JISC DEVELOPMENT PROGRAMMES

Cover Sheet

OCCAM FINAL REPORT

Project Acronym	OCCAM	**Project ID**	
Project Title	**Open Course Collection and Aggregation Model**		
Start Date	**1 Mar 2007**	**End Date**	**30 Sep 2007**
Lead Institution	**The Open University**		
Project Director	**Viki Stirling**		
Project Manager & contact details	Alan Paull APS Ltd, 58 Norton Wood, Forest Green, Nailsworth, STROUD GL6 0HG 01453 835009 alan@alanpaull.co.uk		
Partner Institutions	APS Ltd		
Project Web URL			
Programme Name (and number)	*e-Learning*		
Programme Manager	Sarah Davies		

Document

Document Title	*Project Plan*		
Author(s) & project role	Alan Paull, Project Manager		
Date	14 November 07	**Filename**	OCCAM-Final-Report.pdf
Access	√ Project and JISC internal		√ General dissemination

Document History

Version	Date	Comments
1.1	14 November 07	Final report

Final Report for OCCAM

Project Outputs

1. The aim of the project was to extend, enhance and standardise the existing methods through which the Open University supplies its courses data to third party aggregators, in order to provide the wider community with a model of good practice using the XCRI-CAP schema.
2. Requirements and deliverables were:
   - A web service giving access to the whole live course catalogue of the OU in XCRI-CAP format.
   - Evaluation of the new system, in terms of cost, quality of provision, flexibility in relation to aggregators and extensibility.
   - Organisational and technical issues, including a 'forward plan' for future development with the aggregators.
   - Evaluation of the success of the XCRI-CAP schema and its usability.
   - Recommendations for future JISC activity.
3. The project addressed the topics through 5 work packages:

   WP1, Process Review: Analyse existing systems and business processes to confirm actions for development.

   WP2, Mapping: Map XCRI-CAP to OU course data structures.

   WP3, Data Publishing: Design and implement the web service; convert outputs to bespoke formats for aggregators where necessary.

   WP4, Testing: System, user and live testing of the new system to enable comparisons with the old.

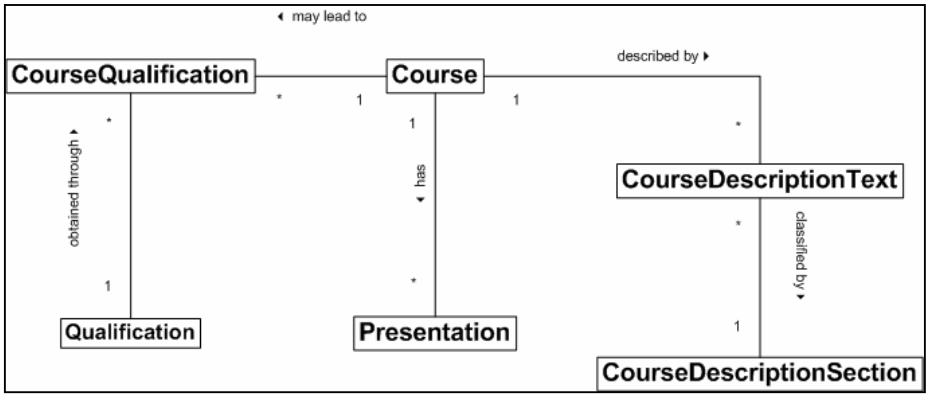
   WP5, Evaluation and reporting: New system evaluated against the old, communication of lessons learned to the XCRI support project, recommendations for future JISC activity.
4. The project outputs are detailed below.

WP1: Process Review

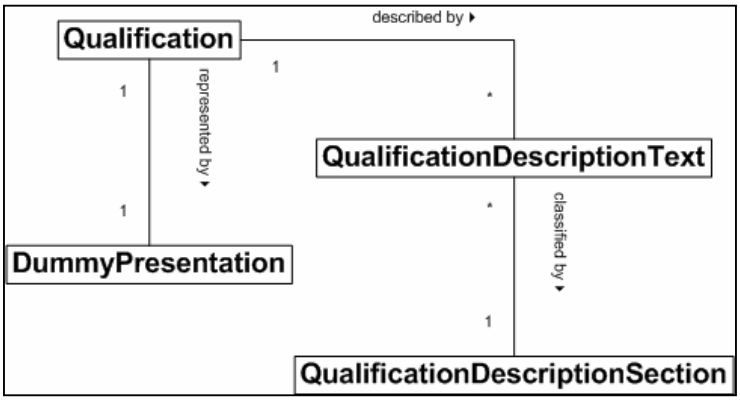
5. Existing processes were modelled using Unified Modelling Language (UML) diagrams, attached at Appendix A. The purpose of the modelling exercise was to aid the project team's understanding of current processes, so that we could confirm intervention points and actions for development. While we intended to carry out a comprehensive business process view, we found that it was sufficient to create only the Process Behaviour and Requirements Views. These give a coherent overview of the current requirements and information flows in relation to courses information delivered to data collecting organisations.

WP2: Mapping

6. The structure of the OU data is depicted in this UML class diagram:



7. While courses and presentations are handled in this way on extraction for data collectors, qualifications are represented as if they are courses, as follows:



8. We have mapped catalogue -> provider -> course -> presentation accurately. Qualification causes more of a problem, because the nature of OU courses is that they equate to modules at other HEIs, rather than to degree programmes. An OU qualification is built up from a number of courses, and a single course would normally lead to a choice of several different qualifications, some of which could be at different levels of study. In contrast, a module at another HEI would normally form part of a single degree programme leading to only one qualification; or at least it would be structurally managed in that way, even if it could theoretically lead to several qualifications.

9. This leads us to model qualification as a course and to have a dummy presentation for each qualification. If this were not the case, then OU degree programmes would not be shown in a way compatible with degree programmes in other universities, which very often equate course with qualification, for example a conventional 3-year History course leading to a BA (Hons) in History would be described as a course with title "BA Honours in History".

The mapping process

10. The OU data was mapped to XCRI-CAP as given in the Mapping Table at Appendix B. The mapping was created using a combination of the Altova MapForce graphical tool and an Excel spreadsheet. MapForce permits 'drag-and-drop' mapping both from and to XML instance documents that meet a DTD or schema, together with XPATH functionality for transformation of nodes of the source data. Source and target files can be XML files, databases or comma-separated value files. In addition to a graphical view of the mapping and transformed data, MapForce can produce XSLT, Java and other programming language outputs, so that transformations can be run independently of the Altova product.

11. The mapping was created using many trial iterations. Output XML files were validated periodically against the XCRI-CAP schema, in order to check the accuracy of the mapping. Altova XMLSpy was the chosen tool for validation. Where particularly complex transformations were required, the mapping was broken into stages, each having a separate MapForce mapping. This process meant that the mapping was clearer visually, and PC performance limits were avoided. An example of the graphical representation from MapForce is shown on page 6.

12. The use of MapForce and XMLSpy in conjunction with a readily understandable Excel spreadsheet has proven of great value. Although not a production process, these tools contribute to an understanding of the detailed transformations required, and they also permit the creation and checking of real outputs alongside the mapping process, rather than separating the mapping from the transformation software. The Excel spreadsheet helped to explain the details of the mapping and transformations to people not intimately involved in the technical work.

13. The XCRI-CAP output files were submitted to the XCRI aggregator website for further validation and inclusion in the test data feeds and helped to test the aggregator. This prompted further comments about XCRI-CAP 1.0, which are indicated below.

WP3: Data Publishing

14. OCCAM's data publishing outputs were divided into two main sections:

    - Design and implementation of an XCRI-CAP web service;
    - Creation of conversion routines for third party data collecting organisations.

15. A web service to enable the harvesting of the data was designed, and then developed and implemented using the .Net framework. The service definition and web service code are available in the project files. The web service can be viewed at: http://www3.open.ac.uk/occam/courseinfo.asmx.

16. The web service was initially trialled using test data and subsequently implemented using the full OU course catalogue for August 2007 in XCRI-CAP format.

17. The project envisaged that after a test phase using a sample data file prepared by APS from the original P12 database, the web service would be populated by data from a new back office system developed via a major project within the OU. However, it did not prove feasible to build the new back office system within the time scales of OCCAM, so the project reverted to a fall back position, whereby the full data set was provided from existing systems. APS used the P12 database and a set of conversion routines developed in Altova MapForce to create XCRI-CAP data.

18. Having built an initial web service, it can in future use XCRI-CAP derived from any new system developed within the Open University. The project envisages that the web service will be developed at a later date to provide both the whole live course catalogue and filtered sets, for

example for individual Faculties or for postgraduate courses. Recommendations are going forward from the project to managers at the OU.

