Complex Systems Informatics and Modeling Quarterly (CSIMQ)
eISSN: 2255-9922
Published online by RTU Press, https://csimq-journals.rtu.lv
Article 253, Issue 45, December 2025/January 2026, Pages 136–159
https://doi.org/10.7250/csimq.2025-45.07

Toward NIS2 Compliance for Multiple Stakeholders with Security Level Evaluation Framework

Mari Seeba[1,2⋆], Tarmo Oja[1,3], Sten Mäses[4], Maria Pibilota Murumaa[3], and Raimundas Matulevičius[1]

[1]Institute of Computer Science, University of Tartu, Narva mnt 18, 51009 Tartu, Estonia
[2]NCSC-EE, Estonian State Information Authority, Pärnu mnt 139a, 15169 Tallinn, Estonia
[3]Cybernetica AS, Mäealuse 2/1, 12618 Tallinn, Estonia
[4]Department of Software Science, Tallinn University of Technology, Ehitajate tee 5, 19086 Tallinn, Estonia

mari.seeba@ut.ee, tarmo.oja@cyber.ee, sten.mases@taltech.ee,
maria.murumaa@cyber.ee, raimundas.matulevicius@ut.ee

Abstract. The revised Network and Information Security Directive (NIS2) aims to achieve a high level of common cybersecurity across the European Union. Several stakeholders, including member states, supervisory authorities, and critical infrastructure service providers, are expected to support this effort by ensuring a high level of information security. Part of increasing security levels involves implementing risk management measures by service providers, but also an evaluation of the security situation and its changes is necessary from the perspective of each stakeholder. Previous research has described NIS2-related activities through user stories that encompass six types of stakeholders with their respective goals in relation to the security level evaluation of organizations. In this article, we examine the real-life implementation of these security evaluation user stories and demonstrate how the framework for security level evaluation (F4SLE) can be utilized to achieve NIS2 compliance within this narrowed scope of security level evaluation. The advantage of F4SLE is that the data can be collected once and then reused to satisfy different stakeholders without imposing an additional reporting burden on entities that must be NIS2-compliant.

Keywords: NIS2 Directive, F4SLE, Security Level Evaluation, Stakeholder, User Stories.

1 Introduction

Organizations are dependent on digital components, and organizations do not exist in isolation. If an organization falls victim to a cyber-attack or in the event of major technical failures, it also affects all related and dependent organizations (e.g., supply chain) or end customers [1]. This relationship serves as an incentive for organizations to implement risk management measures in accordance

⋆ Corresponding author

Reference: M. Seeba, T. Oja, S. Mäses, M. P. Murumaa, and R. Matulevičius, "Toward NIS2 Compliance for Multiple Stakeholders with Security Level Evaluation Framework", Complex Systems Informatics and Modeling Quarterly, CSIMQ, no. 45, pp. 136–159, 2025. Available: https://doi.org/10.7250/csimq.2025-45.07

Additional information. Author ORCID iD: M. Seeba – https://orcid.org/0000-0002-9066-2467, T. Oja – https://orcid.org/0000-0003-3301-7020, S. Mäses – https://orcid.org/0000-0002-1599-8649, M. P. Murumaa – https://orcid.org/0009-0006-4908-637X, R. Matulevičius – https://orcid.org/0000-0002-1829-4794, PII S225599222500253X. Received: 3 October 2025. Accepted: 8 December 2025. Available online: 31 December 2025.

with legal regulations, thereby increasing the overall level of security in the European Union's (EU) digital single market and preventing societal disruptions and economic losses. One such regulation in the EU is the NIS2 Directive [2]. Among other requirements, NIS2 defines tasks that can be directly solved by evaluating the security level. Seeba et al. [3] summarize the corresponding evaluation tasks into ten user stories for six stakeholders: Member State, Supervisory Authority, ENISA, Security Consultant, Entity, and Service Provider & Supplier.

There are several instruments for security evaluation. Maturity model-like instruments are particularly suitable for security evaluation, as they enable the creation of standardized security levels and benchmarks for organizations with varying maturity levels [4], [5]. These maturity evaluation instruments are focused on organizations. The security level evaluation framework F4SLE [6], [7], [8] is inherently designed to meet the needs of organizations and multiple stakeholders, and, following the privacy principles, not to collect centrally sensitive detailed security data [9]. Brezavscek [10] and Rabii [11] highlight the lack of real-life tests and experiments as a significant shortcoming of security maturity assessment instruments. Of the thirty security maturity models examined by Rabii et al. [11], only one-third provided experimental testing results and analysis, with some of the tests being conducted on only one organization. To address this testing gap, we aim to examine the user stories of NIS2 stakeholders [3] in real-world situations, collecting security level data from 284 organizations with the framework for security level evaluation F4SLE. As we already had the artefact F4SLE [6], [7], [8] and user stories [3], we followed the design science research method [12] to demonstrate and evaluate F4SLE's usability.

We followed the principle of collecting the security level evaluation data from organizations once and reusing this data for multiple stakeholders to fulfill the user stories of NIS2. We posed two interrelated research questions:

RQ1. How do different stakeholders fulfill the user stories of NIS2 using the security level evaluation instrument F4SLE? Specifically, we sought confirmation on whether the high-level goals set by NIS2 and the more specific goals set by the stakeholders themselves would be met.

RQ2. What challenges and decisions do stakeholders encounter when applying the security level evaluation instrument F4SLE to NIS2 user stories? Since stakeholders previously lacked similar security level evaluation experiences, we aimed to identify barriers that could either facilitate or hinder the functional and practical adoption of the F4SLE.

In the Background section (Section 2), in addition to the NIS2 Directive context of security evaluation, we introduce the F4SLE framework, along with its data collection and presentation tool, MASS, and usability aspects. Related work (see Section 3) covers European security evaluation tools and incentives. After describing our research method in Section 4, we verify the compliance of F4SLE content to NIS2 Directive risk management measures and describe the user stories testing and usability evaluation results by all six user groups in Section 5. We discuss the evaluation results in Section 6 and, in conclusion, list further research directions in Section 7. The Appendix presents the NIS2 security evaluation user stories from Seeba et al. [3].

2 Background

In this section, we describe the NIS2 Directive in the context of evaluating an organization's security level using the F4SLE framework (see Section 2.1) and the MASS application (see Section 2.2), which is used to implement these user stories. Additionally, we introduce usability principles based on the usability standard (see Section 2.3).

2.1 NIS2 and Security Level Evaluation

NIS2 [2] is an EU directive from 2022 that lists critical sectors (e.g., energy, transport, health, drinking and wastewater, digital infrastructure, ICT service management, public administrations,

space, manufacturing, research). The directive requires establishing the Member States' communication channels and reporting, defines contact points during security incidents, and guides supervisory activities and penalties, but also states risk-management measures for these critical sector entities (organizations). The risk management measures required for entities in these sectors must follow an all-hazard approach. This all-hazard approach can be achieved by applying international or European security management standards, such as the internationally recognized ISO/IEC 27001 [13].

As a result of enumerating specific critical sectors, the number of entities required to implement the necessary measures will increase significantly compared to the previous NIS1 Directive. For instance, in Estonia, with the enforcement of NIS2, this number will increase from approximately 3500 entities to approximately 5500[†]. Obliged entities also include micro-sized, small, and medium-sized enterprises (SMEs), for whom each additional obligation is a financial burden. On the other hand, given the positive purpose of the regulation, by implementing security measures, these organizations protect the functioning of their business processes, customers, business partners, or supply chain [14].

The title of the NIS2 Directive [2] already contains the phrase *high common level of cybersecurity*, which refers to security levels and, in turn, the need to evaluate these levels. Seeba et al. [3] have identified relevant NIS2 security evaluation stakeholders *(Member State, Supervisory Authority, ENISA[‡], Security Consultant, Entity, Service Provider & Supplier)*, and ten user stories that are related to them, which refer to the security level evaluation of NIS2. For instance, Member States need to obtain an overview of the Entities' security level compliance (US1.1) and allocate resources to address the identified vulnerabilities (US1.2). The supervisory authority should prioritize supervisory tasks based on the entity's risk level (US2.1) and assess the entity's compliance (US2.2). A comprehensive list of user stories is in the Appendix. The user stories are described at the same level of abstraction as NIS2. Multiple NIS2 requirements, which share similar goals and apply to the same stakeholder, are consolidated into a single user story. Each Member State should adapt these user stories in accordance with the NIS2 transposing regulations, e.g., in terms of relevant standards and tools, and, if necessary, based on the stakeholders' own functions.

2.2 F4SLE: Framework for Security Level Evaluation

F4SLE is a security evaluation instrument used to evaluate an organization's security [7]. F4SLE concept [7] considers the following requirements: (i) wide coverage of security-related topics, (ii) quantifiable and comparable results, (iii) quick and easy to implement and understand, and (iv) alignment with a security standard(s). F4SLE comprises nearly 200 standard security statements or risk management measures, organized into ten security dimensions (F4SLE version 2023 contained 192 statements, while version 2024 contained 189 statements. According to the F4SLE principles, the number of statements must remain below 200 [15]).

The framework has the following five process dimensions:

- ISMS – information security management, policies and management commitment,
- ORP – organization and personnel, access and rights management,
- CON – concepts and methods, including back-ups,
- OPS – operational work and supply chain management,
- DER – detection and response, incident management,

The framework also has the following five system dimensions:

- APP – applications,
- SYS – IT systems,

[†] Explanatory memorandum on the transposition of the Estonian NIS2 draft law: https://eelnoud.valitsus.ee/main#NMG0WMau

[‡] EUagencydedicatedtoenhancingcybersecurityinEurope.https://www.enisa.europa.eu/

- NET – network and communication,
- IND – IT of industry, robotics systems, and lab equipment,
- INF – infrastructure, cabling, smart buildings, and vehicles.

