Automating Exchange of Educational Certificates Using DRESS

By
Umair Anwar
2011-NUST-MS PhD-IT-049

Supervisor
Dr. Sharifullah Khan
Department of Computing

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in Information Technology (MS IT)

In
School of Electrical Engineering and Computer Science,
National University of Sciences and Technology (NUST),
Islamabad, Pakistan.

(June 2015)

Approval

It is certified that the contents and form of the thesis entitled "**Automating Exchange of Educational Certificates Using DRESS**" submitted by **Umair Anwar** have been found satisfactory for the requirement of the degree.

Advisor: **Dr. Sharifullah Khan**
Signature: _____

Date: _____

Committee Member 1: **Dr. Hamid Mukhtar**

Signature: _____
Date: _____

Committee Member 2: **Dr. Sarah Shafiq Khan**

Signature: _____
Date: _____

Committee Member 3: **Mr. Mujtaba Haider**

Signature: _____
Date: _____

Abstract

Student degree record is exchanged manually between educational institutes in Pakistan. In this paper, we put a push to recommend a data format and architecture for empowering these bodies to inter-exchange degree record digitally. To simplify its implementation, we recommend a mapping tool to semi-automatically create mappings between our standard and the institute data-sets.

Certificate of Originality

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics which has been acknowledged.

Author Name: **Umair Anwar**
Signature: _____

Acknowledgment

Up and above everything all glory to **ALMIGHTY ALLAH**. The Beneficent, The most Merciful and Most Compassionate. It's a great blessing from Almighty Allah that gives me the health and strength to do this research work.

I would like to special thank the **Supervisor**.

Umair Anwar

Table of Contents

List of Abbreviations

Abbreviations	Descriptions
DRES	Document Record Exchange Standard
DRESS	Document Record Exchange Standard Secured
SCHAC	Schema for Academia
LDAP	Lightweight Directory Access Protocol
FVUSPEC	Finnish Virtual University Specifications
MLO	Meta-data for Learning Opportunities
WSDL	Web Service Description Language
EHEA	European Higher Education Area
ECTS	European Credit Transfer and Accumulation System
EA	Exchange Agreement
NQF	National Qualifications Framework
EQF	European Qualifications Framework

List of Figures

List of Tables

Chapter 1

Introduction

The rest of the chapter is organized as follows. In Section 1.1, the problem statement is stated. In Section 1.2, thesis contributions are stated. In Section 1.3, we conclude the chapter with an outline for the rest of the thesis.

1.1 Problem Statement

Description of problem here.....

A formal problem statement is given as:

"....."

1.2 Thesis Contribution

Our research work contributes in ..... areas.

All contributions of ..... are summarizing as follows:

- Finding 1.
- Finding 2.

1.3 Thesis Organization

The rest of the thesis is organized as follows:

Chapter 2 discusses the state of the art related to the current research, and reviews the relevant literature aimed at finding .....

In Chapter 3, the ..... are discussed and then proposed methodology is presented.

In Chapter 4, the results are given along with detailed discussions.

In Chapter 5, the conclusion and future work is presented.

Chapter 2

Literature Review

There already exists a few standards and practices identified with exchanging degree or courses record. It is important to go through these, before proceeding onward to the new standard and the architecture we are proposing. We will review what these standards cover and what we can reuse.

2.1 Bologna Process

It intends to make European educational framework of standards engaging different countries in Europe to compare, contrast and make compatible their educational systems. [**?**]

To improve the mutual recognition of degrees and programs, education ministers from 29 countries signed bologna declaration in 1999. Other partaking countries joined the program later. [**?**] Bologna process is quite often named as European Higher Education Area (EHEA). EHEA focuses on transferability and convergence adaption by 46 countries. This process benefits Europeans and it has its significance for other educational institutes and communities. a) The leading role of European institutes, b) the lessons that are learned in the implementation of the framework of standards, and c) the practices adopted guide the educational communities around the world. 2010 was marked as the deadline across Europe for implementing the agreed specifications. [**?**]

To meet the 2010 deadline, Spain started to implement the convergence of undergraduate engineering degrees that conformed EHEA in 2008. This standardization provided some opportunities for mobility and unified measurements. [**?**]

2.2 Qualifications Exchange Standards

2.2.1 European Qualifications Framework

EQF is an agreed reference framework that helps participating countries to compare national qualifications and make them more clear, readable and understandable across Europe. The point is to advance mobility of workers and learners. This was settled upon by European universities in 2008 to relate their national qualifications to EQF. The new qualifications from 2012 carry a reference to suitable EQF level.

