IADIS International Journal on Computer Science and Information Systems
Vol. 14, No. 2, pp. 59-77
ISSN: 1646-3692

# COMBINING BLOCKCHAIN, SEMANTICS AND DATA ANALYTICS FOR UNIVERSITY PROCESS OPTIMISATION

Christos Kontzinos, Vagelis Karakolis, Stavros Skalidakis, Ourania Markaki, Maro Androutsopoulou and John Psarras
*Decision Support Systems Lab National Technical University of Athens, Greece*

**ABSTRACT**

The aim of Higher Education is twofold: meet the learning needs of individuals through the development of their intellectual abilities, as well as equip them with the necessary skills that will help them enter the labor market. As such, university processes and curricula should be well structured with courses and activities that help students follow their aptitudes, while also providing them with the required level of expertise that the respective job market requires. This paper presents work from the EU funded research project QualiChain that aims to revolutionize the domains of public and private education as well as their connections with the labor market and society at large through the development of a platform that is based on innovative approaches and technologies, such as blockchain, semantics, data analytics and decision support. To achieve its goals regarding education and the labor market, QualiChain involves a pilot targeting student accreditation, curriculum design and process optimization within the School of Electrical & Computer Engineering (ECE) of the National Technical University of Athens (NTUA), which will be the main focus of the present paper.

**KEYWORDS**

Blockchain, Education, Labor Market, Degree Verification, Curriculum Optimization, Decision Support, Data Analytics

## 1. INTRODUCTION

Higher Education in the fourth industrial revolution is a complex, dialectical and exciting opportunity which can potentially change society for the better (Xing, B., & Marwala, T. 2007). In an era that information and communication technologies are transforming much of society and promoting new paradigms in various sectors, Higher Education Institutions (HEIs) are challenged to meet the connectivity demands of prospective students as well as growing

expectations and demands for higher quality learning experiences and outcomes (Garrison, D. R., & Kanuka, H. 2004). Along these lines, to ensure that these demands are met, the modernization of Higher Education learning environments is critical (Williams 2002). According to Young (Young 2005), the adoption of innovative technological solutions by universities can help them facilitate the commercialisation of research results for the public good, reward, retain and recruit faculty, induce closer ties to industry and promote overall economic growth.

At the same time, HEIs need to stay true to their original goals, by helping their students develop their aptitudes and intellectual abilities as well as equip them with the necessary hard and soft skills that will help them enter the job market (South African Council on Higher Education 2013). The former aspect requires guidance from the university to help students make the best use of their talents as well as a way to help them track their skills and qualifications, while the latter requires HEIs to follow closely the developments in their respective domains to maintain a certain level of connection with the job market. Hence, HEIs need to be constantly evolving to respond to the ever-changing needs of society such as growing student demands in certain disciplines, embedding workplace attributes to graduate programs and ensuring that the quality of learning programs is both nationally and globally relevant (Daniel 2015).

Despite the aforementioned challenges and the growing need for universities to become more connected to the job market, oftentimes the curricula of HEIs are shaped with strictly academic criteria without taking into account the demands of the labor market and they are rarely modified to incorporate the latest developments, especially in technology related fields, thereby ending up being obsolete and outdated (Kontzinos et al.). Another major challenge is that education credentials such as degrees, qualifications and other accreditations still require paper documentation in most cases and time-consuming processes for their verification. As a result, holders of such credentials must always be dependent on the institutions that issue them, which can create major issues when they want to transfer to another HEI or send their certifications to a prospective employer. Additionally, the increasing amounts of fraud and corruption around higher education degrees and credentials is shaking the trust in the education system (Mohamedbhai 2016; Chapman and Lindner 2014).

HEIs need to fundamentally change the way they operate to efficiently respond to the expectations of society in the 21st century. However, despite the aforementioned circumstances, the slow digitization of HEIs is a major concern in the research community (Association of Universities in the Netherlands (VSNU) 2017). While university IT infrastructures nowadays support most of the logistic processes, their data are held in centralized databases with little or no interoperability and with limited access to the IT staff of the institution (Turkanovic et al. 2018). However, such data are representative of student life (Koedinger, K. et al. B 2008) and university processes, and can help HEIs realize shortcomings in the way they operate. The lack of structure and analysis in university data is the main concern of this paper as it prevents such institutions to tackle the challenges they face.

It can be surmised from the above that many of the aforementioned challenges could be solved via the implementation of suitable IT infrastructures. Apart from existing IT systems and infrastructures, there are several innovative technologies that HEIs could take advantage of to develop trustworthy and efficient solutions. Data analytics and decision support can help universities analyze their data to draw useful conclusions and receive recommendations on optimizing their operation. Moreover, blockchain technology, as a decentralized, permanent and unalterable store of information can help with the archiving and trust issues around academic credentials (Kontzinos et al.).

Under these circumstances, this paper presents QualiChain, an EU funded project targeting the creation, piloting and evaluation of a distributed platform for storing, sharing and verifying academic and employment qualifications. QualiChain will take advantage of blockchain to create a secure and trustworthy system for degree verification while also promoting transparent student accreditation via smart badge endorsement as an innovative approach to enhance student everyday life. By combining various innovative technologies such as semantics, data analytics and decision support with a blockchain infrastructure, QualiChain aims to show that the projected solution will not only cover student accreditation but has potential for wider university process optimization. To showcase and validate the impact of the platform, the project includes a pilot targeting student accreditation, curriculum design, and process optimization within the School of Electrical & Computer Engineering (ECE) of the National Technical University of Athens (NTUA), which is the main focus of this paper.

Section 1 introduces the scope of this paper by presenting the current situation in Higher Education, the challenges that lead to the need for digitization and how blockchain and other innovative technologies can lead to more effective and secure solutions. Section 2 outlines the methodology that will be followed to realize the pilot case presented in this paper. Section 3 introduces the pilot concept, its stakeholders and initial results as well as the next steps that will be performed. Finally, section 4 presents the conclusions of the paper.

## 2. PILOT METHODOLOGY

The pilot case presented in this paper will focus on the optimization of the qualifications' awarding and curricula design processes and accordingly the university operation of the ECE School of the NTUA and will revolve around the three following axes:

1. Semantification of university data
2. Provision of analytics and recommendations, through the platform's Decision Support System (DSS) for student guidance as well as redesign and update of specific courses and by extension the entire curriculum of the school.
3. Smart Badge student accreditation and Ph.D. student (lecturer) validation and recognition through a trusted, immutable and secure Blockchain ledger.

The methodology that will be followed to fulfil the goals of this pilot can be seen in Figure 1 and will be further explained below.

