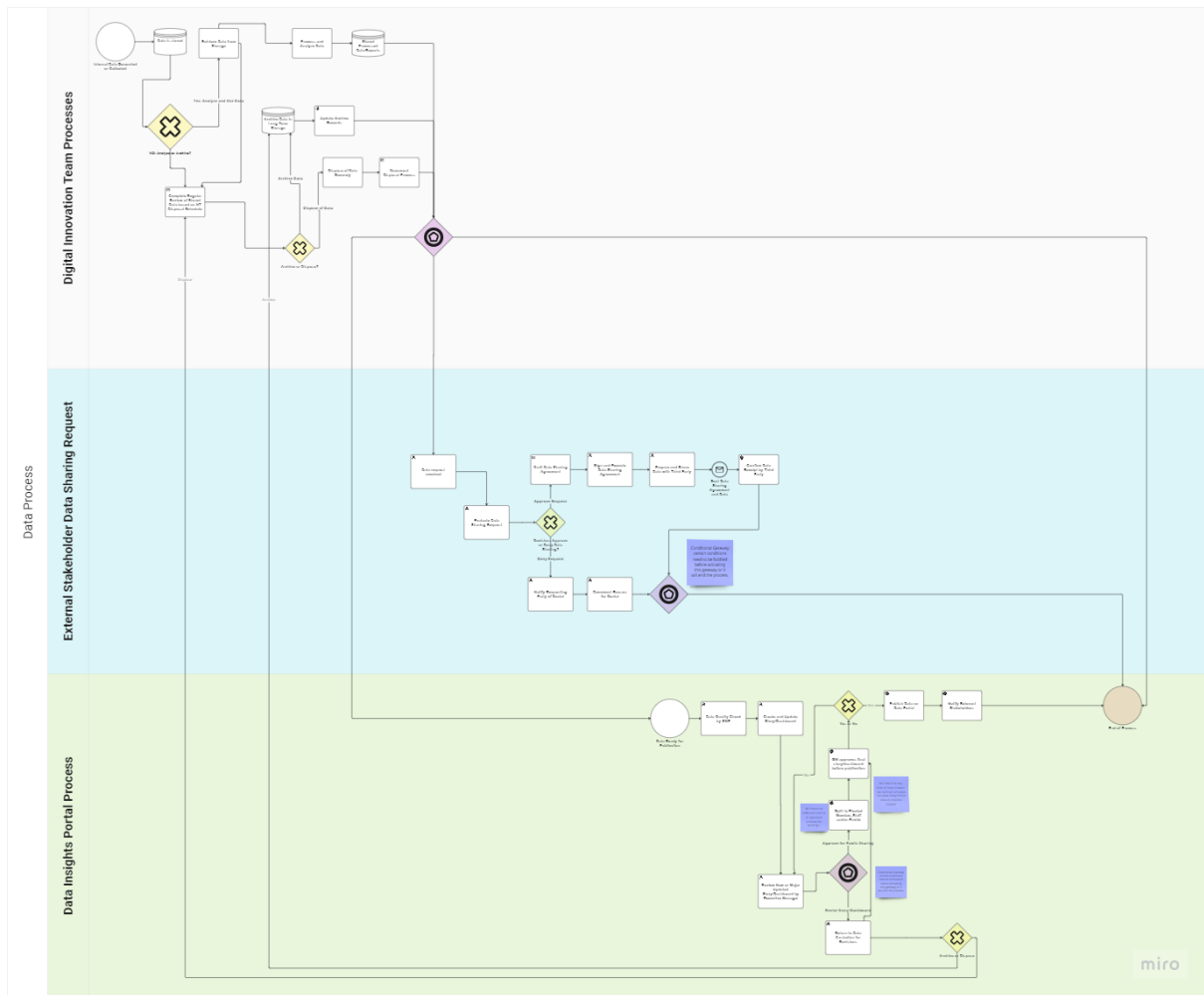## 2.1.4 Phase 4: Advanced Refinement and Preparatory Implementation

Figure 8: Phase 4 Data Flow Diagram

2.1.4.1 Workflow Improvements and Enhancements

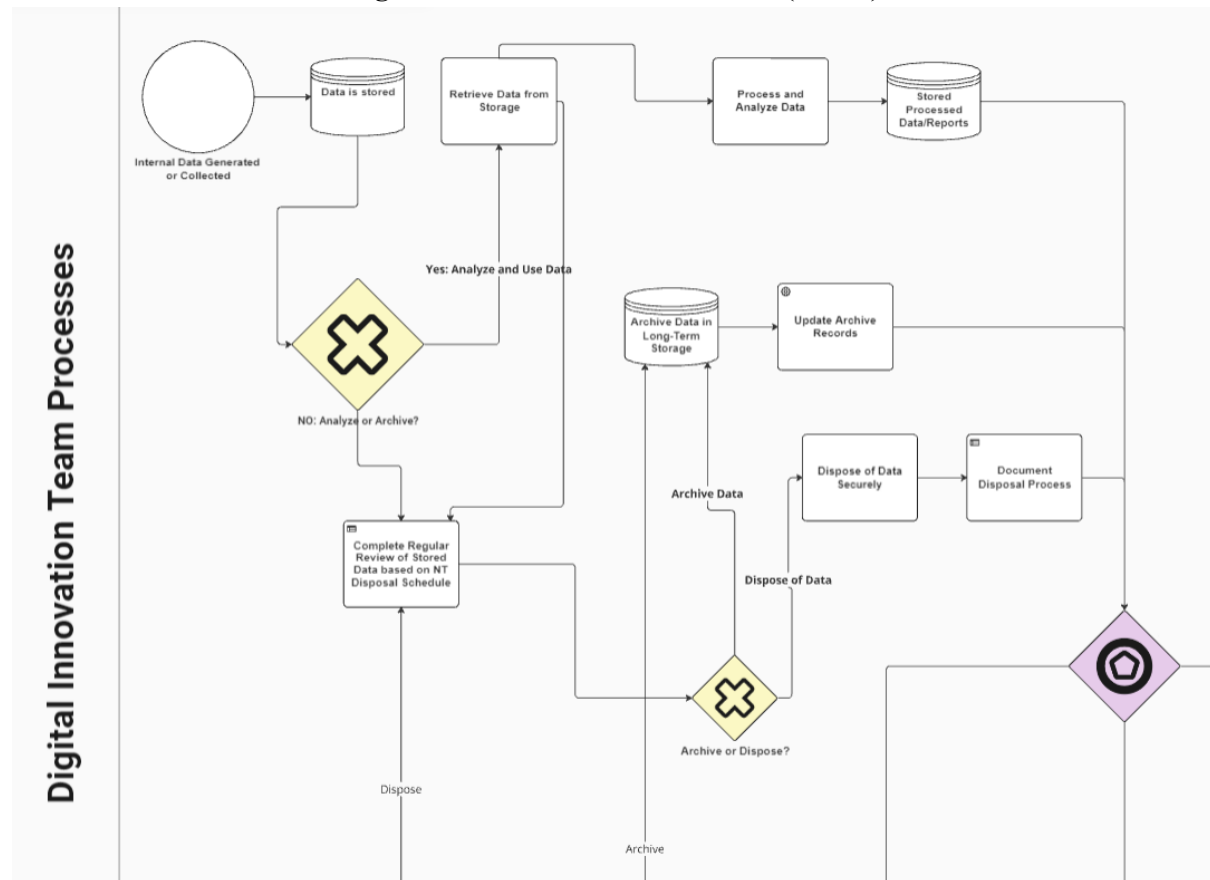**2.1.4.1.1 Enhancements in Digital Innovation Team Processes (Pool 1)**

Figure 9: Phase 4 Digital Innovation Team Processes Pool 1

Optimization of Data Handling Processes:

- **Simplification:** The step "Use Data for Decision-Making/Service Delivery" was taken out to make the process more efficient and allow the focus to be solely on important data management tasks.

Introduction of New Data Review Pathways:

- **Direct Review Connection:** "Retrieve Data from Storage" now leads to "Complete Regular Review of Stored Data based on NT Disposal Schedule." as a new path. With this addition, saved data can be constantly checked to make sure it is still useful and is either used or thrown away based on current needs and legal requirements.

Conditional Decision Points:

A big improvement to Pool 1 is the addition of a conditional decision point that brings together different process results, such as "Stored Processed Data," "Update Archive Records," and "Document Disposal Process." This decision point tells the workflow what to do next based on certain conditions being met. This makes the process more flexible and responsive.

- If the data is still thought to be useful or needs more research, it may go back into the active workflow to be used or processed again.
- If data is no longer useful or fits certain criteria, it is safely archived or thrown away, following strategies for compliance and space optimisation.

The changes that were made to Pool 1 make the data management processes in the City of Darwin's digital innovation team much more efficient and flexible. By getting rid of steps that

aren't needed and adding the ability to make decisions on the fly, the workflow is better able to handle data and make sure that all actions are done on time and in line with legal standards. This step builds a strong base for later phases of the workflow, which will integrate and improve things even more.

