Student Degree Record Exchange Standard

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.

(July 2015)

Approval

It is certified that the contents and form of the thesis entitled "**Student Degree Record Exchange Standard**" 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: **Ms. Sana Khalique**

Signature: _____

Date: _____

Abstract

Bologona process aimed to ease the mobility of students across Europe. Hence, efforts were put developing standards and proposing suitable architectures that fit all across Europe. Inspired by this, the mobility project tried to gather and reuse all the work previously done in related areas. Mobility reused the vocabularies, ideas and focused on the mobility of students. However, it created an opportunity for a semi-automated mapping tool for mapping proposed standard to the heterogeneous schema of different institutes. It lacks enough vocabulary to cover exchange of information for some educational certificates in Pakistan.

We suggest a prototype infrastructure that provides more control and improves this exchange of information between partaking institutes by covering more detailed vocabulary. The term infrastructure includes both the architecture and our proposed standard (DRESS). We avail the opportunity that The Mobility Project provided and suggest a mapping tool that semi-automatically create mappings between our standard and the heterogeneous data-sets of different partaking institutes.

To evaluate this research, we created a test-bed environment and validated the proposed standard XML schema definitions against the generated

XML documents using different evaluation techniques.

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

All gratefulness and wonderfulness to **ALLAH** and with His blessings, I am in good health and he gave me the ability to complete this work. I pray for more of Allah's blessings. It would not be possible without the efforts that were made by my **parents** for supporting me during this duration and throughout my life.

I am truly thankful to my supervisor **Dr. Sharifullah Khan** for his guidance and who had been a great coach. I want to express my gratitude towards him for believing in me and facilitating me for the completion of this work. I appreciate **Dr. Hamid Mukhtar**, **Dr. Sarah Shafiq Khan** and **Ms. Sana Khalique** for suggesting and supporting me.

I am thankful to all my colleagues, co-workers and contributors who were supportive in completing this research.

<div align="right">

Umair Anwar

</div>

Table of Contents

List of Abbreviations

Abbreviations	Descriptions
DRESS	Document Record Exchange Standard of Students
SCHAC	Schema for Academia
LDAP	Lightweight Directory Access Protocol
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
XSD	XML Schema Definition
ECV	Europass Curriculum Vitae
ELP	Europass Language Passport

List of Figures

List of Tables

Chapter 1

Introduction

With time passing, the exchange of official and legal documents digitally is getting more and more importance. In near future, it will be inevitable to transform our systems to facilitate this change. Same is the case with the exchange of educational certificates between different partaking institutes and authorities. Its importance can be imagined by the number of students and institutes involved.

In 1999, it was decided in Bologna declaration to create European Higher Education Area which facilitate to standardize the exchange of information across Europe. Alone in Europe, more than four thousands partaking institutes with more than two million students were involved in this process in academic year 2009-2010.

In April 2015 Bureau of Statistics Pakistan published statistics about Pakistan's population with different levels of education [12]. There are 1,712,308 people having education of BA/BSc & Equivalent and 618,937 people with MA/MSc & Equivalent or Above education. [10]

1

Ministry of Federal Education and Professional Training and USAID published detailed report on Pakistan Education Statistics 2010-2011 [10]. We are mentioning the higher education enrollment statistics from this report to get an idea for need of this research. According to this report there are 135 higher education institutes with total enrollment of 1.108 million. This number includes all bachelors, masters or above enrollments. It is also worth mentioning that 1,558 degree colleges have an enrollment of 0.431 million students.

According to Pakistan Education Statistics 2013-2014 report there are 161 higher education institutes with total enrollment of 1.595 million. This number includes all bachelors, masters or above enrollments. It is also worth mentioning that 1086 degree colleges have an enrollment of 1.336 million. [11]

The statistics we collected from Pakistan Education Statistics reports are represented in table 1.1. These statistics suggest that the number of students are increasing. Although the number of degree colleges has decreased but the number of students have almost tripled in degree colleges from 2010 to 2014. The number of enrollments in universities have also increased. These statistics signify the importance of this research.

Table 1.1: Statistics from Pakistan Education Statistics Reports

Academic Year	Enrollments in Colleges	Enrollments in Universities
2010-2011	0.431 million	1.108 million
2013-2014	1.336 million	1.595 million

It is a well established fact that more and more systems are digitized

every year. The educational institutes are also making their record digital and this trend is increasing. For example, more and more universities are implementing electronic student information systems to keep student courses and credit record. However, the documents often called degrees/certificates issued by these autonomous bodies are still exchanged in paper form.

This manual approach of exchanging physical documents and re-entering the data again manually in digital or physical form by people is error prone and exhaustive. As a result the local data-sets of these autonomous bodies has different record for the same individual.

This creates an opportunity for a common data-exchange standard which bring a common ground for the exchange of information. To resolve the issues facing in manual exchange of documents, the institutional systems must talk to each other using a common standard. This would make the entire process more dependable and less error prone.

This research digs into the currently adopted solutions, standards and the new research in the area of student mobility and exchange of official documents related to students. It suggests a prototype infrastructure including data format and the architecture to exchange degree records digitally.

Each institute is an autonomous body maintaining its data separately. The schema of the data is different in different institutes. To exchange documents, partaking institutes must implement web-services based on our proposed architecture and Document Record Exchange Standard for Students (DRESS). We will use only DRESS in rest of the thesis. This approach is

very time consuming, inefficient, and error prone. This provides us another opportunity to suggest a tool that maps these different types of databases schema from different institutes to DRESS.

To handle the different levels of heterogeneity, we came up with a mapping tool which semi-automatically maps institutes data-sets to our proposed standard and creates web-services to access their data.

The remaining chapter is sorted out as this; In Section 1.1, we explain the problem Statement. In Section 1.2, we discuss what this research has contributed. In Section 1.3, we explain how this research is organized into chapters.

1.1   Problem Statement

The existing standards that are used for exchanging educational certificates partially cover attributes that are required for the exchange of these documents. These are developed mostly as part of other projects and thus only focusing in this domain partially. Hence, most of these lack some basic attributes that are necessary for a standard to be usable.

