Degree Project

UNIVERSITY OF SKÖVDE

DIGITAL MAPPING OF CRITICAL INFRASTRUCTURE

Design of a Component Data Collection Method for Small-Scale Power Grids

Master Degree Project in Informatics

Second Cycle 30 credits

Spring term 2023

Student: Axel Rapp

Supervisor: Sten F. Andler

Examiner: Rose-Mharie Åhlfeldt

ABSTRACT

Critical infrastructures (CIs) distributing water, oil, gas, electricity, etc., to community residents and businesses, leverage cyber-physical systems (CPSs) to supervise and control the physical processes that these services entail. Over recent decades, these systems have moved to implement more modern IT-resembling solutions using Supervisory Control and Data Acquisition Systems (SCADA) for increased reliability, scalability, and remote connectivity. This change exposes these highly critical systems to new threats and vulnerabilities. One approach to mitigate the risks faced by these systems is to perform analysis on digital representations in the form of digital models or digital shadows of the CPSs. However, this is not a trivial task in practice. These practical issues are explored in this design science research through the development of a guidance process to perform the data collection necessary to create a static digital model of a small-scale power grid CPS in Sweden. The results show that it is possible to gather information on the CPS components through the four approaches: SCADA system exports, documentation information, CLI scripting, and network scanning. While the artefact presented in this report demonstrates these results, challenges still remain such as a lack of SCADA export tools, reaching the SCADA network with scanning tools in a responsible manner, and accessing insights into the complete documentation held by the organisations. The researcher suggests these topics for future research directions.

Keywords: SCADA, CPS, cyber-physical system, CI, critical infrastructure, digital model, digital shadow, power grid, design science.

Table of Contents

1 Introduction

The term "cyber-physical system" (CPS) refers to the way contemporary systems use software components to monitor and manage physical mechanisms and processes. Critical infrastructures (CIs) leverage the attributes of these systems to provide vital societal functions, such as the distribution of water, oil and gas, and electricity, to name a few. Within the last decades, CPSs, in general, and CIs, in particular, have experienced a transition from being isolated "off-grid" systems to instead resembling highly connected modern IT systems. While the capabilities of CIs might increase due to this transition, the isolation factor is reduced, wherefore new threats and vulnerabilities are introduced (Stouffer et al., 2015). The allure of increased reliability, scalability, and remote connectivity that comes with an evermore inexpensive cost of internet connectivity continues to drive this transition forward. Thus, increasing the demand for countermeasures to mitigate newly imported risks (Nazir et al., 2017).

By definition, all CIs are essential to society. However, power grids are unique in the sense of being a centrepiece in almost all other forms of CIs as well. Should a power grid be attacked and its service denied, the magnitude of the attack could be devastating due to its cascading effect also affecting other community-supporting operations such as healthcare, telecommunications, resident heating, water supply etc. This is certainly not a speculative scenario; the prime example of such an event is the attack against the power grid in Kyiv in December of 2015, wherein a regional Ukrainian electricity distribution company suffered a breach of its supervisory control and data acquisition (SCADA) system, which caused it to lose connection to a number of substations. The third-party managed to remotely control the SCADA distribution management system, purposefully causing outages. Ultimately, the cyber-attack resulted in approximately 225 000 customers losing power across Ukraine for several hours (Assante et al., 2016). Additionally, the injected malware erased the hard drive on all compromised machines (Pliatsios et al., 2020).

Being central critical infrastructure, a power grid disruption will likely affect significant parts of society. Thus, mitigation efforts to protect them are paramount. One such effort is the European Union project ELVIRA. With funding and collaboration from the EU Internal Security Fund (ISF) and Combitech AB, the ELVIRA project focuses on infrastructure resilience through the development of infrastructure dependency analysis tools for power grids (“Infrastructure Resilience – ELVIRA,” 2020). The project aimed to procure a deepened understanding of the risks and dependencies of smart-grid components. Developed models intend to ease the assessment of vulnerabilities and potential cascading effects of component failure in smart-grid networks (Polismyndigheten, 2021).

The usefulness of the analysis models provided by the ELVIRA project could be utilised not only for general purposes but also on an organisation-specific level, should one create a model of a specific power-grid system. The system could

then be assessed and aid the system owners in decision-making regarding security and resilience. In order to achieve this, however, real-world data about the system is required (Dietz et al., 2022) – a non-trivial task in practice.

1.1 Problem Description

To digitally represent a heterogenous CPS such as a power grid, albeit simply as a digital model or as a digital shadow (a model automatically fed with data of the actual state of an existing physical object), two main objectives emerge (Bergs et al., 2021). Namely, (1) data collection and (2) modelling.

(1) Data must first be collected, manually or automatically, of components and entities that comprise the system, such as IT components, OT (Operational Technology) components, software and operating systems on those devices, physical components, and communication data flows. To this extent, several matters of trouble surface. Data availability and accessibility are not necessarily clear due to data sources often being a mix of analogue and digital nature and under the care of separate organisational structures and people (Jiang et al., 2023). Further, to ensure the scalability and accuracy of the data collection, automatic asset identification is preferred over manual on-site walk-downs. Various tested scanning tools exist and can be separated as passive and active listeners — each type of listener category with its limitations and disruption effect on the targeted network. Other possible means of information extraction for these systems include exporting from the SCADA system and command-line scripting. The availability of the system of focus is essential, and any disruption to its service must be avoided. It should also be mentioned that the information within these systems is security-sensitive, and listener tools must balance the scales of extracting necessary information for the digital model and excluding remaining information not needed for that purpose.

(2) Next, using the collected information, a model is created manually or automatically to represent the system digitally. The analysis models provided by the ELVIRA project are helpful in this regard. For the purpose of this research, only the data collection objective providing qualitative information to use as input to the modelling objective is to be considered.

Digital modelling and simulation are essential tools and potent mechanisms used to understand CPS behaviour better and evaluate risks and threats (Chowdhury and Gkioulos, 2021; Nazir et al., 2017; Pliatsios et al., 2020). Although highlighted as a central concept for CPS defences, few works elaborate on the practical issues of creating holistic models. Moreover, missing or incorrect information on assets and data flows derived from the data collection phase into the model directly affects the model's validity. The difficulties in providing first-rate information are very much a concerning issue (Khalil et al., 2023).

Altogether, the main issues revolve around the accessibility of component data in these heterogeneous systems. With various levels of documentation, the

accessibility to all information needed for a complete digital model is not guaranteed. Additionally, the heterogenous nature of these systems strongly suggests that no all-purpose tool can achieve an active search for all necessary information. An important consideration throughout any process of collecting and gathering data about the system of focus is the essentiality of maintaining its availability and integrity. And of course, that information reaches a high level of quality to ensure real-world representation.

Although large organizations can implement a variety of security measures, businesses with scarcer resources cannot hedge their operations with state-of-the-art protection on all fronts (Deloitte, 2021). Instead, smaller businesses are becoming a target for cyber-attacks while unable to deploy controls and safeguards that larger organisations routinely implement (Tam et al., 2021). Small businesses urgently require effective and appropriate cyber-security solutions.

Overall, these considerations support the notion that as part of small-scale power grid administrators' cyber-physical security efforts, a custom sector best practice method for data collection for digital modelling is required – an artefact not available today. In designing an artefact to fulfil this purpose for a small-scale electricity distributor in the central Gothenburg area, a step toward achieving resource-effective vulnerability analysis for these systems is taken.

1.2 Research Aim & Research Question

This research aims to investigate the practical issues present in the data collection phase of creating a digital model of a small-scale power-grid CPS.

Two research questions are derived from the discussed topic:

- What are some of the challenges and limitations of collecting the data necessary to digitally model a small-scale power grid CPS?
- What are the possible methods and means to overcome the challenges and limitations of collecting data of a small-scale power grid CPS sufficient for an accurate digital model?

By procuring the answer to the stated research questions, this research aims to investigate methods for extracting information about a small-scale power-grid CPS to accommodate the creation of a digital model serviceable to analysis with ELVIRA threat and vulnerability models.

The general contribution of this research to the described problem area is to provide answers to the stated research questions. The particular contribution is the creation of the artefact presented in this research report.

Chapter 2 presents the background of the research area through a description of the central concepts and related works. Next, the methodology and scientific framework utilised to conduct the research are declared and argued for in Chapter 3 before describing the implementation of the framework and the resulting findings in Chapter 4. Discussion of the results and the research validity

is found in Chapter 5. Finally, a concluding Chapter 6 summarises the findings and conclusions drawn from this research along with suggestions for future research directions.

2 Background

The background section of this report consists of two main parts: first, the main concepts of this research are explained, followed by a related works section describing the previous research in the area.

2.1 Main Concepts

A brief explanation of the central concepts is provided to initiate the background section. Namely, a general description of the system of focus, relevant laws, regulations and standards, and the definition of digital models and comparisons to digital shadows and digital twins.

2.1.1 System of Focus – Small-Scale Power Grid

An electrical power grid generates, transmits, and distributes electrical power. A typical power grid can be categorised into three different layers, which it uses to achieve this objective. Namely the physical, control, and communication layers. These are intertwined and work in unison with human operators to maintain and manage the energy network. This work will apply the same domain-specific taxonomy as that of the ELVIRA project, seeing that it should be considered as an extension of said project. As suggested by Jiang et al. (2018), the generation, transmission and distribution of power are supported by the physical layer by, among other components, power generators, transformations, busbars, and circuit breakers, while the control layer concerns the SCADA system which operates, controls, and monitor the operations of power grids through the use of a network of OT devices. The OT network comprises microprocessor-controlled devices communicating with physical sensors (Jiang et al., 2020a). The communication layer is a network of field communication devices and control centres mainly consisting of routers, switches, firewalls and data lines.

SCADA

In distributed systems like electricity distribution systems, SCADA systems control assets spread over broad geographic locations (Stouffer et al., 2015). They integrate data acquisition systems with data transmission systems and Human Machine Interfaces (HMI) to achieve a centralised location of monitoring and control for a human operator or automatic processes. Typical components of a general SCADA system include a control server, communication equipment, data historian, workstations, Master Terminal Units (MTUs), and distributed field sites with Remote Terminal Units (RTUs), Programmable Logic Controllers (PLCs) and Intelligent Electronic Devices (IEDs). The devices communicate with the central control server through specialised software, which can query sensor and telemetry information from different devices. The control centre also collects field station logs that can trigger automatic actions. The field stations often support remote access, allowing operators to perform tasks from afar when necessary. Standard and proprietary communications protocols and communications topologies are used and vary between implementations. A common design consideration of SCADA systems is the redun-

dancy aspect since critical infrastructures that use them need a high fault tolerance.

MTUs are devices that route commands from the SCADA system to RTUs in a master-slave relationship in which only the MTU can initiate bidirectional communication (Jiang et al., 2018).

The RTUs gather field data via sensors and transmit the information to a master station via the communications layer (Ahmed and Soo, 2008). The RTU thus offers an interface with the remote field sensors. The master station shows the data collected and allows the operator to perform remote control activities.

A PLC is a controller device that interprets signals from sensors and, based on a predetermined programme, transmits instructions to actuators, i.e., breakers and switches, which directly affect a physical process in the system (Stouffer et al., 2015).

PLCs and IEDs are similar in many ways; however, an IED can be considered a more sophisticated PLC because it can control multiple aspects of a piece of equipment (Stouffer et al., 2015).