19. A program in PHP was developed to harvest the data from the web service. It is available in the program files.

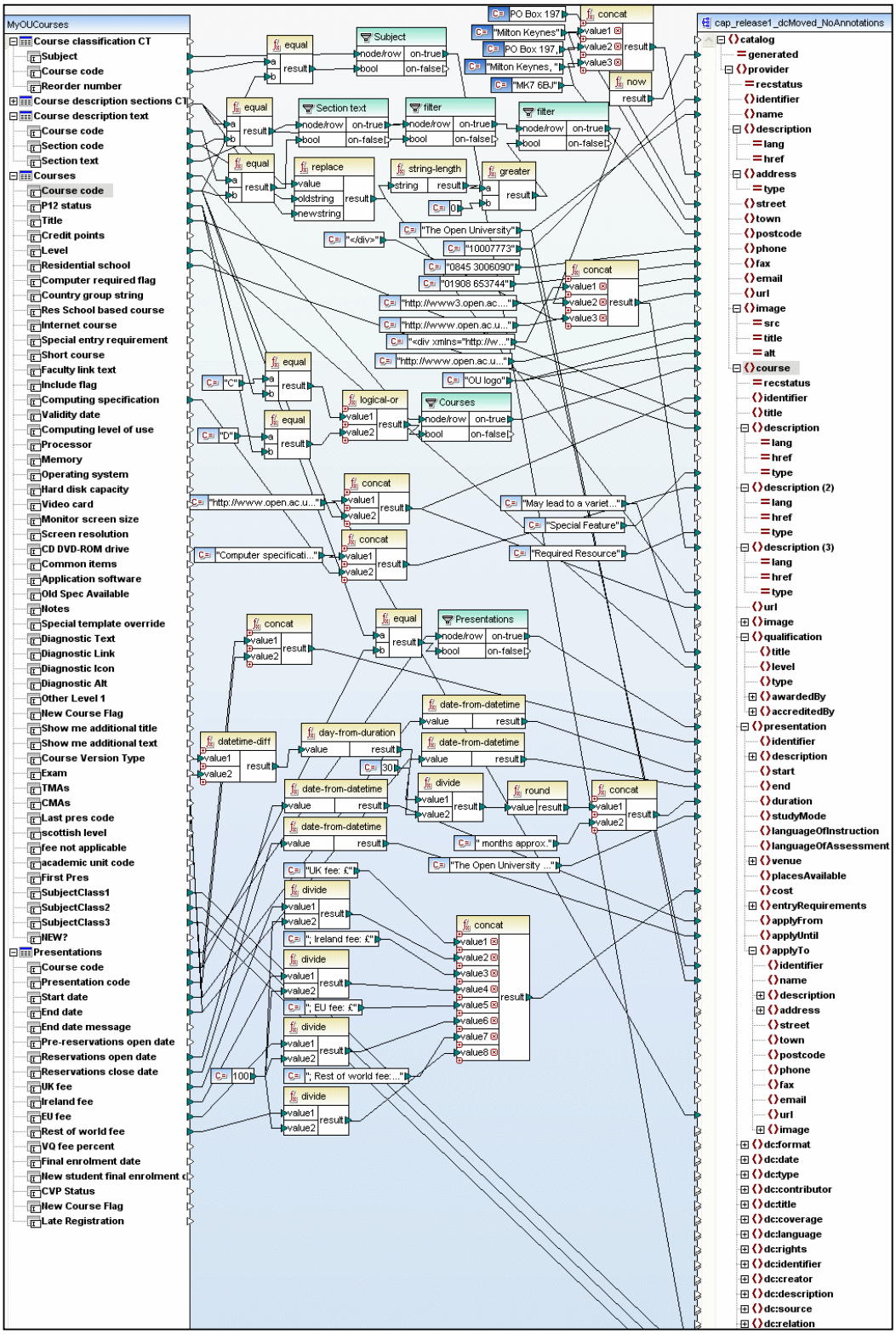
Performance

20. The file size of the full course catalogue was nearly 5 Mb, which presented a potential performance problem. Recognising that the main usage might be an infrequent but regular harvesting of the whole data set and that this is a proof of concept at this stage, the project decided that publishing as a single XML data file was sufficient. If there were additional requirements for searching or filtering, particularly if it was to be the source for a live data aggregation feed, then it would be essential to look at database storage or even storage in RAM. These issues were beyond the project's brief.

21. The PHP program used to harvest the data from the web service was amended to enable it to handle the large file size.

22. The project would suggest that performance issues, particularly when using large XML files, should be carefully considered in the implementation of any XCRI-CAP system.

MAPPING OF OU DATA TO XCRI-CAP – ALTOVA MAPFORCE



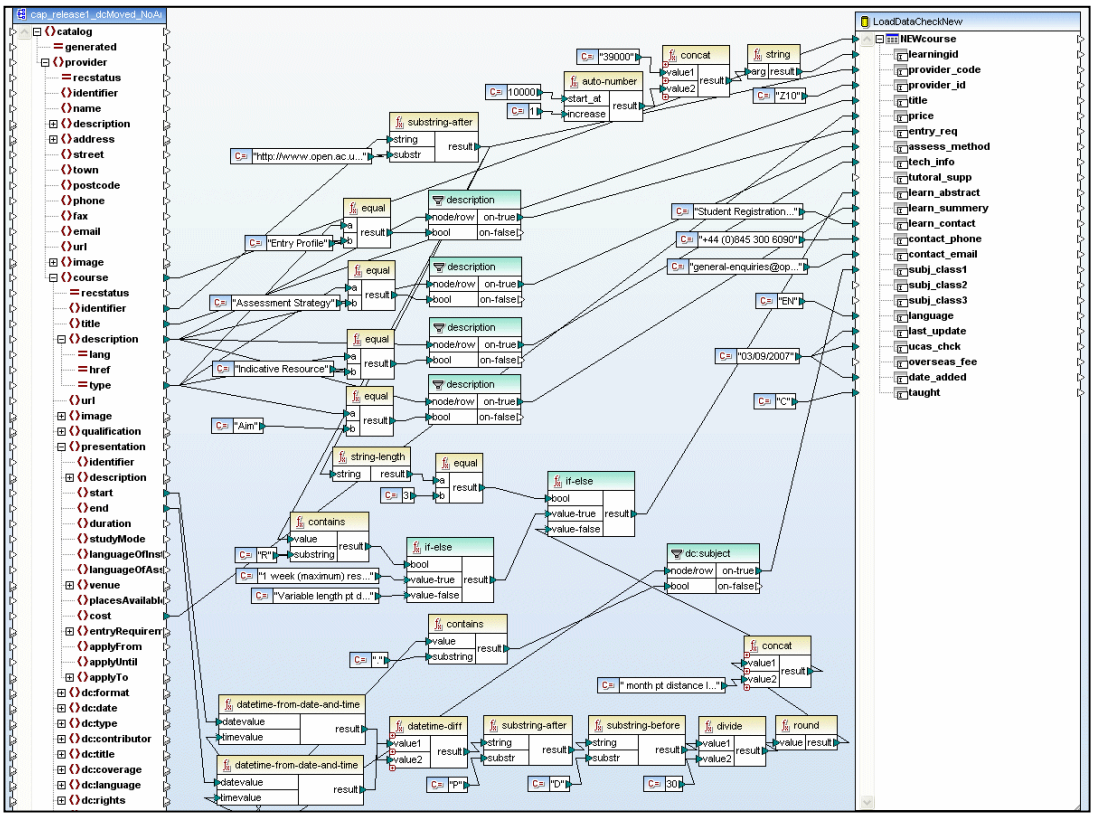
Conversion routines

23. Many data collecting organisations are unlikely to be able to consume XCRI-CAP data without further work to transform it into their own specific data formats. Under the current system operated by the OU via the APS agency, transformations from the OU format into an appropriate format for consumption by learndirect and others is carried out in two steps: step one transforms the data into a base XML format, step two transforms from that into the data collector's format – for learndirect (via UCAS and Hotcourses) a series of pipe-delimited files, which are imported into a UCAS system for onward transmission. The OCCAM replacement system uses XCRI-CAP in place of the original base XML format. A further transformation from XCRI-CAP is needed to ensure that data is usable by the data collecting organisation. APS built a transformation from XCRI-CAP into learndirect format, in order to satisfy this requirement. Altova MapForce was used to create the mapping and the outputs.

24. A graphical representation of the first stage of the mapping is shown on the next page. It shows the mechanism for filtering course description elements dependent on the typed vocabulary values ("Aim", "Entry Profile", "Assessment Strategy", "Indicative Resource"). For this mapping, many values have been inserted as constants (indicated as "C" in the picture); these values would be picked up from an identified permanent source in a fully operational system, for example an OU database or, in some cases, an external web service, such as one of the UKRLP web services.

25. Java output for these transformations has been provided in the project files.

MAPPING FROM XCRI-CAP TO LEARNDIRECT FORMAT – ALTOVA MAPFORCE



WP4: Evaluation and reporting

26. Evaluation of the new system versus the old was carried out primarily by APS. As the back office system at the OU was not the responsibility of the project and had not yet been fully implemented, no attempt was made to evaluate the current back office system involving the P12 database with the proposed new CMS.

27. Two aspects of the systems were examined:

- Efficiency of the processing – addressing cost and time taken;
- Effectiveness – addressing the ability to produce appropriate outputs, quality of provision, flexibility and extensibility.

