REWIRE
CYBERSECURITY
SKILLS ALLIANCE

Co-funded by the
Erasmus+ Programme
of the European Union

REWIRE - Cybersecurity Skills Alliance
A New Vision for Europe

R.2.2.3. Methodology to anticipate future needs

Title	R.2.2.3. Methodology to anticipate future needs
Document description	Document presents methodology to forecast future needs for cybersecurity skills.
Nature	Public
Task	T2.2 Cyber Security Skills Needs Analysis
Status	Final
WP	WP2
Lead Partner	MRU
Partners Involved	All
Date	04 May 2022

Revision history	Author	Delivery date	Summary of changes and comments
Version 0.1	Edmundas Piesarskas (EKT), Paulius Pakutinskas (MRU), Donatas Alksnys (MRU)	27/06/2021	Initial framework of the report
Version 0.2	Edmundas Piesarskas (EKT), Donatas Alksnys (MRU), György Dán (KTH), Jan Jerabek (BUT), Sarra Ricci (BUT), Paulius Pakutinskas (MRU), Viktor Varga (UNICOM)	21/07/2021	Version with inputs from partners
Version 0.3	Edmundas Piesarskas (EKT), Paulius Pakutinskas (MRU), Donatas Alksnys (MRU	16/09/2021	First version for peer review
Version 0.4	Pedro Adao (ULisboa), Vaclav Stupka (MU)	24/09/2021	Comments of internal reviewers
Final Version 1	Donatas Alksnys (MRU)	09/12/2021	Version with updates after comments of internal reviewers
Final Version 1.1.	Regina Valutytė (MRU)	04/05/2022	Final version with updates after EACEA review and comments

Disclaimer:

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

CONTENTS

Executive Summary

The versatile and relevant methodology is a key factor to anticipate future needs in the cybersecurity field. Facing the tremendous speed of new technology development and worrying increase in the number of cyberattacks, it is essential to include and evaluate all possible factors that could contribute to the creation of a long-lasting methodology. One of the main challenges is missing a unified cybersecurity skills framework. Having a shared vision is essential to develop long-lasting model and methodology to forecast future cybersecurity skills needs.

The methodology structure consists of three main parts: (i) methodology applied in relevant documents and research as basis and justification for the following parts, (ii) stakeholder survey, and (iii) automated job ads analysis. Each of these parts are closely related. A taxonomy was derived mainly from NICE competencies and then adjusted to the EU market. It was used both for stakeholder survey and automated job ads analysis. Keeping a shared framework helps to maintain sustainable and consistent methodology where each element supplements the other.

The findings of the pilot projects' (i.e., Concordia, CyberSec4Europe, Echo, and SPARTA) was a basis for the creation of a new methodology. The very first step was to check for existing methodologies related to identification of the skills needs. This focus ensured a continuous process of developing a relevant methodology for anticipating future needs in the cybersecurity field. Valuable insights and identified gaps give a solid ground for further development of the methodology and minimize the need of analysing the same areas.

It is not possible to anticipate future skills needs without a classification of the skills and a common taxonomy. As the ENISA skills framework was not yet available at the moment of writing this report (July 2021), the project group chose the NICE NIST competencies framework and identified the main competencies needed for cybersecurity work roles. The identified competencies were split into three categories: (1) Cybersecurity Skills, (2) IT Skills, and (3) Soft Skills.

With awareness of extensive market growth and sophistication of IT, it is necessary not only to take existing studies but also to introduce and promote flexible and long-term methodological means for coming years. Therefore, the proposed design of future needs anticipation methodology is based not only on previous pilots but also on conducted stakeholders survey, results of dictionary-based job advertisement analysis and automated job adds analysis. These methods allowed the project group to identify the current and near future skills needs and propose a strategy that is not dependent on the period that is applied and, on the taxonomy used.

Combining previous work with a newly created stakeholder survey and automated job ads analysis based on machine learning, methodology for anticipating future needs will also be adjusted to reflect actual market needs. Cooperation with ENISA is also one of the key elements in the creation of a unified EU cybersecurity skills framework and methodology that is based on it.

Introduction

Cybersecurity is becoming increasingly important topic with the rapid development of Information Technology (IT), where roles with particular skills are in high demand in the market. However, due to the lack of shared understanding and taxonomy of cybersecurity roles and skills, it is complicated and sometimes even impossible to coordinate educational programs, share information about market needs, communicate about educational programs and which skills and for which roles could be obtained there. Despite many researches and various frameworks aiming to address this issue, a common vision of the unified framework is still not achieved. This variety of taxonomies in different frameworks and the missing interconnection among them cause difficulties in the creation of a common vision. Though environment is very complex, the project group focused on most relevant documents in this field to assess methodologies used in other projects, come up with methodologies that could contribute in anticipation of future needs and propose new methodology for REWIRE project.

This report was prepared following a three-step process:

(i) an overview of existing skills needs analysis methods. Section 1 presents the methodology applied in relevant documents and research, in particular the findings based on previous pilot projects.
(ii) identification of skills. Section 2 describes the skills that were used in the survey and machine learning model;
(iii) the methodology proposal. Section 3 presents the proposed Methodology to anticipate future needs, Section 4 - the results of work done in this project.

The R2.2.3 Methodology to Anticipate Future Needs is an input to R2.3.2 Cybersecurity Strategy. This methodology of anticipating future needs will be used in Blueprint design (T3.2), Development of European Cybersecurity Skills Framework (T3.3) and Design and Development of the Digital European Cybersecurity Skills Observatory (T5.1) as fundamentals for the mentioned tasks. The relationship among WPs and tasks is illustrated in Figure 1.

The proposed methodology will be used in further development of Cybersecurity Strategy that will lay down fundamentals for proper addressing of market demand and actual respond to existing needs.