The ELVIRA taxonomy can represent all types of devices in a power grid system and the communication paths that connect them (Jiang et al., 2018). Some notable considerations of the taxonomy include the distinction between cyber components and physical components in which the relationship is that cyber components are computer readable code such as data and software, and physical components are physical devices such as computers and routers that can host the cyber components. A similar relationship applies to data streams and physical connections. A data stream is a cyber component hosted by a physical component (e.g., a data line).

The taxonomy can represent three types of relations between cyber components and physical components: (1) embedding, (2) monitoring, and (3) controls.

(1) The embedding relation specifies that a cyber component is hosted on a physical component,
(2) the monitoring relation specifies a logical relation that of a cyber component receiving information from physical components, e.g., sensor readings from a transformer, and
(3) the controls relation represents operation calls sent to physical components, such as actuators of power grid components.

2.1.2 Relevant Laws, Regulations and Standards

For organisations providing services of socially important function in Sweden and implicitly within the European Union, a particularly important document is the NIS directive and its subsequent proposal for a successor repealing the original directive (*Directive (EU) 2016/1148 of the European Parliament and*

of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union, 2016; *Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on measures for a high common level of cybersecurity across the Union, repealing Directive (EU) 2016/1148*, 2020). The directive aims at all public and private entities within the Union and calls to improve resilience and incident response capabilities in cybersecurity and critical infrastructure protection. Through the adoption of *"informationssäkerhetslagen",* the NIS-directive was incorporated into the Swedish legal order (*Lag (2018:1174) om informationssäkerhet för samhällsviktiga och digitala tjänster*, 2018; "Lagar och förordningar gällande NIS," 2022). The regulations of this law provided by the Swedish Civil Contingencies Agency (MSB) include a number of clarifications on how to conduct one's operation:

1. A systematic and risk-based information security work in conjunction with the standards SS-EN ISO/IEC 27001:2017 and SS-EN ISO/IEC 27002:2017[1]
2. A policy describing the organisation's information security goals
3. A well-documented way of working with information classification, risk analysis, and mitigation efforts
4. A solid incident response plan

(MSB, 2018).

While not yet incorporated into Swedish law, the replacement of the NIS directive, NIS 2, has been decided on the union level and will be incorporated in all member states by 17th of October 2024 (*Directive (EU) 2022/2555 of the European Parliament and of the Council*, n.d.).

The primary purpose of the NIS 2 directive is to strengthen further the capacity of information and cybersecurity within the union and to limit the impacts of incidents on critical infrastructures. More effective collaboration between authorities from different member states can align the goals and achieve a more harmonious relationship (MSB, 2023). The directive aims to include more sectors and add more and expanded demands on risk management. These demands include but are not limited to a national strategy and framework for incident response, a national Computer Security Incident Response Team (CSIRT), extended scope of application, stricter security requirements and regulatory measures through administrative sanction fees.

To strengthen national security, the Swedish government proposed a new security law, which was enacted on 1 April 2019 (*Säkerhetsskyddslag (2018:585)*,

[1] *The newer ISO/IEC 27001:2022 and ISO/IEC 27002:2022 versions are not mentioned in the current regulations. They are, however, expected to be included in the regulations related to the NIS 2 directive.*

2018). The law includes clarifying obligations for the organisation conducting security-sensitive operations such as those associated with critical infrastructures. The law has been updated and expanded since and now demands that a certain protective security agreement must be established with any outside party with insights into the security-sensitive operation (*Säkerhetsskyddsförordning (2021:955)*, 2021).

While an assessment should be made for each organisation, it is highly likely that most of an electricity distributor's networks and information systems will fall under the NIS directive scope, ergo, the information security law in Sweden. Parts of the organisation could also be within the scope of Säkerhetsskyddslagen. In collaboration with an outside party, this is especially important to consider.

2.1.3 Digital Model, Shadow or Twin?

There are uncertainties regarding the exact definition and differences between the concepts of digital models, digital shadows, and digital twins (Somers et al., 2023). However, a commonly referred distinction is the nature of data flow and cause and effect between digital and physical objects (Wang et al., 2022). If the change in the state of the physical object does not affect the digital object and vice versa, it should be regarded as a digital model. In such a setup, the data flow between the entities is manual in both directions. A digital shadow, on the other hand, means that a changed state in the physical object automatically causes a change in the state of the digital object, thus having automatic data flow from the physical object to the digital object, but not vice versa. A digital twin achieves automatic dataflow in both directions, meaning that a change in the state of one of the objects, albeit the digital or the physical, changes the state of the other.

As this research is concerned, the focus is developing a method to collect the necessary information feasible to create a digital model of a power grid CPS. Efforts to extend the method to facilitate the creation of a digital shadow or twin are beyond this research's scope. However, since automation is an attribute worth striving for in this regard, the method presented in this paper will be discussed to assess whether or not the suggested solution is applicable to a digital shadow or digital twin creation process.

2.2 Related Work

This section provides an overview of previous research related to the study area. An account of how the information was retrieved is found in Chapter 3.1 – Literature Review.

In providing a general definition for the concepts of digital shadows and digital twins within the manufacturing sector, the authors explain that the concept of a digital shadow can be seen as a precursor or preliminary stage to developing a digital twin (Bergs et al., 2021). For the concept of digital twins, it is noted that data collection is a significant challenge and that the digital model properties need to have asset-specific requirements, which means that only predeter-

mined relevant information is to be collected. The main contribution of this article, however, is the clarification of concepts to create a general understanding which would enable application development.

In mapping the ecosystem of digital shadows in the same sector, manufacturing, a conceptual model of digital shadows intended to favour engineering, combination, and reuse of digital shadows is presented (Brecher et al., 2021). The authors conclude that the main challenge of implementation is collecting and aggregating relevant data for specific needs. Further presented, digital shadow types encapsulate a subset of required data for the specific use case and its purpose. The presented types contain data sources and data specifications to serve as a construction plan in a manufacturing environment.

Several limitations are listed in a thematic review article on digital twins, their evolution, characterisation, and maintenance management purposes (D'Amico et al., 2022). Although the digital twin is of focus, it is of interest to the topic of the stated research questions in this study. That is the high initial costs of implementation, management of large datasets, the need for specific and dedicated solutions, and lack of stakeholders' awareness within the same supply chain. It should be noted that merely six out of the 59 reviews article related to infrastructure and that a clear majority concerned manufacturing.

In an experimental design study, the authors develop a model for the operational optimisation of manufacturing systems (Ehrhardt and Hoffmann, 2020). Part of which is the integration of real-time data from a digital shadow. The authors are able to use homogenous log data from a milling machine to initiate the model construction. Utilisation of the same in a power grid environment is deemed limited due to heterogenous data and limited log data. The article's main contribution is the digital shadow architecture for its applicability in production environments.

In a review article, the authors summarise existing literature on digital twins and a case study of a digital twin implementation in a smart infrastructure railway bridge (Gürdür Broo et al., 2022). One key takeaway concluded by the authors was the importance of including multidisciplinary stakeholders. Non-technical considerations are similar in importance to those of technical nature. While the focus of the article is that of digital twins, this consideration is interesting and relevant to the stated research questions in this study as well.

In a review article on digital twins-based smart manufacturing systems, the authors present definitions, frameworks, major design steps, and key enabling technologies (Leng et al., 2021). Unified Modelling Language (UML) and Model-Based System Engineering (MBSE) are commonly used approaches to describe smart manufacturing systems. Input to this model, however, is not discussed—the main contributions of the article concern design considerations in smart manufacturing systems.

A systematic framework to provide guidelines for modelling and remote control of a Cyber-Physical Production System (CPPS) as a digital twin is proposed (Liu et al., 2020). A relevant passage highlighted by the authors suggests the prerequisite of a unified information model that can represent heterogenous data such as topologies of machines, sensors, and persons. A domain ontology that semantically can denote classes, properties, and data values are needed. By proof of concept, the authors can develop a semantic model of a CPPS.

By in-depth reviewing literature about digital twins, the authors analyse concepts, technologies, and industrial applications of such (Liu et al., 2021). Besides listing common remote communication protocols and other data-related technologies, the article concludes that there is great variety between application domains. Further, data being the basis of a digital twin, mapping and fusion are needed to understand the collected data. The most common technology used in that regard in the literature is Extensible Markup Language (XML).

To add to the solution variety aspect, while proposing a requirement-driven, technology-agnostic digital twin architecture, it is concluded that previously proposed digital twin architectures are largely domain- and technology-specific (Nwogu et al., 2022). While the inspiration for creating a framework certainly can be applied to a new domain, changes from the original framework are to be expected. This aspect is likely to be reflected in the asset identification of the data collection phase of a power grid CPS digital shadow.

Conclusions drawn for the reviewed material is that most literature concern the concept of digital twins – a later stage to the digital shadow. From the digital twin implementation perspective, a sufficient data shadow (including data collection and asset mapping) has already been achieved. Thus, limited information has been found in research regarding the asset identification and mapping of CPSs. Additionally, as shown by the review, most research focuses on the manufacturing sector and less on electricity distribution systems which is the area of focus for this study. As such, the research questions and the aim are well worth pursuing in order to extend the knowledge base.

3 Method and Scientific Framework

To best answer the stated research questions, a mixed methodology consisting of a qualitative literature review and design science research was conducted, each approach presented in subchapters 3.1 and 3.2, respectively.

3.1 Literature Review

In order to provide a relevant and state-of-the-art summarisation of the theoretical foundation and context to the stated research questions, a qualitative literature review based on a down-scaled version of a structured literature review was conducted as part of the background section of the report (Okoli, 2015). The author suggests an 8-step approach including (1) identifying the purpose, (2) drafting protocol and training the team, (3) applying practical screen, (4) searching for literature, (5) extracting data, (6) appraising quality, (7) synthesising studies, and (8) writing the review. Due to the nature of this literature review acting as a theoretical background to a design science research rather than a full-scale structured literature review, a simplified version of this approach was of choice. One difference is the exclusion of step (2) simply because this review only employs one person. Additionally, to adapt to the limited resources available, the quality appraisal step (6) is omitted by only including databases that only provide peer-reviewed content. The databases used are ACM Digital Library, ScienceDirect, and Wiley Online Library. Step (7) is simplified by including all extracted data article-wise instead of thematic results.

The review aims to gain insight into the related works of smart grid and power grid digital shadow creation to provide a level of understanding of state-of-the-art data collection methods.

Some pragmatic considerations are decided upon to apply an initial screening to determine which studies should be included.

1) Content – to the reviewer, the content must be deemed relevant to the stated research questions and the aim of this study.
2) Publication language – the only valid language is English due to limited translation resources.
3) Quality – Exclusively peer-reviewed papers are used by limiting the database usage to those only presenting peer-reviewed materials.
4) Date of publication – for relevancy purposes, only research articles published in 2018 or later will be included.

To form the criteria of search terms, the three main concepts considered are "cyber-physical systems", "digital shadows", and "asset identification". While the main focus of this study is the investigation of power grids, the search terms are extended to include all forms of CPSs since similarities between these systems exist and common courses of action are not to be overlooked. Additionally, when a search using the terms of smart-grid or power grid was conducted, the number of search hits was deemed too low to form an adequate theoretical background. Regarding the concept of digital shadows, this is also extended to include digital twins and digital models. While creating a digital twin or digital

Table 1. Literature search by databases

Database	Search hits	Passing initial screening	Passing all criteria
ACM	4	2	0
Wiley	2	0	0
ScienceDirect	160	31	9
Total	**168**	**33**	**9**

model is not the end goal of this study, the data collection phase overlaps between the concepts. The final search term aims to find any studies related to identifying, discovering, and detecting physical assets in the CPS as part of creating a digital shadow. By synthesising this into a search term, the following query is determined:

("cyber-physical system" OR "CPS") AND ("digital shadow") AND ("identification" OR "discovery" OR "detection")

Once all search results are collected, the articles are initially screened by assessing the content of their title and abstract. The articles passing this initial screening are read in their entirety to assess their relevance. The results from the initial search, initial screening and thorough screening can be seen in Table 1. The final results of the literature review are presented in Chapter 2.2 - Related Work.

