

Surfing Web₂O

The rapid evolution of the world wide web is creating fresh opportunities – and challenges – for chemistry. Richard Van Noorden reports

As chemical reactions go, it was a complete failure. 'Contents of the reaction flask decomposed. Aborted', Drexel University chemist Jean-Claude Bradley and students recorded the day after an attempt to synthesise a catechol aldehyde from adrenaline on 24 January 2006. But the experiment has acquired a peculiar honour: Bradlev chose it to be the first written in his group's new online laboratory notebook, in which all experimental data is made public and freely available to web users - a concept he later christened 'Open Notebook Science'.

Bradley's idea is simple: most failed experiments are discarded, yet their data could be useful to someone else. Even published papers don't always sufficiently explain the workings behind a successful experiment. In contrast, all Bradley's research and raw data is now documented transparently and almost in real-time. Anyone can see it, comment on it, and use it; and the internet is the perfect vehicle for hosting it.

Open Notebook Science is just one of many new routes for chemical information to appear on the internet. From searchable molecular databases to the user-editable Wikipedia; from video recordings of experimental protocols to the informal news, gossip and argument posted on chemistry blogs; a huge amount of chemistry can now be retrieved at no cost.

The emphasis on user-generated content, shared amongst online social networks, is typical of Web 2.0, the umbrella term for the current evolutionary stage of the world wide web. In this 'social web', swamps of data could be powerfully linked together. Search engines can trawl it to pick out whatever another user asks for. And user 'tagging', together with underlying machine-readable descriptions, means that related information can be easily linked. For example, clicking on a molecule could eventually bring up not just a 3D picture and a list of properties, but also the related online articles, experiments, videos and blog posts that refer to it.

Many web users are already familiar with this potential. They post videos to YouTube; they read blogs; they use social networking sites like MySpace and Facebook. But chemists are only just catching on to the possibilities. Some are sure that linking free chemical data on the web will revolutionise the culture and practice of chemistry, aiding collaboration and speedy access to information on an unimaginable scale.

'Mainstream chemistry has no tradition of openness and electronic collaboration. This is a bottom-up movement, largely composed of young researchers,' explains Peter Murray-Rust, a chemical informatics academic at the University of Cambridge, UK, and a keen advocate of what he terms 'Open Chemistry'.

But as Murray-Rust also admits: 'chemistry is the best subject to do this with, but the hardest to sell it to'. The open chemistry model has to prove its worth alongside trusted, high-quality subscription databases and journals. It has to show its volunteered data can be useful and high-quality. And it has to capture the enthusiasm and support of chemists who are, Murray-Rust says, generally apathetic about the possibilities Open Chemistry offers.

Bloggers ahead

Murray-Rust takes heart from the irreverent spirit of the fast-growing chemical blogosphere: the hundreds of online diaries where chemists grumble, gossip, joke, argue, inform and inspire. Writers across the chemical sciences post anything from personal experiences as a post-doc to commentaries on the latest published articles, or reports on drug discovery and software technology. They dissect gossip that would a few years ago have been confined to laboratory corridors or the departmental tea-room. Readers relish the blogosphere's witty opinions and (occasional) reasoned analysis, made all the more frank by the internet's easy anonymity

The blogosphere has a loyal following, but few chemistry professors write blogs; most authors are graduate students or postdocs. As Open Chemistry supporter Steve Bachrach explained about a year ago when interviewed for web-based chemistry magazine, *Reactive Reports*: 'I don't have the time to read random thoughts by random individuals. I barely have time to keep up with the traditional literature in my field. The blogosphere just seemed to me to be filled with the rantings of people who have nothing better to do with their time.' Though there is some information to be found, Bachrach now concedes, he still contests that most chemistry blogs have little content in them useful to the busy researcher.