In order to fulfil the goals of this pilot, the labor market requirements of popular professions among the school's graduates will have to be analyzed and assessed with the current skill/knowledge profile of the average student of the school. For this purpose, the project's technical team will perform data crawling in well-known job posting sites in order to identify the sectors and specific professions that are targeted by Electrical & Computer Engineers. The specific skills and qualifications required by these professions will be mapped and compared with the current skill profile of a student of the School, the latter sketched out by means of interviews and questionnaires. The data produced from these activities will be semantically enriched, given structure and fed into a data analytics module that will produce a number of indicators concerning popular careers among the School's graduates.

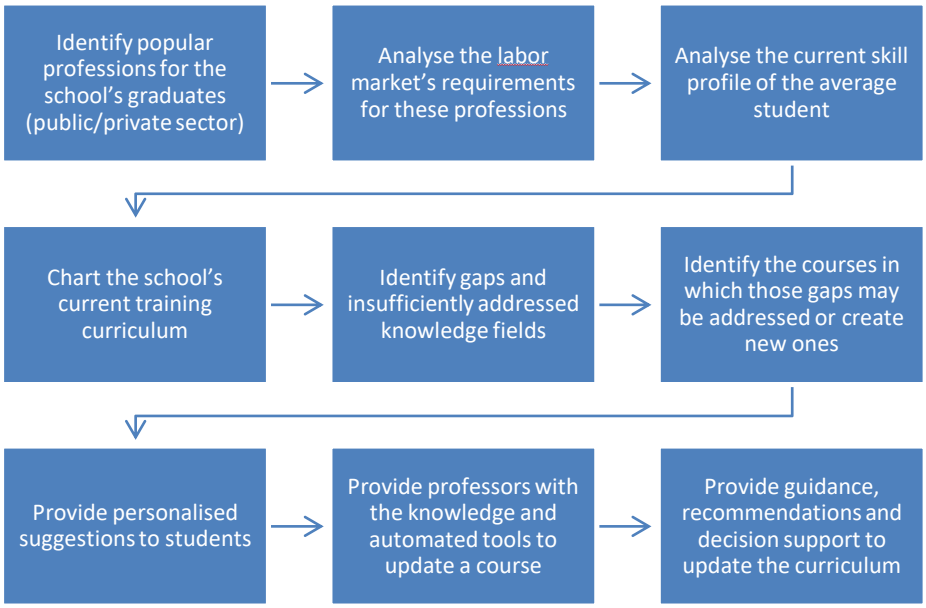



Figure 1. Pilot Methodology

Following that, the pilot will chart the school's current curriculum to identify gaps and insufficiently addressed knowledge fields as well as the courses in which those gaps may be addressed by updating/enriching the course's curriculum. This procedure will have to be performed course-by-course and is projected to be a herculean task given the length and complexity of the School's curriculum. Specifically, the curriculum specifies that a student will have to complete 35 core courses in the first five semesters and 25 specialization courses, selected from three or four out of twelve specialization flows, in the following four semesters. The 60 courses that the student will have to complete are selected from a pool that includes over 200 courses stemming from various technological and scientific fields, such as software and hardware engineering, electrical engineering, energy efficiency, mathematics, physics etc.

The purpose of the previous task is to extract metadata from the analyzed courses and use them in combination with job market required skills as input for the platform's analytics tool that will identify skill and knowledge gaps in the curriculum. This process will generate various metrics and indicators that will feed the platform's DSS tool, which will in turn produce various suggestions for the pilot's stakeholders. These include personalized suggestions for students, recommendations to professors who want to update their courses and overall guidance and decision support for the curriculum update and process optimization of the entire school. The specifics of the above-mentioned procedures and interactions will be further explained in the following section.

## 3. PROGRESS & RESULTS

The QualiChain project is still at an early stage, which means that no technical solution has been developed yet. Consequently, the initial results of the pilot use case presented in this paper include the theoretical groundwork that specifies the methodology for completing the pilot, user requirements, and process flow that will be followed between the various stakeholders of the QualiChain platform. Additional results include mockups that represent the projected user experience of the school's students and professors when using QualiChain and a high-level technical description of the platform's functionalities that are integral for the successful operation of the pilot (blockchain, semantics, analytics and decision support). The following section describes the flow that will be followed during the pilot's operation for the optimization of the various stakeholders' everyday tasks and for the endorsement of smart badges from professors to their students and lecturers. The process flow has been split in three parts to accommodate the different needs of the pilot's stakeholders, i.e. undergraduate student, Ph.D. students and professors, which will be further elaborated below.

An undergraduate student's main challenge with the as-is situation revolves around the lack of recommendations and general guidance for selecting courses and making the choices that will maximize his/her aptitudes and future career plans. The School's curriculum lacks the structure that will allow students to make more informed decisions and there are no added-value tools for filtering and analyzing one's choices. As such, they have to rely on the opinion of older students for selecting courses, which can be in many cases biased. Additionally, especially in the era of digitization, students lack any kind of personal profile in which they could showcase and digitally validate the skills and qualifications received through their education, in a way that is trustworthy for potential future employers.

Ph.D. students/lecturers are a group that is often overlooked in the various approaches that tackle the domain of Higher Education. In the ECE School, Ph.D. students are responsible for various tasks such as teaching classes, grading exams, performing research and other administrative processes, relevant to the wide range of the university's activities. However, their working experience in academia is not often reflected in their professional profile as many of those tasks cannot be digitally validated and recognized with the School's current infrastructure.

Finally, professors of the School, often face difficulties in performing tasks such as updating a course, assessing student evaluations and certifying students' participation in extra-curricular activities that they organize. The main challenge in this case is the lack of automation and structure in the School's data. As a result, professors need to rely on their own perception when trying to identify pertinent advancements in their respective fields in order to update the syllabus of a course. Additionally, students' evaluations offer no possibility for automatic analysis and they require manual work for the extraction of useful conclusions.

In the context of this pilot, data play the most important role. As it was previously mentioned, it is the lack of structure in the School's data that creates most of the challenges for its stakeholders. For that reason, data concerning the university and the respective job market will be semantically enhanced and normalized to provide the much-needed structure for the platform's added value analytics and DSS tools. Such data include but are not limited to the School's curriculum and each individual course, skills developed and students' evaluations, as well as data that will be mined from open repositories, such as job market requirements, popular professions and employment rates for the School's graduates. In the following section, that describes in detail the added value of the QualiChain platform to the pilot's stakeholders, the

semantic enrichment of the above-mentioned data is considered as a given. However, additional information concerning the semantification of the university's data and the QualiChain ontology in general will be presented in section 3.3.