Additional literature searches were performed after the findings were aggregated to complement the slimmed-down version of the SLR. This effort acts to expand the anchor points for the artefact knowledge, and to extend the literature review for any potential research missed in the background, e.g., any articles outside of the time scope set or any work related to the results which fall outside of the search strings presented here. Methods applied in this respect include: consulting literature already deemed relevant to search for additional relevant research, screening other ELVIRA research, and performing additional database searches based on the findings of Chapter 4 – Implementation & Results.

3.2 Design Science

The task presented in this study is proactive as it has less to do with understanding or predicting than developing and applying an artefact for a real-world problem. The problem resides in a relatively new and variable research area where a common best practice is yet to be academically reached. Additionally, an opportunity for improvements presents itself as there are deficiencies in former and current solutions that could be made more secure using currently available tools. Through this line of reasoning, the applied research method of design science stands out as the most appropriate strategy for improving the current practice. In this context, the goal is to produce an artefact that assists in managing a practical issue. Another research method briefly considered was action research since both include problem-solving and evaluation. Both approaches also address practical issues; however, action research does not nec-

essarily do so through the use of an artefact. (Johannesson and Perjons, 2014). Action research can leverage psychological, social, and organisational change in order to find solutions to problems. This is deemed irrelevant to the present problem description. Additionally, design science has the flexibility to apply different research strategies for each activity in the process, in contrast to action research which is just one standalone research strategy. Due to this reasoning, action research was deemed less fit in comparison to design science.

While several types of artefacts exist, a common categorisation is that of (1) constructs, (2) models, (3) methods, and (4) instantiations (March and Smith, 1995). As summarised by Hevner et al. (2004), (1) constructs provide language and communication abilities with regard to the definition of problems and solutions. (2) Models explain the design and solution space of a problem in order to ease understanding for further exploration and real-world condition changes. (3) Methods specify processes and offer direction to search the solution space, and finally, (4) instantiations demonstrate that models, methods, and constructs can be implemented in a real-world system. The problem description of this study searches for a course of action to collect information about a power grid CPS which is why the method type is the most suitable. Methods specify processes and offer direction to search the solution space (Hevner et al., 2004). These directions can be formal, informal, or a combination of both, which is suitable for the outlines of this project.

The methodology framework used to conduct this design science research is provided by (Johannesson and Perjons, 2014). It offers a process consisting of five activities, namely, (1) explicate problem, (2) define requirements, (3) design and develop artefact, (4) demonstrate artefact, and (5) evaluate artefact. Each activity has underlying sub-activities, which will be further detailed in subsequent sections. For each activity, an input is given, and an output is created. The output is then passed to the next activity as input. To exemplify the relationship: the input to the first activity, explicate problem, is an initial problem. The sub-activities: investigate and analyse the problem, formulate it precisely, and justify it as significant for a specific area creates the output – an explicated problem. The explicated problem is then passed to the second activity, defining requirements as input. An important note here is that the activities are not necessarily performed sequentially but more often iteratively. Two additional channels exist that affect the activity in the framework: controls and resources. Controls pertain to the knowledge, including research methodologies, research strategies, and creative processes, which are utilised to direct an activity such as a survey or case study. Resources pertains to the body of knowledge that serves as the foundation for an activity, such as models, theories, stakeholder interests, and previous research in the area.

While the overall research strategy is design science, the framework allows for the use of different research strategies for every activity or even multiple strategies per activity. For example, a case study could be the strategy of choice during the explicate problem activity, while the artefact evaluation makes use of a survey. Not all design science projects thoroughly undertake all five of the tasks in the framework with equal consideration. Instead, a common approach is to

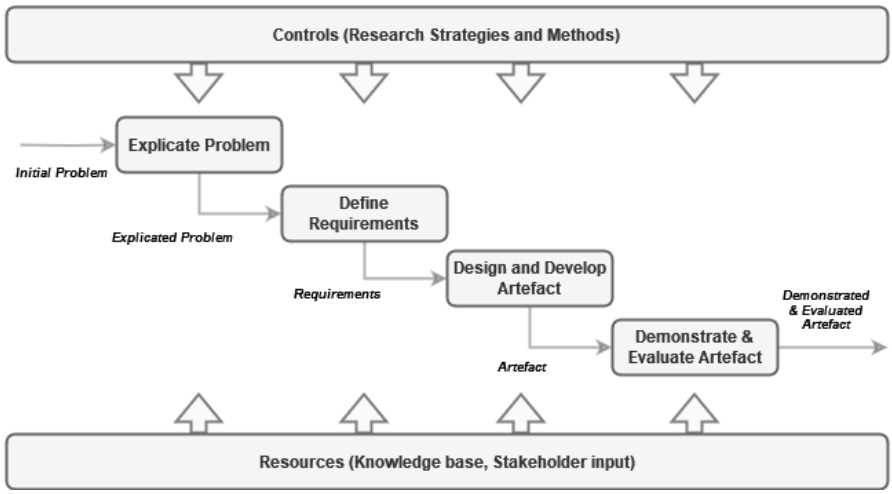

Figure 1. Design Science Framework (Johannesson and Perjons, 2014)

concentrate on one or two of the activities while treating the others more lightly. This rationing of resources certainly applies to this study, in which the focus is placed on the design and development of the artefact.

A lighter focus is placed on explications of the problem as well as the requirement definition due to the collaboration of this research with an organisation that specialises in vulnerability management and cybersecurity for critical infrastructure. Throughout this paper, the company will be referred to by the alias InfraSecure. The problem is evident to InfraSecure and is therefore accepted more or less as is. The attributes needed for the model have also been predetermined by the collaborating partner InfraSecure.

Additionally, the demonstration and evaluation of the artefact will take a lightweight approach as well due to the time constraints under which this study is performed. The activity of artefact demonstration and artefact evaluation is combined for this purpose.

A graphical representation of the framework, including the aforementioned adjustments is available in Figure 1.

The following chapter includes an account of the implementation and result of the four activities to be conducted within the design science framework.

Several data collection methods are mentioned within these sections alongside a stakeholder. The stakeholder term is to be interpreted as a key person in an organisation directly or indirectly involved with systems treated under this research. It can also refer to a person with great insight into systems or a holder of information necessary for the development of the artefact.

Previous research and body of knowledge are to be collected through literature searches. In addition to this form of data collection, the stakeholders are consulted throughout the process to aid the direction of the artefact development.

The stakeholders referred to in this study include:

InfraSecure – an organisation specialising in vulnerability management and cybersecurity for critical infrastructure.

LocalPower Inc. – a small-scale electricity distribution organisation, henceforth referred to as the alias LocalPower Inc., a close collaborating partner with InfraSecure. LocalPower Inc. is a university-led cutting-edge research centre specialising in biomass energy production and carbon capture. Additionally, LocalPower Inc. supplies the campus region with electricity.

SCADA vendor engineer – a SCADA engineer responsible for installing and servicing the SCADA system used by LocalPower Inc.

4 Implementation & Results

The previous chapter explained and argued for the choice of research method. Instead, this chapter focuses on how the design science framework was applied to the context of this study. Each subsection clearly states the goals of each activity, its inputs, outputs, resources, and controls. Additionally, it ties the activities to the practical environment in which the study was conducted, along with the results of each activity.

4.1 Explicate Problem

For the first activity of the framework, the goals are to precisely formulate the problem and justify its importance. As mentioned in the previous chapter, the input to the activity is an initial problem, and the produced output is an explicated problem. Through this activity, the problem is positioned and the practice area wherein it appears clarified. Additionally, the activity aims to describe the problem in an easily understandable manner while still being precise and concise. Being the foundation for the rest of the project, it is also important to ensure that the problem is of general interest and solvable as well as explaining the distinction between the current state and the desired state.

Stakeholder interests and views, together with previous research, build the most important resources throughout the project and certainly apply to this activity. Although the problem's explication is not the project's primary emphasis, any controls that must be utilized in this activity are those of a qualitative type. Should any uncertainties hinder the goals of this activity, consultation and discourse will be held with the concerned stakeholder.

With InfraSecure as the problem owner, discussions were held with the company to explicate the problem. The problem of digitally representing physical systems becomes evident through their perspective, a company aiming to perform various security analyses of power grids through digital models. The problem can be exemplified in a situation wherein a power grid company wants to assess their operation in an effort to ensure cyber resilience. One way of doing this is to create a digital model on which various analyses can be performed. Lacking the expertise themselves, the power grid company hires an outside party to perform this assessment. However, for the outside party, this is a challenging task, mainly due to the difficulties related to collecting information from the system to represent it accurately digitally. An artefact aiding this task would help the cyber vulnerability assessment for the local practice, and mitigation efforts could be taken in order to achieve a more robust operation. While this study focuses on a small-scale electricity distributor, a wider use case can be envisioned to include mid and large-sized organisations both in the energy sector and beyond following the success of a method applicable to CPSs. With InfraSecure as the main contributor to the problem description, the scope of this study is limited to their field of focus, namely small-scale electricity distributors. To represent that field, Local Power Inc. is the subject of investigation in defining a method for possible means of extracting the information necessary to create a digital model of their CPS. InfraSecure's view of the lack of a set method for data collection and extraction of CPSs is supported by Holm et

al. (2014). The authors describe a similar problem definition as the one present in this case, namely that little or no discussion is present regarding the data collection practical issues related to model instantiation, in their case, regarding Enterprise Architecture (EA). As with the work conducted by ELVIRA, most of their analyses are based on the premise that the modelled data is already available. Generally, the main issue is to collect this data in the first place, and that vulnerability analysis models have this as a pre-condition. Therefore, a significant piece of the puzzle is still missing. In a case study performed at a Swedish municipality to digitally model a power grid, it was also concluded that a significant challenge to creating a full and accurate model is the dispersion of information (Jiang et al., 2023). Information relevant to create the digital model can be scattered between different departments of the organisation and vendors of the components, as found in their case study. This dispersion of information was verified when talking to the other stakeholder, LocalPower Inc., which information about their system was not easily retrievable. Furthermore, the documentation that LocalPower Inc. provided had flaws in terms of relevancy and detail.

The resulting problem explication derived from this is that no method exists guiding the process of collecting the data needed to model a power grid CPS digitally. The desired state includes an artefact significantly streamlining the process of digitally representing a CPS.

4.2 Define Requirements

The second activity aims to outline an artefact that can serve as a solution for the explicated problem and what the requirements, both generic and specific, are. The input to this activity is the explicated problem, and the output produced is the artefact outline and requirements. This process aids in deciding what type of artefact to construct, in this case, a method as argued in Chapter 3.2 – Design Science, along with its characteristics. Every requirement is described similarly to the problem explication, namely in an understandable yet precise and concise manner. An important consideration in this activity is that the solution being carved out is realistic relative to the project's available resources.

As with activity one, stakeholder input is also the most important resource to this activity. The requirement definitions are, as previously mentioned, not the main focus of this project. However, consultation and discourse with relevant stakeholders served as the main controls to this activity.