EQF comprises of eight reference levels, each showing what a learner knows and has the capacity to understand it. National qualifications of the partaking countries identify and relate with these eight levels raging from basic (level 1) to advanced (level 8) as shown in figure 2.1. This simplifies qualification comparison in partaking countries supporting mobility of learners and empowering them to not repeat what they have already learned.

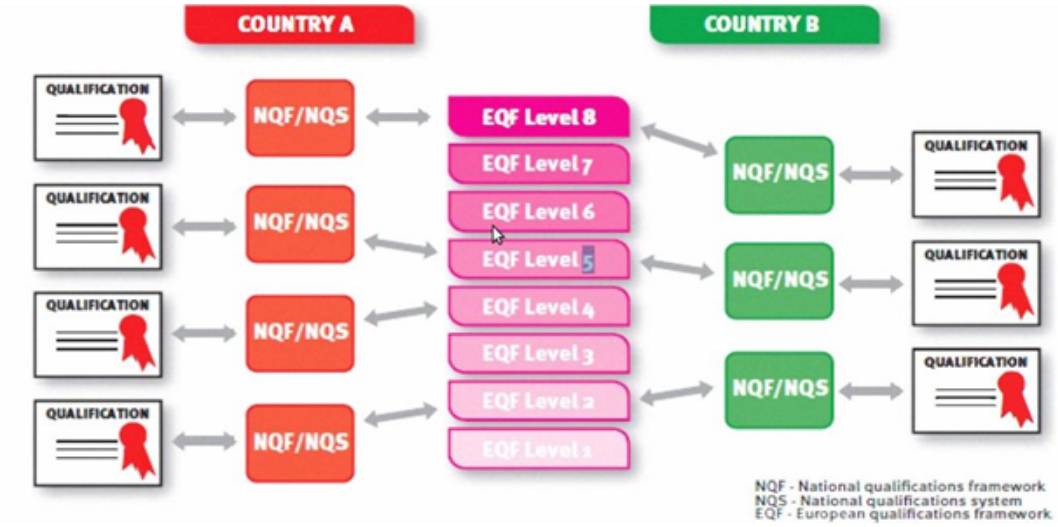

Figure 2.1: EQFs against NQFs [**?**]

EQF concentrates on learning results as opposed to concentrating on learning inputs. It covers all types of education including professional, vocational and school education. It tries to validate formal and in addition informal education.

2.2.2 Europass

Collection of five documents which intend to ease mobility when seeking employment across Europe. These include the Curriculum Vitae, the Language Passport, the Mobility, the Diploma Supplement, and the Certificate Supplement. One can fill himself the Curriculum Vitae, and the Language Passport but the rest of the documents are issued by the related authorities. It follows a standard template format system, a layout. Same format helps to achieve neutrality and transparency while presenting one's skills.

The motto as mentioned on the Europass website's homepage is as follows; "Five documents to make your skills and qualifications clearly and easily understood in Europe"

Europass has defined XML schemas for CV and Language Passport. The documents can be exported in XML format when created on Europass. These exported XML documents can be imported to Europass and converted to HTML, PDF, Microsoft Word or ODT templates.

Europass specifies JSON schema according to Internet Engineering Task Force's JSON specifications draft. The europass JSON vocabulary is close and similar to europass XML schema. The JSON objects for europass documents (CV and Language Passport) can be validated using Europass JSON validator.

All these documents have some common XML schema attributes which describe document type, printed preferences.

Europass does not explain details related to degrees or educational certificates in XML certificate.

Europass Curriculum Vitae

Europass Curriculum Vitae (ECV) is a template which one can create online and it can be exported in xml format. The ECV XML schema contains vocabularies related to document type, printing preferences, personal details, contact details, skills, and educational degrees and institutes. The XML vocabulary related to degree details is very little only to cover the scope of a CV.