## 3.1 Pilot Process Flow

### 3.1.1 Undergraduate Students' Process Flow

In the context of this pilot, students will first have to sign in the QualiChain platform with their verified school credentials and work on their dedicated personal profile. For this purpose, they will use a sub-module of the platform, entitled Profiling & CV Customization module (PCVC). The PCVC will draw various data from QualiChain (credentials, courses etc.) and students will manually fill in additional data (interests, future career projection etc.) to create their profile, which is then saved in QualiChain.

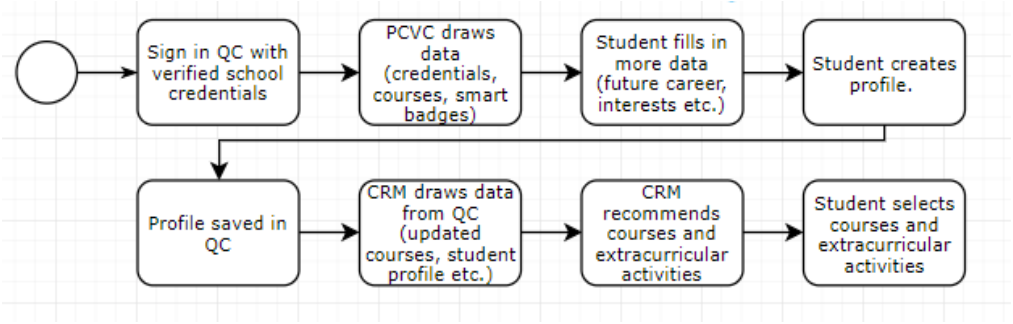
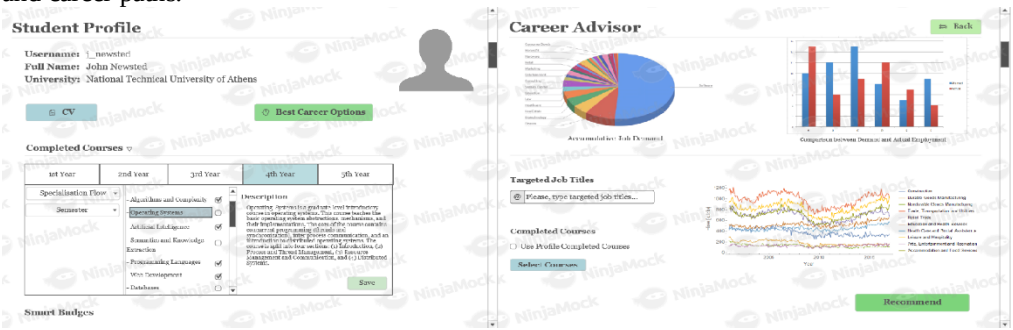



Figure 2. Undergraduate students' process flow

Following that, another sub-module, the Course Recommendation Module (CRM) will take into account the student's profile and the school's curriculum and recommend courses and extracurricular activities. This process flow can also be seen above in Figure 2. Figure 3 below, showcases the initial mock-up screens and provides a more comprehensive image of the process flow described above. The students' profiles will include all pertinent information about them, such as name, CV, completed courses (organized by semester) and smart badges awarded. The CRM will include information on the job market in the form of visualizations and students can receive recommendations on courses and skills that they should acquire for certain professions and career paths.



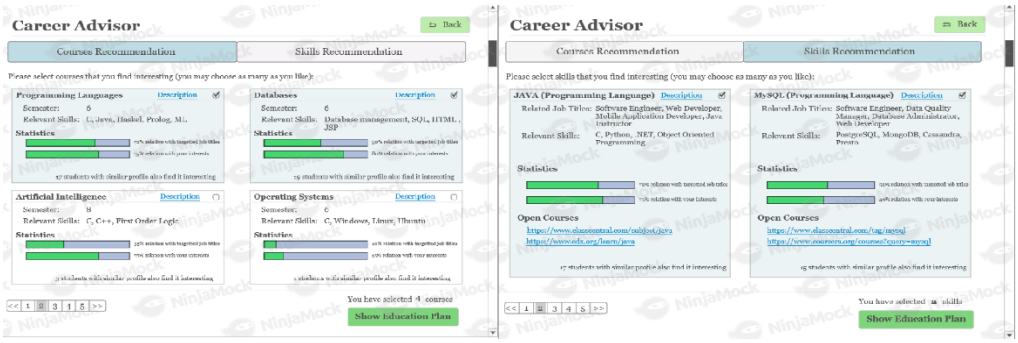



Figure 3. Mockups on students' profiles and recommendation system

### 3.1.2 Ph.D. Students' Process Flow

Ph.D. students, similar to the previous stakeholder group, will sign in QualiChain with their verified credentials and create their personal profile in the platform. The PCVC will take into account the students' degree, course involvement, Ph.D. thesis etc. to help them create their profile. It should be noted that while Ph.D. students will have the capability to use the CRM to receive recommendations for furthering and improving their knowledge, skills and career, their main role in the pilot revolves around receiving tokens and smart badges for their involvement in various university related tasks, through the platform's blockchain ledger. This process flow can also be seen in Figure 4.

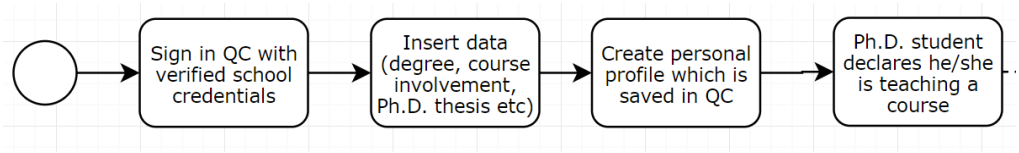



Figure 4. Ph.D. students' process flow

Figure 5 includes the mock-ups that showcase how smart badges will be awarded in the context of QualiChain to Ph.D. students (same process is followed in the case of undergraduate students). Smart badges can only be awarded by professors of the school and will be specific to the courses that the professor is teaching. For Ph.D. students, smart badges will verify their involvement with one or more courses as lecturers. However, they will not receive smart badges from students but rather tokens. The reason for this is to avoid overflowing a Ph.D. student's profile in QualiChain with multiple (potentially hundreds) of smart badges that all represent the same skill. There will be a ratio between tokens and smart badges that will allow Ph.D. students to turn their tokens into smart badges after receiving a certain amount, however this ratio will have to be specified and optimized in real life scenarios of the pilot in later stages of the project.