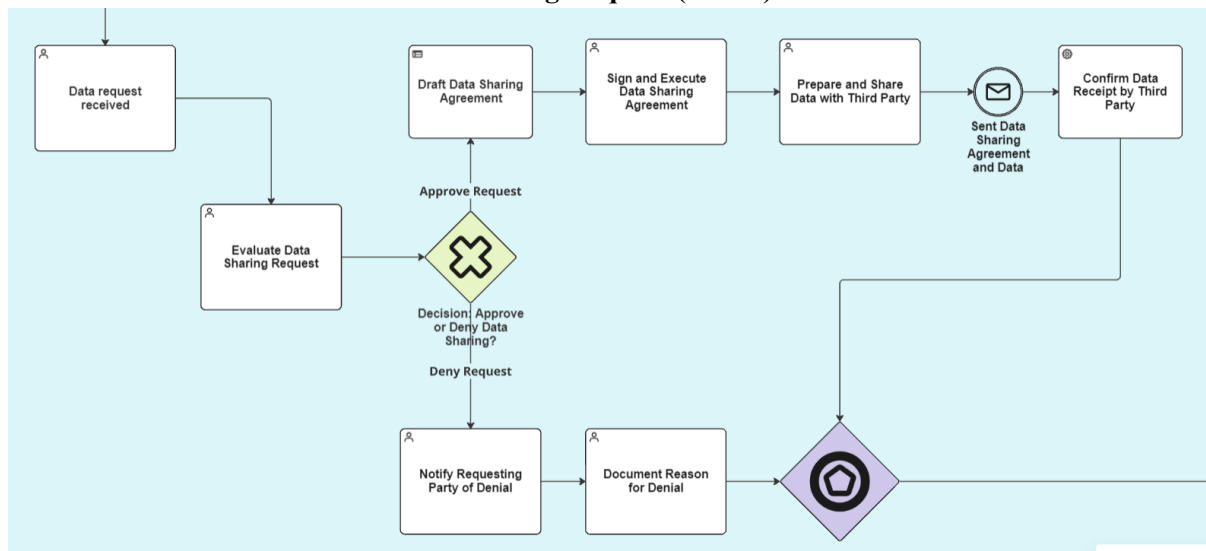**2.1.4.1.2 External Stakeholder Data Sharing Request (Pool 2)**

Figure 10: Phase 4 External Stakeholder Data Sharing Request Pool 2

- Streamlined Initiation:
  - The process now starts right away with "Data Request Received." This is done by the conditional decision point from Pool 1, which makes sure that this workflow is only started by appropriate data requests.
- Process Efficiency Improvements:
  - The "Log and Assign Data Sharing Request to Data Steward" step was taken out to cut down on administrative work and speed up the process of handling requests.
  - "Review Agreement by Legal Team" was taken out to make the process even simpler and help people focus on the most important steps for speeding up data sharing.
- Conditional Integration and Workflow Continuation:
  - The workflow is strongly connected to a conditional decision point that goes straight to the common endpoint. This makes it easier to switch between processes without any problems.
  - Makes sure that all choices about sharing data are made quickly, with clear ways to approve or reject them, and that actions taken afterwards are directly related to the outcomes.

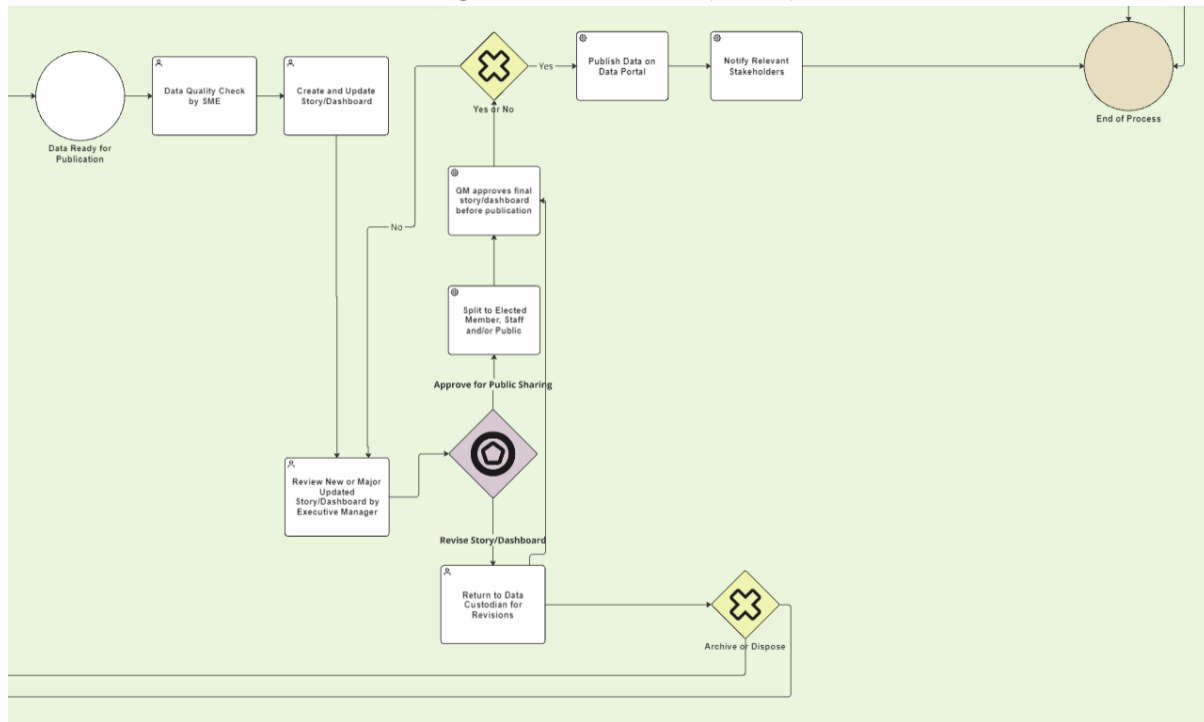### 2.1.4.1.3 Enhancements in Data Insights Portal Process (Pool 3)

Figure 11: Phase 4 Data Insights Portal Process Pool 3

- Streamlined Publication Process:
  - o Got rid of steps that weren't needed so that the focus could be on core tasks that make sure the information being released is accurate and useful.
  - o Linked the beginning of this pool straight to the conditional decision point from Pool 1, making sure that there is a smooth flow from processing data to getting it ready for publication.
- Enhanced Review and Approval Mechanism:
  - o The review process has been changed so that the Executive Manager is now involved with reviewing any new or important changes to stories or dashboards. This ensures high-level oversight and strategic alignment.
  - o Added a flexible system for making decisions about publications that is suited to various groups of people (Elected Members, Staff, and Public).
- Final Approval and Dissemination:
  - o Added one last step for approval by the General Manager before publication. This makes sure that all material meets the highest standards of being correct and appropriate.
  - o Once content is approved, it is put on the data portal and shared with stakeholders, successfully spreading important information.
- Revisions and Archival Strategy:
  - o If the content is not accepted, it is returned to the data custodians for modifications, giving them the opportunity to edit and improve before resubmitting it.
  - o Established a clear method for archiving or securely disposing of data that is unsuitable for distribution, in accordance with compliance and governance norms.

2.1.4.2 Evaluation and Feedback
Feedback for Digital Innovation Team Processes (Pool 1):

- Stakeholders have suggested adding a data quality check to Pool 1 to make sure the data is correct before it is saved or processed further. This addition is seen as very important for making sure that the data stays correct throughout the process. They also suggested a specific change be made to the workflow connections: the conditional decision point should only be tied to the "Data Stored" output. This change makes sure that only processed, relevant data is available for use in other groups. This speeds up the workflow and stops data from moving around that isn't needed.

Feedback for External Stakeholder Data Sharing Request (Pool 2):

- In Pool 2, stakeholders saw that the communication processes were being used twice and considered getting rid of the "Sent Data Sharing Agreement and Data" and "Confirm Data Receipt by Third Party" steps. This feedback shows that the data sharing process needs to be made easier by getting rid of steps that aren't necessary or improve the workflow's usefulness.

Feedback for Data Insights Portal Process (Pool 3):
- Stakeholders proposed numerous major modifications in Pool 3 to streamline the publication process. They noted that after stories or dashboards have been modified and submitted for approval, they should not be returned to custodians for modifications, hence simplifying the approval process. They also stated that informing stakeholders following publication is unnecessary and may be eliminated to speed up the post-publication process. Clarification was also offered regarding the approval process: the General Manager should only examine and approve new or significantly revised dashboards/stories at first, while Executive Managers should review each subsequent update before sharing it with the public or elected members.

