‘Comparing Vocational Education and Training Qualifications: towards a European Comparative Methodology’

Work Assignment 2: Exploring, gathering and analysing national qualifications data

Draft Final Report

Contract notice:
2017-FWC14/AO/DSI/JB/Comparing_VET_Qualifications/010/17

Framework Contract: Comparing Vocational Education and Training Qualifications: towards a European Comparative Methodology

Work Assignment 2: Exploring, gathering and analysing national qualifications data

Draft Final Report

Consortium leader / Contractor
3s Unternehmensberatung GmbH
Wiedner Hauptstrasse 18
1040 Wien
Austria

Consortium partner:
Ockham IPS, NL

Contact Person:
Karin Luomi-Messerer (Project Leader)
Phone: +43 1 585 09 15-41
E-mail: luomi-messerer@3s.co.at
www.3s.co.at

Authors (in alphabetical order):
Monika Auzinger, Simon Broek, Andrew McCoshan, Marye Hudepohl, Karin Luomi-Messerer, Claudia Plaimauer

Vienna, 31 October 2019

Title

Exploring, gathering and analysing national qualifications data

Cedefop project 'Comparing Vocational Education and Training Qualifications: towards a European Comparative Methodology'

Working paper 2

Foreword

This working paper forms part of the Cedefop project *Comparing Vocational Education and Training Qualifications: towards a European Comparative Methodology*.

The purpose of the project is to scale up and strengthen comparative methodologies so as to promote mutual learning between countries and facilitate strengthening of relevance and quality of VET qualifications at national level.

The Framework Contract includes four work assignments

(a) Work Assignment 1 - Exploring and testing a reference point for VET comparison;
(b) Work Assignment 2 - Exploring, gathering and analysing national qualifications data;
(c) Work Assignment 3 - Exploring, gathering and analysing data on the match/mismatch between qualifications and labour market requirements;
(d) Work Assignment 4 - Methodological synthesis.

[Name of the Director]
Director

Acknowledgements

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce eget ipsum vitae risus bibendum consequat et vitae mauris. Duis velit velit, aliquet id egestas id, porttitor quis lectus. Aliquam eget pulvinar sem.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce eget ipsum vitae risus bibendum consequat et vitae mauris. Duis velit velit, aliquet id egestas id, porttitor quis lectus. Aliquam eget pulvinar sem.

Table of contents

List of tables, figures and boxes

Tables

Figures

Boxes

Executive summary

To be developed for the final version of the final report

Lorem ipsum

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce eget ipsum vitae risus bibendum consequat et vitae mauris. Duis velit velit, aliquet id egestas id, porttitor quis lectus. Aliquam eget pulvinar sem.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce eget ipsum vitae risus bibendum consequat et vitae mauris. Duis velit velit, aliquet id egestas id, porttitor quis lectus. Aliquam eget pulvinar sem:

lorem
ipsum;
fusce;
eget.
dolor;
sit.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce eget ipsum vitae risus bibendum consequat et vitae mauris. Duis velit velit, aliquet id egestas id, porttitor quis lectus. Aliquam eget pulvinar sem.

Lorem ipsum

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce eget ipsum vitae risus bibendum consequat et vitae mauris. Duis velit velit, aliquet id egestas id, porttitor quis lectus. Aliquam eget pulvinar sem.

Chapter 1. Introduction

1.1 Setting the scene

1.1.1 Shifting the focus to data sources for national qualifications

In the first working paper of the Cedefop project 'Comparing Vocational Education and Training Qualifications: towards a European Comparative Methodology', potential reference points for comparing national qualifications were analysed, tested and their usability in different contexts of use was discussed ([1]). In the testing phase, country experts identified relevant national documents - describing the national VET qualifications and their learning outcomes - and analysed the extent to which their qualifications corresponded with the selected reference points or systems ([2]). These reference points were:

(a) **ESCO**, the European classification of Skills, Competences, Qualifications and Occupations ([3]): The following occupational profiles were used: healthcare assistant and ICT service technician.

(b) **O*NET**, the Occupational Information Network, the USA's primary source of vocational intelligence ([4]): Nursing Assistants and Network and Computer Systems Administrators;

([1]) This approach was based on a previous Cedefop project on 'The role of learning outcomes in supporting dialogue between the labour market and education and training; the case of vocational education and training' (2015-2017; Contract notice 2015/S 092-164546 of 13/05/2015; Auzinger et al., 2017; Bjørnåvold and Chakroun, 2017; Cedefop, forthcoming).

([2]) In the context of this study, a 'reference point' is understood as conceptual fixed point for mapping learning outcomes included in national qualifications in order to compare them and identify commonalities and differences of their content and profile. Reference points of this kind usually have the form of occupational skills profiles (OSP). In this study, OSP refer to profiles that describe the requirements or essential characteristics of occupations in terms of knowledge, skills, competences, professional interests, work values, etc. They can be independent profiles, e.g. referring only to a specific occupational profile, or they can be part of a more complex 'reference system'. A 'reference system' is a systematic approach to develop and maintain OSP for different economic sectors and occupational fields. It defines how OSP are developed and provides some kind of structuring the content of OSP.

([3]) https://ec.europa.eu/esco/portal/home

([4]) https://www.onetonline.org/

(c) **WorldSkills Standards Specifications** (WSSS) ([5]): WSSS Health and Social Care and WSSS IT Network Systems Administrator;

(d) **The VQTS** (Vocational Qualification Transfer System) based Competence Matrix 'Professional Care' developed in the European project 'HealthCareEurope' (HCEU) project ([6]).

Following the testing phase, the focus of the reflections was on the usability of these reference points for comparing qualifications, i.e. to what extent they meet the requirements identified for this usage context, their strengths and weaknesses.

The data sources for national qualifications - the documents describing the learning outcomes themselves - were not central to the research. However, it became clear that the sources for data on national qualifications and the ways they present the data differ to a large extent across countries. For example, the documents used can have different purposes and functions, in which the qualifications descriptions can either be issued as stand-alone documents, or they refer to programme documents that also include information on inputs and processes (such as the 'framework curricula' in the Austrian school-based VET system). Additionally, there might be a hierarchy of documents, where the higher-level documents (such as national educational standards) inform documents at a lower level of the hierarchy (such as VET programmes and curricula at provider level). The qualification descriptions can also differ in the extent and type of learning outcomes included and the description of learning outcomes themselves can show a different scope and detail of the descriptions ([7]). Moreover, the greater the difference between the way learning outcomes are formulated and structured in a qualification and the way this is done in the reference points, the greater the challenges and limitations of mapping qualifications to reference points will be.

Currently, there are no standardised ways for describing VET qualifications and their learning outcomes across Europe. During the last years, however, several initiatives have been launched to exchange experiences between countries on the interpretation and application of learning outcomes, without leading to harmonisation or standardisation of learning outcomes descriptions. For example, the EQF Note 4 on 'Using Learning Outcomes' was published in

([5]) https://www.worldskills.org/what/projects/wsss/

([6]) https://www.project-hceu.eu

([7]) This was, for example, also identified as a challenge in the 'pilot project on the horizontal comparison of levelled qualifications' (IBE, 2017).

2011 (European Union, 2011) and Cedefop has carried out several studies comparing learning outcomes approaches across countries, has so far organised three policy learning fora on defining and writing learning outcomes ([8]) and has published a handbook for defining and writing learning outcomes (Cedefop, 2017).

A standardised way for providing a rough overview of qualifications profiles is offered by the Europass Certificate Supplement (ECS): Based on a common template, it offers 'information on the content and learning outcomes associated with a qualification and on the education system of the country issuing the qualification' (European Union, 2018, p. 5). The ECS is one of the five Europass documents developed to help European citizens make their skills and qualifications clearly and easily understood in Europe.

Box 1. **Europass Certificate Supplement**

The 'certificate supplement' refers to a document attached to a vocational education and training or professional certificate, issued by the competent authorities or bodies, in order to make it easier for third persons – particularly in another country – to understand the learning outcomes acquired (by the holder of the qualification), as well as the nature, level, context, content and status of the education and training completed and skills acquired.'

Source: European Union, 2018, p. 16.

ECS are in many cases available in the national language as well as in English, and they are based on a common structure, usually including the following six elements:

(a) **Title of certificate** in original language;
(b) **Translated title** of the certificate;
(c) **Profile of skills and competences**;
(d) **Range of occupations accessible** to the holder of the certificate;
(e) **Official basis** of the certificate (Name and status of the body awarding the certificate, Name and status of the national/regional authority providing accreditation/recognition of the certificate Level of the certificate - national or international, Grading scale / Pass requirements, Access to next level of education/training, International agreements, Legal basis);
(f) **Officially recognised ways of acquiring the certificate** (e.g. duration, entry requirements).

([8]) http://www.cedefop.europa.eu/en/events-and-projects/projects/learning-outcomes/events

According to Cedefop (2017, p. 65), 'to date, 26 countries have developed supplements but, to varying degrees, applied a learning outcomes approach.' The revised Europass Decision refers to the potential use of ESCO in the Europass framework: 'Following appropriate testing, and having due regard for the position of Member States, ESCO could be used by the Commission within the Europass framework; the use of ESCO by Member States is on a voluntary basis, following testing with, and evaluation by, the Member States' (European Union, 2018, p. 7). To date, the Europass Certificate Supplements have not yet been revised in this regard.

The following table provides an overview of the strengths and weaknesses of using Europass Certificate Supplements in their current format for comparing qualifications:

Table 1. **Using Europass Certificate Supplements for comparing qualifications – strengths and weaknesses**

Strengths	Weaknesses
Common template/structure provided by Europass	Not all countries have a central inventory
Purpose and currencies are indicated to a certain extent (Range of occupations accessible to the holder of the certificate; Access to next level of education/training)	Not all countries have translated all Europass Certificate Supplements into English
	Profile of skills and competences is presented to a varying degree of detail and by using different formats and structures
	EQF level is not provided in all cases;
	Distribution of types of learning outcomes (general knowledge subjects, transversal learning outcomes and occupational learning outcomes) is not indicated

Source: Authors.

Another emerging solution to overcome the challenges - related to the diversity of data sources on national qualifications - might be the use of national qualifications databases or registers. Several countries have developed national databases for the collection, presentation and sharing of information on qualifications (including their learning outcomes). Such databases have often been set up in the context of NQF implementation, because they can be used to communicate qualifications frameworks and the qualifications contained in NQFs at national level. They are important to increase the transparency of qualifications across Europe, and provide end-users with direct access to information on

qualifications included in the NQF. When information on qualifications - included in the NQF - is based on the same underlying meta-data structure as suggested by the 2017 EQF Recommendation (Council of the European Union, 2017) and shared by all countries at European level, this can help to provide a comprehensive and accessible overview of qualifications at EU level.

Qualifications databases and registers containing qualifications included in NQFs, are at different stages of development; some countries have comprehensive registers in place including all levels and types of qualifications. The EU has supported the development of such databases in the context of the EQF implementation, by supporting EQF NCPs (National Coordination Points) with dedicated grants ([9]). This should also ensure their linkage to European portals, such as the 'Learning Opportunities and Qualifications Portal' (LOQ) ([10]) and ESCO ([11]). Moreover, support was provided in the context of the EQF Advisory Group (AG), by organising a Peer Learning Activity (PLA) in May 2019 - to discuss and exchange experiences, regarding the development and use of national qualifications registers and databases and their connection at European level (European Commission et al., 2019), and by commissioning a mapping study - to provide a comprehensive overview of national practices on developing and using databases and NQF portals in August 2019.

1.1.2 Making use of emerging digital tools

The challenges and limitations identified in the first phase of the Cedefop project 'Comparing Vocational Education and Training Qualifications: towards a European Comparative Methodology', and in previous studies using this comparative approach, did not only point to the **diversity of data sources of qualifications**, but also to **problems of scalability of the method**. While these exercises demonstrate that systematic comparison of national qualifications is indeed feasible, the approach proved time and resource-demanding, making it difficult and costly to repeat for other qualifications and countries. This also points to the challenge of consistency and reliability; the **interpretations and decisions**

([9]) Out of the ten countries covered by this study, only four had applied for EU grants to support database development: Bulgaria, Lithuania, the Netherlands and Austria.

([10]) https://ec.europa.eu/ploteus/en?cookie=no. All information on qualifications in the LOQ portal will be integrated in the new Europass platform to be launched in 2020.

([11]) https://ec.europa.eu/esco/portal. ESCO is conceptually and operationally structured around three pillars: the qualifications pillar, the skills pillar (contains knowledge, skills and competences), and the occupations pillar.

made by individual experts are not always transparent, raising questions of overall consistency of the methodology.

Referring to the above challenges, it is necessary to identify (existing and emerging) solutions for gathering and analysing national data on qualifications. While comparison of qualifications will always, at some stage, and to some extent, involve qualitative interpretation of data (thus involving VET experts), 'automated' data gathering could be used as well. New digital technologies for gathering data on qualifications include in particular: data-mining, web-crawling, text-mining, text analytics (including entity extraction), complex data visualisation, computational linguistics, semantic technologies and, lastly, semantic deep learning and reasoning. Digital technologies might help to develop more efficient procedures, saving time and resources (and thereby lower costs), which can also be used beyond dedicated projects.

This approach is of particular interest in the ESCO context, for linking the qualifications pillar with the skills pillar. Thus, the European Commission commissioned a study in 2018, to investigate the feasibility of the conceptual and technical link between the learning outcomes of qualifications contained in the ESCO qualifications pillar and the ESCO skills pillar. The study investigated the extent to which automated information technology tools can be used for this purpose and discussed different options, including a wholly 'human' solution (i.e. manual linking) and a highly intensive artificial intelligence solution, and also 'shows the potential for a more pragmatic solution which embraces both the best use of human and technologies' (DG EMPL, 2019, p. 8).

1.2 Main research questions, methodological approach and structure of the working paper

1.2.1 Objectives and key research questions

The aim of this paper is twofold: First, to analyse existing data sources for the presentation of qualifications and their relevance for the cross-national comparison of qualifications. Furthermore, to discuss the strengths and weaknesses of the use of important 'official' documents - of different types and with different functions - for comparison purposes. A particular focus will be on national qualifications databases, developed for access to qualifications data. Secondly, it will look at how and to what extent new technologies can help to support the collection, processing and comparison of data on qualifications.

In particular, this report addresses the following questions:

Box 2. **Core research questions**

1.What are the key sources for data on national qualifications, in particular related to their content and profile?

1.1 Which data sources exist and are of relevance for the comparison of national qualifications?

1.2 To what extent can national qualifications databases support comparisons of VET qualifications?

2. How can new digital technologies support automated gathering, structuring (including cleaning, fusion) and analysis of data on qualifications?

3. How can new digital technologies address the linguistic challenges involved in comparing qualifications?

4. What can be the role of the multilingual classification ESCO in supporting gathering, structuring and classifying qualifications data?

Source: ToR.

The study mainly focuses on IVET qualifications from the following ten countries: Austria ([12]), Bulgaria, Denmark, Finland, France, Ireland, Lithuania, the Netherlands, Spain, and United Kingdom-England. If relevant, the qualification of the ICT service technician (also analysed in WA1) was used as an example to illustrate certain aspects related to core research question 1.

1.2.2 Methodological approach

The methodological approach included the following steps and research activities:

In order to explore **key sources for data on national qualifications and qualification databases** in the ten countries covered by this study (core research question 1), a first step was to introduce a set of criteria or conditions that need to be in place to support comparison.

Based on these criteria, a template for collecting information on the following aspects was developed:

(a) 'Reference documents' presenting qualifications and their learning outcomes (information availability and where to find it, descriptions of learning outcomes of IVET qualifications, infrastructure behind the description and storage of IVET qualification information);

(b) National qualifications databases (descriptions of learning outcomes of IVET qualifications included in the database, technical infrastructure).

([12]) With regard to Austria, the focus is on apprenticeship qualifications, but information on school-based IVET qualifications was also included where necessary.

The template also includes a section on tools for comparing learning outcomes of qualifications at national level to examine whether digital tools are already available that can be used to support the automated collection, structuring and analysis of qualifications data (core research questions 2 and 3).

The template was completed by country researchers, mainly based on desk research. In addition, expert interviews were conducted to validate the results and/or fill information gaps. Where appropriate, the national ICT service technician qualification (analysed in WA1) was used as an example. The following table gives an overview of the qualifications used in this study, for the ten countries surveyed.

Table 2. **ICT service technician qualifications used for illustration purposes**

Country	ICT service technician qualification	EQF level
Bulgaria	Application Software Developer	4
Denmark	IT-supporter specialised in infrastructure	4
Spain	Higher Technician in Computer Network Systems Management	5
France	ICT support technician	4
Ireland	Computer Systems and Networks	4
Lithuania	ICT maintenance and service specialists	4
Netherlands	ICT management assistant	3
Austria	Certificate of Apprenticeship Information Technology Specialising In Systems Engineering	4
Finland	Vocational Qualification in Information and Telecommunications Technology. (Competence area in Information and Telecommunications Technology - ICT Technician)	4
UK-England	Level 4 Diploma for ICT Professionals - Systems and Principles	5*

Source: Country templates. *It is a level 4 qualification in the UK and this level is referenced to EQF level 5; however, it can be argued that would fit better to EQF level 4.

The information collected was analysed, to determine whether and to what extent the criteria or conditions supporting a comparison of qualifications were met.

The possible **use of digital technologies** to support the automated collection, structuring and analysis of data on qualifications and the comparison

of qualifications was explored by performing desk research, in order to: (1) propose a workflow which outlines the separate steps that would be involved for the automated comparison of qualifications within a context of text mining, text analysis and machine learning; and (2) to identify existing or new digital technologies (preferably freeware), that could potentially be used for the purpose of automated comparison of qualifications. After identifying potential technologies and selecting the most-promising ones, a prototype was developed from scratch by building upon these tools - using the insights gained from examples of their application, seen on relevant discussion forums ([13]) and insights from expert consultations – by tailoring them to fit the proposed workflow (through trial and error), while simultaneously identifying and addressing / resolving issues as they were encountered. After developing a feasible prototype for (parts of) the workflow, with some adaptations ([14]), a testing exercise was performed – using the (full) ESCO KSC skills pillar and the learning outcomes descriptions included in the Dutch national documentation for the qualification of ICT service technician – in order to analyse to what extent the texts included in these descriptions could be matched to the right skills in ESCO.

1.2.3 The structure of this report

The following chapters present the findings of this study:

Chapter 2 introduces dimensions or conditions for the suitability of data sources of qualifications for the comparison of qualifications and for the automated text processing of qualifications data to support comparison. With reference to these dimensions, it discusses and assesses the key sources for data on national qualifications and their suitability for the comparison of national qualifications. Particular emphasis was placed on examining the presentation of data in national qualifications databases, the sources used for their development, whether they are based on a common metadata scheme and whether they correspond to a common format for the presentation of learning outcomes.

([13]) Within the scope of this study, these are mainly: 'StackOverflow'; Reddit (/r/AskProgramming); and the 'Software Engineering StackExchange'

([14]) Due to challenges encountered during development, and issues regarding the feasibility that became apparent through expert consultations it became clear using machine learning elements in the workflow would require more time and human capacity than feasible for the scope of this study. The difference in approach is that the final prototype is mainly aimed at **supporting** the text analysis and (manual) comparison between qualifications and the full ESCO skills pillar; rather than a fully automated comparison based on the pre-selected OSPs.

Chapter 3 focuses on how qualification data can be gathered and analysed by using digital technologies (technologies for 'automated' data processing). It presents conditions for the operation of an automated workflow for comparing qualifications, including the individual steps and sub-tasks of such a work flow. Furthermore, this chapter discusses the operationalisation of the workflow in terms of the development process of the prototype and the challenges encountered.

Chapter 4 provides insights into the testing exercise that uses the developed prototype to analyse and 'match' learning outcomes descriptions from one selected national qualification document for ICT service technician (NL) to the (full) list of KSCs included in the ESCO skills pillar – while reflecting on the extent to which differences in phrasing (lexical gap) pose challenges in doing so.

Chapter 5 presents conclusions in relation to the key research questions of this study and recommendations in view of the overall aim of the Cedefop project, to prepare methodologies allowing for a systematic comparison of VET qualifications.

Chapter 2. Key sources for data on national qualifications and their relevance for comparison of national qualifications

2.1 Conditions for suitability of data sources for comparing national qualifications

In order to indicate the relevance of qualifications data sources for comparison of national qualifications, we first need to introduce a set of criteria or conditions that need to be in place to support comparison. While work assignment 1 (WA1) focused on reference points and has identified requirements that a reference point needs to meet in different usage contexts (including comparison of qualifications), this study concentrates on the other side of the comparison, namely the qualification descriptions. Nevertheless, we can still learn from what was identified as criteria that need to be fulfilled for the purpose of **international comparison of VET qualifications** and **automated text processing of qualifications data.** The following dimensions have been identified that will be used to assess the suitability of data sources for qualifications for these purposes.

The necessary conditions **('must haves') refer to the following aspects:**

(a) **Unit of analysis**: The source describes (together with other sources in case multiple reference documents are used) the learning outcomes of a qualification.

(b) **Completeness of the learning outcomes description**: The source provides a full account of the learning outcomes related to obtaining the qualification.

(c) **Sentence components:** Learning outcomes statements are preferably composed of the components as suggested by Cedefop (2017, p. 47) ([15]): action verb, object of the verb, statement specifying the depth/breadth of

([15]) The feasibility study on the conceptual and technical link between the learning outcomes of qualifications contained in the ESCO qualifications pillar and the ESCO skills pillar also emphasised the use of action verbs and objects for enhancing results accuracy and concluded that the potential for automated linking is maximised if conformance with the Cedefop guidelines on describing learning outcomes is ensured (DG EMPL, 2019, p. 75).

learning to be demonstrated, indication of the context. This is a necessary condition for the purposes addressed in this study since it would make analysis and comparison of learning outcomes easier because it supports 'part-of-speech tagging' (also called 'grammatical tagging' or 'word-category disambiguation') which is understood as a process of assigning a 'part-of-speech' (such as noun, verb, pronoun, preposition, adverb, conjunction, adjective, and article) to each word in a sentence.

(d) **Information related to 'key comparability criteria'**: Not all aspects of a qualification that are relevant for cross-country comparison can be described with learning outcomes. The study on VET qualifications linked to EQF levels 3 and 4 (Cedefop, forthcoming) identified the following aspects as 'key comparability criteria' for which information should be provided: the EQF level, the distribution of types of learning outcomes (general knowledge subjects, transversal learning outcomes and occupational learning outcomes) ([16]); the purpose and currencies of qualifications and the extent to which qualifications provide access to further learning and (conditional/limited) access to higher education.

(e) **Coverage of qualifications**: For the majority of individual qualifications belonging to a qualification type ([17]), a consistent approach should be used

([16]) 'General knowledge subjects' usually include languages, maths, history, geography, etc. In the higher education (particularly university) context, a polarisation can be observed between 'disciplinary knowledge' (which 'is abstract and generally emphasises conceptual understanding that is defined and legitimised from within the disciplines and forms the ground for disciplinary identity') and 'relevant skills and knowledge'. Some researchers assert a risk that learning outcomes or 'know how' knowledge might replace conceptual disciplinary knowledge (Muller and Young, 2014, p. 137 – in: Prøitz et al., 2017, p. 33). 'General knowledge subjects' could also be considered transversal learning outcomes as these general knowledge subjects are often provided in VET courses that are not attached to occupational learning outcomes. One could however as well argue that within some of these courses, occupation-specific content is integrated (for instance in languages or math). In other cases, this is usually only the case to a limited extent, such as history or geography, but the learning outcomes associated with them are sometimes of high importance, in particular for preparation for higher education. In general, it can be observed that there are no clear distinctions and that the different terminological approaches tend to emphasise different aspects and lead to a variety of different categories and structures (Cf. Note JAG 2-4: Meetings of the EQF Advisory Group and ESCO Member States Working Group, 5-6-7 February 2019: The need for an agreed terminology on transversal skills and competences).

([17]) 'A qualification type is a group or cluster of qualifications within a country that share specific characteristics, for example in terms of subsystem they belong to, legal regulations and regulatory body, purpose, general educational objectives as well as duration of related programmes, access requirements or level of labour market entry. Within a qualification type, there can be many different qualifications with regard to

for presenting them. This ensures that analysis conducted for one qualification is also valid for others belonging to the same type. Thus, the data source needs to cover the majority of qualifications in a similar manner (but not in the same document) to allow comparison within a country and between countries.

(f) **Up to date**: The source needs to provide an up-to-date picture of the qualification. The validity of the description needs to be ensured.

Next to these necessary conditions, other aspects can be indicated that support international comparison of VET qualifications and automated text processing of qualifications data. These **optional conditions ('nice to haves') include the following aspects:**

(a) **Languages**: The source preferably provides the information in the national language and in English. However, this depends on the reference point used: Since ESCO vocabulary is translated into other languages, this is not a necessary condition when using ESCO occupational profiles as reference points.

(b) **Structure for presenting learning outcomes descriptions**: The source preferably applies a structured approach to presenting learning outcomes descriptions. This refers, for example, to the use of headlines for grouping learning outcomes related to the domains of learning (e.g. into knowledge, skills and competence). It is also useful if learning outcomes are presented at different levels of specificity (using headlines presenting them at a more general level and then listing more specific learning outcomes).

(c) **Storage format**: The source is preferably formatted in a way that allows easy extraction and application in software packages.

In order to better understand the characteristics of qualifications, it is important that the source also provides **additional information** that characterises the qualification and its **context**. This includes, for example, information on the awarding body of the qualification or the legal basis of the qualification – in general, the elements that are included in the Europass Certificate Supplement or the elements for the data fields for the electronic publication of information on qualifications with an EQF level.

---

the content: the specific learning outcomes they include can be quite different because they are linked to different fields (such as different technical fields, social and health care, business)' (Cedefop, 2014, p. 142).

It is expected that there will be no readily-available sources that fulfil all necessary and optional conditions and that for all sources, additional steps would need to be taken to prepare them for comparison.

The following sections first discuss the main features of the 'national reference documents' and then the national qualifications databases to assess their suitability for the comparison of national qualifications and the automated text processing of qualifications data. Although this analysis is presented in separate sections, there are some overlaps: In some countries, the descriptions of learning outcomes in the reference documents are identical to those in the databases analysed. In these cases, the analysis of learning outcomes descriptions is only presented in Section 2.2.2 and not repeated in Section 2.3.6.

2.2 Documents as key data sources for presenting qualifications and their learning outcomes ('national reference documents')

This section examines the nature of documents containing qualifications and associated learning outcomes. It considers the documents in terms of the organisations responsible for them, the different types of information they contain, the inter-relationships between them, and how learning outcomes are structured, presented and written. It also examines how learning outcomes are stored. The table below provides an overview of key reference documents containing information relevant to learning outcomes and the responsible organisations in the ten countries studied ([18]).

([18]) Please note: the information provided in the table is based on the ICT service technician qualifications analysed.

Table 3. **Overview of key reference documents containing information relevant to learning outcomes and the responsible organisations**

Country	Learning outcomes descriptions (full and short summary)	NQF & EQF level	Possibilities for further learning (particularly: access to higher education)	Link to occupations/ labour market/ role of the qualification in the occupational context	Additional contextual information (such as the awarding body or the legal basis for the qualification or any other contextual and institutional – input – factors)
Bulgaria	**State Educational Standard** (*Държавен Образователен Стандарт)* National Agency for Vocational Education and Training (NAVET)	**State Educational Standard** (*Държавен Образователен Стандарт)* National Agency for Vocational Education and Training (NAVET)	**State Educational Standard** (*Държавен Образователен Стандарт)* National Agency for Vocational Education and Training (NAVET)	**State Educational Standard** (*Държавен Образователен Стандарт)* National Agency for Vocational Education and Training (NAVET)	**State Educational Standard** (*Държавен Образователен Стандарт)* National Agency for Vocational Education and Training (NAVET) **VET Act & Framework programmes** Ministry of Education and Science.
Denmark	**Executive Order** for the relevant qualification (*Uddannelsesbekendtgørelse for data- og kommunikationsuddannelsen/IT-supporter*) and the relevant **Training Regulation** (*uddannelsesordning for data- og kommunikationsuddannelsen/IT-supporter).* Ministry of Education Local level adaptations are described in **local training plans** *(lokale uddannelsesplaner)* Local Trade Committees (one for each vocational school)	**Europass Certificate Supplement** Danish Agency for Science and Higher Education	**Executive Order** for the relevant qualification (*Uddannelsesbekendtgørelse for data- og kommunikationsuddannelsen/IT-supporter*) Ministry of Education	**Occupational profile** Relevant Trade Committee	**Executive Orders and Vocational Training Act** Ministry of Education

Country	Learning outcomes descriptions (full and short summary)	NQF & EQF level	Possibilities for further learning (particularly: access to higher education)	Link to occupations/ labour market/ role of the qualification in the occupational context	Additional contextual information (such as the awarding body or the legal basis for the qualification or any other contextual and institutional – input – factors)
Spain	**Royal Decree** (*Real Decreto),* **Europass Diploma Supplement** Ministry of Education and Vocational training and Autonomous Communities according to their areas of competence	**Royal Decree** (*Real Decreto),* **Europass Diploma Supplement** Ministry of Education and Vocational training and Autonomous Communities according to their areas of competence	**Royal Decree** (*Real Decreto),* **Europass Diploma Supplement** Ministry of Education and Vocational training and Autonomous Communities according to their areas of competence	**Royal Decree** (*Real Decreto),* **Europass Diploma Supplement** Ministry of Education and Vocational training and Autonomous Communities according to their areas of competence	**Royal Decree** (*Real Decreto),* **Europass Diploma Supplement** Ministry of Education and Vocational training and Autonomous Communities according to their areas of competence
France	**Qualification Standards** *(Référentiel de certification*) Ministry of Education (Ministry of higher education and research when there is one) **National Catalogue of Qualifications** (RNCP: Répertoire national des certifications professionnelles) France compétences (formerly CNCP, Commission nationale de la certification professionnelle; National Commission of Qualifications)	**Qualification Standards** *(Référentiel de certification*) Ministry of Education (Ministry of higher education and research when there is one) **National Catalogue of Qualifications** (RNCP: Répertoire national des certifications professionnelles) France compétences (formerly CNCP, Commission nationale de la certification professionnelle; National Commission of Qualifications)	**The Education Code** Ministry of Education (Ministry of higher education and research when there is one)	**Qualification Standards** *(Référentiel de certification*) Ministry of Education (Ministry of higher education and research when there is one) **National Catalogue of Qualifications** (again) (RNCP: Répertoire national des certifications professionnelles) France compétences (formerly CNCP, Commission nationale de la certification professionnelle; National Commission of Qualifications)	**Qualification Standards** *(Référentiel de certification*). Ministry of Education (Ministry of higher education and research when there is one) **National Catalogue of Qualifications** (again) (RNCP: Répertoire national des certifications professionnelles) France compétences (formerly CNCP, Commission nationale de la certification professionnelle; National Commission of Qualifications)

Country	Learning outcomes descriptions (full and short summary)	NQF & EQF level	Possibilities for further learning (particularly: access to higher education)	Link to occupations/ labour market/ role of the qualification in the occupational context	Additional contextual information (such as the awarding body or the legal basis for the qualification or any other contextual and institutional – input – factors)
Ireland	**Certificate Specification for a Major Award** and **Component Specifications for the Minor Awards** which comprise the Major Award Quality and Qualifications Ireland **Programme Modules** related to Minor Awards Education and Training Board responsible for developing the Programme Descriptor and Associated Programme Modules on behalf of providers	**NQF:** **Certificate Specification for a Major Award** and **Component Specifications for the Minor Awards** which comprise the Major Award Quality and Qualifications Ireland **EQF:** **Europass Certificate Supplement** Quality and Qualifications Ireland	**Certification Specification for a Major Award** Quality and Qualifications Ireland	**Certificate Specification for a Major Award** and **Component Specifications for the Minor Awards** which comprise the Major Award Quality and Qualifications Ireland	**Certificate Specification for a Major Award** and **Component Specifications for the Minor Awards** which comprise the Major Award Quality and Qualifications Ireland
Lithuania	**Occupational standard** *(Profesinis standartas)* **Modular VET curricula** *(Modulinės profesinio mokymo programos)* Centre for Development of Qualifications and Vocational Education and Training *(Kvalifikacijų ir profesinio mokymo plėtros centras)*	**Occupational standard** *(Profesinis standartas)* **Modular VET curricula** *(Modulinės profesinio mokymo programos)* Centre for Development of Qualifications and Vocational Education and Training *(Kvalifikacijų ir profesinio mokymo plėtros centras)*	**Occupational standard** *(Profesinis standartas)* Centre for Development of Qualifications and Vocational Education and Training *(Kvalifikacijų ir profesinio mokymo plėtros centras)*	**Occupational standard** *(Profesinis standartas)* **Modular VET curricula** *(Modulinės profesinio mokymo programos)* Centre for Development of Qualifications and Vocational Education and Training *(Kvalifikacijų ir profesinio mokymo plėtros centras)*	**Occupational standard** *(Profesinis standartas)* Centre for Development of Qualifications and Vocational Education and Training *(Kvalifikacijų ir profesinio mokymo plėtros centras)*

Country	Learning outcomes descriptions (full and short summary)	NQF & EQF level	Possibilities for further learning (particularly: access to higher education)	Link to occupations/ labour market/ role of the qualification in the occupational context	Additional contextual information (such as the awarding body or the legal basis for the qualification or any other contextual and institutional – input – factors)
Netherlands	**Qualification File** (*Kwalificatiedossier*), **Occupational Profile** (*Beroepsbeschrijving*) Foundation for Cooperation on Vocational Education, Training and Labour Market (SBB) *(Stichting Samenwerking Beroepsonderwijs Bedrijfsleven, SBB)*	**Qualification File; Occupational Profile; Europass Certificate Supplement** (*Kwalificatiedossier; Beroepsbeschrijving; Europass Certificaatsupplement*) Foundation for Cooperation on Vocational Education, Training and Labour Market (SBB) *(Stichting Samenwerking Beroepsonderwijs Bedrijfsleven, SBB)*	**Europass Certificate Supplement** *(Europass Certificaatsupplement)* Foundation for Cooperation on Vocational Education, Training and Labour Market (SBB) *(Stichting Samenwerking Beroepsonderwijs Bedrijfsleven, SBB)*	**Qualification File; Occupational Profile; Europass Certificate Supplement** (*Kwalificatiedossier; Beroepsbeschrijving; Europass Certificaatsupplement*) Foundation for Cooperation on Vocational Education, Training and Labour Market (SBB) *(Stichting Samenwerking Beroepsonderwijs Bedrijfsleven, SBB)*	**Europass Certificate Supplement** *(Europass Certificaatsupplement)* Foundation for Cooperation on Vocational Education, Training and Labour Market (SBB) *(Stichting Samenwerking Beroepsonderwijs Bedrijfsleven, SBB)*
Austria	**Training Regulation** (*Ausbildungsordnung*) Federal Ministry for Digital and Economic Affairs	**Europass Certificate Supplement** Federal Ministry for Digital and Economic Affairs	Partly regulated in **the Vocational Training Act** (*Berufsausbildungsgesetz - BAG*) Federal Ministry for Digital and Economic Affairs **Europass Certificate Supplement** Federal Ministry for Digital and Economic Affairs	**Training Regulation** (*Ausbildungsordnung*) Federal Ministry for Digital and Economic Affairs **Europass Certificate Supplement** Federal Ministry for Digital and Economic Affairs	**Training Regulation** (*Ausbildungsordnung*) Federal Ministry for Digital and Economic Affairs **Vocational Training Act** (*Berufsausbildungsgesetz - BAG*) Federal Ministry for Digital and Economic Affairs **Framework curriculum for part-time vocational school** (*Lehrplan für die Berufschule*) Federal Ministry of Education, Science and Research

Country	Learning outcomes descriptions (full and short summary)	NQF & EQF level	Possibilities for further learning (particularly: access to higher education)	Link to occupations/ labour market/ role of the qualification in the occupational context	Additional contextual information (such as the awarding body or the legal basis for the qualification or any other contextual and institutional – input – factors)
Finland	**National Vocational Qualification Requirements** The Finnish National Agency for Education (EDUFI)	**Europass Certificate Supplement** The Finnish National Agency for Education (EDUFI)	**Universities Act 558/2009 Universities of Applied Sciences Act 932/2014 ([19])** Ministry of Education and Culture **Europass Certificate Supplement** The Finnish National Agency for Education (EDUFI)	**National Vocational Qualification Requirements** The Finnish National Agency for Education (EDUFI)	**Act on Vocational Education and Training** (Laki ammatillisesta koulutuksesta) and **Decree on VET** (Asetus ammatillisesta koulutuksesta) Ministry of Education and Culture **Provider-specific curricula** VET providers
UK-England	Relevant **qualification document/specification** (no generic name) Relevant awarding organisation (e.g. City and Guilds)	Information available in the **Regulated Qualifications Framework** administered by The Office of Qualifications and Examinations Regulation Ofqual.	Relevant **qualification document/specification** (no generic name) Relevant awarding organisation (e.g. City and Guilds)	Relevant **qualification document/specification** (no generic name) Relevant awarding organisation (e.g. City and Guilds)	Relevant **qualification document/specification** (no generic name) Relevant awarding organisation (e.g. City and Guilds)

Source: Country templates.

([19]) These are the documents where the legal basis for access into higher education is defined. The VET legislation refers to these two acts when stating the eligibility for higher education studies.

2.2.1 Mapping of reference documents presenting qualifications and their learning outcomes

Across the ten countries covered by this study, a range of types of documents contains full and short summary descriptions of learning outcomes. These are variously called qualification/certification specifications/requirements (FI, IE, UK-EN), qualification standards (FR), qualification files/profiles (NL), VET standards (BG), occupational standards (LT), and training regulatory documents (AT, ES, DK).

