Deliverable 8.3 (D8.3)

Data Publishing and Dissemination Toolbox

M52

Project acronym:	EU BON
Project name:	EU BON: Building the European Biodiversity Observation Network
Call:	ENV.2012.6.2-2
Grant agreement:	308454
Project duration:	01/12/2012 – 31/05/2017 (54 months)
Co-ordinator:	MfN, Museum für Naturkunde - Leibniz Institute for Research on Evolution and Biodiversity, Germany
Delivery date from Annex I:	M52 (March 2017)
Actual delivery date:	M52 (March 2017)
Lead beneficiary:	Pensoft
Authors:	Lyubomir Penev, Teodor Georgiev, Petar Geshev, Seyhan Demirov, Iliyana Kuzmova, Pavel Stoev (Pensoft)

This project is supported by funding from the specific programme 'Cooperation', theme 'Environment (including Climate Change)' under the 7th Research Framework Programme of the European Union

Dissemination Level

PU	Public	✓
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 308454.

All intellectual property rights are owned by the EU BON consortium members and protected by the applicable laws. Except where otherwise specified, all document contents are: "© EU BON project". This document is published in open access and distributed under the terms of the Creative Commons Attribution License 3.0 (CC-BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Executive Summary

Introduction

The ARPHA-BioDiv Toolbox for Scholarly Publishing and Dissemination of Biodiversity Data is a set of standards, guidelines, recommendations, tools, workflows, journals and services, based on the ARPHA Publishing Platform of Pensoft, designed to ease scholarly publishing of biodiversity and biodiversity-related data that are of primary interest to EU BON and GEO BON networks. ARPHA-BioDiv is based on the infrastructure, knowledge and experience gathered in the years-long research, development and publishing activities of Pensoft, upgraded with novel tools and workflows that resulted from the FP7 project EU BON.

Progress towards objectives

The toolbox comprises a comprehensive set of standards, guidelines, recommendations, tools, workflows, journals and services towards improvement of data publishing in the biodiversity domain. Advanced semantic markup layer for each publication within the ARPHA-BioDiv tool facilitates discoverability, re-use and archiving. Based on the technologically advanced ARPHA Platform, the toolbox features novel data publishing workflows that allow streamlined publication of occurrence records and metadata from major data repositories. Novel article templates were developed to meet the needs of biodiversity data publishing.

All developments of ARPHA-BioDiv to date are also published in an open access paper within the Open Science Pilot Collection in Research Ideas& Outcomes (RIO) journal.

Achievements and current status

To achieve wider outreach and ensure sustainability, this report was also published as a scientific paper, within the EU BON Open Science Collection, hosted in the innovative Research Ideas & Outcomes (RIO) Journal:

Penev L, Georgiev T, Geshev P, Demirov S, Senderov V, Kuzmova I, Kostadinova I, Peneva S, Stoev P (2017) ARPHA-BioDiv: A toolbox for scholarly publication and dissemination of biodiversity data based on the ARPHA Publishing Platform. Research Ideas and Outcomes 3: e13088. https://doi.org/10.3897/rio.3.e13088

Future developments

Being part of the ARPHA Platform, all features and tools within ARPHA-BioDiv will be continuously updated. Moreover, news tools and workflows will be added in future.

Table of content

1. Data publishing and dissemination toolbox

1.1 What is ARPHA-BioDiv?

The transformation from human- to machine-readability of published content is a key feature of the dramatic changes experienced by academic publishing in the last decade. Non-machine readable PDFs, either digitally born or scanned from paper prints, require significant additional effort of post-publication markup and data extraction into a structured form, in order to address issues of interoperability and reuse of publications and data (Agosti 2006, Penev et al. 2010, Agosti 2016). A partial solution to the problem is the pre-publication markup which can be generic (e.g., for the article metadata and the standard division into article sections such as Introduction, Material and Methods and others) and domain-specific (e.g. markup of taxon names or biological collection codes). The open access journal ZooKeys was the first to implement both generic and domain-specific markup which was adopted thereafter by PhytoKeys, MycoKeys, Journal of Hymenoptera Research, Deutsche Entomologische Zeitschrift, Zoosystematics and Evolution and other Pensoft journals (Penev et al. 2010, Penev et al. 2012). The domain-specific, pre-publication markup was possible thanks to the TaxPub XML schema, developed by Plazi and later endorsed as an extension to the Journal Archival Tag Suite (JATS) by the National Library of Medicie of the USA (Catapano 2010). The pre-publication markup required creation of some tools to facilitate the process (for example, Pensoft Markup Tool and Pensoft Wiki Convertor) and also other tools to visualise the results of it (for example, Pensoft Taxon Profile, or PTP).

The next stage of development of integrated narrative and data publishing was landmarked by the Biodiversity Data Journal (BDJ) and its associated authoring tool, ARPHA Writing Tool (AWT), launched within the ViBRANT EU Framework Seven (FP7) project (Smith et al. 2013). The Biodiversity Data Journal was the first ever journal that provided a fully Web- and XML-based life cycle of a manuscript, starting from authoring to submission, peer review, publishing and dissemination. Later, the BDJ workflow was upgraded to the "ARPHA-XML journal publishing workflow" which itself is a part of the ARPHA Journal Publishing Platform (Penev 2017). The ARPHA-XML workflow came with several tools and workflows developed by Pensoft, such as ReFindit for discovery and import of literature and data references, import/export of tabular data and also of Darwin Core occurrence records, conversion of Ecological Metadata Language (EML) metadata into manuscripts, automated archiving of articles and sub-article elements in Zenodo and others (for details, see next section).

The third stage of Pensoft's effort towards open science publishing was the launch of the Research Ideas and Outcomes (RIO) journal that publishes all outputs of the **research cycle**, beginning with research ideas; project proposals; data and software management plans; data; methods; workflows; software; and going all the way to project reports; research and review articles, using the most transparent, open and public peer review process (Mietchen et al. 2015). The RIO Journal publishes open science collections of various project or research cycle outcomes, with the EU BON project collection, entitled Building the European Biodiversity Observation Network (EU BON) Project Outputs, being a fine example.

Eventually, all these years spent in development of novel approaches to publication of biodiversity data resulted in a set of standards, guidelines, workflows, tools, journals and services which we define here as ARPHA-BioDiv: A Toolbox for Scholarly Publishing and Dissemination of Biodiversity Data (**Fig.1**). The toolbox is designed to ease scholarly publishing of biodiversity and biodiversity-related data with special emphasis on the EU BON and GEO BON networks.

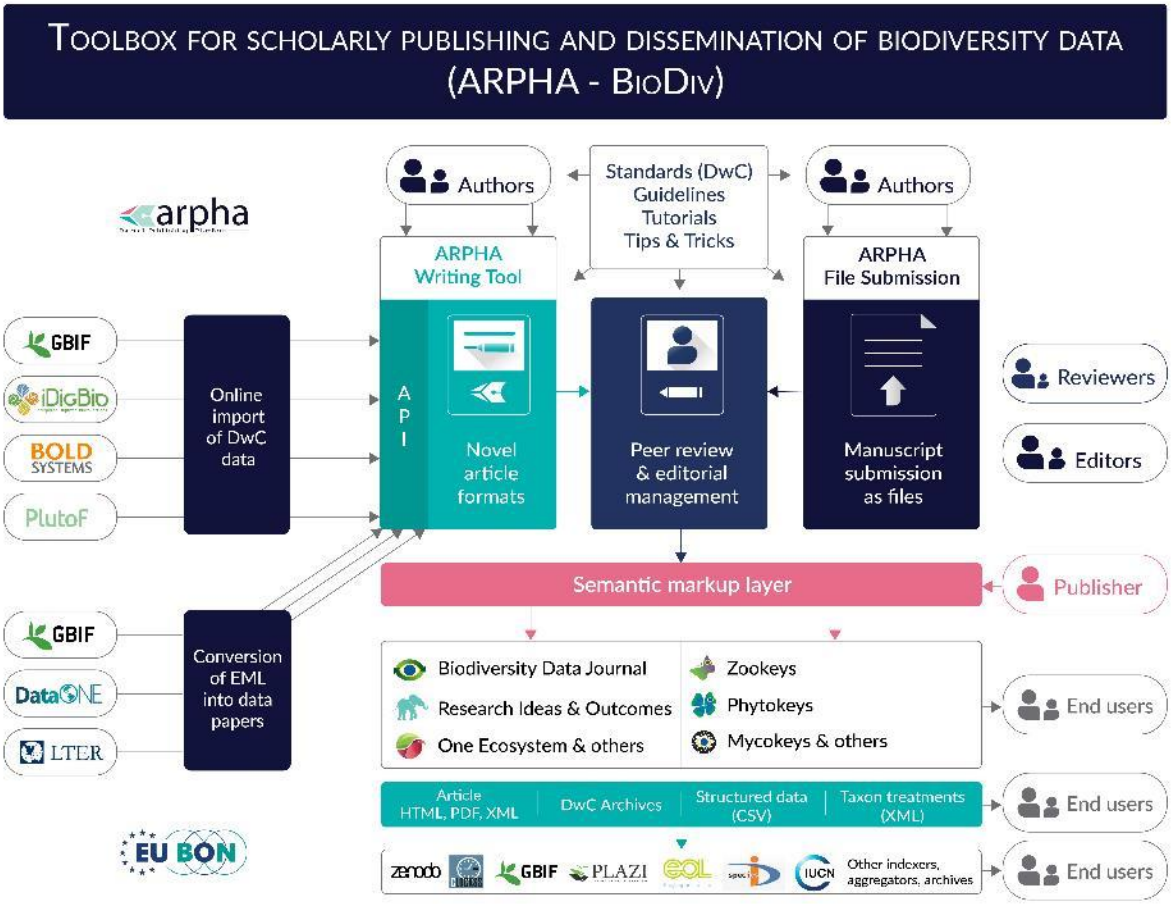

Figure 1. ARPHA-BioDiv is a set of standards, guidelines, tutorials, tools, workflows, journals and services, designed to facilitate the scholarly publication and dissemination of biodiversity data.

2. ARPHA Journal Publishing Platform

The market for online collaborative writing tools has long been dominated by Google Docs. However, as it is too generic, it has not met the specific demands of academic publishing and, in recent years, some start-ups have developed platforms and services to fulfil this increasing gap in the publishing market. Some examples include OverLeaf (originally WriteLaTeX), Authorea, ShareLatex and others, most of them being based on LaTeX, but differing in the level of complexity and features for manuscript writing. For people unfamiliar with LaTeX, the learning curve is steep which explains the comparatively restricted usage, mostly centred around the LaTeX community. Currently, none of the above-mentioned tools provides all the components of an end-to-end authoring, peer review and publishing pipeline. For instance, most tools lack a peer review system and rely on integrations with well-established platforms, such as Editorial Manager, ScholarOne, or others.

ARPHA has emerged as the first ever publishing platform to support the full life cycle of a manuscript, from authoring through submission, peer review, publication and dissemination, within a single, fully Web- and XML-based, online collaborative environment. The acronym ARPHA stands for "Authoring, Reviewing, Publishing, Hosting and Archiving" - all in one place, for the first time. The most distinct feature of ARPHA, amongst others, is that it consists of two interconnected but independently functioning journal publishing platforms. Thus, it can provide to journals and

publishers either of the two or a combination of both services by enabling a smooth transition from the conventional, document-based workflows to fully XML-based publishing (**Fig. 2**):

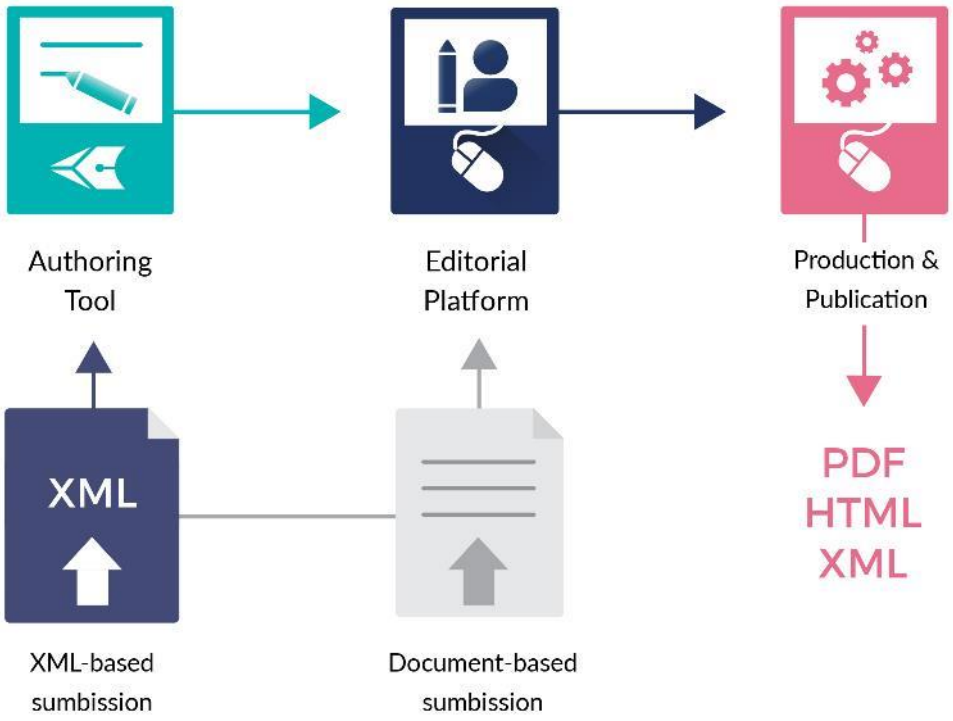

Figure 2. ARPHA consists of two independent journal publishing workflows: (1) ARPHA-XML, where the manuscript is written and processed via ARPHA Writing Tool and (2) ARPHA-DOC, where the manuscript is submitted and processed as document file(s).