Europass Language Passport

Europass Language Passport (ELP) is a template. One can create it online and export it in europass xml format. It contains XML vocabulary related to language skills and the scale of six values to score proficiency.

2.2.3 Schema for Academia

Schema for Academia (SCHAC) describes vocabulary related degrees and courses. The schema is written for LDAP (Lightweight Directory Access Protocol). It aims at promoting a common framework to inter-exchange data between educational institutes. It defines attributes that describe individuals and their LDAP profile

2.2.4 Dublin Core

The Dublin core is a simple meta-data standard consisting of set of elements to describe information resources on the network. There are two type of elements; simple and qualifiers. It has 15 simple elements and qualifiers which have additional three elements namely Audience, Provenance and RightsHolder. Qualifiers help in resource discovery.

2.3 European Learner Mobility

Some related work has been done recently and systems have been proposed based on the above mentioned standards. These are "The Mobility Project" and "The REST Mobility" projects.

2.3.1 The Mobility Project

It aimed to provide a platform and infrastructure for exchange of electronic data exchange between educational institutes. Infrastructure includes data format, architecture and the prototype software. The system will be called The Mobility later in this paper.

The Mobility is peer to peer like architecture. Nodes exchange data using SOAP base web service. Other web services like XML-RPC and REST were not used due to their limitations. XML-RPC not have developer defined datatypes and character set. REST does not imposes a standard specification, instead it follows set of rules and is used for speedy development of web service interface.

The nodes represented the universities, and their number tends to change. So there was a need for system to maintain this record and UDDI was used. He did not recommend the central or delegated private registry instead gave advantages and disadvantages of both. Central single registry has all information at one place but also it a single point of failure.

The software has two transport modules and each have web interface.

Nagrozki proposed a new standard, defined its vocabulary re-using ideas taken from SCHAC to leverage ISO and RFC rules. Some like grade, credits were taken in inspiration from Eropass Mobility.

Although The Mobility project was started by MUCI and CINECA, two European Higher Education Consortia. Many universities consortia, individual universities and companies joined in later on.

2.3.2 The REST Mobility

This is alternative implementation of The Mobility. Nagrozki's system used SOAP web service for data exchange. Karol created a RESTful implementation of the Mobility. The Mobility lacked data model. In The REST Mobility a data model is proposed since REST is resourceful. The model proposed not represents or intends to be a standard.

2.4 Information Manifold

Providing a uniform interface for querying data from many sources is the aim of Information Manifold. It enables a simple user to not worry about locating sources and manually combining results. This leads to concept of Deep Web. Data integration systems give users a common global schema called mediated schema for posting queries. To answer these queries semantic relationships called mappings are needed between mediated schema and the sources schema.

2.5 MAPQFTOOL

This tool helps comparing National Qualification Frameworks against European Qualifications Framework in Europe. This automates the process of creating mappings between these frameworks and stores the mappings in the database.

Related to the [**?**] research related to the regression models for predictions is concerned, Ali et al. [**?**] discussed the application of linear regression for future prediction using *SPSS* (Statistical Package for the Social Sciences). They found the P-values, beta scores, R^2, mean and standard deviation parameters that helped to learn good models for future prediction.

These regressions were found using response variable y and predictor variable x as shown in equations 2.1-**??**.

$$y = w_0 + w_1 * x \text{ (Linear)} \tag{2.1}$$

$$\mathbf{D} = \mathbf{AV} \tag{2.2}$$

where

$$\mathbf{D} = \begin{bmatrix} d_1[n] & d_2[n] & d_3[n] \end{bmatrix} \mathbf{A} = \begin{bmatrix} a_x[n] & a_y[n] & a_z[n] \end{bmatrix} \tag{2.3}$$

Chapter 3

Requirements Analysis

3.1 Definitions

We define the basic terms that are used in exchange of documents. It is necessary to understand these before we go through the requirements.

Exchange Agreement: It is an understanding between partaking institutes, between the requester and the provider, for exchanging the educational certificates. This agreement comprises of; a) web-service access point b) authentication credentials

Requester: The partaking institute asking for the exchange.

Provider: The partaking institute providing the details.

Coordinator: A person responsible for signing and exchange of agreements between institutes.

3.2 Business Process

This section describes the business that is involved in the execution of exchange of different educational institutes. The process is explained using the sequence diagrams.

3.2.1 Make Exchange Agreement

For two universities to exchange data, they have to create an exchange agreement first. The agreement will have the web-service access point and exchange secrets. These details are used for requesting exchange and for authenticating the requester.

After the implementation of web-service by the provider. The requester fills a form and asks for the credentials. The emails the access point and

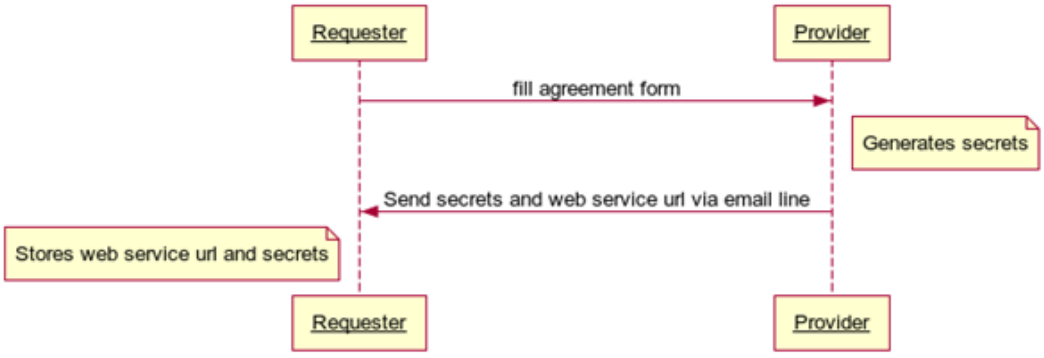

Figure 3.1: Making Exchange Agreement

exchange secrets to the Coordinator.

3.2.2 Find Student Data

To find a student record, the requester asks a provider from the agreed providers list for a student record. The provider sends back list of documents associated with the student.

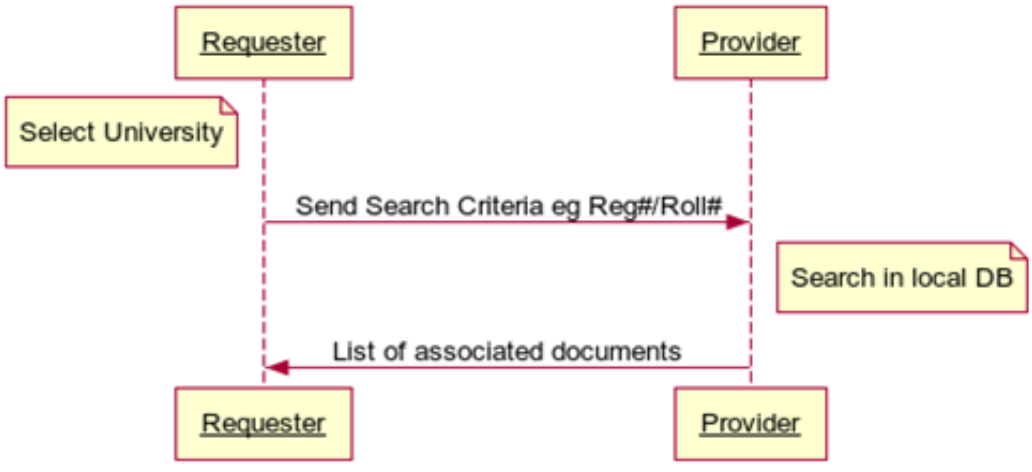

Figure 3.2: Finding Student Data

3.2.3 Exchange a Document Details

To exchange a document details, the requester asks a provider with search criteria and document type. The provider sends back the document details

using DRESS.

