Deliverable Report

Agreement number: SI2.488704 (ECOKT2016-1)
Project name: Learner-centred digital ecosystem of competence development (Compleap)
Date:30.5.2019
Work package:WP3
Title: D25 Technical documentation, including glossary, model and architecture
Author:EDUFI

Description of the action:

This document describes the technical basis and implementation of CompLeap prototype application.

The CompLeap prototypes have been developed in the EDUFI test environment. The technical documentation presents the technological choices of the prototype. The framework architecture presents the conceptual strategic framework within which the prototype functions, and the vision towards which the prototypes act as a proof of concept.

The work has compiled team-wide planning and preparatory workshops for the framwork architecture as well as guided technology development with our prototype subcontractors. The glossary works as a key for interpreting both the framework architecture and the prototype functionalities.

Outcome of the action:

This document describes the technical basis and implementation of CompLeap prototype application.

The document first presents the Glossary, a brief overview of the general concepts behind the prototype. Following that, the core application components are described in more detail. Finally the Framework Architecture compiled in the project is presented.

The interoperable prototype modules presented in this document can be accessed through this link:

https://poc.compleap.testiopintopolku.fi/

Table of Contents

I Glossary

Agile methodology: Highly interactive and flexible management method with short-termed delivery cycles (sprints), agile requirements, dynamic team culture, less restrictive project control and emphasis on real-time communication

Authentication: The process of determining whether someone is who it declares itself to be

Strong authentication:

Competence : Ability to apply learning outcomes adequately in a defined context (education, work, personal or professional development)

CompLeap: A two-year project funded by the European Union, which aims to create solutions to improve the matching of supply and demand of competence. It draws its name from the words "Competence Leap". Through the service, the user can map their competence and build their own personal competence profile.

CSC: IT Center for Science

Decision makers: Decide if CompLeap have to be used as an facilitating tooling for end-users

Deployment: All of the activities that make a software system available for use

Education recommendation: Calculated and based on similarity of content between user's profile data and education descriptions

EDUFI: Finnish National Agency for Education, a national development agency. It is responsible for developing education and training, early childhood education and care and lifelong learning, as well as for promoting internationalisation. EDUFI is subordinate to the Ministry of Education and Culture and its tasks and organisation are set in the legislation.

End-user: A person who are going to use the CompLeap Framework

eRequirements: service that contains qualification requirements for the vocational qualifications for which an authorisation to provide education in English has been granted.

Facilitating users: who are going to help the end users and supports CompLeap

Formal learning: Learning that occurs in an organised and structured environment (such as in an education or training institution or on the job).

Framework:

Informal learning: Learning resulting from daily activities related to work, family or leisure. It is not organised or structured in terms of objectives, time or learning support.

Interest: Activities that one enjoys doing and the subjects that one likes to spend time learning about

Interested parties: Interested or involved in supporting Lifelong Learning

Interoperability : Ability of organisations to interact towards mutually beneficial goals, involving the sharing of information and knowledge between these organisations, through the business processes they support, by means of the exchange of data between their ICT systems.

Technical interoperability: means the ability of information and communication technology systems to interact so as to enable the sharing of information, achieved through agreement by all parties and owners of the information

KOSKI: The National Student Records (in Finland) comprising educational data on all Finnish students as well as exchange students in Finland and vocational student

Learner: Any citizen of EU, a role in CompLeap project

Learner path: Personal services for users to map competences throughout their life; includes a Competence Profile, which ties together user competences, skills and interests, with possibilities of guidance and comparing existing education opportunities

Lifelong learning: All learning activities undertaken throughout life, which result in improving knowledge, knowhow, skills, competences and/or qualifications for personal, social and/or professional reasons

Mock-up prototype: Visual screenshots of the modules for testing a concept

Neural networks: A set of algorithms that are designed to recognize patterns

Non-verified education: No authoritative repository of study records

Open source: Software that uses an open development process and is licensed to include the source code

Prototype: An early model of a product built to test a process

PoC, Proof of Concept: A demonstration, the purpose of which is to verify that certain concepts or theories have the potential for real-world application. POC is therefore a prototype that is designed to determine feasibility, but does not represent deliverable.

Python: An interpreted, high-level, general-purpose programming language

Qualification: Formal outcome (certificate, diploma or title) of an assessment process which is obtained when a competent body determines that an individual has achieved learning outcomes to given standards and/or possesses the necessary competence to do a job in a specific area of work. A qualification confers official recognition of the value of learning outcomes in the labour market and in education and training. A qualification can be a legal entitlement to practise a trade.

R: A programming language and free software environment for statistical computing and graphic

Skill: Ability to apply knowledge and use know-how to compete tasks and solve problems

Verified education: Authoritative repository of study records that contains information about learners' prior education is required

Word embeddings: Techniques which uses multidimensional floating point vectors to represent words where semantically and syntactically similar words are close together in geometric space. For every word in models corpus, there is exactly one vector. This approach should help learner to realise new areas where one's competences and interests could be combined in a new way.

II Technical Documentation

This document describes the technical basis and implementation of CompLeap prototype application. A brief overview of the general concepts behind the prototype is offered first. Following that, the core application components are described in more detail.

The prototype application demonstrates a learner profile that makes use of information about learner's skills, competences, and interests. With the aid of this information, learning opportunity recommendations are offered to help the learner realise new areas where the competences and interests could be combined.

On a logical level, the application therefore consists of three interlinked components: prior education and competences, current interests and aspirations, and recommendations for finding new ways to combine these. While each of the components demonstrates a particular viewpoint to CompLeap, the components also work in tandem to form a learner profile for the user.

Technology Overview

Technical documentation	https://github.com/Opetushallitus/compleap
Database description	Only browser cache
Open source code	https://github.com/opetushallitus/compleap
Development phase	Service is in proto/poc-phase.
Dependences	Tarjonta-service, Koski-service. NB. Service is using only test data.
Technologies	JavaScript, React, AWS (infra), Scala
Release notes	https://jira.oph.ware.fi/jira/projects/CL?selectedItem=com.atlassian.jira.jira-projects-plugin:release-page
Bugs	https://jira.oph.ware.fi/jira/secure/RapidBoard.jspa?rapidView=163
Link to service	https://poc.compleap.testiopintopolku.fi/
Licence	EUPL (Lisensiointi only in Finnish)

Technical Prototype Descriptions

1. Prior education

The prototype for prior education demonstrates two different scenarios: verified education and non-verified education. The distinction between the two is that in the verified education scenario, a central, authoritative repository of study records that contains information about learners' prior education is required, whereas in the non-verified scenario this requirement is removed.

In the prototype, ***Koski,*** the Finnish national repository for study records, is used as a demonstrative example of a provider of verified education information. Koski repository contains study records for verified education done in Finland for multiple different levels of education. These records can be used together with textual descriptions of the studies, which offers a way to examine learner's competences. In the prototype, the usage of the repository is demonstrated with a mock implementation and test learner profiles that correspond to key use cases.

For non-verified education, the lack of such authoritative repository must be compensated for. This is essential for example in the use case where a learner has done their studies outside of Finland – e.g. when the user is an expat. In this case, the user can input the prior education in more generic terms by selecting the appropriate level and domain of the education. This information is then used in a manner roughly similar to the verified information.

The verified education scenario also offers an additional level of customisation for the user. Whereas fetching the learner's prior education data is an automated process, the way in which this information is used for the recommendations can be customised by the learner. The learner is offered an option to mark certain units either "liked" or "disliked" by a simple thumbs up / thumbs down voting mechanism. This provides the learner an opportunity to adjust the profile and enrich the prior education information by e.g. emphasising topics they would like to pursue more deeply in the future or by marking themes they would like to steer away from.

2. Current interests

The second logical component – gathering user's current interests – complements the education information. The aim of this prototype component is to enable the learner to express their aspirations more fully.

The interests are gathered by first presenting the user a wide range of potential fields of interest. From this pool of topics, the user can select the relevant ones. For each selected top-level field of interest, more relating topics are displayed. This allows the user to elaborate on their interests. As a result, an interest profile is formed that describes learner's motivational topics that may have their basis on e.g. the non-formal activities such as hobbies, in addition to formal education.

3. Learning opportunity recommendations

The final component – the recommended education opportunities – ties together the previous two components and completes the learner profile by offering the user meaningful insight based on their profile data. This prototype component demonstrates a concrete use case for the learner profile: by filling in the information, the user is offered recommendations for the next step on their learner path.

The recommendations use semantic matching for finding opportunities that could offer the learner new ways to combine their competences with their interests. The most suitable matches are

presented for the user. The user can then examine the matches further, and also do some additional filtering such as restricting the results to a particular province.

Application Components

In terms of source code, the prototype application is divided into two main components:

- Client-side application (model directory)
- Recommendations application (src directory)

As such, on a physical level, the system has a conventional client–server architecture that consists of two main applications: the client-side application with the graphical user interface and the server-side application with the recommendations model. These two main application components implement together the core concepts described above.

The user-facing client-side application provides the front-end of the system. The application is implemented with the single-page application model, which offers good support for highly interactive applications. A substantial amount of application logic is handled in the client-side application, and additionally, the user profiles together with mock data for verified education are also packaged into the application. This eliminates the need for separate backend processing in the prototype implementation for areas other than the recommendations.

The core of the recommendations application is the semantic matching of user profiles to learning opportunities. The rest of the application consists of a simple application programming interface (API) wrapper, which exposes the recommendations service as a REST-style endpoint. This allows the recommendations model to be queried over HTTP protocol in the client-server architecture in an established manner.

The Client-side Application

The main responsibility of the client-side application is to provide a graphical user interface for the service and coordinate the distinct prototype features: the utilisation of prior education data, user interests, and learning opportunity recommendations.

The application is implemented in JavaScript, making use of modern ECMAScript 2015 features. This provides an easily readable example implementation of the prototype. For increased runtime compatibility with different web browsers, the application is transpiled into an older version of ECMAScript.

The application uses React[1] as the view library for the user interface. React is a highly common view library that makes both building user interfaces and reading the implementation code easy due to its declarative nature. An essential part of the client-side application is the state handling of the learner profile. To make examining this aspect straightforward, the application uses state charts using xstate[2] to describe this behaviour.

[1] https://reactjs.org/

[2] https://xstate.js.org/

In addition to the user interface and business logic, the application contains predefined test user profiles and a mock integration to the Koski repository, which provides test study records. This demonstrates how a centralised repository can be used for users' verified education information. This information is communicated to the recommendations application.

Similarly, the client-side application bundles a set of topics that are presented to the user in the user interface in order to gather user's interests. The application keeps track of user's selection and stores this in the user's learner profile. This information is communicated to the recommendations application along with the education data. The basic structure of the prototype application is shown below (Figure 1).

