

Prospero Scoping Discussion Paper:
**Perspectives and Models relating to a National Facility to support
Deposit of Pre- & Post-Prints under Terms of Open Access**

Introduction

This is the first document to be issued from the scoping activity that forms part of the Prospero Project Preparatory Phase (due to complete at the end of July 2006). It serves as a supporting document, as an Appendix, to the Outline Proposal for the Main Phase to be considered by the JIIE in May 2006. It also has a role in assisting contribution by members of the Advisory Group. It is based upon an earlier draft circulated to, and upon comments received from an 'Oversight Group' at the end of April. (Comments by project staff or members of the Oversight Committee appear in footnotes.)

The initial approach made to EDINA & SHERPA by the JISC Executive was a request that they work together to scope, design, implement and then to populate what was termed 'an interim repository'. The title of present document reflects the need to step back and achieve clarity of purpose about what is required and what could work, both as project activity and as a set of services. The timescale requires decision: we have taken a five-year planning horizon, with three-year funding.

Proposed scope

Our presumption is that the overall aim is to support exposure (and implicitly, discovery and use) of scholarly literature by Open Access (OA). Our focus here is on whether and how this strategic aim is served by a JISC-sponsored, national (UK) repository for digital pre- & post- prints of published journal articles and the like¹ that can be made available on Open Access terms.

Necessarily it is not to be an Institutional Repository (IR), and could be thought to share some characteristics of Jorum, already designated as a JISC national repository, albeit one for learning materials made available under licence terms other than Open Access.

As a repository, this would be deployed as a centrally-managed facility at a JISC national data centre, and support three types of service: ingest, access and transfer. Like Jorum, it will be necessary to pre-define specific designated use communities for each of these services.

- It will be important that that is done in ways that involved positive interaction with institutional repositories, especially with respect to encouraging deposit into extant IRs and the emergence of new IRs.

Like Jorum, there needs to be an exit strategy, which is positive with respect to institutional provision and responsibility.

- It is important to recognise that there is not presumption of continued funding beyond the initial three years, with policy expectation that UK-wide IR provision might then be sufficient, both with respect to institutional responsibility and success in deposits to support the OA agenda. The extent of that sufficiency in advancing Open Access requires independent appraisal – both in terms of the number of IRs and their OA deposits. Is the over-riding aim of this national repository to support OA or IRs, or are these presumed to be the same?

What could become distinctive about the scope for this national facility is that it acts as 'UK Repository Junction', for 'appropriate deposit'. This would provide a WAYF-like re-direct

¹ Mention has been made by publishers and others of the need for a place of deposit for associated data that support visualisation and other summary in journal articles, eg images and spreadsheets. Were this to come into scope, and there is some merit, then there would need to be definition that excluded overlap with the remit for data archives with respect to large-scale datasets.

that would present the potential depositor at the web-door of their institution's IR, or else the national repository facility would act as that appropriate repository where the institution is not yet supporting an IR.

An important, and as yet unresolved part of the scoping activity is the focus for this repository: author/researcher or grant-holder; institution or subject. Set out in later sections of this document are a review of stakeholders (Stakeholder Perspectives). This is to assist best 'value-added' use of effort and funds, to assist public understanding, and to mitigate against 'mission creep'.

A related point is the role that this national repository facility should have in assisting the creation of IRs, within the institutions or by providing the technology solution, bearing in mind both other strands in the JISC Repositories and Preservation Capital programme, and other potential providers of such services. Some consensus is required, and it is hoped that this document assists that to emerge.

Success also depends upon an understanding of the operational characteristics of any such repository facility. A set of abstract models (entity and relationships) is also presented below. This throws light upon the role and authority of the depositor in making assertions about the deposited object, the transient relationship between author and institution, and hence the indirect, mediated relationship between institution and deposited object.

What's in a name?

The project name is Prospero, even though we now discover another JISC project with that name. The national repository facility will require a name, one that conveys some sense of its purpose - the term 'interim' is regarded as being unproductive as a prominent label, or badge, that would encourage depositors to use the ingest service. The name Prospero has literary connotations; 'Open Access deposit' is the key message – suggesting BROADS, XROADS and INROADS, alluding to a British commons, to being at the cross-roads and to making inroads. However, this is not the occasion for extended discussion.

Background

The original idea for an interim repository had come from SHERPA in the form of a bid (the PROSPER proposal) submitted under the JISC 3/05 Circular; while the bid was unsuccessful, it had been identified by JISC as being of strategic importance:

"The last year has seen the rapid growth in the number of institutional repositories in UK HE. For example, 16 of the 19 Russell Group Universities currently have live repositories for their academics: the remaining 3 have plans. However, while the major research-led universities either have or will establish repositories in the near future, it will take some years for all HE institutions to build archives to disseminate the work of their authors. While this UK network of institutional repositories is developing, there will be a large number of academic authors without access to a repository.

"This is a personal disadvantage for those authors who will be unable to gain the readership and citations of their peers who use repositories. This is an institutional disadvantage, as there will not be the exposure of the research effort in these institutions. This is a disadvantage for research funders, who will not gain the increased dissemination of their funded research.

"... SHERPA has identified a clear need for a national interim repository, to allow every academic author in UK HE immediate access to a trusted, secure open access archive. This would give the advantages of open access dissemination for these authors, their institutions, their research funders and so create a level playing field for all academics in repository use."

