

Publish your Data and Model Code: Research Output is More Than "Just" a Research Paper

Martin Rasmusen^{a,1}

^a*Copernicus Publications (Copernicus GmbH), Bahnhofsallee 1e, 38081 Göttingen*

Abstract. Research output is not only research articles. To provide outlets for publishing other outputs than articles, Copernicus Publications, an innovative open access publisher based in Göttingen, Germany, currently publishes the journals *Earth System Science Data* and *Geoscientific Model Development*. The first journal is dedicated to the peer-reviewed publication of articles on original research data sets in the Earth System Sciences. The second journal is dedicated to publish the description, development and evaluation of numerical models of the Earth System and its components. Both journals apply an innovative interactive open access peer-review with public referee reports, public comments from the community prior to editor's decision, and public author's responses. The motivation is to make the whole research output from data, to models, to the scientific findings and novel interpretations freely accessible, to foster scientific discussion, to increase transparency in scientific quality assurance, and to give credit to all involved contributors.

Keywords. Data publication, model code publication, open access, public peer-review

Introduction

In the Earth System Sciences, as well as in many other disciplines, the final interpretation of new scientific findings is the result of a long process of data collection, data interpretation, model calibrations, model runs, interpretation of these results, and conclusions regarding novel aspects. And it is teamwork of many people contributing to these results, not only scientists, but also engineers, data specialists, and many other groups of learned contributors.

When open access publishing started, the idea has been raised quickly to expand this principle to many other scientific sources than "just" to the final revised, peer-reviewed article. For decades, readers of scientific articles had to settle for graphs resulting from data interpretation or model runs without knowing much more about the data provenance and structure, without getting access to this data, without a broader understanding of the used models, and without a deeper insight into the model code. Neither reviewers nor readers could have ever reproduced the work of the author of a scientific manuscript.

¹ Corresponding Author.

Fortunately, the open access principle became a politically widely accepted strategy and liberal copyright and license agreements like Creative Commons' CC-BY license fundamentally reinvented the idea of access to scientific work and options for the reuse of, in most cases, research outputs financed through taxpayers money.

In 2008, two groups of scientists raised, independently, the ideas of a data publication journal on the one hand and a model development journal on the other hand. Copernicus Publications started these two titles applying the interactive open access publishing approach with public peer-review and interactive public discussion, established in 2001. Public referee reports, public comments from the community prior to editor's decision, and public author's responses are published alongside the discussion paper, an access-reviewed version of the author's manuscript. The motivation was to make the whole research output from data, to models, to the scientific findings and novel interpretations freely accessible, to foster scientific discussion, to increase transparency in scientific quality assurance, and to give credit to all involved contributors.

The following sections describe the journals Earth System Science Data (ESSD) and Geoscientific Model Development (GMD) in more detail, and explain the concept of interactive open access publishing.

1. Earth System Science Data (ESSD)

1.1. Aims, Scope, and Motivation

Earth System Science Data (ESSD) is an international, interdisciplinary journal for the publication of articles on original research data(sets), furthering the reuse of high (reference) quality data of benefit to Earth System Sciences. The editors encourage submissions on original data or data collections which are of sufficient quality and potential impact to contribute to these aims. The journal maintains sections for regular length articles, brief communications (e.g., on additions to datasets) and commentary, as well as review articles and "Special Issues".

Articles in the data section may pertain to the planning, instrumentation and execution of experiments or collection of data. Any interpretation of data is outside the scope of regular articles. Articles on methods describe nontrivial statistical and other methods employed, e.g. to filter, normalize or convert raw data to primary, published data, as well as nontrivial instrumentation or operational methods. Any comparison to other methods is out of scope of regular articles. Review articles may compare methods or relative merits of datasets, the fitness of individual methods or datasets for specific purposes or how combinations might be used as more complex methods or reference data collections.

This journal aims to establish a new subject of publication: to publish data according to the conventional fashion of publishing articles, applying the established principles of quality assessment through peer-review to datasets. The goals are to make datasets a reliable resource to build upon and to reward the authors by establishing priority and recognition through the impact of their articles.

The peer-review secures that the data sets are at least plausible and contain no detectable problems, that they are of sufficiently high quality and their limitations are clearly stated, that they are open accessible (toll free), well annotated by standard

metadata (e.g., ISO 19115) and available from a certified data center/repository, and that they are customary with regard to their format(s) and/or access protocol, however not proprietary ones (e.g., Open Geospatial Consortium standards), expected to be useable for the foreseeable future.