While many post for fun and

In short

 The internet is becoming flooded with free chemical information: from blogs to videos and databases
Linking this data together and interacting via the 'social web' could revolutionise the practice and teaching of chemistry
So-called 'Open

Chemistry' faces many challenges: not least maintaining data quality and co-existing with trusted subscription databases interest, riffing around the culture of lab-based chemistry, blogs such as Paul Docherty's *Totally Synthetic* provide useful summaries of the latest organic syntheses, effectively acting as global online journal clubs where researchers all over the world chip in with constructive criticism. As blogger Andrew Sun argues, the blogosphere's content is a product of its authors, and it would surely change if more chemistry professors bothered to blog – as happens, to an extent, in other sciences.

Indeed, a few chemists are using blogs to discuss their own research, as well as comment on others'. Bradley's Open Notebook works on a 'wiki' (a site that any user can edit quickly - the name derives from a Hawaiian word for 'fast'), but he discusses higher-level thinking about the project on a related blog. A few Open Notebook converts, including Cameron Neylon of Southampton University, UK, are attempting Bradley-like online records, and discovering the difficulties of keeping a faithful up-to-date log. Many researchers working in cheminformatics regularly discuss their own research directly on blogs - prominent among them Murray-Rust, and Egon Willighagen at Wageningen University in the Netherlands. In his spare time, Willighagen runs Chemical blogspace, a site automatically picking useful information out of the blogs which regularly discuss research in the core chemical sciences.

Some websites collect chemistry information submitted voluntarily by other researchers. The most famous of these, of course, is Wikipedia - the user-editable encyclopaedia. 'Chemistry is an ideal subject for recording factual information and Wikipedia will soon become acknowledged as the primary chemical reference for undergraduate study,' insists Murray-Rust. Less well-known are such websites as Organic Syntheses, a database of protocols for organic chemists, and Synthetic Pages, a similar, smaller, database which records the personal experiences of chemists attempting particular reactions.

ChemTube

But as scientists are discovering, the internet allows communication beyond the limits of a traditional printed journal article. Thanks to high-capacity broadband internet, it is easy to watch video and audio clips

avid supporter of Open Chemistry

Peter Murray-Rust:



Blogroll

There are hundreds of blogs related to chemistry on the internet. *Chemistry World*'s own blog (www. chemistryworld.org/blog), which brings you news, opinion and discussion on the chemical sciences, lists some of our favourites, including:

Totally Synthetic

totallysynthetic.com/blog

The journal club taken global. Paul Docherty guides organic chemists through the latest syntheses in a beautifully presented, high quality blog which attracts plenty of informed comment.

Chembark

blog.chembark.com

With the atmosphere of a lab group's night out in the pub, Paul Bracher coordinates irreverent chemistry gossip.

of experiments or lectures, often on video-sharing sites like the Googleowned YouTube, where anyone is encouraged to post content.

Chemistry videos – with their crowd-drawing explosions and bright colours – attract a popular following on YouTube. Hundreds of chemistry educational initiatives use YouTube to upload videos; so many that each risks being swamped by competing content. The observer interested in the explosive consequences of shaking Mentos mints in Coca Cola can take their pick of over 10 000 YouTube clips, for example, although a ratings system helps to sort out the most eye-catching examples.

In July this year chemists working on a European nanoscience research project, called Nano2Hybrids, began recording their progress with a series of weekly video diaries, intended both to inform researchers and explain to the general public what being a scientist is like. The project works in partnership with the Vega Science Trust, established by buckyball-discoverer and former RSC president Harry Kroto in 1994. It freely broadcasts a variety of science programmes over the internet, including interviews and recorded lectures from Nobel prize winners.

Indeed, many universities now post audio and video of research discoveries and chemistry lectures online, in some cases free to access. 'Whether chemistry professors like it or not, students will be using these tools,' says Bradley.

Innovative web content aimed solely at professional scientists has arrived a little more slowly. But researchers are beginning to post

In The Pipeline

corante.com/pipeline

Insightful and stylish analysis from Derek Lowe – *Chemistry World* columnist and medicinal chemist – keeping readers up to speed on the development and culture of drug discovery.

post doc ergo propter doc

propterdoc.blogspot.com Charts the ups and downs of life as an ex-pat postdoc in Canada. Chemistry's very own Bridget Jones?