Efficiency of the processing

28. The existing APS system for transforming the OU data is based on a non-standard XML schema and a series of specially written VB and Perl programs, which are tightly coupled to the P12 database at the OU. Changes to the P12 database and changes in specifications from data collecting organisations have led to extensive adjustments to these programs. Using the XCRI-CAP format, especially when standardised, will greatly reduce the need for ad hoc changes at this initial transformation stage.

29. The maintenance of the current suite of VB and Perl programs becomes unnecessary under the new system, which will be built on a robust transformation into XCRI-CAP as a standard output from the OU's new back office system. Individual transformations for data collecting organisations can be made via the MapForce graphical tool and can be upgraded in the future to fully automated procedures managed by the data collecting organisations. APS is committed to

taking this approach and is working closely with UCAS, Hotcourses and Scottish Ufi to develop these solutions, building on the outputs of the project.

30. In addition changes to the OU's back office system can be made without jeopardising the outward facing processes, as long as any revised back office system can continue to produce XCRI-CAP. The outward facing processes, currently operated by APS, would be only loosely coupled to the XCRI-CAP production systems. Therefore maintenance of the OU processes and APS processes can be carried out without undue impact on each other.

31. The efficiency of the new system appears greater than that of the existing one, for the following reasons:

- It uses the XCRI-CAP standardised information model;
- It has much less reliance on proprietary software;
- It de-couples the stages of the processing;
- It is extensible.

32. It has a slight disadvantage in that full implementation will require more development work both at the OU and for data collecting organisations.

Effectiveness - ability to produce appropriate outputs

33. Effectiveness was evaluated by comparing outputs for learndirect from the new system with those from the old. A regular scheduled update of OU data for August 2007 was compared through both methods (see Appendix C).

34. Three main difficulties with the new system were encountered:

- Some formatting was not entirely correct, in particular that for qualification titles, which is currently held in a bespoke format (re-keyed); it is likely that this will have to continue, because an automatic transformation would be too complex.
- Presentations are created for learndirect, such that each start record contains all start dates for one year of entry; this proprietary format has not been replicated in the transformation from XCRI-CAP to learndirect format. It is proposed that APS discusses this issue with the data collectors.
- Study mode details are held in a coded format for learndirect. In the old system these were derived by analysing a number of textual description fields for appropriate key words. This complex process has not yet been replicated in the transformation from XCRI-CAP to learndirect format, but will be prior to a full implementation.

35. Although the above issues remain outstanding, none are significant stumbling blocks to full implementation, and the new system does provide appropriate outputs.

36. Outputs for other data collecting organisations have not been finalised, owing to time constraints. Outputs for learndirect were selected, because these are the most complex and the most vital for the OU.

Findings

37. Findings are divided into 4 sections:

    - Technical;
    - Organisational;
    - Evaluation and usability of XCRI-CAP;
    - Summary of Recommendations.

Technical

Security

38. For the OCCAM Project we envisaged that the group of users would be strictly controlled, that the pilot version would not require extensive security, and that security would be re-examined on full implementation. For the project, accesses to the web service were logged, but no security system was enacted.

39. For XCRI-CAP web service implementations, each provider will need to decide on an appropriate security model, so that access to the data is available to authorised users only.

Organisational

Engagement of technical staff

40. It proved time consuming to engage the relevant technical staff for the small scale development required by OCCAM, because the staff were involved in larger, more mission critical work and were reluctant to commit enthusiastically to other work.

Project approval

41. Approval for this small externally funded project was time consuming and required a lot of administration out of proportion to the income, even though potentially the project outcomes would be very important. The project had good support from senior management within the University, and the OU might seek a different funding model for this scale of project in the future.

Implications for data collection responsibilities

42. There is an issue regarding the responsibilities of provider, data collector and any agent; where are the boundaries of these responsibilities? For example, at the present time some data collectors insist or encourage providers to key to specific data entry or formatting standards. With an XCRI-CAP web service, the provider could point the data collector to the data, but it might not meet the data collector's data structures, formatting or style requirements. The nature of the relationship between the provider and the data collecting organisation changes. For the OCCAM Project we replicated the current system, so that APS continued to act as an agent both for the OU and for the data collecting organisations. With XCRI-CAP it would be possible for a provider to supply the data in XCRI-CAP format and place the onus on the data collecting organisation to carry out any other transformations.

43. Data collecting organisations currently operate by harvesting via paper forms, paper prospectus mailings, by spreadsheets or increasingly via electronic forms on web pages. For all these methods little or no technical expertise is required of providers and comparatively little is required of the data collecting organisations. The data can be input using tried and trusted technologies, not yet excluding pen and ink. Extension to using fully electronic automated methods of bulk updating requires much greater technical input from both the provider and the data collecting organisation, both of which need to have an understanding of the technologies involved. Therefore there will need to be a shift of culture within the organisations, towards interoperability and efficient information management and away from less efficient low technology systems that duplicate effort.

Evaluation and usability of XCRI-CAP

44. The project's work demonstrates that XCRI-CAP can be a suitable vehicle for an intermediate base format between the provider's and the data collecting organisation's internal storage of course advertising information. Although there are some significant issues with version 1.0, rehearsed below, the project shows the efficacy of the information model at 'proof of concept' level. OCCAM has used the XCRI-CAP model to facilitate transformation of the OU's data into a format suitable for consumption by data collecting organisations.

45. The project team identified a range of issues, recorded in its issue log. Some of these issues relate to internal OU concerns and are not relevant to a wider audience. Others relate directly to XCRI-CAP, and we list them here. Many have already been referred to the XCRI Support Project and are recorded on the published XCRI Issue Log.

46. Significant design issues regarding the contextualisation of data and credit accumulation and transfer are discussed below. In addition a major issue regarding the responsibilities of the provider and the data collecting organisation was identified and is discussed in the 'organisational' section above.

TABLE OF XCRI-CAP ISSUES FROM THE OCCAM PROJECT

ID	*Title of Issue*	*Description*
OCCAM/I006	Presentations with no specific start date	How do we deal with presentations with no specific start date, e.g. CPD courses which say 'start at any time'?
OCCAM/I007	Summer school	How do you use XCRI-CAP to represent a summer school (with dates) within a course which has much longer dates?
OCCAM/I008	Courses which are just a summer school	How do we represent courses which are just a summer school? Details of residential courses are not held in P12.
OCCAM/I012	Security Model: who uses it and how?	We envisaged that the group of users would be strictly controlled, that the pilot version would not require extensive security, and that the full version should be secure.
OCCAM/I014	identifier	No attribute to state what the identifier is.
OCCAM/I015	provider.name	Type of name is not prescribed; it could be official name or trading name.
OCCAM/I016	provider.addressGroup.street	Mandatory. OU has a PO Box not street. Clash is between address as a postal location and a geographical location. Especially as OU students do not have campus-based tuition.
OCCAM/I017	MIAP policy on attributes	Attributes shall not be defined globally, but defined locally to the element where they are used, using a globally defined simple type.
OCCAM/I018	MIAP IDs in Schemas	The id attribute shall be used to show the CDD identifier. Where an element is defined globally, the id shall be used in the element definition. Where the identifier applies to an attribute, the id may be used in the definition of the simple data date used by the attribute definition.
OCCAM/I019	Address format issue	MIAP is no help. UKRLP format is not a recommended MIAP format. We should investigate what bits are mandatory.
OCCAM/I020	Validation of OU vocabularies	Do we want to do this? Could be via an imported OU schema.
OCCAM/I021	course.description@type vocabulary	Some vocabs should be recommended but external and imported?
OCCAM/I022	course.description	Is mandatory in CAP but annotation says optional; should be optional.
OCCAM/I023	course.xs:any	Is mandatory in latest CAP but annotation says optional; should be optional.
OCCAM/I024	faculty / school in CAP?	Where?
OCCAM/I025	course.qualification	Is mandatory, perhaps shouldn't be? Reference qualifications as courses records
OCCAM/I026	qualification.level	Should be on a course as well as qualification.
OCCAM/I027	qualification title	Qualification titles are changed around to give the subject first and the qualification type afterwards when they are provided to aggregators as 'course' titles. When will this be done in the new system and by whom?
OCCAM/I028	href in source data	hrefs are all relative not absolute values. These will not resolve properly when collected.
OCCAM/I029	Courses offered every 2 years	Some courses are offered every 2 years, so are 'live' courses, but with no 'live' presentations.

OCCAM/I030	Email	Email element is in the information model as a conventional email address, but the OU implements a webmail URL
OCCAM/I031	Description elements are formatted by XHTML	It would be unusual for a provider to store the information as XHTML; and if provided as XHTML, the data collector must be able to handle it, or strip it out sensibly.
OCCAM/I032	CourseID to contain provider ID?	The recommendation for a CourseID in XCRI-CAP is to use a persistent identifier, preferably a URI. Is it expected or desirable that this identifier should have a provider ID with it, for example "http://www.open.ac.uk/courses/12345"?
OCCAM/I033	Responsibilities between provider and data collector	There is an issue regarding the responsibilities of provider, data collector and any agent; where are the boundaries of these responsibilities? For example, at the present time some data collectors insist or encourage providers to key to specific data entry or formatting standards. With an XCRI-CAP web service, the provider could point the data collector to the data, but it might not meet the data collector's formatting or style requirements. The nature of the relationship changes.
OCCAM/I034	Future courses	Filtering by date, for example to obtain 'future courses' for paper publications to be published some time in the future. This suggests a requirement for visibleFrom and visibleUntil dates.

