

Don't keep it under your hat

A Fez design description and case study

Christiaan Kortekaas

Lead Developer for UQ eScholarship, Australian Partnership for Sustainable Repositories

Senior Library Systems Programmer, Library Technology Service

University of Queensland

c.kortekaas@library.uq.edu.au

ABSTRACT	3
1. INTRODUCTION	4
1.1 UQ LIBRARY REPOSITORY HISTORY	4
1.2 AUSTRALIAN PARTNERSHIP FOR SUSTAINABLE REPOSITORIES	4
1.2.1 <i>The Early Days</i>	4
1.2.2 <i>Gateway yes, Sustainability no</i>	5
1.2.3 <i>Launched into eSpace</i>	5
1.3 ESCHOLARSHIPUQ OPEN SOURCE SOFTWARE OUTCOMES	6
2. FEZ DESIGN	6
2.1 CONTENT MODELS	6
2.2 SECURITY – FEZACML	7
2.2.1 <i>FezACML History</i>	7
2.2.2 <i>FezACML Design: Rules, Roles, Conditions</i>	7
2.2.2 <i>XACML the future?</i>	8
2.3 WORKFLOW	9
2.3.1 <i>Workflow History</i>	9
2.3.2 <i>Components</i>	9
2.4 SEARCH INDEX	12
2.5 SECURITY INDEX	12
2.6 CONTROLLED VOCABULARIES	13
2.7 CHALLENGES	13
3. CASE STUDY: UQ RESEARCH ASSESSMENT EXERCISE 2006	13
3.1 CONTENT MODEL AND DOCUMENT TEMPLATE CUSTOMISATION	14
3.2 MANAGING SECURITY AND ACCESS FOR RAE STAKEHOLDERS	16
3.3 CSV EXPORT	16
4. PREPARING FOR THE UQ RESEARCH QUALITY FRAMEWORK 2007	16
4.1 LESSONS LEARNT FROM RAE TRIAL RUN	16
4.1.1 <i>Data Entry Quality Problems</i>	16
4.1.2 <i>Non-Technical Solutions</i>	17
4.1.3 <i>Technical Solutions</i>	17
5. OTHER UQ USES FOR FEZ	17
5.1 MIGRATION FROM EXISTING SYSTEMS	17
5.1.1 <i>ePrints data migration</i>	17
5.1.2 <i>METS object import</i>	18
6. CONCLUSION.....	18
ACKNOWLEDGEMENT.....	19
REFERENCES	20
HYPERTEXT	20

Abstract

Fez is the University of Queensland's (UQ) new digital repository management and workflow system. Robust and scalable, the open source system is Fedora-based and provides easy access to electronic content like theses, book chapters, articles and research output. *Fez* is based on PHP and MySQL and provides a user-friendly front-end to Fedora 2.1.1+. Highly flexible and configurable, *Fez* is an advanced administration and content management tool for digital repositories.

This case study will show how *Fez* can be easily customised to manage a large, multi-user content entry project with complex security requirements. This capability is highlighted in UQ's current Research Assessment Exercise (RAE), which will see international expert panels assessing multiple research articles from the University's academic staff. Academic staff from different schools and centres must nominate their top three research works from the last five years to be deposited into *Fez* for review. Specific content models were developed to meet the reporting requirements of the RAE, including a 'citation view' facility, with links to either the object's digital object identifier or to a locally housed version of the file to grant reviewers full access to materials. To simplify metadata creation, *Fez* was integrated with data feeds from UQ central human resource systems to provide rich form controls based on cutting-edge AJAX technology. For example, an 'author suggest' control was developed to assist users by predicting an author's name as it is typed, based on an institutional academic staff list.

Fez is also being prepared for federated authentication and authorization based on the international standard eduPerson attributes, implemented with Shibboleth technology. This will assist future RAE processes by allowing external reviewers to authenticate using their own institutions central identity provider and *Fez* will be able to apply access control rules for these external reviewers based on their individual eduPerson attributes.

1. Introduction

The first part of this paper details how Fez [HREF33] was developed, explaining its design and the decisions made during development.

The latter half of the paper describes how Fez and Fedora [HREF32] software supported the UQ [HREF31] Research Assessment Exercise.

1.1 UQ library repository history

UQ has had some prior experience with digital repositories. UQ library staff implemented an ePrints repository in 2002. The library website describes ePrintsUQ as[HREF1]:

"a deposit collection of papers that showcases the research output of UQ academic staff and postgraduate students across a range of subjects and disciplines, both before and after peer-reviewed publication. The ePrintsUQ archive was set up in 2002 as part of a [Group of Eight](#) initiative to make Australian university scholarship and research collections more centralised and visible. The ePrintsUQ server will provide free, searchable access to this research and manage its long-term archiving."

ePrintsUQ has been a successful institutional repository with a (growing) count of 2956 publications [HREF2], which are primarily available publicly under open access. The system now has a strong user base of self-submitting authors and a steady influx of content. But the need for a repository to house research outputs other than publications helped drive the development of the Fez/Fedora project. Such outputs can include raw research data in the form of datasets, survey data, spreadsheets, multimedia, digital image collections and other forms of digital research output.

The UQ Library is also a member of the Australian Digital Theses (ADT) [HREF34] project. ADT repository software was specifically designed to manage digital versions of theses produced by postgraduate research students. It contains theses from participating Australian universities, with a federated search facility at the central ADT website [HREF35].

In the past, postgraduate students were encouraged to deposit their papers as an optional publishing method, but as of 2006, digital thesis submission has become mandatory.

