

RSCPublishing

**Response to Office of Science and Technology Policy (OSTP) Request for  
Information: Public Access to Digital Data Resulting From Federally Funded  
Scientific Research**

**On behalf of the Royal Society of Chemistry, UK**

To: Office of Science and Technology Policy (OSTP)  
Washington, DC 20502, USA

via e-mail to: [digitaldata@ostp.gov](mailto:digitaldata@ostp.gov)

9<sup>th</sup> January 2012

From:

*James Milne PhD*  
Editorial Director & Acting Managing Director  
RSC Publishing  
Royal Society of Chemistry  
Thomas Graham House  
Science Park, Milton Road  
Cambridge, CB4 0WF, UK

## About RSC Publishing

RSC Publishing is one of the largest and most dynamic publishers of chemical science information in the world. We publish 34 international peer reviewed scholarly journals, approximately 95 scientific books per annum, two highly acclaimed magazines, and a number of successful databases.

### Not-for-profit

We are a not-for-profit publisher wholly owned by the Royal Society of Chemistry. Our authors, readers and customers are truly international and our publishing activity dates back to 1841.

### Authoritative

RSC Publishing is a member of ALPSP, the Association of Learned and Professional Society Publishers, and we adhere to the ALPSP principles of scholarly-friendly journal publishing practice.

All research articles published by the RSC are peer reviewed. The journals are considered to be of the highest standards in their field, with an average impact factor of an impressive 5.4. Through the professional management of the publishing process, from submission through to publication, RSC content satisfies the pillars of scholarly publishing:

- Certification (validation of quality and integrity)
- Registration (recognition of achievement)
- Accessibility (unparalleled online access, worldwide)
- Archiving (reliable perpetual accessibility)
- Navigation (industry leading services to identify content)

### Award-winning

RSC Publishing has been recognised by a number of prestigious awards, including the 2011 ALPSP Best New Journal Award for the high impact journal *Chemical Science*, and several innovation Awards for its free online chemical database ChemSpider.

### Professional

The publishing operation is based in Cambridge, UK, and employs around 275 people on the Science Park. These professional publishing staff engage in the preparation, peer review, selection, editing, production, marketing and distribution of information in the chemical sciences. Additional international publishing staff are based in Philadelphia and Raleigh, USA; Beijing and Shanghai, China and Tokyo, Japan.

### Investing for the Research Good

As a Not For Profit organization, the RSC sustains its proven and established publishing activities primarily through subscription revenue. This model also enables the RSC to invest in new highly valued services for the community, generally at no additional cost to the user.

By way of example, during 2009 RSC Publishing acquired ChemSpider, a structure centric database for chemists. ChemSpider provides searchable access to over 26 million chemical structures and is considered to be one of the richest single sources of structure-based chemistry information worldwide. RSC Publishing provides free access to this service, as part of its publishing operations. Ref: [www.chemspider.com](http://www.chemspider.com)

*We welcome the opportunity to respond to the Office of Science and Technology Policy (OSTP) Request for Information (RFI): Public Access to Digital Data Resulting From Federally Funded Scientific Research.*

*Our comments are presented below, in response to the questions posed in the RFI.*

DATA - Preservation, Discoverability, and Access

**(1) What specific Federal policies would encourage public access to and the preservation of broadly valuable digital data resulting from federally funded scientific research, to grow the U.S. economy and improve the productivity of the American scientific enterprise?**

- We broadly support Government and funding agencies encouraging and endorsing data preservation as part of grant applications and awards
- Consideration should be given to acknowledging the differences between data types or needs, by: subject, level of curation, collection types, etc.
- Additional benefit could be gained from accreditation of a variety of approved repositories
- It is likely to prove important for policies to encourage and incentivise data deposition in approved repositories
- Thoughts should also be given to who may wish to access this data (publicly) and how to make the data discoverable
- Measures should be put in place to ensure ongoing data integrity
- Such activities are likely to support the US and international economies, and encourage international collaboration.

**(2) What specific steps can be taken to protect the intellectual property interests of publishers, scientists, Federal agencies, and other stakeholders, with respect to any existing or proposed policies for encouraging public access to and preservation of digital data resulting from federally funded scientific research?**

- Policy and statements should be clearly defined to differentiate what data may be openly (publicly) shared, and what may not.
- Through the creation and adoption of individual repositories, appropriate recognition should be given to the authors, and curators. Suitable referencing (citation) between the data and associated works (e.g. journal articles) will also be both beneficial and necessary.
- The long-term sustainability of the system must be preserved, changes which impact on intellectual property rights must consider the impact on existing policies, processes and systems
- Policies should acknowledge the costs involved in hosting, maintaining, preserving and making available data sets. The hosts should be entitled to recover these costs to ensure sustainability of the systems

**(3) How could Federal agencies take into account inherent differences between scientific disciplines and different types of digital data when developing policies on the management of data?**

