



**III Międzynarodowa
Konferencja OA**

**III Międzynarodowa Konferencja Open Access w Polsce
„Otwarta nauka i edukacja”**

13-14 marca 2012, Bydgoszcz, Polska

**III International Conference Open Access in Poland
“Open learning and education”**

March 13-14, 2012, Bydgoszcz, Poland

**Peter Murray-Rust, PhD
Unilever Centre for Molecular Informatics
Cambridge University, United Kingdom
Open Knowledge Foundation**

Open data and the scholarly literature

The phrase “open data” was almost unknown 6 years ago but is now universal and heavily promoted by governments, funders and some academics. Some disciplines such as biosciences (genes, proteins, etc.) have been practicing openness for many years, enhanced by the *Bermuda Declaration*. Others such as astronomy have built community systems such as the virtual observatory, while the high energy physics programs have invested very large amounts of effort in information capture and processing. Many central facilities such as synchrotrons are also heavily committed to making their outputs open.

However most current scientific data is still unpublished in this manner. We estimate that most of it is never published at all, and that a lot of the rest is probably only accessible in theses and dissertations. For the rest published in scholarly publications, most of the actual data is omitted. A small amount may be printed as numbers or tables in text, and some in tables but this is generally far less than required to replicate the experiment or re-use the data in a productive manner. A few journals require the co-publication of “supporting information” and somewhat more permit it, but other journals actively refuse to publish data as opposed to full-text.

There are useful prototypes of how data can be co-published with journal articles in specialist repositories and an increasing number of these are appearing. But still the biggest problem is the scientists themselves. Most scientists find the labour of preparing public data to be heavy (we agree) and not cost-effective. There is little reward for doing so and there are perceived dangers.

The problem is compounded in “long-tail” science where researchers work in small groups (say 6-20) and do not have any community supporting data publication. Several things have to happen before they will publish data:

Incentives. There may be sanctions, such as failure to renew grants but it is more effective to provide positive incentives.

Infrastructure. One component is domain-specific repositories. But there also need to be tools and ontological support.

Training and human support. This is a complex field and researchers need help — they cannot work it out themselves.

Culture of openness. Until the process of publishing data is open and results in open data the system will not work efficiently.

Unfortunately relatively few publishers are actively supporting the use of the scientific literature for publishing data and most of the larger ones deliberately make it hard (technically and legally) to extract data from the current literature.



III Międzynarodowa
Konferencja OA

III Międzynarodowa Konferencja Open Access w Polsce
„Otwarta nauka i edukacja”

13-14 marca 2012, Bydgoszcz, Polska

III International Conference Open Access in Poland
“Open learning and education”

March 13-14, 2012, Bydgoszcz, Poland

This presentation will be given in absentia. There will be a supporting paper which carries the first publication of our studies on publisher practices regarding text and data extraction from the literature. This will be supported by a video surveying some of the current practices, especially those which show promise of taking this forward in a positive manner.

Biography

http://en.wikipedia.org/wiki/Peter_Murray-Rust
<http://www.ch.cam.ac.uk/person/pm286>



Peter Murray-Rust is a contemporary chemist born in Guildford in 1941. He was educated at Bootham School and Balliol College, Oxford. After obtaining a Doctor of Philosophy he became lecturer in chemistry at the (new) University of Stirling and was first warden of Andrew Stewart Hall of Residence. In 1982 he moved to Glaxo Group Research at Greenford to head molecular graphics, computational chemistry and later protein structure determination. He was professor of pharmacy in the University of Nottingham from 1996-2000, setting up the Virtual School of Molecular Sciences. He is now reader in molecular informatics at the University of Cambridge and senior research fellow of Churchill College.

In 2002, Peter Murray-Rust and his colleagues proposed an electronic repository for unpublished chemical data called the World Wide Molecular Matrix (WWMM). In January 2011 a symposium around his career and visions was organized, called *Visions of a Semantic Molecular Future*. In 2011 he and Henry Rzepa were joint recipients of the Herman Skolnik award of the American Chemical Society.