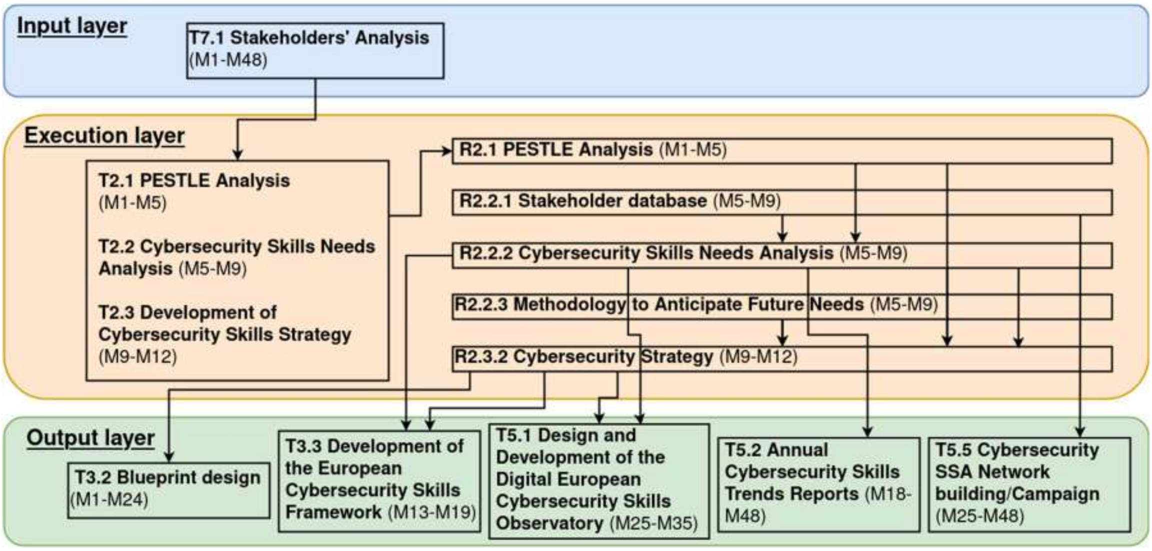

Input layer
T7.1 Stakeholders' Analysis (M1-M48)
Execution layer
T2.1 PESTLE Analysis (M1-M5)
T2.2 Cybersecurity Skills Needs Analysis (M5-M9)
T2.3 Development of Cybersecurity Skills Strategy (M9-M12)
R2.1 PESTLE Analysis (M1-M5)
R2.2.1 Stakeholder database (M5-M9)
R2.2.2 Cybersecurity Skills Needs Analysis (M5-M9)
R2.2.3 Methodology to Anticipate Future Needs (M5-M9)
R2.3.2 Cybersecurity Strategy (M9-M12)
Output layer
T3.2 Blueprint design (M1-M24)
T3.3 Development of the European Cybersecurity Skills Framework (M13-M19)
T5.1 Design and Development of the Digital European Cybersecurity Skills Observatory (M25-M35)
T5.2 Annual Cybersecurity Skills Trends Reports (M18-M48)
T5.5 Cybersecurity SSA Network building/Campaign (M25-M48)

Figure 1. Relationship with other WPs and Tasks in REWIRE project

1. Review of relevant documents

This section reviews the pilot projects' (i.e., Concordia, CyberSec4Europe, Echo, and SPARTA) findings and their deliverables as a basis for the creation of a new methodology. This step ensures a continuous process of developing a relevant methodology for anticipating future needs in the cybersecurity field. Valuable insights and identified gaps give a solid ground for further development of the methodology and minimize the need of analysing the same areas.

1.1. Review of the Pilot projects' findings

1.1.1. SPARTA

After analysing the state-of-the-knowledge on skills management, the SPARTA project delineated the way forward to a European Skills Framework[1]. In particular, the SPARTA Framework is based on the Joint Research Centre (JRC) Cybersecurity domains taxonomy and the US-based National Initiative for Cybersecurity Education (NICE) since they provide a comprehensive structure that can incorporate the EU-specific realities and emerging skills landscape. The framework is then connected to education for the development of Good-practice Cybersecurity Curricula[2].

In particular, NICE Competencies were mapped to SPARTA topics which represent the content usually taught in cybersecurity university study programs. This list was created considering the existing curricula guidelines and the deliverable D9.1.

Moreover, SPARTA Topics are classified in either Fundamental or Cyber Security categories. Fundamental subjects are those that are not directly linked to the Framework, but which serve as a prerequisite for later studies. SPARTA topics can be viewed as skills needs analysis from the education point of view. Their relevance was analysed by comparing a selection of existing cybersecurity study programs

1.1.2. ECHO

The ECHO project aims to address the needs and skills gaps of cybersecurity professionals by aligning with know-how from existing frameworks (e.g., the e-Competence Framework, ECSO classification, and European Qualification Framework). They developed the ECHO Cyberskills Framework[3] to target the identified gaps in cybersecurity talent, contextualizing the approach to fit organizations from various sectors of industry, and leveraging the currently fragmented efforts in the field of cybersecurity skills, knowledge, and competencies. The end goal is to provide a flexible mechanism to design and develop of learning outcome-based training program.

[1] Edmundas Piesarskas, D9.1 Cybersecurity skills framework, January 2020, https://www.sparta.eu/deliverables/

[2] Jan Hajny, D9.2 Curricula descriptions, July 2020, https://www.sparta.eu/deliverables/

[3] Pavel Varbanov, D2.6 ECHO CYBERSKILLS FRAMEWORK 2021 https://echonetwork.eu/wp-content/uploads/2021/03/ECHO_D2.6_Cyberskills-Framework.pdf

Within this process, they identify the necessity of a stakeholders' skills analysis. In particular, the ECHO project proposes that the identification of competence needs could start from the comparative analysis of multi-sector and inter-sector challenges, but also from the analysis of the entire supply chain of any given organization and its interrelation with its sector of operation, as well as with other sectors in terms of suppliers and customers.

