Catena-X

THE FIRST OPEN AND COLLABORATIVE DATA ECOSYSTEM

Data Governance Guide

This Guide gives a Rough Overview of Basic Governance Principles for Cross Company Data Exchange

Release V1, January 2023

TABLE OF CONTENTS

Supported by Federal Ministry for Economic Affairs and Climate Action

ABSTRACT

From an ecosystem perspective, this document highlights platform-based governance mechanisms for building data and analytics use cases with the purpose of delivering and receiving data assets to or from third parties. We focus on the people and their roles in the organization, as well as the processes required to govern and manage a Data and Analytics Ecosystem that spans functions, departments and third parties.

In the broader sense, the target audience this document wants to address includes data consumers and data producers participating in the ecosystem in general. In the narrow sense, the target audience includes specifically

1) Data & Analytics Roles
2) Governance Roles.

Thereby **Data & Analytics Roles** represent roles directly involved within the Data & Analytics Value Chain, either in a core or supporting role. Therefore, Data & Analytics Roles perform operational activities (e.g., developing and maintaining Assets and Use Cases), directly contributing towards a sustainable data ecosystem. Supporting **Data Governance Roles** provide guidance for Data & Analytics Roles on how to perform activities of the Data Value Chain in a compliant manner. Hence, those roles translate internal and external norms and regulations into binding policies and standards which can be applied in practice.

1. INTRODUCTION

Organizations can capitalize on collaborating with partners in ecosystems, but there is also a risk of losing control on data, granting unsecured access to information, or providing low-quality information.

To deal with these issues and support the objectives of a shared ecosystem**, inter-organizational data governance mechanisms** need to be established. These mechanisms can be grouped into the four areas of data management processes, data governance policies, IT governance policies, and digitization strategies.

Data management processes encompass various aspects of handling data. This includes defining specific roles and their respective responsibilities. In addition, data policies provide rules for the creation, control, management, and audit of data. The development of data standards is particularly important for the inter-organizational exchange of data. These standards define, for example, how data is handled and how it is represented to ensure that the required quality criteria are met.

Data governance policies target the methods to regulate data inside and outside the organization. On the one hand, this involves establishing processes and procedures for data use and data flow. On the other hand, the setting for data provisioning and data sharing is defined. In this way, the sharing of data between two or more organizations is controlled, including the descriptions of the data, the data flow, and the obligations for providers and users through legal and data governance terms.

Supported by Federal Ministry for Economic Affairs and Climate Action

IT governance policies are established to address the complexities of managing digital documents and ensuring appropriate security. This can be achieved by leveraging technology with its ability to automate and scale the implementation of standards, processes, and rules.

Digitization strategy embrace transformation in terms of the adoption and leverage of digital technologies. They are essential for organizational success. This has a strong impact on the organization and, to this end, includes interaction with competitors, suppliers, and customers across organizational boundaries.

Supported by Federal Ministry for Economic Affairs and Climate Action

2. BUILDING A DATA EXCHANGE USE CASE

Besides data, specific analytics tools, methods, and skills are needed for value generation. Data usage in the context of Data & Analytics (D&A) comprises all elements that are required to build and operate D&A Use Cases. A D&A Use Case is characterized by any kind of data processing for direct or indirect value generation for a specific purpose.

This section focuses on the use case of sharing data with third parties. For a brief role description we refer to Appendix A.

The process is initiated by the need for a D&A Use Case within a business unit. The first step is to check whether there is already a use case that meets the requirements of the business unit. If this applies, a request can be made to gain access to this corresponding use case. Otherwise, the process must be resumed and a new use case has to be developed.

If an existing use case needs an extension, such as additional data or a change of purpose, similar steps must be performed as for building a new use case.

The first steps in building the use case are initiated by the creation of a **use case profile**. For this goal, the use case owner first defines the needed data. In the next step, the business value generated by the use case is estimated and compared to the expected costs. Aspects that must be considered include, for example, whether the needed data is available in the company data lake prepared according to the data requirements. Furthermore, the use case owner must check for the availability of the required personnel capacity. After the use case profile has been created, it must be determined which data management governance function (DMGF) will be the leading and responsible function for the use case. Usually, this is the DMGF assigned to the department of the respective use case owner.

