A publication of the Society of Systematic Biologists

Enhanced monography in a collaboratively evolved hub for systematic biology

Jennifer C. Girón[1, 2], Eugenio Valderrama[3], Patrick M. O'Connor[4,5], Nancy B. Simmons[6], Deborah L. Paul[7], Matthew J. Yoder[7]

Published: 27 January 2022

Keywords: method; monograph; workflow; publishing; taxonomy

Affiliations:

[1]Department of Entomology, Purdue University, West Lafayette, IN, USA; [2]Natural Science Research Laboratory, Museum of Texas Tech University, Lubbock, TX, USA; [3]School of Integrative Plant Science, Section of Plant Biology and the L.H. Bailey Hortorium, Cornell University, Ithaca, NY, USA; [4]Department of Biomedical Sciences, Heritage College of Osteopathic Medicine; [5]Ohio Center for Ecological and Evolutionary Studies, Ohio University, Athens, OH, USA; [6]Department of Mammalogy, Division of Vertebrate Zoology, American Museum of Natural History, New York, NY, USA; [7]Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, IL, USA.

Correspondence:

Jennifer Girón

Email: entiminae@gmail.com

1 INTRODUCTION

Monographs are comprehensive, integrative summaries of information on a particular subject. Their authors undertake multiple, highly iterative steps to find, integrate and analyze information, and then present their conclusions to the scientific community. Here, in the spirit of previous work (e.g., Maxted 1992), we define a model that explains and characterizes the monographic process as it typically applies in systematic biology, and advocate for its use as a guiding framework on which new technical, communal, and financial support for monography can be built. Our model contains four kinds of actions: gathering, examination, rendering, and propagation (Fig. 1, Table 1), all of which are iterative and interconnected. We also call for the creation of a community of monograph workers wherein we can develop and share best practices to improve the reusability of the extensive information compiled in monographs and argue that establishment of a hub containing tools and other resources for monographers would greatly enhance both monographs and accessibility of the data and insights that they contain.

2 GATHERING

Gathering materials is typically the first step of any project, but it is also an activity that must be repeated numerous times in the process of building a monograph as the researcher learns more about the study organisms. Building