(extracted from the earlier PROSPER Proposal)

Principal services to be supported by Prospero

1. Ingest – to support deposit (of pre-determined information objects, eg eprints)
 - a. by authors
 - b. by those otherwise authorised
2. Access – to support access by Open Access, and therefore wide community
 - a. by readers via a human computer (user) interface
 - b. by other service providers via a machine-to-machine interface
 - exposing metadata and/or the information objects
3. Transfer – to support export/migration of deposited objects (and their metadata) to extant and emerging Institutional Repositories
4. Other - to support successful operation and to ensure successful exit strategy
 - e.g. Re-direct service, for (potential) depositors that have an extant IR

Each of these pre-defined services is geared to specific designated communities of use. They form part of an evolving infrastructure to underpin the operation of a wider set of repositories and services. The requirements for each service have inter-dependence. In particular, the terms of deposit must support Open Access terms of access and facilitate transfer and preservation. Whilst there is not intent to be a 'subject repository' with respect to Ingest, the quality of any access service, wherever delivered, depends upon the extent and quality of metadata achieved at ingest (by whatever means).

Those who deposit material should gain some assurance that this repository facility must be seen to exercise 'archival responsibility', even if this is acquitted through transfer to an appropriate IR, to some place of last resort; much would need to be determined about that but this adds to the requirements for the Transfer Service.

Subject-based searching implies both machine-to-machine search via such as an RDN/Intute cross-search facility and web-based direct search/browse by end-users; subject-based presentation implies potential for discipline-based use communities, perhaps related to research council domains. This requires more scoping activity before decisions are made.

Prospero and Jorum

It may be instructive, in passing, to look at some of the characteristics of Jorum, the JISC-sponsored national repository for e-learning objects, see www.jorum.ac.uk. There is far from an exact match but it too has an ingest service (Jorum Contributor) and an access service (Jorum User), for which there is clarity with respect to the community of use, of metadata requirements and licence terms at deposit: both services require institutional licence and authentication/authorisation before use. There is no Transfer service as such, although the 'user licence' does support download (ie transfer) into an institutional (learning object) repository, for subsequent use in a VLE. Jorum is also regarded as a place where objects are 'kept safe' but not, as yet, offering a long-term Preservation solution/service. It is important to achieve a comparable level of clarity for the national repository facility.

Analysis

Use Communities for the Ingest Service

Clarity about the target client, or designated community, for the repository also serves to qualify, rather than define, this repository of digital content. This has implications for the design, implementation and operation: typically of the user interface and (Open Access) licensing, including authentication/authorisation, for the ingest service.

Figure 1

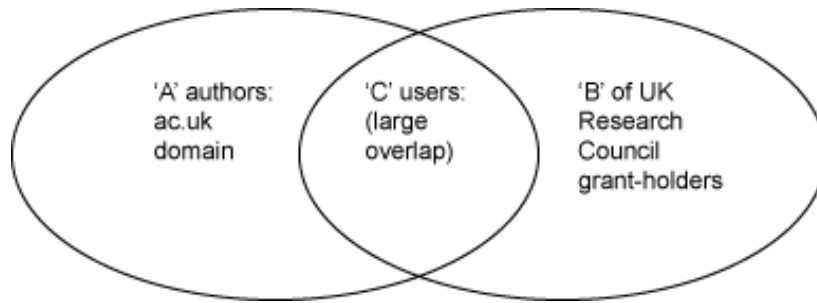


Figure 1 shows there to be two major classes of potential users for an ingest service, termed here as 'depositors'. It is a moot point just who may act as agents with authority to deposit on behalf of these users.

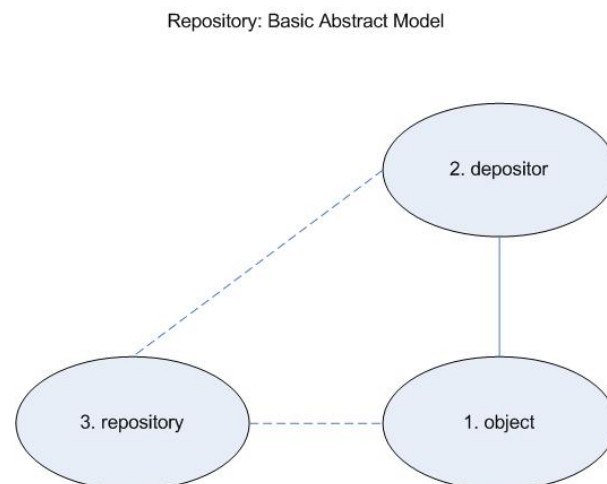
There are two principal classes. The first group, 'A', are authors of scholarly literature who are staff, and perhaps students, in UKHE & FE; these could be said to be in the ac.uk domain.

The second group, 'B', are Principal Investigators, who hold grants from Research Councils and perhaps other funding agencies. It is these that might be obliged (encouraged or required) to deposit articles associated with their grants.

There is an overlap between the two sets A and B, where it is the PIs who are the authors. However, not all scholarly literature that is produced in grant-funded project activity is authored by PIs. If there is to be deposit by authors associated with PI's of grant-holders, then there may need to be control by the PI, with use of appropriate grant reference numbers in a form that allows linkage.

