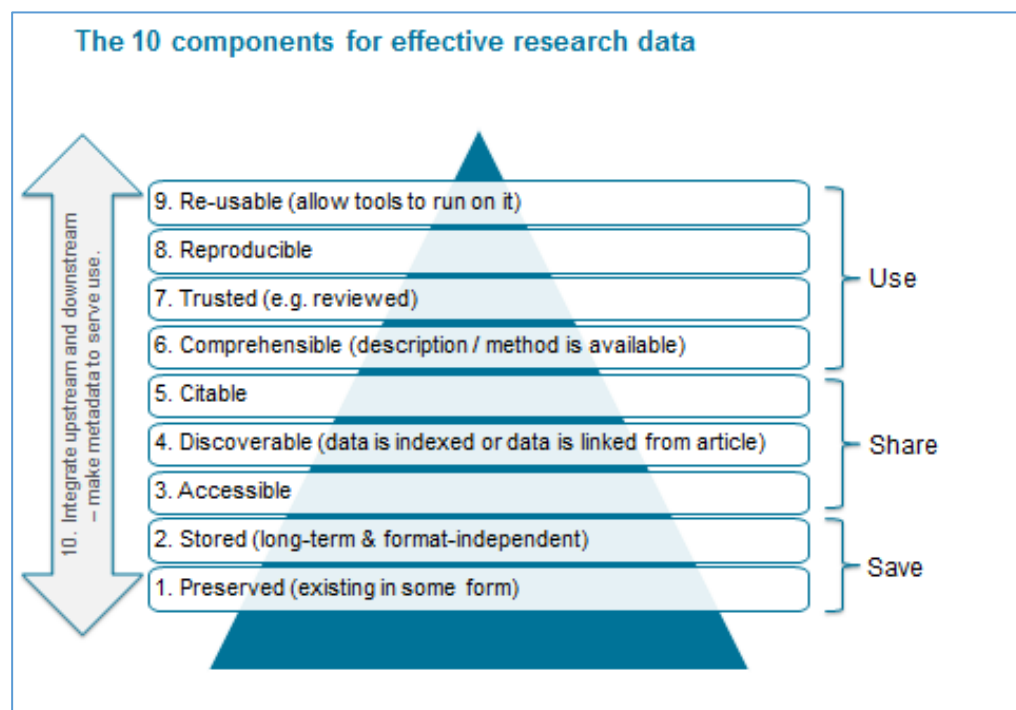# Elsevier response to Concordat on Open Research Data

Elsevier welcomes the opportunity to express our support for the UK Concordat on Open Research Data, published by the Concordat Working Group. The principles expressed in the Concordat align with Elsevier's support for open research data, as exemplified in our research data policy and in the tools and services we have developed to help make open research data a reality. We include some specific comments on the principles below.

Elsevier supports the intention to the Concordat to establish sound principles of good practice rather than to mandate, codify or require specific activities establishing open research data. Expectations and resources for data curation and use vary across the broad range of academic disciplines, so a more flexible approach is welcome, in view of these differences.

We also welcome the Concordat's recognition that others, including publishers, play a vital role in helping make the principles expressed in the Concordat a reality and that there needs to be cooperation between stakeholders to achieve this. We agree, and Elsevier already has a number of partnerships and collaborations in the research data space. We would welcome the opportunity to explore further partnerships with all stakeholders to implement these principles. As noted in the Concordat, open data is more than mere disclosure; data must be discoverable and accessible. The below graphic demonstrates the steps that need to be taken, in collaboration with stakeholders, to ensure real open data is a reality:

The 10 components for effective research data

9. Re-usable (allow tools to run on it)
8. Reproducible
7. Trusted (e.g. reviewed)
6. Comprehensible (description / method is available)
Use

5. Citable
4. Discoverable (data is indexed or data is linked from article)
3. Accessible
Share

2. Stored (long-term & format-independent)
1. Preserved (existing in some form)
Save

10. Integrate upstream and downstream – make metadata to serve use.

We have included below a number of examples as to how we are already supporting and implementing the principles laid out within this Concordat, and where we see further opportunities for making these principles a common reality.

Principle #1 Open access to research data is an enabler of high quality research, a facilitator of innovation and safeguards good research practice.