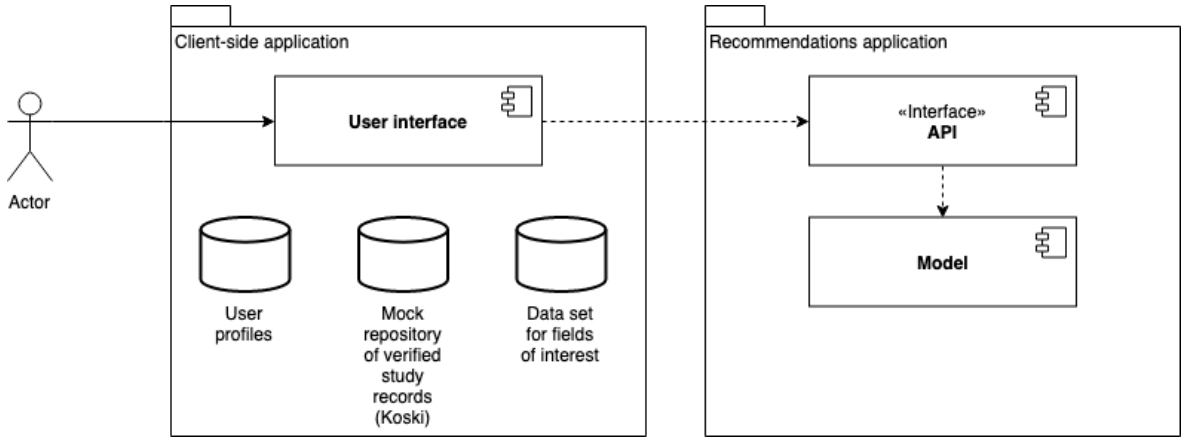


Figure 1. Structure of the prototype application

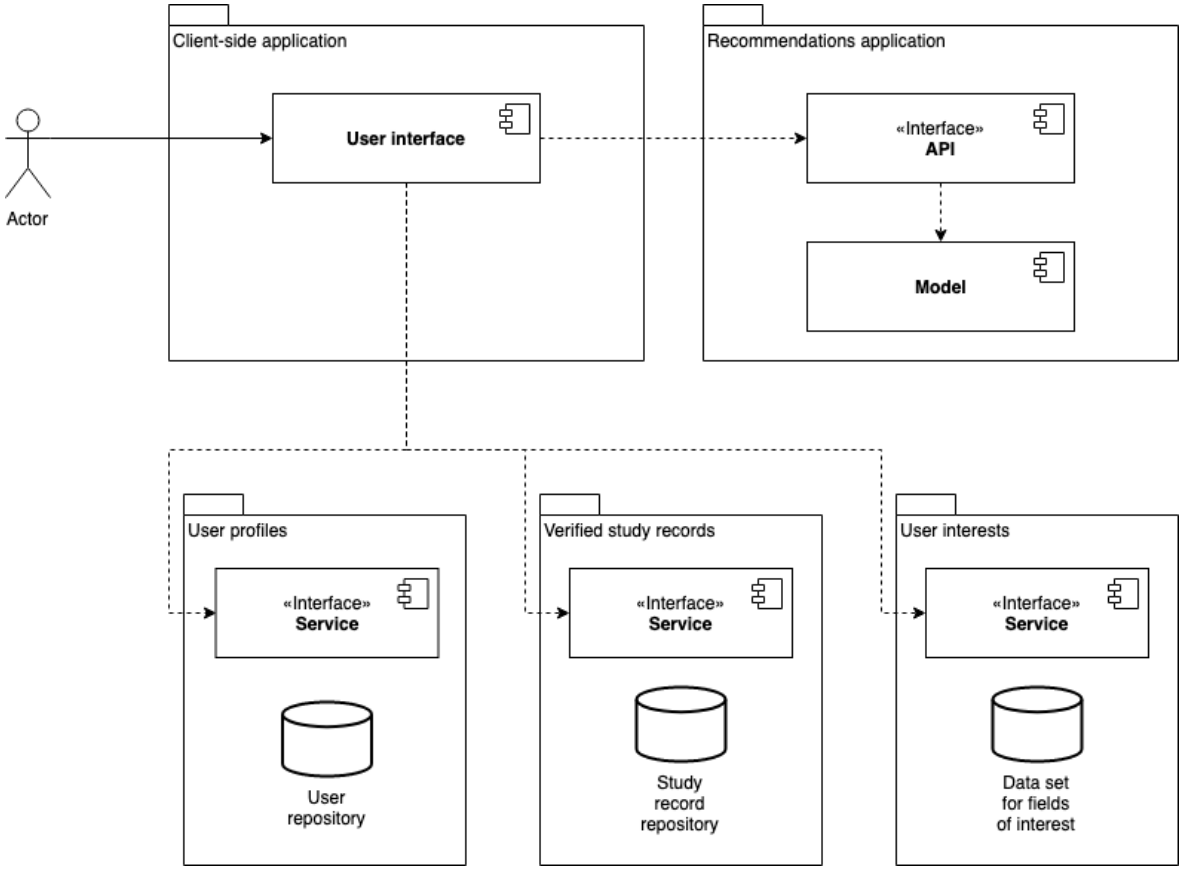
When the user has filled in enough detail to the learner profile (i.e., selected at least the specified number of interest topics, and optionally e.g. adjusted the study units by marking them with thumbs up or down attributes), the recommended learning opportunities are negotiated with the recommendations application and the results are presented to the user. The user can continue adjusting the profile information, which will update the profile and recommendations accordingly. In the prototype, the learner profile is persisted only in the client-side application, i.e., in the web browser's internal storage.

Generalising the application structure

In the prototype, test user profiles are included in the application, as described above. The test profiles can be used with a mock log-in flow, which enables the piloting users to try out different scenarios. This log-in flow could be replaced with a strong authentication method to support the use of actual personal information. With such modifications, the application would be able to act in behalf of the user when accessing e.g. personal study records from a production repository.

The study record repository is contained within the application as a mock implementation in a similar manner. The fetching of the study records is wrapped as an internal service, which acts as a variation point for possible implementations in another environments or locations. For example, when replacing the mock implementation with data from a production repository, this service would be modified accordingly. The modifications could include e.g. specifying a different endpoint for fetching the data or communicating with a mediating service.

Some deployments could also benefit from modifications more substantial than the ones described above. In such cases, the general approach can be supported with different underlying implementations. For example, supporting a study record repository other than the Finnish Koski repository would involve two distinct aspects: First, the internal education data model of the application would we modified to match the structure of the repository data. Second, the recommendations component would be modified to cater for the different data. The fields of interest could be adjusted in a very similar way: the data set used in the prototype application could be replaced with another one and the model could be modified correspondingly. The generalised



structure is shown below (Figure 2).

Figure 2. Generalised structure of the prototype application

The Recommendations Application

The idea is to use semantic matching recommendations using sophisticated, neural networks based, natural language processing technique (NLP) called ***word embeddings*** (Mikolov et.al., 2013).

Word embeddings based techniques are superior compared to keyword based searches by catching semantically and syntactically similar words, and documents, not only exact or partial matches to given search parameters. This approach should help learner to realise new areas where one's competences and interests could be combined in a new way. Additionally, it gives us a possibility to use a large wealth of information included in descriptions of verified competences. The idea is to use as much information as possible from verified competences to get a comprehensive ***competence***

profile of learner's skills, competences and interests. That competence profile is used for recommendations.

Recommendation application takes information in through API from client side, uses pre-calculated document embedding vectors to find semantics similar learning opportunities, and returns them back to client side application (Figure 3).

In this specific prototype case, we gathered information about completed study program units (i.e. verified competences) from eRequirements –service, that gathers information about national qualification requirements in one place. For demonstrative purposes we included only vocational education and qualifications, because they have the most coherent and complete information available in Finland. However, it is still quite easy to add other learning opportunities too.

The current prototype is developed using R. There are some existing packages for word and document embeddings for R. However, Python has more established and scalable packages for production use. The initial reason choosing R over Python was its familiarity to the development team and considerably short time period reserved for development of recommendation prototype.

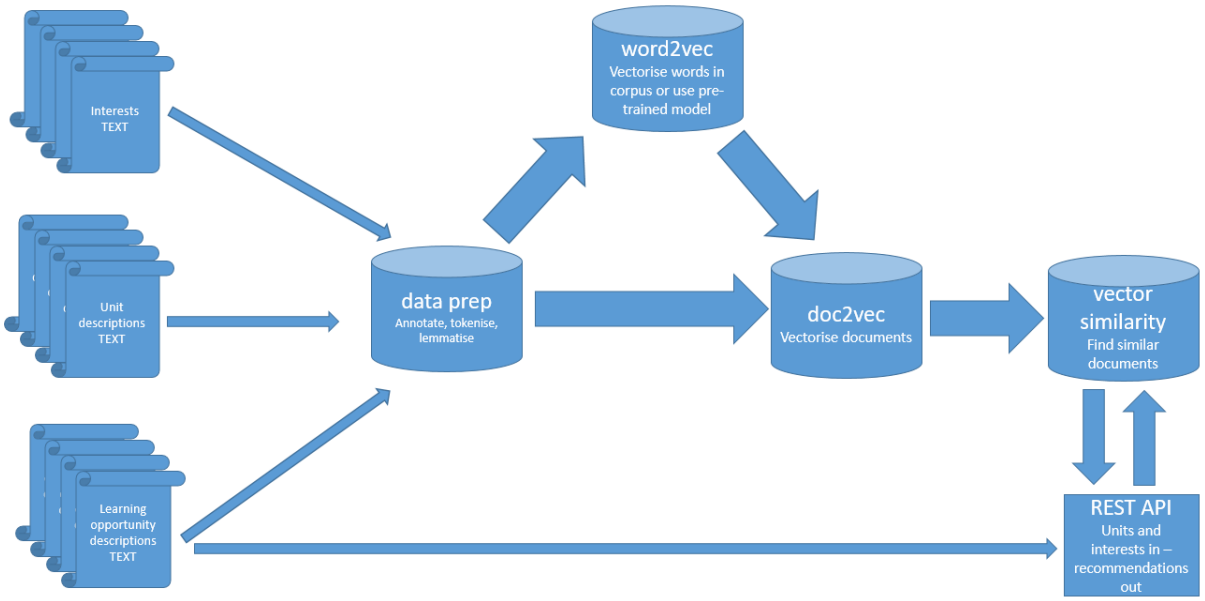


Figure 3. Overview of components in recommendation application prototype

Data preparing and embedding models

Typically, every NLP tasks starts by making a corpus of all words in the input text. It is the same in this case. In addition, the prototype uses pre-calculated document models both for inputs (units and interest terms) and outputs (learning opportunities), those need to be prepared beforehand. It is possible to do it on-fly but it would decrease the performance in the client side application – that's why it is wise to calculate all that is possible beforehand.

Corpus

In this case all of the input materials are in Finnish. Finnish is not an easy language for NLP applications because it's complex morphology and having relatively low number of native speakers. The text needs to be annotated, tokenised and lemmatised. For this tasks we used Finnish

dependency parser pipeline (Kanerva et.al. 2018, Pyysalo et.al. 2015). Whether to use lemmas, stems or full words depends on task at hand, corpus size and word vector model. In the prototype, we have tried both lemmas and full words, and tests are still ongoing to find the best performing model and pre-processing option.

We made separate corpuses for all data sets used: program study unit descriptions, learning opportunities and interests. Corpuses are needed for training of word embedding model and making document embeddings.

Word embeddings

Word embeddings uses multidimensional floating point vectors to represent words where semantically similar words are close together in geometric space. For every word in models corpus, there is exactly one vector.

There are different options for word embedding vectors. It is possible to train your own vectors based on corpus or use pre-trained models, freely available on the internet For example, fastText (https://fasttext.cc/) has multi-lingual pre-trained models including one for Finnish. Pre-trained models may offer large corpuses but still it is possible to miss some domain specific terms.

In the prototype we have tried both pre-trained models and built our own using corpus from study program units and learning opportunities descriptions. Training your own word vectors need some knowledge about the technique. In our prototype we have use the original word2vec algorithm (Mikolov, 2013) with Skip Grams (i.e. every word is predicted based on its surroundings). Skip Gram should work better with small amount of data and rare words. The number of nodes in neural network define the dimension of the vector. It is typical to use networks with nodes between 50 and 300. We have used 200 nodes which gives vectors of same lengths.

Also, we have experimented with pre-trained models of Finnish internet parse bank which have large word2vec models both for lemmatised and full words (Luotolahti et.al. 2015). Using large pre-trained models put some requirements for hardware and development environment. For the testing purposes, we built a version of the prototype where results of different recommendation approaches can be compared. Evaluation is still going on and the best technique will be selected after that.

Document embeddings

To semantically match larger entities than words, we need to be able present sentences, text paragraphs and full documents in vectors. There are couple of techniques how to vectorise documents but basically they are founded on word embedding vectors. A simple approach is just to take all words within the document, find their corresponding word vectors from pre-trained vector model and take an average of those, which results in one document vector for the document in case. Also, there are more sophisticated methods that take into account words frequency and rarity, like document vectors through corruption using doc2vecC algorithm (Chen, 2017).

In the prototype, we made documents vectors simply by averaging over all words in the documents. Document vectors were made for all different vocational education and qualification program unit descriptions (over 8 000 in total) and for descriptions of all different vocational education opportunities (124 in total).

Recommendations

In the prototype, vocational educations and qualifications may have one or more competence areas under them. If that is the case, the recommended unit is the competence area. Altogether we had

124 recommendable programs and those are offered in 2 890 different applicable competence areas or qualifications in vocational education institutions.