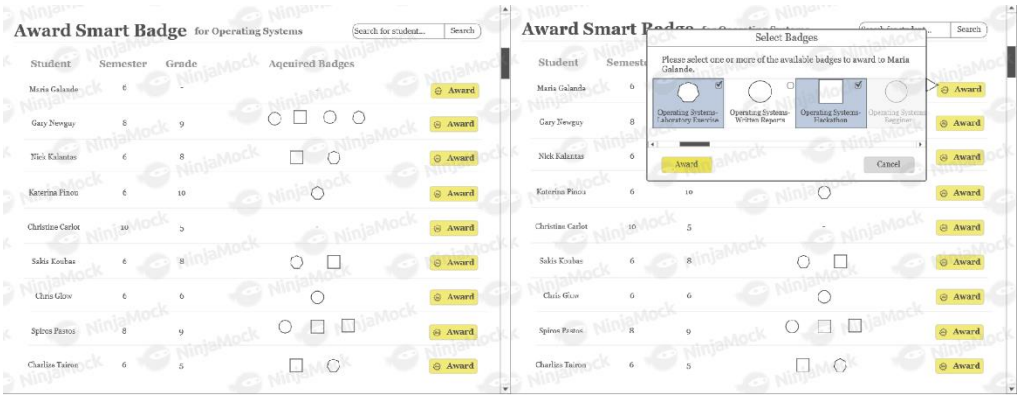



Figure 5. Smart Badge Endorsement

### 3.1.3 Professors' Process Flow

Finally, professors, after signing in QualiChain, will insert data into the platform that are relevant to their courses (course metadata, related skills, course badges, students' evaluations etc.). When a professor wants to update a course, such data will be fed into the platform's analytics tools along with curriculum data, job market data, technological developments for the respective domain and job market requirements to produce various analyses. The platform's Multi-Criteria Decision Support Module (MCDM) will take as input these analyses and produce recommendations that will help the professors update a course. The updated course is then saved in QualiChain and in case that some of the knowledge gaps identified in the pilot's previous steps were addressed, QualiChain will update the recommendations produced for the other courses as well as the recommendations that students receive through the Course Recommendation Module can also be seen in Figure 6.

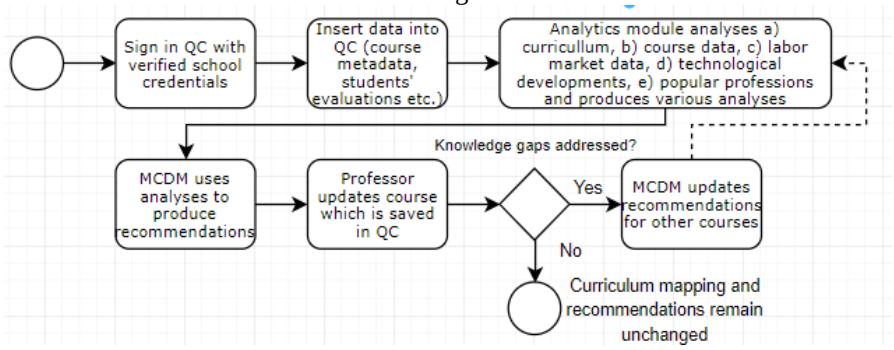



Figure 6. Professors' process flow

In Figure 7, the mock-ups that represent the professors' process flow in the pilot can be seen. Apart from personal and professional information, professors can list the courses that they are involved with in their profile. Each course will involve a short description, related skills and smart badges that students can be awarded by attending the said course. Via their profiles, professors will be able to either award smart badges to students or redesign their courses by modifying the description and/or syllabus, organize events (special lectures, hackathons etc.) and update the list of lecturers. Moreover, for the purpose of updating courses, professors will

have at their disposal a number of useful visualizations related to the school's curriculum, such as job market data, skills' gap analyses, missing skills and recommendations as to the courses in which these skills can be added.

-

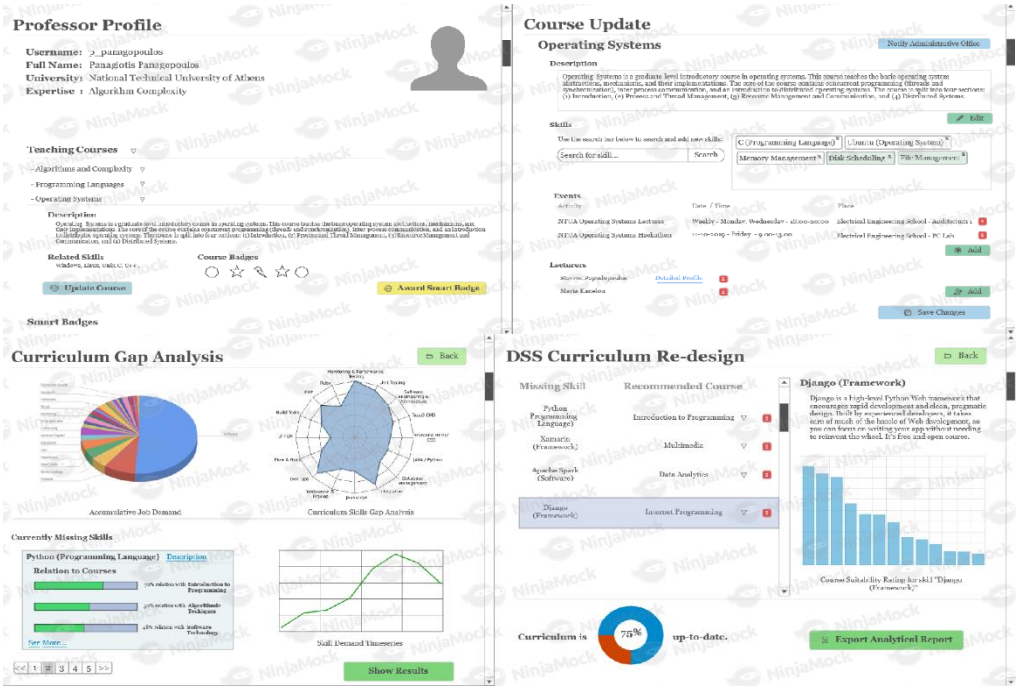



Figure 7. Professors profile and course/curriculum redesign

## 3.2 Smart Badge Endorsements

As already mentioned, QualiChain will take advantage of the platform's blockchain ledger for validating, through tokens and smart badges, skills and accreditations that students receive by attending courses, completing assignments and participating in hackathons, seminars and other various extracurricular activities. The same mechanism will be employed to award Ph.D. students and lecturers with smart badges for teaching courses or contributing to various university-related tasks. The process flow that will be followed for smart badge endorsement can be seen in Fig. 8 and is described below.