1.2 Australian Partnership for Sustainable Repositories

1.2.1 The early days

The Australian Partnership for Sustainable Repositories (APSR) [HREF36] is funded by the Australian Federal Government's Department of Education, Science, and Training (DEST) [HREF37]. APSR was allocated \$2.4m over three years from January 2004 to December 2006 to develop a centre of excellence in digital collections management. APSR is a coordinated set of programs and projects that address strategic issues of digital sustainability, eResearch facilitation, repository management and system development.

The UQ library successfully joined the APSR project with the aim to "develop an integrated entry point to the vast body of under-reported research, initially within a single institution and then nationally" [HREF3].

Initially, the eScholarshipUQ testbed was only aiming to provide a gateway service to existing repositories. The gateway was accomplished reasonably early with the installation and customisation of the Public Knowledge Project (PKP) OAI-PMH Harvester software [HREF4] to launch UQ Research Finder [HREF5].

ePrintsUQ data was indexed by harvesting the ePrints OAI-PMH service provider. ADT was more difficult and required the creation of a custom ADT OAI-PMH service provider. UQ's other major relevant repository was the Research Master database, which contains metadata for every research publication in the university. A custom OAI service provider was also written to allow its indexing by UQ Research Finder.

1.2.2 Gateway yes, Sustainability no

After adding the ePrintsUQ, ADT and Research Master repositories to the UQ Research Finder gateway search service, the eScholarship team analysed the location of other research material that needed to be indexed. The majority of this content was hosted on academics' personal websites and school-maintained web servers. While these websites and documents could be linked to and harvested from the UQ Research Finder gateway, the maintenance of these links and the sustainability and preservation of the content could not be guaranteed.

The eScholarship team decided a central repository was needed to store this information, preserve it, increase its accessibility and control security where necessary. By this stage the library had been running ePrints for some time and while it solved an immediate need for a repository, it was not very customisable or flexible and any changes to the source code were difficult and time consuming.

The eScholarship team evaluated the other available software options for a repository (see "Software Options" matrix) [HREF6] and while Fedora was not a "turn-key" solution, it appeared to have the best designed framework for flexibility of the content model. It was also selected over DSpace for similar reasons to those given by the ARROW project [HREF29].

As the eScholarship team had strong programming skills and a clear understanding of what was required in a repository from its experience with ePrints, it decided to create a front-end interface to Fedora. The team started with a prototype with some PHP API code kindly provided by Elly Cramer from Cornell University (which also appears in the NSDL Fedora CMS example application [HREF8]).

After initial success with the prototype, the team went ahead with the full design and development of the Fez software.

1.2.3 Launched into eSpace

UQ launched its own Fez repository in January 2006 and branded it 'UQ eSpace'. The Fez software had been named eSpace during development, but it was changed following concern it would be confused with DSpace. It was renamed Fez as a nod to Fedora's hat theme and because it was a short and catchy moniker for the program.

Soon after the site launch, it was planned to migrate ePrintsUQ, ADT and other repositories' content into UQ eSpace over the following weeks. At the same time, the library was looking for a solution to manage the Research Assessment Exercise (RAE), which was due to start in April. In discussions with the UQ Office of Research and Postgraduate Studies (ORPS), it was decided the eSpace repository could manage this process. As the RAE required substantial eSpace customisation to

make content submission easy, the migration of ePrintsUQ, ADT and other repositories has been given lower priority and postponed until June-July 2006.

1.3 eScholarshipUQ open source software outcomes

UQ's APSR eScholarshipUQ testbed project reached an important milestone during 2005 - its Fez digital repository management software for the Fedora platform was released. eScholarshipUQ aimed to produce free, open source, cross-platform software any institution could deploy.

Fez 1.0 Beta was released in October 2005 on the open source software hub SourceForge.net under a General Public Licence (GPL) [HREF6]. Since then it has become more refined, with subsequent releases up to version 1.2 Beta made available in April 2006.

Throughout 2006 Fez has been further enhanced to meet a long list of functionality requirements, including customising and adding features to support the UQ Research Assessment Exercise (RAE).

eScholarshipUQ has also released an Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) service provider feed to METS object generator called OAI2METS and an ADT OAI service provider called ADTOAIProvider. Both of these software script packages are available free to download on the Fez Sourceforge.net website under GPL licences.

Fez will continue development until the end of the APSR funding grant period, finishing at the end of 2006. By this time all the major features are to be implemented and UQ's own repository will be heavily populated.

2. Fez Design

When assessing all the digital repository options Fedora stood out due to its potential for highly flexible customisation of content models and its solid XML backend object schema.

When designing Fez it was decided the front-end application to Fedora should also follow the customisable design philosophy. Rather than restrict Fez to hard set document templates, content models, security rule sets and workflow processes, the design vision was to allow a repository administrator to make small or even large modifications to these core subsystems easily on levels of scope from global down to datastream.

The challenge to the developers was to make a built-in administrative interface to manage these potentially complex core components that would be easy-to-use, without sacrificing functionality.

At the beginning of the Fez development the programming team had recently finished implementing and customising the Eventum helpdesk software from MySQL AB [HREF9]. The code structure and open-source framework of Eventum was carried over to Fez as the programming team were very impressed with its object oriented methodology and separation of business logic from presentation with the Smarty template system [HREF10]. Using this programming set of design standards the eScholarship programming team was able to rapidly design and implement Fez.

2.1 Content models

Fez is designed around the idea of the hierarchical layers of communities, collections and records. Records can belong to multiple collections and collections can belong to multiple communities. This

grouping of records fits in with the traditional organisation of repository object relationships and also allows for (multiple) inheritance of security and content controls. These relationships are mapped by the Fedora standard RELS-EXT datastream [HREF18].

