Bring data to life: data management for the biomolecular sciences

Marta Teperek [1], Yasemin Türkyilmaz-van der Velden [2], Ewan Birney [3], Elisabeth Busch-Nentwich [4], Nancy George [5], Keeva Cochrane [6], Maria Levchenko [7], Sirarat Sarntivijai [8], Sarah Jones [9], Rob Hooft [10], Sarah Morgan [11], Melissa Burke [12]

- Cross-domain
- Beginner
- 3 hours

Most researchers know that data management is an important part of their work and is essential for getting the most out of your data. But what exactly is data management and what can you do to ensure that you’re managing your data well?

From June – August 2018 EMBL-EBI ran a webinar series on data management for the biomolecular sciences. This series introduces key concepts in data management, highlights common myths and provides biological researchers with hints and tips on how to approach data management in their own research.

Learning objectives:

- Describe the data management cycle
- Discuss benefits and challenges of data sharing
- Explain what happens to data after it is shared
- Identify appropriate data repositories and their submission requirements
- Describe how ontologies and standards are used to annotate biological data
- Identify appropriate resources and tools for data management
Managing and making the most of your data

In this webinar Marta Teperek & Yasemin Türkyilmaz-van der Velden provide an overview of what data management is highlighting common misconceptions, pitfalls and best practice. They pose the questions 'Is there a reproducibility crisis in science'? and 'How much data would lose if your laptop was stolen?'. This is followed up with an exploration of the selfish benefits of data management and a few hints, tips to get you started on your data management journey.

For further reading Marta and Yasemin recommend:

For advice on data management:

- Digital Curation Center [13]
- UK Data Service [14]

For information on Metadata standards:

- FAIRsharing [15]

As a tool for data management plans:

- DMPonline [16]

Data sharing:

- Registry of data repositories [17]
- EC report on evaluation based on Open Science Practices [18]

This webinar took place on 28 June 2018. The slides from the webinar can be downloaded here [19].

Why share your data?

Presented in two parts, this webinar will explore the value of sharing your data from an institutional and researcher perspective.

Our first speaker, Ewan Birney, is one of the directors of EMBL-EBI. He speaks about how the growth in data over the past decade has changed the way that we approach biology and how EMBL-EBI provides some of the infrastructure to allow data sharing. He finishes up by highlighting the benefits of data sharing to the EMBL-EBI and shares his top four reasons for sharing data.

Our second speaker, Elisabeth Busch, is the head of the Vertebrate Genetics and Genomics Group at the Wellcome Sanger Institute and the University of Cambridge. She tells us about her experiences in sharing her own research data, and the reasons that she felt compelled to share her data including the advantages for her own research.

This webinar was recorded on 2 July 2018 and is for scientists at all levels who need to manage data or build data.
management plans as part of their research.

The slides from this webinar are available through the following links:

- Ewan's slides [20]
- Elisabeth's slides [21]

**What happens to your data?**

This webinar focuses on what happens to your data after you submit it to a public repository and addresses the question of data ownership.

Keeva and Nancy introduce the concepts of FAIR data, metadata [22] and standards. They then show you where you can find an appropriate place to share your data and give examples of how metadata and standards are used when you submit data to Metabolights and ArrayExpress [23]. They also answer a few frequently asked data submission-related questions.

Maria wraps up the webinar with a closer look at data ownership and discussion on conditions for reusing data, citing data and getting credit for the data that you have produced.

Keeva, Nancy and Maria mention a number of resources that you might find useful when sharing your data:

- To learn more about FAIR data standards have a look at this article on The FAIR Guiding Principles for scientific data management and stewardship [24].
- EMBL-EBI Data submission wizard [25] can be used to find an appropriate database.
- ORCIDs [26] are unique researcher identifiers can be used to claim both articles and data.

This webinar will take place on Thursday 12 July 2018. The slides from the webinar can be downloaded here [27].