Summary of issues

47. While many of the issues listed above are points of detail, the major observations can be summarised as follows:

.1 Contextualisation of data: Should information be presented in a generic descriptive format for immediate readability by end users, or broken down into component parts and presented in a form suitable for re-use by computer systems (see below).

.2 Shift of responsibilities between provider and data collecting organisation: This observation has been included in the organisational findings. It is worth noting here that data collecting organisations will have to make arrangement to handle XHTML within their data feeds; many have not yet developed this capability (see above).

.3 Credit accumulation and transfer: A recommendation for a revision of XCRI-CAP in relation to credit accumulation and transfer is made above. This is a significant issue in relation to progression and linking of course components (see below).

.4 Address format: There is no generally accepted standard format for addresses; it will be important for XCRI-CAP to be able to support emerging standards for both geo-location, which may be used by data collection organisations for searching, and for postal communication.

.5 Link to wider initiatives, for example MIAP: Communication with MIAP and other initiatives will help to ensure that XCRI-CAP can be integrated with other information systems.

.6 Vocabularies: The project decided that it would not publish OU vocabularies or validate against them within the XCRI-CAP process. These vocabularies are generally proprietary and would not be used by external agencies, so the team preferred to publish the information in a human-readable form. The use of vocabularies in conjunction with XCRI-CAP was regarded as vital for transformation purposes, and the project recommends that vocabularies are kept separate from XCRI-CAP, but are referenced from it.

.7 Identifiers: The use of identifiers is likely to be a key issue for the wide adoption of XCRI-CAP. It will be important to be able to label an identifier, so that its context is recorded. In addition the use of persistent identifiers (for example persistent URIs) is not yet wide-spread; specific action to encourage this should be considered.

.8 Supported open and flexible modes of learning versus 'traditional' face-to-face teaching: A major outcome of the project is to suggest the reduction in the number of mandatory elements in the model. Many elements have been indicated as mandatory in version 1.0, for example qualification, start date, venue address, that are not relevant for the types of course that are offered by the OU.

Contextualisation of the data: 'cost' element

48. The project team's work on the OU data was specifically aimed at presenting the data for consumption by other data processing systems. Under the current system there is a lot of pre-processing, so that much of the data is transformed into contextualised information. However,

there may be a fundamental difference between presenting data for re-use and presenting information directly for consumption by the human reader, for example using the information directly in the OU website.

49. For the human reader, data is extracted from a data source and presented within a known context, for example within the OU website. The data is transformed into information via the context provided by the framework of the website, for example a structured series of headings within a web page with a title, as a result of a subject search.

50. However, collection of the data for utilisation in other data processing systems means that the context of the information is less clear. If the OU data is provided as if it is "information in context", but used for a different purpose, then there is a danger that the context will be changed or lost, leading to confusion when the end user eventually reads it.

51. One approach would be to provide well-defined data without contextual refinement in the element, for example provide the numeric data in the UK Fee field (e.g. "51300"), but no formatting in the text. The implication is that the data should be decomposed as much as possible and presented in the service layer in the finest possible level of granularity and with the formatting separated out, so that its context and format can be re-built from the elements and attributes by the data collector or the information system in which it is presented. A snippet of this approach would be:

Approach 1

```
<cost type="UK fee" currency="pounds sterling">513</cost>
<cost type="EU fee" currency="euro">307</cost>
```

52. The second approach would be to provide the data, context and formatting in the element, or in other words to provide the full information in the text. So for the Cost example above, you might provide

Approach 2

```
<cost>UK fee: £513; EU fee: €307</cost>
```

53. With Approach 1 we have a rigorous definition for XML data for output from the OU. The data could then be transferred to another organisation and re-used, as long as they could handle the rigorous definition. Approach 2 has a more relaxed definition, but on transfer to another organisation, it could only be used as free text; for example it could easily be published on a website under a heading "Cost", but it could not sensibly be used under a specific heading "UK Fee".

54. Should XCRI-CAP be confined specifically to contextualised information for direct publication to users?

Credit Accumulation and Transfer in XCRI-CAP

55. It is suggested that the course element be amended to include credit as an element. This element would provide details about the quantified value of the learning awarded in recognition of the verified achievement of designated learning outcomes at a specified level[1].

56. The credit element would contain two mandatory items:

- creditValue, the number of general credit points associated with successful completion of the course. This is the total number of general credit points for the course, which will not necessarily equate to the number of specific credit points that a learner would earn, if the value is used to satisfy the entry requirements of another course. Specific credit cannot be specified here, because it depends on the nature of the course into which the learner is moving.
- creditLevel, the level of the credit points indicated in creditValue, typically a statement of where the credit points sit within the credit scheme's hierarchy or framework.

and two optional items:

[1] From SEEC website; http://www.seec.org.uk/

- creditScheme, the name of the credit scheme to which the creditValue and creditLevel apply, typically a national credit framework;
- description, any additional material about the credit associated with the course; specified in the same way as other description elements in XCRI-CAP.

57. The credit element is designed to handle information specifically about the quantified value of the learning. The use of the term 'level' should not be confused with the level of any qualification associated with the course, which is conceptually different from credit level.

AN EXAMPLE OF CREDIT DATA TYPE IMPLEMENTATION IN XCRI-CAP XSD SCHEMA

```
<xs:complexType name="creditDType">
        <xs:annotation>
                <xs:documentation>

                creditDType is a complex type that provides details of
                the quantified value of the learning awarded in
                recognition of the verified achievement of designated
                learning outcomes at a specified level. It supports the
                guidelines recommended by CATS bodies in England,
                Wales and Northern Ireland through the Higher
                Education Qualifications Framework.

                It uses the standard description element.

                </xs:documentation>
        </xs:annotation>
        <xs:sequence>
                <xs:element name="creditValue" type="xs:integer"/>
                <xs:element name="creditLevel" type="xs:string"/>
                <xs:element name="creditScheme" type="xs:string" minOccurs="0"/>
                <xs:element name="description" type="courseDescriptionDType" minOccurs="0" maxOccurs="unbounded"/>
        </xs:sequence>
</xs:complexType>
```

An XML instance of the above credit element from OU data

```
<course> …
        <credit>
                <creditValue>60</creditValue>
                <creditLevel>HE level 2</creditLevel>
                <creditScheme>The Open University credit scheme</creditScheme>
                <description><div xmlns="http://www.w3.org/1999/xhtml"><p>You need:</p>
                        <ul>
                                <li>60 points for a certificate,</li>
                                <li>120 or 130 points for a diploma,</li>
                                <li>240 points for a foundation degree,</li>
                                <li>300 points for a degree without honours,</li>
                                <li>360 points for a degree with honours.</li>
                        </ul>
                        <p>Our certificates and diplomas are a way of marking your achievement on the way
to gaining a degree. Most students study part-time and do 30 or 60 points a year. If you are able to study full-
time, the maximum points you may study in a single year is 120. In your first year we advise you to take no more
than 60.</p></div>
                </description>
                <description><div xmlns="http://www.w3.org/1999/xhtml"><p>AA305 is a specified course in
our</p>
                        <ul>
                                <li>BA (Hons) Humanities</li>
                                <li>BA (Hons) History</li>
                                <li>BA (Hons) Literature</li>
                                <li>BA (Hons) or BSc (Hons) European Studies</li>
                                <li>BA (Hons) Language Studies</li>
                                <li>BA (Hons) Modern Language Studies</li>
                        </ul>
```

It can also count towards most of our other degrees at bachelors level, where it can help to weight your degree towards a BA. We advise you to refer to the relevant award descriptions for information on the circumstances in which the course can count towards these qualifications because from time to time the structure and requirements of a qualification may change.
</description>
</credit>
… </course>

Summary of findings and outcomes

58. The findings and outcomes of the OCCAM Project are summarised below.

.1 Mapping of OU course advertising data to XCRI-CAP was successful and led to some recommendations for change.
.2 Altova MapForce, XMLSpy and Excel spreadsheets were good tools for the mapping process.
.3 A web service and a means of harvesting the data have been created giving access to the OU course catalogue in XCRI-CAP format. The data was also submitted successfully to the XCRI aggregator.
.4 A conversion routine from XCRI-CAP into learndirect format was developed and tested against the current live system.
.5 Evaluation of the new system showed that it was more efficient and effective than the old one.
.6 The project demonstrated that XCRI-CAP can be a suitable vehicle for an intermediate base format between the provider's and the data collecting organisation's internal storage of course advertising information.
.7 Significant design issues for further discussion included the contextualisation of data, the inclusion of a structure for credit accumulation and transfer, links to wider initiatives, the use of vocabularies, the need to encourage the use of permanent identifiers and the reduction in the number of mandatory data elements to take account of supported open and flexible modes of learning.
.8 An important organisational issue was the change in the responsibilities of the provider and the data collecting organisation.
.9 Matters of detail in respect of the content of XCRI-CAP were also listed in the OCCAM Issue Log.
.10 Performance issues should be carefully considered in the implementation of any XCRI-CAP system.