The most important contribution of the blockchain ledger for this pilot is that professors will be able to award students with smart badges. On the one hand, given that several courses include personal or group assignments, professors at the end of the semester will be able to award smart badges to the student(s) that achieved the best results. An example could be the most efficient algorithms in software related courses. Additionally, oftentimes professors organize extracurricular activities such as hackathons, seminars and special lectures. Students that attend such activities will be awarded with participation badges. Smart badges are a secure and trustworthy way to validate that students possess a certain skill or qualification and will be displayed on their personal profile.

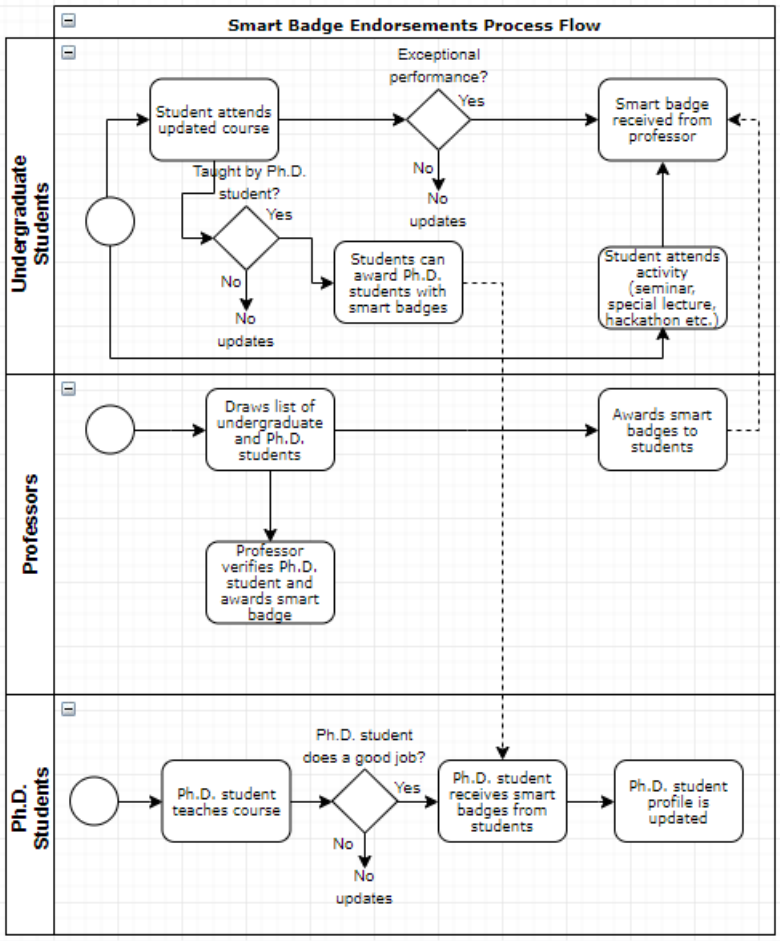



Figure 8. Smart badge endorsement process flow

On the other hand, Smart Badges are an innovative way to verify Ph.D. students' and lecturers' participation in courses and other university related activities. Focusing on courses, a Ph.D. student, who is also a user of QualiChain will be able to declare via the platform his involvement with a specific course. The professor, who is responsible for the said course will validate that declaration by offering a smart badge to the Ph.D. student. Through their involvement in a course, Ph.D. students will also be able to receive tokens from the students for performing well during the lectures. For example, students will award Ph.D. students (lecturers) with tokens for being informative, cooperative, communicative etc. and there will be a set ratio of tokens to smart badges (for example 30 tokens for being communicative equal one smart badge). As already mentioned in section 3.1.2, the reason that for this case, tokens are used instead of smart badges, is to keep the smart badge accreditation portion of this pilot under control and avoid overemphasizing the contributions of Ph.D. student in courses with hundreds of similar smart badges awarded to them.

## 3.3 QualiChain Ontology

### 3.3.1 Semantic Lifting of Course Descriptions

Semantic technology is a major cornerstone of the project, since most data to be processed, analyzed and used for value-added services will be subject to the process of 'semantification'. In other words, data behind QualiChain will be semantically lifted onto a knowledge graph that can be uniformly queried and processed by various QualiChain platform components. QualiChain will promote and enforce semantic interoperability through domain-specific standards and ontologies – some of which are already available, and some that are being extended or designed anew. These vocabularies will be brought together through the design of a QualiChain Ontology, which despite its name will not only be relevant and of value for the immediate needs of the project but will be independent and can represent any efforts using Blockchain to verify smart contracts and certificates. As such, the QualiChain Ontology will be a wide mixed set of existing, extended or new vocabularies that represent the following sub-domains:

- Personal Accomplishments – as often described in Curriculum Vitae documents.
- Job Postings – centered around the concepts of skills and qualifications that they require.
- Educational Course Descriptions – also centered around skills that candidates will accomplish.
- Blockchain Entity Descriptions – describing properties of verifiable objects (certificates, acquired skills and expertise) on the Blockchain.

These sub-domains, and other ones that might be identified in future stages of the project, can be considered as modules of the integrated QualiChain ontology. The descriptive concepts and relations of each of these sub-domains will have to be identified to modify existing ontologies and potentially define new classes and properties. Specifically, for the pilot presented in this paper, while there are vocabularies that sufficiently describe and cover course descriptions, skills and qualifications, such vocabularies need to be extended and connected to the specific data of the ECE School. For that reason, the courses of the school's curriculum will have to be analyzed to extract data from them and use them to extend existing vocabularies. An example of the information available for each course is presented in Fig.9.

Information like the course's unique code, semester, flow and lecturers will help in the semantic structuring of the school's curriculum, while information in the course's description will help distil the hard and soft skills that students will possess after completing the course. The data extracted from the courses will be analyzed to distil the knowledge gaps of the curriculum, better represent the skills and qualifications of students in their personalized profiles and used as input in the platform's DSS tools to provide suggestions and guidance for optimizing the university's processes.


Figure 9. Course related data

### 3.3.2 Existing Vocabularies

As already mentioned in the previous section, the QualiChain ontology will include several sub-domains, each of which will be represented by one or more existing vocabularies. At this stage of the project, the most relevant vocabularies per sub-domain have been identified as well as the major missing descriptions that will need to be added as local extensions within the core QualiChain Ontology.