As described in Chapter 3 – Method and Scientific Framework, the artefact of choice is a method applicable to the problem explication in the previous activity. Similar to problem explication, the requirements for the artefact are significantly affected by the experience of InfraSecure. Hence, the company was asked to provide requirements for the information they needed to collect in order to create a digital model. Through this consultation, InfraSecure provided a set of requirements describing what components and properties are used when creating and representing the model digitally and whether or not those components

are mandatory in the context of creating a model in ConceptBase using the ELVIRA taxonomy. This way, it is confirmed that the desired vulnerability analyses can be performed using the required data. While digital models can be extended into absurdity, these components and properties are considered the most essential in order to perform fruitful vulnerability analyses.

The requirements are divided into seven categories with corresponding properties: (1) software components, (2) software component communication, (3) IT and OT components, (4) physical components, (5) connections between physical components, (6) subcomponent tables, and (7) other containers. First, (1) software components present in the system are to be identified. This includes all software and applications installed on computers. For each software component, an identifying name will be set, and additional property information such as the software vendor, the model of the software as well as version and build numbers need to be collected. The software component will also have a type, e.g., firmware, operating system, or application software, and finally, a relationship that determines on what hosting computer the software component resides. Next, (2) software component communication deals with the communication between software components. They, too, are given an identifying name. The properties related to the software component communication are the IDs of the sender and the receiver components, together with the protocol used. Additional property information includes the purpose of the communication, e.g., sensor reading, and the communication type, e.g., if the communication is unidirectional or bidirectional. For (3) IT and OT components, besides setting an ID, the vendor, model, and version are required, along with the type of component. Examples of the component types are server computers and PLC. Components such as machines, transformers or other physical components controlled by IT and OT are placed in the category (4) physical components. Here, the properties ID, vendor, model, version and type are needed. Examples of the type property include busbar, circuit breaker and transformer. The next category, (5) connections between physical components, describes the physical connections between components, either physical or IT and OT components. Property information needed is the IDs of the connected components and the connection type (data, power, or physical connection). (6) Subcomponent tables are needed to describe, for example, software installed on a host machine or computers residing on a particular network. Finally, (7) other containers are needed to describe areas or networks that contain other physical components. Each container is given an ID, a type (network, substation, etc.) and a descriptive name.

(7) Other containers, (2) software component communications, and (3) connections between physical connections were marked as non-mandatory in regard to their role in creating the digital representation. However, since information regarding their properties could still impact security, they are included in the study.

Table 2. Defined Requirements

Software components	Software component communication	IT and OT components
ID	ID	ID
Vendor	Sender ID	Vendor
Model	Receiver ID	Model
Version	Protocol	Version
Build	Purpose	Component type
Software type	Type	
Hosting Computer		

Physical components	Component connections	Subcomponent tables	Other containers
ID	ID #1	Host ID	ID
Vendor	ID #2	Software ID	Container type
Model	Connection type		ID
Version			
Component type			

The required seven categories with corresponding properties are assembled into a table format for an easier overview in Table 2.

As per the functional requirement of the artefact, it is to map the components and their properties to a means of extracting that information from a power grid CPS.

This set of requirements was the sole set of requirements desired by the company, meaning that no requirements were set on the artefact's structure. Due to this, a simplistic approach was decided upon where the artefact is in the form of a document containing directions for the user of said artefact. Additionally, although not explicitly expressed, a few structural qualities to aim for during the design and development of the document are the coherence, modularity and conciseness of the artefact. Coherence refers to the logical ordering and relationship between the different parts of the artefact, essentially how naturally easy the artefact is to use and understand. Modularity refers to the degree the different parts of the artefact are separated and combined and how easily modules can be edited and replaced. Conciseness refers to the cropping of redundant and unnecessary content of the artefact.

4.3 Design and Develop Artefact

The third activity entails the creative process of creating an artefact that provides a solution to the explicated problem and fulfils the requirement definitions from the previous activity. The input consists of the artefact outlines and requirements, and the output produced is the artefact itself, along with artefact knowledge. Artefact knowledge refers to prescriptive knowledge integrated into

the artefact as well as descriptive knowledge about the design decisions and the reasoning behind them. The activity processes assist in describing each artefact component, its justification, and what the artefact's intended use case looks like.

The design and development of the artefact are two of the focal points of this project, which is why the effort in this activity was raised compared to the previous activities. Being the least traditional scientific activity due to the creative nature of design and development, the controls and resources of this stage combine the stakeholder views and ideas through consultation and discourse, as well as previous research.

Subchapter 4.3 – Design and Develop Artefact of the implementation and results section describes the results found used to populate the artefact. First, the artefact structure and functionality outlines were used to create a template for the subsequent population of the artefact. Next, accounts are made as to the results found regarding the possible data extraction methods. The subchapters are organised into different approaches to get hold of the required information explored throughout the design and development process.

4.3.1 Artefact Structure and Functionality

To align the requirement with the structural aims, an initial draft displaying the outline of the structural characteristics of the artefact was produced and agreed upon together with InfraSecure, which can be seen in Figure 2. Each component type to be included in the digital model will have at least one extraction method for how to collect information of that specific component's presence in the CPS. Additionally, each component type will have a subset of properties that further describe the component, each with at least one extraction method for how to collect that specific information. The artefact can be further populated by all component types and their subsequent properties according to the requirements provided by InfraSecure and elicited in the previous section. The usage of the artefact is to choose a component type of which data needs to be collected, and by following the arrows, the artefact will lead the user to a suggestion for an extraction method to be used for that purpose.

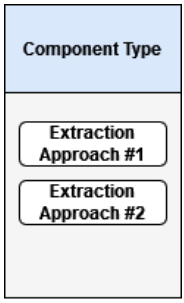

Figure 2. Structural and functional characteristics of the artefact (authors own)

4.3.2 Documentation Results

As part of the collaboration with LocalPower Inc., they provided documentation of their system, including inventory lists, single-line diagrams, and an overview map of high-voltage power lines of the service area. The inventory list is a combined inventory list and service interval spreadsheet that is manually kept by the staff members. It is made up of the physical power grid objects in the system and is divided into internal location codes which symbolise its physical location, e.g., a specific room. Each object has six additional columns further describing the object. The object name specifies what kind of component it is, for example, a transformer, an arc guard, or a high-voltage switch. Although not present for all objects, it also has an internal alphanumeric denotation to identify the physical location further. For almost all objects in the inventory, a vendor is specified, and an additional column further specifies the component with either a model number or a more specific object type is detailed here. This is true for most but not all objects in the list. Each object also has a year column specifying the installation year as well as a quantity specification. In addition to this information, an additional field allows for a brief description, such as an alarm unit or control unit, although this is only prevalent on a few objects in the list. Finally, another column is utilised on about half of the objects. Here, specific voltages are specified or other characteristics, such as the number of racks/compartments in a switchgear.

An obfuscated line from the inventory list is exemplified in Table 3.

While some details of the objects are present, there is a lack of detailed information on many devices. Moreover, the inventory list does not include any objects not directly a part of the physical power process, meaning that no IT devices are present in the list, and OT functionality is unclear to what can be seen in the inventory list.

In addition to the inventory list, LocalPower Inc. provided a single-line diagram (SLD) over the high-voltage physical grid. An SLD is a power engineering block diagram which provides a graphical, symbolic representation of the conductors, whether they are connected or not, circuit breakers, transformers, busbars or other devices in a power system (“Single-line diagram,” 2023). An SLD does not necessarily match the physical size or location of the apparatus presented; it does, however, organise them in relation to one another and provides a high-level abstraction for the PLC control system. While there is no explicit mapping between the inventory list and the SDL, it is possible to match the two documents in order to create a topology of the power grid and determine each object’s location in the topology. A challenge presents itself here as the lack of a direct mapping between the two almost certainly requires internal expertise to merge the information between the two documents. As with the inventory list, the SDL does not include IT components in its representation, and the detail to

Table 3. Inventory list entry

Object	Code	Vendor	Type	Desc.	Year	Quantity	Other
Transformer	A54	Hitachi	ES	n.a.	2008	2	1000 kVA

which components are shown is limited to the device type. No details regarding vendor, model or version are available in the SDL. The power connections between the devices are, however, rather clear.

4.3.3 Command Line Scripting

One approach to obtaining system configuration information is through command-line scripting and querying. While vendor-specific, this approach can extract property information of IT components and detailed information regarding what software is installed on those machines. Compounding an aiding tool containing the commands used to query the sought information quickly becomes a complex task due to the variability of semantics and syntax between different vendors (Jiang et al., 2023). Moreover, should the companies outsource their IT and OT services to third parties, gaining access to the systems used to perform the queries is non-certain within a shorter timeframe since the additional party must be cooperative in order to allow the data extraction to be performed. The functionality can differ depending on the active environment since different vendors provide different solutions and possibilities for command-line interfaces. However, a Windows environment could be examined to exemplify the possibilities for command-line scripting. Windows-based systems offer an infrastructure for the management of data and operation, which one can leverage in Windows Powershell to write scripts to automate specific tasks on local and remote machines (“WMI,” 2023). By using the cmdlet (a command used in Windows PowerShell environments) Get-WmiObject, instances of WMI classes or information about them can be outputted to the terminal or a comma-separated value file (“Get-WmiObject - PowerShell,” n.d.). An example query is:

```
Get-WmiObject -Query "SELECT * FROM Win32_Product"
```

This query returns all installed software products on the machine. The same method can be applied to a remote machine, and by passing the output to the Export-CSV cmdlet, the output can be stored in a file. This lets the operator collect information about the software component properties, including an ID, the vendor, model, version, and build. The missing property, the hosting computer, can be mapped with the local computer (if performed locally) or the ID of the remote computer if one were to perform the query remotely. To create lists of all computers to query in an active directory domain, the cmdlet Get-ADComputer can be utilised, allowing for information collection about all computers in a domain or organisational unit (“Get-ADComputer - ActiveDirectory,” n.d.). This information can be compounded into lists which one can loop over in a script to perform software component extraction of all computers. Another option is to make use of the windows registry to output the same information by using the cmdlet Get-ItemProperty.

In Linux environments, similar commands are available but depend on what package manager the OS uses. Examples include: apt list --installed, and dpkg --list, which lists all software installed using the package manager.

Command-line interfaces can be used to extract information about software and IT and OT components with the limitation of what tools are available from the vendors. A collection of commands used to output software and OS information for Windows and Linux environments are available in Appendix A.

4.3.4 Network Scanning

An analysis of network scanning on SCADA systems shows that, even though the tools might be successful in terms of device discovery, several issues arise when applying standard scanning tools in an ICS system (Coffey et al., 2018). Among the tools analysed (Nmap, Zmap, Nessus, Passive Vulnerability Scanner, and Shodan), all use one or a combination of ICMP ping sweeping, TCP connecting, ARP scanning, and packet sniffing. Furthermore, it concludes the main differences between SCADA networks and traditional commercial IP networks into three main themes: (1) the network implementation, (2) the architecture structure of each node and subnet, and (3) the severity of failing network consequences. The article further highlights the issues that can arise from the discrepancy in communication protocols between the scanning tools and the end devices. The most significant issues are that the returning information of a scan may not solicit a correct response, additional stress put on sensitive SCADA equipment, and that the scans will not return any information or even cause a denial of service or connection disruption. Using these tools without the correct configuration could cause substantial damage to the SCADA devices connected to an ICS rather than helping to inventory and audit them, which is why a network scan on a SCADA system with an all-purpose tool is not a feasible solution.

Holm et al. (2014) showed that, in their EA model, through network scanning on IT networks using the network scanner NeXpose, device information such as MAC addresses, IP addresses and software information such as operating systems versions and application information including port usage, protocol, type, and version for both client and server could be collected with high accuracy. They further conclude that automatic data collection is preferable since it reduces the manual modelling efforts as well as increases the quality of the data since human error is reduced in the loop.

