JISC Final Report

Project Information			
Project Identifier	*To be completed by JISC*		
Project Title	Course Data Intelligence		
Project Hashtag	None		
Start Date	1st January, 2012	**End Date**	31st March 2013
Lead Institution	University of Bolton		
Project Director			
Project Manager	Stephen Powell		
Contact email	s.j.powell@bolton.ac.uk		
Partner Institutions	None		
Project Web URL	http://coursedataintelligence.wordpress.com/		
Programme Name	Course Data Programme		
Programme Manager	Ruth Drysdale		

Document Information			
Author(s)	Stephen Powell		
Project Role(s)	Project Manager		
Date 28th January 2013		**Filename**	
URL			
Access	This report is for general dissemination		

Document History		
Version	**Date**	**Comments**
1	28/01/13	Feedback received from Jisc (Sandra Winfield, 3rd March 2013)
2	15/03/13	Jisc feedback incorporated

Table of Contents

1 Acknowledgements

Project funded under the Jisc Course Data Programme[1], September 2011 – March 2013. Work primarily undertaken by Dave Hagan, Wilbert Kraan, James Sherborne and Stephen Powell.

2 Project Summary

This project sought to improve the capability of the University of Bolton (UoB) to capture, store, and re-use course related data for marketing, efficiency, and intelligence purposes. As well as using established techniques, the project sought to explore the use of innovative approaches that are more flexible in allowing for the combining of data from different sources; this may be poorly structured data and/or from outside the institution. This is of particular relevance in the current HE context, where there is increased reporting and use of data for different purposes. The XCRI-CAP specification was used for standardising course related information so that is available for marketing and advertising purposes and, where appropriate, aggregation with other providers' data.

Context

The University of Bolton is a relatively small HEI (302 FTE academic staff, 54 research staff, 251 support staff, and 5151 FTE students). The institution was involved in a previous Jisc XCRI-CAP demonstrator project[2] (2007), and through this work had a reasonable starting point in terms of its systems and processes to undertake work that involved course related information. However, little development or maintenance work had been undertaken since the end of the demonstrator project. At the start of this project course related information was located in the course marketing database, the student management system, an online module database (developed in 2003), and Programme Specifications documents in text document formats that were circulated using email.

3 Main Body of Report

3.1 Project Outputs and Outcomes

Output / Outcome Type	Brief Description and URLs (where applicable)
1. Module & Programme Specification Databases	Replacement Module Database https://modules.bolton.ac.uk/ New Programme Specification Databasehttp://programmes.bolton.ac.uk/
2. Deepening understanding of Learning Analytics contributed to the following Jisc Cetis reports	CETIS Analytics Series: Infrastructure and Tools for Analytics http://publications.cetis.ac.uk/2013/535 CETIS Analytics Series: Institutional Readiness for Analytics http://publications.cetis.ac.uk/2012/527
3. Course related information common vocabulary	Module and Programme Specification Documents http://coursedataintelligence.wordpress.com/2012/06/26/wp-2-develop-a-common-vocabulary-programme-specification/
4. XCRI-CAP feed	http://xcri.ictbenchmark.org/xcri/ (needs migrating to COOL URI)
5. Developing Enterprise Architecture (EA) capability supported by Archimate modelling language	Staff attended TOGAF training and used the project as a vehicle to explore the development of Enterprise Architecture and further develop the use of Archimate modelling language. http://coursedataintelligence.wordpress.com/2013/03/03/archimate-case-study/

[1] Jisc funded project of 10K phase 1 and 70K phase 2 to develop better approaches to use and re-use of course data http://www.jisc.ac.uk/whatwedo/programmes/elearning/coursedata.aspx

[2] The demonstrator technical case study can be accessed here http://www.xcri.co.uk/bolton.html - journey

3.2 How did you go about achieving your outputs / outcomes

The previous XCRI-CAP project (footnote 2) contributed to a good level awareness in Internet Services and Technology (IST) and other parts of the university of the importance of course related data. The demonstrator project created a database that could be extended to accommodate future developments, and was mapped into XCRI-CAP specification. The database was developed so that departments could directly enter their own marketing information with appropriate authentication and structures, whilst the Marketing team retained overall control.

The key lesson learned from this early work was that challenges faced through the implementation process were not generally associated with the technology, but rather the pre-conceptions, competing ideas, priorities and working practices of staff. For example, although the database has the facility to enable departments to enter information this is not used, and instead remains a function of marketing. At the start of the project two further databases held course related information: the module database developed in 2003; and the SITS:Vision[3] student management system.

There is an institutional aim to move towards a Service Oriented Architecture (SOA) which fits well with fundamental aims of the Course Data Programme, which is to promote interoperability, reuse and re-purposing of data through standards based approaches. To support this approach, the university used the project to help develop its Enterprise Architecture (EA) capability and used the Archimate modelling language to support analysis of the existing as-is and develop plans for the desired to-be state.