Recommendations are based on semantic matching of learner's vectorised competences and interests to similarly vectorised learning opportunities. The same word embedding model must be used in training of document vectors for all data sets for the results to be meaningful. After that, finding similar documents in vector space is quite straightforward by calculating cosine distance between vectors. This is called **cosine similarity**.

Several possible approaches for the best solution to combine recommendations is still under test in the fore-mentioned recommendation testing version of the prototype and the alternatives include:

1. Sum up the thumbed up study program unit vectors and interest term vectors to one vector, which is used to find the best matching learning opportunity vector
2. Sum up the thumbed up study program unit vectors and interest term vectors and subtract thumbed down study program unit vectors from them to have to one vector which is used to find the best matching learning opportunity vectors.
3. Calculate average of best matching learning opportunity vectors calculated separately for each thumbed up study program unit vectors and selected interest term vectors
4. Find separately for all thumbed up study program unit vectors and interest term vectors the best matching learning opportunities, put them together, order by decreasing cosine distance and return the best matching.

Finally, all the functionalities of recommendations and meta information wrapping are packaged behind REST API. After successful API call, recommendations including meta information about learning opportunities and institution specific application info are returned in JSON format to client side application.

References

Chen, M., 2017. Efficient vector representation for documents through corruption. *arXiv preprint arXiv:1707.02377*.

Kanerva, J., Ginter, F., Miekka, N., Leino, A., Salakoski, T., 2018. Turku Neural Parser Pipeline: An End-to-End System for the CoNLL 2018 Shared Task. *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*.

Luotolahti, J., Kanerva, J., Laippala, V., Pyysalo, S., Ginter, F., 2015. Towards Universal Web Parsebanks. *Proceedings of the International Conference on Dependency Linguistics (Depling'15)*.

Mikolov, T., Chen, K., Corrado, G. and Dean, J., 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Pyysalo, S., Kanerva, J., Missilä, A., Laippala, V., Ginter, F., 2015. Universal Dependencies for Finnish. *Proceedings of NoDaLiDa 2015*.

III CompLeap Framework Architecture Design

1. Background

CompLeap Framework Architecture i.e. reference architectures purpose is to visualize the aims of the project in the contextual level, describing users and services in conceptual level, creating learners pathway and designing the most wanted parts of application architecture in logical level. The framework could be deployed across the EU-member state countries. The framework architecture could be used as a help when developing digital services for citizens

It is one of the key outcomes for the project: to study and develop an integrated and holistic learner-centred digitalised ecosystem framework that will look beyond existing, often siloed, structures

1.2. Enterprise Architecture as a Method

Enterprise architecture (EA) is a coherent planning method and a uniform way to describe practices and models for organizations in different stages of the development process. The objective of the EA work is to improve interoperability of activities and services of public administration and private organizations.

EA is a structure that defines an entity consisting of activities, processes, services, data and information systems as well as the services produced by them. The EA method is a systematic working method and procedure for identifying, analyzing, planning and describing elements of a given entity and their interdependencies.

In general, the main reasons for using EA are putting the strategy on live and visualizing future state of an organization. EA is also a tool for decision making, communication and co-operation. Thus, it realizes interoperability in all architecture levels from contextual level to implementation.

The main purposes of the CompLeap Framework Architecture are listed below:

1. Visualizing the aims of the project in contextual level

We modeled and visualized a strategy map which includes our strategic goals, strategic outcomes and capabilities that we need for achieving the goals and outcomes of the project.

2. Describing users, stakeholders and services in conceptual level

With the help of business architecture, we described stakeholders and main actors and their roles in order to map out who are the customers and users of the CompLeap services.

3. Creating Learner's pathway in conceptual level

The learner's pathway can be seen as a value stream which increments learner's competences and, thus, value in a way. While the value stream is described from learner's perspective, the services that the learner uses on the path are produced by other actors.

4. Designing the most wanted parts of application architecture in logical level

The purpose of the layered architecture views is to show what kind of information systems and application services we need for realizing new digital services – competence mapping, guidance et cetera.

5. Interoperability

Interoperability is a crucial point in all levels. We have paid special attention to designing integrations and data flows between the most important information systems which in return helps us to understand and notice the necessary technical interfaces and APIs.

2. Way of Working

With the above perspectives, we created the CompLeap framework architecture. We modelled and visualised the strategy map, which includes strategic goals, strategic outcomes and capabilities, abilities what we need for achieving the outcomes and goals.

In business architecture we described the stakeholders and main actors and their roles to realize who are the customers and users of the CompLeap services. Also, the learner's path was designed and the main business services utilized by the learner was identified.

Beside these, we modelled the most needed CompLeap services and their elements relationships in different architecture layers (business, information and application). Our application integration model describes the importance of collaboration of different information systems and shows in a very concrete way the importance of the co-operation between organisations. If we cannot break down silos and co-operate across different administrative branches and between public and private sectors, we will not have enough power and ability to develop useful and usable services for learners. And without proper services we cannot succeed in the digitalised world.

EA is a crucial tool for planning and achieving the learner-centred digitalised ecosystem of competence development. We can not only be focusing on digital business optimization i.e. improve productivity or just try to improve better customer experience but put a lot of power into the development of real digital business transformation – new products, services and brand new business models.

3. Methods and Tools

Enterprise Architecture is a coherent planning method and a uniform way to describe practices and models for organizations in different stages of the development process.

The project has used JHS 179 as a framework tool. It is a Finnish public sector recommendation for enterprise architecture planning and development. The recommendation is to a great extent based on TOGAF® (v. 9.1), which is the most common international open enterprise architecture framework used in both the public and the private sector.

Used language ArchiMate® is an open and independent modeling language for Enterprise Architecture, supported by different tool vendors and consulting firms. Users of ArchiMate benefit from sharing a common language.

Used modeling software is QPR EnterpriseArchitect by Finnish vendor QPR.

As a way of visualising the architecture and vision further, a s a HTML-prototype has been developed to exemplify the user flow and the full architecture: **https://compleap-proto.testiopintopolku.fi/**

The HTML prototype works together with the architecture images, bringing to life the vision level of the planned interoperable framework. Further information about the HTML-protype can be found in CompLeap Deliverable 20: Running prototypes in cooperation with WP3.

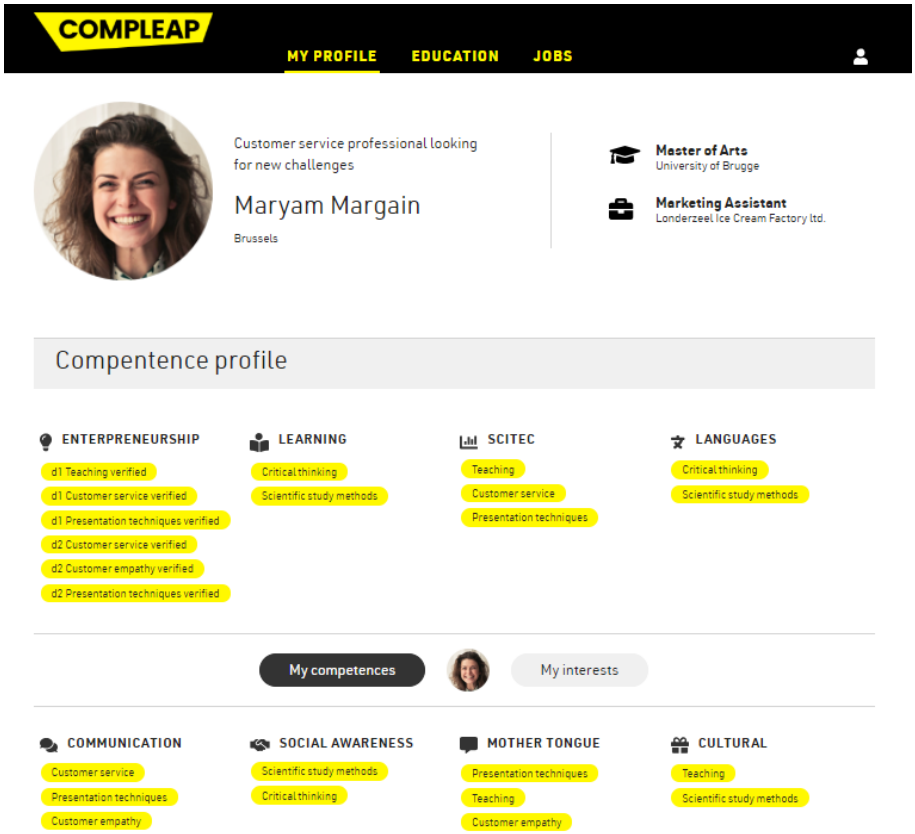


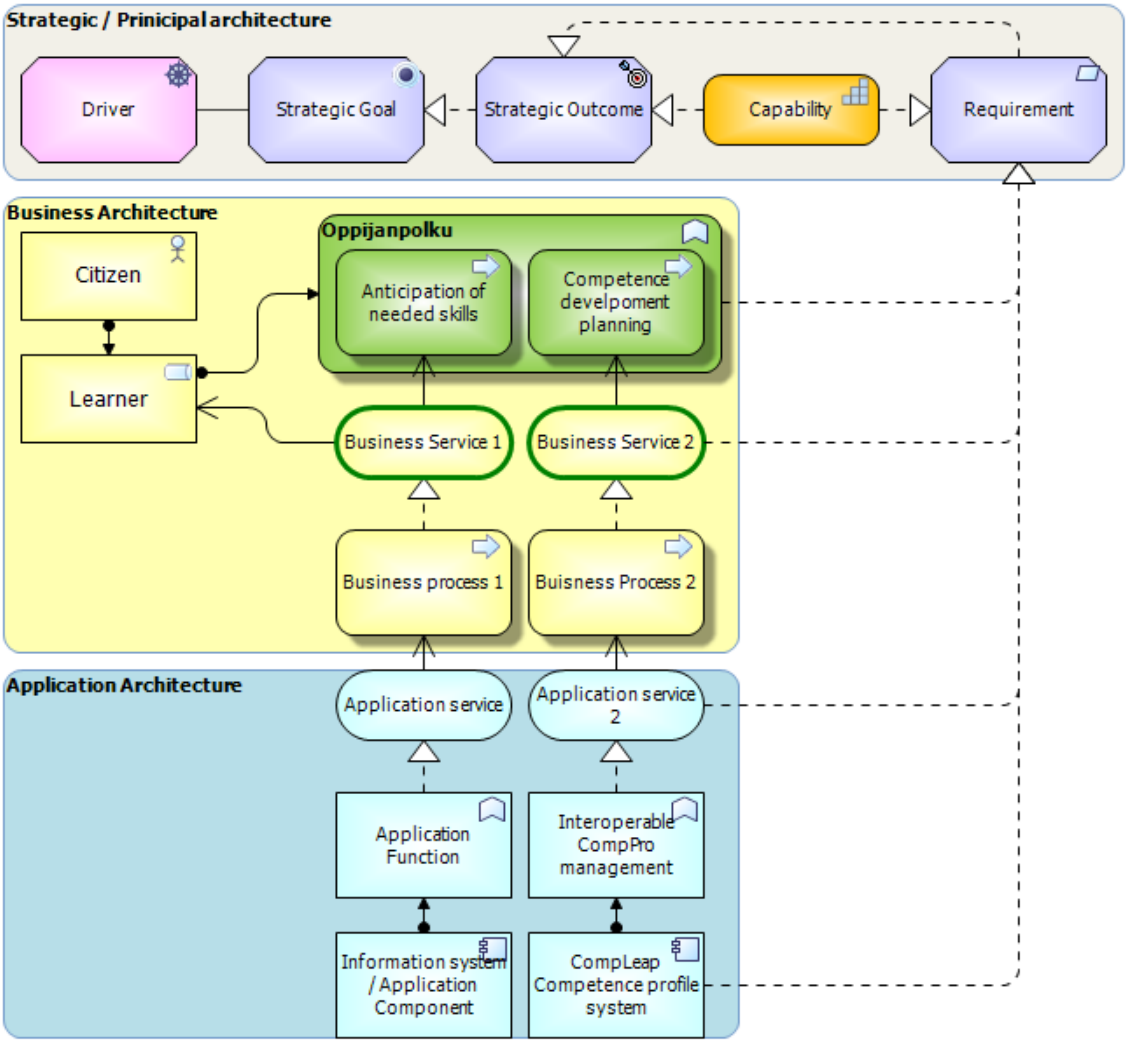
Figure 4: Screenshot form HTML-prototype depicting Competence Profile

4. Purpose

The framework architecture's purpose is to describe and visualize learner's pathway and available digital services at a very generic level so that the framework can be deployed across the EU-member state countries. Such an ecosystem framework could be used as a roadmap to help and guide in developing digital services for citizens.