1. ARPHA-XML: Entirely XML- and Web-based, collaborative authoring, peer review and publication workflow;
2. ARPHA-DOC: Document-based submission, peer review and publication workflow.

The two workflows use a one-stop login interface and a common peer-review and editorial manuscript tracking system. The XML-based workflow in use at Biodiversity Data Journal (BDJ) was the first of its kind back in 2013 and has since seen continuous refinement over the course of more than three years of active use by the biodiversity research community. It is also now used by the Research Ideas and Outcomes (RIO), One Ecosystem and BioDiscovery journals. The second, file-based submission workflow, is currently used by ZooKeys, PhytoKeys, MycoKeys, Journal of Hymenoptera Research, Nature Conservation, Deutsche Entomologische Zeitschrift, Zoosystematics and Evolution, NeoBiota and other journals, published by Pensoft.

At the core of the ARPHA-XML workflow is the collaborative online manuscript authoring module called ARPHA Writing Tool (AWT). AWT's innovative features allow for upfront markup, automisation and structuring of the free-text content during the authoring process, import/download of

structured data into/from human-readable text, automated export and dissemination of small data, on-the-fly layout of composite figures and import of literature and data references from online resources. ARPHA-XML is also perhaps the first journal publishing system that allows for submission of complex manuscripts via a dedicated API.

The generic and domain-specific features of ARPHA (used for publication and dissemination of biodiversity data via the ARPHa-BioDiv toolbox) are listed in **Table 1** and **Table 2** respectively.

Table 1. Generic features of the ARPHA Journal Publishing Platform

FEATURE	ARPHA-DOC	ARPHA-XML
ARPHA is a combination of software platform and a wide range of associated services.	X	X
ARPHA serves individual journals or multiple journal platforms.	X	X
Integrated with the industry leading indexing and archiving platform (see list) through web services, APIs and data exchange protocols.	X	X
Individual journal website design.	X	X
Customisable submission module.	X	X
Peer review and editorial management system.	X	X
Peer review process customisable by journal. It can be conventional (either single-blind or double-blind), community-sourced, or public.	X	X
Online collaborative authoring tool (ARPHA Writing Tool, abbreviated AWT, formerly Pensoft Writing Tool, abbreviated PWT), closely integrated with submission, peer review, production and dissemination tools.		X
Collaborative work on a manuscript with co-authors; external contributors, such as mentors; pre-submission reviewers; linguistic and copy editors; or colleagues. The external contributors are not listed as co-authors of the manuscript.		X
Large set of pre-defined, but flexible article templates covering many types of research outcomes.		X
Online search and import of literature or data references; cross-referencing of in-text citations; import of tables; upload of images and multimedia; assembling images for display as composite figures.		X
Automated technical validation step (it can be triggered by authors any time) checks the manuscript for consistency and for compliance with the JATS standard as well as the journal's requirements.		X
Human-based, interactive pre-submission technical check and validation tool helps authors to proceed with their manuscripts to a form almost ready for publication.		X
Pre-submission external peer review(s) performed during the authoring process. The pre-submission peer reviews are submitted together with the manuscript to prompt editorial evaluation and publication.		X

For editor's convenience, peer reviews in ARPHA are automatically consolidated into a single online file that makes the editorial process straightforward, easy and comfortable.		X
In the ARPHA-XML workflow, authors can publish updated versions of their articles anytime.		X
Automated archiving of all published articles in Zenodo and CLOCKSS on the day of publication.	X	X

Table 2. Domain-specific features of the ARPHA-BioDiv toolbox used for publication and dissemination of biodiversity data

FEATURE	ARPHA-DOC	ARPHA-XML
Markup and visualisation of all taxon names used in the text.	X	X
Markup and visualisation of taxon treatments following the TaxPub XML schema (an extension of the Journal Archiving Tag Suite (JATS) used by PubMed and PubMedCentral).	X	X
Markup and automated mapping of geo-coordinates of geographical locations.	X	X
Markup and visualisation of biological collection codes against the Global Registry of Biological Repositories (GRBIO) vocabulary (Schindel et al. 2016).	X	X
Pre-publication registration of new taxa in ZooBank, IPNI or Index Fungorum (as relevant).	X	X
Dynamic, real-time creation of online profile for each taxon name mentioned in an article through the Pensoft Taxon Profile tool.	X	X
Automated linking through the Pensoft Taxon Profile tool of each taxon name mentioned in an article to various biodiversity resources (GBIF, Encyclopedia of Life, Biodiversity Heritage Library, the National Center for Biodiversity Information (NCBI), Genbank and Barcode of Life, PubMed, PubMedCentral, Google Scholar, the International Plant Name Index (IPNI), MycoBank, Index Fungorum, ZooBank, PLANTS, Tropicos, Wikispecies, Wikipedia, Species-ID and others).	X	X
Workflow integration with the GBIF Integrated Publishing Toolkit (IPT) for deposition, publication and permanent linking between data and articles of primary biodiversity data (species-by-occurrence records), checklists and their associated metadata.	X	X
Workflow integration with the Dryad Data Repository for	X	X

deposition, publication and permanent linking between data and articles of datasets other than primary biodiversity data (e.g. ecological observations, environmental data, genome data and other data types).		
Export of XML-based metadata and TaxPub XMLs of the papers to PubMedCentral.	X	X
Automated export of all taxon treatments (new taxa and re-descriptions, including images) to Encyclopedia of Life. Example: http://eol.org/pages/21232877/overview.	X	X
Automated export of all taxon treatments (new taxa and re-descriptions) to Plazi TreatmentBank. Example: http://tb.plazi.org/GgServer/html/B07E9CD77F60DCC65C10A381F6E3BBF0	X	X
Automated export of all taxon treatments (new taxa and re-descriptions), including images, keys, etc. to the Wiki repository Species-ID. Example: http://species-id.net/wiki/Spigelia_genuflexa.	X	X
Automated export of the occurrence data published in BDJ into Darwin Core Archive (DwC-A) format (see also Baker et al. 2014) and its consequent ingestion by GBIF. The DwC-A is freely available for download from each article's webpage that contains occurrence data.		X
Automated export of the taxonomic treatments published in BDJ into Darwin Core Archive. The DwC-A is freely available for download from each article's webpage that contains taxonomic treatments data.		X
Automated export and archiving of images from the published articles in Zenodo. Images from biodiversity journals are imported into the Biodiversity Literature Respository (BLR) of Zenodo.	X	X
Import of Darwin Core-compliant primary biodiversity data from spreadsheet templates or via a manual Darwin Core editor and consequent publication in a structured downloadable format (Smith et al. 2013, Robertson et al. 2014, Wieczorek et al. 2012).		X
Direct online import of Darwin Core-compliant primary biodiversity data from GBIF, Barcode of Life, iDigBio, and PlutoF into manuscripts (Senderov et al. 2016).		X
Multiple imports of voucher specimen records associated with a particular Barcode Index Number (BIN) (Ratnasingham and Hebert 2007) from the Barcode of Life.		X
Automated generation of data paper manuscripts from Ecological Metadata Language (EML) metadata files stored at GBIF Integrated Publishing Toolkit (GBIF IPT), DataONE and the Long Term Ecological Research		X

Network (LTER) (Senderov et al. 2016; for details, see also Pensoft's blog).		
Novel article types in ARPHA Writing Tool: Taxonomic Paper, Data Paper, Software Description, Monitoring Schema, Ecosystem Inventory, Ecosystem Service Inventory, Ecosystem Service Models, Species Conservation Profile, compliant with the IUCN Red List (Cardoso et al. 2016), Alien Species Profile, compliant with the IUCN Global Invasive Species Database (GISD) and others.	X	X
Nomenclatural acts modelled and developed in BDJ as different types of taxonomic treatments for plant taxonomy.		X
Automated archiving of all biodiversity articles in the Biodiversity Literature Respository (BLR) of Zenodo.	X	X

3. Novel Article Formats

Research articles have traditionally been containers for scientiifc results for several centuries and this holds even more for research books. The Internet era brought disruptive changes to academic publishing and one of these is that the notion of the research article as the only valid output for scientific endeavours was challenged. Resulting from this, novel article formats started to proliferate in an attempt to publish extra research objects from across the research cycle, such as methods, data and software. Pensoft pioneered several novel article formats with the launch of the Biodiversity Data Journal. Currently, the ARPHA Writing Tool supports nearly fifty article formats (**Fig. 3**), used in the Biodiversity Data Journal, Research Ideas and Outcomes, One Ecosystem, and BioDiscovery. The article formats can be generic, e.g. used within almost any domain (for example, research idea, research article, data management plan and others), or domain-specific, such as the article formats described below.

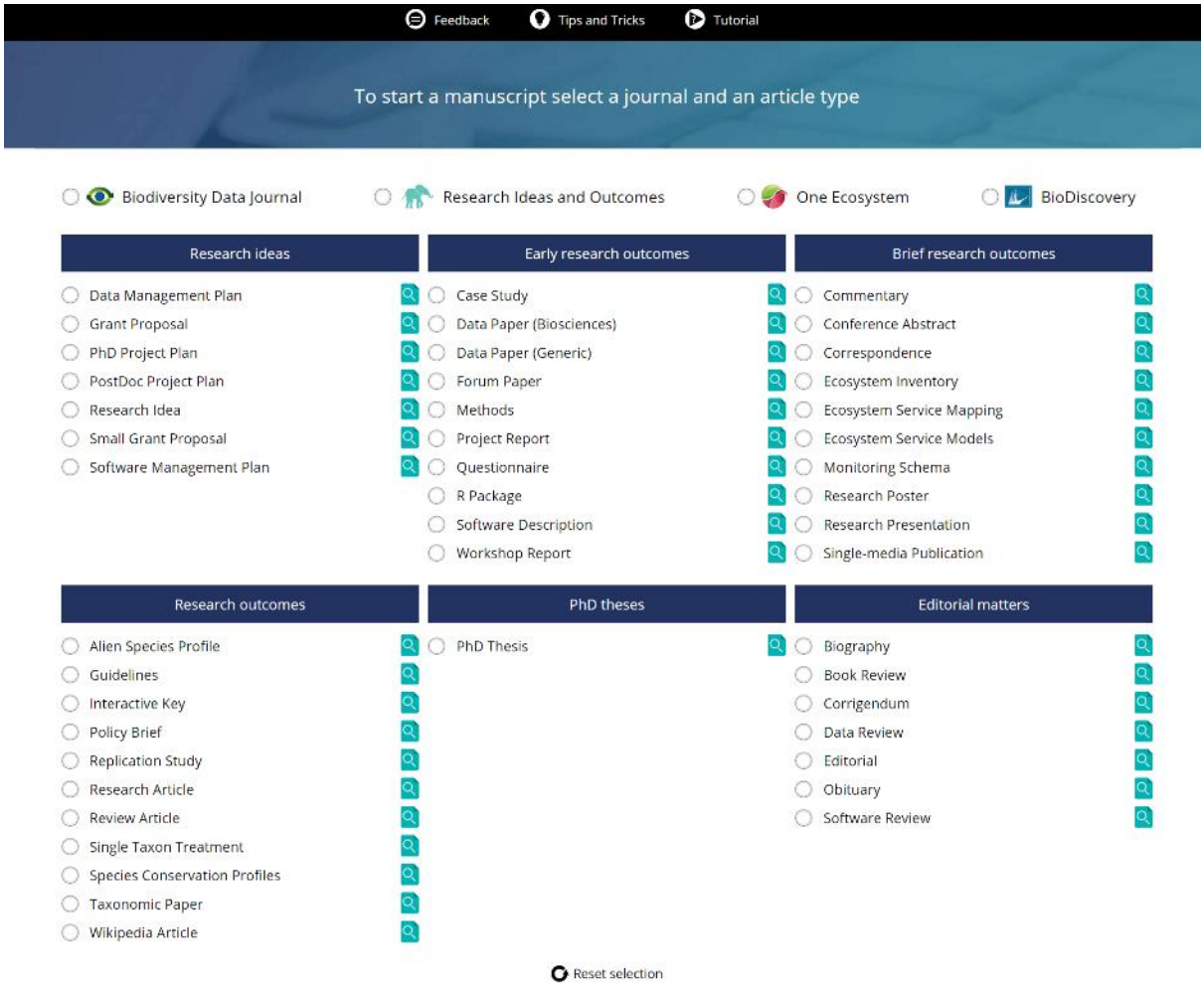

Figure 3. Article formats available in ARPHA Writing Tool.