https://doi.org/10.18061/bssb.v1i1.8340

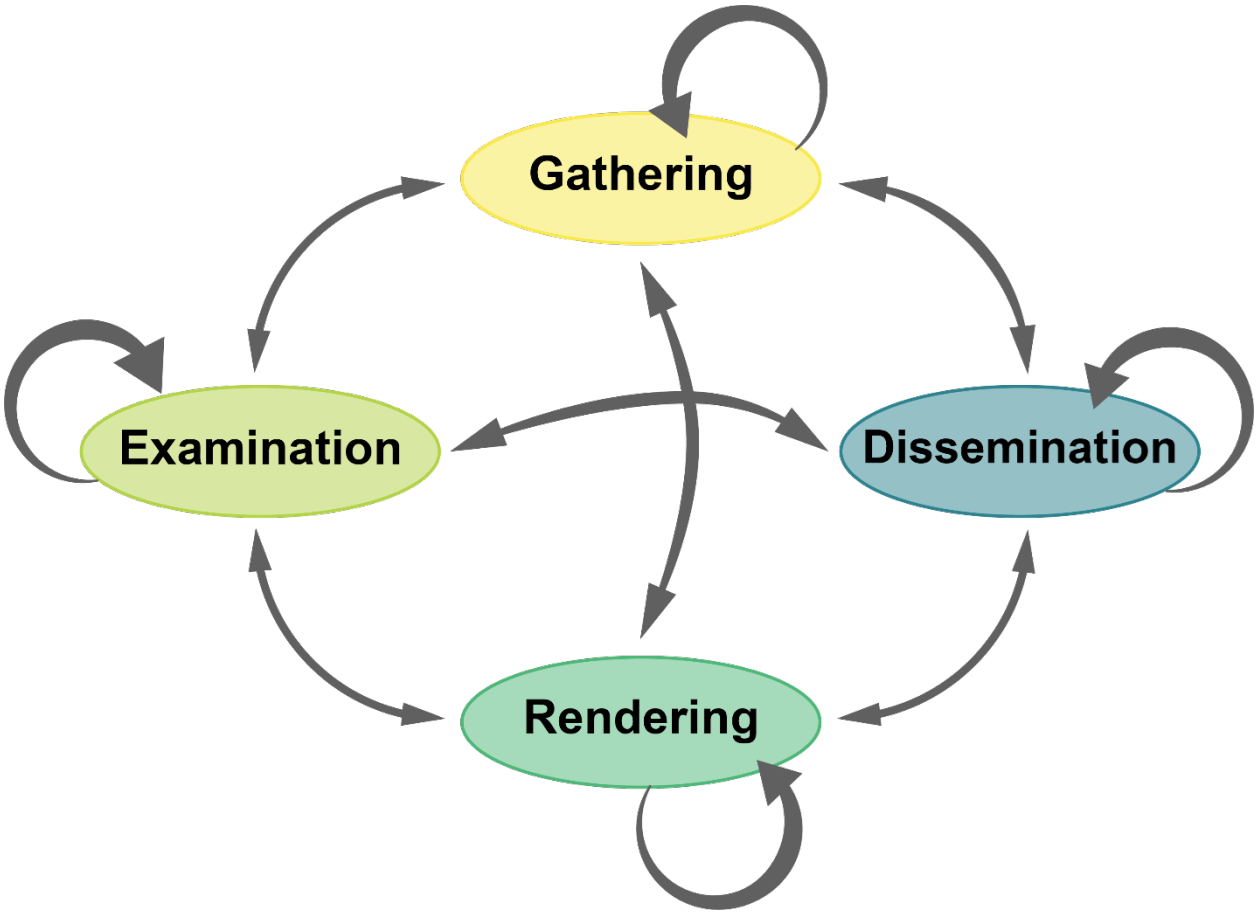

Figure 1. Schematic of the processes involved in the production of a monograph in systematic biology highlighting the iterative interactions between four primary processes: gathering, examination, rendering, and dissemination. **Gathering** includes bringing together literature, specimens, and existing data (e.g., phenotypic information, molecular, taxonomic keys, phylogenies, distributions [geographic, temporal], life history, associations, methods, processes, metadata support). **Examination** includes processes through which many different sources of information are generated, integrated, and interpreted, potentially including taking measurements, CT scanning, producing images, making observations, and generating sequences, genomes, proteomes, and phylogenetic trees. **Rendering** may include creating text, tables, figures, maps and any other derivative files or media in other formats, e.g., Darwin Core, Nexus files, tree files, taxon descriptions, nano-publications, or CT/Surface scan data. **Dissemination** is the distribution of results of rendering e.g., the published monograph, via traditional publication, archiving, APIs, data aggregators, nano publications, real-time web-portal (taxon pages), and self-publishing datasets. A workflow can start or continue at any node. For example, it may begin when specimens are **gathered** from the field or from collections, or both. These specimens are **examined** to obtain morphological, physiological, and behavioral traits and/or genetic data that can be **rendered** as taxonomic descriptions, data matrices, and phylogenies. Data matrices can be **disseminated** not only as a 'final' publication or PDF, but also on websites. At each step of the process of creating a monograph, the researcher can proceed to any other step: e.g., gather another specimen, (re) examine a specimen, render figures using a different layout, or deliver data to a shared repository. Each step can be repeated multiple times, both prior to traditional publication and afterwards if the authors use dynamic tools for managing and disseminating their data.

a set of historical literature and notes is an important component of any monograph, as it represents the conceptual foundation for systematic proposals. Previous observations and analyses (e.g., phenotypic data, molecular data, identification keys, phylogenies, temporal and geographic distributions, life history traits, habitat, associations with other organisms), published or unpublished, must be gathered to establish the history and context of the study. These data and observations must be stored in such a way that they stay linked to citations from which they were drawn. Specimens to be examined must be located in collections or collected in the field, the latter sometimes being the initial stage for previously untreated taxa. Analytical tools must also be located and learned, both physical (e.g., measurement devices, CT scanners) and digital (e.g., statistical, or phylogenetic software).