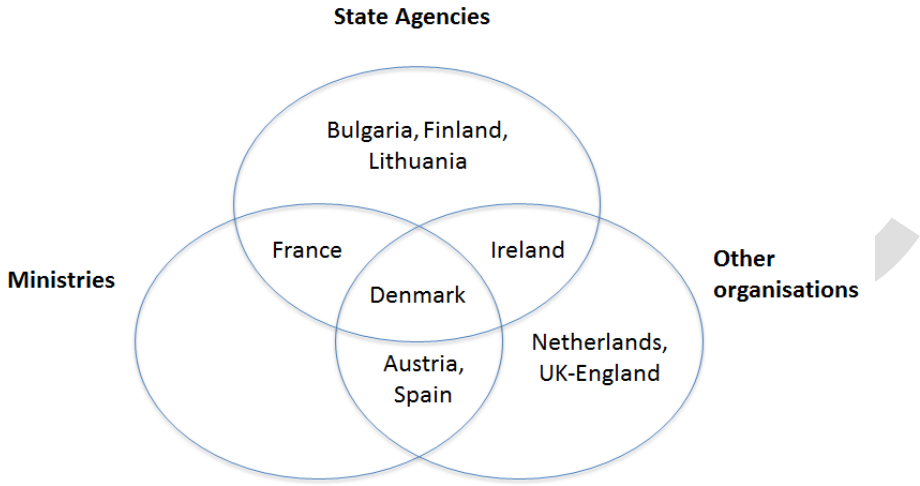
A range of organisations has responsibility for these documents. Ministries are mainly involved where learning outcomes descriptions are contained in a training regulation/executive order (AT, DK) or Royal Decree (ES). Across the ten countries, however, other organisations, most notably state agencies, tend to be the most important organisations in terms of responsibility for the documents containing information relating to learning outcomes (see figure below).

In **Austria**, whilst the Ministry is involved, so too is the ibw (ibw Austria research & development in VET) which also drafts the Europass Certificate Supplements (ECSs). In **Spain**, as well as the Ministry of Education and Vocational training, Autonomous Communities may have responsibility for documents containing information relating to learning outcomes, according to their areas of competence. In **France**, a Ministry is responsible for the issuing of what is termed a 'qualification standard', whilst a state agency, France Compétences, maintains a catalogue of qualifications. In **Denmark**, trade committees of social partners develop the occupational profiles that provide the basis for executive orders that specify learning outcomes issued by the Ministry of Education, and the Danish Agency for Science and Higher Education issues the ECS where the NQF and EQF levels are recorded. In **Ireland**, Quality and Qualifications Ireland (QQI) is responsible for the key documentation relating to most IVET qualifications; it is an independent State agency responsible for promoting quality and accountability in education and training services. At the same time, 16 Education and Training Boards (ETBs) are responsible for IVET programmes that also contain information relating to learning outcomes ([20]). In the **Netherlands**, all documentation containing information relating to learning outcomes is the responsibility of the Foundation for Cooperation on Vocational

([20]) The learning outcomes presented in the specifications for which the QQI is responsible at national level are 'expected' learning outcomes which the ETBs should use to devise 'minimum intended' learning outcomes at programme/curriculum level (see also section 2.2.1.1).

Education, Training and Labour Market *(Stichting Samenwerking Beroepsonderwijs Bedrijfsleven, SBB)***.** SBB is a tripartite organisation (government, employers, VET sector) that carries out tasks commissioned by the Ministry. **In the UK-England**, independent awarding organisations are responsible for documentation containing learning outcomes.

Figure 1. **Responsibility for documents containing information relating to learning outcomes**



Source: Authors, based on country templates.

Full and short summaries of learning outcomes may also occur in other documents, e.g. curriculum documentation. In **Lithuania**, for example, the *occupational standard* contains the list of the units of qualification – corresponding to the core work areas for the occupation - composed of competences, described as core work tasks; whereas the introductory part of the profile of the qualification contains the outline of the core work tasks. Meanwhile, modular VET curricula documentation contains extensive and detailed descriptions of learning outcomes, based on the descriptors of qualifications of the occupational standards: competences defined in the units of qualification are split into learning outcomes of the module: knowledge, skills and values (attitudes).

A range of information is important to fully understand and contextualise the written descriptions of learning outcomes, including: the NQF and EQF level of the qualification, possibilities for further learning, particularly access to higher education; the link to occupations/ labour market/ role of the qualification in the occupational context; the distribution of types of learning outcomes (general knowledge subjects, transversal learning outcomes and occupational learning outcomes); and additional contextual information (such as the awarding body or the legal basis for the qualification or any other contextual and institutional – input

– factors). Regarding the distribution of types of learning outcomes (in terms of the percentage of each type), it should be noted that this proved to be unanswerable in most cases, owing to the interwoven nature of the different types.

A general key question is whether this additional information is located in the same documents as the descriptions of learning outcomes, or whether other documents need to be sourced. The findings are as shown below. Quite often, these other documents are the responsibility of authorities different to those responsible for documents containing the full or short summaries of learning outcomes:

(a) **NQF and EQF levels** are indicated either in the same documents as the full and short descriptions of learning outcomes (BG, LT) or in the relevant Europass Certificate Supplement (AT, DK, FI ([21]), NL ([22])) or in both (FR, ES, IE) or in a national register (UK-EN). The NQF and EQF are given in the same document, except in Ireland where the NQF level is in the Certification Specification and the EQF in the ECS.

(b) **Possibilities for further learning** are indicated either in the same documents as the full and short descriptions of learning outcomes (BG, ES, IE, LT, UK-EN) or other documents, notably Europass Certificate Supplement (AT, NL, FI) ([23]). Such information is also sometimes provided in wider over-arching documentation, such as the Education Code in France and the Vocational Training Act in Austria.

(c) **Information on links to occupations/the labour market tends** to be provided in the documents that also contain the full and short learning outcomes descriptions. This is the case in all countries except Denmark, where such information is included in the occupational profiles developed by trade committees that provide the basis for the Executive Orders, in which the learning outcomes are described. Sometimes, information on links to occupations/the labour market additionally appears in ECSs (AT, NL).

(d) Many countries do not explicitly contain **information on different types of learning outcomes**, e.g. general, occupationally specific, transversal (e.g.

([21]) In Finland, the ECS is provided as part of the qualification documentation.

([22]) In the Netherlands, the levels are explicit in the ECS and implicit in the documents that contain the full and short learning outcomes descriptions.

([23]) In the case of the Spanish ICT service technician qualification, this information is provided in the Europass Diploma Supplement – see: http://www.todofp.es/dam/jcr:525bf081-bd23-482f-bd89-1151f2498373/tsadministracionsistemasinformaticosreden-pdf.pdf

AT, ES, IE, LT, NL, UK-EN). In Bulgaria, although there is no clear distinction between occupational and transversal learning outcomes, VET qualifications include *general modules* common to all VET qualifications (e.g. work safety, entrepreneurship), *sector-specific modules*, which contain both occupational and transversal learning outcomes, and *occupation-specific modules*. In Denmark, learning outcomes for general knowledge subjects and transversal learning outcomes are included in the Vocational Training Act (*erhvervsuddannelsesloven*) and the General Executive Order, insofar as these are relevant to all IVET qualifications. In individual qualifications, general knowledge subject and transversal learning outcomes particular are usually listed after the occupational learning outcomes.

2.2.1.1 *Relationship between reference documents*

In many countries, information on qualifications and their learning outcomes are contained in a number of different sources, which are interlinked or aligned to some extent. When defining the content and profile of IVET qualifications, countries or VET sub-sectors have different choices on how to combine these documents. There can be different logics for developing these documents and the learning outcomes, and therefore differing (hierarchical) relationships between these key documents. In general, such interlinked documents can be identified on five levels of the education system: the national, the institutional, the study programme, the module and the level of structuring of teaching, learning and assessment (Prøitz et al., 2017, p. 35). The 'higher-level' documents usually present learning outcomes in a more general way and inform the more specific ones at the 'lower levels'.

As shown above, in several of the ten countries covered by this study, information on qualifications and their learning outcomes are contained in a number of different sources which are interlinked or aligned to some extent. The countries studied here reveal that there are two aspects to consider: (i) documents at national level and (ii) sets of documents that together span the 'vertical' institutional/governance dimension of VET systems.

Documents at national level tend to be described as non-hierarchical, in the sense that they form equally important parts of a whole, all being necessary to form a full understanding of learning outcomes. As described above, there tends to be main document(s) that contain the learning outcomes and other documents, most noticeably the ECS, that contain other information important for a full understanding of the learning outcomes. The box below shows the relationship between documents in the Netherlands.

Box 3. **Relationship between documents related to learning outcomes at national level in the Netherlands**

In the Netherlands, three reference documents have a role in respect of learning outcomes: a qualification file; an 'occupation in short' leaflet; and the ECS. All three are drafted at national level, and there is no hierarchical relationship between them. Even though all three documents provide at least an overview of the relevant learning outcomes by listing the core tasks and work processes a qualification holder can perform, only the qualification profile provides any further descriptions in terms of learning outcomes. Even though the qualification profile and 'occupation in short' leaflet are linked in that they are retrieved from the same SBB (Foundation for Cooperation on Vocational Education, Training and Labour Market) registry, the 'occupation in short' leaflet is considered a 'leaflet', which makes the qualification file the main document to consult in terms of learning outcomes. The ECS is not linked at the database level, though can be searched for in a separate database using the qualification's unique identifier. As the name suggests, it is a supplementary document and is mostly so in providing additional information, such as the legal context, the range of professions accessible to the qualification holder and to what extent the qualification provides access to further education. There is reference to the profile of skills and competences, though it simply lists the learning outcomes included in the qualification in the same way as the 'occupation in short' leaflet.

Source: Country template Netherlands.

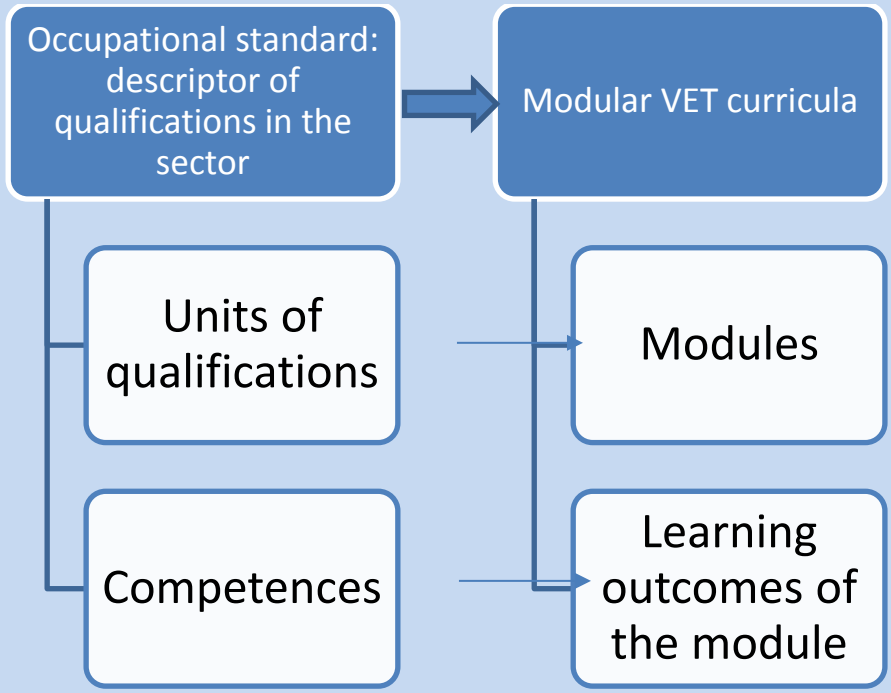
As noted above, in Lithuania, national VET curriculum documentation plays a role regarding learning outcomes in conjunction with occupational standards, and is described further in the box below.

Box 4. **The relationship between occupational standards and modular VET curricula in Lithuania**

Occupational standards in Lithuania are regulatory documents that specify the qualifications in an economic sector, their content and how they are acquired. They are unified across the country, aligned with national employers' and trade union organisations, approved by state authorities and used by all training and qualification assessment and recognition bodies. The occupational standard is one of the most important elements of the qualifications system, serving as a guide to the process of obtaining, evaluating and recognising qualifications. The main function of the standards is to specify the needs and requirements of the world of work - in terms of qualifications and their content and level. Currently, there are officially five approved occupational standards for the sectors of hotels and restaurants, machinery production and metalworking, production of electric and electronic equipment, public service, woodworking, furniture and paper production. A further 20 occupational standards are expected to be approved in the near future.

The **national modular VET curriculum** is a short, structured and consistent descriptor of the content of theoretical and practical vocational training that make up the whole and guarantee the acquisition of a specific qualification. The starting point for each modular curriculum is a qualification based on competences, defined by the occupational standard and referenced to a level of the Lithuanian NQF. The main unit of the modular curriculum, the module, is based on the unit of qualification defined in the descriptors of qualifications. Units of qualifications describe the competences

needed for the execution of the core work process in a given occupation which correspond to the learning outcome in modules, as shown in the diagram below.



Source: Country template Lithuania.

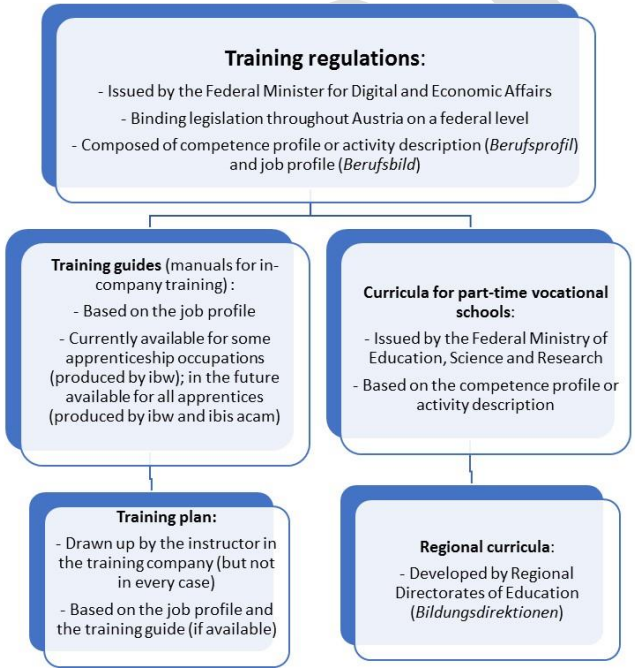
Another feature of documents in some countries is that some documents are more generalised and provide an 'umbrella' in the form of an over-arching framework of learning outcomes. The 'higher-level' documents typically provide more generalised learning outcome descriptions that inform more specific learning outcome descriptions at 'lower levels'. This aspect is explored in section 2.2.2.1 below.

Turning to the second aspect of the relationship between documents, documents spanning the governance dimension tend to reflect hierarchies within VET systems, in terms of the extent to which responsibilities are devolved. At local/regional levels, national qualification specifications are incorporated into programmes/curricula. Systems vary in the degree of autonomy, in respect of learning outcomes at the local/regional level programme level and also the extent to which that autonomy is exercised:

(a) **In Austria**, the main document is the training regulation. Based on the competence or activity profile presented in the training regulation framework curricula for part-time vocational schools for each apprenticeship trade are issued. The framework curricula are then further developed by the Regional Directorates of Education (*Bildungsdirektion*; formerly: *Landes- and Stadtschulräte*) into regional curricula. The so-called regional curricula are generally more detailed than the rather general framework curricula,

however, the framework curricula clearly specify the type and extent of changes that Regional Directorates of Education are allowed to make. In terms of learning outcomes descriptions, there is very little difference between what is described in the regional curricula and the framework curricula, although Regional Directorates of Education are able to attribute learning outcomes and teaching materials to individual school years/grades and further specify the level of performance within learning outcomes descriptions and attribute them to the different school years/grades (rarely done). Training guides for in-company trainers (manuals for in-company training; *Ausbildungsleitfäden*) are also produced for some apprenticeship occupations which include learning outcome-oriented formulations; and training plans (*Ausbildungsplan*) drawn up by instructors (most commonly in large rather than small companies). The relationships between the documents are shown diagrammatically below.

Figure 2. **Hierarchy of documents in Austria – apprenticeship training**



Source: Country template Austria.

(b) In **Finland**, three documents are key in VET and are related to one another as follows: 1. national vocational qualification requirements define, inter alia, the structure (composition) of a vocational qualification, its units or learning outcomes and criteria for assessment; local curricula, drawn up by VET providers, have to follow the national vocational requirements in terms of learning outcomes and assessment criteria, although they can include

learning outcomes which are necessary for local working-life, but which are not explicitly included in the national vocational requirements, along with optional modules from other qualifications (including study programmes of Universities of Applied Sciences); 2. individual personal competence development plans (PCDP) based on the national vocational requirements, 3. provider specific curricula and the needs of the student in question.

(c) In **Ireland**, the learning outcomes presented in the specifications available nationally are intended to be 'expected' learning outcomes which 16 local Education and Training Boards (ETBs) should use to devise 'minimum intended' learning outcomes at programme/curriculum level. As such, the learning outcomes in the published specifications are described by the national qualifications agency, QQI, as an 'approximation' of the learning outcomes actually contained within programmes/curricula. However, in practice, at programme level - the vast majority of learning outcomes are the same as in the national specifications, though it is intended that in future ETBs will take greater ownership of the learning outcomes with less 'copy-pasting' of learning outcomes, whilst having regard to national 'expected' learning outcomes.

(d) In **Spain**, Autonomous Communities play a role according to their areas of competence, though in practice this possibility to adapt learning outcomes is seldom exercised.

(e) In **UK-England** a large number of independent awarding organisations are responsible for learning outcomes specification within broad parameters set by the qualifications agency. In the main occupational areas, there is typically more than one awarding organisation providing a qualification, which means a competitive 'market' exists in qualifications. Providers have complete curriculum autonomy, but tend to follow the unitised structure of the qualification specification in determining curricula and assessment.

2.2.1.2 *Functions of qualifications documents*

Data on the content and profile of qualifications can be extracted from a variety of sources that serve different purposes. The intended learning outcomes described in these sources differ based on these purposes and a qualification's orientation. The functions of documents range from an orientation on informing on-the-job practice (such as occupational standards, which are normally set outside education and training systems, are firmly embedded in the labour market and can ideally serve as a link between education and training and the needs of the labour market) to an orientation on informing educational delivery (such as educational standards, curricula or learning programmes) or assessment. One

challenge for comparing VET qualifications and their learning outcomes is linked to the different functions of documents and instruments countries use to define and describe the intended learning outcomes (Cedefop, forthcoming; Bjørnåvold and Chakroun, 2017, p. 100).

It is difficult to classify the core documents relating to qualifications (those containing the learning outcomes descriptions) according to this classification of functions since frequently they contain a mix of information. The titles of the documents generally do not signal much about their function, using terms such as qualification, certificate/certification or, still less specifically, training regulation or executive order. In Bulgaria, the main reference document is called the State Educational Standard (*Държавен Образователен Стандарт)* and contains information that functions as occupational, educational and assessment standards. The clearest distinction in documentation between functions is made in Lithuania, where a new system is being introduced in which the core documents are called the Occupational standard *(Profesinis standartas)* and Modular VET curricula *(Modulinės profesinio mokymo programos)*.

The core qualification documents are likely to contain some information related to the occupational, educational and assessment standards functions; the amount of information will usually depend on the nature of devolution of responsibilities within a country. Information relating to the curriculum or learning programmes is least likely to be included in the core qualification documents, since this is most likely to be a responsibility devolved down the administrative hierarchy to local/regional and provider levels.

It is common for there to be overlap between documents so that learning outcomes and information relating to educational and assessment standards may be repeated and/or then further elaborated, and additional material added in. This is also often related to vertical administrative hierarchies and the way in which responsibilities are distributed across them. Denmark provides a good example of this, as shown in the box below.

Box 5. **Overlap between documents – example from Denmark**

In Denmark, the **Vocational Training Act** (*erhvervsuddannelsesloven*) is passed by Parliament, and it sets the overall framework for VET. More operational information concerning all VET programmes is given in a **general executive order** (*bekendtgørelse om erhvervsuddannelser*), and **individual executive orders** are elaborated for each programme, giving more detailed information (including learning outcomes) for these.

The social partners (employers' associations and trade unions) are in Trade Committees (*faglige udvalg*) that are responsible for drawing up **occupational profiles** for each trade/programme, containing the relevant learning outcomes. These profiles are adjusted every time it is deemed relevant by the trade Committees, and at

least once per year. On the basis of these profiles, the Ministry of Education draws up an **Executive Order** for each programme (*uddannelsesbekendtgørelse*). This is the official catalogue of learning outcomes.

On the basis of the relevant Executive Order, the Trade Committees then draw up a **Training Regulation** (*uddannelsesordning*) for each programme, which contains detailed information on how learning outcomes are achieved (subjects to be studied, duration, level etc.), It also contains detailed LOs for each subject (so called "*målpinde*"/yardsticks).

Finally, local Trade Committees (*lokale uddannelsesudvalg*), of which there is one for every vocational school, will draw up **a local training plan** (*lokal uddannelsesplan*), which contains adaptations to the training regulations to streamline this with specific local conditions. Local specifications do not affect the learning outcomes of the Executive Order, but only the levels at which individual subjects are taught during school periods, and which optional subjects are available.

The learning outcomes of the Executive Order are also embedded in a so called 'apprenticeship declaration' (*praktikerklæring*) for employers, where they have to indicate whether and at what level they have achieved these during the placement.

Source: Country template Denmark.

2.2.1.3 *Availability of documents*

National documents specifying the learning outcomes in VET qualifications are in most cases public regulatory documents, and therefore are (generally) publicly available and published online on the websites of the responsible organisations. An exception to this is the UK-England, where the qualification specifications are owned by the awarding organisations (AO), who design and validate them and issue certificates; the AOs are independent of government and have charitable status. Qualification specifications need to be purchased through the relevant AO. Although they are listed in the public part of the Register of Regulated Qualifications, maintained by the state standards monitoring body, they are not accessible from there.

Public accessibility to documents containing learning outcomes becomes less certain below national level, where VET providers or regional/local authorities may be responsible for formulating and maintaining curricula, and thus have some autonomy in respect of learning outcomes. For example, in Ireland, Education and Training Boards are responsible for designing VET programmes, and aim to adjust the learning outcomes whilst respecting national awards standards, but programme documentation needs to be accessed from the ETBs themselves. In Austria, there is no systematic recording of regional curricula: some Regional Directorates of Education publish their regional curricula (e.g. Upper Austria, Tyrol), while others do not. In Finland, all national qualification requirements are publicly available online on the eRequirements (*ePerusteet*) portal; VET providers are legally required to publish their curricula, and hence one can find their curricula

on providers' webpages – they may also use the eRequirements portal and currently there are a few different VET providers doing so in regard to the ICT service technician.

2.2.1.4 *Link to Europass Certificate Supplements*

It is common for there to be a close link between national reference documents and Europass Certificate Supplements. In Austria, the competence or activity description (*Berufsprofil*) of the training regulation is used for the section on 'profile of skills and competence' in the ECS. In France, the National Catalogue of Qualifications contains a 'fiche' for each qualification officially registered and this forms the ECS. In Ireland, the ECS is derived directly from the relevant award standard (though for the ICT qualification in question, the ECS does not display all relevant information).

In contrast, in Bulgaria there are not yet any ECS for vocational qualifications, and in Lithuania so far the link between the profiles of qualifications from the occupational standards and the ECSs is not yet established although the content of ECSs is based on the current descriptors of the VET standards and /or national modular VET curricula.

Practice varies in terms of how regularly information in ECSs is updated. In the Netherlands, any qualification changes are incorporated into the relevant ECS every three months, whereas in Denmark the ECS is not automatically revised every time there is a new Executive Order.

2.2.1.5 *Coherence across IVET qualifications*

Regarding the question of whether all IVET qualifications are covered in the same way by the reference documents, **in most countries there is no variation between types of qualifications**. Variations were found in three countries:

(a) In **Austria**, the approaches in relation to learning outcomes are in general completely different between school-based IVET qualifications and apprenticeship qualifications (which are also IVET). This applies in particular to training regulations, both in terms of structure and the process of their development. However, the curricula of part-time vocational schools are written in a similar manner as those offered in the school-based IVET, since the Ministry of Education is responsible for both types of curricula. Yet another approach is being pursued in the healthcare sector. The qualification profile 'Diploma nursing assistance (level 2)' is included in the Ordinance of the Federal Minister of Health and Women's Affairs on training and qualification profiles for care assistant professions (Care assistant profession

training ordinance – *Pflegeassistenzberufe-Ausbildungsverordnung / PA-PFA-AV*), which is based on the Health and Nursing Care Act 2016.

(b) In **Ireland**, apprenticeships involve different approaches/documentation to most IVET qualifications; and new ways of setting standards are being developed which have been applied initially in Early Learning and Care. Apprenticeships draw on 'Professional Award-Type Descriptors' (PATDs), that were originally developed for the Irish national framework of qualifications (NFQ) levels 7-9 (EQF levels 6-7) for recognition of professional awards, but were then extended to levels 5 and 6 to be used specifically for apprenticeships. The apprenticeship most similar to the qualification in question, is based on accredited modules from City & Guilds and lead to City & Guilds awards combined with vendor certification from CompTIA – rather than using the awards used in the qualification in question. In Early Learning and Care, the state agency responsible for qualification standards is currently phasing in changes to how standards are set as a model of a general approach. PATDs are used to set generic standards, but unlike apprenticeships and other IVET qualifications, these are accompanied with detailed guidelines on how they should be interpreted. There are also prescribed programme validation (programme approval) conditions, for example a minimum number of hours in practice placement.

(c) In the **UK-England**, there is no standard way of structuring or writing learning outcomes, which means there can be some important differences in the design of qualifications between awarding organisations – in terms of both how learning outcomes are structured and how they are written. With the ICT qualification in question, the qualification studied is provided by City & Guilds and a notable difference with the other major qualification in the field, provided by BTEC, is that the BTEC qualification can count as the knowledge element (Technical Certificate) of the related apprenticeship.

2.2.2 Features of learning outcomes descriptions

2.2.2.1 *Structuring the presentation of learning outcomes*

In terms of the degree to which learning outcomes are structured, the countries cover a broad range, from Denmark - where there is no systematic structuring and learning outcomes are not given in any specific order in the main reference document - through to highly structured systems, such as in the Netherlands (see box below):

Box 6. **'Nested' learning outcomes in the Netherlands**

Learning outcomes are described separately for the Base (B) and Profile (P) parts of the qualification file, the key learning outcomes reference document, and then grouped by their 'core tasks'. This is made visible through the use of 'identifiers' – a combination of letters and numbers. To illustrate:

For the ICT service technician, the Base part contains two core tasks (K), which carry the identifiers 'B1-K1' and 'B1-K2'. The first core task: *'Installs and maintains hardware, software and connections'* (i.e. 'B1-K1'), consists of three learning outcomes or 'work processes' (W), which are subsequently identified as 'B1-K1-W1' thru 'B1-K1-W3'. This structure is applied for all learning outcomes – resulting in a 'nested' structure of learning outcomes.

Each core task (K) is then described separately, in terms of:

- Complexity of the professional activities;
- Role(s), responsibilities and degree of autonomy expected of the qualification holder;
- The accumulated professional knowledge and skills (listed);
- Additional points of interest (such as cross-sectoral knowledge and skills required).

Additionally, the corresponding learning outcomes are described as work processes (W), including information on the expectations in terms of the qualification holder's professional behaviour, underlying competences and the results to be delivered. To illustrate:

The first learning outcome described (B1-K1-W1) is *'Preparing for use systems, peripherals and applications'*, and expects the delivered systems to work accordingly, as well as the development and upkeep of any related documentation and administration.

The description of learning outcomes included in the Profile part of the qualification file are structured in the same way as those in the Base part, though the number of profiles that is distinguished between may vary between qualifications. For the ICT service technician, for example, there are two profiles: ICT assistant and ICT management assistant. The former is the basis of the qualification, whereas the second profile includes an additional core task *(P2-K1; 'Offers support to users')*, which consists of three learning outcomes. These are described in the same structure as discussed for the Base part.

The only exceptions to this structure are the 'generic' learning outcomes, described separately in the Base part of the qualification file. These learning outcomes do not carry an 'identifier', as they are considered national learning outcomes (i.e. to be included in every VET qualification). For Dutch language and Mathematics, the required level is specified for qualifications at level 2 thru 4. The qualification standards for the exam components 'Career and Citizenship' are formed at national level and can be accessed separately from the database of national reference documents.

Source: Country template Netherlands.

In France, learning outcomes are presented in Employment, Activities, and Competences Standards which comprise three inter-related parts as described in the boy below:

Box 7. **How the presentation of learning outcomes is structured in France**

Employment, Activities, and Competences Standards or *référentiel emploi activités compétences du titre professionnel* (REAC) comprise three parts. The first part provides general pieces of information, and the next two parts provide details. The expected competences are general in the first part, and detailed in the third part.

The first part (**Typical Occupation Sheet, *fiche emploi type***) contains general pieces of information on:

- Place;
- List of activities (with an update from the previous version of the qualification if need be);
- Equivalencies with other qualifications;
- A synoptic view of the typical occupation (in the case of ICT Assistant, three typical activities that are described with eleven vocational competences; i.e. four, three, four);
- Transversal competences necessary in the occupation; and
- Level and/or domain of activity.

There is one Typical Occupation Sheet, even if the qualifications may lead to several occupations.

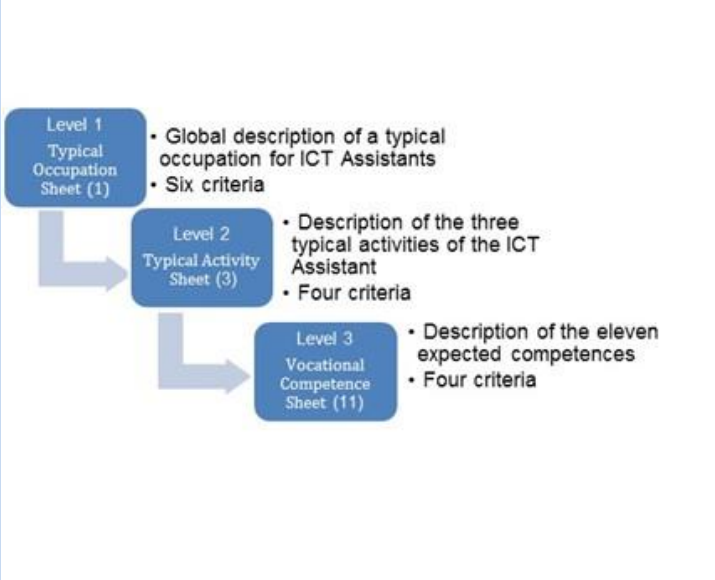
The second part (**Typical Activity Sheet, *Fiche activité type***) provides details regarding the previous part:

- Definition of the typical occupation, and of the conditions for practising;
- Regulations concerning the activities;
- List of typical activities, and of vocational competences; and
- Transversal competences necessary in the occupation.

There are three typical activities in the case of ICT Assistant, and four functions, broken down into seventeen sub functions.

The third part (**Vocational Competence Sheet, *Fiche compétence professionnelle***) details the expected competences. It includes all eleven vocational competences that are composed of:

- Definitions and description of the typical activity, and conditions for practicing;
- Occupational context for the implementation of competences;
- High quality performance criteria; and
- Technical, organisational, interpersonal know-how, as well as knowledge.



Source: Country template France.

It is quite common for **learning outcomes to be grouped into modules in a structured manner**. For instance, in Finland each qualification comprises vocational study units, core subjects and free choice modules. A similar approach is followed in Ireland (see box below). Such 'core plus options' approaches present challenges for international qualification comparison, as it requires decisions to made as to what is included in the comparative analysis, given that – at the level of the individual – there might be quite some variation in the actual content of the same qualification.

Box 8. **Use of modules to structure learning outcomes in Ireland**

In Ireland, learning outcomes are presented in the Certification Specification for the Major Award and also in Component Specifications for the Minor Awards which comprise the Major Award (QQI, 2014a). Computer Systems and Networks is a Major Award comprising 'a significant volume of learning outcomes' and is described in a 'Certificate Specification' (5M0536). In the Irish system, a Major Award is achieved by completing a given set of 'Minor' or 'Special Purpose' Awards. To obtain the Major Award certificate in Computer Systems and Networks, learners have to complete 3 core Minor Awards and then a range of optional Minor Award components. To achieve the Major Award learners have to achieve 120 credits by completing: the 3 core Minor Awards, which are occupation-specific, worth 45 credits; at least 15 credits from 3 options related to maths and computation (15 credits each, except 'Maths for STEM' worth 30 credits); at least 15 from Work Experience or Work Practice (15 credits each); at least 15 credits from 4 options related to transversal competences e.g. teamworking, communication (15 credits each); and at least 30 credits from 5 occupation-specific options worth 15 credits each e.g. mobile

technologies (unless 'Maths for STEM' has been taken under the maths and computation group of Minor Awards).

Source: Country template Ireland.

In Lithuania, each module of a qualification is related to a given 'competency', each of which comprises a number of learning outcomes.

Table 4. **Structuring learning outcomes in Lithuania: examples of modules/competences from the 'adjuster of computer equipment' qualification**

Title of the module	Competency	Learning outcomes
Computer network technologies	To design computer network.	To be aware of the principles and parameters of the functioning of computer networks and systems. To be aware of computer network technologies. To be able to analyse network technologies.
IT systems and products	To execute adjustment of IT systems	To be aware of the purpose, classification and functions of information systems. To be aware of the potential of testing software. To be aware of the possibilities of IT design. To be able to work with IT tools. To be able to use testing software. Ability to design IT systems

Source: Country template Lithuania.

In the UK-England, **assessment criteria** (AC) in the qualification specification are at a similar 'level' to the learning outcomes in Lithuanian qualifications, so that they are, in effect, specific 'sub- learning outcomes'. However, this does not necessarily mean there is a cognitive hierarchy - at least in the use of verbs between the learning outcomes and dependent ACs (see box below).

Box 9. **Assessment criteria functioning as learning outcomes in UK-England**

In the UK-England, learning outcomes are presented in the Certification Specification for the *Unit 335 IT Consulting Skills:*

First, there is a **unit aim** (UA) which could arguably serve as an overall learning outcome: 'to enable learners to identify the key characteristics, techniques and methods associated with successful consulting skills. To practice those skills whilst engaging with internal and external clients at all stages of the consulting cycle in order to bring about change within an organisation.'

Then Learning Outcome 2. 'The learner will structure a client assignment and specify the key management deliverables involved.'

Assessment Criteria are framed as 'the learner can …'

AC1: 'Use a consulting cycle to structure a client assignment to deliver an IT system.'

AC2: 'Produce and agree a statement of work for the assignment – business outcomes, products to be delivered.'

AC3: 'Construct an outline plan of work for the assignment.'

AC4: 'Negotiate, agree and document quality acceptance criteria for the conduct of the assignment.'

AC5: 'Identify key stakeholders and engage with them'.

Source: Country template UK-England.

Where modular approaches are not used, **alternative methods of structuring learning outcomes** pertain, such as in Austria (see box below).

Box 10. **Structuring learning outcomes in Austrian apprenticeship qualifications**

In the Training Regulation, which governs the employer's side of apprenticeships in Austria, two features are noteworthy:

(a) The competence profile or activity description lists the necessary professional competence ('*berufliche Handlungskompetenz*') apprentices should have acquired by the end of their training, usually in an unstructured way. In modularised apprenticeships it is structured according to modules. The Information Technology apprenticeship, which is not a modularised one, has been set up as an apprenticeship with the following two specialisations: 1. systems engineering, 2. industrial engineering. These specialisations form the structuring element of the competence profile / activity description.

(b) Learning outcomes statements in the job profile are presented in table format, with four columns that represent the four years of the training programme, and are further structured according to thematic areas (*Berufsbildpositionen*) ([24]). The job profile of the Information Technology apprenticeship is structured based on the following thematic areas: The general part (for both specialisations) includes: training company, business basics, technical basics, interdisciplinary training (key

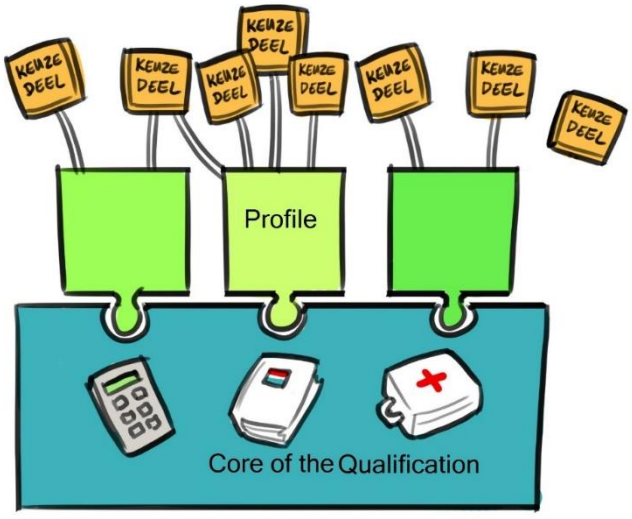
([24]) In modularised apprenticeships they are structured into modules.

qualifications), user terminal devices and peripherals, networks, quality and project management, customers. The specialisation 'systems engineering' is structured in the following three thematic areas: User terminal devices and peripherals, networks, and customers; the specialisation 'industrial engineering' includes the following three thematic areas: 'User terminal devices and peripherals, solution building (IT solutions), applications'.

Source: Country template Austria.

Another approach is for qualifications to be structured so as to reflect **groups of related occupations, with progressive specialisation** 'beneath'. In Finland, for example, vocational qualifications are conceived quite broadly with the possibility for specialisation, e.g. in the 'Vocational Qualification in Social and Health Care (OPH-2629-2017)'. Students may specialise in one of eight 'competence areas': Care for the Disabled, Nursing and Care, Podiatric Care, Mental Health and Substance Abuse Work, Care and Rehabilitation for Elderly People, Basic Life Support (BLS), Oral Health Care, Children's and Youth Education and Care. In the Netherlands, one **qualification file** can include more than one qualification - while sharing some learning outcomes, different qualifications within one file have different learning outcomes in their profile parts, and may differ in the optional parts that are accessible to learners (and thus a qualification file can correspond to several, similar occupations). The figure below illustrates this structure:

Figure 3. **Structure and number of VET qualification files in the Dutch education system**



The Dutch (VET) education system consists of a total of:

- 1000 optional or elective courses
- 479 qualifications
- 176 qualification files

Source: Country template Netherlands.

