Dissemination or Publication?

Some consequences from smudging the boundaries between research data and research papers

Graham Pryor – University of Edinburgh
StORe guide

• Vision and defining features
• Relationship with scholarly publishing
• StORe in an e-Science landscape
• How StORe works
• What else might be in StORe
The StORe vision

- Seamless transport from research data to research publications and vice versa
- Bi-directional links proven in social science e-research but capable of export to other disciplines
StORe middleware – defining features

Web 2.0 approach, similar to FOAF web services such as Flickr or MySpace, gives control to the researcher

- Permanent links between publications and underlying data
- Federations of institutional repositories and data archives
- Simple process for assignment of metadata
- Searchable metadata assigned at collection level inherited by items within the collection
- Collection owners to add individual items with titles as minimum additional metadata
- Owners to determine which items are public / private
- Owners to form collaborations with colleagues / ‘friends’
- Owners to select items for deposit and for publication
To publish

• The production and dissemination of literature or information
• The activity of making information available for public view
• Traditionally, the distribution of printed works
• Recently expanded to include electronic resources – e.g. the electronic versions of books and periodicals, as well as websites and blogs
• Includes development, acquisition, marketing, production (printing and its electronic equivalents) and distribution
Traditional scholarly publication

- Hypothesis
- Undertake research/experiment
- Produce/manage data
- Select and organise evidence
- Write/assemble paper
- Submit to peer review
- Apply revisions and produce final draft
- Publish
Making e-data public

- Hypothesis
- Undertake research/experiment
- *Produce data/commence data curation*
- *Publish data ←*
- Select and organise evidence
- Write/assemble paper
- Submit to peer review
- Apply revisions and produce final draft
- Publish paper ←
- Activate data-publication links

[Peer review of data?]
Two-way links - benefits and risks

Opportunities to
- Explore a deeper level of detail
- Validate experiments (viz Chang’s nightmare)
- Track the use and improvement of research output
- Identify collaborators
- Confirm completeness of information searches
- Supplement published papers

Potential risks from
- Uncertainty of peer review
- Premature dissemination
- Subversion of scholarly paper
- Scavenging
- Lack of interpretative data
Current trends

• Big science
  – Astronomy: Virtual Observatory
  – Biosciences: the Wellcome mandate

• Federations and consortia
  – IRIScotland
  – EU policy: global knowledge economy (⅔ of €50billion FP7 funding for ‘Cooperation’)

• StORe pilot (UKDA/LSE)
How StORe works

The StORe pilot demonstrator
• Allows publication only if data has been deposited or identified
• Groups items (data and publications) as projects
• Accepts projects based on either primary or secondary analysis
  – Primary analysis creates new data
  – Secondary analysis is based on existing data and may or may not create new data
• Source repository - the UK Data Archive
• Output repository - the LSE’s Research Articles Online
• Pilot federation - includes a test institutional repository at the University of Essex
# How StORe works

<table>
<thead>
<tr>
<th>Research Articles Online (Output)</th>
<th>University of Essex (Institutional)</th>
<th>UKDA (Source)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSE’s Own Site</td>
<td>LSE collaborations</td>
<td>Researchers Area</td>
</tr>
<tr>
<td>Peer-reviewed items in own house style, additional metadata and ID</td>
<td>Items for review <em>(have IDs of data used and data produced)</em></td>
<td>Items for review <em>(have IDs of data used)</em></td>
</tr>
<tr>
<td></td>
<td>LSE approves <em>(ID added as part of process)</em></td>
<td>UKDA verifies (acquisition number added) and approves <em>(ID added)</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Value-added items in own house style, additional metadata and ID</td>
<td></td>
</tr>
</tbody>
</table>

CERN - OAI5 - April 2007
How StOREe works

PRESENT SEARCH is across metadata in Essex

• Find an article and you can
  – find the associated data
  – move to the official versions at LSE or UKDA
  – list all articles and related items of the author

• Find data and you can
  – find all articles associated with them
  – e-mail data owner to request access to approved but embargoed/private data
How StORe works

EXAMPLE FUTURE

Essex as multi-disciplined institutional repository in large federation of

• source & output repositories plus
• StORe-enabled institutional repositories at other HE/FE institutions

Providing

• local searches in institutional repositories
• cross-Federation access to all repositories
• individual institutional search indexes merged at the StORe portal
A user can search across all or specific collections without being logged in, using a simple Google-type search...
How StORE works

......or by employing more advanced options
How StORe works

Registered users can also view data within
• Solely owned private or public collections
• Collaborative collections within a federated source or output repository
• Collections to which they are contributor
Making data public with StORE – parallels with traditional processes

• Data deposit in institutional repositories until accepted by source repository
• Source repository verifies data authenticity
• Publication dependent on data deposit and subject to output repository controls
• Access to non-public objects can be restricted and requires authentication
• Release of data can be embargoed
Implications for scholarly publication

• Regulation
  – not strictly open access
  – extent of ‘publication’ is controlled
• Access control
  – Authentication of users
  – Not based on paying the price
• Researcher rule
• Potential for verifiable research output
• Managed forum for collaborations
• Increased expectation for delivery of research ‘package’
Opportunities for ‘small science’

• Role for institutional repositories
• Federations – regional or discipline
• StORe middleware