A high-level project aim was to support the development of the SOA approach outlined above through the specific project objectives:

1. building a programme and module specification database to meet the needs of administrators, academics and learners that is XCRI-CAP compatible;
2. speculative work exploring the use of triple stores and RDF technology for combining unstructured data sets for exploring course related matters; and
3. implementing an XCRI-CAP feed for marketing and advertising purposes.

The work to build the programme and module databases were undertaken in an ASP.NET, Model View Controller[4] (MVC) environment that was a new departure for the university. Although requiring a significant amount of up front effort to develop institutional capabilities, there was substantial advantage from the re-use of components between the course and module specification databases. The MVC approach achieves this by separating the development work for data and applications, rather than combining it together. Taking the two databases together, the resource required, compared to previous working practices, was significantly reduced (Director IS&T interview).

The institution is seeking to develop its use of learning analytics by developing both technical and operational capabilities and one focus of this work is student progression and retention. The triplestore and RDF technology using SPARQL query language are seen as possible route for this development, enabling combined sets of course and student related data to be retrieved and manipulated. A critical challenge beyond the technical one is the identification of questions that will yield ‘actionable insights’. For example, academic leads have identified the need for a better understanding of the longitudinal life cycle of students combining information that includes UCAS entry points, assessment outcomes, and progression

[3] http://www.tribalgroup.com/technology/sitsvision/Pages/default.aspx
[4] http://coursedataintelligence.wordpress.com/2012/12/21/wp-1-develop-programme-specification-management-system-programme-database-application-architecture/

patterns[5]. It is hoped that better analysis of data will help the development of understanding of this complex area and gain insights into the student experience so that interventions can be made.

The demonstrator work at the university had encountered time-out issues with the XCRI-CAP feed, because of excessive load, so a different approach was required. To address this issue, it was initially anticipated that a range of different feeds would be produced for different end purposes, e.g. CPD, subject domains, etc. However, the Jisc programme demonstrator services expressed a preference for the simplicity of a single place that is well signposted where all feed data could be collected. The Bolton XCRI-CAP feed is generated using an intermediary staging database[6] for advantages of speed and efficiency. This approach means that the existing course marketing database that is also serving the website is not overloaded.

The project used an action research approach to engage with stakeholders. In practical terms this meant that the developers worked in conjunction with users to identify requirements, using an iterative or agile development methodology. This approach used an ongoing evaluation of work against user needs to produce a result that does what is needed by users who may, at the outset, have a limited understanding of the kind of functionality they require. As an approach, it is more demanding of the development team and does not always fit with their preferred ways of working that in many cases have been specification lead. However, we believe that end result is better applications.

During the lifetime of the project, there were significant structural and staffing changes both within the project team and across the wider university that impacted significantly on the work being undertaken by the project. However, the project work was largely delivered and this could be attributed to the flexible, action research approach taken.

3.3 What did you learn?

The University of Bolton experience over a number of years indicates that there is ongoing resource required to maintain the data structure and integrity. Administrators, academics and technical staff require ongoing education, training and development. Unless the XCRI-CAP feed is actively consumed externally, it would be difficult for institutions to maintain or invest in developing XCRI-CAP further.

From an institutional perspective, the experience working in a MVC environment has proved to be valuable in increasing the capability of the organisation to deliver on software projects as it makes the re-use and re-purposing of code and the addition of new data more straight forward.

Organisational data ownership is a difficult issue to navigate and is one that we have not yet solved. Who can do what with particular sources of data and who can decide what can be done can take a long time to work through. This applies at a business process level as well as a technical one. The establishing of the Management Information Working Group is probably a good step for any institution to consider, if they wish to better coordinate activities around institutional data. At Bolton, the group is comprised of heads of non-academic departments, individuals with relevant knowledge and expertise, and is chaired by the Executive Dean - Market and Corporate Intelligence who is located in the Vice Chancellor's Office. Thus far, matters addressed have arisen from operational issues faced by

[5] http://coursedataintelligence.wordpress.com/2012/10/17/workpackage-5-application-of-analytics-to-university-processes/
[6] An intermediate storage area between sources of information before further processing is carried out

departments. A good example of this are the issues as a result of structural reorganisation of the university which impacts across the organisation: on finance; human resources; teaching departments; course related marketing information; and timetabling, etc. Having a group to think about solving these current issues is important, but developing this group to take a proactive approach to issues such as data policy and governance and procedures will be essential.

3.4 Immediate Impact

Module and programme databases have informed the collection of the KIS information and are now being used as a part of the course validation process and are more readily available to potential applicants to base judgements on about programmes and courses.

XCRI-CAP capacity has been retained at a time of significant staff turnover. Without the project it is possible that much of this intuitional knowledge would have been lost.