3.1 Data Paper

A data paper is a scholarly journal publication whose primary purpose is to describe a dataset or a group of datasets, rather than report a research investigation. As such, it contains facts about data, rather than hypotheses and arguments in support of those hypotheses based upon data, as found in a conventional research article (for details, see Newman and Corke 2009, Chavan and Penev 2011, Penev et al. 2017).

Examples from: ZooKeys, Biodiversity Data Journal, PhytoKeys, Nature Conservation.

The Article template is available for Biodiversity Data Journal, One Ecosystem, Research Ideas and Outcomes (RIO), BioDiscovery.

3.2 Software Description

A publication that describes software or an online platform. It contains a link to an openly accessible code (for details, see Penev et al. 2017).

Examples from: Biodiversity Data Journal.

Customisable templates are available for Biodiversity Data Journal, Research Ideas and Outcomes (RIO), One Ecosystem and BioDiscovery.

3.2 R Package

A description of an R Package including information on its purpose, installation and usage. The code should be openly available and a link to it should be present in the article.

The Article template is available for Biodiversity Data Journal, One Ecosystem, Research Ideas and Outcomes (RIO).

3.3 Monitoring Schema

A brief description of a monitoring schema including information on the monitored system component; its location; indicators used; spatial and temporal scales; purpose of the monitoring programme; and potential application of the resulting data.

The Article template is available for Research Ideas and Outcomes (RIO) and One Ecosystem.

3.4 Species Conservation Profile (SCP)

A publication of a single or multiple IUCN species assessment report(s) imported and edited in an IUCN-compliant species template.

Examples from: Biodiversity Data Journal.

The Article template is available for Biodiversity Data Journal.

3.5 Alien Species Profile (ASP)

An assessment report of alien or invasive species following an IUCN-compliant species template. After publication, the article can be exported to the Global Invasive Species Database (GISD).

The Article template is available for Biodiversity Data Journal.

3.6 Ecosystem Inventory

A brief description of a specific ecosystem type; its structures; processes and functions; abundant species; biodiversity; anthropogenic pressures; and management options. Data could result from, for example, direct observations, monitoring programmes, modelling or literature and database reviews.

The Article template is available for One Ecosystem.

3.7 Ecosystem Service Mapping

A brief description of an ecosystem service mapping study or application including information on the purpose of the map; data and methods used (biophysical, economic, social); mapped ecosystem service; mapped beneficiary (ecosystem service potential, flow, demand); spatial and temporal scale and indicators. The resulting maps should be included in the manuscript or uploaded to the ESP Visualisation Tool.

The Article template is available for One Ecosystem.

3.8 Ecosystem Service Models

A brief description of an ecosystem service mapping study or application including information on the purpose of the map; data and methods used (biophysical, economic, social); mapped ecosystem service; mapped beneficiary (ecosystem service potential, flow, demand); spatial and temporal scale and indicators. The resulting maps should be included in the manuscript or uploaded to the ESP Visualisation tool.

The Article template is available for One Ecosystem.

4. Semantic Tagging of the Article Content

In 2010, ZooKeys published its 50th issue Taxonomy shifts up a gear: New publishing tools to accelerate biodiversity research in a new format based on pre-publication tagging of biodiversity-specific terms in the article XML and semantic enhancements to the published paper (Penev et al. 2010b, Penev et al. 2010a). ZooKeys implemented the TaxPub XML schema, developed by Plazi, later endorsed as an extension of the Journal Archiving Tag Suite (JATS) standard (Catapano 2010). Since then, all life science journals published by Pensoft use the semantic markup workflow in their everyday editorial work to "atomise" and disseminate the content at sub-article level. A list of tools and features for semantic tagging and enhancements of the article content is available in Table 2; implementation and use cases are reviewed by Penev et al. (2012). Examples of the use of the domain-specific markup are illustrated in **Fig. 4 a-d**.

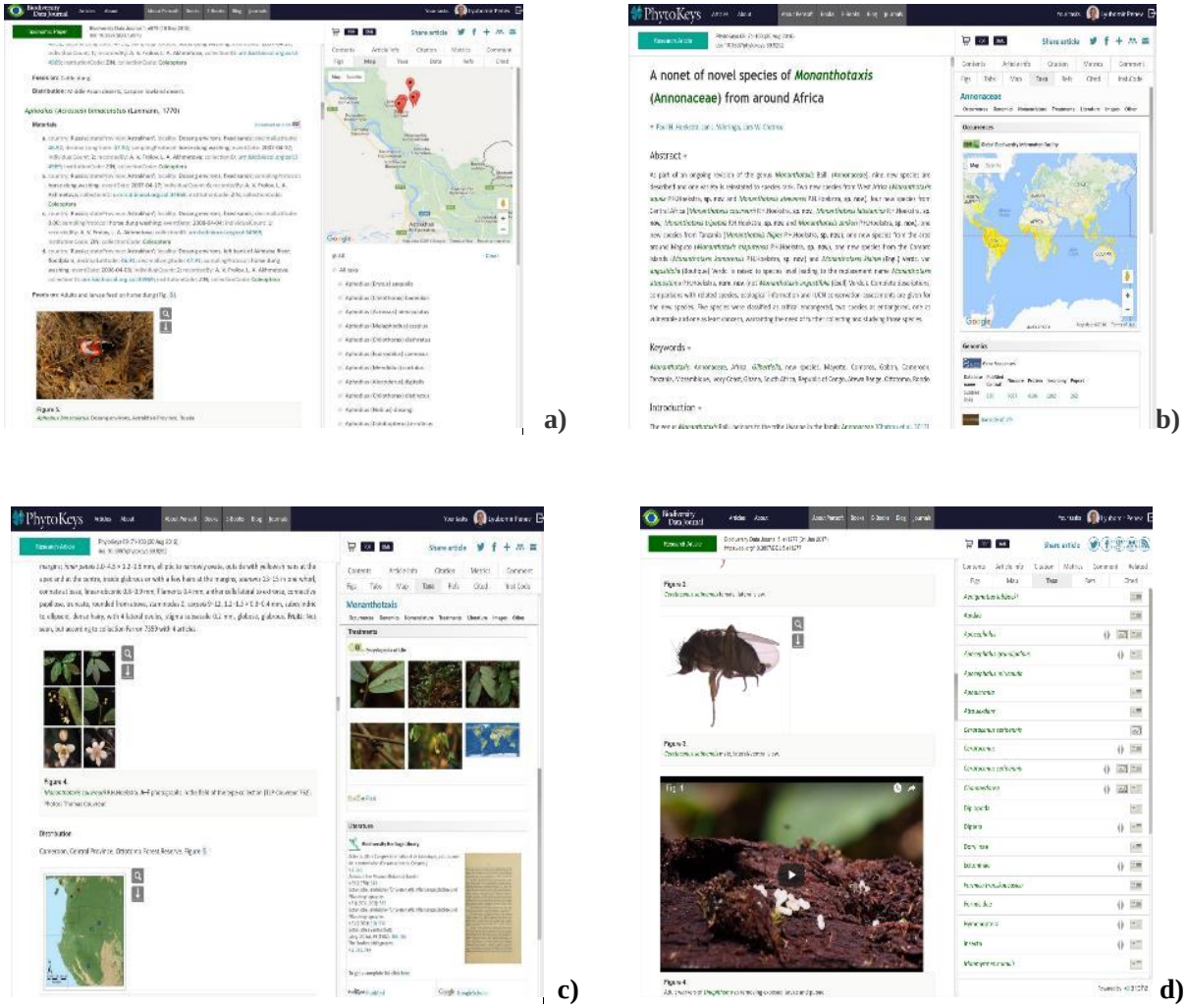

Figure 4 a-d. Examples of use of the domain-specific XML markup in the published artices.

a) Interactive mapping of geo-coordinated species occurrences (example from Frolov and Akhmetova 2013).
b) Pensoft Taxon Profile (PTP) is created in real time by clicking on any taxon name mentioned in an article (in this case Annoniaceae from Hoekstra et al. 2016).

c) Images and pages from historic literature where a taxon name has been mentioned are available from various sources (e.g. Encyclopedia of Life and the Biodiversity Heritage Library via Pensoft Taxon Profile (PTP) (in this case Annoniaceae from Hoekstra et al. 2016).
d) All taxon names usages (TNU) in an article are indexed and matched to their type of use (e.g. citations in the text, heading a taxon treatment, associated to images or present in identification keys, example from Brown et al. 2017).

5. Integrated Narrative and Data Publishing

The "integrated narrative and data publishing", or "integrated data publishing", is a relatively new approach, assuming that data or code are imported in a structured form in the manuscript text and are downloadable from the published article. In biodiversity science, this term has been coined and first demonstrated by the Biodiversity Data Journal (BDJ), developed in the course of the EU-funded project ViBRANT (Smith et al. 2013, see also **Fig. 5**). Publishing of an executable code, also known as "literate programming", in an article was proposed back in 1984 (Knuth 1984), but only recently did we see this practice in journals (Veres and Adolfsson 2011). Another example of integrated data publishing is the linking of a standard article to an external platform that hosts all data associated with the article and provides additional data analysis tools and computing resources; this approach is believed to have been pioneered by the GigaDB and the GigaScience journal (Edmunds et al. 2016). Various kinds of implementing 3D or other multimedia visualisations in an article can also be considered as integrated narrative and data publishing; a good example of that in the biodiversity domain is the paper of Stoev et al. (2013).

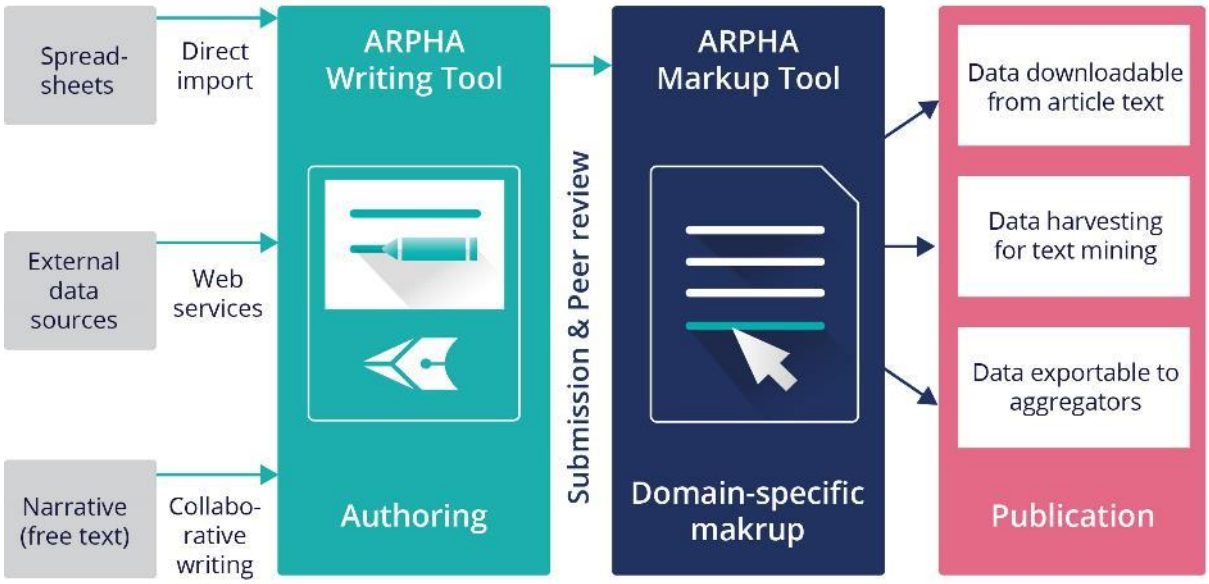

Figure 5. Integrated data and narrative publishing in the ARPHA-XML journal workflow.

