Bulletin of the SSB

Monographs as a nexus for building extended specimen networks using persistent identifiers

Makenzie E. Mabry[1], Felipe Zapata[2], Deborah L. Paul[3], Patrick M. O'Connor[4], Pamela S. Soltis[1], David C. Blackburn[1], Nancy B. Simmons[5]

Affiliations:

[1]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611 USA; [2]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095 USA; [3]Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign IL 61820 USA; [4]Department of Biomedical Sciences, Heritage College of Osteopathic Medicine, Ohio Center for Ecological and Evolutionary Studies, Ohio University, Athens, OH USA; [5]Department of Mammalogy, Division of Vertebrate Zoology, American Museum of Natural History, New York NY 10014 USA

Correspondence:
Makenzie Mabry
Email: mmabry44@gmail.com

1 INTRODUCTION

Taxonomic monographs compile, integrate, and interpret the information associated with specimens stored in biological collections. This information traditionally includes phenotypic, genetic, geographic, and ecological data as well as, increasingly, phylogenetic trees; cumulatively, these data form the basis of taxonomic opinions and decisions made by systematists (Winston 1999; Grace et al. 2021). Specimens thus serve as the fundamental unit of monography, and they are critical for biodiversity science (NASEM 2020). With ongoing advances in digitization of biodiversity collections, more and more specimens have a digital representation (e.g., metadata, images, DNA sequence data) that is available online. The community notes the ongoing evolution of the "extended specimen" (Webster 2017; also called the "holistic specimen"; Cook et al. 2016), which consists of the physical specimen plus all associated data and derivative products, from traditional physical objects. Together these form a series of increasingly interconnected resources creating an extended specimen network (ESN; Lendemer et al. 2020), digital or otherwise. Potentially limitless, these additional resources range from standard physical attributes, such as phenotypic measurements or genetic data, to novel digital resources, such as

multimedia components (e.g., CT-scan images, sonograms) or biotic interaction networks.

During a recent NSF-sponsored workshop, participants addressed the question: *How might we modernize and revitalize monography?* We propose here that since taxonomic monographs already act as hubs through which specimen data are centralized, enhanced, and communicated, future taxonomic monographs should become portals to the growing nexus of comprehensive records of linked biodiversity data embodied by the ESN. Realizing this potential requires a shift from traditional taxon-based monographs (i.e., a monograph where the fundamental unit linking biodiversity data is the taxon, such as the species) to specimen-based monographs (i.e., a monograph where the fundamental unit linking biodiversity data is the specimen). This implies the adoption of a standard of practice that includes citing both known identifiers (e.g., traditional collection catalog numbers, field numbers) and using persistent identifiers (*see below*) for specimens and all data and metadata related to a given specimen. Persistent identifiers require a robust infrastructure and changes in community practice. An infrastructure that supports universal adoption and curation of persistent identifiers would facilitate enhanced citation, attribution, and credit as well as enable dynamic linking of extended specimen metadata, establish connections between different data types/sources, enable discovery of these data across research domains, and allow tracking of specimen usage across the scientific community worldwide. These goals will become increasingly attainable as monographs transition to entirely electronic (if still printable) formats because persistent identifiers can facilitate the effective use of semantic web approaches

(see also Parr et al. 2012).

2 DISCUSSION

1.1 What is a Persistent Identifier?

A persistent identifier (PID) is a long-lasting and globally unique reference to a digital resource that cannot be reused to refer to a different object (Güntsch et al. 2017; McMurry et al. 2017; Nelson et al. 2018). Borrowing terminology from software development, PIDs are sometimes treated as synonyms of globally unique identifiers (GUID) or universally unique identifiers (UUID), but persistence of identifiers remains challenging for both social and technical reasons. Since both persistence over time and uniqueness are key features, we prefer the term "persistent identifier" sensu McMurry et al. (2017). Familiar examples include digital object identifiers (DOIs) and Open Researcher and Contributor IDs (more commonly known as ORCIDs), although even these do not always consistently recover exactly the same data (see Elliott et al. 2020). However, there are many other types of PIDs that may be used to permanently link monographs and other scientific contributions to underlying specimens, data, notes, images, and other media. PIDs are increasingly recognized as critically important in biodiversity science because linking specimen data via unique specimen identifiers would make the data findable, accessible, interoperable, and reusable (i.e., FAIR data; Wilkinson et al. 2016; Miralles et al. 2020), but consensus on the optimal form of PIDs remains elusive. Current collection management software and online biodiversity data aggregators accept a variety of identifier types.

Persistent identifiers are not the same thing as museum catalog numbers because the latter are often not unique. For

example, "AMNH 12345" may be understood to refer to a specimen in the collection of the Department of Mammalogy at the American Museum of Natural History when that number appears in a paper on mammals, but another "AMNH 12345" may exist in other collections at the same institution (e.g., collections of birds, fish, fossils) and also may appear in the biodiversity literature. Similarly, "CAS 12345" might refer to the California Academy of Sciences, Chinese Academy of Sciences, or Chicago Academy of Sciences. Online databases and text-mining tools are often not capable of distinguishing among these different traditional identifiers. Even the Darwin Core Triplet, which is a concatenation of values for institution code (e.g., AMNH), collection code (e.g., Mammalogy), and catalog number (e.g., 12345), may not be sufficient for permanently linking specimens and data due to problems in how they are constructed, deployed, and curated (Guralnick et al. 2014). Even when deployed consistently, they may not be unique; while "CAS:Herp:12345" might indicate the collection within the institution, exactly which institution remains ambiguous. However, the clear value of PIDs makes it essential that researchers make every effort to incorporate PIDs in their specimen-based publications. Best practices should additionally include the publication of digitized data and PIDs in recordsets (e.g., online collection catalogs) by the institutions that house the specimens, which can then be harvested by aggregators such as GBIF (the Global Biodiversity Information Facility; gbif.org) and iDigBio (integrated Digitized Biodiversity Collections; idigbio.org). If monographers have collected specimens themselves, every effort should be made to coordinate with a museum collection where they plan to eventually reposit the specimens to ensure that PIDs are assigned and/or recorded before or

at the time of cataloging.