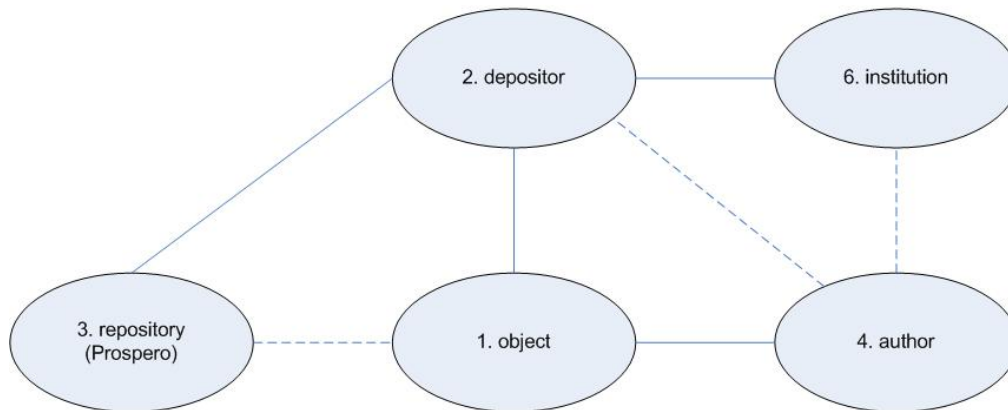
Modelling Relationships

At its simplest, the ingest service for the repository involves a depositor 'who has' an object. This is the Basic Abstract Model, an entity-relationship model shown as the first graphic below. A (potential) depositor contacts the repository, a relationship is established with the (potential) depositor ["who are you, where are you from?"], which clarifies the relationship that s/he has with the object ["what have you got, is it your's, is it OK to make available under OA terms?"] and allows the repository to ingest the repository [with whatever metadata].



Next is an Extended Abstract Model, in which the object has an author (or set of authors); at its simplest, the depositor is one of those authors and has a defined relationship to an institution: perhaps as an authenticated member of an institution; perhaps as someone authorised to act with relation to the object. It is shown as Extended Abstract Model I to correspond to the first of the stakeholder perspectives (discussed later), namely that of the researcher as author.

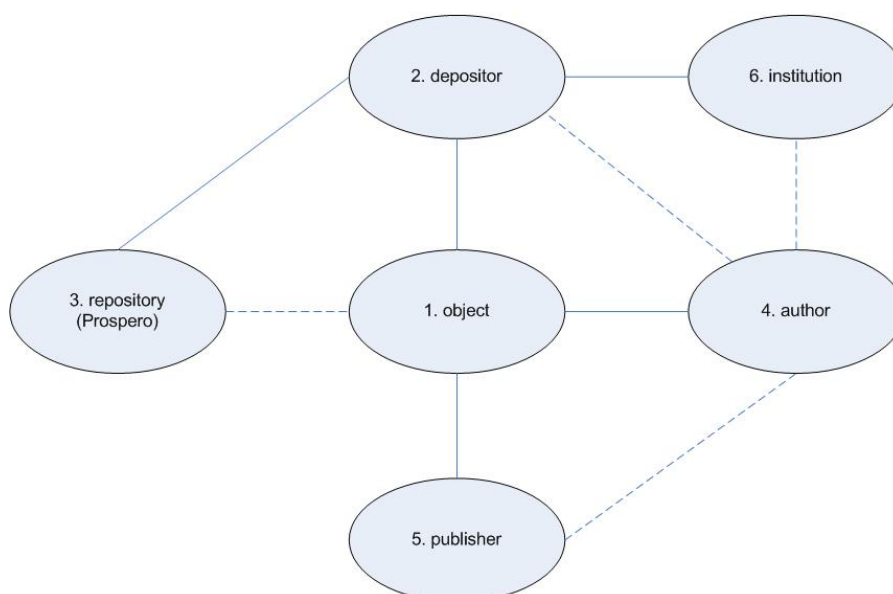
Repository: Extended Abstract Model I (researcher as author)



In passing it is interesting to note from this entity-relationship model that the Institution has no direct relationship with the Object. How then is this mediated? The relationship with the Depositor is defined by the deposit event; that with the Author(s) of the Object requires definition, and could be regarded as transient, as the Author can move Institution, complete with bibliography of authored objects.

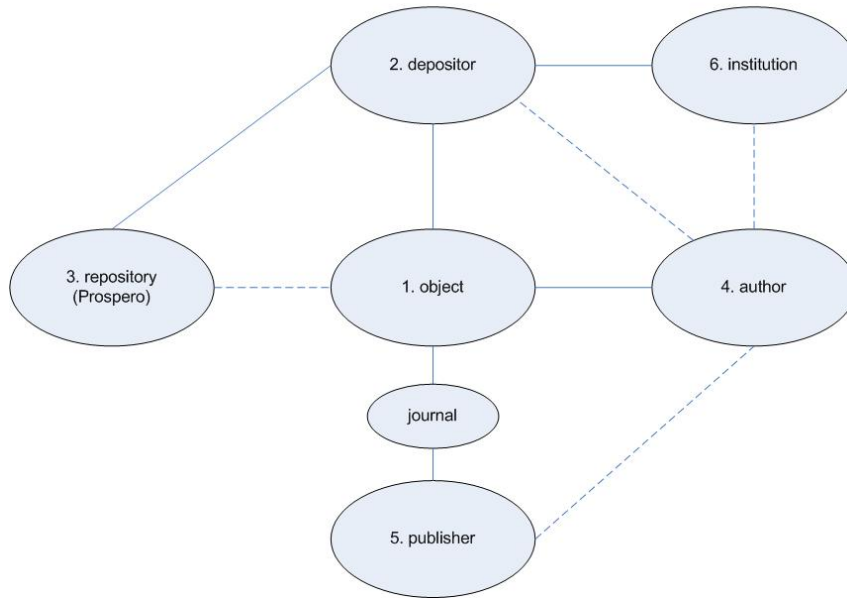
In a further version of the Extended Abstract Model, Model 1a, account is taken of the publisher, the party that made the object public. The publisher has defined relationship with the author of the object (handover of copyright), and defined relationship with the object – typically over post-prints but extending to versions of pre-print.

