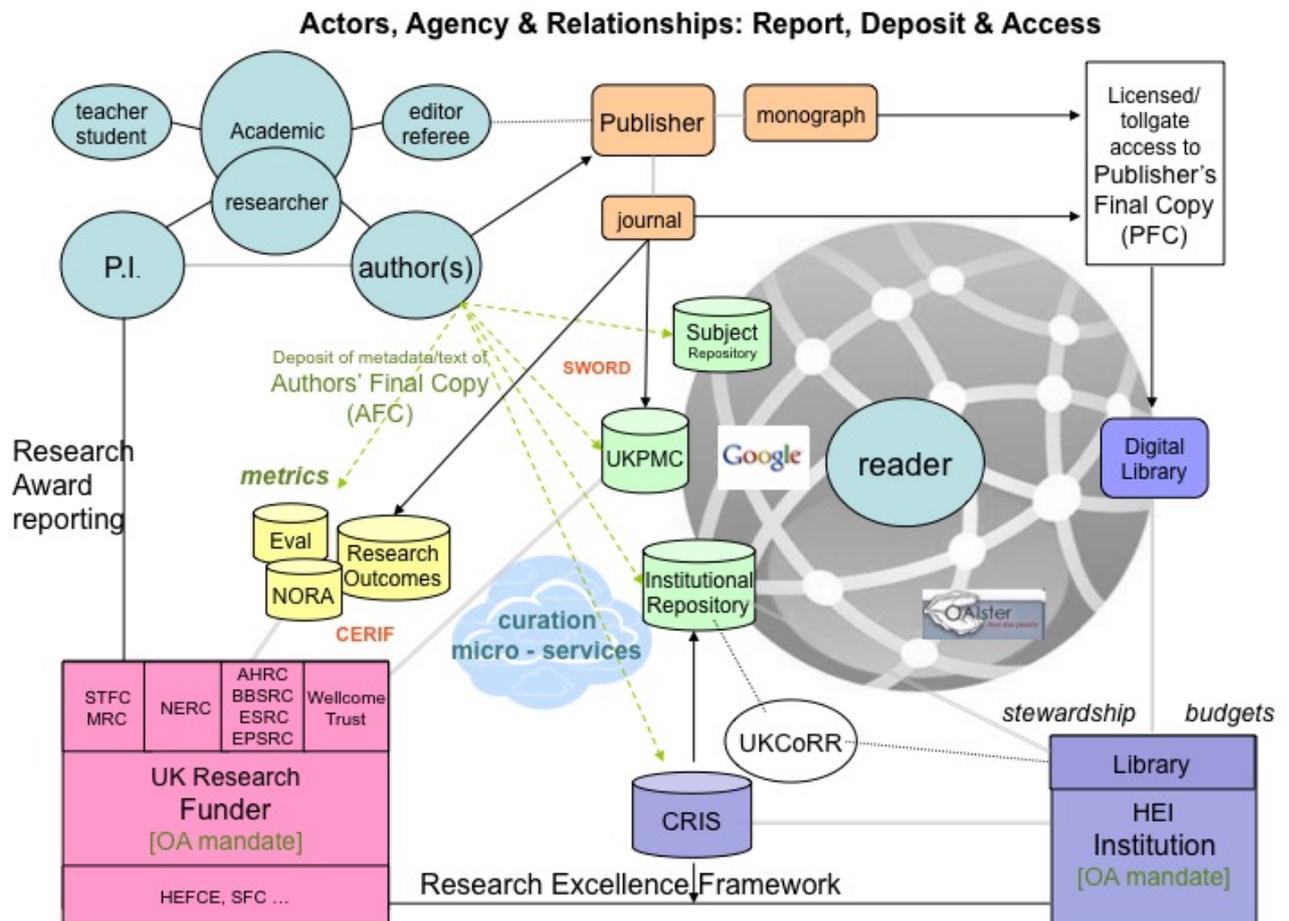


## Mapping the Repository Landscape

1. We want to test our understanding of the ecosystem in which repositories for research literature have now been operating since their emergence onto the scene<sup>1</sup>. The main actors and stakeholders each have varying purposes and drivers, and the following graphic is intended to help us review some of the main workflows and explore both what is missing as 'information deficit' and the ways in which there can be 'cross-pollination' between workflows.



2. The story begins, top left, with research proposal and award of grant giving obligation of report by the **Principal Investigator (PI)** to the **Funder**, and the research award reporting workflow, connecting in part, as 'outcome of research', with submission of the (multi-) **authors' manuscript to a journal**, regarded as the flow of publication from funded research<sup>2</sup>.
3. Building upon work done by the SONEX Group on deposit opportunities<sup>3</sup>, the focal point of interest for Open Access repositories is the **deposit of the Authors' Final Copy (AFC)**<sup>4</sup>, shown as green dotted line, into the **Institutional Repository (IR)**.
4. Noting the (growing) significance of the Current Research Information System (**CRIS**), **Institutions** and the **Funders** represent two key stakeholder groups, with some variety of motive, noting that all institutions are not of one type: ranging from the large research intensive

<sup>1</sup> Jones, R, Andrew, T and MacColl, J. *The institutional repository*. Chandos Publishing, Oxford, 2006.

