

Managing research data: a guide to biocuration

May 2010

A Research Information Network factsheet

www.rin.ac.uk

Over the last few years there has been a growing belief that publicly funded research should be made as widely available as possible. Though in the past this has been limited to published research outputs, such as journal articles, there has been a recent movement for the inclusion of the underlying research data to be made available as well.

Research funders and policy-makers believe that data are a key part of the scholarly record and that making data available will stimulate innovation and drive new discoveries. As a result most UK research funders have policies which encourage or require grant holders to manage their data effectively and make them openly accessible, using suitable formats with established standards.

Effective data curation depends on good support and guidance for researchers to help them in managing large volumes of data and information (RIN 2009). There is also a need to develop standards (for things such as data collection, nomenclature and tagging etc.), user-friendly tools, and better quality metadata and annotation of data of all kinds. Greater support for, and acknowledgment of, the individuals involved in these activities is essential.

Ontologies: An ontology is a controlled vocabulary of well-defined terms with specified relationships between those terms, capable of interpretation by both humans and computers.

Data curation: The selection, preservation, maintenance, collection and archiving of research data.

Nomenclature: A system of words used to name things in a particular discipline.



What is a biocurator?

In the simplest terms, biocurators are professional scientists who collect, annotate, and validate information that is housed within biological databases. They are also likely to be involved in the design, development and dissemination of biological databases, including developing ontologies, setting nomenclature standards, designing database interfaces, analyzing data to elucidate more information, scripting queries, education and training of database users. Biocurators may also be involved in developing software and tools to be used by the database users or other biocurators.

What does a career in biocuration look like?

The day-to-day tasks of biocurators depend largely on the type of resource on which they work. Some of the most common subsets of resources include: genome/protein sequence databases, pathway/interaction databases, protein structure databases, model organism databases and expression databases. The range of resources available to today's scientists is wide and varied, and is constantly growing, such that almost every area of biological sciences is covered.

In most cases, a primary source of the data populating the resource is the literature, and so one of the key jobs of biocurators is to read the publications and extract the relevant information to be included in the database. Data may also come directly from a data provider (researcher) or another database, in which case the role of biocurators is to liaise with providers to ensure the data received are validated and formatted appropriately for presentation in the resource. In all cases, this will involve the use of specialist tools and software alongside biocurators' expertise. Biocurators will assign, or validate, IDs and standard naming conventions to the data.

Required skills

Biocurators should have experience in scientific research as well as being competent in database management systems, multiple operating systems and scripting languages, and bioinformatics. Most biocurators have come to the field from either an informatics or a bioscience background. Some formal training options for those considering a career in biocuration are now available. In the US, the University of Illinois at Urbana-Champaign has a Masters Degree in Biological Information Specialist (specialization in data curation) and the University of Michigan has a Master of Science in Information (specialization in Preservation of Information). In the UK, there are courses offered by the Digital Curation Centre based in Edinburgh. Biocurators play an incredibly important role in streamlining submission to

databases and data curation, as well as enriching the content and usability of databases. As more and more data of varying types are produced, and more emphasis is placed on re-using existing data in innovative ways, biocurators are increasingly needed as an integral part of the research landscape.

Research funders and research institutions need to support and promote biocuration as a viable career option, to attract the highly qualified individuals needed in this field.

Resources and useful links

International Society of Biocuration
www.biocurator.org and email intsocbio@gmail.com

Courses and advice

Digital Curation Centre (DCC)
www.dcc.ac.uk

University of Illinois at Urbana-Champaign Masters Degree in Biological Information Specialist
<http://cirss.lis.illinois.edu/SciCom/bis.html>

University of Michigan, Master of Science in Information
www.si.umich.edu/msi/pi.htm

Further reading

Connotea biocuration articles
www.connotea.org/tag/biocuration

DATABASE: The journal of Biological Databases and Curation
<http://database.oxfordjournals.org>

Nature: Big Data Issue
www.nature.com/nature/journal/v455/n7209/full/455047a.html

Nucleic Acids Research Annual database issue
http://nar.oxfordjournals.org/content/vol38/suppl_1/index.dtl

Patterns of information use and exchange: case studies of researchers in the life sciences (RIN 2009)
www.rin.ac.uk/patterns-information-exchange

PLoS Collections: Biocurator
www.ploscollections.org/article/browseIssue.action?issue=info%3Adoi%2F10.1371%2Fissue.pcol.v03.i05

To share or not to share: research data outputs (RIN 2008)
www.rin.ac.uk/to-share-research-data-outputs

Some examples of databases
<http://colleagues.biocurator.org/affiliations>

UK research funders data policies
www.dcc.ac.uk/resources/policy-and-legal/funders-data-policies

Get in touch with us

If you have any questions or would like to get involved in our work, please contact Branwen Hide, Liaison and Partnerships Officer (email branwen.hide@rin.ac.uk or telephone 020 7412 7759).

About the Research Information Network

The Research Information Network has been established by the higher education funding councils, the research councils, and the UK national libraries. We investigate how efficient and effective the information services provided for the UK research community are, how they are changing, and how they might be improved for the future. We help to ensure that researchers in the UK benefit from world-leading information services, so that they can sustain their position as among the most successful and productive researchers in the world. All our publications are available on our website at www.rin.ac.uk