Illustration inspired by the work of John Maynard Keynes

ADVANCING
DISCOVERY

BETTER SCIENCE
THROUGH BETTER DATA
2018 PROGRAMME

Springer Nature & The Wellcome Trust
14th November 2018 | 09:00 to 17:00

The Natural History Museum
Flett Theatre, Cromwell Road | SW7 5BD
London United Kingdom

Introduction

Iain Hrynaszkiewicz
Head of Data Publishing, Springer Nature

With this being the fifth edition of Better Science through Better Data, it seems pertinent to reflect on the last few years' events, and share a few "data points" with you. Since our first small event in 2014, combined, the conference has seen:

- Nearly 400 in-person attendees, with attendance growing strongly each year
- 99% of attendees saying they would recommend the event to a colleague
- Nearly 7000 tweets on the conference hash tags, with a potential online audience of tens of millions
- Awards for 20 previously unpublished science writers, with several of these past winners since establishing careers in science writing and communication

We have also heard from an attendee of our second event, in 2015, who was inspired by the event to pursue their postdoctoral career in applications of open science in their field.

On to 2018, and I thank the programme committee for their advice and support in crafting – and active participation in – this year's programme. Two themes that emerged in planning this year's conference were "mentoring open science" and "making data Findable, Accessible, Interoperable and Reusable (FAIR) throughout the research lifecycle".

To help us explore these themes, we will hear from both senior and junior researchers about their approaches to practicing and supporting open, reproducible research – in particular in our closing panel. We have, also, arranged the keynote talks to take us on a journey, from data generation and analysis through to data management and stewardship, in our morning session. Followed by, in the afternoon, publication of research data, onto, finally, reuse of research data - with data-driven approaches to journalism with data visualisation.

We are sticking largely to the same format as last year, as your feedback suggests we have found the right the formula. Because we received so many exciting proposals – and because hearing from researchers "on the ground" is one of the most important parts of this event – we have squeezed 14 lightning talks into the sessions that straddle our lunch break.

The event is again brought to you through a partnership between Springer Nature, publishers of the Nature Research journal *Scientific Data*, and the Wellcome Trust, a global biomedical research charity that supports thousands of researchers. We're grateful to Wellcome for their continued support, and to our venue hosts for 2018, here at London's iconic Natural History Museum.

As always, we are as open as possible at the event. The talks are being live-streamed online, and will be shared under open licences afterwards as will our amazing digital scribe images. Debate on social media, on the #scidata18 tag, is strongly encouraged.

Iain Hrynaszkiewicz

Event Schedule

Registration and coffee	**09:00**
Welcome/conference opening Iain Hrynaszkiewicz, Springer Nature	**09:30**
Keynote #1: The Rise of the Data Generalist Rebecca Boyles, Senior Manager, Bioinformatics and Data Science, RTI International	**09:45**
Keynote #2: On a (cultural) journey towards FAIR data Marta Teperek, Data Stewardship lead, TU Delft	**10:30**
Coffee Break	**11:00**
Lightning talks session #1:	**11:30**

- Overcoming data barriers for regional-scale coastal-impact analysis
  Claudia Wolff, Christian-Albrechts-University Kiel

- Share for Rare: Promoting Data-Sharing through Japan's Initiative on
  Rare and Undiagnosed Diseases (IRUD)
  Takeya Adachi, Japan Agency for Medical Research and Development
  (AMED)

- Mapping population distributions, demographics and dynamics
  Andrew Tatem, University of Southampton: WorldPop

- Gridded birth and pregnancy datasets for Africa, Latin America and
  the Caribbean
  Natalia Tejedor Garavito, University of Southampton

- Multi-centre Epilepsy Lesion Detection Project conducting clinical
  research in an open-science framework
  Sophie Adler, University College London

- Consent for data archiving in end of life care research: reflections on
  two qualitative studies
  Jane Seymour, University of Sheffield

- Sharing Electrical Impedance Tomography and neuroimaging data
  from stroke patients
  James Avery, University College London

Lunch	**12:30**

Event Schedule

Lightning talks session #2:   13:15

- Compact Identifier resolution services
  Sarala M. Wimalaratne, EMBL-EBI: Identifiers.org

- Live Audit and Feedback for Trials Transparency
  Nicholas J. DeVito, University of Oxford

- STRENDA DB – Monitoring the completeness of information in data reports
  Carsten Kettner, Beilstein-Institut