Lessons Learned

Aims and Objectives

59. The project achieved it aim to "extend, enhance and standardise the existing methods through which the Open University supplies its courses data to third party aggregators, in order to provide the wider community with a model of good practice using the XCRI Course Advertising Profile (CAP) schema."

60. It also achieved its objectives:

- "The project will contribute to the understanding and usage of XCRI in a live situation, so will increase the knowledge and skills about the emerging standard. Contributions will be in the form of written documents, diagrammatic models and open source service models and specifications to be published by the end of the project.
- The project will produce a model of good practice that is transferable to other HEIs. It is expected that this model of a web service coupled with a validating aggregator, will assist in the take up by other aggregators of the consumption of web services for course advertising materials.
- The project will test the use of the XCRI-CAP schema in a live situation between one university and a range of aggregators."

Overall Approach

61. It would have been preferable to have obtained earlier buy-in and resource commitment from technical staff with more formal statements of deliverables and precise details of time slots available with individual developers as early as possible. Owing to conflicting commitments and a range of tasks with higher priority, it meant that we were sometimes not able to gain access to staff at a convenient time.

62. With more resources from the OCCAM Project, we could have spent more time and effort briefing staff at the OU, with the objective of ensuring a more integrated approach with the wider CMS project. This might have been operationally and politically difficult, but might have been helpful in the transmission of our aims and objectives to the wider community.

63. A useful lesson here is to spend as much time as possible on achieving commitment of technical resources as early as possible.

Project Outcomes

64. Key project outcomes:

    .1 Mapping of OU course advertising data to XCRI-CAP was successful and led to some recommendations for change.
    .2 Altova MapForce, XMLSpy and Excel spreadsheets were good tools for the mapping process.
    .3 A web service and a means of harvesting the data have been created giving access to the OU course catalogue in XCRI-CAP format. The data was also submitted successfully to the XCRI aggregator.
    .4 A conversion routine from XCRI-CAP into learndirect format was developed and tested against the current live system.
    .5 Evaluation of the new system showed that it was more efficient and effective than the old one.
    .6 The project demonstrated that XCRI-CAP can be a suitable vehicle for an intermediate base format between the provider's and the data collecting organisation's internal storage of course advertising information.
    .7 Significant design issues for further discussion included the contextualisation of data, the inclusion of a structure for credit accumulation and transfer, links to wider initiatives, the use of vocabularies, the need to encourage the use of permanent identifiers and the reduction in the number of mandatory data elements to take account of supported open and flexible modes of learning.
    .8 A further implementation issue was the change in the responsibilities of the provider and the data collecting organisation.
    .9 Matters of detail in respect of the content of XCRI-CAP were also listed in the OCCAM Issue Log.
    .10 Performance issues should be carefully considered in the implementation of any XCRI-CAP system.

65. Main lessons learned:

    - Obtain early buy-in and resource commitment from technical staff.
    - If possible, commit more time and effort on linking with other relevant work within partner organisations.

66. Further opportunities have arisen to engage The Open University in work within the XCRI domain.

Stakeholders

67. The beneficiaries of the outcomes of the project are expected to be the XCRI community in general, and new institutions implementing XCRI in particular. OCCAM provides a case study of the implementation of data exchange between a provider and third party data collecting organisations.

Project Partners

68. The project partners worked effectively together, on the basis of a long-standing relationship. The project team was greatly assisted by good collaboration with XCRI Support Project, in particular very helpful detailed input from Mark Stubbs.

Project Management

69. Regular communication from and to the other mini-projects would have been useful. The project team was not able to learn from the other mini-projects during its life. We would encourage all future XCRI projects to make use of the XCRI forum and blogging facilities.

Programme Support

70. The project had good support from the XCRI Support Project.

71. It would have been useful to have had earlier publication of XCRI-CAP version 1.1, so that we could have discovered whether the new version resolved issues raised. We started work with version 1.0, but we found out very quickly that the nature of the OU's supported open and distance learning courses was in conflict with mandatory elements in the schema, and some extra structures might usefully be added. We knew that version 1.1 was likely to resolve these issues, but publication of version 1.1 was not within the time scale of the project.

Future Work

72. Characterisation of providers by technical capability, process maturity, and courses information management development might enable JISC to provide a service usage model in this domain, and a resource of 'likely problems and likely solutions' to help other HEIs implement XCRI.

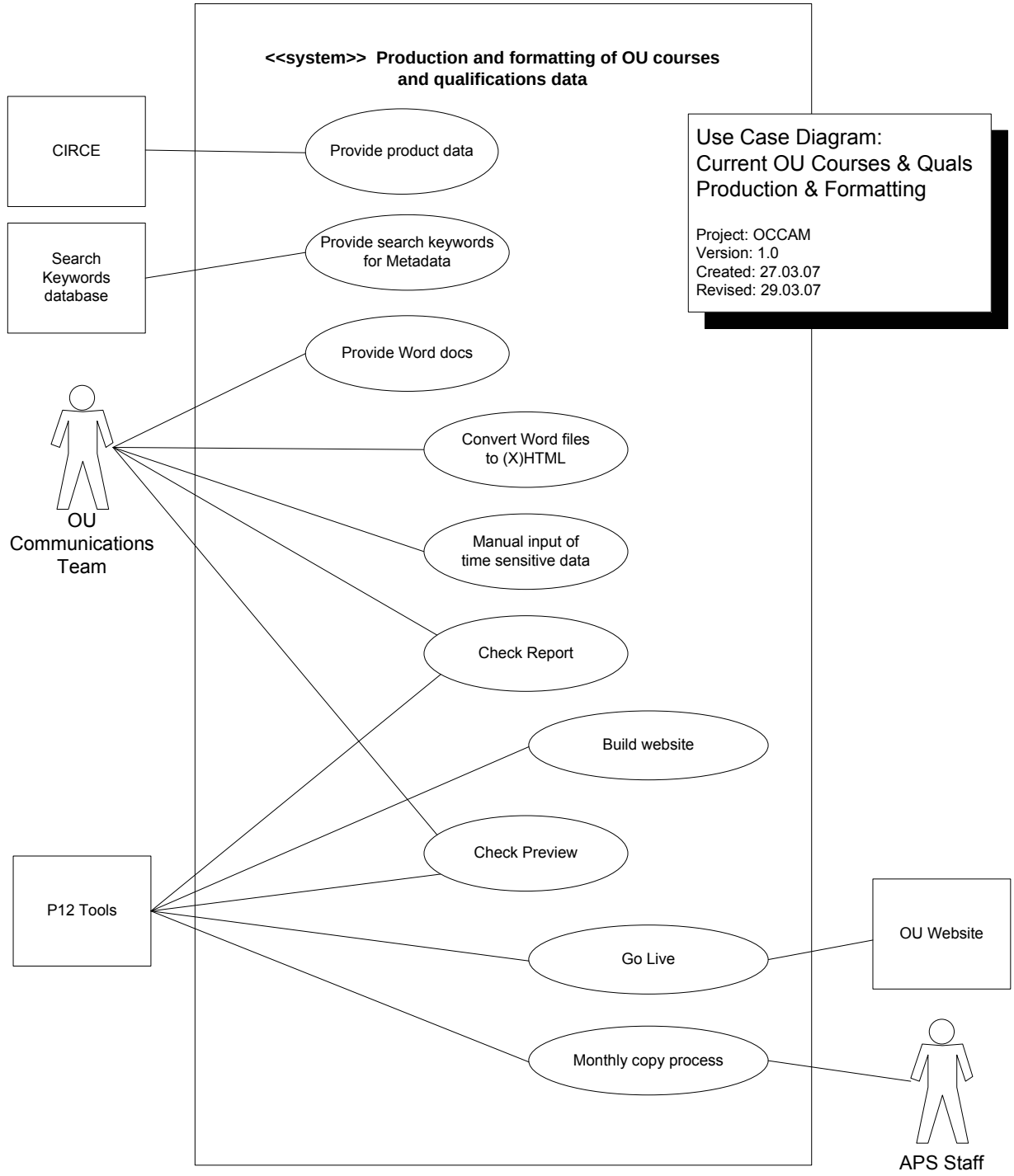
73. The Open University would like to offer leadership and resources for further investigation of the use of XCRI to describe in a structured way the hierarchy of Higher Education > Part Time Education > Distance Learning > Qualifications > Courses > progression information, and so on. In effect a schematic way in which institutions offering distance learning can be equitably represented.

