Availability of data, material and methods

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. A condition of publication in a Nature journal is that **authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications**. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must **also** be disclosed in the submitted manuscript.

After publication, readers who encounter refusal by the authors to comply with these policies should contact the chief editor of the journal. In cases where editors are unable to resolve a complaint, the journal may refer the matter to the authors' funding institution and/or publish a formal statement of correction, attached online to the publication, stating that readers have been unable to obtain necessary materials to replicate the findings.

See sections below for details on:

Reporting requirements

Reporting requirements for life sciences research

Since May 2013, Nature journals require authors of life sciences research papers that are sent for external review to include in their manuscripts relevant details about several elements of experimental and analytical design. This initiative aims to improve the transparency of reporting and the reproducibility of published results. It focuses on **elements of methodological information** that are frequently poorly reported. During peer review, authors will be asked to confirm that these elements are included in the manuscript by filling out a **checklist** that will be made available to the editors and reviewers.

Guidance and resources related to the use and reporting of statistics are available **here**.

Reporting and materials availability requirements for Earth sciences research

Details of geological samples and palaeontological specimens should include clear provenance information to ensure full transparency of the research methods. Samples should always be collected in a responsible manner and in accordance with relevant permits and local laws. Any submission detailing new material from protected sites should include information regarding the requisite permission obtained. Palaeontological and type specimens should be deposited in a recognised museum or collection to permit free access by other researchers in perpetuity.

Further reading

Top of page

Availability of data

Supporting data must be made available to editors and peer-reviewers at the time of submission for the purposes of evaluating the manuscript. The preferred way to share large data sets is via public repositories. (Details about how to share some specific data sets can be found in the sections below.)

Some of these repositories offer authors the option to host data associated with a manuscript confidentially, and provide anonymous access to peer-reviewers before public release. These repositories then coordinate public release of the data with the journal's publication date. This option should be used when possible but it remains the author's responsibility to communicate with the repository to ensure that public release is made on time for online publication of the paper. **For information about suitable public repositories, see sections that follow.**

Unstructured repositories like **figshare** and **Dryad** are suitable alternatives if no structured public repositories exist. As a less desirable alternative, data sets can be made available as Supplementary Information files, which will be freely accessible on nature.com upon publication. In rare cases when data files cannot be deposited in an accessible repository for technical reasons, authors must make the data available to editors and peer reviewers if requested. After publication, authors must likewise arrange to make the data available to any reader directly upon reasonable request.

Nature journals encourage authors to consider the publication of a Data Descriptor in ***Scientific Data*** to increase transparency and enhance the re-use value of data sets used in their papers. Data Descriptors are designed to be complementary to a primary paper and

can be published prior to, simultaneously, or after publication of the primary paper. Nature journals will not consider prior Data Descriptor publications to compromise the novelty of new manuscript submissions, as long as those manuscripts go substantially beyond a descriptive analysis of the data and report important new scientific findings appropriate for the journal. (This policy does not necessarily extend to journal articles whose primary purpose is to describe a new data set or resource.)

Further reading

Mandates for specific datasets

For the following types of data set, submission to a community-endorsed, public repository is mandatory. Accession numbers must be provided in the paper. Examples of appropriate public repositories are listed below.

Mandatory deposition	Suitable repositories
Protein sequences	**Uniprot**
DNA and RNA sequences	**Genbank**
	DNA DataBank of Japan (DDBJ)
	EMBL Nucleotide Sequence Database (ENA)
DNA and RNA sequencing data	**NCBI Trace Archive**
	NCBI Sequence Read Archive (SRA)
Genetic polymorphisms	**dbSNP**
	dbVar
	European Variation Archive (EVA)
Linked genotype and phenotype data	**dbGAP**
	The European Genome-phenome Archive (EGA)
Macromolecular structure	**Worldwide Protein Data Bank (wwPDB)**
	Biological Magnetic Resonance Data Bank (BMRB)
	Electron Microscopy Data Bank (EMDB)
Microarray data (must be MIAME compliant)	**Gene Expression Omnibus (GEO)**
	ArrayExpress
Crystallographic data for small molecules	**Cambridge Structural Database**

Special considerations

DNA and protein sequences: When publishing reference genomes, the assembly must be made available in addition to the sequence reads. Sequence must be deposited even for short stretches of novel sequence information such as epitopes, functional domains, genetic markers, or haplotypes. Short novel sequences must include surrounding sequence information to provide context. The sequences of all small RNA probes central to the conclusions of the paper must be provided.

Linked phenotype and genotype data for human subjects: should be submitted to a public repository with appropriate access control (see above). Any restrictions on data access for sensitive data (for example electronic medical records, forensic data, and personal data from vulnerable populations) require an explanation of the nature of and reasons for the restrictions, and details of the conditions under which the data can be accessed or reused. (See the related *Nature Genetics* **Editorial** discussing privacy issues.)

