Mapping Open Access requirements to Pure and Fedora: a technical case study
Authors: Masud Khokhar \textsuperscript{a}, Adrian Albin-Clark \textsuperscript{b}

Abstract
Recent developments in the Open Access environment require institutions to develop their systems in a manner that supports external mandates while keeping internal workings of these systems intact. In the UK, EPrints is the most widely used repository system followed by DSpace with Fedora sitting in a distant 4\textsuperscript{th} place with 5 deployments\textsuperscript{1}. However, Fedora as a system is used by over 330 organisations worldwide\textsuperscript{2} and has gone through a major transformation over the course of last two years. We believe that Fedora will play a significant role in many institutions repository infrastructures in the UK in the future. This technical case study charts our initial work on mapping Open Access requirements to Fedora. A separate non-technical case study from the E2EOA project provide details on managing Open Access via EPrints\textsuperscript{3} and shows considerable time savings and reduction in errors when compared with doing similar work with spreadsheets.

Background
Lancaster University has a well-established centralised Open Access (OA) and Research Data Management (RDM) service. This allows us to provide a comprehensive service and a single point of contact for requirements that often jump between OA and RDM. We currently use Pure (Elsevier) for managing our OA publications and the long tail of research data. Pure serves the OA requirements well and provide a familiar interface to our academics (Figure 1).
However, Pure also has its issues which has made us think of alternative back-up systems to manage certain tasks. These are highlighted below for version 5.5 of Pure:

- Pure does not fully comply with all optional RIOXX metadata fields.
- It severely restricts the level of customisation that can be done to provide guidance on Open Access or to change the User Interface.
- Pure portal provides UUID based URL structure to records, hampering discoverability in search engines such as Google.
- The current APIs for Pure are restrictive in their usage.
- New developments can take a long time to accomplish.
- It is a proprietary system, which can be a reason of concern for some institutions.

For these reasons, we started investigating alternative systems that can provide a good secondary repository in addition to Pure. We already use EPrints which is currently synchronised from Pure but is only a backup mechanism as it does not serve our needs for storage of other outputs of the institution. After discussions and analysis, we have decided to use Hydra (based on Fedora) as this secondary system for OA.

**Why Fedora?**

Fedora is a powerful, modular, flexible, extensible open source repository system that allows organisations to manage, disseminate and preserve their digital content. It is not an Out of the Box solution (OOTB) but rather provides the underlying architecture that organisations can build repositories on. It supports complex semantic relationships between objects both within and outside of the repository. It can scale to millions of objects and provides interoperability with other applications and services.

As mentioned before, Fedora has gone through a major transformation in the past couple of years, moving from version 3 of the software to version 4. Technically speaking, this means that Fedora 4:

- Provides web resources (objects and datastreams) which have associated RDF properties
- Provides support for XML datastreams
- Removes support for FOXML objects and inline XML
- Provides hierarchy for objects and datastreams from a root node/resource to enhance performance
- Provides datastreams in a PairTree organisation at filesystem level and objects in a database which defaults to LevelDB
- Replaces object’s Persistent Identifiers (PIDs) with an internal ModeShape UUID. Additional identifiers such as DOIs can also be used and attached to objects as a property.
- Provides a pluggable security framework
- Has semantic functionality built directly in the repository architecture

Fedora 4 can also provide projection over content in an external system. This is especially useful from an RDM point of view when you have large data in an external system that you
don’t want to ingest. However, this does not allow any data enhancement, only discovery for these objects from Fedora point of view.

RIOXX and Pure

During the course of the project and discussions that we had at the Pure Interoperability working group, the question of Pure’s compatibility with RIOXX has come several times. RIOXX metadata application profile allows institutional repositories to comply with RCUK requirements on OA². Software components for supporting RIOXX already exist for EPrints and DSpace [4] but not much has been done on analysing Pure’s compatibility with RIOXX. Appendix A provides a current state of affairs by mapping RIOXX elements with Pure UI and OAI (DC) elements. This shows significant shortcomings in the compatibility between Pure’s OAI output and full compatibility with RIOXX. As far as we know, this is the first time such work has been done and we will also pass this on to Pure’s development team for further enhancements of Pure. This is further illustrated by the current listing of repositories providing RIOXX 2.0 metadata records via OAI-PMH where none of the 63 repositories listed is powered by the Pure software⁸.