The learner's pathway can be seen as a value stream which increments learner's competences and, thus, value in some sense. While the value stream is from learner's perspective, the services the learner uses on the path are produced by other actors. The services the learner uses are pinned on the path to tell when the learner is able to use that particular service. By expanding achitecture of the services the framework guides what is needed to implement the services and it is up to a user of the framework architecture to define how the services are actually implemented. To guide in implementation the framework architecture includes an example how this framework is implemented in Finland.

5. Conventions in the document - Compleap Framework Architecture Meta Model



The metamodel describes conventions and structure of the CompLeap framework architecture. It consits of three architecture levels: strategic, business architecture and application architecture.

1.Strategic level

a) drivers represent an external or internal condition that motivates our project to define its goals and implement the changes necessary to achieve them.

b) strategic goals are reflecting and responding to drivers

c) strategic outcomes realize strategic goals

d) capabilities represent abilities that we need to realize the business outcomes

e) requirement represents a statement of need that must be met by the architecture

f) business and application architecture levels realize the needed requirements

2.Business architecture level includes

a) actors and their roles i.e. our customers

b) descriptions learner's path i.e. life cycle,

c) business services used by learner

d) business processes of education providers realize business services

3.Application architecture level includes

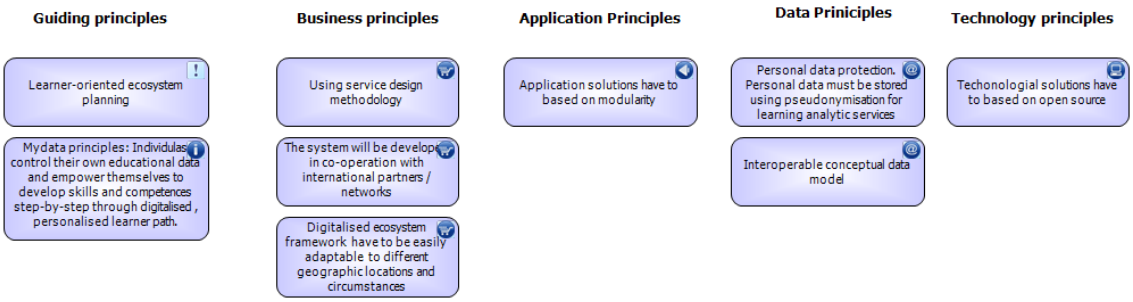
a) needed application services used by business process

b) application functions realize the application services

c) information systems provide application function and application services

6. Architecture Principles

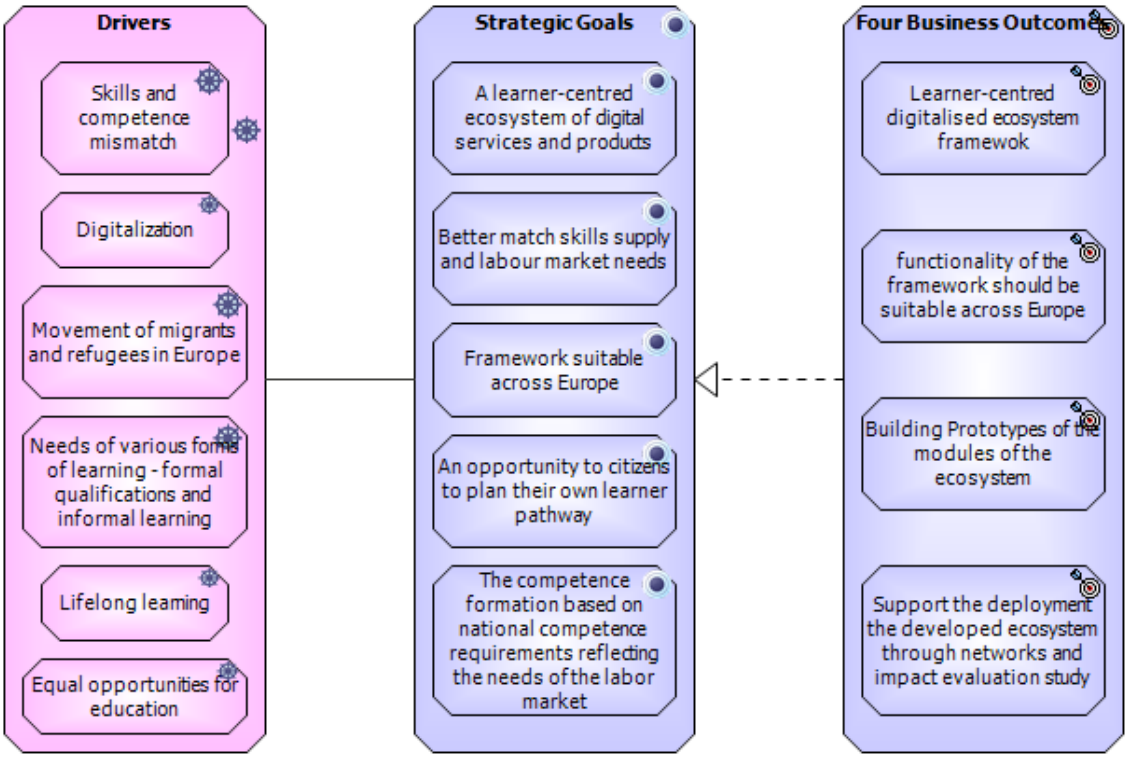
Architecture Principles



Architecture principles guide organization development towards strategic goals. Every development decision should be aligned with principles. The power of principles is that everyone can utilized them in their everyday work and work towards strategic goals.

The project aims to build a learner-centred ecosystem of digital services and products around skills and competences to better match skills supply and labour market needs, serving individual citizens, employers, decision-makers and, ultimately, the society at large. More specifically, the project aims to empower the citizen to take ownership of their skills development by offering them an opportunity to plan their own learner pathway based on enhanced self-awareness, improved recognition practices and responsive education offer. Furthermore, the skills formation will be based on national competence requirements reflecting the needs of the labor market.

6.1. Strategy Map

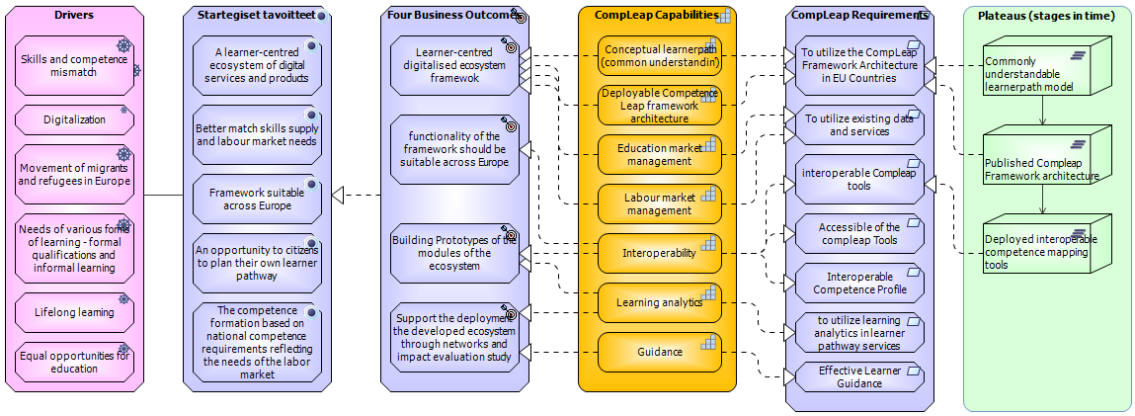


Strategy map links drivers, strategic goals, business outcomes and capabilities together. Drivers motivate for business shift and strategic goals are long term development targets meeting the environment change. Business outcomes are concrete results realizing the strategic goals in short term development.

Drivers represent an external or internal condition that motivates our project to define its goals and implement the changes necessary to achieve them. Strategic goals and and outcomes are aims of the project. Drivers, goals and outcomes are directly from EU application documents.

Description of digitalization driver. Digitalization consist of digital transformation and digital optimization: a) digital transformation: brand new services and products, new operating models and new business models b) digital optimization: improved productivity and better customer experience.

Strategy Map with Needed Capabilities and Development Requirements

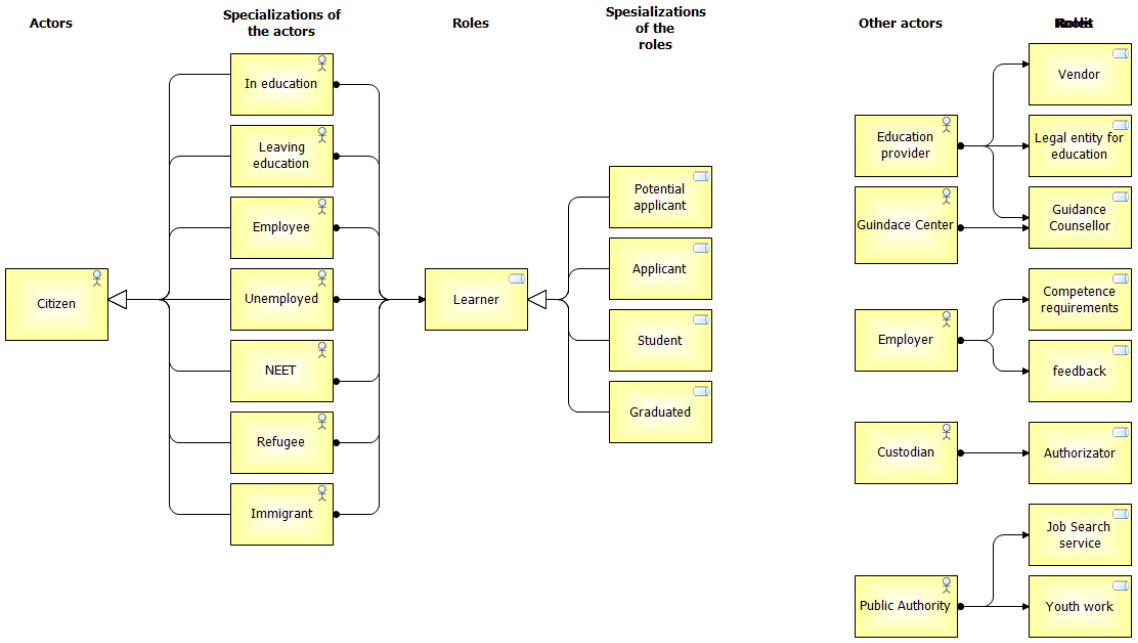


Needed capabilities tell what capabilities organization needs to realize business outcomes. Capability appears when its different dimensions are developed. The different dimensions are people, processes and material. Development of people dimension includes development of personal competences, development of processes include processes, concepts and information management and development of material dimension includes infrastructure, information technology and equipment. CompLeap project development requirements describe what is needed to increase the recognized capabilities.

7. Business Architecture

Business architecture describes business and operative functions of the project. Business architecture represents real world aspects and interaction between people. It includes actors and their roles, learner's path i.e. life cycle, business services and business processes. Business processes realize business services.