Macromolecular structures: Atomic coordinates and related experimental data (structure factor amplitudes/intensities for crystal structures, or restraints for NMR structures) must be provided upon request. Electron microscopy-derived density maps and coordinate data must be deposited in EMDB. Accessibility in repositories must be designated "for immediate release on publication."

Crystallographic data for small molecules: Manuscript reporting new three-dimensional structures of small molecules from crystallographic analysis should include a .cif file and a structural figure with probability ellipsoids for publication as Supplementary Information. The structure factors for each structure should also be submitted. Both the structure factors and the structural output must have been checked using the IUCR CheckCIF routine, and a PDF copy of the output must be included at submission, together with a justification for any alerts reported.

Recommendations for other datasets

In addition to these mandates, the preferred way to share any data sets is via public repositories. *Scientific Data*, a sister publication to Nature journals, maintains a **list of approved and recommended data repositories** organized by discipline. Please consult this list to identify an appropriate repository for your data sets.

When repositories do no exist for a particular data type, authors can deposit and share data via **figshare** or **Dryad**, two general-purpose scientific data repositories.

Further reading

Top of page

Availability of materials

A condition of publication in a Nature journal is that authors are required to make unique materials promptly available to others without undue qualifications.

It is acceptable to request reasonable payment to cover costs of distribution and reagents may be made available via commercial or non-commercial third party providers. If materials are to be distributed by a for-profit company, this must be stated in the paper.

Authors reporting new chemical compounds integral to the conclusions of the paper must provide the chemical structure, synthesis and characterization of the compounds with sufficient experimental details to allow other researchers to reproduce the synthesis and characterization.

Mandates and recommendations for specific materials
For biological materials such as mutant strains and cell lines the Nature journals require authors to use established public repositories when one exists (for example, **Jackson Laboratory**, the European Mouse Mutant Archive (**EMMA**), the **European Conditional Mouse Mutagenesis Program** (EUCOMM), the **Knockout Mouse Project** (KOMP), **Addgene**, **RIKEN Bioresource Centre**, the **Mutant Mouse Regional Resource Centers**, **American Type Culture Collection**, and provide accession numbers in the manuscript.

Cell lines: We strongly encourage deposition of new cell lines in repositories that will distribute them with certificates of authentication. Alternatively, we recommend that authors establish a profile of their new cell lines to allow future authentication. The distribution of human cell lines used in research should not be hindered by restrictions from donors. Researchers developing cell lines must investigate and disclose any restrictions associated with the tissue they are using (see this *Nature* **Editorial** for further explanation.) Cell line misidentification and cross-contamination is a common problem with serious consequences. Authors are asked to report on the source and authentication of their cell lines (relevant resources are listed under **Further Reading**).

Top of page

Availability of computer code

Authors must make available upon request, to editors and reviewers, any previously unreported custom computer code used to generate results that are reported in the paper and central to its main claims. Any practical issues preventing code sharing will be evaluated by the editors who reserve the right to decline the paper if important code is unavailable.

Upon publication, Nature Journals consider it best practice to release custom computer code in a way that allows readers to repeat the published results.

For all studies using custom code that is deemed central to the conclusions, a statement must be included in the Methods section, under the heading "Code availability", indicating whether and how the code can be accessed, including any restrictions to access.

Further reading

Top of page

Experimental protocols

Authors are encouraged to share their specific step-by-step experimental protocols on **Protocol Exchange**, a free-to-use and open resource maintained by Nature Publishing Group. Links to these protocols can appear in the online methods section of the published article.

Top of page

Clinical Trials

Authors reporting phase II and phase III randomized controlled trials must refer to the **CONSORT Statement** for recommendations to facilitate the complete and transparent reporting of trial findings. Reports that do not conform to the CONSORT guidelines may need to be revised before formal review.

Authors reporting prognostic studies with tumor markers are encouraged to follow the **REMARK guidelines** for complete and transparent reporting.

Prospective clinical trials must be registered before the start of patient enrollment in **www.clinicaltrials.gov** or a similar public repository that matches the criteria established by ICMJE . The trial registration number must be reported in the paper. (Trials in which the primary goal is to determine pharmacokinetics are exempt.)

For describing human biospecimens, we recommend referring to the BRISQ reporting guidelines (Biospecimen Reporting for Improved Study Quality) and ensuring at least Tier 1 characteristics are provided (**doi: 10.1002/cncy.20147**).

Top of page

Further Reading

Reporting requirements and reproducibility

Nature Special: **Challenges in irreproducible research** (continually updated) - collection of articles from Nature journals addressing various aspects of irreproducibility.