Appendixes

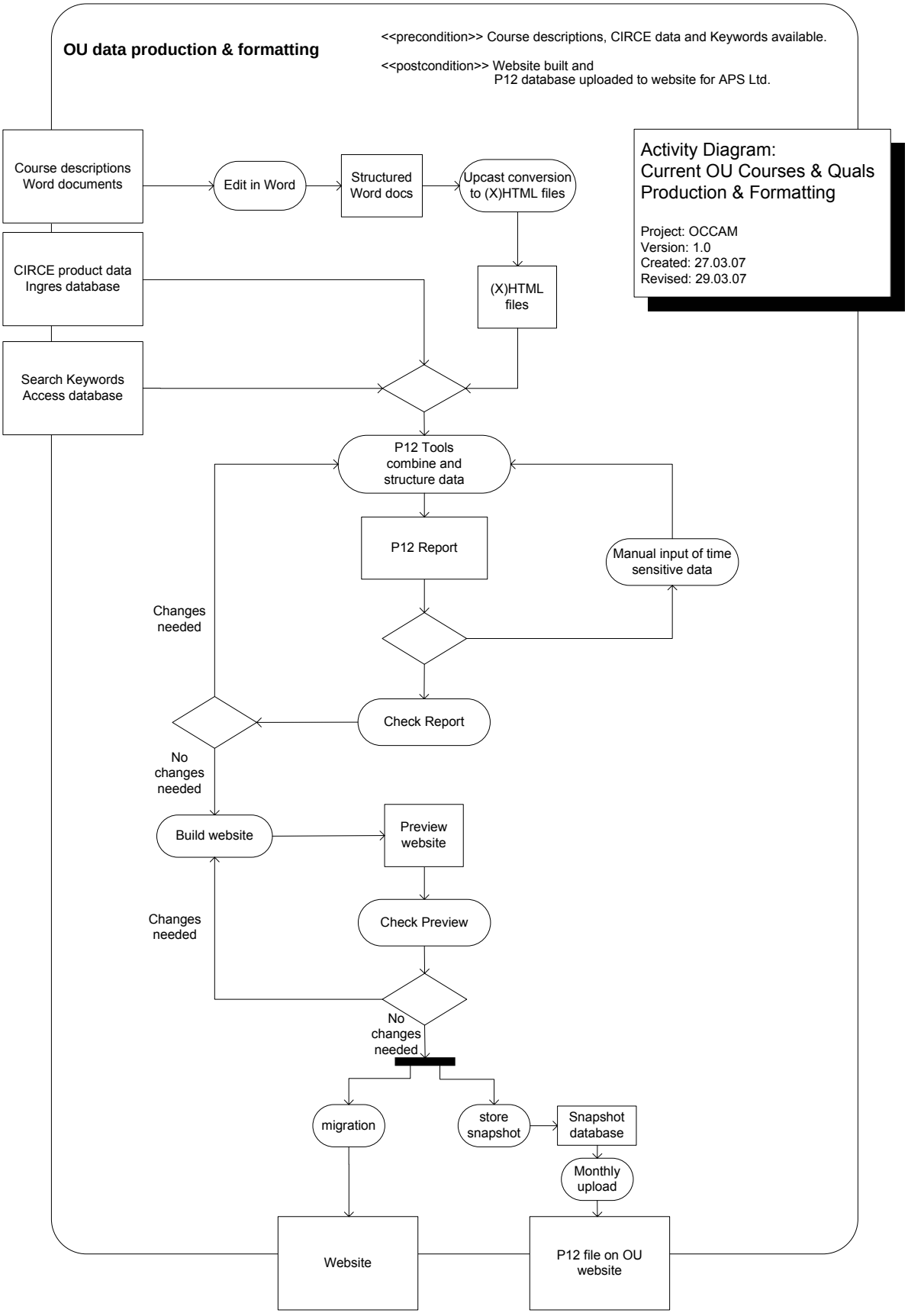
Appendix A: UML Model of current processes

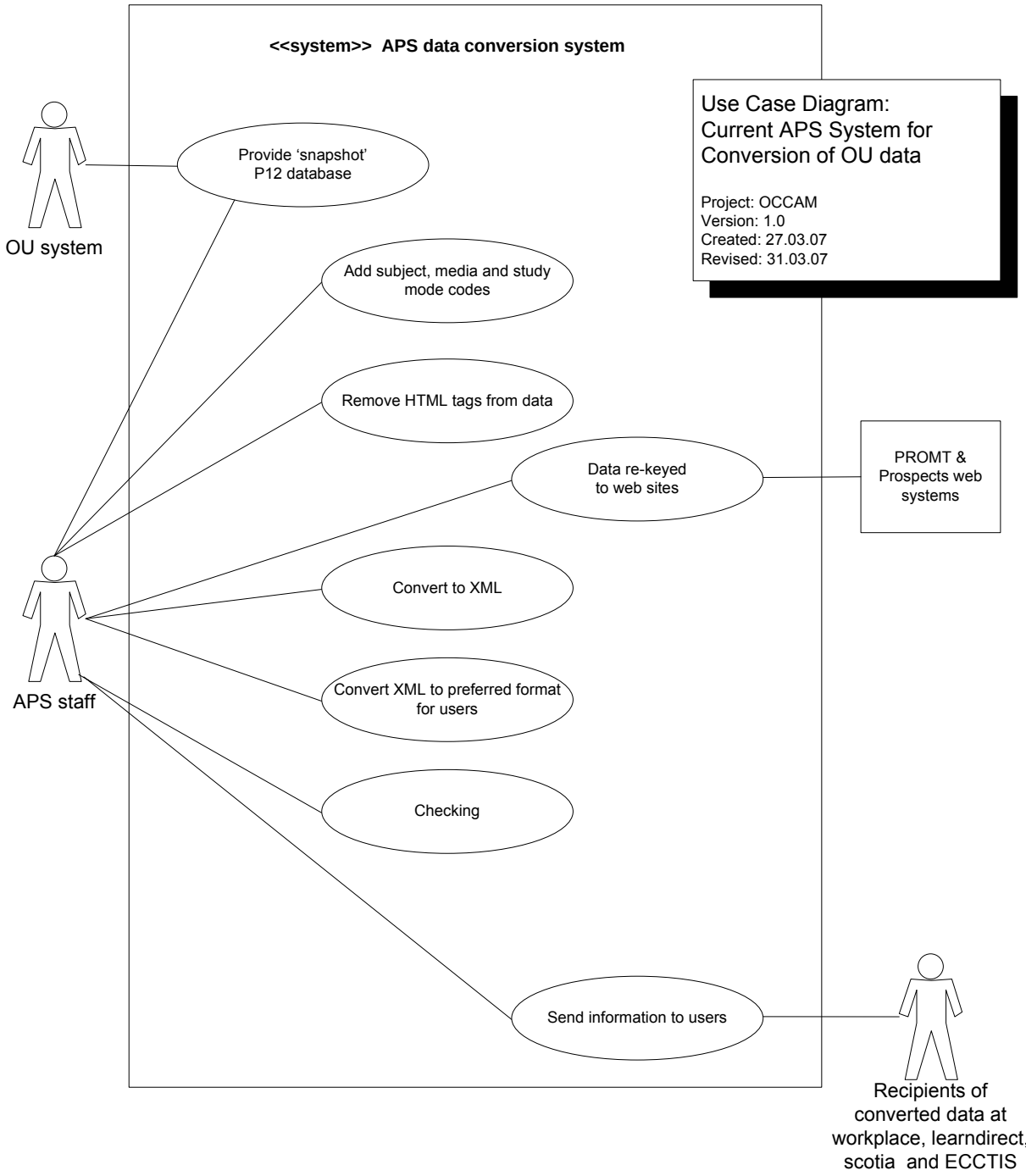
- OU Use Case Overview
- OU Activity Diagram
- APS Use Case Overview
- APS Activity Diagram