- Meaningful and reproducible statistics: Does my data hold what it promises?
  Andrej-Nikolai Spiess, University Hospital Hamburg-Eppendorf

- Materials Cloud, An Open Science Portal for FAIR Data Sharing
  Aliaksandr Yakutovich, École Polytechnique Fédérale de Lausanne

- Counting reuse to Make Data Count
  Helena Cousijn, DataCite

- Nine good things about open science (and one bad thing)
  Alasdair Rae, University of Sheffield

Keynote #3: Theme: Publishing research data   14:15

Magdalena Skipper, Editor in Chief, *Nature*

Coffee Break   15:00

Keynote #4: Theme: data reuse and visualisation   15:20

John Burn-Murdoch, Senior Data-Visualisation Journalist, *Financial Times*

**Panel discussion: The responsibility of reproducibility:
whose job is it to change the status quo?**   15:40

Moderator: Kirstie Whitaker

Panellists:

Paola Quattroni, Cancer Research UK
Natalia Tejedor, University of Southampton
Sue Fletcher-Watson, University of Edinburgh
Zaheer-Ud-Din Babar, University of Huddersfield

Conference closing remarks   16:45

Conference chair: Iain Hrynaszkiewicz – Head of Data Publishing,
Springer Nature

**Drinks sponsored by Springer Nature at Angelsea Arms
(10 min walk from conference venue)**   17:30

Speaker Biographies

Iain Hrynaszkiewicz

Head of Data Publishing, Springer Nature

Iain is Head of Data Publishing in the Open Research Group at Springer Nature, where he develops new areas of research data publishing and data policy. This includes efforts to make research data more visible and reusable by developing links with data repositories, and data journals such as *Scientific Data*. He has led various initiatives and published numerous articles related to data sharing, open access, open data and reproducible research.

Dr Kirstie Whitaker

Alan Turing Institute

Kirstie completed her PhD in Neuroscience at the University of California, Berkeley in 2012 and holds a BSc in Physics from the University of Bristol and an MSc in Medical Physics from the University of British Columbia. She was a postdoctoral researcher in the Department of Psychiatry at the University of Cambridge from 2012 to 2017. Dr Whitaker uses magnetic resonance images to understand the changes in the brain's structure and function that underlie the emergence of schizophrenia and depression. She is particularly passionate about ensuring that work is reproducible and can be replicated in independent data sets. She is a Fulbright scholarship alumna and 2016/17 Mozilla Fellow for Science. Kirstie was named, with her collaborator Petra Vertes, as a 2016 Global Thinker by Foreign Policy magazine.

Rebecca Boyles

Senior Manager, Bioinformatics and Data Science, RTI International

Rebecca has nearly 20 years of experience in environmental health science research with a focus on data science applications, including data management technologies and data integration. She has particular expertise designing data-driven research methods and architectures to support research collaborations. She is currently the Co-Director of Infrastructure for the Environmental Influences on Child Health Outcomes (ECHO) Data Analysis Center and a member of the North Carolina Precision Health Collaborative Steering Committee.

Dr Marta Teperek

Research Data Stewardship Lead, TU Delft

Marta completed a PhD in molecular biology at the University of Cambridge. She then 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 leads the Data Stewardship project.

Speaker Biographies

Magdalena Skipper

Editor in Chief, *Nature*

Magdalena is Editor in Chief of *Nature* and Chief Editorial Advisor for Nature Research. She has considerable editorial and publishing experience, having started in Nature Publishing Group in 2001. She was Chief Editor of Nature Reviews Genetics, Senior Editor for genetics and genomics at Nature, Executive Editor for the Nature Partner Journals and Editor in Chief of Nature Communications. A geneticist by training, she obtained her PhD from University of Cambridge for the studies on sex determination at the MRC Laboratory of Molecular Biology, Cambridge, UK. She continued her research at the ICRF Laboratories (CRUK today), London, before turning her attention to scientific publishing. She is passionate about mentorship, transparent science and clarity in science communication. She has a keen interest in innovation in science.

John Burn-Murdoch

Senior Data-Visualisation Journalist, *The Financial Times*

John Burn-Murdoch, 29, is a senior data visualisation journalist at the Financial Times, where he uses statistical analysis and graphics to find and tell stories on a range of subjects including politics, demographics and sports. Before joining the FT in 2013, John was a data journalist with *The Guardian's* Datablog, which he joined after studying geography at Durham University and completing a master's degree in interactive journalism at City University, London.