Table 1

| Sub-domain | Fitting Vocabularies | Required Additions |
|---|---|---|
| Personal Accomplishments | ResumeRDF<br>OpenBadges Vocabulary | - Accomplishments per se, and not their embodiment in Resumes/CVs.<br>-Work Experience metadata |
| Job Postings | SARO<br>Job Posting (schema.org) | -Mostly Complete |
| Course Descriptions | SARO | -Descriptions of Course Providers |
| Blockchain Descriptions | Ethon | -Mostly Complete |

ResumeRDF[1] covers personal accomplishments as described or contained within CVs, but not as independent concepts themselves. Since in QualiChain actual accomplishments (e.g. work experience, or a course certificate) need to be verified, additional elements will have to be created to appropriately describe them. Moreover, OpenBadges[2] is another vocabulary that contains detailed metadata about achievements and will have to be extended to include additional accomplishments and better serve the needs of the project.

[1] http://rdfs.org/resume-rdf/
[2] https://www.imsglobal.org/sites/default/files/Badges/OBv2p0Final/index.html

The Skills and Recruitment Ontology (SARO)[3] is heavily based on other vocabularies (ESCO[4], schema.org, etc.) to model professional skills, skillsets, awarding bodies, certifiers and as such is complete in terms of QualiChain requirements for the Job Posting sub-domain. For the Course Descriptions part of the ontology, SARO will be extended to describe course providers at a more fine-grained level. SARO is particularly relevant since it was used in the H2020 European Data Science Academy project (EDSA[5]) to model and extract information from online job postings from a number of APIs for online portals. The schema.org JobPostings module will be used to extend SARO with additional descriptions that are required by QualiChain.

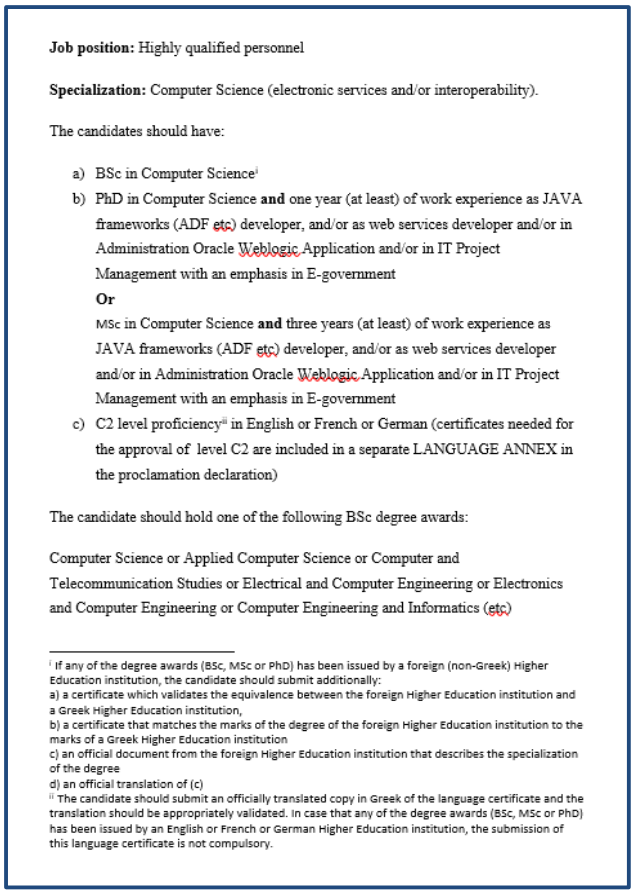
The Ethereum Ontology (EthOn[6]) for Blockchain is a valuable vocabulary that models Blockchain concepts independently of Ethereum (thus QualiChain is not limited to one platform). It provides enough descriptions, that can be partially re-used in the ontology integration exercise to link accomplishments, as verification objects, to their Blockchain verification methods.

### 3.3.3 Examples of Use

This section includes an indicative example of structuring raw data based on the QualiChain Ontology modules. Figure 10 shows a text excerpt from an online job posting that will be structured using the identified concepts from SARO and schema.org Job Posting. The reason that a job posting is used as an example is because the SARO and Job Posting ontologies are the most complete out of the ones identified in the previous section. As such, they can be efficiently used at this stage of the project to semantify data without having to define too many new concepts and relations. Additionally, job posting data are integral for the success of the pilot described in this paper as they will be analyzed to showcase popular professions in the job market as well as the skills, knowledge and experience required for individuals to be considered qualified for these positions. Consequently, students can use such data as input for the analytics and decision support modules of QualiChain to receive recommendations and develop an education plan that can best equip them to pursue a career in specific fields.

An example showing the result of the semantic lifting of content from the Job Posting is shown below (table 2). The results are in RDF-compliant triple format (subject, property, object) and link multiple objects to one subject using different types of properties (relationships). The blue text refers to existing elements (relationships and concepts), derived from the identified vocabularies (saro:<element> indicates that the element is defined in SARO, whereas so:<element> refers to schema.org Job Posting vocabulary elements). Finally, the red text refers to vocabulary elements that required addition.

[3] https://vocol.iais.fraunhofer.de/saro/visualization
[4] https://ec.europa.eu/esco/resources/data/static/model/html/model.xhtml
[5] http://edsa-project.eu/overview/
[6] https://github.com/ConsenSys/EthOn



**Job position:** Highly qualified personnel

**Specialization:** Computer Science (electronic services and/or interoperability).

The candidates should have:

a) BSc in Computer Science[i]
b) PhD in Computer Science **and** one year (at least) of work experience as JAVA frameworks (ADF etc) developer, and/or as web services developer and/or in Administration Oracle Weblogic Application and/or in IT Project Management with an emphasis in E-government
**Or**
MSc in Computer Science **and** three years (at least) of work experience as JAVA frameworks (ADF etc) developer, and/or as web services developer and/or in Administration Oracle Weblogic Application and/or in IT Project Management with an emphasis in E-government
c) C2 level proficiency[ii] in English or French or German (certificates needed for the approval of level C2 are included in a separate LANGUAGE ANNEX in the proclamation declaration)

The candidate should hold one of the following BSc degree awards:

Computer Science or Applied Computer Science or Computer and Telecommunication Studies or Electrical and Computer Engineering or Electronics and Computer Engineering or Computer Engineering and Informatics (etc)

[i] If any of the degree awards (BSc, MSc or PhD) has been issued by a foreign (non-Greek) Higher Education institution, the candidate should submit additionally:
a) a certificate which validates the equivalence between the foreign Higher Education institution and a Greek Higher Education institution,
b) a certificate that matches the marks of the degree of the foreign Higher Education institution to the marks of a Greek Higher Education institution
c) an official document from the foreign Higher Education institution that describes the specialization of the degree
d) an official translation of (c)
[ii] The candidate should submit an officially translated copy in Greek of the language certificate and the translation should be appropriately validated. In case that any of the degree awards (BSc, MSc or PhD) has been issued by an English or French or German Higher Education institution, the submission of this language certificate is not compulsory.

