

# JISC eLEARNING PROGRAMME

## 14-19 Prospectus Pilot in the East Midlands

### Final Report

Project			
Project Acronym	N/A	Project ID	
Project Title	14-19 Prospectus Pilot in the East Midlands		
Start Date	1 November 08	End Date	31 March 09
Lead Institution	University of Nottingham		
Project Director	Alan Paull		
Project Manager & contact details	Alan Paull, alan@alanpaull.co.uk		
Partner Institutions	University of Nottingham Centre for International ePortfolio Development (CiePD) Aimhigher in the East Midlands Through CiePD team involvement, the LEAP AHEAD Lifelong Learning Network (LLN) for Nottinghamshire and Derbyshire 14-19 Area Prospectuses for Nottinghamshire, Leicestershire, Derbyshire, Northamptonshire and Lincolnshire & Rutland APS Ltd		
Project Web URL	<a href="http://www.XCRi.org/wiki/index.php/14-19_Area_Prospectus_XCRi_Aggregation_Pilot">http://www.XCRi.org/wiki/index.php/14-19_Area_Prospectus_XCRi_Aggregation_Pilot</a>		
Programme Name (and number)			
Programme Manager	Paul Bailey		

Document			
Document Title	<i>Final Report</i>		
Author(s) & project role	Alan Paull, Project Director		
Date	21 May 09	Filename	14-19report-Final.doc
Access	<input type="checkbox"/> Project and JISC internal		<input checked="" type="checkbox"/> General dissemination

Document History		
Version	Date	Comments
1.1	4 May 09	Draft
2.0	21 May 09	Final

## 14-19 Prospectus Pilot in the East Midlands

### Final Report

#### Executive Summary

The purpose of the JISC-funded 14-19 Prospectus Pilot in the East Midlands was to investigate any issues for 14-19 Area Prospectus organisations and the wider information management community in the implementation of data aggregation using the eXchanging Course Related Information<sup>1</sup> model. This document is a final report to JISC on the usage of XCRi for this purpose, and it includes wider recommendations in respect of using the XCRi Course Advertising Profile<sup>2</sup> standard<sup>3</sup> within the 14-19 Area Prospectus sector nationally.

The project aggregated 14-19 Area Prospectus course information from five different areas in the East Midlands into Aimhigher in the East Midlands' Progression Pathways database, as part of a wider project to amalgamate regional courses information. The project demonstrated that courses information can be supplied by many providers and Area Prospectus organisations using XCRi-CAP and can be aggregated into a single repository, the Progression Pathways database. The data is soon to be published on the live Progression Pathways service to show that the 14-19 course advertising information can be aggregated via XCRi-CAP and searched.

The project concentrated on co-ordinating the organisations involved to produce XCRi-CAP outputs and to aggregate them. It looked at the management of the processes to produce the course information feeds, as well as technical issues related to the compatibility of information models, data content and structures across the partners. It also touched upon data maintenance issues. It has produced a case study of XCRi-CAP implementation in the 14-19 sector that can be used as a source of good practice guidance for the sector as it pursues the LSC's 14-19 vision of interoperability and orienting data around learners.

The project achieved its central objective of aggregating data from Area Prospectuses into the Progression Pathways repository using XCRi-CAP 1.1<sup>4</sup>. Processes for producing the data feeds were documented and are supplied in the report in diagrammatic form.

As part of the work of the project, the data issues that were encountered in the process of mapping and transforming the data into XCRi-CAP were noted, and solutions were suggested. Major data issues included a lack of standard vocabularies in use in the sector, the variability in the types and comparability of the data held across the multiple data sets, and in some cases relatively poor quality of data in terms of the use of HTML and XHTML mark-up and consistency across the records. Use of the UKPRN to identify providers uniquely was problematic in 10% to 20% of cases, owing to name changes, mergers, federations or partnerships, and national providers with a local presence.

The project recommended that a set of clear, simple, and preferably national, interoperability guidelines in respect of good practice in information management be developed, covering in particular vocabularies, data entry standards and the re-use of existing resources, so that Area Prospectus organisations and other interested parties can share their courses data, building on the XCRi-CAP information model.

---

<sup>1</sup> <http://www.XCRi.org/>

<sup>2</sup> eXchanging Course Related Information - Course Advertising Profile

<sup>3</sup> <http://www.escs-isb.org.uk/news/news/Newsletter-February-2009.htm>

<sup>4</sup> [http://www.xcri.org/wiki/index.php/XCRI-CAP\\_XML\\_Schema](http://www.xcri.org/wiki/index.php/XCRI-CAP_XML_Schema)

## Table of Contents

### Contents

Executive Summary .....	2
Table of Contents .....	3
Part 1: Background .....	4
Introduction.....	4
Background: XCRi in the East Midlands Project.....	4
Background: Aimhigher in the East Midlands, Progression Pathways .....	4
Part 2: Aims and objectives .....	5
Part 3: Methods .....	6
Data collection.....	6
Mapping and transformation.....	6
Processes .....	7
Data aggregation methods .....	7
Issues addressed .....	8
Scope.....	8
Critical Success Factors .....	8
Technical development.....	8
Part 4: Work packages.....	9
Part 5: Area Prospectus Case Studies.....	10
Derbyshire .....	10
Leicestershire.....	10
Lincolnshire and Rutland .....	10
Northamptonshire.....	11
Nottinghamshire.....	11
Part 6: Review of data .....	12
Progression Pathways mapping.....	12
General data issues.....	13
Identifiers.....	13
Vocabularies .....	14
Part 7: Dissemination .....	17
Part 8: Conclusions .....	18
Aggregation.....	18
Process models.....	18
Data review and issues .....	18
Software vendors.....	20
Recommendations for the development of XCRi-CAP .....	20
Cross boundary issues .....	20
Good practice guidance for the 14-19 sector .....	20
Appendix A: Data issue log .....	22
Appendix B: Process Models .....	23
Appendix C: Institution mapping .....	27
Glossary of terms .....	28

## Part 1: Background

---

### Introduction

---

1. The project provided specific information and recommendations about live implementation of the emerging XCRi-CAP 1.1 standard within the 14-19 Prospectus arena.
2. The project was put forward by the Centre for International ePortfolio Development (CiePD, University of Nottingham), which has a long track record in supporting regional development work in the realm of interoperability to enhance the learner experience. The CiePD aimed to deliver the project by working with regional partner organisations, including up to five 14-19 Area Prospectus organisations, the Nottinghamshire and Derbyshire Lifelong Learning Network (Leap Ahead) and Aimhigher in the East Midlands. The project was fairly small in scope with a modest amount of funding available to support input from the Area Prospectus managers.
3. Area prospectus organisations involved in the project:
  - 14-19 Area Prospectus for Nottinghamshire
  - 14-19 Area Prospectus for Leicestershire
  - 14-19 Area Prospectus for Derbyshire
  - 14-19 Area Prospectus for Northamptonshire
  - 14-19 Area Prospectus for Lincolnshire & Rutland

---

### Background: XCRi in the East Midlands Project

---

4. The broader project, of which the 14-19 Area Prospectus formed a part, aimed to use XCRi to aggregate courses information from all learning providers in the East Midlands into the Progression Pathways database. A number of regional learning providers and Area Prospectus leads had already been exposed to XCRi through meetings and presentations from the CiePD's project team who also work on the technical and ePortfolio strands for Leap Ahead (the Lifelong Learning Network for Nottinghamshire and Derbyshire), and groundwork was prepared for a pilot involving some of the Leap Ahead partner providers and Progression Pathways.