They also combine a top-down (mapping the tasks, skills, knowledge, and competencies to professional profiles and knowledge domains) with a bottom-up approach (mapping the competence descriptors to real sector scenarios, assets, threats, and responsible professionals). In this way, the user can compare all existing sources of information and, based on an expert view, decide the most critical knowledge and skills that they should develop or deliver to the market.

ECHO approach supports clear provision of career and educational pathways. They connect the learning outcomes with skills supported by specific practical training scenarios. In this way, the ECHO system of solutions and services contributes to the faster building of practical skills and knowledge transfer to protect systems and digital assets more effectively.

Another need identified by the ECHO Consortium is that cybersecurity-related training, education, and certifications must be comparable. The significance and complementarity of fundamental and practical education should be balanced and transferable.

They also identified the need to create a more flexible method for short-term training in operational environments to develop purpose-based acquisition of top-priority skills. Therefore, ECHO creates conditions and infrastructure to support the timely identification of gaps and methods to fill those gaps continuously with hands-on tools that facilitate collaboration and knowledge sharing.

The ECHO assessment method provides hints for detailed analysis of missing knowledge and skills, and the pathways for filling the gaps. The method does not result only in a positive or negative assessment of the demonstrated abilities but identifies weaknesses and suggests learning objectives that address them.

1.1.3. CONCORDIA

The CONCORDIA project recognizes several issues affecting cybersecurity education: (a) lack of cybersecurity educators, (b) poor interaction with industry, (c) poor understanding of the labour market, (d) outdated platforms in the educational environment, and (e) difficulty in keeping the pace with the outside world.

In order to overcome the aforementioned issues and complement the existing ENISA Good Practice Guide on Training Methodology[4], CONCORDIA proposed the following methodology to create the optimal course[5] which is mapped to the following individual process stages: ENGAGE, DEFINE, PRODUCE, VALIDATE and DELIVER.

[4] https://www.enisa.europa.eu/news/enisa-news/good-practice-guide-on-training-methodologies-published-by-enisa

[5] Felicia Cutas, CONCORDIA: Methodology for the creation and deployment of new courses and/or teaching materials for cybersecurity professionals https://www.concordia-h2020.eu/wp-content/uploads/2020/06/CONCORDIA-methodology-courses-professionals-for-publication.pdf

In fact, the ENGAGE and DEFINE stages consist of initial inputs on the learning objectives and associated competencies to be developed during the course. A multi-stakeholder workshop with representatives of different companies from the same industry sector is used to identify cybersecurity topics. Following, the PRODUCE stage develops a specific curriculum and related content materials, the VALIDATE stage contains a pilot version of the course which is offered to professionals in the DELIVER stage.

Moreover, the methodology looks into addressing the following topics (1) understand your target audience, (2) look into their needs, (3) the content of the course, (4) choosing the lecturers, (5) lesson design, (6) deliver strategy, (7) consider the evaluation strategy, (7) the importance of certification, and (8) looking into partnering.

Of our interest is the "look into their needs" process which deals with skills needs analysis and it is associated to ENGAGE, DEFINE and VALIDATE/DELIVER course stages.

They propose several strategies on how to run the skills needs analysis:

- Do market research by checking for the existing courses addressing the different needs of the target audience.
- Check for trends either by market analysis or by considering new research areas mentioned in EU calls which are a good indicator of the needs for competencies.
- Cluster the needs per industry. In fact, the needs could differ from one industry to another.
- Check for needs per profile. Note that job profiles are characterized by a set of competencies.
- Seniority matters. The more professionals advance in their career, the more they need to include in their formation soft skills and business skills.
- Consider feedback loops. The needs may change overtime.

1.1.4. CyberSec4Europe

CyberSec4Europe and its Work package 6 is dedicated to the identification and prioritization of the cyber skills needed for education at University level, and the investigation of existing cybersecurity curricula.

Work package 6 – Cybersecurity Skills and Capability Building[6] of the CyberSEc4Europe project deals with:

- identification and prioritization of the cyber skills needed for security professionals and professionals in general;
- specification of training programs and professional assessment for different target groups in comparison to already existing industry programs (ISACA, CISSP, ISC2, etc.);
- design of a methodology to develop such programs;
- implementation of the capabilities (i.e. training and skill assessment) required to run such programs.

[6] https://cybersec4europe.eu/wp-content/uploads/2020/02/D6.2-Education-and-Training-Review-V1.2-Submtted.pdf

The task responsibilities are split among partners.

- Leading the task and being responsible for the initial review of existing programs at the European Level and the final development of assessment mechanisms for the general cybersecurity capabilities of the workforce across all Demonstration cases.
- Designing and providing content and assessment approaches for the specific areas of Security Intelligence, Adaptive Security and Cross-Border Authentication.
- Providing workforce assessment methods through serious games for PSD Demonstration cases.
- Providing assessments mechanism for non-ICT workforce (lawyers etc.).

Providing workforce assessment for Federated IdM scenarios on Public Sector.

1.2. Other documents

1.2.1. Eurostat

Eurostat's overview of Digital economy and society[7] focuses on mobile internet access, social media use, e-commerce, internet security, cloud services, digital skills and employment of ICT specialists. The statistics on Information and Communication Technologies (ICT) in this section are available separately for households/people and businesses/enterprises. Though cybersecurity skills and cybersecurity roles are not analysed in these reports, overall growth of ICT specialists and high level of ICT specialists' employment correlates with increasing demand for cybersecurity specialists. As increasing number of businesses rely on ICT for their daily operations and threat of cyber-attacks become more and more expected phenomenon, the demand for cyber security specialists could potentially grow at even higher rates than those of demand for ICT specialists.

