Data sharing at the RSC

- Open Science and positive science culture
- Open Data Standards
- Data Sharing
- Data Availability Statements
- Digital Discovery examples
Open science and a positive science culture

One of the three foundations of a great science culture is Scientific Practice.

Good scientific practice is achieved when science and innovation are conducted in a rigorous, ethical and responsible way. Effective science is also open and collaborative, involving the sharing of knowledge and expertise across disciplines, sectors and countries.

Scientific data is deposited in an accessible repository whenever it can be useful to others. Scientists ensure that practices such as peer review are fair and transparent, that tools and data are presented in a useful way, and that data is FAIR (Findable, Accessible, Interoperable, Reusable).
What is Open Science?

“scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society” – United Nations

“The principle and practice of making research products and processes available to all, while respecting diverse cultures, maintaining security and privacy, and fostering collaborations, reproducibility, and equity.” - U.S. Office of Science and Technology Policy and the National Science and Technology Council

“as open as possible, as closed as necessary.”

Figure from: UNESCO Recommendation on Open Science © 2021 by UNESCO is licensed under CC BY-SA 4.0
Our Commitment to OS

Our purpose is to help the chemical science community make the world a better place; we envision a world in which the chemical sciences fulfil their potential as a force for good. We strive to work with our community to break down barriers to advancing the chemical sciences and to ensure appropriate equitable, global access to knowledge and data.

Read the full statement here:
https://www.rsc.org/journals-books-databases/open-science/
Supporting open data standards

We are actively supporting the development of open data standards - we developed open chemistry ontologies, and are a founding member of InChI Trust, a structure-based chemical identifier.
Data sharing in RSC journals

We provide recommendation and guidance on data sharing on our website:

**Our data sharing policy**

*The Royal Society of Chemistry believes that where possible, all data associated with the research in a manuscript should be Findable, Accessible, Interoperable and Reusable (FAIR), enabling other researchers to replicate and build on that research.*

*We strongly encourage authors to deposit the data underpinning their research in appropriate repositories.*

*For all submissions to Royal Society of Chemistry journals, any data required to understand and verify the research in an article must be made available on submission.*

*To comply, we suggest authors deposit their data in an appropriate repository. Where this isn’t possible, we ask authors to include the data as part of the article Electronic Supplementary Information (ESI).*

*Some journals may have additional subject requirements for both sharing and/or publishing supporting data, so please ensure you check the journal-specific guidelines.*

**Follow FAIR guidance**

**Deposit data where possible**

**Data supporting the reported findings must be shared at submission**

**Choose repositories over SI**

**Look out for journal-specific guidance**
Choosing a repository

Selection criteria (basic)
- Register datasets with a persistent identifier (e.g. DOI)
- Provide support to include as much documentation as possible
- Facilitate linking the dataset to related items, such as a publication
- Support attribution and credit, e.g. by linking to ORCID, ROR, RAiD, etc.
- Have licensing options that are as open as possible

Tools and support
- Explore community repository registries such as re3data.org and FAIRsharing.org
- Check publisher recommendations and requirements
- Consult institutional guidelines, local librarians or research data support services
- Check funder policies and guidance

ROR: Research Organisation Registry; RAiD: Research Activity Identifier
Repository guidance at the RSC

Subject specific repository guide

<table>
<thead>
<tr>
<th>Data type</th>
<th>Repository</th>
<th>URL</th>
<th>File / standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystal structure</td>
<td>Cambridge Structural Database (CSD) - managed by the Cambridge Crystallographic Data Centre (CCDC)</td>
<td>Visit the CCDC website</td>
<td>Crystallographic information file ( cif)</td>
</tr>
<tr>
<td>Structural Materials</td>
<td>Required for all RSC journals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural Omics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural Sequence data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural Computational simulations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural Spectroscopy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural Images</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Choosing a repository

Where possible, choose discipline-specific, community-recognised repository

Alternatively, explore institutional or generalist repositories

Generalist repository guide

<table>
<thead>
<tr>
<th>Repository Name</th>
<th>Information on costs</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dryad Digital Repository</td>
<td>Fees apply</td>
<td>Visit the Dryad Digital Repository website</td>
</tr>
</tbody>
</table>
Data Availability Statements at the RSC

Why DASs?

- To summarise which data is available to support your findings, where it is stored and how it can be accessed
- To demonstrate compliance with relevant funder, institution or journal data policies
- To support reproducible research and ensure availability of data for reuse

“To maintain high standards of transparency, research reproducibility, and to promote the reuse of new findings, we **strongly encourage** authors to include a Data Availability Statement (DAS) as part of the final published article.”
Publishes research on high-throughput methodologies for accelerated discovery of chemical and biological entities.

- AI and machine learning
- Quantum computing
- Automation

Find out more: rsc.li/DigitalDiscovery
Data availability and reproducibility

Accessibility of data and code is absolutely essential for the reproducibility of computational research.

• authors strongly encouraged to deposit as much data and code
• data and code must be made available for Editors and Reviewers at the time of peer review
Data reviewer

A third researcher in addition to two regular reviewers who checks the data and code and their usability.

**Data reviewer checks:**

- accessibility of data
- data cleaning
- data representation
- model training and validation
- reproducibility of the code
Examples – data availability

Faster and more diverse de novo molecular optimization with double loop reinforcement learning using augmented SMILES

The authors could not provide their data and code for the Editor and reviewers.
Examples – data availability

Accelerated materials language processing enabled by GPT

Work builds on the proprietary models from OpenAI
Examples – data availability

Data available

All the training data: conditionally gene

https://github.com

Conditionally and

1ub: