An Empirical Examination of the Technical Aspects of Data Sovereignty

Julia Pampus[1][a] and Maritta Heisel[2][b]

[1] *Institute for Software and Systems Engineering ISST, Dortmund, Germany*

[2] *Paluno - The Ruhr Institute for Software Technology, University of Duisburg-Essen, Duisburg, Germany*

Keywords: Data Sharing, Data Sovereignty, Requirements Engineering, Empirical Study, Goal Modeling.

Abstract: Self-determination and autonomy in data sharing, in recent research also referred to as data sovereignty, arouses increasing interest in the context of industrial ecosystems. Its practical implementation considers organisational, regulatory, legal, and particularly technical aspects. Previous work has not yet focused on the structured analysis of technical characteristics of systems used in data sharing concerning data sovereignty. In this paper, we therefore elicit what system requirements help the data sovereignty of a data sharing participant, starting from privacy protection goals, FAIR principles, and ISO/IEC 25010:2011. To address this, we conducted a qualitative study in the form of an online questionnaire. We asked 18 domain experts to evaluate selected system criteria for their impact and relevance to the implementation of data sovereignty. Our work has resulted in a set of 22 functional requirements that can be used for designing data sharing systems. Subsequently, we discuss our findings, compare them with related work, and address further research.

1 INTRODUCTION

Nowadays, digital innovation and transformation benefit from data-driven value chains that include data usage across company boundaries (Brauner et al., 2022). A fundamental concept in this context is the principle of *data sovereignty*. Data sovereignty "refers to the self-determination [and autonomy] of individuals and organisations with regard to the use of their data" (Jarke et al., 2019, p.550). Recent work focuses on general aspects of data sovereignty and its conceptualisation within the scope of information systems. It shows that the data infrastructure is critical for creating a trustworthy environment for data sharing (von Scherenberg et al., 2024). Here, in addition to organisational, regulatory and legal requirements, especially technical requirements need to be met by the data sharing participants.

Yet, there is an existing research gap in the detailed examination of these technical requirements (Hellmeier et al., 2023). To address this, we pose the following research question (RQ): *What requirements does a system have to fulfil to ensure the data sovereignty of a data sharing participant?*

In this context, a system refers to a software application or a group of applications potentially deployed on multiple infrastructures. To answer the RQ, we conduct a qualitative empirical study and analyse the impact of common system characteristics on the assurance of data sovereignty. Analysing the study results, we consider data sovereignty as a non-functional requirement (NFR). The identified functional requirements (FRs) are presented as goal models using the i* modelling notation.

The remainder of this paper is structured as follows: Section 2 introduces the foundations of our work. Section 3 presents our research method, including preparatory tasks, study design, and data collection and analysis. We then outline our study results in Section 4, compare these to related work in Section 5, and discuss them in Section 6. Finally, Section 7 summarises our findings and the relevance of our work.

2 FUNDAMENTALS

This section introduces the fundamentals for understanding the following work.

2.1 Privacy Protection Goals

The protection goals for privacy engineering describe generally applicable criteria for "the legal, technical, economic, and societal dimensions of privacy

[a] https://orcid.org/0000-0003-2309-6183

[b] https://orcid.org/0000-0002-3275-2819

Pampus, J. and Heisel, M.
An Empirical Examination of the Technical Aspects of Data Sovereignty.
DOI: 10.5220/0012760600003753

In *Proceedings of the 19th International Conference on Software Technologies (ICSOFT 2024)*, pages 112-122
ISBN: 978-989-758-706-1; ISSN: 2184-2833

and data protection in complex IT systems” (Hansen et al., 2015, p.159). They are composed of three security protection goals (confidentiality, integrity, availability), known as the CIA triad, and three further aspects that specifically address the issues of privacy and data protection (unlinkability, transparency, intervenability). *Unlinkability* describes the property that privacy-relevant data cannot be linked to other privacy-relevant data beyond the context of use, e.g., to conclude persons. *Transparency* means understanding all data movements, e.g., during processing, at any time to reconstruct them. *Intervenability* is defined as the ability to observe and actively interrupt or modify data processing (Hansen et al., 2015).

2.2 FAIR Principles

The FAIR principles define guidelines for the management of data and associated metadata for re-use by third parties. First, (meta-) data should be *findable* for users and systems. Next, the discovered data must be *accessible*, including appropriate authentication and authorisation. To use data for analyses, it must be *interoperable* with other data or interfaces of systems. Last, (meta-) data should be properly prepared so that it can be *reusable* (Wilkinson et al., 2016).

2.3 Software Quality Characteristics

The ISO/IEC 25010:2011 (International Organization for Standardization, 2011) provides software quality characteristics and describes an evaluation process model. The model specifies eight quality properties with some sub-characteristics each: *Functional suitability* means that a system works as expected and required in a specified context. The system could do that with a certain *performance efficiency* and *reliability*, and being *compatible* with other systems in the same hardware or software environment. Next, *usability* describes how effective, efficient, and satisfying a user can interact with the system. The system can be *secure* concerning confidentiality, integrity, non-repudiation, accountability, and authenticity. In addition, its degree of *maintainability*, i.e., the ability to maintain functionality and the *portability* to another hardware, software, or operational environment, can be determined.

2.4 I* Modelling Notation

The i* modelling notation is a goal- and actor-oriented framework for modelling requirements. It focuses on actors, their intentions, and the strategies to achieve goals (Dalpiaz et al., 2016). The language consists of several entity types: actors, actor associations, intentional elements, intentional element links, and social dependencies. Figure 1 visualises the elements used in our work: Each goal model has at least one actor and an associated actor boundary, shown as a grey circle in the background. There are different actor types; we use Roles in the following. A *Role* is an actor with an abstract characterisation within a domain, in our case, data sharing.