All Nature journals publishing life sciences introduce measures to improve reporting standards. *Nature*. **Announcement: Reducing our irreproducibility**, April 2013.

New rules for presentation of statistics. *Nature Cell Biology*.**Reproducible methods** and **How robust are your data?** June 2009.

Nature Methods ushers in an Online Methods section, fully integrated with the paper, for all original research articles. *Nature Methods*. **Methods section remake**, May 2009.

From now on, Nature authors will be able to include more experimental details in their papers. *Nature*. **Methods in full**, February 2007.

Availability of data

In our continued drive for reproducibility, *Nature* and the Nature research journals are strengthening our editorial links with the journal*Scientific Data* and enhancing our data-availability practices. *Nature*. **Data-access practices strengthened**, November 2014.

Datasets can now be published, shared — and cited — in *Scientific Data*. *Nature Physics*. **It's good to share**, July 2014.

Data must be accessible to support the conclusions of scientific publications and for the research to have impact. *Nature Genetics*. **It's not about the data**, February 2012.

A recent report highlights the urgent issues regarding the preservation of large datasets. *Nature Neuroscience*. **Ensuring data integrity**, October 2009.

Reference datasets should be accessible independently of scientific papers in a citable form, allowing attribution. *Nature Cell Biology*. **Sharing data**, November 2009.

Datasets released to public databases in advance of (or with) research publications should be given digital object identifiers. *Nature Genetics*. **Data producers deserve citation credit**, October 2009.

The scientific community needs to develop better incentives to encourage compliance and reward those who share. *Nature Neuroscience*. **Got data?** August 2007.

Procedures for microattribution need to be established by journals and databases. *Nature Genetics*. **Compete, collaborate, compel**, August 2007.

Online publishing should have made 'data not shown' largely a thing of the past. *Nature Cell Biology*. **Nothing to hide (data not shown)**, June 2006.

Discipline-specific data sets

Data from genome-wide association studies should be reported and deposited, even if the data does not reach genome-wide levels of significance. *Nature Genetics*. **Asking for more**, June 2012.

Human exome sequencing data should ideally be archived in appropriate repositories before submission and authors must explain their data management plan before peer review. *Nature Genetics*. **Capture and release**, September 2011.

Reference datasets should be accessible independently of scientific papers in a citable form, allowing attribution. *Nature Cell Biology*. **Sharing data**, November 2009.

Proposes a universal tagging system that connects databases with authors. *Nature Biotechnology*. **Credit where credit is overdue**, July 2009.

Genetically modified mouse strains must be made available. *Nature*. **The sharing principle**, June 2009.

Deposition of proteomics data. *Nature Methods*. **Thou shalt share your data**, March 2008.

Describing the Nature journals' Creative Commons license for genome sequences. *Nature*. **Shared genomes**, December 2007.

Scientists coin new terms, or neologisms, at a tremendous pace, but name choice can have unforeseen results.*Nature Structural & Molecular Biology*. **Name that gene!** August 2007.

Raw proteomics and molecular interaction data should be deposited in repositories at submission **Democratizing proteomics data** (March 2007), and researchers should embrace minimum information reporting guidelines**Time for leadership** (August 2007). *Nature Biotechnology*.

The tools for genome-wide association studies are now available. Here we present the journal's current criteria for manuscripts in this area of research. *Nature Genetics*. **Framework for a fully powered risk engine**, November 2005.

Biological research must provide the data necessary for replication. *Nature Medicine*. **Structural Integrity**, February 2005.

How to discuss ancestry and ethnicity. *Nature Genetics*. **The unexamined 'Caucasian'**, June 2004.

Clarifying the Nature journals' policy on data deposition for chemical structures. *Nature*. **Crystal Clear**, June 2005.

Required controls for studies involving RNAi. *Nature Cell Biology*. **Whither RNAi?** June 2003.

Controls for studies involving microarrays. *Nature Immunology*. **Microarray policy**, February 2003.

Data requirements for studies involving microarrays. *Nature Cell Biology*. **Microarray data standards**, November 2002.

The microarray community has issued guidelines that will make their data much more useful and accessible. *Nature*. **Microarray standards at last**, September 2002.

Any paper containing new structural data will not be accepted without an accession number from the Brookhaven Protein Data Bank. *Nature*. **New policy for structural data**, July 1998.

Statements by presidents of countries and societies highlight the concern that human genome data be publicly accessible, and quickly. *Nature*. **Rules of genome access**, March 1990.

Availability of materials

There must be no restrictions on the redistribution of patient-derived cell lines or other tissue. *Nature*. **Common consent**, 20 August 2009.