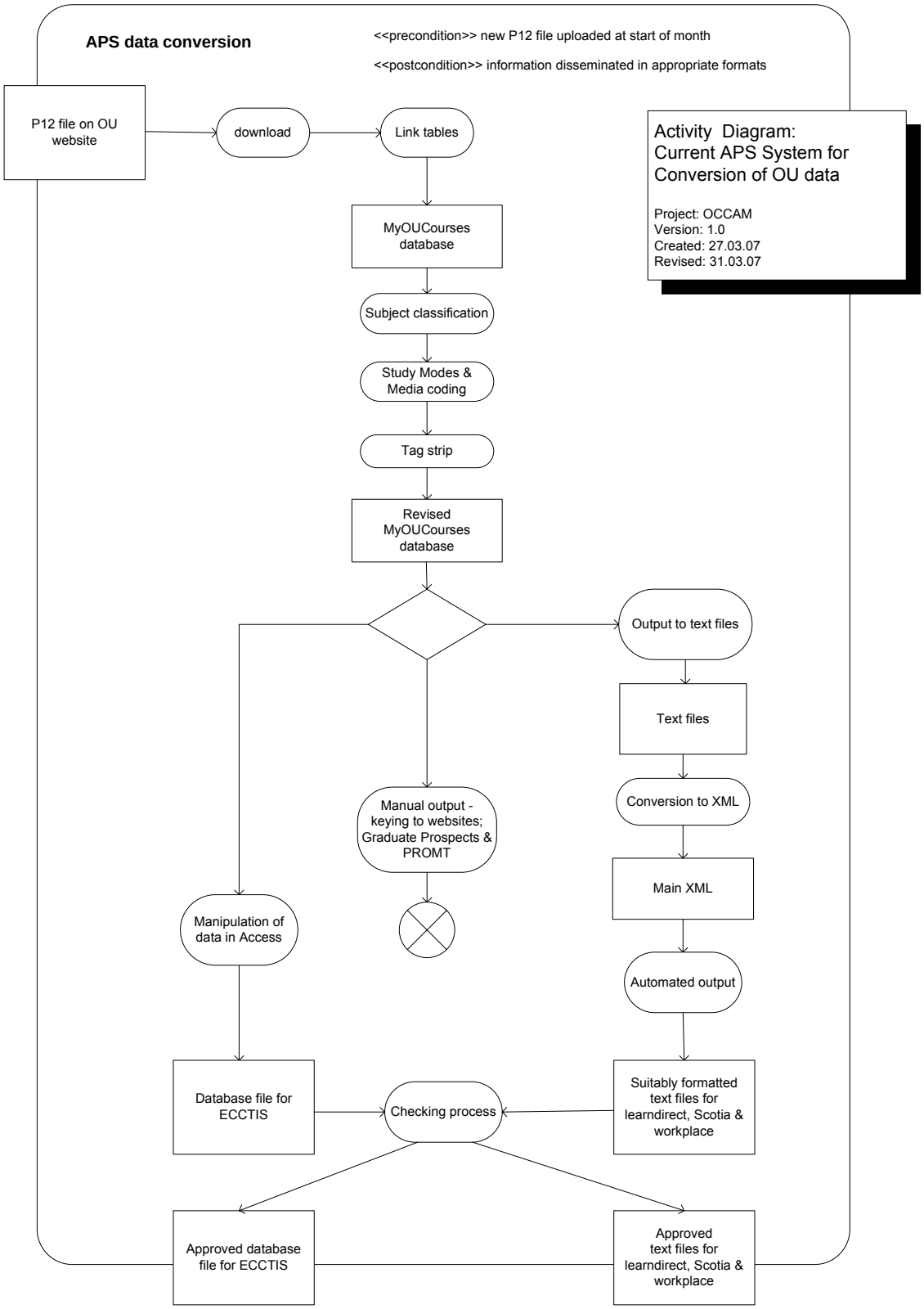
The UML Model is available in the project files as navigable HTML and XML with descriptions for each Use Case. For the sake of space just the diagrams are given here.


<<system>> Production and formatting of OU courses and qualifications data
CIRCE
Search Keywords database
Provide product data
Provide search keywords for Metadata
Provide Word docs
Convert Word files to (X)HTML
Manual input of time sensitive data
Check Report
Build website
Check Preview
Go Live
Monthly copy process
OU Communications Team
P12 Tools
OU Website
APS Staff
Use Case Diagram:
Current OU Courses & Quals Production & Formatting
Project: OCCAM
Version: 1.0
Created: 27.03.07
Revised: 29.03.07

Read Me


OU data production & formatting
<<precondition>> Course descriptions, CIRCE data and Keywords available.
<<postcondition>> Website built and
P12 database uploaded to website for APS Ltd.
Activity Diagram:
Current OU Courses & Quals
Production & Formatting
Project: OCCAM
Version: 1.0
Created: 27.03.07
Revised: 29.03.07
Course descriptions
Word documents
Edit in Word
Structured
Word docs
Upcast conversion
to (X)HTML files
CIRCE product data
Ingres database
(X)HTML
files
Search Keywords
Access database
P12 Tools
combine and
structure data
P12 Report
Manual input of time
sensitive data
Changes
needed
Check Report
No
changes
needed
Build website
Preview
website
Changes
needed
Check Preview
No
changes
needed
migration
store
snapshot
Snapshot
database
Monthly
upload
Website
P12 file on OU
website


<<system>> APS data conversion system
Use Case Diagram:
Current APS System for
Conversion of OU data
Project: OCCAM
Version: 1.0
Created: 27.03.07
Revised: 31.03.07
OU system
Provide 'snapshot'
P12 database
Add subject, media and study
mode codes
Remove HTML tags from data
Data re-keyed
to web sites
PROMT &
Prospects web
systems
APS staff
Convert to XML
Convert XML to preferred format
for users
Checking
Send information to users
Recipients of
converted data at
workplace, learndirect,
scotia and ECCTIS


APS data conversion
<<precondition>> new P12 file uploaded at start of month
<<postcondition>> information disseminated in appropriate formats
P12 file on OU website
download
Link tables
MyOUCourses database
Subject classification
Study Modes & Media coding
Tag strip
Revised MyOUCourses database
Output to text files
Text files
Conversion to XML
Main XML
Automated output
Manual output - keying to websites; Graduate Prospects & PROMT
Manipulation of data in Access
Database file for ECCTIS
Checking process
Suitably formatted text files for learndirect, Scotia & workplace
Approved database file for ECCTIS
Approved text files for learndirect, Scotia & workplace
Activity Diagram:
Current APS System for Conversion of OU data
Project: OCCAM
Version: 1.0
Created: 27.03.07
Revised: 31.03.07

Appendix B: Mapping Table

XCRI-CAP column: nesting of elements is shown by use of full-stops and one level of indent. Elements with collapsed sub-elements are shown with an ellipse (…) to the right.

O / M: optional or mandatory component

OU data item or constant value column: Name of data item from OU perspective, or indicates a constant value (shown in double-quotes).

OU definition column: Defines the OU data item.

Notes column: Issues, explanations of transformations, comments on XCRI-CAP v1.0, etc.

XCRI-CAP	O / M	OU data item or constant value	OU definition	Notes
catalogue …	M	N/A		root element; holds all the data.
catalogue@generated="[datetime]"	M	N/A		date or datetime of generation of the XML file
catalogue.provider …	M		OUBS would have different contact details	Holds data at provider level.
.provider@recstatus	O	N/A		Optional. Record change of status (Added; Updated; Deleted).
.identifier	M	"10007773"		UKPRN (UK Provider Reference Number); not (yet) held in OU systems Need a good reason NOT to use UKPRN for this. Rest of provider details could come from the UK Register of Learning Providers. No attribute to state what the identifier is.
.name	M	"The Open University"		OU question: where to get the data item from? Type of name is not prescribed; it could be official name or trading name. Suggest it's picked up from UKRLP through a web service.
.description …	O			Textual description of provider facilities.

				Currently considering where to get this from
.address …	O	"PO Box 197 Milton Keynes MK7 6BJ"		From Contact Database This data is for students (i.e. not a contact for data aggregators)
.street	M	"PO Box 197"		-ditto-
.town	M	"Milton Keynes"		-ditto-
.postcode	M	"MK7 6BJ"		-ditto-
.phone	O	"+44 (0)845 300 60 90"		-ditto-
.fax	O	"+44 (0)1908 654914"		-ditto-
.email	O	"general-enquiries@open.ac.uk"		-ditto-
.url	O	"http://www.open.ac.uk/"		-ditto-
.image …	O	OU logo		
.image@src	M	URI of OU logo		http://www.open.ac.uk/includes/oulogo_med.gif
.image@title	O	"OU logo"	Text constant	
.image@alt	O	"The Open University"	Text constant	
.provider.course …	M	courses and qualifications COURSES		Live courses and qualifications only; courses status = C or D THIS PART OF THE MAPPING HAS BEEN REPEATED FOR OU QUALIFICATION RECORDS TREATED AS COURSES.
..course@recstatus	O	N/A		
.identifier	M	Code	Key to course, e.g. A103	
.title	M	Title	Full course title	
.description …	M	course section codes	course description section headings	XCRI-CAP controlled vocabulary for type; optional, but if used, must have this vocabulary (currently mandatory, but this looks like human error in the schema design). Is it a good idea to put the description types

				vocabulary in CAP? Should be recommended but not in CAP.
.description@type="Aim"	O	section code 19	section heading "Summary" 100 word summary	
.description@type="Assessment Strategy"	O	section code 14	section heading "Assessment"	
.description@type="Career Outcome"	O	section code 2	section heading "Vocational Relevance"	
.description@type="Entry Profile"	O	section code ?	section heading ?	
.description@type="Indicative Resource"	O	section code ?	section heading ?	
.description@type="Learning Outcome"	O	section code 63	section heading "You will learn"	
.description@type="Policy"	O	section code ?	section heading ?	
.description@type="Provided Resource"	O	section code 7	section heading "Included in the course"	
.description@type="Regulations"	O	section code ?	section heading ?	
.description@type="Required Resource"	O	section code 8	section heading "You will need" Also section heading "Set books"	How do we map both? Could be done by concatenation of strings.
.description@type="Special Feature"	O	section code 36	section heading "Associated residential school course"	
.description@type="Support"	O	section code 41	section heading "Study support"	
.description@type="Structure"	O	section code ?	section heading ?	
.description@type="Syllabus"	O	section code ?	section heading ?	
.description@type="Teaching Strategy"	O	section code 13	section heading "Tuition and counselling"	
.url	O	URL from course description in C&Q		
.image …	O	N/A		