The OAI endpoints used for this analysis are available here:
Lancaster University: http://pure.lancs.ac.uk/ws/oai (protected)
University of Edinburgh: http://pure.ed.ac.uk/ws/oai

RIOXX and Fedora

As Fedora does not force any fixed structure on its user, any custom schema is supported. For the purpose of this case study, we will work on an article in our repository available at: http://www.research.lancs.ac.uk/portal/en/publications/entrepreneurial-orientation-and-the-franchise-system(1780fa-c98a48e4-bbd0-c5160ebea171).html

In order to create a similar record in Fedora which is RIOXX compliant, we will create an RDF file which can then be posted directly to Fedora.

Filename: 6142578_rioxx.rdf

PREFIX ali: <http://ali.niso.org/2014/ali/1.0>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX rioxxterms: <http://www.rioxx.net/schema/v2.0/rioxxterms/>
<> ali:license_ref "http://www.rioxx.net/licenses/all-rights-reserved".
<> ali:start_date "2013".
<> dc:description "Purpose - The purpose of this paper is to understand the franchisor's perception of the role of entrepreneurial strategic orientation (EO)—innovative, risk-taking, and proactive actions—within the special case of franchised firms, given the opposing forces for standardisation/uniformity and system innovation/adaptation. Design/methodology/approach - A cross-sectional research design, involving a mail questionnaire survey, was employed to collect data from a sample of franchisors operating in the UK. The hypotheses specified in the study were tested using a path model. Findings - The results show that for franchise organisations, EO is positively related to performance. The recent meta-analysis conducted by Rauch et al. (2009) also demonstrated that the correlation of EO with performance is fairly large (r=0.242). This correlation is consistent with the significant value reported in our study of 0.234. In addition, franchise system support structures were found to be important in fostering EO within the organisation. Research limitations/implications - Future studies may consider including other dimensions of EO, notably competitive aggressiveness and autonomy, in exploring the relevance of EO to the franchise system. It would be interesting to also use a longitudinal analysis to examine the long-term effect of EO on the performance
of franchised firms. Originality/value - The study extends our knowledge of the EO-performance outcomes relationship to the franchising context where the role of EO is presently underexplored. This could be attributed to the fact that the franchise concept is built on standardisation and uniformity, notions that run counter to the flexible strategies involved in fostering EO.

This is then posted to Fedora. We are using Sufia’s Jetty Fedora in this example. The username and password for default Fedora instance is fedoraAdmin. The Content-Type element defines whether our content is in turtle or n-triple format.

Creating a new record in Fedora


The Content-Type format can be changed to determine the type of the output too, e.g. in the next two examples, we are requesting response in the format of “n-triples” and “turtle” respectively. To keep the report brief, we are providing these responses as Appendix B and Appendix C respectively.

Requesting n-triples from Fedora

$ curl -H "Accept: application/n-triples"
"http://q8983/fedora/rest/rioxx/ff/63/13/f9/ff6313f9-3e9e-4847-bc3e-bd65ae0a7090#u" -u fedoraAdmin:fedoraAdmin > rioxxGetNTriples

Requesting turtle from Fedora


In the above requests, we need to send the UUID of the object that was returned back to us when we initially created the object in the first place.
Conclusion

This technical case study highlights two important aspects. The first aspects showcase the issues that currently exist in Pure’s compliance with RIOXX. While Open Access can be managed with Pure easily, it is not yet interoperable enough to provide RIOXX compliant records via its OAI-PMH interface. Furthermore, the lack of enhanced customisation makes it difficult to change Pure’s OAI provider to incorporate more RIOXX compliant fields. The current levels of compliance will be provided to Elsevier under the umbrella of Pure’s interoperability working group and development will be asked for to make Pure RIOXX compliant.