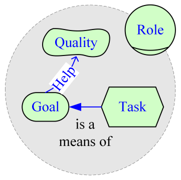


Figure 1: Overview of i* Elements Used in This Work.

A goal model following the i* notation consists of intentional elements. We use Goals, Qualities, and Tasks. A *Goal* describes a state that an actor wants to achieve. A *Quality* represents the desire of an actor. In many applications of the notation, this is an NFR. *Tasks* describe actions that an actor performs to achieve a *Goal*. Links connect intentional elements. In our work, we only use links between Goals and Qualities, referred to as Contribution, and links between Tasks and Goals, referred to as Refinement. There are four types of *Contributions*: make, help, hurt, and break. *Make* means that an element alone can ensure the fulfilment of a Quality; *break* can prevent the fulfilment. *Help* and *hurt* denote general negative and positive influences. If a *Refinement* has a Goal as its parent, there can be an AND or OR relationship. The arrows used in Figure 1 imply OR relationships and allow the fulfilment of a parent with “the fulfilment of at least one child” (Dalpiaz et al., 2016, p.10) that is a ‘means’.

3 RESEARCH METHOD

To identify technical aspects of data sovereignty, we conducted an empirical study. Figure 2 visualises its research design. Our work involved four steps: First, we designed the questionnaire and selected participants based on selection criteria. Second, we conducted the study to, next, analyse the responses. Last, the results of this analysis form a selection of requirements for the implementation of data sovereignty.

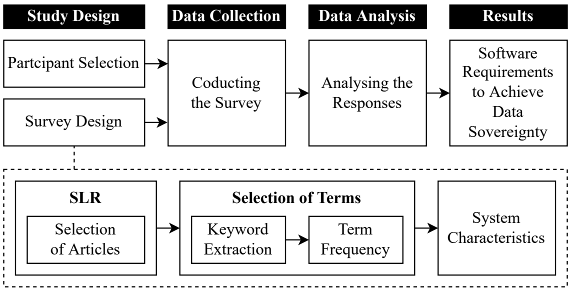


Figure 2: Research Design for the Survey.

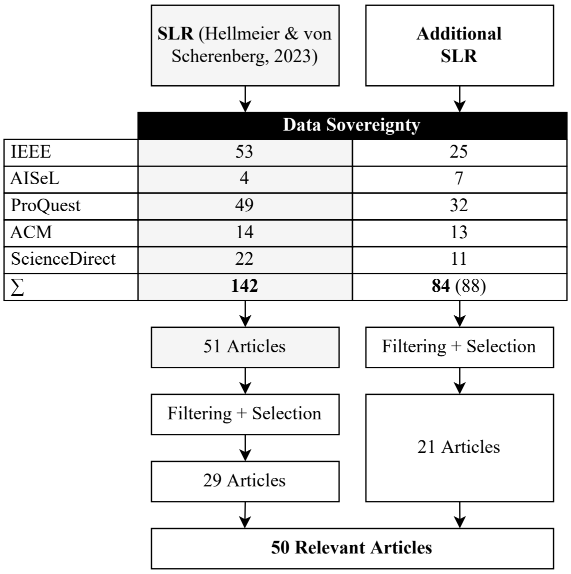
3.1 Preparatory Work

The pre-selection of system characteristics used for our questionnaire required some preparatory work. Therefore, as shown in Figure 2, the creation of the questionnaire comprised the following steps: First, we conducted a Systematic Literature Review (SLR) to find relevant articles. Next, we selected terms for an analysis of these articles. Last, we used the most frequent terms, i.e., system characteristics, as input for the questionnaire design.

3.1.1 Systematic Literature Review

We built on the work of Hellmeier and von Scherenberg (2023) as a basis for our literature review since they already provide an up-to-date literature review on data sovereignty. In their work, the authors focussed on distinguishing data sovereignty from digital and technical sovereignty. For this, they conducted an SLR and identified publications that "give a concrete definition, discussion, implementation, or explanation" (Hellmeier and von Scherenberg, 2023, p.5) of the examined terms. In sum, 142 articles form their final result set, of which 51 deal with data sovereignty. Figure 3 visualises the described process (shaded grey). As we wanted to analyse the selected articles automatically afterwards, we filtered the 51 articles according to whether they deal with data sovereignty in-depth and technically (inclusion criteria) or only provide a brief insight or definition (exclusion criteria). That resulted in a set of 29 articles of the original 51. To add more recent ones to the dataset, covering April 2022 to October 2023, we conducted a Multivocal Literature Review, a form of an SLR that includes grey literature (Garousi et al., 2019). Following Hellmeier and von Scherenberg (2023), we considered scientific and industrial articles to examine the chosen subject from a practical point of view. We used the following abstracted search string:

(Title: *X* OR Keywords: *X* OR Abstract: *X*),
X = "data sovereignty"



	Data Sovereignty	
IEEE	53	25
AISeL	4	7
ProQuest	49	32
ACM	14	13
ScienceDirect	22	11
∑	**142**	**84** (88)

Figure 3: SLR Process for 'Data Sovereignty'.

Figure 3 depicts the overall process. We searched a total of five databases, listed there. The additional search resulted in another 84 articles (without duplicates), of which 21 remained after filtering using the previously mentioned exclusion/inclusion criteria. Overall, the result set of the combined SLR comprised 50 relevant articles for the subsequent data analysis.