Network scanning tools often offer CLI versions or functionality, which is why the CLI scripting approach also is attributed the same benefits as the network scanning approach (“Nessus,” 2023; “Nmap,” 2023; “The ZMap Project,” 2023).

4.3.5 SCADA System Exports

Another approach is to collect information using SCADA exports. Supervised administrative privileges were obtained through collaboration with the main stakeholder parties of InfraSecure and Local Power Inc. The system documentation for the SCADA system is not publicly available but instead incorporated into the software package of the paying customers of the SCADA vendor. Through access to the system, the documentation became available, which accommodated the possibility of identifying files and system tools interesting from the research perspective. Stemming from this, this project investigated

two main system exports to clarify what information is extractable in the working environment of Local Power Inc., namely, the configuration files and process overview images.

With the aid of a vendor-employed SCADA engineer responsible for the installation and configuration of the SCADA system of focus, the contents of the exportable configuration files were contextualised, and the information available in them was explained. It became apparent that information such as vendors, models, and versions of the physical components or their specific software and firmware details were not included in the exported configuration files simply because that information is not fed to the system during installation and setup. In terms of what was available in the exported configuration files, it was mainly the address and node number for the central data processing service system as well as addresses for all devices included in each communication network, in this case, a total of four separate networks. The networks are communication lines on serial networks and serial COM ports. There were no further details of the devices the addresses were connected to in the configuration files.

The second export was a text file representing a topology overview image of the physical power grid process and the objects involved. The detail of each object is limited to its object ID, which is a globally unique identification name. The overview images, and any other configuration, are manually built by the SCADA engineer based on information provided by Local Power Inc. The text file was easily exported using the SCADA user interface but serves little purpose outside the confinement of the SCADA software tools without a proper interpreter. Should such an interpreter be developed, the topology of the power grid process with object IDs could be translated into ELVIRA representation to explain the device's connections and relationship to one another.

Besides the exported material, insights into the workings of the system are available through the SCADA operator interface. The SCADA system provides navigation of the system objects, defining the physical and logical connections as well as parameters for the software and hardware of the central data processing service system as well as the communication objects and their attributes, which specify the configurations for the handling of the process communication on the serial networks. Using this interface, any configuration changes can also be applied. Granted access, this is a source of great information. However, no embedded export tools accommodate the extraction of this information.

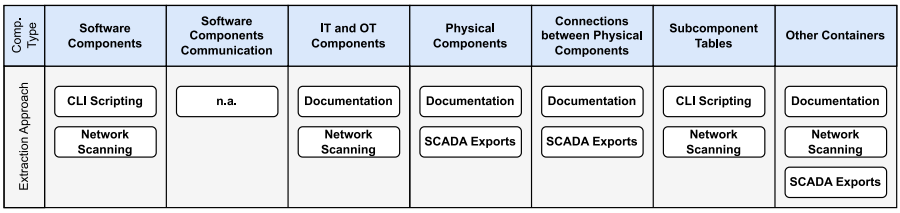

Figure 3. Resulting artefact (authors own)

4.3.6 Resulting Artefact

Combining the results from Chapter 4.3.2 to 4.3.5 with the artefact structure presented in Chapter 4.3.1, the possible data extraction methods are mapped to each component's requirement, as shown in Figure 3. A full-size figure of the artefact is available in Appendix B. In seeking to extract a specific component type's information, the user can refer to the extraction approach mapped to that component type.

4.4 Demonstrate and Evaluate Artefact

The fourth activity combines the demonstration and evaluation of the artefact into one coherent activity. The combined activity revolves around proving the artefact's feasibility, i.e., how the developed artefact can be used to provide a solution to the explicated problem in a case scenario and to evaluate how well the artefact solves the explicated problem per the defined requirements. The input to the activity is simply the artefact and the produced output is a demonstrated and evaluated artefact.

The activity is to be viewed as a proof-of-concept process in which the developed artefact is demonstrated and evaluated using a particular setting. This involves the selection of a case or environment against which the artefact is evaluated. For this project, a suitable base is that of LocalPower Inc. and their business that concerns the campus region with electricity, which is a suitable setting for the artefact demonstration and evaluation. The primary resources needed for this activity are access to and knowledge about the case that will be used to apply the artefact.

Evaluation of the artefact can be divided into two categories for this project: formative and summative evaluation. First, formative evaluation is an ongoing and iterative evaluation that will be performed during the creation of the artefact so that knowledge may be gathered on how to make it better throughout its development, and second, a final summative evaluation seeks to evaluate an artefact after it has undergone final design and development in order to gain a final evaluation of the artefact's usefulness. For both approaches, an *ex-ante* artificial evaluation is deemed the most suitable for this project. An *ex-ante* evaluation signifies that the artefact is evaluated without being utilized in an authentic, real-life setting or even completely developed. Instead of launching a product to be used in an authentic setting and observing how well it performs (as in *ex-post* evaluations), the artefact is evaluated through informed

argument, an evaluation method consisting of the researcher's evaluation of the artefact through reasoning and arguing for the artefacts fulfilment, or non-fulfilment, of the defined requirements (Johannesson and Perjons, 2014). This evaluation method is deemed suitable due to its low resource usage and the fact that informed argument is an often-used method for immature artefacts. It is important to note that this evaluation method is a weak form of evaluation due to the researcher's strong influence on the outcome. A certain risk of bias and false positives emerges from this evaluation selection. This is further discussed in Chapter 5.2 - Validity.

To test the feasibility of the artefact, each data extraction approach for every requirement category in the artefact is mapped to each individual requirement based on the results presented in this implementation and results chapter. This will provide an overview of how the artefact manages in terms of fulfilling the defined requirements. The overview is presented in Table 4. The first column represents the requirements accounted for in Chapter 4.2 – Define Requirements. The subsequent column each corresponds to the approaches explored in the design and development of the artefact. Each cell in the table clarifies whether or not a particular data point could be collected using the approach as explained in this chapter. If the result show that a particular data point could be collected using the approach, the cell is marked with "Yes". If the result does not show that a particular data point could be collected using that approach, the cell will be marked with "No". This does not necessarily mean that the full extent of that approach is exhausted, but rather that this study did not manage to show that the approach could be utilised to collect the sought after information in this scenario. Additionally, in those circumstances where the result showed that a particular data point could be collected but with some limitations, the cell is marked with "Yes" followed by a specification of that limitation.

Table 4. Artefact demonstration overview

Requirements	Documentation	CLI Scripting	Network Scanning	SCADA exports
Software components				
ID	No	Yes, excl. SCADA	Yes, excl. SCADA	No
Vendor	No	Yes, excl. SCADA	Yes, excl. SCADA	No
Model	No	Yes, excl. SCADA	Yes, excl. SCADA	No
Version	No	Yes, excl. SCADA	Yes, excl. SCADA	No
Build	No	Yes, excl. SCADA	Yes, excl. SCADA	No
Software type	No	Yes, excl. SCADA	Yes, excl. SCADA	No
Hosting Computer	No	Yes, excl. SCADA	Yes, excl. SCADA	No

Software component communication				
Sender ID	No	No	No	No
Receiver ID	No	No	No	No
Protocol	No	No	No	No
Purpose	No	No	No	No
Type	No	No	No	No
IT and OT components				
ID	Yes, incomplete	Yes, excl. SCADA	Yes, excl. SCADA	Yes
Vendor	Yes, incomplete	Yes, excl. SCADA	Yes, excl. SCADA	No
Model	Yes, incomplete	Yes, excl. SCADA	Yes, excl. SCADA	No
Version	Yes, incomplete	Yes, excl. SCADA	Yes, excl. SCADA	No
Component type	Yes, incomplete	Yes, excl. SCADA	Yes, excl. SCADA	No
Physical component				
ID	Yes, incomplete	No	No	Yes
Vendor	Yes, incomplete	No	No	No
Model	Yes, incomplete	No	No	No
Version	Yes, incomplete	No	No	No
Component type	Yes, incomplete	No	No	Yes, incomplete
Connections between physical components				
Component ID#1	Yes	No	No	Yes
Component ID#2	Yes	No	No	Yes
Connection type	Yes, only power connection	No	No	Yes, only power connection
Subcomponent tables				
Host machine ID	No	Yes	Yes	No
Software component ID	No	Yes	Yes	No
Other containers				
ID	Room, building	IT-networks	IT-networks	Room, building
Container type	Room, building	IT-networks	IT-networks	Room, building
Description	Room, building	IT-networks	IT-networks	Room, building

The evaluation for the approaches documentation and SCADA exports is based on the empirical information gathered from LocalPower Inc. However, due to resource limitations, the approaches for CLI scripting and network scanning were not applied to the live systems of LocalPower Inc., why the evaluation for these approaches instead relies on the theoretical results shown in Chapter 4.3.3 – Command Line Scripting, and 4.3.4 – Network Scanning.

Neither the documentation nor the SCADA export provided insights into the software components of the power grid CPS. However, CLI scripting and network scanning provide a possibility to acquire the component requirements for the IT networks of the systems. However, the results show no successful results in this regard for the SCADA network. As for software component communication, none of the explored approaches yielded a successful result.

The approaches gave mixed results when collecting IT and OT components data. The documentation included data for all requirements; however, this data was incomplete, meaning that some components could be missing, and some components did not have all properties documented. For exported SCADA files, only the ID of the components could be derived through the topology and configuration files. Furthermore, network scanning was shown to be able to provide component data, excluding components of the SCADA network, e.g., only the office IT network. The CLI scripting approach is attributed the same results with regards to IT and OT components as the network scanning due to the fact that scanning software offer CLI functionality as well.

Neither network scanning nor CLI scripting was shown to provide the property data required for the physical components. Similar to IT and OT components, the documentation did provide all properties for the physical components; however, the documentation was incomplete in this regard as well. This means that some components could be missing and some components did not have all properties documented. The exported SCADA files only contributed with the ID of the components as well as some instances of component type, e.g., transformers. The connections between the physical components were not shown to be able to map through either network scanning or CLI scripting. On the contrary, the documentation and the SCADA export files both provided data on the physical connection; however, only for power connections.

Subcomponent tables are concerned with the software components and on what host machines they reside. This information was not included in the SCADA export files or the documentation. On the contrary, both CLI scripting and network scanning provides this possibility.

The SCADA exports and documentation provided by LocalPower Inc included other containers, such as rooms and buildings. For network scanning and CLI scripting, only the possibility of mapping the IT-networks were found.

Finally, the structural qualities of the artefact are evaluated as well. The coherent design of separating all component types in columns and the corresponding extraction approaches being places in rows underneath the component types makes the logical connection between the two clear. The modularity of the artefact is consistent with the requirements of the ease of which the extraction approaches could be separated and edited. Finally, in terms of conciseness, omitting the exact courses of action to take depending on circumstances in different settings increases the conciseness and makes sense for the current maturity of

the artefact. By aggregating the results from Table 4, the artefact subsequent result for all requirements can be seen in Table 5.