Repository: Extended Abstract Model 1a



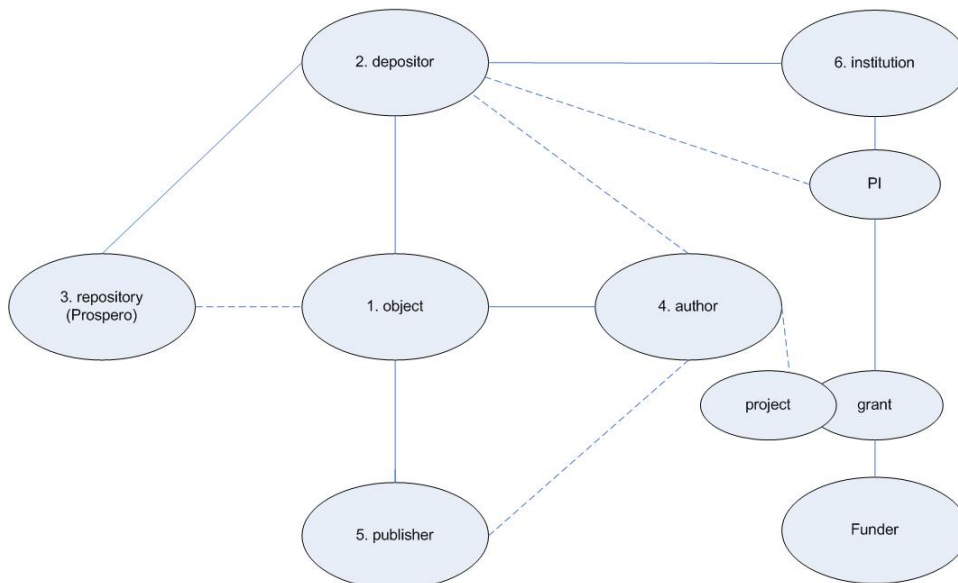
Extended Abstract Model 1b contains acknowledgement that the deposited object is an article within a journal/serial (or paper in proceedings, chapter in book, etc), as that frames the relationship that the publisher has with the object.

Repository: Extended Abstract Model 1b



Discussion of Extended Abstract Model II, which features the grant-holder, or Principal Investigator (PI), is deferred until discussion of Stakeholder Perspectives. However, it is clear that the PI does have clear relationship to the Institution (the grant is made to an Institution, defining the PI) but that the PI, who may be required to comply with obligation by a Funder, does not have a direct relationship with the Object – that too must be mediated in some way.

Repository: Extended Abstract Model II (Pis: grant-holders)



Stakeholder Perspectives

The analysis above has exposed a number of different stakeholders, with interests in advancing the Open Access agenda. Some of these may coincide in part, others conflict. These stakeholders include: researchers as authors (I); grant-holders (II); grant funding bodies (II a); universities and other research institutions (III); publishers (IV). Stakeholders are not mutually-exclusive in their viewpoint. At this stage in the scoping, we attempt to contrast these view-points in order to air differences in their expectation of services from a national repository facility. We do so for a variety of purposes, not least because these stakeholder perspectives may serve as rationale for this work by members of the policy community that will review the Outline Proposal for the Main Phase. Our hope is that this will assist the building of consensus about the aims and objectives of this national repository facility in order that we can plan appropriately. Gaining early consensus should also assist public presentation of purpose, in such as the project web-site, and help prevent 'mission creep' and help ensure focus and the 'best fit' and effective 'value-added' use of effort and funds.

In this document we examine:

Stakeholder Perspective I: for UK Researchers as Authors

Stakeholder Perspective II: for Principal Investigators as Grant-holders

Stakeholder Perspective III: for Universities as Institutions

Stakeholder Perspective IIIa: for Universities as subscribing Institutions

The Table below contains a checklist for each Stakeholder attempting to draw out relevant aspects of the services and activities for a national repository facility. These are then reviewed and discussed. It seems likely that what is to be designed and built will seek to serve more than one stakeholder.

For the present, we have omitted focus on grant funding bodies (Funders) and Publishers.

Stakeholder Perspectives: Summary Table (working draft)

for UK Repository Facility

	Stakeholder I	Stakeholder II	Stakeholder III-(i)	Stakeholder III-(ii)
Key Characteristics	Author	Grant-holder/PI	Institution	Subscribing Institution
Primary 'customer'	UK researchers as authors (individuals)	UK Research Councils and other research funders (cf RCUK)	UK Universities and other Institutions	Participating, subscribing Institutions
Rationale for national repository service	Allow every UK academic author to expose material under Open Access; has somewhere to deposit. [?Could provide personal 'keepsafe' filestore to encourage motivation to use.] Need to define relation to extant IRs.	Provide repository for grant holders (PIs) so that they can comply with demands of Funders, especially if no extant institutional or subject-based repository	Provide every UK HE Institution with a service: technological capability, that gives 'authorised' authors a place to deposit; and institutions opportunity to focus on procedural matters before taking on full repository functions. Extant IRs can opt for 're-direct' service.	'Vanilla' temporary repository service motivates institutions to commit to building IRs; service provided only to those which 'sign up'.
Ingest Service				
'Producers' (in OAIS language)	Any author from ac.uk	Any author associated with a grant (& reference number) from a UK Research Council or other research funder.	Any author from ac.uk	Only ac.uk authors from participating institutions

