# In the context of Open Access policies, what is a “repository”?

## Some definitions and principles

#### Neil Jacobs. May 2016

## Introduction

This document outlines some principles by which to assess whether arrangements to store and share research outputs constitute a “repository”. The term “repository” is used in several open access (OA) policies, including in the UK for the Research Excellence Framework, so it is important that, at least, the UK academic community has a clear understanding of what a repository is. This document is a thought piece, intended to support discussion and moves toward consensus, not a position statement.

## Background

In 2003, Clifford Lynch of the Coalition for Networked Information defined a university-based institutional repository as "a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members". While this is reasonably clear with respect to institutional repositories, can the word “university” be replaced by other words (eg, “funder”? “subject community”? “learned society”? “publisher”?) and the remainder of the sentence retain its meaning? In some cases the answer is clearly “yes” – arXiv is clearly a repository. Many would argue that Europe PubMedCentral and RePEc are also repositories (among other things). But what about SSRN? Academia.edu? ResearchGate? CHORUS? ScienceDirect, Portico?

That illustrative list, and this paper, are limited to services concerned with research outputs such as journal articles and working papers, not books, research data, software, methods, etc. In fact, this paper is even more limited in scope, it is specifically about **OA repositories** not, for example, “trusted” digital repositories focusing on preservation requirements, which might reference the OAIS model and meet ISO 16363:2012.

Some will argue that making distinctions between OA repositories and other systems is unhelpful. This paper does not address that question. It assumes that some people need to make the distinction between OA repositories and other systems, and it attempts to outline some ways in which they might do so. It does so at operational and strategic levels.

## Operational definitions

There are many, many operational definitions of OA repositories. In order to delineate some features of repositories, the following selection might be helpful:

* As noted above, Clifford Lynch defines an institutional repository as "a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members"[[1]](#footnote-1)
* Sarah Shreeves defines a repository as “a digital assets management system of some kind or a network of systems that allows for the deposit and subsequent distribution of digital files over the internet.”[[2]](#footnote-2)
* Queen Mary University of London’s web pages state that “an open access repository is a digital repository where the content is freely available to download and reuse (sometimes with restrictions), where no login or subscription is required.[[3]](#footnote-3)
* Smita Chandra, Librarian at the Indian Institute of Geomagnetism notes that a repository is an “open access digital archive on open source software; a managed, persistent way of making research, learning and teaching content with continuing value both discoverable and accessible.[[4]](#footnote-4)
* The Irish Research Council notes in its OA policy that “suitable repositories are those, both local and other, that provide free public access to, and make provision for long-term preservation of, published research findings. Suitable repositories should support interoperability with other repositories and with other research information and reporting systems and should be harvestable by national portal/s and international aggregators.”[[5]](#footnote-5)

While these definitions are selected more or less at random, they do represent views from around the world, and from experts, librarians and research funders. Abstracting from them, the following seem to be features of OA repositories:

1. Allow deposit of digital research outputs
2. Manage those digital research outputs
3. Disseminate digital research outputs over the internet
4. No login or subscription required to access outputs[[6]](#footnote-6)
5. Are fully interoperable with other research systems[[7]](#footnote-7)
6. Have some role with respect to preservation

One definition above notes that repositories use open source software. This is not universally agreed, and is addressed under “strategic principles” below.

## Strategic principles

As well as operational features, and sometimes implied by them, OA repositories are arguably also subject to more strategic principles that distinguish them from related systems. The following set of principles draw from the definitions noted above, plus other sources such as the principles proposed by Bilder, Lin and Neylon[[8]](#footnote-8) for open scholarly infrastructures, and the principles proposed by the Authors Alliance following the acquisition of SSRN by Elsevier in May 2016[[9]](#footnote-9). The use of these principles is specific for this purpose only, and is not intended to be a commentary on, or general interpretation of, those principles.