The F4SLE maturity levels for each security dimension align with the recommendations for the maturity models' critique by Pöppelbuß and Röglinger [16], where the levels are not directly sequential waterfall model, but instead indicate directions for organizational development needs. The four levels of F4SLE evaluate organization cybersecurity awareness (Initial level), documentation (Defined level), practical implementation (Basic level), and sustainability (Standard level). To achieve high security, consistent and high performance must be ensured in all dimensions at all levels.

In the context of continuous improvement, technological advancements, and changes in security techniques, the content of F4SLE statements requires regular updates based on changes to baseline standards or suggestions from threat landscape reports. As F4SLE is compiled using the MUSE method [15], the collected results remain comparable even after the F4SLE update. This approach enables data collection using F4SLE to evaluate an organization's security dynamics over time.

The organization's representative (respondent) should respond to nearly 200 statements, divided into 10 security dimensions. Organization representative should respond by choosing between four color-coded options: (1) Nothing significant has yet been done for the situation described in the statement (indicated with red); (2) The statement is partially in accordance with the description of the situation, but still with significant shortcomings (orange); (3) The statement is reasonably addressed by your organization, but with some shortcomings (yellow); (4) The statement is completely true in the context of your organization (green). The colors indicate the four-level traffic-light color scale, where red signifies not implemented, and green signifies fully implemented. Orange and yellow are positioned between red and green, expecting the respondent to decide whether the statement's implementation status should be on the positive (yellow) or negative (orange) side. Additionally, respondents have the option to mark the question as "not applicable"; such a response is excluded from the calculation of results.

All these levels and dimensions, along with their management need support of a suitable software application. The application should collect the data and provide immediate results and benchmarks to the respondent. To maintain the respondent's security management data privacy, the MASS application was specifically designed to meet the needs of F4SLE [9].

2.3 Data Collection with the MASS Tool

MASS [6] is a confidentiality-oriented web browser application that allows the collection of responses to F4SLE statements, providing an immediate report with benchmark comparisons to the organization and an option to send aggregated results to the central data collection server. MASS converts the F4SLE's four-level scale from qualitative values to quantitative values.

Based on the provided responses, the MASS displays the results to the organization. When the organization chooses to send its data to the central server, only the averaged results are sent. Organizations' results (40 numerical values: averaged 10 security dimension results of 4 maturity levels) are used for sectoral benchmarks (an average of corresponding groups of organizations). In addition to the security evaluation, metadata is collected from organizations. The data is stored in a personalized manner on the central server, but cannot be reversed into raw results or specific security measures. An overview of the data processed and collected via MASS is shown in Table 1. The metadata mentioned in Table 1 can be used for various statistical breakdowns. For instance, each respondent could estimate the time spent answering based on six predefined ranges. Requesting users about time to respond (#ID 10 in Table 1) was selected due to the probable nature and possible length of the answering – the users might refer to documentation or colleagues when answering. Additionally, the answering tab might remain open in a web browser for longer than the actual answering time (e.g., starting in the evening and finishing the next day).

The primary principle of the MASS is that detailed answers are never transmitted outside the user's web browser, and the user has control over whether aggregated data is submitted for collection. This approach was selected because detailed data about information security is classified as limited in various jurisdictions. Also, ensuring the user/organization that the detailed information does not leave the organization, helps the user to provide (subjectively) more honest answers, as they do not need to fear direct information leakage or penalties.

The user also has the option to download detailed answers from a web browser to their own workstation for resuming the process, sharing information within an organization, or reviewing results in the MASS user interface (UI) on a later date. The user also has the option to download aggregated results and a PDF report for review or to pass the results to an external party if needed. When a user decides to participate in data collection, the UI presents a form requesting an organization description, which is submitted along with the aggregated data.

Table 1. Data aggregation elements of the maturity measurement in the F4SLE/MASS setting

ID	Description	Number of items	Scale/options	Collected centrally
1	The organization's responses to individual questions	≈200	0, 1, 2, 3, N/A	No
2	Organization's responses average results per 10 security dimensions per 4 maturity levels (calculated from ID#1)	40	0..3, N/A	Yes
3	Organization's responses average per 10 security dimensions (calculated from ID#1)	10	0..3, N/A	Yes
4	Country	1	2 predetermined + free option	Yes
5	Sector information	1	19 predetermined + free option	Yes
6	Number of employees	1	6 predetermined options	Yes
7	Number of IT/security staff	1	8 predetermined options	Yes
8	Implemented security standards	1	6 predetermined + free option	Yes
9	Role of the responder	1	7 predetermined + free option	Yes
10	Time to respond	1	6 predetermined options	Yes
11	Organization name and contact email	2	Free text	Yes
12	General feedback	1	Free text	Yes

The MASS data analyst verifies the quality of the submitted information, removes duplicate submissions, and adjusts information about organizational metadata (e.g., changing the organization sector to a more specific one) if necessary. The duplicates are removed if a submission from the same entity was made within one month of the previous one, retaining only the most recent submission. If using external data analytics is required, then organization identifiers and feedback fields are removed at this step.

Current data analysis uses Jupyter Notebook[§] workflows. The tool provides statistical and graphical output to be published. For confidentiality, the k-anonymity principle [17] is followed: for each statistical value published, there must be at least 5 measurements whose results can be provided.

Our evaluation focused on the MASS functionalities defined by F4SLE, without separately assessing the tool's ergonomics; hereafter, we refer to both simply as F4SLE.

[§] https://jupyter.org/

2.4 Usability Aspects

Usability is defined in ISO 9241-11:2018 [18] as *"extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use"*. In our study, we used usability aspects based on the ISO 9241-11:2018 [18] standard and its explanatory paper by Bevan [19] to evaluate the usability of F4SLE to achieve the goals of the NIS2 user stories:

- **Effectiveness** – whether users can actually complete their tasks and achieve their goals defined in user stories without negative consequences. The criteria to consider are task completed, objectives achieved, and task error intensity. The prerequisite for effectiveness is the verification that F4SLE content is compliant with NIS2 requirements.
- **Efficiency** – the extent to which users expend resources (time, human effort, costs, and materials) in achieving their goals. In our study context, the resource expenditure (in particular, time) allows users to benefit from it and avoid errors in making assessments and decisions. The criteria to consider are task time, time efficiency, cost-effectiveness, and fatigue.
- **Satisfaction** – the level of comfort users experience in achieving their goals. In our context, it encompasses a positive attitude and emotional feedback, as well as whether users are more likely to recommend the instrument to others. The criteria to consider are discretionary usage, proportion of users complaining, and overall satisfaction.

3 Related Work

The NIS2 readiness survey was conducted on the Swedish infrastructure by Rehnstam et al. [20]. Their semi-formal interviews revealed that readiness is weak and that even the lowest level of NIS2 risk management measures is difficult for many organizations to achieve. Similar results were reflected by other Nordic countries [21]. This means that when evaluating the information security situation, it is not appropriate to focus on high maturity levels, but rather to monitor compliance with low-level security controls to gain insight. Additionally, the entrance barrier for evaluation should be as low as possible.

There are many security evaluation instruments. For instance, through the European Cybersecurity Competence Centre and Network (ECCC), SMEs in EU member states are supported by National Coordination Centers (NCCs). Grants are often based on compliance metrics as a primary indicator of the security situation. Such metrics have been developed over the last few years by countries such as Belgium [22], Austria [23], Ireland [24], Luxembourg [25], Portugal [26], Finland [27], Greece [28], and Estonia [6]. All of the mentioned instruments can be used for self-assessment with the lowest possible entry barrier (downloadable Excel-type file [22], [24], [27], [28], a ready-to-respond online solution without logging in, [6], [23], [25], [26]). The security evaluation instruments differ primarily in their structure, method of question presentation, number of questions, and format of presenting the results. Most often, these instruments provide a summary in the form of a radar chart and a printable PDF document. They are also built in accordance with the nation-state's information security requirements. None of them is visibly used for anything other than the organization's self-evaluation. Only Estonian and Austrian instruments provide benchmarks with other respondents' average results. This means that only Estonian and Austrian instruments collect data centrally, which other users can use to fulfill their tasks and goals.

Drivas [29] proposed its maturity assessment instrument to comply with the requirements of the NIS Directive. Still, his instrument was only tested with theoretical table-top testing and was not designed as a multiple-stakeholder instrument.

Brezavscek et al. [10] emphasize the gaps in their analysis of 36 security evaluation instruments: the lack of simplified instruments for SMEs, the lack of cross-organizational comparison, the lack of automation, the lack of integration with existing systems and best practices, financial barriers, testing in real-world environments, and time-consuming assessment avoidance. Furthermore, we

can add to this list of gaps that even if instruments are developed to fulfill specific NIS2 goals, they are not confirmed to fulfill the goals of multiple stakeholders simultaneously by reusing the data.

4 Method

This work aims to evaluate the usability of F4SLE for implementing the NIS2 security level evaluation user stories, as described in the Appendix and in [3]. We demonstrate and test these user stories empirically, and where empirical demonstrations were impossible, in a conceptual way, following the design science research method [12], [30]. Also, we followed the usability evaluation guidelines defined in ISO 9241-11 [18]. The key usability aspects — effectiveness, efficiency, and satisfaction — were introduced earlier in Section 2.4.

Before evaluating the F4SLEs' usability for NIS2 user stories, we had to verify that F4SLE and NIS2 are compatible and that F4SLE covers NIS2 risk management measures mandatory to Entities – the basic prerequisite for security evaluation. Next, we categorized users into groups based on the defined user stories and reviewed all user stories as described in the following subsection. The study workflow is illustrated in Figure 1.

