

Stevan Harnad, Universite du Quebec a Montreal/University of Southampton
Statement for the "First DRIVER Summit", Panel Discussion, 2008-01-16

THE FEEDER AND THE DRIVER: Deposit Institutionally, Harvest Centrally

DRIVER is designing an infrastructure for European and Worldwide Open Access research output, stored in [institutional and disciplinary repositories](#), now increasingly under [institutional and research-funder mandates](#). It is critical for DRIVER to explicitly take into account in its design (as some research funders have not yet done, because they have not yet thought it through) that institutional and disciplinary (central) repositories (IRs and CRs), although they are fully interoperable and at a par in that respect, nevertheless play profoundly different roles.

Universities and research institutions are the FEEDERS-- the primary providers of research, funded and unfunded, in all disciplines -- for both kinds of repositories (IRs and CRs).

This difference in role and function must be concretely reflected in the design of the DRIVER infrastructure. The primary locus of deposit for all research output is the researcher's own institution's IR (except in the increasingly rare case of institutionally unaffiliated researchers). Thanks to OAI-interoperability, the metadata for those deposits, or even the full-text deposits themselves, can also be harvested by (or exported to) any number of CRs -- discipline-based CRs, funder-based CRs, theme-based CRs, national CRs, European CRs, global CRs.

Neither IRs nor CRs will fill without deposit mandates. This is a hard lesson, that has been learned very late (NIH, for example, made the mistake of requesting rather than requiring deposit, the NIH policy failed, and three years of research impact was consequently lost); but the lesson has now at long last indeed been learned. So the number of institutional and funder mandates is now set to grow dramatically. Institutions of course always mandate deposit in their own IRs. Many funders have mandated deposit, indicating that deposit can be in either IRs or CRs. But a few funders still stipulate, dysfunctionally, that deposit must be in CRs.

This is a symptom of not having thought OA through. Funders are of course greatly to be commended for mandating OA, but their short-sightedness on the question of locus and means of deposit needs correction, and DRIVER can and should help with this, pre-emptively, rather than blindly following the unreflective and incoherent trends in the air today. Indeed DRIVER must take a coherent position, if it wants OA content to be provided and OA repositories to be filled, reliably and fully.

The model that DRIVER should adopt in designing its infrastructure is "Deposit Institutionally, Harvest Centrally." That is the way to scale up -- simply, swiftly, systematically and surely -- to 100% OA. I presented the reasons in detail in [my talk](#). Here I only summarise the principle points:

Institutions (i.e., universities and research institutes) are the providers -- the source -- of all research. Institutions have a direct interest in showcasing and managing their own research output, but they have been even more sluggish than funders in adopting mandates. If funders mandate central deposit, they neither cover all of OA output nor do they collaborate coherently with the providers (the institutions) to scale up systematically to providing OA to all of their institutional research output. The OAI protocol makes it possible to harvest content from all

OAI-compliant repositories. That is the coherent, systematic pattern of content provision for which DRIVER should be designed, not an incoherent patchwork of arbitrary institutional and central depositing and repositories that will neither scale up to all of OA nor accelerate its attainment.

Not all research is funded; not all research fits into defined disciplines; disciplines are not all independent. Disciplines, being overlapping and redundant, would entail that discipline-based depositing had to be overlapping and redundant. Depositng can be mandated once, but not multiply. The natural way to ensure that a paper is present in multiply loci (institutional, (multi)-disciplinary, national, etc.) is to deposit it at source – i.e., institutionally – and then *harvest* or *import* its metadata (or both its metadata and the paper itself) into whatever CRs we decide we need. That is what the OAI interoperability protocol itself was designed for.

And, not to put too fine a point on it, the very notion of Central Repositories already betrays something of a misunderstanding of the online medium: Is Google a central repository? Is it a repository at all? Do people deposit directly in Google?

[OAIster](#), [Citebase](#) (and many other central OAI services like them) are an even better model: OAIster and Citebase were explicitly designed to be [OAI service-providers](#) -- functional overlays on the distributed OA content-providers. Do CRs -- disciplinary, interdisciplinary, national and international -- really need to be any more than that?