3.1.2 Selection of Terms

As shown in Figure 2, the SLR was followed by a selection of terms to analyse the found articles. For the selection, we used existing system characteristics, covering terminology from the specifications presented in Section 2. As data sovereignty relates to privacy and data protection, we added each privacy protection goal to the set of terms. Next, we considered the FAIR principles as relevant, whereby 'accessibility' is subordinate to 'availability', as accessibility is part of the availability of data or systems (Hansen et al., 2015). Complementary, we adopted terms from ISO/IEC 25010:2011 (International Organization for Standardization, 2011). From the eight characteristics, we omitted 'functional suitability' and 'compatibility' being too unspecific; 'interoperability' and 'reliability' are included in the FAIR principles. Here, reliability is a part of integrity (Hansen et al., 2015). We added the remaining five terms (performance, usability, maintainability, portability, security) to the list of relevant terms. We considered their child terms as synonyms (if not already contained in the list, e.g., 'confidentiality' or 'integrity'), not in a grammatical sense but in their meaning.

For the selection of relevant system characteristics of our questionnaire, first, we scanned the available articles from our SLR for the frequency of the above-selected terms. We assumed that the relevance of terms increases with rising frequency in a text. We used a Python script to analyse the articles by word stems. Second, we conducted a reverse lookup to prevent missing relevant terms during the pre-selection in the previous processing. Accordingly, we searched the articles for frequent terms using NLTK[1] and KeyBERT (Grootendorst, 2020).

The overall result set included 15 system characteristics: intervenability, transparency, confidentiality, integrity, availability, (data) findability, interoperability, reusability, performance (efficiency), usability, security, maintainability, portability, trustworthiness, and automation. The statistical evaluation of the matches had already indicated one possible direction of our survey: Security was mentioned very often in the analysed publications, while the searches for reusability or portability did not result in any significant matches. A simple examination of the terms and their meaning made us believe that the following characteristics had no significant influence on data sovereignty: maintainability, reusability, findability, portability, and automation. However, we included them in the survey to minimise subjectivity in its design.

3.2 Survey Design

Online questionnaires offer a suitable way of interviewing people in a targeted manner and without much effort. There is no need for an interviewer, and the participant can deal with the discussed subject independently of time and place. That can avoid undesirable methodological effects, facilitate completion, and thus increase data quality (Krosnick, 2018).

3.2.1 Questionnaire Design

We tested our questionnaire in a trial run to check the formulations of the questions, how long it takes to complete the questionnaire, and whether everything works as expected from a technical point of view. As a result, we had to make a few adjustments, including the wording of the questions, which are already incorporated below.

We divided the questionnaire into three parts and 55 questions, summarised in Table 1. At the beginning, some opening words described the further course of the survey and presented an established definition of the term ‘data sovereignty’. That should help to create the same knowledge base as a prerequisite for all participants and avoid any subsequent ambiguities. In addition, we recorded the participant’s name for possible follow-up contacts during the evaluation phase (cf. Q01).

The first part of the questionnaire consisted of closed questions that allowed “respondents to select an answer from a set of choices” (Krosnick, 2018, p.266). It asked the participants to individually assess the presented system characteristics and their impact and relation to the implementation of data sovereignty (cf. Q02-Q46). Here again, we provided some definitions. To answer the guiding questions, we used pairs of 7-point Likert scales with the options ‘strongly negative’, ‘slightly negative’, ‘negative’, ‘neutral’, ‘positive’, ‘slightly positive’, and ‘strongly positive’. The first scale evaluated the existing or strongly positive characteristic (as appropriate); the second assessed the opposite (missing or negative characteristic). We offered an optional comment field to justify the selected answer.

The second part of the questionnaire comprised open questions to identify other characteristics of a system that we might have missed in the study design (cf. Q47-Q49). In addition, a further comment field offered the opportunity to express additional thoughts and opinions relevant to the questionnaire evaluation. The third and final part of the questionnaire asked for general information about the participants for statistical analyses, including the current job title, the current employer, and the extent of experience with the concept of data sovereignty (cf. Q50-Q55).

3.2.2 Participant Selection

The participants were a heterogeneous group of people with different job positions and companies of various sizes based in Western Europe, primarily in Germany. The main selection criterion was the knowledge of the topic of data sovereignty and, thus, the suitability to provide well-founded information. We present an analysis of their ages, job positions, and professional experiences in Section 4.1.

3.3 Data Collection

We requested the participants by personal e-mail. 18 out of 25 contacted persons agreed to take part in our study. We provided the questionnaire with the help of Microsoft Forms[2]. This tool offers the possibility of simple participation without registration and easy administration and analysis of responses.

[1] https://www.nltk.org (accessed on 2023-10-12)

[2] https://forms.office.com (accessed on 2024-02-26)

Table 1: Shortened List of Questions.

ID	Question (Q)	Input Type
Q01	What is your full name?	Text field
Q02	How does the presence of *X* affect data sovereignty?	Likert scale
Q03	How does the absence of *X* affect data sovereignty?	Likert scale
Q04	Justify your answer. *(optional)*	Text field
...	*Repetition of Q02-Q04 with each X* $\in$ *{"intervenability", "transparency", "confidentiality", "integrity", "availability", "data findability", "interoperability", "reusability", "performance efficiency", "usability", "security", "maintainability", "portability", "trustworthiness", "automation"}*	...
Q47	From your point of view, what other system characteristics have a positive impact on the implementation of data sovereignty? Why?	Text field
Q48	From your point of view, what other system characteristics have a negative impact on the implementation of data sovereignty? Why?	Text field
Q49	Is there anything else you would like to share that could be relevant for the evaluation?	Text field
Q50	How old are you?	Choice box
Q51	Who is your current employer?	Text field
Q52	What is your current job title?	Text field
Q53	What is your professional background (education/studies/profession)?	Text field
Q54	How many years of work experience do you have?	Text field
Q55	Since when are you familiar with the concept of data sovereignty?	Text field

3.4 Data Analysis

We analysed the collected data in two ways: we statistically evaluated the inputs via Likert scales and the answers to questions Q51 to Q55, and qualitatively analysed all inputs via text fields.