The articles in this journal should enable the reviewer and the reader to review and use the data, respectively, with the least amount of effort. To this end, all necessary information should be presented through the article text and references in a concise manner and each article should publish as much data as possible. The aim is to minimize the overall workload of reviewers, e.g., by reviewing one instead of many articles, and to maximize the impact of each article. [1]

The initiators of ESSD were David Carlson, director of the programme office of the International Polar Year (IPY) in 2007-2008, and Hans Pfeiffenberger, head of IT infrastructure at the Alfred Wegener Institute for Polar and Marine Research (AWI) in Bremerhaven, Germany.

1.2. Manuscript Submission

The precondition to submit a manuscript for publication in Earth System Science Data is that the data sets referenced in the manuscript are submitted to a long-term repository. Such a repository has to fulfill the following basic criteria under all circumstances [1]:

- Persistent Identifier: The data sets have to have a digital object identifier.
- Open Access: The data sets have to be available free of charge and without any barriers except a usual registration to get a login free-of-charge.
- Liberal Copyright: Anyone must be free to copy, distribute, transmit and adapt the data sets as long as he/she is giving credit to the original authors (equivalent to the Creative Commons Attribution License).
- Long-term Availability: The repository has to meet the highest standards to guarantee a long-term availability of the data sets and a permanent access.

1.3. Review Criteria

For future reuse and reinterpretation it is mandatory for the user to be assured about research data quality. It is the aim of ESSD to provide the quality assessment for datasets which already reside in permanent repositories. Is the article itself appropriate to support the publication of a dataset? Is the dataset significant – unique, useful and complete? Is the dataset publication, as submitted, of high quality? Reviewers are asked to decide how well the respective datasets presented by an article and the article itself meet the criteria significance, data quality, and presentation quality. [1]

1.4. ESSD Facts & Figures

By the end of March 2014, ESSD had 127 manuscripts submitted from which 99 have been published in the discussion forum of ESSD and 85 in ESSD as final revised journal articles. The final articles have an average length of 12 pages (median) and the

review takes on average (median) 29 days from submission to publication of the discussion paper, and 33 days from revised submission after public discussion to publication of the final revised and fully peer-reviewed paper. In the discussion forum, 433 comments have been posted, 207 of which are referee comments, 194 are author comments, eight comments are published by the journal editors, and 24 comments by members of the scientific community prior to the final acceptance of the manuscripts [2]. ESSD is indexed by Scopus.

2. Geoscientific Model Development

2.1. Aims, Scope, and Motivation

Geoscientific Model Development (GMD) is an international scientific journal dedicated to the publication and public discussion of the description, development and evaluation of numerical models of the Earth System and its components. Manuscript types considered for peer-reviewed publication are [3]:

- Geoscientific model descriptions, from box models to GCMs;
- Development and Technical papers, describing development such as new parameterisations or technical aspects of running models such as the reproducibility of results;
- Papers describing new standard experiments for assessing model performance, or novel ways of comparing model results with observational data;
- Model intercomparison descriptions, including experimental details and project protocols.

GMD is owned by the European Geosciences Union (EGU, <http://www.egu.eu>) and started in 2008. The main drivers and Executive Editors have been in alphabetical order James Annan and Julia Hargreaves, both from the JAMSTEC Research Institute for Global Change in Yokohama, Japan; Dan Lunt, University of Bristol, UK; Robert Marsh, University of Southampton, UK; Andy Ridgwell, University of Bristol, UK; Ian Rutt, Swansea University, UK; and Rolf Sander, Max Planck Institute for Chemistry in Mainz, Germany.

Since the scale and complexity of computer modelling tools increased, it was no longer practicable to describe models in papers. Furthermore, the normal journal peer-review focusses on the scientific results and the model description, and the technicalities are less well presented. However, the GMD initiators saw the needs to fully describe models and model developments in peer-reviewed publications. They aimed to guarantee reproducibility, traceability, transparency, and access [4]. Two nice quotes given on the GMD website are:

"I believe that the time is ripe for significantly better documentation of programs, and that we can best achieve this by considering programs to be works of literature."
(Donald E. Knuth, *Literate Programming*, 1984)

"Essentially, all models are wrong, but some are useful."
(George E.P. Box, *Robustness in the strategy of scientific model building*, 1979)

2.2. Review Criteria

Reviewers are asked to rate on the scientific significance, the scientific quality, the scientific reproducibility, as well as on the presentation quality. The reviewers decide whether substantial new concepts, ideas, or methods are described, whether the approaches and applied methods are valid, whether the models have the potential to perform calculations leading to significant scientific results, and to what extent the modelling science is reproducible. Fellow scientists have to be able to reproduce the sciences. Therefore, a focus is the completeness and the preciseness of the descriptions [3].