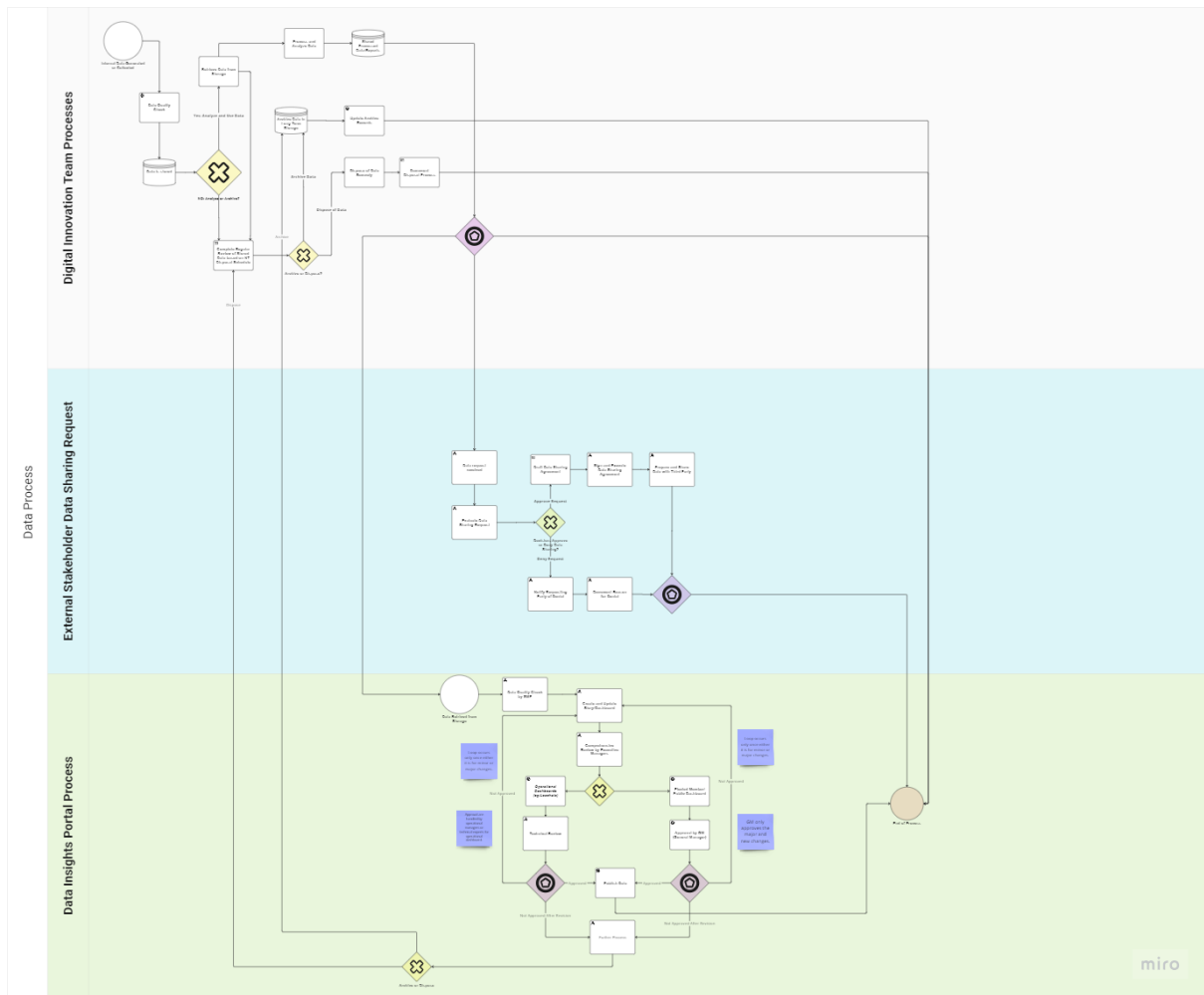## 2.1.5 Phase 5: Final Implementation and Optimization

Figure 12: Phase 5 Dataflow Diagram

2.1.5.1 Workflow Improvements and Enhancements

2.1.5.1.1 Enhancements in Digital Innovation Team Processes (Pool 1)
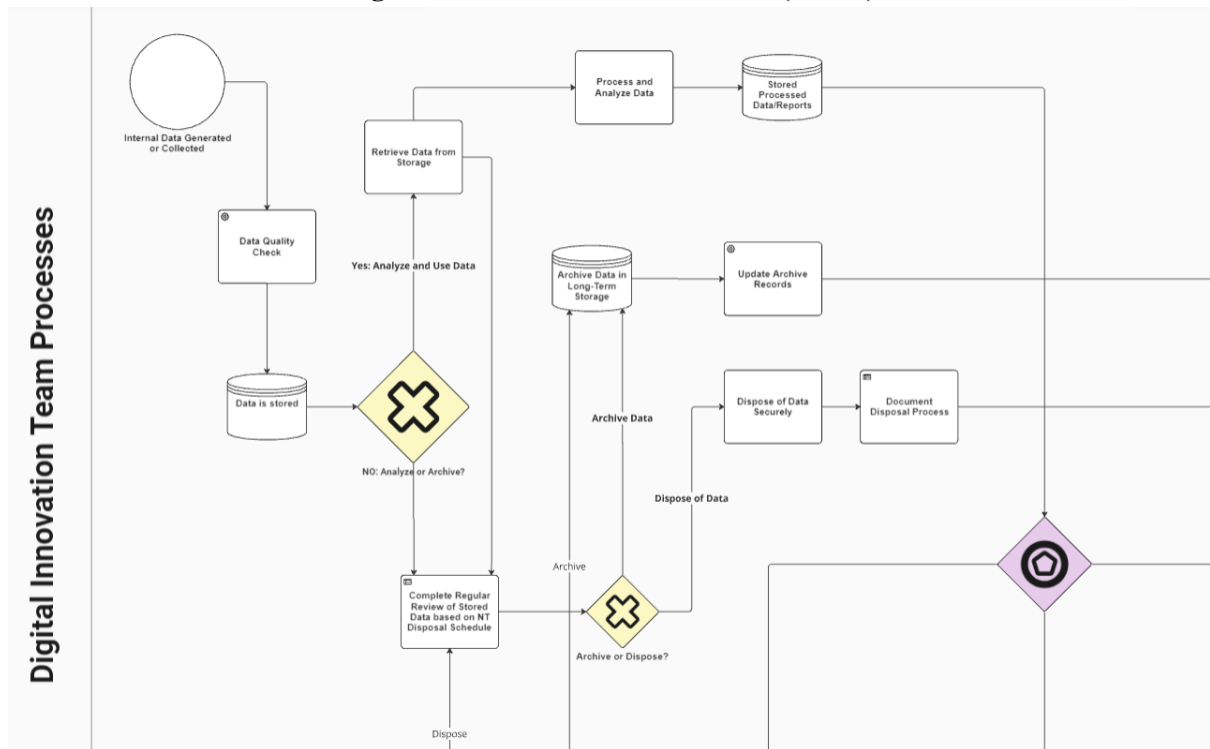

Figure 13: Phase 5 Digital Innovation Team Processes (Pool 1)

- **Refinement and Optimization**:
  - The conditional decision-making point is now only linked to "Stored Processed Data/Reports." This point decides the next steps based on specific operation needs.
  - 'Update Archive Records' and 'Document Disposal Process' have been streamlined so that they connect directly to the same endpoint. This makes the process easier and cuts down on decision points that aren't needed.
  - A new step called "Data Quality Check" has been added to the process to make sure that the data is correct and reliable before it moves on to the next step.

2.1.5.1.2 External Stakeholder Data Sharing Request (Pool 2)
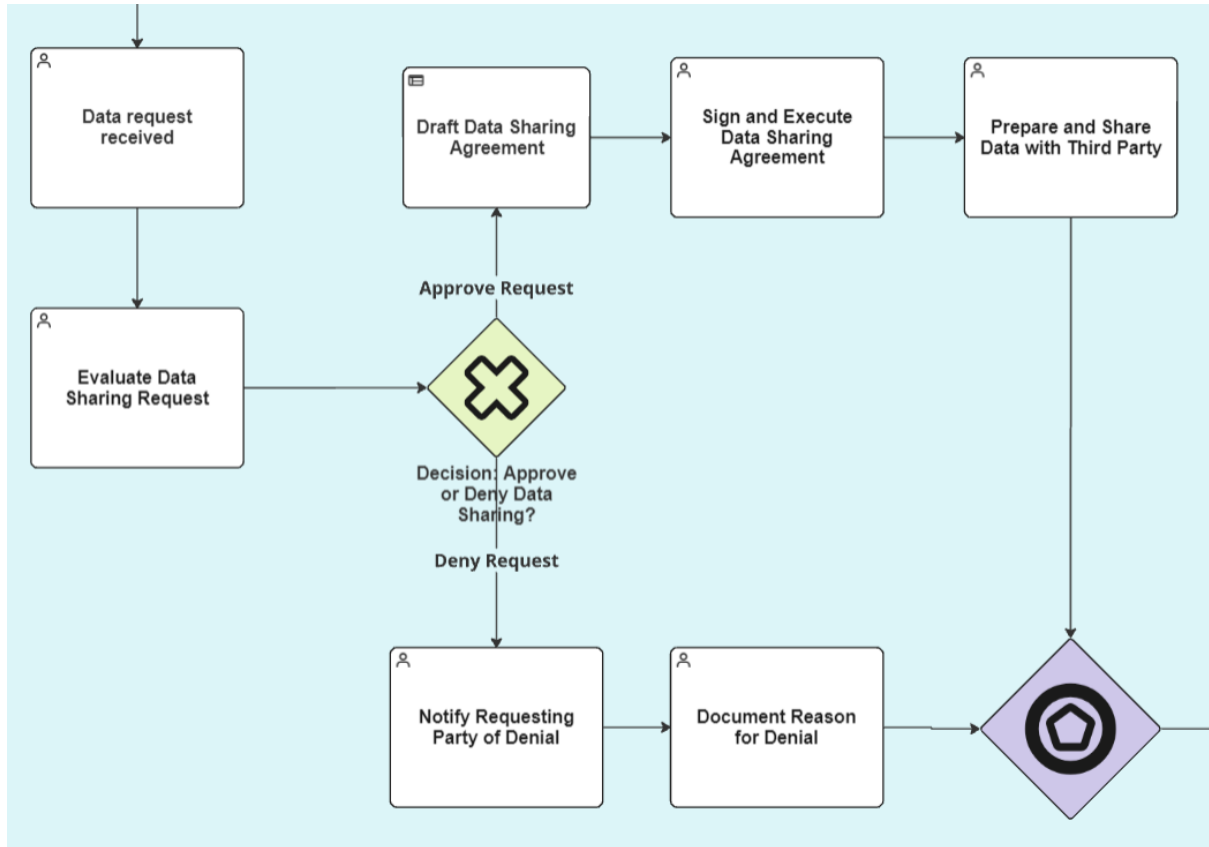
Figure 14: Phase 5 External Stakeholder Data Sharing Request (Pool 2)

- **Streamlined Operations**:
  - o All suggested deletions have been made, including "Data Receipt Notification" and "Confirmation of Data Receipt." The focus is now on key activities that add value and make sharing data easier.