This type of approach may be linked to VET programmes/apprenticeships that are structured so as to give learners broad experience in an occupational area initially, before they specialise. In Denmark this approach is also tied into devolution of responsibilities to social partners at local level. VET learners start with an introductory period (1/2 to 1 year), with very broad learning outcomes which often cover several related trades. After this, they start the main programme where learning outcomes are tuned to a particular trade. In the Executive Orders so called 'final competences' (*slutkompetencer*) are given, i.e. tasks a learner must be able to carry out at the completion of the programme. They are consequently broad, leaving space for local trade committees to tune these to local conditions, if needed.

Only in some cases are learning outcomes structured according to the **domains of learning of knowledge, skills and autonomy/responsibility**. Learning domains refer to the horizontal dimension of learning outcomes statements, and are used to clarify the object and the scope of the intended learning. The EQF level descriptors use the following domains: knowledge, skills and autonomy/responsibility. In general, the following learning domains can be distinguished:

(a) Cognitive (knowledge) – What will students know?
(b) Psychomotor (skills) – What will students be able to do?
(c) Affective (attitudes, values or habits of mind) – What will students value or care about?

Within the ten countries examined in this study, domains are not often specified separately, as in the UK-England example above. In the Netherlands, professional knowledge and skills are separately mentioned under one heading (*vakkennis en vaardigheden*) for each core-task. In a similar manner, Responsibility and autonomy are mentioned (*Verantwoordelijkheid en zelfstandigheid*). These, however, do not form the principal structure for organising the learning outcomes.

In contrast, in the Lithuanian example above, each module/competency comprises two types of relevant learning outcomes (cognitive, representing knowledge and psychomotor, representing practical skills) although they are not put into separate fields as such so they would have to be read and interpreted if they were to be placed into these categories (e.g. verbs like 'to be aware' are used rather than 'have knowledge of'). In Bulgaria, VET standards usually list first the units of learning outcomes (modules) included in the standard. For each module the learning outcomes are specified in the form of a table in which knowledge, skills and competences are differentiated. Other information is also included, i.e. assessment methods (e.g. solving a test/case, performing a

practical task), examination conditions (e.g. classroom, real work place or training room) and assessment criteria related to the examination conditions for the unit (module) as a whole (see figure below).

Figure 4. **Structuring of learning outcomes into domains of learning in Bulgaria**

Name of the Units of learning outcomes	
NQF level:	
EQF level:	
Name of the profession:	
NQF level:	
EQF level:	
Learning outcome 1:	
Knowledge	
Skills	
Competences	
Learning outcome 2:	
Knowledge	
Skills	
Competences	
Assessment methods:	
Exam conditions:	
Assessment criteria:	

Source: Country template Bulgaria.

In Ireland, there is something of a hybrid model between no differentiation of learning domains and explicit classification: classification occurs at the level of the over-arching 'Major Award' Certificate Specification, with variable use of the same classification at the level of the component 'Minor Awards'. To elaborate, the overarching Certification Specification for a Major Award provides a summary description of the learning outcomes to a common format. For the qualification 'Computer Systems and Networks 5M0536 Certificate Specification NFQ Level 5', the corresponding summary learning outcomes are presented in the table below.

Table 5. **Common format for the summary description of the learning outcomes in Ireland**

Strand	Sub-strand	Nature of learning	'Computer Systems and Networks 5M0536 Certificate Specification NFQ Level 5'
			Learners will be able to:
Knowledge	Breadth	Broad range of knowledge	Demonstrate an understanding of key concepts, terminology and functionality associated with computer hardware, operating systems, networks and peripherals and their configuration.
	Kind	Some theoretical concepts and abstract thinking, with significant depth in some areas.	Demonstrate an appreciation of the roles and functions associated with the management and administration of computer systems and networks.
Know How & Skill	Range	Demonstrate a broad range of specialised skills and tools	Utilise methods and tools to contribute to the implementation, configuration and maintenance of computer systems and networks under supervision and as a member of a team.
	Selectivity	Evaluate and use information to plan and develop investigative strategies and to determine solutions to varied unfamiliar problems	Utilise standard computer systems and networks skills as required in the workplace, while taking responsibility for the nature and quality of their own output.
Competence	Context	Act in a range of varied and specific contexts, taking responsibility for the nature and quality of outputs; identify and apply skill and knowledge to a wide variety of contexts	Exercise some initiative both as an individual and as a team member in solving defined problems arising in computer systems and networking environments.
	Role	Exercise some initiative and independence in carrying out defined activities; join and function within multiple, complex and heterogeneous groups	Learn to take responsibility for their learning within a managed environment. Demonstrate a capacity to think laterally and strategically in a computer systems and networking environment.
	Learning to Learn	Learn to take responsibility for own learning within a managed environment	Relate their comprehension of computer systems and networks to society and potential career opportunities.
	Insight	Assume full responsibility for consistency of self-understanding and behaviour	Assume full responsibility for consistency of self-understanding and behavior.

Source: Country template Ireland (Extract from 'Determinations for the Outline National Framework of Qualifications': NQAI; Computer Systems and Networks 5M0536 Certificate Specification NFQ Level 5).).

At the level of the component Minor Awards, the same three-fold division of (i) knowledge, (ii) know-how and skill, and (iii) competence is only sometimes used. Where it is not used, the learning outcomes are simply listed. For the qualification in question, two of the three mandatory Minor Awards use the structured approach, one does not, the reason for this probably being that Minor Awards do not need to have all types of learning outcomes.

In terms of the extent to which learning outcomes are **structured in terms of different levels of specificity**, as already indicated above, it is not uncommon for learning outcomes to be structured so that they are presented at a more general level which provides an over-arching structure for the listing of more specific learning outcomes 'beneath'. Examples of this exist in Ireland, Lithuania and Bulgaria, as already mentioned, and also in the Netherlands. In the latter country, a 'qualification file' provides an overview of all the core tasks, profile parts and their corresponding work processes (learning outcomes) for a qualification in a set of tables, as presented below.

Table 6. **Overview of tables presenting learning outcomes in qualification files in the Netherlands (translated)**

Profile name	MBO-level (EQF level)	Legal professional requirements (Y/N)	Qualification type	Length of education to obtain qualification (hours)
P1 ICT Management Assistant	3	No	Vocational Training	4800
B1-K1 Installs and maintains hardware, software and connections		B1-K1-W1 Preparing for use systems, peripherals and applications B1-K1-W2 Replacement, repair and/or (dis)assembly of systems and peripherals B1-K1-W3 Realisation of connections		
B1-K2 Treats incident reports		B1-K2-W1 Working and registering of incidents B1-K2-W2 Resolves and/or escalates incidents		
Profile part				
P1 ICT Assistant				
No extra core tasks or work processes				
P2 ICT Management Assistant				
P2-K1 Offers support to users		P2-K1-W1 Drafting of instructions P2-K1-W2 Providing the users with an oral presentation P2-K1-W3 Performing after services		

Source: Country template Netherlands (Qualification file for ICT service technician - CREBO 25191).

In Ireland, the relevant ICT Major Award contains just seven learning outcomes (see table above) which are expressed in a general way, whilst much more detailed learning outcomes are contained in the constituent Minor Awards. These learning outcomes sit within a wider hierarchy, which can be represented as follows:

Table 7. **Summary learning outcomes for the 'Computer Systems and Networks 5M0536 Certificate Specification NFQ Level 5'**

Scope	QQI expected learning outcomes	Programme: intended learning outcomes	Assessment task designer
Macro	NFQ grid of level indicators		
Macro	NFQ award-type descriptors		
Macro-meso	Broad standards contained in Certificate Specifications of Major Awards		
Meso	Narrow standards contained in Component Specifications of Minor Awards		
Meso		Minimum Intended Programme Learning Outcomes (MIPLOs)	
Meso-micro		and Minimum Intended Module Learning Outcomes (MIMLOs)	
Micro			Assessable LOs.

Source: Country template Ireland (provided by QQI for completion of template).

In some countries, more detailed learning outcomes might occur within what are termed 'assessment criteria'. This is the case in the UK-England and Finland, for example ([25]).

2.2.2.2 *Components of learning outcomes statements*

'Components of learning outcomes statements' refers to the syntactic (and/or conceptual) structuring of the text on learning outcomes: When writing a learning outcomes statement, the Cedefop handbook (Cedefop, 2017, p. 47) suggests to

([25]) In Finland, assessment criteria are clearly specified in terms of learning outcomes. They are specified at five 'levels': Satisfactory 1 and 2, Good 3 and 4, Excellent 5. However, this five-level structure is not used in every qualification, probably because for a certain time only three 'levels' were distinguished (e.g. https://eperusteet.opintopolku.fi/#/en/esitys/3689879/reformi/tutkinnonosat/3708462).

focus on the learner and start with an action verb, followed by the object of the verb as well as a statement specifying the depth/breadth of learning to be demonstrated, and complete with an indication of the context (which can be related to learning, work or other relevant social contexts). These components (action verb/object/context) are considered as forming the basic structure of learning outcomes statements.

In the countries studied, **verbs and objects are common components of individual learning outcomes statements, although it is not uncommon for nouns to be used instead of verbs,** as noted in Austria. Context is only sometimes given in individual learning outcomes statements. Bulgaria provides a fairly typical example. Here, learning outcomes statements are most often composed of action verb and object, sometimes there is an indication of the context. For example, in relation to the ICT unit/module the learning outcome 'use of different information and communication technologies' is broken down as follows:

(a) knowledge: 'know the most frequently used ICT within the work process'; 'know the special terminology used for working in the ICT field';
(b) skills: 'use programme products and specialised software'; 'prepare, register and archive documents in an electronic format';
(c) competences: 'work independently with appropriately selected ICT in accordance with the concrete work activity' (VET Standard for application software developer).

It seems **most common for context to be handled by being included in over-arching statements, rather than individual learning outcomes statements.** For example, in Ireland, under 'Purpose' in the Major Award it states: 'The purpose of this award is to equip the learner with the fundamentals to install, configure and maintain computer systems and basic networks under supervision leading to employment in a range of sectors. It also facilitates progression in education including to further and higher, education or training.'

This finding hints that it might be quite efficient to include context statements in this way, since it avoids repetition at the level of individual learning outcomes. However, this would need to be taken into account in comparative analysis of qualifications using digital tools.

2.2.2.3 *Vertical dimension*

The vertical dimension of learning outcomes refers to the expression of the increasing complexity of learning. 'Action verbs play a key role in defining and articulating this vertical dimension but need to be supported in this by clarifying the object of learning and the occupational and/or social context in which the

learning takes place and where the outcomes are to be used' (Cedefop, 2017, p. 48).

In most countries, it seems that it is the object and context of a learning outcome statement that most expresses the vertical dimension, rather than the verbs used. A notable exception is in the Austrian apprenticeship system where learning outcomes are spread across the four years of the apprenticeship (see box below). It is also in Austria where the Common European Reference Framework for languages is used and which has an inherent vertical dimension.

Box 11. **Expression of learning complexity in the Austrian apprenticeship system**

Training regulation component:

The job profile positions show the knowledge and skills to be imparted, broken down by apprenticeship year. Some of the learning outcomes or competences described in the job profile are not allocated to a single apprenticeship year, but cover the entire training period – they are 'spread' over all four columns (i.e. four years). Others, however, are divided into apprenticeship years and describe an increasing level of proficiency over the years. For instance, 'Knowledge of the importance of documentation of all work performed and tests carried out in accordance with company-specific quality management' becomes 'Document all work performed and tests carried out in accordance with company-specific quality management.

The increasing level of ambition is expressed in different ways for example: ([26])

(a) increasing autonomy in carrying out an activity: e.g. '2nd & 3rd year: 'Involvement in the design and planning of data storage systems'; 4th year: 'Design and planning of data storage systems';
(b) increasing specification of knowledge: basic knowledge at lower levels;
(c) increasing use of knowledge: e.g. 1st year: 'Knowledge of offered cloud services, their evaluation, and possible integration into existing networks', 3rd year: 'Participate in the evaluation, selection and integration of cloud services into existing networks';
(d) increasing complexity of activities: Solution Building (IT-Solutions): 2nd - 3rd year: 'Working on projects (creating your own time and resource planning, accepting partial orders, presenting solutions, comparing the project status with other team members, creating partial project reports)' '; 4th year: 'Carry out projects (create time and resource planning, place partial orders, present solutions, compare project status with other team members, create project reports)'.

Curricula for part-time vocational schools component:

Different levels of complexity are expressed mainly by the use of specific action verbs. For example, in some cases the verbs 'describe' is used, in others 'distinguish'

([26]) Auzinger & Luomi-Messerer, 2017, p. 34.

and in others ‘select’ and ‘argue’. And other statements use verbs which actions and tasks referring to an even higher degree of complexity - e.g. 'build and configure networks with current network protocols and detect and correct errors'. However, the increasing of the complexity is not systematically expressed in the curriculum (as it is the case in the training regulation).

Source: Country template Austria.

In most countries, there is likely to be an implicit assumption that learning outcomes statements take into account where the qualification of which they are part sits on the NQF and/or EQF and thus should reflect the types of terms used in framework descriptors. This is explicit in Ireland, where awards standards are expected to be consistent with the NFQ (National Framework of Qualifications) levels: the NFQ has 10 levels with a progression in the complexity of learning from level 1 to level 10. Similarly, in Austria, the formulation of learning outcomes in new training regulations is based on EQF/NQF descriptors. And in Lithuania, the descriptors of the NQF include some guidance words that might be used to formulate competences. But the concept of complexity is perhaps treated most comprehensively in the Netherlands, where the vertical dimension is described at the level of the core tasks, in a sub-section on complexity in the qualification file. This section includes information on (1) the relevant contexts in which the qualification holder works, (2) their role within the organisational structure and (3) which factors determine and/or affect the level of complexity of their profession. For the ICT service technician (ICT Managing Assistant), for example, specific mention is made of the fast-paced nature of technological developments within ICT technology – leading to increased ‘ageing’ of their knowledge on base processes, materials, tools and terminology, which requires qualification holders to quickly adapt and learn to use products and tools, in order to keep their knowledge and competences up-to-date.

Cedefop (2017, pp. 33-36) notes that the choice of action verbs for expressing different complexity of learning, might refer to the taxonomies developed by Bloom and colleagues or the SOLO taxonomy. These taxonomies classify learning in terms of the complexity of skills and knowledge to be acquired. However, **the use of taxonomies to reflect the vertical dimension was found to be rare in our sample. Also rare was the provision of guidance regarding the vertical dimension.** A notable exception is Austria, where in curricula for part-time vocational schools, the foundation of the educational standards is a two-dimensional competence model which represents the different levels of action of the competences to be developed as well as the central content; for the level of action, different complex dimensions are given following the taxonomy of Bloom. Also, the manual for the educational standards in VET curricula for part-time vocational schools provides guidelines for writing

learning outcomes. The action dimension includes five elements: reproduce, understand, apply, analyse and develop. These five stages describe different processes, some of which are dependent on the underlying characteristics. The manual includes some examples of verbs for expressing each of the five stages. In Spain, there is no national taxonomy, although the taxonomies of Bloom and SOLO influenced the initial process of defining diplomas and certificates. However, this influence was very generic, for example using infinitive verbs, and neither taxonomy was used systematically. In Denmark, in the Act on Vocational Training, four levels of competence are described - beginner, routine, advanced and expert level – which is inspired by the model of skill acquisition put forward by Dreyfuss and Dreyfuss. These are used to grade the level of complexity of individual subjects in a given programme.

It is important to note that **not only were taxonomies not much in evidence but reservations were sometimes expressed about using them**. In Lithuania, the methodical guidelines for the design of modular VET curricula do not indicate or recommend the application of particular taxonomies in order to leave the freedom of choice to experts. In Ireland, one interviewee commented: 'We are cautious about taxonomies but aware of their popularity. We stress that people need to try to express their ideas about the intended learning outcomes in ways that will be understood by users of the statements.' It was stressed that the Irish NQF provides a broad framework and that 'currently we avoid suggesting lists of action verbs. We don't think single words fully capture the progression in complexity.'

2.2.2.4 *Coherence of learning outcomes descriptions*

In each of the ten countries, it was found that **learning outcomes are described in a similar manner with no differences between, for instance, occupational and transversal learning outcomes**. It should be noted, though, that the treatment of transversal learning outcomes is highly variable, and in some countries, they are not fully expressed or well written into learning outcomes. In Spain, for example, it was noted that transversal learning outcomes are hardly described and when they are, the descriptions are very generalised. It should also be noted that there can be great variation between the way learning outcomes are described between qualifications and also between modules within the same qualification, reflecting different authorships, as in Ireland.

2.2.2.5 *Language*

Regarding the question of the availability of learning outcomes in English - in addition to the native language - none of the main documents containing learning

outcomes are systematically available, although ECS in English are available in Austria, Denmark, Lithuania, the Netherlands, Spain, and Finland. Only in Finland are a limited number of qualifications available in English, with plans to increase the number ([27]); and in France some fiches are available in English, as well as German and Spanish.

2.2.2.6 *Coherence across qualifications*

Overall, the learning outcomes of IVET qualifications tend to be described in the same way, without different approaches being applied to different (types of) IVET qualification, although there are general exceptions to this pattern as already described in section 2.2.1.4. In Denmark, this is due to all Executive Orders being drafted according to the same template, issued by the Ministry of Education.

2.2.3 Infrastructure behind the description and storage of IVET qualification information

2.2.3.1 *Storage of the learning outcomes descriptions of IVET qualifications*

Learning outcomes descriptions are, in all countries, 'stored' in PDF versions of the documents described in section 2.2.1.1. The PDFs are accessed through databases or registries (for a detailed discussion of the structure and functions of databases/registries see section 2.3) on websites belonging to the organisations responsible for designing the qualifications and learning outcomes. Sometimes they are also published on the websites of other stakeholders, e.g. in Denmark Executive Orders and training regulations are available on the Ministry of Education website and also trade committee websites. VET providers may also publish their own curricula on their webpages, as in Finland. In Ireland, the learning outcomes presented in the specifications available nationally are intended to be 'expected' learning outcomes which local Education and Training Boards (ETBs) should use to devise 'minimum intended' learning outcomes at programme level. As such, the learning outcomes in the published specifications are described by the national qualifications agency, QQI, as an 'approximation' of the learning outcomes actually contained within programmes/curricula. However, in practice at programme level, the vast majority of learning outcomes are the same as in the national specifications, though it is intended that in future ETBs will take greater ownership of the learning outcomes with less 'copy-pasting' of

([27]) In Finland, all formal qualifications, including VET qualifications, are available in Swedish and some of them are also available in the Saame and Roman languages.

learning outcomes, whilst having regard to national 'expected' learning outcomes. Currently, ETB programme documents need to be obtained from the ETBs themselves and the 'minimum intended' learning outcomes are not available from the national qualifications agency, QQI, but it is hoped to include in the in the newly launched Irish Register of Qualifications in the future ([28]). In the UK-England, the qualification specifications are owned by awarding organisations and need to be purchased through the relevant awarding organisation; they are listed in the public part of the Register of Regulated Qualifications but are not accessible from there.

In the Netherlands, the database includes PDF documents, but it is also intended for purposes of analysis, as described in the box below.

Box 12. **Learning outcomes infrastructure in the Netherlands**

The Foundation for Cooperation on Vocational Education, Training and Labour Market (SBB) is responsible for developing new qualifications and enables access to learning outcomes in two ways at national level:

1. An online 'registry' of national reference documents – further referred to as the **'SBB Registry'** – which is accessible to the public through a webservice that allows them to search for specific qualifications. Each qualification has its own 'landing page', which provides some basic information (a short description, the CREBO number, etc.) and the download links to the corresponding reference documents (qualification file, job description) in PDF format. It is important to note that the landing page itself does not provide an overview of the learning outcomes included in the qualification, only a list of optional curriculum elements students can opt for which have been considered applicable to this qualification ([29]).

2. One XML-based database – further referred to as the **'XML database'** that can be accessed ([30]) by students, teachers, organisations or others for acquiring data for analysis (usually for studies concerning education, labour market and/or mapping exercises). To access the XML database the user first needs to acquire an access code (which can be requested at SBB), and needs to set up a so-called 'client' ([31]) on their device (laptop, PC) to 'call' the service end of the registry. This will generate a set of files that are to be used as the base coding script to build on, using whichever 'client application' the user would prefer (i.e. one that can read and execute XML-

([28]) The new register is in development and currently available to view at irq.ie.

([29]) See for example the landing page for ICT technician at https://kwalificaties.s-bb.nl/Details/Index/2571?type=kwalificatie&item_id=957853&returnUrl=%2F%3FResultaatType%3DAlles%26AardKeuzedeel%3D%26SBU%3D%26Niveau%3D%26Wettelijkeberoepsvereisten%3D%26Cohort%3D%26Schooljaar%3D%26Certificaat%3D%26Trefwoorden%3D25191

([30]) The XML database can only be accessed with an access code provided by SBB.

([31]) Which is downloadable from a link in the documentation on the service: https://kwalificaties.s-bb.nl/services/ZoekRegister.svc

based scripts). Depending on the requests sent (through code), the retrievable information from the registry includes:

- Lists of changed files (qualification files, curriculum elements and certificates) for a given period;
- Whether a qualification is still valid as of date;
- The XML or PDF documents of a specific qualification file or curriculum element;
- Information on the overlap between curriculum elements or qualifications (if available);
- The XML or PDF file of a qualification's certificate / diploma.

Source: Country template Netherlands.

In terms of **processes for storing learning outcomes**, these tend to be part of the usual processes of designing/updating qualifications: when a qualification description is agreed, it is normally just published online, normally in a registry. Similarly, when learning outcomes are updated, the revised version of the specifications/standards is simply published, without any special arrangements. There can be a time lag in this process with delays in publishing updated specifications. For example, in Austria, it is intended that training regulations will be published on the Ministry website as soon as they are decreed. However, there is no systematic process to ensure 'automatic' publication here and sometimes a training regulation is missing or not published. Training regulations are automatically published in the Legal Information System, RIS (Legal Information System of the Republic of Austria – *Rechtsinformationssystem der Republik Österreich*) but this is a huge database that contains the entire documentation of legislation in force in Austria and it would probably not be suitable for automated gathering of information on qualifications.

Only the Netherlands appears to update documentation on a periodic rather than 'as necessary' basis. The process followed is that when any changes to existing qualifications or newly developed ones are confirmed, the related documents are added to the SBB registry roughly every three months. Changes to the XML database are made separately as they need to be sent in to Kennisnet for conversion (Kennisnet is a public organisation – Foundation - that provides and maintains a national ICT structure for the educational sector). Interestingly, after linkage to the ESCO-portal is finished, the goal is to have any new qualifications included into the database directly (i.e. without waiting for the next 'update' to be exercised).

2.2.3.2 *Format of stored learning outcomes descriptions of IVET qualifications*

As noted, learning outcomes are often 'stored' in PDF documents. Such documents can only be accessed and downloaded individually, providing no opportunity for any automated comparison. Learning outcomes are not available outside these documents, in terms of being in a database or Excel spreadsheet (or similar) that would facilitate the searching of individual learning outcomes. Often, the websites on which the PDF documents are published have a search function, as in Denmark where the Ministry database allows the public to search for programmes and qualifications past and present (covered in more detail in the section below on national qualifications databases).

As shown in section 2.2.1.1, in order to obtain a full understanding of learning outcomes, an array of information on their context as well as the descriptions themselves would need to be consulted. The two principle documents in this respect, are the documents containing the learning outcomes and the corresponding ECS, but there is much variation across the countries regarding which documents have to be consulted for which pieces of information, making for a complicated picture in respect of the automation of international comparative analysis.

2.2.3.3 *Coherence across IVET qualifications*

In line with the finding above – that in almost all cases countries treat all (types of) IVET qualifications – the same was found in respect of learning outcomes: no major differences were found, in terms of the infrastructure behind the description and storage of qualification information. The most salient exceptions to this include Ireland, where outside the Common Awards System - which covers the vast majority of IVET qualifications - information is stored in different ways, and must be sourced from individual providers.

2.3 National qualifications databases

This section looks at national qualifications databases that present qualifications and their learning outcomes, and discusses some important features that are considered necessary to use for comparing national qualifications and for automated text processing. Particular attention will be paid to examining the presentation of data on qualifications in these databases, the sources used for their development, whether they are based on a common metadata scheme and whether they correspond to a common format for presenting learning outcomes.

2.3.1 Availability of a national qualifications database that includes learning outcomes descriptions of IVET qualifications in the countries analysed

The starting point for identifying national databases is the regularly updated overview on national databases prepared for the EQF AG and the NQF Inventory developed by Cedefop. According to the 2018 update of the NQF Inventory, '24 countries have included levels in their national qualifications databases' (Cedefop, 2019a, p. 3): AT, BE (FL, FR), CZ, DK, EE, FR, DE, EL, HU, IE, XK, LV, LT, MT, ME, NL, MK, PL, PO, RO, SI, SK, TR, and the UK. However, these databases do not necessarily include detailed information on levelled qualifications or types of qualifications and their learning outcomes. Sometimes, there is rather a register that merely lists qualifications allocated to NQF levels, such as in Malta ([32]) or Switzerland ([33]).

Nevertheless, in the past two years, important progress has been made in establishing qualifications databases. They exist and are available in most countries in a mix between 'pre-NQF databases', which sometimes focus on programmes rather than qualifications, and databases reflecting the NQFs (Pevec Grm & Bjornavold, 2019). Furthermore, as the results of this study show, the databases also differ in their scope and the information provided and in particular whether they contain learning outcomes (full description, short summary or not at all). The table below provides an overview of the databases available in the ten countries covered in this study.

It should be noted that **a clear definition of the concept of a 'qualifications database' has not yet been developed in the EQF context, and that there are therefore different perceptions of what can be described as such in different countries**. For example, in the most general sense, a database is an organised collection of data. Thus, any website that simply lists qualifications available in a country which can be downloaded (as a PDF-file) when clicking on the title of the qualification, could already be considered to be a qualification database. The Merriam-Webster dictionary additionally refers to a search and access function and defines a database as 'a usually large collection of data organised especially for rapid search and retrieval (as by a computer)' ([34]).

([32]) http://ncfhe.gov.mt/en/register/Pages/register.aspx

([33]) https://www.admin.ch/opc/de/classified-compilation/20151046/index.html

([34]) https://www.merriam-webster.com/dictionary/database

Table 8. **Overview of qualifications databases or registers (including IVET qualifications) in the countries studied**

Country	Database	Responsible body	Short description	All individual IVET qualifications incl.	Learning outcomes of IVET qualifications displayed	NQF/ EQF level incl.	Relationship with reference docs. described	Link to LOQ and /or ESCO
Bulgaria	Portal on education and training opportunities in secondary and tertiary education and lifelong learning in Bulgaria ([35])	Human Resource Develop-ment Centre	It supports career counsellors, learners, teachers and parents to quickly find information about VET as well as HE providers and the education and training opportunities they offer.	Yes	No	No	Loose	No
Denmark	Education Guide (*UddannelsesGuiden)* ([36])	Danish Ministry of Education	It presents information about educational programmes, adult education and continuing training, as well as job and labour market conditions in Denmark. Its aim is to supply precise and updated information for everyone in Denmark seeking job, education, business or labour market information.	Yes (programmes)	No	Yes	Close	No
Spain	All VET (*TodoFP*) ([37])	Ministry of Education and Vocational Training	The database serves as guidance tool for students, teachers and companies and as a National Catalogue of VET programmes from the Education System.	Yes	Yes: short summary and full description (PDF)	No	Close	No

([35]) http://euroguidance.bg/Public/EducationInfo#/ProfessionalQualification/Search.

([36]) www.ug.dk

([37]) http://www.todofp.es

Country	Database	Responsible body	Short description	All individual IVET qualifications incl.	Learning outcomes of IVET qualifications displayed	NQF/ EQF level incl.	Relationship with reference docs. described	Link to LOQ and /or ESCO
France	National register of vocational and professional qualifications (*Repertoire national des certifications professionnelles, RNCP*) ([38])	France compétences ([39])	Its purpose is to provide individuals and companies with constantly updated information on professional diplomas and titles as well as on the certificates of qualification appearing on the lists drawn up by the national joint employment committees of the professional branches.	Yes	Yes: summary & comprehensive descriptions	Yes	Close	No
Ireland	QQI qualifications database ([40])	QQI – Quality and Qualifications Ireland	It was developed to provide public list of further education and training awards made by QQI. It contains the awards specifications of the CAS (Common Awards System), i.e. the Certification Specifications and Component Specifications.	Yes ([41])	Yes: short summary (PDF) & full descriptions (PDF)	Yes	Close	Database linked to LOQ
Lithuania	Qualifications database (AIKOS) ([42])	Ministry of Education, Science and Sport and the Centre of Information Technologies of Education	Its aim is to help citizens choose a marketable profession that can be acquired for the first time or through retraining at Lithuanian or European HE and vocational schools by collecting, processing and making easily accessible information to a wide range of users.	Yes	Yes: short summary	Yes	In development (close)	Database linked to LOQ and ESCO

([38]) http://www.rncp.cncp.gouv.fr/

([39]) *France compétence* is responsible since 2018. Formerly, the RNCP was placed under the responsibility of the National Commission of Qualification (*CNCP*, *Commission nationale de la certification professionnelle*).

([40]) http://qsearch.qqi.ie/WebPart/Search?searchtype=awards

([41]) excluding apprenticeships

([42]) https://www.aikos.smm.lt/en/Pages/Default.aspx

Country	Database	Responsible body	Short description	All individual IVET qualifications incl.	Learning outcomes of IVET qualifications displayed	NQF/ EQF level incl.	Relationship with reference docs. described	Link to LOQ and /or ESCO
Nether-lands	Vocational Education Qualifications (*Kwalificaties mbo*) ([43])	Foundation for Cooperation on Vocational Education, Training and Labour Market (SBB)	It has been developed for the national dissemination of information and documentation on vocational qualifications, in particular to provide insight (for students and employers) into the knowledge, skills and competences contained in a qualification.	Yes	Yes: full descriptions (PDF)	Yes	Close	No
Austria	NQF Register ([44])	NQF-National Co-ordination Point (NCP) in Austria	It supports the general objective of making all qualifications assigned to the NQF in Austria comparable and transparent. It is also intended to make a valuable contribution towards greater visibility of qualifications offered.	No: types & examples	Only for examples of IVET qualifications: short summary	Yes	Loose	
Finland ([45])	eRequiremets (*ePerusteet*) ([46])	Finnish National Agency for Education	It covers all formal qualifications from early childhood education to upper secondary education. National vocational qualification requirements are published and made available on this platform. The organisers of education and training also publish local curricula and execution plans.	Yes	Yes: short summary & full descriptions (PDF)	No	Close	No
UK-England	Register of regulated qualifications ([47])	Ofqual	It shows the qualifications and awarding organisations regulated by Ofqual ([48]) and CCEA Regulation ([49]). The site is shared by the two regulators.	Yes	No	Yes	Loose	No

([43]) http://kwalificaties.s-bb.nl

([44]) https://www.qualifikationsregister.at/public/home#

([45]) According to Cedefop (2019a, p. 3), Finland is not among the 24 countries that have included levels in their national qualifications database and also the country report for Finland states that FiNQF and EQF levels are currently not included in the qualifications databases. However, since the country report also mentions that 'information about qualifications, the qualification requirements, and their FiNQF and EQF levels is available on the e-perusteet/e-grunder website' (Cedefop, 2019b, p. 9), we are referring to this website here as 'qualifications database'. We consider it – at least – partly for the analysis here because it is an electronic platform that allows searching for qualifications and includes learning outcomes descriptions of qualifications.

([46]) https://eperusteet.opintopolku.fi/#/fi

Source: Authors – based on country templates.

([47]) https://register.ofqual.gov.uk/

([48]) Ofqual is the independent qualifications regulator for England.

([49]) The Council for Curriculum, Examinations and Assessment (CCEA) has responsibility for the regulation of qualifications in Northern Ireland.

The table above shows that some countries do not have a qualifications database available or not one that includes learning outcomes descriptions of qualifications:

(a) In **Bulgaria**, the database was developed by the Human Resource Development Centre and launched in 2017 ([50]). Data is collected through different registers maintained by the Ministry of Education and Science and the National Agency for Vocational Education and Training (NAVET). A procedure has been put in place enabling educational and training providers to update and supplement information on the database by themselves. However, the portal has not been updated since 2017 and it does not include learning outcomes descriptions.

(b) In **Denmark**, the only database that includes all officially recognised qualifications in Denmark – which means that the qualifications are part of the Danish NQF (GE, VET, HE, AE etc.) – is *UddannelsesGuiden,* but it does not include learning outcomes of qualifications ([51]). It is designed as a guidance tool to help people clarify educational choices, and contains short general descriptions of all IVET/CVET programmes, including the NQF level, and practical information (such as entry requirements, how and where to apply etc.). There are also links to further sources of information (including relevant executive orders for VET qualifications, i.e. the legal act forming the basis of the qualification and containing learning outcomes descriptions), and there is a function where one can ask questions to guidance counsellors. Guidance counsellors are, in fact, a secondary target group, and it is also possible to access a 'knowledge centre' for guidance counselling through the website. However, all training regulations with detailed information on learning outcomes can be found on another database ([52]). This is a database for VET-professionals for use in the ongoing administration of the VET-system. However, only a part of the database is accessible without a password. There are no plans to develop any other type of national qualifications database for the time being.

(c) The Register of Regulated Qualifications from **UK-England** includes individual qualifications, but not their learning outcomes. The database gives

([50]) http://euroguidance.bg/Public/EducationInfo#/ProfessionalQualification/Search

([51]) It is also referred to as a 'database' in the evaluation of the implementation of the Danish NQF (EVA, 2013, p. 9). Moreover, the responsible ministry considers ug.dk to fulfill all requirements to a qualification database for now, even though it does not directly include learning outcomes.

([52]) https://www.eud.uddannelsesadministration.dk/

the title, level and some other minimal information. It provides a link to the awarding body, but access to the qualification document is not available without payment and/or permission from the awarding body. All information at the appropriate level of detail, such as learning outcomes and unit aims, can only be found within the qualification documents themselves. Ofqual does not currently have plans to develop a new qualifications database and does not seem to have any particular current interest in matching their database with the various European policy tools.

Ireland is a specific case, since there is one qualifications database available and a new Irish Register of Qualifications (IRQ) in development. The new Register has had only a 'soft' launch and cannot yet be found on internet search engines. Currently it contains qualifications at Level 1-6 on the Irish Framework made by QQI. By the end of 2019 it will have data from all the national awarding bodies that have qualifications on the NFQ (including universities, which are national awarding bodies).

Since this study is about comparing VET qualifications and their learning outcomes, the analysis in this section focuses on those countries that have a qualifications database in place that directly presents IVET qualifications and their learning outcomes. Information related to other databases is added only where specific information is available.

According to the results of a recent PLA on national qualifications databases (European Commission et al.; 2019, p. 7), such databases serve a broad range of purposes: 'They are used to document the qualifications, which have been assigned an NQF level. In many countries, the main function of the databases is to be a communication tool that provides information to the public about qualifications and educational systems. The information is primarily for learners, but the databases are also used as a communication tool for HR, employers, career guidance and parents. Finally, the databases also serve as sources for statistics, and ENIC/NARICs are using the databases for recognition purposes.' The following stakeholders were identified as the users of the databases: 'learners, education and training providers, awarding bodies, employers, qualification recognition bodies and governmental institutions.' In general, 'there was consensus about the main purpose of the qualifications databases, i.e., to establish a better understanding of qualifications to create trust, transparency, and comparability. In a European perspective, the databases are largely used for mobility purposes for learners but also to compare qualifications (for awarding bodies) (European Commission et al., 2019, p. 9).

The Austrian NQF-Register was specifically developed to support the general objective of making all qualifications mapped to the NQF in Austria

and Europe transparent and comparable. It is also intended to make a valuable contribution towards greater visibility of qualifications offered. Another (future) aim of the register is to link the data in this national database to the European database and make it available to end users in a timely manner.

For the other databases analysed in this project, the focus is much more on serving national stakeholders. For example, the **Bulgarian database** intends to support career counsellors, learners, teachers and parents to quickly find information about VET as well as HE providers and the learning programmes they offer. Also, the **Lithuanian database** intends to present easily accessible information on qualifications to a wide range of users and similarly, the current database in the Netherlands was developed mainly to provide insights (for students and employers) into the knowledge, skills and competences included in a qualification. The **Spanish database** intends to serve as guidance tool for students, teachers and companies and it serves as a National Catalogue of VET programmes from the Education System. The current **Irish search engine** was developed to provide the public with a list of further education and training awards made by QQI as part of the Common Awards System. Under QQI's national legislation there is a requirement to develop a register of qualifications and details of providers with courses on the NFQ. Thus, a new register is in development.

In Finland, the main reason for the development of the platform is to improve the digitisation of public services, inter alia in order to reduce the costs of the public sector. It was developed as part of the national SADe programme (Action Programme on eServices and eDemocracy) during the years 2009–2015. The aim of the SADe programme was 'to provide interoperable, high-quality public sector services via digital channels. These services will improve cost-efficiency, create savings, and generate benefits to citizens, businesses, organisations and local and government authorities. Special attention will be paid to the achievement of cost benefits to municipalities' ([53]). The more specific aims for the educational sector were the need to build a common digitalised structure for qualification requirement documents in order to enhance their usability and actualisation in regard to the implementation of VET in Finland. Hence, ePerusteet contains also other digital tools to be used in implementing the VET qualifications, e.g. tools for designing individualisation/personalisation plans, student paths, provider specific curricula.