The second aspect showcase how a RIOXX compliant record can be easily created in Fedora 4. As Fedora does not bound you to a fixed structure, it makes it easier to define your own fields. We have also developed a RIOXX playground that can help people understand how RIOXX compliant records can be created in Fedora with ease. This playground provides example RDF triples which can be posted to Fedora to create a new record. The playground is available at: http://dams.lancs.ac.uk:8080/rioxxpg/ and provides output both in JSON and RDF triples.

While Fedora provides a lot of flexibility, there is still a lot to be desired with it. The technical difficulty to implement Fedora is still steep and requires system administrator role in your team. The documentation is not always up-to-date and it can be difficult to understand who is responsible for what. One example of this is the difficulty we faced in implementing an OAI-PMH provider with Fedora 4. Implementing Fedora 4 with Tomcat (7 or 8) is fairly simple but neither the Fedora Labs version of OAI provider, nor any fork can be deployed with it. This has been raised with the core development team of Fedora.

We believe that Fedora’s future is bright but most institutions will opt for a near OOTB (there is no true OOTB solution for Fedora yet) solution for it. This could be in the form of Hydra or Islandora. Lancaster University is moving forward with Hydra and we are also part of the newly formed Hydra/Fedora UK consortium. We are working with other partners to make this consortium an Interest Group within the larger Hydra community, giving it an official status.

Competing interests

The authors have declared no competing interests.

Acknowledgement

This work was supported by Jisc (grant number 3267).

Authors

a Corresponding Author, Head of Digital Innovation, masud.khokhar@lancaster.ac.uk
b Digital Developer, a.albin-clark@lancaster.ac.uk

The Library, Lancaster University, Lancaster, LA1 4YH, United Kingdom
References

[4] RIOXX Metadata Application Profile
### Appendix A

<table>
<thead>
<tr>
<th>RIOXX element</th>
<th>Cardinality</th>
<th>Pure UI element</th>
<th>Pure OAI element/attribute</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ali:free_to_read</td>
<td>Zero or one</td>
<td></td>
<td></td>
<td>Pure does not provide this in a URI form, therefore not fully conforming to RIOXX requirement for this element</td>
</tr>
<tr>
<td>ali:license_ref</td>
<td>One or more</td>
<td>License to document</td>
<td>rights</td>
<td>Pure does not provide this in a URI form, therefore not fully conforming to RIOXX requirement for this element</td>
</tr>
<tr>
<td>dc:coverage</td>
<td>Zero or more</td>
<td></td>
<td></td>
<td>Only exists for datasets in Pure</td>
</tr>
<tr>
<td>dc:description</td>
<td>Zero or more</td>
<td>Abstract</td>
<td>description</td>
<td></td>
</tr>
<tr>
<td>dc:format</td>
<td>Zero or more</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dc:identifier</td>
<td>Exactly one</td>
<td>UUID</td>
<td>identifier</td>
<td>The identifier element is repeatable in the metadata section but seems to be unique in the header section</td>
</tr>
<tr>
<td>dc:language</td>
<td>One or more</td>
<td>Original language</td>
<td>xml:lang attribute on title and description</td>
<td>Wrong interpretation of RIOXX element as it represents language of title only. The attribute on description seems to be statically fixed on ‘eng’. Not all languages are represented, e.g. Swedish or Portugese are represented by ‘und’.</td>
</tr>
<tr>
<td>dc:publisher</td>
<td>Zero or more</td>
<td>Publisher</td>
<td>publisher</td>
<td></td>
</tr>
<tr>
<td>dc:relation</td>
<td>Zero or more</td>
<td>Relations</td>
<td></td>
<td>None of the relations are exposed via OAI-PMH interface</td>
</tr>
<tr>
<td>dc:source</td>
<td>Zero or more</td>
<td>Journal</td>
<td>source</td>
<td>Not a direct match based on description of RIOXX element</td>
</tr>
<tr>
<td>dc:subject</td>
<td>Zero or more</td>
<td>Subject areas</td>
<td>subject</td>
<td></td>
</tr>
<tr>
<td>dc:title</td>
<td>Exactly one</td>
<td>Title</td>
<td>title</td>
<td></td>
</tr>
<tr>
<td>dcterms:dateAccepted</td>
<td>Exactly one</td>
<td>Publication state</td>
<td></td>
<td>No dates other than current version exposed, which could be published version. Exposed in “date” element.</td>
</tr>
<tr>
<td>rioxxterms:apc</td>
<td>Zero or more</td>
<td>APC paid</td>
<td>Authors with ORCID IDs do not explicitly add an id attribute as per RIOXX recommendation.</td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------</td>
<td>-----------------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>rioxxterms:author</td>
<td>One or more</td>
<td>Authors</td>
<td>creator</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Authors can have different roles in Pure but our Pure instance does not provide a Contributor role for a journal article or a Book chapter. However, Edinburgh’s pure instance does release contributor as an element for contributionToPeriodical type.</td>
<td></td>
</tr>
<tr>
<td>rioxxterms:contributor</td>
<td>Zero or more</td>
<td>contributor</td>
<td>Authors can have different roles in Pure but our Pure instance does not provide a Contributor role for a journal article or a Book chapter. However, Edinburgh’s pure instance does release contributor as an element for contributionToPeriodical type.</td>
<td></td>
</tr>
<tr>
<td>rioxxterms:project</td>
<td>One or more</td>
<td>Project</td>
<td>Edinburgh’s Pure exposes project information but that is also not RIOXX compliant</td>
<td></td>
</tr>
<tr>
<td>rioxxterms:publication_date</td>
<td>Zero or more</td>
<td>Publication state (Published)</td>
<td>No dates other than current version exposed, which may or may not be published version. Exposed in “date” element.</td>
<td></td>
</tr>
<tr>
<td>rioxxterms:type</td>
<td>One or more</td>
<td>Implicit as type is selected to begin with</td>
<td>type</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>The values in Pure does not conform to the RIOXX proposed controlled list of types.</td>
<td></td>
</tr>
<tr>
<td>rioxxterms:version</td>
<td>Exactly one</td>
<td>Document version</td>
<td>Controlled list in UI does not conform to RIOXX proposed controlled list of versions.</td>
<td></td>
</tr>
<tr>
<td>rioxxterms:version_of_record</td>
<td>Zero or one</td>
<td>Final published version</td>
<td>identifier</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Not RIOXX compliant as multiple identifier elements are often present.</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