**Giving data context, structure and meaning**

Metadata [22], standards and ontologies are common ways of providing data with context, structure and meaning. In this webinar Sira explores the idea of once data are ontologised (mapped to standard structured vocabulary), they become linkable to build a bigger infrastructure of knowledge network and provides a few scientific examples of how knowledge are drawn from linked data. She also provides a brief overview of tools that can be used to give data context, structure and meaning at different stages of the data lifecycle.

To learn more about ontologies you can watch Sira's lecture Ontologies 101 [28] (requires Adobe Flash Player) or read James Malone's blog Common Ontology Questions #1: What is it you do again? [29].

This webinar took place on Thursday 19 July 2018. The slides from this webinar can be downloaded here [30].
Having a data management plan is essential if we're to make the most of our own data and share it with others. Just like planning an experiment, we should plan what we are going to do with our data and how we are going to do it to ensure that we (and others) can use and re-use the data. If you have never made a data management plan before the process may initially seem daunting, but fortunately there are many tools out there to help you think through the what, where, how, why and when of data management. In this webinar Sarah Jones from the Digital Curation Centre and Rob Hooft from ELIXIR Netherlands show us two tools - DMP online [16] and the ELIXIR Data Stewardship Wizard [31] - and explain how they can help you apply data management and FAIR data practices in your own research.

This webinar took place on 26 July 2018. The slides from this webinar can be downloaded here [32].

**Myths and best practice**

The final webinar of this series Melissa Burke and Sarah Morgan provide a summary of the series. They point out key hints and tips for managing your own data and address some of the most burning data management questions that have arisen while making the series.

To learn more about data management you can:

- Speak to the data management team at your own institute.
- Visit Fairsharing.org [33] to learn about metadata [22], standards, databases and data policies.
- Use re3data.org [34] or the EBI's data sharing wizard [25] to find a repository that is appropriate for the type of data you are creating.
- Read about FAIR guiding principles [24].

This webinar took place on 2 August 2018. The slides from this webinar can be downloaded here [35].

**Contributors**

Yasemin Türkyılmaz-van der Velden [2]

TU Delft
Data Steward
Yasemin Turkyilmaz-van der Velden works as the Data Steward at the TU Delft Faculties of Applied Sciences, and Mechanical, Maritime and Materials Engineering. She obtained her MSc degree at Radboud University Nijmegen from the program ‘Molecular Mechanisms of Disease’. Then, she did her PhD at the Molecular Genetics Department of Erasmus MC Rotterdam where she studied UV-induced DNA damage repair using proteomics and live cell imaging approaches. During her academic journey, she daily managed data during acquisition, analysis, and publication stages and therefore she has a good understanding of researchers’ perspective and experiences about data management. As a Data Steward, she provides customized and discipline-specific data management support to researchers at every stage of the research lifecycle, with the eventual goal of achieving a cultural change in research data management at daily practice.

Marta Teperek
TU Delft
Data Stewardship Coordinator

Dr Marta Teperek did a PhD in epigenetics at the University of Cambridge. Having first-hand experience of problems that researchers face on a day-to-day basis, with the journals’ impact factor, and not the quality of the research process, dictating the future of their academic career, she got professionally involved in advocating for Open Research and for better transparency in science. In 2015 she joined the University of Cambridge and led the creation and development of the Research Data Management Facility, supporting researchers at the University of Cambridge in good management and sharing of research data. While at Cambridge, Marta initiated and oversaw the Data Champions programme and the Open Research Pilot. In August 2017 Marta moved to TU Delft in Netherlands, where she is leading the Data Stewardship project.

Marta serves on the Editorial Board of the Data Science Journal and Co-Chairs the Research Data Alliance Libraries for Research Data Interest Group. She is a frequent speaker at international conferences and events related to Open Science and to Research Data Management, regularly publishes blog posts and peer-reviewed publications on these topics, and is professionally active on Twitter.

Ewan Birney
EMBL-EBI
Director of EMBL-EBI and Senior Scientist
Ewan Birney is Director of EMBL-EBI with Dr Rolf Apweiler, and runs a small research group.