Lightning Talk Abstracts

Overcoming data barriers for regional-scale coastal-impact analysis

Claudia Wolff, Christian-Albrechts-University Kiel

A pre-requisite for analyzing the potential impacts of sea-level rise and for planning associated adaptation strategies is the availability of consistent, multidisciplinary data on current and future conditions of the coast. Developing such datasets is a challenging task due to the lack of primary data or to difficulties in merging data of different formats, resolution, quality and accuracy in a consistent manner. To meet the needs of regional-scale impact analysis we have developed an open access spatial database for the Mediterranean that provides consistent information on 160 physical and socio-economic characteristics. The structure of the database relies on a linear representation of the coast with associated spatial assessment units. This data model enables us to depict the highly dynamic and complex coastal system with a lean data format. We envisage this database as a first step in overcoming barriers for regional coastal studies of future change in the region. We believe that coastal impact analysis can greatly benefit from open, peer-reviewed and well-documented datasets that enable us to replicate and update studies as new data become available. These characteristics increase the validity and understanding of the produced data and therefore help overcome some of the main barriers for data-driven science.

Share for Rare: Promoting Data-Sharing through Japan's Initiative on Rare and Undiagnosed Diseases (IRUD)

Takeya Adachi, Japan Agency for Medical Research and Development (AMED)

Japan has taken measures for specially defined rare diseases, called Nan-Byo (literally "difficult" + "illness") in Japanese since 1972. This governmental support have surely benefited Nan-Byo patients; however, those suffering from medically undiagnosed conditions do not fall into this scheme and thus cannot receive subsidies nor chances for consultations.

To identify such rare – and often undiagnosed – diseases, we must integrate systematic diagnosis by medical experts through phenotypic and genetic data matching, thereby solving the "N−of−1" problem. Initiative on Rare and Undiagnosed Diseases (IRUD), having been launched and coordinated by AMED since 2015, is an ambitious challenge to construct a comprehensive medical network with more than 400 hospitals and establish databases storing patients' clinical/genetic data gathered in a globally compatible data-sharing system, which has strongly contributed to delivering the precise diagnosis.

At its core, IRUD will continue to expand the concept of microattribution; that is, proper credit and evaluation shall be given not only to the representative research institutes but also to local hospitals where primary care doctors, co-medicals and research-collaborators are engaged for the practical data registration. Furthermore, AMED is aiming to advance research achieved by IRUD also towards novel treatment(s) upon our international collaboration including global data-sharing.

Lightning Talk Abstracts

Mapping population distributions, demographics and dynamics

Andrew Tatem, University of Southampton: WorldPop

The rapid growth in computing power, availability of satellite imagery and expansion of geospatial analysis tools over the past decade are providing new opportunities for data integration to improve demographic mapping. WorldPop (www.worldpop.org) develops methods for data integration and disaggregation, drawing on census, survey, satellite and cell phone data among others to produce open and consistent demographic maps. The methods build on advances in geostatistics, machine learning, big data processing and sample survey design to produce spatial demographic estimates with appropriate uncertainty metrics. Openly available estimates of population counts, age and sex structures, births, pregnancies and rates of literacy and poverty per 100x100m grid cells across countries are produced, together with data on population mobility, covering displacement following disasters, to long-term migration flows. All methods and data undergo peer-review and are published openly to ensure rigour, reproducibility and equity in data access. Research methods and outputs are often tailored to the needs of stakeholders, and recent outputs have, for instance, been used by multiple UN agencies for estimating populations affected by disasters, adopted as official statistics in Afghanistan following approval by President Ghani, and used to plan polio elimination efforts in Nigeria.

Gridded birth and pregnancy datasets for Africa, Latin America and the Caribbean

Natalia Tejedor Garavito, University of Southampton

Understanding the fine scale spatial distribution of births and pregnancies is crucial for informing planning decisions related to public health. This is especially important in lower income countries where infectious disease is a major concern for pregnant women and new-borns as highlighted by the recent Zika virus epidemic. Despite this the spatial detail of basic data on the numbers and distribution of births and pregnancies is often of a coarse resolution and difficult to obtain with no co-ordination between countries and organizations to create one consistent set of subnational estimates. To begin to address this issue under the framework of the WorldPop program an open access archive of high resolution gridded birth and pregnancy distribution datasets for all African, Latin American and Caribbean countries has been created. Datasets were produced using the most recent and finest level census and official population estimate data available and are at a resolution of 30 arc seconds (approximately 1 km at the equator). All products are available through WorldPop.