We take a step and put an effort in suggesting a standard that is generic, reusable, extendable and based on technologies that are platform independent and widely usable accompanying tools that make it easy to implement and use. So, our formal problem statement becomes;

"To propose a document record exchange standard for students that en-

ables the partaking institutes to exchange educational certificates while hiding the heterogeneous details. The proposed standard must reuse the ideas already suggested in existing standards and must be generic, extendable, reusable and easy to use. The standard must accompany the tools that make it easily usable."

1.2 Contributions

This research made following contributions;

- It explored existing standards and tools that are available for the exchange of educational certificates. It identified the attributes that are lacking in these standards.

- It suggested DRESS, a new educational certificates exchange standard, that reuses the ideas in previous standards and cover some attributes that were missing in these standards.

- It suggested a mapping tool that maps different heterogeneous schema to DRESS semi-automatically.

- It suggested an algorithm for the exchange of information over the proposed architecture using DRESS.

1.3 Goals

- Propose a generic and extendable standard for the exchange of educational certificates with well defined vocabulary.

- Suggesting a suitable prototype architecture for the exchange of educational certificates such that every data owning entity is the owner of its own data to build trust and have built-in control for authenticity.

- Implementation of a mapping tool to ease mapping of heterogeneous schema with our proposed standard DRESS.

1.4 Thesis Organization

Let us now discuss how we have organized the remaining research in chapters:

In Chapter 2, we go through the literature review. We discuss the work that has already been done for exchanging student information. We look into the already existing standards, a few implementations and related projects. We will go deep in details in this section, to understand the ideas and technologies that are already in use.

In Chapter 3, we discuss the business process in detail using sequence diagrams and define the functional and non-functional requirements. In Chapter 4, the proposed architecture and design are discussed in detail. In Chapter 5, our proposed standard "DRESS" is presented. We discuss its vocabulary and the possible values. In Chapter 6, we discuss the implementation and the test-bed environment we created. In this chapter, we also evaluate whether the goals are achieved. In Chapter 7, the research work is concluded and possible 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. [2]

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. [8] Bologna process is quite often named as European Higher Education Area (EHEA). EHEA focuses on transferal and convergence adaption by 46 countries. This process benefits Europeans and it has its significance for other educational institutes and

communities. The significance of EHEA is due to these reasons;

1. The leading role of European institutes,

2. the lessons that are learned in the implementation of the framework of
   standards, and

3. the practices adopted guide the educational communities around the
   world.

2010 was marked as the deadline across Europe for implementing the
agreed specifications. [5] To meet the 2010 deadline, Spain started to imple-
ment the convergence of undergraduate engineering degrees that conformed
EHEA in 2008. This standardization provided some opportunities for mobil-
ity and unified measurements. [5]

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 un-
derstandable 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.
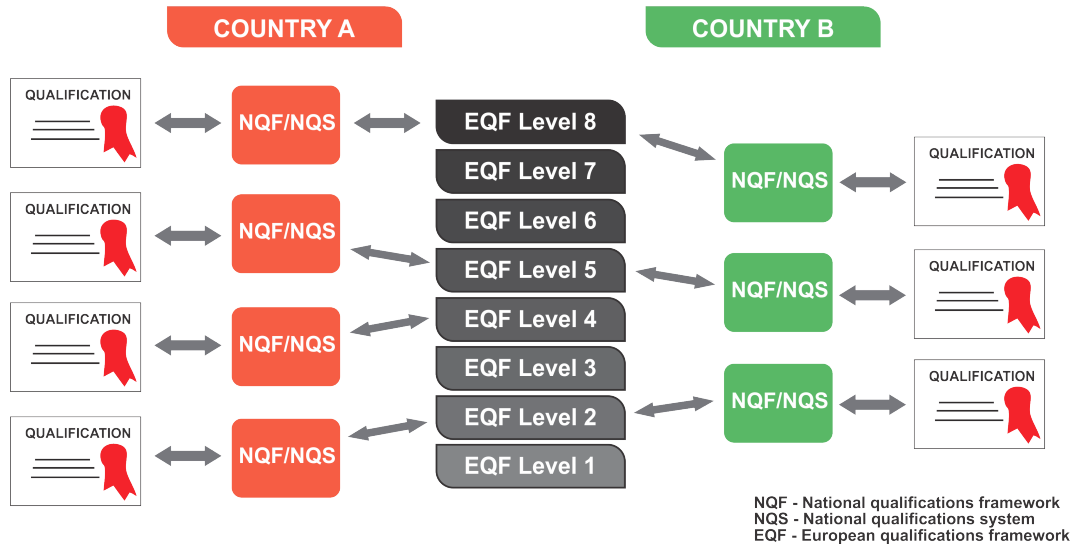
Figure 2.1: EQFs against NQFs [3]

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" [13]

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.

2.2.2.1 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.

2.2.2.2 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 [9]

The need for the inter-exchange of information between institutes across Eu-
rope has highlighted the importance of common attributes for this exchange
to take place. Schema for Academia (SCHAC) is the result of the attribute
coordination between different institutes. It plans to define and advance com-
mon attributes in the field of higher education to encourage data-exchange
between institutions. It does not plan to supplement different the national
schemas, rather it suggests a coordinated framework on top of the different
national schemas.

Schema for Academia (SCHAC) describes vocabulary related degrees and
courses. The schema is not technology dependent and 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.

It is a collection of schemas which can be classified into following cate-
gories;

- Personal Characteristics

- Location Information

- Student Information

- Employee Information

- Linkage Identifiers

- Administration Information

- Confidentiality (Visibility)

- Authorization, Entitlements

- Group-related Attributes

SCHAC has a clearly defined meta-information for defining an attribute.
To discuss the ideas used by SCHAC, let us look in detail the schacGender
attribute as an example shown in Table 2.1. It uses ISO-standards where-
ever possible. In schacGender attribute, it is using values from ISO-5218.
However, it lacks hierarchy which prevents from reuse-ability and it can be
considered its disadvantage.

Table 2.1: Example of SCHAC attribute: schacGender [9].

Name	schacGender
Description	Male or Female, specify the legal gender
Format	0 - Not Known, 1 - Male, 2 - Female, 9 - Not Specified
Values	Single
References	ISO-5218
Example	schacGender = 1

It is also worth mentioning that SCHAC has a category student information for curriculum, major and degree but no attribute is defined. This is because SCHAC is not completed till now and it is in progress.

