Research Presentation

From Open Access to Open Science from the viewpoint of a scholarly publisher

Lyubomir Penev ‡

‡ Pensoft Publishers & Bulgarian Academy of Sciences, Sofia, Bulgaria

Corresponding author: Lyubomir Penev (penev@pensoft.net)

Reviewable　v1

Abstract

Background

The open access publishing model led to dramatic changes in the way scientists communicate their results. Open access also challenged the traditional business models of academic publishers that have been maintained for hundreds of years. Open access to article content, however, soon appeared insufficient as far as access to underlying data was concerned. Opening research data came as the logical second stage of this challenge which was soon put on the agenda of scientific communities, funding organisations and governments. Open data, by itself, raised the question how we can re-use data and reproduce research results, how transparent is the peer-review and, more generally, how scientific evaluation is being performed. Over time, these and other similar developments morphed into what we now call "open science" or, in more general terms, transforming research into a primarily collaborative rather than a primarily competitive endeavour.

New information

The present lecture summarises the key milestones of the movement from open access through open data to open science from the viewpoint of an academic publisher. It is also illustrated by the ARPHA Biodiversity Data Publishing and Dissemination Toolbox (ARPHA-BioDiv) which is a set of standards, guidelines, tools, workflows and journals, developed by

Pensoft within its ARPHA Journal Publishing Platform. The history of development of ARPHA-BioDiv largely resembles the evolution of the open access to open science which started with the pre-publication semantic markup of important domain-specific terms and relationships between them, as implemented in 2010 by the ZooKeys open access journal and then followed by others, for example: PhytoKeys, MycoKeys, Nature Conservation, NeoBiota, Journal of Hymenoptera Research, Deutsche Entomologische Zeitschrift, Zoosystematics and Evolution. The next stage of integrated narrative and open data publishing was pioneered in 2013 by the Biodiversity Data Journal and its associated authoring tool, the ARPHA Writing Tool (AWT), launched as the first ever journal publishing workflow that supported the full life cycle of a manuscript, from writing through community peer-review, publication and dissemination within a single, entirely Web- and XML-based, online collaborative platform. The latest stage of open science publishing is demonstrated by the Research Ideas and Outcomes (RIO) journal that publishes all outputs of the research cycle – including project proposals, data, methods, workflows, software, project reports and research articles – together on a single collaborative platform, with the most transparent, open and public peer-review process.

Keywords

open access, open data, open science, scholarly publishing, semantic publishing, XML tagging, data re-use, science communication

Introduction

Open access to research articles was a disruptive change in the academic publishing paradigm that took place and, for about the last 20 years, is still continuing (Suber 2012). While open access was still in a process of establishing itself as a publishing model, it soon became clear that only opening the narrative, human-readable content (especially, if the latter is presented only as PDF), is far from sufficient to utilise its huge potential for scientific progress.

Open access plus open data were indeed great steps forward in both research and publishing practices. Nonetheless, even before open access and open data succeeded in becoming a significant publishing model, scientists, funders and governments started to realise that there were many other issues in the whole ecosystem of research production and communication that should be performed in a much more open and transparent manner than they are now. This is how we all arrived at the concept of open science (Nielsen 2011, Pontika et al. 2015, see also the TED talk video of Michael Nielsen). Open science refers to a whole range of issues around opening up the research life cycle, the most important of which are: (1) Open access, (2) Open data, (3) Free and Open-source software, (4) Reproducible research, (5) Open peer-review, (6) Open science policies, (7) Open funding, (8) Open science evaluation, (9) Open science tools and (10) Open education. A critical requirement of open science is the transparency in methodology,

observation and collection of data, open access and re-usability of research objects covering the entire research cycle, public accessibility and transparency of scientific communication – including the open peer review process – and using web-based open tools for scientific collaboration and communication. In brief, open science builds on collaboration rather than competition between researchers (European Commission 2016b).

The process of transformation of open access into open science academic publishing is the main focus of the current presentation held within the iDiv Seminar Series at the Biodiversity Informatics Unit of the German Centre for Integrative Biodiversity Research (iDiv), Leipzig, on 15th of February 2017. The presentation claims that the way we publish most of the scientific results nowadays creates some bottlenecks that hamper the otherwise extraodrinary rapid progress in science. It illustrates the transition from open access to open science in the field of biodiversity publishing which is the main area of expertise of the author and the publishing company he has established, Pensoft Publishers.

