



**Appendix to the Subject and Institutional Repositories
Interactions Study Report**

Institutional Managers' survey

November 2008

Catherine Jones, Robert Darby, Linda Gilbert and Simon Lambert

Information Services Team, eScience Centre,

Science and Technology Facilities Council

SIRIS Survey: comment and results

This survey formed part of the fact-finding phase of the SIRIS project. It was advertised to institutional repository managers and support staff by a posting to UKCORR-discussion@jiscmail.ac.uk, the discussion list for UK research repository administrators, on 16 July 2008. A reminder was posted to the list on 24 July and survey was closed on the morning of 29 July. From 125 eligible discussion list subscribers (excluding project members), a total of 32 responses were received, comprising 24 completed surveys and 8 partially – though substantially – completed surveys. This represents a satisfactory 25.6% response rate. The survey was produced using the tool SurveyGizmo (<http://www.surveygizmo.co.uk/>).

Relationships with other repositories

The general impression is that most of these repositories are not interacting at a content level with other repositories. 82% expose the default metadata set of their repository software – which in most cases is likely to be simple Dublin Core. 2 respondents used Scholarly Works Application Profile (SWAP); 1 used Metadata Object Description Schema (MODS). Only 19% currently collected from other repositories, although there was strong interest among the others in collecting in the future, as well as in exposing content. The respondents were most interested in interacting with were UKPubMed (67%) and ESRC's Society Today (63%); other favourites were RePEc and arXiv.

The most significant technical factor influencing the choice to harvest was, by a considerable way, ability to identify output by institution. Respondents also valued minimal post-transfer processing of information. The trust factor was also rated highly as an influence on the establishment of transfer relationships, closely related to issues of source metadata quality and clarity about rights and permissions associated with transferred material.

Open questions

The following represent the main points made by respondents in the free text sections of the survey. For full details see the sections within the survey results.

Deposit policy

There appears to be a progress in institutions towards a defined deposit policy; but most institutions seem currently to be at a fairly early stage in policy development, and where a policy does exist, compliance may be limited.

Subject coding

Subject coding does not come across as a concern. One respondent reported rejection of subject coding by the user community early in the repository development. Most respondents appeared to use the default software option of Library of Congress subject headings, coding at a fairly broad level.

Versions

Version policy tends largely towards the most definitive available version, i.e. post-peer review, either preprint or publisher version, where available. Only a few repositories indicated a more comprehensive multiple version policy.

Repository relationships

- Authors should be able to deposit once-only, whether to IR or SR, and trust the repository system to manage distribution;
- The technology and infrastructure are simply not sufficiently developed and integrated for co-ordinated deposit and distribution of metadata/content to be practicable on a significant scale;

- National leadership in co-ordination of repository interactions is required;
- Transferred content should have clear identifiers and an explicit audit/provenance trail;
- Clarity about copyright and permissions is desperately needed.

Practical steps

- More national leadership and co-ordination of repository network and intercommunications;
- Co-ordination of Research Councils and IRs to develop repository transfer working models;
- Clearer definition of institutional, funder and subject repository roles and relationships;
- More joined-up thinking informing repository development, e.g. institutional research management, national REF working with libraries;
- Metadata exchange standardization/co-ordination (e.g. SWAP);
- Implementation of version identification standards in all repositories;
- Improved cross-repository search and retrieval mechanisms
- Engagement with publishers to develop automatic deposit process;
- Addressing problems of author identification and institutional affiliation;
- Looking beyond the UK: the problems are global;
- Simplify deposit and distribution process for authors (e.g. SWORD);
- Clarify rights issues

Evolution of repository landscape

- Convergence at architectural/technical level is probable, while repositories remain distinct at user/service level;
- SRs will be more user-focused; IRs will be more like data sources;
- IRs will become more strongly defined by the research management function;
- Whether or not they converge is not relevant as long as they are interoperable and so, from the user perspective, unified: it should not matter to the end user where content is located;
- IRs will need to become integrated into academic workflow, as successful SRs, e.g. arXiv, are;
- IRs will come to dominate in time once embedded;
- Unnecessary duplication of effort and content needs to be avoided;
- Some pessimism about the evolution of effective transfer relationships with subject and funder repositories;
- Deposit in IR and distribution to other repositories sounds like an attractive model, but given the established subject and funder repositories this is unlikely to become the dominant model. The mixed model is likely to be confusing;
- IRs often seem to lack clarity in the definition of their objectives and policies (research dissemination, management, preservation).
- There is some concern about duplication of research outputs and whether this will cause end-users problems.