7.1. CompLeap Actors and their Roles

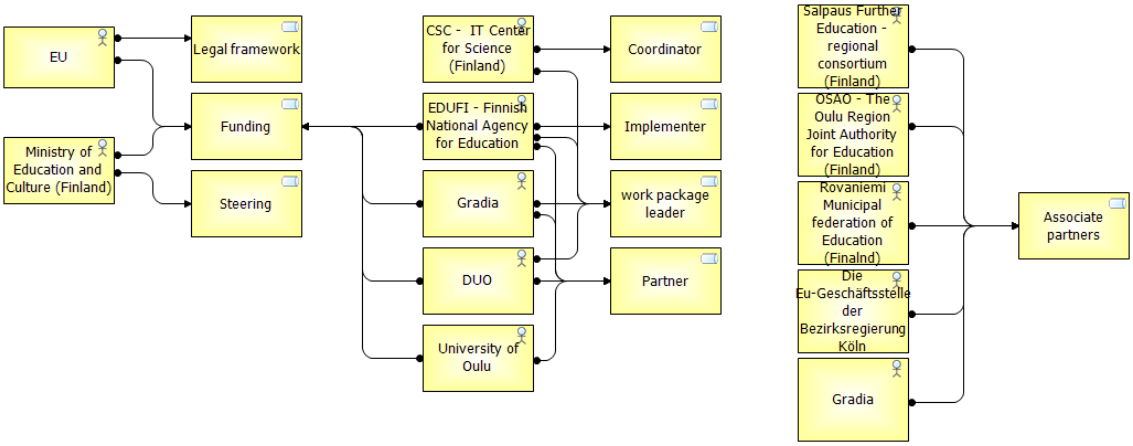


The learner's pathway actors and roles are described above. Learner is the center role in the framework architecture and the the learner acts in different specialized roles in the leaner's pathway. The learner can potential applicant, applicant, student or graduated depending on at what stage the learner is in the learner's pathway. Learner can be potentially any citizen of European Union and the citizens are divided into different categories to focus services for them depending on their life situation. From the value stream point of view main actors producing more value to the learner are public authorities, education providers and employers. Education provider offers services to increase learners competencies and employer provides possibility to utilize the competences and turn competences to actual value.

Education provider acts in different roles, namely in vendor, legal entity of education and guidance, depending on what education provider's service the learner is using.

Employer acts in different roles to provide competence requirements and feedback.

7.2. CompLeap Project Actors



Compleap project actors are described above. From the framework point of view the roles are meaningful and the implementer of the framework should consider what roles are required to implement the framework and what are the actors or stakeholders in each role. A stakeholder analysis and stakeholder management plan has been implemented in the CompLeap project and is available in Deliverable No. 35: Stakeholder management plan.

7.3. Business Service Map

Framework's business services are described above. In this framework business service produces value to the learner. The learner starts using the business service by contacting the actor producing the service. The contacting may occur via different channels which are up to an implementer to define. Some of the services require that the learner has to give some information as input to the service and the learner gets information back. Some of the services produce information for the other services to be utilized in the learner's pathway and some services give actual value increment to the learner. The while the business services may be fully automated there is still need for person-to-person interaction and, thus, service design should take other actors in account when applications are defined.

Compleap project focuses on the business services with green ovals around, namely Career and study guidance, Education offer, Comparing opportunities and Competence mapping. The project aims to create application prototypes for the named business services and the applications gives an possibility to use services via World Wide Web. The framework architecture and its implementation helps prototype implementation to give holistic view of all services and possible data streams so that data can utilized everywhere in the learner's path and the learner does not need to input same data many times on the path.

Guidance and Personalisation service area includes business services to help individuals to identify their competences and what competences they are missing and what course of actions they should take to improve their competence.

Admission service area includes business services to admit into education or other competence development.

Learning and studying service area includes business services to improve competence.

Feedback service area includes business services to give feedback about the competence development at any stage.

Employment service area includes business services to collect information about employment offer. This area could contain business service for applying to work but it is left out intentionally because of the perspective and scope of the framework.

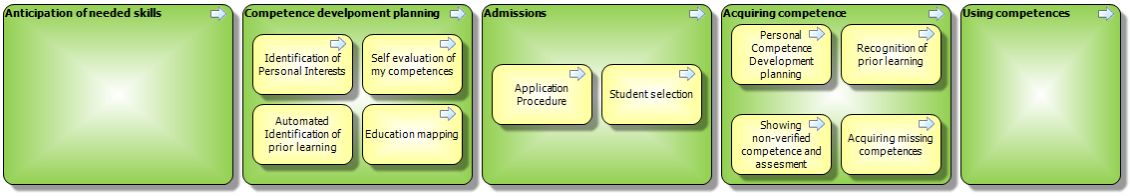
Identity and myData service area contains business services to identify learner, to manage learner's personal data and consents.

Statistics and analytics service are contains business services to manage statistical and analytical data to provide information to different management levels.

7.4. Digital Learner Pathway - Target Phase

Learner's pathway is described above. The learner's pathway is main business process or value stream in the framework. The value stream defines the steps the learner takes to increment hers or his competence. The value stream begins from anticipation of needed skills where the learner somehow recognizes the need for new skills or competences. The learner may use guidance at every stage of the leaner's pathway and the guidance could give impulses to the learner at every stage to guide the learner to right direction. After the anticipation of needed skills the learner may do planning or do pre education personal competence development planning. The difference between the planning and pre education competence development planning is that the planning is not well structured process while the pre education personal competence development planning process is. The planning is more ad hoc and flexible and depends on the learner how to planning is done or if it is done at all. The pre education personal competence development planning is defined by legal entity of the education and is formal way to do planning at that stage. After planning the learner does education mapping and after finding proper education, the learner enters to admissions. After admissions the learner acquires competences and finally uses the competences. During acquiring the competences and using the competences the learner does personal competence development planning all the time. The personal competence development planning is also formal process defined by legal entity of the education which guides the learner to do planning all the time. While the learner may do other kind of planning also, the formal planning provides common tools and information to do planning in competence development scope.

7.5. Digital Learner Pathway - Main and Sub Processes



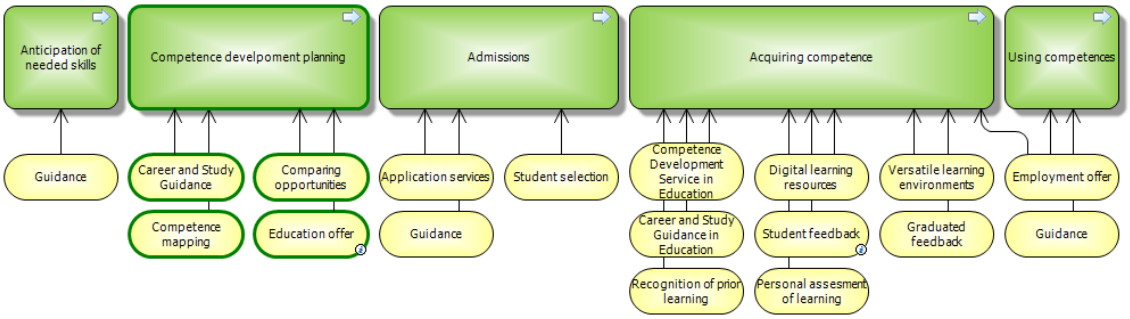
Learner pathway's sub business processes are described above. The sub processes clarify the main processes and tell in more detail what steps the learner takes in the learner's path.

Planning process is divided into three steps. Namely, identification of personal interests, self evaluation of my competences and automated identification of prior learning.
The sub processes are described in more detail later.

Admissions are divided into application procedure and student selection sub processes.

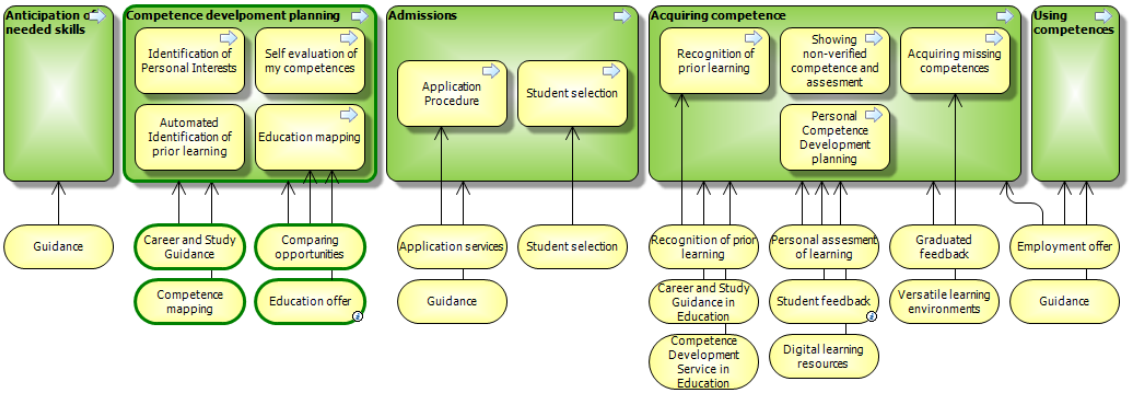
Acquiring competence is divided into recognition of prior learning, showing non-verified competence and assessment, and acquiring missing competences sub processes.

7.6. Digital learner pathway: Main Business processes and service



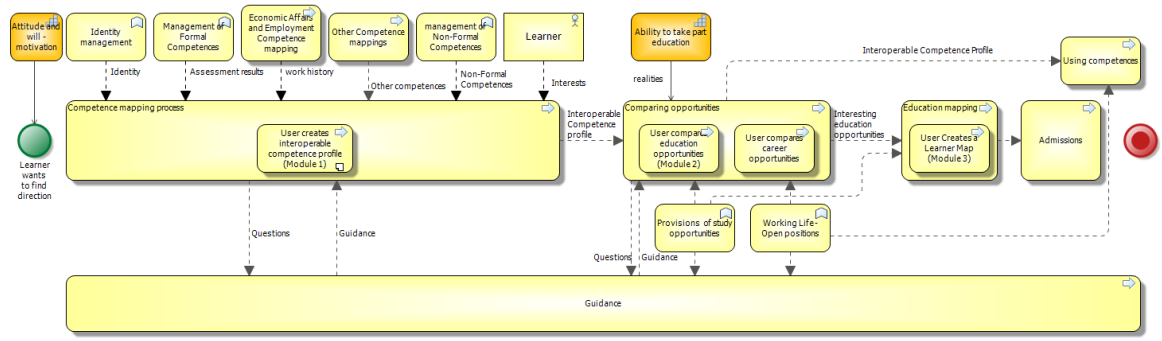
Learner's pathway supporting business services are described above. The business services are pinned to the learner's pathway using the arrow notation. The notation describes at which stage of the learner's pathway the business service is utilized. For example learner uses competence mapping business service during planning process.

7.7. Main and Sub Business Processes and Services

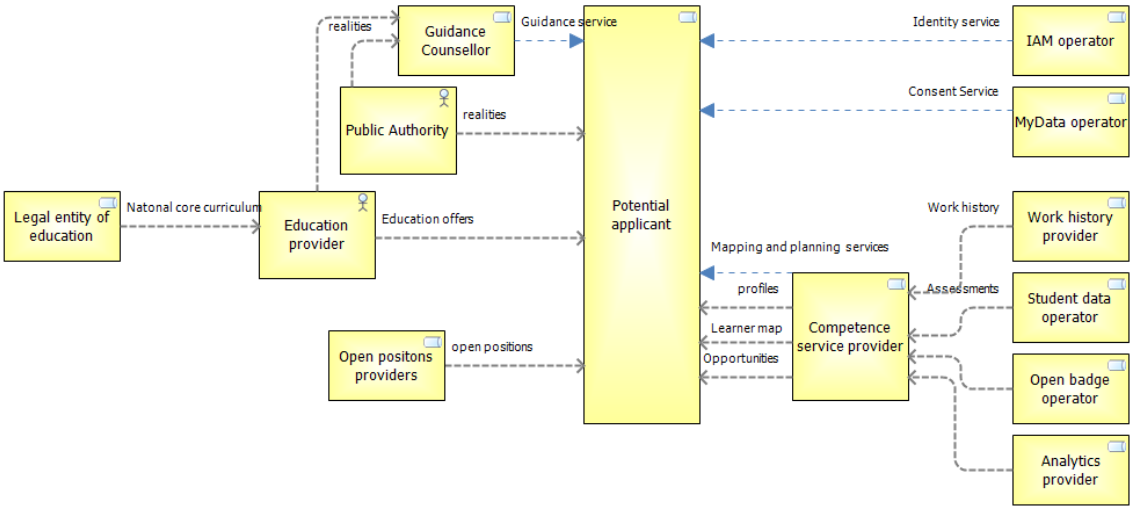


7.8. Process Integration

Framework process integration is described above. The diagram shows how data concepts flow between processes.

