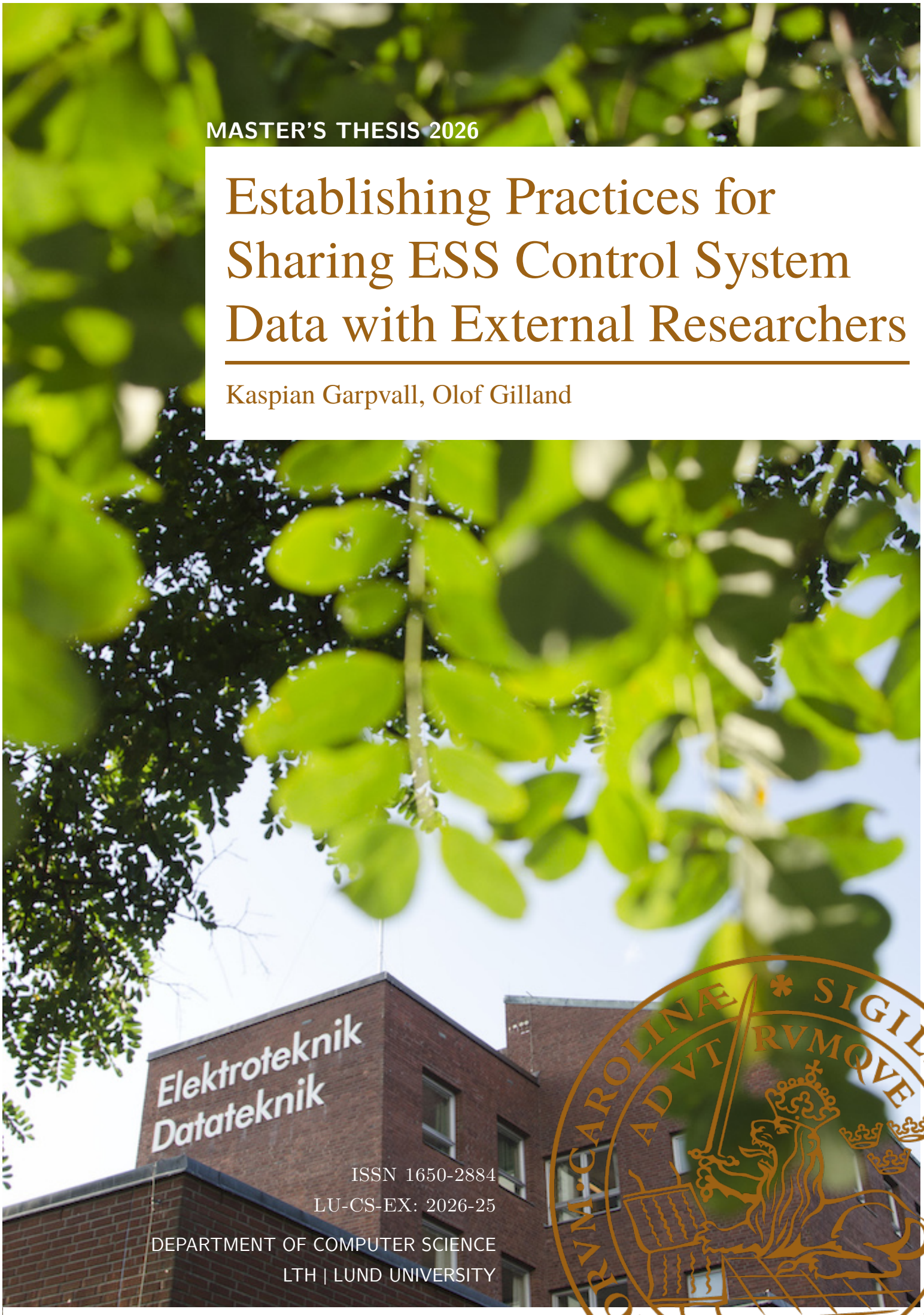


MASTER'S THESIS 2026

Establishing Practices for Sharing ESS Control System Data with External Researchers

Kaspian Garpvall, Olof Gilland

ISSN 1650-2884
LU-CS-EX: 2026-25

DEPARTMENT OF COMPUTER SCIENCE
LTH | LUND UNIVERSITY

EXAMENSARBETE
Datavetenskap

LU-CS-EX: 2026-25

Establishing Practices for Sharing ESS Control System Data with External Researchers

Praktiska rutiner för delning av kontrollsystemsdata från ESS med externa forskare

Kaspian Garpvall, Olof Gilland

Establishing Practices for Sharing ESS Control System Data with External Researchers

Kaspian Garpvall
ka4440ga-s@student.lu.se

Olof Gilland
ol5030gi-s@student.lu.se

June 16, 2026

Master's thesis work carried out at the European Spallation Source (ESS).

Supervisors: Fredrik Edman, fredrik.edman@innovation.lu.se
Karin Rathsman, karin.rathsman@ess.eu
Timo Korhonen, timo.korhonen@ess.eu

Examiner: Per Runeson, per.runeson@cs.lth.se

Abstract

As control systems increasingly rely on data-driven and machine learning approaches to improve reliability, efficiency, and operational insight, the demand for data from real-world operational environments has grown in both research and industry. Large-scale research infrastructures generate data with high analytical value, but external sharing can be challenging due to unclear incentives, limited processes, and sensitivity concerns. This thesis investigates how operational control system data from the European Spallation Source can be shared with external researchers in a secure, useful, and repeatable manner.

Using a design science research approach, the study combines literature review, document analysis, exploratory and evaluative interviews with ESS stakeholders, solution design, and prototype development, to create and empirically validate a data sharing framework. The framework includes recommendations for a role-based workflow, an authorisation scheme, sensitivity and shareability classification, a release request form, and an upload template. The findings suggest that data sharing should be enabled by workflows with clear authorisation schemes, sensitivity and shareability classification for data generated by subsystems then complemented with dataset-level review, standardised templates for internal and external communication, and follow lightweight technical implementation grounded in organisational practice.

Keywords: data sharing, control systems, public research infrastructure, operational control system data, control system data, operational data, data governance, design science research, sensitivity classification, European Spallation Source, ESS

Acknowledgements

We would like to thank the European Spallation Source for the opportunity to explore this fascinating topic at the institute, including all who participated in the interviews. In particular, we are deeply grateful to our supervisors, Karin and Timo, who have supported us greatly with planning, staying on track, and provided their ESS experience.

We would also like to express our sincere thanks to our academic supervisor, Fredrik, for his strong engagement and invaluable help in guiding us through the project, experiment structure and thesis writing.

Finally, we would like to thank Per for his valuable feedback and effort as examiner.

Contents

1 Introduction **11**
- 1.1 Research Questions . . . 12
- 1.2 Scientific Contribution . . . 13
- 1.3 Statement on Generative AI . . . 13
- 1.4 Thesis Structure . . . 13

2 Background **15**
- 2.1 Previous Work at ESS . . . 15
- 2.2 ESS Ecosystem . . . 16
  - 2.2.1 System Architecture and EPICS . . . 16
  - 2.2.2 Operational Control System Datasets . . . 17
  - 2.2.3 Internal Documentation and Policy . . . 17
  - 2.2.4 Operations Data Steering Board . . . 17
  - 2.2.5 WARA-Ops . . . 18
- 2.3 FAIR Principles . . . 18
- 2.4 Benefits of Data Sharing . . . 19
- 2.5 Challenges of Data Sharing . . . 19
- 2.6 Sensitive Data in Control Systems . . . 20
- 2.7 Open Data Ecosystems . . . 20
- 2.8 Data Governance and Data Stewardship . . . 21
- 2.9 Security and Privacy Frameworks . . . 21
  - 2.9.1 NIST Publications . . . 21
  - 2.9.2 The Five Safes Framework . . . 22
  - 2.9.3 RACI Model . . . 22

3 Method **23**
- 3.1 The Design Science Paradigm . . . 23
  - 3.1.1 Design Artifact . . . 24
  - 3.1.2 Technological Rule . . . 25
- 3.2 Design Science Activities . . . 25

3.3 Boundaries and Scope 29

4 The Artifact: The Framework **31**
4.1 Problem Conceptualisation 33
4.1.1 P1: Insufficient Data Sharing Maturity 33
4.1.2 P2: Varied Attitudes Toward Data Sharing 34
4.1.3 P3: Ambiguous Data Responsibility 34
4.1.4 P4: Organisational Risk and Impact 35
4.1.5 P5: No Record of Shared Datasets and Systems Classification . . 36
4.1.6 P6: Legal and Regulatory Limitations 36
4.1.7 P7: Coordinating the Internal Workflow 36
4.1.8 P8: Sharing Useful Data that Attracts Research 37
4.1.9 P9: Turning Raw Data into Workable Datasets 37
4.2 Solution Design 38
4.2.1 S1: Layered Data Sharing Process with Roles and Responsibilities 38
4.2.2 S2: A Comprehensive Authorisation Scheme 43
4.2.3 S3: Adopting a Data Governance Structure 45
4.2.4 S4: System and Dataset-Level Sensitivity and Shareability Grading 47
4.2.5 S5: Saving System Classification Status and Data Sharing Processes 49
4.2.6 S6: Investigation into Custom Data Licenses 49
4.2.7 S7: Release Request Form with Document Handler 50
4.2.8 S8: Upload Template 51
4.2.9 S9: Data Pre-processing and Transform Scripts 51
4.2.10 Technological Rules 52
4.2.11 Prototypes 53
4.3 Empirical Validation 54
4.3.1 Evaluating the Data Sharing Process (S1) 55
4.3.2 Evaluating the Authorisation Scheme (S2) 55
4.3.3 Evaluating the Data Governance Model (S3) 57
4.3.4 Evaluating the Sensitivity classification and Shareability Grading (S4) 58
4.3.5 Evaluating the Release Request Form with Document Handler (S7) 60
4.3.6 Evaluating the Upload Template (S8) 61

5 Findings **63**
5.1 RQ1: Processes, Roles and Authorisation Steps 63
5.1.1 Layered Data Sharing Process with Roles and Responsibilities (S1; P1, P3) 63
5.1.2 A Comprehensive Authorisation Scheme (S2; P1, P2) 64
5.1.3 Adopting a Data Governance Structure (S3; P3) 64
5.2 RQ2: Identifying and Handling Sensitive Data 64
5.2.1 System and Dataset-Level Sensitivity and Shareability Grading (S4; P4) 64
5.2.2 Saving System Classification Status and Data Sharing Processes (S5; P5) 65
5.2.3 Investigation into Custom Data Licenses (S6; P6) 65
5.3 RQ3: Technical Support for Data Sharing 66

5.3.1 Release Request Form with Document Handler (S7; P7) 66
5.3.2 Upload Template (S8; P8) 66
5.3.3 Pre-processing and Transformation Scripts (S9; P9) 66

6 Discussion **67**
6.1 Research Relevance, Rigour and Novelty 67
6.2 Implications for Research 68
6.3 Implications for Practice 69
6.4 Scope of Validity 70
6.5 Validity Considerations 70
6.5.1 Construct Validity 71
6.5.2 Internal Validity 71
6.5.3 External Validity 72
6.5.4 Reliability 72

7 Conclusion and Future Work **73**
7.1 Conclusions 73
7.2 Future Work 74

Appendix A A: Interview Material **83**

Appendix B B: Technological Rules **95**

Appendix C C: Empirical Validation Summary **99**

List of Abbreviations

AI Artificial Intelligence

EPICS Experimental Physics and Industrial Control System

ESS European Spallation Source

ICS Integrated Control System

ML Machine Learning

ODE Open Data Ecosystem

ODSB The ESS Operations Data Steering Board

PV Process Variables

WARA-Ops Wallenberg AI Autonomous Systems and Software Program (WASP) Research Arena for Operational Data

WASP Wallenberg AI Autonomous Systems and Software Program

Chapter 1
Introduction

Control systems increasingly rely on data-driven methods to improve reliability, efficiency, and operational insight. This has led to a growing demand for datasets from real-world operational environments in both research and industry. Such data supports the development of anomaly detection [1], [2], predictive maintenance [3], [4], [5], autonomous control, and benchmarking of analytical methods [6], [7]. When shared, it can also increase the social return on public-funded research infrastructure by enabling reproducibility and external innovation [8]. However, the sharing of data from the control system is complex. Barriers include security and privacy concerns, immature sharing practices and technical support, as well as differences in organisational priorities between research-oriented and industrial actors [9]. Furthermore, obstacles such as data preprocessing needs and mismatched data formats further complicate sharing. These challenges highlight the need for structured and thorough approaches to data sharing adapted to operational control system environments.

The European Spallation Source (ESS) in Lund is a pan-European research facility under construction, set to become the world's most powerful neutron source, enabling research in areas such as materials science, life science, and energy research [1]. Although the facility is not yet fully operational, its control systems already generate substantial volumes of operational data. This data is valuable for external research, particularly through collaboration with the WASP Research Arena for Operational Data (WARA-Ops), an ESS partner initiative connecting Swedish academia and industry [2].

Despite the interest and value of this data, external sharing of data from the ESS control system remains limited. Uncertainty surrounding responsibility and risk, combined with inconsistent sharing and authorisation procedures, makes it difficult to determine what data can be shared, by whom, and under what conditions. In addition, ESS must be confident that datasets do not expose sensitive information, while remaining useful to researchers.

[1] `https://ess.eu/about`, accessed on 28 January 2026

[2] `https://wasp-sweden.org/industrial-cooperation/research-arenas/wara-operational-data/`, accessed on 23 January 2026

This thesis investigates how ESS control system datasets can be shared with external researchers in a safe and repeatable manner. We begin by examining the current data sharing situation at ESS, before designing and evaluating recommendations and prototypes to support the sharing process. By focusing on control system data and on WARA-Ops as the external collaboration context, the thesis proposes a practical framework for sharing data under real organisational constraints.

1.1 Research Questions

The study is structured around three research questions:

Research Questions

RQ1 What processes, roles and authorisation steps are needed for sharing data generated by the ESS control system?

RQ2 How can sensitive data be identified and handled?

RQ3 What technical solutions and software tools can streamline the identification and handling of sensitive data (RQ2), as well as authorisation and sharing of ESS control system data (RQ1)?

RQ1 addresses the organisational side of data sharing by investigating how a repeatable release process can be structured, who should participate in it, and how authorisation decisions should be made. RQ2 addresses data sensitivity by investigating how ESS can approach handling of sensitive control system data. RQ3 addresses the practical support needed to make the process usable.

1.2 Scientific Contribution

This thesis contributes to research on data sharing in large-scale research infrastructures.

The first contribution is an empirically grounded characterisation of the data sharing problem at ESS, identifying barriers related to responsibility, authorisation, data sensitivity, organisational risk and impact, legal uncertainty, documentation, and external reuse.

The second contribution is a set of recommendations and prototypes designed to support controlled external data sharing in organisations with low data sharing maturity. The recommendations connect identified problems with solutions such as role-based workflow, authorisation scheme, sensitivity and shareability classification, and standardised templates for internal communication and documentation, as well as for external uploads.

The third contribution is the empirical evaluation of the recommendations and prototypes in the ESS context, identifying the solutions practitioners perceived as most feasible and relevant. Together, the findings extend knowledge on structuring secure and useful sharing of control system data in complex environments by considering both technical and organisational practices.

1.3 Statement on Generative AI

We used generative Artificial Intelligence (AI) tools during the writing of this thesis. Neutro, an in-house ChatGPT 5.0 large language model, and ChatGPT 5.5 were used for grammar and style assistance, text restructuring, LaTeX figure generation, and limited support in identifying relevant sources. All generated material was critically reviewed and validated by the authors.

1.4 Thesis Structure

The thesis is organised as follows. Chapter 2 introduces the study context. It covers previous work at ESS, the ESS control system environment, WARA-Ops, as well as literature on data sharing, including the FAIR principles, data governance, open data ecosystems, data licensing, and security and privacy frameworks. Chapter 3 describes the research method. It includes the design science approach, the design artifact, technological rules, design science activities and interview studies, as well as the boundaries and scope of the research. Chapter 4 presents the data sharing framework. It describes the identified problems, solutions, technological rules, prototypes, and the results from the empirical validation. Chapter 5 presents the findings structured around the research questions. It distinguishes between validated solutions as findings and non-validated solutions as supporting recommendations. Chapter 6 discusses the research relevance, rigour and novelty of the recommendations. It also includes the research and practical implications of the findings, as well as the scope and limitations of their validity. Chapter 7 concludes the thesis and outlines future work, including framework piloting, approval criteria, handling external data requests, and a broader ESS policy for sharing control system data.

Chapter 2
Background

This chapter provides the background needed to understand the data sharing problem studied in this thesis and the design choices made. It positions the work in relation to prior ESS studies and the ESS organisational context. It then summarises relevant literature on data sharing, governance, licensing, metadata, and reuse. It also introduces existing policies and frameworks as well as how they are applied in this thesis.

In this thesis, *control system data*, *operational data* and *operational control system data* all refer to data generated during operations by the industrial control systems at ESS. The terms *control system data* and *operational data* are used interchangeably, while *operational control system data* is used where additional clarity is appropriate.

2.1 Previous Work at ESS

Previous work at ESS provides a starting point for this thesis. Runeson and Söderberg [10] present a pre-study on open collaboration on control system data at ESS, outlining recommendations on tools, ecosystem governance and a general vision for the ESS Control System Data Lab. Their work highlights the importance of strategic commitment, stakeholder alignment, and common standards. Andersson et al. [11] further demonstrate the research value of ESS operational data and stress that shared datasets must remain understandable and useful even for actors without ESS-specific expertise.

Källström [12] examined the ESS control system from the perspective of data quality, metadata, and Machine Learning (ML)-readiness. The study identified problems such as inconsistent naming, noisy and redundant data, incomplete metadata, and ambiguous data responsibility. Källström recommends richer metadata practices, clearer archiving guidelines, and more explicitly defined governance roles separating data producers from data users. These findings are highly relevant to this thesis because they identify data responsibility, metadata quality, and governance as prerequisites for later sharing. They are therefore used as contextual grounding for the governance-oriented parts of the framework.

However, Källström's study primarily addresses internal data readiness, whereas this thesis focuses on the external sharing process and the organisational decisions required before release.

Taken together, these studies establish both the value and the challenges of sharing ESS control system data. This thesis builds on this work by shifting the focus from data readiness and ecosystem vision to the practical question of how such data can be shared in a secure and repeatable way.

2.2 ESS Ecosystem

This section outlines the relevant background on the ESS organisational context, including system architecture, communication systems, data handling and the WARA-Ops initiative. This background informs the framework, supporting both problem understanding and solutions.

At a general level, a control system monitors and controls physical processes by collecting measurements, communicating system states, and issuing commands to equipment. For ESS, operational data is data produced as part of the systems that operate, supervise, and coordinate technical equipment across the facility.