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. [4]

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 [4]. 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

data-types 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.

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. [6]

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. [7]

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.

2.6   Summary

We explored existing standards and technologies for exchanging information regarding educational documents. The bologna declaration initiated the standardization of educational systems across Europe. European qualification framework, Schema for Academia and Europass are efforts that were put in place to standardized exchange of qualifications. The Mobility Project and The REST mobility are implementations to ease mobility of students across Europe. These details helped us better understanding the problem domain.

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.

1. Exchange Agreement:

   It is an understanding between partaking institutes, between the requester and the provider, for exchanging the educational certificates. This agreement comprises of;

   - web-service access point
   - authentication credentials

2. Requester:

   The partaking institute asking for the exchange.

3. Provider:

   The partaking institute providing the details.

4. 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.
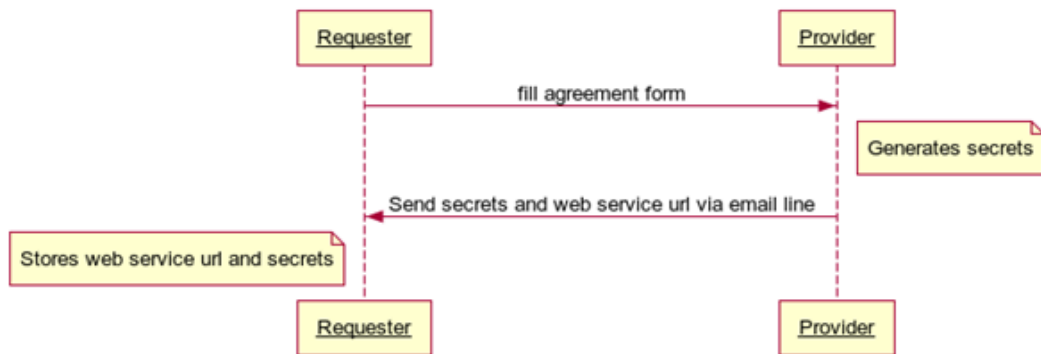

Figure 3.1: Making Exchange Agreement

After the implementation of web-service by the provider, the coordinator sends an endpoint URL and API access credentials to the central authority. The central authority store these in the URL registry.

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.
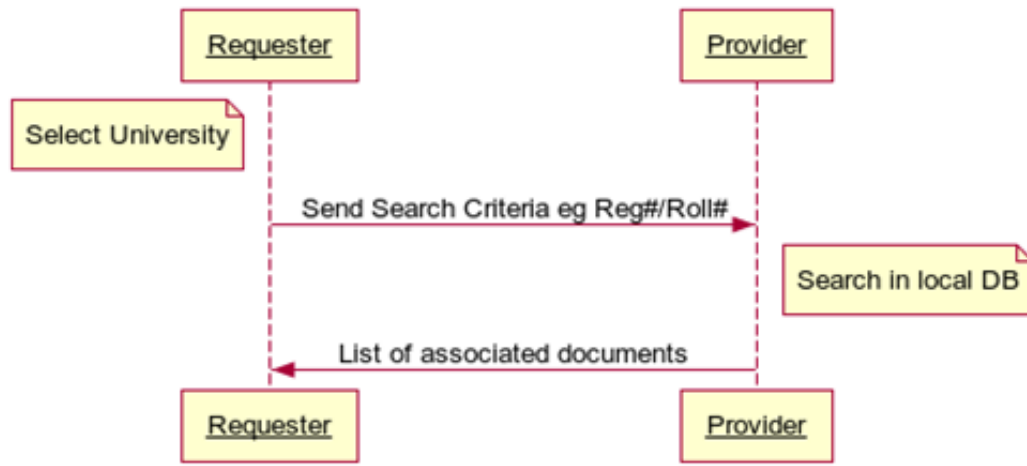

Figure 3.2: Finding Student Data

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.

3.3 Mapping and Web Service Challenge

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
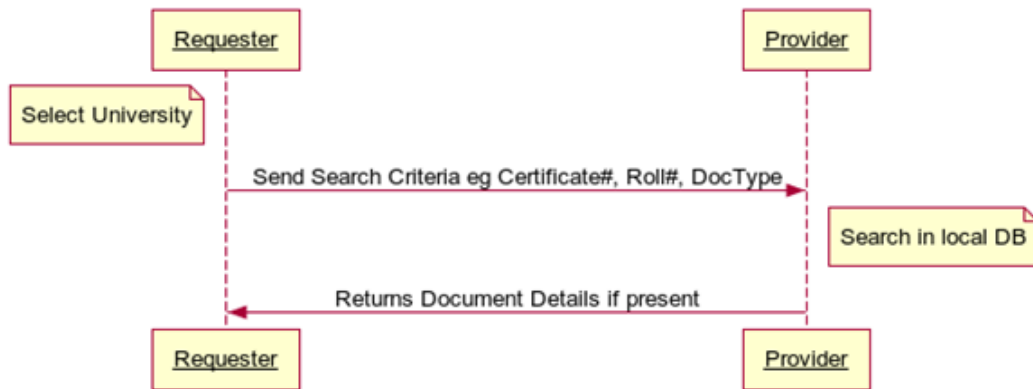
Figure 3.3: Exchanging a Document Details

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

1. The client must be able to fetch educational certificates from different providers with database schemas of different heterogeneous levels. This exchange must satisfy all the use cases discussed in section 3.2.

2. Provider must authenticate the client.

3. The system ought to be testable with sample data.

4. The addition of new providers into the system must not involve any

coding.

3.4.2 Non-functional Requirements

1. The proposed standard and the supporting software must be generic. This means using this standard and the proposed architecture, the software must be able to exchange documents with different attributes and types.

2. It should provide control and authority.

3. It should build trust among the data sharing entities.

4. It should be easy to integrate and use.

5. The tools and technologies required should be freely available, widely used and independent of any specific hardware or operating system.

3.5 Summary

We explored the business process involving getting list of documents against a student and exchanging the documents. This presents us with a challenge to fetch this data from heterogeneous data sources. Considering the business cases and the challenges, we have set functional and non-functional goals.

Chapter 4

Architecture & Design

From the requirements analysis chapter 3 and the goals discussed in section 1.3, we conclude that the educational certificate system will be a distributed system. This system is based on DRESS and logically layered architecture.

4.1   Distributed System

We suggest an educational certificate exchange system that is distributed. As partaking institutes are autonomous and themselves maintain their data. They sign agreements independently for exchanging data with other universities. Each can be a requester plus a provider of data. The circles/nodes in the figure 4.1 represent universities. The arrows represent exchange of data.

This peer to peer like distributed architecture has benefits over adding a middle agent or central server in the system.

1. Avoidance from single point of failure.
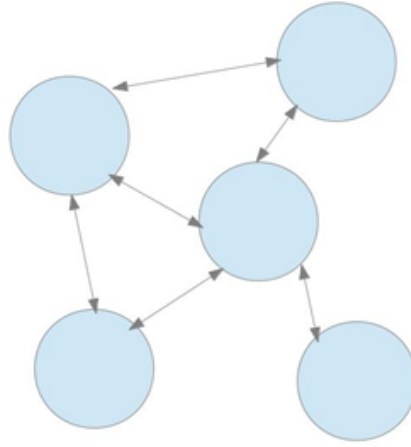
2. Lesser load.

Figure 4.1: Multiple Nodes Exchanging Information Independently [4]

3. Each university having control over its own data and thus building
   trust in the system.

However, we also want to have control and authenticity. With fully dis-
tributed architecture we lack these attributes. Hence, we suggest student
exchange system that will have distributed architecture with a central node
that acts as a central authority as shown in figure 4.2. Each university has
its own data and signs agreements with the central node for exchanging data
with other universities. This central node is a requester while the rest are
provider of data. The circles/nodes in the figure below represent the partak-
ing institutes.

The introduction of a central node has a few benefits as well as a few
compromises. Now we discuss these one by one.

1. Although we compromised avoidance from single point of failure by
   introducing the central node but gained more control and authority.

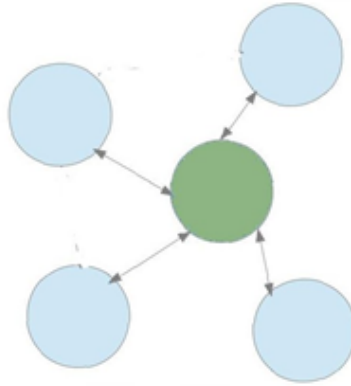2. Load is still distributed across the providers.

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

3. Each university having control over its own data, thus building trust in the system. The introduction of central authority even helps to gain more trust in the system.

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 requesting node can have its own private URL registry.

As we have made the necessary decisions, let us look the system design in more detail. The figure 4.3 shows the components in a provider and a requester.

4.1.1 Requester

The requester is the central node requesting the providers for the exchange of certificates. The requester has two components;

1. Client-agent

2. URL Registry

Whenever the requester receives a request, its client-agent finds the web-service URL from URL Registry and sends a request to the corresponding provider using DRESS. The URL Registry is used as private registry for the requester, since there is only one requester node.

4.1.2 Provider

Every node that owns data and serves the educational certificates to the requester is a provider. Maps are generated at these nodes using the mapper and stored at the provider. These mappings are later used by the server-agent for accessing data from the database using the query builder. The query builder builds queries to access different heterogeneous databases for the mapper and for the server-agent.
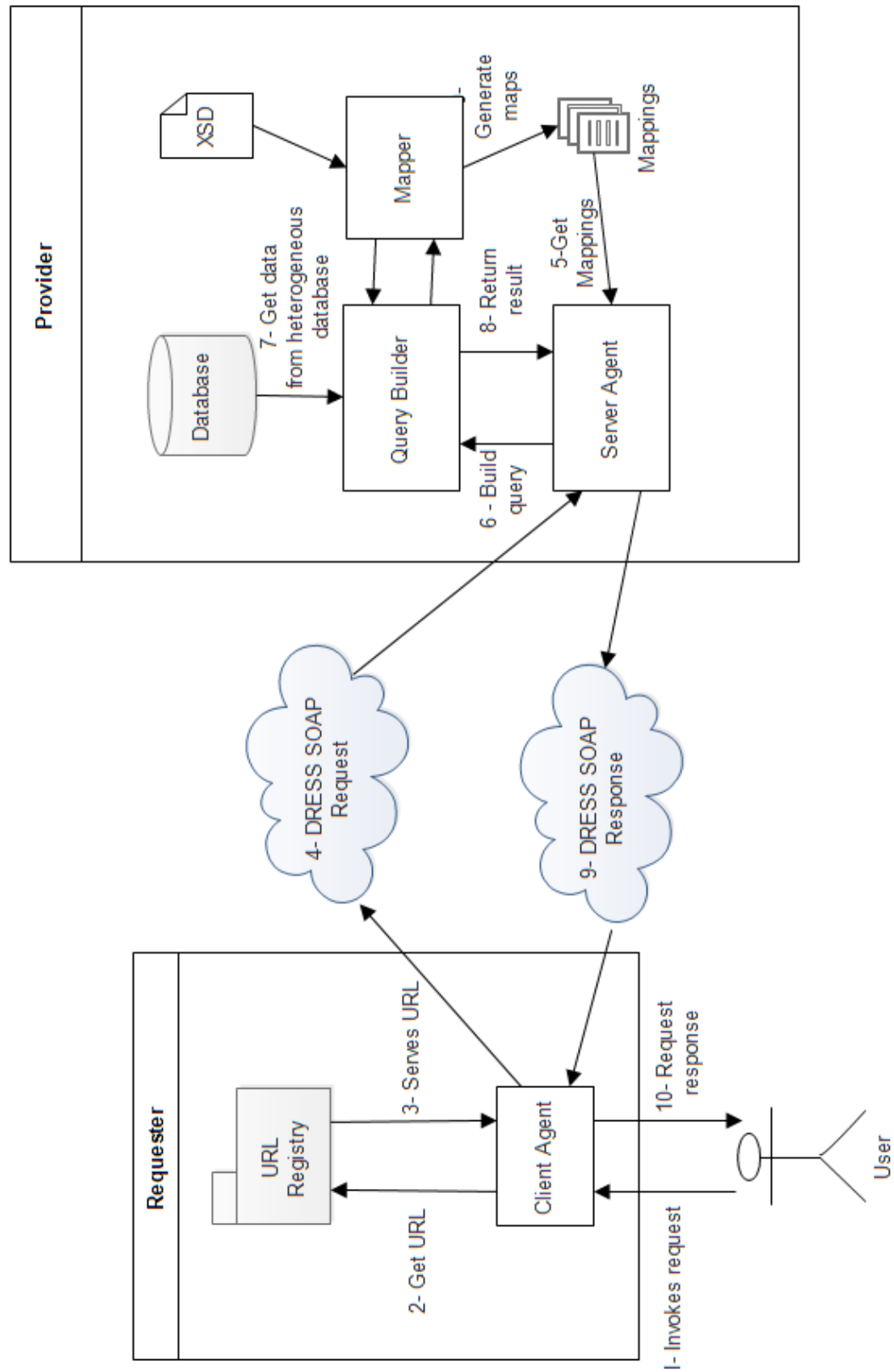
Figure 4.3: Architecture Diagram

4.2   Summary

We presented a distributed system where every node is a provider that serves the requests expect one central node which is the requester. We see that by having central node over distributed architecture, we have more control and authority. SOAP is chosen over REST and RPC since it forces to follow a formal standard.

Chapter 5

Document Record Exchange Standard for Students

In this chapter, we discuss the vocabulary of the Document Record Exchange Standard for Students abbreviated as DRESS. It is an information exchange standard for the exchange of educational certificates. In the coming sections, we will discuss the vocabulary including the attributes and the possible values for each of the schema definitions that DRESS provide.

5.1   Vocabulary

Before discussing DRESS in detail, there are some important considerations to remember regarding a standard. While proposing the vocabulary, we have focused on the following objectives;

1. The vocabulary of the standard must address the problem domain well.

2. It must be easy to understand.

3. It must be reusable and extendable.

4. It must be in conformance with the existing standards and norms as much as possible.

We have reused some of the vocabulary of the mobility project [4]. We have reused the ideas from the SCHAC [9] and used ISO standard wherever possible. Figure 5.1 represents partial representation of a schema document.

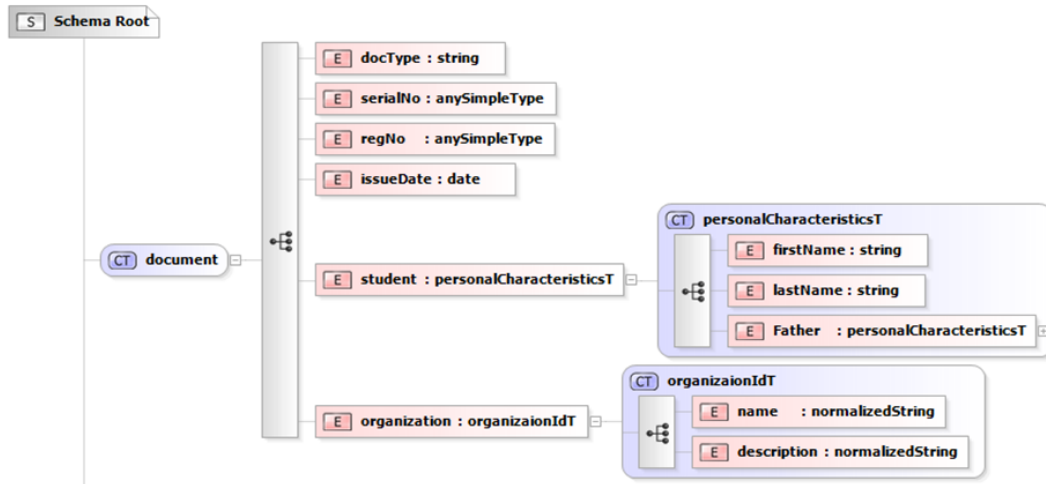Figure 5.1: Fragment Representation of a Document Schema

- docType: Enumeration of string values shown in Table 5.1 describing the document that is under consideration.

Table 5.1: Possible values for docType

Value	Description
REGISTRATION	Registration document
DEGREE	Degree document
TRANSCRIPT	Transcript document

- docTitle: string containing title of the document under consideration.

- serialNo: string containing a document serial number.

- regNo: string containing a student's registration number.

- rollNo: string containing a student's roll number.

- issueDate: date containing a valid issue date for the document. The date must follow ISO 8601 e.g. 2015-07-14

- student: It is a complexType personalCharacteristicsT and comprises of three fields as shown in figure 5.2.

    - firstName: string containing student's first name.

    - lastName: string containing student's last name.

    - Father: It is a complexType personalCharacteristicsT.
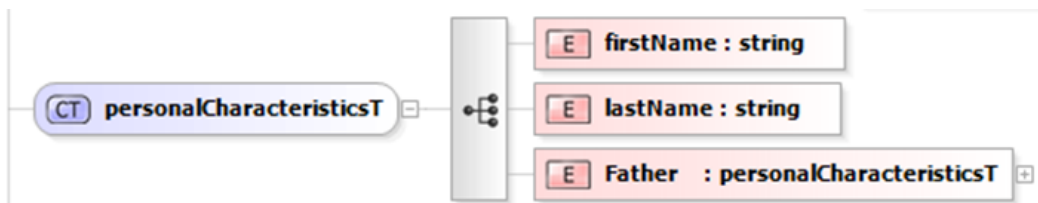

Figure 5.2: Representing Type Personal Characteristics

- organization: It is a complexType organizationIdT and comprises of two fields as shown in figure 5.3.

    - name: string containing organization's name.

    - description: string containing organization's details.

- marks: It is a complexType gradeT and contains an attribute "gradingScheme" and a value as shown in figure 5.4.
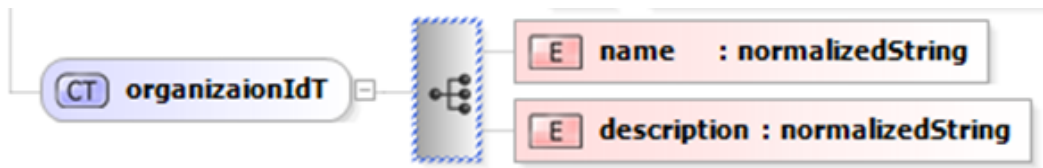
Figure 5.3: Representing Type OrganizationIdT

- Attribute gradingScheme: It is an enumerated list of gradingSchemes.