---

### Background: Aimhigher in the East Midlands, Progression Pathways

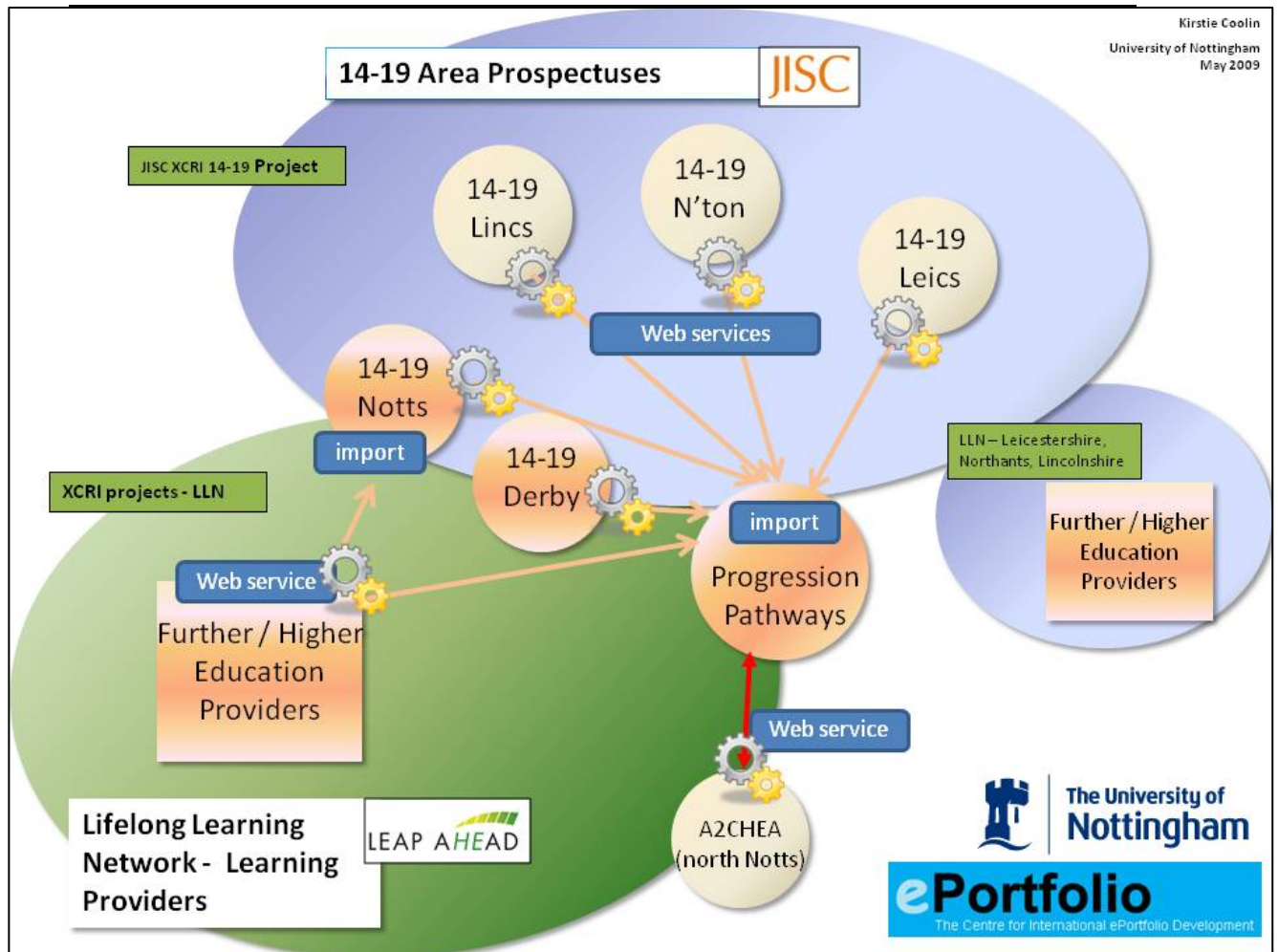
---

5. The vision of the Progression Pathways initiative is to develop a web-based information, advice and guidance (IAG) tool focusing on educational progression. A website funded by EMDA, Aimhigher and the Skills for Sustainable Communities and Leap Ahead Lifelong Learning Networks, the tool will include a comprehensive and searchable database of qualifications and courses (from basic skills/skills for life, to postgraduate level) that are available throughout the East Midlands. Beyond this, by linking courses, the tool will be able to generate customised progression maps, allowing learners, along with practitioners and employers, to discover routes into higher education that are specific to their needs. These routes will cover vocational, work-based and general academic pathways. They will feature case studies of learners who have taken comparable progression routes. Links to sources of subject and sector-specific careers information will also be available within each generated map. The tool will contain contact details of, and links to, all colleges, universities and private training providers, as well as acting as a sign-posting tool for information on funding opportunities, guidance on application processes and other useful information. Sector profiles will also be presented, providing users with detailed up-to-date labour market information for the East Midlands.

## Part 2: Aims and objectives

6. The project aimed to:
- Engage at least four 14-19 Area Prospectus data suppliers;
  - Support them in the planning and implementation of XCRi for output to Progression Pathways, as a first step towards national aggregation;
  - Demonstrate 14-19 Area Prospectus course advertising information aggregation;
  - Make recommendations to JISC on wider roll-out for 14-19 Prospectus purposes.

FIGURE 1: OVERVIEW OF THE EAST MIDLANDS PROJECTS



## Part 3: Methods

### Data collection

7. The Area Prospectuses already had a remit to collect the data they needed from institutions (including schools) and therefore already had a relationship with learning providers under this remit. This meant that institutions would not be asked more than once for the same information and were not asked to provide information individually to Progression Pathways. Problems of data collection would thereby be minimised.
8. Sharing of school data was restricted to school sixth form information, not all school data. The Progression Pathways team collects college and higher education data itself.

### Mapping and transformation

9. The project used many of the techniques already successfully developed for earlier XCRi mini-projects, centred on the XCRi Implementation Models (XIM) method<sup>5</sup>. Owing to the short time scale, relatively small funding for the project and the relatively small organisations involved, we did not attempt to carry out a rigorous assessment of XCRi readiness for each partner. Instead we engaged a leading manager and a leading technical expert in each of the organisations and provided technical support through the ClePD team with Leap Ahead LLN support.
10. The project built on earlier work on 14-19 Prospectus data carried out by the XCRi Support Project, particularly through Pan-London data, and APS Ltd through Greater Manchester area data. This work had involved an examination of scaling of the volumes of the data using XCRi-CAP 1.0 and some basic mapping work to establish that XCRi-CAP 1.1 was likely to be useful in this context. The early work had shown the feasibility of using XCRi-CAP 1.1, but had suggested that a full pilot would be necessary, in order to examine data compatibility and aggregation issues.

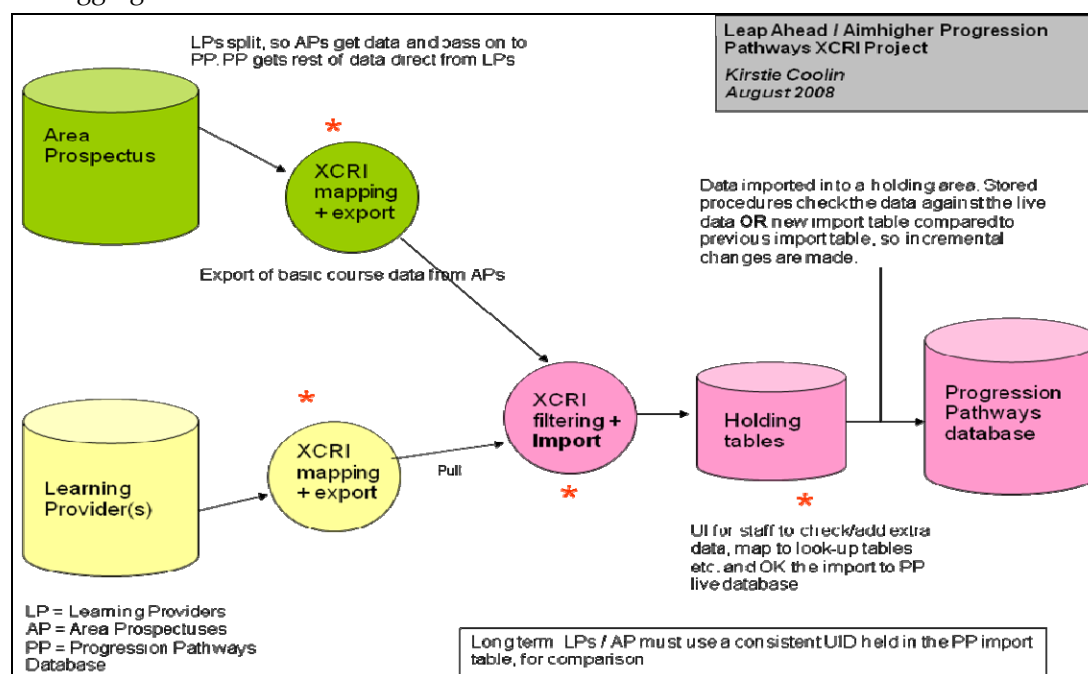


FIGURE 1: OVERVIEW OF THE DATA AGGREGATION PROCESS, SHOWING AREA PROSPECTUS WORK IN GREEN

<sup>5</sup> See [http://www.XCRi.org/wiki/index.php/XIM\\_-\\_XCRi\\_Implementation\\_Models](http://www.XCRi.org/wiki/index.php/XIM_-_XCRi_Implementation_Models)

11. APS carried out pilot mappings and transformations from samples of the existing data formats for each Area Prospectus to XCRi-CAP 1.1 and from there to the Progression Pathways data requirements. The mapping output was described in a standardised spreadsheet<sup>6</sup> with the following headings:

Data item or constant value	AP definition (where different)	XCRi CAP	Progression Pathways	Notes	In feed (vs Mandatory PP)
-----------------------------	---------------------------------	----------	----------------------	-------	---------------------------

12. APS created sample XCRi-CAP files from the sample data using a proprietary visual data mapping tool called Altova MapForce<sup>7</sup>. This commercial tool has a graphical interface that enables the user to 'drag-and-drop' data items from one format to another format, for example from a database to XML, and has built-in transformation components based on XPath together with customisable and user-defined components.
13. APS supplied the pilot mapping outputs back to the Area Prospectus organisations, so that they could convert and refine them for their implementation of the XCRi-CAP feed. The intention then was for the Area Prospectus organisations, or their software suppliers, to provide their data via web services or via static XML files, using interfaces designed and developed by the project team or by local in-house or software vendor technicians.
14. The project reviewed the data content and structures of the existing course advertising information from the Area Prospectuses, so that any issues could be discovered, recorded and resolved. Data issues are listed at Appendix A. The project intended to take existing data, if necessary improving the quality of coded, searchable fields, then converting it to XCRi-CAP format, so that it could be aggregated automatically from separate information flows into the pilot service.