We agree that research data is the foundation on which scientific, technical and medical knowledge is built and as a publisher we realize how important it is that good research data underlie the publications that are used to report scientific progress. However, there are challenges in making data accessible and shareable. Elsevier is addressing these challenges by creating solutions that support researchers to store, share, discover and re-use data. Examples of these are:

- **Database Linking**

Database Linking1 is available for researchers and data repositories as one method to ensure that data that is relevant for scientific, technical and medical research can be easily discovered and accessed. The bidirectional links between data repositories and online articles on our publication platform, ScienceDirect provides readers with one-click access to relevant, trusted data that may help to validate research or drive further investigations. We currently collaborate with over 40 research data repositories, including PDB, GenBank, and NIST. We support various mechanisms to set up such links, including data DOI's, banners next to the article and relevant data accession numbers from supported databases, which automatically link in text references to data deposited in public databases.

- **Data journals**

Not all research leads to publications, while the data may still be useful. In order to lower the barriers for data sharing, Elsevier has started a program to launch new Data Journals. These allow researchers to share (and receive credit) for methods, data, software, without having to publish a full article. (Open) data journals, and data sections in existing journals, enable authors to have their research data peer-reviewed and cited. It will also make sure readers can find, use and analyse the data hosted in external databases or submitted as supplementary data. Examples of recently launched data journals are Genomics Data and Data in Brief.

- **Common Data Citation Principles**

Elsevier is a signatory of the Joint Declaration of Data Citation Principles. Elsevier endorsed these principles in 2014 to help make research data become an integral part of the scholarly record, properly preserved and easily accessible, while ensuring that researchers get proper credit for their work.

Principle #2 Good data management is fundamental to all stages of the research process and should be established at the outset.

It is crucial for data sharing that researchers manage their data well during the research process. Elsevier is therefore working on data management tools that should make this as easy as possible for researchers. We are working on an Electronic Lab Notebook (ELN), so that researchers can keep track easily of their experiments, in digital format, in their ELN. We believe tools such as these will help with a more structured data acquisition, making data more easily shared and reused. This ELN is now available for selected launching partner institutions.

In addition to linking to existing open data repositories, we offer our own open data repository via Mendeley, (data.mendeley.com), and are seeking pilot partners to help steer its development. It accommodates any number of files up to 2GB each, and supports viewing and creating datasets (including photos) on any device. The data uses DOIs and versioning following Force11 guidelines and will be preserved in dark archive storage. Mendeley Data allows researchers to share data with collaborators, but also to make data public and thereby accessible and discoverable.

---

1 Data Base Linking: https://www.elsevier.com/books-and-journals/content-innovation/data-base-linking

Our PURE platform can also be used to capture, publish and cite research datasets.

Principle #3 Data must be curated so that they are accessible, discoverable and useable.

We fully agree that in order to make data open it must be not only accessible, but also discoverable and useable. In addition to those tools and services outlined under principle 1, we support principle 3 in the following ways:

- **Research Data Discovery**
We continuously expand the number of indexing agreements with search and discovery tools such as abstracting and indexing databases, library discovery services, web search engines and other innovative search and discovery tools such as Mendeley, ScienceScape and PubChase. In addition, we are improving the discoverability of content from ScienceDirect on third party platforms.

- **DataLink**
Elsevier has also developed the DataLink platform2 to lower the barriers that make it difficult for genomics researchers to promote and discover data. DataLink consists of a database search engine which facilitates genomic data discovery by retrieving results from all the major public databases and articles indexed in leading archives. These include data sets from GenBank, Gene Expression Ominbus and ArrayExpress as well as articles from ScienceDirect and PubMed. In addition, we are searching a series of data repositories including DataVerse, Dryad, Columbia's EarthChem, and PubMed Central.

- **Data Visualization**
In close collaboration with selected data repositories, Elsevier has developed a number of data integration and visualization applications that are shown next to the article on ScienceDirect. These applications build further on tagged entities or banner links to visualize data and integrate it into the online reading experience. This provides readers with deeper insights through applications such as 3D models; some examples include the Protein Viewer (with PDB), the PANGAEA data visualization tool, and the Genome Viewer (with NCBI).

- **Data curation**
Elsevier already works in several areas to help curate data to common taxonomies (e.g. health, life sciences, engineering, chemistry to name a few). We would welcome partnerships to leverage existing Elsevier taxonomies more broadly, or conversely to use existing taxonomies as a standard in a field – an example of the latter being the NIH Mesh taxonomy that has been adopted by Elsevier.

Principle #4 Open access to research data carries a significant cost, which should be respected by all parties.