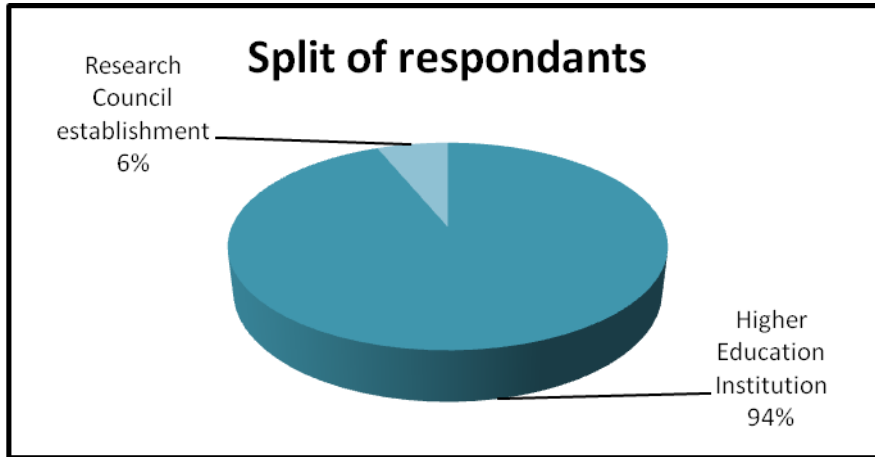
We would like to thank all those who took the time to participate in this survey. The results provided a very useful overview of the views of UK Repository Managers.

It should be noted for most of the questions it was possible to choose more than one answer. The “other” answer feature was used extensively to give more details.

SIRIS - Institutional Repository Managers' survey July 2008

Output produced on September 11, 2008

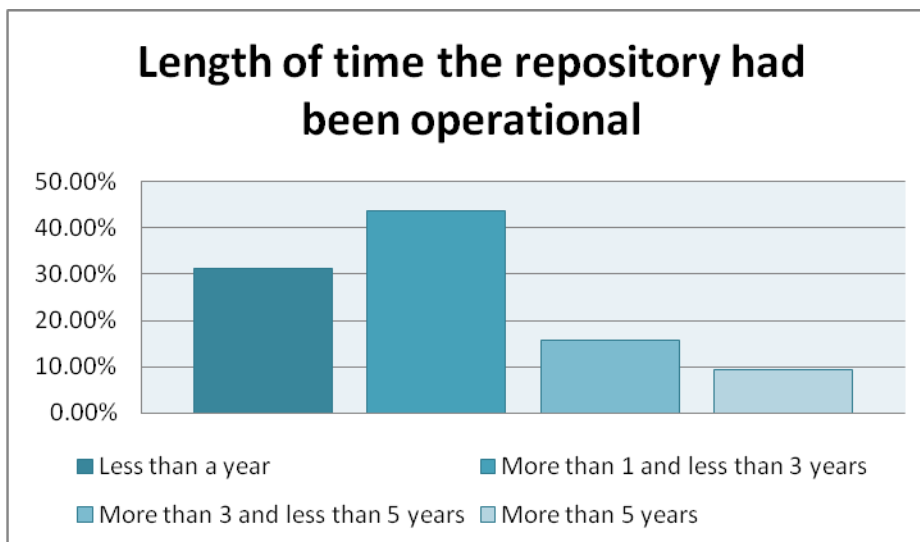
1. What type of institution do you work for?