For the interpretation of the results, we use goal models. As described in Section 2.4, goal modelling focuses on actors and expresses their intentions and strategics. When analysing the requirements for a system to implement data sovereignty, we consider the desires of a data sharing participant as an actor who uses the system. The created goal models help to derive appropriate software requirements.

4 RESULTS

The following subsections present the study results including the descriptive findings and our interpretation and analysis of the participants' responses.

4.1 Descriptive Findings

In total, we interviewed 18 persons, aged between 25 and 44. They represented seven industrial companies and two research organisations. The distribution shows that over 50 percent of the respondents have a research and development (R&D) background. As Table 2 shows, the participants have different professions but are all technically orientated, from software development to research to mid-level management. Around 75 percent of the respondents have completed a higher academic degree, a significant amount with a specialisation in computer science or related subjects. More than half have ten or more years of professional experience, and all are familiar with data sovereignty for at least one year, most even four years or more. When presenting and discussing our results, we refer to the participants and their citations using numbers to ensure their anonymity.

Table 2: Overview of Survey Participants and Job Positions.

Job Position	Participants (P)
Development	P01, P05, P18
IT Management	P04, P08-10, P15, P17
R&D	P02-03, P06-07, P11-14, P16

The participants spent an average of 70 minutes completing the questionnaire. As expected, most selected system characteristics were rated positively in their presence and negatively in their absence. Nevertheless, there were also characteristics whose presence was rated positively and their absence was not rated negatively, and therefore, according to our interpretation, not considered a risk. Overall, we observe the following correlation: the more positive the impact of the presence of a system characteristic, the more negative its absence. The participants assessed the presence of confidentiality, integrity, security, and trustworthiness as particularly relevant for the realisation of data sovereignty and their absence as particularly risky. Also, the absence of interoperability, transparency, availability, usability, and maintainability was rated negatively, although not as much. The assessment of intervenability was controversial. The Likert scale ranges to either side for presence and absence. Findability, performance efficiency, portability, and automation were considered 'nice-to-have', i.e., positive in their presence but not critical in their absence. The participants did not consider reusability as relevant. The evaluation largely coincides with our assumptions from Section 3.1.2.

All additionally collected characteristics (cf. Q47 and Q48) can be allocated to the previously selected terms. For example, controllability, observability, and modifiability belong to intervenability and transparency. Most participants intensively used the comment fields to justify and discuss their responses. Thus, we derived the following requirements from evaluating the Likert scales and from a qualitative analysis of the comments.

4.2 Definition of Requirements

Figure 4 depicts a goal model focusing on the strategic rationales for achieving data sovereignty. We see 'data sovereignty' and 'trustworthiness' as soft goals as both are NFRs whose achievement is not clearly defined and measurable. We consider all previously surveyed system characteristics as measurable NFRs, i.e., as goals. The model illustrates that some goals have a direct influence on data sovereignty (intervenability, security, interoperability), while others have an indirect one by supporting other goals (integrity, confidentiality, usability, transparency, findability). No goal has a *make* relation as no requirement can achieve the soft goal independently.

Figure 4 highlights reusability, maintainability, performance, portability, availability, and automation white as we are not further considering them in the derivation of FRs. *Re-usability* is a requirement that cannot be externally assessed during an initial requirements engineering process. *Availability* and its sub-tasks are affecting elements on intervenability with little relevance for the establishment of data sovereignty. *Automation* can have a direct influence on data sovereignty, either positive or negative, however, it is more an extension of other FRs and does not stand for itself.

The following subsections analyse each goal by defining strategics and tasks as goal models. We derive one FR for the system under consideration from each task. The set of FRs is listed in Table 3. The formulation of the FRs uses MoSCoW prioritisation (must, should, could, would). For simplicity, in the following, we refer to the data sharing participant as such and only concretise its role if necessary. For concretisation, we will only use the terms data provider and consumer. A *data provider* includes the data rights holder, data providing agents, and third parties like data intermediaries. A *data consumer* means data consuming agent, also known as a data recipient or user.

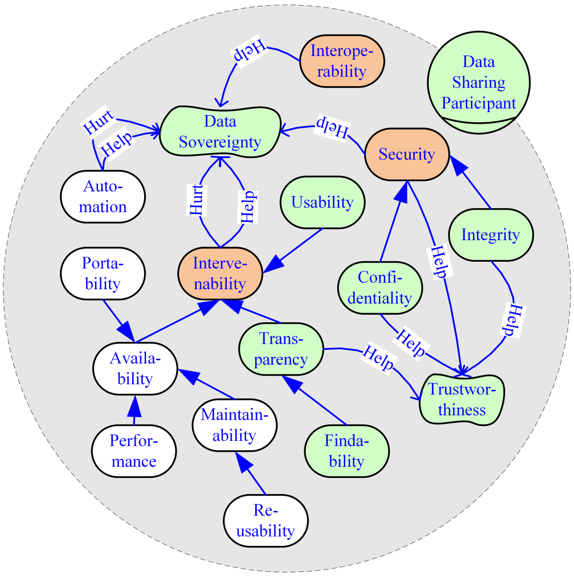


Figure 4: Goal Model for Data Sovereignty.

4.2.1 Intervenability

Intervenability is the "degree to which a system, product or component prevents unauthorized access to, or modification of, computer programs or data" (International Organization for Standardization, 2011). Figure 5 depicts a goal model that illustrates the relations between intervenability, its means, and data sovereignty.