([53]) See: https://vm.fi/sade/perustiedot

The French RNCP, which is also mainly aimed at the national context, is an example of a database that already existed before the introduction of the EQF:

Box 13. **RNCP**

The RNCP is a transversal tool established in 2002 to provide visibility/readability to users regarding qualifications. The 2002 Law of social modernisation established the Validation of Experiential Learning Outcomes (*Validation des acquis de l'expérience, VAE*) as a route to qualification. The VAE was the fourth route to a qualification, in addition to initial education and training, continuing education and training, and apprenticeships. There were some concerns that the qualifications system may become somewhat complex, especially due to the high number of ministries awarding (sometimes similar) qualifications. The need for increased transparency of the French national qualifications system and better readability of qualifications for users was obvious to stakeholders and the RNCP was an important response. It was placed under the responsibility of the newly created National Commission of Qualifications (*Commission nationale de la certification professionnelle*, CNCP). Since 2018, *France compétence* is responsible for the RNCP.

The RNCP is organically linked up to French Operational Catalogue of Occupations and Trade Jobs (*Répertoire operationnel des métiers et des emplois, ROME*), of the Public Employment Service.

Source: Country template France.

All databases include individual qualifications. In Austria, the database contains individual qualifications and their learning outcomes but also refers to qualification types: Formal qualifications are levelled by using a 'combined' approach: mapping requests were submitted for a sample of individual qualifications (e.g. for ten apprenticeship qualifications and for eleven qualifications from schools for intermediate vocational education), but the mapping decision is applied to all qualifications belonging to the respective type. However, the NQF Register displays information (and learning outcomes descriptions) only for those individual qualifications for which mapping requests were submitted.

In the following countries, the database already has a full coverage of IVET qualifications: Bulgaria, Denmark, Spain, France, Ireland ([54]) (in the current search engine), the Netherlands, Finland, UK-England ([55]):

([54]) The Irish database contains all qualifications under the Common Awards System but some qualifications that might be considered to be IVET, notably apprenticeships, are not included. However, apprenticeships represent a small fraction of people doing IVET qualifications.

([55]) In all of these databases, the ICT service technician qualification is also included.

(a) In France, since the early days, all IVET qualifications have been in the RNCP. However, none of the qualifications awarded by the Ministry of Health are in the RNCP and there is no plan to add them in the near future. One reason is that all qualifications in the RNCP have to offer Validation of Experiential Learning Outcomes (*Validation des acquis de l'expérience*, VAE) and none of the Health qualifications are available through VAE.

(b) In Ireland, both the current search engine and the new register are currently populated with QQI awards Levels 1- to 10 of the NFQ. Since QQI make the only VET awards on the NFQ, the details from Level 1 to 6 on the new register is complete.

(c) In the Netherlands, all IVET qualifications at NQF/EQF levels 1-4 that are currently valid are included any new qualifications developed are added periodically ([56]) after confirmation by the Ministry of Education, Culture and Science.

(d) In Finland, the database already has a full coverage of formal IVET qualifications. The database includes all formal qualifications from early childhood education to the upper secondary level education which lie under the jurisdiction of the EDUFI, the Finnish National Agency for Education. Only formal qualifications are included and currently there are no plans to extend the scope. However, what will be extended is the number of available local curricula, i.e. provider-specific curricula made available at the site.

(e) In the UK-England, only those qualifications that qualify for registration are included. There may well be many others that are unregulated. As quoted above, Ofqual does hope to regulate newly emerging qualifications conforming with government policy on qualifications. But there is also a 'grey' sector of non-registered qualifications that are based on informal and non-formal learning.

The situation is different, however, in Lithuania and Austria ([57]):

(a) In Lithuania, the database currently includes vocational qualifications that are designed by following the 'old' or previous model of VET standards, developed in 1997 and updated in 2008. But with the approval and registration of the new modular VET curricula - designed according to the

([56]) Roughly every three months.

([57]) In both databases, the ICT service technician qualification is not included. In Lithuania, since neither the occupational standard of the ICT sector nor the modular training programme for the ICT maintenance and service specialists has been approved yet, this qualification is not yet included in the database. In Austria, only the qualification type apprenticeship qualifications and example qualifications are included; the ICT service technician qualification is not among these examples.

new qualifications described in the occupational standards - these VET curricula and qualifications will be gradually included in the database.

(b) In Austria, the main types of IVET qualifications are already included (but not as individual qualifications). IVET qualifications from the health care sector are not yet included.

In most countries, the database including IVET qualifications (analysed in this study) only includes formal qualifications ([58]) (Bulgaria, Denmark, Spain, Lithuania, the Netherlands, Finland). In Ireland, the current search engine includes formal qualifications, the further education and training awards made by QQI. The new register currently includes only formal qualifications that are on the NFQ. However, there is scope to include other qualifications as well.

The French RNCP contains all qualifications formally agreed (i.e. delivered in the name of the State), whether CVET, IVET, or otherwise (e.g. tertiary education), regardless of the ministry in charge of awarding the qualification (e.g. Ministry of Education, Ministry of Labour, Ministry of Agriculture) or regardless of the pathway (academic or vocational, although, in fact, they all are called 'vocational' anyway). Since the Law of 2018 for the 'Freedom to choose a vocational future', the general Baccalaureate and the PhD are also included. All qualifications awarded in the name of the State are directly registered. Other qualifications may be registered on demand ([59]). The database in the UK-England includes all qualifications, but those that rely exclusively on non-formal or informal learning will probably not be able to register on the RQF under the regulations that apply from 2016 onwards.

In Austria, according to the NQF Act, all qualifications / qualification types included in the NQF will be published in the NQF Register (*NQR Register*).

([58]) There are no commonly agreed definitions for formal or non-formal qualifications. 'Formal qualifications' are usually defined as those qualifications that are awarded within the formal qualification system (usually regulated by law), whereas 'non-formal qualifications' are those awarded outside the formal system, for example by private training providers (see also Sgarzi & Debowsky, 2019).

([59]) In this case, a case is constructed and a decision is made on the basis of the quality of the case.

Box 14. **The Austrian NQF Register**

The mapping of qualifications started in 2017 with formal qualifications (i.e. qualifications with a legal basis). Currently, only VET qualifications are included.

A procedure for mapping non-formal qualifications is currently being prepared; mapping is expected to start in 2019.

'Bologna qualifications' are not included because mapping requests are not required since they are assigned by the NQF Act ([60]).

Any decision related to formal qualifications from general education is still pending. It has to be noted that qualifications from general education including the *Reifeprüfung* certificate from AHS schools (upper secondary school leaving certificate from general education which gives access to higher education) were not included as 'reference qualifications' to illustrate the NQF levels in the EQF referencing report.

Source: Country template Austria.

The population of the NQF Register is an ongoing process: as soon as a new qualification is included in the NQF (mapping process completed), qualifications / qualification types will be published in the NQF Register. Currently (as of July 2019), the following qualifications / qualification types are included: ([61])

Table 9. **Qualifications included in the Austrian NQF Register (July 2019)**

NQF/EQF level	Qualifications	
	IVET	Other
1		-
2		-
3		-
4	-Apprenticeship qualifications (10 individual qualifications) - School for Intermediate Vocational Education qualifications (5 individual qualifications) - Agricultural technical school qualifications (6 individual qualifications)	- Military non-commissioned officer (MBUO, first use)
5	- College for Higher Vocational Education qualifications (10 individual qualifications)	- E2a basic training for the executive service in the use group E2a in the justice department ('basic training for the executive service' - middle management)

([60]) see: https://www.qualifikationsregister.at/nqr-register/nqr-zuordnungen/

([61]) https://www.qualifikationsregister.at/nqr-register/nqr-zuordnungen/

6		- Master craftsperson qualifications (5 individual qualifications) - engineer qualification (*Ingenieur*)
7		-
8		- Clinical Psychology - Health Psychology

Source: Country template Austria.

2.3.2 Relationship between the data on qualifications included in the database and the reference documents (for IVET qualifications)

In some countries, the relationship is rather loose, i.e. the database does not include the full description of qualifications as provided in the reference documents:

(a) In **Bulgaria**, the titles of the IVET qualifications (as defined in the reference documents) are used as search categories in the database. They are translated into English, however, details about a specific qualification is provided in Bulgarian language only.

(b) The database in the **UK-England** gives the title, level and some other minimal information. It provides a link to the awarding body, but access to the actual qualification document is not available without payment and/or permission from the awarding body. All information at the appropriate level of detail such as learning outcomes and unit aims can only be found within the qualification documents themselves.

(c) In **Lithuania**, the AIKOS qualifications database is linked to the National Register of the Training and Higher Education Study Programmes and Qualifications; thus, all registered curricula and qualifications have to be present in the AIKOS database. However, the vocational and professional qualifications defined by the occupational standards are not yet included in the database, because the occupational standards are still in the process of design, approval and implementation. Upon the completion of this process, the linkage between the occupational standards and AIKOS should be provided. However, so far there are no clear technical decisions on this issue. Therefore, currently the database includes vocational qualifications that are designed following the 'old' model of VET standards developed in 1997 and updated in 2008. But with the approval and registration of the new modular VET curricula designed according to the new qualifications described in the occupational standards these VET curricula and qualifications will be gradually referenced in the database.

(d) In **Austria**, the relationship depends on the individual organisations requesting a mapping (i.e. inclusion of their qualification in the NQF): They need to complete a 'mapping request template' which requires the description of learning outcomes. The description of learning outcomes has to be based on those presented in 'official' sources but these do not have to be adopted one-to-one. The learning outcomes presented in the NQF Register are usually an excerpt or summary of the learning outcomes included in the official sources. If the summary is not readily available, it needs to be developed specifically for the mapping request. The new mapping request template (valid since the beginning of 2019) asks for a description of the qualification with approx. 10-15 subject-related and interdisciplinary (transversal) learning outcomes at a higher level of abstraction (i.e. as a summary of the qualification of maximum 400 words) ([62]).

A rather close relationship can be stated for the remaining countries covered in this study:

(a) In **Denmark**, the database focuses on programmes and contains short general descriptions and practical information but not qualifications with their learning outcomes. There are, however, links to further sources of information (including relevant executive orders for VET qualifications including learning outcomes – they can be accessed under the category 'Laws and executive orders').

(b) In **Spain**, the reference document for each qualification (*Real Decreto*) is incorporated in the TodoFP database and can be downloaded as PDF-file.

(c) In **France**, the reference documents provide most of the content included in the databases:

  (i) The qualification standard of the vocational Baccalaureate is stored as such in the Ministry of Education database. The RNCP contains a summary of this Ministry of Education sheet.

  (ii) The RNCP contains a summary of REAC (standard for the vocational training of adults) and RC (qualification standard that describes the assessment process) which are stored in the database of the National Agency for the Vocational Training of Adults (AFPA).

([62]) As of July 2019, however, no mapping procedure has been carried out using the new template. The old mapping request template (valid until the end of 2018) asked for information on the main learning outcomes for publication in the NQF Register (max. 1,000 words).

(d) In the **Netherlands**, the reference documents are available online through a searchable database ('SBB Registry') that allows for retrieval of a qualification's basic characteristics in terms of level, validity and unique identifier. Furthermore, the database allows for retrieval of the main reference documents corresponding to the qualification (qualification file and occupational profile): The reference documents are downloadable from the landing page of the selected qualification, supplemented by an overview listing the optional curriculum elements (with links to their landing pages) that are available for said qualification. However, the database itself only includes a short description of responsibilities and autonomy (which is not a summary of learning outcomes) and information on the level, validity and CREBO-code ([63]) of a qualification.

(e) The current search engine in **Ireland** contains the awards specifications of the CAS, i.e. the Certification Specifications and Component Specifications – as PDFs (for the new register this is unclear at this stage).

(f) The **Finnish** reference documents, the national vocational qualification requirements, are published and made available at the *ePerusteet*-platform. Thus, it is the national source for IVET qualification requirements.

2.3.3 Connection to European portals

According to the European Commission, qualifications databases and registers should be linked to the European 'Learning Opportunities and Qualifications Portal' (LOQ) ([64]) and to ESCO ([65]). **As part of an effort to improve communication, the Commission is currently developing the new Europass portal (online platform) to be launched in 2020.** All information on qualifications in the LOQ portal will be integrated in the new Europass platform.

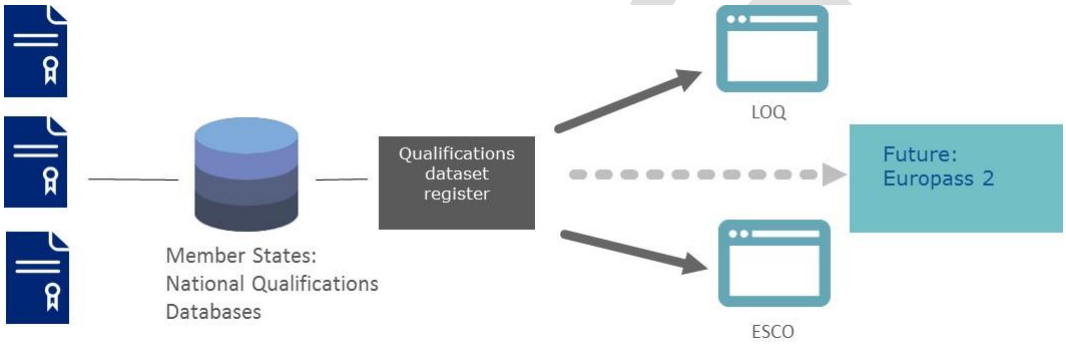
([63]) Each qualification is assigned a unique identifier ('CREBO'-number), which can be used to search for its Europass certificate supplement through a separate database. The Europass supplements can be accessed and downloaded through https://cs.s-bb.nl/

([64]) https://ec.europa.eu/ploteus/en?cookie=no

([65]) The ESCO qualifications pillar also aims to collect existing information on qualifications. The final objective of the pillar is to provide a comprehensive list of qualifications relevant for the European labour market: (a) qualifications included in NQFs that have been referenced to the EQF will be indirectly included in ESCO via the linkage of national qualifications databases of Member States to the ESCO qualifications pillar; (b) other qualifications that are not part of NQFs but are also relevant for the European labour market (such as private, sectoral and international qualifications) will be directly provided to ESCO by awarding bodies. https://ec.europa.eu/esco/portal/qualification

For publication of information at European level, the Commission has developed the Qualifications Dataset Register (QDR), software that allows the exchange of data between different stakeholders and European portals (currently LOQ and ESCO, and in the future Europass). Countries can either upload their dataset in the QDR or create a dynamic link with the national database. If countries have more than one database or register for qualifications reflecting their NQF, all can be connected with the QDR separately, without a need for integration. Currently an automatic link exists between the QDR and ESCO, and the same link between QDR and LOQ is in the process of development. The figure below illustrates how the connection to the European portals (LOQ and ESCO) works:

Figure 5. **Connecting national qualifications databases to EU portals**



Source: Moyes, 2019.

Despite grants being available (since 2014) from the Commission for the development of qualifications databases and their linkage to European portals, so far, only a few countries have linked their databases to the LOQ portal: As of July 2019, links between national databases and the LOQ portal – Information on qualifications ([66]) are in place for the following countries: BE (vl), BE (fr), DE, EE, EL, HU, IE, LT, LV, PT, SI. The ESCO qualifications pillar includes data from the following countries ([67]): BE, EE, EL, HU, LV, LT, PT, SI.

From the ten countries covered in this study, as of July 2019, learning opportunities from the following countries are presented at the European portal: Austria, Denmark, Finland, France, Ireland, and Lithuania ([68]). However, the information on learning opportunities provided there is far from complete, and it seems it is not regularly updated ([69]).

([66]) https://ec.europa.eu/ploteus/search/site?f%5B0%5D=im_field_entity_type%3A97
([67]) https://ec.europa.eu/esco/portal/qualification#
([68]) https://ec.europa.eu/ploteus/search/site?f[0]=im_field_entity_type%3A96#
([69]) For example, it was not possible to find the qualification if ICT Assistant (TAI) from the French Ministry of Labour (AFPA). And the qualification of the Ministry of

So far, among these ten countries, only Ireland and Lithuania have qualifications in the database of the European portal, presenting qualifications included in NQFs. However, only the Lithuanian qualifications presented there include learning outcomes descriptions. In Ireland, it is planned that the new Irish Register of Qualifications will also share data on learning outcomes.

A link to the ESCO qualifications pillar has been established only for the Lithuanian database.

Some countries have, however, plans to connect to the European portals:

(a) The National Agency for VET in Bulgaria has discussed with the Ministry of Labour and Social Policy possibilities for linking the database with ESCO in the future, but no deadlines or concrete action plan have been set so far ([70]). The database is in the process of being exported to the LOQ portal to inform foreign students coming to Bulgaria.

(b) There are plans to link the Spanish database to the European portal: The responsible ministry plans to improve the 'TodoFP' portal to provide a better information service on all aspects of VET. One of the planned aspects is a better connection to the information available in the European Union. The Ministry of Labour is currently updating the qualification classification system through the PES to bring it into line with ESCO. Once this work is completed, they will analyse how their databases can be linked to European portals, both ESCO and LOQ.

(c) In the Netherlands, it is planned to link the new database of learning outcomes currently being developed to the ESCO portal. Besides a full listing of the learning outcomes in IVET, this database is to include information on which learning outcomes are mapped to which qualifications, thus, allowing for insights into the extent to which similar qualifications 'overlap' in terms of learning outcomes.

(d) In Austria, a link between the NQF Register and the 'LOQ portal – Information on qualifications has not yet been established. However, the linked open data approach is used for the NQF Register, and data could be transferred easily. A connection is planned for 2019-2020 (Moyes, 2019). But information on qualifications (learning outcomes descriptions) are currently only available in the German language.

Education called Vocational Baccalaureate 'Digital System' is there but under its old denomination (Electronic and Digital Systems), and with the old options (four instead of three for the Vocational Baccalaureate 'Digital System').

([70]) Based on an interview.

2.3.4 Construction of the databases and information provided for users

2.3.4.1 *Search function*

The analysed databases differ in the search function offered. The following table gives an overview of the categories used, and below the table, some special features of the search function are commented on.

Table 10. **Search functions in databases – categories used**

Country	Categories used
Bulgaria	Professional direction, profession (the name of the IVET qualification), keyword (speciality), IVET/CVET, region, location, school name
Denmark	Interests, admission requirements, region, NQF level
Spain	Sector (occupational area) and level (basic, medium, higher level, and specialised qualifications)
France	Occupational field (85 fields; by alphabetical order, or directly inputting the name of the field), keywords; or advanced search: RNCP specific code, the name of the full qualification, the ROME code (*Répertoire opérationnel des métiers et des emplois*), the NSF code (*nomenclature des spécialités de formation*), the level (French 1969 nomenclature or EQF)
Ireland*	Code, title, award class (e.g. major, minor), NFQ level, field of learning
Lithuania	Programme name, levels of studies, required education, municipality, field of education, sub-sector of education, institution name, subordination of institution, whether accommodation facilities are required ([71])
Netherlands	File type (qualification, qualification file, optional curriculum element or cross-over qualification ([72])), version (based on validity, whether it was changed/reviewed or is a cross-over qualification), level (MBO level, which is synonymous to NQF/EQF level), whether there are legal professional requirements (Y/N), which educational period a qualification file was published in or valid for ('academic year')
Austria	NQR level (1-8), type of qualification (one of the types already mapped can be selected), categories for presenting information on qualifications (title of the qualification, NQF level, qualification provider, key learning outcomes, authorisations (rights/entitlements in the labour market and for further education pathways), areas and sectors, entry requirements, duration)
Finland	title of the qualification requirements, qualification title, competence area or unit that is part of the qualification; valid, expired, in progress, in transition
UK-England	Qualification title, Qualification status, Qualification Type, Qualification level, Qualification sub-level, Organisation name, Sector Subject Area, Grading Type, Assessment methods, Offered in England / Northern Ireland, Regulated By CCEA Regulation, Total Qualification Time

Source: Country templates. * current search engine.

The **Spanish database** offers a simple text search tool embedded into the 'TodoFP' portal (Google), which however does not include any settings for advanced search (by category). The website is however structured in a way that allows users to search VET qualifications by sector (occupational area) and level

([71]) Three main categories are provided at the landing page: 'Want to learn' (general education and initial VET), 'Want to train in' (continuing VET and adult education), 'Want to study' (higher education); the categories presented in the table are available in the section 'Want to learn'.

([72]) A cross-over qualification is new type of qualification, currently considered an experiment by the Ministry of OCW that runs until 2025. This type of qualification consists of a selection of parts from existing qualifications and aims to cover the intersection between sectors. VET providers, in cooperation with labour market organisations, can compose such a qualification and request it to be accepted by the Ministry. A cross-over qualification can only be offered by the VET provider that has requested it. See also https://www.s-bb.nl/onderwijs/kwalificeren-en-examineren/cross-over-kwalificaties

(basic, medium, higher level, and specialised qualifications) – this is however done through website navigation and not a separate search feature. When selecting the occupational area of '*Informática y Comunicaciones*' (Informatics and Communications), for instance, users will arrive at a page summarising all VET qualifications from that sector, structured into levels (basic, medium, higher). When clicking one example, this will open a separate page with detailed information (including a summary of learning outcomes) on this particular qualification.

In the current **Irish search engine**, searching is straightforward given the number of ways of searching. The 'Title' search function, however, is not intelligent, meaning that one needs to use an exact word contained in the title, e.g. searching using 'computing' for 'Computer systems and networks' leads only to an exact match, i.e. 'Business computing'.

In the **Lithuanian database**, the search function does not allow for directly searching for a qualification. Access to information about the concrete qualification to be awarded on completion of a programme is possible as part of the description of the programme. This description includes the NQF and EQF levels and also a short summary of learning outcomes ([73]).

In the **Netherlands**, searching for a qualification in the database will lead to the corresponding landing page, which provides a set of tables that provide the qualification's name/title, a short description, the level, validity and download links to the reference documents (e.g. the qualification file and occupational profile). As there are cases of qualifications with the same or similar name, the easiest way to ensure the 'right' qualification is tracked down is to use its unique identifier, the CREBO-code, to search for it.

The **Austrian database** does not only provide the possibility to search for qualifications mapped to the NQF but also allows for comparing the learning outcomes and other descriptions of two qualifications included in the Register. However, this only means that the information is displayed side by side.

The **Finnish database** also offers a simple text search function to search for national qualification requirements (*ePerusteet*). One can search for qualifications by the name of the qualification or by word search and one can decide whether to search for qualifications that are currently valid or (also) for qualifications that have expired, are in progress or are in transition. In addition, it is possible to tick the fields 'qualification titles', 'competence areas' and 'units' to

([73]) Example: 'Computer hardware adjuster': https://www.aikos.smm.lt/en/Learn/_layouts/15/Asw.Aikos.RegisterSearch/ObjectFormResult.aspx?o=KVAL&f=KvalEn&key=1001&pt=of

define the scope of the text search. By clicking the desired named of a qualification, one is directed to an opening page of the qualification which provides basic information on the qualification. The following categories are used for providing information on a qualification: Title of the qualification, Journal number of the regulation, Date of the regulation, Valid from, Regulation document (PDF-file), Statistics Finland codes for education, competence areas, qualification titles, Competences acquired after completion, Work tasks that can be executed, Descriptions of the competence areas, Document web page, Certificate supplement. Moreover, there is a menu for finding more detailed information on a qualification: Units (all units are listed, including learning outcomes and assessment criteria), Composition of a qualification (specifying the total number of competence points for the qualification, the compulsory and optional vocational units and their competence points for each competence area, and the common units and their competence points: Communication and interaction competence, Skills in mathematics and natural sciences, Citizenship and working life competence), Assessment scale, Competence areas (overview).

In the **UK-England**, the easiest way to access a qualification (including those no longer or not yet valid) is to key in the qualification title. The advanced search option is more difficult to use since it includes various elements.

2.3.4.2 *Elements for data fields for the electronic publication of information on qualifications with an EQF level*

One of the new elements of the revised EQF Recommendation includes the invitation to Member States to make the results of the referencing process publicly available, and to ensure that information on qualifications and their learning outcomes is accessible and published (using data fields included in Annex VI of the Recommendation). The revised Recommendation also refers to a common format for presenting qualifications. Such a format could be applied for international communication and comparison purposes, without affecting learning outcomes descriptions in national qualification documents. However, the participants at the Budapest PLA on national qualifications databases concluded that one ‘of the major issues during the linking process to the European portals has been data fields. Not all requested information was available for all qualifications and collecting this information has therefore been time consuming. Some countries are already technically prepared to link to the European portals, but are still waiting for information from the qualification providers’ (European Commission et al. 2019, p. 11).

The elements for data fields ([74]) (six required ones and twelve optional ones) for the electronic publication of information on qualifications with an EQF level are presented in the table below.

Table 11. **Elements for data fields for the electronic publication of information on qualifications with an EQF level**

Data field	Required/Optional
Title of the qualification	Required
Field (ISCED FoET2013)	Required
Country/Region (code)	Required
EQF Level	Required
Description of the qualification (either as 'Knowledge, Skills, Responsibility and autonomy' or as Open text field describing what the learner is expected to know, understand and able to do)	Required
Awarding body or competent authority	Required
Credit points/notional workload needed to achieve the learning outcomes	Optional
Internal quality assurance processes	Optional
External quality assurance/regulatory body	Optional
Further information on the qualification	Optional
Source of information	Optional
Link to relevant supplements	Optional
URL of the qualification	Optional
Information language (code)	Optional
Entry requirements	Optional
Expiry date (if relevant)	Optional
Ways to acquire qualification	Optional
Relationship to occupations or occupational fields	Optional

Source: Council of the European Union, 2017, Annex VI.

The participants at the PLA on national qualifications databases agreed that 'Annex VI [of the EQF Recommendation] generally fulfils its role by providing relevant information for users and that it also inspires new databases' (European Commission et al. 2019, p. 9). This is true, for example, in the Austrian case: Since the Austrian NQF Register was developed in the context of the EQF implementation, it is in general based on the common data model. Also in case of the new Irish register, the data model formed the basis for the data fields populated by QQI and required from the national awarding bodies.

([74]) A 'data field' is the smallest part of a database (with a unique name) where data can be entered, stored and displayed. The term is often used to refer to a column in a database or a field in a data entry form or web form.

In the other countries covered by this study, this data model was not used (at least not explicitly) but some of the categories (data fields) are reflected. The following table presents an overview of the 'required' elements covered in the databases of the countries covered by this study:

Table 12. **Required elements for data fields for the electronic publication of information on qualifications with an EQF level – used in the databases in the countries analysed (and visible for users)**

Data field	Used in the databases	
	Yes	**No**
Title of the qualification	BG, DK, ES, FR, IE*, IE-new, LT, NL, AT, FI, UK-England	
Field (ISCED FoET2013)	IE, LT, FI ([75])	BG, DK, ES, FR, IE-new ([76]), NL, AT, UK-England
Country/Region (code)	BG, DK, FR, IE-new, LT	ES ([77]), IE, NL, AT, FI, UK-England
EQF Level	FR, IE, IE-new, LT, NL, UK-England	BG, DK, ES, AT, FI
Description of the qualification (either as 'Knowledge, Skills, Responsibility and autonomy' or as Open text field describing what the learner is expected to know, understand and able to do)	ES, FR, IE, LT, AT***, FI	BG, DK, IE-new**, NL ([78]), UK-England
Awarding body or competent authority	FR, IE, IE-new, LT, AT, UK-England	ES ([79]), BG, DK, NL, FI

Source: Country templates. *IE: refers to the current search engine; IE-new refers to the new register. **in development; ***in case of VET, only for examples and not for each individual qualification.

The title of the qualification is the only data field that is available in all databases (although not always as a search category, as in the case of

([75]) For each qualification the 'Statistics Finland codes for education', which correspond to ISCED FoET2013, are included (but not as a search category) – see: http://www.stat.fi/meta/luokitukset/koulutus/001-2016/index_en.html.

([76]) The new search engine has a 'Field of Learning' filter but it currently has no content.

([77]) The database only includes a link to the provincial curricula.

([78]) The database includes a field called '*Omschrijving*' which only refers to attitudes / personal requirements of the holder of the qualification and to responsibilities and autonomy. For example, the (translated) description for ICT service technician (here: ICT management employee): 'The IT management employee shows his own initiative within set frameworks. He is able to work independently and shows his own insight when applying standard procedures and methods.' The learning outcomes descriptions are not present in the searchable database as text fields, as they are included separately through the reference documents (the qualification file and qualification profile).

([79]) Information available only indirectly through linked documents on the site.

Lithuania) ([80]). The data field 'Country/Region (code)' is not always explicitly included, because in most cases, the database was developed for national purposes only. Although several databases contain information on sectors or occupational fields, in most cases the ISCED fields of Education and Training 2013 (FoET 2013) are not used; only in a few cases can the correspondence be identified, even though usually the ISCED FoET2013 are not explicitly referred to ([81]).

As mentioned above, according to Cedefop (2019), 24 countries have included levels in their national qualifications databases ([82]): AT, BE (FL, FR), CZ, DK, EE, FR, DE, EL, HU, IE, XK, LV, LT, MT, ME, NL, MK, PL, PO, RO, SI, SK, TR, and the UK. Currently, some countries only include NQF levels. This is the case, for example, in the Austrian and in the Danish databases, where the NQF levels equal the EQF levels. In Bulgaria, only the degrees of IVET qualifications are visible which correspond to NQF/EQF levels ([83]). Thus, although the NQF/EQF levels are not used in the database, Bulgarian people who work with qualifications are aware of the respective correspondence between degree levels and NQF/EQF levels. In Finland, the EQF level is not directly visible on the website providing information on qualifications, but it is included in the ECS which is made available as PDF document on the same page (via link).

A data field for the 'description of the qualification (either as 'Knowledge, Skills, Responsibility and autonomy' or as Open text field, describing what the learner is expected to know, understand and able to do)' is not available in all databases. However, most of the databases offer the opportunity to download a full description of learning outcomes (in PDF-format), or a link to such descriptions. This is also often the case for information on the 'awarding body or competent authority'.

In some countries, data fields related to the 'optional' elements are also provided; some examples are presented below:

([80]) In the Danish database, the title used refers to programmes.

([81]) http://uis.unesco.org/en/topic/international-standard-classification-education-isced

([82]) Please note: This analysis does not necessarily refer to the same database as discussed in this chapter. For example, the Cedefop publication refers to the NLQF (Dutch qualifications framework) database which includes only private (non-regulated) qualifications. There is reference to CREBO/CROHO by including only the names of qualifications.

([83]) For example: 1st degree IVET qualifications correspond to NQF/EQF 2; 2nd degree IVET qualifications to NQF/EQF 3; 3rd degree correspond to NQF/EQF 4; 4th degree correspond to NQF/EQF 5.

(a) **Credit points/notional workload needed to achieve the learning outcomes**: These are included in the databases in Spain (only for qualifications at higher levels; for IVET qualifications, 'duration' in hours is indicated); in Ireland (both in the current search engine and in the new database), and in Lithuania (in the description of the curriculum). In the UK-England, guided and total learning hours are provided. Credit points seem to be calculated on a basis of 'ten hours of total learning time to one credit point', although this is not openly stated. In Austria, credit points are not used but 'duration' is presented instead. In Finland, 'competence points' are used. There is no explicit data field for competence points on the website that provides information about a qualification, but the information about competence points is available via the menu items 'Units' or 'Composition of the qualification'.

(b) **Internal quality assurance processes**: This information is available in the current Irish search engine (validation process links to provider quality assurance), as well as in the new database (there are links to quality assurance reports). This is not the case in the Austrian NQF Register but information on quality assurance of the assessment process is provided in the mapping request template and could be made visible. In Finland, this information is not directly available on the website that provides information about a qualification, but providers (awarding bodies) are invited to present their own 'execution plans' (i.e. local curriculum, including information on quality assurance procedures) on another part of the Eperusteet website (however, this is not compulsory and the information available is not complete) ([84]).

(c) **External quality assurance/regulatory bod**y: Related information is provided in the French and both Irish databases.

(d) **Further information on the qualification** is available in several databases: the Bulgarian database (professional direction, keyword - speciality, IVET/CVET, location, school name, form and duration of learning as well as financing (public/private school); the Spanish database; the Danish database (e.g. application and admission, how to find an internship, how to combine vocational education with a high school diploma, vocational training for adults, financial issues, further education opportunities, job opportunities); the French one (validity of the acquired components); both Irish ones; the

([84]) Information can be found here: https://eperusteet.opintopolku.fi/#/fi/selaus/kooste/ammatillinenkoulutus?hakutyyppi=jarjestajat

Austrian Register (duration of qualification acquisition, rights/entitlements in the labour market and for further education pathways); the Finnish one (see information above); and the database in the UK-England: Qualification Status (availability to learners), Qualification Type, Sector Subject Area, Grading Type, Assessment methods, Offered in England / Northern Ireland, Regulated By CCEA Regulation, Guided Learning Hours, Total Qualification Time.

(e) **Source of information**: This is indicated in the databases from Spain, France, Ireland (both), Finland, and UK-England. Again, this is not included in the Austrian Register, but information is provided in the mapping request template and could be made visible.

(f) **Link to relevant supplements**: Such links are available in the databases of the following countries: Denmark, Spain, France, Ireland (both), Lithuania, the Netherlands (download links to reference documents), Finland, UK-England.

(g) **Information language (code)**: This is provided in the Spanish database, the new Irish database and in the Lithuanian database.

(h) **Entry requirements**: Respective information is available in the databases from France, Denmark, Spain, Ireland (both), Lithuania, and Austria.

(i) **Expiry date**: In the current Irish search engine, 'Status' information is given (e.g. qualifications may be 'under review'). Relevant information is also included in the Irish database and the Lithuanian one. In the UK-England, 'Operational end date' and 'Certificate End date' are indicated but it is not explained what the distinction between these two is.

(j) **Ways to acquire qualification**: These are indicated in the databases from Spain (the database includes an external link to institutions offering the respective programme), France and Lithuania, as well as in the new Irish database. In Finland, this is not a data field on the website that provides information about a qualification, but the information is available in the description provided by providers (in the 'execution plans', i.e. local curriculum).

(k) **Relationship to occupations or occupational fields**: This is indicated in the databases from Spain, France, Lithuania, Austria (in Austria, information on areas and sectors in which qualification holders can typically be active is included, as well as on entitlements which are linked to the qualification in the labour market), and Finland (work tasks that can be executed are described).

2.3.4.3 *Availability of information of particular relevance for the current study*

For comparing qualifications and their learning outcomes, some information elements are of particular relevance. The table below shows that the ones identified for this project are not commonly included in all databases:

Table 13. **Information relevant for the current study is available in the database in the countries analysed**

Information	Available in the database	
	Yes	**No**
Full learning outcomes descriptions of qualifications including detailed learning outcomes descriptions in units/modules	FR, IE, IE-new**, FI ([85])	ES ([86]), BG, DK, LT, NL ([87]), AT, UK-England
Short summary of learning outcomes (sometimes called 'qualification profile')	ES, FR, IE, IE-new**, LT, AT***, FI	BG, DK, NL ([88]), UK-England
NQF level	DK, FR ([89]), IE, IE-new, LT, NL, AT, FI ([90]), UK-England	BG, ES
EQF level	FR, IE, IE-new, LT, NL, FI ([91]), UK-England	BG, DK, ES, AT
Possibilities for further learning (particularly: access to higher education)	DK, ES, IE-new, AT	BG, FR, IE, LT, NL, FI ([92]), UK-England

([85]) Learning outcomes descriptions of each unit included in a qualification are available.

([86]) Full learning outcomes descriptions are accessible from the qualification page in the database.

([87]) The learning outcomes descriptions are not present in the searchable database as text fields, as they are included separately through the reference documents (the qualification file and qualification profile).

([88]) The database includes a field called '*Omschrijving*' which only refers to attitudes / personal requirements of the holder of the qualification and to responsibilities and autonomy. For example, the (translated) description for ICT service technician (here: ICT management employee): 'The IT management employee shows his own initiative within set frameworks. He is able to work independently and shows his own insight when applying standard procedures and methods.'

([89]) 1969 *Nomenclature*, i.e. 'exit levels'.

([90]) The NQF level is not directly visible on the website providing information on qualifications but it is included in the ECS which is made available as PDF document on the same page (via link).

([91]) The EQF level is not directly visible on the website providing information on qualifications but it is included in the ECS which is made available as PDF document on the same page (via link).

([92]) This information is not available directly at the eRequirements platform but at the main site of the platform at which eRequirements platform is situated one can find all

Link to occupations/ labour market/ role of the qualification in the occupational context	DK, ES, FR, LT, AT, FI	BG, IE, IE-new, NL ([93]), UK-England
Distribution of types of learning outcomes (general knowledge subjects, transversal learning outcomes and occupational learning outcomes) – in percentages		BG, DK, ES, FR ([94]), IE, IE-new, LT, NL, FI ([95]), UK-England
Additional contextual information (such as the awarding body or the legal basis for the qualification or any other contextual and institutional – input – factors)	FR ([96]), DK, IE ([97]), IE-new, LT, AT ([98])	ES, BG, NL, UK-England, FI ([99])

Source: Country templates. *IE: refers to the current search engine; IE-new refers to the new register. **in development; ***in case of VET, only for examples and not for each individual qualification

relevant information in regard studying in Finland, including VET: https://studyinfo.fi/wp2/en/

([93]) This information is not present in the searchable database, but in the qualification file.

([94]) Not a heading but sometimes indicated in the 'Assessment procedure' provided in the document of origin.

([95]) Information on the distribution of the types of learning outcomes is not available as a percentage, but indications can be found under the menu item 'Composition of a qualification'. For example, the qualification 'Vocational Qualification in Mechanical Engineering and Production Technology' is composed of 'vocational units' with 145 competence points and 'common units' (Communication and interaction competence, Skills in mathematics and natural sciences, Citizenship and working life competence) with 35 competence points. However, as transversal learning outcomes are also integrated into vocational units, the distribution of types of learning outcomes cannot be clearly derived from the information presented.

([96]) Application orders, statistics, place(s) of qualification.

([97]) Major Award Certificate Specifications show e.g. 'Purpose of the award', 'Certificate Requirements' (sets out the structure in terms of mandatory and optional modules and credit values), 'Specific Validation Requirements' (e.g. 'Where Maths for STEM is used, the additional credit can be drawn from the pool of electives'), 'Field of learning'. In addition to the above, Minor Award Component Specifications show: 'Assessment' - provides information on assessment processes, 'Recognition of prior learning' – validation of non-formal and informal learning, 'Supporting documentation' (typically this says 'None').