Multi-centre Epilepsy Lesion Detection Project conducting clinical research in an open-science framework

Sophie Adler, University College London

Focal cortical dysplasia (FCD) is a congenital abnormality of cortical development and a leading cause of surgically remediable drug resistant epilepsy. Machine learning offers a powerful framework to develop automated and individualized tools that aid the detection of lesions. However, machine learning continues to improve with increasing numbers of examples necessitating multi-centre collaboration. Here we have created an open infrastructure, with freely accessible pipelines and protocols for parallel post-processing of clinical MRI scans across multiple sites internationally. This approach equips individual sites to replicate our pipeline locally with a standardized preprocessing enabling multi-centre data sharing.

Consent for data archiving in end of life care research: reflections on two qualitative studies

Jane Seymour, University of Sheffield

This talk offers reflections on experiences of gaining consent for data archiving in two projects involving qualitative interviews about end-of-life care. The projects were funded by the ESRC and took place a decade apart. The first study focused on how older people understand technologies used in end-of-life care entailing the development of novel interview techniques. Data from this study included interview transcripts and field notes as well as consent and information sheets. The second study focused on decsion-making about continuous sedation for terminally ill cancer patients with refractory symptoms (i.e. symptoms that cannot be controlled by other means). This study was part of an international collaboration and sought to provide insights into inexplicable variation in practices by country and ethical complexities. Data collection took place in hospices hospitals and primary care. Data included interview transcripts from doctors nurses and bereaved family care givers organised as patient centred case studies. Consent challenges related to different formations and approaches of the NHS research ethics committees involved as well as the differing levels of perceived by various categories of participant whether older member of the general public bereaved relatives or health care professional. These shaped participants' responses to requests for data sharing.

Sharing Electrical Impedance Tomography and neuro-imaging data from stroke patients

James Avery, University College London

The ability to differentiate ischaemic or haemorrhagic stroke without the need for neuroimaging could decrease the time to treatment, substantially improving patient outcomes. New, portable diagnostic tools are necessary to achieve this, such as the Multi Frequency Electrical Impedance Tomography methods developed by the UCL EIT group. These techniques are particularly sensitive to modelling errors, or require strong *a priori* assumptions, and so far, none have translated successfully from simulations or phantoms to human studies. Many EIT researchers are connected to mathematics or engineering departments, with no simple pathway to collect patient data. Connection to a specialised unit such the UCLH Hyper Acute Stroke Unit is an even rarer commodity. Therefore, many talented researchers are unable to test the robustness of their algorithms against anything representing realistic clinical data. To address this problem and aid future development of EIT imaging methods, we have published a rich neuro*scientific data*set (*Sci Data* 5, 180112) collected from 23 patients at the UCLH HASU, including MRI/CT, EEG and MF-EIT data. The accompanying repositories were written to lower the barrier for entry as much as possible. We hope it can serve as a platform for future research into stroke classification.

Lightning Talk Abstracts

Compact Identifier resolution services

Sarala M. Wimalaratne, EMBL-EBI: Identifiers.org

The Identifiers.org resolution system provides consistent access to life science data using Compact Identifiers (CIDs). CIDs consist of an assigned unique prefix and a local provider designated accession number (prefix:accession). The resolving location of CIDs is determined using information that is stored in the Identifiers.org Registry, which contains high quality, manually curated information on over 700 data resources.

The prefix registration process involves assigning a unique prefix to individual life science databases, together with storing a variety of useful metadata, including a description of the data resource, accession identifier pattern and a list of known resolving locations. When a CID is presented to the Identifiers.org resolver, redirection can be accomplished in either a location independent manner (pdb:2gc4) or resource specified (pdbj/pdb:2gc4). In addition, we have formally agreed with N2T resolver, part of the California Digital Library, to share a common prefix registry. This enable users to resolve CIDs using Identifiers.org or N2T resolvers.

Besides resolution, Identifiers.org provides a number of additional services, including the ability to harvest and display (bio)schema.org metadata markup associated with datasets and data records by presenting a CID to the Identifiers.org metadata service.

