Crystallographic EPrints

Simon Coles, Jeremy Frey, Michael Hursthouse, Les Carr, Christopher Gutteridge
Combechem 'Associated' Project, University of Southampton

The Publication Problem

Recent advances in crystallographic instrumentation and computational resources have caused an explosion of crystallographic data, as shown by the recent exponential growth of the Crystallographic Structural Database (CSD). The traditional peer review methods of dissemination of chemical data are unable to keep up with this new pace of data generation and are resulting in a publication bottleneck. This problem will become even more severe with developments in high throughput chemistry (Combichem) and the impact of eScience chemistry (Combechem). As a result of this situation, the user community is deprived of valuable information, and the funding bodies are getting a poor return for their investments!

The Open Archive Initiative (OAI) approach of EPrints offers a solution to this problem through EPrints archives, which are currently a method for disseminating scholarly output that cannot enter the public domain through conventional routes, by means of publically accessible archives of information.

Data Publication @ Source

Crystallographic EPrints use the OAI concept to make available ALL the data generated during the course of a structure determination experiment.

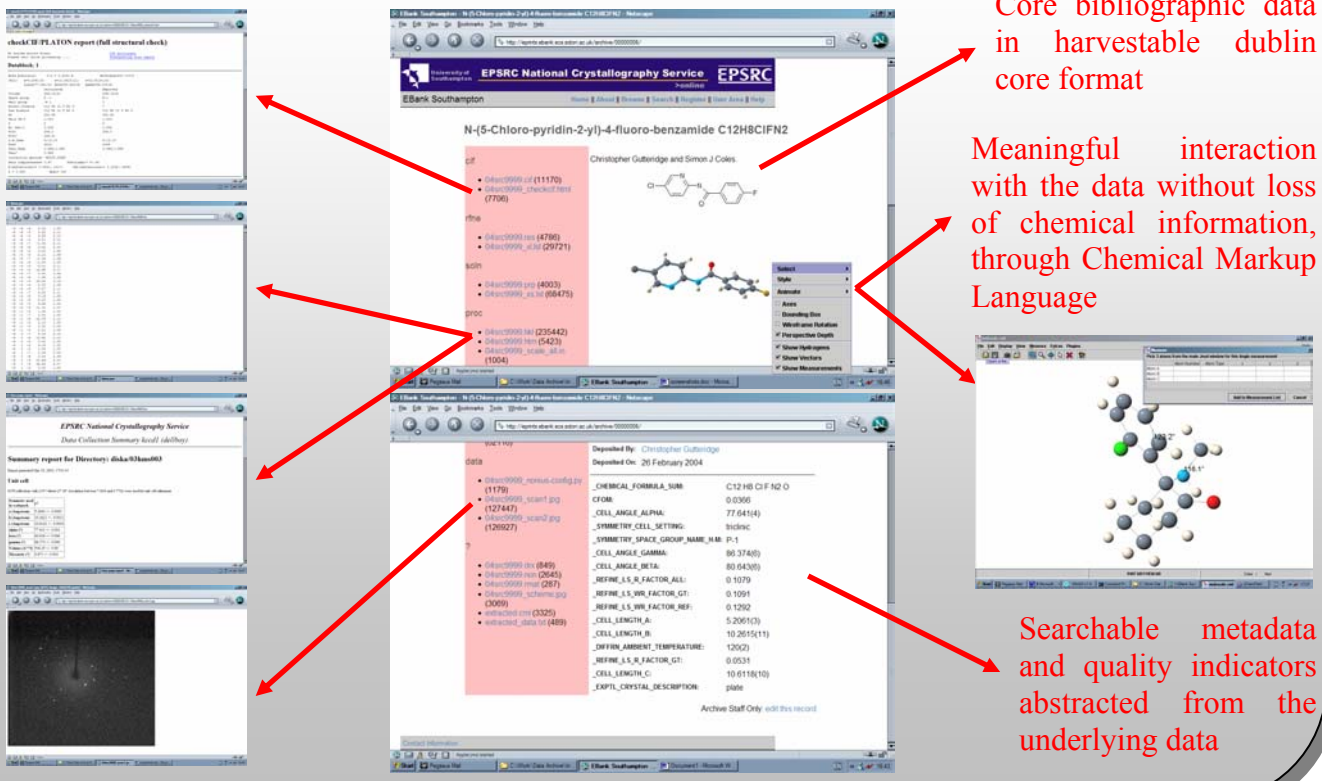
That is: the publishable output is constructed from all the raw, results and derived data that is generated during the course of the experiment.

This presents the data in a searchable and hierarchical system. At the top searchable level this metadata includes bibliographic and chemical identifier items which allow access to a secondary level of searchable crystallographic items which are directly linked to the associated archived data.

Hence the results of a crystal structure determination may be disseminated in a manner that anyone wishing to utilise the information may access the entire archive of data related to it and assess its validity and worth. This way the world becomes the peer reviewers!

The Crystallographic EPrint: Proving the Concept

Making the data available



Core bibliographic data in harvestable dublin core format

Meaningful interaction with the data without loss of chemical information, through Chemical Markup Language

Searchable metadata and quality indicators abstracted from the underlying data