5.1 Import of Data into Manuscripts

The ARPHA Writing Tool provides online direct import from external databases using community-accepted standards (e.g. within the biodiversity community, these are Darwin Core, TaxPub JATS extension and others - see http://www.tdwg.org/standards/). Initially, data import was from CSV

spreadsheets or manually via a Darwin Core HTML editor (Penev et al. 2017). A new functionality of the integrated data publishing system in ARPHA is the online import of specimen records from GBIF, Barcode of Life, iDigBio and PlutoF (**Fig. 6**). The workflow is described in Senderov et al. (2016). Stepwise guidelines on how to use the feature are also available from Penev et al. (2017) and a blog post.

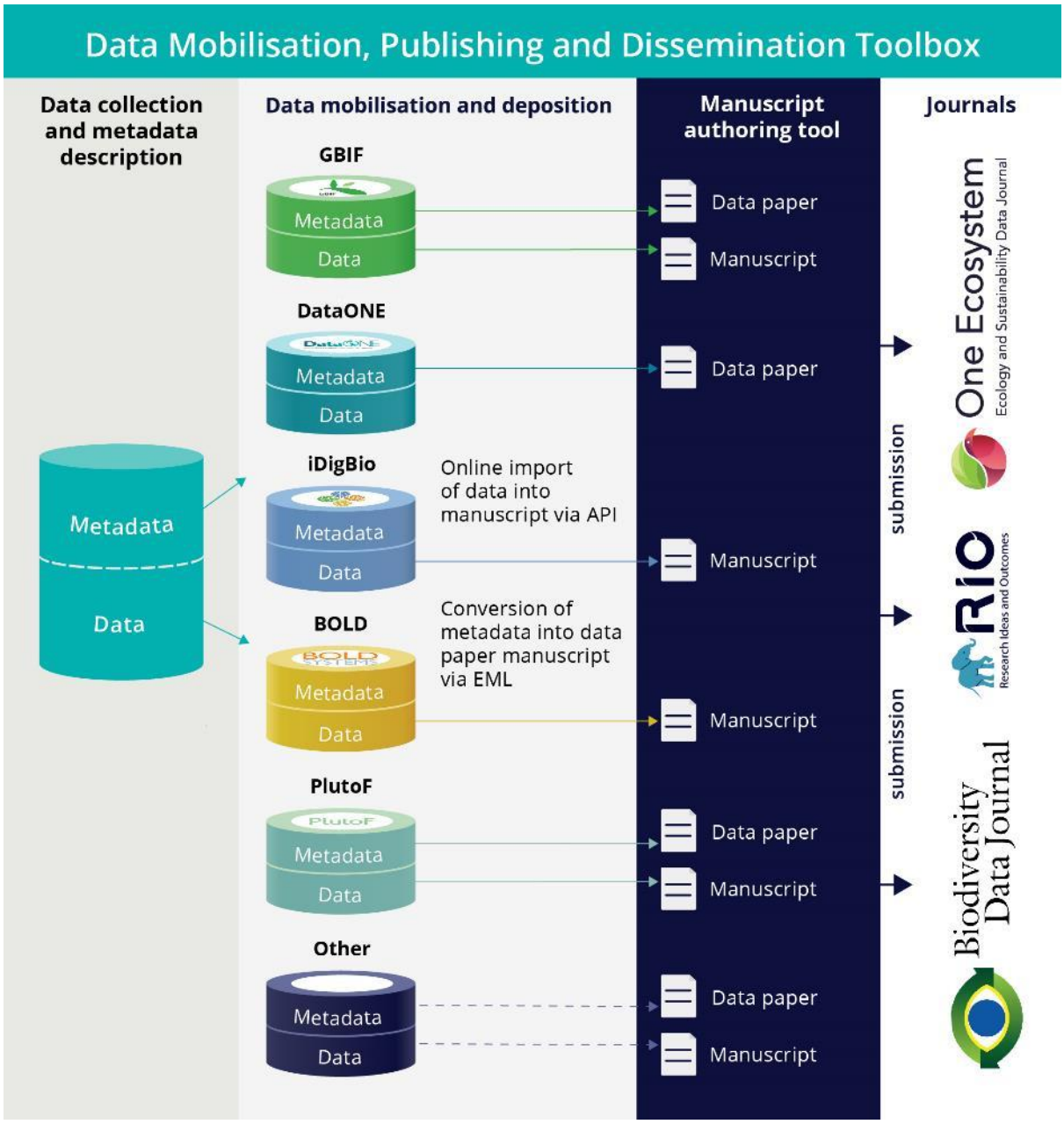

Figure 6. Data and metatada import into manuscripts in ARPHA Writing Tool.

Another example of online import of structured text is the ReFindit tool which exists both as a stand-alone application and a plugin in ARPHA Writing Tool. ReFindit locates and imports literature and data references from CrossRef, DataCite, RefBank, Global Names Usage Bank (GNUB) and Mendeley.

5.2 Content and Data Export from Published Articles

Article content that is tagged and available in TaxPub XML can be harvested by aggregators which can select and pick sub-article elements, such as metadata, taxon treatments, occurrence records, images and others. Several of these aggregators are major players in biodiversity data preservation and management, for example, GBIF, Encycopedia of Life, Biodiversity Heritage Library, Plazi, Biodiversity Literature Repository at Zenodo, ZooBank, International Plant Names Index, MycoBank, Index Fungorum and many others. The data export in some cases is provided by a featured outbound API. The workflows and aggregators that use the semantically enriched article XMLs are listed in Table 2, and illustrated in part on **Fig. 7**; the initial core set of features was also reviewed by Penev et al. (2010b) and Penev et al. (2012).

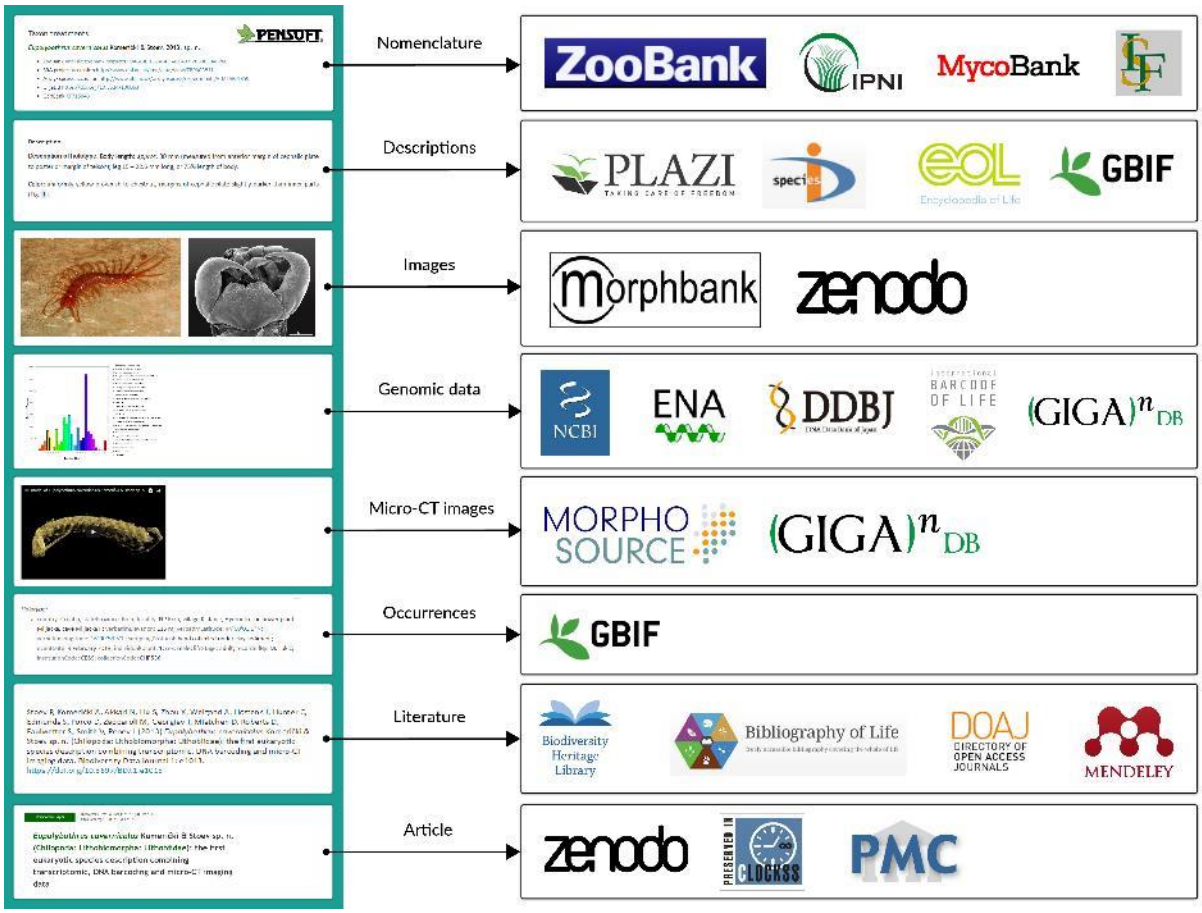

Figure 7. Extraction and delivery of data and content from published articles to aggregators, nomenclators, archives, and indexers.

All data published in the Biodiversity Data Journal can be downloaded in tabular format (CSV) straight from the article text and re-used by anyone, provided that the original source is cited (**Fig. 8**). Upon publication, the primary biodiversity data (for example, species occurrence records, species descriptions and taxon checklists) are also automatically exported into machine-readable Darwin Core Archives and become available for harvesting and indexing by aggregators (**Fig. 8**). Furthermore, species occurrences are indexed and made available as a separate dataset in GBIF bearing the article's DOI (**Fig. 9**) which increases the visibility and citation probability of both the article and the underlying data.