2.3. GMD Facts & Figures

By the end of March 2014, GMD had 596 manuscripts submitted from which 495 have been published in the discussion forum of GMD and 351 in GMD as final revised journal articles. The final articles have an average length of 16 pages (median) and the review takes on average (median) 33 days from submission to publication of the discussion paper, and 46 days from revised submission after public discussion to publication of the final revised and fully peer-reviewed paper. In the discussion forum, 2,129 comments have been posted, 1,018 of which are referee comments, 963 are author comments, 67 comments are published by the journal editors, and 81 comments by members of the scientific community prior to the final acceptance of the manuscripts [5]. GMD is indexed by Scopus and the Web of Science, and received the Thomson Reuters Impact Factor of 5.030 in 2012 [3].

3. Interactive Open Access Publishing

The Interactive Open Access Publishing aims to bring more transparency into scientific quality assurance by publishing the reviewer reports and the author's response freely accessible. In the first stage, the submitted manuscript is access-reviewed by one of the topical editors of the journal. It is a rapid review and involves only technical corrections. Then, the manuscript is typeset and published as so-called Discussion Paper. It is fully citable, receives a classical citation and pagination, as well as a DOI. The publishing platform is called the discussion forum.

The Discussion Paper is then subject to Interactive Public Discussion, during which the referees' comments (anonymous or attributed), additional short comments by other members of the scientific community (attributed) and the authors' replies are also published in the discussion forum alongside the Discussion Paper. Different from other initiatives experimenting with Public Peer-Review, the comments in this concept are also fully citable, paginated, typeset automatically by an online application, and remain online permanently.

In the second stage, the peer-review process is completed and, if accepted, the final revised papers are published in the journal itself. The latter is then the fully peer-

reviewed publication platform which is subject to indexing in the Web of Science, Scopus, and other databases.

The concept of Interactive Open Access Publishing started in 2001 and traces back to Ulrich Pöschl and Nobel laureate Paul Crutzen, both at the Max Planck Institute for Chemistry in Mainz, Germany. It was first applied to the journal *Atmospheric Chemistry and Physics* (ACP) [6], a very successful title owned by the European Geosciences Union (EGU) and published by Copernicus Publications.

Ulrich Pöschl described his concept in many publications [7], [8], [9].

References

- [1] ESSD journal website, available at: <http://www.earth-system-science-data.net>. Access on 27 March 2014.
- [2] ESSD paper statistics, taken from Copernicus Publications' manuscript review system Copernicus Office Editor. Access on 27 March 2014.
- [3] GMD journal website, available at: <http://www.geoscientific-model-development.net/>. Access on 27 March 2014.
- [4] GMD Executive Editors: Editorial: The publication of geoscientific model developments v1.0, *Geosci. Model Dev.*, 6, 1233-1242, doi:10.5194/gmd-6-1233-2013, 2013.
- [5] GMD paper statistics, taken from Copernicus Publications' manuscript review system Copernicus Office Editor. Access on 27 March 2014.
- [6] ACP journal website, available at: <http://www.atmospheric-chemistry-and-physics.net/>. Access on 27 March 2014.
- [7] Pöschl, U., Interactive journal concept for improved scientific publishing and quality assurance, *Learned Publishing*, 17, 105-113, 2004
- [8] Pöschl, U., Interactive Open Access Publishing and Peer Review: The Effectiveness and Perspectives of Transparency and Self-Regulation in Scientific Communication and Evaluation, *Liber Quarterly*, 19, 293-314, 2010
- [9] Pöschl, U., Multi-stage open peer review: scientific evaluation integrating the strengths of traditional peer review with the virtues of transparency and self-regulation, *Frontiers of Computational Neuroscience*, 6, 33, doi:10.3389/fncom.2012.00033, 2012