2.1.5.1.3 Data Insights Portal Process (Pool 3)
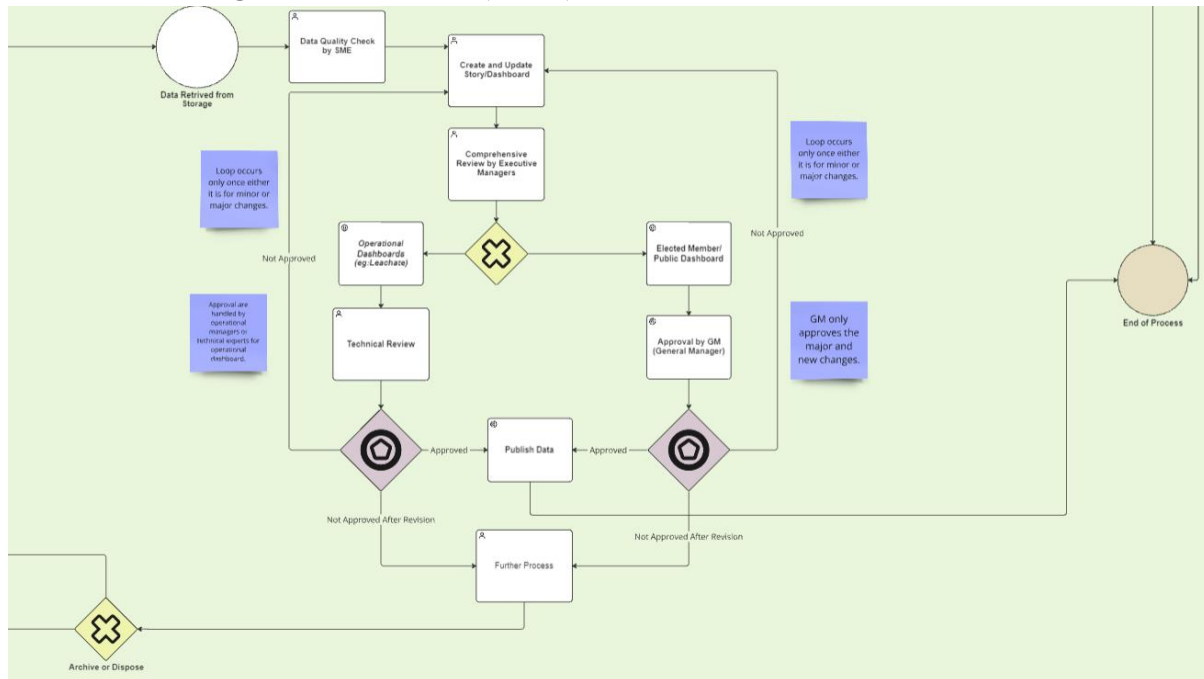
Figure 15: Phase 5 Data Insights Portal Process (Pool 3)

- **Enhanced Review and Approval Processes**:
- When a dashboard is made or updated, it goes through a "Comprehensive Review by Executive Managers" to make sure that all of its material meets operational and strategic standards.
- At this point, a choice is made about whether the dashboards are "Operational Dashboards" for internal use or "Public-Faced Dashboards" for elected officials and the public.
- After an operational dashboard goes through a technical review, it may be published right away, sent back for minor changes (with one loop allowed for revisions), or sent to a more in-depth analysis step if major problems still exist. This depends on the conditional decision point.
- After initial clearance, public-facing dashboards are reviewed by GMs before being published. If changes are required, a similar looping method enables one-time adjustments. If the problem persists, it is submitted for further analysis, which will determine whether the material is stored or discarded.

2.1.5.2 Evaluation and Feedback

In Phase 5, stakeholders looked over the suggested final implementations of the data management workflows across all pools and were happy with the anticipated enhancements and improvements. Even though the changes to the ideas have been approved, the data flow has not yet been put into action. This acknowledgement shows how thorough the planning and design stages were. In theory, they met the city's strategic and operational expectations. Stakeholders' approval means that the workflows are ready to be put into action in the future, which builds a strong foundation for their ultimate operational deployment.

2.2 Hypothetical Testing of Organizational Data Flow

This part talks about the hypothetical tests that were done on the company's new data flow process, which has three groups that are all connected: the Digital Innovation Team Processes, the External Stakeholder Data Sharing Request, and the Data Insights Portal Process. This data flow has never been put into reality, so the results of the tests and the suggestions are based on simulations, theoretical analysis, and best practices in the industry.

2.2.1 Testing Methodology

The testing process involved the following steps:

- **Scenario Development**: To make sure that all possible operational problems were covered, hypothetical examples were made to show how the data flow might work in real life and in common use cases.
- **Efficiency and Decision Accuracy Evaluation**: Every situation was looked at to see how well the system handles data, makes choices, and follows rules for compliance.
- **Predictive Analysis**: Using assumptions and logical conclusions, possible problems, bottlenecks, and improvement areas were found. This gave a prediction of how the workflow would work.

2.2.2 Scenarios Tested and Hypothetical Outcomes

2.2.2.1 Scenario 1: High Volume of Data Requests

- **Objective**: To assess the workflow's ability to manage many simultaneous data sharing requests efficiently.
- **Hypothetical Execution**: Ten data sharing requests were expected to arrive simultaneously.
- **Hypothetical Observations:**
  - **Bottlenecks:** Inappropriate prioritising systems could cause possible delays at the "Evaluate Data Sharing Request" level.
  - **Decision Errors:** Potential misrouting of requests depending on unclear or inconsistent use of evaluation criteria
- **Recommendations:**
  - Create and apply a set of well-defined priorities.
  - Standardise the procedure of requesting evaluation to lower mistakes in decision-making.

2.2.2.2 Scenario 2: Discrepancy in Data Quality

- **Objective**: To replicate a situation in which data quality calls for intervention after compromise.
- **Hypothetical Execution**: Presumed a 15% variation in data quality throughout inspections.
- **Hypothetical Observations**:
  - **Bottlenecks:** Delays in fixing data affect later steps, like making the dashboard and getting approval.
  - **Decision Errors:** Risk of incorrectly marking data as corrected due to oversight or SME fatigue.
- **Recommendations:**
  - Set up a dual-verification method to fix problems with the quality of the data.
  - Improve the teaching for SMEs and set strict rules for quality control.

2.2.2.3 Scenario 3: Sensitive Data Disposal

- **Objective:** To see how well data disposal rules are being followed in a situation where a lot of data needs to be thrown away.
- **Hypothetical Execution**: Designed a scenario in which 20% of stored data needs to be thrown away.
- **Hypothetical Observations**:
  - **Bottlenecks**: Delays in approval because of complicated multi-level authorisation rules
  - **Decision Errors**: Classification mistakes could lead to wrong data being marked for removal.
- **Recommendations**:

o Make the process for getting permission to delete data easier or set up automatic ways to get permission.
o Set up a strong review system to make sure that data is properly classified before it is thrown away.

3. Results

This section displays the results of the hypothetical testing performed on the newly constructed organisational data flow. Each scenario was designed to represent distinct operational difficulties inside the digital innovation framework, with an emphasis on efficiency, compliance, and decision accuracy.

3.1 Findings by Scenario

3.1.1 Scenario 1: High Volume of Data Requests

- **Outcome**: The hypothetical execution expected that 10 data requests could be handled at the same time. Prioritising and managing these requests were easy for the data flow, but there were some delays seen at the "Evaluate Data Sharing Request" step.
- **Data Handling**: Handled quickly and without any major delays, showing that the process can handle a lot of work at once.
- **Bottlenecks Identified**: Potential slowdowns may develop due to a lack of specified prioritisation rules, indicating a need for process improvements.

3.1.2 Scenario 2: Discrepancy in Data Quality

- **Outcome:** As a result, the system found and reported problems that needed to be fixed, with an assumed 15% discrepancy rate during data quality checks.
- **Effectiveness of Response:** The system had a strong way of finding mistakes, but it might take a while to fix these problems, which could affect later steps of data use.
- **Observations on Corrections:** Delays in corrections could affect the creation of dashboards and the general adherence to deadlines.

3.1.3 Scenario 3: Sensitive Data Disposal

- **Outcome:** Modelled a situation in which 20% of archived data was planned for removal while complying to NT guidelines.
- **Compliance with Disposal Regulations:** The process was compliant, with all scheduled data being disposed of in accordance with regulatory guidelines.
- **Efficiency of Disposal Process:** Potential delays were identified as a result of multi-level authorisation procedures, indicating the need for more efficient approval processes.

The hypothetical testing found weaknesses in the system's compliance with demanding data disposal rules and handling of high volume of data requests. The tests did, however, also point up areas needing work, including the necessity of more effective discrepancy resolution techniques and better prioritising guidelines. These findings offer insightful analysis of the possible practical use and areas where more improvement can improve the compliance adherence and system efficiency.

4. Discussion

4.1 Efficiency and Operational Scalability

- The workflow's ability to handle several simultaneous requests proves its suitability for high-volume operations. However, there is an obvious need for stronger prioritisation rules to avoid unnecessary delays and improve responsiveness.