Table 5. Aggregated results

Requirements	Artefact results
Software components	
ID	Yes, excl. SCADA
Vendor	Yes, excl. SCADA
Model	Yes, excl. SCADA
Version	Yes, excl. SCADA
Build	Yes, excl. SCADA
Software type	Yes, excl. SCADA
Hosting Computer	Yes, excl. SCADA
Software component communication	
Sender ID	No
Receiver ID	No
Protocol	No
Purpose	No
Type	No
IT and OT components	
ID	Yes
Vendor	Yes, incomplete/Yes excl. SCADA
Model	Yes, incomplete/Yes excl. SCADA
Version	Yes, incomplete/Yes excl. SCADA
Component type	Yes, incomplete/Yes excl. SCADA
Physical component	
ID	Yes
Vendor	Yes, incomplete
Model	Yes, incomplete
Version	Yes, incomplete
Component type	Yes, incomplete
Connections between physical components	
Component ID #1	Yes
Component ID #2	Yes
Connection type	Yes, only power connection
Subcomponent tables	
Host machine ID	Yes
Software component ID	Yes
Other containers	
ID	Room, building, IT-networks
Container type	Room, building, IT-networks
Description	Room, building, IT-networks

5 Discussion

The discussion chapter of this report first elaborates on the results presented in the previous chapter in Chapter 5.1 – General Discussion. Secondly, the research validity is discussed in Chapter 5.2 – Validity, before the ethical and societal aspects are discussed in Chapter 5.3 – Ethical and Societal Aspects.

5.1 General Discussion

This first section of the discussion concerns the results found in this study and how they compare to previous literature. A general section of the discussion is implemented based on the four data collection approaches presented in the previous chapter.

5.1.1 Documentation

Regarding the two approaches closely connected to the collaboration stakeholders, namely the documentation approach and the SCADA system exports approach, a discrepancy emerged between the documentation provided by LocalPower Inc. and the detail richness present in the SCADA system. This indicates that the information provided to the SCADA engineer used in the process of installing and configuring the system, as well as creating the topology overview image, differentiates from that of the information of the system provided as documentation to this research. As this question was raised to LocalPower Inc., the staff did not know what information was provided to the SCADA engineer during set-up. The same issue is highlighted by Jiang et al. (2023) who conclude that the data availability and accessibility often are unclear due to organisational structures, which make the data inaccessible since there is a lack of directions on who possesses it.

In a study concerning power grid infrastructure and IT and OT networks, the same phenomenon occurs. The study concludes that the information regarding various components of the system is dispersed among both departments in the organisation as well as information held by the components manufacturers, which is a challenge in the development of digital network models (Jiang et al., 2023). To add to this challenge, it is evident from a case study on vulnerability assessment using system information of a large data centre that creating these models heavily rely on documentation as well as systems owners for contextualising the data, a scenario similar to the one presented in this research (Jiang et al., 2023). Hopefully, legislation aids in this regard. In proposing a framework for monitoring performance for NIS directive adherence, asset management is highlighted as a key principle to the objectives the directives aim to achieve (Wallis and Johnson, 2020). They further conclude that while completely achieving all objectives of the NIS directive and its Swedish implementation *Säkerhetsskyddslagen* (2018:585) is a dream scenario, the ambition is that better asset management should be an effect of the directive. In such a case, it would raise the quality of the documentation the organisation could provide to an analysing party, thus making the possible derivatives from that documentation less lacklustre.

5.1.2 SCADA Exports

The results show some of the challenges and limitations of exporting information from the SCADA system. In which the environment the artefact was demonstrated, component properties collected were scarce due to information not being added to the system upon configuration. Instead, component information was limited to possibly entail what type of component it was, and the main focus of the environment was to provide functionality in terms of defining processes and communication. This includes the timing of communication, protocol usage, frequency, etc. (which are also requirements for the artefact). This research was not able to tap into this information since no export tools available through the system could export this information. Thus, more relevant information than what was extracted is likely available in the system, which speaks to the need for more custom extraction tools.

The same problem of information availability regarding data flow and software communication is highlighted in a study where a digital power grid model was constructed (Jiang et al., 2023). Power grid components, including OT and control network components, were modelled. However, the software communication between the OT components and the central control network was omitted due to insufficient information. Getting access to this information would be of great benefit to the creation of holistic digital models.

In this research, component information about mainly physical components and their connections, some instances of IT and OT components, and information regarding their whereabouts in rooms and buildings were available in the export files. The documents and files that were extracted also contained other information, either deemed irrelevant for the purpose of this study or, at the very least, not further investigated in terms of component properties information. This pinpoints a key issue highlighted by Bergs et al. (2021), namely that a data collection challenge is that of asset-specific requirements, which means that only the required data should be collected. Any overreach in this regard is a potential security risk. This study, too, came across this issue with regard to the SCADA Export. This further endorses the need for custom SCADA export tools which can manage to specify more precisely what information to extract.

The extractable topology information can aid the digital model's creation processes, even though it is not an explicit requirement. Instead, the connection between components is the implicit information which can be interpreted from the topology overview.

5.1.3 CLI Scripting and Network Scanning

For the two approaches, CLI scripting and network scanning, similar results were shown with regard to their ability to collect the required data. Their similar performances are not a surprise since network scanning and CLI scripting often offer similar interfaces, making them interchangeable (“Nessus,” 2023; “Nmap,” 2023; “The ZMap Project,” 2023). As stated by Jiang (2023), performing CLI scripting can evolve into a rather tricky task once various vendors are present in a system, each with their syntax and semantics. In summary, the findings indicates that the approaches perform well with software and IT com-

ponents. On the contrary, devices on the SCADA network exposed limitations in the approaches' reach or suitability. This, along with the picture painted by Coffey et al. (2018), highlight the risk associated with using scanning tools developed initially for IP-based networks on ICSs or SCADA systems. For traditional IP-based networks, several node discovery tools exist. However, their suitability to be used on SCADA systems is questionable since they are operationally fragile in terms of service uptime and use other sets of communications protocols compared to traditional IP-based networks.

However, as mentioned, scanning-based tools are performing well in the IT setting, as made evident by Holm et al. (2014) in scanning for IT and software components with great accuracy. While negative impacts, such as service disruption, is risk associated with scanning SCADA networks, future exploration of this approach can still be seen with optimism due to the automation benefits a secure scanning solution would offer, either through responsible active probing or the less intrusive passive scanning or packet sniffing. As concluded by Jiang et al. (2023), achieving automation in vulnerability analysis is of key importance. The work presented in this report concerns the beginning stage of vulnerability analysis, but achieving automation in the initial stages of the process is equally as crucial in terms of not creating bottlenecks for the process as a whole. In the study by Jiang et al. (2023), it also shows that manually modelling an electrical grid CI is time-consuming, calling for automation as a virtue in digital modelling. Nevertheless, the study also concludes that most subsystems share similar architectures and that information about just one subsystem could be collected and reused to save resources.

5.1.4 Summary

In combining the discussion of the above points, it became apparent while designing the artefact that several phenomena were encountered, also mentioned in previous literature. Even before design and development, the premise of problem explication and that no method for data collection and extraction of CPS exists was also encountered by previous research (Holm et al., 2014). While the artefact is relatively immature in completeness, some results are evident from the working artefact. The artefact depends on a high standard of system documentation, expertise in terms of system knowledge to contextualise and connect different data sources to the CPS, and a significant degree of manual data collection and aggregation – a testament to the challenge that a heterogenous system poses for collecting information about it. Heterogeneity is a central challenge to model CPSs, as concluded by Derler et al. (2012). Jiang et al. (2020b) presents a framework supporting the collaboration between actors in the production process of a CPS. Their work shows the weight of keeping stakeholders of different levels included in the vulnerability assessment and highlights the benefit of collaboration in this regard. From the point of view of this work, connections can be made to the difficulties found with documentation. Approaching one or two staff members and asking for documentation is not enough. Instead, different stakeholders from different parts of the organisation must be approached, partly because they might hold different information and documentation but also because they might have other insights into possible vulnerabilities or peculiarities of the system.

Despite providing guidance on how to collect data to create a digital model, the design and development of the artefact came across several challenges in acquiring high-quality and accurate information — certainly a concern for those undertaking similar goals as this project, as agreed on by Khalil et al. (2023).

5.2 Validity

In this subchapter, the validity of the research method and its implementation is discussed in Chapter 5.2.1 – Method and Implementation Validity, followed by a discussion regarding the result validity in Chapter 5.2.2 – Result Validity.

5.2.1 Method and Implementation Validity

The two initial activities of problem explication and requirement definition were heavily influenced by the perception of one of the key stakeholders InfraSecure. InfraSecure provided the initial problem and a pre-determined set of requirements. This scenario introduces participation bias in the sense that InfraSecure's agenda could guide the direction of the results based on the foundation for the design science research method, which is set in these two initial activities. To combat this bias, the problem explication also considers previous literature on the subject to confirm and triangulate the core issues. Similarly, for the requirement definition, while the same concerns regarding participant bias occur here, the validity and reliability of the resulting requirements remain high since they were based on previous ELVIRA research and ConceptBase interoperability.

For the third activity, design and development, both participant bias and researcher bias are introduced naturally due to the creative process of the chosen research method. The resulting artefact is significantly shaped by the researcher, the stakeholders and the subject SCADA environment, which was the testing ground. The risk this introduces is that the direction of the development and, subsequently, the resulting artefact suffers in external validity in how it is generalisable. To combat this potential risk, various resources such as consultation and discourse with key stakeholders, directed literature searches, and empirical data were combined to form a triangulated conclusion and analysis of the findings.

Regarding stakeholder participation and data collection, consultation and discourse with the stakeholders were deemed the most suitable approach. This is because the complex systems under scrutiny were deemed unsuitable for traditional interviews. Instead, the importance was for the researcher to understand what could be possible from different contexts and questions. From the outside looking in, it is challenging to understand the intricacies of the SCADA systems and why it is argued that more traditional interviews with prewritten questions and answers would be an ineffective data collection method. Importance was instead placed on the ability to understand the systems through the eyes of the stakeholder in order to lead the research forward towards potential data collection approaches. Again, no conclusions were made solely on this consultation and discourse. Instead, it is to be seen as a compassing tool during the design

and development activity to develop ideas that were later validated using either empirical data (SCADA system exports and documentation) or literature.

The fourth and last activity was the combination of both the demonstration and the evaluation of the artefact. The execution of this activity includes a few matters that should be discussed in terms of validity and reliability. First, no empirical data was collected in the live environment of LocalPower Inc. through the two approaches, network scanning and CLI scripting. Two main reasons are attributed to this decision, namely that there are risks associated with scanning and probing SCADA systems and that, for the more risk-free IT environment, the time constraints under which this study was performed by those responsible for the IT at LocalPower Inc. were unable to give clearance to allow the researcher to conduct any network scanning or probing. For this reason, the report does not go further in purposing empirical data in this regard. Instead, supporting results in the form of previous literature conclusions yielded the findings that laid the foundation for the artefact demonstration and, further, the evaluation.

As mentioned in Chapter 4.4 – Demonstrate and Evaluate Artefact, informed arguments are used to evaluate the artefact in which the researcher evaluates whether or not the artefact fulfils the defined requirements. This places the researcher at the centre of the process and thus demands a high level of scientific integrity and truthfulness in the evaluation. The main risk this entails is research bias and false positives (Johannesson and Perjons, 2014). This researcher bias risk is still assessed as low, particularly when it comes to the component requirement, as there is a simple mapping between what the result showed to the previously defined requirement. The significance of the researcher's involvement in this regard is negligible. The structural and functional qualities, which were added as soft requirements for the artefact, could be more susceptible to researcher bias in the evaluation since the result include no data on the coherence, modularity, and conciseness of the artefact. For this reason, this result will not affect the conclusions but rather serve as a contribution for anyone willing to develop the artefact further. The non-empirical results (network scanning and CLI scripting) face the most significant risk of false positives since the researcher did not implement the data collection approach to see the results first-hand from the target system. This was combated with supporting evidence from literature, as described previously in this section. Regarding false positives for the empirical data, there is a risk that the approach worked because it managed to collect data to some extent but not for all components in the system. For this reason, the result section clearly states the instances where the documentation or SCADA exports lacked in completeness.

