The Royal Society of Chemistry and the data publication landscape

Serin Dabb
Executive Editor, Data
Royal Society of Chemistry

The world’s leading chemistry community

- Over 54,000 members
- Leading international not-for-profit publisher
Leading international publisher

- Our publishing activities span books, e-books, journals, databases and magazines

Our activities

Royal Society of Chemistry

- International not-for-profit Publisher
- Conferences & Events
- Professional Body Qualifications
- Library and Information Centre
- Learned Society Charity
- Education Facilitator
- Science Policy - campaigning organisation
- Global Membership Organisation
Who am I?

Databases

- The Merck Index Online
- MarinLit

Literature Updating Services

- Analytical Abstracts
- Natural Product Updates

Synthetic Reaction Updates
Advancing excellence in the chemical sciences

As a global learned society:
• How can we facilitate data best-practice (sharing, discoverability, transparency, standardisation, …)?

As a publisher:
• How does data publication align with the traditional journal article publication process?

As a membership body:
• How can we support our members and their data needs?
Other stakeholders

- Funders
- Institutions
- Other societies
- Government agencies
- Community driven initiatives
Funder

Data enthusiast

Researcher
RCUK Common Principles on Data:

• Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner.
• Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice.
• Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.
• It is appropriate to use public funds to support the management and sharing of publicly-funded research data.

Role of institutional repositories?
“UKCoRR has a vision of the work of repository management as a professionally recognised and supported role within UK research institutions”
## UK landscape

<table>
<thead>
<tr>
<th>Funder</th>
<th>Requirement for data sharing</th>
<th>Support from the funder</th>
<th>Support required from Cambridge</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPSRC</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>BBSRC</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>ESRC (Economic &amp; Social Research Council)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>NERC (Natural Environment Research Council)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>EPSRC (Engineering and Physical Sciences Research Council)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>MRC (Medical Research Council)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Science &amp; Technology Facilities Council</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Wellcome Trust</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>EU (European Union)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>ERC (European Research Council)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Cancer Research UK</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Requirement for data sharing</th>
<th>Financial</th>
<th>Dedicated advisory team</th>
<th>Repository</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPSRC</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>BBSRC</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>ESRC (Economic &amp; Social Research Council)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>NERC (Natural Environment Research Council)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>EPSRC (Engineering and Physical Sciences Research Council)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>MRC (Medical Research Council)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Science &amp; Technology Facilities Council</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Wellcome Trust</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>EU (European Union)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>ERC (European Research Council)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Cancer Research UK</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

Support required from Cambridge:
- Repository: ✔
- Advice on data plans & policies: ✔
- Advice on data management: ✔

---

**Attribution:** Dr Danny Kingsley, Head of Scholarly Communication, University of Cambridge

Blog: [https://unlockingresearch.blog.lib.cam.ac.uk/](https://unlockingresearch.blog.lib.cam.ac.uk/); Website: [http://osc.cam.ac.uk](http://osc.cam.ac.uk)
What can we do?
ChemSpider
Search and share chemistry

- Free database
- Data downloadable
- Allows users to deposit and share data

44 Million chemical structures
509 Data sources
National Compound Collection

• Questions we wanted to answer:
  – *Is there ‘lost chemistry’ in PhD theses that can be retrieved through a data mining process?*
  – *Is that chemistry significant in volume, novel and potentially industrially useful?*

• Some(!) of the partners:
  – University of Bristol, University of Huddersfield, British Library

The creation and characterisation of a National Compound Collection: the Royal Society of Chemistry pilot
*Chem. Sci.*, 2016, Accepted Manuscript, DOI: 10.1039/C6SC00264A
Validation and standardisation tools

Open PHACTS

Platform to ensure normalisation of chemical structures to rules derived from the FDA structure standardisation guidelines and modified based on input from consortia members.

Chemical Validation and Standardisation Platform
http://cvsp.chemspider.com/

• Submit structure files, check for errors and inconsistencies
• Suite of databases and tools provided for UK academic institutions

Vision:
• To build a data repository for the UK academia, to deposit and share chemistry research
• Milestones: compounds → reactions → spectra
Data in the publication process
The challenge

Electronic Supplementary Information

TEMPO-Mediated Allylic C-H Amination with Hydrazones

Xu Zhu and Shunsuke Chiba

Division of Chemistry and Biological Chemistry, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore.

E-mail: shunsuke@ntu.edu.sg

(Z)-1-[(1-Benzylcyclohexyl)phenoxy]methyl]-2-phenylhydrazine (3a):
Journal advises on data standards policy

Most authors submit appropriate data & files to journal as doc/pdf

Data published within article & supplementary information

Reader views, but can’t interact, with data in published article & SI
Our current data policies

- With an article submission, we request all data necessary to understand and verify research to be provided
  - Crystallography, macromolecular and sequence data in appropriate repository
  - All other data in the article or supplementary information

- Other, journal specific, guidelines

- Authors are encouraged to deposit all associated data in a relevant repository

- Aim: develop these data requirements appropriately for each field
Synthesis and protective effect of new ligustrazine derivatives against CoCl₂-induced neurotoxicity in differentiated PC12 cells. Part 2†

Bing Xu†*, Yan Gou†*, Xin Xu†, Chenze Zhang†, Yuzhong Zhang†, Fuhao Chu†, Hongbing Liu‡, Penglong Wang** and Haimin Lei**

†School of Chinese Pharmacy, Beijing University of Chinese Medicine, Beijing 100102, China. E-mail: hm_lei@126.com; wpm581@126.com
‡Department of Pathology, Beijing University of Chinese Medicine, Beijing 100102, China.

*College of TCM, Xinyang Medical University, Xinyang 464001, China.

Received 9th December 2014, Accepted 1st February 2015.

First published on the web 12th February 2015

Ligustrazine, ligustrazine, and related promising neuroprotective activities have previously been reported. To further improve the neuroprotective effect of ligustrazine, a new ligustrazine derivative was synthesized and evaluated for their protective effect against cobalt chloride-induced neurotoxicity in differentiated PC12 cells. Our results showed that compound (E)-(2,5,6-trimethylpyrazin-2-yl)methyl 13-(2,3,4-trimethoxyphenyl)acrylate (9) had significant neuroprotective activity and did not induce cell death in ligustrazine, of which, we found compound (E)-(2,5,6-trimethylpyrazin-2-yl)methyl 13-(2,3,4-trimethoxyphenyl)acrylate (9) could protect differentiated PC12 cells damaged by CoCl₂ (EC₅₀ = 0.719 μM). Structure–activity relationships are briefly discussed.
Next stages....?
Thank you

Serin Dabb
Executive Editor, Data

dabbs@rsc.org

@SerinDabb