4.2 Data Integrity and Process Continuity

- The system did a good job of finding data differences, but the time it took to fix these problems shows where major improvements are needed. To keep things going and make sure that data is used on time, the editing process needs to be sped up.

4.3 Regulatory Compliance and Process Efficiency

- Following the rules for disposal proves that the process can follow the rules. Still, the delays in the process show that making the authorisation process simpler could greatly improve working efficiency.

4.4 Industry Benchmarking

- Comparing these features to industry standards will assist contextualise the system's performance, offering a better understanding of its relative strengths and opportunities for improvement.

5. Limitations

5.1 Hypothetical Testing Constraints

- Using hypothetical circumstances makes it harder to predict all problems that might happen in the real world. Even though these situations are helpful, they might not fully show how complicated or variable real processes are.

5.2 Scope and Generalizability

- Since only a few situations were tried, it's possible that more testing would reveal other operational problems. Because of this, the data should only be used with care in bigger situations.

5.3 Potential Biases

- Selection and interpretation biases during scenario development and analysis may have impacted the results. Future testing should try to reduce these biases in order to better reflect actual operating settings.

6. Conclusion

This thesis thoroughly developed a framework for data governance for Darwin City. This framework is based on the specific Data Governance Guidelines that can be found in the appendices. The strict rules of the Information Act 2002 and other Australian privacy laws have been considered when making these recommendations. The framework has a strong theoretical base and is meant to help Darwin City reach its goal of becoming a smart city, but it hasn't gone beyond the idea stage yet. The system is supposed to make operations much more efficient, strengthen data security, and make things more open to the public. Nevertheless, its real usefulness can't be proven until it's used in real life. This exploratory work is not only a plan for Darwin City, but also an indicator for other cities that want to use the changing power of data governance to help with better urban ecosystems and sustainable development.

7. Appendices

7.1 Literature Review

1. Introduction

The literature review looks at the latest developments in smart city models, with a focus on how technologies like IoT and big data analytics make life better in cities. It talks about important topics in smart city development, like how to handle data, run a city, and use technology standards, with the goal of making cities more efficient, environmentally friendly, and welcoming to everyone. In the parts that follow, we'll talk more about individual frameworks and what they've added to the field.

2. Smart City Frameworks and Methodologies

Recent smart city research offers a variety of concepts and methods for urbanisation and data management. (Osman, 2019) proposes a big data analytics architecture for smart cities that integrates and analyses data from several municipal domains to improve service delivery and decision-making. This shows how big data is being used to regulate population growth. (Paskaleva et al., 2017) examine data governance in smart cities, emphasising sustainable development and providing a paradigm that balances technology-driven growth with social inclusion. (Lai et al. ,2020) also emphasise the need of technical standards in smart cities for technology implementation and system interoperability. Smart governance and the orchestration of smart city data ecosystems are discussed in (Pereira et al., 2018) and (Gupta et al., 2020), which advocate for stakeholder collaboration through ICT and identify key factors like openness and shared vision for coordinating data-driven initiatives in complex urban environments.

3. Integration of IoT and Big Data in Smart Cities

Smart city research emphasises the need of IoT integration for city management and service delivery. To create a full urban information system, (Jin et al. ,2014) present a comprehensive IoT architecture that includes sensory data gathering and cloud integration. Infrastructure, strategic, and technical integration are emphasised in the Smart City Transformation Framework (SCTF) (Kumar et al., 2020). (Hashem et al. ,2016) emphasise the importance of big data in urban decision-making, whereas (Sta ,2017) addresses data quality issues that might limit these technical benefits. Smart city growth raises cybersecurity concerns, requiring strict security and privacy measures to secure urban data infrastructures (Elmaghraby and Losavio, 2014).

4. Enhanced Data Management in Smart Cities

Smart city data management requires robust access restrictions, regulatory compliance, and high-quality data maintenance. Role-based and time-based access controls reduce unauthorised access and ensure data is utilised for approved purposes under the Australian Privacy Act 1988 (Office of the Australian Information Commissioner, 2022). Compliance with the Northern Territory Information Act 2002 and continuing compliance checks protect personal data and provide transparency and integrity (Northern Territory Government, 2022). To promote trustworthy decision-making data, the Australian National Data Quality Framework emphasises standardised data collection and validation (Australian Bureau of Statistics, 2020). The Australian Government's Open Data Principles require data to be anonymised and summarised to prevent misuse during public and third-party sharing (Department of Prime Minister and Cabinet, 2015). This complete strategy supports smart city projects by governing data handling from collection to dissemination with robust frameworks that prioritise security, compliance, and utility (National Archives of Australia, 2021).

5. Similar Works in Smart City Frameworks and Data Governance

Recent research emphasises the necessity of integrated frameworks and strong governance in smart city development. Osman (2019) proposes "SCDAP," a big data analytics system that analyses data across city domains to improve urban service delivery and decision-making. (Paskaleva et al., 2017) promote sustainable urban development through data governance that balances technology and social participation. Technical standards are essential for smart city operations and system interoperability (Lai et al., 2020). (Pereira et al., 2018) promote smart governance by using ICT to improve city stakeholder collaboration, decision-making, and citizen involvement. These contributions show a trend towards

sophisticated data management and governance frameworks to solve urbanization's difficulties and improve smart cities.

6.Conclusion

The literature assessment agrees that smart cities need sophisticated frameworks and governance systems to control urban development. Big data analytics and careful data governance can improve urban planning and operational efficiency, as shown by (Osman,2019) and (Paskaleva et al.,2017). Technology is essential to smart city operation and stakeholder participation, as (Lai et al., 2020) and (Pereira et al., 2018) emphasise standards and smart governance. These works suggest a balanced approach that combines technical innovation and good governance to create durable, efficient, and inclusive environments in cities.

7.1.7 References

Gupta, A., Panagiotopoulos, P. and Bowen, F., 2020. An orchestration approach to smart city data ecosystems. *Technological Forecasting and Social Change*, *153*, p.119929.

Pereira, G.V., Parycek, P., Falco, E. and Kleinhans, R., 2018. Smart governance in the context of smart cities: A literature review. *Information Polity*, *23*(2), pp.143-162.

Lai, C.S., Jia, Y., Dong, Z., Wang, D., Tao, Y., Lai, Q.H., Wong, R.T., Zobaa, A.F., Wu, R. and Lai, L.L., 2020. A review of technical standards for smart cities. *Clean Technologies*, *2*(3), pp.290-310.

Paskaleva, K., Evans, J., Martin, C., Linjordet, T., Yang, D. and Karvonen, A., 2017, November. Data governance in the sustainable smart city. In *Informatics* (Vol. 4, No. 4, p. 41). MDPI.

Osman, A.M.S., 2019. A novel big data analytics framework for smart cities. *Future Generation Computer Systems*, *91*, pp.620-633.

Sta, H.B., 2017. Quality and the efficiency of data in "Smart-Cities". *Future Generation Computer Systems*, *74*, pp.409-416.

Hashem, I.A.T., Chang, V., Anuar, N.B., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E. and Chiroma, H., 2016. The role of big data in smart city. *International Journal of information management*, *36*(5), pp.748-758.

Kumar, H., Singh, M.K., Gupta, M.P. and Madaan, J., 2020. Moving towards smart cities: Solutions that lead to the Smart City Transformation Framework. *Technological forecasting and social change*, *153*, p.119281.

Jin, J., Gubbi, J., Marusic, S. and Palaniswami, M., 2014. An information framework for creating a smart city through internet of things. *IEEE Internet of Things journal*, *1*(2), pp.112-121.

Abraham, R., Schneider, J. and Vom Brocke, J., 2019. Data governance: A conceptual framework, structured review, and research agenda. *International journal of information management*, *49*, pp.424-438.

Northern Territory Government. (2022). *Information Act 2002 Overview*. Retrieved from Northern Territory Government website.

Australian Bureau of Statistics. (2020). *National Data Quality Framework*. Retrieved from Australian Bureau of Statistics website.

Australian Government, Department of the Prime Minister and Cabinet. (2020). *Australian Government Public Data Policy Statement*. Retrieved from Australian Government website.

National Archives of Australia. (2021). *Data release and publication guidelines for public sector organizations*. Retrieved from National Archives website.

Office of the Australian Information Commissioner. (2022). *Australian Privacy Principles Guidelines*. Retrieved from OAIC website.

7.2 Data Governance Guidelines

1. **Purpose**

The purpose of this document is to provide guidelines and procedures for managing, transforming, and publishing the data. The policy aims to balance transparency and public access to data with the protection of privacy and confidentiality, in compliance with Northern Territory (NT) and Federal legislation. This policy establishes data governance principles that align with relevant Council procedures, policies, and legislation.

2. **Scope**

This guideline applies to all categories of data that is collected and generated by the council; every set of data is to be considered as sensitive data unless verified by the council. This policy covers all staff and processes involved in the acquisition, handling, preprocessing, approval, publication, and disposal of City of Darwin data.

3. **Guideline Statement**

This guideline outlines our commitment to safeguarding data, ensuring data utility, data quality and complying with relevant data protection laws, data policy and standards. The key principles of our policy are:

3.1 Data Storage

The City of Darwin treats all data as sensitive. To keep it safe, these steps are followed:

Treating All Data as Sensitive

Even though data may be used differently (for the public, internal staff, or specific departments), always handle it with care.

Manual Data

Converting any data collected manually (e.g., on paper) into digital form quickly.

Make sure the data entered is accurate and securely stored.

Automated and Semi-Automated Data

For data from automated or semi-automated systems (e.g., sensors or online forms), making sure it is securely transferred to a central storage system.