---

### Processes

---

15. The pilot activity aimed to confirm the suitability of the data for the outlined purposes and to suggest amendments to the data entry processes. Although recommendations were made to partners about changes in order to improve data quality and consistency, it was a primary principle of the pilot project that existing processes and data would not be disturbed. Therefore the pilot processes were to an extent implemented alongside live activities.
16. The data aggregation and transformation activity concentrated on identification and some resolution of issues around the vocabularies to be used and data entry standards, building on earlier 14-19 work carried out by APS Ltd in the Greater Manchester area.
17. Process modelling was carried out using the ArchiMate Enterprise Architecture modelling language. Results for three of Area Prospectus supply processes are given at Appendix B and in separate diagram files. Northamptonshire and Leicestershire supply processes were not completed by the time of reporting, so are not given separately. It is likely that both will be similar to the Nottinghamshire one, using a web service.

---

### Data aggregation methods

---

18. The project worked through Aimhigher in the East Midlands to import the raw data, validate it for errors in the mark-up and vocabularies, code and classify the data where necessary to enable accurate searching and consistency across the data set, and carry out data maintenance and updating tasks. Import services were developed by the software supplier for Aimhigher in the East Midlands, Winona.
19. Organisations were encouraged to implement a web service for providing the data, with assistance from the CIEPD project team, but it was deemed acceptable for the data to be supplied via a standalone static XCRi-CAP file.

---

<sup>6</sup> See '14-19\_MappingSpreadsheetTemplate.xls'

<sup>7</sup> See [http://www.altova.com/products/mapforce/data\\_mapping.html](http://www.altova.com/products/mapforce/data_mapping.html)

20. Course information is published in the Progression Pathways service for use by learners. Review of the collection of this data from colleges and HEIs is being carried out by the wider Leap Ahead XCRi in the East Midlands project, but not by this pilot.

---

**Issues addressed**

---

21. Specific issues that were addressed included:
- Data quality across the data sets, and any implications for data creation and maintenance
  - Choice of vocabularies, including any coding and classification issues
  - Representation of vocabularies in XCRi
  - Mapping of different data structures from the different data sets to a single repository

---

**Scope**

---

22. The mini-project's scope was limited to the processing of 14-19 Area Prospectus data, its aggregation and publication, and any organisational, technical and process issues related to this work. The wider Nottinghamshire and Derbyshire LLN project handled issues related to the data from other Learning Providers.

---

**Critical Success Factors**

---

23. The mini-project depended on the co-operation of all project partners, particularly a sufficient number of the Area Prospectus organisations. Our earlier work through the XIM Project demonstrated that technical issues, though possible, are rarely show-stoppers, but organisational, political or resource elements can be problematic, and this project was no exception. Knowing the difficulties in advance, we addressed these at the earliest stages of the project, gaining management buy-in through an early project-wide meeting and technical buy-in through a series of hands-on workshops. Software vendors for each of the Area Prospectuses were invited to send technical representatives and both VIS Communications (Northamptonshire's and Lincolnshire & Rutland's supplier) and S-Cool (Leicestershire's supplier) did so.

---

**Technical development**

---

24. Technical work to be carried out by the Area Prospectuses, by their software vendors and by the ClePD team on their behalf, consisted of:
- Deciding on both business and technical processes to establish how the data would be exported and developing the corresponding XCRi export of course advertising data.
  - Participating in the review and testing of the XCRi-CAP feed and the Progression Pathways import module.

The import module was developed by the wider ClePD project team and Winona, the Aimhigher in the East Midlands software vendor. It was designed to be used both for this project and further developed for the corresponding Leap Ahead project.

25. Technical support was provided through telephone and site visits.

## Part 4: Work packages

<i>Work Package</i>	<i>Description</i>
<b>WP1:</b> Project initiation	Project plan, product planning, agreement to detailed activities by each partner
<b>WP2:</b> Review data	Identification of data to be transferred, review of vocabularies in use, data quality. Trial mapping of each data set to XCRi-CAP.
<b>WP3:</b> Implement XCRi	Implementation of XCRi mapping, transformation and export
<b>WP4:</b> Data maintenance	Design of data maintenance processes
<b>WP5:</b> Import to Progression Pathways	Develop import functions at Progression Pathways
<b>WP6:</b> Develop user interface	Development of UI for Progression Pathways staff and to demonstrate searching across the data
<b>WP7:</b> Evaluation and reporting	Write report (including case study) and recommendations to JISC
Management, evaluation and dissemination throughout	LLN subgroup to disseminate progress and outcomes: presentations at appropriate events; regional workshops

## Part 5: Area Prospectus Case Studies

26. The data mapping for each of these brief case studies is presented in a consolidated table at Appendix C, which also includes a mapping to the Progression Pathways data format.

---

### Derbyshire

---

27. The 14-19 Area Prospectus is run by Connexions Derbyshire and is published via the Routes Ahead website at <http://www.routes-ahead.org/>. The data is managed by its own staff, not by a commercial software vendor, in a MySQL database.
28. Sample data was mapped and transformed to XCRi-CAP using Altova MapForce. The mapping and an instance file were reflected back to Connexions Derbyshire. The only difficulties with the mapping were in relation to ill-formed HTML and the use of a non-date format in the start information. Some of the data had been sourced from Word documents (via cut-and-paste), which resulted in control characters appearing in the data; these had to be eliminated by a series of automatic and manual data cleansing processes carried out by Connexions Derbyshire prior to the conversion.
29. The course, qualification and presentation structure mapped well to XCRi-CAP 1.1.
30. Connexions Derbyshire's Web Developer has produced a feed in the form of an XCRi-CAP export available on the web at [http://www.routes-ahead.org/utils/ra\\_xcri\\_export.asp](http://www.routes-ahead.org/utils/ra_xcri_export.asp).

---

### Leicestershire

---

31. The 14-19 Area Prospectus is run by Connexions Leicester Shire and is published via the CourseFinder website at <http://www.connexions-leics.org/learning/>. Backend services to the course advertising data are provided by S-Cool.
32. Sample data was received in XML format, initially from Connexions Leicester Shire who extracted it from the CourseFinder system and from the LeCAP system; the two data sets could be linked via a common identifier. The project was then encouraged to work directly with S-Cool to enable a more flexible data provision.
33. Sample data was mapped and transformed to XCRi-CAP and reflected back to Connexions Leicester Shire and S-Cool with an instance file. Production of the sample XML file proved not to be technically difficult, but there was a significant challenge in fitting in the development of a feed within S-Cool's existing schedule. The sample data was well-formed.
34. Courses data was held at presentation level only, so there was a one-to-one mapping between course in the input data and course plus presentation in XCRi-CAP.
35. S-Cool was able to provide a large and comprehensive XCRi-CAP data file towards the end of the project. There were initial minor problems with validation of the file, but the project was able to overcome them with little difficulty. This file represented a repeatable stand-alone file rather than a dynamic feed.

---

### Lincolnshire and Rutland

---

36. The 14-19 Area Prospectus is run by Connexions Lincolnshire and is published via their 14-19.info website at <https://www.14-19.info/>. Backend services to the course advertising data are provided by VIS Communications (now part of UCAS Media), and the data is supplied via an Application Programming Interface (API) that obtains search results from the Learning and Skills Council's National Learning Directory (NLD), managed by Hotcourses Ltd.
37. There was a one-to-one mapping between course and course plus presentation.
38. It was not until comparatively late in the project that it became clear that the data was obtained on a course by course basis in the form of search results and could not be readily obtained as a data set without extra development work. Fortunately the Hotcourses team was already working on another XCRi-CAP project and were able to develop a relevant mechanism to output the data. XCRi-CAP data was obtained as a separate XML instance file for each school or college from Hotcourses. The instance files used a Hotcourses extension to

21 May 2009

XCRi-CAP 1.1 that enabled NLD-controlled vocabularies to be used and validated. The data was transformed into the Progression Pathways XCRi-CAP version via a transformation on import.

39. This data was obtained as a 'once off' operation. Further discussion and resourcing will be necessary before a regular feed can be organised.

---

### **Northamptonshire**

---

40. The 14-19 Area Prospectus is run by Connexions Northamptonshire and is published via their Xplore website at <https://www.xplore.uk.com/>. Backend services to the course advertising data are provided by VIS Communications (now part of UCAS Media). The data is provided by Connexions Northamptonshire's 14-19 team.
41. Sample data was supplied and mapped and transformed to XCRi-CAP. Provider and venue information were not differentiated in different tables, so it is not currently possible to identify the provider details, except for name and, in some cases, UKPRN. There is an acknowledged problem over identifying duplicate providers and venues, which is currently being addressed. For a regular supply of data, this difficulty will have to be resolved.
42. The data supports the proposed new recDateTime attribute at presentation (course instance) level, so that data maintenance can be addressed. The course, qualification and presentation structure maps well to XCRi-CAP 1.1.
43. In the event it did not prove possible to implement an XCRi-CAP feed within the time scales of the project. During the project VIS Communications was bought by UCAS Media, and in addition the staff's concentration was on the new version of their common application process product. However, there is an undertaking that a feed will be provided at the end of May 2009, initially via an XML flat file and subsequently via a web service.