- value: string containing a measurement of student's marks according to the grading scheme.
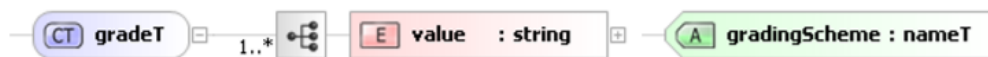
Figure 5.4: Representing Type gradeT

- course: It is a complexType compromising of following fields.

  - name: string containing title of the course.

  - description: string containing a short description about the course.

  - weight: It is complexType with a string value and an attribute "weightScheme".

  - grade: It is a complexType gradeT and contains an attribute "gradingScheme" and a value.

- session: It is a complexType containing two fields.

  - startDate: the date student registered with the institute. The date must follow ISO 8601 e.g. 2015-07-14

  - endDate: the date student completed a qualification. The date must follow ISO 8601 e.g. 2015-07-14

- examSystem: Enumeration of string values shown in Table 5.2 describing the examination system of the issuing authority.

Table 5.2: Possible values for examSystem

Value	Description
QUATERLY	exams after 4 months
SEMESTER	exams after 6 months
YEARLY	exams after 1 year
2YEARLY	exams after 2 years

5.2   Summary

We explored the vocabulary required for the exchange of educational documents. We discussed the elements, their types and the possible values that can be assigned. This covers required vocabulary for registration, degree and transcript XML schema definitions.

Chapter 6

Implementation & Evaluation

6.1 Implementation Stages

Automating exchange of educational certificates using DRESS can be divided into four stages of implementation. The very first stage begins with defining the schema, that is our proposed standard and writing an XSD. The second stage involves implementation of a mapping tool that generates mapping and filters over heterogeneous schemas. The third stage is the development of a web-service that uses the generated mappings and filters to serve requests using DRESS. In this stage we implement a soap service describing the possible operations in a WSDL that satisfy the business processes described in chapter. The fourth stage is the last stage which includes development of a GUI interface for a Client which is a central authority to send and retrieve information from different institutes/nodes.

6.1.1 Defining the Schema

DRESS defines vocabulary for educational certificates. Using this vocabulary, XML documents are created for the exchange of documents to take

place. There are different types of documents like DEGREE, REGISTRA-
TION and TRANSCRIPT. Each document has a corresponding XML docu-
ment. At this stage we create the XSDs for these XML documents.

6.1.2   Map Generation

Using the defined schema, we create mappings at stage 2. A semi-automated
wizard is run on the provider node to create mappings and filters. These
mappings and filters are stored in JSON format as properties. The Mapping
Tool and generated mappings reside on the provider node.

6.1.3   DRESS based Web-Service

At stage 3, a SOAP based web-service uses the defined mappings and filters.
This piece of code fetches data from the provider database using the map-
pings. The fetched data is then served in DRESS format to the request/client.
The SOAP base web-service runs on the provider node.

6.1.4   Client

At stage 4, a request mechanism is developed on the central Authority. A
client tool is developed that enables us to request the DRESS based web-
services using a graphical user interface. It then displays the exchanged
document in readable format.

6.2   Tools & Technologies

In this section, we discuss some important implementation decisions, and
the technologies we chose for implementation. These decisions are based on

following concerns;

1. The tools we developed will be freely available, so we choose to use open-source technologies as much as possible.

2. The implementation and deployment involves many parties. For example, every institute has to use our mapping tool to generate mappings. This piece of software must be easy to use and should be able to run with zero configuration. Hence, we chose mappings to be stored as properties on the secure server disk in JSON format. This enables us to run this mapping wizard tool on a web-server with almost zero configuration.

6.3   Evaluation

In this section we present the evaluation techniques that can be used and the ones we used to evaluate this research against the target goals discussed in section 1.3. We list down these techniques and later discuss these in detail.

1. Evaluating against functional & non-functional requirements.

2. Using program validation technique. [16]

3. Fault-based testing of XML Schema.

All the above mentioned techniques require us to create a test bed environment. Hence, we created an environment of three nodes for the evaluation purpose.

6.3.1 Test-bed

Keeping in mind the end goal to mimic a true domain and to simulate more realistically, it is better to setup separate servers for each node/entity. So we setup separate server for each node with different technologies and services running on these machines. We take three nodes as shown in figure 6.1. The two nodes (Node1 with horizontal lines and Node2 with vertical lines) are the providers and the node (Node0 with grid pattern) in the center is the node that acts as a central authority. Both provider nodes are running database management systems and schema that are different from each other. Thus simulating a heterogeneous environment.
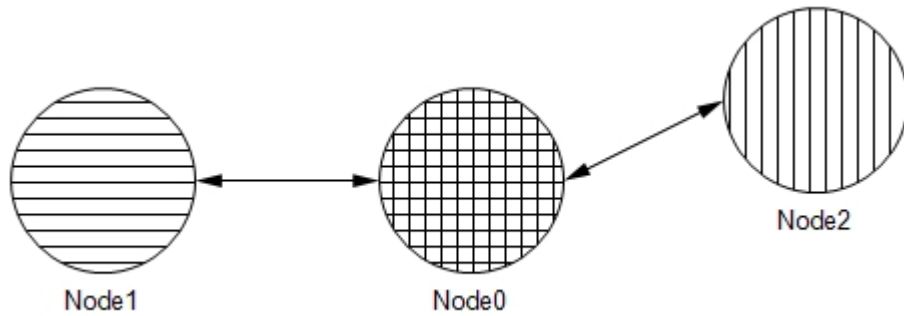

Figure 6.1: Three Nodes Exchanging Information Usign DRESS with one Node as Central Control Authority

We begin with setting up the providers first by generating mappings over the heterogeneous schema and then installing the web-services that enable these providers to exchange educational certificates. Once the providers are ready, we setup the client to use the provider web-services for exchanging information using DRESS.

1. At Node1, we installed MS-SQL as database management system with a university database schema. We generated mappings using the map-

ping tool and then setup the DRESS based provider web-service.