The Sceptical Chymist

blogs.nature.com/thescepticalchymist

Editors working at *Nature* and its research journals post interviews and their analysis of developments in chemistry and chemical biology.

videos of their own experiments on sites such as the Journal of Visualised Experiments and SciVee: both video-sharing sites that directly target scientists, not the general public. SciVee promotes the idea of launching a video or audio presentation along with a published paper, which it calls a 'pubcast' much like giving a conference talk which is accessible to all internetusers. Meanwhile, this year has seen a proliferation of podcasts from chemistry journals and science magazines, including Chemistry World, which highlight the latest discoveries in the chemical sciences.

Chemists are also actively researching together in online forums, though numbers are small. Medicinal chemistry is a noted frontrunner: The Synaptic Leap is just one website that promotes collaborative biomedical research, focusing on building communities for diseases such as malaria and tuberculosis. Bradley's Open Notebook Science research also looks into the synthesis of antimalarial agents - as he explains, collaborative involvement in this kind of research can only work where intellectual property squabbling is less important, because little profit can be made from the results. One of the most

speculative projects includes blogger Mitch Volunteered free chemistry content is flooding the social web

Nano²Hybrid

Molecules

First

ChemBark



Many chemoinformatics researchers and web-technology enthusiasts are excited by the prospects of chemistry communication in internet-based virtual worlds, such as Second Life. This online world already hosts a few islands of chemistry activity, where chemists can gather to discuss science, aided by virtual rotating molecules, conference papers and videos.

Data drive

But the greatest source of established free information for research chemists on the internet are the 60+ small open-access journals (see p12), some delayed open-access archived journals and, especially, the free online chemistry databases that aggregate together information on millions of molecules.

PubChem, 'the granddaddy of all free chemistry databases', as former medicinal chemist Rich Apodaca

puts it, allows users to search almost 11 million compounds. It is maintained by the National Center for Biotechnology Information (NCBI), part of the United States National Institutes of Health, and takes data from over 80 sources, including subscription-access journals such as *Nature Chemical Biology*. Other free online databases provide particular

useful molecular details: NMRShiftDB contains over 20000 NMR spectra for

organic compounds, while ChemExper and eMolecules link a molecule searcher to commercial suppliers. SureChem picks out more than 7 million chemical structures held under US and European patents. And medicinal chemists are particularly well supplied with numerous free databases for drug discovery.

Project Prospect

One effective example of how the web can enhance content has been provided this year by the RSC's Project Prospect, whereby electronic journal articles are enriched with extra computerreadable metadata. It means readers can click on named compounds, scientific concepts and experimental data in an article to download structures, understand topics, or link through to electronic databases like lupac's Gold Book.

The applause for this project – which recently marked up its 1000th paper and won the 2007 ALPSP/Charlesworth Award for Publishing Innovation – has demonstrated the potential of machine-readable articles. For the moment, authors are not asked to provide anything special when submitting the material; the work falls to the technical editors. As yet, Prospect lacks the interactivity that would allow users to add extra data to a molecule's pop-up information box, for



example.

Still, the project may help to bolster the tiny numbers of those chemists aware of how the web can enhance chemistry. 'The average head of the average chemistry department probably thinks we are just playing games,' concedes Murray-Rust.

All have to compete against the authority – and guaranteed quality – of subscription journals and databases. Elsevier's Beilstein database, and CAS (the chemical abstracts service, a division of the American Chemical Society (ACS)) along with its delivery services SciFinder and STN, are still the undisputed gold standards.

Revolution in bits

Open chemistry advocates are frustrated by the way chemical data is fragmented between different closed databases. They reluctantly concede that gaining 'Open Access' to chemistry journals is a tough cause to fight (see p12). But 'Open Data' is quite a different proposition – publishers could well restrict access to journal papers while still freeing online records of their molecules and spectra, for example.

The possible benefits of this approach to the chemical community are already apparent, via an online service called ChemSpider. which launched in March 2007. It promises chemists free access to almost 18 million compounds, sourced from free chemistry databases. Plans are to turn the service into a search engine - a chemical version of Google - which automatically 'spiders' across web chemistry publications and databases looking for structures, much as Google trawls through the text of the internet.