3 EXAMINATION

Once the necessary specimens, literature, notes, and other data are gathered, researchers extract and analyze information from these sources and integrate it in the context of the review, including treatment within the author's theoretical framework for defining taxa. Activities might include describing, coding and/or measuring phenotypic traits, obtaining, and aligning sequences, performing phylogenetic analyses, and imaging specimens. During examination, it is common to go back to previously processed data (e.g., to re-examine specimens for a newly found trait) as understanding evolves, new discoveries are made, and new hypotheses generated. It is at this stage, which is increasingly collaborative and interdisciplinary in nature, that derivative and novel data emerge and are slowly organized. These activities are perhaps the most likely to benefit from new digital tools and workflows (Table 1).

4 RENDERING

When a researcher feels that a component or the whole of their work is sufficiently refined or nearing completion, part, or all of it can be summarized and/or presented to others. We describe the formatting of gathered data for the purposes of sharing it with others (and thus enabling iterative examination by the community) as "rendering". The goal of rendering is to present information in an understandable format. In a taxonomic revision, rendered information may take many forms including distribution maps, graphs, tables and matrices, phylogenetic trees, and figures with captions. Standardized taxonomic descriptions (text including synonyms, diagnosis, and description) are also a form of rendered information.

Modern technology offers new and more dynamic ways of rendering information than were available to monographers in previous centuries. If the data associated with other steps in the monographic process (i.e., gathering and examination) are accessible and constantly or regularly updated, then rendered products can be more dynamic and therefore reflect the current taxonomic hypotheses. For example, a taxonomic description can be presented in a wiki knowledge base (e.g., https://species-id.net/wiki/) or in the form of a nano-publication (e.g., Groth et al. 2010), summarizing information for the taxon of interest, linked to distributional records in a Darwin Core standard formatted file that is accessible in a repository (e.g. GitHub, OSF, DRYAD). Some journals, (those published by Pensoft, in particular) are already automating the generation of these kinds of products. As data

Table 1. Example tools and processes indexed within our conceptual model. Within a hub serving monographers we can use our model (Fig. 1) to organize resources into categories: "G" - gathering, "E" - examination, "R" - rendered and "D" dissemination. Monographers looking for aid can self-assess what they are doing and seek tools and guidance in the appropriate part of the hub.