If an existing use case needs to be extended, the use case profile must be adjusted accordingly, and the steps just mentioned have to be taken regarding the new conditions.

Based on the use case profile, the responsible DMGF first of all decides on whether the use case may be developed further. To better assess the effort and benefits, the DMGF may ask for the support of the data stewards who are affected.

Once the use case has been approved by the DMGF, the use case owner creates the use case including its technical environment on the D&A platform/data lake and documents it over the use case lifecycle. In addition to the information in the use case profile, this documentation also comprises details of the current situation and the objectives to be achieved. In addition, the use case owner maintains important metadata, such as the relevant information for "transfer of data to third parties" in the case of data exchange use cases.

In the next step, a **data assessment** has to be conducted. This ensures compliant data usage in the use case. The use case owner is responsible for the correct and complete implementation and should consult specialized departments within the organization for support in the individual tasks.

To combine and analyze large data sets within the data lake in a compliant way, individual-related data needs to be protected against misuse. According to data protection laws data may only be collected and processed for a valid purpose. A privacy impact analysis is needed to ensure that the

Supported by Federal Ministry for Economic Affairs and Climate Action

data processed by the use case is compliant with data protection requirements. Depending on the purpose of the use case, it can be checked whether it belongs to a particular cluster, fostering the reuse of data and facilitating the assessment. To this end a set of specific rules has already been established as to how personal data must be processed through anonymization or pseudonymization. Particularly for use cases with the purpose of transferring data to third parties, adequate data preparation is essential. If necessary, the data protection officer in the organization will provide support for this task. Together with the data stewards and, if necessary, the legal department, the use case owner will also have to consider possible violations of antitrust laws when making the data available to competing organizations.

A use case must meet the required protection level of the information processed and the potential business value at risk. Therefore, the use case owner determines which security class is to be assigned to the information obtained in the use case. The data steward should be consulted for this task. The final step of data assessment is to verify that the data is being used in accordance with the company's data strategy.

After data assessment, the use case owner is in the position to request the necessary data. For data located inside the organization on the D&A platform, the responsible DMGF, supported by the data stewards, will review the data requests based on the given data assessment. The result of the review is communicated to the use case owner. If not all data requests have been approved, the owner can adjust the use case and thus the data requirements. If the use case owner chooses to adjust the use case accordingly without the intended purpose disappearing from view, the DMGF will reassess whether the required data for the modified use case can be approved.

Once the data requests have been approved, the actual technical implementation of the use case comes up. Thereafter, it is released in the productive system.

If external data, that is data from an external partner in the network, is to be processed in the use case, a review by the DMGF is not required for this data. In this case, however, the use case owner must ensure that the data is only consumed according to the usage policies set up by the external partner.

Once the use case is finally created and live, the use case owner is responsible for managing the access requests. Access requests can be divided into internal and external requests. For internal requests, that are made via the D&A platform/data lake, the use case owner can also accept or reject the request directly on the platform.

If a request for a use case is made by an external partner, the owner has to generate a contract offer for this use case. In this contract offer, access policies define which persons or user groups will be granted access to the use case. In the usage policies, the use case owner defines how the data may be processed by the external partner. After having been authorized and agreeing to these access policies, this partner is allowed to capitalize on the use case.

The introduced process flow is visualized in Figure 1 using the BPMN 2.0 [1] notation.