([98]) Qualification provider, entry requirements, duration.

([99]) While this information is not directly available at the webpage providing information on a qualification, further information is available in the sections presenting provider-specific curricula (https://eperusteet.opintopolku.fi/#/fi/kooste/3397335) and accompanying instructions and materials, e.g. for assessment, skills demonstrations, personalisation (https://eperusteet.opintopolku.fi/#/fi/selaus/kooste/ammatillinenkoulutus?hakutyyppi=oppaat)

2.3.5 Link between the databases and Europass Certificate Supplements (ECS)

The ECS's 'Profile of skills and competences' would be a logical source for summarising the learning outcomes of IVET qualifications that can be presented in qualifications databases. In Germany and Slovenia, for example, the ECS are used as the basis for the description of VET qualifications (European Commission et al., 2019). Such close links between ECS and the information on qualifications, presented in the databases analysed here, can only be identified in a few cases:

(a) In the **French database** there currently is a link: When the RNCP was managed by the CNCP (National Qualifications Commission, *Commission nationale de la certification professionnelle*) until 2018, there was a clear policy – at the RNCP – that the RNCP Sheet would directly form the Certificate Supplement (*Supplément descriptif du certificat*) to be sent to Europass through the Erasmus Office in Bordeaux ([100]). The Certificate Supplement is therefore a copy of the RNCP Sheet with the Europass logo on the top right of the first page (Box A.1), and is directly available on the RNCP web page ([101]) where it says '*Fiche Euro*' (Box A.2). It is not clear, at the time of the drafting, whether this policy has been systematically maintained with the transfer of the RNCP from the CNCP (now dismantled) to the new body *France compétences*.

(b) In **Spain** ([102]) and **Ireland** (both the current search engine and the new IRQ), there is a direct link to the ECS from the database, and in **Lithuania**, a link to the ECS is provided as part of the descriptors of the registered VET programmes (curricula).

(c) In the **Netherlands**, the ECS profile of skills and competences is not presented at the landing page of the SBB registry for any qualification, though it is included indirectly through the reference documents (qualification file, 'occupation in short' leaflet) included as downloadable files (PDF) at the page. As stated above, the 'occupation in short' leaflet does include the same 'profile' as the ESC does, describing learning outcomes as a list of core tasks and work processes. The qualification file, on the other hand,

([100]) http://www.agence-erasmus.fr/page/europass

([101]) http://www.rncp.cncp.gouv.fr

([102]) In Spain, for the ICT service technician, the European Diploma Supplement is included as downloadable file (PDF). While the database presents a short summary of learning outcomes, the European Diploma Supplement includes a description of the professional modules and their learning outcomes.

presents a more tabular overview of the learning outcomes included in the qualification – using separate tables for the base part and profile parts of the file that list the corresponding learning outcomes. Additionally, the qualification file includes separate sections to further describe each learning outcomes as Open text fields.

2.3.6 Descriptions of learning outcomes of IVET qualifications included in the databases

In some countries, the descriptions of learning outcomes accessible through the database are identical to descriptions of learning outcomes in the reference documents (discussed in Chapter 2.2) or the databases do not include learning outcomes at all. Thus, the following databases are not fully covered in this section: the databases from Bulgaria, Denmark, Finland, Ireland, the Netherlands, UK-England. The analysis related to the description of learning outcomes in this section therefore mainly refers to the databases from Spain, France, Lithuania and Austria. Moreover, some key results of the PLA on national qualifications databases (European Commission et al., 2019) are included.

The PLA on national databases (European Commission et al., 2019) observed that some 'countries use LOs at a very detailed level for description of qualifications in databases/registers. While other countries use shorter 'synthetic' descriptions of qualifications by using LOs' (European Commission et al., 2019, p. 7). The learning outcomes descriptions, available in the databases analysed here, are usually not complete, meaning a short summary is provided instead. Thus, compared to other national sources (the 'national reference documents' analysed in Chapter 2.1), these descriptions are usually less detailed. Moreover, there is usually no structured approach of grouping learning outcomes and learning outcomes descriptions are usually not structured in a hierarchical way.

The Cedefop (2017, 62pp.) handbook on defining, writing and applying learning outcomes suggests the use of common principles for presenting learning-outcomes-based qualifications. These common principles are linked to the agreed common 'data model' for the collection, presentation and sharing of information on qualifications in national databases or in qualification supplements. It is suggested to develop short summaries (extracts) of national learning-outcomes-based descriptions of qualifications. In order to 'be accessible and comparable, the following technical requirements can be considered:

(a) for this summary/extract to be used in qualifications databases and/or supplements, it should be short (± 500 to 1 500 characters). This volume-indication, while flexible, reflects existing practices, for example related to Europass Certificate Supplements;

(b) it should follow a predefined structure and syntax. This is critical for ensuring comparability of presentations;
(c) it should refer to agreed but flexible learning domains. While some countries may choose to use the EQF domains (knowledge, skills and autonomy/responsibility) as a basis for their descriptions, countries and institutions should choose the distinctions they find most appropriate;
(d) it must be supported by a standardised terminology, including lists of action verbs' (Cedefop, 2017, p. 63).

The figure below shows the basic structure that can be used as a starting point for developments.

Figure 6. **Principles supporting the presentation of learning outcomes**

The learning outcomes description should be 500 to 1 500 characters and be written considering the following elements			
It should present the qualification from the perspective of the learner and what he/she is expected to know, be able to do and understand.	It should use action verbs to signal the level of learning expected, normally with an (explicit or implicit) reference to the levels of the national qualifications framework and/or the EQF.	It should indicate the object and scope of the expected learning outcomes. This description should capture the main orientation of the qualification and the depth/breadth of the expected accomplishment. It can, if deemed appropriate, use domains as defined by NQFs/EQF.	It should clarify the occupational and/or social context in which the qualification operates.

Source: Cedefop, 2017, p. 65.

Participants at the PLA on national qualification databases (European Commission et al., 2019, p. 7) 'did not favour a common format for describing learning outcomes. Instead more discussions should take place on good practices concerning the use of learning outcomes in databases at national levels. The participants welcomed the existence of guidelines for formulating LOs at EU level but stressed that it is important to consider that there are national differences and traditions as well. Consequently, guidelines should allow for national particularities.' Moreover, it was suggested to have a stronger focus on capacity building in relation to how learning outcomes are formulated and to share good practices at European level (European Commission et al., 2019, p. 11).

The common principles proposed by Cedefop are only partly reflected in the databases analysed:

In Spain, Ireland and France, there are no specific **regulations regarding length of text** or number of characters for presenting learning outcomes descriptions. In France, for example, the RNCP system provides a template ([103]) for the sections to fill in, but without indicating the length. Nevertheless, a careful review of many RNCP sheets clearly shows they all have the same approximate length, which means there was some sort of harmonisation during the preparation process, between the RNCP staff and the provider of qualifications.

In Austria, there is an indication of the required length for presenting learning outcomes descriptions in the database: The old mapping template asked for a maximum of 1,000 words, the new template asks for a maximum of 400 words for describing key learning outcomes that are displayed in the database.

The **proposed structure and syntax for describing learning outcomes** is only partly used in the databases analysed.

The short summary of learning outcomes in the Spanish database (under the heading *'¿Que voy a aprender y hacer?* / What will I learn and do?) includes statements that are composed of action verbs and objects and there is usually some reference to the context, for example: 'Manage server operating systems, installing and configuring the software, in quality conditions to ensure the operation of the system'; 'Evaluate the performance of the hardware devices identifying possibilities of improvements according to the needs of operation'; 'Diagnose system dysfunctions and take corrective measures to restore their functionality'.

In the French RNCP, the expected competences are described using action verbs; the context is usually not mentioned. The summary of the standard, however, uses the form: action verb + subject + context. Learning outcomes descriptions in the AIKOS database (from Lithuania) are composed of action verbs (for describing the tasks) and objects. However, a statement specifying the depth/breadth of learning to be demonstrated or an indication of the context is usually missing. In Austria, there is no systematic approach to writing learning outcomes or to composing statements of these components. Learning outcomes descriptions of apprenticeship qualifications usually use action verb and objective of the verb and sometimes also some context information. In learning outcomes descriptions of some School for Intermediate Vocational Education qualifications, not even action verbs are used consistently. The learning outcomes statements

([103]) http://www.cncp.gouv.fr/sites/default/files/media/cncp_notice_daide_cas_general_02_06_2017.pdf

of Colleges for Higher Vocational Education qualifications sometimes include action verb and object of the verb – as the following example shows:

Box 15. **Secondary College of Business Administration (*Handelsakademie*) (NQF level 5) – excerpt**

He/She

- can make well-founded decisions for a company on the basis of given information;
- can understand the conclusion and contents of relevant contracts (e.g. purchase contract);
- can track the fulfilment of contracts and take appropriate action;
- can apply various legal aspects to the employer-employee relationship;
- can describe the processes of dissolution of companies.

Other learning outcomes statements from the same qualification include other components as well:

The graduate is able to plan and critically question a marketing mix for a company's range of services on the basis of concrete market and company information. He/She

- can use various instruments of market and opinion research to collect relevant information in order to design a marketing mix.
- can efficiently implement process steps relevant to sales with the support of standard commercial software.
- can create, implement and evaluate a marketing plan and analyse its effects.
- can critically question a marketing mix from the point of view of a consumer.

Source: Country template Austria.

The learning outcomes statements included in the analysed databases also do not systematically reflect the **vertical dimension of learning,** and in none of the countries are taxonomies used for expressing the increasing complexity of learning nor is any guidance provided in this regard.

In some cases, specific words are used for signalling the complexity. For example, in the French RNCP, the term 'autonomous' is frequently used and sometimes complemented with some element of context information ('reporting to the line manager', 'respecting the commission and the context of the intervention', or 'first level intervention'). However, the vertical dimension does not appear clearly in the RNCP when it comes to the IVET qualification under study. Also, in the Lithuanian database such signalling words are used (e.g. 'basic'), but rather rarely and randomly. Similarly, in the Austrian database such words are not used systematically or consistently. Some examples can be identified: 'basic knowledge', 'familiar situations', 'comprehensive and specific skills and knowledge', 'on one's own responsibility'.

Although there is no systematic approach to writing learning outcomes in the databases analysed, the **learning outcomes are usually described in a similar**

manner, or at least no specific differences across qualifications can be observed. In Austria, different approaches are used – depending on the awarding body (qualification provider) and the department or person developing the learning outcomes descriptions for the mapping template that will be published in the NQF Register. However, these different approaches are not applied in a systematic way across different types of qualifications.

In France, differences between occupational and transversal learning outcomes can be observed: In the RNCP, transversal competences usually do not appear but whenever they are included, they are described in a very short way, as opposed to technical competences that are described in greater details.

The balance between information on learning outcomes in the database and data referring to other information is also assessed differently: While information on learning outcomes constitutes the main part of the presentation of qualifications in the Austrian NQF Register, there is an equal balance in the French database and in the Spanish and Lithuanian databases, the information on learning outcomes is presented in a brief way and constitutes about 20-25 per cent of the total information provided.

While the Lithuanian database presents **learning outcomes in English**, this is only sometimes the case in France, and not at all in Spain (there, English descriptions are available only in the ECS that can be downloaded from the database). The Austrian NQF Register has been designed to provide information and descriptions, both in German and English language, with the English language version currently being work in progress: Currently, learning outcomes are only presented in national language (German). In the near future, a translation of the learning outcomes descriptions of the formal qualifications already included in the NQF Register is planned. The EU grant is to be used for this purpose.

Participants at the Budapest PLA on national qualification databases (European Commission et al., 2019, p. 11) emphasised the need to translate information on qualifications into English to make them better understood outside the own country. However, since this is time-consuming and costly, it was suggested to use the information presented in ECS which are ‘already available in most countries in a widely spoken European language and could therefore reduce the translation work for VET qualifications.’

2.3.7 Technical infrastructure

In those databases that include learning outcomes descriptions, they are available and accessible to all users, including the general public. However, one of the features ideally needed to help achieve the goals of automated

qualification comparison is that data sources are formatted to be easily extracted and applied in software packages. Such a technical infrastructure is usually not available and the qualification data displayed in the database is prepared in various ways.

In Denmark, for example, there are no underlying databases feeding ug.dk. The ministry keeps the database updated continuously. In France there is preparatory work for the registration of all qualifications in the RNCP, namely the creation of the 'RNCP Sheet', which is done in collaboration between the RNCP and the provider (thereby it is required that the description of the qualification enables the general public to understand the content of the qualification). Similarly, in Austria, the description of a qualification displayed in the Register (including the learning outcomes) is based on a document (the 'mapping request') that is submitted by the qualification provider during the application for assignment to an NQF level.

None of the countries studied have experience with the use of digital technologies for automated collection, structuring (including cleaning, fusion) and analysis/comparison of qualification data - this usually has to be done manually. In the Austrian NQF Register, two qualifications can be displayed next to each other; this feature may support a comparison, but any comparative work still has to be performed by humans.

For the AIKOS database in Lithuania, it is stated that it does not support the use of digital technologies for this purpose. However, such functions will probably be available at the National System of Human Resource Monitoring (*Nacionalinė žmonių išteklių stebėsenos sistema*) which is currently being developed on the basis of linking and integrating different existing statistical databases on education and the labour market.

Similarly, for the Finnish database, it is stated that it is not suited or planned for automated gathering of data. However, one can also aggregate data to a certain extent. Moreover, the base language is Finnish (as well as Swedish, as the second official national language) and there will be no resources for the translation of ePerusteet in the near future.

The SBB registry in the Netherlands can be considered to be a web-based version of the database and may allow for digital technologies such as web crawlers to extract information from the landing pages (although these pages do not present learning outcomes). To what extent this allows for automated downloading of the files provided depends on the digital tool, though in theory this should be possible. The XML database would be superior in terms of gathering multiple files, as it allows for selecting cases from the database based on conditions (such as 'valid between year X and Y' or 'CREBO number X to Y').

The Austrian NQF Register ([104]) is based on new input and does not include entries from existing databases. Data transfer, however, is possible due to the linked open data approach.

2.3.8 Current activities or future plans for the further development of the qualifications databases

While there are currently no plans to further develop the databases in Bulgaria, Denmark, or UK-England, such plans are being discussed or already implemented in other countries and, in Ireland and in the Netherlands, new databases are being developed:

(a) The Lithuanian database has been updated and improved several times, by introducing different new functions, such as: assistance for vocational guidance and career design, personalised services, links to online applications for training and study programmes, links to the preparation of CVs, portfolios, career plans etc. Currently, there are discussions on the introduction of an online database with the structured descriptors of qualifications (occupational standards).

(b) The Austrian NQF Register was only recently relaunched to enhance usability. It now uses Wordpress, and the website has been developed according to the principles of barrier-free accessibility. In the near future, a translation of the learning outcomes descriptions of the formal qualifications (that are already included in the NQF Register) into English is planned.

(c) In France, it is difficult to say anything about future plans at the moment as the harbouring body of the RNCP has just changed (from the CNCP to France Compétences); and it seems that the main focus currently is on switching from five to eight levels in the NQF. The establishment of the new NQF has not changed the central place of the RNCP (and even less the place of the REAC, RC, and qualification standard of the Ministry of Education for IVET). This probably goes without saying, as the RNCP is the National Catalogue of Qualifications, the natural companion of the NQF. In other words, a national qualifications framework would be useless without a catalogue of qualifications ([105]).

(d) The new Irish register is in development to complete the publication of a comprehensive register - to meet QQI's national statutory obligation, and also to comply with EU requirements.

([104]) It is based on an Apache server with Apache Tomcat and is a Java program. PostgreSQL 9.2 is used as the database.

([105]) https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000038200990

(e) In the Netherlands, a project is currently underway aimed at creating a separate database (also maintained by SBB) containing all learning outcomes of all IVET qualifications and linking them simultaneously to the ESCO portal. The project is coordinated by CINOP and uses the SBB database of qualification files to extract and 'map' learning outcomes. Kennisnet is the body responsible for streamlining the data formats (from unstructured XML to an ESCO structure). This project has not yet been completed (planned for the end of 2019).

(f) Following the principle of continuous improvement, the Finnish platform ePerusteet platform, its functions, usability, user satisfaction etc., is under constant monitoring and development. However, as it is a relatively new platform at this stage no major developments are planned. As part of the new features a digital tool for designing of personal study plans has been recently introduced to the ePerusteet. Another new feature will be a digital tool for mapping of current competences of a student, i.e. there will be a digital tool available with which one can assess/compare his/her skills and competences against the ones presented in VET qualification requirements (learning outcomes), and as a result get some sort of competence profile which one can then use e.g. for seeking a suitable VET qualification, for designing personal study plan etc.

2.4 Assessment of sources against conditions for suitability of data sources for comparing national qualifications and automated text processing of qualifications data

Reference documents

The relevance of reference documents as data sources for qualifications suitable for the comparison of national qualifications and automated text processing of qualifications data is discussed in this section. This assessment is based on reflections related to the necessary conditions ('must haves') defined for this purpose that can be derived from section 2.2:

(a) **Unit of analysis**: One necessary condition is that the source describes the learning outcomes of a qualification. This is the case with the reference documents analysed. However, learning outcomes descriptions are available in a range of different types of documents (with different functions) across countries.

(b) **Completeness of the learning outcomes description**: A complete description is required for the context of this study. The reference documents for IVET qualifications from the countries covered by this study usually include a full description of learning outcomes of qualifications.

(c) **Sentence components**: Learning outcomes statements included in the reference documents are only partly composed of the components suggested by Cedefop (2017, p. 47) ([106]): In the countries studied, verbs and objects are common components of individual learning outcomes statements, although it is not uncommon for nouns to be used instead of verbs. Context is only sometimes given in individual learning outcomes statements; more commonly, context is provided in over-arching statements, which may be a more efficient way to indicate context, from the point of view of writing learning outcomes, than repeating it in individual learning outcomes. Moreover, it is usually the object and context of a learning outcome statement that expresses the vertical dimension rather than the verbs used.

(d) **Information related to 'key comparability criteria'**: Information on the EQF level is only in few cases provided in the same documents as the full and short descriptions of learning outcomes. The same is true for information on possibilities for further learning. Information on links to occupations/the labour market tends to be provided in the documents that also contain the full and short learning outcomes descriptions. Explicit information on the distribution of different types of learning outcomes is usually not provided.

(e) **Coverage of qualifications**: In general, all IVET qualifications within a country are covered by the same types of reference documents in a similar manner. Variations were only found in three countries (Ireland, Austria, UK-England).

(f) **Up to date**: The source needs to provide an up-to-date picture of the qualification. The validity of the description needs to be ensured. There are not necessarily specific processes for updating learning outcomes descriptions in reference documents that are publicly available when they are revised in qualifications: the revised version of the

([106]) The feasibility study on the conceptual and technical link between the learning outcomes of qualifications contained in the ESCO qualifications pillar and the ESCO skills pillar also emphasised the use of action verbs and objects for enhancing results accuracy and concluded that the potential for automated linking is maximised if conformance with the Cedefop guidelines on describing learning outcomes is ensured (DG EMPL, 2019, p. 75).

specifications/standards is simply published, without any special arrangements. Sometimes there are delays between finalising the revised descriptions and making them publicly available.

The following aspects refer to the optional conditions ('nice to haves') that could support the use of reference documents as sources for the comparison of qualifications:

(a) **Languages**: None of the main documents containing learning outcomes are systematically available in English (in addition to the native language).

(b) **Structure for presenting learning outcomes descriptions**: There are great variations across countries in whether and how learning outcomes descriptions of IVET qualifications are structured in the reference documents analysed. It is quite common for learning outcomes to be grouped into modules in a structured manner or to be structured so as to reflect groups of related occupations, with progressive specialisation 'beneath'. Only in some cases are learning outcomes structured according to the domains of learning of knowledge, skills and autonomy/responsibility. Sometimes, they are structured in terms of different levels of specificity.

(c) **Storage format**: The source is preferably formatted in a way that allows easy extraction and application in software packages. Learning outcomes descriptions are, in all countries, 'stored' in PDF versions. Such documents can only be accessed and downloaded individually providing no opportunity for any automated comparison (without additional preparatory (manual) work).

In order to better understand the characteristics of qualifications, the source should also provide additional information that characterises the qualification and its context. However, there is much variation across the countries regarding which documents have to be consulted for which pieces of information, making for a complicated picture in respect of the automation of international comparative analyses.

Databases

The relevance of qualification databases (identified in the ten countries) as data sources for qualifications suitable for the comparison of national qualifications and automated text processing of qualifications data is discussed in the following paragraphs. This assessment is based on the reflections related to the necessary conditions ('must haves') defined for this purpose:

(a) **Unit of analysis**: One necessary condition is that the source describes the learning outcomes of a qualification. This is not the case in all databases analysed. While in some of them, short summaries of learning outcomes are

provided, other databases offer only a link to the full description of learning outcomes (either to another website or a document that can be downloaded) but do not contain any learning outcomes descriptions directly.

(b) **Completeness of the learning outcomes description**: Although a complete description is required for the context of this study, the degree of completeness varies across databases. However, as mentioned above, a document with full descriptions of learning outcomes of qualifications can usually be accessed via one of the databases examined.

(c) **Sentence components**: Learning outcomes statements included in qualifications databases are rather rarely composed of all components suggested by Cedefop (2017, p. 47) ([107]): While in most cases, action verb and objective of the verb can be identified, a specification of the depth/breadth of learning to be demonstrated, or an indication of the context is usually not included.

(d) **Information related to 'key comparability criteria'**: While the EQF level is in most cases indicated in the databases (sometimes it is even one of the search categories), or can be easily 'translated' from the NQF level, this is less often the case for the purpose and currencies of qualifications and the extent to which qualifications provide access to further learning and (conditional/limited) access to higher education. The distribution of types of learning outcomes (percentages of general knowledge subjects, transversal learning outcomes and occupational learning outcomes included in a qualification), is not at all included in any of the databases analysed.

(e) **Coverage of qualifications**: Not all databases already have a full coverage of IVET qualifications and in one case (Austria), not all individual IVET qualifications are included. Usually, although systematic approaches cannot be clearly identified, the learning outcomes are to a certain extent described in a rather consistent way within each country.

(f) Up to date: There is not necessarily a regular process of updating the data included in the databases (i.e. the descriptions of qualifications).

([107]) The feasibility study on the conceptual and technical link between the learning outcomes of qualifications contained in the ESCO qualifications pillar and the ESCO skills pillar also emphasised the use of action verbs and objects for enhancing results accuracy and concluded that the potential for automated linking is maximised if conformance with the Cedefop guidelines on describing learning outcomes is ensured (DG EMPL, 2019, p. 75).

The following aspects refer to the optional conditions ('nice to haves') that could support the use of databases as sources for the comparison of qualifications:

(a) **Languages**: In most countries, the information included in the databases is presented in national languages only. Translation of (learning outcomes) into English is available only in a very few cases.

(b) **Structure for presenting learning outcomes descriptions**: The summary of learning outcomes available in some databases is not structured in a systematic way.

(c) **Storage format**: The source should preferably be formatted in a way that allows easy extraction and application in software packages. This is usually not the case, since only a few databases include directly accessible descriptions of learning outcomes of qualifications. Mostly, the databases include links to downloadable documents (usually PDFs) which include the learning outcomes descriptions.

In order to better understand the characteristics of qualifications, the source should also provide additional information that characterises the qualification and its context. The analysis of the databases, available in the ten countries covered by this study, shows that such additional information is provided to a varying extent and also the descriptors used differ across countries. While some databases are closely based on the elements for the data fields defined for the electronic publication of information on qualifications with an EQF level (or at least reflect these data fields), other databases still need to be further developed in this direction.

Overall, it can be concluded that **the suitability of the databases analysed in the ten countries for comparing qualifications across countries is currently rather limited**.

2.5 Emerging issues and conclusions

Reference documents

Analysis of qualification-related documentation shows that learning outcomes are structured and expressed in a wide variety of ways, which raises challenges for the comparison of qualifications. In relation to different types of learning outcomes (general, occupationally specific, transversal) these are seldom identified separately. Indeed, even within countries there may be mixed approach so that some modules of qualifications integrate, say, occupationally specific and transversal learning outcomes whereas there also occupationally specific

modules without any transversal outcomes and vice versa. Such variation can also be found between qualifications in the same country, where national regulations do not cover this dimension and where different groups responsible for the writing of learning outcomes have adopted different approaches. Similarly, when it comes to the domains of learning (knowledge, skills and competences) these are not often separately specified, and, when they are used for structuring, different approaches are used across countries. Learning outcomes are also structured in terms of different levels of specificity so that it is quite common to find an "upper" level of learning outcomes written in a more general way with more detailed learning outcomes "below". Any procedure for comparing qualifications will need to be able to cope with this diversity in the absence of a common format between countries for presenting learning outcomes.

Variation also exists related to the degree of autonomy within systems to adapt elements of qualifications, e.g. through optional modules and local autonomy to adapt learning outcomes to local labour market needs. This variation means that there will be differences between individuals holding the same certificate for the same qualification in terms of the learning outcomes they have sought through the qualification. In some cases, such as Ireland, these differences may be quite substantial. This suggests that any method for comparing qualifications will need to determine the core and optional elements of qualifications, and, more than this, it may need to acknowledge that in certain circumstance it may make sense to compare only the 'core' elements of a qualification.

Databases

First of all, the analysis of the qualifications databases, available in the countries covered by this study, clearly revealed that there is a need for a commonly agreed definition of what can actually be considered as 'qualifications database' or what is a 'data field'. Up to now, manifold developments in national qualifications database development can be observed with huge variations across countries. Moreover, the common principles for presenting learning-outcomes-based qualifications in databases suggested by Cedefop (2017) are rarely used, the elements for data fields for the electronic publication of information on qualifications with an EQF level as presented in the EQF Recommendation are only used in a few cases and only few databases are linked to European portal. In general, the databases analysed in the ten countries covered by this study support the cross-country comparison of qualifications, and particularly the used of digital tools for this purpose only to a very limited extent. This needs to be taken into account in future discussions on what the purpose of qualifications databases should actually be. The participants at the Budapest

PLA on qualification databases (May 2019), for example, ‘considered that the main focus of databases should be on transparency of learning outcomes, more than on the comparability of qualifications per se’ (European Commission et al., 2019, p. 7).

Chapter 3. Digital technologies and their potential for supporting automated gathering, structuring and analysing of data on qualifications

3.1 Introduction

This part of the report focuses on the analysis of existing and emerging digital technologies, their potential contribution to a proposed workflow in supporting the gathering and processing of national data on qualifications (including providers offering them), as well as their potential use in the comparison of qualifications and on the requirements for their application in this context. Specific attention is being paid to the use of ESCO. Thus, the key research questions are:

(a) How can digital technologies support automated gathering, structuring (including cleaning, fusion) and analysis of data on qualifications?
(b) How can new digital technologies address the linguistic challenges involved in comparing qualifications?
(c) What can be the role of the multilingual classification ESCO in supporting gathering, structuring and classification of qualifications data?

The overall goal of the study at hand **is to design and test an automated workflow (prototype) for comparing detailed overviews of learning outcomes by qualification (longer texts displaying different structures) with preselected 'reference points' (those occupational skills profiles displaying strongest resemblance with the qualification's learning outcomes profile).** Thus, with regard to ESCO, a qualification's set of learning outcomes would be matched with a very small, preselected set from ESCO's KSCs pillar only; namely, exclusively those KSCs highlighted as required or optional for carrying out this particular occupation, ESCO occupational (skills) profiles (OSP).

This automated workflow would support the analysis and comparison of qualifications in Europe. Considering most countries have a large number of competent authorities and awarding bodies engaged with qualifications and the fact that the vast number of educational programmes often correspond to more than one qualification, it is difficult to estimate the total number of qualifications currently existing across Europe that *could* be linked and compared by making use of the ESCO OSPs. A recent study provides some insights: 'One studied

Member State (population ~5 Million) revealed 55 competent authorities/awarding bodies engaged to address 11,000 qualification/certificate-providing programmes, each having multiple qualifications with several learning outcomes. Across all MS there would be significantly greater complexity and heterogeneity' (DG EMPL, 2019, p. 59).

In terms of the potential saving in processing time of an automated workflow, it was estimated that mapping a qualification to a reference point takes between half a day and one day – including the consultation with an expert. Thus, doing so manually already require a large amount of time and human capacity for each qualification in one country would – let alone for all qualifications across Europe. In this regard, the proposed workflow aims to reduce the time and human capacity needed for analysing and comparing qualifications across Europe.

3.2 Conditions of the workflow

Exploring possibilities did not reveal any direct fit-for-purpose digital solutions readily available given the complexity of the task at hand. Rather, one would have to be developed from scratch, building upon already available tools, preferably freeware. It is however possible to take inspiration from existing vacancy-analysis systems ([108]) that focus on comparing 'requirement texts' from vacancies that incorporate the use of ontologies, thus identifying knowledge, skills and competences (KSCs) or occupations mentioned therein.

Some of the desired conditions for this automated workflow for comparing qualifications are described below:

(a) **Ability to process different text formats (prior conversion or automated)**: Member States use different (file) formats to store the learning outcomes of qualifications in reference documents. The automated workflow for comparison must be able to deal with this diversity of formats.

(b) **Allow extraction or labelling of key words** signalling the vertical dimension of learning outcomes (i.e. levels of complexity). The level of complexity specified in individual learning outcomes of otherwise largely identical qualifications may vary between countries.

(c) **Ability to deal with natural language (text parsing):** Words need to be assessed in their linguistic context to derive meaning. The sources for the comparison will be national reference documents for qualifications. They

([108]) https://www.cedefop.europa.eu/en/data-visualisations/skills-online-vacancies/ [accessed 30-08-2019]

describe learning outcomes in national languages, coding the meaning of a sentence or phrase by a combination of words (i.e. use of action verbs, object of the verb, etc.).

(d) **Ability to process different languages**: National reference documents are mostly drawn up in national languages, learning outcomes are hardly available in English (see Chapter 2). The automated workflow needs to process these national languages.

(e) **Use Open Access software packages to maximise inclusivity**: There is not yet a fit-for-purpose system available that is able to automatically conduct the comparison of learning outcomes of qualifications. There are however many tools, methods, packages and codes that were developed for different purposes, which could be applied at least for parts of the workflow (e.g. commercial tools to automatically compare job requirements from vacancies with CVs). Besides being economical, using Open Access software has two additional advantages:

   (i) Open Access software has a large user-base (forums) which can be exploited to solve specific application problems. Furthermore, depending on its distribution platform (such as GitHub), open source software often comes paired with a highly active community of (tech) experts and hobbyists providing additional ideas and feedback on modules or packages (i.e. continuous development of ways to perform tasks that are not yet available in the 'general' programmes used);

   (ii) Open Access software enables easier development, testing and application after project completion (this is important with regard to the scalability of the methods developed).

(f) **Operation of the automated workflow is not too demanding**: After being set up by experts, the tool/device should be operable by the 'layman'.

3.2.1 Necessary conditions of the reference system(s)

An automated comparison between the occupational skills profile of any reference system and a qualification's learning outcomes description will render better results, if the reference point:

(a) Lists all essential and optional KSCs comprehensively (it even makes those KSC explicit, that can normally only be assumed by reading the learning outcome descriptions together with additional context information) and to an appropriate level of detail ([109]);

([109]) As stated in the report for WA1 of this study, this aspect can be used to provide insights into the match between reference point and national qualifications: 'An

(b) Highlights occupation-specific as well as transversal components;
(c) Structures the learning outcomes within the profile into a hierarchy (for example into broad areas of knowledge or competence);
(d) Indicates the level of complexity associated with individual KSCs.

3.2.2 Some notes on the terminology

As the development of the digital tool involves programming, which comes with its own extensive vocabulary of technical terms, it would be difficult to describe the benefits and limitations of the tools used in the protype related to the workflow without using some of the more common terms. The following paragraphs introduce the main terms that will be used throughout this chapter.

In general, regardless of the coding language used when programming, the development of any workflow is done through writing 'scripts', consisting of lines of code that tell the programme which files to address and which functions to perform. These 'command scripts' are written and executed ('run') by the 'user' (or developer) through the selected programme's 'command script', which will interpret the script written by the user and return any output(s) requested. For the scope of this study, the term 'user' refers to the authors of this study (unless preceded by 'any'). 'Running script' refers to the script that is in development.

Furthermore, as this study aims to incorporate existing solutions into the workflow to perform as many steps of the workflow as possible without manual supervision, additional tools need to be 'imported' into the base programme. These additional tools are usually referred to as 'packages' or 'libraries' and consist of collections of multiple pre-coded scripts which can be used as a 'function' in the development of any new script. However, as 'libraries' can also refer to dictionaries – lists of words used in text analysis – we will refer to any specific tools used in the prototype with their given name (e.g. NLTK, pandas), followed by 'package' (i.e. NLTK package).

indication of how well a reference point is able to reflect the learning outcomes of a national qualification is whether all learning outcomes of the qualification are represented in the reference point (reference point is comprehensive). A second aspect is whether the reference point does not exceed too much the learning outcomes of national qualifications (reference point is relevant).'

3.3 Workflow steps

In our understanding, the workflow for automated comparison of learning outcomes (of VET qualifications), contains several sub-tasks, in a more or less chronological sequence (i.e. pipeline). These sub-tasks or steps include:

(a) Provide access to national qualifications in machine readable form; pre-processing of reference point(s), reference systems and national qualifications;

(b) Parse the learning outcomes of national qualification descriptions (text segmentation and POS tagging ('chunking') of learning outcomes' descriptions);

(c) Normalise detected text segments for each national qualification by mapping it onto the reference points' or system's vocabulary (compare to preferred as well as non-preferred terms) ([110]):

  (i) Full matches between learning outcomes and terms of the reference point or system → no human action required;

  (ii) Fuzzy / multiple matches between learning outcomes and terms of the reference point or system → human action required: choose most appropriate KSC concept from suggestions, or add free text (if no suggestion is considered appropriate);

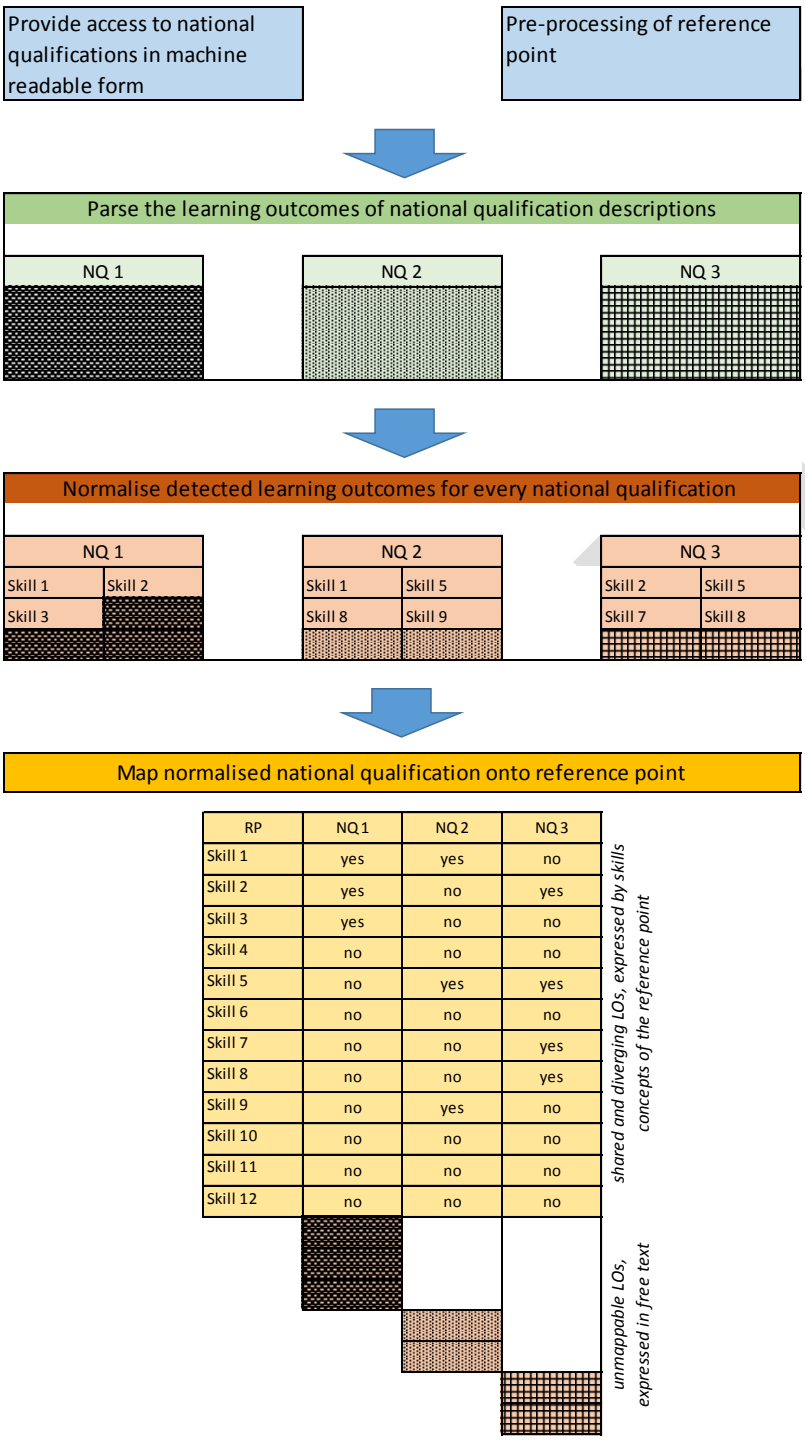
  (iii) Unmappable learning outcomes → human action required: check whether there really is no suitable KSC contained in the reference point and, if yes, add free text to express learning outcomes (should be forwarded to reference system's maintenance team); if no, manually link to suitable KSC.

(d) Mapping of most suitable occupational skills profile (OSP, also referred to as 'reference point') with normalised national qualifications, registering overlap and divergence.