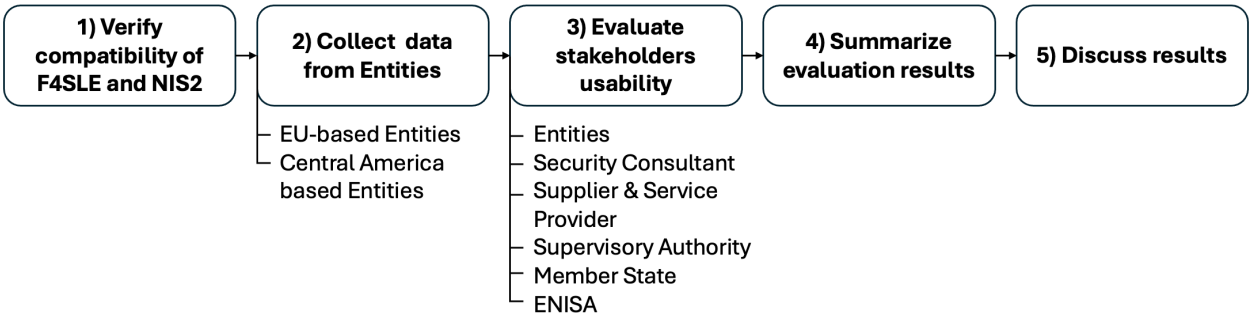


Figure 1. Workflow of evaluating F4SLE usability for NIS2 user stories

User Groups and Data Collection. The user stories of Seeba et al.[3] define six user groups: Member State, Supervisory Authority, ENISA, Security Consultant, Entity, and Service Provider & Supplier (see Appendix). Of these, only *Entities* provided primary security evaluation data (user story US5.1). User groups (Member State, Supervisory Authority, ENISA) relied solely on processing and analyzing this data; no separate data collection was carried out for them. A *Security Consultant* group collected the data in cooperation with the Entity.

Our demonstration and evaluation focused on 284 responses from *Entities* user group participating in the pilot project, divided into two groups:

- **EU-based Entities:** respondents of 225 EU-based entities. These entities were subject to the NIS2 Directive, required to implement risk management measures and comply with relevant national cybersecurity regulations.
- **Control group (Entities of Central America):** respondents of 59 small and medium-sized enterprises (SMEs) from Central America, providing comparative feedback from organizations not subject to NIS2 obligations.

After completing the F4SLE questionnaire, both group representatives were able to provide open-ended feedback and indicate the time required for completion. Additionally, representatives from 18 entities in the control group were invited to provide oral feedback to one of the authors and later anonymously rate their likelihood of recommending the F4SLE to partners, clients, or peers on a 5-point scale.

In the *Member State* user group, the Estonian Ministry responsible for cybersecurity commissioned an analysis of the data collected with F4SLE. The national standards development unit, also acting as a *Member State*, used the results to inform outreach and engagement planning.

Insights from the *Security Consultant* perspective were obtained from two practitioners who had applied F4SLE in real-world consulting engagements.

Based on conceptual demonstration feedback from the *Supervisory Authority* user group, user stories were obtained through an interview with the head of the supervisory division at the Estonian National Cyber Security Centre (NCSC). Due to legal and regulatory constraints, this feedback remained conceptual in nature.

For the user groups *Service Provider & Supplier* and *ENISA*, no direct demonstration took place; instead, for the *Service Provider & Supplier* user group, we collected feedback from the *Entity* group, and about ENISA, we referred to ENISA reports containing aggregated input from *Member States*.

5 Usability Evaluation Results

This section demonstrates the use of F4SLE and assesses its usability in fulfilling NIS2 user stories across different user groups. We begin with verification that the F4SLE context is compliant with NIS2 risk management measures for organizations. We then describe the use of F4SLE for each user group in completing user stories, along with their corresponding evaluation results. Entities contributed the most detailed data and feedback during the pilot observations, making them central to this evaluation.

5.1 F4SLE Content Compliance with NIS2 Risk-Management Measures

Table 2 summarizes the compliance verification mapping for each NIS2 [2] risk-management measure. The NIS2 clause risk-management measures content is followed by the next four columns, which count (in brackets) F4SLE statements per security dimension in each security level. The last line sums up the unique security dimensions of F4SLE for each maturity level. Section 2.2 described that F4SLE has a total of ten security dimensions and summing the unique dimensions shows that NIS2 risk management measures are included at each maturity level.

NIS2 risk management measures are at a very high level of abstraction (e.g., *"Article 21.2.g basic cyber hygiene practices and cybersecurity training"*); NIS2 measures map to multiple security dimensions in F4SLE. For instance, it is essential that cyber hygiene training is covered by personnel management in the ORP security dimension and conceptualized under the CON dimension, while also aligning with the organization's asset management and secure use training (APP, SYS, NET, INF). Therefore, F4SLE includes specific training in several security statements in different security dimensions and maturity levels.

The empty fields in Table 2 can be explained as follows: Article 20.1 of NIS2 requires the management bodies of entities to approve the risk management measures outlined in Article 21 of the Directive. This requirement is addressed through the awareness and accountability of the management and is therefore already implicitly represented across the security dimensions. As such, there is no need to duplicate this general obligation at every maturity level of the F4SLE.

Similarly, Article 21.2.h requires the use and definition of principles and procedures related to cryptography. However, the initial awareness of this topic for employees is expected to be covered under cyber hygiene training, as required by Article 21.2.g. Thus, cryptographic practices are not included as a separate statement at the initial level of awareness in F4SLE.

In summary, the mapping confirms that at every maturity level of the F4SLE, all security dimensions are addressed, and these in turn fully cover the risk management obligations defined in Articles 20 and 21 of the NIS2 Directive.

Mapping each F4SLE statement to the relevant NIS2 clause is available to respondents via the MASS tool [6], providing transparent traceability and legal context for each security measure. The content of F4SLE is also harmonized with ISO/IEC 27001 [13], both in terms of the principles (see

Table 2. NIS2 [2] clauses corresponding to security dimensions and levels of F4SLE

F4SLE Security Dimensions divided into Levels			
Initial	**Defined**	**Basic**	**Standard**
Art.20(1) Member States shall ensure that the management bodies of essential and important entities approve the cybersecurity risk-management measures taken by those entities in order to comply with Article 21, oversee its implementation and can be held liable for infringements by the entities of that Article.			
ISMS(1)			
Art.20(2) Member States shall ensure that the members of the management bodies of essential and important entities are required to follow training, and shall encourage essential and important entities to offer similar training to their employees on a regular basis, in order that they gain sufficient knowledge and skills to enable them to identify risks and assess cybersecurity risk-management practices and their impact on the services provided by the entity.			
ORP(1)	ISMS(1)	ORP(1)	
Art.21(1) Member States shall ensure that essential and important entities take appropriate and proportionate technical, operational and organizational measures to manage the risks posed to the security of the network and information systems which those entities use for their operations or for the provision of their services, and to prevent or minimize the impact of incidents on recipients of their services and on other services.			
	ORP(1)	ISMS(1)	
Art.21(2.a) policies on risk analysis and information system security;			
CON(1), INF(1), ISMS(1), NET(2), OPS(1), SYS(2)	APP(3), CON(3), IND(1), INF(2), ISMS(1), OPS(2)	APP(4), IND(2), INF(1), NET(2)	APP(2), IND(1), INF(1), ISMS(1), ORP(1), SYS(2)
Art.21(2.b) incident handling;			
DER(1), OPS(1)	DER(2)	DER(3), OPS(3), ORP(1)	DER(3), NET(1), OPS(2)
Art.21(2.c) business continuity, such as backup management and disaster recovery, and crisis management;			
CON(1)	CON(1), DER(2), ORP(1), SYS(1)	CON(1), DER(1), INF(2), NET(1)	CON(1), DER(1), IND(1), OPS(1), SYS(2)
Art.21(2.d) supply chain security, including security-related aspects concerning the relationships between each entity and its direct suppliers or service providers;			
INF(1)	IND(2), OPS(4), SYS(1)	CON(1)	CON(1), OPS(2)
Art.21(2.e) security in network and information systems acquisition, development, and maintenance, including vulnerability handling and disclosure;			
APP(2)	APP(1), CON(1), IND(1), NET(2), OPS(3), SYS(1)	CON(1), DER(1), IND(1), NET(2), OPS(3), SYS(3)	APP(3), CON(1), DER(2), IND(1), OPS(2), SYS(6)
Art.21(2.f) policies and procedures to assess the effectiveness of cybersecurity risk-management measures;			
DER(1)	OPS(1)	ISMS(2)	DER(2), NET(2), SYS(1)
Art.21(2.g) basic cyber hygiene practices and cybersecurity training;			
ORP(1)	NET(1)	CON(1), INF(1), OPS(1), ORP(1), SYS(1)	APP(1), INF(1), ORP(1)
Art.21(2.h) policies and procedures regarding the use of cryptography and, where appropriate, encryption;			
	CON(1)	NET(2), OPS(1), SYS(1)	APP(1), CON(2), ISMS(1), SYS(2)
Art.21(2.i) human resources security, access control policies and asset management;			
APP(1), IND(1), OPS(1), ORP(1), SYS(2)	APP(1), INF(4), ISMS(2), ORP(5), SYS(1)	APP(3), IND(3), INF(1), ORP(3), SYS(4)	APP(2), CON(1), IND(1), INF(3), NET(2), ORP(2), SYS(2)
Art.21(2.j) the use of multi-factor authentication or continuous authentication solutions, secured voice, video and text communications and secured emergency communication systems within the entity, where appropriate			
ORP(1)	ORP(1)	APP(2), ORP(2)	CON(1)
Art.21(4) Member States shall ensure that an entity that finds that it does not comply with the measures provided for in paragraph 2 takes, without undue delay, all necessary, appropriate and proportionate corrective measures			
		ISMS(1)	
Mentions of unique F4SLE security dimensions per maturity level:			
10	10	10	10

the Appendix of [15]) and the references within the MASS tool. F4SLE content with mappings is also available [31].

Based on these mappings, we verified that the F4SLE framework is consistent and aligned with the risk management measures of the NIS2 Directive, which are obligatory prerequisites for ensuring the efficiency (see Section 2.4) of all user stories for all stakeholders' user groups.