One of the main complaints from ePrints administrators at UQ was the difficulty customising existing document templates and creating new ones for different document types with specific requirements.

Fez can handle any user definable content model by the mapping of XSDs (eXtensible Schema Definitions) [HREF38] against system “XSD Displays”. These XSD Displays map XSD XML elements against HTML form elements and other Fez controls to generate GUI content entry and update forms. Submitting a Fez user HTML form for an XSD Display will make Fez generate a Fedora Object XML [HREF11] instance which can then be saved into the Fedora database and indexed into the Fez index.

Up until the changes made to Fez for the RAE, there was one community template, one type of collection and many different types of records. The records could be of many different template types, all of which were defined and configured using the Fez administrative interfaces. The administrative interfaces were designed to be easy enough for non-programmers to use.

Most of the document types that come with Fez closely mirror ePrints document templates, as Fez was originally focused on migrating from ePrints.

2.2 Security – FezACML

2.2.1 FezACML history

During the early planning and design phases of Fez, the team realised temporary security control would be needed until Fedora 2.1 arrived with XACML support (at that stage an unknown timeframe).

FezACML was designed to solve the urgent security needs, with the potential for it to be replaced by XACML controls once they had been implemented in the core Fedora system.

2.2.2 FezACML design: rules, roles, conditions

FezACML is designed around a simple rule creation based on roles. Those roles are granted to a user when they satisfy one of the conditions in a role-condition pair.

The Fez common roles are:

- **Lister**: Whether the users will be able to view an object in listings and search results. If no viewer security rights are set on the object it will be assumed anyone can list or search the object.
- **Viewer**: If no viewer security rights are set on the object it will be assumed anyone can view the object. If viewer role is granted to a user they will also gain the Lister role by default.
- **Creator**: Grants the user access to create child objects under this object. Gaining this role grants the Viewer role as well.
- **Editor**: Grants the user access to edit an object and child objects inheriting security. Gaining this role grants the Viewer role as well.
- **Approver**: Grants the user access to publish an object from the submission buffer. Gaining this role grants the Viewer role as well.

These roles are granted when a user's attributes match against values in these possible attributes:

- Organisation Active Directory / LDAP group membership
- Organisation Active Directory / LDAP username
- Organisation Active Directory / LDAP distinguished name (Organisational Unit)
- Internal Fez groups
- Internal Fez usernames
- In AD/LDAP: whether the user is in the Active Directory / LDAP
- In Fez: whether the user is a user in the internal Fez user management system

With the release of Fez 1.2 the FezACML security model could also base rule conditions on the higher education standard eduPerson [HREF17] attributes, when the user logs in with Shibboleth [HREF26] authentication:

- **eduPersonTargetedID**: A unique, anonymous ID from the user's Identity Provider
- **eduPersonAffiliation**: eg staff, student, faculty, alum, affiliate, employee (multiple)
- **eduPersonScopedAffiliation**: eg student@uf.edu.au, staff@uf.edu.au
- **eduPersonPrimaryAffiliation**: eg staff, student, faculty, alum, affiliate, employee (single)
- **eduPersonOrgUnitDN**: Will allow access to this role if any text matches eg Library text would match for value of ou=Library, o=The University of Fez (multiple)
- **eduPersonPrimaryOrgUnitDN**: eg Library

The security rule-set for a Fez Fedora object is saved in its own FezACML datastream. FezACML is designed so security can be inherited from parent hierarchy objects. This allows rules for an entire community or collection to be set in one place.

If a collection or record belongs to more than one parent and is set to inherit security, a user can gain roles for that object from any of their authorisation matches in its parent object FezACML rules (thus achieving multiple security inheritance).

Security can be set at any level of hierarchy from community, to collection, to record, even down to the datastream level (for managed content eg images, PDFs). For managed content datastreams the FezACML rules are saved in a datastream with a "FezACML_" prefix and then the filename as the suffix, without the file extension. While this naming convention will probably stay with future versions of Fez the logic governing the relationship will be moved to be controlled by a REL-INT datastream instead.

FezACML will also contain other security control rules concerning non-role-authorisation aspects for objects in future versions of Fez. For example for an image datastream (eg a TIFF image), the record creator will be able to watermark and add copyright statements to the preview and web display versions and define specific web and preview image format dimensions. Record creators can specify separate security for access to the archival (originally submitted) version of the file for image mime-type datastreams.

2.2.2 XACML: the future?

Initial performance problems with the Fedora XACML implementation [HREF20] and more recently other development priorities have delayed a study into replacing FezACML with XACML. During such a study the XACML engine will have to be able to handle security inheritance and at least match the performance and functionality of the FezACML security model before a migration will be assessed.

Another challenge is that XACML is lacking a strong GUI tool for editing security rules. Some basic prototypes exist in various stages but none are user friendly, even for IT administrators.

The major benefit of moving to XACML is it has some standardisation and will (in theory) allow for cross-portability of Fedora objects between a Fez-Fedora repository and Fedora repositories governed by other front-ends. Although XACML is a form of standard, the uses for XACML attributes will vary between deployments. The ARROW [HREF19] project (with some participation from eScholarshipUQ) is in the process of creating a guideline for the use of XACML attributes. The finalisation of such a guideline would spur further action to move Fez to support XACML.

2.3 Workflow

2.3.1 Workflow history

In addition to having flexible document types and security, Fez is designed to have a flexible, easily customisable interface for controlling the flow of user GUI forms and document management.

