Fedora at the Wellcome Library, progress and work to date

DPC Digital Repositories Meeting

5 July 2006

Dave Thompson
Web Archiving Project Officer
Wellcome Library

wellcometrust

Wellcome Library and its Strategic Plan

The Library is assuming that born digital materials will be a feature of its collections in the future

'…to maintain its position as a major research library in an increasingly digital environment'

The Digital Curation in Action project is our first attempt to turn this strategy into action

wellcometrust

Dealing with digital stuff

The Library will store born digital material in a repository

Assuming that descriptive metadata will be generated and stored using existing Library systems and not using Fedora.

Data exchange by harvesting, mechanics remain unclear

Looking to normalisation, eg XML, of textual material for economic preservability and repurposeability

Metadata is a key management tool for the future and the automatic creation of this metadata is essential if life cycle management is to be economically sustainable

wellcometrust

Fedora as a test vehicle

The use of a repository is essential if material is to have a secure home and is to be permanently associated with its management metadata

Fedora provides a low entry cost approach to repository experimentation, it's flexible, well supported, expansible and with low hardware requirements, we don't have resources to evaluate many systems

We have created a dummy body of 'typical' material that we can use as standardised testing objects, concentrating at this stage on Simple and Simple/Complex objects

The Library has implemented Fedora as a test/experimental repository– Fedora V2.0

Complex relationships and Fedora

Wellcome Library and National Library of Wales collaborated on a Relationship Builder project

Fedora has in-built support for hierarchical relationships that are essential to collection building, but…

No front end access – so we built a web delivered tool that provides a simple user interface to underlying Fedora functionality

Allows construction of hierarchical relationships between existing Fedora objects, expressed in the METS as RDF

No means as yet to re-create/retrieve by relationship but the metadata exists

Building complex relationships with Fedora

Fedora relationship builder example

The library is developing an e-phemera 'spam' collection, the collection has the identifier dcia:44

The individual email '*Antidote found in crocodiles*' - identifier dcia:55 - belongs to this collection

```
<rdf:RDF xmlns:fedora="info:fedora/fedora-system:def/relations-external#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <rdf:Description rdf:about="info:fedora/dcia:55">

    <fedora:isMemberOfCollection rdf:resource="info:fedora/dcia:44"/>

  </rdf:Description>

</rdf:RDF>
```
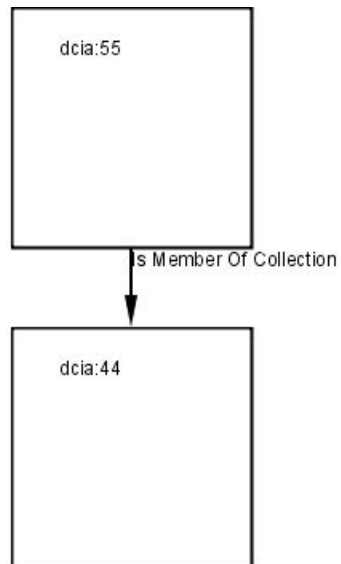
Fedora relationship builder

dcia:55

s Member Of Collection

dcia:44

Relationships can be
viewed graphically, the
hope is that a fully
graphical drag and drop
interface can be built

Generating preservation metadata

Wellcome Library has worked with Systems Simulation Ltd to integrate the National Library of New Zealand Preservation Metadata Extract Tool into the Fedora workflow

The metadata extract tool can extract technical / preservation metadata from commonly used 'office', still image and internet format file types and output it as XML

A proof of concept development, not a production tool but production requires little additional development

Technical metadata, the output of the extract tool, is stored in Fedora as managed content

Issues with/around emergence of PREMIS as presmet 'standard'

Our first collection

'Medical' spam emails of the sort we all get, 21st century e-phemera

Easy to use, no relationship issues, an excellent introduction to life cycle management of email

Useful sandbox/test object, can test normalisation approaches such as the use of Xena as well as ingest workflows and processes

Test emails are both Simple and Simple/Complex objects and have been normalised to XML and text