Item	Count	Percent %
Higher Education Institution	30	93.75%
Research Council establishment	2	6.25%

Total responses to this question: 32

2. How long has your repository been operational?



Item	Count	Percent %
Less than a year	10	31.25%
More than 1 and less than 3 years	14	43.75%
More than 3 and less than 5 years	5	15.63%
More than 5 years	3	9.38%

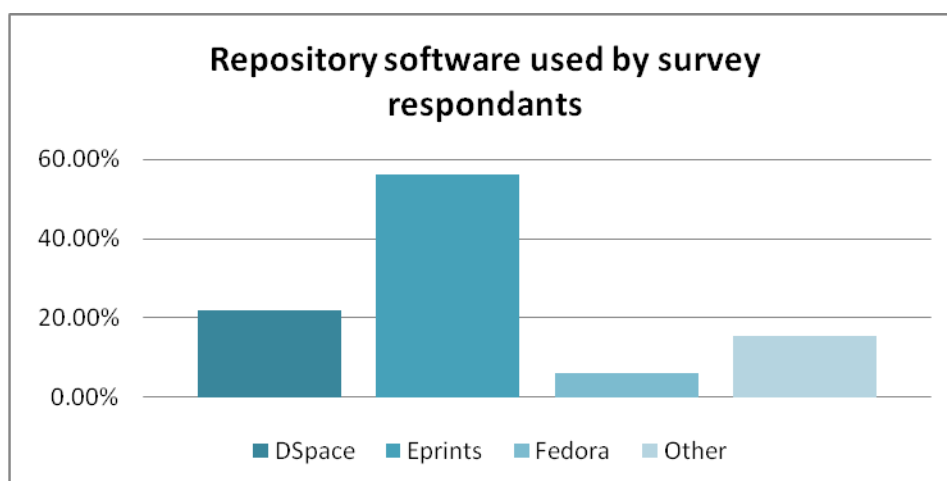
Total responses to this question: 32

3. What are the aims of your repository? Please rate 1-5, 5 being most important.

Item	Most imp	4	3	2	Least imp	Not considered	Average	Total
To be an open access repository with full-text content	56.3%	31.3%	9.4%		3.1%		4.4	32
To be a management tool for institutional research outputs	40.6%	25.0%	21.9%	9.4%		3.1%	3.9	32
To increase visibility and dissemination of the institution's research outputs	71.9%	18.8%	6.3%		3.1%		4.6	32
To provide a mechanism for digital object preservation	3.1%	37.5%	34.4%	12.5%	6.3%	6.3%	3.0	32

Total responses to this question: 32

4. What repository software do you use?

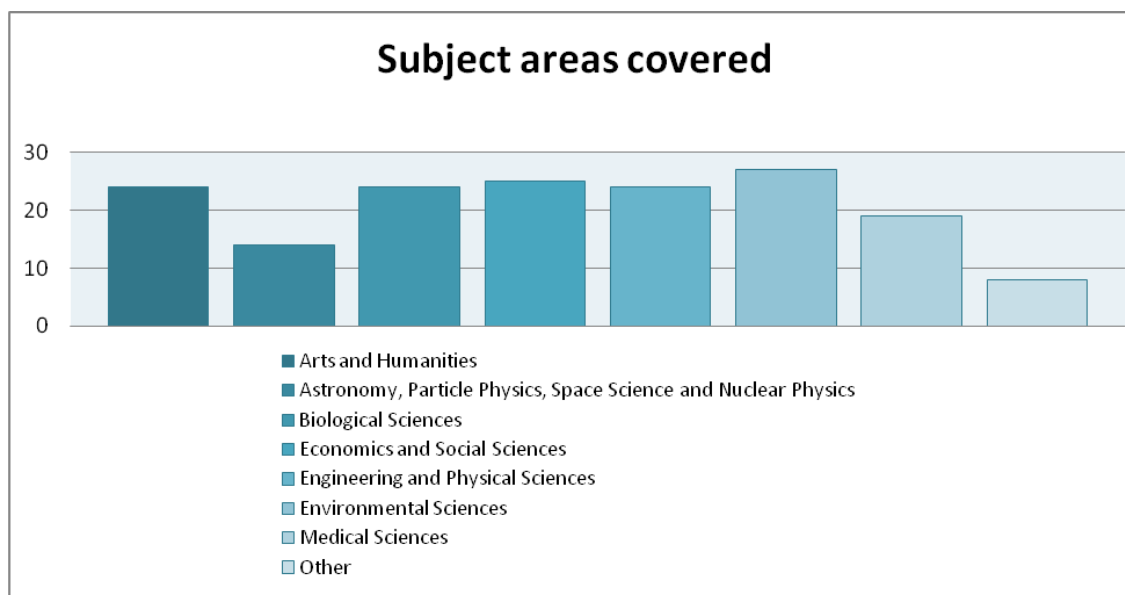


Item	Count	Percent %
DSpace	7	21.88%
Eprints	18	56.25%
Fedora	2	6.25%
Other	5	16.00%