n-triples from Fedora

Request

$ curl -H "Accept: application/n-triples" 
"http://q:8983/fedora/rest/rioxx/ff/63/13/f9/ff6313f9-3e9e-4847-bc3e-bd65ae0a7090" -u fedoraAdmin:fedoraAdmin > rioxxGetNTriples

Response


<http://purl.org/dc/elements/1.1/publisher> "Emerald Group Publishing".

"bypassAdmin""^^<http://www.w3.org/2001/XMLSchema#string> .

"Journal\Article/Review""^^<http://www.w3.org/2001/XMLSchema#string> .

"all.niso.org/2014/all/1.0license_ref" "http://www.rioxx.net/licenses/all-rights-reserved""^^<http://www.w3.org/2001/XMLSchema#string> .
Purpose - The purpose of this paper is to understand the franchisor's perception of the role of entrepreneurial strategic orientation (EO) within the special case of franchised firms, given the opposing forces for standardisation/uniformity and system innovation/adaptation. Design/methodology/approach - A cross-sectional research design, involving a mail questionnaire survey, was employed to collect data from a sample of franchisors operating in the UK. The hypotheses specified in the study were tested using a path model. Findings - The results show that for franchise organisations, EO is positively related to performance. The recent meta-analysis conducted by Rauch et al. (2009) also demonstrated that the correlation of EO with performance is fairly large (r=0.242). This was found to be important in fostering EO within the organisation. Research limitations/implications - Future studies may consider including other dimensions of EO, notably competitive aggressiveness and autonomy, in exploring the relevance of EO to the franchising system. It would be interesting to also use a longitudinal analysis to examine the long-term effect of EO on the performance of franchised firms. Originality/value - The study extends our knowledge of the EO-performance relationship to the franchising context where the role of EO is presently underexplored. This could be attributed to the fact that the franchise concept is built on standardisation and uniformity, notions that run counter to the flexible strategies involved in fostering EO.