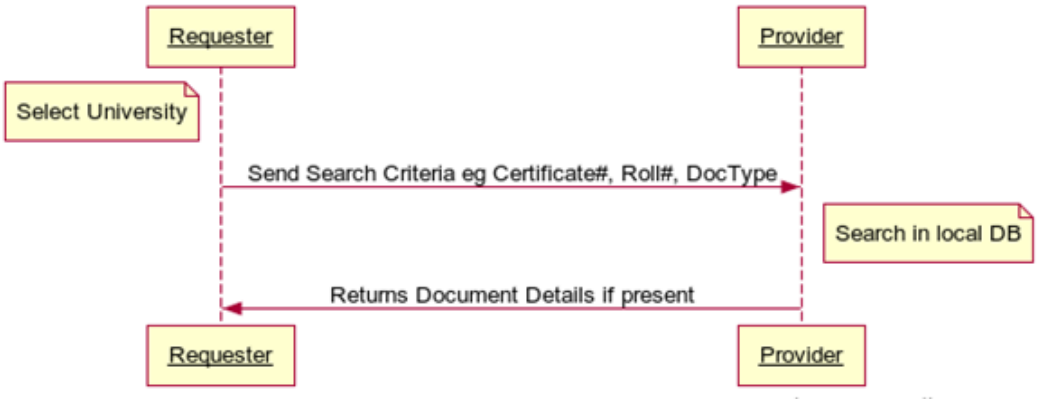

Figure 3.3: Exchanging a Document Details

3.3 Mapping and Web Service Challange

It is very important to understand that to implement the scenarios mentioned in section 3.2, every provider must implement a web-service based on DRESS. This is a challenging task since every partaking institute (provider) has different schemas and different database management systems. This schema is required to be mapped to DRESS. This must be served to the requester using a web-service. This creates an opportunity for a semi-automated tool which maps the schemas to DRESS and generate a web-service automatically.

3.4 Software Specifications

Based on the business process and the challenges we discussed, functional and non-functional requirements are;

3.4.1 Functional Requirements

3.4.2 Non-functional Requirements

Chapter 4

Architecture & Design

From the requirements analysis, we suggest student exchange system should have distributed architecture. As partaking institutes are autonomous and themselves maintain the its own data. It signs agreements independently for exchanging data with other universities. Each can be a requester plus a provider of data. The circles/nodes in the figure below represent universities. The arrows represent exchange of data.

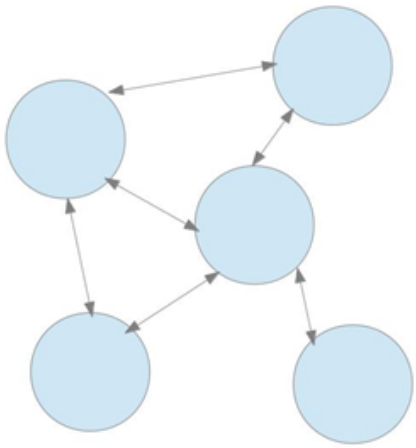

Figure 4.1: Multiple Nodes Exchanging Information Independently in Distributed Architecture [**?**]

This peer to peer like distribute architecture has benefits over adding a middle agent or central server in the system. 1. Avoidance from single point of failure. 2. Lesser load. 3. Each university having control over its own data and thus building trust in the system.

We suggest student exchange system will have distributed architecture. Each university has its own data and signs agreements with HEC for exchang-

ing data with other universities. Each can be a requester plus a provider of data. The circles/nodes in the figure below represent universities

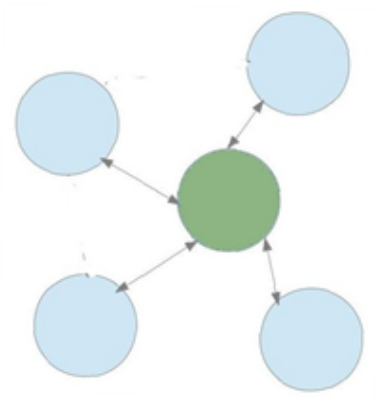

Figure 4.2: Multiple Nodes Exchanging Information Through a Central Control Authority in Distributed Architecture

There are some choices to be made at this point. We will be using web services for exchanging data as they provide a high abstraction from network issues and use well known standards like XML over HTTP. There are some XML based data exchange protocols on web. These are XML-RPC, SOAP, and REST.

The nodes will exchange data using SOAP based web service in our system. We chose SOAP as it forces to follow a formal standard and supports developer defined data types.