Intervenability *Helps* or *Hurts* Data Sovereignty (Cf. FR01-FR07). Intervenability is a "core construct" (P08) for the self-determination of data sharing participants. Both data providers and consumers should be able to modify data flows anytime. Yet, there is some "potential for error and failure" (P09) and a risk of misuse of provided intervenability mechanisms (P05; P08; P09). The data consumer must not be able to bypass the data usage conditions defined by the data provider (P05). Therefore, mechanisms for intervenability should not be applied by default or without consent by involved data sharing participants (P10). If applicable, the data sharing participants should be able to negotiate the data usage conditions.

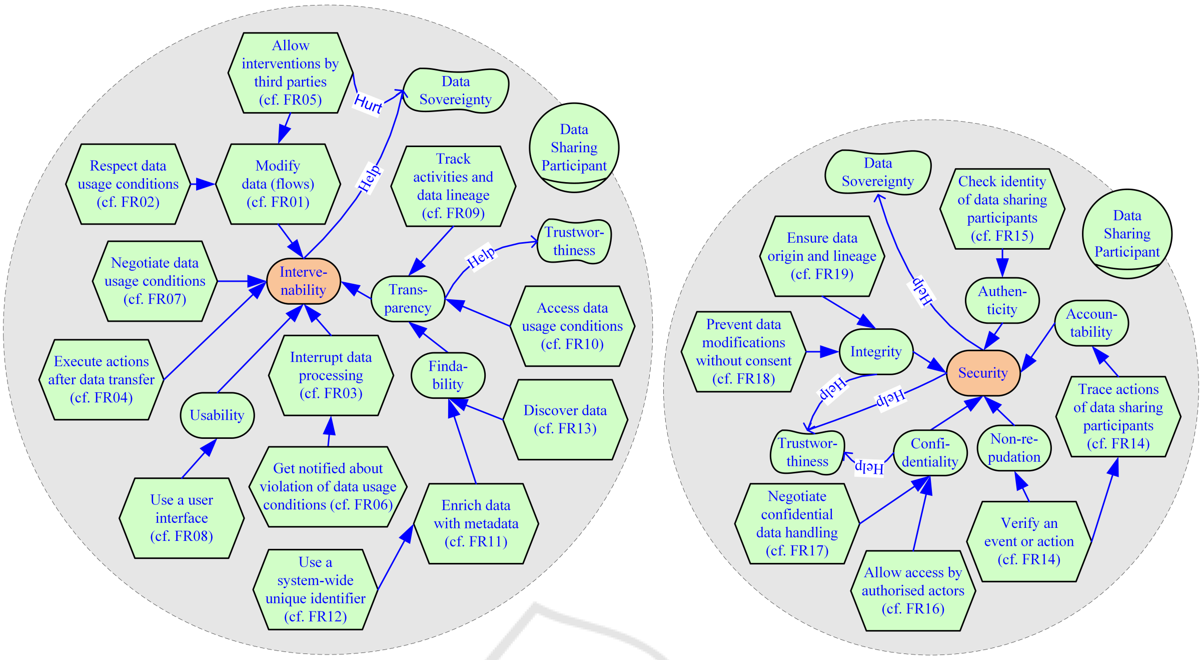

Allow interventions by third parties (cf. FR05)
Hurt
Data Sovereignty
Data Sharing Participant
Respect data usage conditions (cf. FR02)
Modify data (flows) (cf. FR01)
Help
Track activities and data lineage (cf. FR09)
Trustworthiness
Help
Negotiate data usage conditions (cf. FR07)
Intervenability
Transparency
Access data usage conditions (cf. FR10)
Execute actions after data transfer (cf. FR04)
Interrupt data processing (cf. FR03)
Findability
Usability
Discover data (cf. FR13)
Use a user interface (cf. FR08)
Get notified about violation of data usage conditions (cf. FR06)
Enrich data with metadata (cf. FR11)
Use a system-wide unique identifier (cf. FR12)
Data Sovereignty
Check identity of data sharing participants (cf. FR15)
Data Sharing Participant
Ensure data origin and lineage (cf. FR19)
Help
Authenticity
Accountability
Prevent data modifications without consent (cf. FR18)
Integrity
Security
Trace actions of data sharing participants (cf. FR14)
Help
Help
Trustworthiness
Confidentiality
Help
Non-repudation
Negotiate confidential data handling (cf. FR17)
Verify an event or action (cf. FR14)
Allow access by authorised actors (cf. FR16)

Figure 5: Goal Models for Helping Data Sovereignty with Intervenability (left) and Security (right).

Usability *Is a Means of* intervenability (Cf. FR08). The better the usability of the applied systems, e.g., by providing a graphical user interface with a good user experience, the easier it is to access essential utilities such as the definition of data usage conditions (P06; P07; P13). Missing or reduced usability of interfaces, e.g., due to complexity, also on a technical level, "might hinder market adoption" (P10) or increase the risk of incorrect use (P18).

Transparency *Is a Means of* Intervenability and *Helps* Trustworthiness (Cf. FR09-FR10). Transparency is the primary means of proving that data usage conditions are respected (P07; P12; P17), thus also strengthening the trust of data sharing participants (P01; P05). That may be a must from a legal perspective (P13); however, it is up to the data sharing participants and their requirements to decide where transparency is required (P09). During implementation, it is crucial to restrict access to logging or audit trails, as transparency can otherwise quickly result in abuse of confidential information (P10; P18).

Findability *Is a Means of* Transparency (Cf. FR11-FR13). Data sovereignty and transparency require data to be traceable (P12) and thus findable. Most importantly, allocating data helps to establish references to existing data usage conditions (P05; P06). In this context, it is essential to distinguish between internal and external findability. Internal findability is achieved, e.g., by unique and persisted identifiers, whereas external findability, in the sense of discoverability of data offerings, must remain under the control of the data provider (P09; P10).

4.2.2 Security

Security is the "degree to which a product or system protects information and data so that persons or other products or systems have the degree of data access appropriate to their types and levels of authorization" (International Organization for Standardization, 2011). Figure 5 depicts a goal model that illustrates the relations between security, its means, and data sovereignty.