CAS doesn't permit web search engines to scour its database, even though searchers would have to pay for any CAS information they were pointed towards. Indeed, ACS doesn't allow Google to index its Still, Murray-Rust's own CrystalEye project is aggregating x-ray crystal structures, from the CIFs (crystallographic information files) that publishers demand as supplementary material for online articles. These don't fall under copyright laws, so it is possible to build up a free online database of crystal structures, even though they

journal articles. so for the moment

the search engine's power is limited.

belong to closed-access papers. Turning vast quantities of online data into a useful resource poses two significant problems. First, maintaining quality is crucial. No chemist wants to be faced with hundreds of incorrect chemical structures, or directed to blog posts of drivel, but that is the inevitable result of allowing search engines to pick through unchecked data. This is exactly what publishers and database owners guarantee to avoid by paying staff to oversee their publications, although some Open Chemistry advocates point to the success of Wikipedia as proof that community editing can establish acceptable levels of quality control.

The second problem is making the data searchable. Google searches only by plain text, which is not always ideal for the chemical community. They need a chemical version of the XML metadata that, unseen, holds the regular internet together. Many computer chemists are developing machinereadable languages that represent molecular structures – systems include SMILES, Chemical Mark-up Language (the chemical version of HTML) and the Iupac International Chemical Identifier (Inchi). And 'Mainstream chemistry has no tradition of openness and electronic collaboration. We may simply have to wait for the dinosaurs to die' of course, customised RSS (Really Simple Syndication) feeds can automatically trawl users' favourite information resources to gather the latest updates into one place.

And yet, Bachrach insists, the biggest problem is cultural – persuading chemists that they would benefit from access to other people's data is not easy, particularly as many chemists already have access to paid-for databases. 'Chemistry is a conservative subject,' fumes Murray-Rust. 'The chemical information market is now holding back opportunities.'

Bachrach agrees, pointing out that because some established journals refuse pre-published submissions, he would never publish original research on a blog or wiki. And chemists need their work to appear in those journals, because they determine career progress as viewed by university faculty and funding bodies.

But while the kinetics may be slow for open chemistry supporters, the thermodynamics are on their side. The next generation of professional chemists are far more likely to be in tune with web-based chemistry, treating blogs and social networking sites as professional tools in the same manner as email. For Open Chemistry advocates, the inevitable passage of time may be enough to usher in their revolution. Or, as Bachrach puts it: 'We may simply have to wait for the dinosaurs to die.'

Web references and further reading Open Notebook Science: usefulchem.blogspot. com; usefulchem.wikispaces.com Peter Murray Rust (blog): wwmm.ch.cam.ac.uk/ blogs/murrayrust Reactive Reports: reactivereports.com Chemical blogspace (summary of chemistry blogs): ch openmolecules net Wikipedia Chemistry portal: en.wikipedia. org/wiki/Portal:Chemistry Organic Syntheses: orgsyn.org Synthetic Pages: syntheticpages.org Nano2hybrids: nano2hybrids.net Vega Science Trust: vega.org.uk SciVee: scivee.tv Chemistry World podcast: chemistryworld. org/podcast Journal of Visualised Experiments: jove.com The Synaptic Leap: thesynapticleap.org Chemmunity: chemmunity.com Second Life: secondlife.com PubChem: pubchem.ncbi.nlm.nih.gov ChemSpider: chemspider.com CrystalEye: wwmm.ch.cam.ac.uk/crystaleye Project Prospect: rsc.org/Publishing/Journals/ **ProjectProspect** P Murray-Rust, Nature Precedings, 2007, in press (open access manuscript): precedings. nature.com/documents/1200/version/1 P Ertl and S Jelfs, Curr. Top. Med. Chem., 2007, 7.1491

Thirty-two free chemistry databases (Rich Apodaca blog): depth-first.com/articles/2007/01/24/ thirty-two-free-chemistry-databases