Decision Criteria for Selecting Data Infrastructure Design Options in the Private Sector
Online Appendix B:
A Practical Methodology for Selecting Data Infrastructure Design Options Based on the Catalog of Criteria

General Information:

This appendix to the paper *"Decision Criteria for Selecting Data Infrastructure Design Options in the Private Sector"* presents a practical methodology for applying the criteria catalog to identify and evaluate design options for private-sector data infrastructures for data sharing. It follows established enterprise architecture procedure models such as the Generalized Enterprise Reference Architecture. The methodology assumes a greenfield approach to building a data infrastructure. The starting point for applying the model is the motivation to implement at least one data-sharing use case supported by the data infrastructure.

Figure A depicts the four phases needed to identify the most suitable data infrastructure design option. Four phases are needed to identify the most suitable design option:

- Environmental analysis: Identifies and evaluates key stakeholders, influencers, and their expectations.
- Business model analysis: Identifies the relevant business model and use case characteristics that influence the data infrastructure development (e.g., the relevant data and systems).
- Requirements definition: Derives and prioritizes data infrastructure requirements based on the environmental and business model analyses.
- Preliminary design: Develops potential design options for the data infrastructure and selects the most suitable alternative based on the defined requirements.

While preceding phases define standards, requirements, and boundary conditions for work in the succeeding phases, the succeeding phases evaluate the feasibility of these mandates and give feedback to their predecessors. We elaborate on each of the phases in greater detail below. Lastly, we describe the application of the criteria catalog within this procedure.

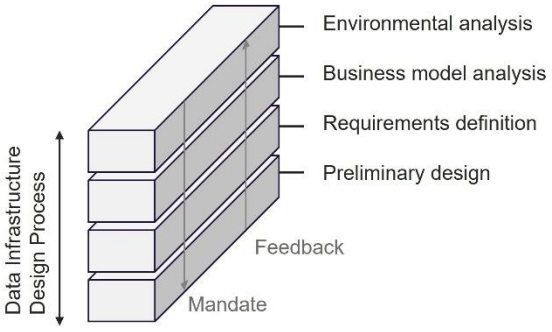

Figure A. Data Infrastructure Design Process

Environmental Analysis

Goal: Identification of the stakeholders and other influencers of the data infrastructure initiative.
Inputs: Conceptualization of the envisioned use case(s).
Outputs: Structured overview of stakeholders, their concerns and the related risks and opportunities.

The environmental analysis identifies and assesses the stakeholders, influencers, and concerns in the context of the data infrastructure that serve as a basis for selecting decision criteria. This involves three dedicated steps. First,

relevant stakeholders need to be identified. Stakeholders of data infrastructure initiatives are, on the one hand, actors who directly influence the design of the data infrastructure and, on the other hand, actors who are affected by the data infrastructure. The environmental analysis focuses on stakeholders outside the data infrastructure initiative. The external environment is usually structured into political, economic, social, technological, ecological, and legal (PESTEL) domains.

After the relevant stakeholders have been identified, their concerns need to be derived. Typically, concerns emerge from a stakeholder's knowledge and experience, their responsibility and their mission. Concerns can have a general nature or refer to specific aspects of data infrastructures, such as their resource consumption. A stakeholder can have one or more concerns.

Lastly, the extent to which stakeholders can legitimately influence the design of the data infrastructure is evaluated. Only important claims should be considered in the subsequent phases. The assessment draws on the stakeholder's role and their financial, social, or legal status. It can be framed in terms of risks and opportunities of complying with or neglecting a specific stakeholder concern. All identified information should be presented in a neutral tone and summarized.

Business Model Analysis

Goal: Identification of business model and use case characteristics.
Inputs: Conceptualization of the envisioned use case(s) and business model(s).
Outputs: Structured overview of the business model and use case characteristics relevant to the development of the data infrastructure.

Since the data infrastructure is implemented to support at least one use case and its associated business model, alignment with the business model must be ensured. This phase therefore derives the key concerns for data infrastructure development from the business model. To this end, the key components of the business model need to be identified. The business model canvas can be employed to describe the core business model components, including the needed data assets and data providers, key partners, and monetary characteristics that influence the data infrastructure design. Subsequently, data assets, value streams, and information flows should be examined in greater detail, as they constitute core components of the data-driven business model and directly inform the implementation of the data infrastructure. Example characteristics include data types, frequency, and sensitivity, as well as the bargaining power of the involved data-sharing partners and possible data-sharing incentives. This information should be clearly documented to serve as input for the next phase.

Requirements Definition

Goal: Identification and evaluation of relevant decision criteria.
Inputs: Stakeholder overview and technical business model characteristics.
Outputs: Catalog of weighted decision criteria.

The aim of the requirements definition phase is to identify and evaluate the relevant decision criteria for the selection of data infrastructure design options based on the concerns derived from the data infrastructure stakeholders, business models and use cases. This approach allows decision criteria and their evaluations to be traced to underlying motivational factors and adjusted as these factors change. In a first step, the relevant decision criteria are derived. Based on the experiences obtained during their research, the authors recommend using so-called "Joint Requirements Planning Sessions". These sessions bring together data infrastructure stakeholders from diverse backgrounds to discuss and collaboratively plan the decision criteria for selecting the optimal design option. Engaging stakeholders cooperatively in identifying decision criteria prevents conflicting or redundant requirements, which may emerge in other possible methods such as individual interviews. Further, when identifying the relevant decision criteria, it must be defined whether a criterion is positively or negatively connected to data infrastructure success. For example, high transparency in a data infrastructure may not always be desired.