Ewan completed his PhD at the Wellcome Sanger Institute with Richard Durbin. In 2000, he became Head of Nucleotide data at EMBL-EBI and in 2012 he took on the role of Associate Director at the institute. He became Director of EMBL-EBI in 2015. Ewan led the analysis of the Human Genome gene set, mouse and chicken genomes and the ENCODE [38] project, focusing on non-coding elements of the human genome. Ewan’s main areas of research include functional genomics [39], DNA algorithms, statistical methods to analyse genomic information (in particular information associated with individual differences in humans and Medaka fish) and use of images for chromatin structure.

Ewan is a non-executive Director of Genomics England, and a consultant and advisor to a number of companies, including Oxford Nanopore Technologies, Dovetail Genomics and GSK. Ewan was elected an EMBO member in 2012, a Fellow of the Royal Society in 2014 and a Fellow of the Academy of Medical Sciences in 2015.

He has received a number of awards including the 2003 Francis Crick Award from the Royal Society, the 2005 Overton Prize from the International Society for Computational Biology and the 2005 Benjamin Franklin Award for contributions in Open Source Bioinformatics.

---

Nancy George [5]

EMBL-EBI
Scientific Curator - Papatheodorou team: Gene Expression

In her role as a curator [40], Nancy works with scientists, supporting the submission of gene expression [41] data to ArrayExpress [23].

---

Keeva Cochrane [6]

EMBL-EBI
Scientific Database Curator - O’Donovan team: Metabolomics

Keeva is a technical officer within the Metabolomics [42] team at EBI, managing curation [43] and outreach. She
has a degree in biochemistry and PhD in Molecular Therapeutics. Keeva has significant experience of working with small and large biological and chemical datasets having worked in drug development R&D in academia and industry for many years. As part of her role at in the metabolomics team she supports submission of metabolomics data to the MetaboLights database.

Maria Levchenko [7]
EMBL-EBI
Europe PMC Community Manager - McEntyre team: Literature services

Maria is the Community Manager for Europe PMC [44], a literature repository and partner in PubMed [45] Central International. In her role Maria engages with all stakeholders of Europe PMC to communicate their feedback to the team of developers so that they can deliver the best tools for scholarly publishing and communication. She has a PhD in Biochemistry from The University of Göttingen.

Sirarat Sarntivijai [8]
ELIXIR
ELIXIR Interoperability Platform Coordinator

Sira works in ELIXIR Technical Team to support the scientific research across ELIXIR's 21 member states. Prior to joining ELIXIR in March 2018, she had been part of the Samples, Phenotypes, and, Ontologies Team (SPOT) at the EMBL-EBI as an ontologist. Her expertise lies in the area of biomedical big data integration by the implementation of ontologies, linked data, and semantic web technologies. She is a developer of the Cell Line Ontology [46], the Ontology of Adverse Events, and the Experimental Factor Ontology [47] (EFO). She is also an active member of the OBO Foundry community.

Prior to joining the EMBL-EBI, Sira received her PhD in Bioinformatics from the University of Michigan, Ann Arbor, and an ORISE post-doctoral fellowship from the US FDA's Office of Clinical Pharmacology, Office of Translational Sciences.
Sarah Jones [9]
Digital Curation Centre - University of Glasgow
Associate Director

Sarah coordinates work on the DCC's Data Management Planning tool - DMPonline - and undertakes research on data policy and data management planning. To learn more about Sarah's work please visit her profile [48] on the DCC website.

Rob Hooft [10]
Dutch Techcenter for Life Sciences
Programme Manager

Rob is working for DTL managing the Dutch tasks in the European ELIXIR infrastructure for life science data. As part of these roles, Rob has been building up a body of knowledge on FAIR research data stewardship since early 2014. To learn more about Rob's work please visit his website [49].

Sarah Morgan [11]
EMBL-EBI
Training Coordinator
Sarah has a BSc and MSc in Biomedical Sciences, a PhD from Cranfield University and an MA in Healthcare Law and Ethics from the University of Manchester. Over a 10 year academic career her research has focused on tumour biomarker characterisation and cell-surface interaction. Beyond research she took an active role in the provision of postgraduate taught courses for the University as Lecturer in Molecular Medicine, leading the development and direction of a number of MSc Courses including the MSc in Translational Medicine, the MSc in Molecular Medicine, and more recently acting as director for the MSc Programmes in Advanced Biosciences. She has lectured on a variety of biomedical techniques and topics, including ethics and governance of research.