<sup>2</sup> Burnhill, P., & Tubby-Hille, M. (1994). On measuring the relation between social science research activity and research publication. *Research Evaluation*, 4(3), pp130-152. See page 8 of text available online as [eig.sdss.ac.uk/projects/rapid.pdf](http://eig.sdss.ac.uk/projects/rapid.pdf)

<sup>3</sup> Burnhill, P, Castro, Pablo de, Downing, J, Jones, R, Sandfær, M, Handling Repository-Related Interoperability Issues : the SONEX Workgroup, <http://hdl.handle.net/10016/9257>

<sup>4</sup> The 'multi-authored and multi-institutional work' is the default object

<http://sonexworkgroup.blogspot.com/2011/03/jisc-repository-deposit-programme.html>

universities to less well resourced small to medium sized institutions. Each of these two key stakeholders generates workflow and controls elements of metadata needed by the other.

5. That is also true of Publishers, who wish the Authors' Final Copy to contain DOI link to the **Publisher's Final Copy (PFC)**, the citation to this published version is also wanted by the authors for purposes of impact. The **Research Excellence Framework (REF)** is noted for its importance for the institution, alongside its need to satisfy **compliance** with the requirements of **Open Access mandates** by Funders and Institutions.
6. **The Reader** must be an essential part of the picture - justification for the re-use of repository content, and of this initiative. We deliberately include the arrangements made by libraries and the industry for access to the Publisher's Final Copy, as context for the role of repositories in providing access to the Authors' Final Copy. Metrics on usage, as well as deposit, are key, but what is plain is the **role played by Google** and the like as *the* route to repository content: it is 'discoverability' of repository content by Google (etc) that is mainstream, keeping in mind the tack taken by harvesters and exposure within other aggregation clusters, eg Discovery.ac.uk.
7. Reminded of the *machine-as-user* of repository content format of content<sup>5</sup>, metadata format and modes of metadata disclosure also matter. We have therefore added SWORD and CERIF into the picture. Use of RDF and well as crosswalk to protocols other than OAI-PMH could be added. *Publishing to the machine* as strategy also reminds us of continuity of access and preservation as libraries decide how to address their mission of stewardship, exercising archival responsibility for the IR contents.
8. At some future date we will look at what serve as common service components, addressing a range of functions: registries, authority files and identifiers; deposit tools and protocols; aggregation and discovery services; metrics; and services intended to ensure continuity of access to repository content. SHERPA RoMEO is an obvious example; clearly there are others. Our focus will be the UK but there is global context and much 'positive externality' in the international role that UK-based service components play in the repository space, and vice versa, eg OAlster. EPrints, DSpace and Fedora, as the main software platforms, can also themselves be thought of as contributors of cross- platform components.
9. Our sketch of the landscape has firm focus upon *research literature*, especially that resulting from funded research activity and made available under OA via the enabling role of IRs<sup>6</sup>. But there is the wider context in a broader definition of 'research output', to include e-theses, grey literature (PPTs and working papers) and newspaper articles – the latter signalling the importance of the 'impact' agenda<sup>7</sup>. With supplementary data (multimedia and datasets) in enhanced publication there are no hard lines in this.
10. We want to test this understanding with developers and managers of institutional repositories<sup>8</sup> (UK-CORR<sup>9</sup> and DevCSI<sup>10</sup>), and variety of other research managers (ARMA<sup>11</sup>). A Roundtable at Repository Fringe 2011<sup>12</sup> seemed a good place to start.

Contributors at EDINA & Edinburgh University Data Library:

Theo Andrew, Peter Burnhill, Sheila Fraser, Stuart Macdonald, Nicola Osborne, Christine Rees, Robin Rice, Adam Rusbridge, Ian Stuart, Robin Taylor and Gareth Waller. August 2011

---

<sup>5</sup> For example, pdf2text to convert pdf file to XML, but also with EPUB in mind.

<sup>6</sup> What is deposited is not always AFC; UKPMC requires 'version of record' to be deposited

<sup>7</sup> Hicks, D "The Four Literatures of Social Science" *Handbook of Quantitative Science and Technology Research*. Ed. Henk Moed. Kluwer Academic, 2004, [http://works.bepress.com/diana\\_hicks/16](http://works.bepress.com/diana_hicks/16)

<sup>8</sup> As well as a range of existing tools more directly associated with life-cycle preservation practices, and potential in deployment of Private LOCKSS Networks, there is need for services and components relating to high-availability hosting and backup facilities for repositories to support service continuity.

<sup>9</sup> United Kingdom Council of Research Repositories <http://www.ukcorr.org>

<sup>10</sup> Developer Community Supporting innovation <http://devcsi.ukoln.ac.uk/blog/about/>

<sup>11</sup> Association of Research Managers and Administrators <http://www.arma.ac.uk/>

<sup>12</sup> <http://repositoryfringe.org>, 3-4 August 2011, Edinburgh