No attempt to incorporate header data or any log data

Basic DC part of the METS but more interested in technical metadata

e-phemera examples in Fedora

Collection issues broader than just a repository

Need to design workflows for identifying and bringing material into collections

Digital material still needs selection, A&D, description etc

Assigning 'proper' responsibilities for material

Need to re-write policies and manuals

Need to re-word donation/loan/gift agreements

Need to educate users in using new materials

Ideally need to work more closely with creators/donors

Key lessons so far

Collaboration is a viable means to developing approaches and solutions

Everything must be economically sustainable

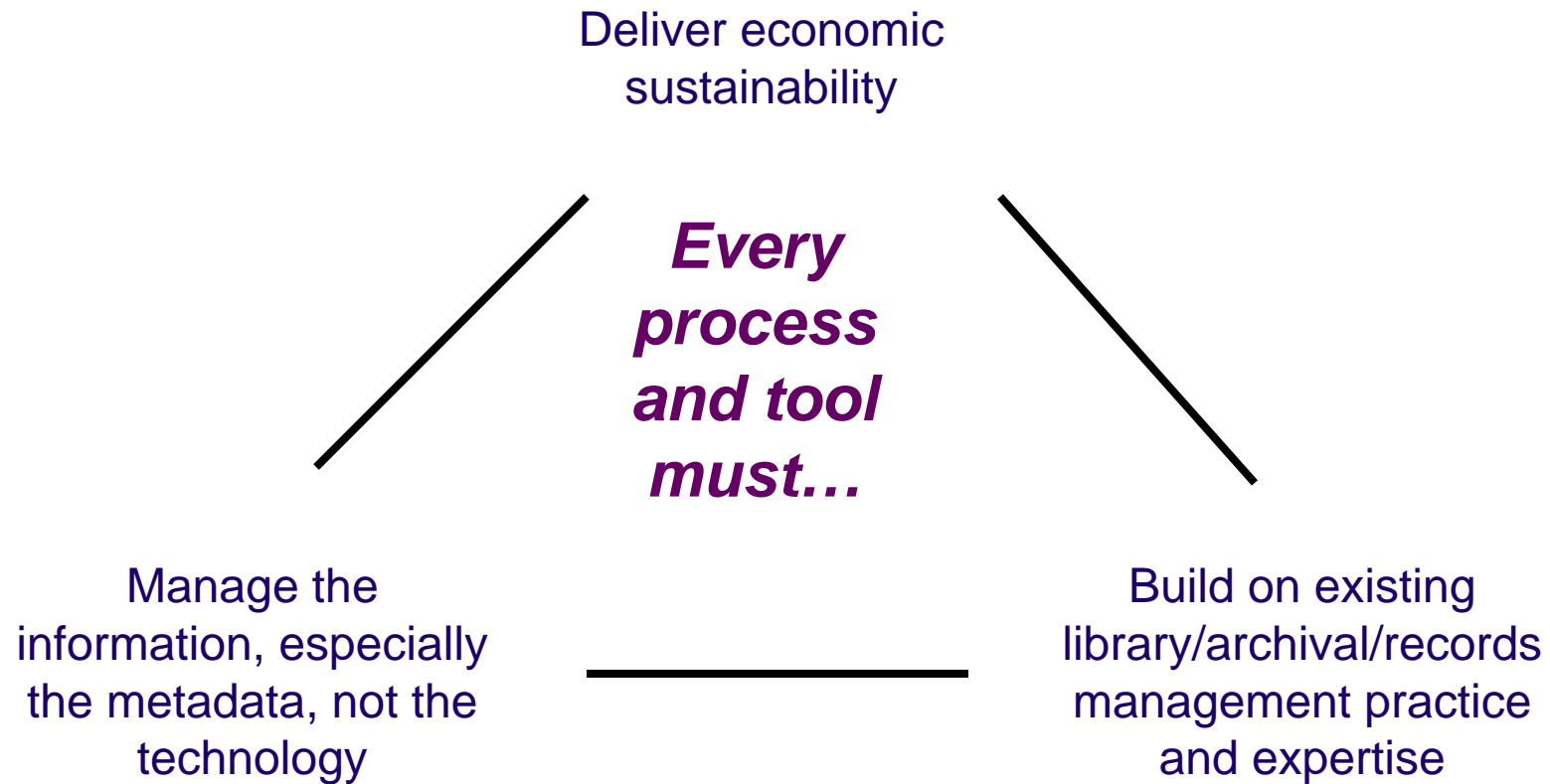Metadata is essential to future life cycle management

Archivists/Librarians shouldn't be asked to become technical experts, they need automated tools to help them apply existing expertise– who wants to edit METS files by hand?

Many of the changes required are 'social' and not repository specific

Lots of theory has been written but nothing beats actually working with the stuff

wellcometrust

In summary…

Deliver economic
sustainability

*Every
process
and tool
must…*

Manage the
information, especially
the metadata, not the
technology

Build on existing
library/archival/records
management practice
and expertise

wellcometrust

For more information

On the web

- Official Fedora web site – www.fedora.info

- Xena, National Archives of Australia - xena.sourceforge.net

- METS, Library of Congress - www.loc.gov/standards/mets

- National Library of New Zealand -
  www.natlib.govt.nz/en/whatsnew/4initiatives.html#extraction

General reading

- D-Lib Magazine – www.dlib.org

- Ariadne Magazine - www.ariadne.ac.uk

Thank you

Dave Thompson
Web Archive Project Officer
Wellcome Library

d.thompson@wellcome.ac.uk

wellcometrust