---

### **Nottinghamshire**

---

44. The 14-19 Area Prospectus is run by Connexions Nottinghamshire and is published via their My Choice website at <http://www.mychoicenotts.com/>. Backend services to the course advertising data are provided by Connexions Nottinghamshire.
45. Sample data was received from Connexions Nottinghamshire, and it proved to be relatively simple to map the data to XCRi-CAP. It was necessary to carry out some work to remove data redundancy, for example 'level' information was held in flagged fields rather than via reference data, and basic provider data had to be extracted from the course samples.
46. There is a one-to-one mapping between course and course plus presentation.
47. An XCRi-CAP web service feed was implemented by the CIEPD project team on behalf of Connexions Nottinghamshire. Details can be found here: <http://80.249.109.45/passportfolio/export/service.asmx>. UKPRNs have been added manually to the data as an aid to provider identification.

## Part 6: Review of data

48. In June 2006 the National Foundation for Educational Research issued national guidance on "How to write a 14-19 Prospectus". Our starting point for reviewing the data content of the Area Prospectuses was this document, and our initial mapping was as follows:

<i>Course data items from National Guidelines</i>	<i>Course data items in XCRi-CAP</i>
Course Title (QCA / S96 recommended)	title (multiples permitted)
Age range (14-16; 16-19)	description (age range type)
Qualification	qualification.title
Level	level
Topics / modules	description (topic; structure)
Entry requirements	entryRequirements
Cost	cost
Awarding body	awardedBy
Assessment methods	description (assessment)
Progression routes	relation

<i>Course data items from National Guidelines</i>	<i>Course data items in XCRi-CAP</i>
Title	title (multiples permitted)
Location	town
Post code	Postcode
Provider type	description
Contact details	address, phone, fax, email, url
Description	description

49. While this mapping shows a basic level of compatibility between the data items, each data item was open to interpretation, and in many cases a data definition peculiar to the particular Area Prospectus was adopted. The National Guidelines did not address interoperability issues or data aggregation. The project team was therefore conscious that it was essential to address data compatibility issues and that analysing and mapping each Area Prospectus data set to XCRi-CAP and to the Progression Pathways data format were necessary.

---

### Progression Pathways mapping

---

50. Progression Pathways required relatively few mandatory data items, as follows:

#### Provider record

- Provider
- ProviderID

#### Course record

- Course (course title)
- Course Web-page
- First Subject Area (look up)
- First QCA Subject (look up)
- Second Subject Area (look up)
- Second QCA Subject (look up)

- CollaborativeProvider
- Course Web-page
- Award Qualification
- Award Level (look up)
- Award Type (look up)
- Provider Course Code
- Duration Amount
- Duration Unit (look up)
- Attendance (look up)
- ER\_Entry Requirements

51. From these data items it can be seen that comparatively few mandatory descriptive textual items would be imported, but that much of the data was made up from vocabularies. Only the QCA subject vocabulary was not proprietary to Progression Pathways.

52. We reviewed sample data from each of the five Area Prospectuses and the three colleges and single university involved in the wider data aggregation project. The data sets were mapped to XCRi-CAP and then to Progression Pathways, so that we could determine those data items for which we had data from all or most of the learning providers.
53. At provider and venue level we were able to obtain most of the basic contact and location data that was required. See 'providers' section below for issues in this area.
54. The sample data showed that for course and presentation level data we had only two data items from all the providers, namely 'course title' and 'course identifier', noting that in some instances these data items also doubled up as presentation title and identifier. The following data items at course and presentation level were obtainable from all or most suppliers:
  - Course
    - identifier
    - title
  - Qualification
    - title
    - level
    - type
  - Presentation
    - identifier
    - attendance mode
    - duration
    - entry requirements
    - venue title

---

### **General data issues**

---

55. Common difficulties were caused by poor data quality and included:
  - Lack of consistency in identifiers for course, presentation, venue and / or provider, particularly in relation to permanent identifiers
  - Mark-up (generally HTML) that was not well-formed
  - Start date information being in free text rather than date format
  - Data not normalised, so that, for example, Venue Names are repeated with the danger that different names are used for the same venue
  - Gaps in searchable data items, including subject, duration, start details.
56. For linking to additional courses information over and above that held in the Progression Pathways system, we had hoped to be able to import course-based URLs. However, in most cases this did not prove possible, because permanent course URLs were not often held by the Area Prospectuses. Where this was not the case, the Progression Pathways system could link back to the main Area Prospectus URL.

---

### **Identifiers**

---

57. XCRi-CAP recommends the use of Uniform Resource Identifiers (URIs) as permanent identifiers. This did not prove practical for the project, because URIs were not widely used for this purpose in the source data sets. Permanent identifiers for provider and course elements were particularly important, and most data sets used their own internal IDs for them. Progression Pathways used its own internal IDs for these too.

### **Provider identifiers**

58. The main issue was: how could the importing system know uniquely and reliably that the stated provider was in fact the one in the existing data set? A subsidiary problem was: how

should we address the fact that for some providers more than one data source might supply updated information (for example national providers or those close to county borders)?

### **Use of the UK Provider Reference Number (UKPRN)**

59. The only comprehensive use of MIAP's<sup>8</sup> provider identifier, the UKPRN in suppliers' data was in that from Lincolnshire and Rutland, which was scarcely surprising since Hotcourses developed the UK Register of Learning Providers (UKRLP), which promotes this identifier. The Northamptonshire data also held the UKPRN for a proportion of its records.
60. Major difficulties were encountered in using the UKPRN as a unique identifier in the context of course advertising information. The major reason for this was that the UKPRN identifies the legal entity, not the entity that is marketed to students. This issue raised the question of whether the UKPRN is suitable as a unique identifier for course advertising information. The following problems were noted:
- Where a provider has changed its name, but the aggregator only has the original one, it can be difficult to track down the UKPRN, because historical data is not retained on the UKRLP (for example Sherwood Hall School & Sixth Form College is now called The Samworth Church Academy).
  - Merged providers or those with more than one main venue might be promoted as if they were separate brands, but would only have a single UKPRN (for example Newark College is now part of Lincoln College; High Pavement Sixth Form College is part of New College Nottingham).
  - A federation or partnership made up from provision from several schools or colleges may be marketed as one provider, but will have no UKPRN, because it may not be a legal entity in its own right (for example White Hills Park if a federation of Alderman White School, Bramcote Hills Sport and Community College, Bramcote Park Business and Enterprise School, all three of which have their own UKPRN).
  - National providers may have only one UKPRN, but market themselves through local or regional entities.
61. A possible resolution of this issue would be to concentrate marketing information purely on the venue element, retaining the provider element purely for identifying the legal entity. However, this idea might be simply displacing the issue of uniquely identifying the entity from the provider to the venue.
62. For course advertising information the learning provider will usually be identified as "the company or organisation that provides the learning opportunity"<sup>9</sup>. Its name will usually be "the trading name and for open or flexible learning may be the originator, publisher or distributor as appropriate"<sup>10</sup>. For uniquely identifying the provider with a public ID, it is suggested that data collectors and learning providers use the provider's URL. The advantage of this identifier is that it is readily discoverable, and sub-units of over-arching providers can be readily identified separately from the host organisation. The disadvantage of this identifier is that a small minority of private providers do not have a URI; however, this disadvantage should reduce to negligible proportions over time.

---

### **Vocabularies**

---

63. As we examined Area Prospectus data, it quickly became apparent that importing specific common vocabularies was not going to be possible for most search fields, because the data was either not present or not amenable to automatic transformation. As the principles of the project were to disturb existing processes as little as possible, coding and classification of the newly aggregated data was work undertaken by Aimhigher in the East Midlands.

---

<sup>8</sup> Managing Information Across Partners, see: <http://www.miap.gov.uk/>.

<sup>9</sup> From the National Learning Directory Data Entry Standards, version 6

<sup>10</sup> *ibid*

64. Earlier work had suggested vocabularies as follows:

<i>Data Item</i>	<i>Possible vocabulary</i>
Subject	Sector Subject Area
Level (of qualification)	National Learning Directory, Qual Aim
Level (of qualification), alternative	Enhanced QCF level
Duration	National Learning Directory, Duration Code
Awarding Body	Composite list from QCA, LAD and HMRC lists
Attendance Pattern, Attendance Mode and Study Mode	XCRi-CAP draft vocabularies

### **Subject**

65. Progression Pathways uses the QCA Sector Subject Area system, though with some modifications. Most of the data sources had an approximation to this classification system, so it is recommended that the Sector Subject Area classification be adopted for 14-19 Area Prospectuses. If more differentiation for subject searching is required, then it is recommended that the Learndirect Classification System (LDCS) is used<sup>11</sup>.