Future studies may consider including other dimensions of EO, notably competitive aggressiveness and autonomy, in exploring the relevance of EO to the franchising system. It would be interesting to also use a longitudinal analysis to examine the long-term effect of EO on the performance of franchised firms. Originality/value - The study extends our knowledge of the EO-performance relationship to the franchising context where the role of EO is presently underexplored. This could be attributed to the fact that the franchise concept is built on standardisation and uniformity, notions that run counter to the flexible strategies involved in fostering EO.
Appendix C

Turtle RDF from Fedora

Request


Response

@prefix premis: <http://www.loc.gov/premis/rdf/v1#> .
@prefix rixxterms: <http://www.rioxx.net/schema/v2.0/rioxxterms/> .
@prefix nt: <http://www.jcp.org/jcr/nt/1.0> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix ns004: <info:fedora/fedora-system:def/relations-external#> .
@prefix ns003: <http://purl.org/dc/terms/> .
@prefix ns002: <http://scholarsphere.psu.edu/ns#> .
@prefix ns001: <info:fedora/fedora-system:def/model#> .
@prefix xsi: <http://www.w3.org/2001/XMLSchema-instance> .
@prefix mode: <http://www.modeshape.org/1.0> .
@prefix ns007: <http://www.w3.org/ns/auth/acl#> .
@prefix ns006: <http://id.loc.gov.vocabularies/relations/> .
@prefix ns005: <http://pcdm.org/models#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix fedoraconfig: <http://fedora.info/definitions/v4/config#> .
@prefix mix: <http://www.jcp.org/jcr/mix/1.0> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix fedora: <http://fedora.info/definitions/v4/repository#> .
@prefix ebucore: <http://www.ebu.ch/metadata/ontologies/ebucore/ebucore#> .
@prefix ldp: <http://www.w3.org/ns/ldp#> .
@prefix xs: <http://www.w3.org/2001/XMLSchema> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .


fedora:lastModifiedBy "bypassAdmin"^^<http://www.w3.org/2001/XMLSchema#string> ;

rioxxterms:type ""Journal Article/Review""^^<http://www.w3.org/2001/XMLSchema#string> ;

<http://ali.niso.org/2014/ali/1.0license_ref> "http://www.rioxx.net/licenses/all-rights-reserved"^^<http://www.w3.org/2001/XMLSchema#string> ;

dc:publisher "Emerald Group Publishing Limited"^^<http://www.w3.org/2001/XMLSchema#string> ;


fedora:createdBy "bypassAdmin"^^<http://www.w3.org/2001/XMLSchema#string> ;

fedora:created "2015-09-14T10:30:00.286Z"^^<http://www.w3.org/2001/XMLSchema#dateTime> ;
Purpose - The purpose of this paper is to understand the franchisor's perception of the role of entrepreneurial strategic orientation (EO)-innovative, risk-taking, and proactive actions-within the special case of franchised firms, given the opposing forces for standardisation/uniformity and system innovation/adaptation. Design/methodology/approach - A cross-sectional research design, involving a mail questionnaire survey, was employed to collect data from a sample of franchisors operating in the UK. The hypotheses specified in the study were tested using a path model. Findings - The results show that for franchise organisations, EO is positively related to performance. The recent meta-analysis conducted by Rauch et al. (2009) also demonstrated that the correlation of EO with performance is fairly large (r=0.242). This correlation is consistent with the significant value reported in our study of 0.234. In addition, franchise system support structures were found to be important in fostering EO within the organisation. Research limitations/implications - Future studies may consider including other dimensions of EO, notably competitive aggressiveness and autonomy, in exploring the relevance of EO to the franchise system. It would be interesting to also use a longitudinal analysis to examine the long-term effect of EO on the performance of franchised firms. Originality/value - The study extends our knowledge of the EO-performance outcomes relationship to the franchising context where the role of EO is presently underexplored. This could be attributed to the fact that the franchise concept is built on standardisation and uniformity, notions that run counter to the flexible strategies involved in fostering EO.

fedorawritable "true"