The number of universities can increase when agreements are signed with new universities for exchange data. The web service URLs need to be saved so that requester can retrieve this URL and request that university. This can be achieved by developing a custom system "URL registry" for saving web services URLs. Now we have to make a choice. URL registry can be global or each requesting node can have its own private URL registry. We will use private registry to avoid single point of failure and to minimize load.

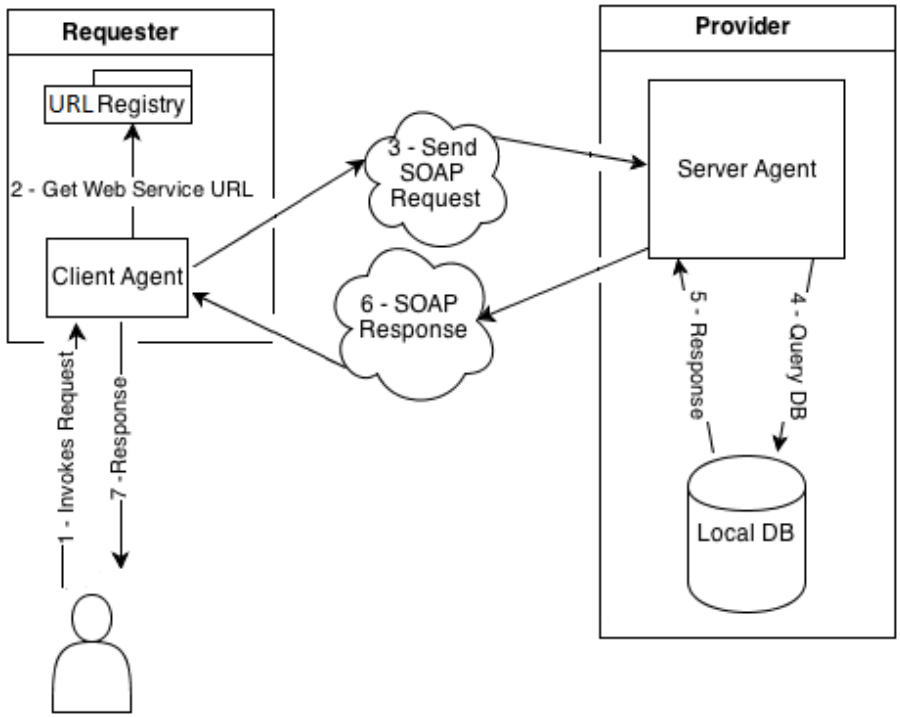

Figure 4.3: Architecture Diagram

Chapter 5

Implementation

In this chapter, the methodology that is used for modeling is explained.

Chapter 6

Conclusions

6.1 This is the End

In this research we reported the design and implementation of.

Table 6.1: Detailed Accuracy By Class (Naïve Bayes).

Class	TP Rate	FP Rate	Precision	Recall	ROC Area
Walking	0.789	0.031	0.675	0.789	0.818
Running	1.000	0.000	1.000	1.000	0.999
Climbing Stairs	0.450	0.054	0.731	0.450	0.807
Descending Stairs	0.833	0.023	0.814	0.833	0.919
Driving	0.933	0.018	0.897	0.933	0.934
Cycling	1.000	0.000	0.880	1.000	0.992
Inactive	0.933	0.000	0.996	0.933	0.961
Weighted Average	**0.847**	**0.040**	**0.846**	**0.847**	**0.891**

Figure 6.1: Feature rank by information gain.

Table 6.2: Confusion matrix (Naive Bayes).

a	b	c	d	e	f	g	← Classified As Actual Activity ↓
59	0	6	5	1	0	0	a ← Walking
0	**50**	0	1	0	0	0	b ← Running
5	0	**34**	4	0	0	0	c ← Climbing Stairs
4	0	3	**29**	5	0	0	d ← Descending Stairs
0	0	0	0	**32**	0	4	e ← Driving
2	0	0	1	0	**20**	0	f ← Cycling
0	0	0	0	5	0	**30**	g ← Inactive

Appendix A

Feature Extraction Walking

Features plotted in Matlab.

Bibliography