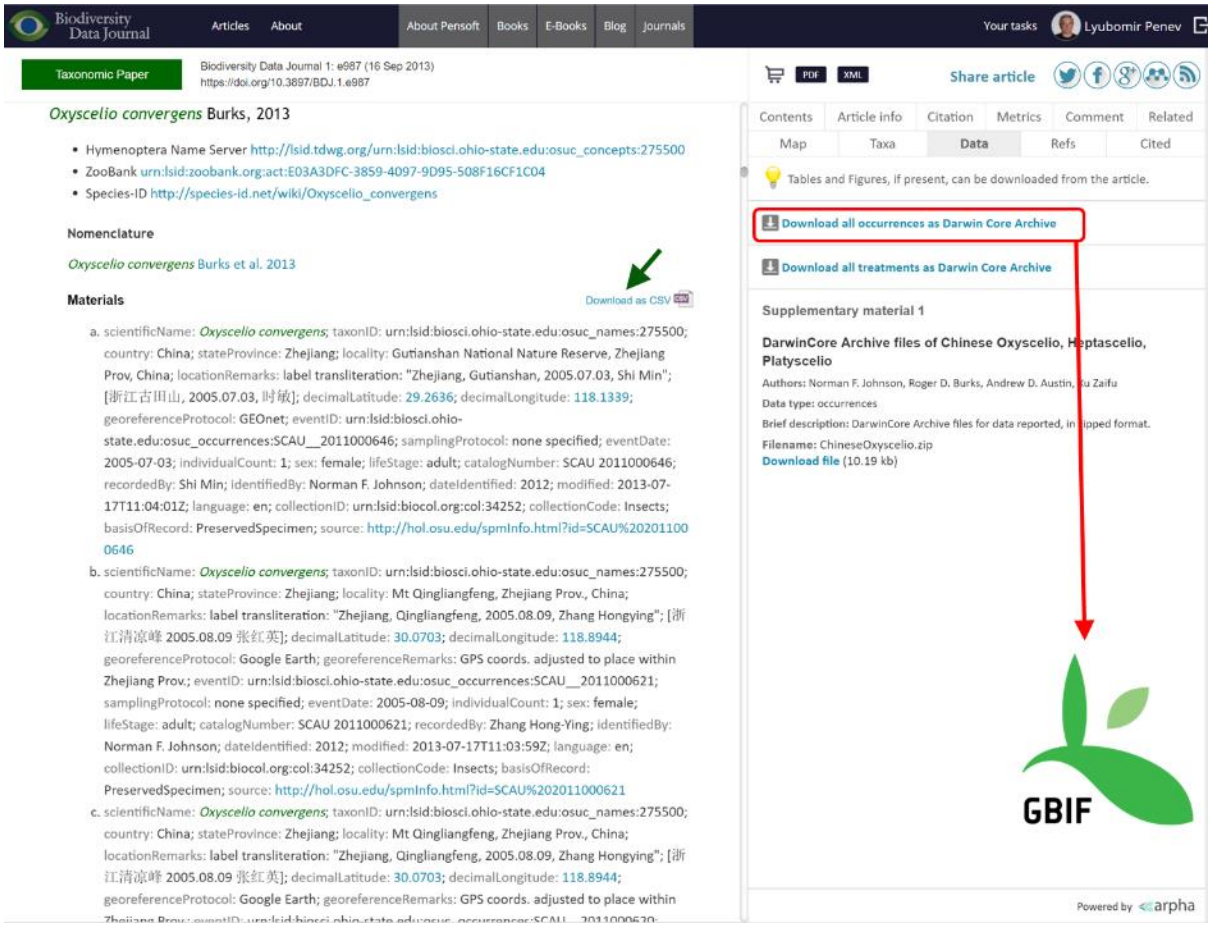

Figure 8. Export of data from articles published in Biodiversity Data Journal. Species occurrences and other structured data tables can be downloaded in CSV format (green arrow); all species occurrences are also available as Darwin Core Archives and are automatically harvested and indexed by GBIF (red box and arrow).

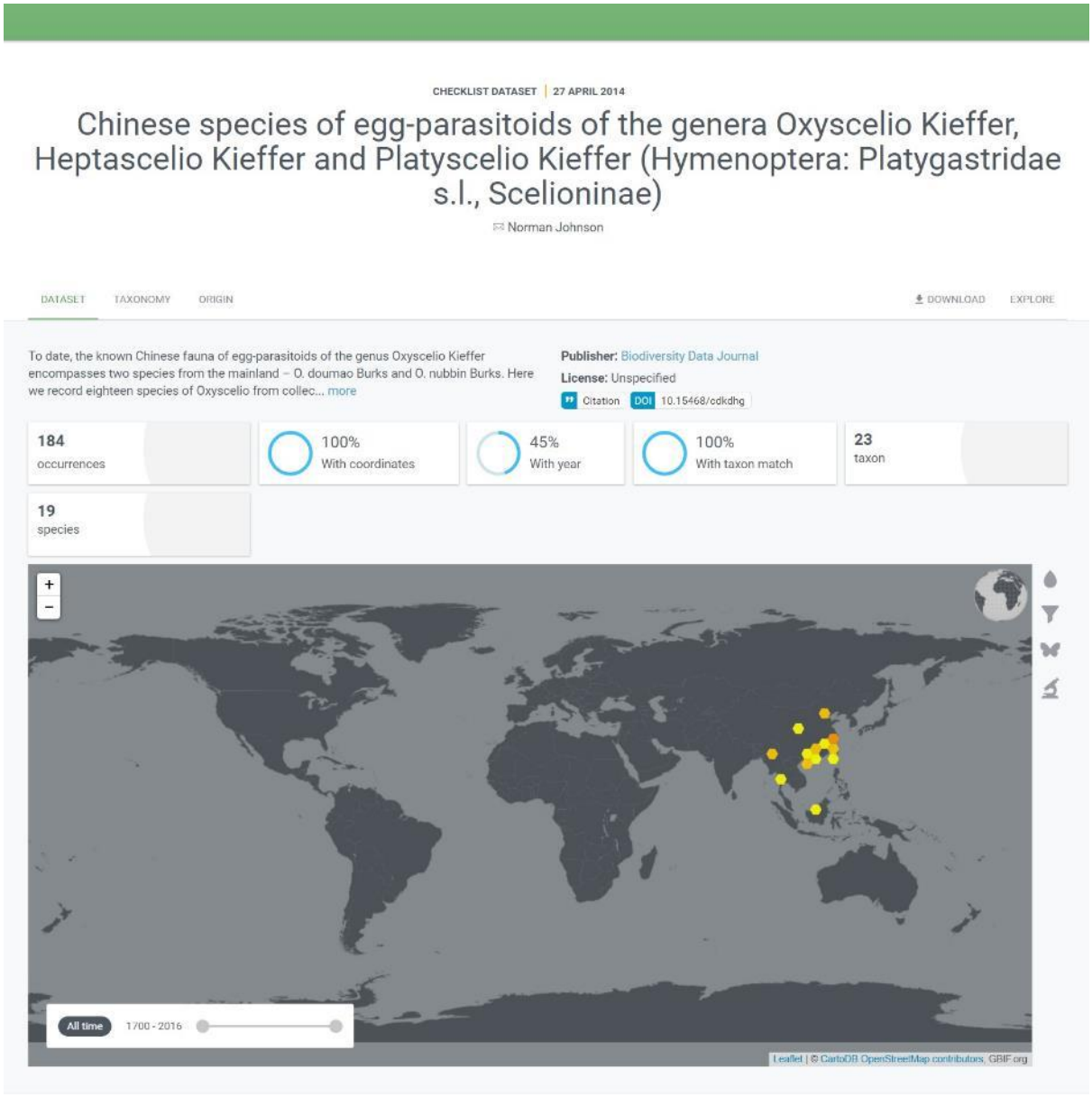

Figure 9. The occurrence data from articles published in the Biodiversity Data Journal (in this case from the paper of Johnson 2013) are automatically indexed via Darwin Core Archive in the GBIF Integrated Publishing Toolkit.

5.3 Data Extraction and Re-publishing Workflow

The present workflow has been created and tested with three different book titles with the support of the EU-funded projects pro-iBiospehere, SCALES and EU BON. It resulted in the launch of the Advanced Books platform of Pensoft, designed to (re-)publish historical or new books in semantically enhanced open access. The workflow is illustrated on the main homepage of Advanced Books at http://ab.pensoft.net and in **Fig. 10**. Of particular interest was the text and data extraction and conversion to XML of the historical book of Winch (1831) *Flora of Northumberland and Durham. Trans. Nat. Hist. Soc. Northumberl., Durham and Newcastle upon Tyne 2: 1-149. Printed by T. and J. Hodgson.* The data extraction and conversion has been processed by Quentin Groom from the Botanical Garden Meise, Belgium. The source document was scanned by Ernst Mayr Library of Harvard University for the Biodiversity Heritage Library. The digitised text was uploaded

to Wikisource, where it was proofread. The corrected text was then marked-up into XML and semantically enhanced with additional details, including links to the original citations and coordinates of the mentioned localities (for details, see Pensoft blog).

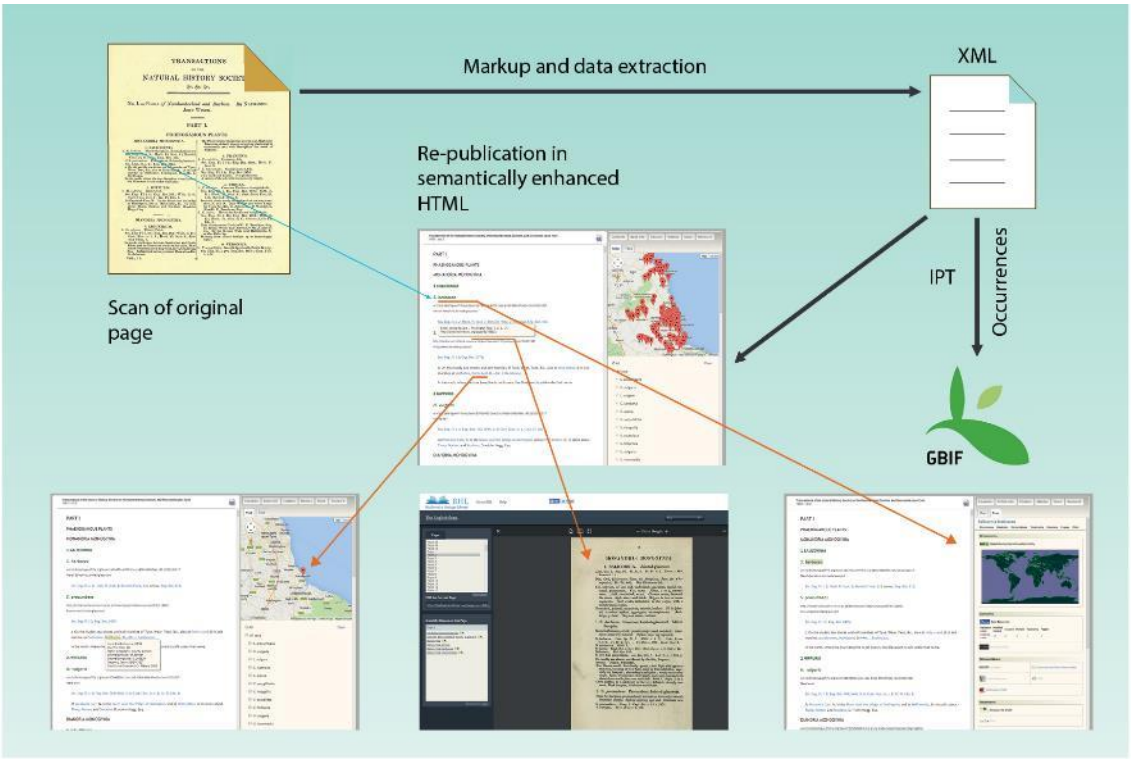

Figure 10. Data extraction and re-publishing workflow of the Advanced Books platform

6. Submission of Manuscripts through an Application Programming Interface (API)

A distinct feature of the ARPHA-XML publishing workflow is the possibility to import complex manuscripts, including metadata, text figures, tables, references, citations and others, via an API available in ARPHA Writing Tool (**Fig. 11**, documentation at http://arpha.pensoft.net/dev/). A working example of the workflow is described in the next section.

