# R4L: The Repository for the Laboratory

S.J. Coles[a*], J.G. Frey[a], M.B. Hursthouse[a], L.A. Carr[b].

[a]School of Chemistry, University of Southampton, UK.; [b]School of Electronics & Computer Science, University of Southampton, UK.

## The Data Overload Problem

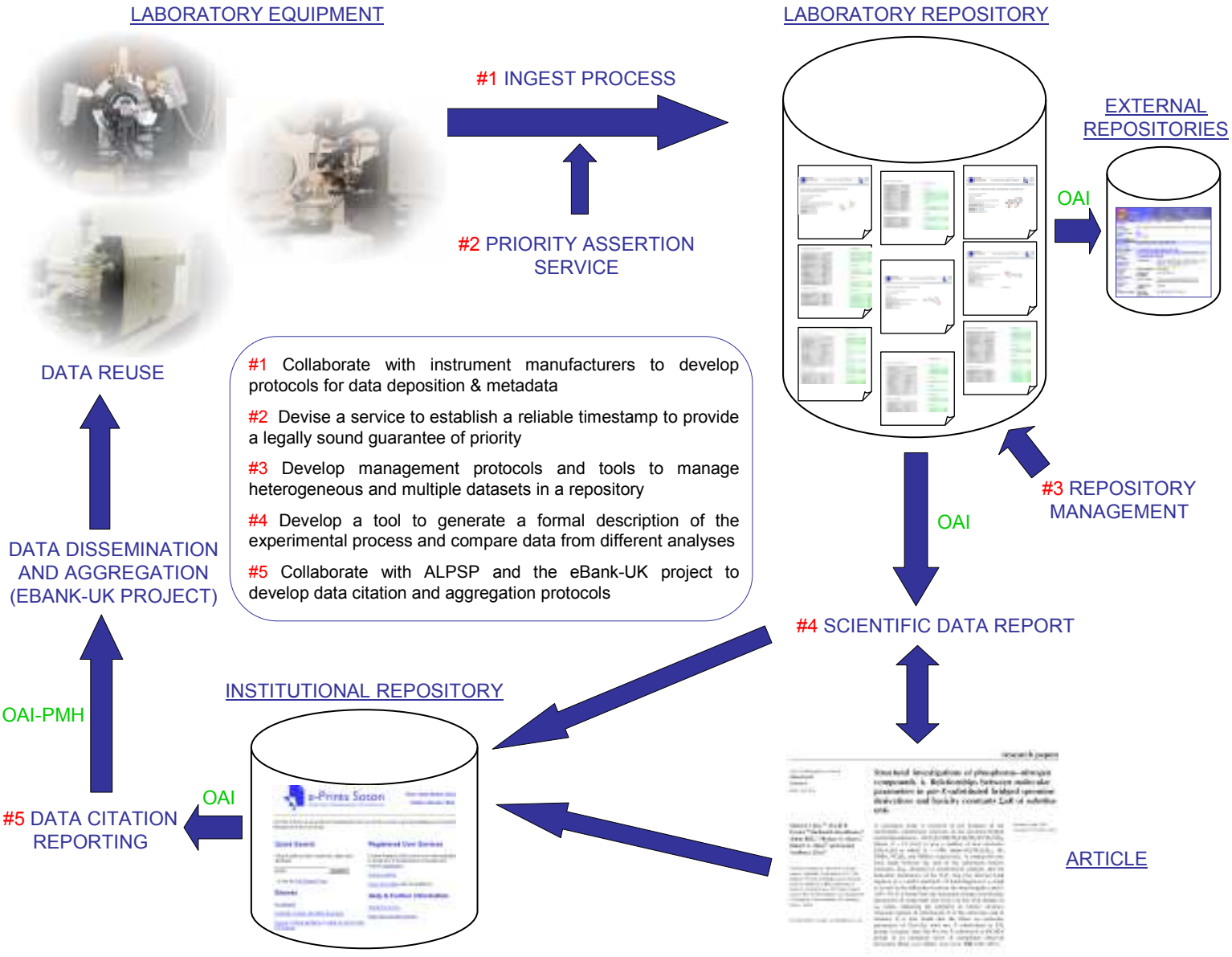
Recent advances in scientific instrumentation and computational resources have caused a huge increase in experimental data generation. Conventional methods for management and publication of scientific data are unable to match this new pace, causing a publication bottleneck. This problem will become more severe with developments in high throughput experimentation aided by eScience. The user community is therefore being deprived of valuable information, and the funding bodies are getting a poor return for their investments!

The Open Archive Initiative (OAI) approach offers a solution to this problem through publically accessible archives. Currently a method for disseminating, via the internet, scholarly and research output in the form of articles, the OAI approach also provides a potential mechanism for dissemination of, and unhindered access to, the experimental data underpinning these articles.

## R4L as a Solution

The data cycle outlined below describes the R4L approach to the data overload problem. The # comments indicate the focus areas where the project will develop procedures to generate outputs to complete the cycle.

Methodologies will be developed in collaboration with instrument manufacturers for the capture of scientific information and generation of metadata, as it is created. The R4L priority assertion service will timestamp the data to establish time, methodology and institution. Data is then deposited in a repository containing a schema describing the data and experiment. This repository would be capable of hosting many different experimental schemas and thus would have to manage a very heterogeneous data holding. A report generation tool would then be developed that could aggregate datasets from different types of study (i.e. different analyses on a chemical compound) and present salient information for interpretation. This report would be deposited in an Institutional Repository and protocols and tools for citation reporting of the data would be developed. Methods for the data to enter the scholarly knowledge cycle will be developed in collaboration with eBank-UK.




## Project Outcomes and Benefits

The project will have significant impact on a number of different parties and the scientific community as a whole. R4L will use chemistry experiments as its exemplar and will build on the experience and outcomes gained from involvement with eBank-UK. A number of instrument manufacturers will benefit from the development of automated links to a managed repository. The institution will be able to protect its rights over data by means of the priority assertion service. The repository will hugely benefit laboratory technicians and managers by providing a total, automated and accurate infrastructure for the gathering and management of data. Researchers will benefit from immediate and unhindered access to data pertaining to a study and will be able to generate reports that will facilitate analysis, write-up and dissemination. Thus through data citation and OAI publication the scientific community will benefit enormously from being able to discover and access ALL digital data arising from research.