5.2 EU-Based Entities: Empirical Demonstration and Evaluation

Between March 2023 and May 2025, 225 EU-based respondents completed the F4SLE questionnaire and submitted their results to a central server. The activity sectors of respondents' organizations and the versions of F4SLE used as part of the collected and aggregated data (see Table 1) are summarized in Table 3. Sector classification is based on the NIS2 Annexes [2]. All respondents who sent their data to the server received their individual all-hazard approach security status results for each security dimension, showing strengths and improvement needs, as well as the benchmark for comparison with others. It should be noted that we excluded from the scope of this study the organizations that did not share their data with the F4SLE server. Still, these excluded organizations were able to receive individual security status results, but were unable to get a benchmark, as they did not contribute to it.

Table 3. Distribution of organizations located in the EU by sector of activity of the organization and by F4SLE content version to which the organization's representative responded.

Sector	F4SLE ver2023	F4SLE ver2024	Total
Education and research	21	77	98
Municipality	14	18	32
Public administration	6	25	31
Water and district heating	9	8	17
ICT	3	9	12
Healthcare	11	0	11
Energy	2	3	5
Non-profit	4	3	7
Transport	6	0	6
Manufacturing	0	2	2
Other	1	3	4
Total	77	148	**225**

Representatives of the Entities were invited to participate voluntarily in a sectoral cybersecurity seminar organized by the Estonian National Cybersecurity Centre (NCSC-EE) to support data collection with F4SLE before the seminar. In addition to other seminar topics, we interpreted and discussed the results, focusing on the improvement needs. Seminar participants agreed with the aggregated results trends, which support the reliability of self-assessment. Sector-specific seminars were conducted for healthcare, education, municipal, transportation, water supply, and district heating, as well as public sector institutions. No specific seminars were organized for other sectors; therefore, no sufficient data could be collected from them.

To characterize the efficiency aspect of F4SLE usability, Table 4 and Figure 2 shows completion times for the F4SLE questionnaire. Statistically, the mode of time spent was approximately one hour (45%), which is shorter than security standard introduction training (half a day to 4 days) or security audits (a few working days), but comparable to national-level cybersecurity surveys conducted by the Statistics Authority (e.g., 112 minutes in Estonia).

Free-text feedback was optional and used by 45 of the 225 respondents. Verbal, not formal, feedback collected during seminars indicated that organizations answering the questionnaire multiple times reported increased completion times (from 30 minutes initially to 2 hours later) due to a deeper understanding of their security posture and the need to consult supporting evidentiary documents. This led to more accurate responses over time.

Written feedback criticized the terminology, noting that many SMEs lacked familiarity with cybersecurity vocabulary. Frequent comments included: *"We have service providers, we don't know!"* (similarly 10 times), insufficient maturity ("We would like additional explanations" – 9 times), and unclear language (*"Very specific vocabulary, we don't understand it"*, *"Complicated sentences"* – 7 times).

Nevertheless, 29 respondents noted that the questionnaire offered a useful, quick overview of standard security controls and started to give suggestions for improvement (*"Perfect questions showed me what to focus on in our security strategy. I am glad to participate."*) Suggestions for F4SLE improvement included: adding a comment field, filtering questions when services are outsourced, and including references and direct links to legal frameworks or standards or adding additional explanations or examples.

In conclusion, by verifying the F4SLE all-hazard approach in the context of NIS2 (see Section 5.1), the results of the security evaluation provide organizations with an overall report of their security status, their strengths, and options to highlight improvement needs. Based on the time spent, respondents found that using F4SLE to evaluate the security level of the Entity was faster (in terms of efficiency) and more informative (in terms of effectiveness) than traditional manual gap analysis methods (specific training or IS audit). However, additional interpretation support (such as seminars and sectoral summaries) was necessary to interpret the results and use them for their security implementation plan. Rather than being driven by spontaneous interest, participation was motivated by the added value of seminars, training, benchmarking, funding options, or the possibility of audit substitution (in terms of usability satisfaction came through external benefits).

Table 4. EU-based Entities (n = 225) and Entities of Central America (n = 59) distribution by time to respond F4SLE questionnaire.

Time	EU-based Entities (n = 225)		Entities of Central America (n = 59)	
	Respondent count	Of total %	Respondent count	Of total %
Approx. 30 minutes	30	13%	16	27%
Approx. 1 hour	102	45%	25	42%
2 hours	56	25%	14	24%
2–4 hours	26	12%	3	5%
4–8 hours	9	4%	1	2%
More than 1 working day	2	1%	0	0%

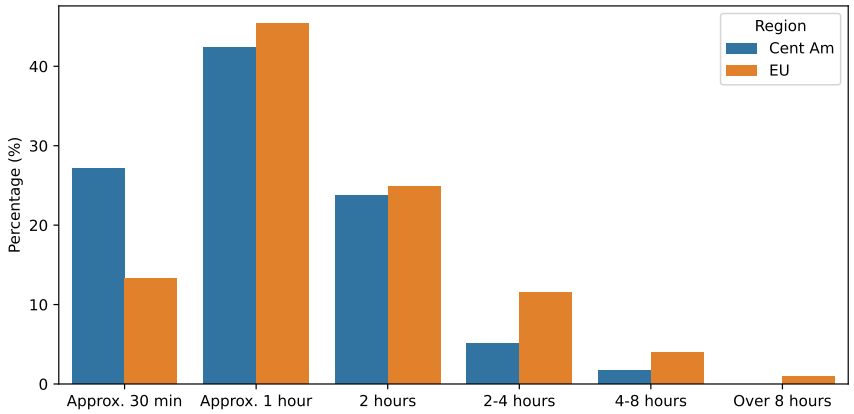


Figure 2. Entities distribution by time to respond F4SLE questionnaire, EU n = 225 and Central America n = 59

5.3 Entities of Central America: Empirical Demonstration and Evaluation

As a control group, representatives of 59 SMEs from Central America participated in the F4SLE pilot project. The pilot project was independent of EU legislation to minimize the bias introduced by regulatory obligations. Participation was linked to a 3-day cybersecurity training program organized by LAC4 of CyberNET[¶], where completing the F4SLE questionnaire was a prerequisite for the training.

Out of 59 participants' responses to F4SLE, 24 provided written feedback, primarily offering general company information. Only nine responses directly addressed the evaluation, all of which were positive (e.g., *"Evaluation was exciting and covered all security issues!"* or simple *"Thanks!"*). However, since completing the F4SLE was a prerequisite for participating in the seminar, we cannot consider the feedback provided as independent feedback.

Similar to EU-based Entities, the mode of time to complete was approximately 1 hour (42%). The detailed distribution is illustrated in Table 4 and Figure 2.

In the classroom training, 18 SME representatives participated, with each SME represented by a management board member and a security or administrative staff member, for a total of 36 participants. These SMEs, operating across six Central American countries (Belize, Costa Rica, El Salvador, Guatemala, Honduras, Panama) and various sectors, had limited prior exposure to security standards, and only one organization reported experience with regulatory compliance. Unlike EU respondents, they framed cybersecurity primarily in terms of business risk rather than regulatory adherence. Each of the 18 SMEs received individual F4SLE result interpretation sessions during the training. The interpretation was reflected in the final part of the training, where all 18 organizations introduced their security implementation plans and improvement priorities based on their F4SLE results.

Additional feedback on F4SLE was collected verbally during the 18 private sessions in the form of open-ended conversations. During these conversations, talking about the possible benefits or shortcomings of F4SLE, all participants recommended F4SLE to others, especially subcontractors (7 mentions), other SMEs (7), parent companies (1), and customers (1). Eight participants noted that the evaluation revealed previously unrecognized vulnerabilities, even among organizations that considered their security posture strong. Two acknowledged that security management had been *ad hoc*, with limited resource planning or management involvement. Seven SMEs reported difficulties understanding the questions due to a lack of prior knowledge, but also stated that the process was educational. They appreciated discovering unfamiliar security topics and felt encouraged upon recognizing controls they had already implemented. One SME highlighted that F4SLE questions were more precise than ISO/IEC 27001 controls. One CEO noted that the radar diagram format was intuitive for C-level stakeholders: a smooth, circular shape was considered ideal, prompting targeted investment in weaker dimensions if the shape was uneven. One participant emphasized that the F4SLE results could justify security-related budgeting, as they provided a transparent and interpretable overview of the organization's security posture.

All participants valued the opportunity for a private consultation, which helped them understand their results and the interrelation of security dimensions. Many suggested repeating the security evaluation after six months, which is when they expected to be able to interpret the results independently.

An anonymous feedback from control group participants was collected by LAC4 of CyberNET after the training. In an anonymous survey, 27 of all 36 participants rated the necessity of F4SLE and their likelihood of recommending it to partners, clients, or subcontractors, using a 5-point Likert scale (1 = "would not recommend", 5 = "would strongly recommend"). Of these 27 participants, 23 rated it as "5", and four rated it as "4", resulting in an average score of 4.85. This reflects a high

[¶] https://www.lac4.eu/event/training-for-strengthening-cybersecurity-skills-for-smes/

level of approval and a strong perception of its value and applicability to other organizations (in terms of all three usability aspects).

However, it is worth mentioning that there were no additional voluntary respondents in the control group after the training, as it did not offer any additional benefits (e.g., the opportunity to participate in the training, as the control group did).

Based on feedback from both EU-based respondents and the control group, we can confirm that user stories for evaluating the security situation and identifying vulnerabilities that require improvement are achievable. Still, Entities require additional external motivation to become an F4SLE respondent, which is the prerequisite for user stories that rely on the entity's security data.

5.4 Security Consultant: Empirical Demonstration and Evaluation

We collected F4SLE usage experience from two consultants. One security consultant utilized MASS and F4SLE for two Entities in the healthcare sector during the security consultation service aimed to implement the Estonian security framework E-ITS [32]. Relevant F4SLE statements were evaluated with the Entity's representative to gather an understanding of the technologies utilized and the state of implemented security measures. When necessary, the security consultant provided additional clarifications (e.g., definitions of technical terms or illustrative examples) to ensure the participant's full understanding and accurate responses.