	Stakeholder I	Stakeholder II	Stakeholder III-(i)	Stakeholder III-(ii)
Object types Pre-prints & Post-prints	Initial limits on open access items to non-proprietary (or open format) document formats, e.g. PDF, XML, RTF, and possibly one or two standard image types. No filetype limit on private store, but guidelines given.	Initial limits to document types, e.g. PDF, XML, RTF, and possibly one or two standard image types. Future expansion to further data types based on policies of RCUK and of Research Councils	Initial limits to document types, e.g. PDF, XML, RTF, and possibly one or two standard image types. Future expansion to further data types based on institutions' policies	Limits to document types, e.g. PDF, XML, RTF, and possibly one or two standard image types.
Authorisation	Lightweight use of email (ac.uk) verification; Athens or Shibboleth authentication; all ac.uk authorised to use	Does authorisation to deposit needs to come from either PIs, allowing compliance check by Research Council staff? Possible non ac.uk domain authors.	Authorisation controlled by institutions: lightweight or Athens or Shibboleth authentication; may need designated Rep from each Institution, cf Jorum	Athens or Shibboleth authentication; authorisation controlled by institutions
User interface	As national eprint repository (& personal filestore?); as simple as possible (no expectation of user support at local site)	As national eprint repository by subject or 'partitioned' by research council	As national repository 'partitioned' by institution with 'switchboard' links to existing repositories	As national academic repository with 'switchboard' links to existing repositories
'Editor' (gatekeeper) control	Minimal checking by project staff - document opens correctly, mandatory fields filled out, legal deposit for OA appears correct.	Research council staff or PI may need to authorise document as outcome of grant award; else project staff verify award exists manually. Minimal checking by project staff	Institution plays a role. Minimal checking by project staff until institutional staff are on board to take over.	Minimal checking by project staff until institution is ready for wholesale transfer into own IR.

	Stakeholder I	Stakeholder II	Stakeholder III-(i)	Stakeholder III-(ii)
End-user license issues	<p>ROMEEO database used to ensure open access status for published items.</p> <p>Creative Commons licensing options provided for non or pre-published items.</p>	<p>ROMEEO database used to ensure open access status for published items.</p> <p>OA non or pre-published items dependent on research council's policies and by arrangement with Repository (such as datasets).</p>	<p>ROMEEO database used to ensure open access status for published items.</p> <p>Creative Commons licensing options provided for non or pre-published items.</p>	<p>ROMEEO database used to ensure open access status for published items.</p> <p>Creative Commons licensing options provided for non or pre-published items.</p>
Embargo	<p>Embargo service for delayed open access, including option to never allow open access set by depositor.</p>	<p>Embargo service for delayed open access.</p>	<p>Embargo service for delayed open access</p>	<p>Only open access items can be deposited (no embargo service).</p>
Important ingest functions	<p>Keyword 'tagging' with basic subject classification input by author; non-OA 'dark archiving' with embargo time-periods to set; personal filestore for broader range of filetypes.</p>	<p>Confirmation of award numbers through research councils' databases (automatic or manual); possible additional email authorisation by PI per item deposited; subject classification based on each research council; varying ingest policies per research council.</p>	<p>'Repository junction' to lead authors to deposit in appropriate institutional repository; 'virtual repository' customisation by institution if they choose to do so; possibly varying ingest policies per institution.</p>	<p>'Repository junction' to lead authors to deposit in appropriate institutional repository; plain vanilla functionality so as to motivate institutions to build their own IRs for customisation and local control.</p>

	Stakeholder I	Stakeholder II	Stakeholder III-(i)	Stakeholder III-(ii)
Access Service				
Discovery aids for 'Consumers' (in OAIS language)	Basic subject classification (author-assigned) to support search & browse. Search enhanced by author's keywords and full-text.	Extensive subject classification to support search and browse, matching pre-existing research council's taxonomies or funding programmes. Search enhanced by author's keywords and full-text.	Extensive, standard subject classification to support search & browse (e.g. UDC, Dewey, LOC). Perhaps only for post-prints where information from publishers could be used(?) Institutional staff encouraged to take over subject assignment for their authors.	Basic subject classification (author-assigned) to support search & browse. Search enhanced by author's keywords and full-text.
Access for M2M	OAI-PMH metadata	OAI-PMH metadata	OAI-PMH metadata	OAI-PMH metadata
Data Management				
Backups, object IDs, checksums, other routine QA	Procedures will be established for project staff to ensure integrity of objects deposited	Procedures will be established for project staff to ensure integrity of objects deposited	Procedures will be established for project staff to ensure integrity of objects deposited	Procedures will be established for project staff to ensure integrity of objects deposited
Metadata packaging and editing rights	No complex metadata(?) User can edit own metadata & remove own items.	Likely adoption of METS or DIDL as standard metadata wrapper(?) Research council designated staff may be able to edit metadata & remove items.	Likely adoption of METS or DIDL as standard metadata wrapper(?) Institutional designated staff may be able to edit metadata & remove items.	No complex metadata(?) Only Institution staff may edit metadata & remove items.