Live Audit and Feedback for Trials Transparency

Nicholas J. DeVito, University of Oxford

The EBM DataLab at the University of Oxford is built around the principles of multidisciplinary collaboration between software engineers, clinicians and academics. We develop and openly share tools and data with the public, as well as publish papers, in the areas of health informatics and research integrity. Our TrialsTracker initiative has seen the launch of two live trackers: the FDAAA TrialsTracker (fdaaa.trialstracker.net) and the EU TrialsTracker (eu.trialstracker.net). Together these websites provide live statistics on which sponsors are complying with their legal requirement to report results on two of the largest trial registries in the world. To date this initiative has also led to a peer-reviewed publication in the BMJ with additional papers forthcoming. For each project we have openly shared our code and methods, provided the ability to access and download the underlying data, and maintained public-facing, frequently updated websites that allow stakeholders to monitor compliance in real-time. This presentation will introduce attendees to these trackers, their features, impact to date, and briefly explain how they were created with public data in an open and transparent manner.

STRENDA DB – Monitoring the completeness of information in data reports

Carsten Kettner, Beilstein-Institut

Scientific research is in part a creative act producing new results or theories through observation and experimentation and part verification through reproduction and comparison. This requires the experimental data to be reported completely including all necessary meta-data. Discussions with scientists have shown up many deficiencies in the way that the data are currently reported, resulting often in incomplete and even unusable data sets that are not suitable for subsequent research and knowledge generation. Since more than a decade ago, the Beilstein-Institut supports the STRENDA Commission (Standards for Reporting Enzymology Data) which has developed community based recommendations to authors for the reporting of enzymology data – the STRENDA Guidelines. Today, more than 55 biochemistry journal recommend their authors to refer to these guidelines when reporting enzyme kinetics data. In parallel, the Commission has developed STRENDA DB, a robust web-based validation and storage system for functional enzyme data that incorporates the STRENDA Guidelines allowing authors to easily check manuscript data for compliance with the Guidelines prior to or during the publication process. The data is stored in an open access database which will provide a knowledgebase for researchers. This talk highlights the benefits of STRENDA DB for journals, authors, reviewers, and readers.

Meaningful and reproducible statistics: Does my data hold what it promises?

Andrej-Nikolai Spiess, University Hospital Hamburg-Eppendorf

There are no doubts that science is facing a reproducibility crisis. A survey with 1500 scientists revealed that these were unable to replicate 40-80% of their own and others' results, while a large-scale data replication study in the psychological sciences demonstrated that 30% of results were not reproducible or even led to contrary conclusions. In these studies, the authors merely replicated the statistical analysis steps based on the original raw data, demonstrating that by just repeating simple statistical steps, one can obtain completely different analytical (and hence interpretational) outcomes. We are strongly convinced that the lack of understanding statistical prerequisites frequently results in highly unstable and irreproducible results.

We present an online tool that checks one of the most widely applied statistical methods, linear regression, for "reversers". These are single data points whose removal results in the reversal of the significance statement ("p = 0.05 dichotomization"). We further demonstrate that many linear regressions in high impact journals (Science, Nature, PNAS) are highly sensitive to the presence of single data points. We discuss similar problems with other statistical methods and advocate the development of an online platform that validates statistical robustness prior to publication.

Lightning Talk Abstracts

Materials Cloud, An Open Science Portal for FAIR Data Sharing

Aliaksandr Yakutovich, École Polytechnique Fédérale de Lausanne

Materials Cloud is a web platform enabling seamless sharing of resources for computational research, with current focus on Materials Science: educational materials, interactive tools, virtual hardware and simulated data. Through its rich web interface, users can run complex calculations and browse the provenance of the generated data (i.e. the trail leading from the inputs to the results of a computational work flow), while maintaining fine-grained access to the underlying simulations.

Materials Cloud is powered by AiiDA, a python framework for managing simulations. AiiDA records automatically the data provenance and stores it in a local database. Sharing an AiiDA database on Materials Cloud thus provides access not only to the final results of the calculations, but to every step along the way. The data is stored in a persistent and citable way (DOIs), implementing the data management practices recommended by major funding bodies. Peers can browse the database interactively, download individual files or the whole database, and start their research right from where the original author left off. Further features include lightweight web apps integrating with AiiDA to run and manage simulations in the cloud.