As not all criteria might be equally relevant to a specific data infrastructure endeavor, they must be prioritized to enable subsequent decision making. Usually, methods from the field of multi-criteria decision-making are employed for this purpose. Examples of such methods include the Analytical Hierarchy Process (AHP), the Hundred-Point Test,

the Weighted Sum Model, or simpler approaches like High–Medium–Low evaluations. Figure B presents an exemplary result of the requirements definition phase.

Data infrastructure requirement					U
Category	Designation	Definition	Connotation (positive/negative)	Relevance points	Proportionate relevance points
Business	Sustainability	The sustainability of a data infrastructure refers to the ability to use it in the long term without compromising economic or social aspects.	positive	1	0,083333333
Business	Cost amortization	The costs of a data infrastructure refer to the total financial expenditure required to build, set up, operate and maintain the data infrastructure. Amortization refers to when the costs are refinanced through profits and depreciation.	positive	3	0,25
Organizational	Trustworthiness	Trust refers to the willingness of one party to expose itself to the actions of another party, regardless of the ability to monitor or control that other party. In the context of data sharing, trust concerns the actions of the data sharing partners, the provisioning of data and the behaviour of the data infrastructure itself. It is governed by the technological design choices such as the implementation of zero-trust mechanisms.	positive	2	0,166666667
Organizational	Data sovereignty	Data Sovereignty describes the ability of a data provider to be self-determined about the usage of its data assets (including data and metadata) along their whole lifecycle, from data creation to data deletion and including actions conducted by third parties. Data Sovereignty requirements are usually formulated in contracts or policies.	positive	2	0,166666667
Technical	Portability	Portability is the ability to easily move data or software services from one system to another.	positive	3	0,25
Technical	Scalability	Scalability of a data infrastructure refers to the ability to support potentially increasing sizes of data, numbers of transactions, or numbers of supported participants as while ensuring constant performance.	positive	1	0,083333333

Figure B. Exemplary result of the requirements definition phase

Preliminary Design

Goal: Selection of the most suitable design option.
Inputs: Catalog of weighted decision criteria.
Outputs: Design options and their total evaluation points.

In the preliminary design phase, the most suitable data infrastructure design option is identified. This involves two dedicated steps. First, the suitable architecture design options are derived based on the given decision criteria and stakeholder concerns. For example, data infrastructure design options that are not able to fulfill legal requirements are excluded at this point. Such a preselection reduces decision-making complexity as it limits the number of assessments to be made. Second, the selected design options are compared against the predefined criteria. Using the same (multi-criteria) evaluation system as in requirements prioritization reduces comprehension issues and stakeholder effort in the evaluation process. Finally, a ranking of design options is established based on the prioritization. Figure C depicts an exemplary final result of the preliminary design phase.

Data infrastructure requirement				Utility value analysis					
Category	Designation	Definition	Connotation (positive/negative)	Relevance points	Proportionate relevance points	Rating centralized	Proportionate rating centralized	Rating federated	Proportionate rating federated
Business	Sustainability	The sustainability of a data infrastructure refers to the ability to use it in the long term without compromising economic or social aspects.	positive	1	0,083333333	2	0,666666667	1	0,333333333
Business	Cost amortization	The costs of a data infrastructure refer to the total financial expenditure required to build, set up, operate and maintain the data infrastructure. Amortization refers to when the costs are refinanced through profits and depreciation.	positive	3	0,25	2	0,666666667	1	0,333333333
Organizational	Trustworthiness	Trust refers to the willingness of one party to expose itself to the actions of another party, regardless of the ability to monitor or control that other party. In the context of data sharing, trust concerns the actions of the data sharing partners, the provisioning of data and the behaviour of the data infrastructure itself. It is governed by the technological design choices such as the implementation of zero-trust mechanisms.	positive	2	0,166666667	1	0,25	3	0,75
Organizational	Data sovereignty	Data Sovereignty describes the ability of a data provider to be self-determined about the usage of its data assets (including data and metadata) along their whole lifecycle, from data creation to data deletion and including actions conducted by third parties. Data Sovereignty requirements are usually formulated in contracts or policies.	positive	2	0,166666667	1	0,5	1	0,5
Technical	Portability	Portability is the ability to easily move data or software services from one system to another.	positive	3	0,25	3	0,75	1	0,25
Technical	Scalability	Scalability of a data infrastructure refers to the ability to support potentially increasing sizes of data, numbers of transactions, or numbers of supported participants as while ensuring constant performance.	positive	1	0,083333333	2	0,666666667	1	0,333333333
						Rating centralize	0,590277778	Rating federated	0,409722222

Figure C. Exemplary result of the preliminary design phase

Application of the Catalog of Criteria

During this procedure, relying on the catalog of criteria can have multiple advantages, especially in the requirements definition and preliminary design phases. First, it helps to identify potentially relevant criteria and supports practitioners in ensuring core criteria are not neglected in the identification process. Second, the provided definitions reduce time and consolidation efforts for involved stakeholders needed to precisely describe the intention of a specific criterion. Lastly, it can potentially foster multi-criteria decision-making by providing a structure and hierarchy of decision criteria that can be incorporated into processes such as AHP.