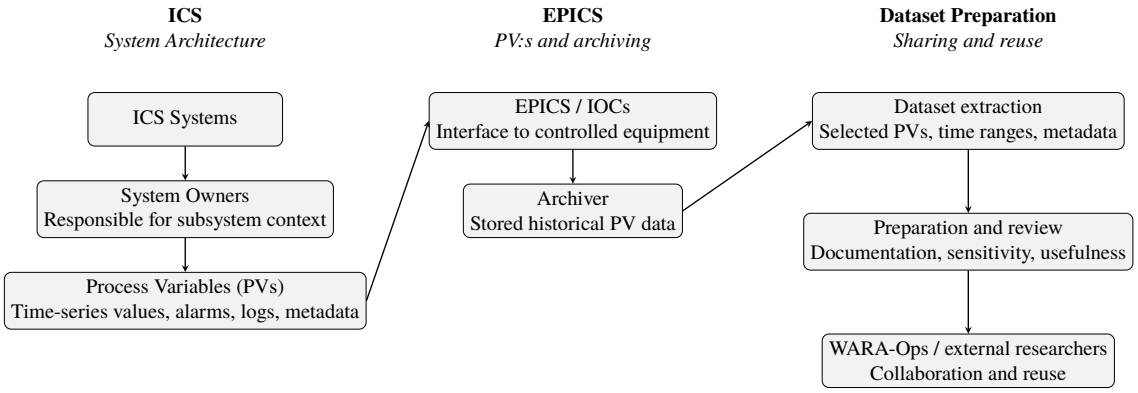


Figure 2.1: Conceptual overview of the ESS control-system data sharing context. Different parts of the Integrated Control System are associated with system owners and generate EPICS process variables through IOCs. These PVs are stored in the archiver and may later be extracted, prepared, reviewed, and shared as datasets through WARA-Ops.

2.2.1 System Architecture and EPICS

The Integrated Control System [1] forms a set of control and communication subsystems comprising the overall control architecture at ESS. Based on the distributed Experimental Physics and Industrial Control System (EPICS) framework [2], the Integrated Control System (ICS) is characterized by high degrees of independence in processes and behaviours

[1] `https://ess.eu/controls#overview`, accessed on 22 January 2026

[2] `https://docs.epics-controls.org/en/latest/`, accessed on 22 January 2026

of the individual systems [13]. EPICS enables the facility-wide data communication layer of ICS, integrating technical systems with communication interfaces to implement input/output controllers - the software components bridging hardware processes and the communication layer.

EPICS operates using addressable units of data known as Process Variables (PV)s [2], generated by the input/output controllers. Each PV is identified by a unique name that follows internal naming conventions and may include information about its originating subsystem and function. The ICS currently monitors on the order of 100 000 devices in the facility that generate significant amounts of operational data. The ICS context grounds the thesis in a real operational environment where complex, large-scale systems generate data that can support external analysis, while also introducing risks related to system dependencies, operational patterns and safety-relevant behaviour.

2.2.2 Operational Control System Datasets

In the context of industrial control systems, operational data typically consists of time-series process values, logs, alarms, control signals, status values and metadata generated by facility systems. When sharing data from the ICS, the relevant unit of sharing is therefore not an individual PV, but rather processed datasets prepared for external use or research. Such datasets may combine several PVs, metadata, time ranges, pre-processing steps, and explanatory material. Sensitivity and usability are properties of the dataset as a whole, rather than of individual signals.

2.2.3 Internal Documentation and Policy

ESS has internal policies relevant to information handling, including rules for intellectual property, personal data protection, export control, scientific data management, information security, and employee conduct. However, no dedicated policy currently governs the external sharing of control system data. This indicates a potential governance gap for sharing ICS data that is further addressed in this thesis.

2.2.4 Operations Data Steering Board

The ESS Operations Data Steering Board (ODSB) is an existing forum concerned with acquisition, storage, and governance of operational data ESS data [14]. While it highlights operational data as an already recognised concern, its focus is primarily on acquisition and storage rather than external sharing. How the proposed framework should relate to ODSB structures is outside the scope of this thesis, since assigning or extending formal decision authority is an organisational mandate issue for ESS rather than a design choice that can be settled within this study. The ODSB is therefore treated as relevant organisational context and a possible future integration point, rather than as the direct object of redesign.

2.2.5 WARA-Ops

The WASP Research Arena for Operational Data (WARA-Ops) is a Swedish research arena and collaborative partner of ESS, connecting Swedish academia and industry for data-driven operations[3]. It serves as the external collaboration context for this thesis.

The WARA-Ops data portal hosts partner-provided datasets and offers environments for analysis, including Python- and Jupyter-based workflows[4] [5]. The platform provides a relevant environment for the data sharing context studied in this thesis.

2.3 FAIR Principles

The FAIR principles describe how research data and metadata should be made *Findable, Accessible, Interoperable,* and *Reusable* [8]. FAIR does not mean that all data must be openly available. Rather, it means that data should be described, managed, and made usable under clearly defined conditions.

In this thesis, FAIR provides a useful reference for evaluating data sharing practices, even when access remains restricted. It supports the view that data sharing involves more than just moving files, underscoring the need for shared datasets to have sufficient metadata, documentation, and usage conditions to remain reusable outside their original context.

[3] `https://wasp-sweden.org/industrial-cooperation/research-arenas/wara-operational-data/`, accessed on 23 January 2026

[4] `https://wasp-sweden.org/new-data-portal-connects-swedish-industry-and-academia/`, accessed on 23 January 2026

[5] `https://jupyter.org/`, accessed on 23 January 2026

2.4 Benefits of Data Sharing

Control system data is valuable for data-driven research because it captures behaviour of complex technical systems during operation. This is reflected in three use cases that rely on time-series data, metadata and system context generated by the ICS.

First, anomaly detection research uses time-series data to identify abnormal system behaviour. Real-world, operational datasets are valuable since they include noise, correlations, and other system interactions that may be difficult to reproduce synthetically [1], [15], [16].

Second, predictive maintenance and long-term performance analysis rely on historical data to identify degradation patterns, forecast failures, and improve maintenance planning [3], [4], [5]. Such datasets may help external researchers study trends and system behaviour over time.

Third, operational datasets can support benchmarking and method comparison. The data is then used as a test-bed to determine the effectiveness of data-driven algorithms [7].

There is precedence for releasing datasets from large-scale research organisations for these purposes. Mogensen et al. [6] aims to support benchmarking efforts through the release of an ESS-related dataset with a reference causal graph. Similarly, the BOOSTR dataset released from Fermilab, was released demonstrate aspects on accelerator control system data to researchers [17].

Beyond research outcomes, data sharing can create organisational value by improving data reuse practices, encouraging consistent workflows, and strengthening long-term data stewardship [8]. At an inter-organisational level, it may also support innovation and new services, although such initiatives require coordination across incentives, responsibilities, and risk perceptions [18].

2.5 Challenges of Data Sharing

Data sharing is constrained by organisational, technical, legal, security, privacy, and ethical concerns [9]. For stakeholders, the benefits of sharing may appear uncertain, while the risks and costs are often immediate. Barriers include unclear incentives, data reliability concerns, large data volumes, licensing conditions, geopolitical restrictions, and the effort required to prepare and maintain shared datasets.

Privacy regulations, such as the European General Data Protection Regulation (GDPR) [19], and user information embedded in datasets create further obstacles for inter-organisational data sharing [20]. These challenges are amplified when data is shared for AI and ML research. ML applications often require high-quality, well-prepared, and well-labelled data at scale [21], [22]. In control system contexts, this creates a tension between data quality, analytical usefulness, and the presence of sensitive information in metadata, labels and system context.

2.6 Sensitive Data in Control Systems

For this thesis, sensitive data is understood as data that may create security, safety, privacy, legal, or organisational risks if shared externally. In control systems, sensitivity extends beyond personal information and trade secrets. Shared data may reveal system behaviour, operating conditions, incidents, interlocks, and other safety relevant functions.

Furthermore, even if individual signals are harmless, aggregation of the data may enable inference about system vulnerabilities [23]. Privacy risks in the datasets may arise through usernames, access logs, operator identifiers, free-text fields, timestamps, or metadata identifying individuals [24].

Sensitivity therefore depends on context. The same dataset may be low-risk in one situation but sensitive in another, depending on recipient, purpose, time of release, aggregation level, metadata, and combination with other datasets. This motivates the later distinction between system-level classification and dataset-level review before release.

2.7 Open Data Ecosystems

The Open Data Ecosystem (ODE) provides a useful lens for developing and understanding data sharing processes. Although data ecosystem research is still maturing and lacks a single settled definition, this thesis builds on prior work that conceptualises open data ecosystems as socio-technical arrangements involving data providers, users, intermediaries, governance structures, and supporting infrastructure [10], [25], [26].

In an ODE, data may be understood as a digital common: a shared resource available to all partners, but that requires collective maintenance responsibility to prevent deterioration [26]. While the data is openly accessible to ecosystem actors, it may still be subject to access restrictions preventing public sharing. Industry interest in ODEs is on the rise though complications exist: ODEs lack maturity, organisations are often hesitant to allow others to use their data as their own [26], and there is a need for greater standardisation in practices and tools to support collaborative setups [25].

For ESS, the data ecosystem perspective implies that data sharing and governance should focus on lowering barriers to safe sharing and enabling reuse. Drawing on Ostrom's design principles for open data ecosystems [26], the most actionable guidance in this study concerns clear scope and boundaries, locally defined and scalable rules, and the need for collective rule-making involving both data producers and data consumers. In practice, this entails maintaining clarity regarding who may access shared datasets and under what obligations, defining usage conditions for consumers that balance risk and research value, and involving stakeholders from both the producing and consuming parties: ESS and WARA-Ops.

2.8 Data Governance and Data Stewardship

Data governance concerns decision rights, roles and accountability for data-related matters. Otto [27] frames data governance around decision-making, data authority roles, and participation in governance processes. Ladley [28] similarly emphasises authority, roles, and accountability in ensuring proper management of information assets. Together, these perspectives emphasise data governance as the structured assignment of decision rights and responsibilities for managing data assets.

Data stewardship is the operational side of data governance. It concerns the people and processes that maintain data, metadata, definitions, and practical knowledge needed for responsible use [27], [29]. Data governance and stewardship are highly relevant to the thesis, as decision structures and data maintenance are central to enabling long-term, sustainable data sharing.

In this thesis, the choice was made to refer to *data responsibility* rather than *data ownership*. Ownership is a weakly supported definition that may imply legal property rights or exclusive control, whereas responsibility better captures the practical need for accountable decisions about sharing, documentation, and risk.

2.9 Security and Privacy Frameworks

This section outlines guidelines, policies and recommendations from external industry and research settings.

2.9.1 NIST Publications

The National Institute of Standards and Technology (NIST) is an agency of the United States Department of Commerce. Three NIST publications were particularly relevant as they translate broad security and privacy concerns into practical review criteria.

NIST SP 800-82r3 provided guidance for operational technology and industrial control system security, emphasising safety, reliability, and operational constraints [23]. It supports the idea that control system data should be assessed in relation to subsystem function and misuse potential.

NIST SP 800-53 Rev. 5 provided a catalogue of security and privacy controls [30]. This thesis does not implement the catalogue, but uses its control-oriented logic to motivate review, approval, access restriction, traceability, and role separation.

NIST SP 800-122 provided guidance for identifying and protecting personally identifiable information [24]. It supports dataset-level review of identifiers, logs, free-text fields, and metadata combinations that may create privacy risk.

2.9.2 The Five Safes Framework

The *five safes* framework is a set of principles for enabling safe access to research data. More specifically, the framework centres around the five safe principles: *project, people, data, setting,* and *outputs*[6]. In this thesis, the five safes functioned as a screening lens for release decisions, where its principles regarding project, people, data, settings, and outputs, informed the contents of forms and templates.

2.9.3 RACI Model

The RACI model is a project management tool for defining clear areas of responsibilities, enabling security and accountability for common assets and tools within organisations [31]. The model consists of four roles:

- **R - Responsible:** Individual or group responsible for conducting the practical work of data collection, building data sets and analysis of data.
- **A - Accountable:** Individual or group responsible for ensuring correct procedure and ensuring that the responsible party has the tools at their disposal to conduct their work.
- **C - Consulted:** Individual or group that carries the knowledge of the data and who can be consulted regarding the data.
- **I - Informed:** Individual or group that must be informed of the work, regardless of knowledge or direct involvement with the common or data.

In this thesis, RACI is used as a lightweight design lens rather than as a full matrix. The model informed the separation between actors who prepare the dataset, actors who approve the release, actors who provide knowledge or process support, and actors who must be informed of the decision. This logic is used primarily in S1 to structure the operational sharing workflow and in S3 to motivate the separation between stewardship, responsibility, and governance authority.

[6] `https://fivesafes.org/`, accessed on 29 January 2026

Chapter 3
Method

Design science is an empirical research paradigm focused on developing and evaluating solutions to real-world problems, in their contexts [32]. Considered a good fit for software engineering [33], research conducted within the design science research framework produces prescriptive design knowledge, such as principles, guidelines and actionable frameworks [32], [34], [35].

Design science research emphasises generalisable design knowledge to a greater degree than other design-oriented research paradigms, such as action research [36]. At the same time, it investigates problems situated within specific contexts in order to develop more general solution approaches [36]. By combining context-specific problem exploration and evaluation with broader design knowledge, design science research forms an appropriate foundation for this thesis, which explores the practical challenge of sharing control system data within the ESS context. This chapter presents the design science paradigm, design artifact, technological rules, and reflections regarding study scope.

3.1 The Design Science Paradigm

Design science research is primarily concerned with the development and evaluation of artifacts intended to improve practice in real-world settings [34], [35]. In this thesis, the artifact is a data sharing framework for ESS control system data. The framework consists of prescriptive recommendations: it guides action by specifying roles, workflows, classification approach, and supporting technical solutions.

Research contributions in design science are expressed through artifacts and interventions [32]. Artifacts capture the design knowledge produced by the study, while interventions constitute recommendations that achieve the desired outcome in practice [34]. The contributions are articulated as *technological rules*: general prescriptive statements that link interventions to an expected effect in a defined situation [32], [36].

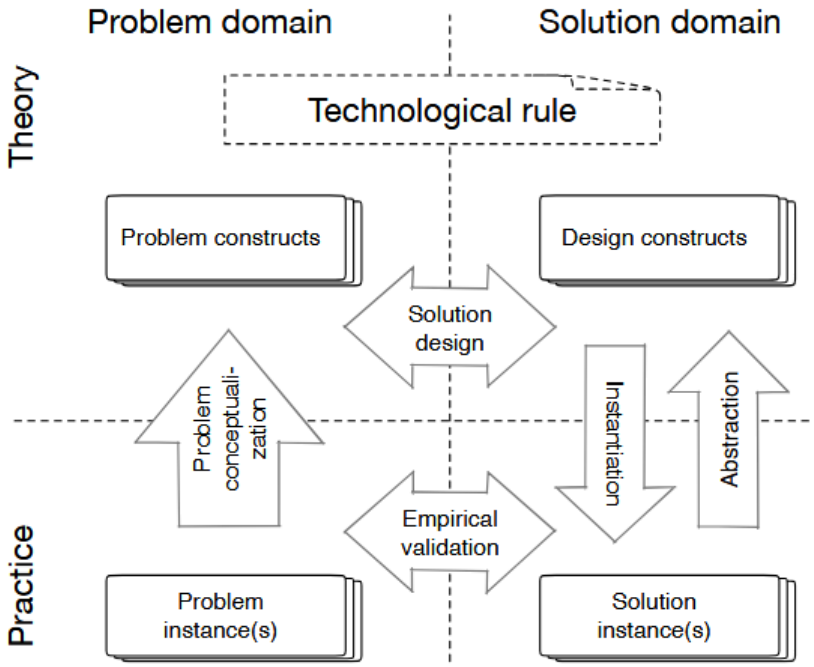


Figure 3.1: Visualization of design science research activities within software engineering [36].

This thesis follows the adaptation of design science research for software engineering formulated by Runeson, Storey, Engström, and colleagues [33], [36], [37]. In this framing, the work progresses through the stages of *problem conceptualisation*, *solution design*, *abstraction*, *instantiation*, and *empirical validation* (see Figure 3.1). In addition, practical guidelines for design science research in software engineering are employed, including defining the design artifact early and collaborating closely with practitioners [33], [38].

3.1.1 Design Artifact

This thesis delivers a composite design artifact in the form of a data sharing framework. The framework comprises a set of empirically grounded solutions and early-stage prototypes that enable organisational data sharing practices.

In this thesis, the design artifact constitutes a role-based sharing workflow, a comprehensive authorisation scheme, a data governance structure, a sensitivity classification approach, recommendations for custom license agreements and data registers, and finally templates facilitating internal and external communication. In addition, practical prototypes supporting the internal workflow and external data upload are included as part of the artifact.

While the composite design artifact formally constitutes the study deliverable, the term is comparatively abstract and unintuitive in repeated use. For this reason, this thesis refers to the design artifact simply as the *framework*. The accompanying prototypes constitute sub-artifacts within the overall design artifact and operationalise parts of the proposed solutions. These are referred to simply as *prototypes* for consistency and readability.

3.1.2 Technological Rule

Fundamentally, the desired outcome of design science research is the creation of field-tested and grounded *technological rules* [32], [36]. These rules describe general solutions derived from concrete problem-solution instances. They reflect the core steps of design science research by linking the problems with the solutions (see Figure 3.1). The rules are framed in a working situation and thereafter gain maturity and scientific rigour through empirical validation.

Technological rules can be expressed at varying levels of abstraction and may be structured hierarchically, with more specific rules forming broader, more comprehensive ones. The level of abstraction reflects the breadth of the rule's applicability: narrow rules are limited to specific settings, such as a single company, while broader rules are applicable across multiple contexts. The technological rules in this thesis are organised hierarchically across multiple levels of abstraction. At the highest level, the framework itself constitutes a broad technological rule, which is then supported by the rules addressing individual solution areas. These are thereafter further decomposed into more granular sub-rules that provide more concrete guidance. To ensure consistency, the study formulates each technological rule according to the template proposed by Engström et al.:

"To achieve *Effect* in *Situation*, apply *Intervention*"[33].

Similarly to the design artifact, the term technological rule may feel abstract in repeated use. This thesis therefore refers to instantiated technological rules as *solution candidates* throughout. The term technological rule is retained for when the solution candidates are presented in their rule-based representation.

3.2 Design Science Activities

This section describes how the study activities were carried out and how they map to the core activities of design science research. Table 3.1 summarises the relationship between the practical study steps and the corresponding activities.

In this thesis, *problem conceptualisation* was informed by literature review, document analysis, and exploratory interviews. *Solution design* addressed the problems and *instantiation* adapted the solution designs to the ESS and WARA-Ops setting as solution candidates. These solution candidates were subsequently evaluated from practitioner perspectives during *empirical validation*. Finally, *abstraction* involved assessing the extent to which the solution candidates may apply beyond the immediate study setting, i.e., the study's scope of validity.

Problem Conceptualisation

Prior software engineering research emphasises that weak initial problem understanding leads to poorly targeted recommendations [33], [37], and how practitioners often lack cross-disciplinary perspectives for shared problems such as data sharing [38]. Taking this into account, the problem understanding was built using three primary sources: literature, internal documentation, and interviews with relevant stakeholders.

Table 3.1: Research activities and their mapping to the design science research (DSR) paradigm.

Phase	Study activity	DSR activity
1.1	Literature review	Problem conceptualisation
1.2	Document review	Problem conceptualisation
1.3	Exploratory interview study	Problem conceptualisation
2.1	Synthesis of solution designs	Solution design
2.2	Adapting designs to ESS as solution candidates	Instantiation
3.1	Evaluative interview study	Empirical validation
4.1	Discussing scope of validity of solution candidates	Abstraction

The interviews followed a semi-structured format, involving both internal and external stakeholders, to balance open participant responses with discipline-specific questioning, consistent with established qualitative data collection practices [39], [40]. Questions were organised into categories, and participants received different question sets depending on their expertise (see Table 3.2 for the role-question set mapping and Tables A.1–A.5 in appendix A for the question sets).

The collected material was analysed by identifying concrete *problem instances*: concise descriptions of data sharing problems expressed in the ESS context. These were subsequently abstracted into broader *problem constructs* to guide solution design. As a result, this phase of the study produced a structured problem formulation suitable for design oriented research. To illustrate the distinction, two examples of observed problem instances are:

- Engineer hesitates in sharing a set of data due to it being unclear who is responsible for the data.
- There is no review procedure for datasets before sharing.