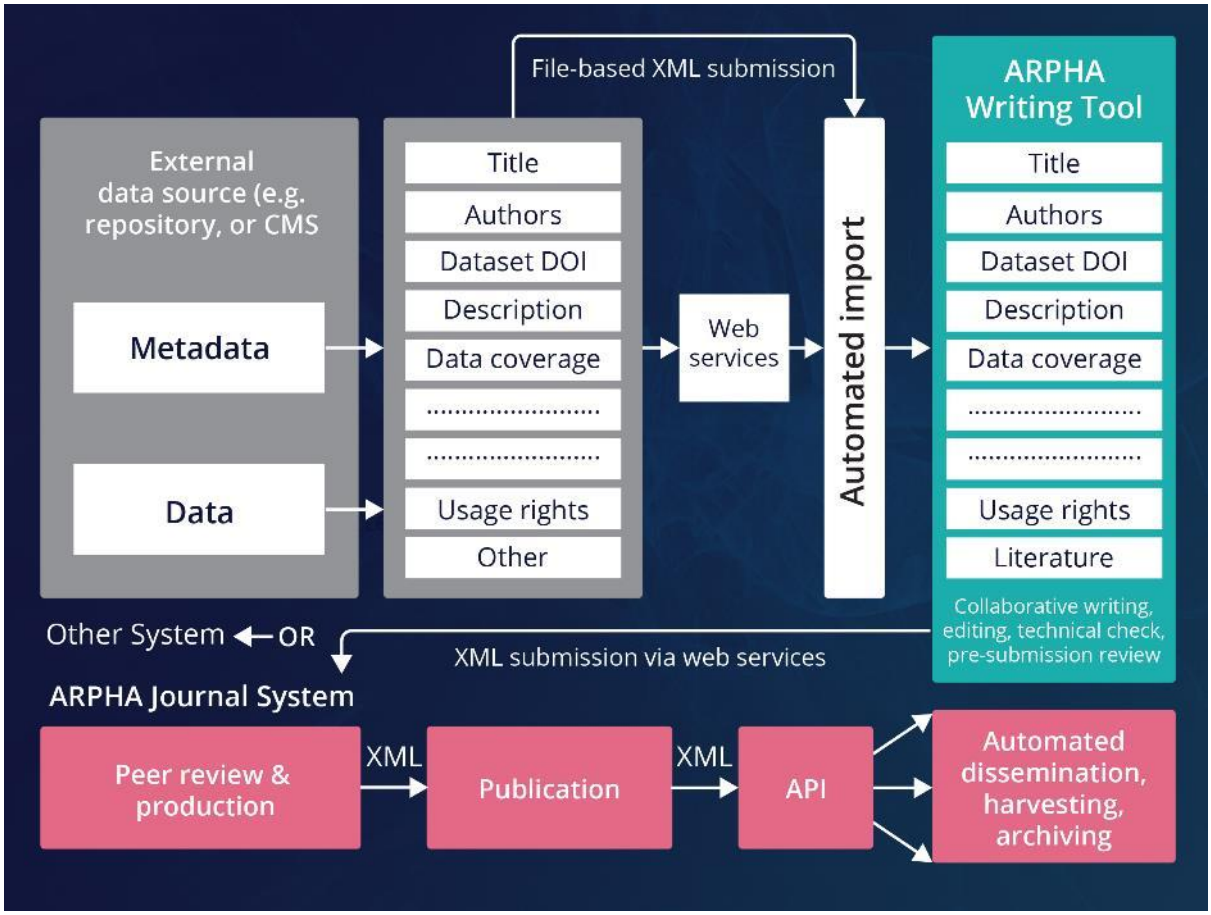

Figure 11. Submission of manuscripts to ARPHA Writing Tool through Application Programming Interface (API).

In order to submit an article via the Pensoft RESTful API, one has first to prepare an XML file according to the Pensoft XML schemas or according to the Ecological Metadata Language (EML) standard (information listed in the link above). An authentication token is obtained from the settings dialogue in the ARPHA-BioDiv which is supplied together with the XML file to the endpoint. If the document is imported successfully, it is created in the respective journal's ARPHA Writing Tool instance, where it can be further edited manually and submitted to the journal.

7. Creation and Publication of Data Papers from Ecological Metadata Language (EML) Metadata

Data papers, often called also “data articles”, “data notes”, or similar, were first established by the journals Ecological Archives (published by the Ecological Society of America) and Earth System Science Data (ESSD) (published by Copernicus) (see Newman and Corke 2009, Chavan and Penev 2011). According to the definition of Chavan and Penev 2011, data papers are “scholarly publications whose primary purpose is to describe data, rather than report a research investigation. As such, data papers contain facts about data, not hypotheses and arguments in support of those hypotheses based on data, as found in a conventional research article. Their purposes are threefold: to provide a citable journal publication that brings scholarly credit to data publishers; to describe the data in a structured

human-readable form; and to bring the existence of the data to the attention of the scholarly community."

The data paper should include several important elements (usually called metadata, or "description of data"), for example:

- Title, authors and abstract;
- Project description;
- Methods of data collection;
- Spatial and temporal ranges and geographical coverage;
- Collectors and owners of the data;
- Data usage rights and licences;
- Software used to create or view the data.

These metadata, if available and deliverable in machine-readable form (XML, JSON, etc.), can be used to produce a "data paper manuscript" that can be submitted to a journal for peer review and publication. The ARPHA approach to data paper publishing was first demonstrated in 2010 in a joint project of the Global Biodiversity Information Facility (GBIF) and Pensoft. As a result, this partnership created a workflow (**Fig. 12**) between the GBIF's Integrated Publishing Toolkit (IPT) and Pensoft's journals (ZooKeys, Phytokeys, Nature Conservation and others). A special module at IPT generates data paper manuscripts into RTF files from the extended metadata descriptions automatically, at the click of a button. Thereafter, manuscripts can be submitted to a journal for peer review and publication. After publication, the data paper's DOI is linked back to the dataset's DOI at IPT. In less than three years, more than 100 data papers have been published in Pensoft journals this way (for examples, see the Data paper subsection above).

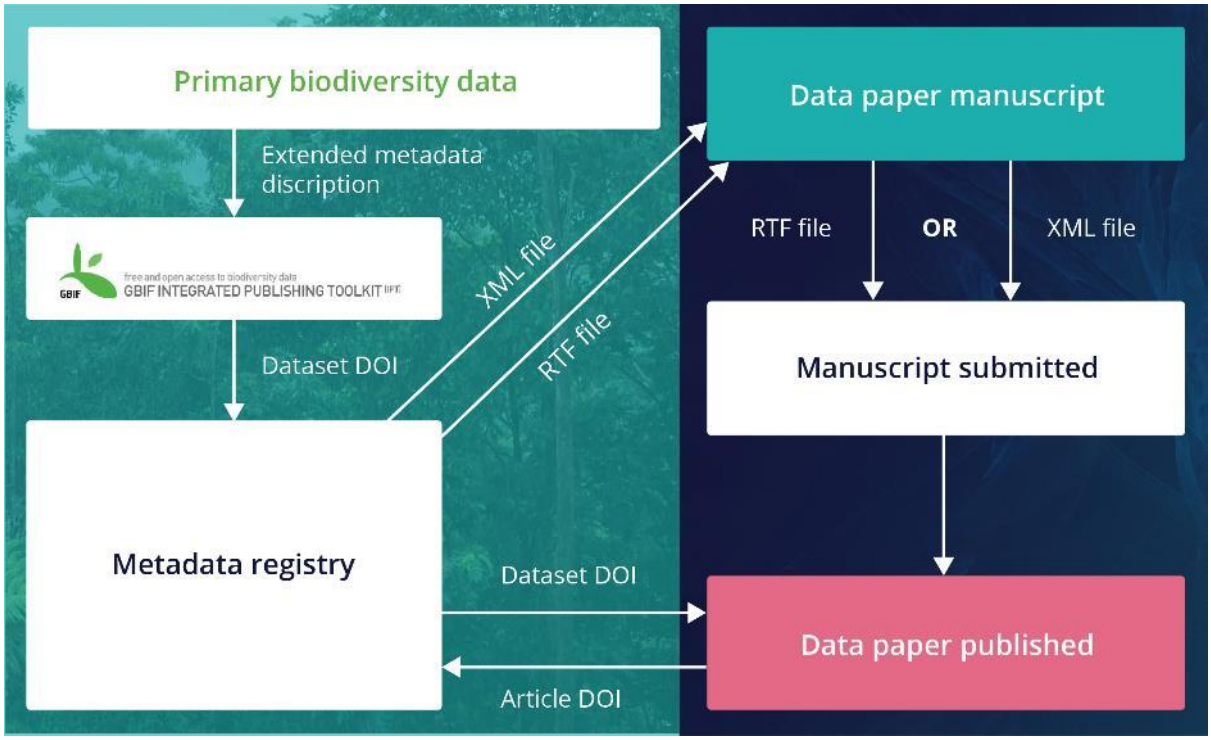

Figure 12. Creation of data paper manuscripts from Ecological Metadata Language (EML) metadata hosted at the GBIF IPT

Recently, the workflow was amended by a direct import functionality of EML metadata downloadable from GBIF, LTER and DatONE networks on to a data paper manuscript in ARPHA Writing Tool (Senderov et al. 2016, Penev et al. 2017, see also **Fig. 13**). The workflow has been thoroughly described in a blog post, while stepwise instructions are available via ARPHA's Tips and tricks guidelines.

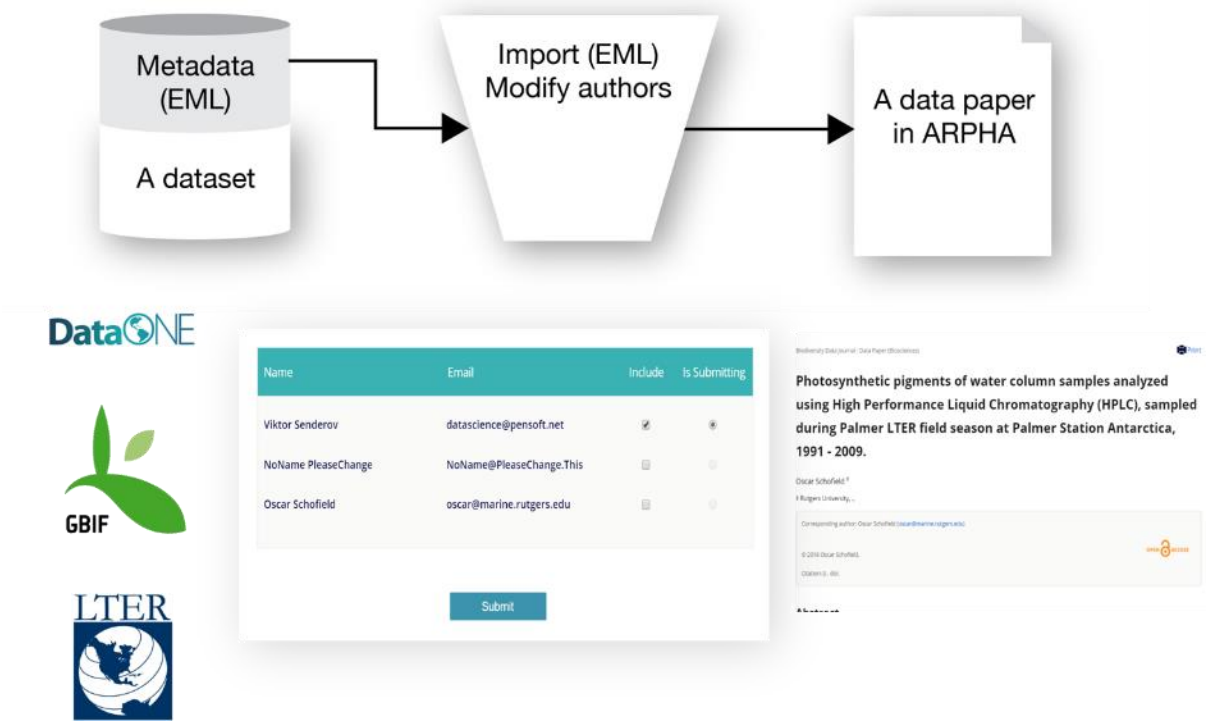

Figure 13. Conversion of Ecological Metadata Language (EML) metadata into data paper manuscripts in ARPHA Writing Tool.

8. Use Cases

The ARPHA-BioDiv toolbox has been developed in the course of several years and its tools, workflows and journals are used routinely by thousands of authors, reviewers, editors and readers worldwide. It is virtually impossible to list here the numeroususe cases and approaches that have been tested and succesfully implemented over the years (see Penev 2017 and Penev et al. 2017 for review). Below we describe three publishing use cases that have been elaborated during the EU BON project.

8.1 Expert and Data Mobilisation through the Fauna Europaea Special Issue

One of the major data mobilisation initiatives realised by ARPHA and the Biodiversity Data Journal is the publication of data papers on the largest European animal database 'Fauna Europaea' within a new series "Contributions on Fauna Europaea", launched in 2014. This novel publication model was aimed at assembling in a single collection data papers on different taxonomic groups of higher rank covered by the Fauna Europaea project and accompanying papers highlighting various aspects of this project (gap-analysis, design, taxonomic assessments etc.) (Jong et al. 2014). Altogether, eleven artciles have been published so far.

8.2 Expert and Data Mobilisation through the LifeWatchGreece Special Issue

The LifeWatchGreece special collection LifeWatchGreece: Research infrastructure (ESFRI) for biodiversity data and data observatories was published in the Biodiversity Data Journal and currently contains twenty-three papers organised in four sections:(1) Electronic infrastructure and software applications; (2) Taxonomic checklists; (3) Data papers and (4) Research articles (Arvantidis et al. 2016). The Biodiversity Data Journal was chosen because it is a "community peer-reviewed, open access, comprehensive online platform for publishing part of the up-to-date outcomes of LifeWatchGreece and enables the publication of a wide variety of papers (e.g. software descriptions, data papers, taxonomic checklists and research articles) along with the accompanying datasets and supporting material" (Arvantidis et al. 2016).

