

# The Royal Society of Chemistry

## Harboring the Genius of Innovation

*An Interview With Richard Kidd by Svetla Baykoucheva (Svetla Baykousheva)*



*Richard Kidd, manager, informatics,  
RSCPublishing, Royal Society of Chemistry*

The Royal Society of Chemistry (RSC) has always had some interesting, innovative ideas ([www.rsc.org](http://www.rsc.org)). I remember how we were fascinated with the periodic table that artist Murray Robertson had created, according to his vision for the chemical elements ([www.chemsoc.org/viselements](http://www.chemsoc.org/viselements)). In the past few years, the RSC Library & Information Centre has undergone a transition to a virtual library ([www.rsc.org/Library/LIC](http://www.rsc.org/Library/LIC) Member), which is now providing online services to the society's 44,000 members. The beautiful old library in London, Burlington House, will soon be converted to a center for chemistry.

And now comes the RSC Project Prospect ([www.rsc.org/Publishing/Journals/ProjectProspect](http://www.rsc.org/Publishing/Journals/ProjectProspect)), an innovative service that in its first incarnation allows readers of the RSC journals to just click on a name of a chemical or concept and link to an external source that provides further information. Richard Kidd, RSC's manager, informatics, and manager of the RSC Project Prospect, explains.



**Svetla Baykoucheva:** Could you briefly tell us what the RSC Project Prospect is about and where the idea for this enhancement has originated from? How was it planned and developed? What kind of technology lies behind the screen?

**Richard Kidd:** Our internal development project started about a year before launch and had a number of aims—to use open standards for subject and chemical terms that would allow better identification of relevant content by searchers, to develop the display and reuse of structured experimental data within publication workflow, and to apply this across our published content.

Along with other publishers, we have sponsored summer students at the Unilever Centre of Molecular Informatics at Cambridge University (supervised by Peter Murray-Rust) for a number of years. Out of these projects evolved a number of software tools, such as the Experimental Data Checker ([www.rsc.org/publishing/resource/authorguide/lines/authoringtools/experimentaldatachecker/index.asp](http://www.rsc.org/publishing/resource/authorguide/lines/authoringtools/experimentaldatachecker/index.asp)), as well as a greater understanding of the possibilities of using structured data within the publication process. The development of the IUPAC InChI ([www.iupac.org/inchi](http://www.iupac.org/inchi)) identifier as an open standard for representing a chemical substance also enabled us to deal with chemical compounds in a sensible manner. We realized what the oppor-

tunities were and started this project to implement them in a sustainable form into our production workflows.

The technology is mostly standard publishing technology, although we do use the OSCAR3 text mining package (part of the SciBorg collaboration between the Unilever Centre and the Computer Laboratory), to identify some of the chemical compounds and subject terms within the articles. The main novelty is that we have identified the compounds and subject terms *in situ* within the text, store the information in the XML, and reuse this information in different ways to help the reader.

Several things really helped us in the development of the project—we have had an XML workflow in place for a number of years, and our technical editing teams already had a good understanding of XML structure and context. They are all chemical science experts (as are our development team), and they have been invaluable in the development of the project—they ask difficult questions for the developers to fix, they make sure the markup is correct, and they also suggest new enhancements that we should be looking at. It has been an advantage that, although we publish a range of titles, the range of subject areas we have to deal with is relatively focused.

**SB:** Do you implement these enhancements to whole journals or to selected articles, and how do you decide

which articles or journals should go first? Are all RSC journals going to have these enhancements?

**RK:** We have applied the enhancements to selected articles from all journal titles—initially, the ontology terms we use are most useful for the articles with a biomedical slant, or for those that are based on organic compounds. The articles selected for enhancement are those that will benefit most from this process. The current selection represents about a quarter of our annual journal output (about 1,400 articles so far in 2007). As we expand the Project Prospect enhancements to cover other subject areas, we are looking to enhance all the RSC journal content within the next year. We will also be looking at our book and database content, along with the journal backfiles.

**SB:** The project has been described as a technological leap that will bring chemical publishing closer to the semantic web. What does this mean?

**RK:** Project Prospect is still the first real application of semantic enhancement to primary research literature. The problem that the semantic web will help solve is a pretty basic one—it will allow searchers to find the right content. In the case of the chemical sciences, though, this is quite hard to achieve, as there are a lot of compounds that are never even given names in our publications—they are represented by structure images and boldface numbers. Using open standards such as InChI and the Open Biomedical Ontologies (www.obofoundry.org) will remove the ambiguity of searching. I won't pretend this is well-integrated with the search engines, yet, but we now have the information in a structured form that can make this happen. We firmly believe that our future in publishing is embedded within the web of related information and that the developments within Project Prospect are going to enable this in some really exciting ways.