As Fez was designed to follow standards where possible the development team assessed the Business Process Execution Language (BPEL) [HREF39] as a model for describing workflows. At that stage BPEL and its tools appeared very complicated and potentially very expensive depending on what tool to implement. The Fez programming team decided it would require too many development resources to implement a true BPEL conforming system so instead derived strategic ideas from the model and implemented a model that would meet requirements.

Fez workflow design will be briefly explained in the following sub-headings, however a more detailed documentation exists and will be added to a Fez Wiki in the near future.

2.3.2 Components

A design based on workflow states, triggers and behaviours was created which can be somewhat explained by the following diagram:

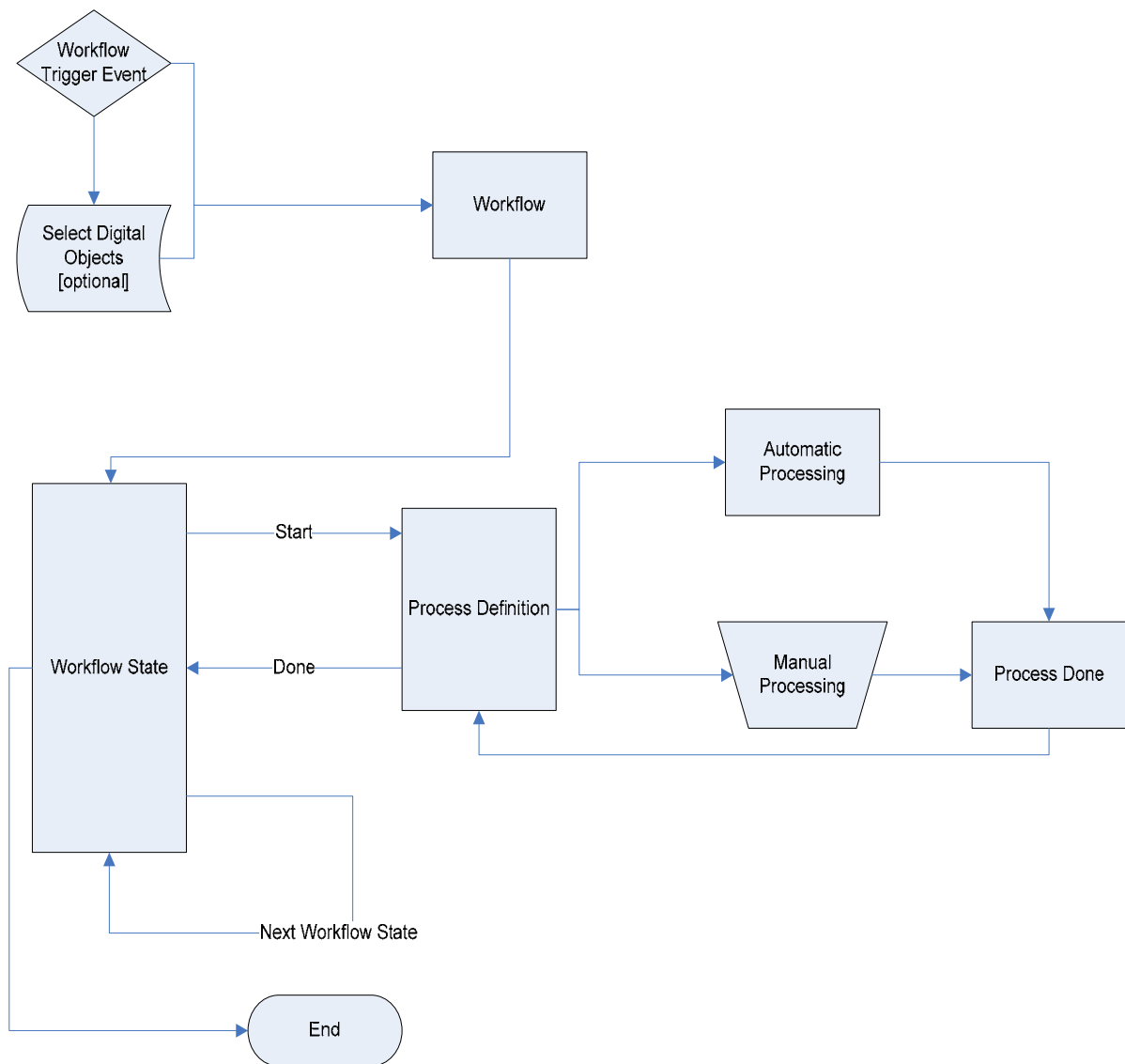


Figure 1: Fez Workflow Components Diagram.

Workflows are a state-based system so they have a process associated with each state. The process definition could be automatic or require manual completion. On completion of the workflow state process, the workflow moves to the next state. Each state has an associated action and can have access privilege restriction based on the user's FezACML role.

Workflows are triggered by events in the system. A set of default workflow triggers applies to all Fez objects, although custom workflows can be defined on any hierarchical level of object similarly to FezACML security.

A workflow has sets of behaviours (essentially web services) that occur when the item moves to each state – these might be automatic like generating thumbnails or creating preservation metadata (with JHOVE [HREF21]), or they might be manual such as entering and reviewing metadata. Each workflow behaviour is implemented as a PHP web service script, accepting passed variables as URL query string arguments and performing an action.

The workflow administration panels in Fez provide an easy way of defining and understanding this complex system. An example of a “create record” workflow admin interface can be seen in the following screenshot:

Don't Keep it Under Your Hat

Christiaan Kortekaas – The University of Queensland, Australia
Conference Paper for the Fedora Users Conference 2006, Virginia USA

Manage Workflow States for Workflow Template 'Create Record' ([New](#) | [back to Workflow Templates list](#))

Name:

Description:

Roles:

Options: ☐ Scripted ☐ Start State ☐ End State ☐ Transparent

Select Workflow Form:

Linked States:

Previous:

- None
- Select Collection
- Enter Metadata
- Publish
- Submit for Approval
- Preview
- Review Metadata

Next:

- None
- Select Collection
- Enter Metadata
- Publish
- Submit for Approval
- Preview
- Review Metadata

Note: Scripted states can only have one 'next' state.