Tool/process	G	E	R	D	Description
Create a Darwin Core Archive (DwC-A) (Wieczorek et al 2012, GBIF 2021)			X	X	Besides the *material examined* section of a monograph, specimen and nomenclature data can be rendered in this format. When disseminated, this core data is usable by a wide range of existing tools and data aggregators.
Create a Catalogue of Life Data Package (CoLDP) names list			X	X	Taxonomic monographs often revise and update the status of proposed taxonomic names and concepts. This information can be rendered as a fully annotated checklist in this format.
Build an observation matrix for phenotypic data (spreadsheet, Mesquite (Maddison and Maddison 2019), TaxonWorks (Species File Group et. al 2021), MorphoBank (O'Leary and Kaufman, 2011)	X	X	X		These kinds of matrices are usually created during the examination process. The choice of tool that is used depends largely on availability and author's preference. Matrices often get "printed" as flat text to the final PDF during the editorial process, which reduces the re-usability of this information. It is ideal to store matrices, in their original file format, in appropriate repositories.
Georeference specimens (Google Maps)		X	X		Geographic coordinates are not always included in specimen data labels, which requires researchers to approximate the coordinates of given localities by using gazetteers. This information gets rendered as data points in distributional maps. It can also be included in DwC-A files for dissemination.
Interactive maps of georeferenced occurrences (e.g., leaflet and Shiny Apps; Paleobiology Database)	X	X	X	X	Mapping georeferenced occurrences is the first step to understanding distribution patterns. Interactive maps can be rendered to include details of the available records and linked to images, museum specimens and other data produced by examination activities. These interactive maps can be disseminated as websites (e.g., https://eugeniovaes.shinyapps.io/Proof/) and guide more efficiently gathering and examination activities.
Image specimens (MorphoBank (O'Leary and Kaufman 2011, 2012); MorphoSource (Boyer et al., 2016)	X	X	X	X	Specimen images are often organized as annotated plates, video files and/or polygon/ volumetric mesh files to showcase particular phenotypic traits. Independent images/datasets can also be uploaded to and disseminated via online repositories, which in some cases generate a DOI for each independent file.
Molecular data (BOLD - Barcode of Life Data System (Ratnasingham and Hebert 2007), NCBI (NCBI Resource Coordinators 2016))	X	X	X	X	This information is rarely rendered as part of the text in a monograph. More often it is part of the data that is analyzed in the context of the monograph and is disseminated through appropriate repositories and linked back to the monograph, usually as part of tables including codes assigned within the repository.

Information on biotic interactions (GloBI, Poelen et al. 2014)	X		X	X	Information about species interactions can be reported to platforms that aggregate and integrate it.
Phylogenetic trees (Open Tree of Life https://opentreeoflife.github.io/)		X	X	X	Phylogenetic trees are often rendered as part of figures within monographs. Tree files should also be disseminated in their original format (not just as an image) through online repositories and be incorporated in databases.
Literature cited	X			X	It should be standard practice to include DOI links for every cited paper that is available online. Besides a list of bibliographic resources, researchers could provide BibTeX-formatted files to be hosted along with the treatments included in the monograph.
Upload all your research files into a single repository (DRYAD, Zenodo, Figshare, OSF)	X	X	X	X	Resources uploaded to these repositories are openly accessible and citable via DOI or other persistent URL.

in the repository is updated, so too could the taxon description be updated. A fluid transition from examination and gathering to rendering can efficiently guide the workflow and indicate areas for new investigations.

5 DISSEMINATION

We take for granted (see other papers in this issue) that monographs in systematic biology are of potential interest to many different user communities. A strategy to ensure that monographs and their supporting data are maximally useful over time and across scientific disciplines is to focus on their propagation and dissemination. Our model suggests that this should not be conceived as something done only once at the point of completion, but rather as a multipart process that can occur at any time during the development of a monograph. For example, a researcher can share a new map online after it is rendered and use community feedback on that map to immediately identify new specimens to examine. Dynamic feedback, increased through real-time dissemination, can improve both the completeness and quality of a monograph and its "reach" in the scientific world and beyond, especially when persistent identifiers are linked throughout the monograph (Mabry et al. 2022). It is also important to employ a variety of dissemination approaches. This strategy increases the chance others will both see and re-propagate the data, further extending the reach and importance of the monograph.

6 CONCLUSION

The conceptualization of the four kinds of actions outlined above as iterative, dynamic, and interactive highlights the importance of identifying appropriate tools and pipelines to make the monographic workflow more efficient, more translatable, and more fruitful. The number of examples of monographs implementing the use of the tools presented here (i.e., GBIF, NCBI, MorphoBank, MorphoSource, DRYAD, etc.) continues to increase. A guide for monographers seeking relevant contemporary tools is urgently needed, particularly for the uninitiated, but also for the more experienced systematists who want to implement the most recent software, tools, and online resources. Ideally, this resource hub (i.e., clearinghouse of useful tools for monographers) must be embraced by the community of systematic

biologists and should be maintained and updated. Already established societies and their networks could help to ensure awareness by supporting this initiative and to consolidate the community around it.