Figure 10. Job Posting Example

Table 2

<JobPosting1232> is a saro:JobPosting
saro:describes <Highly qualified personnel> which is a saro:JobRoleOrType
so:employmentType "Temporary"
saro:advertisedIn <Public> which is a saro:Sector
so:occupationalCategory <Computer Science/Code> which is a so:Occupation
listingOrganisation <ASEP> which is a so:Organization
so:hiringOrganisation <Ministry for the Prot. of Citizens> which is a so:Organization
so:jobLocatio <Athens> which is a so:Place
saro:requiresCapability <Electronic serv./interoperability> which is a saro:Capability
requiresWorkExperience <Java>, <…>, <…>
requiresQualification <BScCS>, <EnglishC2>, <…>, <…>
<Java> is a WorkExperience
so:experienceRequirements "3 years"
……
<BScCS> is a so:EducationalOccupationalCredential
field <ComputerScience> which is a FieldOfStudy
so:educationLevel "Bachelors"
<EnglishC2> language <English> which is a NaturalLanguage
score "C2"

## 3.4 QualiChain Analytics & Decision Support

Analytics and Decision Support, along with Visualizations, constitute a significant part of the QualiChain architecture and the specific pilot presented in this paper. For the purposes of the pilot presented in this paper, four modules will be developed; the Education Analytics module, the Career Analytics module, the Multi-criteria Decision Support module and the Visualisation engine.

### 3.4.1 Education Analytics Module

The Education Analytics Module is responsible for the analyses that concern university data and can be divided into two sub-modules; the Course Recommendation and the Curriculum Design module.

The Course Recommendation sub-module is projected to be a personalized advisor for students interested in selecting courses or acquiring new skills. It aims at recommending suitable courses, seminars and other activities to students, taking into consideration their personal career and education plans, thus targeting an efficient way to acquire certain skills and become more competent for a certain goal. The module uses as input the Course Description sub-domain of the QualiChain ontology, mainly comprising of a curriculum-like structure. Depending on the use case, the module matches this structure to required qualifications for the targeted career or education plans and suggests the optimal way to cover all (or most) of the requirements as well

as alternatives which include choices that could help build a strong CV. The latter will require the Course Recommendation sub-module to interact with the Multi-criteria Decision Support Module, to develop alternative plans that minimize skill repetition, widen the variety of the acquired qualifications and aim at producing competent graduates that are suitable for a wide variety of jobs with high demand.

The Curriculum Design sub-module facilitates the process of updating a university curriculum. It helps educators spot obsolete skills in the current curriculum (or specific courses), and recommends updates and additions to it, taking into account the current status of the job market in terms of popularity and demand. It uses as input the curriculum of a university, job market data and technological developments (relevant to technical courses), analyses them, and proposes modifications for the curriculum in order to update it and equip it with skills that are in high demand in the job market. Job market data will be acquired by the Job Market Data Extractor, a tool designed to perform web scraping (harvesting) on several predefined job posting sites and extract information concerning job market demands and skill requirements of professions relevant to a specific university's field of studies. The data will be occasionally extracted, cleansed and stored keeping the record as up to date as possible (given the nature of the job posting procedure). The final result is suggested additions and updates on existing courses, sorted by the impact they would have on the curriculum along with meaningful visualizations created by the Visualisation Engine.

### 3.4.2 Career Analytics Module

The Career Analytics Module consists of two sub-modules; the Job Matching Module and the Profiling and CV Customization Module.

The Profiling and CV Customization Module will help users build their personal profile in a semi-automatic way based on their personal information (list of qualifications, skills, courses etc.). This module will also allow users to export their personal profile in the form of a CV.

The Job Matching Module will help students (or graduates) identify the best jobs matching their profiles or CVs. As such, it will match users' skills and qualifications with requirements from various professions that are relevant to the university's field of studies. Taking into account the student profiles, the module goes through a comparison and ranking procedure, in terms of completed courses, acquired qualifications and other skills. The module performs all the necessary analytics to the available data and interacts with the Multi-criteria Decision Support Module for the evaluation process. This process will help students identify potential future careers so that they can optimize courses selection and skills acquisition to be better equipped in pursuing certain professions.

### 3.4.3 Multi-criteria Decision Support Module

The Multi-Criteria Decision Support Module will be responsible for every decision-making process that takes place on the QualiChain platform. It is projected to be a stand-alone general-purpose tool that receives all the data and parameters required to make a decision, takes into account every criterion, weighs it accordingly and produces a number of suggestions to decision makers while also helping them decide on the most suitable one. On an algorithmic level, the module consists of several methods that stem from various Decision Support Theories. The data entered in the module are analyzed by selecting one of the available algorithms and the procedure is calibrated based on the parameters defined by the user. Once the analysis is complete, the result is sent to the module that requested it in the first place.

As multi-criteria decision-making is a broad scientific field, hence there are plenty of methodologies that have been identified for use under the context of QualiChain. However, most of them stem from two general theories; the Multi Criteria Utility Theory and the Outranking Relations Theory. The Multi Criteria Utility Theory tries to define a total evaluation score for every alternative option taking into account every criterion. This approach provides an evaluation score for every option and allows ranking of the alternative decisions. Indicative methods based on this theory are the MAUT, the Weighted-sum method and the AHP (Dyer et al. 1992). Contrary to the Multi Criteria Utility Theory, the Outranking Relations Theory's goal is not to develop a metric for every alternative option but rather to develop a methodology that enables the quantification of bipartite comparisons between all the alternatives. In other words, this theory searches for alternatives that outrank other alternatives. Indicative methods of this theory are Promethee, ELECTRE I, ELECTRE III, ELECTRE Tri (Bouyssou and Vincke 1997). Out of all the methodologies that were mentioned, the most suitable ones will be selected and developed according to the pilot's requirements.

### 3.4.4 Visualisation Engine

The visualisation engine will provide users with advanced visualizations to facilitate a complete overview of the data they investigate. It will be a general-purpose tool that can be used in several services of QualiChain. The specific visualizations will be decided according to the pilot's requirements and can be bar charts, pie charts, scatter plots, comparison visualizations, heatmaps etc. The engine can be used both combined with different services (e.g., embedded in the Dashboard) and as a standalone tool, and allows the user to draw conclusions from the presented data and make more informed decisions. After the module is fed with the data that will be visually represented, it processes and converts them into a suitable format, depending on the visualization that is requested. This step is the most significant in the process, since, in most cases, the data that will enter the Visualisation Engine will need further alterations so that the final result is rendered and makes actual sense.