Counting reuse to Make Data Count

Helena Cousijn, DataCite

Data sharing is becoming a normal part of the research process, but there is limited insight into how shared datasets are being reused. While researchers are rewarded when other researchers cite their articles and thereby show that the work in being built on, for research data there are very few recognized metrics that lead to credit for researchers. The Make Data Count project is working on addressing this. In this project, standards have been developed to count views and downloads of datasets to provide insight into interactions with datasets. In addition, information about data citations is collected from publishers and data repositories through collaboration with the Scholix initiative and used to assess data reuse. These different data metrics are made openly available so that data repositories, research institutions, funders, and publishers can display them. This way, researchers can see and show the impact of their research data and be rewarded for the work they did to make data reusable.

Nine good things about open science (and one bad thing)

Alasdair Rae, University of Sheffield

Open access, open data, open source, open science: all terms which should engender warm, fuzzy feelings and positivity. Yet this is not always the case. In fact, lots of people really aren't on board yet with 'open' as a concept. For those of us who might consider ourselves advocates of open science and better data, this might be disappointing. But I see it as an opportunity. An opportunity to ask ourselves what we are doing, and why, but also to build up examples and case studies which help make the case for an open approach. So, this talk is basically an advert for an open approach in research and data sharing. I'll also mention at least one bad thing that being open can lead to. I'll include examples from my work on US commuting and UK land cover, talk about tools and data sharing and show you some nice maps.

Programme Committee

To ensure the successful development of the 2018 conference programme and provide expertise in postgraduate education, research funding, and publishing, the following people sat on the conference programme committee:

Iain Hrynaszkiewicz – Programme Chair

Head of Data Publishing, Springer Nature

Iain is Head of Data Publishing at Springer Nature, where he develops new areas of research data publishing and data policy. This includes efforts to make research data more visible and reusable by developing links with data repositories, and data journals such as *Scientific Data*. He has led various initiatives and published numerous articles related to data sharing, open access, open data and reproducible research.

David Carr

Programme Manager – Open Research, The Wellcome Trust

David is Programme Manager for Open Research at the Wellcome Trust, where he is responsible for developing and taking forward a range of activities to maximise the availability and re-use of research outputs – including publications, datasets, software and materials. Previously, David worked as a Policy Adviser at Wellcome – leading on work to develop and communicate policy in several areas – including data sharing, open access publishing, biosecurity and genomics. In 2001, David worked on secondment at the World Health Organisation in Geneva, where he assisted in the preparation of the Advisory Committee on Health Research (ACHR) report on Genomics and World Health. Prior to joining the Trust in 1999, David worked as a project researcher at a scientific consultancy firm in Cambridge. He has undergraduate and master's degrees in genetics from the University of Cambridge.

Louise Corti

Associate Director, UK Data Archive

Louise leads the UK Data Service's Collections Development and Data Publishing teams. Louise actively researches and publishes on key aspects of data management, sharing and reuse of social science data. She was instrumental in helping operationalise the ESRC's Research Data Policy from 1995 and extending this to fully accommodate qualitative data. She teaches regularly, acts as a consultant for qualitative data archives setting up round the world, and is currently helping UKDS to scale up for managing new and novel forms of data.

Programme Committee

Dr Kirstie Whitaker

Research Fellow, The Alan Turing Institute

Kirstie completed her PhD in Neuroscience at the University of California, Berkeley in 2012 and holds a BSc in Physics from the University of Bristol and an MSc in Medical Physics from the University of British Columbia. She was a postdoctoral researcher in the Department of Psychiatry at the University of Cambridge from 2012 to 2017. Kirstie uses magnetic resonance images to understand the changes in the brain's structure and function that underlie the emergence of schizophrenia and depression. She is particularly passionate about ensuring that work is reproducible and can be replicated in independent data sets. She is a Fulbright scholarship alumna and 2016/17 Mozilla Fellow for Science. Kirstie was named, with her collaborator Petra Vertes, as a 2016 Global Thinker by Foreign Policy magazine.