This is illustrated in the following chart:

([110]) Outcomes of this processing step are the (detected/identified) learning outcomes descriptions of national qualifications, in terms of the reference point or system (plus unmappable learning outcomes as free text).

Figure 7 **Overview of sub-tasks in the automated workflow for comparing qualifications**



Source: Authors.

The individual steps of the workflow are further discussed below.

3.3.1 Step 1: Provide access to national qualifications in machine readable form; pre-processing of national qualifications (and reference point for comparison)

Before conducting any text analysis, rigorous data processing will have to be conducted. Firstly, depending on the tools to be used in the workflow, the gathered data will need to be streamlined into a readable format (i.e. plain text or simple tabular formats) that can be fed into a pipeline of tasks. PDF files, for example, are often unreadable for digital tools, due to the use of special 'fields' throughout the text, such as graphs and images ([111]).

After ensuring the data can be fed into the pipeline, further pre-processing steps are necessary to prepare the actual texts for analysis. Besides cleaning the text of punctuations, capitalisation, special symbols, hyphenation, additional whitespace, this involves for instance ([112]):

(a) **Tokenisation:** Break up text stream into meaningful elements (e.g. words, phrases);
(b) **Lemmatisation / Stemming**: Reduce words to their basic morphological form (e.g. conjugated verbs to their infinitive form or their word stem);
(c) **Stop-word elimination**: Removal of all words having minimal informative power in quantitative evaluation (e.g. frequently occurring words like articles, certain prepositions, conjunctions, modal verbs, pro-nouns, etc.);
(d) **Resolution of polysemy/synonymy**: Nowadays, when processing large corpora, dictionaries used for resolving polysemy/synonymy are often generated automatically;
(e) **Part-of-Speech (POS) tagging:** Word category disambiguation to identify word type (e.g. noun, verb, adjective) and sentence function (e.g. subject, predicate, object). Further processing could highlight verbs which potentially reveal autonomy and responsibility involved in carrying out occupational tasks.

Whichever pre-processing tools are chosen, it will likely be necessary to use 'training corpora' to develop a model for the analysis part of the workflow. Although self-development is an option and there are guides online on how to prepare such a training corpus from scratch, there are numerous pre-defined

([111]) This has been a barrier for comparison of apprenticeship standards and national occupational standards in the UK; ways of circumventing the problem are being explored by NESTA and the Gatsby Foundation

([112]) Based on Maheswari, 2017; Salloum et al., 2018; Talib et al.

training corpora available online ([113]). Furthermore, using annotated data (POS-tagged) would lead to more accurate predictions on the pattern matching of descriptions of qualifications to the reference point and between qualifications.

After pre-processing, the resulting tokens can be used as units of analysis (for counts, building term-document matrices, identifying patterns), but may also serve a purpose in increasing the accuracy of the analysis by:

(a) Identifying similar terms to those used in the reference point, allowing for **feature extraction** and selection – i.e. enrichment of the topic list (such as the ESCO skills list) to be used in pattern matching ('topic matching');
(b) Further resolving **polysemy** and **synonymy**;
(c) Allowing for the construction of word combinations or **'n-grams'**, which can be used as a secondary requirement in pattern matching (i.e. only 'counting' terms if a second related term is nearby in the text).

3.3.2 Step 2: Parse the learning outcomes of national qualification descriptions

After Part-Of-Speech Tagging 'chunking', a shallow parsing technique for segmenting and labelling multi-token sequences, could be used to group the words in 'chunks' of entities and their relationships. This means identifying the text segments that should be compared to the vocabulary of the reference point described in Step 3 below.

Alternatively, qualification text could be analysed by several other techniques which make implicit textual information explicit and reveal meaning relations, usually via analysis of frequency distribution and joint occurrence of words. Here below some techniques are briefly explained:

(a) **Cluster analysis** is an unsupervised process to classify text documents into groups, based on the hypothesis that relevant documents must have more

([113]) However, it is important to note that most of these lexical databases and pre-developed functions are aimed at the English language only – for example the 'Porter stemmer', 'Snowball stemmer' and the 'Wordnet' database for stop-word removal. The NLTK corpus 'stopwords', on the other hand, allows for 'defining' which language to use when building a stop-word removal function (See section 3.4.4.3), Regardless, a drawback here is that there are some pre-processing issues that may arise when using these existing resource(s). This is mainly due to the nature of the texts to be processed within the scope of this studies (Qualifications, Skills descriptions ESCO) – as they are generally more occupation-specific than the texts being processed in more commonly seen uses for text analysis (tweets, webpages, movie/book reviews, etc.). This means it may still be necessary to further develop (i.e. train, enrich) these lexical resources to fit for our purpose, and/or to develop our own resources for languages that are currently not included in the NLTK corpus 'stopwords'.

similarities with one another than non-relevant ones. Similar terms or patterns extracted from a collection of documents are grouped in a cluster. This results in sets of tokens, related to the topics in the corpus and facilitates the method of topic analysis. Different techniques can be used to perform the analysis, such as hierarchical, distribution, density, centroid and k-mean clustering (Salloum et al., 2018).

(b) **Co-occurrence analysis** is simply the counting of data 'pairs', or in this case the occurrence of pairs of tokens. This analysis would count the number of times a certain token occurs together with another token from the 'bag of words' and repeats this process for all tokens in the collection or text. These counts are put into a so-called co-occurrence matrix which can then be analysed for similarity (or dissimilarity). When items co-occur, this indicates an association between them. If the paring only occurs once in the collection, this association can be considered spurious, while each additional pairing increases the strength of the association.

(c) This co-occurrence matrix can also be interpreted as a network with each token or element in the collection as a 'node' and the connections as 'links' with a value. This network can be visualised as a vector in an n-by-n grid (n being the number of items or tokens in the collection) or as a **self-organising map (SOM),** also referred to sometimes as a 'Kohonen map' (Buzydlowski, 2015). To train a SOM, the network of nodes is firstly initialised to random values. Then, using the 'best matching unit' approach, a sample row is chosen and compared to each node to determine the row with the closest Euclidian distance (i.e. the physical distance between words in the text). The value of the selected node will be adjusted to become more like the sample vector's (row) and the same is done for neighbouring nodes. This process is repeated until all rows have been adjusted. The visualisation can be done through a colour grid (or 'heat map'), showing lighter and darker shades depending on the average distance between nodes or through assigning different symbols depending on which cluster of nearest neighbours a node belongs to.

3.3.3 Step 3: Normalise detected learning outcomes for every national qualification by mapping it onto the reference system's vocabulary

The previous processing step identified part-of-speech chunks representing meaningful units. This step compares them to the vocabulary of the reference system, in order to identify:

(a) Text strings which equal preferred terms or non-preferred terms of the reference system;

(b) Fuzzy matches, i.e. text string contained in the reference system in a slightly different way; the user in the loop chooses the most suitable match;
(c) Text strings which do not even match in a fuzzy way; these unmappable learning outcomes are recorded 'as they are' ([114]).

The necessary technical/organisational tasks of this step include:
(a) Build a normalisation system for the chosen reference system (a 'classifier' based on machine learning) to translate learning outcomes into skills concepts via mapping onto its preferred and non-preferred terms;
(b) Build an editorial interface for learning outcomes 'fuzzy matches' (text strings resembling several of the reference system's terms), enabling humans to select the most appropriate one; this editorial interface could also be used to attempt an editorial allocation of text strings without any resemblance to any ESCO KSCs (only if also this conceptual mapping can't be done, the text string might be filed as a potentially new concept).
(c) Expand reference system (expanded keywords), or alternatively organise an amendment workflow for the reference point, using input / feedback from the normalisation system: process potentially 'new' skills concepts and their names as well as naming alternatives for already existing skills concepts to improve the reference system's suitability for NLP purposes.

3.3.4 Step 4: Mapping of most suitable occupational skills profile (OSP, 'reference point') with normalised national qualifications, registering overlap and divergence

If the learning outcomes of a qualification could be matched onto one of the skills of the reference point – either automatically (full match) or semi-automatically (fuzzy match + human choice) – a comparison with the respective OSP would identify shared and differing skills.

Text strings which could not be normalised would be recorded as additional 'differing features' or tagged for supplementary manual processing.

The result should be an overview of KSCs which are covered in national qualification data, and which are not. For this purpose, a uniform format will have to be developed:
(a) That is easy to understand and interpret by a layman;

([114]) Even for these 'unmappable' learning outcomes, it would be possible to have humans manually select suitable KSCs, for example from the rest of the ESCO vocabulary (e.g. outside the respective OSP).

(b) That also highlights learning outcomes not accounted for by the reference point or the qualification;
(c) That facilitates further, more fine-grained comparison between different national qualifications (e.g. with respect to level of complexity of individual learning outcomes, or with regard to methods or tools used).

3.4 Operationalisation of the workflow (lessons learnt during initial prototype development)

The approach in terms of development of the prototype (digital tool) consisted of selecting a set of base tools to work with and tailoring solutions provided on discussion forums to fit the workflow through trial and error – identifying bottlenecks as development progressed. Simultaneously, the study team reached out to experts to provide suggestions and feedback on the proposed workflow, as well as on the use and feasibility of incorporating machine learning aspects into it.

3.4.1 Base tools for development

Following the suggestions from text analysis forums and online comparisons between these programmes and their main uses, the decision was taken to explore **Python as the main starting point for the prototype**. Firstly, because Python is often discussed in forums and tutorials related to text processing and text analysis as the more suitable option for beginners/laymen, as the language used for the scripts is easier to write and understand without prior knowledge. Secondly, it was found that there is an open-source tool available (**Anaconda Navigator**) that allows the user to work through a user interface ([115]) which is easier to use than the standard Python command script. Apart from this, Anaconda Navigator provides several additional advantages in terms of ease of use of the prototype, as it allows the user to:

(a) **Search for Python packages in a cloud database and install them** by simply making a selection – whereas in the standard Python installation, each package would have to be installed manually (i.e. through separately 'run' lines of script) and in a specific order, which can be an arduous task as many packages require other packages to be installed first ('dependencies').

([115]) This refers to the visual aspects of the programme that the user interacts with in order to use a programme. For example, the side bars and menu options of any commonly used programme (such as Word, Excel, etc.) can be considered part of a user interface.

Furthermore, the Navigator automatically detects any packages considered dependencies, allowing to include them in the installation, as well as any possible updates to previously installed packages;

(b) **Create mutually exclusive environments**, which means packages would only be installed in a temporary file – addressed only when working on the package – which reduces the capacity needed on the workstation running the Navigator. It also allows the user to work with different versions of packages for separate environments. As not all packages are updated at the same speed, some dependencies may require a specific version of a package to function. This removes the need to manually check online repositories for new versions of packages and their compatibility with existing packages on a regular basis.

(c) **Save the script and run multiple lines or sections at a time**, whereas the standard Python command script would require manual re-entry of the script (i.e. re-typing line for line) and thus keeping a separate script file that has to be edited for each attempted alteration to the script. This reduces the time needed during development of the script, as the development of the code is done through a temporary script that is linked to the project environment, loaded automatically when starting the Python command script.

(d) **Display results in a separate window** – as opposed to results being displayed in the command script, between the lines of code. Being able to view results separately makes it easier to check for anomalies in the results throughout the development process.

The second main resource included in the base set of tools is considered the most-used collection of packages related to working with text data, called the 'Natural Language Toolkit (NLTK)'. Technically, NLTK is a platform for building Python programmes to work with human language data ([116]). It provides access to over 50 corpora and lexical resources, along with a multitude of text processing libraries that can be used for specific elements within the workflow, such as: classification, tokenisation, stemming, POS-tagging, and parsing. The official website provides their own guide to text analysis and maintains an active discussion forum. Additionally, the resources of this toolkit are often referred to when providing coding solutions for text analysis on other forums ([117]).

([116]) For more information see: http://www.nltk.org/

([117]) Within the scope of this study, these are mainly: 'StackOverflow'; Reddit (/r/AskProgramming); and the 'Software Engineering StackExchange'.

However, before (pre-) processing texts through the functions included in the Natural Language Toolkit, the text files themselves will need to be imported into the project environment in a way they can easily be processed. Desk research showed **using tabular data formats (such as Excel) or comma-separated text files (CSV) is recommended** in terms of the file formats for 'input' at the start of the workflow. This is preferred as it would require less work to manually fill out a pre-determined Excel table (XLSX/CSV) based on the reference files, than to identify and extract the information from other formats into the right structure through coding. **Since such a function is not included in the NLTK package, a third main resource was needed in order to work with tabular data in Python.**

In this regard, the online community indicates that the best package to use is **'pandas'**. This package can be used to read CSV or Excel files into so-called **'dataframes'** (Willems, 2019) – keeping the existing structure of the table intact (i.e. columns, headers), while still allowing for the base transformations that are available in their original programmes (i.e. renaming headers; flipping rows to columns and vice versa; adding columns based on functions using existing cells, etc.). Furthermore, this allows for the pre-processing tasks to be performed only on the selected rows and columns (i.e. exclusion of the headers for tokenisation) and allows for loading multiple tables into one (i.e. merging datasets, analysing multiple qualifications).

3.4.2 Subtasks performed during development

In terms of the workflow discussed in section 3.3, development of the prototype has mainly focused on developing a running script for structurally pre-processing the reference point for comparison (i.e. workflow Step 1). After ensuring the necessary subtasks can be performed, this running script should require less tailoring to perform the same process on full qualification texts.

3.4.2.1 *Selecting a test file*

For this base script, an Excel test file was selected that consists of the KSC terms and descriptions (as separate columns), for the first 10 rows of learning outcomes included in the healthcare assistant reference point (resulting in a 10x2 table). **Selecting a small set of data fields as a testing file is preferred in the early stages of coding**, as development goes hand in hand with unexpected results. This way, the results are easier to display between subtasks, and thus easier to check for anomalies. Additionally, this approach tackles working with tabular data from the start while keeping the original structure intact – as opposed to loading the full text of one qualification file into one 'cell' first,

requiring further transformation into a table through coding. The table below represents the KSC terms and descriptions that were included in the test file.

Table 14. **Selection of (10) healthcare assistant learning outcomes as included in the test file**

KSC term	Description
advise on healthcare users' informed consent	Ensure patients/clients are fully informed about the risks and benefits of proposed treatments so they can give informed consent, engaging patients/clients in the process of their care and treatment.
convey medical routine information	Convey routine information to patients, relatives, and members of the public.
identify abnormalities	Identify what is normal and abnormal concerning the well-being of patients, through experience and instruction, reporting to the nurses what is abnormal.
manage healthcare users' data	Keep accurate client records which also satisfy legal and professional standards and ethical obligations in order to facilitate client management, ensuring that all clients' data (including verbal, written and electronic) are treated confidentially.
monitor basic patients signs	Monitor basic patient vital signs and other signs, taking actions as indicated by the nurse and report to her/him as appropriate.
use e-health and mobile health technologies	Use mobile health technologies and e-health (online applications and services) in order to enhance the provided healthcare.
evaluate older adults' ability to take care of themselves	Assess the condition of an older patient and decide if he or she needs assistance in taking care of him- or herself to eat or to bathe and in meeting his/hers social and psychological needs.
disability types	The nature and types of disabilities affecting the human beings such as physical, cognitive, mental, sensory, emotional or developmental and the specific needs and access requirements of disabled people.
geriatrics	Geriatrics is a medical specialty mentioned in the EU Directive 2005/36/EC.
older adults' needs	The physical, mental, and social needs of frail, older adults.

Source: Extracted from the ESCO Occupational Profile for Healthcare Assistant.

3.4.2.2 *Setting up the workstation*

Before the data can be processed by the Spyder interpreter (available through Anaconda Navigator), some base code is needed to set up the project environment: to determine the location where the project files are to be loaded from and exported to, as well as the alphabet the programme is to use for reading text characters ([118]). Additionally, **it is recommended to perform the**

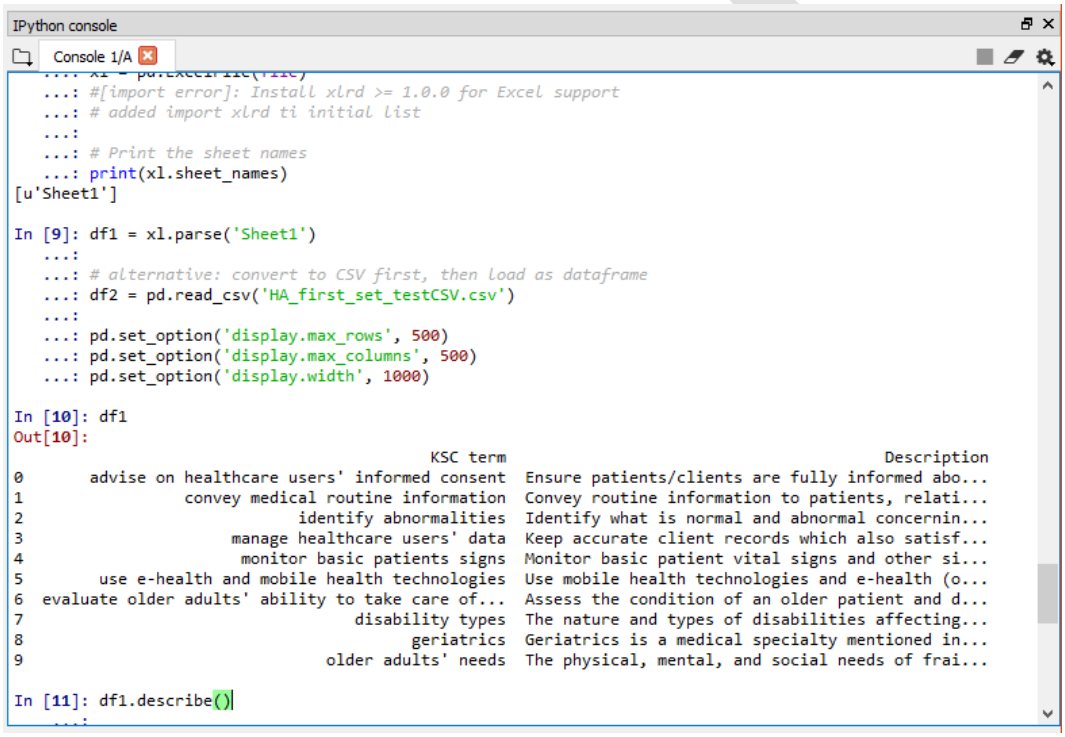
([118]) The main ones being: The American Standard Code for Information Interchange (ASCII) and the Unicode Transformation Format (UTF). As the 'language' used in

installation ('importation') of all packages and libraries that are likely to be incorporated within this early stage. This allows for early identification of dependencies – meaning additional packages that are required to run the ones selected for the prototype – as attempting to run (sections of) the script without installing them will result in errors specifying which function/package is missing ('undefined') ([119]).

3.4.2.3 *From test file to dataframe*

In order to read Excel files into the project environment, functions from the pandas package were used, resulting in a dataframe that keeps the existing structure of the Excel table intact (i.e. columns, headers). The figure below shows the output as displayed in the Spyder interpreter after running the script to display the dataframe (df1).

Figure 8. **Dataframe as seen in Spyder, presenting the (10) selected learning**



files is not easily identified and often unspecified, this may present readability or conversion issues when using the script on text files from other languages, especially those that use symbols for characters (Chinese, Arabic, etc.) or languages using more characters in their alphabet than the commonly used ABC alphabet does.

([119]) In our case, this led to the inclusion of the following dependencies: xlrd, numpy, openpyxl, xlsxwriter.

outcomes

Source: User installation by Authors, installation files available via https://www.anaconda.com/distribution/

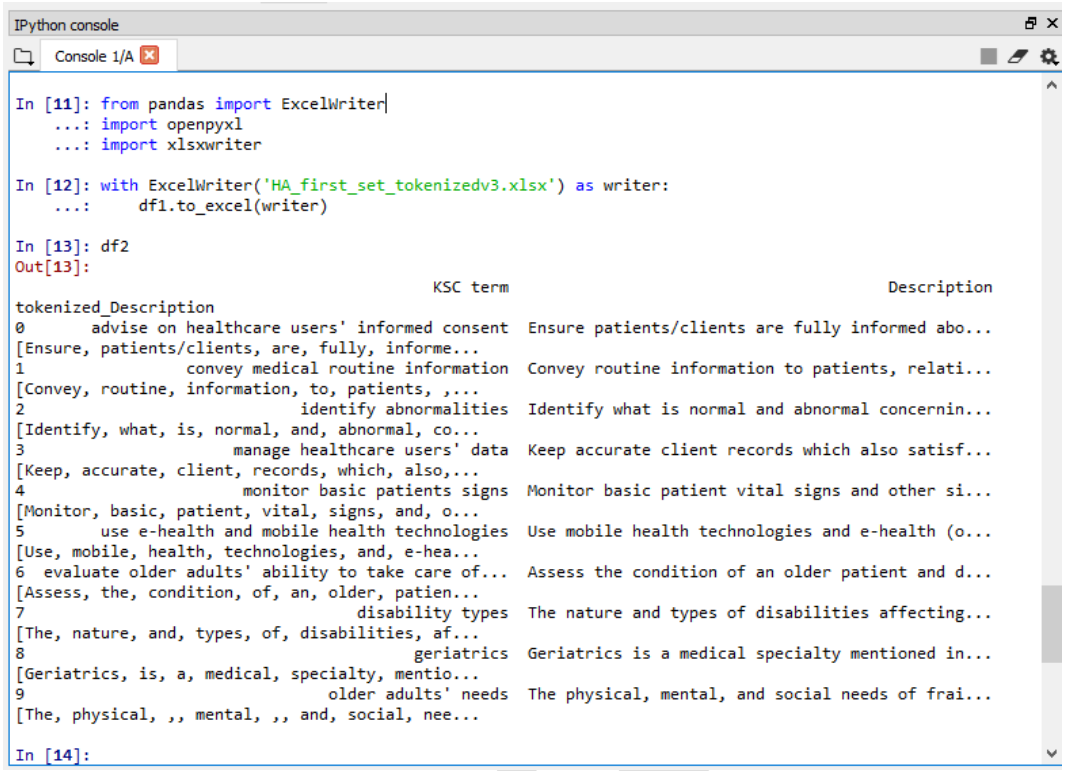
Note that the viewing window still poses limitations in terms of visibility of the results, depending on the length of the text being processed – especially for the column 'Descriptions'. However, as this is also the case when working in the regular Python command script (where results are displayed between lines of code), this issue is not linked to the use of this specific programme. The workaround identified was to export the dataframe back into an Excel file between subtasks, for manual inspection of outliers or non-functioning solutions (i.e. tasks not being performed, without declaring an error in the programme itself) ([120]).

3.4.2.4 *Tokenisation of the descriptions*

The next phase in pre-processing the data was to combine functions provided in the pandas and NLTK packages and perform the tokenisation of one column of learning outcomes (the descriptions) – while making sure to save the newly created tokens as a new column into the existing dataframe. This is done to keep the original column intact (i.e. not corrupting the original data), the ability to retrace from which learning outcome a single token originated and for manual detection of anomalies in the results. The following figure shows the results as they are displayed after performing tokenisation on the test file's descriptions.

Figure 9. **Dataframe with tokenised descriptions as 'new column'**

([120]) One additional issue is that when saving the dataframe back to an Excel file for inspection, the tokens or bigrams are exported with added text – in which each token is presented as (u'Token') and each bigram as (u'Token1', u'Token') so long as they are contained within one data row, as 'lists' (see also a related issue in section 3.4.4.2 on 'unnesting'). Even though the functions needed for the analysis automatically resolved this issue for single tokens, this was not the case for bigrams. Thus, in order to present the resulting data on bigrams it would be necessary to manually edit the resulting texts.



Source: User installation by Authors, installation files available via https://www.anaconda.com/distribution/.
Note: Due to the limitations in window size, it is difficult to distinguish between the first two columns (the tokenized descriptions and the KSC terms).

The newly created tokens are displayed in the left-most column ('tokenised_Description'), more specifically the lines preceded by '['. As seen in the figure before, however, the displayed results do not show the complete texts and the columns are cluttered to some extent. Manual inspection revealed that they are correctly processed in the dataframe itself, just not displayed properly due to limitations in window size (indicated by …).

After performing tokenisation, the development would move on to subtasks related to cleaning the tokens in order to reduce the number of unique tokens identified (by removing punctuation marks, English stop words, resolving polysemy / synonymy, etc.), during which it would be expected to start incorporating or developing 'dictionaries' – thus moving towards machine learning elements of the prototype. Thus, at this stage, development was put on hold while experts ([121]) were consulted, in order to determine possible approaches and provide some insights on the feasibility of our prototype.

([121]) With representatives from Cedefop's RTLMI project (Jiří Braňka) and Textkernel (Panos Alexopoulos and Jakub Zavrel).

3.4.3 Expert consultations – issues regarding feasibility

Through the expert consultations ([122]), it was determined that there were several issues regarding the scope of this project. These issues correspond to three overarching lessons learned, regarding the feasibility of developing a tool for *automated* gathering, structuring and analysing data on qualifications (as well as for the comparison with ESCO data). These lessons resulted in a change of approach regarding the intended purpose of the tool in development – more specifically to **focus on delivering a tool to support** the data gathering, structuring and analysis, **rather than one automating the process fully**. Section 3.4.4 discusses these three lessons, as well as the resulting implications for the initial prototype that led to an adapted approach – of which the change in approach and further development performed will be discussed further on, in Chapter 4.

3.4.4 Taking stock: Mid-term stocktaking on development, lessons learned and changes to the approach

3.4.4.1 *Lesson 1: Tailoring existing examples considerably slows down the development process*

Although it was expected to run into minor issues during development, some were easier to resolve than others. Minor issues were mostly related to: (1) packages or their functions not working due to dependencies on other packages; (2) results not being displayed properly, thus requiring manual inspection (by exporting the data back to Excel) at regular intervals; (3) the difficulties in identifying coding examples that are suitable for tailoring to our project.

Whereas the first two types of issues were relatively easy to resolve – including dependent packages and developing a small script for exporting data back into Excel – the third type was mainly responsible for **the development progressing at a slow pace**. Although generally, a multitude of examples can be identified on any of the sub-tasks mentioned above (5-10 coding examples on average), most of the examples tested were not found suitable/tailorable for our intended purpose on their own. This is mainly due to differences in:

(a) The type(s) of data examples were provided for – most examples using the same resources (especially those regarding dataframes) were for numerical data instead, or a mix of text and numerical;

([122]) With representatives from Cedefop's RTLMI project (Jiří Braňka) and Textkernel (Panos Alexopoulos and Jakub Zavrel).

(b) The structure in which data was to be introduced into the function (input structure);
(c) The desired structure in which they were to be delivered after processing (output structure).

Following these issues, **most of the subtasks performed for the prototype thus required 'piecing together' a workaround**, ultimately incorporating coding 'tidbits' from multiple approaches. Overall, this tailoring process showed that, even though there are many ways to approach any given task with different available resources, it also showed that every function (existing, tailored or self-developed) generally has a set of requirements in terms of which datatypes are readable and/or writable – and thus analysable. To some extent, this can be taken into account by setting requirements on how the data to be processed would need to be delivered into the prototype at the start. However, given that **each subtask performed could potentially change the way data is stored for further processing**, the tailoring process is especially important in making sure not to end up with a datatype that is 'unreadable' for later functions ([123]).

3.4.4.2 *Lesson 2: Pre-processing the texts for proper feature extraction requires more resources in terms of human capacity than anticipated*

When comparing our approach to a related project working with more labour market-specific texts (Cedefop's RTLMI project), it became clear they started out with a set of skills provided by ESCO as well. In that case, it was found that the ESCO vocabulary (as it is) is phrased quite differently from the vacancy market – resulting in only a limited number of KSCs matching. Therefore, during the development process, the University of Milan generated alternate labels (from vacancy data), in order to provide better access to ESCO concepts for NLP purposes – meaning they had to develop their own lexical database to yield accurate results in linking skills descriptions to ESCO KSCs. Through this studies' prototype development and the expert consultations, the study team came to the same conclusion as the University of Milan and TextKernel – that **much more processing was needed beforehand (especially the generation of large synonym dictionaries to improve ESCO's suitability for NLP) to improve data processing**. This aspect is relevant to comparing qualifications as

([123]) This was found to be the case at several stages in development, especially in first attempts to construct word pairs (bigrams) instead of single tokens. The details of these issues are further discussed in their respective sections (

well, as the way skills are phrased in national documents / vacancy data is not always at the same level of specificity as in ESCO. Furthermore, ESCO's KSCs pillar shows massive conceptual overlap, which complicates the mapping of National KSCs' taxonomies onto ESCO. Another concurring issue may be that the lemmatisation causes:

(a) Misclassifications - when trying to classify two similar (but considered separate) skills in ESCO);
(b) Disambiguation - one lemma pointing at too many ESCO skills, expressing almost the same content.

Another point addressed in the consultations (that relates to this issue in particular) was that **major work on the ESCO classification would be needed before it can be used in a satisfactory manner for automated text processing,** e.g.:

(a) Enrich vocabulary – with stemmed versions of skills phrases; by generating skills phrases from existing texts (vacancies, for example); or through incorporating such resources from existing projects that process such texts ([124]);
(b) Supplement semantic structure to enable aggregations;
(c) Dissect complex skills into enabling skills components to make implicit components visible and to gain insight into performance levels;
(d) Consolidate terminology (e.g. summarise skills expressing the same meaning with different words under one concept),

3.4.4.3 *Lesson 3: Training an automatic classifier for our project would require more manually pre-processed 'training documents' than we would ultimately aim to classify after completion*

Furthermore, it became apparent from the expert consultations that **the scalability of the method needed to be taken into account as well**. Incorporating machine learning aspects (in this case a classification model) into any project requires a large number of manually pre-processed and tagged 'training documents' as well as a large amount of human resources for evaluating the outcome of machine learning (e.g. identifying outliers and for improving the model itself).

Considering that an automated comparison of national VET qualifications can only render results of suitable quality if the basis of comparison – individual

([124]) Textkernel, for example, generates skills phrases from millions of vacancies, thus creating a vocabulary bottom-up. Unfortunately, this vocabulary cannot be structured by machines alone and it would require human intelligence to connect terms.

national qualifications – has been processed comprehensively and accurately, it **can be assumed that humans will also need to verify the results of automated tools. Thus, even once all technical issues have been resolved, this automatic classifier would have the status of a suggestion tool rather than an alternative to human processing.** Additionally, this would become even more of an issue when attempting to process multiple languages since it would require training a separate model for each language involved. Considering that - even when looking at all qualifications within all countries selected – 'only' a few thousand qualifications would be available and even then they would still vary in language used and/or structure of the documents to process. This would result in an unrealistic level of capacity requirements for this project (e.g. even if we have only around 10,000 qualifications to work with).

3.4.4.4 *Implications for the prototype*

Following the issues raised during the prototype development and expert consultations, **it became apparent that the effectiveness of using machine learning for the intended project really depends on the amount of data being processed after completion**, and how much it would reduce the workload. It is most effective when one is likely to process *even more* data *after* finishing the training of a model, as the development of a classification model (for POS-tagging, resolving polysemy / synonymy, etc.) already requires a substantial amount of resources.

Since the number of qualifications to be processed – at least within the scope of this study – is far too low to incorporate machine learning aspects into the tool (and thus the prototype for testing), it was necessary to review the current approach and determine an alternative in which a prototype could still be developed for testing, within the context of comparing data on qualifications. For this prototype, further discussed below (sections 3.4.5.1 until 3.4.5.6), it was agreed to **explore possibilities for technically supporting human processing,** rather than building a fully automated qualification mapper.

3.4.5 Subtasks performed during further development of the prototype

The adapted approach is still focused on extracting learning outcomes from qualifications in order to compare them to ESCO's KSC pillar – under the assumptions:

(a) That the information to compare with will be structured prior to automated processing (i.e. be delivered in a specific (file) format and structure); and

(b) Instead of using separate preselected (ESCO) reference points for each qualification, the prototype will use the full list of KSCs – a total of 13,485 skill descriptions.

This would allow for identifying patterns within qualifications, comparison with the ESCO classification, as well as for analysis of the ESCO classification itself – for example by identifying terms (or groups of terms (bigrams)) that may cause the classification issues discussed before (misclassification, disambiguation). This section will discuss the new files used for the prototype and the additional subtasks performed during development.

3.4.5.1 *Selection of the alternative files*

The prototype builds on the steps performed in development so far (as discussed in 3.4.2), since **the code was developed in such a way that one can simply indicate a different file to be selected at the start, with minor tailoring of the code where needed** (column names, variable names etc.). As the list of ESCO KSCs (and their respective unique identifiers) was already provided in Excel ([125]), no structuring was necessary beforehand for the ESCO test file. For the qualification test file, 11 core tasks and work processes (further referred to as 'CT/WP') identified within the Dutch qualification for ICT service technician were extracted from the official documentation (PDF), translated and stored in an Excel database. Like the approach at the start, only the shortest descriptions / labels were used for either test file ([126]). The following table provides the CT/WP of the new qualification test file, with their official (Dutch) identifiers as 'ID'.

Table 15. **ICT service technician core tasks and work processes (NL and ENG) and unique identifier, as included in the test file.**

ID	CT/WP description (NL)	CT/WP description (ENG)
B1-K1	**Installeren en onderhouden van hardware, software en verbindingen**	**Installing and maintaining hardware, software and connections**

([125]) This was done already in WA1.

([126]) After finalising the prototype, tailoring it to use the more extensive descriptions would merely be a matter of specifying a different data column (and manual extraction only for the qualification test file) to perform the pre-processing and analyses tasks on a larger set of terms.

B1-K1-W1	Gebruiksklaar maken van systemen, (rand)apparatuur en applicaties	Making systems, (peripheral) equipment and applications ready for use
B1-K1-W2	Vervangen, repareren en/of (de)monteren van (onderdelen van) systemen en (rand)apparatuur	Replacement, repair and / or (dis) assembly of (parts of) systems and (peripheral) equipment
B1-K1-W3	Realiseren van verbindingen	Realize connections
B1-K2	**Behandelen van incidentmeldingen**	**Handling of incident reports**
B1-K2-W1	In behandeling nemen en registreren van incidenten	Handling and registering incidents
B1-K2-W2	Oplossen en/of escaleren van incidenten	Resolving and / or escalating incidents
P2-K1	**Ondersteunen van gebruikers**	**Supporting users**
P2-K1-W1	Opstellen van instructies	Preparation of instructions
P2-K1-W2	Mondeling toelichting geven aan gebruikers	Give users oral explanations

Source: Excel database, based on the information from the Dutch qualification file for ICT service technician ([127]).

As the programme used for development allows for **working with multiple scripts at the same time**, the steps completed up until now could be performed on both files side-by-side (ESCO and qualification), thus accelerating the process of bringing them to the point where initial development had stopped (at tokenisation). After this, some further pre-processing was necessary before moving onto the testing/analysis, as discussed below.

3.4.5.2 *'Unnesting' tokens*

In the current dataframes, the tokens included for each skill are still 'nested' within one cell – meaning they are considered as a list by the programme, and thus the respective column as a collection of lists. However, since most of the functions to be performed for text cleaning and further analysis are designed to use strings (equivalent for text), they were found to be unable to read these lists. Solving this issue required a method referred to as 'unnesting', which essentially:

(a) Picks a row in the dataframe;

(b) Extracts the individual tokens from the corresponding list;

([127]) 'Medewerker_beheer_ICT (25191).pdf', obtained from the SBB registry of Dutch VET qualifications at https://kwalificaties.s-bb.nl/Details/Index/2571?type=kwalificatie&item_id=957853&returnUrl=%2F%3FTrefwoorden%3D25191

(c) Duplicates the row for each token.

This resulted in **an expanded dataframe, where each individual token has its own data row – thus allowing the tokens to be extracted into functions** – with the original data in other columns copied, in order to prevent data loss. This expansion resulted in a total of 48,859 tokens for the ESCO file (contained within 13,485 skills), and a total of 41 tokens for the qualification file (contained within 11 skills) to be further pre-processed for testing.

3.4.5.3 *Cleaning text / tokens*

Initially, the cleaning up of text was performed after tokenisation, based on examples seen during desk research. When using the running script developed up until now, however, manual inspection of the resulting dataframes indicated that some of the pre-processing steps were not being performed as desired – resulting in tokens not being matched properly (for example: incorrectly considering 'Installing' and 'installing' as two separate tokens). This issue was solved by restructuring the existing code to perform the separate cleaning functions before moving on to tokenisation and unnesting, and by identifying texts to adapt in order prevent data loss ('2D', for example, when removing numbers). Aside from this restructuring, the necessary text cleaning steps were performed as a series of small functions (lambda functions) ([128]), tailoring existing functions from the string and corpus packages (within NLTK) to fit our data. These cleaning steps included:

(a) Lowercasing all text;
(b) Removing any digits;
(c) Removing special characters, including punctuation marks ([129]);
(d) Stop word removal ([130]).

([128]) Python allows you to create **anonymous functions** (i.e. function having no names) using a facility called lambda function. Lambda functions are small functions, usually not more than a line. They are generally used when you need a function for a short period of time, when you want to pass a function as an argument to higher-order functions, that is, **functions that take other functions as their arguments.** In the development of the prototypes, lambda functions were also used for the 'cleaning functions' aimed at lowercasing and stop word removal.

([129]) More specifically, the following characters: ! " # $ % & ' ( ) * + , - . / : ; < = > ? @ [\ ] ^ _ ` { | } ~

([130]) For the scope of this prototype, only the stop words for the English language were necessary. The NLTK resource called 'stopwords' (callable through NLTK's 'corpus' package) does contain **lists of stop words for a total of 21 languages**: Arabic, Azerbaijani, Danish, Dutch, English, Finnish, French, German, Greek, Hungarian, Indonesian, Italian, Kazakh, Nepali, Norwegian, Portuguese, Romanian, Russian,

To illustrate to what extent this cleans up the data being processed: Manual inspection of the resulting dataframe showed that this had reduced the total number of tokens identified from 48,859 to 42,873 (for the ESCO file) – thus resolving 5,986 incorrectly identified tokens ([131]). It is important to note that this is *not* the number of unique tokens/terms within ESCO, as this will be determined in the testing phase, through frequency analysis (i.e. Chapter 4).