## 3.5 Coping with Legal GDPR-related Implications

Given that QualiChain's activities will include the gathering, processing and management of personal data (personal achievements, skills, qualifications, fields of study etc.), it needs to follow the general guidelines set by the General Data Protection Regulation (GDPR). Despite the fact, that QualiChain will not collect and process data that are considered sensitive (racial data, health-related data etc.), it still has to respect the rights of its users and have a clear strategy that concerns data governance. As such, in order to assess any privacy risks related to individuals in the collection, use and disclosure of information, a Data Protection Impact Assessment (DPIA) has been performed. The DPIA details the project's information flows and risk management and identifies potential privacy risks and their respective solutions.

One of the key risks that were identified in QualiChain's DPIA is that of compliance with the users' right for data erasure. In other words, a platform user has the right to request that his data are deleted from QualiChain. However, QualiChain's solution is heavily influenced by blockchain, which provides an immutable ledger, meaning that data cannot be removed from the database. To solve the problem of data deletion in QualiChain, the blockchain ledger will only be used to store information and hashes related to user transactions within the system (i.e. verification of a degree) and these hashes will point to the location of the actual data, outside

of the blockchain. As such, upon request for data erasure, the actual data will be deleted and all that will remain is a hash that points to an empty block of data. QualiChain's DPIA is considered as a living document and will be used to evaluate and assess the project's data management strategy and potential risks throughout its duration.

Furthermore, a consent form has been drafted to ensure legal and ethical compliance. The informed consent form, which each participant will be asked to complete prior to their participation in the pilots, aims at ensuring that the user accepts participation and is informed about all relevant aspects of the research project; it will be collected in written form after the users have been provided with clear and understandable information about their role (including rights and duties), the objectives of the research, the methodology used, the duration of the research, the possibility to withdraw at any time, confidentiality and safety issues, risks and benefits. The basic elements of the QualiChain consent form include information about the following: Data collected, Usage of users' data by third parties, Users' rights concerning their data, Explanation of why QualiChain processes user data, Cookie & Contact details.

## 4. CONCLUSIONS

In a rapidly evolving era, Higher Education is still struggling to find the appropriate level and methods of digitization, while the lack of suitable IT infrastructures is perpetuating yearly problems faced by universities and other institutions. University processes are far from being optimized, validation of certificates still requires a lot of manual labor in most cases and there is a vast pool of university data that are not being exploited. QualiChain, the project presented in this paper recognizes these issues as well as the potential of combining innovative technologies such as blockchain, semantics, data analytics and decision support for resolving them. Along these lines, the paper at hand presents a pilot case that will be applied in the ECE School of the NTUA, focusing on semantically enhancing the university's data and integrating blockchain and analytics in the school's processes.

By adding structure, security and automation in the school's processes, students will have a more efficient procedure for selecting courses and career trajectories and their own personalized profile where they can showcase validated skills and qualifications. Furthermore, professors will be able to more effectively update their courses and perform other administrative tasks. Finally, the updated curriculum that is the end-goal of the pilot is projected to increase the reputation of the university and the skillset of its students and decrease the time it takes for graduates to find employment.

## ACKNOWLEDGEMENT

The research presented in this paper has been funded by the European Union's Horizon 2020 research and innovation programme under the QualiChain project, Grant Agreement No 822404.

# REFERENCES

Association of Universities in the Netherlands (VSNU) (2017): Digitisation in academic education. Our agenda for a future-proof range of degree programmes. Available online at https://www.vsnu.nl/files/documenten/VSNU%20Digitisation%20in%20academic%20education.pdf.

BOUYSSOU, DENIS; VINCKE, PHILIPPE (1997): Ranking Alternatives on the Basis of Preference Relations. A Progress Report with Special Emphasis on Outranking Relations. In *J. Multi-Crit. Decis. Anal.* 6 (2), pp.77–85. DOI: 10.1002/(SICI)1099-1360(199703)6:2<77::AID-MCDA144>3.0.CO;2-I.

Chapman, David W.; Lindner, Samira (2014): Degrees of integrity. The threat of corruption in higher education. In *Studies in Higher Education* 41 (2), pp. 247–268. DOI: 10.1080/03075079.2014.927854.

Daniel, Ben (2015): Big Data and analytics in higher education. Opportunities and challenges. In *Br J Educ Technol* 46 (5), pp. 904–920. DOI: 10.1111/bjet.12230.

Dyer, James S. et al. (1992): Multiple Criteria Decision Making, Multiattribute Utility Theory. The Next Ten Years. In *Management Science* 38 (5), pp. 645–654. DOI: 10.1287/mnsc.38.5.645.

Garrison, D. R., & Kanuka, H. (2004): Blended learning: Uncovering its transformative potential in higher education. In *The internet and higher education* 7 (2), pp. 95–105.

Koedinger, K. et al. (2008): An open repository and analysis tools for fine-grained, longitudinal learner data. In *1st International Conference on Educational Data Mining, Montreal*, pp. 157–166.

Kontzinos, C. et al. University process optimisation through smart curriculum design and blockchain-based student accreditation. In *Proceedings of 18th International Confernece on WWW/Internet 2019, Cagliari, Itali 7-9 November [to be published]*.

Mohamedbhai, Goolam (2016): The Scourge of Fraud and Corruption in Higher Education. In *IHE* (84), p. 12. DOI: 10.6017/ihe.2016.84.9111.

South African Council on Higher Education (2013): The Aims of Higher Education. Pretoria - South Africa: Council on Higher Education (9). Available online at https://www.che.ac.za/sites/default/files/publications/kagisano9.pdf.

Turkanovic, Muhamed et al. (2018): EduCTX. A Blockchain-Based Higher Education Credit Platform. In *IEEE Access* 6, pp. 5112–5127. DOI: 10.1109/ACCESS.2018.2789929.

Williams, Christina (2002): Learning On-line. A review of recent literature in a rapidly expanding field. In *Journal of Further and Higher Education* 26 (3), pp. 263–272. DOI: 10.1080/03098770220149620.

Xing, B., & Marwala, T. (2007): Implications of the fourth industrial age for higher education. In *The Thinker* 73.

Young, Terry A. (2005): Academic Technology Transfer. In *International Journal of Intellectual Property -Law, Economy and Management* 1 (1), pp. 13–18. DOI: 10.2321/ijip.1.13.