ITEMBANKING INFRASTRUCTURE:

A PROPOSAL FOR A DECOUPLED ARCHITECTURE

Mhairi McAlpine and Linn van der Zanden

Itembanking Infrastructure: A Proposal for a Decoupled Architecture

Mhairi McAlpine and Linn van der Zanden
Scottish Qualifications Authority
Computer Assisted Assessment
The Optima Building
58 Robertson Street
Glasgow
G2 8DQ

Mhairi.McAlpine@sqa.org.uk
Linn.vanderZanden@sqa.org.uk

Abstract

The paper aims to provide a comprehensive outline of the elements which make up an Itembanking system and through the use of basic workflows and diagrams create a visual of the overall system and user interaction. In particular it will provide an overview of the proposed Itembanking Infrastructure that SQA is currently developing, and steps which have been taken towards its realisation. Our aims in developing this are to promote more flexibility in assessment, improve on access, increase efficiency, cost-effective processes, enhancement of validity and reliability and improve possibilities for feedback and reporting. The functionality of an Itembanking system will be explored in light of the ways that institutions may use such a technical structure along with the challenges and issues surrounding its implementation.

We have divided the system into four main elements:

- The itembank itself which stores the items and facilitates searching
- The item production elements which generate items suitable for entry into the bank
- The test delivery elements which control the delivery, marking and reporting of the results
- The test generation elements which control items being selected from the bank and concatenated into tests.

The paper will focus on a 'reference' diagram which will provide an overview of the elements and associated software, the relationships between them and the overall interaction of the system. Within the four main elements, sub elements will be identified; including the storage of items, the generation of items, item description, item delivery, marking, result processing, item analysis and test construction. These will be explored with a view to defining the functionality of each element independently to allow autonomous development – fitting in with a standards based decoupled system. Existing projects and recommended standards in these areas will also be highlighted.

Role profiles and workflows are discussed in terms of how different users may interact with the system and roles may be transferred onto an electronic banking system. Future plans to establish user requirements for each component of the Itembanking Infrastructure will be discussed in the conclusion.

Introduction

With recent developments in educational technology and emerging standards, item banking is coming to the fore as an efficient and cost effective method of recycling expensively produced examination material. Although awarding bodies have used itembanking for decades and computerised systems for over 30 years; the advent of XML and a standardised way of describing assessment data encoded in the IMS QTI specification - together with the technological possibilities opened up by large scale distributed systems, have given people confidence in the stability of this approach. SQA in particular has been itembanking in a paper form since the 1960s. In the 1980s this was enhanced by the introduction of a computer database to hold the item usage data. The migration of current paper-based itembanks within SQA to a computerised, internationally recognised format is now progressing and SQA is looking toward developing a technical and organisational infrastructure to support the surrounding activities such as test construction, item analysis, delivery and marking. The underlying aim is to achieve more flexibility in assessment; improved access; more efficient cost effective processes; enhanced reliability and validity; advanced possibilities for feedback and reporting; and the provision of both paper and computer based assessment.

Itembanking may provide an intermediate step between fully on-line assessment and fully paper driven assessment as the items from the bank could always be printed and distributed in a traditional format, before CAA was introduced. It is also noted that this may ease the introduction of CAA, providing a bedrock from which CAA may be launched, with the major burden of this intermediate change being borne by the SQA rather than on candidates and centre staff. This may also prove a beneficial process for HE institutions. Although computerised itembanking is nothing new, existing systems tend to be monolithic entities with fixed functionality. Revamping the system features or adding additional is prohibitively expensive, however as we enter the brave new world of computerised assessment – ensuring that we take advantage of the increases in assessment approaches and validity enhancements that this will bring requires a flexible technological architecture.

This paper proposes a decoupled architecture, based on international standards and a webservices approach to the integration of functionality. The paper envisages the itembank itself as a unit made up of two components; a database which facilitates metadata storage, retrieval and search functionality, and a repository which facilitates the storage of items, resource files and manifest files. However, this bank is merely the datastore for a larger system which sits around the bank, feeds into it, interrogates it and exports from it.

Rationale for a Decoupled System

The concept of a decoupled, open source and standards based infrastructure is becoming increasingly popular[1]. For example, the JISC e-Learning Framework (http://www.elframework.org) adopts such a model in providing a networked service, typically using Web Services[2], referencing open specifications and standards that can be used to implement the service, and providing open-source implementation toolkits.

Diagram 1 suggests a potential architecture for a decoupled itembanking system. At its centre is the core itembank – comprised of a linked database and repository together with content unpackaging functionality. To the right are services associated with the generation and input of items; to the left are services associated with the export and delivery of items and at the bottom area are services associated with test construction.

Most existing systems conflate a number of these pieces of functionality into one software system, locking a user in to one provider and therefore requiring compromise to achieve the best overall fit. Decoupling the system in this way facilitates a "mix and match" approach. So long as each element inputs and outputs data in accordance with international standards and specifications, the integration of vendors, open source and custom built solutions are possible. In this system, elements may be replaced on a rolling basis with continual evaluation and thus providing the capability to keep up with new assessment approaches without periodic major upgrade.

---

[1] IBM (Leymann, Roller, and Schmidt, 2002) uses the service-oriented architecture (SOA) approach, which is the latest in a long series of attempts in software engineering that try to foster the reuse of software components and where programs are broken down into smaller programs through functional decomposition.

[2] An official standard for WSDL (Web Services Description Language) was released in 2001 by the World Wide Web Consortium. For a further description of web services see http://www.ariadne.ac.uk/issue29/gardner/.

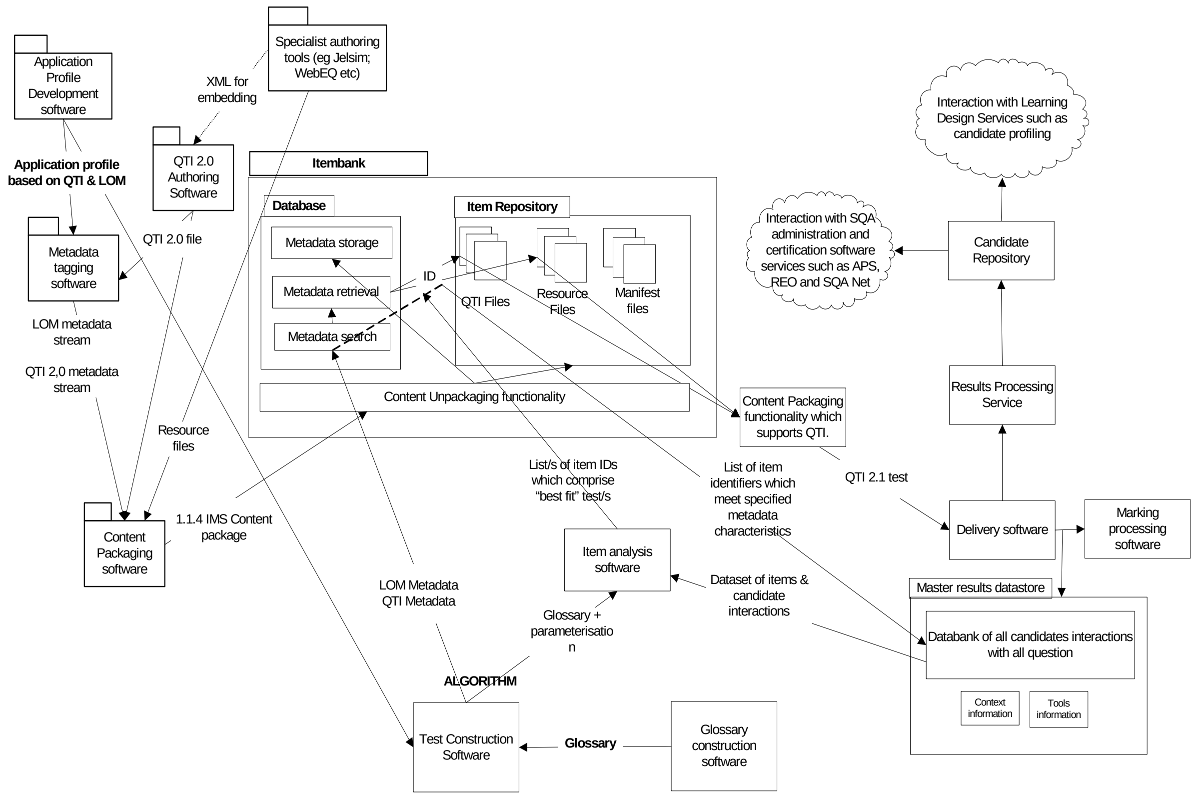

Diagram 1

Particular advantages of a decoupled architecture include:

- It can be easily adapted to accommodate a change in the model or workflow processes.
- A central team within an institution should be able to control the changes in workflow and software packages without major impact on the user.
- With a modular approach small chunks can be built and used immediately, with existing processes used to fill in the gaps until the next pieces are built.
- As the specifications grow and develop, pieces can be upgraded in line. Furthermore it can be developed cross-institutionally ensuring community involvement.
- As CAA 'beds-in' and people become more sophisticated in the way they use itembanking, additional demands will be placed on the system. A modular architecture allows for these demands to be slotted in at the appropriate points.

However the consequences and implications of adopting a modular approach include:

- The need to ensure standards compliance. Not just to the strict specifications, but where the specifications are loose to ensure that the manner implemented is in line with existing, emerging best practise.

- As a workable system will not be developed in one go, manual processes and pre-existing software systems will have to be built into the workflow. This may require developing additional functionality that will not be required once the full system is in place.
- Where a piece of the architecture is faulty, the possibilities for computer interaction with little user input may lead to difficulties with early detection of errors. System testing of each piece must be highly robust before it goes live.

What does Itembanking Entail?

We have divided the system into four main elements:

- <u>The itembank itself:</u> which stores the items in the repository and facilitates searching through the linked database
- <u>The item production elements:</u> which generate items suitable for entry into the bank
- <u>The test delivery elements:</u> which control the delivery, marking and reporting of the results
- <u>The test generation elements:</u> which control items being selected from the bank and concatenated into tests

The following section will identify the associated components to each element, a brief overview of the functionality all components, together with references to existing projects in this area and standards used.

Storage of Items

Within the proposed decoupled architecture the storage of the items is facilitated by two elements: the database and repository. The repository stores the QTI files, any associated resources from those files and the manifest files from the imported content packages as well incorporating content ‘un-packaging’ functionality in order to de-aggregate the elements of a submitted content package and deposit them into the repository and database as appropriate. The database stores both the metadata and QTI metadata, as well as housing the search and retrieval functionality.

The separation of the repository from the database allows for faster processing of search and retrieval requests. This is because the search is carried out on the metadata first before the item files are retrieved from the repository on the basis of the ID’s of the selected items. Another advantage is the afforded greater control of security and access permissions, as the metadata can be made more available than the actual item files in the repository.

Examples of existing systems incorporating itembank activities are the Hamlet itembank used by BTL among others; the TOIA system developed at the University of Strathclyde and the itembanking functionality incorporated into Perception from Questionmark. These are, however, integrated closed systems.

Generation of Items

The software associated with the generation of items for an itembank would depend on the content of these items. The IMS Question and Test Interoperability (QTI) specification is a standardised format for exchange of assessment item data –making it the most desirable form for storage and hence native authoring output. There may also be a requirement for additional specialist software associated with authoring items which have particular requirements – such as the inclusion of mathematical notation or multimedia elements.

Once the questions themselves have been authored, they must be tagged with standardised metadata to facilitate the search and retrieval processes. The international standard for describing learning objects is the IEEE LOM, however the full specification is very extensive. Application profiles can simplify data entry particularly where there are a large number of similar questions being entered at once, increasing both speed and accuracy of metadata entry. Such profiles could be generated by an additional piece of architecture.

The QTI Authoring Software allows questions to be created and exported in QTI 2.0 format and is complimented by Specialist Authoring Software. This develops parts of the items which cannot be directly encoded in QTI 2.0 but are instead either embedded or called from within the item. The Metadata Tagger enables metadata to be entered and attributes data to the items held in the bank, such as the author, the subject area of the question and the type of item. Application Profile Development Software facilitates the development of application profiles (or customised templates) based on the LOM, which pre-fill or restrict the entries that are allowed into the fields.3 The Content Packager packages together the elements of QTI 2.0 items according to the specifications of the IMS content packaging guidelines, facilitating import into any repository which recognises such standards.