	Stakeholder I	Stakeholder II	Stakeholder III-(i)	Stakeholder III-(ii)
Preservation planning	<p>To be undertaken after exit strategy/continued funding is clear (nearer to end of project). Limited filetypes for OA items ensure short-term usability.</p> <p>Possible alerting system notifies depositor that (OA and non-OA) item's filetype requires migration</p>	<p>To be undertaken after exit strategy/continued funding is clear (nearer to end of project). Limited filetypes for OA items ensure short-term usability.</p> <p>Preservation planning for other types as part of 'contract' with specific research councils.</p>	<p>To be undertaken after exit strategy/continued funding is clear (nearer to end of project). Limited filetypes for all items ensure short-term usability.</p> <p>Possible alerting system notifies institutional staff that item's filetype requires migration</p>	<p>No preservation planning because exit strategy is clear and project is short-term. Limited filetypes for all items ensure short-term usability.</p>

	Stakeholder I	Stakeholder II	Stakeholder III-(i)	Stakeholder III-(ii)
Transfer Service & Exit Strategy				
Transfer	Duplicates of OA items deposited will be transferred to IRs upon their request and if deposit license allows, but will not be removed from repository until closure of repository.	Transfer to research-council supported subject repositories is envisaged as they are established. Award reference number is key to linking items to individual research councils.	Transfer to IRs is envisaged as they are established. Institutional ID is key to matching items to appropriate IR.	Transfer to IRs is envisaged after a certain period of the project. Institutional ID is key to matching items to appropriate IR.
Exit Strategy	Either continued funding (by JISC?) or arrangement sought for wholesale transfer of items to a trusted repository.	Either continued funding (by research councils?) for tailored services or transfer of items to relevant subject repositories.	Either continued funding (by JISC?) with partial income by subscribing institutions for 'virtual repository' services, or arrangement sought for wholesale transfer of items not already transferred to IRs, to a trusted repository.	All items transferred to new IRs by agreement by end of project.

Stakeholder Perspective I: for UK Researchers as Authors

Key Characteristics

The intention here is to assist UK researcher-as-authors expose their scholarly literature to the benefits of Open Access, by providing for all, that is to add provision for those who do not already have an Open Access (OA) institutional repository (IR) at their university of employment, or access to a relevant subject repository. The key here is how to determine who already has, and is aware of institutional provision.²

This is institutional-repository-neutral. It leaves open the form of relationship to be established with extant IRs – such as a re-direct option.

Ingest Service / Authorisation

As a repository made up of objects created/originated 'by researchers, for researchers' the deposit interface has to meet certain requirements, as well as having desirable design characteristics. Academic authors have varying motivation and competence when it comes to metadata, for example.

Authorisation to ensure provenance and to permit OA:

One requirement would be that the author-as-depositor warrants the provenance of the information object, demonstrating the authority both to do so and to make the deposit. The related requirement is that the depositor is able to warrant that there is licence to provide access to the information object by 'Open Access', respecting any third party rights.

In this stakeholder perspective it is the author that provides the warrant and asserts clearance of rights.

Minimally, use could be made of confirmation by (ac.uk) email address. Use could be made of Athens ID (and subsequently Shibboleth-enabled) to limit access to the ingest service to those authors in UKHE/FE, but access would not be mediated by institutions. Prospero staff effort may be needed to double-check the submitter's digital rights assertions for open access items, such as in the Romeo publisher's database. Nominal checking by staff to rule out 'spam' or spoofs as determined by a set of guidelines defining legitimate materials may also be necessary.

The intention is that all content would be available by Open Access, possibly subject to embargo conditions. A creative commons license form could be made available for open access materials *not* owned by a publisher (along lines of Flickr). Another possible customisation is that private items may be able to have their metadata visible (or not), for searching by others.

[Re-direct options: 1. automatic re-routing, preventing central deposit in the UK Repository. 2. A list of existing repositories would be made available once they were made aware that their institution had a repository, e.g. they would have the option to choose.]

[Subject classification at ingest. 1. Could be done by individuals without any mediation. Basic subject classification (drop-down short list) and open-ended keyword 'tags' could be submitted by the depositor. 2. Could be added by professional staff. 3. Could look to automatic generation from text mining of full text.]

² This definition of scope could be tightened to exclude those who have a subject-based open access repository in their discipline in which to deposit their research outputs but such definition might be harder to operationalise.

Access Service / Discovery Metadata

Searching will be enhanced by author's keywords and full-text.³ Browsing by subject will be at a basic level only.⁴

Storage / Management

There needs to be a 'take-down' mechanism to remove items deposited. This could also be offered for use by the depositor. A facility to correct/enhance metadata is also needed. A filestore limit may need to be imposed for the personal deposit items.

Additional service: As an added incentive to individuals to encourage deposit, a personal repository service could be offered, in which, in addition to those papers deemed open access, a private store could be offered for papers that are either embargoed by the publisher temporarily, or even for other objects deemed worthy of saving by the individual, accessible only to themselves, for the life of the project. Only the open access items would have a strict restriction on filetype. (Takeup on the additional service might be low if the service is billed as 'interim'.)

Transfer Service / Exit Strategy

Items may be deemed to belong to the depositor, so for export to an IR the owner would need to agree - preferably in advance rather than at time of repository closure, else effort would be made to contact the depositor before closure to return their items to them or export to the repository of their choice.⁵

Preservation planning could come into affect if the Repository gets continued funding to remain in service, but only for the open access items of the agreed file-types. Items stored by depositors for their own use only would probably not be migrated, though a backup of the bit-stream would remain available.