Security *Helps* Data Sovereignty and Trustworthiness (Cf. FR14-FR15). Security helps the trustworthiness of a system and, with this, affects the data sovereignty of a data sharing participant (P05; P09). However, only selected requirements can have a direct impact. For instance, a system should be able to verify the identity of a data sharing participant or guarantee the indisputability of actions. In addition to confidentiality and integrity, all other security requirements depend on the type of data, the respective use case, and the data usage conditions (P02; P17; P18). Accountability, e.g., is often a legal aspect (P18).

Confidentiality *Is a Means of* Security and *Helps* Trustworthiness (Cf. FR16-FR17). Confidential-

Table 3: List of FRs.

ID	**Functional Requirements** (The system . . .)
FR01	. . . SHOULD enable data sharing participants to modify data (flows).
FR02	. . . MUST ensure that data (flow) modifications are consistent with the data usage conditions.
FR03	. . . SHOULD enable a data provider to interrupt data processing activities on the data consumer side.
FR04	. . . COULD enable a data provider to execute operations on shared data on the data consumer side.
FR05	. . . MUST prevent interventions by third parties without the consent of the data sharing participants.
FR06	. . . SHOULD provide notifications if data usage does not comply with the data usage conditions.
FR07	. . . COULD enable data sharing participants to negotiate data usage conditions.
FR08	. . . COULD provide a graphical user interface to lower the barriers for non-experts.
FR09	. . . SHOULD provide features to keep track of data processing activities and data lineage at any time.
FR10	. . . MUST ensure that data usage conditions are accessible to all data sharing participants.
FR11	. . . SHOULD provide features to enrich any data with metadata, at least the data usage conditions.
FR12	. . . MUST assign a system-wide persisted unique identifier to each data set.
FR13	. . . COULD provide features that enhance the discoverability of data.
FR14	. . . MUST provide mechanisms to ensure the indisputability of occurring events and actions.
FR15	. . . MUST incorporate mechanisms to authenticate and verify the identity of data sharing participants.
FR16	. . . MUST ensure access to data and metadata only by authorised actors, i.e., systems and users.
FR17	. . . MUST enforce the confidential handling of data following the agreed upon data usage conditions.
FR18	. . . MUST prohibit changes to data by a data sharing participant without the data provider's consent.
FR19	. . . MUST not remove any reference to the data origin without explicit consent of the data provider.
FR20	. . . COULD implement common data formats to facilitate data transfers.
FR21	. . . MUST implement common vocabularies for data usage conditions.
FR22	. . . SHOULD implement common protocols for data sharing.

ity is crucial for building trust (P01) and positively impacts data sovereignty, as it prevents the general misuse of data by third parties. Accordingly, data sharing participants are always well-advised to act according to the "Need-to-Know" (P08) principle. Nevertheless, confidentiality only needs to be ensured to the extent required by the data sovereign (P04; P12; P14). For example, when it comes to open data, confidential handling would not be part of the data usage conditions and thus irrelevant.

Integrity *Is a Means of* Security and *Helps* Trustworthiness (Cf. FR18-FR19). Integrity is critical when implementing security and trust (P05), as "no secure and/or trustworthy environment can be created on a compromised system" (P06). In this context, it is particularly important that data modifications and the removal of the data origin must not be executed without prior consent.

4.2.3 Interoperability

Interoperability is the "degree to which two or more systems, products or components can exchange information and use the information that has been exchanged" (International Organization for Standardization, 2011). Figure 6 depicts a goal model that illustrates the relations between interoperability, its means, and data sovereignty.

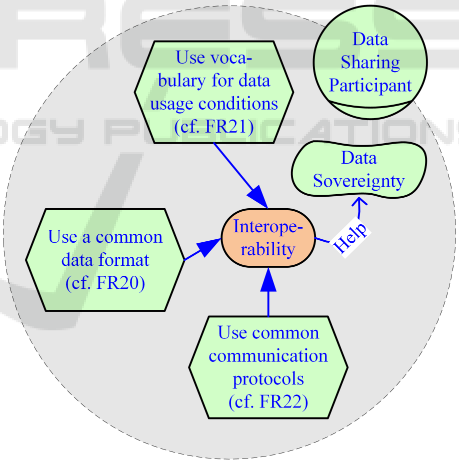

Use vocabulary for data usage conditions (cf. FR21)
Data Sharing Participant
Data Sovereignty
Use a common data format (cf. FR20)
Interoperability
Help
Use common communication protocols (cf. FR22)

Figure 6: Goal Model for Helping Data Sovereignty with Interoperability.

Interoperability *Helps* Data Sovereignty (Cf. FR20-FR22). Interoperability "enables the capability to conduct [. . .] data sovereignty" (P04). It is a fundamental condition for data sharing. Concerning data sovereignty, interoperability and thus the use of mutual communication protocols, including defined processes and a common vocabulary, ensures a standardised understanding among all data sharing participants (P02; P15). That allows for an equal harmonisation, interpretation, and implementation of data usage conditions (P05; P06). At the same time, interoperability can increase the system's scalability, efficiency, and automation (P15).

5 RELATED WORK

Recent research papers have already dealt with the exploration of requirements for data sovereignty in industrial data sharing, also using empirical methods. For example, Biehs and Stilling (2024) and Hellmeier et al. (2023) conducted interview studies to identify requirements for data sharing and, in this context, primarily considered the implementation of data sovereignty. Nevertheless, most works have one thing in common: they focus on specific aspects determined by the research domain and influenced by certain use cases. For instance, Opriel et al. (2021) concentrate on the exchange of sensitive data. The work of Larrinaga (2022) considers data sovereignty from a manufacturing perspective and equates it with usage control. We addressed this by asking people from various industrial domains, from energy, mobility and logistics, manufacturing and automotive to healthcare. By focusing on common system characteristics and deriving system features, not the content of data usage conditions, we demonstrate that establishing data sovereignty is not only about implementing access and usage control.