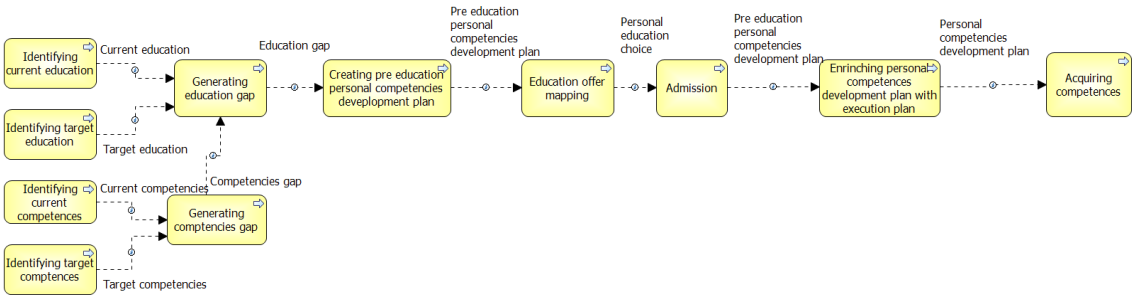


7.9. Potential Applicant Interaction



7.10. Planning in Learner's Pathway

Planning is part of learner's pathway presented above. In the learner's pathway there is three types of planning: planning (non-formal), pre education competence development planning and personal competence development planning. The document describes planning in learner's pathway in more detail. What are process stages and data concepts inputted and outputted at every stage.



Learner's education gap is first identified on higher level and creates first plan which answers to the question "What education learner requires?" Pre education personal competence development planning process identifies learner's current education and target education and the difference between current and target education is the gap which is filled by Acquiring competence process. The same education gap can identified by Planning and Education mapping process but the process is more less defined by the learner herself or himself. After Admission the learner makes more detailed plan of her or his studies and the plan answers to the question "How the required education is acquired?"

First stage pre education personal competence development planning is identification of learner's current competencies. Current competencies are identified in competence mapping process. Competences can come from different sources and before the admission it is up to the learner what competences are relevant and how to emphasize the competences to achieve learner's interests and goals. Before admission the reliability of the competences are up to the learner her- or himself. The reliability is validated later in the learner's pathway process.

Current education is identified by comparing identified competences in national core curriculum. Some of the education is not formally recognized at this stage but can be recognized during Acquiring the competence process. Some of the current competences may not be found from national core curriculum but should be stored for later identification.

Target education is identified by the learner based on the learner's opportunities and interests.There is several ways to identify target education. The learner may choose education from national core curriculum, or education offer, or identify needed education through chosen competences or through chosen goals.

Learner chooses education from education offer straight. The selected education becomes target education. This requires the education offer data is available to the learner.

The learner may choose education from national core curriculum, also. The national core curriculum describes what education contains and competences can be learnt from the education. Before admission the learner has to map selected education to education offer to find education provider.

Learner can also identify single competences which she or he would like to learn and then find education providing those competences. This requires the competencies learner identifies are mapped to education. Learner can be interested in competencies not linked into education or not even provided by education, and in that case there should be a process to guide learner further.

Learner can identify the goal, for example career, and if there is data available between the goal and competencies, then goal's competences can be compared to education and find the target education through this kind of mapping. This requires data about goals and link between goals and competences.

Education gap is determined by comparing current and target education. In summary, first checking what education the learner already has comparing to current competencies and then checking what education learner should have by comparing to the target competencies. The education gap is the difference between the current and target educations. The key data is the link between competencies and education which in this case is defined to be in national core curriculum.

Instead of comparing the current and target competencies into national core curriculum which to get current and target education, it is also possible to get difference on current and target competences first and then find the education matching the difference.

After the learner has identified education gap, the learner can compare the education gap to the education offer and find the suitable education provider to acquire the education.

To map education offer to education gap, there should be either link between education offer and national core curriculum or link between education offer and competencies.

Education offer mapping can be done using just competence gap if there is a link between education offer and competencies. If education offer is linked to national core curriculum, the link from education offer to competencies can be determined through the national core curriculum.

Admission is controlled process which validates the learner's data. Each organization has its own admission process but some admission processes are controlled by national education authority. The input for the process is learner's recognized education and competencies and some other data for selection and output is rights to study, if the learner qualifies the selection criteria.

After admission PrePCDP (what education is needed to fill the education gap) can planned further in more detail together with education provider for exact information how, when and where the learner participates in education.

Planning and Education Mapping has same outcome as Pre education personal competence development process. Learner may have same choices as in pre education personal competence development process but available tools and services can come from anywhere.

Guidance can support planning at every stage. Before admissions the guidance providers can use their own tools to provide guidance services or they can help learner with Compleap tool. Either way the end result should be same from the learner's point of view, the learner gets enough guidance to admit to education according to hers or his interests and possibilities.

On pre education personal competence development planning process on different stages: competence mapping, comparing opportunities and education mapping. If the learner goes through non-formal planning process the same services can be provided by other tools available in the

market. The end result should be still the same, the learner knows what education the learner should acquire and the learner enters into admission process to fill in the education gap.

7.11. Impact and Socioeconomic Effect

Learner's life situation is modeled in the framework architecture with a concept of realities. The concept includes anything concerning the learner's capabilities to work or study or to achieve learner's goals in general. Measuring the learner's capabilities is not an easy task but some can be approached with available data. For example if the potential applicant applies to studies and is accepted, the learner might get some student benefits like study allowance or cheaper accomodations or cheaper public transport tickets. On the other hand when the learner starts to study full-time, the learner might need to leave current work or start doing part-time work.This kind of realities can be modeled as currency flows in the learner's life. The changes in the currency flows are one sort of impact in the learner's life situation.

In general, impact can be thought as the changes in the realities in time. The impact in the learner's life is different depending on at what stage the learner is in the learner's pathway. While the learner can utilize the knowledge of impact when doing short term planning, the actual goal is to reach impact after the studies. In other words, long term impact.

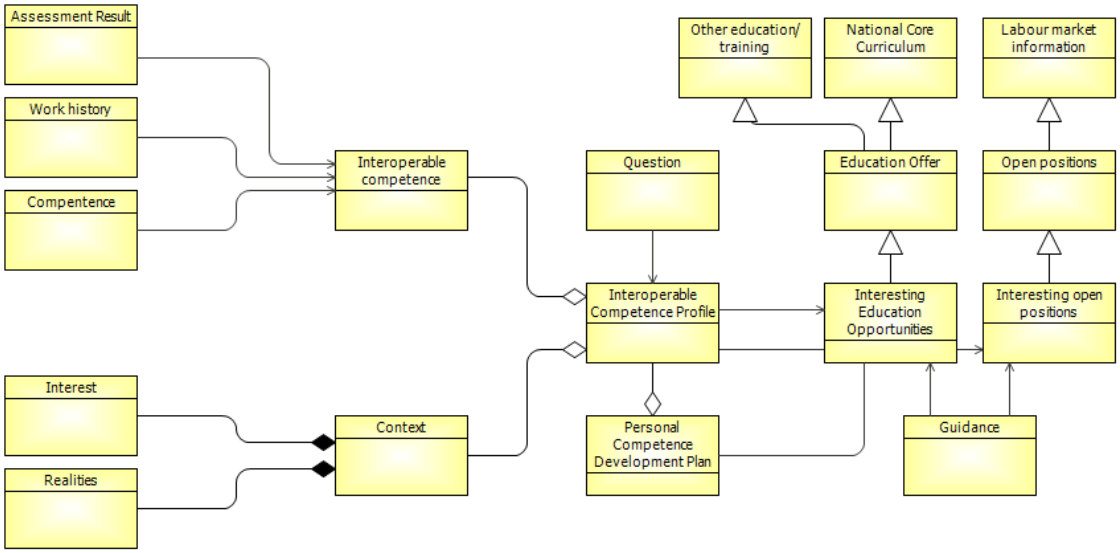
In the framework architecture the long term impact can be defined as a change in learner's realities when learner begins the learner's pathway and ends the learner's pathway. The impact is not immediate when the learner begins utilizing the new acquired competences, so there should be some reasonable time to assess the impact after the studies.

Predicting the learner's short term impact might be possible if there is enough realities data available on different life situations. Predicting the learner's long term impact requires vaste amount of information like prediction about the job market situation after the studies, prediction of employment possibilities and so on. The system could also learn by time what kind of impact the learners have after traveling through learner's pathway. Comparing similar learner's meaning similar competence profiles, intents and goals, the impact estimation could be based on those. The impact estimation could be used in simulation how the choices in the learner's pathway would affect in short term and in long term. The simulation could help learner to make decisions for her or his good. If the prediction is not reliable and brings some ethical issues whether this kind of guidance is acceptable

The socioeconomic effect in this scope can be thought as a sum of all learner's long term impact in a certain time period. This kind of collective impact can be used to help decision making while planning the national core curriculum or budgeting and funding. The prediction of socioeconomic effect could be possible in similar way as in individual's impact prediction and also simulation would be possible in the same way.

8. Information Architecture

8.1. Conceptual Data Model

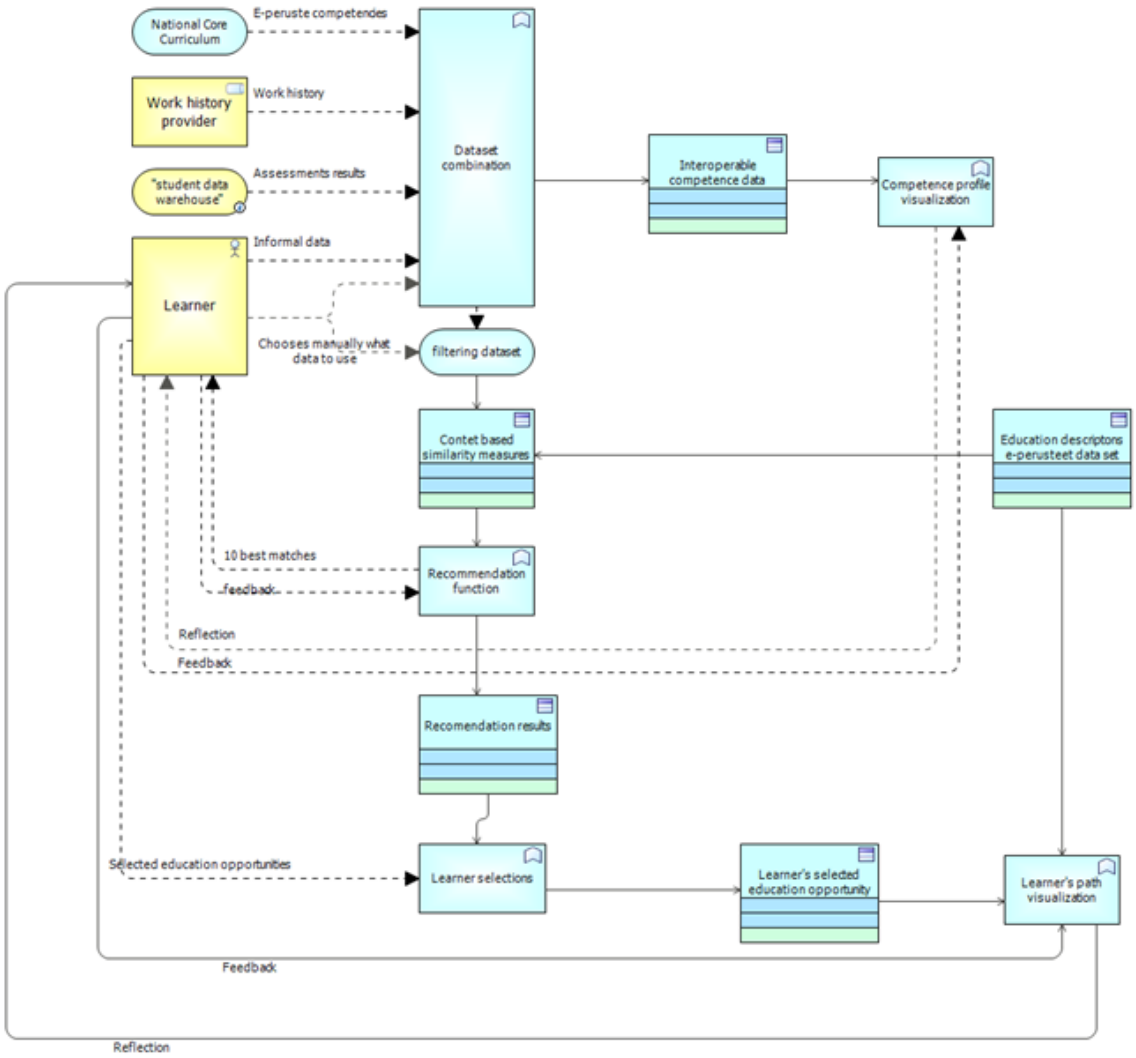


The framework's conceptual data model is described above. Main concept is interoperable competence profile (ICP). The interoperable competence profile contains set of interoperable competences and contexts and it is refined by questions. Interoperable competence is built from assessment results, work history, competences and badges. The context is build from interests and realities.