Nature Chemical Biology is committed to enhancing interdisciplinary communication and features online content to increase the accessibility of chemical information for our readers. *Nature Chemical Biology*. **A new look for chemical information**, June 2007.

Note to biologists: submissions to Nature should contain complete descriptions of materials and reagents used. *Nature*. **Illuminating the black box**, 16 July 2006.

Pragmatically adapting our sharing-of-materials policy. *Nature Cell Biology*. **Sharing science**, May 2006.

There is a great need for community standards for sharing data, materials and information between chemists and biologists. *Nature Chemical Biology*. **Molecular cross-fertilization**, February 2006.

On sharing materials to foster reproducible research. *Nature Cell Biology*. **Policy update**, March 2005.

On sharing materials to foster reproducible research. *Nature Genetics*. **'Good citizenship' or good business?**, October 2004.

Resources on cell line identity

To help curb the inadvertent use of cross-contaminated or otherwise misidentified cell lines, authors are asked to check their reagents against the list of known misidentified cell lines maintained by the **International Cell Line Authentication Committee** (ICLAC) and also accessible through the NCBI **BioSample database**. If using a cell line that is on this list, authors should provide a scientific justification and state the identity issue in the Methods section. Editors reserve the right to demand that the data be removed from the paper if the justification is deemed unsatisfactory. In addition, authors must identify the source of cell lines (with catalog number if obtained from vendor or cell bank) and report whether the cell lines have been authenticated. They should include the method used, the results and when authentication testing was last performed for that cell line. Authentication test results must be provided upon request. Mycoplasma contamination testing status must also be reported. As of May 2015, these requirements are particularly emphasized for cancer research where the issue of cell line misidentification has been well documented, but authors in all disciplines are strongly encouraged to comply with these reporting criteria. It is good practice to obtain cell lines from reputable repositories, to routinely authenticate cell line stocks and test them for mycoplasma contamination. Resources on cell line authentication follow.

Authentication of Human Cell Lines: Standardization of STR profiling (American National Standards Institute)

Yu M. et al. **A resource for cell line authentication, annotation, and quality control**. *Nature*, 520, 307–311, doi:10.1038/nature14397 (April 2015).

Capes-Davis, A. et al. **Match criteria for human cell line authentication: Where do we draw the line?** *Int J of Cancer*, doi:10.1002/ijc.27931 (2012).

STR profiles and misidentified cell lines databases:

Database of Cross-contaminated or Misidentified Cell Lines (International Cell Line Authentication Committee)

BioSample Database (NCBI/NIH)

COSMIC Cell Line Project

ATCC STR Profile database

DSMZ STR Profile Verification database

Background and guidance:

Announcement: Time to tackle cells' mistaken identity. *Nature* 520, 264, doi:10.1038/520264a (16 April 2015).

Geraghty R. J. et al. **Guidelines for the use of cell lines in biomedical research**. *British J of Cancer* 111, 1021–1046, doi:10.1038/bjc.2014.166 (2014).

Lorsch J. R., Collins F. S., Lippincot-Schwartz J. **Fixing problems with cell lines**.*Science* 346, 1452-1453, doi:10.1126/science.1259110 (2014).

Dirks, W.G. and Drexler, H.G. **STR DNA typing of human cell lines: detection of intra- and interspecies cross-contamination**. *Methods Mol Biol* 946, 27-38, doi: 10.1007/978-1-62703-128-8_3 (2013).

Masters, J. R. Cell-line authentication: **End the scandal of false cell lines**. *Nature* 492, 186, doi:10.1038/492186a (2012).

Cell line misidentification: the beginning of the end. *Nat Rev Cancer* 10, 441-448, doi:10.1038/nrc2852 (2010).

Identity crisis. *Nature* 457, 935-936, doi:10.1038/457935b (2009).

The websites of the **International Cell Line Authentication Committee** and **ATCC** provide useful additional information.

Availability of computer code

Sharing data is key for efficient scientific progress. More open code would be beneficial too. *Nature Geoscience*. **Towards transparency**, November 2014.

Papers in Nature journals should make computer code accessible where possible. *Nature*. **Code share**, October 2014.

The usefulness of computational methods can be improved by releasing code and designing software that supports reproducible research.*Nature Methods*. **Software with impact**, March 2014. (See also this **comment**)

Review, replication, reuse and recognition are all incentives to provide code. *Nature Genetics*. **Credit for code**, January 2014.

Improving the integration of computational analysis into biology will require better documentation, validation and accessibility of software associated with papers. *Nature Biotech*. **In need of an upgrade**, October 2013.

Software that is custom-developed as part of novel methods must be made available to readers upon publication. *Nature Methods*. **Social software**, March 2007.

Top of page

COPE