8.3 EU BON Open Science Collection in RIO Journal

The journal Research Ideas and Outcomes (RIO) was designed to publish all outputs of the research cycle, from research ideas and grant proposals to data, software, research articles and research collaterals, such as workshop and project reports, guidelines, policy briefs, Wikipedia articles and others (Mietchen et al. 2015). In the RIO Journal, EU BON realised one of the first ever open science collections of publications, entitled Building the European Biodiversity Observation Network (EU BON) Project Outcomes. To date, the collection contains 15 publications.

9. Guidelines, Policies and Licences for Scholarly Publishing of Biodiversity Data

9.1 Legal Framework and Policies

The legal framework and policies for publishing and re-use of biodiversity data is a subject of primary interest to the biodiversity community and policy-makers. Several EU BON teams and tasks worked on various aspects of the subject which resulted in the following set of documents:

- Open Exchange of Scientific Knowledge and European Copyright: The Case of Biodiversity Information (Egloff et al. 2014)
- EU BON Policy Brief on Open Data (Egloff et al. 2015)
- Biodiversity Data Publishing Legal Framework Report (Milestone MS841), published as a supplementary file 1 to Egloff et al. (2016a) Egloff et al. (2016b)
- Data Sharing Agreement (Milestone MS971), published as a supplementary file 2 to Egloff et al. (2016b)
- Data Policy Recommendations for Biodiversity Data (Milestone MS972) (Egloff et al. 2016b)
- Section "Data Publishing Licenses" within the *Data Publishing Strategies and Guidelines for Biodiversity Data* paper (milestone MS842) (Penev et al. 2017)

The last two documents summarise the effort and can serve as guidelines and recommendations in the work Group on Earth Observation's Biodiversity Observation Network (GEO BON) and beyond.

The paper of Egloff et al. (2016b) was published as part of the data publishing recommendations in the EU BON Biodiversity Portal. However, its importance goes far beyond EU BON. The document deals with the following issues: (i) Mobilising biodiversity data, (ii) Removing legal

obstacles, (iii) Changing attitudes and (iv) Data policy recommendations. It is targeted at legislators, researchers, research institutions, data aggregators, funders and publishers.

9.2 Licences for publishing and re-use

This section from the paper of Penev et al. (2017) builds on the fundamental principles of open data publishing and re-use, known as Panton Principles and their biodiversity-specific interpretation in the Bouchout Declaration for Open Biodiversity Knowledge Management. The document is supported by a wide range of previously published research and review papers, as well as the data publishing practices of Pensoft and other publishers (Penev et al. 2011a, Hagedorn et al. 2011, Egloff et al. 2014).

The recommended data publishing licence used by Pensoft is the Open Data Commons Attribution License (ODC-By), which is a licence agreement intended to allow users to freely share, modify and use the published data(base), provided that the data creators are attributed (cited or acknowledged). This ensures that those who publish their data receive the academic credit that is due.

Alternatively, other licences, namely the Creative Commons CC0 (also cited as “CC-Zero” or “CC-zero”) and the Open Data Commons Public Domain Dedication and Licence (PDDL), are also STRONGLY encouraged for use in the Pensoft journals. According to the CC0 licence, "the person who associated a work with this deed has dedicated the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighbouring rights, to the extent allowed by law. You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission."

9.3 Strategies and Guidelines for Scholarly Publishing

The S*trategies and Guidelines for Scholarly Publishing of Biodiversity Data* (Penev et al. 2017) have been elaborated during the Framework Program 7 EU BON project on the basis of an earlier version published on Pensoft's website in 2011 (Penev et al. 2011a). The document discusses some general concepts, including a definition of datasets, incentives to publish data and licences for data publishing. Further, it defines and compares several routes for data publishing, namely as (1) supplementary files to research articles which may be made available directly by the publisher or (2) published in a specialised open data repository with a link to it from the research article or (3) as a Data Paper, i.e. a specific, stand-alone publication describing a particular dataset or a collection of datasets or (4) integrated data publishing through online import/download of data into/from manuscripts, as provided by the ARPHA Writing tool and its associated journals (Biodiversity Data Journal, RIO Journal, One Ecosystem).

The paper also contains detailed instructions on how to prepare and peer review data intended for publication, listed under the Guidelines for Authors and Reviewers, respectively. Special attention is given to existing standards, protocols and tools to facilitate data publishing, such as the GBIF Integrated Publishing Toolkit (IPT) and the DarwinCore Archive (DwC-A).

Here, we include the table of contents of the document which will give the reader a comprehensive overview of its content (Penev et al. 2017):

- Data Publishing in a Nutshell
    - Introduction
    - What Is a Dataset
    - Why Publish Data

    - How to Publish Data
    - How to Cite Data
- Data Publishing Policies
    - Data Publishing Licences
- Open Data Repositories
- Guidelines for Authors
    - Data Published within Supplementary Information Files
    - Import of Darwin Core Specimen Records into Manuscripts
    - Data Published in Data Papers
    - Data Papers Describing Primary Biodiversity Data
    - Data Papers Describing Ecological and Environmental Data
    - Data Papers Describing Genomic Data
    - Software Description Papers
- Guidelines for Reviewers
    - Quality of the Manuscript
    - Quality of the Data
    - Consistency between Manuscript and Data

The *Strategies and Guidelines* are referred to in the Author Guidelines of Pensoft's journals and are used in their everyday publishing practices.

9.4 Tutorials, Manuals and Supporting documentation

The current article describes the rationale, overall structure and the key elements of ARPHA-BioDiv. The various elements of ARPHA-BioDiv have been featured in several papers (cited in the respective sections of the present document), guidelines, blog posts and tutorials. Below, some important supporting documentation are listed to assist the users to access this complex system.

- Overall description of ARPHA-BioDiv: ARPHA website (http://arphahub.com) and current paper.
- Guidelines for scholarly publishing of biodiversity data: Penev et al. (2017).
- Guidelines for authors, reviewers and editors: see each journal's webpage from http://journals.pensoft.net and also Penev et al. (2017).
- Stepwise welcome tutorial for ARPHA Writing Tool: avaiiable from the AWT menu; appears automatically to all first-time users.
- Tips and Tricks guidelines for ARPHA Writing Tool: availabe from the AWT menu.
- Promotional video for ARPHA
- Promotional video for RIO Journal
- Posts at Pensoft blog:
    - Introducing the ARPHA Writing Tool

- How to import data papers from GBIF, DataONE and LTER metadata
- How to import occurrence records into manuscripts from GBIF, BOLD, iDigBio and PlutoF
- In a nutshell: The four peer review stages in RIO explained
- The 5 Most Distinct Features of ARPHA
- We ask. We listen. We innovate!
- Faster, Better, Stronger: New batch of updates now available in ARPHA Writing Tool
- Guidelines for scholarly publishing of biodiversity data from Pensoft and EU BON

10. Future of ARPHA-BioDiv

In the future, we want to reimagine and reinvent the academic publishing process. At the dawn of academic publishing, papers had been written especially for human consumption. The human mind alone was expected to crunch the data. Now humans rely on computers to store and manipulate the data and verify the correctness of numerical algorithms, whereas our minds focus on the big picture and the story behind the data.

With ARPHA-BioDiv, we have already taken the first few steps in creating articles that can be read both by humans and computers, as has been described so far in this article. However, more can be done. One area of innovation in academic publishing lies in creating linked content - embedding machine-readable database records in each publication that are linked to the world-wide network of linked knowledge hubs. To achieve this goal, we are currently working towards exporting content that has been semantically enriched in a knowledge graph called the Open Biodiversity Knowledge Management System or OpenBioDiv for short (pro-iBiosphere 2014, Senderov and Penev 2016).

This will enable the reader of an aritcle, for example, to connect published occurrence data to portals such as GBIF and geographic repositories such as GeoNames. An illustration of the use-value of this integration will be, for example, an accelerated creation of various models, such as species distribution models, based on the article data. Thanks to the linking of the occurrence data in the article to databases, it will be possible to assemble all the elements needed for a species distribution model of the discussed taxon programmatically in an environment such as R. Moreover, the links in themselves are valuable information and can point to "hot" topics, such as "hot" taxa or "hot" figures, having many incoming links to them (Page 2016). Or, the user may choose to investigate the genetics of the taxon, the occurrence of which they had just seen, through a link to GenBank.

We also believe that a large portion of tradional academic publishing, even if enriched with Linked Data, will be supplemented by nano-publications (Groth et al. 2010, Mons et al. 2011, Chichester 2013). More and more academic research reveals stories and data that cannot be published in the traditional seven-figure-paper. Imagine that the research team you are leading has just discovered 500,000 gene-disease associations across the genome of an important domestic animal. You want all of these findings to be first class research objects - with DOIs, just as publications - and not to be relegated only to a database record that can be altered or deleted. Towards this goal, we are working on nano-publications: first class research objects with DOIs and metadata including author, publisher, etc. which are published as a regular publication, but nevertheless formatted primarily as a machine-readable fact that can be ingested by a database without any alterations.

Finally, we believe that publishers are stewards of the worlds' scientific information and there is knowledge in the totality of the published articles that is not part of any article alone. We are working on artifical intelligence algorithms both from the machine logic domain and from the machine learning domain to discover this hidden knowledge. The authors of tomorrow will have at their disposal not only a tool to format their manuscript, add citations and mark-up their data, but also tools that will discover additional information relevant to the authors' ideas and suggest similar research during the authoring phase. And, if we can dream very big, why not have artificial intelligence algorithms sophisticated enough to act as a research assistant during the authoring phase? What a marvelous thought!

11. 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

Agosti D, Egloff W (2009) Taxonomic information exchange and copyright: the Plazi approach. BMC research notes 2: 53. https://doi.org/10.1186/1756-0500-2-53

Altman M, King D (2007) A Proposed Standard for the Scholarly Citation of Quantitative Data. D-lIb Magazine 13 (3): 1.

Arvanitidis C, Chatzinikolaou E, Gerovasileiou V, Panteri E, Bailly N, Minadakis N, Hardisty A, Los W (2016) LifeWatchGreece: Construction and operation of the National Research Infrastructure (ESFRI). Biodiversity Data Journal 4: e10791. https://doi.org/10.3897/bdj.4.e10791

Baker E, Rycroft S, Smith V (2014) Linking multiple biodiversity informatics platforms with Darwin Core Archives. Biodiversity Data Journal 2: e1039. https://doi.org/10.3897/bdj.2.e1039

BioMed Central (2010) BioMed Central's Position Statement on Open Data. https://blogs.biomedcentral.com/wp-content/image_archive/opendatastatementdraft.pdf

Brown B, Hash J, Hartop E, Porras W, Amorim D (2017) Baby Killers: Documentation and Evolution of Scuttle Fly (Diptera: Phoridae) Parasitism of Ant (Hymenoptera: Formicidae) Brood. Biodiversity Data Journal 5: e11277. https://doi.org/10.3897/bdj.5.e11277

Cardoso P, Stoev P, Georgiev T, Senderov V, Penev L (2016) Species Conservation Profiles compliant with the IUCN Red List of Threatened Species. Biodiversity Data Journal 4: e10356. https://doi.org/10.3897/bdj.4.e10356

Catapano T (2010) TaxPub: An extension of the NLM/NCBI Journal Publishing DTD for taxonomic descriptions . Proceedings of the Journal Article Tag Suite Conference. URL: http://www.ncbi.nlm.nih.gov/books/NBK47081/

Chavan V, Penev L (2011) The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC Bioinformatics 12: S2. https://doi.org/10.1186/1471-2105-12-s15-s2

Chavan VS, Ingwersen P (2009) Towards a data publishing framework for primary biodiversity data: challenges and potentials for the biodiversity informatics community. BMC Bioinformatics 10: S2. https://doi.org/10.1186/1471-2105-10-s14-s2

Chichester C (2013) Open_PHACTS and NanoPublications. EMBnet.journal 19: 17. https://doi.org/10.14806/ej.19.b.746

CODATA/ITSCI (2013) Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data. Data Science Journal 12: CIDCR1-CIDCR75. https://doi.org/10.2481/dsj.osom13-043

Costello M (2009) Motivating Online Publication of Data. BioScience 59 (5): 418-427. https://doi.org/10.1525/bio.2009.59.5.9