The information in ICP links to interesting education opportunities and open positions which are refined by guidance. Interesting education opportunities is realization of education offer which is realization of national core curriculum. Interesting open positions is realization of open positions.

The conceptual data model does not tell how the information is actually structured or modeled. From the framework perspective the concepts are recognized as something which is needed to make data flow smoothly among processes.

8.2. Analytics Data flow



Data flow model in the Compleap project illustrates how data flows in the Compleap user services. Information about formal education, previous work history is gathered automatically from other systems and databases. Information about non-formal and informal education as well as interests comes from the manual user (learner) input. Specific data set is extracted and used as a data combination to visualize competences for the user and at the same time stimulate his reflection on his current competence situation. User can also manually select data combination and filters to get a education recommendation. This is content based recommendation. 10 best results are presented to the user where he can further select his favorites and proceed to learners path visualization where information about education option is visualized for the user again to promote his understanding of the studying and also self-reflection. User is also able to give feedback about competence and learners path visualizations as well as education recommendations.

Information flow described above leads to four main services provided: Competence visualization, education recommendation, education comparison and study guidance.

Competence visualization - there are certain competences acquired during the study process. These competences are usually described in the national curriculum and are at the core of the modern world of work and study. Although they are learnt and specified in documentation many students

don't have enough information about them or maybe haven't even heard about them at all. Compleap service provides an opportunity to see clearly what competences have been studied during the study process in educational institution. Also what groups they make and what are the main or key competences as described by EU.

Education recommendation - education recommendations are calculated and based on similarity of content between user's profile data and education descriptions.

Education comparison - closest matches of between user's profile and education descriptions are presented to the user in ranking. Starting from the closest one to the less closer and so on. User can mark some recommendations as favorite and this way generate more suitable suggestion.

Study guidance - the whole service is seen as a study guidance and support for decision making. Information about own competences, previous experiences and interests are gathered in one place and presented in a user centered way to promote his reflection about studying possibilities. Education recommendation is provided not as a solution but as an encouragement for the person to think and reflect on his interest and future goals.

User is able to give feedback on the usefulness and accuracy of the the profile and education recommendations to further develop these services.

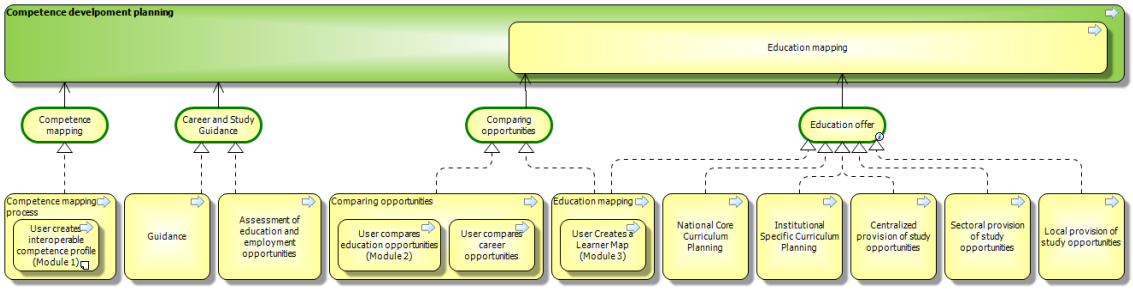
9. Information Systems Architecture (Application Architecture)

The application architecture is used to model the information systems architectures of the CompLeap project. It describes the structure and interaction of the needed applications.

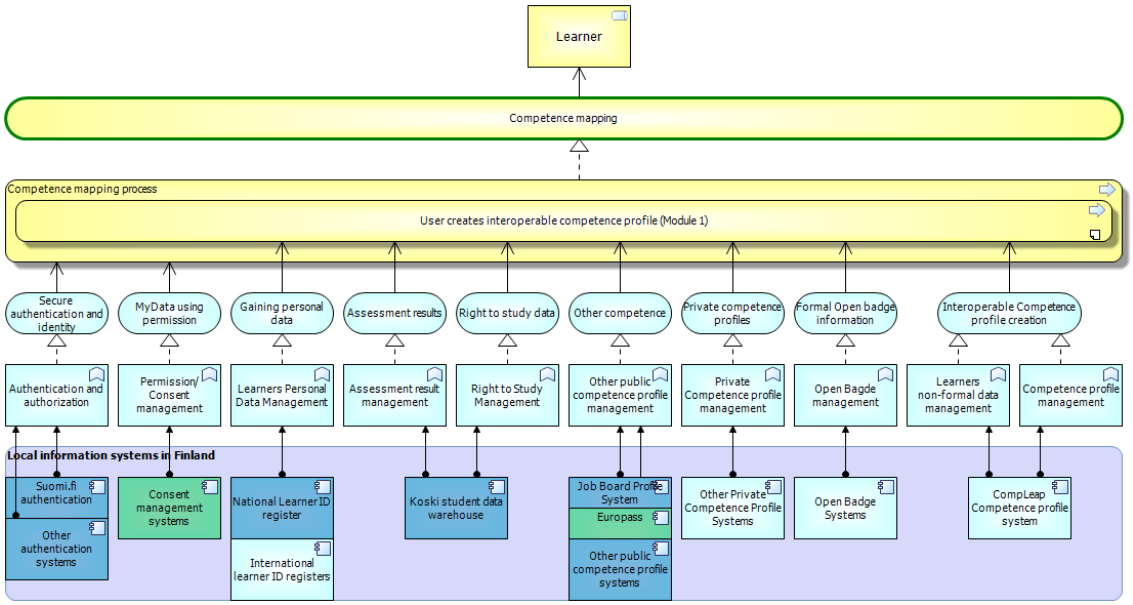
9.1. Layered Architecture views

Purpose of the models: Show processes, application services and information systems needed for providing CompLeap services.

This first diagram illustrates overall process of competence development planning and it's subprocesses in CompLeap referece architecture. The rest of the diagrams in this chapter are subdiagrams that drill deeper into smaller architectural components ("Competence mapping", "Career and Study Guidance", "Comparing opportunities" and "Education offer")



9.1.1. Competence Mapping layered architecture view



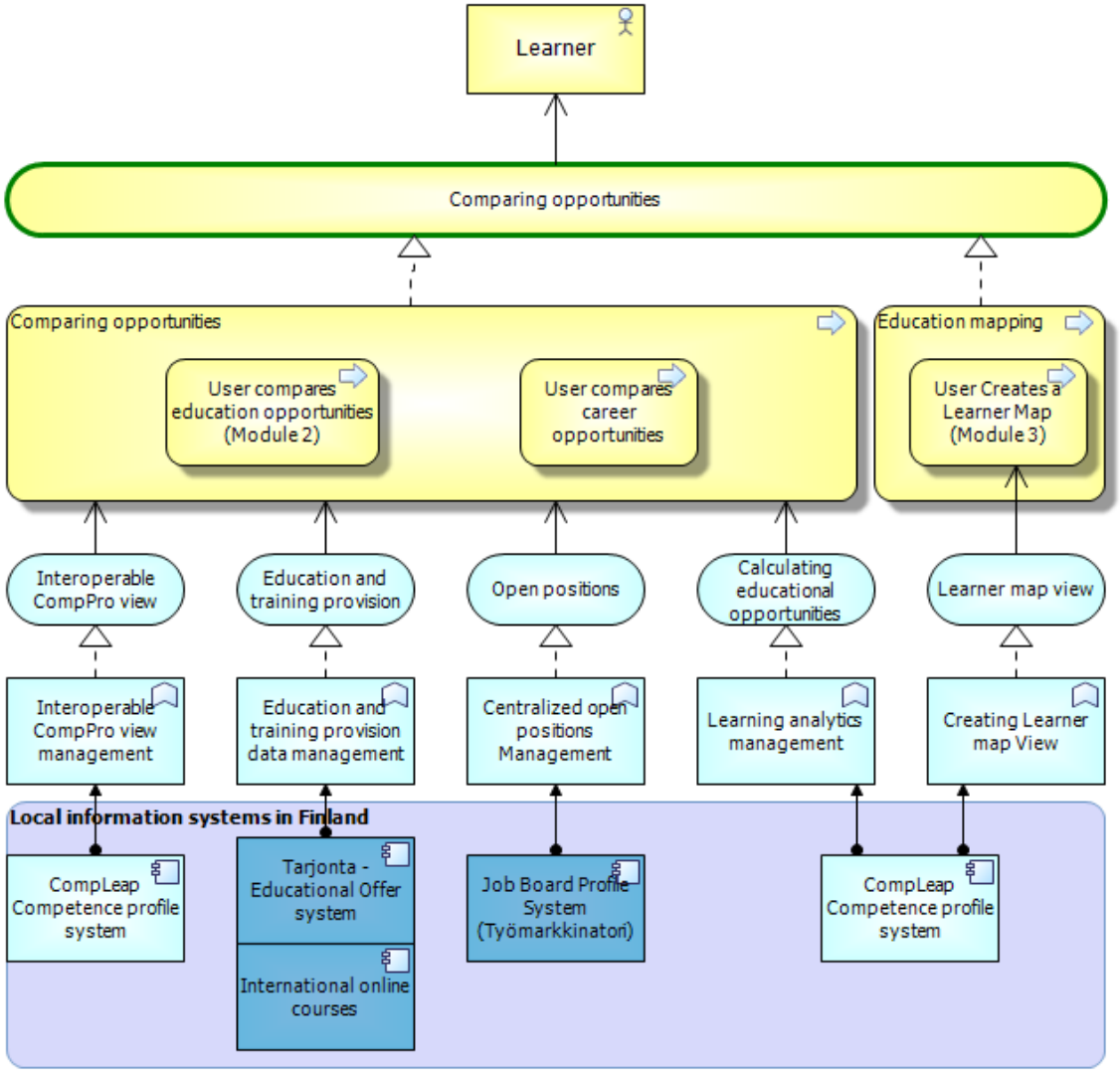
Competence mapping layered architecture view is described above. The diagram describes the competence mapping service is implemented by competence mapping process. When learner uses the service the competence mapping process is executed. The learner gives information about her or himself and process outputs hers or his competences back to the learner.

In CompLeap context the whole competence mapping process is about to be digitalized by module 1 and the process of using digitalized tool is put inside the competence mapping process. In this case, "User creates interoperable competence profile"-process implements the whole competence mapping process but in other implementations there can be manual stages also.

The blue ovals describe the application services supporting the competence mapping process. The application functions are implementations of application services and there could be many application functions which are categorized as same application service. In the diagram only the relevant in the context of the framework are shown.

The lowest layer are systems to which the application functions are assigned in Finland. Thus, the system layer is implementation specific.