The security consultant's expectation was to facilitate an efficient and structured approach for understanding the security posture of the external Entity, including identifying and prioritizing areas of improvement, reassessing measures after changes, tracking progress, and demonstrating results. Both US4.1 and US4.2 goals were achieved.

The security consultant expressed overall satisfaction with F4SLE functionality but stated that certain enhancements could improve the efficiency and satisfaction. Particularly, adding a feature for exporting all attributes with their corresponding states would improve the overall workflow of task prioritization.

The second consultant used F4SLE after the interview and observations to confirm the all-hazard approach of the conducted review of a manufacturer. The consultant responded to the F4SLE questionnaire himself based on the observation and interview results. He used the radar diagram from the F4SLE result page to illustrate the organization's security posture in the final consultation report and, based on the F4SLE results, he created a security improvement plan. Both consultants' expectations were to facilitate an efficient and structured approach for understanding the security posture and identifying and prioritizing areas for improvement (US4.1). The second consultant did not have the option to follow the improvement process and see the changes (US4.2) because this was not in the scope of the project. However, he reflected that the use of F4SLE helped him quickly identify the most vulnerable areas and suggest risk management measures for improvement (using F4SLE statements and references to ISO27002 [33]). He related the findings to the MITRE ATT&CK framework [34] to justify the cyberattacks and risks associated with the identified vulnerabilities. Additionally, he requested the option to use F4SLE in his next projects, even when they are not the subject of NIS2, because F4SLE's all-hazard approach and the low entrance barrier to its use are relevant to all organizations that depend on the digital networked assets.

5.5 Member State: Empirical Demonstration and Evaluation

Member State representatives emphasized their expectations when analyzing the F4SLE results:

- overview across different sectors, sector-specific results;
- results should be comparable to each other so that good practices can be found and shared;
- find problematic areas to focus on;
- observe changes through time to evaluate the effect of support measures and policies;
- easily understood but detailed enough to guide the efforts.

Those expectations were well-aligned with NIS2 user stories (see the Appendix), as the main focus was on raising general awareness and ensuring effective resource allocation.

The data analysis and visualization were done in iterations in collaboration with the representatives of the Member State. Significant effort was put into ensuring the fast understanding of the visualizations (e.g., radar diagram, red-yellow-green areas to simplify the interpretation of problematic dimensions). For that, the maturity levels were given simplified labels:

- initial = awareness,
- defined = documentation,
- basic = practicality,
- standard = maturity.

Those labels helped quickly communicate the essence of different levels to people unfamiliar with the F4SLE framework.

Visualizations were divided into three graphs: General (average); defined and basic (comparing documentation with practically implemented security measures); and initial and standard (comparing the essential activities to the more difficult measures). Those three graphs were well received, as it was quite easy to generate an interpretive storyline based on them (a narrative to give meaning to the generalized results).

As a general result, the areas of ISMS and DER were outlined as needing additional resources and attention. It was also evident which type of organizations (what sector) need support to raise their security levels. Based on these findings, NCSC-EE prepared e-training for top management[‖] and conducted engagement seminars for NIS2 subjects on digital forensics and incident management.

As general feedback, the Member State representatives were satisfied with the analysis and looked forward to gathering the same metrics regularly to observe changes over time and thereby estimate the effect of their supportive measures. The Member State representative started to communicate and disseminate the results in his presentations.

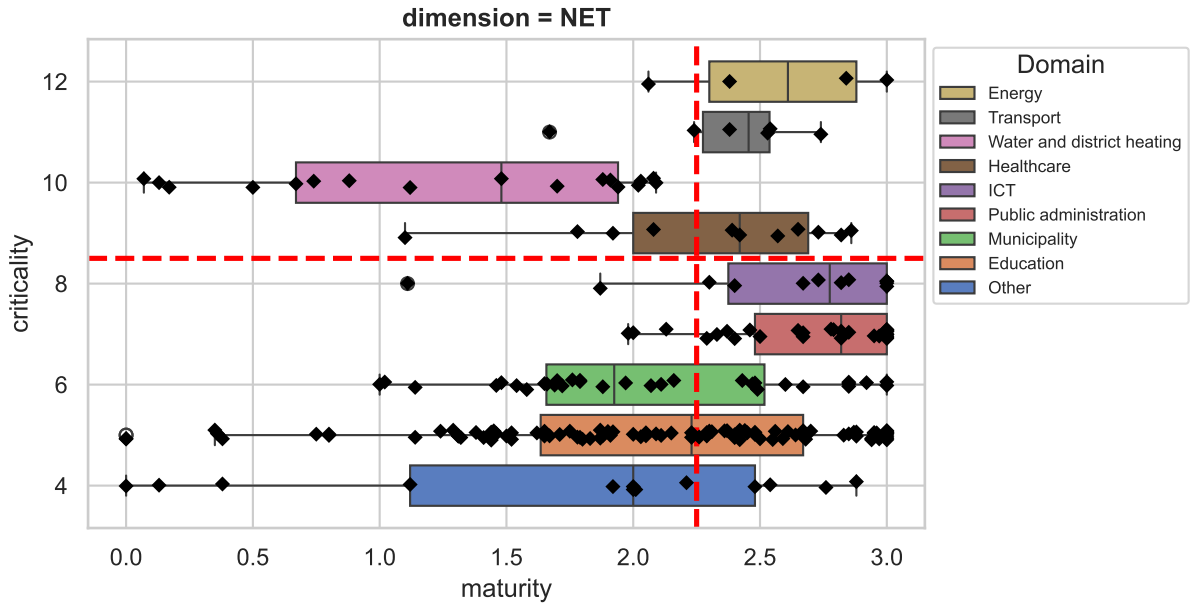
5.6 Supervisory Authority: Conceptual Demonstration and Evaluation

We explored a conceptual demonstration and evaluation of how the Supervisory Authority might utilize the results of the F4SLE. The Supervisory Authority must rely on facts and has the right to request all security-related documents from the Entity, conduct interviews and observations, testing, and scanning. On the one hand, the Supervisory Authority is interested in streamlining and speeding up procedures, focusing on weaknesses to identify them in a timely manner and eliminate deficiencies. At a moment, they walk through the entire security management system of the Entity via sampling, with no optimization options implemented.

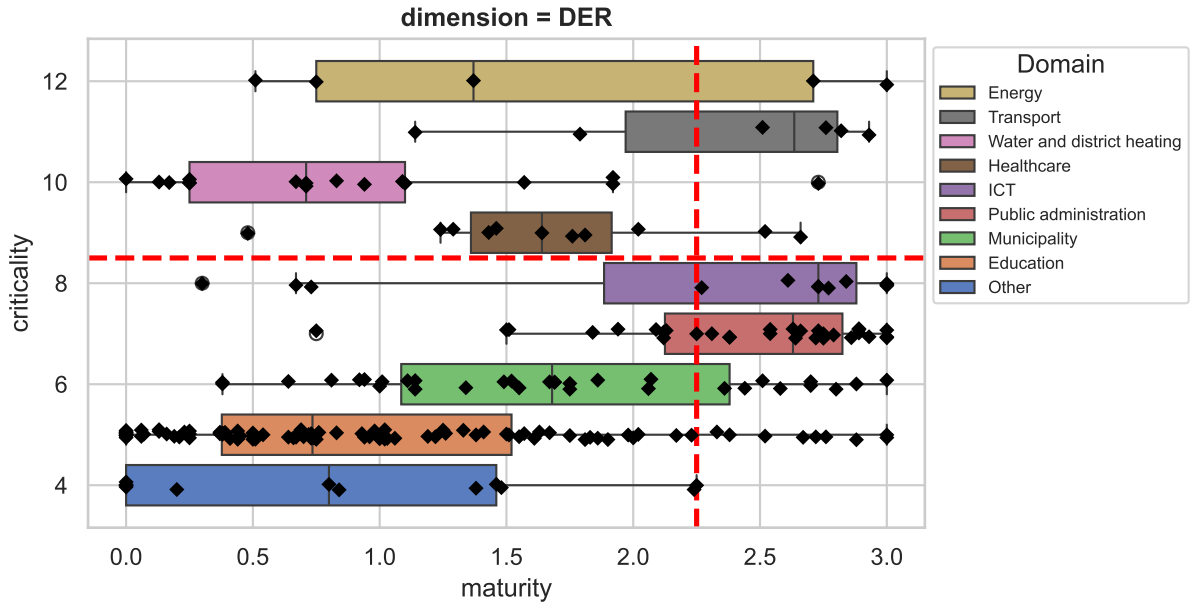
We conducted the interview with the NCSC-EE supervisory authority leader to demonstrate the F4SLE results in the context of how the Supervisory Authority could use them. The leader of the Supervisory Authority confirmed that in the near future, the Supervisory Authority plans to focus on NIS2 obliged Entities by sector and is interested in specific potential vulnerable topics, e.g., network (covered by the NET dimension of F4SLE), incident management (DER dimension), access management (ORP dimension), supplier management (OPS dimension), etc. This aligns with the expectations described in the ENISA report, ‘Cybersecurity Maturity & Criticality Assessment of NIS2 sectors’ [35]. Following the ENISA’s ideas, we presented the F4SLE results to the leader of the Supervisory Authority using box-plot diagrams of sectors and security dimensions’ maturity (see Figure 3). The assessment of the criticality of the NIS2 sectors was outside the scope of this study; the Member State should define the criticality of the sectors. In our demonstration, we used the adapted ENISA’s [35] criticality assessment results. We divided the graphs into four sections with red dotted lines (see Figure 3). We used the adapted sectors’ criticality value (more than 8), and the F4SLE collected data aggregated security maturity value, which should exceed 2.25. This allows

[‖] https://digiriigiakadeemia.ee/

for prioritizing the need for supervisory intervention of the Entities' activity sectors on a specific security dimension. Highly critical and low-maturity areas of the graph have the most significant impact on societal functioning (top-left section in both figures shown in Figure 3). They are followed by lower criticality and lower maturity (bottom-left). The third priority is a high-criticality and high-maturity area (top-right), and the fourth priority for supervision intervention is the Entities and their security dimensions, which have lower criticality but high maturity. The leader of the Supervisory Authority agreed that these priorities can be used when planning supervisory work to focus on sectors and their vulnerable areas (security dimensions), thereby reducing the overall supervisory workload in conducting procedures.