These form the following problem constructs:

- Ambiguity in data responsibility.
- Lacking process for dataset approval.

Solution Design

During solution design, the previously identified problem constructs were used to develop solution designs. This step shifts the focus from problem understanding to developing the recommendations for improving practice in design science [32]. The solution design was informed both by findings from the problem conceptualisation, as well as concepts from prior research on data sharing, governance, metadata, security, and stewardship. This approach ensures that solution designs capture reusable design knowledge while remaining grounded in the study context.

Table 3.2: Participant ID, associated discipline role at ESS and corresponding question set for the exploratory interview study.

ID	ROLE	QUESTION SET
P1	Information Security	General, Security
P2	Information Security	General, Security
P3	ICS Engineering	General, Engineering
P4	ICS Engineering	General, Engineering
P5	Legal Officer	Legal
P6	Legal Officer	Legal
P7	External Representation	External

To illustrate how problem constructs inform solution design, consider the following simplified examples:

- Ambiguity in data responsibility.
- Lacking process for dataset approval.

These problem constructs could be addressed through corresponding design constructs:

- Adopting a data governance structure to ensure all data has a responsible party.
- Establishing a set of dataset approval steps.

Instantiation

Instantiation adapted the solution designs to the specific organisational and technical context of ESS and WARA-Ops, forming concrete solution candidates for role responsibilities, workflows, classification approaches and designs for supporting prototypes. Further, prototypes enabling the sharing process were implemented in practice as well. This step enables validation in the original problem context, ensuring that solution designs are assessed not only on their theoretical quality, but also on how well it works in the intended operating environment [36].

Empirical Validation

Empirical validation evaluates the solution candidates in a real-world context. In this thesis, it was conducted through evaluative interviews with relevant personnel at ESS.

During the evaluative interviews, the problems identified during problem conceptualisation were presented to relevant stakeholders at ESS followed by the associated solution candidates. After each problem-solution candidate pair had been presented, participants were given the opportunity to discuss and provide feedback regarding feasibility and relevance at ESS. The interviews followed a semi-structured interview format, but were more structured than the exploratory rounds to allow a stronger focus on evaluation. To ensure consistency across the interviews, the solution candidates were presented to participants with a slide-show accompanied by a script read verbatim, with short breaks after each solution candidate to allow for questions and discussion. During the subsequent analysis,

participant responses were examined for sentiments regarding feasibility and general attitudes, with supporting quotes included to substantiate the sentiments and link the attitudes to the responses.

Solution candidates 5, 6, and 9 were excluded from the empirical validation for reasons of scope and dependency toward other parts of the framework. S5, which recommends registers for shared datasets and sensitivity classifications, has limited value before sharing itself is in place. S6, which addresses legal licensing, is broadly stated by design, and meaningful validation thereof is better conducted after legal experts first establish a starting position. S9 supports dataset preparation and would expand sharing potential, but is not as foundational to the sharing framework as the others. The non-validated candidates are therefore presented as supporting recommendations.

An overview of the participants and their respective interview roles can be found in Table 3.3. For each interview, participants were given a printed summary of the solution candidates. These can be found in appendix A.

Table 3.3: Participant ID and associated discipline at ESS for the evaluative interview study.

ID	ROLE
PA	Information Security
PB	Information Security
PC	Managerial Representation
PD	External Representation
PE	External Representation
PF	Managerial Representation
PG	ICS Software
PH	Systems Engineering

Evaluation Dimensions

Evaluation of a solution candidate in design science encompasses three dimensions: relevance, rigour, and novelty [33], [34]. Relevance concerns the severity of the problem and the general applicability of the solution from a practitioner perspective, as well as its alignment with established research challenges. Rigour reflects the extent to which the solution is grounded in prior research and the degree to which alternative solutions have been considered. Novelty refers to how the solution compares to existing approaches in similar contexts.

In this thesis, the empirical validation focused on practitioner feedback as the basis for evaluating practitioner relevance. Research relevance was assessed based on how well the solution candidates address problems identified in prior literature. Rigour was assessed based on grounding in established literature and relevant theoretical frameworks. Novelty was assessed through qualitative comparison with existing approaches to data sharing. Research relevance, rigour, and novelty are addressed separately in the discussion section.

Abstraction

Abstraction refers to the generalisation of design knowledge from a specific study context to broader settings. For this thesis, abstraction meant assessing the general applicability of the solution candidates toward other industrial and research contexts. The discussion therefore addresses the scope of validity for the solution candidates and study implications for both practice and research.

To ensure maximum relevance, rigour and novelty, technological rules are typically abstracted to the highest feasible level without losing realism [36]. Overly vague or narrow rules offer limited value: the prior lose actionable guidance, while the latter see restricted applicability to other contexts [33], [36]. These principles informed the assessment of the scope of validity of the solution candidates.

3.3 Boundaries and Scope

Design science research is inherently iterative, where repeated cycles of problem investigation, solution design, and evaluation progressively refine both results and problem understanding [38], [41]. However, due to the time and resource constraints of a master thesis, conducting multiple iterations was not possible. Instead, this study followed a well-grounded single-cycle approach, with thorough problem exploration and subsequent empirical evaluation. The study primarily relied on qualitative data rather than quantitative data, to support in-depth understanding of practitioner feedback. Due to scope constraints, not all solution candidates could be empirically validated. The most critical recommendations were prioritised for validation, while the remainder are included as supporting recommendations.

Chapter 4
The Artifact: The Framework

This section outlines the study phases of problem conceptualisation, solution design and empirical validation used to develop the data sharing framework. First, the problems identified during problem conceptualisation are described. Next, the solution candidates are presented, followed by their formal representation as technological rules. The prototypes are described including their adaptations to fit in the ESS context. Finally, the results of the empirical validation are presented. An overview of this section and the general research study is provided in Figure 4.1.

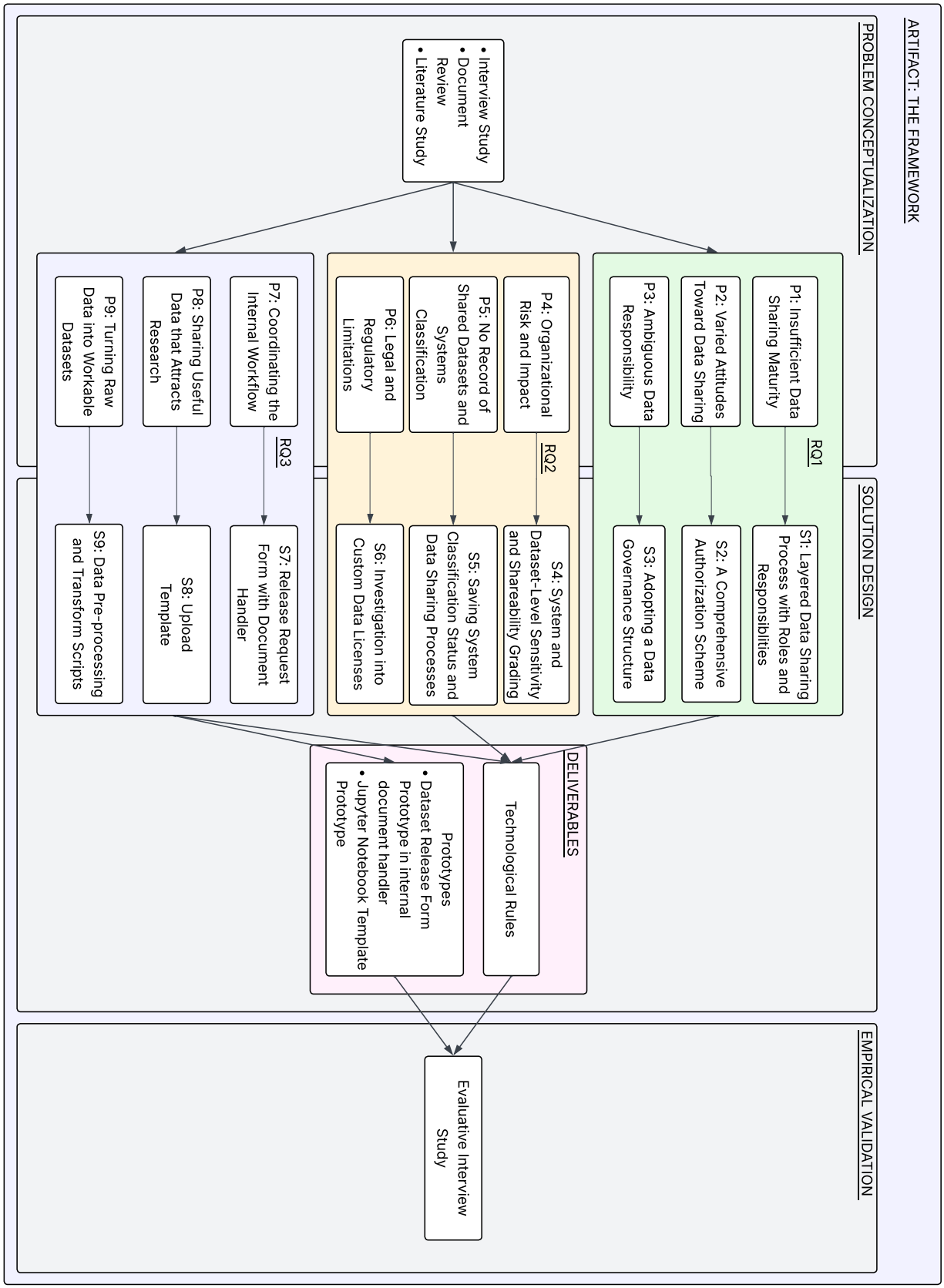


Figure 4.1: Overview of the creation of the framework: S1-S3 address RQ1, S4-S6 address RQ2, and S7-S9 address RQ3.

4.1 Problem Conceptualisation

The problems related to data sharing were identified through a literature study alongside stakeholder interviews. They were subsequently generalised as problem constructs, then grouped according to broader problems (P1-P9) to be addressed during solution design.

The categories were extrapolated through thematic analysis of interview responses and separated as cleanly as possible. Within the design science research framework, it was done through creation of one-sentence problem instances, that concretely described problems identified by the interviewees. These were subsequently generalised and abstracted away from the context of ESS into problem constructs. The problem constructs were thereafter grouped according to the thematic and broader categories P1-P9. Each problem construct was grouped according to the category that described it the best, but they are not entirely distinct, and some overlap was accepted.

A literature study was conducted at this stage to situate the thesis in established research, informing problem conceptualisation and solution design. Literature was collected through iterative searches in digital libraries, using keywords including *control systems, data sharing, data governance, data sensitivity, data-driven control systems, data licensing, anomaly detection*, and *data ecosystems*. Sources were selected for their applicability to ESS, assessed for quality and recency, then complemented with ESS internal documentation where applicable.

A total of seven stakeholders participated across five interviews. Interview categories included **Information Security**, **ICS Engineering**, **Legal** and **External Representatives**. Information security provided insight into data sensitivity, ICS engineering contributed practical perspectives on sharing challenges from those closest to the data, legal clarified the implications of sharing, and external representatives offered the perspective of receiving parties. All interview questions are available in appendix A. Recordings were deleted after the study's conclusion in accordance with European research ethics guidelines [42].

4.1.1 P1: Insufficient Data Sharing Maturity

Information security, engineering and legal stakeholders identified the lack of a unified cross-disciplinary procedure and called for greater clarity in processes. Information security stakeholders described the current ad hoc, case-by-case approach, and noted alongside engineering stakeholders the absence of a clear starting point for data sharing at ESS. Although all groups agreed on the need for sign-off practices, none identified an established approach for defining participants or responsibilities. External representatives noted that limited data sharing maturity is common across other research partners also, not just at ESS.

Document review found no practical guidelines for defining the scope of a data sharing solution for operational data from the control system at ICS. These findings reflect broader obstacles of organisational immaturity and lack of procedure regarding data sharing at ICS.

4.1.2 P2: Varied Attitudes Toward Data Sharing

Benefits and risks of data sharing from the control system were perceived differently across disciplines. Information security and legal stakeholders expressed greater concern for risk assessment than engineering and operations. All stakeholders agreed that no data sharing solution would satisfy everyone. Information security stakeholders emphasised the challenge of managing stakeholder expectations. External representatives viewed the data sharing as a potential research catalyst for ESS.

Overall, the interviews suggest that differing attitudes and preconceptions constitute barriers to data sharing, and that any solution must accommodate varying perceptions of risk and benefit.

4.1.3 P3: Ambiguous Data Responsibility

Responsibility over data from the control system was inconsistently interpreted in practice. Information security stakeholders noted that data authority rested with the information systems owner, in practice, the head of the division. They further remarked that formal responsibility did not necessarily align with technical expertise and that the distinction between formal and practical data authority remained unclear. In contrast, stakeholders representing engineering emphasised the system owners as responsible for the data in practice. For the external representatives, data responsibility belonged to whoever uploaded the dataset to the external platform.

Document review highlighted data responsibility as concentrated towards the information systems owner. This does not appear to reflect how the data responsibility is handled in practical terms. Unclear responsibility creates friction for data sharing, as it becomes ambiguous who should manage, maintain and make decisions regarding the data. Further, the interviews confirmed that, despite the ODSB's existence, data responsibility for sharing decisions remains ambiguous in practice.

4.1.4 P4: Organisational Risk and Impact

Concerns were raised by information security and engineering that data sharing externally could lead to reduced internal research opportunities, in particular if shared datasets saw use in closed industrial settings. Information security stakeholders emphasised the need for sensitivity assessment before release to avoid exposing operational security or personal information, and to ensure correct handling of data tied to third parties. Engineering stakeholders considered operational data to generally contain little sensitive information, but cautioned that releasing poor-quality data could damage ESS's reputation. Security and legal stakeholders both echoed concerns over reputational harm, particularly if sensitive data were mishandled, and raised additional concern over data being accessed by malicious actors.

The document review provided additional context and support for the participant responses. For example, the ESS rule-set for personal data handling affirmed ESS's adherence to GDPR and ethical rules, as well as the redistribution limits for data tied to ongoing research in cooperation with third parties. Internal code-of-conduct further explicitly states the responsibility of ESS employees to understanding and applying data handling rules, and to always act in the best interest of ESS over personal benefit. Furthermore, it was found that there were existing confidentiality and security classification standards that would apply to control system data as an information asset.

The concerns and problems identified regarding risk and impact reflects broader issues regarding data, whether to structure it based on datasets or systems, and how to incorporate such efforts with existing sensitivity rule-sets within organisations.

4.1.5 P5: No Record of Shared Datasets and Systems Classification

Engineering stakeholders mentioned the lack of internal storage for processed datasets and the need for an indexable and structured way of handling the shared data. They emphasised that control system data management efforts were focused primarily toward acquisition, with less attention to sharing. Post-interview follow-up with information security stakeholders confirmed that no solution was in place for maintaining or indexing security classifications for information assets at the ICS.

Review of internal documentation and the ESS data archival systems highlight that no record for processed datasets used in research is maintained. Previous operations were focused primarily on acquisition of scalar raw data in large quantities. No evidence suggested that existing confidentiality and security classifications were systematically maintained for operational data.

The problems highlight the broader issue of lack of lifecycle management of shared datasets as well as the management of sensitive classifications of operational subsystems and processed datasets.

4.1.6 P6: Legal and Regulatory Limitations

Information security stakeholders raised concerns regarding intellectual property, personal data handling, and third party contractual obligations, while emphasising that data sharing must take into account existing agreements between third parties and ESS. Legal and security stakeholders both identified a lack of a dedicated license for control system data. Further, the legal stakeholders brought up the potential for use of operational data for dual-use purposes in military applications development. External representatives mentioned user permission-based security mechanisms for their platform, but noted their limited effectiveness.

Document review contextualised the problems through description of ESS responsibility to limit dual use potential and handle personal information with care. It was found that the Creative Commons attribution (CC-BY) license had been assigned at ESS for future sharing of scientific data.

These issues reflect broader barriers to data sharing related to legal procedures and the difficulties of licensing operational data for external use.

4.1.7 P7: Coordinating the Internal Workflow

Engineering stakeholders noted the lack of mechanisms for cleansing metadata or ensuring datasets were free from sensitive information. There was unanimous agreement across all groups that resource-intensive or complex solutions would not be practically applicable. No participant knew of any existing process that could facilitate data sharing.

Broadly, the interviews highlighted the need for simple solutions in coordinating a data sharing process in large-scale and complex organisations such as ESS, with necessary steps to ensure sensitive data is not present.

4.1.8 P8: Sharing Useful Data that Attracts Research

External representatives emphasised that while dataset uploads did not need to follow a strictly enforced standardised structure, they should include enough context to motivate research and re-use. Ideas were raised that prompts or challenges to stimulate innovation could be included with the data. Engineering stakeholders highlighted that there could be internal conventions unclear to external parties in the control system data, such as PV naming conventions.

The interviews highlight the challenges of adapting internal conventions and data representation to external needs and the need for uploads to give enough context and motivation to encourage future research.

4.1.9 P9: Turning Raw Data into Workable Datasets

Engineering stakeholders noted the lack of a process for producing processed datasets, as the archival software only stores raw sensor data as scalars or arrays. External representatives emphasised researcher-friendly data formats compatible with standard data analysis environments in Python, R, or Julia.

An overview of how the EPICS implementation at ESS stores control system data highlighted that data formats were incompatible with the data format requirements of the WARA-Ops portal. Previous research at ESS had combined operational datasets with machine metadata stored elsewhere to create interpretable, processed datasets.

The interviews reflect the broader issue that ESS lacks easy and repeatable methods for processing and producing research-ready operational datasets.

4.2 Solution Design

This section presents the solutions candidates developed from the problems P1–P9. Each solution candidate S1–S9 is informed by exploratory interview findings, document review, and literature on data sharing, governance, and information security.

The solution candidates constitute instantiated design constructs. They were developed at a general level to capture reusable design knowledge and subsequently adapted to the ESS and WARA-Ops context. Together, the solution candidates and prototypes form the data sharing framework.

The solution candidates in this section are presented as they were shown to practitioners during the latter stage of empirical validation. The practical value of the framework is fully established only when the solution candidates are interpreted together with the empirical evaluation results and the corresponding suggestions for future improvements. Not all solution candidates were evaluated by practitioners. Consequently, solution candidates S5, S6 and, S9 should only be considered supporting recommendations.

4.2.1 S1: Layered Data Sharing Process with Roles and Responsibilities

To reduce uncertainty and distribute responsibility in an immature data sharing environment (P1), where data responsibility is ambiguously distributed (P3), we propose a clear layered data sharing process built around the four operational roles of *sharer*, *approver*, *admin*, and *receiver*. The four roles were selected to cover the minimum set of functions needed in a complete release process. These roles are functional process roles rather than organisational job titles, meaning multiple roles may be held by the same individual where responsibilities overlap. The role structure is intended to provide a clear starting point, a repeatable release path (see Figure 4.2), and visible responsibility boundaries. This solution candidate operationalises RQ1 by defining a repeatable workflow that reducing ad hoc decision-making. Validation status: *This solution candidate is empirically validated and considered feasible within the ESS context.*

Sharer

Description

The sharer initiates a data sharing case by identifying a candidate dataset and proposing it for release to an external actor. The sharer prepares the dataset release and supporting documentation, but does not approve the release.

Responsibilities

- **Define the dataset.** Compile the dataset and provide necessary metadata.
- **Explain purpose and context.** State why the dataset is shared, who the intended recipient is, and the context of sharing.
- **Propose release conditions.** Specify preferred access conditions, such as retention limits or restrictions on onward sharing.
- **Identify risks.** Provide an initial assessment of security, privacy, third-party, reputational, or other sensitivity concerns.
- **Apply mitigations.** Carry out reasonable minimisation measures before review, such as filtering, aggregation, downsampling, or metadata sanitisation.
- **Submit release request.** Complete the release form and submit it to the admin for an intake check of completeness, purpose clarity, and reviewer inclusion.