Presentation

This presentation consists of four main blocks: (1) Open access, (2) Open data, (3) Open Science and (4) The Future. The first block presents the story of the flagship journal of Pensoft, Zookeys, established as a conventional **open access** journal in 2008. Soon thereafter, we realised that the continuing emphasis on formats that make it difficult to extract the content algorithmically, e.g. paper or PDF, was – and still is! – one of the increasingly worrying impediments in data and content (re-)usablity (Agosti 2006, Agosti 2016). Compensating for this lack of machine readability requires significant additional effort of post-publication markup and data extraction into a structured form, in order to make publications and data inter-operable and re-usable. One of the solutions to this problem was the pre-publication markup of important domain-specific terms and relationships between them which has been implemented in ZooKeys in 2010 (Penev et al. 2010) and subsequently in other journals published by Pensoft, e.g. PhytoKeys, MycoKeys, Journal of Hymenoptera Research, Deutsche Entomologische Zeitschrift, Zoosystematics and Evolution (Penev et al. 2012). For the pre-publication markup, the TaxPub XML extension to the Journal Archival Tag Suite (JATS) developed by Plazi was used (Catapano 2010).

Open access to journal articles gave birth to a quickly growing baby now known as "**open data publishing**" which normally takes place as: (1) publishing data supplementary files to the article, (2) deposition of data in repositories and linking these to and from the article, (3) stand-alone description of the data as "data papers" or "data notes" and (4) publication of data integrated in the narrative content of the article. This last stage of publication of machine-readable, integrated structured biodiversity data and narrative was piloted by the Biodiversity Data Journal (BDJ) and its associated authoring tool, the ARPHA Writing Tool (AWT), launched within the ViBRANT EU Framework Seven (FP7) project (Smith et al. 2013). The Biodiversity Data Journal realised in practice the first ever journal publishing

workflow that supported the full life cycle of a manuscript, from writing through community peer-review, publication and dissemination within a single, entirely Web- and XML-based, online collaborative platform. Over the course of the years since its inception, the BDJ workflow has been continuously improved, e.g. by way of an upgrade to the ARPHA-XML journal publishing workflow as an integral part of the ARPHA Journal Publishing Platform.

Pensoft's response to the **open science** challenge was the launch of the Research Ideas and Outcomes (RIO) journal that publishes all outputs of the **research cycle** – including project proposals, data, methods, workflows, software, project reports and research articles – together on a single collaborative platform, with the most transparent, open and public peer-review process (Mietchen et al. 2015, see also the RIO video). The scope of the journal encompasses all areas of academic research, including science, technology, medicine, humanities and the social sciences. A good example of a collection of papers that covers a wide range of research outcomes is the one produced by the EU BON FP7 project: Building the European Biodiversity Observation Network (EU BON) Project Outputs .

What is next? At Pensoft, we believe that academic publishers will soon face another disruptive change in their everyday publishing practices which will be provoked by the need to handle, publish and export semantically enhanced content into Linked Open Data (LOD). Since 2015, Pensoft – together with our partners from Plazi – instigated an Open Biodiversity Knowledge Management System (OBKMS). OBKMS aims at converting and amalgamating RDF data extracted from legacy, prospectively published literature, and unpublished sources, together with ontologies and vocabularies, into a Graph database, in order to ensure cross-domain inter-operability and new horizons of data re-use in the semantic Web space (pro-iBiosphere 2014, Senderov and Penev 2016).

Data resources

The presentation that is described in the current article is available from Slideshare (http://www.slideshare.net/pensoft/from-open-access-to-open-science-from-the-viewpoint-of-a-scholarly-publisher-72128076), the Biodiversity Literature Repository at Zenodo (DOI: http://doi.org/10.5281/zenodo.291411), and also as a supplementary file to the article (Suppl. material 1).

Conclusions

Academic publishing and science communication, in general, experience disruptive transformations that can be summarised as follows:

1. **From open access to open science**. Open access and open data publishing models are quickly being transformed into open science practices that affect the whole ecosystem of producing, communicating and re-using research results.

2. **From human-readable to machine-readable content**. Machine readability of the content is now at least as important as human readability as it facilitates the automated harvesting, text mining and re-use of content.
3. **From open data to data re-use.** Publishers should strive to implement technologies that integrate structured data into the narrative to the highest possible degree.
4. **From traditional publishing to technology-driven service.** Technological innovations become critical for the proper publishing and dissemination of scientific content, hence for the survival and sustainability of scientific journals and publishers.
5. **From semantic enrichment of content to semantic publishing**. Semantic tagging and enrichment of content is seen as a transitional step towards the next stage of transformation of the published content into Linked Open Data (LOD).