We recognize that the implementation of some of the tools suggested in this contribution might represent a challenge, especially when financial support and recognition for developing monographic work are scarce (Gorneau et al. in this issue), but we are also convinced that using some of these tools, besides facilitating some tasks, will increase the accessibility, visibility, reusability, and citability of the large amount of data contained in monographs.

Acknowledgements

We thank Felipe Zapata (UCLA), Meg Daly (The Ohio State University), and all participants of the NSF-sponsored workshop on "Modernizing Monography" DEB-1839202. Sarah Jacobs and Nate Shoobs participated in initial discussions on the topics treated here. We thank Jesus Martinez-Gomez for commenting and discussing on a draft of the manuscript.

References

Boyer, D.M., Gunnell, G.F., Kaufman, S., & McGeary, T.M. (2016). Morphosource: archiving and sharing 3-D digital specimen data. The Paleontological Society Papers, 22, 157-181. https://doi.org/10.1017/scs.2017.13

GBIF (2021). Darwin Core Archives – How-to Guide, version 2.1, released on 8 February 2021, (contributions by Remsen D., Braak, K., Döring, M., Robertson, T., Blissett M. Copenhagen: Global Biodiversity Information Facility, accessible online at: https://github.com/gbif/ipt/wiki/DwCAHowToGuide (accessed 13 April 2021)

Gorneau, J.A., Ausich, W.I., Bertolino, S., Holly, B., Daly, M., Demissew, S., Donoso, D.A., Folk, R., Freire-Fierro, A., Ghazanfar, S.A., Grace, O.M., Hu, A.Q., Kulkarni, S., Lichter-Marck, I.H., Lohmann, L.G., Malumbres-Olarte, J., Muasya, A.M., Pérez-González, A., Singh, Y., Siniscalchi, C. M., Specht, C.D., Stigall, A.L., Tank, D.C., Walker, L.A., Wright, D.F., Zamani, A., Esposito, L.A. Framing the Future for Taxonomic Monography: Improving Recognition, Support, and Access. Bull Soc Syst Biol. 2022;1(1):8328. https://doi.org/10.18061/bssb.v1i1.8328

Groth, P., Gibson, A., & Velterop, J. (2010) The anatomy of a nanopublication. Information Services & Use 30.1-2: 51-56. https://doi.org/10.3233/ISU-2010-0613

Mabry, M. E, Zapata, F., Paul D.L., O'Connor, P.M., Soltis, P.S., Blackburn, D.C., Simmons, N.B. Monographs as a nexus for building extended specimen networks using persistent identifiers. Bull Soc Syst Biol. 2022;1(1):8323. https://doi.org/10.18061/bssb.v1i1.8323

Maddison, W.P., & Maddison, D.R. (2019). Mesquite: a modular system for evolutionary analysis. Version 3.61. http://www.mesquiteproject.org

Maxted, N. (1992). Towards defining a taxonomic revision methodology. In: Taxon 41.4, pp. 653–660. https://doi.org/10.2307/1222391

NCBI Resource Coordinators (2016) Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 44: D7–D19. https://doi.org/10.1093/nar/gkv1290

O'Leary, M.A., & Kaufman, S.G. (2011) MorphoBank: phylophenomics in the "cloud." Cladistics 27: 529–537. https://doi.org/10.1111/j.1096-0031.2011.00355.x

O'Leary, M. A., & Kaufman, S.G. (2012) MorphoBank 3.0: Web application for morphological phylogenetics and taxonomy. http://www.morphobank.org.

Poelen, J.H., Simons, J.D., & Mungall, C.J. (2014). Global Biotic Interactions: An open infrastructure to share and analyze species-interaction datasets. Ecological Informatics. https://doi.org/10.1016/j.ecoinf.2014.08.005

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

Species File Group and Collaborators. (2021) TaxonWorks. https://taxonworks.org.

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

Submitted: 27 May 2021

Editor: Felipe Zapata

Managing Editor: Dinah Ward