Approver

Description

The approver decides whether a dataset may be shared. This role must include security and management perspectives while considering the needs of engineering and data producers. The approver role is not necessarily one individual - several approvers can review a request.

Responsibilities

- **Review request.** Assess whether the purpose is justified, risks are understood, and mitigations are adequate.
- **Decide outcome.** Return the request as approved, requiring revision, or rejected.

Admin

Description

The admin performs intake, routes complete requests to relevant approvers, executes approved releases, and maintains records of shared data.

Responsibilities

- **Check request completeness.** Verify that purpose, context, documentation, and reviewer information are sufficient before formal review.
- **Route requests.** Forward complete requests to approvers and return incomplete requests to the sharer for revision.
- **Record decisions.** Document approvals, required revisions, and rejections, including reasons where relevant.
- **Prepare release package.** Ensure the approved dataset conforms to receiver technical and descriptive conventions.
- **Verify release readiness.** Check that the package is ready for upload and request clarification or reformatting when needed.
- **Publish dataset.** Release the dataset according to the approved access conditions.
- **Record data sharing-processes.** Record what was shared, when, to whom, under which conditions, and with which approvals and transformations.
- **Coordinate access requests and incidents.** Handle later access requests, approval history, documentation, and incident-response support.

Receiver

Description

The receiver is the external actor that either receives the dataset directly or acts as a platform host to provide access to the dataset externally.

Responsibilities

- **Enforce access conditions.** The receiver applies agreed access controls and usage restrictions.
- **Handle data under agreed terms.** The receiver manages the dataset according to the stated conditions of access, use, and retention.
- **Support responsible reuse.** Where relevant, the receiver helps reduce risks of misuse or inappropriate disclosure.

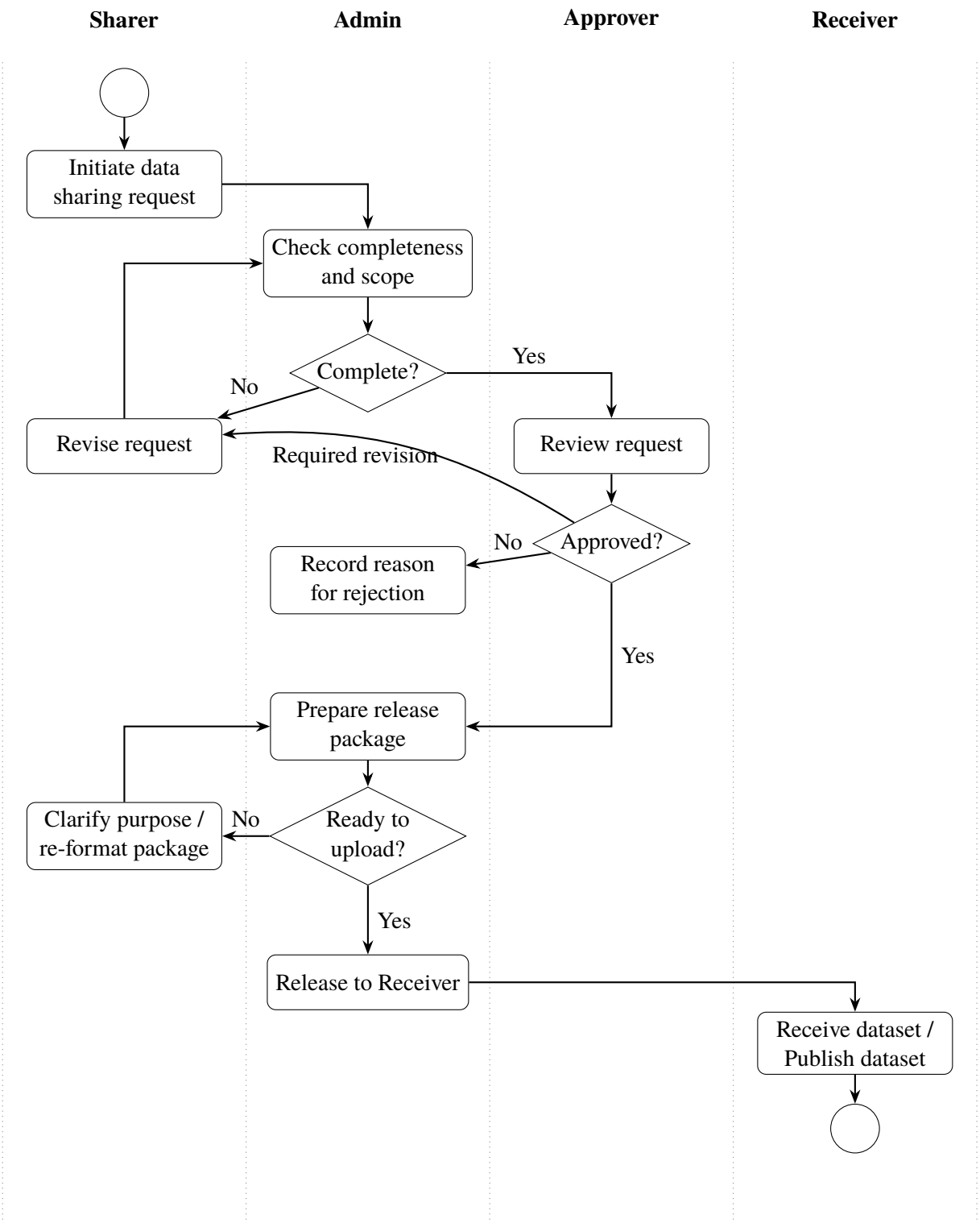


Figure 4.2: Overview of the data sharing workflow. Approver review is described in greater detail in S2, with a potential decision body structure outlined in S3. S4 and S6 inform sharing decisions regarding data sensitivity, while S5 covers recording of decisions and rejection reasoning. S7-S9 address the technical underpinning of the data request initiation and release packaging.

The proposed gated workflow (see Figure 4.2) consists of the following steps:

1. **Initiation.** The sharer compiles a dataset for release, prepares the release request, and assembles the dataset package with relevant metadata.
2. **Admin intake.** The admin checks the request for completeness and scope. At this stage, the admin determines whether the request is sufficiently documented, whether its purpose is clear, and whether it is ready to be sent for formal review. Incomplete requests are returned to the sharer for revision. The admin recommends reviewers and approvers.
3. **Review.** The approvers assess the request and return one of three outcomes: approved, required revision, or rejected.
    - **Approved:** The request proceeds to release preparation by the admin.
    - **Required revision:** The request is returned to the sharer for revision and must pass the admin intake step again before renewed review.
    - **Rejected:** The admin records the reason for rejection.
4. **Release preparation.** After approval, the admin prepares the release package and checks whether it is ready for upload. If needed, the sharer is asked to clarify purpose or re-format the package.
5. **Release to receiver.** Once the package is ready, the admin uploads or otherwise transfers the dataset to the receiver under the approved access conditions.
6. **Request handling.** Subsequent access requests for restricted datasets are handled by the admin in consultation with the original sharer and according to the stored approval conditions.

The role structure follows the RACI management model, ensuring that each release has clearly designated responsible, accountable, consulted, and informed parties [31]. The sharer assumes the primary responsibility, while the admin serves as the consulted party, bringing their knowledge of the data sharing process. The approvers subsequently fulfil both the accountability and informed aspects of the model, ensuring formal sign-off and that necessary parties are informed. Beyond application of the RACI model, the solution candidate also reflects that data sharing requires long-term stewardship rather than one-off exports [9].

4.2.2 S2: A Comprehensive Authorisation Scheme

To address varying attitudes and risk perceptions toward data sharing (P2), we propose a repeatable authorisation scheme that integrates technical, organisational, and security perspectives. The scheme is intentionally comprehensive to compensate for the organisation's current data sharing immaturity (P1) and the need to involve multiple perspectives in order to build trust. This solution candidate operationalises RQ1 by specifying roles and steps in the authorisation scheme, thereby supporting broader safety assessment across diverse expertise and perspectives. It contextualises the approval process through cooperation between reviewers providing domain-specific assessments and a final approver acting as the accountable role responsible for the final decision. Validation status: *This solution candidate is empirically validated and considered feasible within the ESS context.*

The comprehensive authorisation scheme should include the following positions:

- **Managerial Representation**. Managerial participation provides organisational legitimacy and oversight by assessing benefit to the organisation and potential reputational impact. Where appropriate, this role may be substituted by another authority (e.g. an Information Security Officer) and should, if applicable, be formalised in the Delegation of Authority and the decision-making role for approvals and role delegation within the authorisation scheme.
- **Information Security**. Information security provides expertise on misuse potential, security and confidentiality requirements, identifying vulnerabilities and sensitive infrastructure exposed in the data, and determining whether escalation to legal or other specialists is required.
- **Data Responsible**. This role should represent the perspective on whether data sharing aligns with the system's intended use and sensitivity. For ESS, this role could be filled by a system owner.
- **Engineering and System Operations Representation**. Engineering assessment is required to evaluate how the data relates to real system operation, configuration, and dependencies.
- **Data Steward** [1]. In cases where formal approvers and reviewers lack familiarity with the data structure or practical use, a steward may provide support. For ESS, this role could be filled by a control system integrator.

[1] Data Steward refers to the theoretical definition of the term, meaning a knowledgeable party that maintains the particular data being shared.

The authorisation steps should capture complementary perspectives without becoming unnecessarily bureaucratic. The request should be reviewed by all relevant roles, with each role assessing it from its own perspective. The authorisation steps are as follows:

1. **Completeness and scoping.** The sharing party must check that everything required for approvers to judge dataset sensitivity is included.
2. **Initiation by sharer**. A formal release request is submitted by the sharing party.
3. **Review**. The request is reviewed by relevant reviewers.
    - **Operational and technical review.** Data responsible and engineering representatives assess what the data represents, how it is produced, and whether disclosure may reveal internal logic or system-sensitive information.
    - **Security review.** Information security evaluates the risk that the data could enable misuse, hostile inference, or exposure of vulnerabilities.
4. **Managerial approval.** The final approval decision is taken by the managerial representative, weighing the expected benefit against the remaining risk. Valid approval outcomes are *approved*, *required revision*, *rejected*, or *escalated*. Required changes may include aggregation, removal of sensitive material such as images of persons or material exposing major incidents, and removal of free text containing personal information.
5. **Documentation and precedent.** All decisions, conditions, and rationales must be documented internally after authorisation to support traceability.

The authorisation scheme distributes expertise across different roles rather than relying on single experts, combining multiple perspectives in a controlled and documented process. Given data sharing immaturity at ESS, a broader scheme is justified to build trust and reduce person-dependent judgement. The scheme may, however, be streamlined over time as more cases accumulate, for example by reusing decisions or reducing the number of involved parties for well-understood cases.

4.2.3 S3: Adopting a Data Governance Structure

To address problems stemming from ambiguous delegation of data responsibility (P3), we propose adoption of a high-level data governance structure for the data from the ICS. At ESS, the ODSB appears to fulfil such a function, but closer problem examination showed that the board did not fulfil all aspects of data governance. Primarily, whereas the board suggests broad governance, our suggestion instead sees the data from ICS divided up in more granular responsibilities based on subsystems. Furthermore, external sharing of operational data is not currently within the scope of the ODSB.

Our suggestion is intended as a possible alternative governance structure. Whether implemented as a parallel body, a sub-function, or a future extension of the ODSB's scope is an organisational decision for ESS outside the scope of this thesis. The proposed roles are designed to cover the essential entities in data governance that would aid data sharing from ESS. This solution candidate addresses RQ1 by enabling organisational data handling that supports external data sharing. Validation status: *This solution candidate is empirically validated and considered low in feasibility within the ESS context.*

Data Steward

Description	A role that supports the documenting and maintaining of data and metadata of a subsystem.
Responsibilities	• Managing the data and associated machine metadata such as units, pv descriptions, and alarm limits, from a subsystem or several subsystems. • Acting as a contact for knowledge on data quality, characteristics and requirements.

Data Responsible

Description	An authority role that has responsibility over the data produced by a specific subsystem.
Responsibilities	• Responsible for data generated by a certain subsystem or PV in the control system. • Acting as an authoritative party for the specific data, making decisions on how it may be handled including whether data can be shared or not.

Data Council

Description	A decision making authority on how data should generally be handled and with what principles, given system sensitivity and other requirements.
Responsibilities	• Authority to assign data stewardship and data responsible for data generated by subsystems. • Responsible for resolving issues arising from ambiguous data responsibilities by assigning stewards and responsible parties to data from ICS subsystems, may also resolve conflicts for data where no responsible party is defined. • Balances the need of data producers and users, data responsible, and business needs from the ESS perspective.

The proposed data governance model aims to balance simplicity and usability for an organisation with a still-maturing approach to data sharing. Separating the governance into data stewards, data responsible and data council is designed to support formalised data stewardship and encourage data sharing while balancing organisational needs.

As ESS is still in an early stage of data-oriented management of control system data, some reduction in granularity is necessary to keep the approach simple. The transition to the governance structure should formalise existing organisation responsibilities that implicitly deal with data matters already into the governance roles, such as system owners, data producers and users, system integrators and information security. This aligns with the view that data governance is best formalised through developed principles [27]. Furthermore, while the high-level structure provides clarity in responsibility assignment, the roles are not intended to be rigid. As data sharing becomes established practice, the organisation can transition toward other governance models with more detail, such as the scheme proposed by Källström [12] previously.

The data governance is is grounded in theoretical support from the RACI model for accountability and responsible governance, with accountable and informed parties represented in the data council, responsibility through the data responsible, and the consulted roles filled by the data steward [31].

The threefold structure is justified as consisting of the essential parties necessary for a functional data governance solution [27], with an expressed acknowledgement that the governance solution will need to see refinement as difficulties in assigning authority and responsibilities for data arise.

4.2.4 S4: System and Dataset-Level Sensitivity and Shareability Grading

To address the need for organisational risk and impact assessment for data shared from the ICS (P4), we propose a system-level and a dataset-level sensitivity classification approach that expands upon existing ESS baselines for information identification and classification.

Sensitivity classifications are to be made at the level of ICS systems or subsystems. Outputs from the same physical process and operational context generally share sensitivity characteristics; classifying them at PV-level would be costly and error-prone. Employing system-level classification provides a robust approach consistent with established OT security thinking where protection needs are determined by the operational role and environment of components [23]. This solution operationalises RQ2 in defining an approach to sensitive data handling on system- and dataset levels, thereby forming a scalable solution accounting for both the classification need of processed data and the broader purpose of classifying all operational data generated at the division. Validation status: *This solution candidate is empirically validated and considered feasible within the ESS context.*

System-Level Classification

In our proposal, each ICS system is decomposed into reasonable data-producing sub-parts. As an example, the system for cryogenics could consist of refrigeration-cycle instrumentation, valve actuation control, vacuum pumping, interlock-related channels, alarm and event logging. A shareability grading is included within the classification to capture common organisational constraints such as ongoing research. The grading operationalises the organisational constraints beyond security and should be decided on jointly by system owners and key data users. The data produced by the sub-parts are thereafter assigned a:

- **Confidentiality class**: Public, Internal, Confidential, Strictly Confidential, which is an already existing ESS confidentiality classification scale for information assets.
- **Security level**: SL1–SL4, which is an already existing ESS security level scale for information assets at ESS describing plausible impact if data were disclosed improperly
- **Shareability grading**: Shareable now, Shareable after embargo, Shareable on case-by-case assessment, Not shareable.

The classification is performed jointly in classification sessions by information security personnel, responsible parties such as system owners, and the head of division. Confidentiality and security level are graded according to existing ESS rulesets, with an additional shareability grade included to support data sharing. Datasets inherit the classification of their source by default, enabling scalable classification.

Dataset-Level Review

Although system-level inheritance is the default, we propose a mandatory dataset-level review before actual release. The review focuses on whether the requested dataset contains, or enables inference of information that should be restricted regardless of system shareability. In practice, release gating requires inspecting both signals and metadata.

- **Safety and protection signals.** Data that expose interlocks, trip logic, protection thresholds, alarm rationales or other safety-critical behaviour should be treated as higher sensitivity than general process telemetry, since misuse can have disproportionate safety and security impacts [23].
- **Linkage and inference risk.** Combinations of signals can enable sensitive operational inference regardless of individual risk toward individual signals. When inference risk is plausible, dataset must be transformed through minimisation, aggregation, downsampling or masking until residual risk is acceptable [9], [24].
- **Personal or access-related data.** Operator identifiers, access logs, usernames, hostnames, free-text fields, or quasi-identifiers require sanitisation [24].

4.2.5 S5: Saving System Classification Status and Data Sharing Processes

To address the problem of there being no record of system classification status or the datasets that have been shared (P5), we propose creating and maintaining registers for both the sensitivity classification of data generated by subsystems as well as the datasets that have been shared. By maintaining and regularly checking such registers, ESS gains an overview of the research conducted externally and build precedent for the sharing process. The sensitivity classification register should be indexable at subsystem level and include the security, confidentiality and shareability gradings. Each entry should record the assigned classification status, the responsible actors that set the status, the time of classification, and the rationale behind the given status. The shared data register should document the dataset that was shared, with whom, under what access conditions, and who approved the external sharing.

The need for persistent governance of shared datasets aligns with prior research on data sharing infrastructure, which emphasises long-term findability, sustainability and high quality metadata to support reuse [8], [9]. The solution candidate operationalises RQ2 by clarifying the handling of sensitivity classifications and addresses RQ1 by maintaining records of previously shared datasets alongside process information that may inform future decisions. Validation status: *This solution candidate is not empirically validated and is included as a supporting recommendation.*

4.2.6 S6: Investigation into Custom Data Licenses

To address the problem of licensing for control system data shared externally (P6), we propose a thorough investigation into a creation of custom license agreements or a joint Terms of Agreement (ToA) suite covering operational data shared from large-scale research organisations. While the initial interview study reflected a desire for a license supporting the data, it was discovered during literature review that current license frameworks in the Open Data Commons and Creative Commons were generally not drafted to protect control system data. To mitigate, drafting a custom ToA or analogous legal contract would allow for enforceable demands that is tailored to operational data from large-scale research organisations such as ESS, rather than adapting poorly matching existing license frameworks. The proposed investigation would need to evaluate measures to mitigate harmful external use of control system data in dual-use applications development. For this solution candidate, it was deemed outside the scope of this thesis to further investigate license agreements.

The suggestion operationalises RQ2 by emphasising the need for enforceable agreements describing permitted external handling of sensitive data. Validation status: *This solution candidate is not empirically validated and is included as a supporting recommendation.*

4.2.7 S7: Release Request Form with Document Handler

To address the challenge of coordinating the internal workflow (P7), we propose a template for a dataset release request form. The form would be a document artifact, usable within document handler software. The document would be shared document between the actors in the sharing process, allowing for distributed review. This solution candidate operationalises RQ3 by outlining a tool that supports the repeatable data sharing process while enabling a quicker and simpler dataset-level review step. Validation status: *This solution candidate is empirically validated and is considered feasible within the ESS context.*

The document handler solution should support comments from reviewers, alongside a final approver, enabling informed sign-off. The document handler also ensures traceability and auditability of the process. The shared document should be a structured form that contains the following:

- **Purpose Statement:** A clearly stated purpose for sharing, including benefits to the organisation.
- **Intended Recipients:** The intended recipients for the dataset release, described in as much detail as possible.
- **Background:** A description of the dataset characteristics and where the data is from at ESS, with a focus on what characteristics make it a good fit for external sharing. If publishing a dataset after a study, attach or describe the associated research.
- **Processing steps:** The transformations applied to the data to generate the dataset, or if minimisation or similar has been done.
- **Sensitivity information:** The inherited subsystem sensitivity classification for the dataset, alongside confidentiality and shareability grading. If subsystem classification is unavailable, assess the processed dataset directly and allow for case-by-case review.
- **Access conditions:** The access conditions under which the data is to be shared. This includes specifying what environment the data is shared to, whether onward sharing should be limited, retention rules for the dataset, and whether the dataset is shared on an on-request basis.
- **Sign-off specifics:** The reviewers that should be included for the release, based on the approval scheme from earlier solutions, on other shared data precedent, or other reasoning.