**SB:** How will this innovation transform the discovery process and the way scientists gather information?

**RK:** Scientists need to find the relevant information without having to read the whole article, and this is difficult to achieve in cases where the compounds are not even named. In this case, the expanding use of InChI will help to eliminate hidden information and reduce ambiguity. As the search engines index standard identifiers and links propagate from other internet sources, the ability to find very relevant and unexpected information should develop quickly. Users have already reported finding related articles via the compound and subject links that they otherwise would not have seen.

Another unique innovation in Project Prospect is the enhancement of our RSS feeds—all our journal feeds have been upgraded to include compound and subject information for the enhanced articles. So, along with the bibliographic information and a graphical abstract, the reader gets a list of the ontology terms and the structures of the primary compounds. The metadata in the RSS feeds also contain the InChI and the ontology identifiers, which allows anyone (even those who are not subscribers) to put our feeds straight into a database and get a very good idea not only of the compounds we are publishing but also about their biological activity. We consider this a very powerful development, and we will further tailor the RSS feeds in the near future.

**SB:** What benefits do you think the RSC might get, as a whole, from developing such innovative ideas?


**RK:** We are a learned society with a publishing division, and one objective in our charter is to promote the dissemination of chemical science. Our council and management boards have been very supportive of Project Prospect, as chemical science as a whole will benefit from such innovation. The inclusion of more experimental data within the publishing workflow will be very important to both the original researcher and the reader, as a lot of chemical data are currently lost from the publication record. If we can preserve these data in a standard structured form that can be downloaded and interpreted by the reader, this will exploit

 The Open Biomedical Ontologies

Ontologies Resources Participate About

The OBO Foundry is a collaborative experiment involving developers of science-based ontologies who have established a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain. Currently the OBO Foundry ontologies form a part of the wider Open Biomedical Ontologies family, as listed below. In the longer term it is intended that the OBO Foundry will form one collection of ontologies alongside other such collections within the NCBO Bioportal.

In addition to a listing of OBO ontologies, this site also provides a statement of the OBO Foundry principles, discussion forum, technical infrastructure, and other services to facilitate ontology development. We welcome feedback and encourage participation.

Click any column header to sort the table by that column. The  link to the request trackers for the listed ontologies.

Domain	Prefix	File	Format
Biological imaging methods	FBIm	image.obo	obo
Biological process	GO	gene_ontology.obo	obo
BRIN/US tissue / genome source	BT0		obo
C. elegans development	WBdt	ncrm_development.obo	obo
Cellular genes anatomy	WBdt	cell.obo	obo
C. elegans phenotype	WBPhenotype	ncrm_phenotype.obo	obo
Cell type	CL	cell.obo	obo
Cellular component	GO	gene_ontology.obo	obo
Genetic test trait	TD	plant_trait.obo	obo
Chemical entities of biological interest	CHEBI	chebi.obo	obo
Common Anatomy Reference Ontology	CARGO	car.obo	obo
Distycolium discoidium anatomy	DDANAT	distycolium_anatomy.obo	obo
Drosophila development	FBdv	fly_development.obo	obo
Drosophila gene anatomy	FBdt	fly_anatomy.obo	obo
Environement Ontology	ENVDO	envt.obo	obo
Event (InChI pathway ontology)	IEV		obo
Evidence codes	ECCO	evidence_codes.obo	obo
EVDC (Expressed Sequence Annotation for Humans)	EV		obo
Fly taxonomy	FBtax	fly_taxonomy.obo	obo
Foundational Model of Anatomy (ubase)	FMA	fma_obo.obo	obo
Fungal cross anatomy	FAO	fungal_anatomy.obo	obo

Using the Open Biomedical Ontologies will remove the ambiguity of searching.

RSC Publishing

Online Shop Contact us Advanced search

Other RSC site areas: [Home](#) [Site Search](#)

Home » Publishing » Journals » Project Prospect

**Project Prospect**  
See science come alive - structured science within RSC journal articles

Home Features Examples FAQ Meet the team Contact Feedback News archive Alerting Services RSS Feeds E-Alerts Service Tools Email this to a friend Email your librarian

**Features**  
How to access the enhanced features of our articles

**Examples**  
View sample enhanced articles

**What can Project Prospect do for you?**  
Download a demonstration of the powerful HTML enhancements available in journal articles from RSC Publishing, including new updates and RSS feed details

**Project Prospect News**  
Latest news items relating to Project Prospect

**RSC Prospect in the spotlight**  
27 November 2007  
From scientific information professionals to top publishing managers - the award-winning RSC Prospect continues to win plaudits from a variety of audiences.

