

RSpace + iRODS

A scalable, flexible and versatile solution that facilitates data and metadata interoperability and is suitable for deployment in conjunction with a wide range of e-infrastructures and Research Commons

Rory Macneil¹<https://orcid.org/0000-0002-8429-096X> and Terrell Russell²

¹ Research Space, Edinburgh, Scotland

² Renaissance Computing Institute, University of North Carolina at Chapel Hill, USA

Abstract.

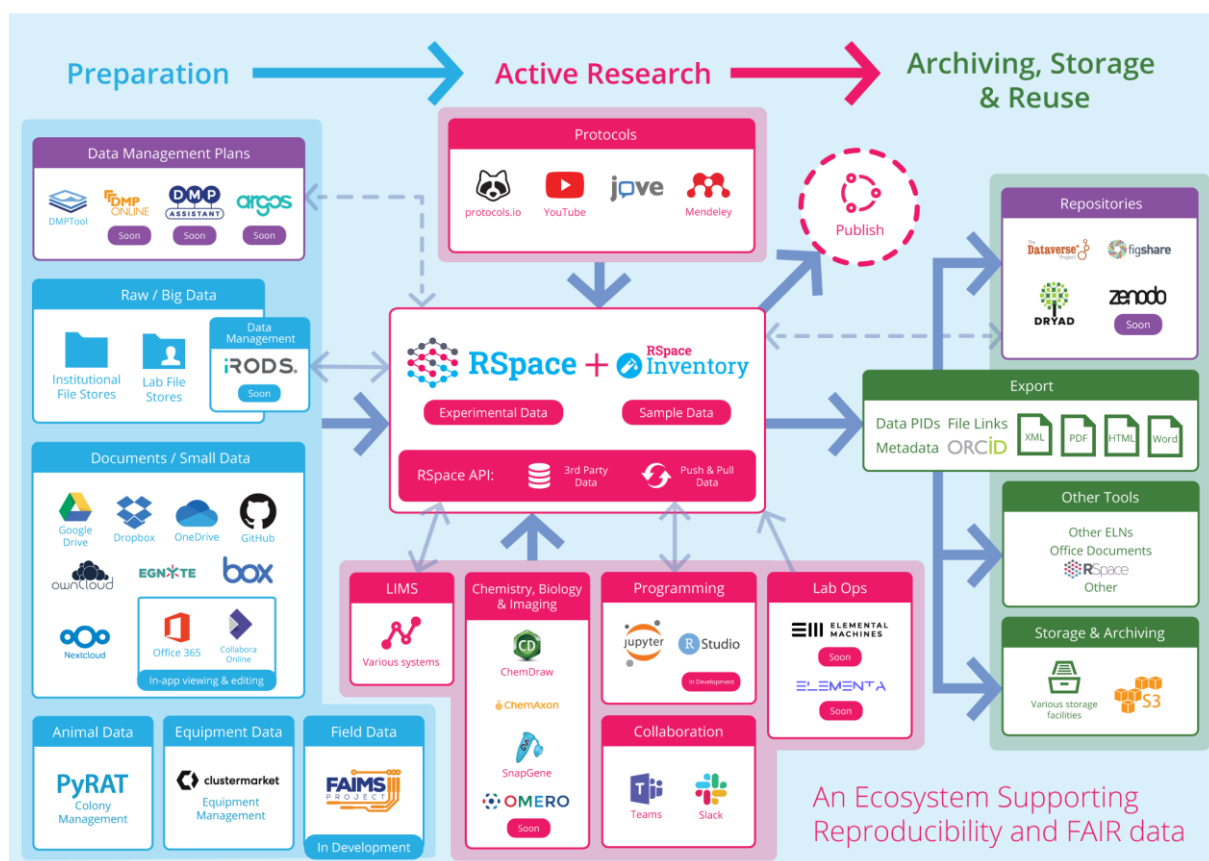
Research infrastructures enabling scalable sharing of data are proliferating, at the institutional, project/consortium and national/international levels. Many of these are domain-specific, but there also is a growing focus on Research Commons that enable general data sharing across domains. Examples include the EUDAT Collaborative Data Infrastructure in Europe, the Gakunin RDM platform in Japan, and the Research Commons planned by Canada's Digital Research Alliance (DRA).

In some cases, such as Gakunin RDM, Research Commons are architected as an integrated set of complementary services, but in most existing and planned Research Commons a disparate set of unconnected services is offered. This paper describes the integration between [RSpace](#), an active content management digital research platform, and [iRODS](#), a policy-driven data management platform, and explains how it is designed to serve as a flexible connecting component that ties together other services that make up a Research Commons.

The paper discusses how, by tying together otherwise unconnected services, inclusion of RSpace + iRODS in Research Commons enables streamlined flows of data and metadata between services, enhancing FAIR principles. This will be illustrated by considering inclusion of RSpace + iRODS in the two specific examples of Canada's proposed Research Commons and the EUDAT Collaborative Data Infrastructure. In both cases the interaction between RSpace, iRODS, and individual services provided by the Commons will be discussed, and a comparison will be made between the two Commons and the benefits derived from the inclusion of RSpace + iRODS.