Comments:
 but planning to change to DSpace
 Digitool (ExLibris)
 in-house developed software
 intraLibrary
 Open Repository

Total responses to this question: 32

5. What subject areas does your repository cover?

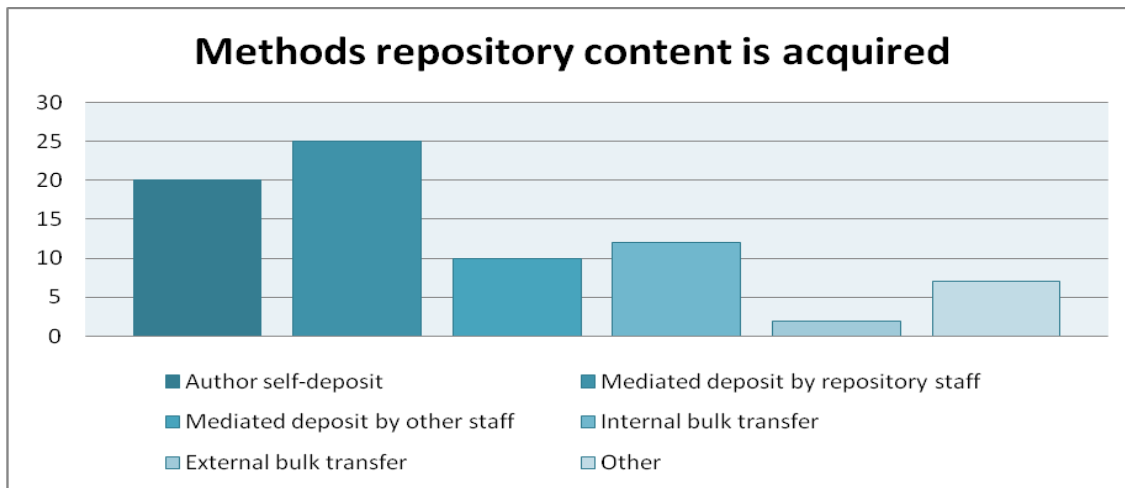


Item	Count
Arts and Humanities	24
Astronomy, Particle Physics, Space Science and Nuclear Physics	14
Biological Sciences	24
Economics and Social Sciences	25
Engineering and Physical Sciences	24
Environmental Sciences	27
Medical Sciences	19
Other	8

Comments:
 Agriculture, architecture, law, dental sciences
 All faculties/subjects (4)
 Are you including Education, Law, Health (not Medicine) & Business in the above?
 Business and Management Studies, Education
 Law

Total responses to this question: 32

6. How does your repository acquire content?



Item	Count
Author self-deposit	20
Mediated deposit by repository staff	25
Mediated deposit by other staff	10
Internal bulk transfer	12
External bulk transfer	2
Other	7

Comments:

Author and research manager self-deposit (with some mediation) is in development

Just about to trial author self deposit

May do internal bulk transfer in the future

Nearly all mediated deposit by repository staff

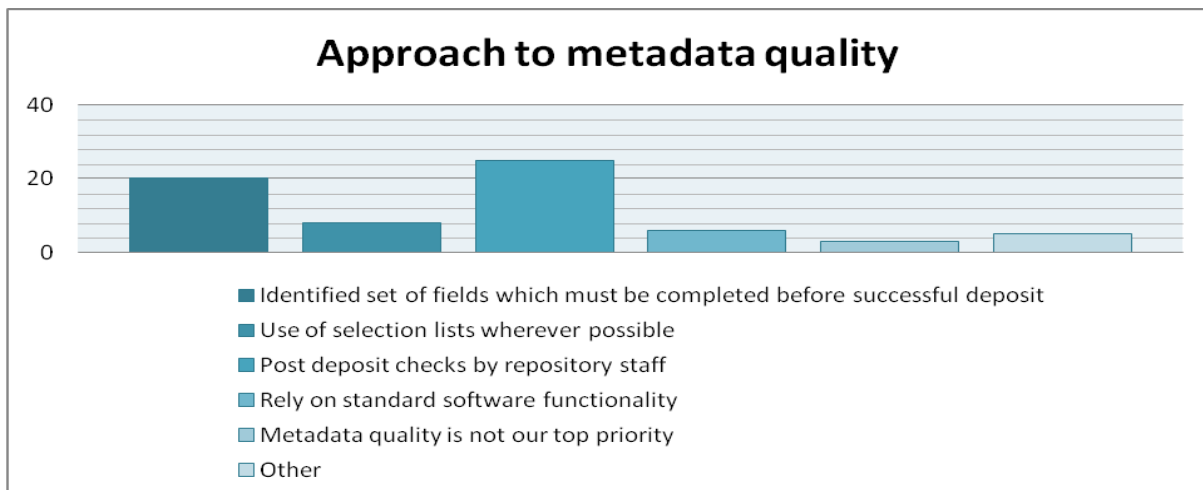
One off import from RAE

Project officer adding retrospective items

We do expect authors to deposit to us, but not on the repository software, so we are mediating deposit. We're also using alerts and looking on our website for material.

Total responses to this question: 30

7. What is your approach to ensuring appropriate metadata quality in your repository?

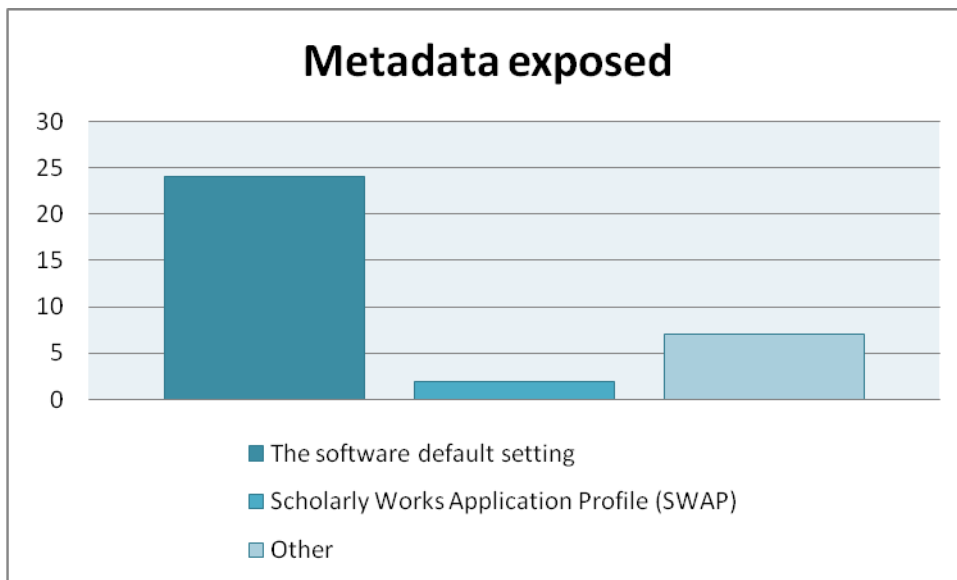


Item	Count
Identified set of fields which must be completed before successful deposit	20
Use of selection lists wherever possible	8
Post deposit checks by repository staff	25
Rely on standard software functionality	6
Metadata quality is not our top priority	3
Other	5

Comments:

As part of self deposit we will have repository staff checking metadata.
 Quality metadata is harvested from several online databases such as PubMed
 Some academic units identify research output via retrospective searching of external sites (eg. Web of Science) and import bibliographic details from that source
 Use automated metadata capture
 We've configured ePrints to use SWAP and ETD
 Total responses to this question: 30