(a) NET security dimension



(b) DER security dimension

Figure 3. Supervisory Authority's view of security evaluation results for individual security dimensions: (a) NET and (b) DER.

Our example demonstration to the leader of the Supervisory Authority illustrates two security dimensions of F4SLE: NET and DER. Figure 3 shows that the network and communication (NET)

of water and district heating companies should be among the top priorities for supervision to fix the possible vulnerable issues (see Figure 3a). However, in the case of detection and response, incident management (DER), there is a very high degree of dispersion, and it is recommended to include incident management-related topics in the supervision of entities across all sectors (see Figure 3b).

In addition, the Supervisory Authority leader is interested in utilizing the other characteristics described in Table 1 to pre-analyze its work, if the sample of responding organizations is more representative. The Supervisory Authority may use an Entity's F4SLE result report or request the full raw dataset (all 200 questionnaire responses) to support its assessment. Repeated completion of F4SLE is not required; the existing results file can be reused.

The Supervisory Authority leader asked for the reliability of the collected data. We provided access to the organization's responses to individual F4SLE statements that allowed the Authority to validate the self-assessment through a small sample, including documentation review, technical testing, or scanning. If the results are consistent and meet the expected criteria, the procedure can be concluded efficiently. In cases of inconsistency, only the relevant areas needed further scrutiny – full system reassessment being unnecessary, thereby reducing the burden on both the Entity and the Supervisory Authority.

While F4SLE streamlines the process, self-assessment carries inherent risks, such as over-reporting or focusing only on included controls. Therefore, although the Supervisory Authority cannot rely solely on F4SLE results, the instrument still offers substantial potential for procedural efficiency.

5.7 User Stories without Demonstration

5.7.1 Service Provider & Supplier

In practice, we could not directly observe the user stories of *Service Provider & Supplier*, because we did not reach the pair of partner Entity and its Supplier. However, our control group consisted of entities that identified themselves as service providers, and they noted that they could use F4SLE to evaluate the security level of their services to share the results with their partners. This, however, would require sufficient awareness and motivation from those partners. For instance, in the EU group, at least ten respondents in their feedback expressed a desire to involve the *Service Provider & Supplier* in the response process to fully understand the organization's security posture fully. This feedback indicates a limited awareness of service provider and supply chain risks; service providers and suppliers are often trusted by default.

5.7.2 ENISA

According to its user story, *ENISA*'s goals largely overlap with those of *Member State*s, but their additional goal is to compare the Member States. From this perspective, Member States can share collected data with ENISA.

In December 2024, ENISA published the *"2024 Cybersecurity report"* [36] to fulfill the NIS2 Art. 18 [2] reporting obligation, which is defined as the user story US3.1 by [3]. The report is based on the EU Cyber Security Index (EU-CSI) method [37] for collecting and interpreting data. The method for integrating Member State Entities' security-related information considers the EUROSTAT, Shodan, and ISO certification databases [37]. Each database has its limitations. For instance, Shodan can access and assess only services connected to public networks (not internal security management issues); EUROSTAT is based on national statistics surveys, giving answers to only limited, separate questions, not an all-hazard approach; ISO issued certifications do not consider compliance with national standards and related audits, and were outdated (results from 2021). ENISA admits in the report: *"A harmonized approach for collecting sector-relevant data could be developed. Member States are encouraged to assess and monitor the maturity and criticality of sectors at the national level."* [36].

Therefore, ENISA needs a relevant standardized instrument, such as F4SLE. It is also worth mentioning here that the ENISA NIS2 360 report [35] (mentioned in Section 5.6), which assesses the criticality and maturity of NIS2 sectors, does not compare Member States. In contrast, Table 1 shows that F4SLE also collects metadata for statistics by country. F4SLE provides a sector maturity assessment for Member State user stories (see Section 5.5) to illustrate the aggregated performance of organizations in cybersecurity capabilities and awareness, as well as to identify gaps. Therefore, F4SLE may have the potential to ensure and fulfill the goals of ENISA's user story; however, a detailed analysis of ENISA's needs, opportunities, and capabilities, as well as those of Member States, is required.

5.8 Usability Evaluation Results Summary

Based on the definition of usability in Section 2.4, we evaluated whether a specific user was able to achieve their specific goal described in the NIS2 user stories (see the Appendix) with the help of F4SLE, and whether the use of F4SLE ensured effectiveness, efficiency, and satisfaction.

The prerequisite for achieving the effectiveness of the user stories was the compliance of the F4SLE content with the NIS2 all-hazard approach so that all risk management measures described at a high level of abstraction in NIS2 were covered by F4SLE statements. This prerequisite assumes that the use of F4SLE comprehensively demonstrates the security maturity of organizations, thereby ensuring that essential topics are not excluded from the security level evaluation. From the perspective of effectiveness (avoidance of errors and support for achieving the goal), it is essential for all user stories, with US1.2, US2.1, US4.2, US5.1, and US6.1 specifically mentioning the all-hazard approach or risk management measures. We evaluated the fulfillment of F4SLE effectiveness from this perspective using the mapping in Table 2. The result confirmed that the F4SLE statements comprehensively cover the risk management measures of NIS2 at a lower maturity level than the standard level. Since the standard level of F4SLE supports the achievement of resilience and higher maturity in the scope of risk management measures outlined in NIS2 Art 21(2), F4SLE also enables the evaluation of progress and change.

The user stories where we achieved the **goal** were US1.1, US1.2, US4.1, US4.2, and US5.1. Conceptually, F4SLE also enables us to meet the goals of US2.1. The real tests of US2.2 were not completed, but conceptually accepted by the Supervisory Authority leader conditionally, if the sample they base their decisions on is more representative than the current demo. The user stories regarding the relationship with the Service Provider & Supplier were not completed and require separate research.

To assess **efficiency**, we collected feedback on time consumption (see Table 4 and Figure 2) and free-text evaluation from Entities. The respondents estimated the time required to answer the F4SLE questionnaire to be approximately 1 hour (statistical mode). As a result, responding to F4SLE, Entities received an overview report on their organization and a benchmark against the security levels of sectoral averages. Organizations could use this report to set priorities in their implementation plans or justify the need for resources. We observed the preparation of implementation plans in the control group (Section 5.3), but we provided F4SLE results interpretation seminars to the EU group (Section 5.2), enabling organizations to consider the results in their security implementation plans independently. The feedback indicated that responding to F4SLE was perceived as a security training, where a structured evaluation of security measures provided a quick insight into the security standards within the respondent's organization.

Efficiency from the Member State's perspective depends on the development resources of the F4SLEs' analytics module, how it works in cooperation with the Member State's requirements (see Section 5.5), and the frequency of necessary improvements (additional resources). However, it must also be taken into account that the role of the Member State is to motivate Entities, set deadlines to comply with the F4SLE questionnaire, and direct focused resources based on the results

to increase the security level. In our demonstration process, both motivation and the development of the analytics module, as well as planning activities based on the results, succeeded.

The resource cost of the Supervisory Authority (see Section 5.6) is directly related to the reliability of the results in the context of a specific organization. Suppose the current results are based on self-assessment, where the risk of errors and fraud is real. In that case, supervision must implement additional control measures at least for the user story US2.2 when introducing F4SLE. User story US2.1 would already create efficiency through planning focused supervisory tasks.

For Security Consultants (see Section 5.4), a structured approach and references to standards were essential, allowing them to direct the organization to implement specific risk management measures immediately. This resulted in significant time savings for consultants and organizations. It is worth noting that the consultant can complete F4SLE either in collaboration with the client or independently. Security Consultants also emphasized the low entrance barrier and free access to F4SLE.

We evaluated the **satisfactory** aspects of usability based only on the results of empirical demonstrations. We collected satisfactory evaluation data from the Entity control group in both verbal and written forms. Based on verbal feedback, they recommended the use of F4SLE to their partners, friends, and subcontractors. The anonymous written recommendation index scores 4.85 out of 5 (Section 5.3). This score shows a very high level of satisfaction. However, it is worth noting that Entities did not voluntarily use F4SLE; they required an external motivator for this. Member State representatives were also satisfied with the initial use and expressed a desire to continue collecting and analyzing data with F4SLE on a yearly basis. Security Consultants wished to continue using F4SLE in their future work. Both the representatives of the Member State and the Security Consultant express their discretionary usage of F4SLE.

Our study demonstrates that F4SLE is suitable for completing the Member State (US1.1, US1.2), Security Consultant (US4.1, US4.2), and Entity (US5.1) user stories. Using F4SLE is also functional when planning the work of the Supervisory Authority (US 2.1). Since we did not have real data to complete the Service Provider & Supplier user stories (US61, US6.2), we cannot provide direct confirmation of usability here. We can only refer to the willingness of some organizations that have used F4SLE to share their results with their partners. The Supervisory user story US2.2 can provide valuable results if the correctness of the Entities data is checked separately. In the case of ENISA (US 3.1), it is necessary to understand the fundamental and detailed objectives as preliminary work to provide a final evaluation.