### **Level of qualification**

66. Some Area Prospectuses used the QAN number to identify the qualification, so level could be picked up from national resources, such as the National Database of Accredited Qualifications or the QAN website. However, there is not consistency here, and the national resources only contain information relating to the National Qualifications Framework (NQF) or the Qualifications and Credit Framework (QCF), neither of which is comprehensive for all learning. Use of this numerical system also requires a level of explanation appropriate for the NDAQ that may not be appropriate for course advertising information.
67. It is recommended that the National Learning Directory qualification aim coding system be used<sup>12</sup>.
68. An alternative would be to enhance the QCF level (1 to 8) from the NDAQ, but augmented by the following additional values:
- 1 / 2 GCSE courses that could lead to levels 1 or 2 dependent on grade achieved;
  - V 'Various', for instances where other multiple levels were necessary;
  - N/K NOT KNOWN

The additional values are necessary, because there are courses available at multiple levels, and some courses do not have an associated level.

### **Duration**

69. There was no consistency across the data sets for duration. It is recommended that the National Learning Directory duration code system<sup>13</sup> be used.

### **Awarding Body**

70. As noted in the MIAP common data definitions, there is no authoritative and comprehensive directory of awarding bodies. For earlier 14-19 work in Greater Manchester, a composite list

---

<sup>11</sup> ibid

<sup>12</sup> ibid

<sup>13</sup> ibid

from notable sources, including QCA, LAD and HMRC lists (total of 784 entries) was constructed. An extension of this work could be to validate the data against this type of list.

71. It is recommended that an appropriate national authority take on this work.

**Attendance Pattern, Attendance Mode and Study Mode**

72. There was no consistency in the usage of these terms, and in fact many approaches to them resulted in mixed data fields (for example “two years full time”). Attendance Mode (for example “full time”) was the only one in which there was data available from a majority of the sources.
73. The vocabularies given below were draft ones from the XCRi-CAP 1.1 refinement schema and have been further refined as a result of earlier work, and extended to cover more types of learning. If this type of information is collected, it is recommended that these three vocabularies be refined, agreed and used, although it is also recognised that many data collectors will not wish to collect information to this level of detail.

***Attendance Pattern (alternatively this could be named ‘Delivery Pattern’)***

Daytime

Sandwich (definition: a significant, continuous, period of the course is spent in work placement / experience or studying elsewhere)

Evening

Twilight

Day / Block release

Weekend

Short (definition: the learning is up to one week in length and takes place in a continuous block or period of study occurring only once; can be as short as 1 hour.)

Customised (definition: the method of and time taken for learning can be negotiated between the learner and the deliverer)

To suit the learner (definition: pattern entirely controlled by the learner)

Not known

***Attendance Mode (alternatively this could be named ‘Delivery Mode’)***

Campus

Distance with attendance

Distance without attendance

Mixed mode

Work-based

Online

Not known

***Study Mode***

Full time

Part time

Flexible (could be FT or PT dependent on the learner)

Not known (necessary if the Study Mode is mandatory)

## Part 7: Dissemination

74. Recording of our collective experiences was a very important aspect of dissemination of the project outcomes. This was carried out during the project through attendance at several relevant events, including:
- BECTA 14-19 Diplomas Conference, 24 November 2008
  - JISC e-Learning Programme event, 3 March 2009
  - Lifelong Learning Symposium, 4 March 2009
  - Lifelong Learning Network National Forum, e-Systems event, 5 March 2009
  - UCAS Data Matters Conference, 25 February 2009
  - JISC Annual Conference, 24 March 2009
  - HudXCRi dissemination event, Leeds, 27 March 2009
  - UCAS Admissions Conference, 7-9 April 2009
  - Skills for Sustainable Communities LLN XCRi-CAP workshop, 21 April 2009
75. The project team has hosted three workshops for the wider LLN and regional community of learning providers, in November 2008, January 2009 and March 2009. The project manager briefed the LSC National Learning Directory team on 19 March 2009.
76. In addition information about the project has been disseminated via the project website at [http://www.XCRi.org/wiki/index.php/14-19\\_Area\\_Prospectus\\_XCRi\\_Aggregation\\_Pilot](http://www.XCRi.org/wiki/index.php/14-19_Area_Prospectus_XCRi_Aggregation_Pilot) and via the other XCRi community resources associated with the XCRi website. The project will also feature in the new website of the University of Nottingham, Centre for International e-Portfolio Development.
77. The project manager is currently seeking further ways to disseminate project outputs and outcomes to the 14-19 Area Prospectus community and to Lifelong Learning Networks via national networks.
78. The project team produced a range of Powerpoint presentations and material drawn from and upgraded from the earlier XCRi Implementation Models work. Specific materials we have produced include a briefing sheet showing the overview of the projects, a revised version of the original XIM mapping spreadsheet and various presentations.

## Part 8: Conclusions

---

### Aggregation

---

79. The project has achieved its central objective of aggregating data from Area Prospectuses into the Progression Pathways repository using XCRi-CAP 1.1. Three of the five Area Prospectus organisations have provided full data sets for inclusion in the repository, while the remaining two, Derbyshire and Northamptonshire, will be supplying shortly after the project has reported. Data has been supplied in XCRi-CAP 1.1 format, but very much 'as is', in other words without the extra quality control and manual intervention services that would result in high quality and consistency across the new data set. The project outputs therefore represent an indication that the aggregation process can be achieved, and that further work, as expected, would be needed to resolve the issues that have been identified.
80. The project itself did not address the service levels that might be involved in the continuous provision of the data. Agreement was reached between the Area Prospectuses and Aimhigher in the East Midlands that the former would supply school sixth form data for the Aimhigher in the East Midland's current geographical area, while college data would be obtained through the normal Aimhigher route. This would avoid duplication of data.

---

### Process models

---

81. Process models have been included at Appendix B, and published data feeds are noted under the Area Prospectus Case Studies section and will also be included in the project website<sup>14</sup>.

---

### Data review and issues

---

82. Data from the five Area Prospectuses was reviewed and mapped to XCRi-CAP 1.1 reasonably successfully. Data issues were noted and have been discussed in this report. It is clear from the review that useful 'headline' information at provider, venue, course, qualification and presentation level can be imported into the Progression Pathway repository. In addition entry requirement information and a small and variable amount of data for search fields can also be gathered.
83. In many cases extra textual descriptive information, such as information about course aims, career outcomes, progression, structure or assessment, was also available, but not mandated by the Progression Pathways system, though for example course aims could be used optionally. These extra data items could be included in future feeds and would make them more useful to other third party data collection agencies, such as the Hotcourses and LSC for the National Learning Directory. It is a useful spin-off of the project that Hotcourses is now working with New College Nottingham and Derby College on the mechanisms and processes required for using XCRi-CAP to supply data for the NLD.

---

<sup>14</sup> [http://www.XCRi.org/wiki/index.php/14-19\\_Area\\_Prospectus\\_XCRi\\_Aggregation\\_Pilot](http://www.XCRi.org/wiki/index.php/14-19_Area_Prospectus_XCRi_Aggregation_Pilot)

## Vocabularies

84. Work on vocabularies suggests the following relationships between XCRi-CAP entities in the 14-19 domain:

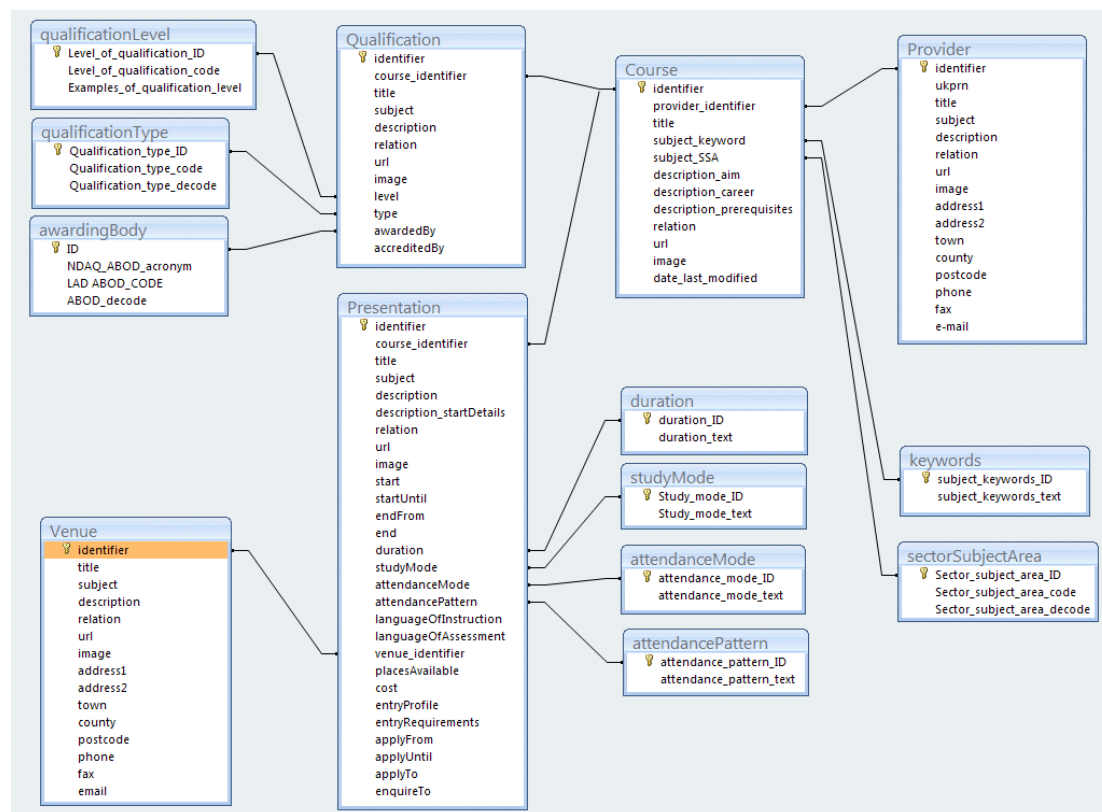


FIGURE 2: VOCABULARIES AND XCRi-CAP ENTITIES

85. It is unfortunate that the National Guidelines for setting up Area Prospectuses did not address information management and data entry or vocabulary standards at the time of issue in 2006. This has meant that without a clear national and authoritative steer on the use of vocabularies, each Area Prospectus had been free to develop its own systems without regard to similar systems in the same locality, region or the wider world. Subsequent attempts at data aggregation and interoperability are therefore much more difficult. These issues were covered in detail in the 1988 government-sponsored Coopers and Lybrand report “Coherence and Compatibility of Databases on Education and Training” (1988) and its subsequent ‘Using Learning Information’ standards initiative, whose lessons seem to have disappeared, despite commitment from all relevant government departments and national agencies.