Costello M, Michener W, Gahegan M, Zhang Z, Bourne P (2013) Biodiversity data should be published, cited, and peer reviewed. Trends in Ecology & Evolution 28 (8): 454-461. https://doi.org/10.1016/j.tree.2013.05.002

Edmunds SC, Hunter CI, Smith V, Stoev P, Penev L (2013) Biodiversity research in the “big data” era: GigaScience and Pensoft work together to publish the most data-rich species description. GigaScience 2 (1): . https://doi.org/10.1186/2047-217x-2-14

Edmunds SC, Li P, Hunter CI, Xiao SZ, Davidson RL, Nogoy N, Goodman L (2016) Experiences in integrated data and research object publishing using GigaDB. International Journal on Digital Libraries https://doi.org/10.1007/s00799-016-0174-6

Egloff W, Agosti D, Patterson D, Penev L (2015) Eu Bon Policy Brief On Open Data. Zenodo https://doi.org/10.5281/ZENODO.188391

Egloff W, Patterson D, Agosti D, Hagedorn G (2014) Open exchange of scientific knowledge and European copyright: The case of biodiversity information. ZooKeys 414: 109-135. https://doi.org/10.3897/zookeys.414.7717

Egloff W, Agosti D, Kishor P, Patterson D, Miller J (2016a) Copyright and the Use of Images as Biodiversity Data. bioRxiv https://doi.org/10.1101/087015

Egloff W, Agosti D, Patterson D, Hoffmann A, Mietchen D, Kishor P, Penev L (2016b) Data Policy Recommendations for Biodiversity Data. EU BON Project Report. Research Ideas and Outcomes 2: e8458. https://doi.org/10.3897/rio.2.e8458

Frolov A, Akhmetova L (2013) A contribution to the study of the Lower Volga center of scarab beetle diversity in Russia: checklist of the tribe Aphodiini (Coleoptera, Scarabaeidae) of Dosang environs. Biodiversity Data Journal 1: e979. https://doi.org/10.3897/bdj.1.e979

Green T (2009) We Need Publishing Standards for Datasets and Data Tables. OECD Publishing White Paper 1: 1. https://doi.org/10.1787/603233448430

Groth P, Gibson A, Velterop J (2010) The Anatomy of a Nanopublication. Inf. Serv. Use 30 (1-2): 51-56. URL: http://dl.acm.org/citation.cfm?id=1883685.1883690

Hagedorn G, Mietchen D, Morris R, Agosti D, Penev L, Berendsohn W, Hobern D (2011) Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information. ZooKeys 150: 127-149. https://doi.org/10.3897/zookeys.150.2189

Hardisty A, Roberts D, Informatics Community TB (2013) A decadal view of biodiversity informatics: challenges and priorities. BMC Ecology 13 (1): 16. https://doi.org/10.1186/1472-6785-13-16

Heller K, Jong Yd, Bohn H, Haas F, Willemse F (2016) Fauna Europaea – Orthopteroid orders. Biodiversity Data Journal 4: e8905. https://doi.org/10.3897/bdj.4.e8905

Hoekstra PH, Wieringa JJ, Chatrou LW (2016) A nonet of novel species of Monanthotaxis (Annonaceae) from around Africa. PhytoKeys 69: 71-103. https://doi.org/10.3897/phytokeys.69.9292

Johnson N (2013) Chinese species of egg-parasitoids of the genera Oxyscelio Kieffer, Heptascelio Kieffer and Platyscelio Kieffer (Hymenoptera: Platygastridae s. l., Scelioninae). Biodiversity Data Journal 1 https://doi.org/10.3897/BDJ.1.e987

Jong Yd, Verbeek M, Michelsen V, Place Bjørn Pd, Los W, Steeman F, Bailly N, Basire C, Chylarecki P, Stloukal E, Hagedorn G, Wetzel F, Glöckler F, Kroupa A, Korb G, Hoffmann A, Häuser C, Kohlbecker A, Müller A, Güntsch A, Stoev P, Penev L (2014) Fauna Europaea – all European animal species on the web. Biodiversity Data Journal 2: e4034. https://doi.org/10.3897/bdj.2.e4034

Klump J (2011) Criteria for the Trustworthiness of Data Centres. D-Lib Magazine 17 (1): 0. https://doi.org/doi:10.1045/january2011-klump

Knuth DE (1984) Literate Programming. The Computer Journal 27 (2): 97-111. https://doi.org/10.1093/comjnl/27.2.97

Martone, M (Ed.) (2014) Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. San Diego CA: FORCE11. https://www.force11.org/group/joint-declaration-data-citation- principles-final

McQuilton P, Gonzalez-Beltran A, Rocca-Serra P, Thurston M, Lister A, Maguire E, Sansone S (2016) BioSharing: curated and crowd-sourced metadata standards, databases and data policies in the life sciences. Database: the journal of biological databases and curation 2016 https://doi.org/10.1093/database/baw075

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

Mons B, Haagen Hv, Chichester C, Hoen P', den Dunnen JT, Ommen Gv, Mulligen Ev, Singh B, Hooft R, Roos M, Hammond J, Kiesel B, Giardine B, Velterop J, Groth P, Schultes E (2011) The value of data. Nature Genetics 43 (4): 281-283. https://doi.org/10.1038/ng0411-281

Newman P, Corke P (2009) Editorial. The International Journal of Robotics Research 28 (5): 587-587. https://doi.org/10.1177/0278364909104283

Page R (2016) Towards a biodiversity knowledge graph. Research Ideas and Outcomes 2: e8767. https://doi.org/10.3897/rio.2.e8767

Penev L (2017) From Open Access to Open Science from the viewpoint of a scholarly publisher. Research Ideas and Outcomes 3: e12265. https://doi.org/10.3897/rio.3.e12265

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 . Journal Article Tag Suite Conference (JATS-Con) Proceedings. National Center for Biotechnology Information (US, Bethesda (MD) [In en]. URL: http://www.ncbi.nlm.nih.gov/books/NBK100351/

Penev L, Roberts D, Smith V, Agosti D, Erwin T (2010) Taxonomy shifts up a gear: New publishing tools to accelerate biodiversity research. ZooKeys 50: 1-4. https://doi.org/10.3897/zookeys.50.543

Penev L, Erwin T, Miller J, Chavan V, Moritz T, Griswold C (2009) Publication and dissemination of datasets in taxonomy: ZooKeys working example. ZooKeys 11: 1-8. https://doi.org/10.3897/zookeys.11.210

Penev L, Mietchen D, Chavan V, Hagedorn G, Remsen D, Smith V, Shotton D (2011) Pensoft Data Publishing Policies and Guidelines for Biodiversity Data. Zenodo 1: 1-34. https://doi.org/10.5281/zenodo.56660

Penev L, Sharkey M, Erwin T, Noort Sv, Buffington M, Seltmann K, Johnson N, Taylor M, Thompson F, Dallwitz M (2009) Data publication and dissemination of interactive keys under the open access model. ZooKeys 21: 1-17. https://doi.org/10.3897/zookeys.21.274

Penev L, Hagedorn G, Mietchen D, Georgiev T, Stoev P, Sautter G, Agosti D, Plank A, Balke M, Hendrich L, Erwin T (2011) Interlinking journal and wiki publications through joint citation: Working examples from ZooKeys and Plazi on Species-ID. ZooKeys 90: 1-12. https://doi.org/10.3897/zookeys.90.1369

Penev L, Mietchen D, Chavan V, Hagedorn G, Smith V, Shotton D, Tuama ÉÓ, Senderov V, Georgiev T, Stoev P, Groom Q, Remsen D, Edmunds S (2017) Strategies and guidelines for scholarly publishing of biodiversity data. Research Ideas and Outcomes 3: e12431. https://doi.org/10.3897/rio.3.e12431

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

pro-iBiosphere (2014) Open Biodiversity Knowledge Management System (OBKMS). Zenodo https://doi.org/10.5281/ZENODO.191785

Ratnasingham S, Hebert PN (2007) BARCODING: BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7 (3): 355-364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Rauber A, Asmi A, Uytvanck Dv, Proell S (2016) Data Citation of Evolving Data: Recommendations of the RDA Working Group on Data Citation (WGDC). Research Data Alliance https://doi.org/10.15497/RDA00016

Robertson T, Döring M, Guralnick R, Bloom D, Wieczorek J, Braak K, Otegui J, Russell L, Desmet P (2014) The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet. PLoS ONE 9 (8): e102623. https://doi.org/10.1371/journal.pone.0102623

Rougerie R, Lopez-Vaamonde C, Barnouin T, Delnatte J, Moulin N, Noblecourt T, Nusillard B, Parmain G, Soldati F, Bouget C (2015) PASSIFOR: A reference library of DNA barcodes for French saproxylic beetles (Insecta, Coleoptera). Biodiversity Data Journal 3: e4078. https://doi.org/10.3897/bdj.3.e4078

Sansone S, Sansone S, Field D, Santarsiero A, Maguire E, Rocca-Serra P, Taylor C, Harland L, Communities TB (2011) BioSharing Overview. Nature Precedings URL: http://dx.doi.org/10.1038/npre.2011.5936.1

Schindel D, Miller S, Trizna M, Graham E, Crane A (2016) The Global Registry of Biodiversity Repositories: A Call for Community Curation. Biodiversity Data Journal 4: e10293. https://doi.org/10.3897/bdj.4.e10293

Schindel DE, Stoeckle MY, Milensky C, Trizna M, Schmidt B, Gebhard C, Graves G (2011) Project description: DNA barcodes of bird species in the national museum of natural history, smithsonian institution, USA. ZooKeys 152: 87-92. https://doi.org/10.3897/zookeys.152.2473

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

Senderov V, Georgiev T, Penev L (2016) Online direct import of specimen records into manuscripts and automatic creation of data papers from biological databases. Research Ideas and Outcomes 2: e10617. https://doi.org/10.3897/rio.2.e10617

Smith AM, Katz DS, Niemeyer KE, Working Group FSC (2016) Software Citation Principles. PeerJ Preprints URL: http://dx.doi.org/10.7287/PEERJ.PREPRINTS.2169V4

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

Smith VS (2009) Data publication: towards a database of everything. BMC Research Notes 2 (1): 113. https://doi.org/10.1186/1756-0500-2-113

Starr J, Castro E, Crosas M, Dumontier M, Downs R, Duerr R, Haak L, Haendel M, Herman I, Hodson S, Hourclé J, Kratz JE, Lin J, Nielsen LH, Nurnberger A, Proell S, Rauber A, Sacchi S, Smith A, Taylor M, Clark T (2015) Achieving human and machine accessibility of cited data in scholarly publications. PeerJ Computer Science 1: e1. https://doi.org/10.7717/peerj-cs.1

Stoev P, Komerički A, Akkari N, Liu S, Zhou X, Weigand A, Hostens J, Hunter C, Edmunds S, Porco D, Zapparoli M, Georgiev T, Mietchen D, Roberts D, Faulwetter S, Smith V, Penev L (2013) Eupolybothrus cavernicolus Komerički & Stoev sp. n. (Chilopoda: Lithobiomorpha: Lithobiidae): the first eukaryotic species description combining transcriptomic, DNA barcoding and micro-CT imaging data. Biodiversity Data Journal 1: e1013. https://doi.org/10.3897/bdj.1.e1013

Talamas E, Masner L, Johnson N (2011) Revision of the Malagasy genus Trichoteleia Kieffer (Hymenoptera, Platygastroidea, Platygastridae). ZooKeys 80: 1-126. https://doi.org/10.3897/zookeys.80.907

Thessen A, Patterson D (2011) Data issues in the life sciences. ZooKeys 150: 15-51. https://doi.org/10.3897/zookeys.150.1766

Veres SM, Adolfsson JP (2011) A natural language programming solution for executable papers. Procedia Computer Science 4: 678-687. https://doi.org/10.1016/j.procs.2011.04.071

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7 (1): e29715. https://doi.org/10.1371/journal.pone.0029715

Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, Silva Santos LBd, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J,

Heringa J, ’t Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, Schaik Rv, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, der Lei Jv, Mulligen Ev, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3: 160018. URL: http://dx.doi.org/10.1038/sdata.2016.18

Winch NJ (1831) Flora of Northumberland and Durham. Trans. Nat. Hist. Soc. Northumberl., Durham, and Newcastle upon Tyne 2: 1-149. [In Printed by T. and J. Hodgson].