**RSC Prospect count reaches 1000**  
17 October 2007  
Another milestone for the project delivering enhanced HTML journal articles

**And the winner is ...**  
14 September 2007  
RSC Project Prospect scoops prestigious award for publishing innovation

**RSC Publishing announces pioneering work on RSS feeds**  
31 May 2007  
Project Prospect's latest development provides enhanced RSS feeds.

**RSC quicklinks**  
I want information on:  
Select a subject

Information for:  
Select a role

I am interested in:  
Select a product

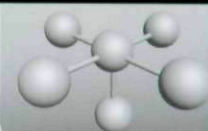
**Future Energy**  
Informed opinion on the issues that matter ...

**Shedding light on crystal engineering**  
CrystEngComm

**See Science Come Alive**  
ProjectProspect.org

Advertisements  
analytics

The homepage for the Royal Society of Chemistry's Project Prospect



*Project Prospect is housed in Cambridge's Science Park at the Thomas Graham House.*



*The beautiful old library in London will soon be converted to a center for chemistry.*

the advantages of electronic publishing to greatly benefit the chemical sciences and change the way chemistry is published on the web.

**SB:** Project Prospect was the winner of the 2007 ALPSP/Charlesworth Award for Publishing Innovation ([www.alpsp.org/ngen\\_public/default.asp?ID=251](http://www.alpsp.org/ngen_public/default.asp?ID=251)). What is this award given for, and why did the RSC get it?

**RK:** The award recognizes a significantly innovative approach to scholarly publication. The judging panel considers the originality and innovative qualities of each application, together with their utility, benefits, and long-term development prospects. In their report, the judges said that Project Prospect "was the clear winner ... with an elegant and intuitive on screen manifestation of the advantages of including metadata. As a result, sophisticated and effective searching of the literature is greatly improved and the value gained from reading each article is significantly enhanced ..." We were very proud and honored to win this award, and it has certainly helped to raise the profile of the project with other publishers.

**SB:** From your preliminary impressions, what are the things that users like most about this new enhancement? What is the value added for users?

**RK:** As a publisher, we want our high-quality research to be found and read, so we will be trying further implementations of the data to see how readers will use the different views. The current representation is just one application of the data we have marked up, and it has been very gratifying to get some great feedback on this, particularly highlighting the enhanced accessibility of the scientific content.

The inclusion of ontology definitions has been seen as a solid educational benefit—this isn't just for students and nonnative English speakers. This functionality contributes to the scientific quality of the work, because it makes it more likely for this work to be retrieved by scientists from other disciplines. It has also added new means to find related articles containing a particular subject term or compound. We have received very positive responses from users

about these enhancements, with one user even comparing them to going from TV to HDTV. Showing authors and readers the initial interface has also provided us with several suggestions for improvements and enhancements that we will be pursuing in the future.

**SB:** What is next for Project Prospect?

**RK:** We are working hard on the next set of enhancements to be launched early in 2008, and I hope we will have a couple of real surprises. The overall intention is to expand the subject coverage and enhance articles from the physical chemistry and materials subject area of our journals portfolio, which will allow us to do more with the available experimental data.

One of the great things about working on this project is that there are really exciting opportunities that open up in terms of adding structure to the published science. While our current enhancements have chimed well with the chemical science community, there is much more that is conceptually easy for us to develop. We are limiting the rate of development at this stage to implement the current enhancements efficiently and sustainably into a live production environment. We have been able to implement Project Prospect without damaging our already leading publication times, which by itself is already a huge achievement.

The interest in our innovations from other publishers and possible collaborations will be another step that would allow us to link between related materials, whether they are journal articles, books, or databases that are based on standard compound and subject terminology. When the reader starts to expect this, and even takes this for granted like they do now with full-text reference linking via CrossRef, we will have moved a step forward in changing the way science is published.

---

*Svetla Baykoucheva (sbaykouc@umd.edu) is head of the White Memorial Chemistry Library at the University of Maryland—College Park. She is also editor of the Chemical Information Bulletin, a publication of the chemical information division of the American Chemical Society.*

*Comments? Send email to the editor (marydee@xmission.com).*

Copyright of Online is the property of Information Today Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.