Dr Emma Ganley

Joint Chief Editor, PLOS Biology

Emma Ganley is Joint Chief Editor of PLOS Biology and she is lead of the PLOS data program. Initially trained as a scientist with a PhD in Molecular Biology from the MRC-LMB in Cambridge followed by postdoctoral research at UC Berkeley, Emma moved into science publishing joining PLOS Biology in 2005. Emma rejoined PLOS in 2010 after some years in New York where she was Executive Editor of the Journal of Cell Biology. Emma has long been involved in efforts to ensure the highest level of scientific integrity via data presentation and making data available alongside publications; she helped develop the JCB DataViewer while at JCB, worked as a Project Manager for the Open Microscopy Environment. She now oversees many projects at PLOS related to Open Science, Open Data, publishing ethics and mechanisms to improve research assessment. Emma is currently Co-Chair of the Advisory Board for biosharing.org, and active participant in the Force11 EG3 Data Citation Implementation Pilot (DCIP) Publisher Early Adopters working group.

Patricia Cruse

Executive Director, DataCite

Trisha leads DataCite with a strong commitment and passion for data sharing. As Executive Director, her role is to advance DataCite's mission, build strategic partnerships and work with a diverse community of stakeholders. Prior to joining DataCite, Trisha was the Director of the University of California Curation Center (UC3) at the California Digital Library (CDL). In addition, Trisha also is a co-PI on the DataONE initiative and co-leads sustainability and governance activities. Trisha holds an MLIS from the University of California, Berkeley.

Dr Fiona Reddington

**Head of Population, Prevention & Behavioural Research Funding,
Cancer Research UK**

Fiona obtained her BSc (Pharmacology) at University College Dublin and her PhD (Neurophysiology) from Kings College London (UMDS). From there, Fiona joined the NHS as a project manager and went on to manage a Cancer Centre at University College London. Management roles at a national cancer network and the National Cancer Research Institute Informatics Initiative followed where Fiona was part of the team to win the inaugural Times Higher Research project of the Year award. Fiona joined Cancer Research UK in 2008. She has responsibility for the Cancer Research UK research portfolio in the areas of population research, prevention and early diagnosis. She represents the charity on matters relating to data sharing and the management boards of a number of external initiatives such as the UK Prevention Research Partnership and UKCRC Public Health Centres of Excellence.

Dr Marta Teperek

Research Data Stewardship Lead, TU Delft

Marta completed a PhD in molecular biology at the University of Cambridge. She then 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 leads the Data Stewardship project.

Springer Nature is committed to supporting researchers in sharing research data and in receiving the credit you deserve. We have developed the following products and services to make sharing your research data faster, easier and more impactful.

Research data helpdesk

Our team of research data editors gives free advice on how and where to share your research data, according to your specific research community. We always respond to your enquiry within two working days.

go.nature.com/helpdesk

Research Data Support service

Knowing how to make your data useful, findable and get the credit you deserve takes a lot of time and effort. Our expert research data editors will organise, curate and deposit your files in a useful and accessible way. We make your data simpler to understand, easier to find by relevant researchers in your field of study and more shareable for just €300/£265/US$340.

go.nature.com/RDS
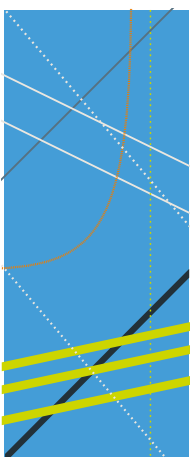
Research data publishing

Nature Research's *Scientific Data* and *BMC Research Notes* are two of our open access, online-only journals that publish article summaries of scientifically valuable datasets.

go.nature.com/publishing

Research data community

Our growing community of advocates for the sharing of research data is a great place to interact with other researchers, read the latest information on research data and help us make research data sharing the new norm.

go.nature.com/community

Springer Nature is committed to opening up paths to discovery for today's researchers in order to accelerate their ability to solve societies' grand challenges. The illustrations we use on our covers celebrate some of the great minds who have shaped our knowledge through history.

John Maynard Keynes (1883–1946)
John Maynard Keynes was a British economist who revolutionised the theory and practice of macroeconomics, reformed economics and had a profound influence on economic policy. His work built upon the causes of the business cycle, and challenged established neoclassical economic ideas. This illust ration represents the Keynesian model that he created as a result of the Great Depression. It showed that in a monetary economy it is possible to have periods of high unemployment unless governments use active monetary and fiscal policy to stimulate aggregate demand.

Springer Nature advances discovery by publishing robust and insightful research, supporting the development of new areas of knowledge and making ideas and information accessible around the world.

For more information, contact us:
researchdata@springernature.com

Follow twitter.com/ScientificData

A64373