4.2.8 S8: Upload Template

To ensure that data shared is useful and attractive to researchers and to make sure external representations of the data matches internal expectation (P8), we propose packaging the dataset release alongside an easy-to-understand upload template. The template can be implemented as a Jupyter notebook. Filling out the template would be required for all external data releases. The solution candidate operationalises RQ3 in enabling a consistent upload format for all external sharing, streamlining the sharing process while ensuring the data is attractive to researchers. Validation status: *This solution candidate is empirically validated and considered feasible within the ESS context.*

The notebook is passed, alongside the release request form to the party responsible for contact with external parties. It should include the following:

- **Purpose:** A clearly stated purpose for releasing the dataset, including its intended research use or an associated research question for future users. The description should be at least as detailed as in the release form.
- **Background:** A description of the background of the dataset, including its origin and data type. Just as for the dataset release form, focus on the characteristics of the data that make it a good fit for sharing. For research conducted, include a thorough but brief description of said research, how the data was used during the research, as well as a link to the paper with publication. The description should preferably be more detailed than in the release form.
- **Guidelines:** A complete set of data usage guidelines. These should cover data extraction, interpretation and any special considerations. The description must enable external researchers to use the dataset independently and include all processing steps necessary to reproduce the dataset to ensure compliance with FAIR reproducibility principles.
- **Examples:** One or more runnable examples, such as a simplified research case, or otherwise describe the usage with pseudo-code. The examples must enable researchers to quickly assess the data and its relevance to their work.

The template requirements are informed by feedback from external representatives and align closely with prior ESS data sharing work by Mogensen [6]. They formalise earlier good practice in a re-usable, repeatable template.

4.2.9 S9: Data Pre-processing and Transform Scripts

To address the problem of data extraction and processing it into workable datasets (P9), we propose lightweight local scripts that merge time-series raw data with metadata into joint datasets. Further, the scripts must ensure the data is in standard data formats for ML, such as Python Pandas, Julia or R-interpretable data frames. The scripts should remain simple and minimal, in line with ESS preferences. For example, the scripts could be developed as Python scripts and distributed via an internal portal.

The solution operationalises RQ3 in the creation of smaller, re-usable scripts for pre-processing data and metadata, thereby enabling research at greater capacity, which is necessary for data sharing at a foundational level. Validation status: *This solution candidate is not empirically validated and is included as a supporting recommendation.*

4.2.10 Technological Rules

This section presents the technological rules representing the instantiated solutions. They are organised hierarchically around the design artifact: the framework. For each technological rule, the more detailed sub-rules can be found in appendix B.

The Framework

TR0. To achieve secure and efficient sharing of control system data **in** large research infrastructures **apply** a structured data sharing framework combining sensitivity classification, role-based sharing process and authorisation steps, and technical solutions.

Layered Data Sharing Process with Roles and Responsibilities

TR1. To achieve clear and repeatable data sharing despite low data sharing maturity **in** data sharing at ESS **apply** a sharing process with roles and responsibilities for the sharing actors within a workflow.

A Comprehensive Authorisation Scheme

TR2. To achieve effective data sharing despite varied attitudes toward data sharing in data sharing at ESS **apply** a comprehensive authorisation scheme including information security, responsible data authority, data stewards, engineering and management.

Adopting a Data Governance Structure

TR3. To achieve consistent interpretations of data responsibility and explicitly assigned authority over data **in** data handling at ESS **apply** adopting a data governance model with roles of data stewards, data responsible, and data councils, while formalising existing responsibilities into the governance structure over time.

System and Dataset-Level Sensitivity and Shareability Grading

TR4. To achieve sensitivity decisions that account for organisational risk factors on both a data characteristics and a processed dataset level **in** data sharing at ESS **apply** a classification scheme for data produced by subsystems that align with existing confidentiality and security standards, extended with a shareability grading and dataset-level review before release.

Saving System Classification Status and Data Sharing Processes

TR5. To achieve indexable storage of data classification and previously shared datasets **in** data sharing at ESS **apply** recording data classification status and shared datasets in shared registers.

Investigation into Custom Data Licenses

TR6. To achieve compliant external data sharing that ensures legal requirements are met **in** ESS data sharing **apply** an investigation into custom license solutions or terms-of-agreement for external parties that is tailored to operational control system data.

Release Request Form with Document Handler

TR7. To achieve an easy-to-follow data sharing workflow that ensures sensitive data is not present **in** data sharing at ESS **apply** a release form with purpose, recipients, background, processing steps, sensitivity information, access conditions and sign-off specifics, within a document handler supporting distributed review.

Upload Template

TR8. To achieve clarity of internal conventions and improved research usability for datasets **in** data sharing from ESS to external users **apply** filling out a standardised upload template for every release containing purpose, background, usage guidelines and descriptive or runnable examples.

Data Pre-processing and Transform Scripts

TR9. To achieve a repeatable and automatic procedure for turning raw control system data into workable datasets formatted for external needs **in** data sharing from ESS to external users **apply** conversion and transformation scripts that combine raw time-series sensor data with metadata and converts file formats.

4.2.11 Prototypes

The prototypes were implemented as a CHESS-compatible[2] release form template and a re-usable Jupyter notebook, to reduce technical debt. When creating the prototypes, implementation details had to be specified as follows:

- The sharer prepares their dataset release package as a Jupyter Notebook.
- The sharer prepares their dataset release request with help and guidance from the Admin, completing and uploading the CHESS form. Approver and reviewers are thereafter set accordingly in CHESS.
- After approval granted, the complete CHESS workflow history, communications and the release form is packaged alongside the Jupyter Notebook to be stored by the Admin.
- The Admin shares the release package (Jupyter Notebook and dataset) externally to WARA-Ops for upload.

The dataset release form CHESS template was inspired by the five safes framework for data access request applications developed by the UK Health Data Research Alliance [43]. The structure emphasises cross-project assessment, people, data, settings and outputs rather than the dataset alone.

The upload template was implemented as a Jupyter notebook designed with the addition of a license section specifying the release license, along with WARA-Ops-provided code snippets for data extraction on their platform.

The prototypes are maintained as internal ESS artifacts and are therefore not included in this report, but available upon request.

[2]CHESS is a document handler that is currently employed at ESS.

4.3 Empirical Validation

This section presents the results of the empirical validation of the solution candidates. For each solution candidate, we summarise the most relevant feedback from the evaluative interviews.

Three interviews were conducted with seven stakeholders present in total, representing **Information Security**, **ICS Software**, **Systems Engineering**, **Managerial Representation** and **External Representation**. Information security stakeholders contributed expertise on data sensitivity and information flows. ICS software stakeholders provided insight into practical feasibility, while systems engineering stakeholders offered a holistic view of system ownership and engineers working closer to the data. Managerial representation offered a broader ESS-wide perspective, and external representation added the viewpoint of data recipients, reflecting the needs of researchers who receive and use the datasets.

A semi-structured interview format was employed in which each problem and its associated solution candidates were presented. Solution candidates 5, 6 and 9 were excluded to focus the validation toward more critical solution candidates. Following research ethics guidelines for European academia, the interview recordings were deleted after study conclusion [42]. A summary of the results can be found in appendix C.

4.3.1 Evaluating the Data Sharing Process (S1)

The data sharing process was generally well received by stakeholders, who highlighted its clarity, lightweight nature, and ease of use (see Table 4.1). Information security stakeholders considered it feasible but noted approver criteria and knowing who should fill the approver roles could be challenging. ICS software stakeholders were largely neutral, noting the approval stage as the main issue. Managerial representatives strongly supported the process but raised concerns about handling externally initiated requests for data.

Overall, participants demonstrated a clear understanding of the solution candidate and considered it feasible and satisfactory. However, uncertainty around approval and request handling may create implementation challenges. Future improvements include providing explicit examples of who could hold approval positions at ESS, establishing approval criteria, and extending the process to support data requests, potentially by routing them through the administrator who delegates sharing responsibility.

Table 4.1: Summary of group sentiments regarding solution 1.

Group	Sentiment
Information Security	Positive. Brought up challenge of approver inclusion and approval criteria.
"It sounds very clear and straightforward" "the approver side, we should maybe investigate further into that. Maybe even look at the criteria on which the approvers make their decision." "I don't think what you're describing is too heavy - it's the minimum required."	
Group	Sentiment
ICS Software	Neutral, brought up challenges of how the approval process could be done and said it was the most interesting part
"The interesting part is going to be the criteria by which approver and admin do their decisions."	
Group	Sentiment
Managerial Representation	Very positive, but worried it does not consider how to handle external requests.
"This seems like a push scenario - someone wants to share something. I imagine it will often be a pull scenario, where someone externally wants to have something." "Say someone wants to share something - then these roles fall into place automatically, but if it's a pull request, then how do you know who is the sharer?"	

4.3.2 Evaluating the Authorisation Scheme (S2)

The comprehensive authorisation scheme was received with mixed to positive views by stakeholders (see Table 4.2). Information security stakeholders viewed the solution candidate positively but stressed the need for clearer paths to put people into authorisation positions. The ICS software stakeholder expressed that the process could become too heavy and that a policy could ensure it was not so. Managerial representatives acknowledged the scheme's benefits but argued that it avoided, rather than addressed, the differing attitudes directly. They also echoed the need for broader organisational policy. External representatives viewed the solution candidate positively.

Overall, the solution candidate was perceived as feasible, but did not appear to be sufficient on its own. Participants repeatedly expressed a wish for broader policy that would outline approval criteria and organisational intent, which they felt would simplify authorisation and directly resolve different attitudes. Future improvements include developing approval criteria based on data characteristics and incorporating them into a top-down control system data policy that would define the facility's stance toward data sharing and guide dataset approval.

Table 4.2: Summary of group sentiments regarding solution 2.

Group	Sentiment
Information Security	Positive, though mentioned that fit at ESS is limited by the lack of knowledge on who to fill the reviewer roles.
"The flow itself seems very logical and clear." "I think it's feasible with the caveat that we don't actually have a clear path to put people into those authorisation roles."	
Group	Sentiment
ICS Software	Neutral. Remarked that it would be heavy to repeat it for each sharing; Expressed the need for a data policy that lays down how data sharing should be done.
"I think this looks good as well." "If you repeat it, it quickly becomes heavy." "What you're missing is a data policy which basically lays down the law of how data sharing should be done. The different perspectives can be moderated because there's a data policy."	
Group	Sentiment
Managerial Representation	Slightly positive. Remarked that it does not solve the core problem of there being different perspectives, but rather works around it.
"It wasn't clear to me how the solution actually solves the problem. It may only solve the problem of getting the different perceptions involved in the authorisation process." "It probably will not help the perception part of things but it may, because perceptions are also formed by engagement"	
Group	Sentiment
System Engineering	Neutral, agreed with all the issues raised by the ICS Software and Managerial Representation groups.
"I agree I don't need to repeat it I think."	
Group	Sentiment
External Representatives	Positive, remarked on the value of protocol to decide on sharing decisions.
"It makes sense, that it is is not the gut feeling who sort of decides whether it's a go or no go."	

4.3.3 Evaluating the Data Governance Model (S3)

The data governance model was received with mixed to negative views by stakeholders (see Table 4.3). While information security stakeholders viewed the solution candidate positively, they raised feasibility concerns related to the granularity of data responsibilities, and questioned if the governance should be limited to the ICS. They also doubted that an organisation of ESS's size warranted a multi-member data council. The ICS software stakeholder expressed low confidence, instead favouring a top-down policy that would treat all operational data as a unified data lake. Managerial representatives saw the solution candidate as problematic and unlikely to resolve tension between individual responsibility and the ESS-wide views. The systems engineering stakeholder found it infeasible, advocating instead for policy governance of a large data entity with potential for separate dataset storage for research purposes.

Overall, the responses suggest that the solution candidate would face significant challenges and resistance at the organisation. While participants generally recognized the benefits in the governance scheme, they raised significant concerns about its implementation. Most participants considered it infeasible to divide data responsibility in this manner at ESS, did not see a clear way to identify suitable role holders, or saw the required organisation change as too extensive to be practical. Instead, participants showed clear preference for broad policy governance over control system data as a data lake. Future work should more closely examine how governance may be integrated with higher-level policy and potentially seek to introduce transitional governance mechanisms with clearer pathways to data governance.

Table 4.3: Summary of group sentiments regarding solution 3.

Group	Sentiment
Information Security	Positive. Mentions difficulties in data aggregation for responsibility role; Whether a managerial function could replace a data council.
"In my professional experience, this only works when you aggregate it to a sufficient level, not too high, but not too low." "I would say there might be more data stewards and few data responsible who make decisions" "Is ESS really so big that we need to have a data council or is this simply a managerial function?"	
Group	Sentiment
ICS Software	Neutral. Felt that the data should be treated as a data lake under a general ESS policy. Authority of data assigned to researchers that aggregated the dataset. Governance solution was in the right direction but not sure this would work.
"Regarding control system data as a data lake, I think there should be a generic policy." "It's a distinction between the whole lake and the data set that you're working on being limited to a group of people that are actually doing the research." "It's in the right direction."	
Group	Sentiment
Managerial Representation	Negative. Remarked it would be problematic and would not resolve differences in personal opinions and facility perspectives.
"This will not be unproblematic." "There will be this tension, I think, and I don't see that this can resolve that."	
Group	Sentiment
Systems Engineering	Negative. Echoed ICS software that ESS handling of the data should follow a joint policy. Expressed how it was not feasible at ESS. Suggest to split data handling in two large lakes, one big data lake for all data collected, and individual personal storage of processed datasets.
"It's ESS data. It needs to follow one policy, one guidance." "Doing this for the data lake, it does not seem feasible." "There is also the possibility with constraints for system owners to have their own private drive on the side for their own analytics."	

4.3.4 Evaluating the Sensitivity classification and Shareability Grading (S4)

Participants responded positively to the sensitivity classification and shareability grading (see Table 4.4). The strongest support concerned the need for a scalable baseline classification of control system data, rather than relying on repeated ad hoc judgements for every release. Participants generally expressed that classification at the subsystem level is a reasonable starting point. At the same time, managerial feedback highlighted the trade-off between simplicity and precision: classifying data at a higher level may make implementation easier, but risks missing important sensitivity details of specific subsystems or datasets.

Information security stakeholders particularly valued the addition of a separate shareability grading. Their feedback confirmed that release decisions cannot be based on confidentiality and security level alone. A dataset may be sensitive at one point in time but

become shareable later, for example after an embargo period, after operational circumstances have changed, or after sufficient transformation has reduced the risk. This supports the proposed distinction between sensitivity and shareability: sensitivity describes the risk characteristics of the data, while shareability describes whether the data can be released under a specific context, recipient, time, and set of conditions.

The participants also emphasised that such classifications should not be treated as static. They noted that data considered unsuitable may become less sensitive later, for example after an embargo period or when organisational conditions have changed. They suggested that classifications should be revisited at reasonable intervals, or when the context of use changes, rather than being regarded as permanent once assigned.

Finally, information security stakeholders saw clear value in retaining a dataset-level review before release. They highlighted that sensitivity cannot always be determined solely from the classification of the originating subsystem, since inference risk may arise when multiple variables are combined or when data is aggregated into larger datasets. Future improvements should define a routine for periodic reassessment of sensitivity and shareability classifications, and provide more explicit guidelines for identifying sensitive data in processed datasets that is supported by concrete examples.

Table 4.4: Summary of group sentiments regarding solution 4.

Group	Sentiment
Information Security	Positive. Considered the classification approach relevant; emphasised that shareability should include a time dimension and be reviewed periodically.
"You could have a time component, like confidential, that might be shareable after an embargo, because it's only sensitive at one point in time and not another." "I feel that maybe a kind of routine to reclassify or re-rate data sets time by time should be put in place."	
Group	Sentiment
Managerial Representation	Positive. Viewed classification level as a pragmatic trade-off; Noted that higher-level grouping could simplify the process at the cost of precision.
"I can't think of any other way. Maybe to group it on some higher level, just to make it simpler - but then we would lose resolution and detail - no, I have no objections."	
Group	Sentiment
ICS Software	Positive. emphasised the need for classification, framed it as an information security baseline that they believed should already exist organisationally.
"This is information security. This is what the organisation should have, but it is not implemented."	

4.3.5 Evaluating the Release Request Form with Document Handler (S7)

Participants generally viewed the release request form template positively, particularly focusing on its lightweight nature and compatibility with the existing ESS document handling systems (see Table 4.5). Although very positive to the solution candidate, information security stakeholders raised concerns that the form header text could more clearly convey their purpose, and questioned whether a document-focused process could lose the benefits of direct data inclusion. Managerial representatives were positive as well, while emphasising that the form and associated prototype would need to be supported by training and guidance to ensure consistent completion, review, approval and record-keeping.

Overall, responses toward the solution candidate were strongly positive. Participants were particularly pleased with how well the implementation of the form fit the existing document handler and perceived the form as a very lightweight way to facilitate the approval steps and general data sharing process. Future improvements could include splitting 'Background' as a section into two: one for the broader study background, and another describing the dataset characteristics - contents, origin, and other relevant information - as well as by establishing channels of communication that allow inclusion of dataset snippets alongside the document when passing it along.

Table 4.5: Summary of group sentiments regarding solution 7.

Group	Sentiment
Information Security	Very positive. Mentioned it as lightweight and feasible. Noted that a document-based process could be insufficient, unless the form is connected to the dataset or relevant attributes. Felt that part of the form communicated meaning poorly.
"I don't think it's super heavy. Probably it would be put in place quite easily within ESS." "Maybe make explicit the section with characteristics of the data." "There may be a weakness in that it's a document-based process, so maybe you lose some advantages of potentially linking the data or attributes."	
Group	Sentiment
Managerial Representation	Positive. Deemed it a very good fit for existing practices at ESS. emphasised that using the form would require guidance, training, and clear review criteria.
"It fits what we're used to working with... ...it's a rather straightforward solution then." "The headings are fairly on the point." "We cannot launch a template then assume that people will automatically understand how to fill it in." "The important thing will become the material that explains what to fill in under them."	

4.3.6 Evaluating the Upload Template (S8)

Participants viewed the upload template positively and considered it a feasible and reasonable way to structure external uploads and communication to external researchers (see Table 4.6). External representatives viewed the upload template very positively and considered the idea of a structured upload template useful for supporting the external understanding and reuse of shared datasets. Managerial representatives were positive toward the features of the template but expressed confusion regarding details over how it could be implemented in practice on the partner platform.

Overall, participants were positive to the elements of the solution candidate. The majority of the feedback touched upon implementation details of the template rather than the template itself. Examples of the concerns regarded whether code sections and descriptive information for the upload could be kept separate, and how to communicate the contents to researchers that did not have access to the data. Future improvements could be to explicitly split up code sections and explanatory sections when implementing the upload template, and to package the it alongside a brief introduction to the dataset that may be accessed when full access to the data itself is not granted.

Table 4.6: Summary of group sentiments regarding solution 8.