Therefore, measure of employed ICT specialists of total employment could be valuable metric to be taken into account for anticipating future needs. Having in mind the significant importance of cybersecurity skills, proposal to Eurostat to include cybersecurity related topics in their questionnaires should also be considered as an option for improving future forecasting of cybersecurity skills demand.

The review of methodologies used in the reviewed documents has shown that the key elements are:

- identification of competencies that could be used as common taxonomy between project partners, and
- the mechanism to monitor need of identified competencies in the future.

2. Identification of skills

During the methodology's creation, the projects group realized the need for a common taxonomy. It is not possible to anticipate future skills needs without a classification of the skills. As the ENISA skills framework was not available at the moment of writing this report

[7] https://ec.europa.eu/eurostat/web/digital-economy-and-society/overview

(July 2021), we considered the NICE NIST competencies framework and we identified the main competencies needed for cybersecurity work roles.

2.1. The role of European Cybersecurity Skills Framework

The role of an agreed Framework for managing cybersecurity skills at the EU-level was discussed in different circumstances. The SPARTA pilot project put significant efforts to describe the importance, role and possible application of cybersecurity skills framework in the deliverable "Cybersecurity skills framework"[8]. The document also aimed to accelerate discussions around this topic. It also recognized some important difficulties in building EU-wide framework, like national legislation of MS's, importance of maintenance and others.

The European Union Agency for Cybersecurity (ENISA) took the initiative on this subject. In 2020 ENISA recognized the importance of skills framework, stating: "The development of an European Cybersecurity Skills Framework that would take into account the needs of the EU and each one of its Member States is considered an essential step towards the Europe's digital future."[9]

In July 2020, ENISA launched a call for an Ad Hoc Expert Group on Cybersecurity Skills Framework with the aim to promote harmonization in the ecosystem of cybersecurity education, training, and workforce development and to develop a common European language in the cybersecurity skills context. The group consists of 15 members, representing different stakeholder groups, including academia, industry, policy making.

The task of the Group is to develop a European Cybersecurity Skills Framework, which permits a common understanding of the roles, competencies, skills and knowledge used by individuals, employers and training providers across the EU Member States. Furthermore, it could also raise awareness by identifying the gaps in the cybersecurity landscape that can be bridged with the creation of a common European Cybersecurity Skills Framework.[10]

During the development of the Framework, there were a few principles to be applied:

- The Framework should fit to European landscape of standardization and legislation. The European Norm (EN) 16234-1 European e-Competence Framework (e-CF)[11] was selected as a reference point. Upcoming Cybersecurity Skills Framework will follow the construction approach of the above-mentioned Norm.
- The Framework should be simple and made for use of SMEs or other non-professionals in the field. This will be reflected in the limited number of profiles.
- The Framework should include only cybersecurity-specific competencies and skills. General capabilities will not be included in the Framework.

Currently the Framework is under development. A draft version is not available for public yet. It is expected to finish activities of the Ad Hoc Expert Group till the end of 2021 and make the Framework available in 2022.

[8] SPARTA - D9.1 - Cybersecurity skills framework
[9] European Cybersecurity Skills Framework — ENISA (europa.eu)
[10] Ad-Hoc Working Group on the European Cybersecurity Skills Framework — ENISA (europa.eu)
[11] European e-Competence Framework (ecompetences.eu)

In the architecture of the REWIRE project, a Europe-wide cybersecurity skills framework is one of the key components. Even if the ENISA Framework will be more of the recommendation nature, presumably it will still be the most-recognized and used across EU. In upcoming period REWIRE will use ENISA European Cybersecurity Skills Framework as it will become publicly available. It will also apply to the methodology of job market analysis.

2.2. The skills identified for further analysis

NIST Special Publication 800-181 revision 1, the Workforce Framework for Cybersecurity (NICE Framework) lists 57 competencies divided into technical, organizational, professional, and leadership categories.

We realized that some of the NICE competencies either need to be adjusted to the European (EU) market, were not relevant for the analysis's purposes, or could be merged. From the NICE competencies, we selected a total of 30 REWIRE competencies as shown in Figure 2. Competencies were carefully selected during WP2 task leaders' meetings and finally agreed on 2021-05-26 meeting. Competencies not included in our selection are visualized as grey. These competencies were split into three families: Cybersecurity Skills, IT Skills, and Soft Skills (Table 1).

Cybersecurity skills	**Other IT skills**	**Soft skills**
Business Continuity	Other IT skills	Communication
Data Analysis	Asset and Inventory Management	Education and Training Delivery
Data Privacy	Database Administration	Organizational Awareness
Data Security	Enterprise Architecture	Policy Development
Digital Forensics	Network Management	Project Management
Identity Management	Operating Systems	Strategic Relationship Management
Incident Management	Software Development	Workforce Management
Information Systems and Network Security	System Administration	
Information Technology Assessment		
Intelligence Analysis		
Law, Policy, and Ethics		
Physical Device Security		
Risk Management		
Testing and Evaluation		
Threat Analysis		

Table 1. REWIRE competencies.

Cybersecurity Skills competencies are those knowledge, skills, and abilities (KSAs) that need to be known in a cybersecurity work role while IT Skills are more general and fundamental information technology knowledge. At last, Soft Skills are related to non-technological KSAs.

IT and Soft Skills need to be considered since they are necessary for a good practice of cybersecurity work roles. We would like to remark that cybersecurity is a multidisciplinary field and therefore, requires knowledge not only from computer science.

Figure 2 depicts the three REWIRE skills families and "Other Skills" group which are those NICE competencies not essential for the analyses.