If the Repository does not get renewed funding past the interim period, the project would seek to transfer the repository contents en masse to an appropriate trusted repository – preferably a national memory institution (for example the British Library). Institutional Repositories (IRs) or subject repositories seeking transfer of content up to the end of the project would need to initiate contact, specify items, and would be accommodated subject to author's wishes specified via the deposit license.

Stakeholder Perspective II: for Grant-holders as Principal Investigators

Key Characteristics

A related group of stakeholders are the grant funding bodies, such as the Research Councils themselves, either as a group (RCUK) or individually, and other funding bodies.⁶ What we look at here is the use community, the Principal Investigators and co-investigators, who may be obliged (encouraged or required) to deposit under Open Access.

³ Consideration needs to be given to full-text searching: straight ASCII text, html-based documents, and pdf file-types are proven to be indexable. Indexing proprietary formats (such as Microsoft Word .doc files) is not always possible.

⁴ There is some evidence that suggests that subject classification by academics themselves has an accuracy of less than 90% (the error rate varies, but tends to be in the 10-20% range). Also, *which* subject classification to use, as some are more complex than others?

⁵ We need to avoid requirements to keep track of contact details for depositors.

⁶ Noting that some Research Councils have OA repositories already (CCLRC), and others are supporting development of new ones (eg MRC/BBSRC supporting UKPubMed).

The repository has to allow deposit only of papers that have a legitimate link to a Research Council grant (e.g. number). In this sense is a service manufactured for the Research Councils, or for Institutions in receipt of grants, rather than only for individual researchers⁷. Like Stakeholder Perspective I, this is institution-neutral, but could be made to assist both extant IRs and those that might emerge, as well as subject repositories.

Ingest Service / Authorisation

The depositor is required to identify the document as an outcome of a grant award, by research council and grant number code. There is some complexity here, as indicated by the Extended Abstract Model II (above). This linking field may require validation.

[It might be possible to have a separate interface taking account of policies etc of each participating research council – say, by partition in D-Space, or with a different domain name and instantiation of e-Prints software.]

Access Service / Discovery Metadata

Searching would be enhanced by author's keywords and full-text.

Subject classification and other fields could be tailored for each funding body, eg each Research Council, allowing a more sophisticated browsing option.

Storage / Management

There is no particular reason to limit filetypes to PDF if any of the Research Councils wish to use the repository for other grant-related materials as a way of moving forward with their data sharing policies. Appropriate filetypes and metadata schema for non-document items would need to be worked out with each research council (possibly in stages, perhaps not even during the life of the initial project).

There may be reasons to build links between the repository functions and the research council's own databases, in time.

Transfer Service / Exit Strategy

It would be possible to export to a research council's repository (if it were to be set up), than to institutional repositories as there may not be an easy way to know which institution any given item should be exported to (e.g. the depositor's institution, who may be a co-author? Lead author / Principal Investigator's institution? There will be changes over time, for example the institution at the time of deposit, or at the present time?)

One possible exit strategy is that the Research Councils take over the funding of the repository or repositories. Another could be to identify suitable subject repositories to export the contents to based on the initial subject classification.

If the repository were to continue past the initial funding, preservation planning would need to be done, including migration strategies for allowed filetypes, or perhaps for published documents only.

⁷ This model seems more feasible if there is open access mandate for deposit.

Stakeholder Perspective III (i): for Universities as Institutions

Key Characteristics

In this stakeholder perspective, the national repository facility aims to serve institutions, both those that currently operate institutional repositories (IRs) and those that do not, both in the short run and in the medium term. Much depends on the form of the engagement. It could be characterised as nurturing the development of institutional repositories through advocacy⁸, liaison, and provision of some sort of repository hosting service to institutions that do not yet have an open access repository covering research outputs.

This may be seen as opportunity to provide a central technological capacity that would encourage library or other appropriate staff at local institutions to focus on the procedural side in setting up their own institutional repository. (The reasoning for this could be based on the notion that an institutional repository is a more natural setting for open access repositories than a central repository because rewards from increased visibility flow not only to the authors but also to their institutions.⁹) Thus it is envisaged that more and more institutions will be motivated to have IRs but the national service could help guide them into it one step at a time.

[It should be noted that EDINA has a network of site representatives at each UK university, although these would not necessarily be the same individuals who have responsibility for IRs they ought to be known to one another.]

Ingest Service / Authorisation

Only UKHE/FE authenticated users would be allowed to deposit. Again, this may include students, or may be limited to particular user groups, depending on whether each institution sets up a limit to authorisation or not (using Athens/Shibboleth). It would be up to institutions to take responsibility for their own policies.

[re-direct option: If a user tries to deposit an item for which there is an appropriate existing IR, the page will redirect the user to their own IR's website automatically. If appropriate staff at institutions choose to take control of deposits to the 'virtual repository', the user interface could be customised so as the look and feel of the web page appears to come from the user's own institution.]

The liaison project staff would encourage local librarians or other appropriate staff to become metadata editors for papers submitted by their users. If no further metadata editing were required because the user had completed it, then preferably it would be the institution's local staff who approve the paper for submission. In this way the repository will become a 'virtual' local repository. Project staff will only 'approve' deposit if no contacts are set up with the local institution of the depositor.¹⁰

⁸ Advocacy is something that SHERPA Plus is doing, so we need to understand where one project ends and the other begins.