Group	Sentiment
External Representatives	Very positive. Concerns regarded implementation details that were platform specific, such as how to separate code and explanations, and how to communicate with researchers that don't have access to the notebook.
"It could be that you update the code and things like that. But they could potentially live separate lives, the information and the code, if you sort of separate them." "Code has a tendency to sort of live and be updated longer than the text documents." "That makes sense. That's good then."	
Group	Sentiment
Managerial Representation	Slightly positive. Judged feasibility highly and was positive towards the template itself. Confusion regarded practical details. Mentioned that they thought practical issues would show themselves during actual sharing.
"I think also that many of these practical issues now will resolve itself when we see it in the next day of validating this, when you actually do it." "Now I see the result of misunderstanding, because I thought that the Jupyter Notebook will live in Git."	

Chapter 5
Findings

This chapter synthesizes the study findings in relation to the three research questions. The findings are based on the empirical validation of the solution candidates presented in Chapter 4. Solution candidates that were not empirically validated did not result in findings and are presented as supporting recommendations.

5.1 RQ1: Processes, Roles and Authorisation Steps

RQ1 asked: *What processes, roles and authorisation steps are needed for sharing data generated by the ESS control system?*

5.1.1 Layered Data Sharing Process with Roles and Responsibilities (S1; P1, P3)

Practitioners perceived the layered data sharing process as highly feasible, clear and sufficiently lightweight. However, the validation also showed that the workflow requires further implementation detail. Practitioners were uncertain about who should act as approvers, which criteria they should use, and how the process should handle externally initiated data requests. Findings include:

- *The layered sharing process based on functional roles is a feasible and relevant way to enable repeatable control system data sharing at ESS.*
- *The practical use of the process depends on mapping the functional roles to actual personnel, defining approval criteria, and handling data requests.*

5.1.2 A Comprehensive Authorisation Scheme (S2; P1, P2)

The comprehensive authorisation scheme received mixed to positive support. Practitioners saw the need to include multiple perspectives in release decisions, but argued that repeated broad review could become heavy unless a higher-level policy clarifies the organisation's general position on data sharing. The scheme therefore accommodates varied attitudes toward sharing, but does not resolve them by itself. Findings include:

- *The authorisation scheme is useful for structuring approval of external data sharing at ESS while data sharing maturity remains low.*
- *It should be complemented by a data policy, as well as approval criteria based on dataset characteristics.*

5.1.3 Adopting a Data Governance Structure (S3; P3)

The proposed data governance structure was not perceived as feasible in its current form by practitioners. Although practitioners recognized the need for clearer responsibility, most considered the proposed division into data stewards, data responsible, and a data council unsuitable for immediate adoption at ESS. Findings include:

- *The data governance model is not suitable as an immediate implementation at ESS.*
- *Data governance should first be approached through broader policy rather than new governance roles.*
- *Detailed governance may become appropriate later, once data sharing practices have matured.*

5.2 RQ2: Identifying and Handling Sensitive Data

RQ2 asked: *How can sensitive data be identified and handled?*

5.2.1 System and Dataset-Level Sensitivity and Shareability Grading (S4; P4)

The sensitivity classification and shareability grading was strongly supported. Practitioners considered subsystem-level classification a practical baseline because individual PV-level classification would be unrealistic, while classifying all control system data at one level would be too imprecise. The distinction between sensitivity and shareability was also supported. A dataset may be sensitive but still shareable under specific restrictions, or non-shareable today but shareable later. Practitioners therefore emphasised the need for embargo periods and periodic reassessment.

Mandatory dataset-level review was also seen as needed before release, since metadata, variable combinations, timestamps, logs, identifiers, or contextual information may change the risk profile of the inherited subsystem classification. Findings include:

- *The sensitivity classification approach combining a baseline for data produced by subsystems with a mandatory dataset-level release review is a good approach to sensitive data handling.*
- *Sensitive data present in datasets should be handled through appropriate release conditions such as minimisation, aggregation, masking, access restrictions, and escalation when risk remains unclear.*
- *Employing a shareability grading is a good way to incorporate perspectives beyond immediate sensitivity of data produced by a subsystem.*
- *Sensitivity and shareability classifications should not be treated as static and should be revisited over time.*

5.2.2 Saving System Classification Status and Data Sharing Processes (S5; P5)

The recommendation to maintain registers for subsystem classifications and shared datasets was not empirically validated. However, it follows from the validated need for repeatability, traceability, and reuse of prior decisions. Our supporting recommendation is to keep records of sensitivity classifications and shared datasets, including rationale, approval history, transformations, access conditions, recipients and dates.

5.2.3 Investigation into Custom Data Licenses (S6; P6)

The recommendation to investigate custom legal terms or a tailored terms-of-agreement structure was not empirically validated. Nevertheless, the underlying problem was identified during the study: existing open-data licenses are not necessarily well suited to controlled sharing of datasets. Our supporting recommendation is for such an investigation to be conducted to ensure sharing of control system data is supported by legal terms adapted specifically to the relationship between a large-scale, publicly funded research institutes and external researchers.

5.3 RQ3: Technical Support for Data Sharing

RQ3 asked: *What technical solutions and software tools can streamline the identification and handling of sensitive data (RQ2), as well as authorisation and sharing of ESS control system data (RQ1)?*

5.3.1 Release Request Form with Document Handler (S7; P7)

The release request form was strongly supported by practitioners. Practitioners considered it feasible because it could fit into existing document-handling practices at ESS and provide a shared reference point for coordinating review and approval. Practitioners also emphasised that the form would require supporting guidance, examples, and review instructions to be used consistently. Findings include:

- *The release request form, when used within a document handler, is a lightweight, practical tool for supporting internal data sharing workflows.*
- *The practical use depends on clear guidance, reviewer instructions, and routines for storing completed forms and decisions.*

5.3.2 Upload Template (S8; P8)

The upload template was perceived positively by the practitioners. It was seen as useful for making shared datasets understandable and reusable by external researchers, especially where internal control system conventions, processing steps, or dataset limitations would otherwise be unclear. Practitioners further noted that descriptive information and executable code may have different lifetimes. This suggests that the upload template should separate stable explanatory documentation from runnable examples where appropriate. Findings include:

- *The upload template improves the usability of externally shared datasets by explaining purpose, context, processing, internal conventions, and example use.*
- *To support long-term maintainability, descriptive documentation should be separated from executable code where appropriate.*

5.3.3 Pre-processing and Transformation Scripts (S9; P9)

The recommendation for pre-processing and transformation scripts was not empirically validated. However, the problem it addresses remains practically important, since raw control system data is not automatically suitable for external research use. Our supporting recommendation is that development of such scripts should be considered at ESS so research-ready datasets can be made more readily available.

Chapter 6
Discussion

This chapter interprets the findings of the study and discusses their implications. The discussion is structured around research relevance, rigour and novelty, implications for research and practice, scope of validity, and validity considerations.

6.1 Research Relevance, Rigour and Novelty

The research relevance is evaluated based on the extent to which the solution candidates address established data sharing challenges. Rigour is judged according to theoretical support of the solution candidates in prior literature, and novelty regards how unique the suggestions are in comparison to other solutions in research.

The problem of collaborating on data in low data-oriented maturity environments and balancing varied attitudes is a well-established challenge in the literature, demonstrating high research relevance for the solution candidates tackling the problem. Linåker et al. [26] describe difficulties on collaborating around shared data, and Runeson et al. [25] highlight the issue when describing obstacles toward ODEs. Similarly, obstacles for establishing reliable data governance are well-documented by Otto's [27] description of the problems facing governing corporate data and Källström's [12] prior work at ESS. Kowalczyk et al. [9] document concerns regarding data sensitivity from the perspectives of research ethics, inter-organisational relationships and aggregated data sensitivity, demonstrating the research relevance of the solution candidates for handling sensitive data as well. Furthermore, the poor fit of existing licenses for research organisations such as ESS underscores the the need for tailored agreements and licensing approaches. In contrast, the solution candidates aimed at technical streamlining show lower research relevance, as they primarily build on problems identified within ESS and during the study.

For the workflow and the authorisation scheme, the rigour for the solution candidates is supported by the RACI model for structuring responsibility across shared resources [31]. The sharer is 'Responsible' for the sharing request. Approvers are the 'Accountable' par-

ties in the process, with those in reviewing capacity fulfilling the 'Consulted' role. The managerial decision ensures necessary parties are 'Informed'. In addition, the governance structure is supported by RACI as well - though with 'Responsible' and 'Consulted' parties covered by the data responsible and the data steward respectively, while 'Accountable' and 'Informed' is satisfied through the data council. The data governance structure furthermore addresses Otto's core questions for structuring decisions, defining roles and how they participate in regards to data handling [27]. The framework altogether supports alignment with the FAIR principles, by enabling data to be managed and shared in a systematic, reusable manner [8]. The NIST publications informed different parts of the framework. SP 800-82r3 [23] was most directly relevant to S4, as it concerns operational technology and industrial control system security. SP 800-53 [30] was used more indirectly to support the need for review, approval, traceability, and separation of responsibilities. SP 800-122 [24] informed the dataset-level review by highlighting privacy risks related to identifiers, logs, free-text fields, and metadata combinations. The baseline approach with dataset-level review is supported by Ostrom's [26] design principles for ODEs, by focusing on lower barriers for safe sharing and re-use, as well as local and scalable approaches to larger systems. The release request is supported by the five safes framework that guided its contents [43]. The upload template was largely informed by prior internal work and had more limited theoretical grounding.

When applicable, solution candidates were evaluated against possible alternatives, strengthening rigour. The process and authorisation scheme considered organisation role-bound implementations and placing greater reliance on managerial authority. Alternative governance structures with more granular role definitions were considered, and direct communication-based mechanisms were weighed against the advantages of the document-handler approach for the release form.

The novelty of the framework is primarily configurational, combining established concepts into a context-specific solution. Concepts such as data governance, sensitivity classification, and controlled access are well established in prior literature. However, we found limited prior research describing concrete workflows for enabling data sharing. Existing work on data sharing often focuses on principles, incentives and general governance challenges, while practical processes tend to be described in less detail. As such, the novelty stems from empirically grounding the configurational approaches within a real operational large-scale research facilities such as ESS.

6.2 Implications for Research

A major research implication from the findings is that data sharing requires more than storage and access mechanisms. Barriers such as unclear responsibilities, varied attitudes and risk perceptions, and limited policy support show that data sharing must be examined as an organisational challenge, rather than a purely technical one. The findings suggest that, within the ESS context, organisational barriers rather than technical limitations are the primary obstacle to data sharing.

The study highlights how theoretically robust governance structures may be perceived as infeasible, particularly when roles cannot be readily mapped to existing practice. Several factors likely contributed to this at ESS. The organisation's prior focus on acquisition

and engineering perspectives may have reinforced a tendency to view all control system data as a single, unified data lake. In addition, as a large, multi-disciplinary research organisation, ESS may tend toward top-down policy governance rather than data-oriented management, leading practitioners to prioritise policy as the solution to governance rather than structural role assignment. Further, the model was inherently cross-disciplinary, requiring coordination across system owners, information security and management. These groups may differ considerably in priorities and authority structures. Finally, an important contributing factor appeared to be the perceived lack of actionable, transitional pathways. A recurring finding across the study was how essential it is for solution candidates to be grounded in practitioner context in both language and function. For S3, the grounding appeared to be lacking in one major way: although practitioners could understand how the governance structure would function once established, they were unable to readily conceptualise how it would be adopted. The distinction between believing a solution could be realistically adopted, and understanding how it would work once established, is an important one. It suggests that future governance proposals in similar contexts should put greater focus on incremental transition mechanisms alongside the governance structure.

A research implication in the context of ODEs is that the study may strengthen WARA-Ops as an ODE by furthering ESS connection to other ecosystem actors. By providing practical approaches and tools, the study also contributes to research on open collaboration and toward research on open data sharing practices.

6.3 Implications for Practice

A major practical implication is the need to treat data sharing as a wider organisational decision, rather than an individual engineering task. Framing data sharing through functional roles made it clear which positions they might hold in the process, while presenting the authorisation scheme with clear responsibilities demonstrated how the framework could reduce reliance on ad hoc approval approaches.

Another practical implication is that repeatable data sharing should start with lightweight solutions aligned with practitioners' conceptualisations. The role-based workflow, release request form and upload template were perceived as feasible because they were low in technical debt, fit existing ways of working, and provided a clear starting point for sharing.

The findings further suggest that examples, prototypes and dedicated resources are essential enablers when introducing new workflows or methods. Practitioner confidence in the solution candidates increased greatly when they could envision concrete action plans and discuss specific individuals to assume responsibilities. Future work should therefore emphasise practical implementation examples of both workflows and role-to-individual mappings.

Another practical implication is the need for a broader policy to underpin organisational attitude and stance toward data-oriented decisions such as sharing. While practitioners understood how authorisation scheme structured the approval steps, they expressed caution to authorisation without a facility-wide policy describing approval criteria and organisational intent. Clear organisational policy appeared to be crucial for building practitioner confidence and legitimising data sharing.

6.4 Scope of Validity

To determine the scope of validity, the solution candidates were abstracted to the highest level possible without losing empirical or logical grounding. For this thesis, this is done by assessing the general applicability of the solution candidates beyond the ESS context.

For RQ1, the solution candidates rely little on organisation-specific roles and demonstrate broad applicability. The sharing workflow specifies functional roles, and responsibilities and disciplines are emphasised over explicit roles for the positions in the authorisation scheme. In contrast, the data governance solution candidate was tailored to resolve responsibility ambiguities at ESS, making it more context-specific and less generalisable.

For RQ2, the sensitivity classification cannot be abstracted fully beyond ESS, as it builds directly on existing classifications. However, the shareability grading addressing additional factors may also be applicable to other low-maturity data sharing environments. While dataset and system classification-registers address specific ESS needs and therefore have limited scope of validity, the development of a tailored terms of agreement or license for large-scale research organisations is broadly applicable to any organisation producing operational data.

For RQ3, the release request form and upload template are broadly applicable with wide scopes of validity, applying to any organisation seeking to share potentially sensitive data. They cover the essential parts of both the internal workflow and the external data upload. The sub-artifact prototypes, however, are far more narrow in scope, due to the ESS-specific adaptations. While the recommendations for transformation and conversion scripts themselves likely have minimal applicability outside ESS, their necessity illustrates how small technical gaps, perceived as too minor to warrant dedicated solutions, can persist as overlooked blockers.

Overall, the framework is most applicable to organisations similar to ESS, including large research infrastructures, industrial research facilities, or technically complex organisations, that generate control system data. It is less directly applicable to organisations with already mature data governance, where formal stewardship roles, metadata catalogues, and approval procedures may already exist. The framework is also less relevant for small organisations where formalised review structures would add unnecessary overhead.

6.5 Validity Considerations

Several validity considerations affect how the results of this thesis should be interpreted. This study evaluates validity using a taxonomy for validity classification that is suited to software engineering case studies, summarised and operationalised toward flexible design studies by Runeson and Höst [44]. It distinguishes between construct validity, internal validity, external validity, and reliability, each addressing a different aspect of the trustworthiness of the study findings.

6.5.1 Construct Validity

Construct validity concerns whether the concepts being studied are interpreted consistently by participants and researchers, and whether the collected data accurately reflects the intended constructs. In this study, construct validity primarily concerns whether the solution candidates were consistently understood during empirical validation. Several central concepts, such as data governance, sensitivity classification, and data management, were unfamiliar to practitioners. It is possible that some feedback reflects communication challenges in getting these concepts across, or that participants misunderstood the scope or implications of the proposed solutions. Construct validity is further challenged by the fact that the ESS approach to handling control system data previously was structured around engineering and operational domains, focusing mainly on acquisition, whereas the framework often adopts a more data oriented perspective. To mitigate the validity concerns, visual material, concrete examples, and opportunities for clarification were provided, though future work should aim to strengthen validation through pilot cases, hands-on walkthroughs, or extended evaluation sessions.

A further validity concern relates to the format of the empirical validation. The solution candidates were presented by the authors using slides and scripted explanations before participants provided feedback. This created consistency across validation sessions, but may also have introduced social desirability bias or researcher expectancy effects, where participants could be more inclined to agree with the proposed solutions out of politeness or because the concepts were framed by the researchers. This risk was mitigated by explicitly inviting critique and by retaining negative and mixed feedback in the analysis.

6.5.2 Internal Validity

Internal validity concerns whether the findings can be attributed to the factors being studied rather than to not accounted for influences. It regards the causal effect of different factors in the experiment. Internal validity is limited because the framework was not implemented and evaluated in operation. While the empirical validation assessed perceived relevance, feasibility and organisational fit, the framework's effectiveness cannot be fully validated without practical implementation. Perceived feasibility depends on several factors beyond the study scope, and it is infeasible to understand it fully without post-implementation evaluation. Further evaluation following implementation would be required to determine to what degree the framework reduces uncertainty and improves data sharing.

The findings are further limited by the study scope. Solution candidates 5, 6, and 9 were omitted from the empirical validation to focus on critical parts of the framework. While they address relevant needs, an evaluation is required before they can be considered empirically grounded findings rather than supporting recommendations. Another limitation is that the study covered only a single iteration of the design science cycle. Additional iterations would have enabled deeper exploration of alternatives to weaker solution candidates, such as solution candidate 3, further refinement of the solution candidates, and extensions to the framework based on practitioner feedback during validation.

6.5.3 External Validity

External validity regards the generalisability of the study findings and how it aligns with researcher interests outside the case. For this thesis, external validity is limited by the ESS and WARA-Ops context. While the study discusses abstraction and scope of validity, restricting the validation environment to these two actors limits the extent to which the findings analytically can be generalised to other contexts. Future evaluations in comparable contexts would strengthen the external validity. Nevertheless, the evaluation benefited from close access to field experts with in-house validation from practitioners, which strengthened the practical relevance of the findings [36].

External validity is also limited by the relatively small interview sample size, and the reliance on qualitative evaluation. The approach reflected the limited number of practitioners with sufficient domain knowledge to meaningfully evaluate the proposed framework, but influences generalisability nonetheless. Future evaluations involving a broader base of participants combining quantitative evidence from operational use would strengthen external validity greatly.

6.5.4 Reliability

Reliability concerns the extent to which the study results depend on the individual researchers conducting the study. In order to improve reliability, both the empirical validation and problem conceptualisation phases followed consistent procedures across all participants. The same problem framing, solution descriptions, and supporting material were used throughout sessions, and participants were provided opportunities to ask questions for clarity when necessary. Feedback from participants reflecting multiple disciplines was taken into account, with both positive and critical viewpoints considered during the analysis, in order to mitigate the effect of researcher bias. While such bias cannot be fully removed when conducting qualitative analysis, these approaches helped improve the consistency and repeatability of the study.

Chapter 7
Conclusion and Future Work

This chapter outlines the conclusions drawn from this study. The conclusions are broadly summarised along the research questions, before they are described in greater detail in regards to the solution candidates as well. Finally, four potential avenues for future work are described.

RQ1. Sharing data generated by the ESS control system requires a repeatable workflow with functional roles, a broad authorisation scheme, and clear approval criteria.
RQ2. Sensitive data can be identified by combining baseline classification of data generated by systems or subsystems with mandatory dataset-level review before release.
RQ3. Technical support should be lightweight and integrated with existing organisational practices. A release request form in an existing document handler can coordinate internal review and approval, while a standardised upload template can make datasets more understandable and reusable for external researchers.

7.1 Conclusions

The conclusions are drawn from the study's findings and subsequent analysis.