.dc:any	O			Any Dublin Core element can be used. Now changed to xs:any (which is better, because can use xsi:type attribute).
.dc:subject	O	Sub-subject	Courses are 'classified' into one or more 'descriptive' subjects, e.g. Art History	Can contain all the subject areas, if more than one.
.dc:subject	O	Subject Class #	LDCS subject classification (up to 3 codes) APS currently codes this; it doesn't come from the OU. We need to work out the most suitable intervention point.	Currently this contains the LDCS classification code, not natural language. Should this be converted? I think we need a better solution for recording classifications, so we can indicate the classification system.
.dc:subject	O	Faculty		Faculty name could be here? Question of where this should be in CAP.
..course.qualification …	M			In XCRI-CAP should not be mandatory?
.title	M	Qualification title		
.level	M	Level	Code is converted to text (P - postgraduate, U - undergraduate). Note, Level is not displayed for C21 and qualification is listed under ITT (neither PG nor UG).	Some OU courses can count towards UG and PG quals. Could have multiple qualification elements, but level is attached to qualification in OU data. SCQF level. Important for Scottish courses; has funding implications. So will need a level element on course.
.type	O	Qualification type	Possible values: Certificates Diplomas Degrees Initial Teacher Training Taught Masters Doctorates	
.awardedBy …	O	N/A		
.accreditedBy …	O	N/A		
..course.presentation	M	Presentation information	Each course can have 0, 1 or	Mandatory in CAP.

			more presentations	
.identifier	O	Presentation code	Uniquely identifies presentation	
.description …	O	N/A		
.start	M	Start date		Need to extract date from datetime for easier reading.
.end	O	End date		Need to extract date from datetime for easier reading.
.duration	O			Often required by aggregators, but not held as a specific field by OU. Implemented as calculation from Start and End Dates and rounded to nearest month; then concatenated string to format: "x months approx." Problem with short courses though.
.studyMode	M			Mandatory in XCRI-CAP, but not held as a specific field by OU. Standard text currently used: " The Open University offers part time distance education - known as supported open learning - which allows students to work at home in their own time, with the support of a personal tutor." However, for Ufi this needs to be classified. APS currently does this by analysing the section text fields.
.languageOfInstruction	O			Optional, not included.
.languageOfAssessment	O			Optional, not included.
.venue …	O			Optional, not yet addressed.
.placesAvailable	O	N/A		
.cost	O	Fee type Pricing area code	Combination of Fee type and Pricing area code	Has UK, Ireland, EU and rest of world Held in pence not pounds! Optionally more structure. Some collectors will want just the UK fee (e.g. Ufi), others will want all the details.

				Current implementation needs to take into account that "0" doesn't mean free.
.entryRequirements …	O	N/A		
.applyFrom	O	Reservations open date		Need to extract date from datetime for easier reading.
.applyUntil	O	Reservations close date (for new students)		Need to extract date from datetime for easier reading.
…presentation.applyTo …	M			Mandatory in XCRI-CAP. Should this be mandatory? Should just be able to reference the provider?
.identifier	M	"10007773"		UKPRN
.name	M	"The Open University"		Constant (could be picked up from URRLP)
.description …	O	"http://www.open.ac.uk"		Use same constant as provider.url
.address …	O	N/A		
.street	O			
.town	O			
.postcode	O			
.phone	O			
.fax	O			
.email	O			
.url	O			
.image …	O			
.provider.course …	M	courses and qualifications QUALIFICATIONS (as courses)		Live courses and qualifications only; qualifications status = C REPEATED FOR OU QUALIFICATION RECORDS TREATED AS COURSES.
..course@recstatus	O	N/A		
.identifier	M	Code	Key to qualification, e.g. B01	

.title	M	Title	Full qualification title	
.description …	O	qualification section codes	qualification description section headings	XCRI-CAP controlled vocabulary for type; optional, but if used, must have this vocabulary (currently mandatory, but this looks like human error in the schema design). Is it a good idea to put the description types vocabulary in CAP? Should be recommended but not in CAP.
.description@type="Aim"	O	section code 1	section heading "Description" 100 word summary	
.description@type="Assessment Strategy"	O	section code ?	section heading ?	
.description@type="Career Outcome"	O	section code 20	section heading " Career relevance"	
.description@type="Entry Profile"	O	section code ?	section heading ?	
.description@type="Indicative Resource"	O	section code ?	section heading ?	
.description@type="Learning Outcome"	O	section code 22	section heading " Learning outcomes"	
.description@type="Policy"	O	section code ?	section heading ?	
.description@type="Provided Resource"	O	section code 7	section heading ?	
.description@type="Regulations"	O	section code ?	section heading ?	
.description@type="Required Resource"	O	section code ?	section heading ?	
.description@type="Special Feature"	O	section code 36	section heading ?	
.description@type="Support"	O	section code 41	section heading ?	
.description@type="Structure"	O	section code 12	section heading " Courses and study lines"	

.description@type="Syllabus"	O	section code ?	section heading ?	
.description@type="Teaching Strategy"	O	section code 13	section heading " Educational aims"	
.url	O	URL from qual description in C&Q		
.image …	O	N/A		
.dc:any	O			Any Dublin Core element can be used. Now changed to xs:any (which is better, because can use xsi:type attribute).
.dc:subject	O	Subject area	Qualifications are 'classified' into one or more 'descriptive' subjects, e.g. Art History	
.dc:subject	O	Subject Class #	LDCS subject classification (up to 3 codes) APS currently codes this; it doesn't come from the OU. We need to work out the most suitable intervention point.	as for courses
..course.qualification …	M		No suitable component, so must create a dummy one; repeats much of the information in the "qualification as course" record	In XCRI-CAP should not be mandatory?
.title	M	Qualification title		
.level	M	Level	Code is converted to text (P - postgraduate, U - undergraduate). Note, Level is not displayed for C21 and qualification is listed under ITT (neither PG nor UG).	Some OU courses can count towards UG and PG quals. Could have multiple qualification elements, but level is attached to qualification in OU data. SCQF level. Important for Scottish courses; has funding implications. So will need a level element on course.
.type	O	Qualification type	Possible values: Certificates Diplomas Degrees	

			Initial Teacher Training Taught Masters Doctorates	
.awardedBy …	O	N/A		
.accreditedBy …	O	N/A		
..course.presentation	M	Presentation information for "qualification as course" record		Mandatory in CAP. Dummy record.
.identifier	O			Leave blank
.description …	O	N/A		
.start	M	"Variable"		
.end	O	N/A		
.duration	O	N/A		
.studyMode	M			Mandatory in XCRI-CAP, but not held as a specific field by OU. Standard text currently used: " The Open University offers part time distance education - known as supported open learning - which allows students to work at home in their own time, with the support of a personal tutor." However, for Ufi this needs to be classified. APS currently does this by analysing the section text fields.
.languageOfInstruction	O			
.languageOfAssessment	O			
.venue …	O			
.placesAvailable	O	N/A		
.cost	O			
.entryRequirements …	O	N/A		
.applyFrom	O			

.applyUntil	O			
…presentation.applyTo …	M			Mandatory in XCRI-CAP. Should this be mandatory? Should just be able to reference the provider?
.identifier		"10007773"		UKPRN
.name		"The Open University"		Constant (could be picked up from URRLP)
.description …		"http://www.open.ac.uk"		Use same constant as provider.url
.address …		N/A		
.street				
.town				
.postcode				
.phone				
.fax				
.email				
.url				
.image …				
		Credit points		Useful information for building a full programme / qual, but no place for it in XCRI-CAP.
		New course	Match First presentation field against a current/future presentation to flag the course as new for duration of first presentation only.	If True, then set recstatus = "Added" recstatus would have to be amended once a data collector has obtained the data; the audit trail process needs to be more sophisticated than this.

Appendix C: Evaluation of output files for August 07

Courses table	
Item	***Notes***
learningid	More records, because filtering not completed
provider_code	constant OK
provider_id	OK
title	OK; but for quals as courses, format is incorrect (we use a differently formatted one from source data - rekeyed).
price	Zero costs need to be transformed
entry requirements	OK
assessment	OK
technical info	OK
tutorial_support	OK
learning_abstract	constant OK
learn_summary	OK
learn_contact	constant OK
contact_phone	constant OK
contact_email	constant OK
subj_class	1 only - could be amended to include all
language	constant OK
last_update	constant OK
ucas_chck	constant OK
overseas_fee	Not supplied
date_added	constant OK
taught	constant OK
Qualifications	
Item	***Notes***
learnid	OK
qual_title	constant OK for courses; OK for quals as courses
qual_type	OK
ward_body	OK
Starts	
Item	***Notes***
	1 per start date; not done by year of entry like previous system
learningid	OK
venueid	constant OK
start_date	OK
start_details	OK
duration_code	OK
duration_desc	OK
timetable	N/A
startid	N/A
year_entry	OK
start_added	N/A

study_mode_1	All standardised at C; old system analysed text fields and therefore had greater detail.
Target Group	
Item	***Notes***
learningid	OK
target_group	constant OK