6 Discussion

In our study, the Entities emerged as the critical stakeholder group, acting as the primary providers of security evaluation data that serves as the essential input for other user groups' user stories. Since the entire ecosystem of F4SLE relies on their contribution, minimizing the administrative burden on them is paramount. Our pilot demonstrated that this can be achieved through a "collect once and reuse" approach, where data is shared across multiple stakeholder groups with periodic updates, effectively eliminating repetitive requests. The main challenge remains obtaining initial data. To overcome this (as also expected by NIS2 [2]), the Member State must actively motivate Entities by linking security evaluations to tangible benefits – such as grant eligibility, access to training, service priority, or a competitive advantage. Therefore, the assumption here is acceptance of a standardized security level evaluation by all user groups and local regulatory support.

A central theme, in addition to the all-hazard approach across all user groups, was the request for comparability. User groups are interested in measuring progress over time, comparing their performance with organizations in the same sector, across sectors, and against broader security trends. For the Supervisory Authority, an additional sector criticality scale is needed for the benchmarking option. NIS2 user stories (see the Appendix) do not mention benchmarking directly;

however, Brezavscek et al. [10] identify its absence as a significant shortcoming of security evaluation instruments. Recognizing this, F4SLE was designed with benchmarking capabilities, and our pilot results confirmed its high relevance for every participating user group.

The results presentation format also influenced engagement. For results, intuitive clarity, most participants preferred radar diagrams with a "traffic light color" scheme (red = high risk, yellow = moderate risk, green = manageable risk). Only the Supervisory Authority accepted more complex box plots, which require statistical literacy. We underutilized analytical options, which our collected metadata enabled (see Table 1), because our stakeholders were unable to accept them due to low maturity levels in both security and statistical literacy. However, the metadata collected by F4SLE holds significant potential for advanced analysis, particularly for Member States, ENISA, and Supervisory Authorities in the next iterations. As the multi-stakeholder security level evaluation concept is novel, users currently lack the experience to envision the full scope of analytical possibilities or articulate their specific needs.

The reliability of self-assessment data was a significant concern for the Supervisory Authority. They prefer supporting evidence from independent sources or their own intervention. Even within organizations, credibility issues arose from misinterpretations, unintentional or intentional errors, rapid situational changes, and gaps between stated policies and actual practices. This highlights the potential of automatic security evaluation controls, which are directly integrated into an Entity's IS and management systems. F4SLE can support automation because of its modular and standardized structure. Research on automated compliance checks against standardized risk management measures is still at a stage where we should keep a human in the loop [38], [39]. Automation supports efficiency, but we must consider that it also increases entry barriers, requiring high security maturity and additional resources at the outset, which may contradict the expectations of Security Consultants. Additionally, we must acknowledge the need to work with the vocabulary of F4SLE and avoid problematic terms or find alternatives to interpret them in simplified forms.

Another shortcoming of security evaluation instruments, as highlighted by Brezavscek et al. [10], is their lack of alignment with best practices. F4SLE addresses this by design, strictly mapping every statement to a relevant standard or regulation. While Security Consultants praised this feature for facilitating immediate recommendations, their focus on the questionnaire itself highlights a potential risk. There is a danger that F4SLE (or any other security evaluation instrument) could be misinterpreted as a standard in its own right, leading users to prioritize optimizing their evaluation score rather than following a genuine risk-based approach grounded in the underlying standards.

We observed a polarization in the Entities' feedback that mirrors Sweller's research in the field of educational psychology [40]. Intermediate users found F4SLE's structured, holistic overview effective in providing a quick overview of the security standard, as it integrated with their existing knowledge, leading to constructive feedback. In contrast, those critical of the terminology and security management issues lacked the necessary prior knowledge. This suggests F4SLE functions as a training tool that must evolve via regular updates, but also confirms that low-maturity, unmotivated users require human or alternative support beyond the instrument itself.

We can briefly list the challenges related to security level evaluation with F4SLE that we met:

- A multiple-stakeholder approach needs standardization acceptance from all user groups.
- There is a need for external motivation for data providers to participate in the security level evaluation.
- Collected data should be reused, and give an option for supplementation with external data for analysis.
- Low statistical literacy can hamper the practical use of the collected data. An iterative approach is suggested.
- Comparability options should consider benchmarking with other sectors, while also tracking progress over time.
- To find options for data reliability issues via automation or additional validation.

- Educational value requires regular updating of the instrument content.

While NIS2 formally requires a security evaluation, its ultimate goal is a common high security level. Research by Thaw [41] and Boggini [14] shows that disclosing non-compliance can be an effective regulatory tool. Therefore, making the Entities' security evaluation results accessible, both to partners and the Supervisory Authority upon request, and proactively through channels such as the organization's website or annual report, could transform the evaluation from a purely formal requirement into a driver of genuine security improvement. Therefore, it may be appropriate in the future to standardize the requirement for regular disclosure of security level evaluation results by Entities and Service Provider & Supplier.

Limitations Our study has several limitations regarding its scope, sampling methods, and data verification processes.

First, the user stories defined by NIS2 are phrased broadly, leaving significant room for interpretation. While local regulations could provide additional clarity by defining specific data or objectives. These national specifics were excluded from the scope of this study to maintain a general focus.

We relied on a convenience sample of volunteers who were already aware of security needs, but mostly were not at a high maturity level. Consequently, we cannot confirm statistical representativeness because we lack usability data for entities that discontinued participation or declined to share their results with F4SLE.

Not all stakeholders could test their user stories empirically; for some, we conducted only conceptual demonstrations or provided assessments based on gathered opinions. However, we can gain input from this study for further experiments.

To keep the F4SLE barrier to entry low for respondents, we did not implement strict authentication solutions to verify the respondent's right to represent their organization. Therefore, we cannot fully rule out the possibility that the organizational data is unverified. To mitigate fraud, access links were distributed only to relevant interest groups, and time limits were set for responses.

Verified data were confirmed for representatives of the control group, comprising 18 SMEs. For this group, metadata was reviewed individually during consultations. To reduce the risk that prior training might have biased verbal feedback, we allowed control group participants to provide anonymous feedback after the training, with a short time delay by LAC4 and CyberNET, not the F4SLE management team.

The collected security self-assessment results rely solely on the opinions of organizational representatives and were not independently audited. We acknowledge that self-assessment can lead to inaccuracies due to differing interpretations or issues with terminology. We excluded strict accuracy verification from the current scope. Future iterations will address this by automating F4SLE data collection with built-in validation checks (see future works in Conclusion), which will help standardize terminology and increase reliability.

7 Conclusion

Our two research questions focused on how the user groups implemented the NIS2 user stories (RQ1), the decisions and challenges we encountered, and the main improvements that need to be addressed in the future (RQ2). We addressed these questions through practical pilot projects, described (organized by stakeholders) in Sections 5.2–5.7 and summarized in Section 5.8. We presented the challenges that emerged in Section 6.

The study demonstrated and evaluated the use of F4SLE for interconnected NIS2-based user stories from multiple stakeholder groups. We conducted both empirical and conceptual demonstrations. The demonstrations did not cover the user stories of ENISA and the Service Provider & Supplier. The results confirm that NIS2 user stories are fulfillable with F4SLE for all user groups participating in the study. Significant improvement is needed in the reliability of

self-assessment (i.e., automation). To ensure efficiency and consistency, standardized involvement of user groups and strong leadership from Member States are essential, enabling data to be collected once and reused across different stakeholders. For this approach to succeed in practice, both regulatory backing and motivational incentives are necessary, especially for the primary data provider, the organization responsible for evaluating its own security level. These elements, when combined, will help create a sustainable, scalable, and collaborative security evaluation ecosystem.

To enhance the overall security level of Member States, including ensuring the security of supply chains, policymakers should consider requiring the publication of summary results at both organizational and sectoral levels.

The results here are generalizable to other centrally managed security level evaluation instruments recommended for supporting NIS2 compliance.

The planned *future works* based on these results are:

- We will continue to work with the regulator to ensure that the benefits for Entities are effective. It would be necessary to study the benefits that motivate Entities and those that are also relevant from a societal perspective.
- At the same time, the F4SLE analytics engine needs to be developed to automatically provide users with results that meet the goals of their user stories (as opposed to the current manual approach).
- We can then focus on the collected security evaluation results and the representative statistical analysis.
- In addition, we are exploring the possibilities of implementing automated checks and plan to analyze the integration of the security level evaluation instrument with other environments, allowing certain responses to occur automatically regardless of the Entity's activity (e.g., security scanner input). This also includes F4SLE content mapping with multiple standards and regulations.

Acknowledgement

This work is part of the Cyber-security Excellence Hub in Estonia and South Moravia (CHESS) project funded by the European Union under Grant Agreement No. 101087529. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] A. Lamba, S. Singh, B. Singh, N. Dutta, and S. S. R. Muni, "Analyzing and Fixing Cyber Security Threats for Supply Chain Management," *International Journal for Technological Research in Engineering*, vol. 4, no. 5, 2017. Available: https://doi.org/10.2139/ssrn.3492687

[2] European Parlament, "Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive)," 2022. Available: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32022L2555

[3] M. Seeba, M. Valgre, and R. Matulevičius, "Evaluating Organization Security: User Stories of European Union NIS2 Directive," in *Advanced Information Systems Engineering*. Springer, 2025, pp. 57–74. Available: https://doi.org/10.1007/978-3-031-94569-4_4

[4] N. T. Le and D. B. Hoang, "Can Maturity Models Support Cyber Security?" in *2016 IEEE 35th International Performance Computing and Communications Conference (IPCCC)*, 2016, pp. 1–7. Available: https://doi.org/10.1109/PCCC.2016.7820663

[5] A. M. Rea-Guaman, I. D. Sánchez-García, T. S. Feliu, and J. A. Calvo-Manzano, "Maturity models in cybersecurity: A systematic review," in *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, 2017, pp. 1–6. Available: https://doi.org/10.23919/cisti.2017.7975865

[6] University of Tartu, NCSC-EE, “Organisation’s information security maturity level evaluation,” 2025. Available: https://mass.cloud.ut.ee/test-massui/. Accessed on June 01, 2025.