⁹ "Self-archiving is done in order to maximize the visibility and accessibility of refereed research, and hence to maximize its usage by researchers and its impact on research. The benefits of maximizing research impact are felt by the researcher and the researcher's institution, rather than by some more central entity (such as the research discipline or learned society). The academic reward system (salaries, research funding) is centered on the researcher's institution. Publishing and impact confer advantages on both researcher and institution. Hence the researcher's institution is the natural one to host self-archiving and ensure that its archives are filled with its annual research output." Self-archiving FAQ on the ePrints website, retrieved 18 April, 2006, URL: <http://www.eprints.org/openaccess/self-faq/#institutional-vs-central>

¹⁰ One suggestion is that a relationship with each institution is negotiated based on an 'contract' or SLA which would include clearly what the institution would provide (eg supply of material, advocacy, metadata) and

Attention needs to be given to allowing partitioning or logical 'views' onto the repository by institution.¹¹ There could be more than one editor per institution looking after papers from different parts of the university. However, simplicity would be sought by the Prospero staff.

Guidelines for what is able to be deposited and what is not can be shaped by separate institutions; local staff can act as gatekeepers for their institutional assets through submission approval. The deposit license would have to grant more power to the local institution than in stakeholder perspectives I and II, and allow the institution to decide ultimately what goes in the repository or not, just as they do when they build their own IR.

Access Service / Discovery Metadata

Searching will be enhanced by author's keywords and full-text.

To encourage best practice, a standard subject classification (e.g. LoC Subject Headings, Dewey, Universal Decimal Classification) will be adopted. Institutional staff will be encouraged to take over subject classification for their authors.

Storage / Management

The ability to identify institution must be maintained, preferably as part of an identifier scheme.

There may need to be some customisation of the repository software to suit the needs of different institutions.¹²

Transfer Service / Exit Strategy

A strategy is needed to ensure an ease of export of repository items into an institution's newly built repository.

There could be a continuum of support for institutions, from interacting directly with their researchers to encourage deposit and check metadata, etc., to allowing local 'editors' to check and approve papers being deposited, to incorporating policies conceived at the institution level to be implemented for that institution's users, to simply providing the storage space/software as a 'virtual' repository which local users may not even know resides off-campus, to wholly transferring all items deposited by users based at their institution (at the time of deposit) to the newly built IR.

The exit strategy varies along this same continuum. Perhaps a charging mechanism per institution could be introduced for those not ready or not wanting to take on full responsibility for an IR at the end of funding.

This stakeholder perspective also calls for training and liaison to institutions at an active level to encourage uptake of the IR responsibilities at institutional level.¹³

what the repository would supply (eg storage, search capacity, advice). The contract would have an end date with possible costed additional services/options.

¹¹ This has possible ramifications for choice of repository software option. For example, D-Space allows 'partitions' by community with a common search interface, whereas e-Prints can have different instances for different communities but it is not cross-searchable as is.

¹² Why would an institution implement their own Repository, when a national repository is tailored to and addresses their needs?

¹³ This might be provided in a kit form.

SP III- (ii): Universities as ‘subscribing’ Institutions

Key Characteristics

This is a variant perspective, with explicit contract with a given Institution. There would be access to a ‘vanilla/basic’ national repository service for those institutions that undertake to implement their own repository within the five year planning horizon. While authors at their institutions begin to gain experience depositing into an OA repository, the institution would be planning their individual policies and services for when they set up and begin running their own IR locally – with assistance from project staff. The subscription would take effect in years three, four and five, assisting an exit strategy for the national repository facility.

Ingest Service / Authorisation

Only members of UKHE/FE institutions that had committed to create their own repository would be allowed to deposit.¹⁴ This might involve a formal agreement (SLA) between the institution, the JISC and the national repository. The decision whether to allow students access would be taken centrally.

There would be a single deposit interface regardless of the institution they belonged to. If a user tries to deposit an item for which there is an appropriate existing IR, the page will redirect the user to their own IR’s website automatically.

Submission approval and added metadata or subject classification and other procedures will be carried out centrally by project staff to ensure consistency across the repository items, but institutions can begin providing services to their users through ‘assisted deposit’ in which library or other institutional staff deposit items on behalf of the author.

Although guidelines for what can be deposited and what not, would be decided centrally, when holdings are transferred into the newly established (contractually required) IRs then local staff can become ‘gatekeepers’ or ‘stewards’ for their institutional assets.

The deposit license would be the same for all participating institutions/authors.

Access Service / Discovery Metadata

Searching will be enhanced by author’s keywords and full-text. Browsing by subject will be made possible, through subject classification assignment at ingest.

Storage / Management

There would be no customisation of the repository software to suit the needs of individual institutions. The functionality offered would be kept to a minimum and would not support institutional processes (for example, RAE).

Transfer Service / Exit Strategy

A strategy is needed to ensure an ease of export of all repository items into an institution’s newly built repository.¹⁵

There could be a continuum of support for institutions, as for Stakeholder perspective III, but is more likely to be based on the project working directly with institutional support staff who will encourage and assist deposit within the institution, with support to the institution for export of their assets at the point the institutional repository is introduced.

This stakeholder perspective also calls for training and liaison with institutions at an active level to encourage uptake of IR responsibilities at the institution but foresees wholesale transfer of holdings to an IR by a specified time rather than a continuum of support levels.

¹⁴ But a national repository is intended for academics who *don't* have access to an Institutional Repository?

¹⁵ Ease of export is however difficult to guarantee: The national repository can only export the metadata it has, which may not fit with the requirements of the particular Institutional Repository. The national repository will also likely need to retain some information (metadata, binary object, or a redirect) after export to the IR.