- **Data sharing workflow with functional roles**. Structuring the sharing workflow around functional roles with clearly defined responsibilities representing actors in the sharing process is a suitable approach for organisations with low data sharing maturity. *See solution 1*.
- **Comprehensive authorisation supported by policy**. Employing a broad authorisation scheme is a practical way to accommodate varied attitudes toward data sharing and risk in low data sharing maturity environments but should be complemented with policy defining approval criteria and organisational intent. *See solution 2*.

- **Introducing data governance structures into organisations can be challenging**. New governance roles and decision bodies are difficult to adopt, where existing responsibilities, authority structures, and ways of working are already established. *See solution 3*.
- **Baseline sensitivity ratings for subsystem classifications combined with complementary release reviews.** Baseline sensitivity classification of control system data complemented by dataset-level reviews is a suitable way of handling sensitive data. *See solution 4*.
- **Shareability gradings for control system data**. Additional shareability categorisation provide useful complementary perspectives beyond sensitivity and confidentiality classifications for control system data. *See solution 4*.
- **Standardised templates for internal and external communication**. Standardised release forms and upload templates are practical tools for improving internal and external workflows. *See solutions 7, 8*.
- **Lightweight solutions grounded in organisational contexts**. Lightweight solutions and technical mechanisms that are easy for practitioners to contextualise within their working environment increase confidence and acceptance while avoiding large-scale organisational change. *See solutions 1, 7, 8*.

7.2 Future Work

This section outlines potential avenues for future work based on the findings from this master's thesis.

Piloting the Framework

Future work should first focus on piloting the framework in real data sharing cases. A pilot should include a complete release request, approval workflow, dataset-level sensitivity review, upload template, and registration of the final decision. Such a pilot would allow ESS to evaluate the framework beyond perceived feasibility and assess concrete outcomes such as time spent, clarity of responsibilities, and approval bottlenecks.

Approval Criteria

Another avenue for future work is the development of approval criteria. The empirical validation showed that practitioners understood the need for approvers but remained uncertain about the basis for approval. Future work should therefore define criteria for approving, rejecting, revising, or escalating sharing cases. These criteria could be structured around factors such as purpose, recipient, dataset characteristics, classification, data and metadata quality, legal constraints, inference risks, and expected organisational value.

Handling Pull Scenarios

The validation highlighted practitioner's interest for supporting external data requests to ESS. Future work could therefore extend the workflow to accommodate for such scenarios. This would likely only require minor additions, for example by allowing administrators to delegate sharing responsibility based on the requested data type.

Data Sharing Policy

A recurring theme in the validation interviews was the absence of a clear policy for external sharing of ICS data. Practitioners described this as a central barrier rather than a minor supporting issue. The organisational position on whether, why, and under which conditions the data should be shared appeared to be an essential enabling step if ESS intends to move from ad hoc sharing toward systematic data sharing practices.

References

[1] P. Filinov, A. Lavrentyev, and A. Vorontsov, "Multivariate Industrial Time Series with Cyber-Attack Simulation: Fault Detection Using an LSTM-based Predictive Data Model," Presented at NIPS Time Series Workshop, Barcelona, Spain, 2016. DOI: 10.48550/arXiv.1612.06676. arXiv: 1612.06676.

[2] L. H. Chiang, E. L. Russel, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems* (Advanced Textbooks in Control and Signal Processing). London: Springer-Verlag London Berlin Heidelberg, 2001. DOI: 10.1007/978-1-4471-0347-9.

[3] N. Mitta and R. Ranjan, "AI-Enhanced Predictive Maintenance Systems for Industrial Equipment: Developing Machine Learning Models to Forecast Failures, Optimize Maintenance Schedules, and Minimize Downtime," *Distributed Learning and Broad Applications in Scientific Research*, vol. 10, pp. 550–591, Jan. 2024.

[4] A. A. Daya and I. Lazakis, "Systems Reliability and Data Driven Analysis for Marine Machinery Maintenance Planning and Decision Making," *Machines 2024, Vol. 12, Page 294*, vol. 12, no. 5, p. 294, Apr. 2024, ISSN: 2075-1702. DOI: 10.3390/MACHINES12050294.

[5] Z. Li, Q. He, and J. Li, "A survey of deep learning-driven architecture for predictive maintenance," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108 285, Jul. 2024, ISSN: 0952-1976. DOI: 10.1016/J.ENGAPPAI.2024.108285.

[6] S. W. Mogensen, K. Rathsman, and P. Nilsson, "Causal discovery in a complex industrial system: A time series benchmark," in *Proceedings of Machine Learning Research*, F. Locatello and V. Didelez, Eds., vol. 236, 2024, pp. 1218–1236. DOI: 10.48550/arXiv.2310.18654.

[7] A. Lavin and S. Ahmad, "Evaluating real-time anomaly detection algorithms – the numenta anomaly benchmark," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, 2015, pp. 38–44. DOI: 10.1109/ICMLA.2015.141.

[8] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al., "The fair guiding principles for scientific data management and stewardship," *Scientific Data*, vol. 3, no. 1, p. 160 018, Mar. 2016, ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18.

[9] S. Kowalczyk and K. Shankar, "Data sharing in the sciences," *Annual Review of Information Science and Technology*, vol. 45, no. 1, pp. 247–294, Jan. 2011, ISSN: 1550-8382. DOI: 10.1002/ARIS.2011.1440450113.

[10] P. Runeson and E. Söderberg, "Tools and Ecosystems for Open Control Systems Data at ESS," Lund University, Tech. Rep., Jan. 2021. [Online]. Available: https://portal.research.lu.se/en/publications/tools-and-ecosystems-for-open-control-systems-data-at-ess/.

[11] P. Andersson, J. E. Larsson, and K. Rathsman, "Control System Data and Meta Data at ESS," Lunds Universitet/Lunds Tekniska Högskola, Tech. Rep. 107, 2021.

[12] C. Källström, "Data Quality and Quantity for Machine Learning at the European Spallation Source," M.S. thesis, Lund University, 2025.

[13] T. Friedrich, "Overview of the Integrated Control System - First Issue," European Spallation Source, Tech. Rep. ESS-0297798, 2018.

[14] European Spallation Source, "ESS CHARTER FOR ESS OPERATIONS DATA STEERING BOARD (Internal Document)," European Spallation Source, Lund, Sweden, Unpublished Internal Document, 2025.

[15] Q. Zhu, Y. Ding, J. Jiang, and S. H. Yang, "Anomaly detection using invariant rules in Industrial Control Systems," *Control Engineering Practice*, vol. 154, p. 106 164, Jan. 2025, ISSN: 0967-0661. DOI: 10.1016/J.CONENGPRAC.2024.106164.

[16] Z. Zamanzadeh Darban, G. I. Webb, S. Pan, C. Aggarwal, and M. Salehi, "Deep Learning for Time Series Anomaly Detection: A Survey," *ACM Computing Surveys*, vol. 57, no. 1, p. 42, Oct. 2024, ISSN: 15577341. DOI: 10.1145/3691338.

[17] D. Kafkes and J. St. John, "BOOSTR: A Dataset for Accelerator Control Systems," *Data 2021, Vol. 6, Page 42*, vol. 6, no. 4, p. 42, Apr. 2021, ISSN: 2306-5729. DOI: 10.3390/DATA6040042.

[18] M. Jagals and E. Karger, "Inter-Organizational Data Governance: A Literature Review," *ECIS 2021 Research Papers*, Jun. 2021. [Online]. Available: https://aisel.aisnet.org/ecis2021_rp/57.

[19] European Parliament and Council of the European Union, *Regulation (EU) 2016/679: General Data Protection Regulation*, Official Journal of the European Union, L 119, 4 May 2016, pp. 1–88, 2016. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[20] K. Malysh, T. Ahmed, J. Linaker, and P. Runeson, "Inter-Organizational Data Sharing Processes - An Exploratory Analysis of Incentives and Challenges," in *Proceedings of the Euromicro Conference on Software Engineering and Advanced Applications, EUROMICRO-SEAA*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 80–87, ISBN: 9798350380262. DOI: 10.1109/SEAA64295.2024.00021.

[21] M. Kim, T. Zimmermann, R. Deline, and A. Begel, "Data scientists in software teams: State of the art and challenges," *IEEE Transactions on Software Engineering*, vol. 44, no. 11, pp. 1024–1038, Nov. 2018, ISSN: 19393520. DOI: 10.1109/TSE.2017.2754374.

[22] A. R. Munappy, J. Bosch, H. H. Olsson, A. Arpteg, and B. Brinne, "Data management for production quality deep learning models: Challenges and solutions," *Journal of Systems and Software*, vol. 191, p. 111 359, Sep. 2022, ISSN: 0164-1212. DOI: 10.1016/J.JSS.2022.111359.

[23] K. Stouffer et al., "NIST Special Publication NIST SP 800-82r3 Guide to Operational Technology (OT) Security," National Institute of Standards and Technology, Tech. Rep., Sep. 2023. DOI: 10.6028/NIST.SP.800-82r3.

[24] E. Mccallister, T. Grance, and K. Scarfone, "Guide to Protecting the Confidentiality of Personally Identifiable Information (PII) Recommendations of the National Institute of Standards and Technology," National Institute of Standards and Technology, Tech. Rep., Apr. 2010. DOI: 10.6028/NIST.SP.800-122.

[25] P. Runeson, T. Olsson, and J. Linåker, "Open Data Ecosystems - an empirical investigation into an emerging industry collaboration concept," *Journal of Systems and Software*, vol. 182, p. 111 088, Dec. 2021, ISSN: 0164-1212. DOI: 10.1016/J.JSS.2021.111088.

[26] J. Linåker and P. Runeson, "Sustaining Open Data as a Digital Common – Design principles for Common Pool Resources applied to Open Data Ecosystems," in *Proceedings of the 18th International Symposium on Open Collaboration*, Aug. 2022. DOI: 10.1145/3555051.3555066.

[27] B. Otto, "Data governance," *Business and Information Systems Engineering*, vol. 3, no. 4, pp. 241–244, Aug. 2011, ISSN: 18670202. DOI: 10.1007/s12599-011-0162-8.

[28] J. Ladley, *Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program*, 2nd ed. Academic Press, 2020. DOI: 10.1016/C2017-0-03353-0.

[29] D. Plotkin, *Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance*, 2nd ed. Academic Press, 2021. DOI: 10.1016/C2019-0-03988-X.

[30] U.S. Department of Commerce, "NIST Special Publication 800-53 Revision 5 Security and Privacy Controls for Information Systems and Organizations JOINT TASK FORCE," National Institute of Standards and Technology, Tech. Rep., Sep. 2020. DOI: 10.6028/NIST.SP.800-53r5.

[31] Project Management Institute, *A Guide to the Project Management Body of Knowledge (PMBOK®)*, 5th. Project Management Institute, 2013, pp. 261–262.

[32] J. E. Van Aken, "Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules," *Journal of Management Studies*, vol. 41, no. 2, pp. 219–246, Mar. 2004, ISSN: 1467-6486. DOI: 10.1111/J.1467-6486.2004.00430.X.

[33] E. Engström, M. A. Storey, P. Runeson, M. Höst, and M. T. Baldassarre, "How software engineering research aligns with design science: a review," *Empirical Software Engineering*, vol. 25, no. 4, pp. 2630–2660, Jul. 2020, ISSN: 15737616. DOI: 10.1007/s10664-020-09818-7.

[34] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quarterly*, vol. 28, pp. 75–106, Mar. 2004. DOI: doi/10.5555/2017212.2017217.

[35] V. K. Vaishnavi and W. Kuechler, *Design Science Research Methods and Patterns : Innovating Information and Communication Technology, 2nd Edition*. CRC Press, May 2015, ISBN: 9780429172205. DOI: 10.1201/B18448.

[36] P. Runeson, E. Engström, and M.-A. Storey, "The Design Science Paradigm as a Frame for Empirical Software Engineering," in *Contemporary Empirical Methods in Software Engineering*, Springer, Cham, 2020, pp. 127–147, ISBN: 978-3-030-32489-6. DOI: 10.1007/978-3-030-32489-6_5.

[37] M.-A. Storey, E. Engstrom, M. Höst, P. Runeson, and E. Bjarnason, "Using a visual abstract as a lens for communicating and promoting design science research in software engineering," in *2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2017, pp. 181–186. DOI: 10.1109/ESEM.2017.28.

[38] E. Knauss, "Constructive Master's Thesis Work in Industry: Guidelines for Applying Design Science Research," in *Proceedings - International Conference on Software Engineering*, IEEE Computer Society, Dec. 2021, pp. 110–121, ISBN: 9780738133201. DOI: 10.1109/ICSE-SEET52601.2021.00021.

[39] A. Lantz, *Intervjumetodik*, 3rd ed. Studentlitteratur, 2013, p. 173, ISBN: 9789144081236.

[40] M. Höst, P. Runeson, and B. Regnell, *Att göra examensarbete*, Release v1.0, Accessed 2026-03-16, 2024. [Online]. Available: https://github.com/lunduniversity/thesis-guide/releases/tag/v1.0.

[41] R. Wieringa, "Design science as nested problem solving," in *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, ser. DESRIST '09, Philadelphia, Pennsylvania: Association for Computing Machinery, 2009, ISBN: 9781605584089. DOI: 10.1145/1555619.1555630.

[42] ALLEA, "The European Code of Conduct for Research Integrity - Revised Edition 2023," All European Academies, Berlin, Tech. Rep., Jun. 2023. DOI: 10.26356/ECOC.

[43] A. Bailey et al., *Five Safe Data Access Request application form*, Working paper, Feb. 2022. DOI: 10.5281/zenodo.5946892.

[44] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," *Empirical Software Engineering*, vol. 14, no. 2, pp. 131–164, 2009, ISI Highly cited. DOI: 10.1007/s10664-008-9102-8.

Appendices

Appendix A
Interview Material

Background and Problem Understanding These interviews enabled comprehensive problem understanding, supporting the problem conceptualisation phase of the study. The general question set was asked to security and engineering (see Table A.1). All groups were asked expertise-based questions (see Tables A.3, A.4, A.2, A.5). The interview manuscript was read verbatim to all participants. The interviews were recorded through the enterprise version of the video-call application Zoom[1] used at ESS. A locally run version of the transcription tool Whisper[2] was used, supplemented by manual review and editing.

Introduction
Hello, and thank you for taking the time to meet with us. Our names are Olof Gilland and Kaspian Garpvall, and we are master students in Computer Science at Lund University. We are currently conducting our master thesis at ESS, focusing on data sharing practices and the prospect of sharing selected control system data with external research partners. The purpose of this interview is to understand the range and depth of challenges related to sharing data from the Integrated Control System (ICS). Your answers will help us understand the problem better and shape proposals for processes and technical solutions.
With your consent, we would like to record the interview to ensure accurate transcription. Recordings and transcripts will be stored securely and privately for the duration of the study and recordings will see deletion at the study conclusion.

C1. Do you consent to this interview being recorded and your answers being transcribed and analyzed for our study?

C2. Would it be okay to contact you again later in the study to follow up on any answers?

[1] `https://www.zoom.com/en/products/collaboration-tools/`, accessed on 2026-03-16
[2] `https://github.com/openai/whisper`, accessed on 2026-03-16

Empirical Evaluation
These interviews took place toward the end of our study and allowed for an investigation into our candidate solutions from the perspective of the practitioner. The interview consisted of an introductory segment presenting the problems and respective candidate solutions for the research questions, with a slide-show segment for each. Solutions 5, 6 and 9 were omitted from the validation. After each solution candidate had been presented, the participants were given an opportunity for discussion and feedback. The evaluation was shaped around the solution candidates, in turn composed of the technological rules, which can be found in append B. During the session, participants were given a short printed summary of the different problem-solution pairs to aid in the evaluation. These are provided below.

The interviews were recorded using the enterprise version of the video-call application Zoom, with a locally run version of the transcription tool Whisper, supplement by manual review and editing.

Solution Candidate 1: Layered Data Sharing Process with Roles and Responsibilities

It was found that several problems occur due to insufficient data sharing maturity. These problems were expressed as **P1: Insufficient Data Sharing Maturity**. It is composed of the following problems (P1.1-P1.3).

P1.1 Data sharing is perceived as unclear with no formalised responsibilities.

P1.2 There is no starting point for sharing a dataset.

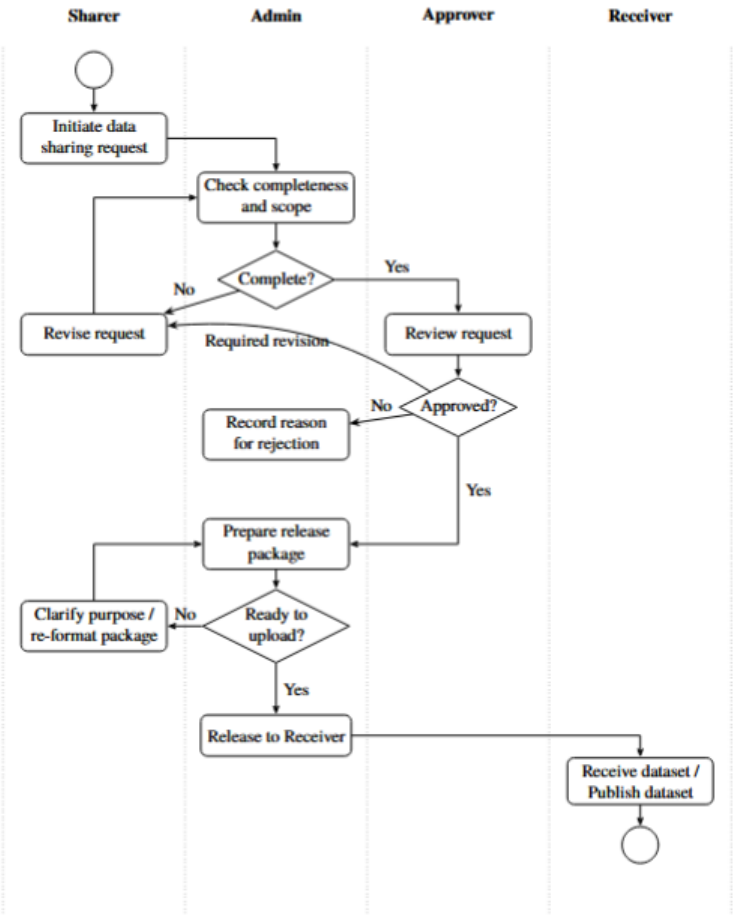
P1.3 Data sharing is done in an ad hoc manner.

We have created a solution candidate to address the problem. It was expressed as **S1: Layered Data Sharing Process with Roles and Responsibilities**. It is composed of the following recommendations (S1.1-S1.3).

S1.1 Define a sharing process with the functional roles of sharer, admin, approver, and receiver, each with distinct responsibilities.

S1.2 Require the sharing party to fill out a release form to pass on for approval.

S1.3 Store decisions and reasoning for data sharing, to use as a base for future decision making.