### UKPRN as a provider identifier within course advertising information

86. There are significant problems, indicated above, in the use of the UKPRN for course advertising purposes. While perhaps 80% to 90% of providers were covered by the UKPRN, significant problems were encountered in respect of name changes, mergers, federations or partnerships, and national providers with a local presence.

### Aggregation and depth of information

87. There is no requirement for a large depth of data in this project’s aggregation into the Progression Pathway system, and Area Prospectuses prefer the aggregation of only headline information coupled with sign-posting back to their own websites. This may be a general principle. As data aggregation services move up a hierarchy of data collectors, the need for depth of data decreases. This is particularly true if the aggregation services collect URIs that give access to greater depth.

88. An implication of this principle is an increased importance for the use of URIs as identifiers, because this is the means that users will need to gain access to the greater depth of data. URIs should be recorded at as great a level of detail as possible. However, there is a caveat here, in that permanence of URIs may be more important than precision; it is worse to provide a broken link than a general link that works.
89. A recommended solution for the location of approved vocabularies in the sector is to use the Information Standards Board's (ISB) Vocabulary Management Service.

---

#### **Software vendors**

---

90. It is clear that vendors need significant notice in order to schedule this type of development, which may not yet be geared to release as part of a software package. Early engagement with vendors is essential and at least 6 months to a year should be the expectation for a level of implementation for pilot work, more time for full implementation.

---

#### **Recommendations for the development of XCRi-CAP**

---

91. The feeds for this project used an extended version of the XCRi-CAP terms schema that included specific vocabulary terms for the Progression Pathways system. The project does not recommend any of these terms for inclusion within XCRi-CAP.
92. It also included the recDateTime attribute for data maintenance, but experimentation with this attribute is as yet inconclusive, because only 'one time' data sets have been imported. Further work is needed to determine whether suggested data maintenance mechanisms will be fruitful.

---

#### **Cross boundary issues**

---

93. Cross boundary issues are a major challenge for the sector, and good practice in interoperability implementation can help to alleviate these issues. Good practice in this context means an emphasis on commonality of data in terms of vocabularies and data entry standards, and a number of points of guidance on good practice are suggested below. While XCRi-CAP implementation can give a basic level of interoperability between aggregated sets of data, having good quality and consistent data across the aggregated data set is essential for viability and credibility in the eyes of users.
94. XCRi-CAP can help to govern standards for vocabulary usage and can assist in the validation of data across multiple providers and feeds. Promotion of good practice in data entry and other issues in data quality will also help. Cross boundary agreements between local authorities or between 14-19 Area Prospectus organisations themselves, particularly in the wider context of the Common Application Process, will help to address issues such as frequency of update, technical synchronisation methods, data quality and the geographical areas to be involved.
95. There may be the possibility of setting up regional hubs in conjunction with other interested groups, such as Regional Development Agencies, Lifelong Learning Networks and other local authority groups. Precise details of regional hubs would depend on local circumstances.

---

#### **Good practice guidance for the 14-19 sector**

---

96. 14-19 Area Prospectus organisations are looking for clear, simple, and preferably national interoperability, guidelines that they can use, so that they can share data between themselves. This brief project has not had a specific focus on providing good practice guidance for the sector in terms of recording and managing courses information. However, it can offer the following guidance in respect of interoperability requirements as an aid to aggregation locally, regionally and nationally.
97. Making more use of existing national resources, such as the QAN service<sup>15</sup> and UKRLP<sup>16</sup>, would strengthen existing information services. It is recognised that more technical expertise

---

<sup>15</sup> [http://data.dcsf.gov.uk/qwsweb/\(S\(cuigqc455ezxqd554wwpjx55\)\)/Default.aspx](http://data.dcsf.gov.uk/qwsweb/(S(cuigqc455ezxqd554wwpjx55))/Default.aspx)

and resources may be required to implement such capabilities. For system-to-system interoperability the skills in the 14-19 sector are limited; although software vendors tend to have them, deployment may be constrained by their own commercial priorities and by other drivers in the sector.

98. The project has found that adherence to the following good practice principles would be of benefit and would increase the quality of aggregated data sets:
- Use plain English at all times and follow standard English grammatical rules.
  - Avoid free text data items, except for verbose descriptive elements, such as course aims, outcomes and the like.
  - Ensure that clear and explicit data entry standards are laid down and adhered to (refer to the National Learning Directory data entry standards<sup>17</sup> for an example), so that the organisation understands exactly the data content of its systems.
  - Record data in well-defined elements that are restricted to one data item only, without concatenating several items into one field. Avoid data entry errors by implementing these data items, and any that use prescribed vocabularies, by using reference data that can be entered using drop-down lists.
  - Wherever possible, use vocabularies that are nationally approved and used widely in existing course advertising information systems, particularly elsewhere in the 14-19 sector. Follow the guidelines in this report with respect to vocabulary implementation and tap into existing published resources rather than inventing new ones.
  - If starting a new courses data structure or revising an existing one, analyse the proposed new or changed structures to ensure that relationships between the data entities are well designed, appropriate and can be mapped to XCRi-CAP.
  - Whenever collecting any courses information, consider how best to make it available on the internet through a system to system connection, for example via HTTP or web service interfaces.