5.2.2 Result Validity

For any research, the question of the generalisability of the findings to extend beyond the scope of the performed study is essential. Perhaps even more so in a study which does not have a population from which a statistically representative subject pool sample has been sampled, as applies to this study. However, as explained in this discussion chapter, triangulation of the design and development activity and literature review findings aims to find the commonalities

between this particular case and other similar CPSs. Any conclusions drawn stem from results appearing in both literatures as well as in this particular case. The extent to which the results are generalisable outside this research scope is, of course, difficult to pinpoint with complete accuracy; however, since the findings are based on multiple sources, it is deemed likely that the results from this design science project would also extend to similar organisations, thus, upholding external validity.

The aim of the research focuses on small-scale power grid CPSs and organisations. With this construct, the LocalPower Inc. case ticks both attribute boxes. Therefore, the construct validity is strengthened by working with the collaborative partner LocalPower Inc.'s system and stakeholders. Furthermore, the produced artefact uses no economically resource-heavy tools or external products, which is an important aspect for smaller-scale businesses that often cannot implement state-of-the-art security solutions (Deloitte, 2021) – a situation this research project aimed to change.

To further add to the topic of result validity, it is important to consider the study's reliability in terms of replicability. Throughout the report, the decision was made to keep the stakeholders and data sources hidden using aliases, and the result data was obfuscated for reasons further discussed in the next chapter, 5.3 – Ethical and Societal Aspects. Consequently, this eliminates the possibility of using the same research subjects for a replicative study and comparing the results. Additionally, due to the naturally explorative and creative process that is the design and development of the artefact, it is most likely that different characteristics of the artefact are to be expected should a similar study be performed. With this in mind, since the results are based on a specific type of organisation and system as well as established through consultation and discourse with relevant stakeholders and literature, similar studies could yield similar results in terms of the challenges and possible solutions to those challenges.

5.3 Ethical and Societal Aspects

When studying security in critical infrastructure, an ethical aspect is the idea of it being a balancing act with risks and rewards. The intended benefit of the study is that bringing attention to possible methods or approaches to essentially perform reconnaissance on critical infrastructure could be information that illicit actors are interested in and possibly lead to malicious attacks on those systems. For this reason, throughout the report, aliases were used to pseudonymize the stakeholder, the SCADA system was kept unspecified, and any presentation of the data collected was obfuscated.

Nevertheless, anyone with malicious intent could go through similar efforts to create similar artefacts, such as the one presented in this report, regardless of this report being publicised or not. The researcher would argue that the societal interest in investigating possible means to gather the data needed to perform vulnerability analysis on critical infrastructure outweighs the risk of focusing malicious intent towards these systems.

The societal interest in protecting critical infrastructures is great since disturbances to the availability of the services they provide could be catastrophic. Therefore, the research community should have access to the information presented in this report to be able to build upon it in a combined effort to make society a more secure place. Furthermore, the main idea behind the information presented in this paper is to serve as a guideline when trying to implement security measures for critical infrastructure, benefiting all organisations within the CPS industry. Any guidance for those responsible for critical infrastructure to improve the security of their system is a step in the right direction, moving towards a comprehensive best practices solution.

To further add to the ethical and societal aspects to which this work contributes, it is linked to the United Nations Sustainable Development Goals (SDGs) – an initiative highlighting 17 separate goals for a global partnership for peace and prosperity for people and the planet (United Nations, 2015). The ninth goal revolves around industry, innovation and infrastructure, and the goal is to build resilient infrastructure, promote inclusive and sustainable industrialisation and foster innovation. There is a strong connection between the report and the ninth goal given that both concern the infrastructure in society. More specifically, targets 9.1 and 9.5 focus on the enhancement of scientific research and technological capabilities invested in industry sectors of all countries as well as the development of reliable and resilient infrastructure. This is closely related to this design science project which aims to, through research, encourage the technological development within critical infrastructure in a resilient manner.

The artefact presented in this report is focused on small-scale power grids. However, extensions of the artefact could be further developed to translate over to other sectors making use of CPSs. For example, SDG 6 – Access to clean water and sanitation, water distribution systems often use cyber-physical functionality. Moreover, SDG 7 – Affordable and clean energy, certainly resides in the same domain as this research, and most targets of this goal are related to the work presented in this report.

Finally, many SDGs targets are aimed at developing countries, and while this project was performed in the setting of a developed country, the availability of the research would make the results usable to anyone in any country since small-scale power grids are present in most, if not all, countries around the globe.

6 Conclusion

As stated in the research aim:

This research aims to investigate the practical issues present in the data collection phase of creating a digital model of a small-scale power-grid CPS.

Two research questions are derived from the discussed topic:

- *What are some of the challenges and limitations of collecting the data necessary to digitally model a small-scale power grid CPS?*
- *What are the possible methods and means to overcome the challenges and limitations of collecting data of a small-scale power grid CPS sufficient for an accurate digital model?*

By procuring the answer to the stated research question, this research aims to investigate methods for extracting information about a small-scale power-grid CPS to accommodate the creation of a digital model serviceable to analysis with ELVIRA threat and vulnerability models.

Through the development of an artefact, the aim has been achieved, and from the study results, answers can be provided to the research questions. By investigating the four data collection approaches making up the artefact, the following challenges and limitations can be concluded:

The heterogeneous nature of the necessary data does not allow for one single approach to be used for all information collection. Furthermore, accessing different sources of data can prove challenging. While documentation may include required information of the CPSs, it can simultaneously have various levels of care and unsatisfactory levels of completeness. SCADA systems are configured with much information required. However, accessing this information is challenging due to its lack of embedded export tools. Taken together, this complicates matters when striving for automation of data collection. On the other hand, network scanning and CLI scripting approaches, which inherently possess greater automation capabilities, were out of reach of the SCADA network and OT devices.

The artefact presented in this report does not provide methods and means to overcome all challenges. However, it shows possible means of collecting data on a small-scale power grid CPS through four approaches to various degrees. In the design and development of the artefact, the four approaches of documentation, SCADA system exports, network scanning, and CLI scripting were investigated and included in the final product. In evaluating the artefact, it showed that, by using the artefact, it was possible to collect at least parts of the data necessary for all requirement categories but software component communication.

The summarised conclusion regarding the artefact is that it can solve parts of the challenges of collecting the data necessary to create a static digital represen-

tation of a small-scale power grid CPS. Furthermore, the artefact needs further development to solve the explicated problem holistically.

6.1 Future Work

At its present state, the artefact is still immature and could be further developed to be even more beneficial to the intended purposes. The development direction can take different aims. For example, one direction is to focus on automating tasks in the artefact to make it more practically viable as a professional tool. Another research direction which would help in this regard is to further research the possibilities and courses of action for performing scanning or sniffing in a CPS environment. This approach showed the most potential in terms of collecting information in an automated fashion on a frequent basis. However, as stated in the report, performing scans on sensitive SCADA equipment is a non-trivial task but, in the researcher's opinion, one worth pursuing.

Another future direction could further investigate the discrepancy that this research came across between the two stakeholders LocalPower Inc. and the SCADA vendor, who had different perspectives on what documentation existed over the power grid CPS. The availability of documentation due to organisational structures and different caretakers of the organisation's information could be further explored in an effort to bring suggestions on how to approach these companies in order to gain access to their complete documentation. If successful, this could be a good starting point in collecting data for a digital model or digital shadow.

Moving forward, the soundness and security aspects of the extraction process have to be considered. Future research could explore what exposure the organisations take on by revealing and handing over this sensitive information. A proper investigation should be conducted to clarify how this process would relate to the NIS directive, as well as *Säkerhetsskyddslagen* (2018:585*)*, and what measures to apply when handling information from critical infrastructures.

Finally, the findings presented in this report conclude that the SCADA system contained considerable information properties on communication flows but that the SCADA systems lacked prebuilt tools to extract it. Future research could explore the possibility of making extraction tools available by partnering with the vendors of the SCADA systems and software developers.

References

Ahmed, M.M., Soo, W.L., 2008. Supervisory Control and Data Acquisition System (SCADA) based customized Remote Terminal Unit (RTU) for distribution automation system, in: 2008 IEEE 2nd International Power and Energy Conference. Presented at the 2008 IEEE 2nd International Power and Energy Conference, pp. 1655–1660. https://doi.org/10.1109/PECON.2008.4762744

Assante, M., Conway, T., Lee, R., 2016. Analysis of the Cyber Attack on the Ukrainian Power Grid. Ind. Control Syst. Secur. Blog 1–26.

Bergs, T., Gierlings, S., Auerbach, T., Klink, A., Schraknepper, D., Augspurger, T., 2021. The Concept of Digital Twin and Digital Shadow in Manufacturing. Procedia CIRP 101, 81–84. https://doi.org/10.1016/j.procir.2021.02.010

Brecher, C., Dalibor, M., Rumpe, B., Schilling, K., Wortmann, A., 2021. An Ecosystem for Digital Shadows in Manufacturing. Procedia CIRP 104, 833–838. https://doi.org/10.1016/j.procir.2021.11.140

Chowdhury, N., Gkioulos, V., 2021. Cyber security training for critical infrastructure protection: A literature review. Comput. Sci. Rev. 40, 100361. https://doi.org/10.1016/j.cosrev.2021.100361

Coffey, K., Smith, R., Maglaras, L., Janicke, H., 2018. Vulnerability Analysis of Network Scanning on SCADA Systems. Secur. Commun. Netw. 2018, 1–21. https://doi.org/10.1155/2018/3794603

D'Amico, R.D., Erkoyuncu, J.A., Addepalli, S., Penver, S., 2022. Cognitive digital twin: An approach to improve the maintenance management. CIRP J. Manuf. Sci. Technol. 38, 613–630. https://doi.org/10.1016/j.cirpj.2022.06.004

Deloitte, 2021. 2021 Future of Cyber Survey.

Derler, P., Lee, E.A., Sangiovanni Vincentelli, A., 2012. Modeling Cyber–Physical Systems. Proc. IEEE 100, 13–28. https://doi.org/10.1109/JPROC.2011.2160929

Dietz, M., Schlette, D., Pernul, G., 2022. Harnessing Digital Twin Security Simulations for systematic Cyber Threat Intelligence, in: 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC). IEEE, Los Alamitos, CA, USA, pp. 789–797. https://doi.org/10.1109/COMPSAC54236.2022.00129

Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union, 2016. , OJ L.

Directive (EU) 2022/2555 of the European Parliament and of the Council, n.d.

Ehrhardt, J.M., Hoffmann, C.T., 2020. The Digital Shadow: Developing a universal model for the automated optimization of cyber-physical production systems based on real-time data. Procedia CIRP 93, 304–310. https://doi.org/10.1016/j.procir.2020.03.069

Get-ADComputer - ActiveDirectory [WWW Document], n.d. URL https://learn.microsoft.com/en-us/powershell/module/activedirectory/get-adcomputer (accessed 5.16.23).