Furthermore, a lot of works mix different perspectives (economic, regulatory, legal, and technical) or provide varying levels of requirements that they do not conclusively detail, such as the combination of access control, GDPR (European Parliament and Council of the European Union, 2016), data quality, and monetisation (Biehs and Stilling, 2024). Alternatively, they elaborate on requirements that have nothing to do with data sovereignty but focus on data sharing, such as "short loading times" (Zrenner et al., 2019, p.484) or data portability (Falcão et al., 2023).

As one of the first, Hellmeier et al. (2023) follow a holistic approach and separate the dimensions of data sovereignty. However, the work also notes that the elicited requirements from their interviewees are very vague. Most of their answers reflect that people see data sovereignty as equivalent to usage control or formulate more general requirements for data sharing. Our study also confirmed this effect. It indicates a lack of understanding or detailed analysis of what is at the core of data sovereignty. Hence, domain experts require tools to help them better understand their requirements.

Ultimately, the existing studies primarily focus on the requirements from the user's perspective. Our study extended these works by concluding implications for the system and corresponding system requirements. Most importantly, we built on well-established concepts like privacy and security for considering technical aspects of data sovereignty. In addition, we use previous approaches from related fields to represent the requirements as part of goal models: For example, Elahi and Yu (2007) have created a goal-oriented approach for analysing security trade-offs; Peixoto and Silva (2018) have used the i* goal modelling notation to model privacy requirements; and Borchert and Heisel (2021) have elaborated on how to resolve trust conflicts using goal models.

Summarising, the presented findings complement previous work with a closer examination of the technical aspects of data sovereignty and form a solid basis for structured requirements engineering for self-determined and autonomous data sharing.

6 DISCUSSION

Initially, we raised the question of which system characteristics have a particular influence on the implementation of data sovereignty of a data sharing participant. The presented study demonstrates that a clear answer to this question is highly controversial for a reason. Referring to Section 5, while some work states that security guarantees data sovereignty, others argue that data sovereignty is addressed by implementing access and usage control. Our assumption that the existing literature is missing a detailed consideration of data sovereignty from a technical perspective (cf. Section 1) is also reflected in the responses of some of the study participants. While all system properties were seen as essential for data sovereignty, a closer examination reveals that the non-fulfilment of most of them was not considered particularly negative. That leads to the question of how significant such characteristics can be for establishing data sovereignty.

Also, the interviewees often assumed that the data provider was the 'data sovereign'. However, in the concept of self-determination and autonomy in data sharing, the consumer has the same rights as the data provider. We have considered this in our interpretation of the survey results and generalised the derived requirements in a way that respects the data provider and the data consumer.

As the core result of our study, it is essential to emphasise that many properties and functionalities of a system support data sovereignty but only lead to a successful implementation in their entirety. As shown in Figure 4, there is no single *make* relation targeting data sovereignty. Consequently, no requirement can ensure the fulfilment of the data sovereignty requirement in isolation. Instead, it is the combination of requirements that provides the technical foundation. The proportions of the three emphasised characteristics (intervenability, security, and interoperability) in implementing a system for sovereign data sharing are determined by the specific use case, e.g., the type of data or its processing (P11). We define half of the 22 derived requirements as a must. Whether the other requirements, or even more, should be fulfilled needs to be defined in the context, e.g., by an authority in a data ecosystem or the data sharing participants themselves. In addition, trust plays a central role: many defined NFRs strengthen the trustworthiness of a system and thus the data sovereignty of the data sharing participants.

We used established system characteristics, including ISO standards, to support our findings. With this, we confirmed that FRs should be embedded in existing concepts closely related to security, privacy, and trust. Besides the already existing NFRs, we worked out system requirements that focus specifically on data sovereignty and, with this, also data sharing, such as FR02, FR07, and more. A particular contribution of our work is the creation of interconnections focussed on the needs of participants in industrial data sharing.

6.1 Further Considerations

With the help of questions Q47 to Q49, we have already involved the study participants in discussing the requirements for implementing data sovereignty. As expected, supported by other studies (Hellmeier et al., 2023; Biehs and Stilling, 2024), not only technical aspects are relevant, but particularly regulatory and legal aspects (P15), which immediately increase the complexity of requirements elicitation (P16). In addition, activities such as standardisation (P18), the use of open-source software (P11), and the establishment of certification processes (P11; P12) can reduce the hurdles to interoperability and trustworthiness and, therefore, the establishment of data sovereignty. After all, not only the system must be trustworthy, but also the actors involved in data sharing (P15).

From a technical perspective, it was suggested that the system architecture, especially decentralised systems, could support the implementation of data sovereignty (P09). It was argued that the involvement of a centralised service necessarily leads to a loss of sovereignty (P09). As a result, data sovereignty often would not be implemented by a single system but by many systems (P08) that require an equal fulfilment of defined requirements, which “may pose additional challenges” (P05).

Ultimately, in terms of implementation, there is always a cost-benefit trade-off (P09). In addition, the market situation (P14) that, e.g., forces a data provider to share their data less restrictively can have a considerable influence on the actual autonomy of the data provider, as well as recent threats such as loss or theft of digital identities (P10).

6.2 Limitations

Limitations to our work are mainly the selection of study participants. First, a qualitative survey is limited to a small number of participants. Next, half of our participants are employees in research. That is because the topic of data sovereignty is currently in the process of being transferred from research to the industry. In addition, there is a possible professional bias due to the topic of dataspaces or data ecosystems. For instance, experts for data privacy will always relate data sovereignty to the guiding principles they are familiar with, while experts for system security will primarily emphasise aspects such as confidentiality, integrity, and availability. Last, all 18 participants are located in Western Europe and are therefore biased by the respective research and industry.