---

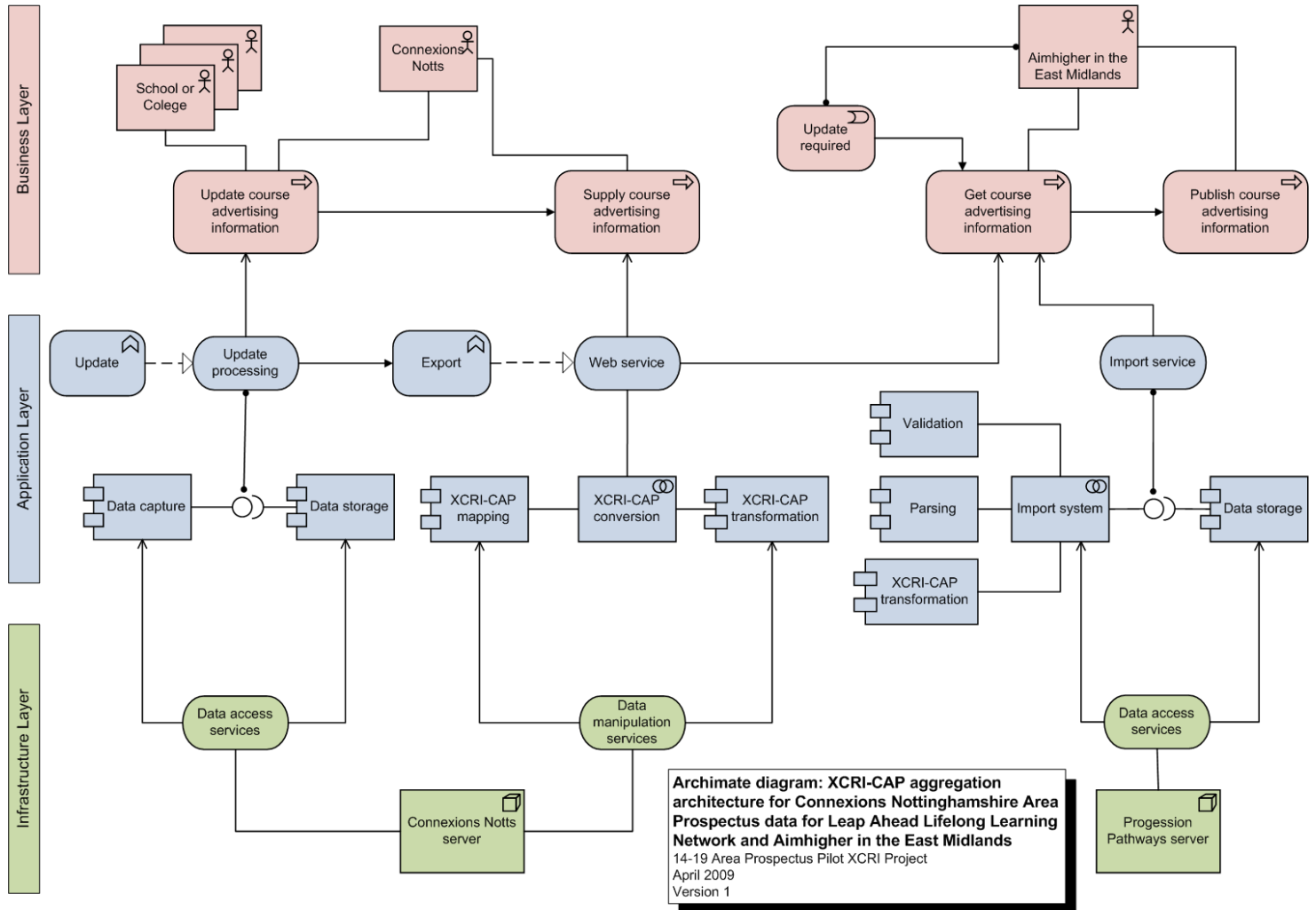
<sup>16</sup> <http://www.ukrlp.co.uk/>

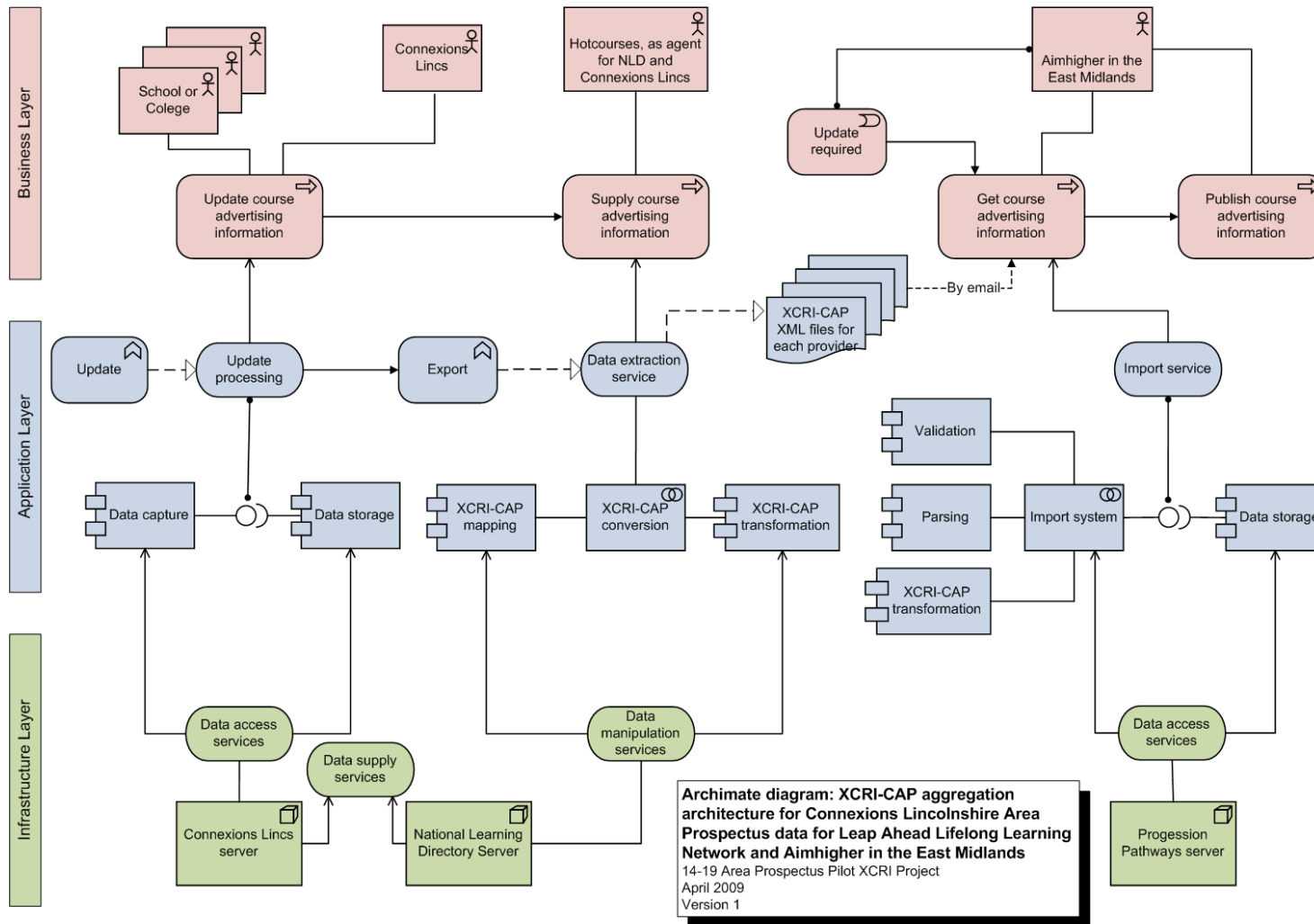
<sup>17</sup> op cit

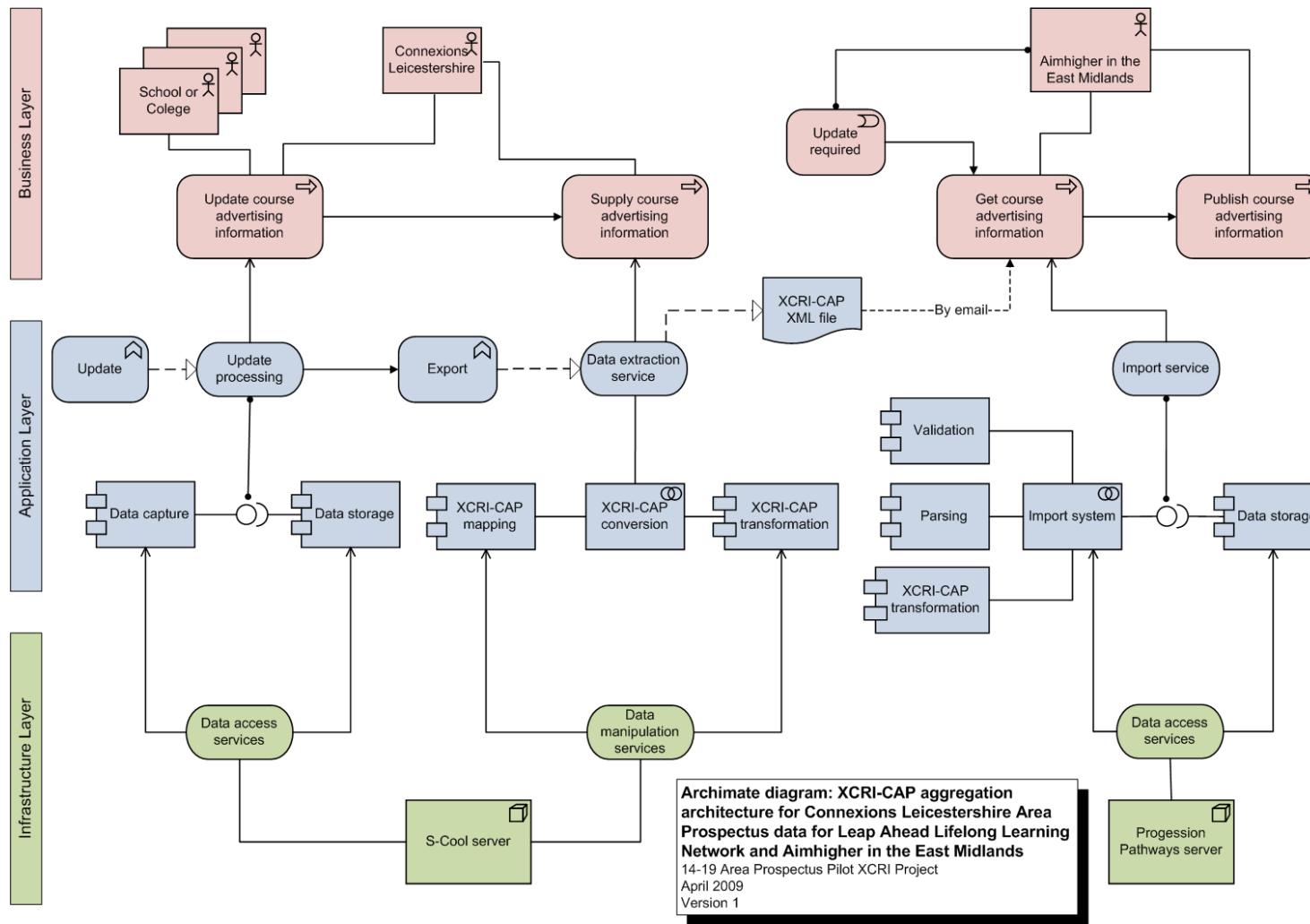
## Appendix A: Data issue log

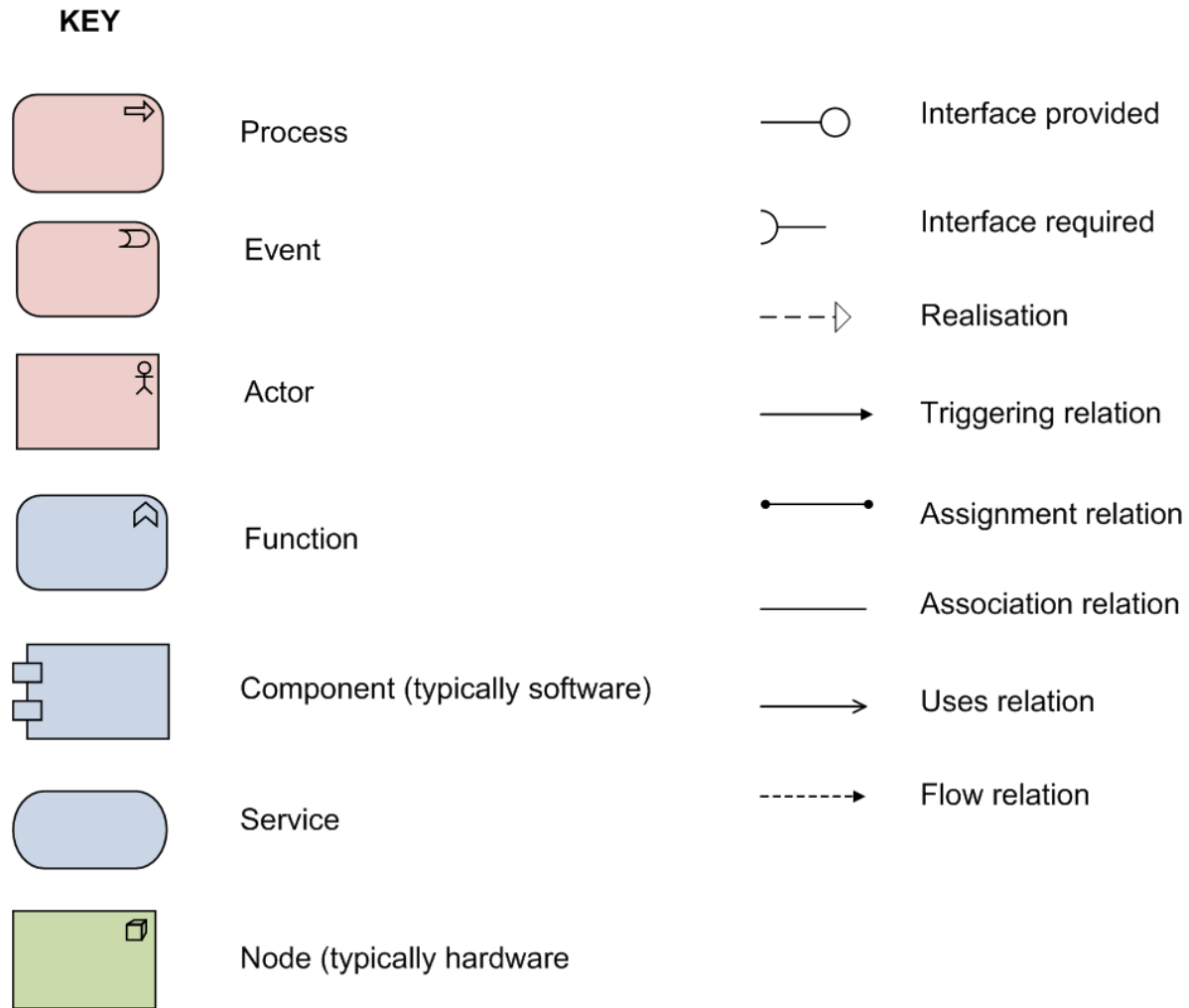
<i>Area Prospectus</i>	<i>Data issue</i>
Derbyshire	Ill-formed HTML and non-printable characters in descriptive text. Currently being worked on and new upload system will eliminate this.
Derbyshire	StartDate: in the input data, this is a month and year (e.g. 'September 2009'). XCRi requires an actual date, so this could be combined with Academic Year to form a pseudo-startDate, such as '2009-09-01'. However, this isn't needed for Progression Pathways, so it isn't a problem for this transfer; for Hotcourses it would be needed. For the moment I've commented out the Start elements, so that the file will validate.
Derbyshire; Leicestershire	StartDate: in the input data, this is a month only (e.g. 'September'). XCRi requires an actual date, so this could be combined with Academic Year to form a pseudo-startDate, such as '2009-09-01'. However, this isn't needed for Progression Pathways, so it isn't a problem for this transfer; for Hotcourses it would be needed.
Leicestershire	One-to-one map between course and presentation
Leicestershire	Data has been normalised as part of the transformation (sample input data had no provider records); the detailed provider data may not be needed for the feed, but at least a provider identifier should be supplied. The preferred longer term solution would be to supply the UKPRN, and for Progression Pathways to hold this; other aggregators should do the same, so that we have (in most cases) an unambiguous understanding of who the provider is.
Leicestershire	UKPRNs have been supplied manually. A local GUID identifier is present.
Leicestershire	URLs for courses are not comprehensive and for some providers may not be unique (they may point just to the provider's website), so cannot at present be used as course or presentation identifiers.
Leicestershire	Subjects: not held in the input data, so would require addition at Progression Pathways.
Leicestershire	Holds QAN number, so might be possible to use NDAQ data for qualification details
Lincs and Rutland	Files received from Hotcourses using nld.xsd for validation of NLD vocabs
Lincs and Rutland	Individual files for each provider, not one consolidated file
Northamptonshire	Unable to complete supply till end of May
Northamptonshire	Some UKPRNs available, but not comprehensively
Northamptonshire	All provider and venue information is at the same level (i.e. they are venues); will need to be addressed for final feed
Nottinghamshire	UKPRNs added manually by Leap Ahead; and matched with imported files
Nottinghamshire	Has QAN number, so qualification details only if taken from the NDAQ

## Appendix B: Process Models









## Appendix C: Institution mapping

This is an Excel spreadsheet showing mappings for each institution involved in the wider Leap Ahead LLN project, including the Area Prospectuses, plus the mapping to the Progression Pathways data format.

	NCN	Castle	Derby	NTU	Derbys	Leics	Lincs	Northants	Notts	XCRI-CAP 1.1	Data from All	Data from most	Confirm used in PP