RSpace and iRODS

As noted, RSpace is a digital content management platform designed to interoperate with other research tools and resources. RSpace and the ecosystem and workflows it supports is captured in this graphic:

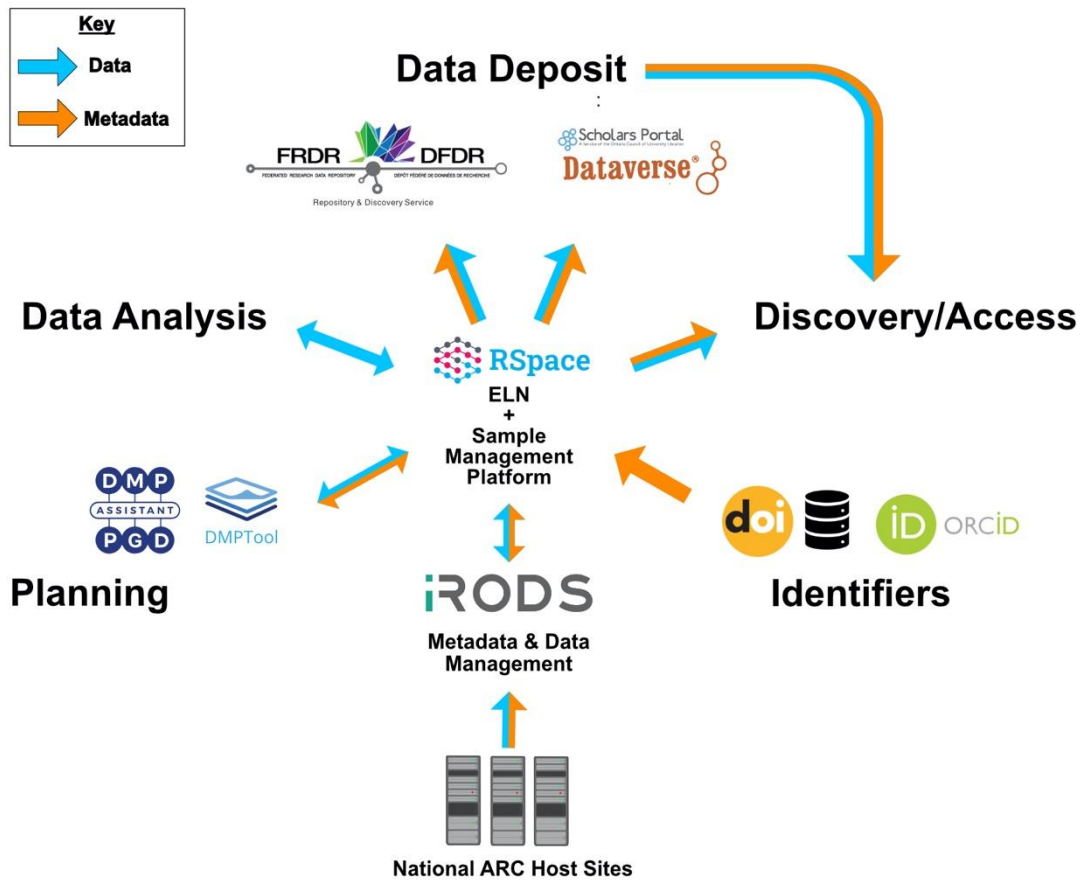


The Integrated Rule-Oriented Data System (iRODS) is open source data management software used by research, commercial, and governmental organizations worldwide. It provides data virtualization with a logical namespace that abstracts the disparate physical storage systems that house the files, a metadata catalog to hold system and user-defined metadata to aid in discovery, and a policy engine to apply admin and user-defined rules to data, based on the metadata in the system. This separation of concerns affords a flexibility to satisfy a wide variety of use cases and scenarios across many diverse domains.

The existing integration between RSpace and the iRODS logical namespace solves the 'broken links' problem by ensuring that, even if files in external resources linked to from RSpace move location the integrity of the link is maintained, thus enhancing the quality and durability of the research record and discoverability. A second phase of the integration is currently being implemented and will be available at the time of the conference in September. This second phase will enable exposure of PIDS and other kinds of metadata from RSpace to iRODS, and association of the metadata with data and metadata in the entire ecosystem of files to which iRODS has access, which could include all the files held by a university, in a large project or collaboration, or a national facility or facilities.