Solution Candidate 2: A Comprehensive Authorisation Scheme

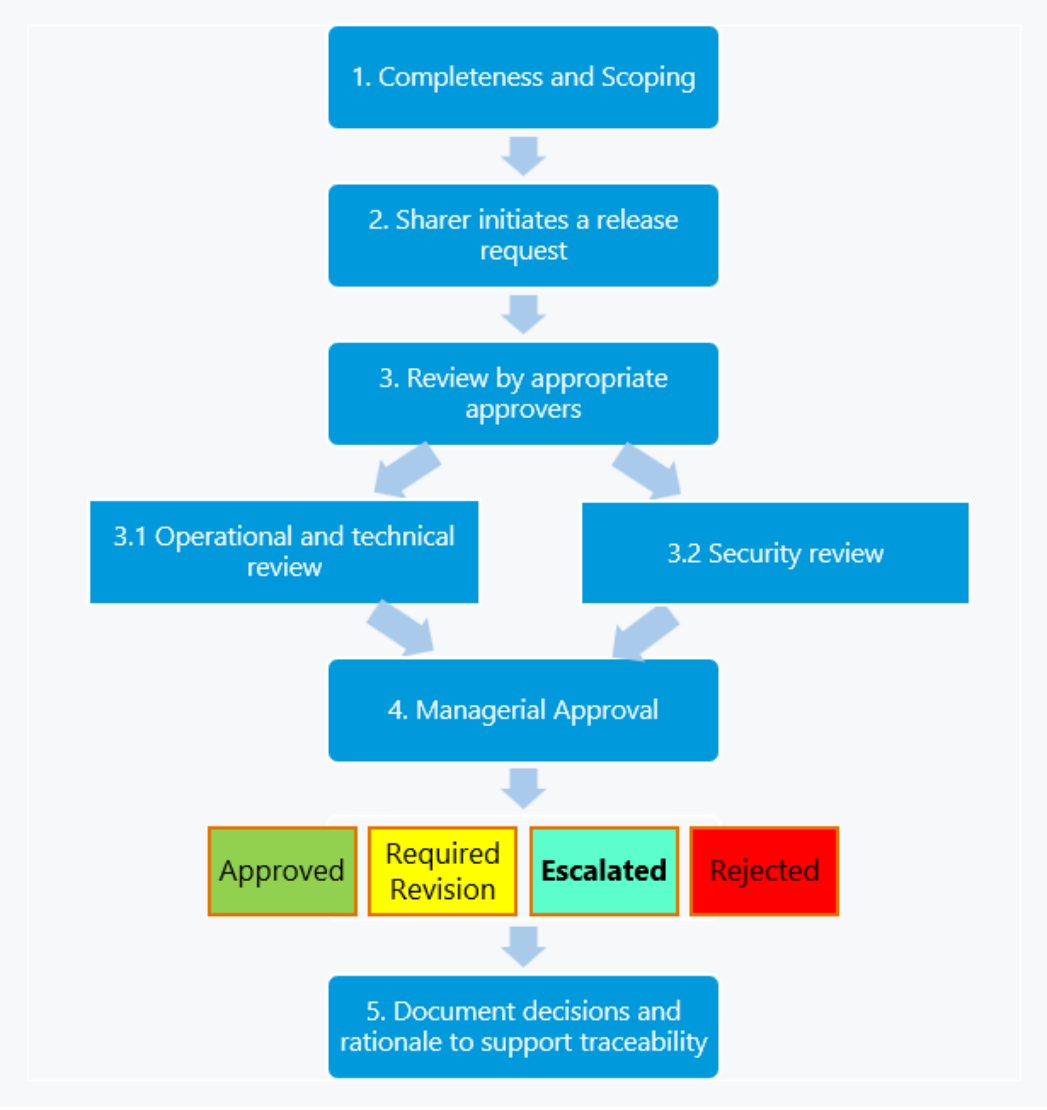
It was found that there are varied attitudes towards data sensitivity and benefits of data sharing between disciplines at ESS. These problems were expressed as **P2: Varied Attitudes Toward Data Sharing**. It is composed of the following problems (P2.1-P2.2).

P2.1 Risk perception varies across disciplines.

P2.2 Opinions differ on the benefits of data sharing.

We have created a solution candidate to address the problem. It was expressed as **S2: A Comprehensive Authorisation Scheme**. It is composed of the following recommendation (S2.1).

S2.1 Introduce an authorisation scheme covering positions in information security, responsible data authority, data stewards, engineering and managerial representation.



Solution Candidate 3: Adopting a Data Governance Structure

It was found that there were inconsistent interpretation of data responsibilities and ambiguities in who held authority over datasets generated by the control system at ESS. These problems were expressed as **P3: Ambiguous Data Responsibility**. It is composed of the following problems (P3.1-P3.2).

P3.1 There is inconsistent interpretation of data responsibility in practice.

P3.2 There is no responsible authority for generated data.

We have created a solution candidate to address the problem. It was expressed as **S3: Adopting a Data Governance Structure**. It is composed of the following recommendations (S3.1-S3.2).

S3.1 Establish data governance using high-level structure of data stewards, data responsible and data council.

S3.2 Formalise existing responsibilities of system owners, data producers, data users, integrators and information security into clear governance roles.

Solution Candidate 4: System and Dataset-Level Sensitivity and Shareability Grading

It was found that there exists sensitivity classifications that are not consistently performed on control system data, that dataset sensitivity depends on more than the data characteristics alone and that data sharing decisions must account for internal research interests and bad faith actors. These problems were expressed as **P4: Organisational Risk and Impact**. It is composed of the following problems (P4.1-P4.3).

P4.1 At ESS mandatory sensitivity classifications are not performed on control system data.

P4.2 Dataset sensitivity depends on more than the data characteristics alone.

P4.3 Data sharing decisions should account for internal research interests and bad faith actors.

We have created a solution candidate to address the problem. It was expressed as **S4: System and Dataset-Level Sensitivity and Shareability Grading**. It is composed of the following recommendations (S4.1-S4.3).

S4.1 Perform a confidentiality and security level grading on the data produced by subsystems with datasets inheriting that classification by default.

S4.2 Dataset-level review with involvement from security representatives.

S4.3 Apply an additional shareability grading (shareable, shareable after embargo, shareable on case-by-case basis, not shareable).

Solution Candidate 7: Release Request Form with Document Handler

It was found that there is a need for lightweight technical solutions to coordinate the internal sharing workflow, and that such a workflow must reliably ensure that no sensitive information is included in the release. The problem was expressed as **P7: Coordinating the Internal Workflow**. It consists of the following problems (P7.1-P7.2).

P7.1 Resource-intensive or complicated solutions will not see implementation.

P7.2 There is difficulty in reliably identifying and excluding sensitive data before datasets are shared.

We have created a solution candidate to address the problem, **S7: Release Request Form with Document Handler**. It consists of the following recommendation (S7.1).

S7.1 A release form within a document handler that supports distributed review.

Solution Candidate 8: Upload Template

It was found that simply uploading data was not enough to drive research interest, and that there exists internal conventions that need to be taken into account before data is shared externally. These problems were expressed as **P8: Sharing Useful Data that Attracts Research**. It is composed of the following problems (P8.1-P8.2).

P8.1 It is difficult to make datasets useable for researchers.

P8.2 Internal conventions must be clarified to external users.

We have created a solution candidate to address the problem, **S8: Upload Template**. It consists of the following recommendations (S8.1-S8.2).

S8.1 Use a standardised upload template including background, purpose, usage guidelines and examples.

S8.2 Consistently describe purpose and background across the dataset release request and the external upload template.

A. Policy and Guiding Documents	
Q1	From your understanding, what is the current policy or principle at ESS regarding sharing ICS data externally?
Q2	Do you know which guiding documents, policies, or agreements govern data sharing at ESS (e.g., internal policies, user agreements, contracts)?
Q3	Have you ever shared ICS data externally, or wanted to share it? If yes, what type of data and what was the context?
B. Governance: who can approve sharing?	
Q4	In practice, who is considered the "data owner" for control system data?
Q5	If ownership is ambiguous or shared across groups, how is that typically handled?
Q6	If you wanted to share ICS data, what would the approval workflow usually look like? *Follow-up: Which roles/functions typically need to sign off (if any)?*
C. Sensitive data and risk	
Q7	How do you determine whether a specific dataset (or part of a dataset) is sensitive? *Follow-up: Who would you contact to assess sensitivity (roles/teams)?*
Q8	Who ultimately decides whether the data is sensitive or shareable?
Q9	What potential security concerns do you see with sharing ICS data externally?
Q10	Do different types of data pose different risks (e.g., metadata/logs vs timeseries, cryogenics, safety-critical systems)?
Q11	What is the worst-case scenario you see if ICS data ends up in the wrong hands?
D. Friction Ranking	
Q12	From your perspective, what are the main obstacles to taking a control system dataset and sharing it with external researchers (e.g., WARA-Ops)?
Q13	Which obstacle could realistically be addressed the easiest with a small change?
Q14	Which obstacle do you consider a hard "no" unless a policy and/or governance decision changes?
E. What would a good outcome look like?	
Q15	For you personally, what would be a beneficial outcome of our project?
Q16	What would make you distrust a proposed solution/process?
F. Closing questions - *These are asked after the role-specific questions.*	
Q17	If you can only pick one: what is the single hardest thing about sharing ESS datasets externally?
Q18	What is the one thing you would need from WARA-Ops (or external partners) to make data sharing feasible?
Q19	What is the one thing ESS should avoid doing, because it will cause trouble?
Q20	Is there anything you would like to add, that wasn't addressed in this interview?
Q21	Are there other people or roles you recommend we interview?

Table A.1: General questions asked to the information security and ICS engineering groups during the semi-structured initial interview study.

Information Security	
S1	Are there legal, contractual, or ethical constraints that apply for "non-personal" control system data?
S2	Are there common cases of sensitive information appearing in logs or metadata (names, emails, proprietary descriptions)?
S3	If a dataset is to be shared externally, what typically needs to be removed or redacted first?

Table A.2: Role-specific questions asked to information security during the semi-structured initial interview study.

Engineering and Operations	
E1	When you say "a dataset" at ICS, what does that mean in practice?
E2	What metadata is guaranteed to exist for every dataset?
E3	What metadata is critical for reuse but is currently missing or inconsistent?
E4	Which metadata is instrument-specific vs facility-wide?
E5	Are there internal conventions that external users would not understand?
E6	What are known failure modes in the data, and how are they annotated currently? This could mean bad calibrations, missing logs, downtime artifacts or similarly.
E7	Do you have data quality flags? Who sets them and when?
E8	Are there situations where the data itself is safe to share but the metadata is not? If yes, what makes it so? (*Follow up: Is there practice for sanitising the metadata in such scenarios?*)
E9	Can you think of scenarios where it would be possible to share data without metadata or vice versa?

Table A.3: Role-specific questions asked to the ICS engineering group during the semi-structured initial interview study.

Legal	
L1	From a legal perspective, what are the main criteria you would use to decide whether a given ICS/machine dataset can be shared externally (and under what conditions)?
L2	Which agreements or legal frameworks most commonly constrain data sharing at ESS?
L3	In our understanding, licensing terms govern parts of the ICS system. In practice, what should we verify when a dataset depends on a specific component/interface?
L4	How do aggregation, filtering, sampling, anonymization, or delayed release change the legal risk profile? *Follow up: Are there transformations that are typically considered meaningful risk mitigations, and what documentation would you expect to justify them?*
L5	If ESS wanted a standardised, repeatable process for external sharing, which roles should be involved in the assessment and sign-off (e.g., Legal, ICS owners, Security, Safety, Data Protection, Export control, Procurement)?
L6	Is there anything you would like to add?
L7	Based on what we discussed, do you think there is someone else we should talk to?

Table A.4: Role-specific questions asked to participants from legal department during the semi-structured initial interview study.

External Research	
W1	From the perspective of the data receiver, what criteria must be satisfied for data to be shared through your platform?
W2	What types of control system data is most useful to WARA-Ops?
W3	What are the notable use case scenarios for data shared from ESS for external researchers?
W4	As you know, control system data can be sensitive in nature. Do you have any strategies in place to ensure secure handling of shared data? *Follow up: How rigorous are these strategies in preventing bad actors access?*
W5	What data formatting and presentation requirement exist for data to be shared within your platform? (formatting, critical metadata, volume...)
W6	Are there any technical requirements that must be satisfied from an organisation looking to share data? (notebooks, annotated datasets, explanatory metadata in text...)
W7	Are there challenges with getting users to actually utilize the data? *Follow up: What is the awareness in terms of user needs in regards to control system data generated by a system such as at ESS?*
W8	Is there anything you would like to add?
W9	Based on what we discussed, do you think there is someone else we should talk to?

Table A.5: Role-specific questions asked to the external representatives group during the semi-structured initial interview study.

Appendix B
Technological Rules

In this appendix, the detailed technological sub-rules are specified for each candidate solution.

Layered Data Sharing Process with Roles and Responsibilities

S1.1. **To achieve** clear data sharing with defined responsibilities despite low data sharing maturity **in** data sharing at ESS **apply** a sharing process with the functional roles of sharer, approver, admin and receiver, each with distinct responsibilities.

S1.2. **To achieve** a starting point for initiating sharing **in** data sharing at ESS **apply** requiring the sharing party to fill out a release form to pass on for approval.

S1.3. **To achieve** less ad hoc practices **in** data sharing at ESS **apply** storing decisions and reasoning made during sharing to use as a base for future decision making.

A Comprehensive Authorisation Scheme

S2.1. **To achieve** effective data sharing despite varied attitudes toward risk perception and perceived benefits **in** data sharing at ESS **apply** an authorisation scheme involving information security, responsible data authority, data stewards, engineering, and management, with outcomes of approval, revision, escalation, or rejection.

Adopting a Data Governance Structure

S3.1. **To achieve** consistent interpretation of data responsibility **in** data handling at ESS **apply** establishing data governance using high-level structure of data stewards, data responsible and data council.

S3.2. **To achieve** explicitly assigned authority over data **in** data handling at ESS **apply** formalising existing responsibilities of system owners, data producers, data users,

integrators and information security into clear governance roles.

System and Dataset-Level Sensitivity and Shareability Grading

S4.1. **To achieve** sensitivity classifications on control system data **in** data sharing at ESS **apply** performing confidentiality and security level grading on the data produced by subsystems with datasets inheriting that classification by default.

S4.2. **To achieve** sensitivity decisions that account for how dataset sensitive depends on more than the data characteristics alone **in** data sharing at ESS **apply** dataset-level reviews with involvement from security representatives allowing for possible escalation.

S4.3. **To achieve** sensitivity classifications that accounts for internal research interests and bad faith actors **in** data sharing at ESS **apply** an additional shareability grading to data generated by subsystems with options shareable, shareable after embargo, shareable on case-by-case basis, not shareable.

Saving System Classification Status and Data Sharing Processes

S5.1. **To achieve** storing of sensitivity classifications for data **in** data sharing at ESS **apply** a register that records classifications for data produced by each subsystem along with rationale, personnel that recorded the classification, and timestamps.

S5.2. **To achieve** storing of shared datasets **in** the ICS division at ESS **apply** a register that stores the datasets along with associated communications and process information from the data sharing process.

Investigation into Custom Data Licenses

S6.1. **To achieve** compliant external data sharing that ensures legal requirements are met **in** ESS data sharing **apply** an investigation into custom license solutions or terms-of-agreement for external parties that is tailored to operational control system data.

Release Request Form with Document Handler

S7.1. **To achieve** an easy-to-follow data sharing workflow that ensures sensitive data is not present **in** data sharing at ESS **apply** a release form with purpose, recipients, background, processing steps, sensitivity information, access conditions and sign-off specifics, within a document handler supporting distributed review.

Upload Template

S8.1. **To achieve** improved research usability of externally shared datasets in ESS data sharing, **apply** a standardised upload template including background, purpose, usage guidelines, and examples.

S8.2. **To achieve** clarity of internal conventions for external users in ESS data sharing, **apply** the same description of purpose and background across the dataset release request form and external upload.

Data Pre-processing and Transform Scripts

S9.1. **To achieve** alignment between internal data representations at ESS and external platforms **in** data sharing from ESS to an external platform **apply** conversion scripts that transform internal datasets into industry-standard formats such as Python Pandas, Julia, and R data frames.

S9.2. **To achieve** research-ready combining raw data with required machine metadata **in** data sharing from ESS to an external platform **apply** transformation scripts that merge raw sensor data with instrument metadata.

Appendix C
Empirical Validation Summary

In this appendix, the results from the empirical validation are provided in summarised form. Table C.1 summarises each solution candidate, its validation status, identified issues and recommended future improvements.

SC	Status	Perceived issues	Future recommendations
S1	Well received	• Approver inclusions • Approval criteria • Handling data requests	• List examples of potential approvers for different data at ESS • List approval criteria for datasets • Admin receives data request and delegates potential sharer
S2	Mixed to positive	• Approval criteria • Potentially too heavy • Need for data policy	• List approval criteria for datasets • Investigate top-down control system data policy
S3	Mixed to negative	• No path to put governance in place • Structure is too different from existing perspective • Too detailed for an organisation of ESS's size	• Examine how governance may be integrated with higher-level policy • Create transition governance mechanisms • Define clearer pathways to data governance
S4	Well received	• Need to reclassify data periodically • Greater need for guidance in assessing dataset sensitivity	• Explore how to revisit sensitivity periodically • Implement a detailed checklist containing examples of sensitive information present in different ESS datasets
S5	Not validated	–	–
S6	Not validated	–	–
S7	Very well received	• Requires guidance and examples	• Conduct pilot of a dataset release • Create guidelines for using the document handler with the form
S8	Well received	• Descriptive information and code may have different lifetimes	• Split up stable descriptive information and code examples in implementation of the template
S9	Not validated	–	–

Table C.1: Overview of the empirical validation results in regards to the solution candidates (SC)

Dela forskningsdata utan att tappa kontrollen

Data från stora tekniska anläggningar kan hjälpa forskare att förstå, förutsäga och förbättra komplexa system. Men innan sådan information kan delas behöver organisationen veta vad som är säkert att lämna ut, vem som får fatta beslut och hur materialet ska hanteras.

Den europeiska spallationskällan, ESS, i Lund är en av Europas stora forskningsanläggningar. När anläggningen styrs, övervakas och testas skapas stora mängder data från dess kontrollsystem. Man kan se denna data som anläggningens tekniska dagbok: den visar hur olika delar av systemet beter sig över tid.

För forskare kan sådan data vara mycket värdefull. Den kan användas för att upptäcka avvikelser, förutse underhållsbehov, testa analysmetoder eller förstå hur stora tekniska system fungerar i praktiken. Det är särskilt intressant inom områden som dataanalys, AI och industriella system, där verklig driftdata ofta är svår att få tillgång till.

Samtidigt går det inte att bara lägga upp filerna på nätet. Data från kontrollsystem kan innehålla information som behöver hanteras varsamt. Den kan till exempel säga något om hur anläggningen är uppbyggd, hur den används eller hur olika tekniska delar hänger ihop. Därför måste delningen ske på ett sätt som gör materialet användbart för forskare, utan att organisationen utsätts för onödiga risker.

Examensarbetet undersökte hur ESS kan dela utvalda delar av sin kontrollsystemsdata med forskare utanför organisationen på ett säkert, tydligt och praktiskt genomförbart sätt. Fokus låg inte främst på att bygga en stor teknisk plattform, utan på den kanske svårare frågan: hur organiseras arbetet så att rätt data delas, på rätt sätt, med rätt personer?

Resultatet blev ett arbetssätt för kontrollerad datadelning. Det innehåller ett flöde med tydliga roller, steg för godkännande, stöd för att bedöma om data är känslig eller möjlig att dela, samt mallar för förfrågningar, dokumentation och uppladdning av data. Målet är att processen ska vara enkel nog att använda i praktiken, men tydlig nog för att skapa ansvar, överblick och förtroende.

En viktig slutsats är att säker datadelning inte bara är en teknisk fråga. Det handlar minst lika mycket om ansvar, kommunikation och gemensamma beslut. Även den bästa tekniska lösningen riskerar att bli oanvändbar om det är oklart vem som ansvarar för frågan, vem som får godkänna delning eller hur känslig information ska upptäckas.

Arbetet kan hjälpa ESS att ta steg mot mer systematisk delning av data från kontrollsystemet. På längre sikt kan det också vara relevant för andra forskningsanläggningar och organisationer som vill dela driftdata utan att tappa kontrollen över risker och ansvar. På så sätt kan data som redan finns komma till större nytta – för forskning, industri och samhälle.