<b>course</b>	identifier title description url qualification.title	identifier title description url qualification.type	identifier title description url qualification.type	identifier title description qualification.type qualification.description	identifier title description.ucasCode qualification.title qualification.level qualification.type qualification.description course.relation subject	identifier title url qualification.title qualification.level qualification.type qualification.description	identifier title qualification.title qualification.level qualification.type subject	identifier title qualification.title qualification.level qualification.type subject	identifier title qualification.title qualification.level qualification.type course.relation subject	identifier title description:topic url qualification.title qualification.level qualification.type qualification.description relation subject	Yes Yes	Yes Yes Yes Yes	Provider Course Course Overview Course Web- Award Qualification Award Level Award Type Award Notes Collaborative Similar to SS/
<b>presentation</b>	identifier  (use provided) duration entryRequirements.description venue.title venue.identifier start	*tbc  (use provided) duration venue.title venue.identifier	identifier attendanceMode cost.examination cost.tuition enquireTo duration entryRequirements.description venue.title venue.description	identifier attendanceMode cost enquireTo duration entryRequirements.description venue.title venue.description	identifier attendanceMode cost duration enquireTo duration entryRequirements.description venue.title venue.description description.startDetails	identifier attendanceMode duration enquireTo duration entryRequirements.description venue.title venue.description	identifier attendanceMode duration enquireTo duration entryRequirements.description venue.title venue.description	identifier attendanceMode duration enquireTo duration entryRequirements.description venue.title venue.description	presentation.identifier attendanceMode cost cost enquireTo duration entryRequirements.description venue.title venue.identifier venue.description start description:startDetails	Yes Yes	Yes Yes Yes Yes Yes Yes	Attendance Course Cost Course Cost Course Content Duration Amc ER_Entry Requirements Venue NOT USED Venue Comments NOT USED NOT USED	

\*Castle currently working on providing further presentation details

## Glossary of terms

<i>Term</i>	<i>Definition</i>
Aggregator	(i) an organisation that collects course advertising information; (ii) a service (for example a web service) that collects XCRi-CAP data feeds, see <a href="http://www.XCRi.org/aggregator">http://www.XCRi.org/aggregator</a>
Catalog	(i) Any store of courses information; (ii) root element of XCRi-CAP document
CiePD	University of Nottingham's Centre for International e-Portfolio Development (project lead institution).
Classification	Process whereby information content is matched to codes drawn from a controlled vocabulary (for example a thesaurus or structured list of values).
Interoperability	The ability of two or more systems or components to exchange information and to use the information that has been exchanged (IEEE)
ISB	Information Standards Board
LAD	Learning Aim Database
LLN	Lifelong Learning Network
MIAP	Managing Information Across Partners
QAN	Qualification Accreditation Number
QCA	Qualifications and Curriculum Authority
UKPRN	United Kingdom Provider Reference Number
UKRLP	United Kingdom Register of Learning Providers
URI	Uniform Resource Identifier. Used as a unique name for something on the internet.
VMS	Vocabulary Management Service
XCRi	eXchanging Course Related Information
XCRi r1.0	First official release of XCRi; covered all types of courses information, not just course advertising
XCRi-CAP	XCRi Course Advertising Profile