Data Governance Policy

Made in reference to https://www.st-andrews.ac.uk/policy/information-governance-and-management-data-governance/data-governance-policy.pdf.

|  |  |
| --- | --- |
| Document type | Policy |
| Scope | Data management in the Data Lake by all staff |
| Applicability date |  |
| Expiry date |  |
| Approver | Christian Mejrup |
| Document Owner | Rune Kennet Christensen |
| Organisation | Datamatiker |
| Document status | Draft |
| Purpose | The purpose of this policy is to understand how to approach data governance in terms of data security, data availability, data accessibility, data quality and how to define clear accountability for each to be managed correctly by all stakeholders. |
|  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Version Number | Purpose / Changes | Document Status | Author of changes, Role | Date |
| 1.0 | For consideration as new policy | Draft | Datamatiker, Rune |  |

Indhold

[1. Policy 3](#_Toc102552788)

[2. Definitions 3](#_Toc102552789)

[3. Requirements 4](#_Toc102552790)

[4. Risks and Threats 5](#_Toc102552791)

[5. Scope and Success 5](#_Toc102552792)

[6. Principles 5](#_Toc102552793)

[7. Data governance framework 5](#_Toc102552794)

[8. Roles and responsibilites in the data steward community. 5](#_Toc102552795)

[9. Data assignment 5](#_Toc102552796)

[10. Data quality capability 5](#_Toc102552797)

[11. Data quality oversight 5](#_Toc102552798)

[12. Data security 5](#_Toc102552799)

# Policy

1. ”Årshjulet” needs high-quality data to manage its assortment of different Sermons and hymns, to sustain for future growth, to properly display the history of these Sermons and hymns.
2. Data Security, Data Availability, Data accessibility & Data Quality are measure of good data governance. This policy will describe “Årshjulet”’s approach to data governance in those terms and outline how clear accountability for each measure is managed.

# Definitions

For this policy the following definitions apply.

* Data is defined as “numbers, word or images that have yet to be organized or analyzed to answer a specific question”
* Data assets is defined as “Sermons / Hymns” along with its associated metadata.
* Data Governance includes the people, processes and technologies used by the “Årshjulet” team to manage and protect its data assets including definitions for how the team assigns accountability and control over the assets and their use.
* Data Quality refers both to the characteristics associated with high quality data (see appendix) and to the processes used to measure or improve quality of data.
* Data Accessibility refers to the retrieval of data in an authenticated manner approved by the “Årshjulet” team. This may be for the purpose of reading, modifying, copying or moving data from a system
* Data Security includes data confidentiality, data integrity and data accessibility.
* Information life cycle is an approach to data & storage management that recognizes that the value of information changes over time and that it must be managed accordingly. See Appendix 2
* Big data refers to extremely large data sets that may be analyzed computationally to reveal patterns, trends, associations. But also, to data that contains greater variety arriving in increasing volumes and with more velocity. Big data should always include the 5 V’s.[[1]](#endnote-1)
* Data Lake is a central repository designed to store, process and secure large amounts of structured, semistructured and unstructured data.
* Storage is the folder with the PDFs on the wordpress platform.
* Personal data
* Reference data

# Requirements

”Århjulet” must ensure the availability and the quality of its data assets to:

* Ensure that data assets are accessible to users.
* The data is secure.
* Produce accurate data assets based on the user’s preferences.

# Risks and Threats

1. ”Årshjulet” suffers, when the data assets aren’t accessible, aren’t based on the users preferences or isn’t following the shared understanding.
2. This can happen due to poor regulation or infrastructure, deficient data accessibility, erosion in data quality and/or disconnection between staff responsible for data collection vs information creation.
3. Symptoms of the data not being accessible are evidence in enhanced risks such as:
   1. Less users
   2. Lack of information
   3. Loss of productivity due to time spent on non-value tasks.
4. Som..

# Scope and Success

1. The successful implementation of this policy will primarily be evidence by the governance of data domains represented in “Årshjulet”, however this policy is not limited to the data in the Data Lake.
2. The scope of this policy covers all data held in the data Lake and in the website ATM. This policy does not cover data from another data owner.
3. Delivery of the objectives in this policy relies on the successful application of data security arrangements to protect data from unauthorized access.

# Principles

1. The principles in Appendix 3 should be applied to the management of all data within the Data Lake and should also be applied to associated operational processes, goals and staff training.
2. Exceptions to these principles must be documented and visible even when principles allow for exception handling (eg, “Data should be collected and recorded once only wherever possible without the need for multiple systems”)

# Data governance framework

# Roles and responsibilites in the data steward community.

1. Every member of staff who interacts with data at any level in the organization has a role to play in the improvement of data accuracy and completeness in completeness in compliance with this policies requirement.
2. Each role in the data steward community must have unambiguous, easily understood and publicly documented responsibilities.
3. The information life cycle recognizes different relationships to data. Data producers (wheather people or systems) control the data they create. Sometimes data are created for one purpose, but used for other purposes by data consumers. Because data producers have knowledge of the purposes and functions of associated processes they own. They can then modify processes to ensure they meet the needs and requirements of the data consumers.
4. Any person (or system) who has access to “Årshjulets” data is a data consumer, therefore data consumers encompass most of “Årshjulets” own staff whether they contribute directly to data collection or edits. Data consumers have a responsibility to follow established guidelines for accessing, sharing and updating data as well as participate in activities that defines data for use.
5. Who has the responsibility for all data?

# Data assignment

# Data quality capability

# Data quality oversight

# Data security

# Appendix

## Appendix 2: Data Lifecycle Management: Tracking your data accurately throughout the information lifecycle helps you determine wher… | Life cycle management, Data, Data security

1. Volume, Value, Variety, Velocity & Veracity. [↑](#endnote-ref-1)