Canada's Research Commons

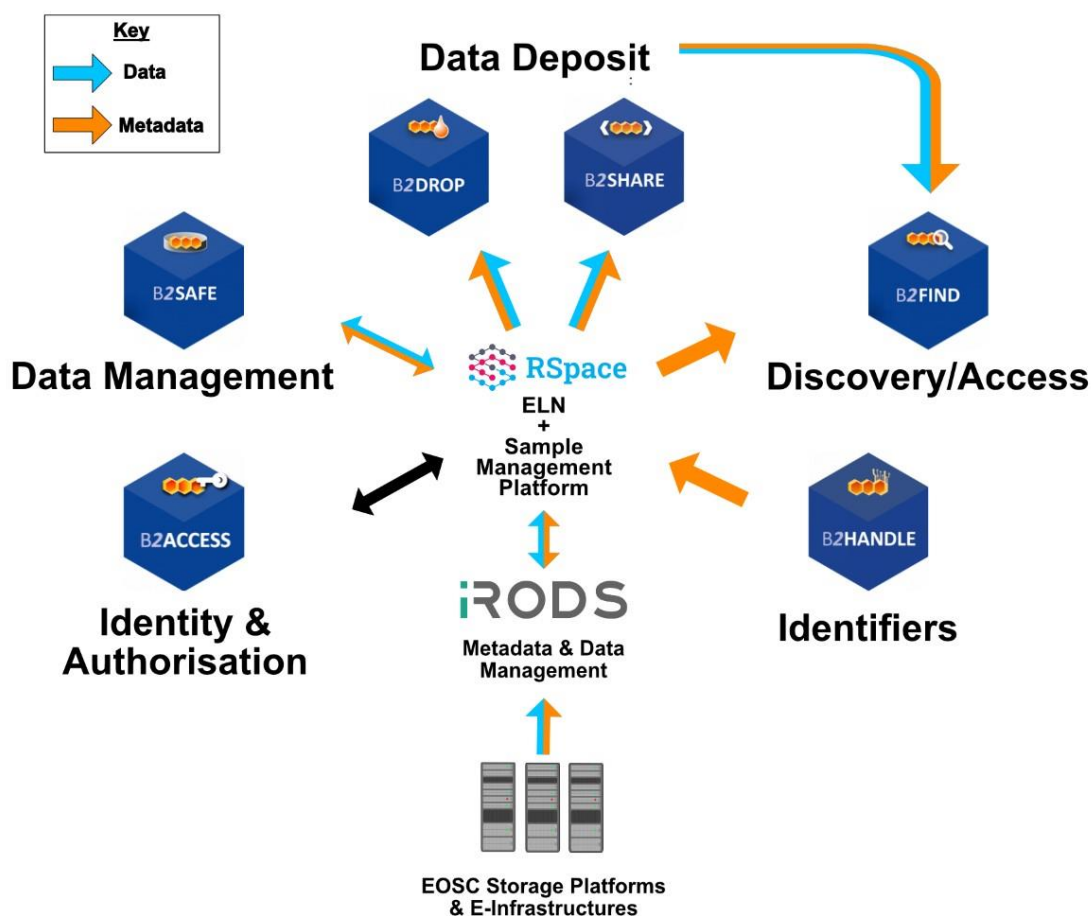
A vision for Canada's Research Commons is set out in this presentation by Mark Leggott, International Director of Canada's Digital Research Alliance, [Digital Research Alliance of Canada as a Research Commons](#). The Commons is to be built around five existing services provided by the DRA: a data storage underpinning in five national data centers; a data management planning tool/service; a national PID delivery service that includes ORCID and DataCite DOIS, and two national data repositories. The Research Commons will be delivered in partnership with Canadian universities and for use by the universities and their researchers. This graphic depicts the vision for inclusion of RSpace + iRODS as a central connecting and catalyzing force in the Commons:



EUDAT Collaborative Data Infrastructure

The [EUDAT CDI](#) consists of six services: B2Share is a data repository; B2Handle is a PID delivery service, B2Drop is a file sharing service; B2Access is an identity and authorization service; B2Safe, a kind of omnibus data policy/management service and actually *is* iRODS; B2Find is a data discovery service. As is the case in the envisioned DRA Commons, the six services are delivered as separate services, although plans are underway to enable some limited integration between the services.

This graphic depicts the vision for inclusion of RSpace + iRODS as a central connecting and catalyzing force in the EUDAT CDI:



Concluding comments

It is interesting and important that all three Commons referred to in this paper, the EUDAT CDI, the DRA Research Commons, and the Gakunin RDM, are built around resources that are (a) pre-existing and (b) intended for general use generally by researchers in multiple domains. The Commons have adopted, and in some cases adapted, existing resources, but they have not attempted to reinvent the wheel by duplicating existing resources, and the Commons are all intended for use in multiple research domains.

A second point of interest is that the EUDAT CDI and the DRA Research Commons both include a similar common set of core services relating to provision of data storage, data management planning, repositories, and provision of PIDS.

We believe that use of pre-existing, generalist tools both represents best practice and is the approach best suited to creating truly scalable RDM solutions. The fact that all three of the Commons mentioned have adopted this approach also facilitates incorporation of RSpace into the Commons, since RSpace already integrates with many of the resources included in both the EUDAT CDI and the DRA Research Commons, including data storage facilities, data management planning tools, data repositories, and PIDS delivery services.

It is, moreover, encouraging that both the EUDAT CDI and the DRA Research Commons share these resources as core components. The inclusion of RSpace + iRODS as an integrating mechanism into these two early attempts to create Research Commons could as described transform these Commons into a more powerful, comprehensive service. It could also provide a model for the development of similar initiatives in other jurisdictions, and because of the shared use of common elements, a series of Commons that are themselves interoperable. Thus this paper is also relevant to the Connecting RDM track.

Keywords: Data, Metadata, Interoperability, FAIR Research Commons