9.1.2. Comparing opportunities layered architecture view

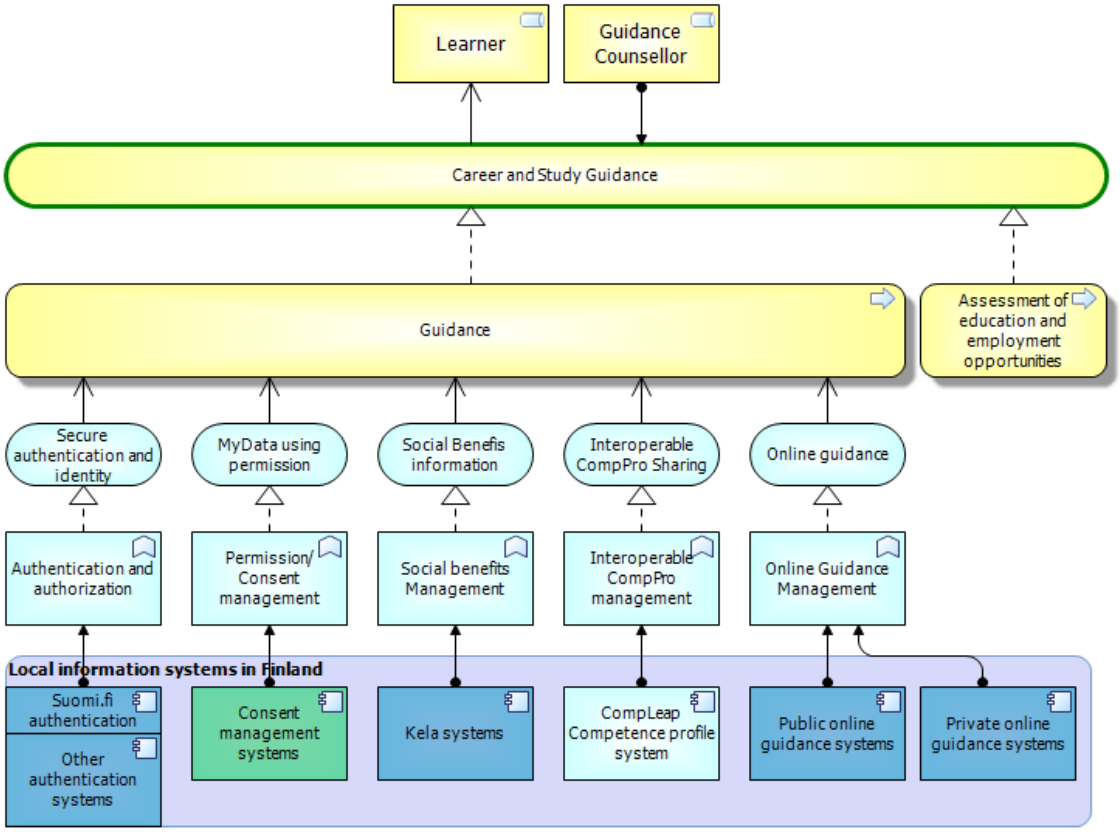


"Comparing opportunities"-service consists of two subprocesses "Comparing opportunities" and "Education mapping". These subprocesses are devided to smaller processes that are supported by

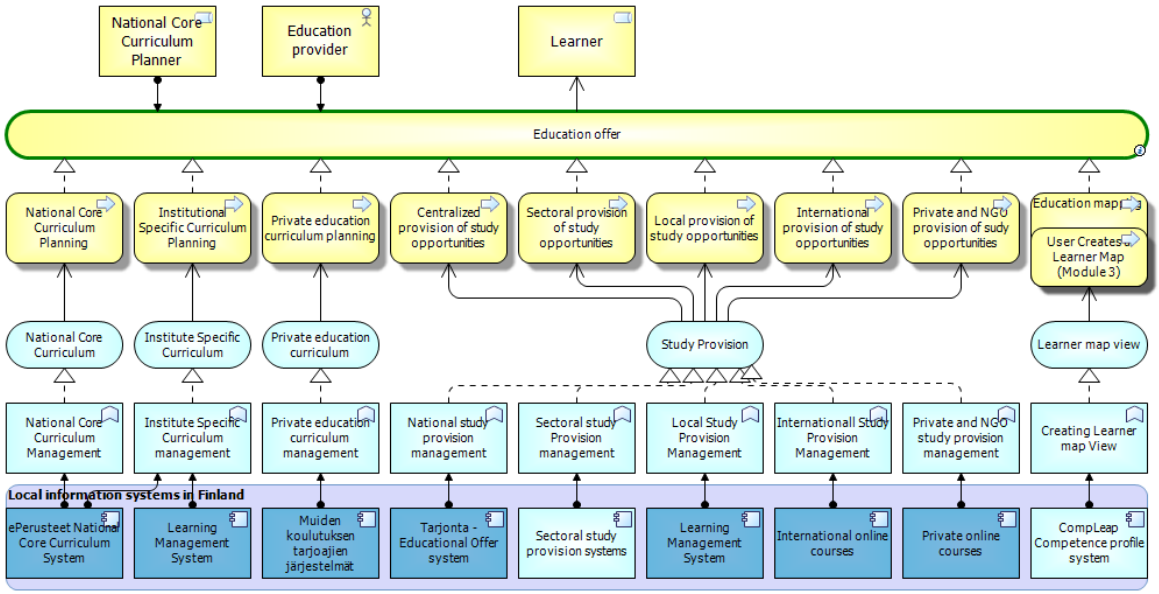
application services (blue ovals) which are implemented by application functions. The lowest layer is country/implementation specific and needs to be replaced with local services.

9.1.3. Career and study guidance layered architecture view

This diagram illustrates "Career and Study Guidance"-service and again the types of the layers are similar to what can be found in previous diagram. Again lowest layer is country/implementation specific.

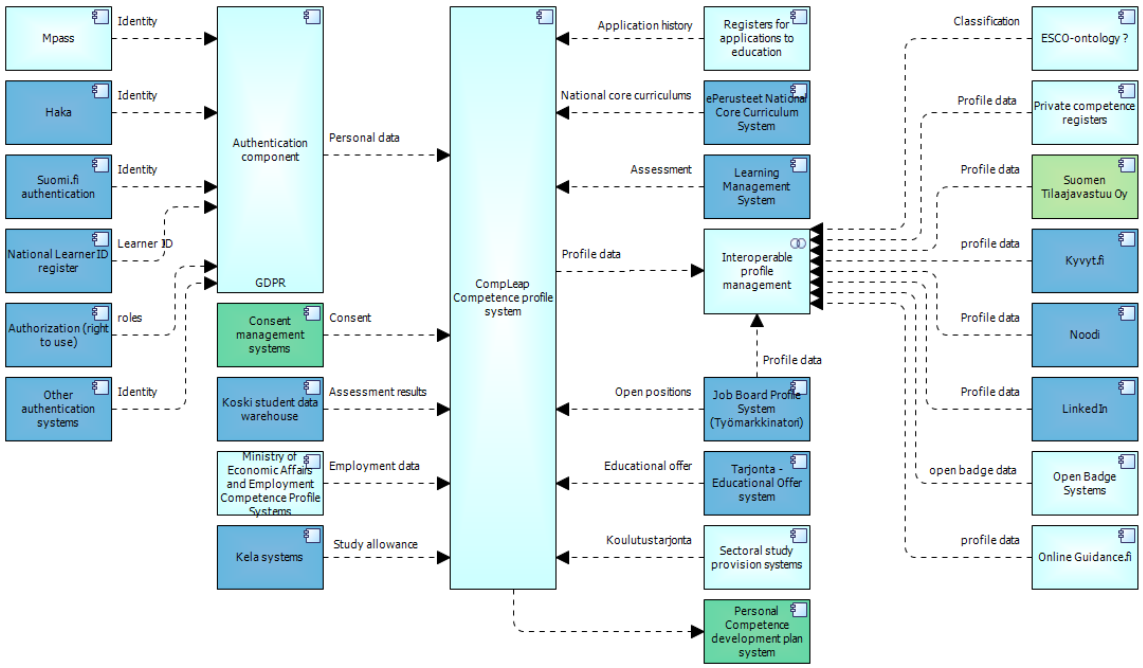


9.1.4. Education offer layered architecture view



This diagram illustrates "Education offer"-service.

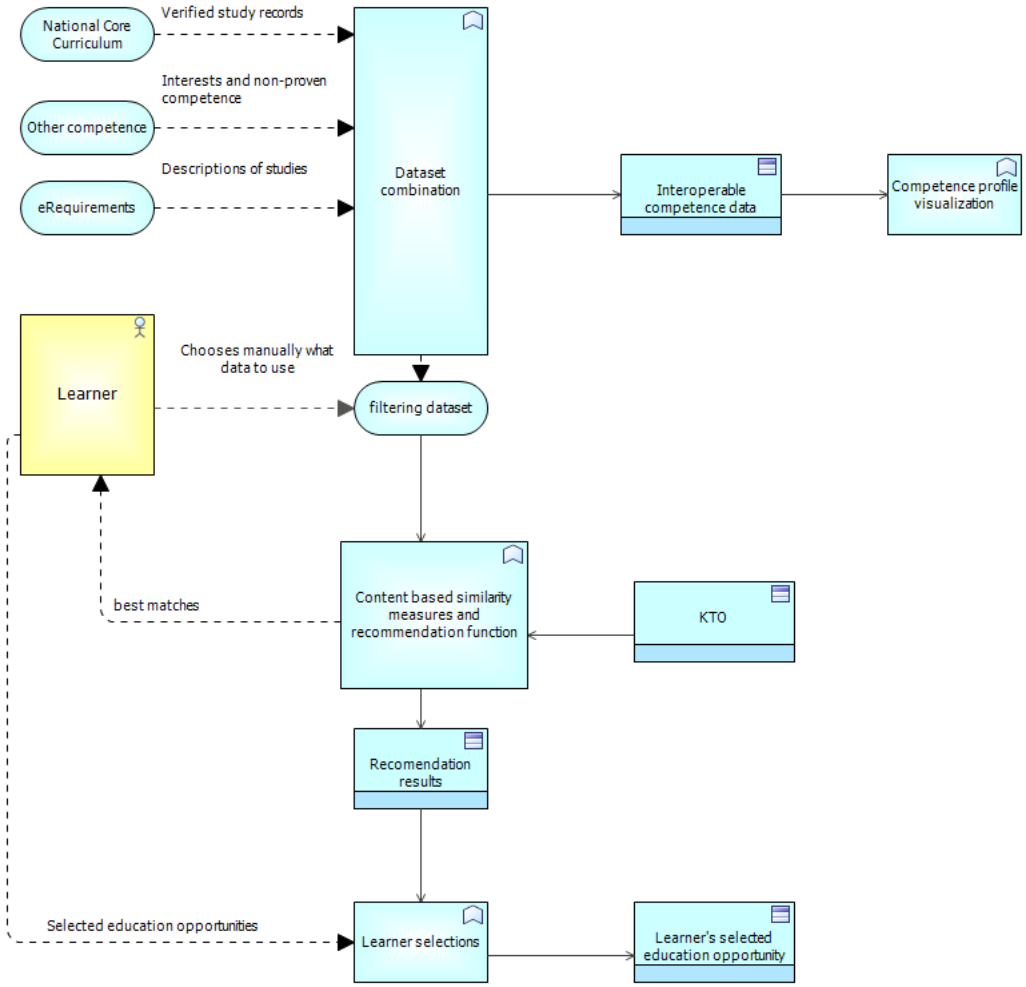
9.2. Application integration: Case Finland



CompLeap Competence profile system Application integration in case of Finland is described above. The system interacts with public authority systems via authentication proxy to meet GDPR requirements. Consent management is new aspect in system development and in Finland learner decides who can access her or his personal data. Interoperable profile management is done in collaboration with other profile related systems.

9.3. Analytics Data Flow (case Finland)

Prototype implements only partial solution of above described reference architecture. Analytics data flow diagram below is modified to describe prototype architecture.



User comes to landing page and either logs in or uses system without logging in. If user logs in and has SSN (HETU) then all user related data is brought automatically from various data sources to dataset combination.

Dataset combination can be filtered by the user.

Filtered dataset is used as basis to semantic matching with KTO (KTO is abbreviation for Finnish words *Koulutusinformaation, koulutustarjonnan ja* [*Opintopolku.fi*](Opintopolku.fi) -uudistus).

After semantic matching phase user will get personal recommendations based on input data.

After dataset combination competence profile is showed to user (ESCO). (This is right from the "dataset combination".)

In case user do not login (s/he does not have Finnish SSN) then only one data set is automatically fetched for the user ("other competences") and this will be used for recommendations. This is a set of competences that user has selected from a list (*national education classification*).