Get-WmiObject - PowerShell [WWW Document], n.d. URL https://learn.microsoft.com/en-us/powershell/module/microsoft.powershell.management/get-wmiobject (accessed 5.16.23).
Gürdür Broo, D., Bravo-Haro, M., Schooling, J., 2022. Design and implementation of a smart infrastructure digital twin. Autom. Constr. 136, 104171. https://doi.org/10.1016/j.autcon.2022.104171
Hevner, A.R., March, S.T., Park, J., Ram, S., 2004. Design Science in Information Systems Research. MIS Q. 28, 75–105. https://doi.org/10.2307/25148625
Holm, H., Buschle, M., Lagerström, R., Ekstedt, M., 2014. Automatic data collection for enterprise architecture models. Softw. Syst. Model. 13, 825–841. https://doi.org/10.1007/s10270-012-0252-1
Infrastructure Resilience – ELVIRA [WWW Document], 2020. URL https://www.his.se/en/research/informatics/distributed-real-time-systems/elvira/ (accessed 1.5.23).
Jiang, Y., Atif, Y., Ding, J., 2020a. Cyber-Physical Systems Security Based on a Cross-Linked and Correlated Vulnerability Database, in: Nadjm-Tehrani, S. (Ed.), Critical Information Infrastructures Security, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 71–82. https://doi.org/10.1007/978-3-030-37670-3_6
Jiang, Y., Atif, Y., Ding, J., Wang, W., 2020b. A Semantic Framework with Humans in the Loop for Vulnerability-Assessment in Cyber-Physical Production Systems, in: Kallel, S., Cuppens, F., Cuppens-Boulahia, N., Hadj Kacem, A. (Eds.), Risks and Security of Internet and Systems, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 128–143. https://doi.org/10.1007/978-3-030-41568-6_9
Jiang, Y., Jeusfeld, M., Atif, Y., Ding, J., Brax, C., Nero, E., 2018. A Language and Repository for Cyber Security of Smart Grids, in: 2018 IEEE 22nd International Enterprise Distributed Object Computing Conference (EDOC). Presented at the 2018 IEEE 22nd International Enterprise Distributed Object Computing Conference (EDOC), pp. 164–170. https://doi.org/10.1109/EDOC.2018.00029
Jiang, Y., Jeusfeld, M.A., Ding, J., Sandahl, E., 2023. Model-Based Cybersecurity Analysis. Bus. Inf. Syst. Eng. https://doi.org/10.1007/s12599-023-00811-0
Johannesson, P., Perjons, E., 2014. An Introduction to Design Science. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-10632-8
Khalil, S.M., Bahsi, H., Dola, H.O., Korõtko, T., McLaughlin, K., Kotkas, V., 2023. Threat Modeling of Cyber-Physical Systems - A Case Study of a Microgrid System. Comput. Secur. 124, 102950. https://doi.org/10.1016/j.cose.2022.102950
Lag (2018:1174) om informationssäkerhet för samhällsviktiga och digitala tjänster, 2018.
Lagar och förordningar gällande NIS [WWW Document], 2022. URL https://www.msb.se/sv/amnesomraden/informationssakerhet-cybersakerhet-och-sakra-kommunikationer/nis-direktivet/lag-forordning-och-foreskrifter/ (accessed 3.20.23).

Leng, J., Wang, D., Shen, W., Li, X., Liu, Q., Chen, X., 2021. Digital twins-based smart manufacturing system design in Industry 4.0: A review. J. Manuf. Syst. 60, 119–137. https://doi.org/10.1016/j.jmsy.2021.05.011
Liu, C., Jiang, P., Jiang, W., 2020. Web-based digital twin modeling and remote control of cyber-physical production systems. Robot. Comput.-Integr. Manuf. 64, 101956. https://doi.org/10.1016/j.rcim.2020.101956
Liu, M., Fang, S., Dong, H., Xu, C., 2021. Review of digital twin about concepts, technologies, and industrial applications. J. Manuf. Syst. 58, 346–361. https://doi.org/10.1016/j.jmsy.2020.06.017
March, S.T., Smith, G.F., 1995. Design and natural science research on information technology. Decis. Support Syst. 15, 251–266. https://doi.org/10.1016/0167-9236(94)00041-2
MSB, 2023. Aktuella EU-regleringar för informations- och cybersäkerhetsområdet [WWW Document]. URL https://www.msb.se/sv/amnesomraden/informationssakerhet-cybersakerhet-och-sakra-kommunikationer/eus-cyberregleringar/ (accessed 4.25.23).
MSB, 2018. MSBFS 2018:8 - föreskrifter om informationssäkerhet för leverantörer av samhällsviktiga tjänster.
Nazir, S., Patel, S., Patel, D., 2017. Assessing and augmenting SCADA cyber security: A survey of techniques. Comput. Secur. 70, 436–454. https://doi.org/10.1016/j.cose.2017.06.010
Nessus [WWW Document], 2023. . Tenable®. URL https://www.tenable.com/node (accessed 5.25.23).
Nmap [WWW Document], 2023. URL https://nmap.org/ (accessed 5.25.23).
Nwogu, C., Lugaresi, G., Anagnostou, A., Matta, A., Taylor, S.J.E., 2022. Towards a Requirement-driven Digital Twin Architecture. Procedia CIRP 107, 758–763. https://doi.org/10.1016/j.procir.2022.05.058
Okoli, C., 2015. A Guide to Conducting a Standalone Systematic Literature Review. Commun. Assoc. Inf. Syst. 37. https://doi.org/10.17705/1CAIS.03743
Pliatsios, D., Sarigiannidis, P., Lagkas, T., Sarigiannidis, A.G., 2020. A Survey on SCADA Systems: Secure Protocols, Incidents, Threats and Tactics. IEEE Commun. Surv. Tutor. 22, 1942–1976. https://doi.org/10.1109/COMST.2020.2987688
Polismyndigheten, 2021. Pågående och avslutade projekt inom Fonden för inre säkerhet (ISF) 2021-06-16. polisen.se.
Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on measures for a high common level of cybersecurity across the Union, repealing Directive (EU) 2016/1148, 2020.
Säkerhetsskyddsförordning (2021:955), 2021.
Säkerhetsskyddslag (2018:585), 2018.
Single-line diagram, 2023. . Wikipedia.
Somers, R.J., Douthwaite, J.A., Wagg, D.J., Walkinshaw, N., Hierons, R.M., 2023. Digital-twin-based testing for cyber–physical systems: A systematic literature review. Inf. Softw. Technol. 156, 107145. https://doi.org/10.1016/j.infsof.2022.107145
Stouffer, K., Pillitteri, V., Lightman, S., Abrams, M., Hahn, A., 2015. Guide to Industrial Control Systems (ICS) Security (No. NIST SP 800-82r2).

National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-82r2

Tam, T., Rao, A., Hall, J., 2021. The good, the bad and the missing: A Narrative review of cyber-security implications for australian small businesses. Comput. Secur. 109, 102385. https://doi.org/10.1016/j.cose.2021.102385

The ZMap Project [WWW Document], 2023. URL https://zmap.io/ (accessed 5.25.23).

United Nations, 2015. Transforming our World: The 2030 Agenda for Sustainable Development.

Wallis, T., Johnson, C., 2020. Implementing the NIS Directive, driving cybersecurity improvements for Essential Services, in: 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA). Presented at the 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), pp. 1–10. https://doi.org/10.1109/CyberSA49311.2020.9139641

Wang, Y., Kang, X., Chen, Z., 2022. A survey of Digital Twin techniques in smart manufacturing and management of energy applications. Green Energy Intell. Transp. 1, 100014. https://doi.org/10.1016/j.geits.2022.100014

Windows Management Instrumentation [WWW Document], 2023. URL https://learn.microsoft.com/en-us/windows/win32/wmisdk/wmi-start-page (accessed 5.16.23).

Appendix A. Software & OS CLI Commands

```
### Using Get-WmiObject ###

# software components
$software = Get-WmiObject -Class Win32_Product | Select-Object Iden-
tifyingNumber, Vendor, Name, Version | Sort-Object Name | Select-
Object -First 5

$software | Format-Table -AutoSize


# OS and build
$os = Get-WmiObject -Class Win32_OperatingSystem | Select-Object
Caption, BuildNumber
$software = Get-WmiObject -Class Win32_Product |  Select-Object Iden-
tifyingNumber, Vendor, Name, Version | Sort-Object Name | Select-
Object -First 5

$os, $software | Format-Table -AutoSize


### Using Windows Registry (Get-ItemProperty) ###

# software components
$software = Get-ItemProperty -Path
'HKLM:\Software\Microsoft\Windows\CurrentVersion\Uninstall\*' |
    Where-Object { $_.DisplayName -and $_.DisplayVersion } |
    Select-Object PSChildName, Publisher, DisplayName, DisplayVersion
|
    Sort-Object DisplayName

$software | Format-Table -AutoSize

# OS and build
$osInfo = Get-ItemProperty -Path 'HKLM:\SOFTWARE\Microsoft\Windows
NT\CurrentVersion' | Select-Object ProductName, ReleaseId, Current-
Build

$osInfo | Format-Table -AutoSize



### LINUX DISTRIBUTIONS


# software components with popular package management systems

# Advanced Package Tool (APT)
apt list --installed  # Documentation:
https://manpages.debian.org/apt/apt.8.en.html

# Yellowdog Updater Modified (YUM)
yum list installed  # Documentation: https://man7.org/linux/man-
pages/man8/yum.8@@yum.html

# Dandified YUM (DNF)
dnf list installed  # Documentation:
https://dnf.readthedocs.io/en/latest/command_ref.html

# Pacman
pacman -Qi | awk '/^Name/{name=$3} /^Version/{version=$3}
/^Packager/{vendor=$3} /^Installed Size/{print
name"\t"vendor"\t"version}'
# Documentation: https://wiki.archlinux.org/title/pacman
```

```
# Zypper
zypper search --installed-only | awk '{print $3"\t\t"$5"\t\t"$9}'
# Documentation: https://en.opensuse.org/SDB:Zypper_manual

# Portage
equery list "*" | awk '{print $1"\t\t"$2"\t\t"$3}'
# Documentation: https://wiki.gentoo.org/wiki/Equery

# Snap https://snapcraft.io/docs/getting-started#heading--listing
snap list --all | awk '{print $1"\t"$2"\t"$3}'
# Documentation: https://snapcraft.io/docs/snap-list

# Flatpak
flatpak list --all | awk '{print $1"\t"$2"\t"$3}'
# Documentation: https://docs.flatpak.org/en/latest/flatpak-command-
reference.html#flatpak-list

# AppImage
# There is no specific command. Manually check the directories where
AppImages are stored.

# Conda
conda list  # Documentation:
https://docs.conda.io/projects/conda/en/latest/commands/list.html

# Debian-based (dpkg)
dpkg-query -l | awk '{print $2"\t\t"$5"\t\t"$3}'
# Documentation: https://manpages.debian.org/dpkg/dpkg-query.1.html



# OS and build

# Ubuntu
lsb_release -a
# Documentation:
https://manpages.ubuntu.com/manpages/focal/man1/lsb_release.1.html

# Debian
cat /etc/os-release
# Documentation: https://www.linux.org/docs/man5/os-release.html

# CentOS
cat /etc/redhat-release
# Documentation: https://access.redhat.com/documentation/en-
us/red_hat_enterprise_linux/7/html/system_administrators_guide/ch-
viewing_system_information

# Fedora
cat /etc/fedora-release
# Documentation: https://www.linux.org/docs/man5/os-release.html

# openSUSE
cat /etc/os-release
# Documentation:
https://doc.opensuse.org/documentation/leap/startup/html/book-
startup/index.html
```

Appendix B. Resulting Artefact

Comp. Type	Software Components	Software Components Communication	IT and OT Components	Physical Components	Connections between Physical Components	Subcomponent Tables	Other Containers
Extraction Approach	CLI Scripting Network Scanning	n.a.	Documentation Network Scanning	Documentation SCADA Exports	Documentation SCADA Exports	CLI Scripting Network Scanning	Documentation Network Scanning SCADA Exports