Enhancing information

Experimental data is the foundation on which science is built. Providing easier ways to find and search it is one way in which new online technologies can help to advance research.

Some love it and some hate it, but online supplementary information is published with most chemistry papers. As recently as the mid 1990s, readers with an interest in this information would have to contact the authors directly or the journal's editorial office to gain access. Online publication has changed that and opinions differ on whether it's a good thing[1].

Many scientific papers do seem to have more and more associated supplementary information, a situation that recently resulted in an announcement[2] from the *Journal of Neuroscience* that it would no longer allow authors to include it with their articles. The authors do have the option, however, of including links to such information hosted on their own websites. The rationale put forth for this change is that the provision of such large amounts of supplementary information is beginning to undermine the peer-review process — mainly because referees are too busy to read and properly review ever-larger files. This, in itself, creates an opportunity for the unscrupulous researcher to hide inconvenient data in the supplementary information and thus have it published in the peer-reviewed literature without the necessary rigorous review.

At *Nature Chemistry*, we consider the inclusion of experimental data to be a vital part of what we publish. Irrespective of whether the primary theme (or novelty) of an article relates to synthetic processes, details of how to repeat the preparative aspects of a study are a vital part of the scientific method. If other researchers are to reproduce results and then build on them, they need to follow in the footsteps of those who first made a new compound or material — and this requires that clear directions are available to signpost the way. This includes not only experimental methods, but also characterization data so that those who do repeat the work know that they are heading in the right direction at each turn. Other technical details that may not be crucial to the narrative of the main text of an article — such as lengthy mathematical derivations, calibration data or basic control experiments — are also ideal candidates for inclusion in the supplementary information. Moreover, we prefer — when possible — that this information be permanently available alongside the original article, rather than subject to the whim of sometimes poorly maintained external websites[3].

That said, the supplementary information — particularly for an article that describes a new synthetic methodology or the synthesis of a complex molecule — can easily run to several hundred pages and quickly become difficult to navigate. Rather than limit the amount of information that can be included, however, we believe that a better approach is to explore new and improved ways in which access to the content in question can be achieved.

Since our very first issue, the HTML versions of *Nature Chemistry* Articles have included links from bold compound numbers to 'compound information pages' that give further detail — compound names, 2D and 3D structures, chemical identifiers and links to PubChem[4]. Our next step along the road to enhanced content in our Articles is to add to these compound pages the synthetic procedures and characterization data or, where appropriate, references to literature procedures. Once again, bold compound numbers within these procedures provide links that enable the reader to follow a reaction sequence should they require details of how to prepare a necessary starting material.

Interested readers can see one example of these new enhancements in the Article by Denmark and Wilson in this issue[5]. Clearly, our ability to provide this sort of information relies on the researchers who publish in *Nature Chemistry* submitting some additional files with the final accepted versions of their manuscripts. That said, just a small amount of extra work in terms of file-preparation and formatting can result in much more informative compound pages. In addition, an unexpected benefit of our process of including synthetic procedures is that it has, in several cases, enabled us to correct small (but important) typographical errors in supplementary information — errors which, because we do not copy-edit supplementary information, would have previously gone unnoticed.

Inclusion of the synthetic procedures in the HTML version of articles not only makes this information more accessible to readers, but it also becomes semantically useful. It is now possible, in principle, not only to identify chemical compounds by text mining — as we do in the full text of our articles — but also to assign a role to them in reactions, which could have a number of benefits. For example, Peter Murray Rust's 'Green Chain Reaction' experiment[6] uses software called ChemicalTagger to identify the solvents used across the chemical literature over time to see whether chemical procedures are becoming 'greener'. In addition, ChemSpider with their 'Synthetic Pages' have shown the utility in linking text-mined procedures with database pages crammed full of useful links and crowd-sourced data[7].

Nature Chemistry compound pages can also display links to associated data files hosted on our website. For instance, we currently mandate the inclusion of crystallographic information files (.cif) for X-ray crystal structures associated with an Article. With the help of those who submit to the journal, we are eager to expand on the types of data files that we host, including those used for spectroscopic data, such as .jdx files[8]. We realize that although large multi-page pdf files of additional data serve well for simple visual inspection, there is huge untapped potential trapped within these pages that could spring to life in an online format.

These changes won't happen overnight, but we are keen to implement those that we feel would be useful for our readers — and adding synthetic procedures to our compound pages is another step along this path. We welcome your suggestions for other potential enhancements of our online content — in particular our compound pages and related supplementary information — and this Editorial will be reproduced on our blog[9] so that you can comment. ❐

References

1. http://go.nature.com/kJm1YG
2. Maunsell, J. *J. Neurosci.* **30,** 10599–10600 (2010).
3. *Nature Chem.* **2,** 697 (2010).
4. http://pubchem.ncbi.nlm.nih.gov/
5. http://www.nature.com/nchem/journal/v2/n11/compound/nchem.857_ci.html
6. http://scienceonlinelondon.wikidot.com/topics:green-chain-reaction
7. http://cssp.chemspider.com/
8. http://www.jcamp-dx.org/
9. http://blogs.nature.com/thescepticalchymist/

© 2010 Macmillan Publishers Limited. All rights reserved