Existing Workflow States:

All	Name	Roles	Start	End	Behaviour
<input type="checkbox"/>	Select Collection		yes		Select Community, Collection and Doc type
<input type="checkbox"/>	Enter Metadata				Enter Metadata
<input type="checkbox"/>	Publish	Approver		yes	Publish (auto)
<input type="checkbox"/>	Submit for Approval			yes	Unpublish (auto)
<input type="checkbox"/>	Preview				Preview
<input type="checkbox"/>	Review Metadata				Edit Metadata

State Links Diagram:
Click to edit nodes in the form above.

```

graph LR
    Start([Select Collection  
(start)]) --> Enter([Enter Metadata])
    Enter --> Preview([Preview])
    Enter --> Review([Review Metadata])
    Preview --> Publish([Publish  
(end|auto)])
    Preview --> Submit([Submit for Approval  
(end|auto)])
    Review --> Publish
    Review --> Submit
  
```

Figure 2: Workflow States for workflow template "Create Record" cropped screenshot.

The above screenshot shows the form for adding and editing workflow states. Of particular notice is the diagram at the bottom of the form. This diagram is generated automatically (with Graphviz [HREF22]) and is based on the above workflow state rules. The user can click on nodes in the

diagram to load the matching state link in the form for editing. There are three other admin panels for managing behaviours, workflow templates, and default workflow triggers.

For the UQ Fez repository, the default workflows have so far suited all collection scenarios and custom workflows have not been heavily used yet.

2.4 Search index

One of the early problems found with the Fedora software was its search service was very basic. The basic search service only allows searching across Dublin Core datastream elements, doesn't provide complicated Boolean logic, and has a limited API interface.

The Fez programming team also looked at the Fedora bundled Kowari Resource Index. It provides Dublin Core and also RELS-EXT searching, however the Fez design specification required the ability to search across any element in any datastream on an object. Also it was found the Kowari implementation at the time didn't fully support all the ITQL syntax specification and documentation for creating queries. It was found to be too basic with some example queries not running as documented.

Due to these limitations, the team decided to create its own Fez index and search system utilising the power of MySQL Full Text Boolean searching [HREF23].

In the XSD HTML matching editor the team added the ability to specify an XSD element is to be indexed. Whenever a Fez object is created or updated the workflow system will scan for XML element matches that are specified to be indexed and either replace or insert an index entry for the element.

A Fez "Search Key" administration panel was added which allows users to specify the common classifications for commonly searched against attributes (eg title, author, subject). This administration panel also configures what types of fields are shown on the advanced search form and what search keys are used in basic searches.

The XSD HTML matching form allows the assignment of a search key against a XSD element. This allows users to search against elements such as "Title" and get results based on not just Dublin Core Title but any XML element or attribute the content model designer wishes to associate with that search key.

Fez searches match against the index and as of Fez 1.2 can return results based on MySQL Boolean logic [HREF23] and are ordered by search relevancy by default.

One key requirement of the Fez index was that if some or all of the Fez MySQL index was corrupted or irretrievable it could be regenerated by a re-index of the backend Fedora repository. This is achieved with the Fez admin panels "Index Fedora Object into Fez" and "Reindex Fez Fedora Objects". The former scans for Fedora objects that have no mention in the Fez index at all, while the latter allows index repairing of selected Fedora objects that have at least some mention in the Fez index. The former also allows for Fedora objects to be ingested for repositories with existing objects by adding Fez administration technical and security metadata: FezMD, FezACML, and RELS-EXT if missing.

2.5 Security index

To increase the performance of FezACML security rules lookups a security index was created following a similar design to the Fez search index. Creating this index avoided the performance problem of looking up potentially many FezACML datastreams in parent objects, by clumping all the security rules for a specific object together.

After heavy content ingestion into UQ eSpace the performance of the security index degraded to a point where it is noticeable (while still useable). A redesign of the security index has been planned for the next release of Fez which will store the rule pairs separately from the rule to PID and datastream ID combination. This will make the security index far more efficient. For example in the production UQ eSpace security index there are roughly 250,000 rows however there are only about 256 different rules. Breaking the relationship between the rule and the object into separate tables will allow MySQL to speed up queries on the index dramatically.

2.6 Controlled vocabularies

Controlled Vocabularies (CVs) was one of the most requested features by the ePrintsUQ content administrator. She wanted to be able to classify objects by many different subject classifications, for example with collection specific subject codes. Some debate in the ePrints community has gone into the necessity for subject classifications as some see it as a thing of the past and a librarian's idea of how users search and browse for content. They argue that search engines (especially Google, Yahoo etc) are the only access point of importance for open access discovery.

There is strong statistical web server access log data backing up this view, however subject classification is also required to group documents for other non-discovery purposes such as reporting research content to DEST [HREF37] and providing meaningful classifications to external federated subject specific harvesters such as Picture Australia [HREF24].

CVs were designed so administrative users could easily define new classifications with the Fez CV admin panel. CVs can be flat lists or hierarchical such as the Australian Research Fields, Courses and Disciplines (the default example in the Fez 1.2 distribution).

Fez will potentially release other default CVs such as the Library of Congress [HREF41] and Dewey Decimal Classification (DDC) [HREF40] in the future, especially if any SQL schemas are contributed back from the Fez user community.