We agree with this principle, as set out in our own research data policy: Researchers invest substantially to create and interpret data and others such as data archives, publishers, funders and institutions further add value and/or incur significant cost. In all such cases these contributions need to be recognized and valued.

Principle #5 There are sound reasons why the openness of research data may need to be restricted but any restrictions must be justified and justifiable

Elsevier recognizes that there are cases where research data cannot be made openly available. Therefore, in addition to our solutions that enable researchers to share their data, we have developed the Data Profile. The Data Profile allows authors to provide a structured summary of the data that has been presented in an article and describe its availability. A Data

---

Profile includes brief descriptions of the data and details about how, if, and where the data is available. When provided by the author it will also include direct links to individual data sets. With the data profile we give authors the option to share their data in a controlled manner but also allow authors to state a reason why data cannot be made available at this time. This ensures authors consider the availability of their data and also in cases where data cannot be shared, still increases transparency.

There are cases where privacy sensitive data (e.g. patient health records) will require even further controlled environments. Elsevier helps research institutions manage, curate, and re-use these datasets by integrating these in a well-controlled Health Analytics database that can be used for data-reuse without compromising on privacy or ethical requirements. See http://www.elsevieranalytics.de/

Principle #6 The right of the creators of research data to reasonable first use is recognised.

We agree that open research data should be formally acknowledged as a legitimate output of the research process and support the principle of formal recognition through the development and use of metrics that allow the collection and tracking of data use and impact. We are already enhancing the ways which we develop and use article level metrics and we are leading the efforts within NISO, as well as working with CASRAI, RDA, FORCE11, COUNTER, BioCADDIE and others, on research data metrics.

Principle #7 Use of others' data should always conform to legal, ethical and regulatory frameworks including appropriate acknowledgement.

We agree with the principle that all data should be a citable object, independent of the paper. As stated above, Elsevier is a signatory of the Joint Declaration of Data Citation Principles. Elsevier endorsed these principles in 2014 to help make research data become an integral part of the scholarly record, properly preserved and easily accessible, while ensuring that researchers get proper credit for their work.

As an example of how we put this into practice, we offer our own data repository via Mendeley, and are seeking pilot partners to help steer its development. It accommodates any number of files up to 2GB each, and supports viewing and creating datasets (including photos) on any device. The data uses DOIs and versioning following Force11 guidelines and will be preserved in dark archive storage. Further, our Open Data Pilot makes raw research data available under a CC-BY license.

To further facilitate good practice, reusability and aid the non-duplication of research, Mendeley Data will allow administrators of institutional data repositories to create custom templates for their datasets. For example, repository manages can easily create a template that states that a dataset must have specific metadata fields and will only allow those file types to be uploaded.

Principle #8 Data supporting publications should be accessible by the publication date and should be in a citeable form.

In line with the statement set out in the Concordat, the first principle of our research data policy is that raw research data should be made available free of charge to all researchers wherever possible and with minimal reuse restrictions. Our comments above demonstrate how we support this principle, alongside helping to ensure data is citable, for example through the use of data DOIs. We also provide specific and additional outlets for research data, through our journals.

Another easy way for authors to store and share research data is to make their data available as a supplementary file associated with their article. Through our Open Data pilot we are

ELSEVIER

providing authors with the opportunity to make their supplementary files with raw research data available open access on ScienceDirect. We invite authors to submit their raw research data, which then we make openly available on ScienceDirect under a CCBY license3. Authors retain copyright of such supplemental raw research data and the journals are able to store, publish, archive, and link to the data as a supplementary data file.

However, not all data can immediately be made available publicly and it might be useful to investigate researcher-controlled or field-dependent embargo periods, after which all data becomes available. Elsevier would welcome the opportunity to participate in any such investigations.

Principle #9 Support for the development of appropriate data skills is recognized as a responsibility for all stakeholders.

Given our frequent interactions with the research community, we feel that we can contribute to the development of data management skills of researchers. We regularly give talks to explain to researchers how they can share their data, for example at conferences and through webinars. We are currently planning a webinar series in 2016 which will be specifically aimed at explaining data management and data sharing practices to early career researchers. In addition, we regularly publish about our data activities to give researchers a better idea of their options so that they can take this into account when carrying out their research and deciding how to deposit their data.

Principle #10
Regular reviews of progress towards open access to research data should be undertaken.

We would fully support and would be happy to participate in any such review.

---

3 Creative Commons license definition: http://opendefinition.org/licenses/cc-by/