8. What metadata do you expose for articles in your repository for others to harvest?



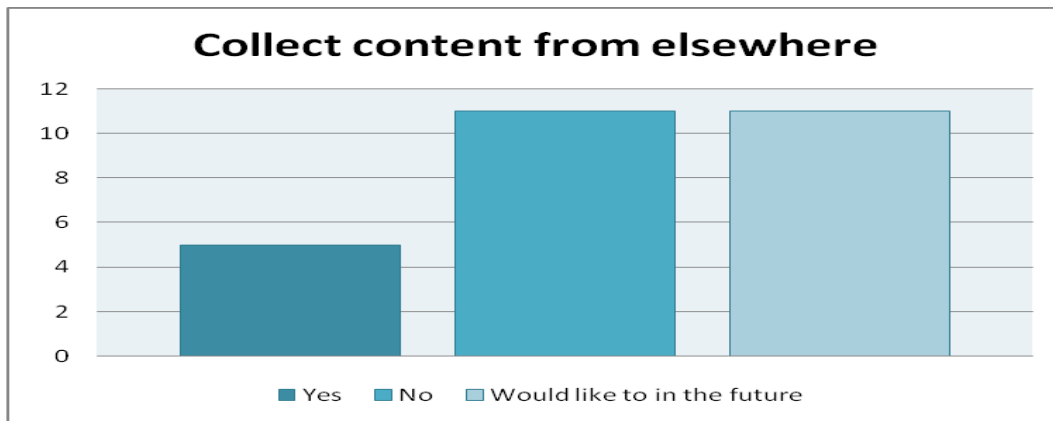
Item	Count
The software default setting	24
Scholarly Works Application Profile (SWAP)	2
Other	7

Comments:

Currently implementing MODS
 Default setting + local modifications (3)
 OAI-PMH DC
 OAI-PMH plus free text notes and abstracts
 Simple DC and UKETD_DC

Total responses to this question: 29

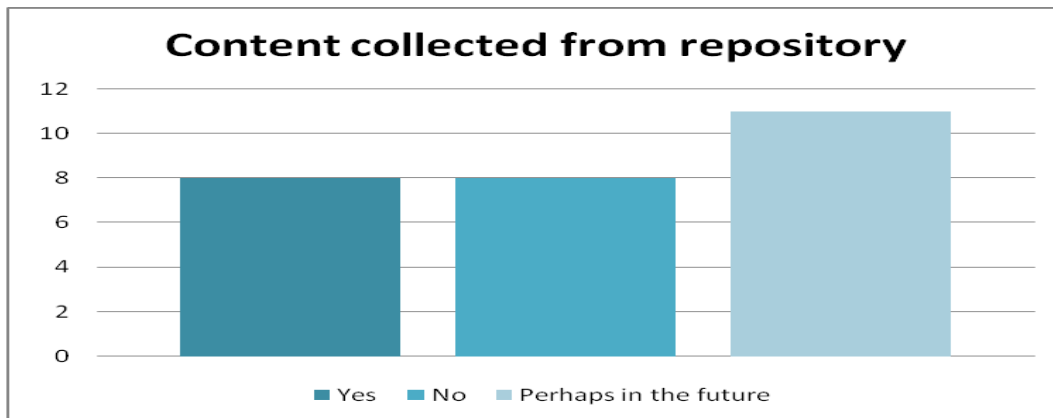
9. Do you collect content from another repository?



Item	Count	Percent %
Yes	5	18.52%
No	11	40.74%
Would like to in the future	11	40.74%

Total responses to this question: 27

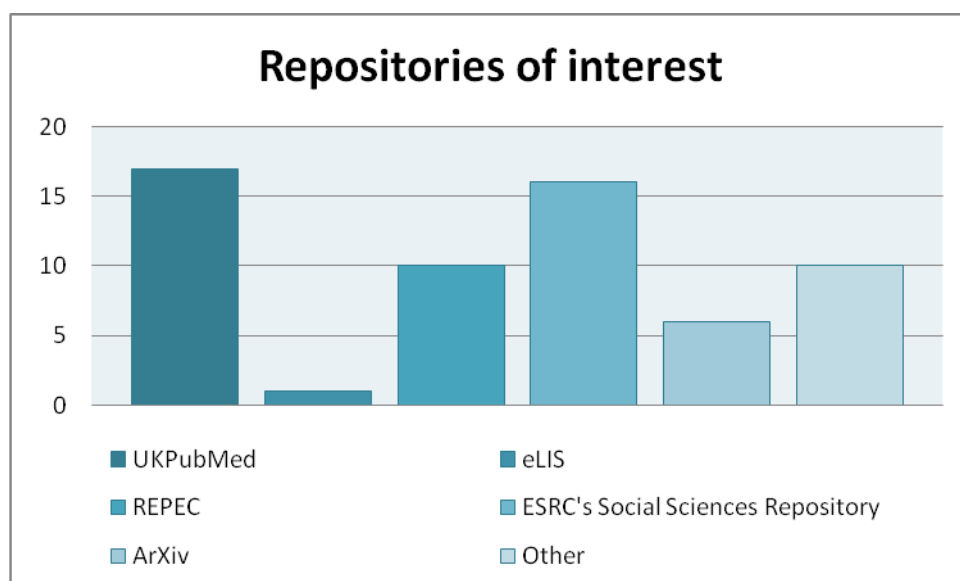
10. Do other repositories collect content from your repository?