- Consultation with the appropriate communities will be able to provide this. Considerable research and understanding is already available on differences between scientific disciplines when it comes to data expectations, needs or opportunities.
- Federal agencies may play a lead role in facilitating such discussions, involving scientists, publishers, and information professionals
- Consideration should be given to supporting the utilisation of existing subject specific data repositories, such as ChemSpider ([www.chemspider.com](http://www.chemspider.com)) for chemical compounds, or CCDC ([www.ccdc.cam.ac.uk](http://www.ccdc.cam.ac.uk)) for crystallographic data.
- Where existing repositories are not available, financial support to create suitable data archives could be necessary

**(4) How could agency policies consider differences in the relative costs and benefits of long-term stewardship and dissemination of different types of data resulting from federally funded research?**

- If the principle is to provide public access to all publicly funded data, then all research should be treated the same.
- However certain areas have more obvious societal benefits in the short to mid term (such as health related areas). For this reason, the fundamental principle may not be truly appropriate, and emphasis should be given to areas where there are greater opportunity benefits to society.
- Attention could therefore be given to areas where there is genuine public interest in accessing research data.
- It may also be worthwhile consulting with other groups who have already undertaken research on data management, such as:
  - DataCite: <http://datacite.org>
  - CoData: [www.codata.org](http://www.codata.org)
  - Opportunities for Data Exchange (ODE): [www.ode-project.eu](http://www.ode-project.eu)

**(5) How can stakeholders (e.g., research communities, universities, research institutions, libraries, scientific publishers) best contribute to the implementation of data management plans?**

- All stakeholders should collaborate to define a sustainable approach for this complex and broad ranging activity
- All stakeholders would then need to engage and adopt the agreed practices
- Areas such as investments, incentives, liabilities, risks, penalties and sustainability would need to be assessed and fully addressed.
- A detailed communication plan, possibly including training is likely to be necessary to the research community, to ensure their support. To this end, the processes required from researchers to deposit their data should be as simple and administratively simple as possible.

**(6) How could funding mechanisms be improved to better address the real costs of preserving and making digital data accessible?**

- Working with established repositories and systems is likely to be the most cost effective solution, where such repositories exist (e.g. ChemSpider, CCDC, as referenced in Q3)
- Investing in new systems that replicate existing systems is likely to be less effective and more expensive
- Adequate funding to support data management systems is necessary to ensure a sustainable solution, especially if data preservation is encouraged or mandated.

**(7) What approaches could agencies take to measure, verify, and improve compliance with Federal data stewardship and access policies for scientific research? How can the burden of compliance and verification be minimized?**

- Defining a minimum/appropriate number of repositories by need and accrediting these internationally would facilitate more effective monitoring of depositions.
- Regular re-accreditation and inspection will ensure the highest standards are maintained

**(8) What additional steps could agencies take to stimulate innovative use of publicly accessible research data in new and existing markets and industries to create jobs and grow the economy?**

- Market forces are likely to lead to innovative use of publicly accessible data
- Repositories openly identifying what data is publicly available, would help facilitate appropriate use or reuse of such data. This would avoid confusion over restricted vs open material which may reside on the same platform.
- We encourage the support of projects, involving researchers, publishers and information professionals, to explore what experimental or innovative uses could be derived from publicly accessible research data.

**(9) What mechanisms could be developed to assure that those who produced the data are given appropriate attribution and credit when secondary results are reported?**

- Established citation and reference standards should be applied to data records.
- Involvement with DataCite, a not for profit organisation aiming to make data easier to access, would help in this regard.

*Standards for Interoperability, Re-Use and Re-Purposing*

**(10) What digital data standards would enable interoperability, reuse, and repurposing of digital scientific data? For example, MIAME (minimum information about a microarray experiment; see Brazma et al., 2001, Nature Genetics 29, 371) is an example of a community-driven data standards effort.**

- Many subject specific formats exist for data management. Working with the individual communities will ensure appropriate (internationally recognised) standards and formats are adopted.

- Once a community-driven standard has been adopted then journal publishers can work with the community to aid compliance to the standard (i.e. that adopting the standard is an expectation or requirement of publication)

**(11) What are other examples of standards development processes that were successful in producing effective standards and what characteristics of the process made these efforts successful?**

- The many chemistry based standards have come around from there being a very strong community desire for a standard to be agreed and adopted. By being community driven, and hence strongly supported by the community, consensus and support is much improved
- Other characteristics for success including fully involving the multiple stakeholders in formulating suitable standards, communicating the process and the outcome to the community (often via societies and/or scientific journals)
- The most important element is to ensure the scientific community, database providers/curators and journal publishers work in partnership to aid the definition of, and compliance to the standard

**(12) How could Federal agencies promote effective coordination on digital data standards with other nations and international communities?**

- There are a number of established organisations seeking to define digital data standards in an international context (including groups mentioned in response to Q4).

**(13) What policies, practices, and standards are needed to support linking between publications and associated data?**

- Through working with publishers and publishing associations, standards and practices can most effectively be defined. Much progress has already been made in this regard, as all parties are supportive of achieving this common goal.