2. At Node2, we installed MySQL as database management system with another university documents database. We generated mappings for this schema and then setup the DRESS based provider web-service.

3. At Node0, we installed the client GUI which enables us to request information regrading educational certificates from the providers.

It is worth mentioning that the mappings generated at Node1 and Node2 are different and correspond to their respective schema only.

6.3.2 Evaluation Techniques

We discuss the evaluation techniques that can be used for evaluating our proposed standard and the proposed architecture. We also provided references of the research projects where these techniques have already been used as an example. We apply these techniques to evaluate against the goals we discussed in sections 1.3.

6.3.2.1 Evaluating against Functional & Non-Functional Requirements

We listed functional and non-functional requirements in section 3.4 for achieving the goals mentioned in section 1.3. In this technique we subjectively evaluate these requirements. Let us begin with functional requirements first.

1. Starting with first functional requirement, we can express that the proposed architecture and DRESS based effective information trade of educational certificates between partaking institutes in the test-bed environment. This worked successfully for the different heterogeneous schema running on different database engines.

2. The test-bed also satisfied the second requirement and only provider data is only shared if a request is made with valid access keys.

3. We collected real educational certificates and added this record to two different test schema on each node. We see that the system is testable with this data. Hence, we met this third requirement.

4. We added two new providers in the environment in a semi-automated way to generate mappings using the mapping tool and made agreements. This satisfies the fourth requirement.

We now evaluate the non-functional requirements which are relatively hard to measure. A subjective evaluation of the non-functional requirements is below;

1. We exchanged different documents and similar documents from different issuing authorities with a few changes at attribute level. However, it is hard to test with every possible document but we were able to test the standard for documents from different issuing authorities. These document contained differences at attribute level and among values. For example, the grade and mark attributes are used both with divisions plus percentage and grades plus CGPA. This successful exchange satisfies the non-functional requirement number 1.

2. The architecture is designed in such a way that it provides both control and authority. The only authority that may request data from the providers in this architecture is the Node0 (the central authority). This satisfies the non-functional requirement number 2.

3. Both providers Node1 and Node2 in the test-bed are the owner of their data. Providers are autonomous bodies approved from a central author-

ity. Since, the data is not copied to some other authority and a document exchange request is always made to the corresponding provider. For example, University 1 at Node1 will always serve exchange requests for its documents and every time request is made to Node1. This makes University1 owner of its own data and thus builds trust in the system. This satisfies the third requirement.

4. The mapping tool made it easy to integrate new providers into the system in a semi-autmated way. This satisfies the fourth requirement.

5. We used open-source technologies that are widely used for setting up test-bed. We used JSON for storing mappings which require no storage configurations. We used PHP as server-side programming language for the development of the web-service and the query builder. HTML and JavaScript are used for front end development. All the technologies are openly available, free and widely used. This satisfies the fifth requirement.

6.3.2.2   Using Program Validation Technique

Program validation technique unit testing and integration testing of the tools that are developed. This technique reports the issues and bugs found in the developed tools. However, in this research our focus is on our proposed standard DRESS and the proposed architecture. Although we developed software tools to evaluate DRESS and the architecture but the testing of the these tools itself is out of scope of this research.

6.3.2.3   Using Fault-based Testing

We used this evaluating technique to test the XML schema for detecting faults. The faults that will be found are related to the development errors or the evolution of the schema [17]. Starting with the creation of a formal model for the schema under consideration, we create fault classes. These fault classes are used to generate the queries. We run these queries against the test-data. The test data are the sniffed XML documents while exchange between during partaking institutes. The results of the queries are then compared with the specified schema to detect faults. The testing process we followed is shown in figure 6.2. It should be noted that instead of mutating the data, we sniffed the traffic and used the captured XML documents as test-data.
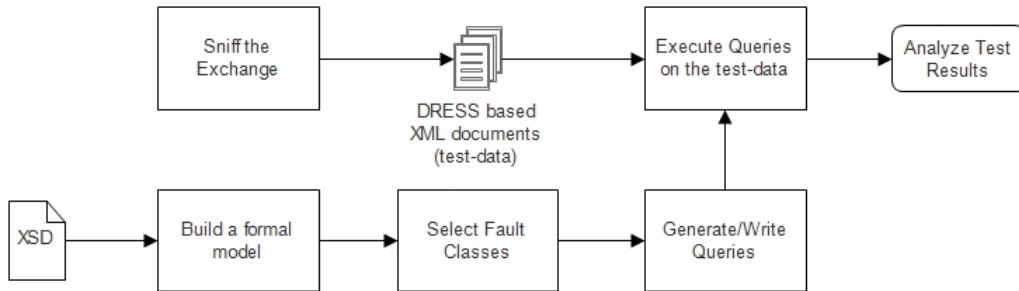

Figure 6.2: Testing process for detecting faults

1. First, we present a formal model to represent the XML documents. A formal model is necessary to formalize the testing of the documents under consideration by executing the queries. We used these following rules to represent the formal model denoted by M.

   M = <E,A,C,P>, where;

   E is set of finite number of elements

   A is set of finite number of attributes

C is set of finite number of constraints

P is set of finite number of associations among E, A and C.

So, the formal model representation of a DEGREE document becomes;

E = {docType, serialNo, regNo, issueDate, ...}

A = {gradingScheme}

C = {type, occurs, pattern, use, ...}

P = { p1(document, docType), ...

2. The second step involves gathering the test-data. Instead of mutating the valid documents as used in [17] [18], we sniffed the documents using Wireshark during the exchange since we need to detect faults in these documents.

3. The third step is the selection of the fault classes. The schemas under consideration are XML schemas. There can be faults specific to XML and can be categorized into three classes;

   (a) class1: Fault class for Elements

   (b) class2: Fault class for Attributes

   (c) class3: Fault class for Constraints

   We associate the elements or attributes of the schemas to the fault classes. These are identified using the elements, attributes and constraints in the formal model.

4. The fourth step involves writing queries based upon the identified fault classes and the associations. These queries are written in XQuery to test the sniffed documents.

5. The fifth step involves executing queries on test-data. In Table 6.1, we execute an XQuery against an element for testing minOccurrence constraint. This falls in class 3 category of the fault classes.

Table 6.1: Xquery on Test data

Test-data	XQuery
<document>	let $doc := document
....	("degree.xml")
<serialNo>012681</serialNo>	for $i in $doc//document
....	let $count := count($i/serialNo)
</document>	return <result>$count </result>

6. At step 6, we analyze the test result. The query returns result 1 and no fault related to minOccurrence is found. However, if the query returned 0, we would have detected a fault.

6.4 Summary

Implementation of DRESS and the supporting tools is completed in multiple stages. We defined the schema documents, developed mapping tool, generated maps, developed web-service and the client. To evaluate DRESS, we set up an environment of three nodes including a requester and two providers. We performed subject evaluation against the goals, the functional requirements and the non-functional requirements. We evaluated the schema by detecting faults in the exchanged documents using a fault-based testing approach.

Chapter 7

Conclusion & Future Work

7.1 Conclusion

Inspired by the standardization of students' information in Europe, we have developed a prototype standard "Document Record Exchange Standard for Students (DRESS)". It is completely focused on exchange of student certificates independently of underlying platform. It currently fulfills vocabulary requirements for three important certificates; REGISTRATION, DEGREE and TRANSCRIPT.

We used DRESS over the proposed architecture to exchange documents. DRESS is evaluated subjectively against the goals, the functional and the non-functional requirements. We used a fault based evaluation for its validation and detecting faults in the proposed schemas.

In DRESS, we reused the ideas that were implemented in already existing standards. It is reusable and conforms to ISO-standards where possible. It covered some vocabulary which was previously missing in existing standard-

s. It focuses on the student educational certificates and defines vocabulary placing the document at the root.

To follow and implement any standard, it must fulfill the requirements as well as it must be easy to use. To ease its implementation we created a semi-automated wizard based mapping tool. This tool enables the providers to easily map their databases to DRESS. These mappings are used by the web-service we created for the providers.

7.2 Future Work

Research in the domain of exchanging educational certificates has lead us to some opportunities for further research in different directions. We go through some of the directions;

7.2.1 Unified Measurements

There exists many measurements differences among educational certificates issued by different issuing authorities. This difference exists in choosing measurement scales and the among scales themselves. A good example related to educational certificates is the use of CGPA or percentage system for evaluating the student results. Class-room based and effort-based measuring units both exist. One credit hour is 10 classroom hours while in European Credit Transfer System (ECTS) the tuition hours have values ranging five to ten. [5]

7.2.2 Improvements in Document Authentication

Although we have introduced a central authority to have control and authority built in to the architecture. However, the exchange document must itself be authenticated and verifiable. The exchange of official and legal documents including educational certificated must be verifiable and authenticated. Some work has already been done in this domain in Australia for exchanging and verifying official documents with the help of signature solution combined with citizen's smart card [14].

Bibliography

[1] M. Ali, H. Qureshi, and M. S. Akhtar (2013), *Analysis of growth in Students Intake and Degree Awarding Contribution: A Comparison of Stanford and MIT*, MLDM 2013: International Conference on Machine Learning and Data Mining, in press.

[2] http://www.ehea.info/Uploads/Irina/Bologna

[3] P. Pouyioutas, H. Gjermundrod, and M. Michael, "MAPQFTOOL: A software tool to support national qualifications frameworks," Information Society (i-Society), 2011 International Conference on, pp. 198-203, Jun. 2011.

[4] Rafal Nagrodzki, "The Mobility Project," Institue of Informatics, University of Warsaw, Warsaw, Master's Thesis 2009.

[5] A. Duran, Y.B. Moon, and E. Giraldo, "Work in progress - the European Higher Education Area ("Bologna process") in Engineering Education in Spain," Frontiers in Education Conference, 2009. FIE '09. 39th IEEE, pp. 1,2, Oct. 2009.

[6] Karol Kanski, "Integration of Services in the Mobility Project," Institute of Informatics, University of Warsaw, Warsaw, Master's Thesis 2011.

[7] Alon Halevy, Anand Rajaraman, and Joann Ordille, "Data integration: the teenage years," In Proceedings of the 32nd international conference on Very large data bases (VLDB '06), pp. 9-16, 2006.

[8] Szentirmai, L.; Radacs, L., "Improvement of academic and research standards of higher engineering education in light of Bologna process," Emerging eLearning Technologies & Applications (ICETA), 2012 IEEE 10th International Conference on , vol., no., pp.381,386, 8-9 Nov. 2012

[9] https://wiki.refeds.org/download/attachments/1606048/SCHAC %2B1.5.0.pdf?version=3&modificationDate=1429195142624&api=v2

[10] http://www.aepam.edu.pk/files/educationstatistics/pakistaneducation statistics2010-11.pdf

[11] http://www.aepam.edu.pk/files/educationstatistics/pakistaneducation statistics2013-14.pdf

[12] http://www.pbs.gov.pk/sites/default/files//tables/POPULATION BY LEVEL OF EDUCATION AND SEX.pdf

[13] https://europass.cedefop.europa.eu/en/home

[14] Neubauer, T.; Weippl, E.; Biffl, S., "Digital signatures with familiar appearance for e-government documents: authentic PDF," Availability, Reliability and Security, 2006. ARES 2006. The First International Conference on , vol., no., pp.8 pp.,, 20-22 April 2006

[15] Maluf, D.A.; McDermott, W.J.; Knight, C.D.; Smith, E.E.; Gurram, M., "Searching Across the International Space Station Databases," Aerospace Conference, 2007 IEEE , vol., no., pp.1,8, 3-10 March 2007

[16] http://hdl.handle.net/1721.1/27091

[17] Emer, M.C.F.P.; Vergilio, S.R.; Jino, M., "A testing approach for XM-L schemas," Computer Software and Applications Conference, 2005. COMPSAC 2005. 29th Annual International , vol.2, no., pp.57,62 Vol. 1, 26-28 July 2005

[18] Emer, M.C.F.P.; Vergilio, S.R.; Jino, M., "A testing approach for XM-L schemas," Computer Software and Applications Conference, 2005. COMPSAC 2005. 29th Annual International , vol.2, no., pp.57,62 Vol. 1, 26-28 July 2005