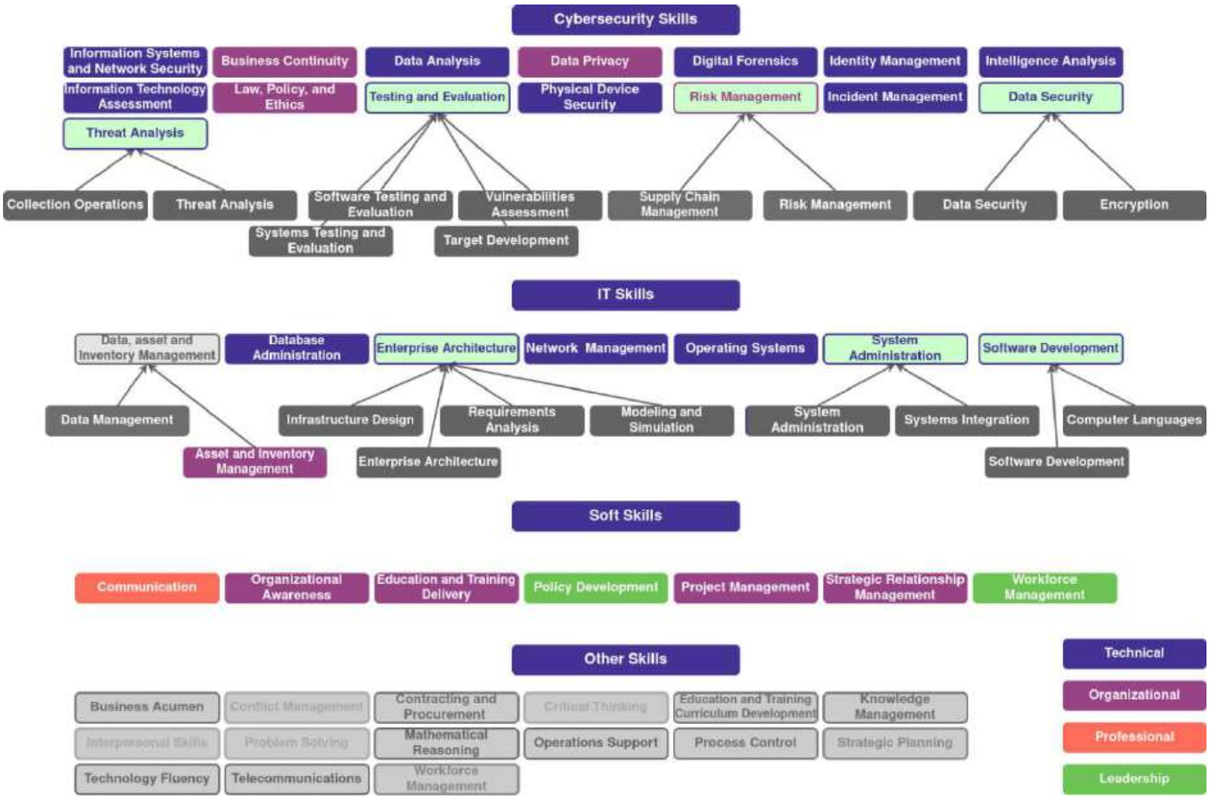


Figure 2. Mapping from NICE to REWIRE competencies

The skills rectangle colour depends on the NICE group of belonging, i.e., blue for Technical, purple for Organizational, orange for Professional, and green for Leadership. Merged competencies have a light-green rectangle with a border of another colour. Moreover, they are connected by arrows to NICE competencies that were merged.

3. REWIRE Methodology to anticipate future needs

With awareness of extensive market growth and sophistication of IT, it is necessary not only to take existing studies into account but also to introduce and promote flexible and long-term methodological means for upcoming years. The proposed design of future needs anticipation methodology is based not only on previous pilots but also on conducted stakeholders survey, results of dictionary-based job advertisement analysis and automated job adds analysis.

3.1. Stakeholder survey

We decided to reach out to the stakeholders identified by the REWIRE partners and entered into the REWIRE Stakeholder Database maintained as part of Work Package 7 activities. We wanted to ask them about their perception of the most sought-after job roles; cybersecurity competencies needed in their organizations and countries; as well as the level of availability and quality of relevant education programs in the cybersecurity domain.

We approached this goal in the following general steps:

1. We extended the REWIRE stakeholder database with additional cybersecurity professionals whom we intended to contact with our survey.
2. We performed a high-level, mostly manual analysis of job ads on relevant platforms, e.g., LinkedIn and national web-based job platforms. This was necessary to identify those job roles, which are mostly sought-after on the job platforms.
3. We analysed the cybersecurity competencies identified by the NIST NICE framework (March 2021 version), consolidated them (as described above) and used the distilled list of competencies throughout the survey.
4. We developed a novel survey covering the above-listed topics.
5. We tested and updated the survey in multiple rounds.
6. We sent out the survey to selected REWIRE stakeholders.
7. We performed an initial analysis of the survey results.

The survey itself consisted of the following sections:

- General information of the respondent;
- Country-level competency level status;
- Organization-level need for cybersecurity professionals;
- Country-level need for cybersecurity professionals;
- Country-level competency needs;
- Country-level coverage of competency needs;
- Different factors having an impact on cybersecurity education on the European level.

Proposed survey template covers both industry and educational perspectives thus contributing to anticipation of future needs.

3.2. Dictionary-based job advertisement analysis

One of the easiest-to-implement text mining approaches is dictionary-based analysis. These methods rely on the creation of novel or the use of existing dictionaries of words or phrases. The dictionary is essentially a mapping of different topics of interest or concepts to a list of words and phrases. A corpus of text can then be analysed and checked whether it, or its parts contain any of the words and phrases in the dictionary and thereby identify whether the texts analysed refer to any of the concepts of interest.

In our case the corpus of texts are sets of job advertisements collected from various, mostly online, digital platforms. The dictionary was created manually by starting from the consolidated list of NICE competencies and looking for related words and phrases in the initial corpus of job advertisements collected by the REWIRE project team members. This exercise was then documented in a spreadsheet, in which the leftmost column contained the NICE competencies, while the columns to the right contained one or more words or phrases which were commonly associated with the competency in the job advertisements analysed. Two such example mappings from this dictionary are shown in Table 2.