2.7 Challenges

Some of challenges facing Fez development are:

- Integration with new releases of Fedora and extension modules
- Integration with ARROW project work including VTLS open source components [HREF8] and interoperability with the Vital [HREF25] content model
- DSpace import and export
- Simplifying user interfaces, while maintaining functionality

3. Case Study: UQ Research Assessment Exercise 2006

The Research Assessment Exercise (RAE) 2006 is a test run of UQ's processes to handle the Research Quality Framework (RQF) which will begin in 2007.

According to the official description on the Australian Government Department of Education Science and Training website the aim of the RQF initiative is to:

develop the basis for an improved assessment of the quality and impact of publicly funded research and an effective process to achieve this. The Framework should:

- be transparent to government and taxpayers so they are better informed about the results of the public investment in research;
- ensure all publicly funded research agencies and research providers are encouraged to focus on the quality and relevance of their research; and
- avoid a high cost of implementation and imposing a high administration burden on research providers. [HREF12]

All Australian research agencies (including all Australian research universities like UQ) must submit to the RQF model as it forms the new basis for research funding. The United Kingdom has used a similar research funding model for many years.

The UQ Office of Research and Postgraduate Studies (ORPS) manage the RAE process and organise the schools, centres and research panels and reporting to DEST [HREF37]. UQ's Fez branded "UQ eSpace" was chosen to house and control the submission, workflow and access to RAE content in discussions between the UQ Library and ORPS.

Fifteen schools and centres were chosen to be part of the 2006 RAE and 'submitting authors' were selected against a set of research programs. School liaison librarians assisted submitting authors in creation of records in UQ eSpace.

3.1 Content model and document template customisation

A UQ eSpace community was created for the RAE with a member collection for each of the fifteen participating schools and centres.

To make the RAE content submission easier for data entry personnel, document templates were customised, new create and edit form controls were implemented and reporting features were added.

ORPS required the entry forms to be as simple as possible and make the input controls as dynamic and user friendly as possible. As each of the collections were essentially an organisational unit of UQ they were able to be mapped to data feeds of lists of each unit's academic staff from the central university human resources databases.

A customised version of the Fez collection XSD display was created (RQF 2006 Collection Display) which had extra fields for faculty, school/centre and research programs.

These new fields were added by a FOXML [HREF11] xmlContent XSD reference to a newly designed XSD for these three extra fields called "RQF2006MD". When a RAE collection is created based on this new template it will get a datastream based on this XSD reference called "RQF2006MD" to store this extra information. The custom collection form elements can be seen in Figure 3:

The screenshot shows the 'Edit Fedora RQF 2006 Collection Display Record' form in the UQ eSpace interface. The form is titled 'Edit Fedora RQF 2006 Collection Display Record' and includes several sections for data entry. At the top, there are navigation links like 'UQ HOME', 'SEARCH', 'MAPS', 'CONTACTS', and 'LIBRARY'. The user is logged in as 'Christiaan Kortekaas'. The form includes a 'Workflow Progress' section with a 'Review Metadata' button. Below this, there are buttons for 'Preview', 'Publish', 'Submit for Approval', and 'Reset Form'. The main form fields are: 'Title *' (School of Business), 'XSD Display Document Types' (a list of document types), 'Options for Research Programs drop down list 1' through '8' (each with a text input field), 'Faculty / Major Area *' (Business, Economics and Law), 'School / Division *' (Business), and 'Member of Communities *' (a list of communities). The 'XSD Display Document Types' list is expanded, showing options like 'Online Journal Article Version 1.0', 'Plain Document Version 1.0', 'RQF 2006 Book Version 1.0', 'RQF 2006 Book Chapter Version 1.0', 'RQF 2006 Conference Paper Version 1.0', 'RQF 2006 Journal Article Version 1.0', 'RQF 2006 Patent Version 1.0', and 'Seminar Paper Version 1.0'. The 'Options for Research Programs' lists are also expanded, showing options like 'Accounting', 'Business Information Systems', 'Finance', 'Innovation and Internationalisation', 'Management', 'Marketing', 'Org Communication', and an empty field.

Figure 3: RAE collection modifications in UQ eSpace - edit form, cropped screenshot.

ORPS decided for the 2006 RAE authors would be able to submit only five different types of content: books, book chapters, conference papers, journal articles and patents. Fez and (therefore UQ eSpace) already had templates for these five document types however ORPS required extra fields, different form validation and field display renaming and reordering.

Clones of the existing five templates were made and form validation and field reordering were easily customised for the RAE using the Fez document type administration GUI panels. The extra fields required were for:

- A list of submitting authors in the selected organisational unit of the parent collection
- A drop down list of the parent collection's research programs
- An "Author Suggest" lookup control of all the academic staff in the entire organisation using AJAX [HREF13] technology to predict and suggest authors as the user types input (similar to Google Suggest [HREF30]).
- A specific link (FOXML contentLocation datastream) to the work's Digital Object Identifier (DOI) if available, prefixing the dx.doi.org URL including the UQ ezProxy [HREF14] prefix.

These fields can be seen in action in Figure 4:

Research Program *	Business Information Systems
Submitting Author *	Bowen, Paul, 620
Author(s) 1 * (More Author(s) input boxes will appear as they are used up.)	Paul Bowen
Author ID:	Paul Bowen (620)
Author(s) 2	bowe
Author ID:	(none)
Title of paper *	
Link to Full Text (DOI) eg. 10.1189/jlb.0104026	http://ezproxy.library.uq.edu.au/login?url=http://dx.doi.org/

Figure 4: RAE extra fields cropped UQ eSpace screenshot.