Regularly checking if the data is up-to-date and tracked properly.

Keeping Data Safe

Only allowing certain people to access the data, depending on their role in the council or in the project they are involved in.

Keeping a record of who accesses or changes the data.

Checking Data for Accuracy

Making sure data from all sources is checked and verified before it is stored to avoid mistakes.

Using Secure Storage

Storing all data in a central, secure location, with proper backups to prevent data loss. Ensure only authorised people can access the data.

Regularly back up data and store copies in safe locations.

Staying Compliant

Ensuring that data storage follows privacy laws and other regulations.

Regularly reviewing how data is stored and improve security measures when needed.

3.2. Data protection

Our data protection strategy will focus on ensuring that sensitive information is handled securely and responsibly. These guidelines outline key measures that employees must follow to protect data from unauthorised access and breaches while ensuring its proper use for legitimate purposes.

3.2.1. Access Control

Access to sensitive data will be restricted based on the role and responsibility of employees. This will prevent unauthorized individuals from viewing or handling data that they do not need.

- **Role-Based Access:**
  Access rights to datasets will be granted according to job roles and the nature of the project. For instance, analysts will have access to non-sensitive datasets, while HR or management will handle confidential personal information.
- **Time-Based Access:**
  Access to sensitive data will be limited to specific working hours or project deadlines. This ensures that data is only available during active work periods, reducing the risk of unauthorised access outside these times.

3.2.2. Data Monitoring

Continuous monitoring will be conducted to ensure data is accessed and used appropriately, detecting any unauthorised or suspicious actions.

- **Audit Trails:**
  All access to sensitive data will be logged. This will record when data was accessed, by whom, and what actions were performed, providing a clear trail of data handling.
- **Anomaly Detection:**
  Regular reviews of data usage will be conducted to detect unusual access patterns, such as employees accessing data outside their role or outside working hours. This will help identify any potential security issues early on.

3.2.3. Authorisation Protocols

Access to sensitive data will be granted only after proper verification, ensuring that only those who need the data for their work can handle it.

- **Data Access Requests:**
  Employees requesting access to sensitive data will need to provide justification, and their request will be reviewed to ensure that it aligns with their role and responsibilities and based on project they are involved in.
- **Data Sharing Approval:**
  Before any sensitive data is shared externally or with third parties, approval from the relevant data governance authority will be required. All shared data will be reviewed to ensure that it adheres to data protection standards.

Benefits of Implementing Robust Data Protection Measures:

a. **Compliance with Regulations**: Governments and industry groups have established strict data protection laws that organisations must follow, such as Australia's Privacy Act 1988. By employing encryption, managing access, and conducting regular monitoring, organisations can prove they are meeting these legal requirements.

b. **Mitigating Risk of Data Breaches**: Data breaches can cause big financial losses, harm to reputation, and a loss of trust. Strong data protection measures help reduce the chances of a breach by limiting unauthorised access and making the data useless to attackers if they get in.

c. **Enhancing Public Confidence**: When individuals know their data is being handled securely and responsibly, it fosters trust in the organisation. For government entities like the City of Darwin, effective data protection reassures the public that their information is used properly and safeguarded against misuse.

d. **Operational Efficiency**: With proper data protection in place, risks can be managed better, data can stay accurate, and our systems can resist attacks. This efficiency allows us to focus on using data for growth instead of constantly dealing with security issues.

3.3. Data Utility

We strive to ensure that, even after data has been transformed, anonymised, or de-identified to remove or obscure sensitive information, it remains valuable and useful for analysis, research, and public use. To protect sensitive information and comply with data protection regulations, we need to follow these steps when working with data. Our goal is to balance data privacy with the ability to derive useful insights.

Identify and Remove Sensitive Information

Before sharing any data publicly, thoroughly review the dataset to identify **sensitive information** (e.g., personal identifiers like names, addresses, contact details). Sensitive data should always be **removed** or **obscured** before public release, as it is not necessary for most public insights.

Action: For each dataset, assess whether specific details are required for the public's understanding. If they aren't, remove them.

Anonymise Personal and Indirect Identifiers

Beyond direct identifiers, ensure that **indirect identifiers** (e.g., dates of birth, postal codes, or job titles) are anonymised or generalized. Even if a dataset doesn't contain names or emails, combinations of data could still lead to re-identification of individuals.

Example: Instead of sharing precise age or exact geographic locations, aggregate ages into groups (e.g., "30-40 years old") and regions into broader areas (e.g., "Northern Territory" instead of a specific suburb).

Apply Aggregation for Broader Insights

To protect individual privacy while still providing useful public information, consider **aggregating data**. Present data at a group or regional level, rather than providing individual-level records. This method helps maintain data utility while safeguarding personal privacy.

Example: Share data on the total number of incidents in each region rather than providing specific details of each incident.

Ensure Data Minimisation

When preparing data for public sharing, follow the principle of **data minimisation**. Only share data that is necessary to achieve the intended purpose or benefit for the public. Avoid sharing any unnecessary fields or details, even if they seem harmless.

Example: If the public report is focused on traffic patterns, avoid including irrelevant details such as user preferences or browsing behaviour.

Review Data for Re-identification Risks

Carefully evaluate whether the combination of data points in the dataset could potentially allow for the re-identification of individuals. Even anonymised datasets can sometimes be linked back to individuals if too many details are provided. Review and reduce the granularity of the data to avoid this.

Action: Regularly conduct re-identification risk assessments before public data release. These assessments evaluate the likelihood that individuals could be re-identified based on the dataset's attributes.

3.4 Compliance

We are fully committed to ensuring that our data management practices comply with all relevant laws, regulations, and standards related to data privacy and protection. This commitment is grounded in the strict adherence to several key legislative frameworks that guide how personal information is collected, used, stored, and disclosed, particularly within the Northern Territory.

Key Compliance Frameworks:

Information Act 2002 (Northern Territory)

The **Information Act 2002** serves as the cornerstone of our data privacy practices within the Northern Territory. It sets forth the **Information Privacy Principles (IPPs)** that specifically govern the management of personal information by government agencies and contractors. This act mandates that personal data must be collected lawfully and fairly, stored securely, and only used for purposes that are clear and justified. By aligning with this act, we ensure that all government-related data is protected, reinforcing responsible handling and management of personal information.

Privacy Act 1988 (Commonwealth)

For any personal data that crosses territorial borders or involves federal entities, we adhere to the **Privacy Act 1988** (Commonwealth). This act, particularly through its **Australian Privacy Principles (APPs)**, provides comprehensive guidelines to ensure the transparency, fairness, and security of

personal data management across Australia. It outlines how personal information should be handled, including proper consent for data collection, secure storage methods, and restrictions on data sharing without proper authorisation. This national framework allows us to maintain privacy and

data integrity beyond the Northern Territory, ensuring compliance across the broader Australian legal landscape.

3.5. Data Quality

We are committed to ensuring the highest standards of data quality across all our data management processes. Our Data Quality Statement is guided by the following key principles.

Accuracy

Goal: Ensuring data is collected, recorded, and maintained to accurately reflect real-world scenarios.

How to Achieve:

- **Automated Data Validation:** By implementing automated checks to identify and correct invalid data entries during data collection or import.
- **Cross-Referencing:** By using trusted external data sources to verify key data points. For example, cross-check customer addresses with postal services to ensure accuracy.
- **Manual Audits:** By scheduling periodic manual reviews or random sampling of datasets to detect and resolve discrepancies.
- **Error Reporting Mechanism:** By Setting up a system where users can report errors or inconsistencies in data, ensuring continuous improvement in data accuracy.

Completeness
Goal: Capturing all necessary and relevant information without missing critical data.

How to Achieve:

- **Mandatory Fields:** In all forms and databases, make critical fields mandatory to prevent incomplete records. For example, ensure that fields like "date of birth" or "region" cannot be left blank if required for analysis.
- **Default Values and Alerts:** Provide default values for non-critical fields and set up alerts for missing critical data during the data entry process.
- **Regular Gap Analysis:** Periodically reviewing datasets to identify missing or incomplete information, and taking action to fill these gaps (e.g., through supplementary data collection).

Consistency

Goal: Ensuring uniformity in data formats, definitions, and standards across all datasets.

How to Achieve:

- **Standardised Data Formats:** Establishing organisation-wide data formatting rules (e.g., date formats, currency, naming conventions) and implement these in data collection and storage systems.
- **Centralised Data Dictionary:** Maintain a data dictionary that defines all key data elements, ensuring consistency in terminology and use across teams. Regularly update and make it accessible to all employees.
- **Validation Rules:** Apply validation rules in databases or data entry forms to ensure data adheres to the defined standards (e.g., dates follow DD/MM/YYYY format).
- **Consistent Reporting Templates:** Use standardised templates for reporting and data entry, ensuring uniformity in structure, format, and presentation across all departments.

Timeliness

Goal: Keeping data up to date to support timely analysis, reporting, and decision-making.

How to Achieve:

- **Automated Data Refreshes:** Where possible, automate data refreshes at regular intervals to ensure that the most current data is available. For instance, set weekly or monthly refresh cycles for time-sensitive datasets like sales or performance metrics.
- **Data Review Schedule:** Implement a scheduled review process, where datasets are regularly updated or archived if no longer relevant.
- **Real-Time Data Feeds:** For critical areas like customer feedback or live monitoring systems, set up real-time data integration to provide up-to-date information.
- **Notifications for Data Expiry:** Set notifications or flags on datasets that require regular updates. Alert responsible teams when data is nearing its expiry or needs refreshing.