Until recently no software capable of supporting the above activities has been available. The JISC-funded SPAID project (Young, MacNeill, Adams, McAlpine, 2005) produced a number of these as part of the Toolkit strand. The SPAID Metadata Tagger facilitates the generation and tagging of both LOM and QTI 2.0 Metadata. This application is customisable through the use of application profiles, while the SPAID Content Packager content packages assessment items in accordance with the QTI 2.0 specification. These are however very much at a prototype stage and further work is required to make them operational.

Delivery, Marking and Result Processing

Delivery, marking and result processing are post-itembank activities. Once tests are constructed they are passed into the delivery system, which then presents these assessments to the candidate and passes the input to the Marking Processing Software. The marked items are fed through the Result Processing Service which informs the Candidate repository and may interact with additional Services such as candidate profiling or administration and certification software. The Delivery system

[3] although it is recommended that the entire data schema is implemented even if many elements remain hidden to the end user, in order to ensure interoperability.

additionally submits all candidate interactions with items to the Master Results Databank for archiving.

The Delivery Software imports the assessments from the itembank in the form of a QTIv2.1 package. On completion of the assessment, the delivery software sends the recorded responses to the Marking Processing software. This consists of several elements which each facilitate the processing of different item types. There are three major approaches to mark processing; the first marks items entirely automatically, the second refers the items to a system where they are entirely human marked and the third uses a mixture of computer based and human marking. Question types which are best marked entirely by computer include Multiple Choice, Multiple Response and hotspot questions- each with their individual response processing template. Questions to be human marked, such as essays, would include a human readable mark scheme, while those using a mixed model- either human marked with a computer check or computer marked with human support- would use both.

The Result Processing Service software aggregates the marked items according to the requirements of the qualification, implementing the pass mark or grade boundaries which may be in force, while the Master Results Databank holds all the candidate interactions with items which are fed out from the delivery software – interacting with the item pools selected from the algorithms produced below.

New forms of marking are anticipated as CAA becomes more sophisticated and the flexibility to change or expand on response processing templates is desirable. The extracting and holding of candidate interactions would allow for more sophisticated analysis and process data, beyond processing scores of candidates on the items.

Test Construction

Test construction is the method by which items are concatenated to produce a test conforming to a particular specification. This has two aspects, firstly metadata searching to identify the questions which meet the descriptive metadata, then statistical analysis of the items and selecting those which meet specified parameters. A list of items comprising a suitable test are then sent back to the bank and a test file is exported for consumption by the delivery software.

Glossary Development Software would produce a glossary which defines the statistics to be used in the test construction system, providing the basis for the item analysis to take place, outputting a glossary in a standardised QTI format**.** Test Construction Software consumes application profiles together with the glossary to produce an algorithm, comprised of metadata (both LOM and QTI) and statistical terms which define the rules for test construction. These are then split – with the metadata first being sent to the bank, identifying an item pool that meets the defined criteria. The items are then matched with candidate interactions from the Master Results Databank, to produce a dataset which is sent to the item analysis software together with the statistical conditions from the algorithms. The Item Analysis Software runs the required analyses from the algorithm, identifying items from the pool that meet the conditions. Those items which meet the conditions of the algorithm are then passed back to the itembank for retrieval and packaging into tests.

The provision of the proposed architecture would increase reliability of item analysis, efficiency, and cost savings through reuse of items. The notion of the service orientated infrastructure includes the generation of usage data at run time. It is envisaged that no pre-testing would be necessary, but instead a small number of live items would be constantly tested. This would negate the requirement to hold static statistics for test generation purposes.

Although there is currently no existing Itembanking software which includes sophisticated test construction and usage data capture, there are a number of Item Analysis software packages available. The CATS project (Tulloch, 2006) is creating a toolkit to support automated assessment construction. It will build upon the outputs of two previous ELF projects – SPAID (Storage and Packaging of Assessment Item Data) and Discovery Plus (D+ - Brokerage for Deep and Distributed e-Learning Resources Discovery). The overall aim of the project is to create a toolkit of loosely-coupled web services which support the various tasks inherent to automated assessment construction e.g. searching for, retrieving and aggregating assessment items held in multiple item banks.

Roles and Workflows

This area is yet to be defined, in particular with regards to an overall system. Previous attempts to capture user roles and processes include the IBIS report (Sclater, 2004) and User requirements for the ultimate online system (Sclater and Howie, 2003). Each of the four elements discussed throughout the paper would have different users interacting with them.