Presented at

Lecture held at the Biodiversity Informatics Unit of the German Centre for Integrative Biodiversity Research (iDiv), Leipzig, on 15th of February 2017, within the iDiv Seminar Series.

References

- Agosti D (2006) Biodiversity data are out of local taxonomists' reach. Nature 439 (7075): 392-392. https://doi.org/10.1038/439392a
- Agosti D (2016) Where Do We Come From, Where Do We Go To? 20 Years Of Open Access To Biodiversity Knowledge. Zenodo https://doi.org/10.5281/ZENODO.165979
- Catapano T (2010) TaxPub: An extension of the NLM/NCBI Journal Publishing DTD for taxonomic descriptions. Proceedings. Journal Article Tag Suite Conference. National Center for Biotechnology Information, Bethesda (MD) URL: http://www.ncbi.nlm.nih.gov/books/NBK47081/
- European Commission (2016a) H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Accession date: 2017 2 09.
- European Commission (2016b) Draft European Open Science Agenda. http://ec.europa.eu/research/openscience/pdf/draft_european_open_science_agenda.pdf#view=fit&pagemode=none. Accession date: 2017 2 09.
- Mietchen D, Mounce R, Penev L (2015) Publishing the research process. Research Ideas and Outcomes 1: e7547. https://doi.org/10.3897/rio.1.e7547
- Nielsen M (2011) Reinventing Discovery: The New Era of Networked Science. Princeton University Press, Princeton, N.J. [ISBN 978-0-691-14890-8]
- Penev L, Catapano T, Agosti D, Sautter G, Stoev P (2012) Implementation of TaxPub, an NLM DTD extension for domain-specific markup in taxonomy, from the experience of a biodiversity publisher. Proceedings. Journal Article Tag Suite Conference (JATS-Con),

Bethesda (MD). National Center for Biotechnology Information URL: http://
www.ncbi.nlm.nih.gov/books/NBK100351/

- Penev L, Agosti D, Georgiev T, Catapano T, Miller J, Blagoderov V, Roberts D, Smith V,
Brake I, Ryrcroft S, Scott B, Johnson N, Morris R, Sautter G, Chavan V, Robertson T,
Remsen D, Stoev P, Parr C, Knapp S, Kress WJ, Thompson F, Erwin T (2010) Semantic
tagging of and semantic enhancements to systematics papers: ZooKeys working
examples. ZooKeys 50: 1-16. https://doi.org/10.3897/zookeys.50.538
- Pontika N, Knoth P, Cancellieri M, Pearce S (2015) Fostering open science to research
using a taxonomy and an eLearning portal. Graz, Austria, 21 - 22 October 2015.
Proceedings of the 15th International Conference on Knowledge Technologies and
Data-driven Business (i-KNOW '15) https://doi.org/10.1145/2809563.2809571
- pro-iBiosphere (2014) Open Biodiversity Knowledge Management System (OBKMS).
http://adm.pro-ibiosphere.eu/getatt.php?filename=oo_4749.pdf.. Accession date: 2017 2
09.
- Senderov V, Penev L (2016) The Open Biodiversity Knowledge Management System in
Scholarly Publishing. Research Ideas and Outcomes 2: e7757. https://doi.org/10.3897/
rio.2.e7757
- Smith V, Georgiev T, Stoev P, Biserkov J, Miller J, Livermore L, Baker E, Mietchen D,
Couvreur T, Mueller G, Dikow T, Helgen K, Frank J, Agosti D, Roberts D, Penev L
(2013) Beyond dead trees: integrating the scientific process in the Biodiversity Data
Journal. Biodiversity Data Journal 1: e995. https://doi.org/10.3897/bdj.1.e995
- Suber P (2012) Open access (The MIT Press Essential Knowledge Series ed.). MIT
Press, Cambridge, Massachusets. [ISBN ISBN 9780262517638]

Supplementary material

Suppl. material 1: From Open Access to Open Science from the Viewpoint of a Scholarly Publisher

Authors: Lyubomir Penev
Data type: PowerPoint presentation (pptx)
Brief description: A presentation held by Lyubomir Penev in the iDiv Seminar Seies at the Biodiversity Informatics Unit of the German Centre for Integrative Biodiversity Research (iDiv) Leipzig, 15 February 2017.
Filename: Open Access to Open Scienve iDiv Leipzig 15 Feb 2017 Pensoft.pptx - Download file (55.94 MB)