Validity

Goal: Ensuring that data is relevant, appropriate, and meaningful for its intended use.

How to Achieve:

- **Predefined Data Rules:** Before data collection, define the acceptable rules and formats (e.g., specific value ranges, valid formats) to ensure that only valid data is entered.
- **Data Type Checks:** Set constraints within databases to prevent invalid data types (e.g., ensure a numeric field only accepts numbers and not text).

- **User Training on Valid Data Entry:** Train staff involved in data entry on the importance of following predefined rules and validating data before it is saved.
- **Periodic Data Cleansing:** Regularly review datasets to identify and correct invalid or irrelevant data points, such as outdated records or data that no longer serves its intended purpose.

Reliability

Goal: Providing trustworthy data for consistent use over time.

How to Achieve:

- **Data Version Control:** Implement version control for datasets to track changes over time, ensuring that updates are documented, and older versions are archived if necessary.
- **Backup and Recovery Protocols:** Establish data backup systems to safeguard against data loss and implement recovery protocols to maintain data integrity in case of unexpected events.
- **Monitoring Data Quality:** Set up a system for continuous monitoring of data quality, where key metrics like accuracy, completeness, and timeliness are tracked. Use this data to assess and improve the reliability of datasets.
- **Regular Audits and Assessments:** Conduct regular audits of data processes to ensure quality control measures are followed. Schedule periodic data quality assessments to identify and resolve emerging issues in data handling.

4. Data Classification:

City of Darwin is dedicated to ensuring the accuracy, appropriateness, and relevance of information shared both internally and publicly. To achieve this, data is categorised and treated based on its sensitivity and use case:

- **Tier 1 Data:** Confidential information that includes sensitive personal or corporate details. This data is to be managed with utmost care using techniques such as Data Suppression, Data Masking, Data Anonymisation, and Pseudonymisation are to be used as open data. This type of data can be used in the Raw form only for internal purpose and should not be shared openly or to the third party without going through transformation or necessary desensitisation process.

- **Tier 2 Data:** Semi-sensitive information that does not directly identify individuals or corporations but may contain detailed data about specific stories. This data requires transformation and de-sensitisation, commonly achieved by aggregating the data and providing summarised information. Also, data falling under this category should not be shared to the public or third parties without necessary transformation and desensitisation process.

- **Tier 3 Data:** Non-sensitive information that does not breach privacy and does not identify individuals or companies. This type of data does not contain any sensitive information and can be shared directly upon approval from the management

5. Definitions

Datasets are a collection of data presented in tabular or non-tabular formats.

Data Steward is responsible for managing critical data, both reference and transactional, created or used by one business function.

Metadata is data that defines and provides information about one or more aspects of the data through summarising, allowing users to discover, manage, control, and understand that data. For example: document creation date, author, last update, spatial information, etc.

Open data is publicly accessible and assumed not to include any sensitive information.

Internal data originates from processes within the Council or in other government agencies. Internal data may or may not contain sensitive information that require further assessment.

Confidential data contains sensitive information. Unauthorised access to this information could lead to significant privacy or security issues.

Data de-identification is removing or replacing potentially identifying information from datasets.

Data aggregation combines individual records into summarised groups based on non-sensitive attributes to reduce the potential for personal identification.

Pseudonymisation: Replacing personal identifiers with pseudonyms to protect identities while allowing data to be used in a controlled manner.

Data quality control includes rectifying errors, standardising data formats, dealing with missing values, etc.

Review and data validation ensure compliance with the Council's data release standards and may involve automated checks or multi-level audits.

Data Suppression: Removing or hiding entire records or fields that contain sensitive information.

Open Data: Information that can be freely shared without any risk of exposing sensitive details such as personal information, detailed financial information or any information that leads to the identification of any personal or a corporate body.

Sensitive Data: Information that could potentially identify individuals or disclose confidential details, requiring transformation or anonymisation before being made public. Sensitive information that cannot be shared openly must be protected from public access.

.

6. Handling &Transformation procedure for Open data.

Data will be treated following all the procedures highlighted in above statements along with the guidelines below which provides a robust ground for a data to be shared in Open portals.

- Data should be summarised at a sector or regional level. Any personal or sensitive information must be supressed.

- Decisions should focus on outcomes rather than personal identities unless these identities are already public.

- Remove or mask any proprietary or confidential business information before public sharing.

- Aggregate contract values and vendor information to avoid revealing sensitive business details. Approval may require before public disclosure.

- Aggregate and anonymize payment data, including personal data related to infringements, such as names and license plate numbers, to publish only aggregated statistics.

- Aggregate usage data to prevent exposure of patterns that could identify individuals.

- Adhere to third-party privacy requirements and aggregate location data, such as Google Places data.

- Aggregate data by region or type of approval (e.g., Building Approvals) and suppress individual application details.

- Summarise data to protect the privacy of individuals and avoid identifying specific animals or owners.

- Aggregate financial data to suppress details that could identify specific recipients.

- Aggregate data to prevent identification of specific planting sites or recipients.

- Generalise location data to avoid exposing specific addresses, such as in tree plantations or greening projects.

7. Data handling process for Internal use

Internal staff can use data that is internally generated, collected, and shared, which is stored either in our databases, SharePoint, any third-party data platforms or local drives.

The Data to be used for internal purpose may fall into any category (i.e. Tier 1, Tier 2, Tier 3) and may contain all kinds of information about a particular department or a sector. This information maybe sensitive and may contain personal information or a detail information about a corporate body. Data for internal purpose must be kept confidential and should only be used for internal purposes.

Purpose: Internal staff can use the data for operational decision-making, service improvements, and strategic initiatives. This includes ongoing projects, reports, performance assessments, and other data-driven actions.

Actions:
- **Accessing Data**: Use internal platforms (databases, SharePoint, third-party platforms, or local drives) to extract relevant data.
- **Data Cleaning**: Ensure the data is clean, organised, and accurate before analysis.
- **Data Analysis**: Utilise analytical tools (such as Excel, Power BI, Python) to draw insights, identify trends, and generate reports.
- **Decision-Making**: Based on analysis, make informed decisions for improving service delivery, optimizing operations, or planning future strategies.

8. Data Guideline for third party sharing.

Data sharing with third parties, such as government organizations (e.g., NTG, CSIRO) and non-government entities (e.g., external event organizers), is conducted only after a formal data-sharing agreement is signed. Before data is shared, the request is carefully assessed, considering the requirements of the requesting party and their data reusability practices. These data are carefully considered and under the assessment of data owners and Data steward to understand the purpose of the data sharing.

The data requested for third party sharing is assessed under 3 terms.
- ✓ If/whether the requested is from Government/Non- Government entities. If they are from non-government entities a background check may be required.
- ✓ What is the purpose of the request or what information they are trying to gain from a dataset?
- ✓ Whether the information they are requesting contributes positive impact to the community?

When a request is received, based on the requirements the data can be customised to match the requirements. The customised data need to be reviewed and processed to make sure the data follows the council standard and does not contain any sensitive, personal or detail information about any individual or a corporate body. The prepared data needs to go through the approval process before being shared.

9. Data approval

Data to be published for the public/elected members must be approved and quality-checked by the relevant program areas' data custodians. After creating a story/dashboard, it will be shared with each program area's data owner (the Executive Manager of each program area), who will review the wording/ story / trends and decide whether it is suitable for public sharing. This process will occur every time council updates the data or story.

The data steward will check and approve any new story/dashboard additions, updates to existing stories/dashboards, or deletions. The data steward will check against a matrix that outlines what can be shared with the public, elected members, or internal staff.

11. Data publication

The City of Darwin ensures that datasets are accurately and timely published, following a structured approval process to guarantee data quality and relevance.

1. Initial Review: Each dataset is thoroughly analysed by a data analyst to assess its alignment with the stories or insights the council aims to convey.

2. Approval Workflow: The dataset then goes through an internal approval process, where relevant departments such as the Executive Manager, Subject Matter Experts (SMEs), and the General Manager of Innovation evaluate the dataset. They determine the appropriateness of sharing the data with the public or Elected Members.

3. Final Approval: Following the review, the dataset can only be published after receiving formal approval from the General Manager of the council.

4. Audience-Specific Publishing: Data that has been approved may be published either to the public or exclusively to Elected Members, depending on the decision made by the Executive Manager, SMEs, and the GM of Innovation.

12. Data disposal

To ensure data security and compliance with regulations, City of Darwin distinguishes two kinds of data disposal:

12.1. Disposal of raw datasets.

City of Darwin follows strict data disposal procedures based on the retention period and the data's sensitivity. For physical storage media (e.g., hard drives, tapes), secure destruction methods such as shredding or degaussing will be used to ensure data cannot be retrieved. In the case of digital data, secure erasure techniques will be employed to permanently delete files, ensuring they are unrecoverable.

All data disposal activities will be carried out in compliance with relevant legal and regulatory frameworks and will adhere to the council's Data policies.

12.2 Disposal of Historical Datasets with Reference Value

When disposing of historical datasets, the council will carefully assess whether the data holds any long-term reference or archival value before proceeding. If deemed significant for historical, legal, or operational reasons, the dataset may be preserved in a secure archive following proper classification protocols.

However, if the data is no longer required and poses a risk due to sensitivity or regulatory obligations, it will be disposed of following the council's secure disposal procedures. Prior to disposal, the data will be evaluated to ensure that no further value can be derived and that its retention does not conflict with data management and privacy standards.

