

BSc Final Year Project Form (2015/2016)

1. Proposal

The student should complete parts 1(a), 1(b) and 1(c) below, and then agree the maximum pocket values with the supervisor and put these in part 2(a) below. An electronic version of this form should be uploaded to the Final Year Project page on Moodle no later than **Monday 2nd November 2015**.

(a) Student details

Name:	Project Type 4 (BUCI027S6)
Email:	

(b) Project details

Title: Mapping XML Schema to Relational Schema - A better way!
Objectives: <ul style="list-style-type: none">• Research, develop and express a deep understanding of the issues relating to the mapping of hierarchical schema (such as XML) into relational schema.• Evaluate solutions already available on the market for their strengths and weaknesses. Propose ways in which they could be improved.• The main object of the project will be to create a utility / code that will perform a 'round - trip' of XML to Relational schema:<ul style="list-style-type: none">○ Create a relational database schema from an XML schema (.xsd) file○ Allow users to populate the relational schema in any way they see fit○ Produce XML from the newly populated relational schema○ Validate the XML against the original schema used to create the relational database

Title: Mapping XML Schema to Relational Schema - A better way!

- **Should time allow, the project will be expanded to develop a utility with a Graphical User Interface (GUI)**
- **Test the solution in a real world environment.**
 - **For example, producing a HESA return for a university**

Description:

Relational databases are still one of the most common ways for organizations to store their data. System interoperability and data sharing between organizations is ever more important. Data is often shared over the internet, and needs to be converted into a system neutral format. XML and JSON are two such formats that are common for the transportation of data over a Wide Area Network.

Data schema are often written in XML, and organizations may be obliged to provide data in this XML format to another organization, such as a statutory body. An example of this is the HESA return that universities are obliged to produce. HESA provide the schema in the form of .XSD files and supporting documentation. This is a complicated schema to replicate in code - a tool to produce the data would be of massive assistance.

However, there are certain difficulties inherent in mapping relational (tabular) data into an XML (hierarchical format). It is tricky to represent a hierarchy as a set of tables, key relationships are often not represented in XML schema and the definition of datatypes in the W3C XML specification is different to the datatypes in the ISO SQL specifications.

This project aims to provide a solution to help resolve some of these common issues.

Title: Mapping XML Schema to Relational Schema - A better way!

Method:

<p>This project is not about re-inventing the wheel, but rather expanding upon the existing research and development in this area. The project will look at utilizing existing tools and aim to provide an easier way for transporting relational data between organizations via XML.</p>

<p>Existing solutions will be heavily tested and evaluated before any work on a solution begins.</p>
--

<p>The project will deliver a middleware solution that will make the mapping of XML into Relational Data easier. This solution may use existing software libraries, command line tools or the tools supplied with major DBMS. Metadata may be stored in order to ‘remember’ the mapping between XML and relational schema.</p>
--

<p>The project will follow some ‘best practice’ guidelines for software development. Any code written be will be saved in Source Control (Git) and unit testing will be used on any code written.</p>

Title: Mapping XML Schema to Relational Schema - A better way!

Work plan:

Stage 1 - Research : October 2015 - December 2015

- **Research and evaluate the many issues around mapping hierarchical data into relational data.**
- **Evaluate commercial attempts and products aimed at solving the problem, such as Altova XML Spy and the tools that come with major RDBMS such as SQL Server, Oracle and DB2.**
- **Evaluate software libraries / components that may be useful for developing a new solution e.g. the Java libraries EclipseLink MOXy and HyperJaxB**

Stage 2 - Proof of Concept / Prototyping : December 2015 - January 2016

- **Based on the research, a prototype will be developed based on the tools that proved most effective / flexible.**
- **Develop a prototype solution (in Java, T-SQL, PL/SQL or similar) to prove that there are better ways of mapping XML to relational data than are available in commercial tools. The prototype will be code only - no Graphical User Interface**

Stage 3 - Evaluate, Develop, Expand : January 2016 - February 2016

- **Evaluate the prototype, what it does well and where it needs improving**
- **Fix any bugs or amend any core functionality**
- **Expand the prototype by adding ‘wishlist’ features such as a graphical interface or the ability for customization**

Stage 4 - Testing and Feedback : February 2016 - April 2016

- **Provide the solution to users who need to map XML - Relational data in their jobs and obtain their feedback.**
- **The aim will be to produce a HESA return by generating a relational schema from the HESA .XSD, populating it with data, and then producing valid XML.**
- **Compare feedback to aims of the project and evaluate how successful the project has been.**
- **Examine what could be improved for the future**

Title: Mapping XML Schema to Relational Schema - A better way!

College equipment required:

The following software will be used in the project:

- **Altova XML Spy**
 - **To evaluate a commercial software approach to the problems**
- **Git**
 - **For source control of any code written**
- **Java 7/8 SDK**
 - **To develop a solution to map XML data into relational data**
- **Netbeans / Eclipse**
 - **For Java development**
- **Microsoft SQL Server / SQL Server Management Studio**
 - **To evaluate the XML capabilities of SQL Server and the .NET framework and test any solutions developed**
- **MySQL / MySQL Workbench**
 - **To evaluate the XML capabilities of MySQL and the .NET framework and test any solutions developed**
- **Oracle Database / Oracle SQL Developer**
 - **To evaluate the XML capabilities of Oracle and the .NET framework and test any solutions developed**
- **Oxygen XML Editor**

The majority of this software is available in the Birkbeck labs. Database access will be required for the college instances of MySQL, Oracle and DB2.

A virtual machine where software not available on the BBK lab computers can be installed (e.g. SQL Developer, Netbeans / Eclipse for Java EE) would be most useful.