[1] https://www.bpmn.de/lexikon/bpmn/

Supported by Federal Ministry for Economic Affairs and Climate Action

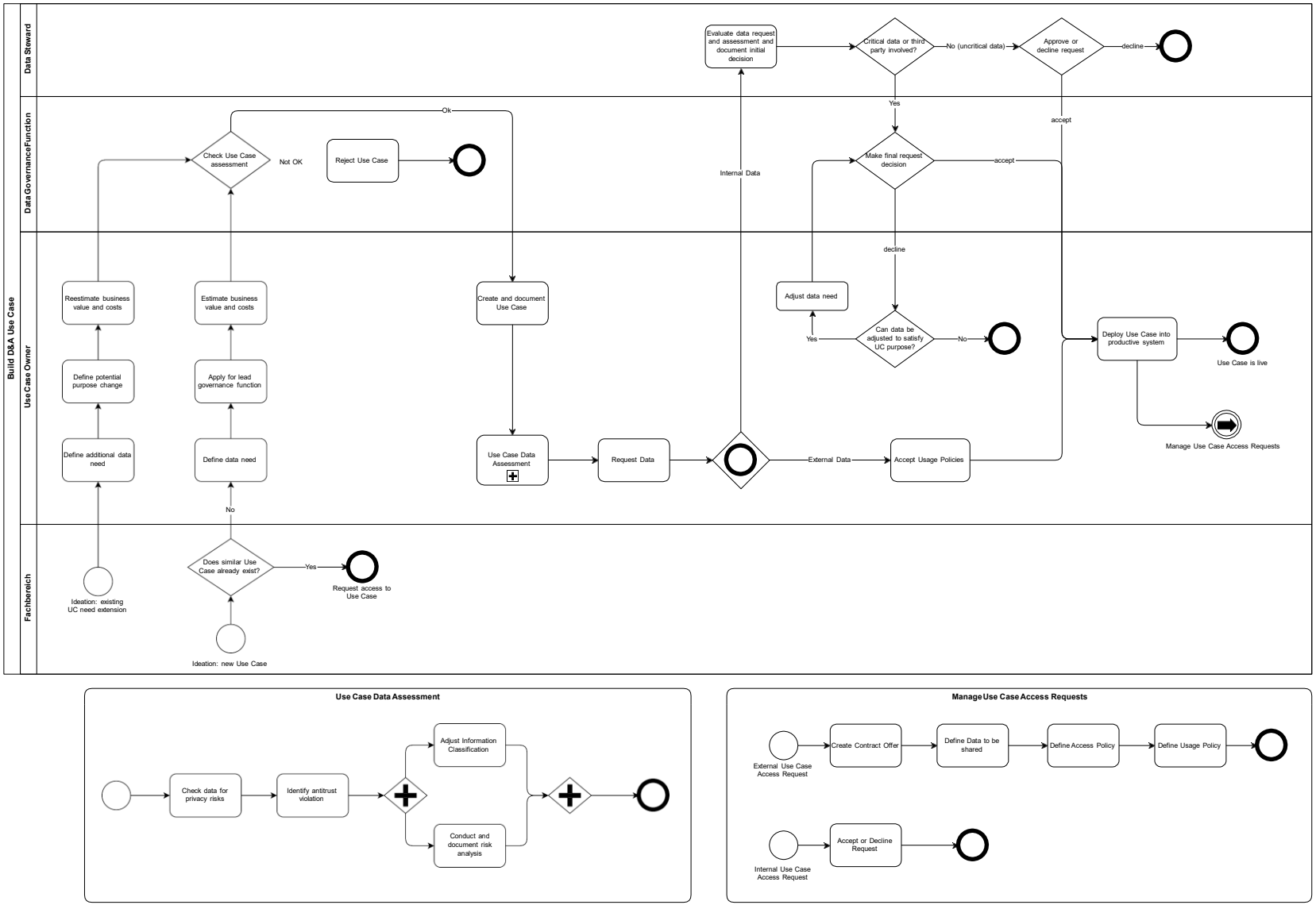

Figure 1: Process Model for Creating a Data & Analytics Use Case

3. APPENDIX A – ROLE DESCRIPTION

This section gives a brief overview and description of the roles involved in the Governance Processes related to Cross Company Data Exchange.

Business Department

The corresponding business unit where the value creation of the use case takes place. This could be procurement, logistics, production, R&D, IT, or similar.

Use Case Owner

A person or institution that is responsible for one or more use cases and the corresponding success. A use case owner can be supported by data scientists, analysts, software engineers, etc.

Data Steward

A person that transforms data from source systems into semantically interpretable data products. As an owner of one or more data products, a data steward is responsible for developing and operating allocated data assets and consults use case owner in how to interpret data belonging to a data product.

Data Management Governance Function (DMGF)

A DMGF is responsible for a decentralized data management and Governance function, defining and providing guidance on data governance, data management and data usage guidelines in a business area (e. g., purchasing, aftersales, customer support, IT, etc.). In addition, a DMGF holder is responsible for identifying and leveraging Data & Analytics (D&A) potential in the corresponding business area.

Data Transformation Board (DTB)

The DTB is a cross-functional, cross-entity leadership committee to provide day-to-day D&A governance, direction and oversight.