Melissa Burke

EMBL-EBI
Scientific Training Officer (e-learning): Training Team

Melissa is the Scientific Training Officer (e-learning) for the Training Team at the EMBL-EBI. She joined the Training Team in July 2016 after having worked as a Scientific Curator for ArrayExpress/Expression Atlas at the EMBL-EBI. She has a PhD in Molecular Parasitology and has worked internationally as a postdoctoral researcher specialising in the functional genomics of infectious diseases.

Your feedback

Please tell us what you thought about this webinar. Your feedback is invaluable and helps us to improve our courses and thus enhance your learning experience.

Source URL: https://www.ebi.ac.uk/training/online/course/bringing-data-life-data-management-biomolecular-sciences

Links
[1] https://www.ebi.ac.uk/training/online/trainers/M.Teperek_17994
[2] https://www.ebi.ac.uk/training/online/trainers/Y.Turkylmaz-vanderVelden_17996
[3] https://www.ebi.ac.uk/training/online/trainers/birney_17992
[4] https://www.ebi.ac.uk/training/online/trainers/emb_18000
[5] https://www.ebi.ac.uk/training/online/trainers/ngeorge_17998
[6] https://www.ebi.ac.uk/training/online/trainers/keeva_14782
[7] https://www.ebi.ac.uk/training/online/trainers/levchmar_10147
[8] https://www.ebi.ac.uk/training/online/trainers/sirarat.santivijai_18098
[9] https://www.ebi.ac.uk/training/online/trainers/Sarah.Jones_18449
[10] https://www.ebi.ac.uk/training/online/trainers/rob.hooft_18451
[11] https://www.ebi.ac.uk/training/online/trainers/sarahm
[12] https://www.ebi.ac.uk/training/online/trainers/mburke
[14] https://www.ukdataservice.ac.uk/
[16] https://dmponline.dcc.ac.uk/
[20] https://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/Ewan_Birney_data_sharing_webinar.pptx
[21] https://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/Elisabeth_Busch_webinar_July2018.pptx
[22] https://www.ebi.ac.uk/training/online/glossary/metadata
[23] https://www.ebi.ac.uk/training/online/glossary/arrayexpress
[24] https://www.nature.com/articles/sdata201618
[25] https://www.ebi.ac.uk/submission/
[26] https://orcid.org/
[27] https://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/What%20happens%20to%20your%20data.pptx
[28] https://connect.umms.med.umich.edu/p9jhc44d8u/?launcher=false&amp;fcsContent=true&amp;pbMode=normal
[29] https://drjamesmalone.blogspot.com/2012/06/common-ontology-questions-1-what-is-it.html?view=classic
[30] https://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/published_SSarntivijai_TrainOnline_GivingDataContext.pdf
[31] https://dmp.fairdata.solutions/
[32] https://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/Tools%20for%20data%20management%20planning%20compressed.pptx
[33] https://fairsharing.org/
[34] https://www.re3data.org/
[35] https://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/Myths%20and%20best%20practice.pptx
[36] https://www.ebi.ac.uk/training/online/glossary/proteomics
[38] https://www.ebi.ac.uk/training/online/glossary/encode
[39] https://www.ebi.ac.uk/training/online/glossary/functional-genomics
[40] https://www.ebi.ac.uk/training/online/glossary/curator
[41] https://www.ebi.ac.uk/training/online/glossary/gene-expression
[42] https://www.ebi.ac.uk/training/online/glossary/metabolomics
[43] https://www.ebi.ac.uk/training/online/glossary/curation
[44] https://www.ebi.ac.uk/training/online/glossary/europe-pmc
[45] https://www.ebi.ac.uk/training/online/glossary/pubmed
[46] https://www.ebi.ac.uk/training/online/glossary/ontology
[47] https://www.ebi.ac.uk/training/online/glossary/experimental-factor-ontology
[50] https://www.ebi.ac.uk/training/online/glossary/biomarker