Although not an intentional output, and the writing up work was funded elsewhere, the project staff on this Course Data Intelligence project used their experience gained in writing two JIS-CETIS publications that offer a starting point for learning analytics in institutions:

- Institutional Readiness for Analytics[7] - this briefing paper is written for managers and early adopters in further and higher education who are thinking about how they can build capability in their institution to make better use of data that is held on their IT systems about the organisation and provision of the student experience. It will be of interest to institutions developing plans, those charged with the provision of analytical data, and administrators or academics who wish to use data to inform their decision making. The document identifies the capabilities that individuals and institutions need to initiate, execute, and act upon analytical intelligence.
- Infrastructure and Tools for Analytics[8] – this briefing paper provides a map of the major categories of tools, and highlights some landmark tools that are available now for analytics work, grouped by tradition, or established approach.

3.5 Future Impact

The most significant medium term, future impact from this project is that around the use of linked data technologies[9] described in output 2, the CETIS Analytics Series: Infrastructure and Tools for Analytics. We believe that this offers a potentially significant way forward for combining different data sets to enable learning analytics approaches to be developed. This is not just the case at our institution, but there are sector wide opportunities to move toward a better use of data for information to inform decision-making processes. Arguably, this is something that the sector doesn't do well currently.

An illustration of the potential advantage of this approach is by way of contrast to the data warehouse also being constructed at our institution. This is being developed for reporting purposes to managers and to personal tutors to provide information about students at risk. An ongoing issue is the changing nature of the source databases making the data modelling for the warehouse problematic and time consuming. Potentially, the linked data approach removes this particular problem as data can be imported in an unstructured way. This is continuing work at the University of Bolton.

[7] http://publications.cetis.ac.uk/2012/527
[8] http://publications.cetis.ac.uk/2013/535
[9] Approaches that enable pieces of information to be connected together to 'build a bigger picture', for example, combining data about bus routes, schools, shops, post offices, etc. to describe a neighbourhood.

4 Conclusions

There are clear and tangible benefits from having data sources in order: more effective and efficient marketing, better data quality, and improved institutional systems and associated costs savings. XCRI-CAP has an important role to play in this alongside better institutional governance of data and the development of new tools and technology for handling and making the most of complex sets of 'messy' unstructured data.

Institutional support of XCRI-CAP feeds into the future will require some evidence of take-up and value to the institution. Hopefully, the environment has changed sufficiently since the early demonstrator projects (2007) to the extent that there will be sector wide take-up of feeds and they become an essential way of doing business. This may come from offerings from providers such as Graduate Prospects or HotCourses who use the data to provide services to prospective students.

5 Recommendations

JISC needs to continue to lobby to get UCAS to adopt XCRI as the method for informing them about undergraduate degree programmes. If this happens, sustainability in institutions will become a non-issue and adoption will become widespread and instantaneous!

A programme that explores learning analytics would be a useful stimulus to supporting instructions in catching up with commercial sectors.

6 Implications for the future

The work required to establish XCRI-CAP feeds could have a significant impact in starting institutions down a route of better handling and use of data. In its own right, XCRI-CAP potentially brings significant advantages to institutions, but the bigger prize may be how institutions handle data more widely once the ideas are introduced and benefits seen of this work. For example, the potential of data recorded and held by institutions that could be put to use through analytics approaches, to understand the enterprise and provide better intelligence for business decisions. Higher education institutions are far behind the commercial sector in using data for these purposes and those institutions that understand how to best use their data will likely gain a significant competitive advantage from it.

Appendix A: Glossary of Terms

Analytics: is the process of developing actionable insights through problem definition and the application of statistical models and analysis against existing and/or simulated future data.

ASP.NET: server-side Web application framework for dynamic web page development

ASP.NET, MVC: it allows software developers to build a web application as a composition of three roles: Model, View and Controller enabling more effective re-use of components.

Enterprise Architecture: a collection of methods and tools used to support an organisation in designing, planning, implementing and governing technical systems, processes and information.

Linked Data: a method of publishing structured data so that it can be interlinked and become more useful, enabling data from different sources to be connected and queried.

MVC: Model View Controller is an approach that separates the representation of information from the user's interaction with it enabling more effective and flexible use and res-use of data and applications.

RDF: data model based upon the idea of making statements about resources in the form of subject-predicate-object expressions, e.g. a subject denoting "the sky", a predicate denoting "has the colour", and an object denoting "blue" (http://en.wikipedia.org/wiki/Resource_Description_Framework)

Service Oriented Architecture: a set of principles and methodologies for designing and developing software in the form of interoperable services. These services have well-defined business functionalities that are built as software components (discrete pieces of code and/or data structures) that can be reused for different purposes (http://en.wikipedia.org/wiki/Service-oriented_architecture)

Triplestore: type of database for the storage and retrieval of triples

SPARQL: Query language for databases using resource description framework format.