1. Mission: An OA repository’s sole mission is to enable the dissemination and preservation of academic research. It does not have other drivers that might, under some circumstances, conflict with this mission. This may limit the range of business models that are appropriate for organisations running OA repositories.
2. Coverage across the research enterprise: While generic infrastructure such as ORCID needs to transcend institutional, national and disciplinary boundaries, repositories do not. However, a repository does need a clear and stated scope, for example, based on an institution, a disciplinary community or those whose research is funded by a particular set of organisations. This scope should be defined by academically relevant criteria.
3. Stakeholder Governed: To ensure they meet their mission (see above), OA repositories are run by organisations that are accountable to the academic community relevant to their scope. That means that those organisations have open and transparent governance arrangements that demonstrate how they meet this accountability principle to the fullest reasonable extent. This includes full disclosure of plans related to the exploitation of works held by the repository, and of data related to those works and their use.
4. Author control of works: The repository acts on behalf of the author(s) of the works held, which implies that the repository requires the minimal non-exclusive permissions from the author(s) to meet its mission, there is author control over any removal or alteration to their work (except to comply with applicable laws), and the repository notifies authors of all its relevant policies, including any changes to those policies.
5. Mission-consistent revenue generation: as an overarching principle, this implies that repositories are services, not projects, and are resourced on the basis of their academic mission. This can include revenue from the provision value-added services, providing those are within the academic mission.
6. Open source: While Bilder *et al*’s principles call for the use of open source software, in the case of OA repositories this seems unnecessary if the other features and principles are strongly applied.
7. Open data: OA repositories should make fully open and practically available, using appropriate standards and without the need for subscription or login, the metadata and content of the research outputs held, respecting any legal constraints. They should also have arrangements in place, or planned, to share usage and other transaction data within the constraints of relevant ethical (eg privacy) frameworks.
8. Path dependency: infrastructure, including repositories, both enables and constrains practice, including research practice. When infrastructure becomes embedded, it creates “path dependency”, that is, it sets a certain trajectory, making some practices easier and some harder. Radical innovation can be necessary to overcome path dependency *en route* to a more open, collaborative and networked science environment, and OA repositories should both be a part of that, and are subject to it.
1. See Wikipedia. “Institutional Repository”: <https://en.wikipedia.org/wiki/Institutional_repository> retrieved 31 May 2016. [↑](#footnote-ref-1)
2. Shreeves, S.L. (2014) The Role of Repositories in the Future of the Journal, in Cope, B and Phillips, A. (2014) The Future of the Academic Journal, 2nd edition. Cambridge, UK: Chandos Publishing. Available from: <https://www.ideals.illinois.edu/bitstream/handle/2142/49968/Shreeves_RoleofRepositoriesChapter12.pdf?sequence=2> retrieved 31 May 2016. [↑](#footnote-ref-2)
3. See: Queen Mary, University of London. “What is a repository”: <http://www.library.qmul.ac.uk/what-repository> Retrieved 31 May 2016 [↑](#footnote-ref-3)
4. Chandra, S. (2013) Institutional Repositories, an overview. <http://www.slideshare.net/smtcd/institutional-repositories-17349083> Retrieved 31 May 2016 [↑](#footnote-ref-4)
5. Irish Research Councils policy relating to the open access repository of published research. <https://www.education.ie/en/The-Education-System/Higher-Education/The-Open-Access-Repository-of-Published-Research.pdf> Retrieved 31 May 2016 [↑](#footnote-ref-5)
6. In fact (4) is probably a subset of (5), since full interoperability cannot be achieved where subscription / login requirements exist, which prevent straightforward machine-to-machine interaction. [↑](#footnote-ref-6)
7. While it is a work-in-progress to achieve full semantic interoperability, this feature is perhaps currently met by using / providing the most open, documented, and appropriate APIs and metadata standards. [↑](#footnote-ref-7)
8. Bilder G, Lin J, Neylon C (2015) Principles for Open Scholarly Infrastructure-v1, retrieved 17 May 2016, <http://dx.doi.org/10.6084/m9.figshare.1314859> [↑](#footnote-ref-8)
9. Authors’ Alliance (2016) Principles for reassuring authors of SSRN-posted papers under Elsevier’s ownership, retrieved 24 May 2016: <http://www.authorsalliance.org/2016/05/19/principles-for-reassuring-authors-of-ssrn-posted-papers-under-elseviers-ownership/> [↑](#footnote-ref-9)