Information Systems and Network Security	VPN	TLS	firewalls	IDS	Cloud security	Web security
Information Technology Assessment	Iperf	Benchmarks	Performance simulation	Authorization tests	Stress tests	

Table 2. Mapping example.

The corpus of job ads was created by manually scraping data from online job advertisement platforms, mainly LinkedIn, and saving the job ad content to text files. These text files were unstructured.

The job ads were analysed in the following steps:

- Open a folder containing the text files with the job ads.
- Iterate through the job ad files.
- Check if the currently open job ad contained any of the words or phrases in the dictionary.
- If a match was found, then map the currently open job ad with the competency or competencies associated with the keywords/phrases.
- Report the top 10 competencies found in the corpus.

3.3. Automated job adds analysis

3.3.1. Natural language processing (NLP)

Machine learning (ML) has been used for various tasks such as image analysis, and time-series forecasting[12,13]. Depending on the tasks, different ML models are used to solve the problems. For example, recurrent neural network (RNN) is used for time series analysis, since RNN sequentially takes input data and compresses the input data into a context vector[14]. After compressing the entire sequence, we expect the context vector has compressed information from the start to the end of the sequence. Thus, the compressed information can be used to perform tasks such as time-series forecasting.

[12] Berg, Stuart, et al. "Ilastik: interactive machine learning for (bio) image analysis." Nature Methods 16.12 (2019): 1226-1232.

[13] Bontempi, Gianluca, Souhaib Ben Taieb, and Yann-Aël Le Borgne. "Machine learning strategies for time series forecasting." European business intelligence summer school. Springer, Berlin, Heidelberg, 2012

[14] Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors." nature 323.6088 (1986): 533-536.

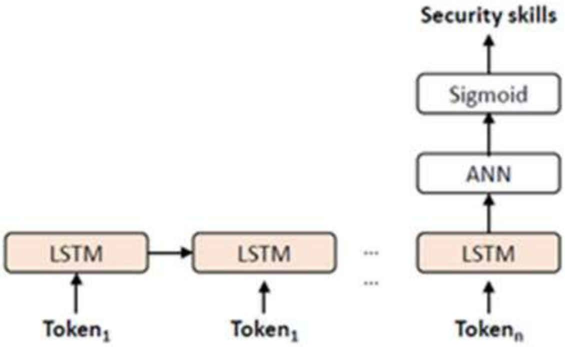
Natural language processing (NLP) is another task where RNNs can be applied[15]. In order to apply RNN to NLP, a sentence is interpreted as a time series by dividing a sentence into words and mapping the words into integers. The series of integers can be input to RNN models, and this approach tends to outperform other algorithms on NLP. Specifically, this approach has been used for sentiment analysis, a task classifying the negative or positive sentiments of a sentence by using ML algorithms and showed great performance.

The classification of required skills for job advertisements is made following analogous to sentiment analysis methodology. In this project, we aim to analyse advertisements with the RNN model and to automate the analysis of job advertisements.

3.3.2. Machine learning (ML) model

Our model uses a prebuilt word-piece tokenizer for Bidirectional Transformer (BERT)[16]. With the tokens from the word-piece tokenizer, two models are used as follows:

- **Model 1**: This model uses the last context vector of the LSTM model[17]. The embedding dimension of a token is 100, the hidden unit of LSTM is 128, and the hidden layer of ANN is 256[18].



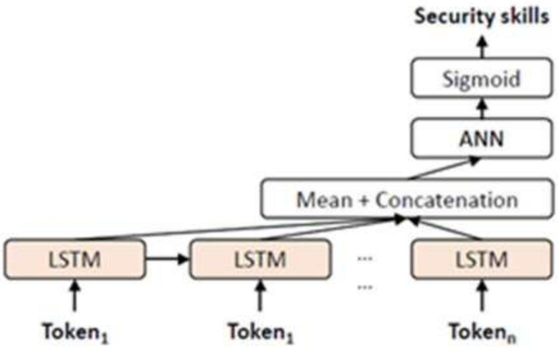
- **Model 2**: This model concatenates the mean of all intermediate context vectors and the last context vector. The concatenated vector is fed to an ANN to classify security skills. The hidden unit of LSTM is 128, and the hidden layer of ANN is 256.

[15] Yin, Wenpeng, et al. "Comparative study of CNN and RNN for natural language processing." arXiv preprint arXiv:1702.01923 (2017).

[16] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[17] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

[18] Wang, Sun-Chong. "Artificial neural network." Interdisciplinary computing in java programming. Springer, Boston, MA, 2003. 81-100



3.3.3. Training and Testing methodologies

In order to train those models, both models used the following settings:

- **Loss**: Cross entropy. This loss is the best fit for the categorical task as LSTM model gets the ads and outputs categories of the ads.
- **Dataset:** 1) Small dataset (31 ads), and 2) Medium dataset (87 ads). This set was collected and according to experts was sufficient to evaluate relevant models.
- **Learning rates**: 0.001. Suggested to be used for natural language processing task by project experts.

  Batch size: 10. Usually, the batch size is chosen to be 32, 64 or 128, but it was not suitable for our case since our dataset is small. Therefore, batch size was reduced to a smaller number 10.
- **Epochs**: 50. For given the batch size, i.e., 10 and 1 epoch, the training would finish within 9 steps (31 /10 < 9 and 87/10 <9). As this was not enough, epoch was iteratively increased by 10 (10 epochs à 20 epochs à 30 epochs and so on) and model test was carried on. At 50 epochs, performance did not increase significantly and this setting was set.