The council will always align this process with applicable legislation and records management guidelines to ensure both compliance and responsible stewardship of information.

7.2.13. Legislative References

Australian Privacy Act 1988 (Cth)

Australian Data Availability and Transparency Act 2022 (Cth)

Information Act 2002 (NT)

Local Government Act 2019 (NT) - Section 292 Public access to information

14. Procedures and Related Documents

City of Darwin 0033.100.E.R Privacy Policy

City of Darwin 2001.100.I.R Records and Information Management Policy

City of Darwin 0086.100.E.R OPEN DATA POLICY

15. Responsibility and Application

The Chief Executive Officer (or delegated authority) is responsible for ensuring the understanding and compliance with this Policy by all relevant City of Darwin staff.

16.Appendices

Appendix A: Existing Dataset Classifications

Appendix B: Data Classification Toolbox

Appendix C: Compliance Checklist

	START			INTERMEDIATE				END
	STANDARD	SUB-PROCESS INTERRUPTING	SUB-PROCESS NON-INTERRUPTING	CATCHING	BOUNDARY INTERRUPTING	BOUNDARY NON-INTERRUPTING	THROWING	STANDARD
Untyped events. Indicates START point, state changes and END point / final state	●						●	●
Messages - Receiving and Sending system messages	●	●	●	●	●	●	●	●
Timer - Cyclic timer events, points in time, time spans and timeouts	●	●	●	●	●	●		
Escalation - Flow to a high level of authority in the process		●	●		●	●	●	●
Conditional - Reacting to changes in conditions or integration of business rules	●		●	●	●	●		
Link - Off-page connectors. Both are required for a legitimate sequence flow				●			●	
Error - Named error process flow		●			●			●
Cancel - Reacting to a cancelled transaction or triggering of cancellation event				●				●
Compensation - Trigger or handling of compensation event		●			●		●	●
Signal - Send and Receive signals across processes.	●	●	●	●	●	●	●	●
Multiple - to catch one event from a set.	●	●	●	●	●	●	●	●
Parallel Multiple - to catch all events from a set	●	●	●	●	●	●		
Terminate - Triggering the immediate termination of a process								●

Figure 16: BPMN - Flow Objects – Events

	Type	Symbol	Description
Tasks	User Task	▢ or ▢	Requires a human user to complete the activity (e.g., approve a document).
	Service Task	▢	Performed by an automated system or application (e.g., calculate a discount).
	Send Task	▢	Sends a message (e.g., email notification).
	Receive Task	▢	Waits for a message to be received (e.g., customer confirmation).
	Business Rule	▢	Interacts with a separate business rules engine to evaluate pre-defined rules based on process data. It sends data to the engine, receives the output, and continues the process flow.
	Script	▢	Executes a pre-written script directly within the BPMN engine. The script can manipulate data, perform calculations, or interact with internal systems.
	Manual Task	▢	Represents a step that requires human intervention outside of the BPMN workflow system. It doesn't involve the process engine and signifies a physical or offline activity.
Sub-Process	Call Sub-Process	▢	This acts as a reusable building block. It references a pre-defined process that exists elsewhere, promoting modularity and avoiding redundancy. The main process flow delegates the work to the referenced sub-process and continues after its completion. Call sub-processes are depicted with a thick border.
	Event Sub-Process	▢	This sub-process is independent of the main flow's sequence. It waits for a specific event to trigger its execution, such as a timer going off or receiving a message. The event sub-process can either interrupt the main process or run concurrently depending on the triggering event type. It's shown with a single, thin dotted line border.
	Transaction	▢	This sub-process focuses on ensuring data consistency within a specific process section. It acts as a unit of work, and its success or failure determines the overall outcome. Transaction sub-processes have a double border and can result in successful completion, rollback, or failure.
Markers	Loop Marker	▢	This indicates the activity or sub-process repeats itself iteratively until a certain condition is met. Imagine a loop marker attached to a 'review document' task, signifying the task keeps looping until the document is approved.
	Multi-Instance Marker (Parallel)	▢	This signifies the activity will be executed for each item in a collection **at the same time**. Imagine a task for 'generating reports' with this marker. The process would initiate report generation for each item concurrently, potentially improving efficiency.
	Multi-Instance Marker (Sequential)	▢	This indicates the activity will be executed for each item in a collection, but one after the other in the order they appear in the collection. For instance, a 'review application' task with this marker would process each application sequentially, ensuring each application receives full attention before moving to the next.
	Compensation Marker	▢	This signals the activity is part of a compensation process. Compensation involves reversing actions taken earlier in the flow if an error occurs later on.
	Ad-Hoc Marker	▢	Indicates that the sub-process contains a set of activities that can be performed in a non-rigid manner. This means the order of execution, the number of times each activity is performed, and even skipping certain activities altogether become possibilities within the sub-process.

Figure 17: BPMN – Activities
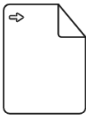
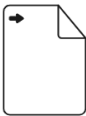|	Symbol	Description
Sequence Flow		Represents the order in which activities or events are performed within a process. It connects two flow objects, indicating the direction of process flow.
Message Flow		Represents the flow of messages between participants or pools in a process. It illustrates the communication paths between different entities.
Conditional Flow		Represents a flow in the process that is taken based on a specific condition or decision.
Default Flow		Default flow is the default branch to be chosen if all other conditions evaluate to false.
Data Associations		Represents the flow of information between data objects, stores, inputs, and outputs.
Assocations		Represents a link between Artifacts with other BPMN elements, and are most commonly used to connect a Text Annotation with activity, gateway or event.

miro

Figure 18: BPMN Connecting Objects

	Symbol	Description
Exclusive		Represents a decision point where only one of the outgoing paths can be taken. The decision is based on evaluating conditions associated with each outgoing sequence flow.
Inclusive		Represents a decision point where multiple paths can be taken based on evaluating conditions associated with each outgoing sequence flow. All paths with true conditions are taken.
Parallel		Represents a point where multiple paths can be taken simultaneously without evaluating conditions. It is used for parallel execution of activities.
Complex		Represents a more complex decision point where conditions and rules may involve a combination of logical operators. It allows for more sophisticated decision-making logic.
Event-Based		Represents a decision point based on events. It is used when the process flow depends on the occurrence of specific events, such as receiving a message or a timer event.
Exclusive Event-Based		Similar to the event-based gateway, but enforces exclusive decision-making based on events. Only one path can be taken, depending on the first event to occur.
Parallel Event-Based		The occurrence of all subsequent events starts a new process instance.
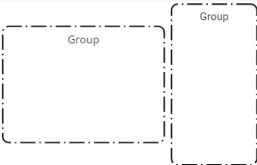
Figure 19: BPMN Gateways

	Symbol	Description
Data Object		Represents data or information used or produced within a process. It helps illustrate the flow of data between activities.
Data Input		Represents the input of data from an activity. It indicates the flow of data into a task or subprocess.
Data Output		Represents the output of data from an activity. It indicates the flow of data out of a task or subprocess.
Data Collection		Represents information collected within a business process.
Data Store		Represents a place where data is stored during the execution of a process. It can be a physical repository or a database.
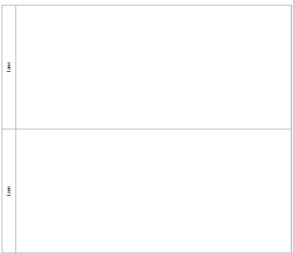
Figure 20: BPMN Artifacts - Data Objects

	Symbol	Description
Group		Groups related elements in a diagram. It is often used to visually organise and highlight specific sections of the process.
Annotation		Provides additional information or comments to enhance the understanding of the process. Annotations are often used to add explanatory notes or documentation.
Assocation		Connects artifacts, data objects, or text annotations to flow objects, indicating a relationship or dependency. Associations help in clarifying connections between elements.
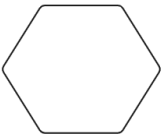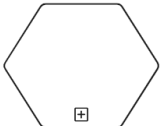
Figure 21: BPMN Artifacts - Group, Annotation and Association

	Symbol	Description
Pool		Pools are used to represent separate organisational entities or participants in a process. Each pool can contain its own set of activities and processes, and communication between pools is depicted using message flows.
Lane		Lanes are subdivisions within a pool and are used to organise and categorise activities further. Each lane typically corresponds to a specific role, department, or system within the organisational entity represented by the pool.

Figure 22: BPMN - Pools and Swimlanes

	Symbol	Description
Conversation		Represents a conversation or communication between different participants in a business process. It provides an overall context for the interactions.
Call Conversation		Represents a more detailed or nested conversation within a larger conversation. It is used to provide a more granular view of interactions within a specific part of the overall process.
Sub-Conversation		Represents a call or sub-process that is initiated from within a conversation to represent a more detailed interaction or subprocess.
Sub-Call		Represents a connection or link between different parts of a conversation or between different participants. It indicates that there is a relationship or interaction between them.

Figure 23: BPMN Artifacts - Group, Annotation and Association

Figure 24: Meeting with Stakeholders

Figure 25: Monday.com

Figure 26: Timeline in Monday.com

Keywords	
Smart City Frameworks	Data Governance in Smart Cities
Urban Data Management	Internet of Things (IoT) Smart Cities
Big Data Analytics Smart City	Smart Urban Infrastructure
Smart City Data Privacy	Smart City Standards
Smart Governance	Data Integration in Smart Cities
Smart City Sustainability	Urban Innovation Technology
Smart City Operational Efficiency	Data Protection Smart Cities
Smart City Compliance and Regulation	

Figure 27: Keywords Searched on Internet