Item	Count	Percent %
Yes	8	29.63%
No	8	29.63%
Perhaps in the future	11	40.74%

Total responses to this question: 27

11. Which repositories are you most interested in interacting with?



Item	Count	Percent %
UKPubMed	17	68.00%
eLIS	1	4.00%
REPEC	10	40.00%
ESRC's Social Sciences Repository	16	64.00%
ArXiv	6	24.00%
Other	10	40.00%
<i>Comments:</i>		
arXiv, EThoS		
Arxive, other RCUK repositories		
Collaborative centres with links to university repositories		
Don't know/need to think about this more (2)		
Economists Online (2)		
EThOS(2)		
WOK, Scopus		
Total responses to this question: 25		

12. What technical factors might influence your decision to source content from another repository? Please rate on a scale of 1-5, 5 being the most important.

Item	5	4	3	2	1	N/A	Average	Total
Easy to implement and use interface	38.5%	30.8%	26.9%			3.8%	4.0	26
Able to get full text as well as metadata	40.7%	44.4%	3.7%	7.4%		3.7%	4.1	27
Clear mechanism for updates (addition, changes & deletion of material)	34.6%	26.9%	34.6%			3.8%	3.8	26
Ability to identify your institution's output easily	74.1%	18.5%	3.7%			3.7%	4.6	27
Minimal processing after information transfer	44.4%	37.0%	14.8%			3.7%	4.1	27

Total responses to this question: 27

13. What policy & cultural factors might influence your decision to source content from another repository? Please rate on a scale of 1-5, 5 being the most important.

Item	5	4	3	2	1	N/A	Average	Total
Quality of metadata in the source repository	40.7%	48.1%	7.4%			3.7%	4.2	27
Trust in the source repository (high quality content, well managed repository etc)	61.5%	30.8%	3.8%			3.8%	4.4	26
Well defined versioning policy	26.9%	34.6%	30.8%	3.8%		3.8%	3.7	26
Clear understanding of the rights & permissions for transferred material	48.1%	44.4%	3.7%			3.7%	4.3	27
Good match between the metadata supplied and your minimum required fields	26.9%	53.8%	15.4%			3.8%	4.0	26

Total responses to this question: 27

Open questions:

Does your institution have a policy or mandate for authors to deposit? How effective is this policy in practice?

Open Text Responses:

Yes (8)

There was no suggestion of compliance “sticks”; much mention of advocacy, and cultural change slowly happening. Several mandates seemed to be being introduced in a low key way. Some mandates covered journals/conference proceedings and some theses only.

No (13)

Several of these institutions were considering the issue or on the road to implementation.

Opt-out process for bibliographic details.

Approval for mandating abstracts

Only work in IR will count for reporting: no mandate but will benefit the authors to comply

Do you use any subject taxonomy to identify articles held in your repository? If so, please briefly indicate standard, level of detail, who applies it etc.

Open Text Responses:

No (5 times)

Library of Congress (13)

- Repository staff./ Applied by cataloguing staff who are also repository staff/professional metadata librarian.

-mostly to 3 levels

- applied by user

JEL (3)

By organisational structure (3)

User driven (2)

MeSH (1)

RAE codes (1) + 1 considering

VERSIONS taxonomy (1)

Considering SKOS standard.

Derived from Dewey (1)

What is your policy on which version(s) of the article your repository collects?