[7] M. Seeba, S. Mäses, and R. Matulevičius, “Method for evaluating information security level in organisations,” in *Research Challenges in Information Science*. Springer, 2022, pp. 644–652. Available: https://doi.org/10.1007/978-3-031-05760-1_39

[8] M. Seeba, T. Oja, M. P. Murumaa, and V. Stupka, “Security Level Evaluation with F4SLE,” in *Proceedings of the 18th International Conference on Availability, Reliability and Security*. ARES ’23, Association for Computing Machinery, 2023. Available: https://doi.org/10.1145/3600160.3605045

[9] M. P. Murumaa, “Designing a Security Sensitive Self-assessment Framework,” 2023, Master’s Thesis, Faculty of Science and Technology, University of Tartu, Estonia. Available: https://thesis.cs.ut.ee/92895428-9fc4-4248-bc78-4a00b3e90101

[10] A. Brezavšček and A. Baggia, “Recent trends in information and cyber security maturity assessment: A systematic literature review,” *Systems*, vol. 13, no. 1, 2025. Available: https://doi.org/10.3390/systems13010052

[11] A. Rabii, S. Assoul, K. Ouazzani Touhami, and O. Roudies, “Information and cyber security maturity models: a systematic literature review,” *Information and Computer Security*, vol. 28, no. 4, pp. 627–644, 2020. Available: https://doi.org/10.1108/ICS-03-2019-0039

[12] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A design science research methodology for information systems research,” *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007. Available: https://doi.org/10.2753/MIS0742-1222240302

[13] “ISO/IEC 27001:2022(en) Information security, cybersecurity and privacy protection — Information security management systems — Requirements,” International Organization for Standardization, Standard, 2022.

[14] C. Boggini, “Reporting cybersecurity to stakeholders: A review of CSRD and the EU cyber legal framework,” *Computer Law & Security Review*, vol. 53, article 105987, 2024. Available: https://doi.org/10.1016/j.clsr.2024.105987

[15] M. Seeba, A. O. Affia, S. Mäses, and R. Matulevičius, “Create your own MUSE: A method for updating security level evaluation instruments,” *Computer Standards & Interfaces*, vol. 87, article 103776, 2024. Available: https://doi.org/10.1016/j.csi.2023.103776

[16] J. Pöppelbuß and M. Röglinger, “What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management,” in *ECIS 2011 Proceedings*, 2011. Available: https://aisel.aisnet.org/ecis2011/28

[17] L. Sweeney, “k-anonymity: A model for protecting privacy,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002. Available: https://doi.org/10.1142/S0218488502001648

[18] “ISO 9241-11:2018 Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts,” International Organization for Standardization, Geneva, Switzerland, Standard, 2018. Available: https://www.iso.org/standard/63500.html

[19] N. Bevan, J. Carter, J. Earthy, T. Geis, and S. Harker, “New ISO Standards for Usability, Usability Reports and Usability Measures,” in *Human-Computer Interaction. Theory, Design, Development and Practice, HCI 2016. Lecture Notes in Computer Science*, vol. 9731, Springer, 2016, pp. 268–278. Available: https://doi.org/10.1007/978-3-319-39510-4_25

[20] E. Rehnstam, W. Winquist, and S. Hacks, “NIS2 Directive in Sweden: A Report on the Readiness of Swedish Critical Infrastructure,” in *Secure IT Systems, NordSec 2024. Lecture Notes in Computer Science*, vol. 15396, Springer, 2025, pp. 176–195. Available: https://doi.org/10.1007/978-3-031-79007-2_10

[21] S. Örri, “NIS2 directive readiness in the Nordics,” Haaga-Helia University of Applied Sciences, Helsinki, Finland, 2025.

[22] NCSC-BE, “Cyberfundamentals conformity selfassessment tool,” 2024. Available: https://atwork.safeonweb.be/tools-resources/cyberfundamentals-framework. Accessed on Jan. 21, 2025.

[23] NCC-AT, “WKO Online Ratgeber,” 2025. Available: https://ratgeber.wko.at/itsafe/. Accessed on Jan. 21, 2025.

[24] NCSC-IE, “Cyber Security Baseline Standards Self-Assessment Form,” 2025. Available: https://www.ncsc.gov.ie/guidance/. Accessed on Jan. 21, 2025.

[25] NC3-LU, “Fit 4 Cybersecurity Assessment Tool,” 2025. Available: https://nc3.lu/assessment-testing-and-training/fit4cybersecurity. Accessed on Jan. 21, 2025.

[26] CNCS Portugal, “CiberCheckUp,” 2025. Available: https://www.cncs.gov.pt/pt/quadro-nacional/#cibercheckup. Accessed on Jan. 21, 2025.

[27] Finnish Transport and Communication Agency National Cyber Security Centre, “Cybermeter,” 2024. Available: https://www.kyberturvallisuuskeskus.fi/fi/palvelumme/tilannekuva-ja-verkostojohtaminen/kybermittari. Accessed on May 13, 2024.

[28] Hellenic Ministry of Digital Governance Government department, “Cybersecurity Self Assessment Tool,” 2021. Available: https://mindigital.gr/wp-content/uploads/2022/03/cybersecurity-self-assessment.xlsm. Accessed on Jan. 21, 2025.

[29] G. Drivas, A. Chatzopoulou, L. Maglaras, C. Lambrinoudakis, A. Cook, and H. Janicke, “A NIS Directive Compliant Cybersecurity Maturity Assessment Framework,” in *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2020, p. 1641–1646. Available: https://doi.org/10.1109/COMPSAC48688.2020.00-20

[30] J. Ralyté, G. Koutsopoulos, and J. Stirna, “Verification, validation, and evaluation of modeling methods: experiences and recommendations,” *Software and Systems Modeling*, 2025. Available: https://doi.org/10.1007/s10270-025-01304-2

[31] M. Seeba, “Framework for Security Level Evaluation (F4SLE) E-ITS based ver 2024,” 2025. Available: https://doi.org/10.23673/re-562

[32] RIA (Estonian Information System Authority), “E-ITS. Portal of Estonian Information Security Standard,” 2022. Available: https://eits.ria.ee/

[33] “ISO/IEC 27002:2022 Information security, cybersecurity and privacy protection — Information security controls,” International Organization for Standardization, Standard, 2022.

[34] The MITRE Corporation, “MITRE ATT&CK,” 2025. Available: https://attack.mitre.org/. Accessed on Jan. 12, 2025.

[35] ENISA, “ENISA NIS360 2024. ENISA Cybersecurity Maturity & Criticality Assessment of NIS2 sectors,” 2025.

[36] ENISA, “2024 Report on the State of Cybersecurity in the Union,” 2024-12.

[37] ENISA, “EU Cybersecurity Index. Framework and methodological note,” 2024. Available: https://www.enisa.europa.eu/sites/default/files/2024-12/eu_csi_methodological_note_v1-0.pdf

[38] A. Shaked and N. Messe, “BridgeSec: Facilitating effective communication between security engineering and systems engineering,” *Journal of Information Security and Applications*, vol. 89, article 103954, 2025. Available: https://www.sciencedirect.com/science/article/pii/S2214212624002564

[39] F. Angermeir, J. Fischbach, F. Moyón, and D. Mendez, “Towards automated continuous security compliance,” in *Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. Association for Computing Machinery, 2024, p. 440–446. Available: https://doi.org/10.1145/3674805.3690748

[40] J. Sweller, “Element interactivity and intrinsic, extraneous, and germane cognitive load,” *Educational Psychology Review*, vol. 22, pp. 123–138, 2010. Available: https://doi.org/10.1007/s10648-010-9128-5

[41] D. Thaw, “The Efficacy of Cybersecurity Regulation,” *Georgia State University Law Review*, vol. 30, p. 287, 2013–2014.

Appendix

User Stories (US) of NIS2 Related to Security Level Evaluation of Entities (extract from M. Seeba, M. Valgre, and R. Matulevičius, "Evaluating Organization Security: User Stories of European Union NIS2 Directive," in Advanced Information Systems Engineering, J. Krogstie, S. Rinderle-Ma, G. Kappel, and H. A. Proper, Eds. Cham: Springer Nature Switzerland, 2025, pp. 57–74. Available: https://doi.org/10.1007/978-3-031-94569-4_4)

US1.1 As a **Member State**, I can oversee the security posture of Entities through structured security level evaluation results, so that I achieve awareness of compliance with regulations.

US1.2 As a **Member State**, I can evaluate an Entity's cybersecurity level using an all-hazards approach, so that I can allocate resources to address directly on identified vulnerabilities.

US2.1 As a **Supervisory Authority**, I can prioritize supervisory tasks by using all-hazard covering security evaluation results so that I can focus supervisory tasks on high-risk entities or areas.

US2.2 As a **Supervisory Authority**, I can ensure (with a security evaluation instrument) that Entities that did not comply with regulatory requirements implement corrective risk-management measures within reasonable deadlines so that supervisory resources are used effectively and unnecessarily hamper the business activities of the Entity is avoided.

US3.1 As **ENISA**, I can collaborate with Member States to assess collected evaluation data on cybersecurity capabilities and awareness, so that I can share cybersecurity best practices and gaps across the European Union.

US4.1 As a **Security Consultant**, I can get an overview of the Entity's security maturity evaluation results so that the most vulnerable areas can be prioritized in a timely manner for an improvement plan.

US4.2 As a **Security Consultant**, I can re-evaluate the Entity's risk-management measures implementation so that tracking characterizes risk-management measures implementation status progress.

US5.1 As an **Entity**, I can ensure the Entity adopts an all-hazards approach to cybersecurity so that the evaluation results show strengths and direct to plan improvements to our security shortcomings.

US6.1 As a **Service Provider & Supplier**, I can provide the risk-management measures in all-hazard evaluation approach results to the partner Entity so that the Entity can choose us as the most suitable secure suppliers.

US6.2 As a **Service Provider & Supplier**, I can regularly evaluate my cybersecurity practices so that I can present my evaluation results to my partner Entity to demonstrate our security.