



Problem: scientific data publication bottleneck

Aim: to enable the publication, discovery and re-use of scientific data

Only a small percentage of data generated in many scientific communities makes it into the published literature. Publication in the mainstream literature still offers only *indirect* (and often expensive) access to this data. As a consequence the user community is deprived of valuable information.

Outputs

- Creation of an Institutional Repository of Scientific Data – ‘publication at source’
- Modified eprints.org software
 - ❖ Supports the submission of scientific (crystallography) data sets
 - ❖ Autogenerates metadata
 - ❖ Provides a navigable interface to the repository
- Demonstrator aggregator service
 - ❖ Harvests metadata using OAI-PMH
 - ❖ Brokers access to scientific data
- Linking made between datasets and articles

Benefits

- Improved access to scientific data and timely dissemination of data results
- Streamlined access to data through a common interface
- Re-use of data by other scientists
- New ways to assess and validate scientific results
- Curation of data built into scientist's workflow
- Requirements for next-generation open-access tools are better understood
- Cross-disciplinary collaboration by exchanging data with wider communities

Outcomes:

- Other Scientific Communities** can adopt approach
- Publishers** build services based on harvested metadata
- Learned Society Journals** focus on discussion, link to datasets

What must happen:

- ❖ Engage crystallography community to encourage use of 'open data archives'
- ❖ Apply the data model to other scientific data
- ❖ Promote standards-based approach to other scientific communities
- ❖ Work with publishers and learned societies to explore mutual benefits and services
- ❖ Investigate impact on the refereeing process
- ❖ Explore business models for institutional ePrint and eData repositories

<http://www.ukoln.ac.uk/projects/ebank-uk/>