There are likely more terms to take into account that can cause incorrect token identification within the ESCO file, such as various field-specific terms, concepts or abbreviations. Especially the names of software programmes or programming languages are of interest in this regard, as they often use capitalised spelling and incorporate numbers or special symbols. To illustrate, some identified examples of potentially problematic terms, currently existing within ESCO were 'N1QL' ([132]), 'OHSAS 18001' ([133]), and C++ ([134]). The cleaning process would affect these terms as follows:

(a) The term 'N1QL' would be returned as two separate tokens ('n' and 'ql');

Slovene, Spanish, Swedish, Tajik and Turkish. *However:* the information file corresponding to this resource **does not specify whether the language titles also function as the 'keys' to be used for each country** (in terms of programming). So, it is not clear if one can retrieve the list of stop words for the other languages by simply replacing the current key for the stop word removal function ('English'), for example with 'Slovene'. Additionally, **the lists for languages other than English may be underdeveloped, as they are mainly based on third-party resources** - thus not specifically developed for the NLTK stop words package. Additionally, these third parties may no longer (actively) update the respective resources and lose accuracy over time (as new stop words naturally emerge in languages over time).

([131]) 'Incorrectly identified', as they would either be mismatched during analysis (due to differences in spelling and use of upper/lowercase), or they would be filtered out at a later stage (when identifying outliers, for example) – leaving 'empty' tokens to be filtered from the dataframe. Due to the change in approach, this final 'filtering' subtask was in fact no longer needed for the scope of this study. Nevertheless, the finalized script will still include the respective function for removing empty tokens from the database – in a separate section for developed (working) functions that became obsolete.

([132]) Preferred label in ESCO for skill (ID): http://data.europa.eu/esco/skill/f597f772-24d3-4cec-813c-cf5a7027c794

([133]) Full label was 'adhere to OHSAS 18001', skill (ID): http://data.europa.eu/esco/skill/2d9aaad3-a3c8-4c5b-bf67-46984d860873

([134]) Which can be found in two separate skills – labelled 'C++' and 'Microsoft Visual C++', where the (implied but not explicit) difference is that 'C++' is referring to the programming language itself, whereas the other refers to an official software programme *using* C++.

(b) The term 'OHSAS' would returned as one token only ('ohsas') – thus losing the numerical part, which refers to a specific edition of standards ([135]);
(c) The term 'C++' would be returned as a single, one-character token ('c').

Therefore, the exact number of unique tokens included in the full list of ESCO KSCs will likely be even lower, if such terms were dealt with beforehand. However, manual inspection of the results for the qualification files indicated that within the scope of this prototype no such terms are included in the ICT Technician's qualification file, and thus, such tokens would automatically be excluded through the matching process. Therefore, no further adaptations were made to the texts (from the ESCO) file prior to testing.

3.4.5.4 *Constructing word pairs - bigrams*

After developing the code for extracting single terms / tokens from both test files while ensuring their texts are 'cleaned', the development process moved onto the extraction of **pairs of words** (bigrams). This was done, considering it is to be expected that using a direct matching approach on single tokens only for analysis would not incorporate enough context to prevent undesired matches (i.e. entire unrelated skills matching only due to that one term). In order to construct bigrams in the prototype, an existing function from the NLTK package was used, called 'ngrams' ([136]). This function builds the bigrams through a series of subtasks:
(a) Calling on the information from the column of already tokenized terms (prior to unnesting);
(b) Linking two words - from the list of tokens - into a new series of lists ([137]);
(c) Storing the results back into the original dataframe as a new column.

As this function did not work on the expanded dataset, this step needed to be performed prior to the unnesting of the tokens. This way, the bigrams are automatically duplicated for each respective token within the same ID, further enriching the data to be used for comparison between files. This resulted in a dataframe ready for further analysis (for both respective test files), each containing the ID's, (unnested) single tokens, and bigrams included within the

([135]) Referring to 'Occupational Health and Safety Management Systems' (OHSAS)

([136]) This is the overarching term used in relation to text processing/analysis, **where 'n' refers to the number of words/terms to 'link' to each othe**r. Thus, this function could be adapted/repeated in further development for building sets of 3 (trigrams) or more words as well – merely requiring the user to specify the number of words to link (n) in the new/adapted function.

([137]) Based on their original position in the sentence (i.e. [Token1 + Token2]; [Token2 + Token3]; etc.).

respective skill or CT/WP descriptions. To illustrate, the resulting dataframe for one of the qualification's core tasks (B1-K1; 'Installing and maintaining hardware, software and connections') is presented in the table below.

Table 16. **Bigrams, constructed based on the selected core task (B1-K1, 'Installing and maintaining hardware, software and connections') from the qualification test file**

ID (CT/WP)	Tokens	Bigrams
B1-K1	installing	[(installing, maintaining), (maintaining, hardware), (hardware, software), (software, connections)]
B1-K1	maintaining	[(installing, maintaining), (maintaining, hardware), (hardware, software), (software, connections)]
B1-K1	hardware	[(installing, maintaining), (maintaining, hardware), (hardware, software), (software, connections)]
B1-K1	software	[(installing, maintaining), (maintaining, hardware), (hardware, software), (software, connections)]

Source: Based on the ICT technician test file (NL) (as described in 3.6.1).

To conclude these pre-processing stages of development (corresponding to workflow Step 1, it is important to note that **throughout the steps discussed in this chapter, a number of dataframes (e.g. datasets exportable to Excel) were constructed within the prototype**. Many of these were created in between pre-processing steps, for development purposes only (i.e. preventing data loss, allowing for manual inspection), and need not be included in final prototype. For the analysis in the testing exercise (discussed in Chapter 4), the actual **inputs** will be the following dataframes, constructed from their respective ESCO and qualification test files:

(a) ESCO dataframe: containing the ID's, tokenized descriptions and bigrams included in the full list of ESCO's KSC pillar;
(b) Qualification dataframe: containing the ID's, tokenized descriptions and bigrams included in the Dutch example for the ICT technician qualification.

3.4.5.5 *Final steps before analysis*

In order to perform the actual analyses – frequency analyses and matching tokens/bigrams in the testing phase – which corresponds (partially) to workflows steps 3 and 4 ([138]), some further development was done in terms of constructing the necessary functions. This included the following steps:

([138]) Workflow Step 2, and some parts of steps 3 and 4, could not be performed with the prototype - due to the changes in approach that followed the expert consultations (see 3.4.3).

(a) Firstly, a 'counter' was developed ([139]), in order to **determine the number of unique tokens within each dataframe**, as well of the number of occurrences across the dataset (term frequency). The results, however, are stored separately, providing only the token and its respective number of occurrences in a new dataframe. Although this allows for some basic insights (such as identifying most/least common terms), the information on the respective skill/LO ID's is not included.

(b) Secondly, in order to **link the frequency information to the skills/ CT/WP they originated from**, an existing pandas function ('merge') was incorporated for further analysis, which writes the information from this 'frequency dataframe' back into the existing one, based specifically on the tokens in the new dataframe. This way, the frequency information will be included as a new column, copying the frequency score to each (corresponding) instance of a token in the existing (expanded) dataframe, thus allowing for retracing the (skill and CT/WP) ID's that the respective token was originally identified in.

(c) Lastly, in order to **compare the information between test files** – in other words to identify which tokens/bigrams in one dataframe are also found in the other dataframe, another existing pandas function ('isin') was used. This function allows for developing one or more conditions to test (such as 'token x in dataframe A is present in dataframe B) while going through all data rows of a given dataframe column. This returns a so-called Boolean series (i.e. true/false), which can then be used to select matching cases and transfer them to a new dataframe. For this prototype, this resulted in two file-specific dataframes:

   (i) **ESCO Tokens** - with all instances of the matching tokens for the qualification file (with corresponding ID's and bigrams);

   (ii) **Qualification Tokens** - with the same selection of matching tokens, linked to their respective ID's and bigrams for the ESCO file.

Following this step, these two 'Tokens' dataframes can be analysed in the same way as discussed before – using a counter to determine frequencies and merging them back into their original dataframes (thus linking tokens to all relevant skill or learning outcome ID's). Additionally, using the merge function

([139]) By embedding an existing function (counter) from the 'collections' package, which constructs a dictionary of tokens and 'counts', increasing the count score by 1 for each occurrence of any token in the dictionary. This dictionary can then be imported as data for a new dataframe.

from before, three more combinations can be made, resulting in a total of five 'combined' datasets:

(d) **Qualification Matches** – which contains the frequencies of matching tokens, linked to all corresponding **work tasks and core processes** within the qualification test file;
(e) **ESCO Matches** - which contains the frequencies of matching tokens, linked to all corresponding **KSC skills** within the ESCO test file;
(f) **Match ID's** – which contains the **matching tokens and their ID's** to each occurring instance, accumulated for both files;
(g) **Match frequencies** - linking the **matching tokens and their frequencies** to each occurring instance, accumulated for both files;
(h) **All Matches** - merging both of the 'matching tokens' dataframes into one, thus containing **all matching tokens, their ID's and respective frequencies** for each occurring instance.

3.4.5.6 *Final issues resolved (bigrams)*

After performing these analysis steps for single tokens, of which the results will be discussed in section 4.2, the same steps would (theoretically) apply to the bigrams extracted in the pre-processing stage. In practice, however, when using the developed functions (for single tokens) on the data containing the identified bigrams, a variety of errors were encountered before a feasible workaround was found. **These issues were similar to those encountered before, related to the way bigrams were enclosed into the dataframe.** In this case, the 'lists' of bigrams were unreadable for the steps needed to expand the dataframe to separate the bigrams within unique skills. In fact, they were stored within each data row as collection or 'array' of tokens (e.g. a long list of smaller lists), instead of as a list.

When attempting to 'unnest' the dataframe based on the bigrams (instead of single tokens), it was found that for the developed function '*all the arrays of the input need to have the same number of dimensions*' – which refers to the (varying) total number of words included in the original skill/ or CT/WP descriptions. Within the scope of this prototype, however, this is not desired given that even for the shortest descriptions, most skill descriptions will vary in their length and the process should allow for this ([140]). **Using the Boolean approach, however, to circumvent the issues in unnesting**, through defining a condition

([140]) This issue was not identified earlier as the 'lists' in which the single tokens were stored do not have to meet a similar length criterium.

for directly linking matching (qualification) tokens to ESCO bigrams (such as 'single token from *Qualification Matches* is present in bigrams of *ESCO dataframe*') resulted in an empty dataframe instead (i.e. a dataframe was created, but the cells could not be filled with data for comparison). This meant the function (incorrectly) determined all cases to be false, thus unable to read the bigrams as well.

Ultimately, a feasible workaround was developed, using another lambda function and some additional functions from the 'itertools' package. For this workaround, firstly, all cleaned descriptions (untokenized) are pulled from their respective dataframes, linked in pairs of two and then stored as a separate list. Secondly, the list is passed through a counter and written back into a new dataframe, resulting in two similar dataframes as seen for single tokens (see section 3.4.4.5; bullet c (I and ii)) – containing a full list of unique bigrams detected for each test file, and the frequency in which they occur throughout the respective skills or CT/WP. **Due to the limitations with the Boolean approach, however, these bigrams cannot be linked back to the skills or CT/WP in the dataframes containing the respective ID's.** Regardless, it would be possible to manually determine which qualification bigrams are present in which ESCO skills, as they are constructed in sequence (and can thus be searched for, separately, in the data once exported to Excel).

The statistical results of all analyses performed (on both the ESCO and qualification test files), will be discussed in the following section for single tokens and bigrams, separately, followed by the insights gained from the additional (combined) dataframes.

3.5 Emerging issues and conclusions following development

Looking back at the development phases described in this chapter, as well as the issues that led to an adaptation in approach, some aspects need to be taken into (further) account.

First of all, it was not feasible to perform all of the envisioned 'workflow steps' for the final prototype, considering that it no longer includes machine learning elements. The conditions that were set for the automated workflow were:

(a) **Ability to process different text formats (prior conversion or automated)**;
(b) **Allow extraction or labelling of key words;**
(c) **Ability to deal with natural language (text parsing);**
(d) **Ability to process different languages;**

(e) **Use Open Access software packages to maximise inclusivity.**

The non-automated approach that was adopted following the issues raised during expert consultations results in a prototype limited to the aspects ((a), prior conversion)); (b); and (e). This means that in particular **the aspects related to dealing with natural languages are not included in the prototype** ((a), when automated); (c); and (d)) – as these were determined to be far too labour-intensive for the scope of the (testing) prototype. This refers to the amount of pre-processing and tagging of textual data that would need to be done simply to train and improve the classifier model(s) – before passing the qualifications to be compared. This increase in human capacity requirements is even more so the case for aspect (d) – as dealing with different language would mean having to build, train and improve a separate classifier model for each language.

Secondly, in terms of challenges and limitations in pre-processing the data, within the context of the digital technologies identified and tested, the following aspects need to be taken into account as well, e.g.:

(a) **Using existing examples of functions to fit the purpose of the prototype requires a considerable amount of tailoring, testing, and identifying / resolving issues before a feasible workaround is identified.** Even then, later stages of development may reveal that the resulting data was delivered in an incorrect data type (as input for other functions), requiring additional tailoring and identifying alternative workarounds. A recommendation here would be to put the code (that was developed so far) online on one of the commonly used code review forums, providing a short description of the intended inputs and outputs and a 'mock' test file to process in order to maintain anonymity ([141]). This would allow the online community to actively 'think along' with the project and present suggestions to finetune / expand the prototype;

(b) Even the relatively simple functions (such as those for cleaning text) are capable of altering the data in such a way, that **more information is lost than preserved or gained**, and using the prototype on other files would

([141]) It is an inexplicit requirement to do so on many such forums, given that the main (if not first) response to any of the coding examples researched throughout development, was that: **identifying coding issues or providing suggestions for approach - *without* providing some sort of dataset - equals asking for inaccurate resolutions**, if not complete non-response.

therefore need manual inspection at several stages during pre-processing, to assure no relevant terms are being lost in the process ([142]);

(c) **Existing lexical resources, such as dictionaries used for stop word removal or stemming / lemmatisation, are mainly focused solely on the English language**. Some additional community and third-party resources could theoretically be incorporated. However, it can be expected that these are less well-developed than the more well-known English ones. Furthermore, the type of texts used to build these resources are generally far less occupation- or field-specific than would be necessary for the scope of this study (even for the English language). A recommendation here would be to look into enriching the selected online resources (such as the NLTK stop words package, or any of the stemmers / lemmatisers) through additional training, using more labour-market oriented texts (vacancies, CV's, job postings, ESCO terms, etc.).

Thirdly, findings show that **the ESCO classification would require major work (e.g. the source data of the final prototype),** before it can be used in a satisfactory manner for automated text processing, e.g.

(a) **Enrich vocabulary** – with stemmed versions of skills phrases; by generating skills phrases from existing texts (vacancies, for example); or through incorporating such resources from existing projects that process such texts ([143]);

(b) **Supplement semantic structure** to enable aggregations;

(c) **Dissect complex skills** into enabling skills components to make implicit components visible and to gain insight into performance levels;

(d) **Consolidate terminology** (e.g. summarise skills expressing the same meaning with different words under one concept).

([142]) For example, when removing special characters and digits of field-specific terms (for programming languages, abbreviations, etc.), or when using stemming (which can result in separate tokens being stemmed so far, they are incorrectly considered as one token).

([143]) Textkernel, for example, generates skills phrases from millions of vacancies, thus creating a vocabulary bottom-up. Unfortunately, this vocabulary cannot be structured by machines alone, and it would require human intelligence to connect terms.

Chapter 4. Testing digital technologies for gathering, structuring and analysing data on qualifications

4.1 Introduction of the testing exercises

The focus in this chapter is on providing insights into the results of the statistical analysis of the various dataframes, constructed from the ESCO and qualification test files. The analysis first focuses on the single tokens (terms) and bigrams (word pairs) identified in their respective test files (ICT technician qualification and full list of ESCO KSCs), followed by the matching between qualification and ESCO. Furthermore, considering the change in approach during development, some insights are provided into the state and characteristics of the finalised prototype – in terms of the data it can process (inputs); the resulting datasets (outputs); and the functions and analyses it can currently perform ([144]).

When considering the prototype as a digital 'tool', it is essentially a tool that can:

(a) Import text data from one (or more) Excel files and store them in a data format that keeps the original structure (of tabular data) intact;
(b) Pre-process the texts for identifying single terms and bigrams within the texts;
(c) Perform basic text analysis (frequencies);
(d) Link frequency results back to the source data (as a new column);
(e) Match the unique terms or bigrams onto those identified in another processed dataset.

However, technically speaking, **the 'tool' is more of a collection of resources** that can be used through an open source Python programme (Anaconda Navigator) ([145]), These resources include: (1) The ESCO and

([144]) A full overview of the existing resources used, as well as an overview of references to online tutorials, guides, examples, forum discussions, and coding examples/reviews will be presented as appendices (@@still in-dev, will also enrich the respective texts in CH3 by adding some references where relevant (discussions on specific errors/issues, for example))

([145]) Although most of the scripts for separate functions should work fine regardless of the 'python interpreter' programme used – the method of importing packages generally differs between programmes and therefore it would be recommended to use the Spyder interpreter found in Anaconda Navigator.

qualification test files; (2) datasets resulting from running the functions included in the full script (see also section 3.4.5.5); and the full script itself – which can be tailored and executed in sections which correspond to the subtasks enclosed within. To provide some additional insight, the following table represents an overview of the main functions and subtasks included in the full script, as well as the existing resources incorporated to perform them:

Table 17. **Main functions and subtasks, as included in the full script of the final prototype**

Main function	Separate Subtasks	Existing packages involved
Setting up workstation	Retrieve current working directory; Change directory to desired folder; List files in current work directory.	'os' (python base).
Test file to dataframe	Read in Excel file (or CSV); Store data in dataframe; Drop unneeded data columns.	'pandas'; 'xlrd'; 'numpy'.
Text Cleanup	Transform specified token to specified alternative. Clean punctuation; lowercasing; removing digits; cleaning ESCO ID's (remove part of URI); stop word removal.	'string'; 'punctuation' (from string); 're'; 'nltk'; 'stopwords' (from NLTK 'corpus').
Tokenization	Tokenize columns, by separate words.	'punkt' (from NLTK); 'sent_tokenize' & 'word_tokenize' (from 'ntlk_tokenize').
Bigrams	Construct bigrams from tokens; Construct bigrams from (cleaned) source text.	'ngrams' (from NLTK.util); 'list' (base python).
Expand dataframe	Unnesting.	N.A. (self-defined).
Base statistics	Define counter; Write frequencies to new dataframe; Defining criteria for subset selection (Ex: 'Freq => 100').	'Counter' (from collections); 'pandas'; 'sort' (base python) .
Matching & merging dataframes	Defining criteria for matching (Ex: 'A is in B'); Use criteria to match / merge.	Self-defined; 'isin' (from pandas); 'chain' (from itertools).
Export data	Write dataframe back to Excel; Write dataframe back to CSV file.	Excelwriter; openpyxl; xlsxwriter.

Source: Based on final script developed for the prototype.

In terms of use, performing the same process on any other qualifications would only require the user to prepare an Excel test file for each qualification with

at least two data columns (skill/LO label and skill/LO ID) and tailor minor parts of the code ([146]). Any additional data columns could still be included and used, for example, in setting criteria for the matching and merging datasets after analysis. This would allow the user to develop subsets of results to provide additional insights (i.e. distinguishing between skill types or professional contexts; including additional (more extensive) skill/learning outcomes descriptions, etc.).

Before moving onto the results of the texting exercise, some general statistics mentioned during pre-processing can be reviewed: The test files contained a total of 13,485 skill descriptions (ESCO's KSC pillar) and 11 'core tasks and work processes' ('CT/WP', Dutch qualification for ICT technician). After the pre-processing in terms of text cleaning, followed by tokenisation and dataframe expansion ('unnesting') - a total of 42,873 (non-unique) terms were used for the ESCO descriptions and 41 for the CT/WP. Furthermore, the counting of steps necessary for the frequency analysis simultaneously provided insight into the number of unique tokens identified for each test file – respectively 6,571 for the full list of ESCO's KSCs, and 35 for the CT/WP in the Dutch ICT technician qualification. In terms of bigrams, 23,852 were identified for the ESCO file, whereas the qualification file only contains 26 bigrams.

4.2 Results of the testing exercise

4.2.1 Frequency analysis results

Before matching tokens or bigrams between ESCO's KSCs and the CT/WP included in the Dutch qualification for ICT technician, some insights can be provided on the full list of tokens in the description of either file. Regarding ESCO, **the frequency distribution of the 6,571 unique tokens is heavily skewed**, in that only a small number of tokens (57) are present in 100 or more cases/descriptions (only 0.87%). When lowering the (filter) condition to 50 or less occurrences, a total of 6,426 unique tokens are returned (97.8%). Most of these are in fact terms that only occur once throughout the entire list of ESCO KSC descriptions – 3,038 cases, which is 46.2% of the total number of unique tokens. This means that **almost half of all terms included in ESCO (excluding stop**

([146]) Such as: which directory (folder path) to work from; names of Excel files to be imported as test file(s); and column names in initial loading where different from test file. After initial loading, the column names are transformed several times following a simple naming structure that is not specific to the content of the qualification and need only be changed if specifically desired by the user.

words) are unique – which is an indication of the high number of field-specific terms used in ESCO ([147]). Examples of terms only occurring once throughout ESCO are: acclimatise, aerodynamics, bronchoscopy, cellular, and chromatography. To provide some insights into the most commonly used terms, the following figure shows the (29) terms and frequencies of those that occur 50 or more times throughout the ESCO test file.

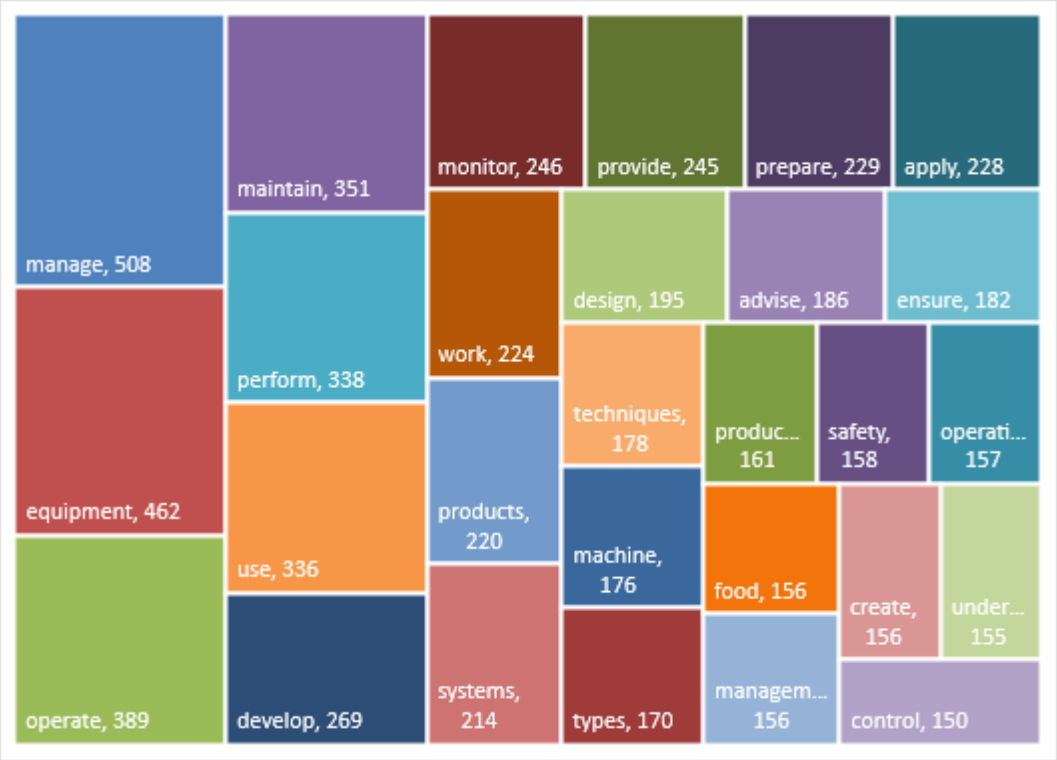


Figure 10 **Frequencies of tokens identified within the full list of ESCO KSCs**

Source: Based on ESCO test file after frequency analysis ('ESCOfreq' dataframe).

As the figure illustrates, the most common term identified in ESCO is 'manage' (present in 508 skills), followed by 'equipment' (462). Overall, **most of**

([147]) There are still some cases where similar tokens are considered separately, due to stemming / lemmatisation not being incorporated into the prototype – although the prototype does allow for extracting a list of all tokens that only occur once for manual inspection in order to identify outliers (for resolving issues with spelling or plural forms, especially). An example currently present would be 'worms' and 'worm', which are considered as separate tokens. However, there are also terms that one would expect to have 'plural' variants, that are not found in the list (thus not in ESCO at all, or occurring more than once).

the common terms in ESCO include action verbs within a variety of contexts – such as various terms related to working with equipment or techniques, as well as managerial functions (e.g. manage, advise, monitor, etc.). **Manual inspection of the mid-frequency terms in ESCO showed that generally correspond to the more context-defining terms** - more specifically, those found in 10 to 50 skills, such as: digital, instruments, legislation, textile, leather, public, environment, electronic, etc.

When performing the same analyses for the qualification test file, the number of unique tokens (35) and their frequencies are (expectedly) much lower – as these descriptions only refer to a small selection of CT/WP (11, compared to 13,485 in ESCP). The following table provides an overview of the terms, grouped by the frequency in which they occur.

Table 18. **Unique terms identified in core tasks and work processes of the Dutch qualification for ICT Technician, by frequency**

Frequency	Terms
1	peripheral, connections, equipment, systems, users, handling, incidents.
2	repair, give, ready, use, registering, parts, instructions, supporting, assembly, applications, incident, realize, maintaining, software, hardware, service, explanations, performing, resolving, installing, reports, preparation, escalating, making, replacement, oral, dis ([148]).

Source: Based on qualification test file after identifying tokens ('TokensQ' dataframe).

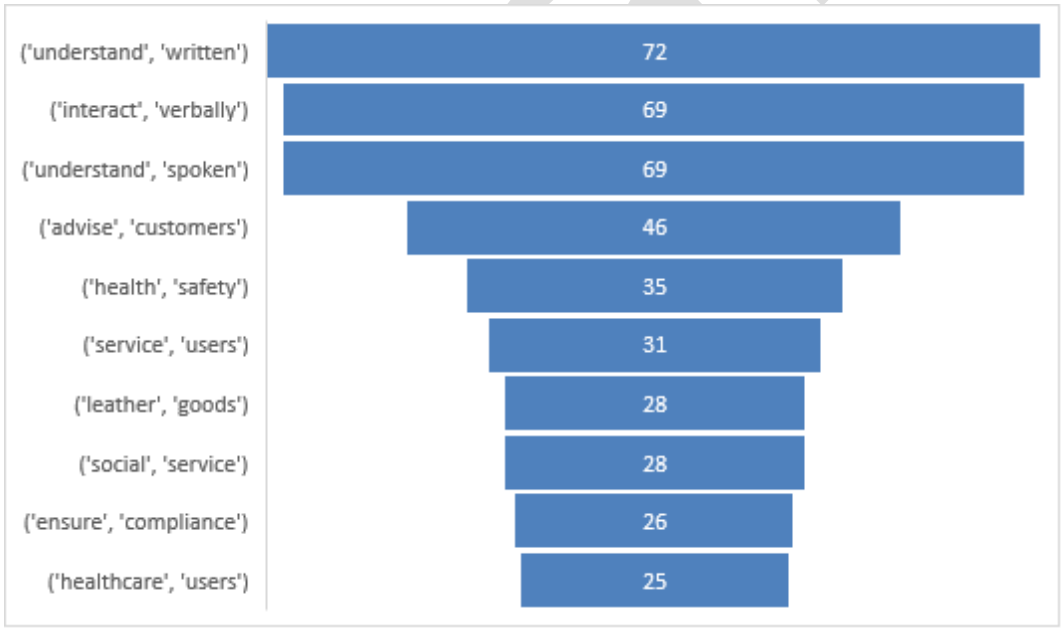
As this file only contains a relatively small number of tokens, their frequency distribution is fairly consistent (since the range is only 1-2). Apart from 'handling', which only occurs once, all action verbs occur twice. The more field-specific terms are found in both groups. **Overall, it can be said that the nature of the terms *within* the qualification test file (seen as a whole) is (highly) field-specific and mainly consist of terms related to the field of ICT technology**. This should benefit the matching, in that excluding all tokens unseen in both test files would filter out many of the less field-specific terms (and/or those linked to some entirely unrelated professional contexts).

([148]) This used to be combined with 'assembly' as (dis)assembly and was turned into a separate token during the cleaning operation. Manual inspection of the ESCO file revealed this term not to be present there, so adding a separate one for disassembly would still result in it being filtered out in the matching. Additionally, as the 'dis' and 'assembly' terms are still next to each other, bigram identification would link these terms back together, thus at least accounting for positive matching in case (dis)assembly did occur as such in the ESCO data.

4.2.2 Paired tokens

As was the case for unique terms (tokens) in terms of frequency distribution, when observing the results for unique bigrams identified in ESCO (23,852 in total), only a small number of high-frequency ones were found. **Only 48 bigrams occur 10 times or more in ESCO (less than 1%), while by far most of them only occur once (87.6%).** To some extent, this is due to the limitations of the bigram function used within the prototype – since this function, even though it does check for inverted versions of the bigram ([149]), it only connects terms that are directly next to each other in the description (thus not accounting for POS). On the other hand, **for identifying all possible combinations of bigrams within a sentence, the required processing capacity of the workstation running the prototype would be significantly higher**, as this would significantly increase the number of both total and unique bigrams for testing. To illustrate, the frequencies of the (10) most-frequent bigrams identified in ESCO are presented in the following figure.

Figure 11 **Frequencies of top 10 bigrams, within the full list of ESCO KSCs (N = 23,852)**



Source: Based on ESCO test file after identifying bigrams ('bigramsESCO' dataframe).

([149]) So that, for example, both 'interact, verbally' and 'verbally, interact' would lead to a positive match.

For the qualification file, a total of 28 bigrams were constructed, of which only two occur more than once in the Dutch qualification for ICT technician (e.g. ‘systems, peripheral’, and ‘peripheral equipment’). The following table presents an overview of all bigrams identified in the qualification test file, grouped by frequency.

Table 19. **Bigrams identified in core tasks and work processes of the Dutch qualification for ICT Technician, by frequency**

Frequency	Bigrams	Bigrams (continued list)
1	('hardware', 'software'); ('handling', 'incident'); ('give', 'users'); ('maintaining', 'hardware'); ('realize', 'connections'); ('preparation', 'instructions'); ('oral', 'explanations'); ('supporting', 'users'); ('ready', 'use'); ('handling', 'registering'); ('installing', 'maintaining'); ('users', 'oral'); ('performing', 'service') (*to be continued --->)	('parts', 'systems') ('assembly', 'parts') ('incident', 'reports') ('software', 'connections') ('making', 'systems') ('escalating', 'incidents') ('equipment', 'applications') ('dis', 'assembly') ('resolving', 'escalating') ('registering', 'incidents') ('applications', 'ready') ('replacement', 'repair') ('repair', 'dis')
2	('systems', 'peripheral') ('peripheral', 'equipment')	

Source: Based on qualification test file after identifying bigrams (‘bigramsQ’ dataframe).

Even though there are a few bigrams that seem ‘off’ or out of context - e.g. (‘users’, ‘oral’) and (‘repair’, ’dis’) – a fair number of these still indicate the ICT or at least a technical or mechanical context and it would be expected to find some interesting matches when comparing bigrams between the qualification and ESCO test files, which will be discussed further on (after matching single tokens).

4.2.3 Matching work tasks and core processes to KSC skills – single tokens

After observing the statistics for each respective test file, the merged dataframes allow for providing insights into the terms that match between the qualification’s CT/WP and ESCO’s KSC skills, and which (KSC) skills in particular they correspond to. Analysis showed that **of the (35) unique tokens identified in the qualification test file, a total of 28 were also present in ESCO’s KSC descriptions**. These 28 single tokens were matched with a total of 2,468 skills. In terms of their occurrences in KSC descriptions, the top 3 terms used most commonly overall, are in fact present in over 200 separate KSC descriptions (‘equipment’ in 462, ‘use’ in 336 and ‘systems’ in 214). **In terms of more field-**

specific matches, however, it is the least-common terms we are most interested in, as they are more likely to correspond to the correct context (when comparing based on single tokens). To illustrate, the following table presents the corresponding ID's and learning outcomes descriptions of a selection of single tokens (matching with only a small number of skills).

Table 20. **Selection of linked CT/WP and KSC descriptions (with professional context), by matching (unique) term**

Token / Term	ID_Q	CT/WP description	KSC description	Linked to ESCO profile(s) ([150]):
connections	B1-K1	Installing and maintaining hardware, software and connections	inspect for unauthorised connections	meter reader.
connections	B1-K1-W3	Realize connections	inspect for unauthorised connections	meter reader.
installing	B1-K1	Installing and maintaining hardware, software and connections	estimate costs of installing telecommunication devices	telecommunications engineer; specialised seller; telecommunications equipment specialised seller.
maintaining	B1-K1	Installing and maintaining hardware, software and connections	assume responsibility for maintaining a safe ship environment	fleet commander; marine surveyor.
peripheral	B1-K1-W1	Making systems, (peripheral) equipment and applications ready for use	set up audiovisual peripheral equipment	camera operator; audio-visual technician; recording studio technician; broadcast technician.
peripheral	B1-K1-W1	Making systems, (peripheral) equipment and applications ready for use	explain characteristics of computer peripheral equipment	specialised seller; computer and accessories specialised seller.

([150]) It is important to note that **since the prototype does not perform retrieval of the professional context, the column indicating the ESCO profiles corresponding to the matched KSC skill was manually constructed**, by retracing each profession's URI (ID) through the corresponding KSC skill URI (using a separate ESCO-database 'occupation-skills relation' for v103).

peripheral	B1-K1-W2	Replacement, repair and / or (dis) assembly of (parts of) systems and (peripheral) equipment	set up audiovisual peripheral equipment	camera operator; audio-visual technician; recording studio technician; broadcast technician.
peripheral	B1-K1-W2	Replacement, repair and / or (dis) assembly of (parts of) systems and (peripheral) equipment	explain characteristics of computer peripheral equipment	specialised seller; computer and accessories specialised seller.
supporting	P2-K1	Supporting users	roles of supporting team for community arts programme	No profile linked.
supporting	P2-K1	Supporting users	work with supporting team in community arts programme	No profile linked.
supporting	P2-K1	Supporting users	evaluate supporting team in community arts program	No profile linked.

Source: Based on merged test file after single token matching and frequency analysis ('mergedDf1labels' dataframe, 'mergedDf2frequencies' dataframe).

When comparing the full descriptions between CT/WP and KSCs for the selected terms, none of these were found to result in a proper match, especially when the professional contexts of the matched KSCs and CT/WP are taken into account. The term 'supporting', for example, is only linked to the context of (community) arts programmes in ESCO, whereas manual inspection of the test file indicates that the alternative term 'support' is in fact linked to another 80 skills in ESCO. To some extent, this issue might be resolved by performing stemming or resolution of polysemy/synonymy, thus effectively reducing the number of 'duplicate'/similar tokens to test for. On the other hand, doing so would simultaneously *increase* the number of incorrect matches, due to some term variation being important within the scope of comparing CT/WP to KSC skills ([151]).

Furthermore, using the KSC ID's to retrace the more ambiguous skill 'inspect for unauthorised connections' – which could potentially refer to an ICT context - it was found to be linked to the 'meter reader' profession, thus within the context of electrical connections (infrastructure). Similarly, the KSC matches for 'maintaining' are within the context of ship environments, optical products and

([151]) For example, when stemming terms such as 'installing' and 'installations', which would both be reduced to 'install' through this addition to pre-processing – losing an important distinction.

bread products – whereas **it would have been expected to get some matches related to systems / software at least**. The only term (in Table 15) that does provide matches in an ICT context is 'peripheral', which is linked to the KSC skill '*explain characteristics of computer peripheral equipment*'. Regardless, this skill is linked to another profession as well, more specifically to 'computer and accessories specialised seller'. Even though it was expected to get a large number of undesired matches using single tokens alone (i.e. 462 matches for 'equipment'), it does seem that **some terms are linked to an unexpectedly low number of skills, regardless of them being used across professional fields** (such as: supporting (3), maintaining (3), connections (2), installing (1)). This is likely the case for more terms within ESCO's KSC pillar, should the same approach be used to compare other qualification files. In terms of recommendations, it might be worthwhile to have the lowest-frequency terms checked manually (on the ESCO side), in order to identify more of such terms (meaning those that one would expect to find in more than one skill, yet fail to be included in more skills)– and determine whether synonymy/polysemy is the underlying issue, or perhaps an indication that some skills are being linked to a low number of professions, whereas they could still be considered transversal.

4.2.4 Matching work tasks and core processes to KSC skills – bigrams

When comparing the ICT Technician's identified list of CT/WP bigrams (28 in total) to the ones identified for KSCs, only three learning outcomes bigrams were found to be present in ESCO's KSC pillar: 'incident', 'reports' (2 matches); 'ready', 'use' (1 match); and 'peripheral', 'equipment' (2 matches). The following table presents the CT/WP and KSC descriptions for all occurrences of the 3 'unique' matching bigrams.