Moreover, our pre-selection of system characteristics might have limited the results. Although the questions for more characteristics did not result in additional criteria, this could change with increasing the number of interviewees. Additionally, the test interview has already shown that the order of the questions may affect the answers. The sequence of presence and absence of the characteristics may lead to the natural behaviour of making an opposite assessment.

6.3 Future Work

As a follow-up of the present study, the presented models and requirements (cf. Section 4) should be evaluated by the study participants as well as a larger group of people. In addition, our presented work allows for the further development of a structured requirements engineering process, including the analysis of the functional requirements from Table 3. In the course of this, it will be desirable to develop an approach for conflict resolution. Next, with the help of goal modelling and other appropriate methods, stakeholders can be guided through a requirements elicitation process. Furthermore, the defined requirements can be used to simplify the derivation of a system design.

7 CONCLUSION

In this work, we have focussed on the technical aspects of data sovereignty and the requirements for its implementation by a system. We evaluated the relevance of selected system characteristics with the help of an empirical study and structured the FRs and NFRs derived from this using goal models. Afterwards, we discussed our findings and compared them to related work. Overall, we have emphasised that data sovereignty is not achieved by implementing a definite list of system features but through a combination of use-case-specific functional and non-functional requirements. As one participant in the study summarised, “[m]odern systems will have [d]ata [s]overeignty by design” (P17). While building on privacy and security, our work has taken a step towards a targeted requirements analysis and reasoned system design by extending research on self-determination and autonomy in industrial data sharing with a more technically refined view.

ACKNOWLEDGEMENTS

This work was partially supported by the German Federal Ministry for Economic Affairs and Climate Action (funding number: 13IK004N). We thank Daniel Tebernum for his valuable input and all participants for contributing to our study.

REFERENCES

Biehs, S. and Stilling, J. (2024). Identification of Key Requirements for the Application of Data Sovereignty in the Context of Data Exchange. In *Proceedings of the 57th Annual Hawaii International Conference on System Sciences*. ScholarSpace.

Borchert, A. and Heisel, M. (2021). Conflict Identification and Resolution for Trust-Related Requirements Elicitation A Goal Modeling Approach. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.*, 12(1):111–131.

Brauner, P., Dalibor, M., Jarke, M., Kunze, I., Koren, I., Lakemeyer, G., Liebenberg, M., Michael, J., Pennekamp, J., Quix, C., Rumpe, B., van der Aalst, W., Wehrle, K., Wortmann, A., and Ziefle, M. (2022). A Computer Science Perspective on Digital Transformation in Production. *ACM Trans. Internet Things*, 3(2).

Dalpiaz, F., Franch, X., and Horkoff, J. (2016). iStar 2.0 Language Guide. *CoRR*.

Elahi, G. and Yu, E. (2007). A Goal Oriented Approach for Modeling and Analyzing Security Trade-Offs. In *Proceedings of the 26th International Conference on Conceptual Modeling*, pages 375–390. Springer.

European Parliament and Council of the European Union (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council.

Falcão, R., Matar, R., Rauch, B., Elberzhager, F., and Koch, M. (2023). A Reference Architecture for Enabling Interoperability and Data Sovereignty in the Agricultural Data Space. *Information*, 14(3):197.

Garousi, V., Felderer, M., and Mäntylä, M. V. (2019). Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology*, 106:101–121.

Grootendorst, M. (2020). KeyBERT: Minimal keyword extraction with BERT.

Hansen, M., Jensen, M., and Rost, M. (2015). Protection Goals for Privacy Engineering. In *2015 IEEE Security and Privacy Workshops*. IEEE Computer Society.

Hellmeier, M., Pampus, J., Qarawlus, H., and Howar, F. (2023). Implementing Data Sovereignty: Requirements & Challenges from Practice. In *Proceedings of the 18th International Conference on Availability, Reliability and Security*. ACM.

Hellmeier, M. and von Scherenberg, F. (2023). A Delimitation of Data Sovereignty from Digital and Technological Sovereignty. In *Proceedings of the 31st European Conference on Information Systems*.

International Organization for Standardization (2011). ISO/IEC 25010:2011, Systems and software engineering, Systems and software Quality Requirements and Evaluation (SQuaRE), System and software quality models. Standard.

Jarke, M., Otto, B., and Ram, S. (2019). Data Sovereignty and Data Space Ecosystems. *Business & Information Systems Engineering*, 61(5):549–550.

Krosnick, J. A. (2018). Questionnaire Design. In *The Palgrave Handbook of Survey Research*. Springer International Publishing.

Larrinaga, F. et al. (2022). Data Sovereignty - Requirements Analysis of Manufacturing Use Cases.

Opriel, S., Möller, F., Burkhardt, U., and Otto, B. (2021). Requirements for Usage Control based Exchange of Sensitive Data in Automotive Supply Chains. In *Proceedings of the 54th Hawaii International Conference on System Sciences*.

Peixoto, M. M. and Silva, C. (2018). Specifying privacy requirements with goal-oriented modeling languages. In *Proceedings of the XXXII Brazilian Symposium on Software Engineering*. ACM.

von Scherenberg, F., Hellmeier, M., and Otto, B. (2024). Data Sovereignty in Information Systems. *Electronic Markets*, 34(1):1–11.

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):1–9.

Zrenner, J., Möller, F. O., Jung, C., Eitel, A., and Otto, B. (2019). Usage control architecture options for data sovereignty in business ecosystems. *Journal of Enterprise Information Management*, 32(3):477–495.