ORPS also required customisation of the views of records the review panellists see in a “citation view”. This was achieved by reordering display elements in browse listings of collections and including extra fields for date, research program and authors, all in specified text formatting. This was implemented easily by manually editing the Smarty templates for listing records. However in future releases of Fez the development team is planning to implement an administration panel so repository administrators can dynamically specify how listings of records should be displayed, including a full citation view allowing designation of order and text font characteristics.

3.2 Managing security and access for RAE stakeholders

The RAE involved hundreds of internal staff and external review panel users with differing access requirements:

- The entire community of records must be hidden from public view, except to authorised staff and review panels, as at least some of the content (eg patents) is commercial in confidence.
- The internal data entry users (mostly liaison librarians) were created as Fez users with LDAP authentication to the UQ central Active Directory [HREF] linked to their UQ usernames.
- The School of Information Technology and Electronic Engineering (ITEE) [HREF16] requested for any of their staff members to be able to create and edit records in their collection. This was achieved by setting their collection editor role to accept anyone with their organisational unit (OU) in their Active Directory OU attribute.
- Submitting authors were grouped together in an Active Directory group and given view access to the RAE community so they could login and check their own records before they were reviewed by the RAE review panel.
- The review panel were given research program generic Fez user accounts with view access to only their reviewing collection.

3.3 CSV export

To aid reporting and tracking documents, ORPS required an Excel spreadsheet of the community and collection listings with selected details such as author ids and status of the documents. This was accomplished by writing a custom workflow to export CSV listings of metadata. A new ‘Export’ workflow trigger was created anticipating future export options from the repository.

4. Preparing for the UQ Research Quality Framework 2007

4.1 Lessons learnt from RAE trial run

4.1.1 Data entry quality problems

Some of the data entry personnel misunderstood:

Don't Keep it Under Your Hat

Christiaan Kortekaas – The University of Queensland, Australia
Conference Paper for the Fedora Users Conference 2006, Virginia USA

- the purpose and limits of the DOI field (not understanding the DOI prefix)
- how to enter authors correctly (one per author field rather than all in one textbox)
- the requirement to prefix external academic subscriber databases with the UQ ezProxy prefix

4.1.2 Non-technical solutions

Better training and explanation of the HTML forms and walkthroughs for data entry staff will need to be carried out for the next research assessment exercise.

4.1.3 Technical solutions

The user interface should have more validation based on regular expressions and make it impossible for the fields to be misused.

To solve the problem with some external links to subscriber articles missing ezProxy URLs a global external reference prefix for Fez was created. Fez administrators can now prefix every redirect datastream with a URL value.

These potential data quality management interfaces would have assisted the 2006 RAE:

- A global search and replace system to match XML elements based on regular expressions (probably using XQUERY). This could have checked all redirect datastream contentLocation XML elements and added the ezProxy prefix if it was missing.
- Global Version Control rollback interfaces based on regular expressions on elements. During one week of the RAE a JavaScript bug was in production which caused loss of some RFCD subject classifications on objects created and modified during this time. A global version control search based on regular expressions could find all the Dublin Core subject XML elements that had changed in a selected time period and rollback that part of the Dublin Core datastream to the pre-bug version.
- Customised advanced search filters to allow for user defined custom reporting. This would be useful for data entry staff to perform reports such as “Which submitting authors have not got all four of their works entered yet?”.
- Federated authentication. By the time of the 2007 RQF research assessment exercises Fez and the Australian shibboleth federation (MAMS Testbed Federation [HREF27]) will be mature enough to allow external review panellists to be able to login with their own institution’s login credentials and UQ eSpace will be able to apply access rules based on their released eduPerson attributes. This will simplify security management and make the login process easier for users. Fez 1.2 can already support this process however the Australian federations are currently still implementing their shibboleth Identity Providers [HREF28].

5. Other UQ uses for Fez

5.1 Migration from existing systems

Fez has been designed to be able to batch import records from ePrints and METS objects (eg from DSpace).

5.1.1 ePrints data migration

ePrints has an export command line tool “export_xml” perl script which will generate an XML file for an ePrints “archive”. This XML file contains metadata for each record in ePrints but does not contain the file attachments or URLs to the files (eg PDFs). However these file links are available in the ePrints OAI-PMH service provider.

When Fez finds an ePrints XML export file during batch import of a selected directory it will create a Fez Fedora object for each ePrints record and do an OAI-PMH getRecord lookup to the ePrints server to get the URL links for each file attachment, download them and add them to the new Fez object. Fez will also read the document type of each ePrints record and match those against Fez document types automatically. If no document type match is found it is created as the “Generic Document” type.

5.1.2 METS object import

When Fez finds a METS XML file in a batch import directory it will extract the Dublin Core records and add them to the Dublin Core datastream for a new Fez Fedora object.

Some discussion between APSR partners (some of which are DSpace core code base contributors) has begun to look at what would need to be done to provide object portability between DSpace and Fez. From initial discussions, DSpace appears to need some work on METS import and export but further testing will be carried out later in 2006.

6. Conclusion

Developing a front-end to Fedora has provided UQ with a highly flexible repository that will support the University’s needs now and into the future. It has assisted greatly with the complex RAE process. Both data entry and review panellists have found the system to be easy and simple to use, giving positive feedback. One UQ RAE 2006 review panellist from Rutgers University commented: “The system appears to work with remarkable ease, given the complexity of the task at hand.”

Performing a trial run of the process has allowed us to identify and iron out both technical and non-technical issues and will prepare us for the first live RQF assessment in 2007.