Table 21. **Linked CT/WP and KSC descriptions and ID's, by matching (unique) bigram**

Bigram	ID_Q	CT/WP description	KSC description	Linked to ESCO profile(s) ([152]):
incident reports	B1-K2	Handling of incident reports	create incident reports	chemical metallurgist; process metallurgist; metal furnace operator; coking furnace operator; heat treatment furnace operator; metallurgist; mine rescue officer.
incident reports	B1-K2	Handling of incident reports	process incident reports for prevention	chemical metallurgist; process metallurgist; metal furnace operator; coking furnace operator; heat treatment furnace operator; metallurgist; mine rescue officer.
ready use	B1-K1-W1	Making systems, (peripheral) equipment and applications ready for use	ensure tableware is ready for use	hospitality vocational teacher; head waiter/head waitress; head sommelier; venue director; restaurant manager.
peripheral equipment	B1-K1-W1 + B1-K1-W2	'Making systems, (peripheral) equipment and applications ready for use' + 'Replacement, repair and / or (dis) assembly of (parts of) systems and (peripheral) equipment'	set up audiovisual peripheral equipment	camera operator; audio-visual technician; recording studio technician; broadcast technician.

([152]) It is important to note that **since the prototype does not perform retrieval of the professional context, the column indicating the ESCO profiles corresponding to the matched KSC skill was manually constructed**, by retracing each profession's URI (ID) through the corresponding KSC skill URI (using a separate ESCO-database 'occupation-skills relation' for v103).

peripheral equipment	B1-K1-W1 + B1-K1-W2	'Making systems, (peripheral) equipment and applications ready for use' + 'Replacement, repair and / or (dis) assembly of (parts of) systems and (peripheral) equipment'	explain characteristics of computer peripheral equipment	specialised seller; computer and accessories specialised seller.

Source: Based on merged test file after bigram matching and frequency analysis ('MatchingbigramsESCO' dataframe, 'MatchingbigramsQ' dataframe), as well as the ESCO v103 file linking occupations to skills.

Similar to the results seen when observing low-frequency matches of single tokens, the identified bigrams also fail to provide matches within the right context. The bigram 'peripheral equipment' in fact links to the exact same skills as were found using 'peripheral' alone – 'setting up audiovisual peripheral equipment' and 'explain characteristics of computer peripheral equipment'. Furthermore, when using the 2 (KSC) ID's matched to the bigram 'incidents reports', it was found that neither was a desired match in terms of professional context. In fact, they are mainly within a context of mining, dangerous goods/chemicals and metallurgy – considering that '*create incident reports*' is only considered an essential skill for the professions 'chemical plant control room operator' and 'mine rescue officer' ([153]), and similarly, '*process incident reports for prevention*' is only considered an essential skill for 'mine rescue officer' ([154]). To some extent, this is unsurprising considering the differences in description of the linked CT/WP and KSC skill, since 'ensure tableware for use' is obviously out-of-context without having to retrace the linked ESCO profile. The fact that not a single bigram match resulted in a KSC within ICT context, however, is an indication of the aforementioned 'lexical gap' between ESCO terminology and the way short definitions of learning outcomes are (in this case, core tasks and work processes) defined. A recommendation would be to further inspect the identified bigrams in the full ESCO list – while using a bottom-up approach (i.e. from occurring least frequent to most) in order to identify bigrams that one could expect to find in more skills than they do currently.

As noted before, using **further pre-processing steps such as stemming, lemmatisation, resolving polysemy/synonymy would allow for a reduction**

([153]) And as optional skill for 'dangerous goods safety adviser', 'steam turbine operator'. See http://data.europa.eu/esco/skill/24310a7b-5db1-4917-a705-4835630227ca

([154]) And as optional skill for 'chemical metallurgist', 'coking furnace operator', 'process metallurgist', 'heat treatment furnace operator', 'metallurgist', and 'metal furnace operator'. See http://data.europa.eu/esco/skill/32ac6bfb-8b86-4312-9421-dcb8e0203ab5

in similar tokens being tested separately – although in doing so, the information lost (in cases where disambiguation is more important) would simultaneously increase, especially when using the most-referred to 'stemmers' / 'lemmatisers' within the NLTK package, as these are built on texts that are far less field-specific than is the case within our prototype. Thus, it may be more worthwhile to **use resources developed (or still in development) in other projects working with more labour-specific terms** – in order to develop a more tailored stemmer / lemmatiser. Examples of such projects would be the vacancy analysis projects mentioned before - Cedefop's RTLMI Project and the one from Textkernel – or from similar projects from other countries (especially if/when the approach is to be expanded to include languages other than English).

4.3 Emerging issues and conclusions following the testing phase

Reviewing the analysis results overall, some notes can be made on emerging outcomes and issues, mainly related to the 'matchability' of the source data (ESCO). Two overarching aspects stand out:

(a) **When focusing on single terms present within ESCO, a highly skewed frequency distribution was found (e.g. a high number of single-occurrence tokens).** The ESCO classification has been designed with a labour market orientation in focus capturing as much occupation-specific detail in skills profiles as possible. Yet it was discovered that ESCO also contains a high number of more generic terms which – contrary to our expectations – did not appear across a large number of skills and hence skills profiles (such as 'peripheral', as seen in Table 20). **Additionally, even when focusing on bigrams (instead of single terms), the same skewed distribution and occurrence of out-of-context matches remained an issue.** A recommendation here would be to use the frequency dataframe for ESCO to inspect the least-frequent tokens and identify to what extent this is due to stemming, or whether some terms are not used in skills descriptions where one would in fact expect to find them.

(b) On the other hand, some more commonly used tokens or bigrams were linked to so many skills (100+ for single tokens) that **manually identifying the 'correct' matches – within their professional contexts – would require a significant increase in human capacity for each separate to be compared**, which is undesirable. A recommendation in this regard would be to include a data column in the ESCO (v103) base file that indicates (as text) the name(s) of the profession(s) each skill is linked to in ESCO. This would

simultaneously allow for selecting subsets of skills in order to perform more specific analyses.

Overall, both aspects are an indication that the ESCO KSC classification uses a language that significantly differs from the language used in the descriptions present in national qualifications, regardless of their (shared) labour market focus and relatively high use of field-specific terms. It would therefore be recommended for ESCO to develop their own databases for dealing with 'real world' terms (i.e. enriched dictionaries), which can be used within the scope of further tool development, such as through incorporating stemming / lemmatising functions for resolving polysemy/synonymy (without losing necessary distinctions); working with texts from non-English languages.

Chapter 5. Conclusions and recommendations

This chapter presents conclusions in relation to the key research questions of this study and in view of the overall aim of the Cedefop project, to prepare methodologies allowing for a systematic comparison of VET qualifications.

5.1 Conclusions related to key research questions

1.What are the key sources for data on national qualifications, in particular related to their content and profile?

1.1 Which data sources exist and are of relevance for the comparison of national qualifications?

National documents, specifying the learning outcomes of IVET qualifications ('main reference documents') – which are therefore relevant for comparing qualifications – are in most cases public regulatory documents, meaning they are (generally) publicly available and published online on the websites of the responsible organisations ([155]). A range of 'types' of documents was found in the ten countries covered by this study, consisting of a variation of full and short summary descriptions of learning outcomes. In terms of what these documents are called, a variation of terms can be identified across the countries studied: qualification / certification specifications / requirements (FI, IE, UK-EN), qualification standards (FR), qualification files / profiles (NL), VET standards (BG), occupational standards (LT), and training regulatory documents (AT, ES, DK).

The reference documents analysed often contain a mixture of information, including information related to the functions of occupational, educational and assessment standards. The amount of information usually depends on the type of division of tasks within a country. Information about the curriculum or learning programmes is least likely to be included in these documents, as this is most likely a task that is transferred in the administrative hierarchy to the local, regional and provider levels.

([155]) An exception to this is the UK-England.

Moreover, contextual information on qualifications and their learning outcomes is often contained in a number of different sources, which are interlinked or aligned to some extent. A distinction can be made between (i) documents at national level and (ii) sets of documents that (together) span the 'vertical' institutional / governance dimension of VET systems. Whereas information on links to occupations / the labour market tends to be provided in documents that also contain the full and short learning outcomes descriptions, it was found that information on the EQF level, and on possibilities for further learning was only in a few cases provided in these documents. Furthermore, findings show that explicit information on the *distribution* of different types of learning outcomes is usually not provided at all in these documents.

The level of detail provided varies and the learning outcomes are presented in different ways in these documents. For example, in terms of the degree to which learning outcomes are structured, the countries cover a broad range from using no systematic structuring of learning outcomes in the main reference document, to highly structured ways of presenting them. The descriptions also differ regarding the sentence components: While verbs and objects are common components of individual learning outcomes statements, it is also not uncommon for nouns to be used instead of verbs. Context is rarely provided in individual learning outcomes statements; more commonly, context is provided in over-arching statements. Moreover, it is usually the object and context of a learning outcome statement that expresses the vertical dimension, rather than the verbs used. The analysis of reference documents for qualifications shows that learning outcomes are structured and expressed in a variety of ways, which poses challenges for comparing qualifications.

None of the main documents containing learning outcomes are systematically available in English (in addition to the national language). Furthermore, learning outcomes descriptions are in all countries stored in PDF documents. Such documents can only be accessed and downloaded individually, thus impeding any automated comparison (without additional preparatory manual work).

1.2 To what extent can national qualifications databases support comparisons of VET qualifications?

During the last couple of years, important progress has been made in establishing qualifications databases. However, as the results of this study show, the databases differ in their scope and not all databases have a full coverage of IVET qualifications yet. They also differ in terms of the type and amount of information provided, and in particular whether they actually contain learning

outcomes. There is also a great variance related to the *categories* used for presenting and structuring information in databases, as well as the search functions provided. Therefore, the databases analysed in the ten countries covered by this study currently only support comparison of VET qualifications to a rather limited extent.

The elements for data fields (five required ones and twelve optional ones) for the electronic publication of information on qualifications with an EQF level (as suggested in Annex VI of the EQF Recommendation) are only used in a few rather recently established databases. Only some databases provide short summaries of learning outcomes, other databases offer only a link to the full description of learning outcomes (either to another website or a document that can be downloaded), but do not contain any learning outcomes descriptions themselves (as a data field). Thus, easy extraction and application in software packages is in most cases not possible, as it requires either manual work to gather the learning outcomes descriptions or extensive and heavily customised instructions to software packages for them to retrieve the right documents and extract the right texts for analysis. In those cases where learning outcomes statements *are* included in qualifications databases, they are usually neither structured in a systematic way nor based on existing taxonomies (or any other clearly identifiable systematic approach), and they are only rarely composed of all components suggested by Cedefop (2017, p. 47). Thus, 'part-of-speech tagging' would only be supported to a limited extent. Moreover, in most countries, the information included in the databases is presented in the respective national language(s) only and a translation (of learning outcomes) into English is only available in very few cases.

2. How can new digital technologies support automated gathering, structuring (including cleaning, fusion) and analysis of data on qualifications?

National data on qualifications is described in text that can, in theory, be gathered, structured and analysed through automated processes. These automated processes could support the comparison of qualifications and potentially reduce the human workload in comparing qualifications. In Chapter 3, the research team presented their experiences with putting theory into practice by designing, developing and implementing a prototype for automated

comparison of qualifications ([156]). The research team tried to implement a workflow consisting of the following steps:

(a) Step 1: Provide access to national qualifications in machine-readable form; pre-processing of national qualifications descriptions (and of a reference point for comparison);
(b) Step 2: Parse the learning outcomes of national qualification descriptions;
(c) Step 3: Normalise detected learning outcomes for each national qualification description by mapping it onto the reference system's vocabulary;
(d) Step 4: Mapping of most suitable occupational skills profile (OSP, 'reference point') with normalised national qualifications, registering overlap and divergence.

As described in Chapters 3 and 4, the research team did not fully succeed in establishing an operational system for an automated comparison of qualifications. In the process, however, an in-depth understanding was gained of the fundamental obstacles for completing this process. This concerns the following:

(a) **Too little qualification data is available for the training of an automated system (i.e. machine learning is not possible).** There is a limited number of national qualifications to be processed ('mapped') for the scope of this study, but equally for the whole set of qualifications in Europe. While there is for example, a constant flow of vacancy texts published every day which provides a rich basis for machine learning and automated analysis, this is not the case for qualifications: Only a limited number of qualifications is published per country and these qualifications are not all that frequently renewed (e.g. every five years). The amount of data needed to train a machine learning model (per reference point, per language ([157])) is far larger than the number of national qualifications available. Furthermore, these national qualifications differ greatly in the way how KSC are described. Finally, the linguistic diversity also greatly complicates the setting up of a machine learning system.
(b) **The national qualification descriptions and the selected reference point differ too much to allow sensible matching based on automated**

([156]) The prototype was tested with data from selected qualification documents. Technically, so long as the information is extracted beforehand, the current prototype is able to process qualification data from other sources the same way.

([157]) For example, when using the different reference points or systems as tested in WA1, the model would need to be trained to map to each one as well (and then again for each language).

processes. Firstly, the ESCO KSC classification contains a language significantly differing from the language used in the descriptions present in national qualifications, regardless of their (shared) labour-market focus and relatively high use of field-specific terms. Secondly, KSC in national qualification descriptions and the reference point are too contextualised (combining KSC elements and very specific contexts within one statement) to allow for a matching based on the occurrence of individual words. Even when the data is enriched through linking individual words with synonyms, matching still leads to imprecise matches. Even if this fundamental obstacle could theoretically be solved, the workload involved in establishing an automated comparison between a national qualification and a reference point is far greater than doing this manually. Furthermore, even though processing more qualifications might eventually reduce the workload needed per qualification slightly, this would not substantially reduce the overall efforts needed because it would not be feasible to put machine learning in place (as explained above).

Therefore, the study concludes that – at this moment of developments – AI technologies can neither be used for gathering, structuring and analysing the content of qualifications, nor do they provide hope for being able to support a manual comparison in a way that makes this endeavour significantly less labour intensive. What the experience with designing, developing and implementing an automated process did provide, however (as shown in Chapter 4), is further insight into ESCO's suitability as a reference point for comparing qualifications.

3. How can new digital technologies address the linguistic challenges involved in comparing qualifications?

Comparing qualifications across Member States and VET systems is associated with considerable linguistic challenges. This firstly concerns the linguistic diversity in Europe, and secondly, the heterogeneity in how KSCs are phrased and structured.

This **linguistic diversity** is a big challenge for an automated comparison of national qualification descriptions, as it would entail repeating all necessary preparatory steps (tokenisation, cleaning texts, eventually adding synonyms etc.) for each language version of ESCO. An alternative could be to perform the comparison in one language only (e.g. English) and focus on training one (classification) model. However, this would make it necessary to translate the national qualification descriptions into English prior to processing.

The **heterogeneity in how KSCs are phrased and structured** presents an even bigger challenge for an automated process. When manually comparing

qualifications (as done in Work Assignment 1), these challenges are met by experts because these do not merely read but also interpret and, if necessary, supplement (e.g. for implicit KSCs) national learning outcomes descriptions and the reference points. Hence, this is not an exact, computed match as it allows for a certain degree of subjectivity (indeed, it is highly likely that two experts looking at the same qualification might come up with varying interpretations of individual or groups of learning outcomes). The automated process cannot conduct an expert assessment of skills profiles but relies on matching words (tokens) and combinations of words. This, however, does not lead to satisfactory and meaningful results providing complete insight into the content of qualifications and their differences and similarities between qualifications from different countries (as shown in Chapter 4).

4. What can be the role of the multilingual classification ESCO in supporting gathering, structuring and classifying qualifications data?

As indicated above, the multilinguality of ESCO does, in reality, not help much in easing the automated comparison, since dealing with different languages would mean having to build, train and maintain a separate classifier model for each language. Another major obstacle is that the national qualification descriptions and the national language versions of ESCO KSCs concepts differ too widely to allow meaningful automation supported matching. The usability of the multilingual classification ESCO in supporting gathering, structuring and classifying qualification data is therefore limited. Furthermore, as Chapter 3 indicates, a lot of work needs to be done on the ESCO classification before it can be used in a satisfactory manner for automated text processing

5.2 Recommendations

The development of an automated process for comparing qualifications using ESCO is only feasible if substantial changes take place, both in the way how national qualifications data is published and presented, and in the way how ESCO KSCs are structured and phrased.

First of all, in terms of **national qualification descriptions**, there is too much diversity to allow for an automated identification of accurate matches. The reference documents for national qualification descriptions are very diverse and often contain a mixture of information, not just learning outcomes. Moreover, the learning outcomes included therein are structured and expressed in a wide variety of ways, but only in the respective national language. In the case of the qualification databases research in the ten countries has shown that these also

differ greatly. In particular, the way in which learning outcomes are included in these databases, if available at all, varies with respect to how they can be accessed, to their level of detail and to the amount of information presented. The main challenge here for automating the process is that extracting both the (PDF) qualification files and the information contained within them, cannot be structured along the same set of rules – such as 'extract the information found under header(s) A, B, C' (which would need to be the case for a web crawler, for example). This would only work if all the files had the information in the same structure/location within their document. The time needed to identify the right locations and tailoring the extraction rules to each qualification document would be more than simply downloading them one by one and manually extracting the texts into a structured format.

Secondly, a short, not too detailed presentation of qualifications (i.e. not the full account of the learning outcomes of a qualification, but a synthesised description of its core profile) based on a common structure could support the automated analysis and comparison of qualifications, though naturally this would lose information, raising questions about how useful it would be in practice ([158]).

Yet Europass Certificate Supplements (ECS) are structured according to a common format, and usually translated into English, thus providing a promising data source for the automated comparison of qualifications. However, they would also need to be improved: ECS are usually available in a PDF-format but not all countries currently have a central inventory of ECS and they are not available for all VET qualifications in all countries. It is also not always clear whether the ECS are updated whenever national sources (core reference documents for qualifications) change. ECS include a 'profile of skills and competences', but the learning outcomes descriptions are not necessarily grouped or presented in a uniform structure. Moreover, the learning outcomes statements in the 'profile of skills and competences' are not necessarily composed of the components suggested by Cedefop (2017, p. 47): action verb, object of the verb, statement specifying the depth/breadth of learning to be demonstrated, indication of the

([158]) For example, if a qualification says 'plumber' on the certificate, it is questionable whether a short summary will help to understand any inter-country differences which can be highly specific – i.e. 'below' the level of the summary, and still 'hidden from view'. The summary might possibly give a hint as to such differences but then people often have to ask specific questions, like does the qualification cover a specific type of pipework. It also has to be kept in mind that these summaries need to be kept updated. It has to be recognised that summaries are used in individual contexts, such as by people looking to move abroad, companies seeking to recruit, guidance people giving individual advice, their usefulness, however, for comparing qualifications in order to identify similarities and differences is limited.

context. Similar to the structure of the ECS, the data fields for the electronic publication of information on qualifications with an EQF level as proposed in Annex VI of the EQF Recommendation (Council of the European Union, 2017) provides the basis for a common structure for presenting qualifications in databases. In both cases (ECS and databases), comparability and also the use of technologies for an automated comparison could be supported by applying common principles for presenting learning-outcomes-based qualifications as proposed by Cedefop (2017, p. 63). It is suggested to develop standardised national learning-outcomes-based descriptions of qualifications (of about 2,500 to 3,500 characters – i.e., a bit longer than suggested by Cedefop). These descriptions should follow a predefined structure and syntax ([159]), refer to agreed but flexible learning domains (EQF domains or other) and should be based on a standardised terminology, including lists of action verbs. This would mean a more harmonised approach for presenting learning outcomes statements which in turn would enhance the comparability of qualifications. However, these common principles refer to the presentations of qualifications in the European context only: in ECS and in national databases with qualifications with an EQF level. But it does not necessarily refer to the national reference documents for VET qualifications which often have a strong traditional basis and are embedded in the structures and approaches of national systems. Thus, this proposal does not conflict with how countries describe their VET qualifications in the core reference documents, nor does it require harmonisation of these. Another benefit could be to use this approach to better promote the EQF across countries, since in both cases the EQF level is indicated. This can also help various users (including learners, VET providers, career counsellors, employers) to better understand qualifications from different countries.

Lastly, in relation to **ESCO**, it would help if this taxonomy would adopt a more structured approach to how KSCs are described (vocabulary control) as well as a better alignment with the expressions found in qualification descriptions or vacancies, e.g. including these as synonyms. This could transform ESCO into a valuable lexical resource for (further) attempts to (automatically) compare qualifications. This includes, amongst others:

([159]) The importance of structuring learning outcomes statements (e.g. verb, object, context) for supporting the mapping of learning outcomes to the ESCO skills pillar was also emphasised by the study on the feasibility of conceptually and technically, linking the learning outcomes of qualifications included in the ESCO qualifications pillar, with the ESCO skills pillar (DG EMPL, 2019).

(a) Enriching the vocabulary – with stemmed versions of skills phrases via generating skills phrases from existing texts (e.g. vacancies or qualifications) or via incorporating suitable resources from existing projects that process such texts ([160]);
(b) Supplementing semantic structure to enable aggregations;
(c) Dissecting complex skills into enabling skills components to make implicit components visible and to gain insight into performance levels;
(d) Consolidating terminology (e.g. summarising skills expressing the same meaning with different words under one concept).

Provided the necessary information is extracted beforehand, it is possible to use the current prototype (resulting from Chapters 3 and 4) for some preliminary analyses, potentially contributing to the improvement of ESCO – with the goal to develop it into a lexical resource also suitable for natural language processing, e.g. by identifying specific terms within ESCO (or between ESCO and learning outcomes descriptions from other sources) which are most likely to cause mismatches (i.e. ambiguous terms, similar but different terms that would be those identified as one after stemming). Only after this challenge has been met, we consider using machine learning elements for (supporting) the mapping of learning outcome descriptions to ESCO KSCs. If the project is limited to one language (English), it would be possible (through translating qualifications in other languages into English) to manually prepare enough data for training a classification model on the full ESCO skills pillar – however, as seen in other projects (vacancy analyses), it is important to note that it is generally an ongoing process of training and updating the model – especially considering terminology changes over time.

All in all, a harmonised language repertory is prerequisite to any form of automation in comparing qualifications. In this context it might be of interest to look at other countries confronted with the qualification-comparison problem for decades, e.g. Canada (francophone versus anglophone parts of the country), and explore to what extent and in which manner they have solved the problem.

([160]) Textkernel, for example, generates skills phrases from millions of vacancies, thus creating a vocabulary bottom-up. Unfortunately, this vocabulary cannot be structured by machines alone, and it would require human intelligence to connect terms.

List of abbreviations

AC	Assessment Criteria
AE	Adult Education
AFPA	National Agency for the Vocational Training of Adults (FR)
AG	Advisory Group
AIKOS	Finnish Qualifications Database *(FI: Atviros Informavimo, Konsultavimo ir Orientavimo Sistemos)*
AO	Awarding Organisations
ASCII	Standard 'code set' for representing (text) characters - 'American Standard Code for Information Interchange'
BTEC	Business and Technology Education Council *(England, Wales, Northern Island)*
BLS	Basic Life Support
CCEA	Council for the Curriculum, Examinations & Assessment (UK-ENG)
CINOP	Dutch Centre for Innovation of Education Programmes (*NL: Centrum voor Innovatie van Opleidingen)*
CNCP	National Qualifications Commission (*Commission nationale de la certification professionnelle*)
CREBO	Dutch Central Registry of Vocational Education and Training Programmes *(NL: Centraal Register Beroepsopleidingen)*
CROHO	Dutch Central Registry of Higher Education Programmes *(NL: Centraal Register Opleidingen Hoger Onderwijs)*
CSV	Filename Extension - 'Comma-Separated Values' (Microsoft Excel; OpenOffice Calc; or Google Docs; other 'spreadsheet programmes')
CT/WP	Core Tasks and Work Processes *(Distinguishment used in Dutch National Qualification files)*
CVET	Continuous Vocational Education and Training
CV	Curriculum Vitae
ECS	Europass Certificate Supplements
EDUFI	Finnish National Agency for Education
ENIC/NARIC	National Information Centre
EQF	European Qualifications Framework
EQF AG	EQF Advisory Group
EQF NCP	EQF National Coordination Point
ESCO	European classification of Skills, Competences, Occupations and Qualifications
ESC	Europass Certificate Supplement
ETB	Education and Training Boards

GE	General Education
HA	Healthcare Assistant
HE	Higher Education
HCEU	'HealthCareEurope' project
HR	Human Resources
ibw	Institut für Bildungsforschung der Wirtschaft (ibw Austria - Research & Development in VET)
ICT	Information Communications Technology.
ID	Identifier
IRQ	Irish Register of Qualifications
KSC	Knowledge, Skills, Competences
IVET	Initial Vocational Education and Training
LOQ	Learning Opportunities and Qualifications Portal
LO	Learning Outcomes
MBO	Upper Secondary Vocational Education (*NL: middelbaar beroepsonderwijs)*
NAVET	National Agency for Vocational Education and Training (BG)
NFQ	Irish National Framework of Qualifications
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
NQF	National Qualifications Framework
Ofqual	The Office of Qualifications and Examinations Regulation (UK-ENG)
O*NET	Occupational Information Network
OSP	Occupational Skills Profiles
PATDs	Professional Award-Type Descriptors
PA-PFA-AV	Care Assistant Profession Training Ordinance *(AT: Pflegeassistenzberufe-Ausbildungsverordnung)*
PCDP	Personal Competence Development Plans
PC	Personal Computer ('Desktop Computer')
PDF	Filename Extension - 'Portable Document Format' (Adobe Acrobat Reader)
PES	Public Employment Service
PLA	Peer Learning Activity
POS	Part-of-Speech
QDR	Qualifications Dataset Register (European Commission)
QQI	Quality and Qualifications Ireland
R	R statistics
REAC	Employment, Activities, and Competences Standards *(FR: référentiel emploi activités compétences du titre professionnel)*
RIS	Legal Information System of the Republic of Austria *(AT:*

	Rechtsinformationssystem der Republik Österreich)
RNCP	National register of vocational and professional qualifications (*Repertoire national des certifications professionnelles*)
ROME	French Operational Catalogue of Occupations and Trade Jobs *(FR: Répertoire operationnel des métiers et des emplois, ROME)*
RTLMI	Real-time Labour Market Information
SADe	Action Programme on eServices and eDemocracy (FI)
SBB	Foundation for Cooperation on Vocational Education, Training and Labour Market *(NL: Stichting Samenwerking Beroepsonderwijs Bedrijfsleven)*
SOLO	Structure of Observed learning (e.g. SOLO taxonomy)
SOM	Self-Organising Map ('Kohonen Map')
UA	Unit Aim
UTF	Standard ‘code set’ for representing (text) characters - 'Unicode Transformation Format'
VAE	Validation of Experiential Learning Outcomes *(FR: Validation des acquis de l'expérience)*
VET	Vocational Education and Training
VQTS	Vocational Qualifications Transfer System
WSSS	WorldSkills Standards Specifications
XLS/XLSX	Filename Extensions - Microsoft Excel ‘spreadsheet files’ (respectively pre/post Excel 2007)
XML	eXtensible Markup Language

References

URLs accessed September 2019

Auzinger, M.; Luomi-Messerer (2017). *Umsetzung der Lernergebnisorientierung in der österreichischen Berufsbildung – Status quo* [Implementation of learning outcomes orientation in Austrian vocational education and training - status quo]. 3s, Wien, April 2017.

Auzinger, M.; Broek, S.; Luomi-Messerer, K. (2017). *The role of learning outcomes in supporting dialogue between the labour market and education and training; the case of vocational education and training*. Final report submitted to Cedefop - 2015-01 29/AO/DSI/JB/Learning outcomes and the labour market /004/15 (not published).

Bjørnåvold, J.; Chakroun, B. (2017). Learning outcomes in comparing VET qualification profiles: a global approach. Cedefop et al. (2017). *Global inventory of regional and national qualifications frameworks 2017 Volume I: Thematic chapters*, pp. 95-112.
http://www.cedefop.europa.eu/files/2221_en.pdf

Buzydlowski, J.W. (2015). Co-occurance analysis as a framework for data mining. *Journal of Technology Research*, vol. 6, pp. 1-19.

Cedefop (2017). *Defining, writing and applying learning outcomes: a European handbook*. Luxembourg: Publications Office.
http://dx.doi.org/10.2801/566770

Cedefop (2019a). *Qualifications frameworks in Europe. 2018 Developments*. Briefing Note April 2019.
http://www.cedefop.europa.eu/files/9139_en.pdf

Cedefop (2019b). *FINLAND. European inventory on NQF 2018*.
https://www.cedefop.europa.eu/files/finland_-_european_inventory_on_nqf_2018.pdf

Cedefop (forthcoming). *Initial VET qualifications at EQF levels 3 and 4*. Contract 2016-0139/AO/DSI/SPEV/Qualifications_EQF_levels_3&4/011/16

Council of the European Union (2017). *Council Recommendation of 22 May 2017 on the European Qualifications Framework for lifelong learning and repealing the recommendation of the European Parliament and of the Council of 23 April 2008 on the establishment of the European Qualifications Framework for lifelong learning* (2017/C 189/03). Official Journal of the European Union, 15.6.2017.
https://ec.europa.eu/ploteus/sites/eac-eqf/files/en.pdf

European Commission; Cedefop; Educational Authority (2019). *EQF AG Peer Learning Activity on National Qualifications Databases*. 21–22 March 2019,

Budapest, Hungary. Summary report

European Union (2011). *Using Learning Outcomes. European Qualifications Framework Series: Note 4.* Luxembourg: Publications Office of the European Union.
https://ec.europa.eu/ploteus/sites/eac-eqf/files/EQF_note4_en.pdf

European Union (2018). *DECISION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a common framework for the provision of better services for skills and qualifications (Europass) and repealing Decision No 2241/2004/EC*.
http://data.consilium.europa.eu/doc/document/PE-70-2017-INIT/en/pdf

EVA (2013). *Evaluering af den danske kvalifikationsramme for livslang læring Status for implementering og anvendelse af kvalifikationsrammen* [Evaluation of the existing qualification framework for lifelong learning. Status of implementation and use of the qualifications framework].
https://www.eva.dk/sites/eva/files/2017-08/Evaluering%20af%20DKKL_implementering%20og%20anvendelse_web%20%282%29.pdf

DG EMPL (2019). *RfS 43 - Support to the Development of the ESCO qualifications pillar Final Report v7 - February 2019*. Framework Service Contract to provide expertise and support for European cooperation in education, training, youth and sport (EAC/07/2015). Authors: Karin Luomi-Messerer (3s), Tine Andersen (DTI), Frank Wilson (Ecorys) and Michael Blakemore (Ecorys).
https://ec.europa.eu/social/BlobServlet?docId=20788&langId=en

IBE (2017). *The pilot project on the horizontal comparison of levelled qualifications. Draft of the final report for discussion at the AG38 meeting*.
http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetailDoc&id=28730&no=11

Maheswari, M. (2017). Text Mining: Survey on Techniques and Applications. *International Journal of Science and Research* (IJSR). Vol.6, issue 6, pp. 1660-1664.

Moyes, N. (2019). *Connecting national databases to EU portals*. Presentation at the EQF AG PLA on 'National Qualifications Databases', 21st-22nd March 2019, Budapest, HU.

Pevec Grm, S.; Bjornavold, J. (2019). *Comparative overview of state of play of national qualifications databases - Based on data from Cedefop's NQF inventory 2018*. PLA Budapest, 21 March 2019

Prøitz, T.S. et al. (2017). Learning outcomes in professional contexts in higher education. *European Journal of Education,* Volume 52, Issue1, March 2017, pp. 31-43.

QQI (2014). *Common Awards System. Restatement of Policy and Guidelines 2014*.
https://www.qqi.ie/Publications/Publications/CAS%20restatement%20of%20

policy%20and%20guidelines%20FET.pdf

Sgarzi, M.; Debowsky, H. (2019). Including non-formal sector qualifications in National Qualification Frameworks. Céreq. *Training and Employment* n°137 May 2019. http://www.cereq.fr/publications/Training-and-employment/Including-non-formal-sector-qualifications-in-National-Qualification-Frameworks-a-challenging-process

Salloum, S., Al-Emran, M., Abdel Monem, A., & Shaalan, K. (2018). Using Text Mining Techniques for Extracting Information from Research Articles. Chapter in: *Studies in Computational Intelligence* (2018). DOI: 10.1007/978-3-319-67056-0_18.

Talib, R., Kashif, M., Ayesha, S., & Fatima, F. (2016). Text Mining: Techniques, Applications and Issues. *International Journal of Advanced Computer Science and Applications*. 7. DOI: 10.14569/IJACSA.2016.071153.

Willems, K. (2019). *Pandas Tutorial: DataFrames in Python*, available at: https://www.datacamp.com/community/tutorials/pandas-tutorial-dataframe-python

Annex 1. The research team

The following table provides a list of the research team who contributed to the study.

Name	Role
Karin Luomi-Messerer	Team leader
	Country expert – Austria
Simon Broek	Core team
	Country expert - Netherlands
Monika Auzinger	Core team
	Country expert – Austria
Maria Kargl	Core team
Andrew McCoshan	Core team
	Country expert – Ireland
Claudia Plaimauer	Core team
Marye Hudepohl	Thematic expert
Christopher Winch	Core team
	Country expert – UK-England
Mariya Dzhengozova	Country expert – Bulgaria
Søren Kristensen	Country expert – Denmark
Jouko Luomi	Country expert - Finland
Patrick Werquin	Country expert – France
Vidmantas Tutlys	Country expert – Lithuania
Oriol Homs	Country expert – Spain
[redacted]	

Annex 2. Online sources and examples used in prototype development

Topics (broad)	Topic (narrow)	URL
Data science - text analysis	Guide to dealing with Text Data (using Python) – for Data Scientists and Engineers	https://www.analyticsvidhya.com/blog/2018/02/the-different-methods-deal-text-data-predictive-python/
Natural Language Toolkit	NLTK official guide/book - practical introduction to programming for language processing	http://www.nltk.org/book/
Dataframes, pre-processing	Workarounds for text clean-up ('punctuations').	https://stackoverflow.com/questions/39782418/remove-punctuations-in-pandas
Dataframes, tokenization	Tokenization using dataframes	https://stackoverflow.com/questions/33098040/how-to-use-word-tokenize-in-data-frame
Dataframes, analysis	Wordcounts, frequencies, creating a matrix of word frequencies	https://stackoverflow.com/questions/43087420/how-to-count-word-frequency-from-a-pandas-dataframe-python
	(e.g. 'Unnesting'); Python Pandas NLTK Frequency Distribution for Tokenized Words in Dataframe Column with a Groupby	https://stackoverflow.com/questions/53528571/python-pandas-nltk-frequency-distribution-for-tokenized-words-in-dataframe-colum
	Coding example using pandas dataframe	https://stackoverflow.com/questions/40206249/count-of-most-popular-words-in-a-pandas-dataframe?rq=1
Dataframes, transforming	Extracting a pd.dataframe column into a list	https://stackoverflow.com/questions/17241004/how-do-i-convert-a-pandas-series-or-index-to-a-numpy-array
	Expand dataframe by adding token counts (as new column)	https://stackoverflow.com/questions/29791785/python-pandas-add-a-column-to-my-dataframe-that-counts-a-variable
Looping/iterating, building functions	'string' module tools *.maketrans()* and *.translate()* functions, how to use *(may be useful for building lambda functions for text cleanup or replacing across full set/column of characters)*	https://www.journaldev.com/23697/python-string-translate & https://www.geeksforgeeks.org/python-maketrans-translate-functions/
Pattern Matching, dataframes	String Matching on 2 Large Data Sets Based on a Condition	https://stackoverflow.com/questions/42847396/fuzzy-wuzzy-string-matching-on-2-large-data-sets-based-on-a-condition-python
	Matching tokens from columns from 2 pandas dataframes	https://stackoverflow.com/questions/37891131/using-fuzzywuzzy-to-create-a-column-of-matched-results-in-the-data-frame
Extracting data from dataframe	Extract rows if column value(s) meets condition(s)	https://www.interviewqs.com/ddi_code_snippets/rows_cols_python https://stackoverflow.com/questions/17424182/extracting-all-rows-from-pandas-dataframe-that-have-certain-value-in-a-specific https://stackoverflow.com/questions/29314033/drop-rows-containing-empty-cells-from-a-pandas-dataframe/56708633#56708633

Data Visualisation	Extracting a pd.dataframe column into a list	https://stackoverflow.com/questions/17241004/how-do-i-convert-a-pandas-series-or-index-to-a-numpy-array
	Extracting least common element (array =NOT= dataframe)	https://stackoverflow.com/questions/4743035/obtaining-the-least-common-element-in-array
	Python Graph Gallery	https://python-graph-gallery.com/
	Matplotlib Guide	• https://realpython.com/python-matplotlib-guide/
	Using matplotlib	• https://stackoverflow.com/questions/19555525/saving-plots-axessubplot-generated-from-python-pandas-with-matplotlibs-savefi • https://stackoverflow.com/questions/51204505/python-barplot-with-colorbar/51205723#51205723 • https://matplotlib.org/gallery/statistics/hist.html#sphx-glr-gallery-statistics-hist-py • https://matplotlib.org/api/pyplot_summary.html • https://stackoverflow.com/questions/332289/how-do-you-change-the-size-of-figures-drawn-with-matplotlib
	Using pandas plot function	• http://jonathansoma.com/lede/data-studio/matplotlib/changing-the-background-of-a-pandas-matplotlib-graph/ • https://pandas.pydata.org/pandas-docs/stable/user_guide/visualization.html • https://mode.com/example-gallery/python_horizontal_bar/ • https://towardsdatascience.com/a-guide-to-pandas-and-matplotlib-for-data-exploration-56fad95f951c • https://stackoverflow.com/questions/44493417/pandas-dataframe-bar-plot-plot-bars-different-colors-from-specific-colormap

*Note: In the case of references to Stack Overflow, there are generally multiple examples provided per issue. However, it is difficult to identify the specific comment / response that was used in the prototype, given that it needed to be 'pieced' together (i.e. the prototype was mostly inspired by the **discussions and comments** that were provided on the forum for specific topics, rather than using the examples 'as presented'.*

Abstract

To be developed

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce eget ipsum vitae risus bibendum consequat et vitae mauris. Duis velit velit, aliquet id egestas id, porttitor quis lectus. Aliquam eget pulvinar

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce eget ipsum vitae risus bibendum consequat et vitae mauris. Duis velit velit, aliquet id egestas id, porttitor quis lectus. Aliquam eget pulvinar sem