When considering user roles, it should be noted that individuals may play one or more roles, both in the existing paper based system and in an online system. To facilitate adequate allocation of permissions however, it is necessary to exhaustively define roles – linking them with bank access and actions. These actions need to be clearly and exhaustively defined before linking with roles.

Established roles within our existing infrastructure include item writers (who write assessment items), qualifications staff (who oversee the assessment administration) and principal assessors (who oversee the assessment process), within those major roles however, there are a number of different functions that they and others perform. These roles may change or be separated into different functions within an operational itembanking system.

Workflow processing also requires further consideration, although some workflows on the generation of items have already been suggested in systems. One of the advantages of a decoupled architecture however, is that workflow processes may be changed as demands placed on the system change over time.

Where workflows are transferred onto an itembanking system, some elements of the workflow will be eliminated, other elements modified and new elements introduced. Workflows for producing a test from an itembanking system may include authoring and moderation of items, authoring and validation of metadata, and manual moderation of an automatically constructed test.

Business Process Execution Language (BPEL) may be used to orchestrate and manage the workflows in which partners in the process are identified and declared, the workflow is designed and defined, and business logic is added using BPEL Constructs before validation and deployment take place. This will produce a clear overview and relationship between the processes in each element of the Infrastructure.

Populating and Monitoring the Bank

Once the infrastructure is developed, the bank will need populating before operational use; consideration of the issues around population and monitoring is desirable at this stage to highlight issues which may impact on the software design.

Question writing procedures will have to be reconsidered to support a banked system. The most popular method of bank population involves content grids. However an extensive system as is being planned here may need a more sophisticated approach. Consideration should be given to parameterised questions, to reduce the impact of any potential security breaches and item exposure effects. This has implications for item analysis and may prove too sophisticated to handle in the short term.

A review is required to determine acceptable item exposure rates and extrapolate minimum bank size for each content area identified. This will help to inform the power of the search functions required and the space needed for storage.

One of the major advantages of banking items is the increased quality assurance it affords. Mechanisms for monitoring and regulating item exposure as well as procedures for discarding/quarantining over exposed items should be developed. Curricular drift (where a content area goes in or out of educational fashion) should be monitored and addressed through the dynamic item analysis including the use of item trend lines. Any unjustified deviation should be monitored and flagged to the relevant subject teams to ensure construct validity.

Problems may arise in banking where items are not independent for example where they have the same source paragraph, or refer to one another. Although complex banking can overcome these issues they require careful consideration. As a preliminary stage, these dependencies should be eliminated as much as possible, or the group should be banked as a single item.

Conclusion

This is an outline of a proposed system to facilitate sophisticated electronic itembanking using a webservices model to enable a decoupled system. Beyond the overview of services and their interactions, the precise definitions, requirements, interactions and data transfers for each of the elements must be scoped. The roles of users interacting with the system need to be defined and the workflow processes likely to be used must be identified.

An overview of the elements is given in this paper and can also be found in McAlpine et. al. (2006), while further work on defining the precise requirements for each of the elements as well as work on user roles and workflows is ongoing within SQA at the moment.

The Scottish Qualifications Authority appreciates the need to engage with the educational community from nursery to higher education to ensure that our technological structures are supportive of the forms of assessment that can best support learning and teaching. We see sophisticated itembanking structures as one component of a well-rounded modern assessment system and are keen to engage with all sectors in scoping, developing and evaluating a system which will be of benefit to all.

References

F. *Leymann*, D. *Roller*, and M.-T. *Schmidt*. Web services and business process management, IBM Systems Journal, Volume 41, Number 2, 2002.

Available at *http://www.research.ibm.com/journal/sj/412/leymann.html*

I. Tulloch, J. Everett, R. Young, M. Watson and R. Taylor. CATS- Constructing Assessments using Tools and Services, 2006.

N. Sclater. IBIS Item Banks Infrastructure Study. HEFCE, 2004.

N. Sclater and K. Howie. User requirements of the ultimate online system Computers & Education 40, 2003

R. Young, S. MacNeill, D. Adams, M. McAlpine. SPAID Final Report, 2005. Available at http://www.jisc.ac.uk/uploaded_documents/SPAIDfinalreport.doc