The Fez software is reaching maturity and will likely grow in deployment in the Fedora community, with the aim to achieve a Fez community of open source contributors. By the end of 2006 a long line of cutting edge features will be implemented making Fez a highly attractive repository option for institutions around the world.

The eScholarship team is excited about the prospects for growth of Fez’s open source community model and the international use of its hard work.

Acknowledgement

eScholarshipUQ is a test bed member project in the Australian Partnership for Sustainable Repositories, as funded by the Australian Government Department of Education, Science and Training (DEST [HREF37]).

The eScholarship team and UQ Information Technology Services received a grant from the Macquarie University Meta-Access Management System (MAMS) project to develop Shibboleth service provider functionality into Fez.

Thanks to the MAMS grant Fez has functionality to act as a “Where are you from?” (WAYF) gateway and base its authentication and authorisation FezACML rules on eduPerson attributes and federated identity delegation.

References

Hypertext

- HREF1. "ePrintsUQ About the Archive": <http://eprint.uq.edu.au/information.html>
- HREF2. "UQ Research Finder" – Count of ePrintsUQ publications as of 14th May 2006: <http://www.library.uq.edu.au/researchfinder/index.php>
- HREF3. "eScholarship@UQ - Original project submission": <http://www.library.uq.edu.au/escholarship/destproposal.pdf>
- HREF4. "PKP Open Archives Harvester (Overview)": <http://pkp.sfu.ca/pkp-harvester/>
- HREF5. "UQ Research Finder": <http://www.library.uq.edu.au/researchfinder/index.php>
- HREF6. "Free Open Source OAI-PMH 2.0 compliant Software": <http://www.library.uq.edu.au/escholarship/software.pdf>
- HREF7. "GNU General Public Licence – GNU Project – Free Software Foundation (FSF)": <http://www.gnu.org/copyleft/gpl.html>
- HREF8. "Fedora / Tools": <http://www.fedora.info/tools/>
- HREF9. "MySQL AB :: Eventum Issue / Bug Tracking System": <http://dev.mysql.com/downloads/other/eventum/>
- HREF10. "Smarty : Template Engine": <http://smarty.php.net/>
- HREF11. "Introduction to Fedora Object XML": <http://www.fedora.info/download/2.1.1/userdocs/digitalobjects/introFOXML.html>
- HREF12. "Research Quality Framework": http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/research_quality_framework/default.htm
- HREF13. "Ajax (programming) – Wikipedia, the free encyclopedia": <http://en.wikipedia.org/wiki/AJAX>
- HREF14. "ezProxy – Overview": <http://www.usefulutilities.com/support/overview.html>
- HREF15. "Windows Server 2003 Active Directory": <http://www.microsoft.com/windowsserver2003/technologies/directory/activedirectory/default.msp>
- HREF16. "School of Information Technology & Electrical Engineering - UQ": <http://www.itee.uq.edu.au/>
- HREF17. "EDUCAUSE | Net@EDU | eduPerson Object Class": <http://www.educause.edu/eduperson/>
- HREF18. "Fedora Digital Object Relationships": <http://www.fedora.info/download/2.1.1/userdocs/digitalobjects/introRelsExt.html>
- HREF19. "Australian Research Repositories to the World": <http://www.arrow.edu.au/>
- HREF20. "Fedora Release Notes": <http://www.fedora.info/download/2.1/userdocs/distribution/release-notes.html>
- HREF21. "JHOVE - JSTOR/Harvard Object Validation Environment": <http://hul.harvard.edu/jhove/>
- HREF22. "Graphviz": <http://www.graphviz.org/>
- HREF23. "MySQL 5.0 Reference Manual :: 12.7.1 Boolean Full-Text Searches": <http://dev.mysql.com/doc/refman/5.0/en/fulltext-boolean.html>
- HREF24. "PictureAustralia": <http://www.pictureaustralia.org/>
- HREF25. "VTLS Inc. | Product Information | VITAL": <http://www.vtls.com/Products/vital.shtml>
- HREF26. "Shibboleth Project – Internet2 Middleware": <http://shibboleth.internet2.edu/>
- HREF27. "MAMS Testbed Federation": <http://www.federation.org.au/FedManager/jsp/index.jsp>
- HREF28. "Shibboleth Technical Overview": <http://shibboleth.internet2.edu/shib-tech-intro.html>
- HREF29. "ARROW Targets: Institutional Repositories, Open-Source, and Web Services: 2.1 Repository foundation: Fedora": <http://ausweb.scu.edu.au/aw05/papers/refereed/treloar/paper.html>
- HREF30. "Google Suggest": <http://labs.google.com/suggest>
- HREF31. "The University of Queensland, Australia": <http://www.uq.edu.au/>
- HREF32. "Fedora": <http://www.fedora.info/>
- HREF33. "Sourceforge.net: Fez": <http://sourceforge.net/projects/fez/>
- HREF34. "Australian Digital Theses Program": <http://adt.library.uq.edu.au/>

HREF35. "Home/Search ADT": <http://adt.caul.edu.au/>
HREF36. "Australian Partnership for Sustainable Repositories": <http://www.apsr.edu.au/>
HREF37. "Welcome to the Department of Education, Science and Training": <http://www.dest.gov.au/>
HREF38. "XML Schema Definition": <http://en.wikipedia.org/wiki/Xsd>
HREF39. "Business Process Execution Language": <http://en.wikipedia.org/wiki/BPEL>
HREF40. "Dewey Decimal Classification": http://en.wikipedia.org/wiki/Dewey_Decimal_Classification
HREF41. "LIBRARY OF CONGRESS CLASSIFICATION OUTLINE":
<http://www.loc.gov/catdir/cpsolcco/lcco.html>