Open Text Responses:

A version which is peer-reviewed and closest to the published version allowing for the publisher's terms & conditions. This may be the author's version or a publisher's pdf (14)

- Publisher's version preferred (3 out of the 14)
- Any post-publication version (5 out of 14)

Multiple versions collected but not necessarily exposed(4)

The version the author wishes to have archived/deposited, within restrictions of copyright (2)

Any version suitable for Open Access (1)

We do not include pre-prints.(1)

Are there any comments you would like to make about the relationship with other repositories?

Open Text Responses:

Improving ingest (5)

- Deposit once & push out info (3)
- Link to existing content (1)
- Retain identifiers for audit/provenance trail for ingested material(1)

A long way in the future (3)

- Need to focus internally first (2)
- Technology too immature(1)

Clarity about copyright for collecting or providing content (2)

Clarity about objectives (1)

Increase in communication would be positive (1)

Needs national approach (1)

Cross-searching rather than content exchange (1)

Interactions with commercial databases more important.(1)

What are the most important practical steps that need to be taken to develop the national repository infrastructure and enhance co-ordination of deposit and content retrieval across repositories?

Open Text Responses:

A clearer definition of the role and relationships between institutional, funder, and discipline specific repositories (5)

- Policy & standards (2)
- Communication forum (2)
- Role for the British Library (1)
- Only one repository for the Research Councils (1)
- Should ensure national service is not more work and has local benefit (1).

Standardised core set of metadata. (5)

- Could include subject classification (1)
- SWAP would allow for interchange of information (1)

Make it seamless for the end-user (5)

- Use SWORD for multiple deposits (2)

Unique author identification. (4)

.Clear descriptions of versions (3)

- what is the metadata describing, the item or where it is published?
- Resolution of duplicate entries
- Is there a need for a canonical version? (1)

Mechanism for assuring copyright permissions have been checked / granted (3)

An active engagement of the academic community is very important and to make sure that they needs are fully analysed and taken into account (3).

IRs need to be the core database for institutional metadata which filters into other organisational research databases and metrics. (2)

Provenance information needed (2).

Shared portal or an improved cross-repository search and retrieval process (by users) (1)

Automatic deposit on publication (get the publishers involved) (1)

Open Access should be understood to allow for items deposited in one repository may appear in others (1).

Please identify any research projects or development activities that are making a particularly valuable contribution to increased co-ordination between repositories. Please explain why you consider any chosen project or activity of particular value.

Open Text Responses:

Names project (4).

Research Support Project (3)

Sherpa (3)

Version identification framework /VERSIONS(2)

SWORD (2)

UKCoRR (1)

- Provides forum for discussions

NEEO project which will feature a subject gateway called Economists Online (1)

ETHOS (1)

Thomson unique identifiers for authors (1).

Intute repository search (1).

EM-Loader project(1)

Fedora/DSpace co-development - will facilitate interoperability between these(1)

Repositories Bridge Project (1)

JISC-CRIG(1)

Deposit Plait (1)

As the repository landscape evolves, do you anticipate a broad convergence between institutional and subject repositories, or do you think they will become more distinctly defined in terms of content and function?

Open Text Responses:

Institutional repositories will develop a greater role in research management, and will always be vital to the individual HEI. (5)

- A copy of everything that is published from my institution should be in its repository because this is an institutional asset. (2)

Different purposes (4)

- Subject repositories may remain quite different with different priorities (1)
- run and managed in distinct ways (3)
- Subject repositories will be user focused and IR more like data sources (1)

IRs and subject repositories do need to collaborate to make life easier for the researchers (3)

Information should be available publicly in a unified and easy-to-find system. (3)

- Interoperability is the issue. (1)
- To the end user through a search engine, it really does not matter where the item is made available(1)

Duplication of material (3)

- There is no problem duplicating materials, although OAI-ORE may make this less necessary (1)
- risk of duplication of effort and presentation(1)
- If duplication can be eliminated the repositories should be harvested together(1)

Institutional repositories need to define what they are doing and set clear policies. (2)

At an architecture/technical level there may well be convergence but there will be different requirements (2)

Deposit in IRs becomes automatic amongst the academic community, and that SRs harvest content from them.(2)

Multiple repository deposit will ease the tension (1)