Note that the batch size is small due to the limited number of datasets (i.e., 31 and 87 ads). After each epoch, we evaluated the accuracy of our trained model as follows:

- Accuracy $= \dfrac{The\ number\ of\ correct\ predictions}{The\ number\ of\ correct\ predictions+Th\ \ number\ of\ wrong\ predictions}$

3.3.4. Experimental results

By using the aforementioned models and methodologies, two models were trained and tested on the two datasets. For each setting, we experimented 10 times to calculate the mean and standard deviation of accuracies. The results from the experiments were as follows:

- Model 1: In Table 3, the model by medium dataset showed lower performance when compared to the model by small dataset. Specifically, from the graphs of training loss, we could observe that the training loss does not decrease after around the 100th step regardless of datasets. However, the training on the medium dataset took longer steps (i.e., around 400) than the steps on the small dataset (i.e., around 140), causing the overfitting of the model. This shows that the model cannot learn more knowledge from different datasets and even suffer from overfitting on the bigger dataset.

Dataset	Train Data		Test Data	
	Mean	Stdev	Mean	Stdev
Small	74.763	0.478	68.056	1.464
Medium	75.922	0.501	68.078	0.694

Table 3. Model 1 test results

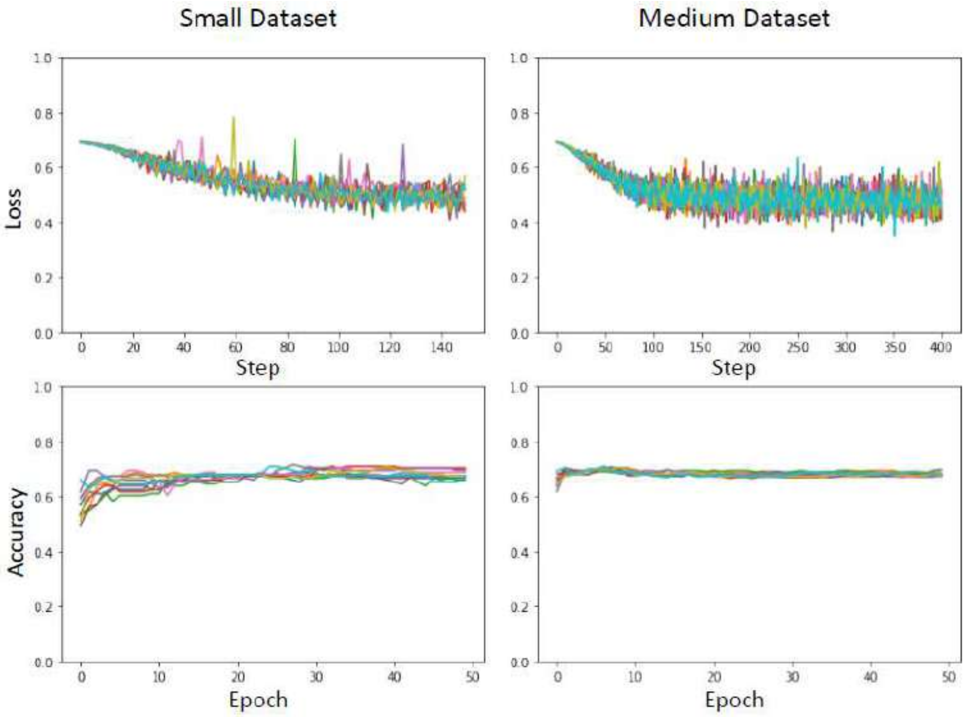


Figure 3. Model 1 small and medium datasets

- Model 2: In Table 4, the model by medium dataset showed higher performance when compared to the model by small dataset. From graphs of training loss, we could observe that model showed consistent improvement. As a result, the model by the medium dataset showed the higher accuracy than the model by the small dataset.

Dataset	Train Data		Test Data	
	Mean	Stdev	Mean	Stdev
Small	80.154	0.983	68.819	2.954
Medium	83.306	0.633	77.027	1.212

Table 4. Model 2 test results

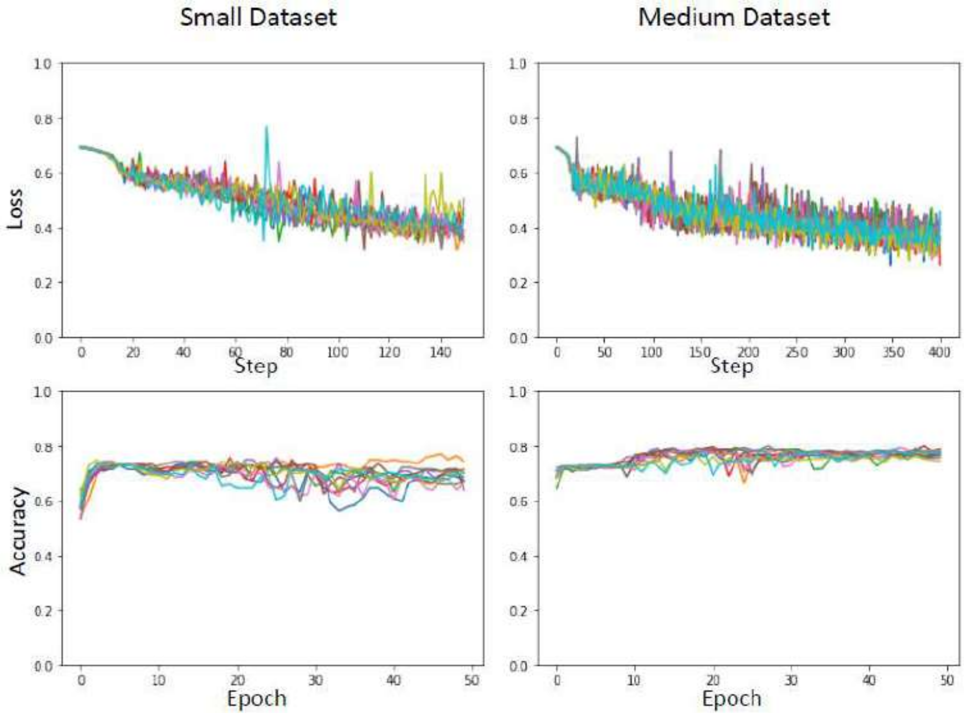


Figure 4. Model 2 small and medium datasets

Following results of experimental run, model 2 would be recommended as providing higher accuracy. The project aims to collect more various data from different countries thus ensuring large date set to be loaded for model training. Therefore, in our proposed methodology we suggest using model 2 as the one capable of predicting more accurately with larger data set. It is rather accurate and can be used efficiently for identifying the required skills in the market. Importantly, the model can be retrained periodically as the terminology and the taxonomy of skills change over time.

4. Conclusions

The analysis of relevant documents and research together with the methodologies applied has shown different angles and approaches to solve the same or very similar problems. A missing unified framework of cybersecurity roles and skills is one of the main obstacles to the development of skills needs methodology. Moreover, this lack makes difficult to efficiently consolidate obtained results and insights from various research. Taking into account and assessing methodologies applied in different projects, the proposal of a methodology to anticipate future needs was presented in this document. It consists of three key parts: (i) methodology applied in the pilot projects and other relevant documents, (ii) stakeholder survey, and (iii) automated job ads analysis.

Results of related projects review have shown that identification of competencies, skills and roles is essential for successful development of methodology to anticipate future needs as it gives common basis and necessary framework. NICE competencies framework was selected as starting point in preparation of REWIRE competencies and skills. From the NICE competencies, a total of 30 REWIRE competencies we selected. These competencies were split into three families: Cybersecurity Skills, IT Skills, and Soft Skills.

The stakeholder survey was based on selected skills and stakeholders, versatile questions helped to collect the market view from industry and educational institutions. To strengthen methodology and automate analysis of cyber security job advertisements posted in various websites, dictionary-based job advertisement analysis and machine learning model with described settings for efficient and accurate cyber security skills demand were introduced.

In general, emphasis is put on the need to have a unified framework of cybersecurity roles and skills and ensuring aligning of future research based on the unified framework. Seizing valuable insights from previous projects and combining it with proposed stakeholder survey and automated job ads analysis gives the opportunity to have a consistent model of future needs anticipation.

In addition, the co-operation with ENISA as one of the most important organizations of Cybersecurity in EU is an essential element in the development of unified framework which is also one of the goals of REWIRE project. Though, initial competencies and skills framework is used in currently proposed methodology, it should be further adjusted. Joining forces in preparation of a unified framework and adjusting it according to findings of each above-mentioned methodologies are key directions for coming future. Because of great variety of different frameworks used in different countries, the proposed methodology has to be used in an agile way, i.e., adjusting to changes and modify itself accordingly that will be done in following project steps.

5. References

1. SPARTA - D9.1 - Cybersecurity skills framework;
2. European Cybersecurity Skills Framework — ENISA (europa.eu);
3. Ad-Hoc Working Group on the European Cybersecurity Skills Framework — ENISA (europa.eu);
4. European e-Competence Framework (ecompetences.eu);
5. Edmundas Piesarskas, D9.1 Cybersecurity skills framework, January 2020, https://www.sparta.eu/deliverables/;
6. Jan Hajny, D9.2 Curricula descriptions, July 2020, https://www.sparta.eu/deliverables/;
7. Pavel Varbanov, D2.6 ECHO CYBERSKILLS FRAMEWORK 2021 https://echonetwork.eu/wp-content/uploads/2021/03/ECHO_D2.6_Cyberskills-Framework.pdf;
8. Good Practice Guide on Training Methodologies published by ENISA https://www.enisa.europa.eu/news/enisa-news/good-practice-guide-on-training-methodologies-published-by-enisa;
9. Felicia Cutas, CONCORDIA: Methodology for the creation and deployment of new courses and/or teaching materials for cybersecurity professionals https://www.concordia-h2020.eu/wp-content/uploads/2020/06/CONCORDIA-methodology-courses-professionals-for-publication.pdf;
10. Cyber Security for Europe - Education and Training Review https://cybersec4europe.eu/wp-content/uploads/2020/02/D6.2-Education-and-Training-Review-V1.2-Submtted.pdf;
11. Digital economy and society - Overview https://ec.europa.eu/eurostat/web/digital-economy-and-society/overview;
12. Berg, Stuart, et al. "Ilastik: interactive machine learning for (bio) image analysis." Nature Methods 16.12 (2019): 1226-1232;
13. Bontempi, Gianluca, Souhaib Ben Taieb, and Yann-Aël Le Borgne. "Machine learning strategies for time series forecasting." European business intelligence summer school. Springer, Berlin, Heidelberg, 2012;
14. Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors." nature 323.6088 (1986): 533-536;
15. Yin, Wenpeng, et al. "Comparative study of CNN and RNN for natural language processing." arXiv preprint arXiv:1702.01923 (2017);
16. Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018);
17. Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780;
18. Wang, Sun-Chong. "Artificial neural network." Interdisciplinary computing in java programming. Springer, Boston, MA, 2003. 81-100.

6. List of Abbreviations and Acronyms

Abbreviation	Explanation/ Definition
CONCORDIA	Cyber security cOmpeteNce fOr Research anD Innovation
CyberSec4Europe	Cyber Security For Europe
ECHO	European network of Cybersecurity centres and competence Hub for innovation and Operations
ENISA	European Cybersecurity Agency
EU	European Union
ICT	Information and Communication Technologies
KSA	Knowledge, Skills, and Abilities
LSTM	Long short-term memory - is an artificial neural network used in the fields of artificial intelligence and deep learning.
ML	Machine Learning
NICE	National Initiative for Cybersecurity Education
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
RNN	Recurrent Neural Network
SPARTA	Strategic Programs for Advanced Research and Technology in Europe

Table 1. List of abbreviations and acronyms