1.2 How Persistent Identifiers can Modernize Monography

Our proposal is that modern monographs should become the nexus for building ESNs, and PIDs are essential to accomplish this goal. Establishing a comprehensive network of extensible specimen data that integrates the wealth of biodiversity and expertise held in natural history collections and repositories is a major, but indispensable, undertaking for the ESN to reach its fullest potential (Lendemer et al. 2019). A robust, comprehensive system of specimen identifiers is thus critical to enable such transparent integration of biodiversity data. We argue that modern monographs can work as effective distributed platforms of PIDs which will facilitate the integration of specimen data and metadata across disparate data resources, especially when connected via appropriate cyberinfrastructure (e.g., an Application Programming Interface, API). This will contribute not only to fulfilling the goals of the ESN (Lendemer et al. 2019) but will also bring monography and revisionary taxonomy to the forefront of systematic biology. Modern monographs linking biodiversity data via PIDs will be essential to more easily identify knowledge gaps for particular taxa, promote biodiversity discovery, as well as properly track and attribute specimen use (Groom et al. 2017). For instance, PIDs can facilitate locating type specimens, enabling efforts to automate the development of type catalogs, which is a resource currently lacking for most taxonomic groups. PIDs could also help to reconcile Linnaean names, taxon concepts, and their identifiers associated with lineages across multiple data sets and analyses (Parr et al. 2012; Sterner et al. 2020). PID-enabled monography could facilitate establishment of databases
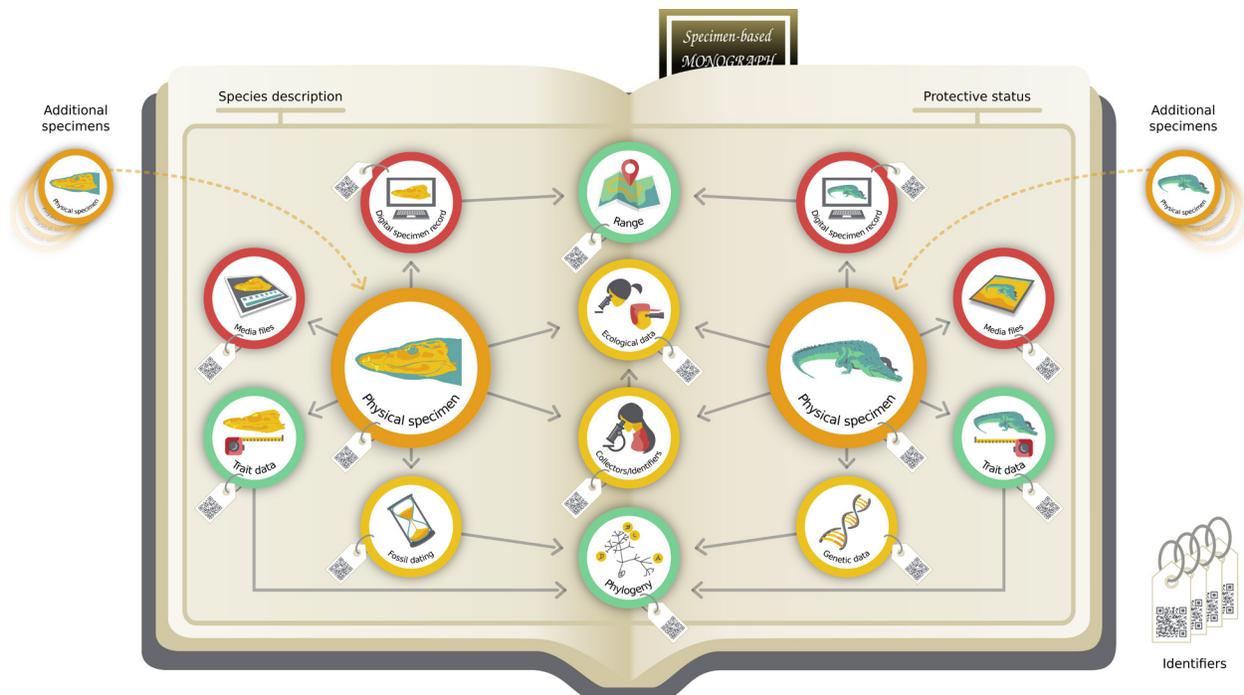
Figure 1. PID enabled Monograph. Physical specimens are centrally linked by associated data. Example products which result from those associated data are then further linked. Colors are the same as the proposed extended specimen by Lendemer et al. 2020 (i.e., red = primary extension, yellow = secondary extension, and green = tertiary extension). Artwork by Jagoba Malumbres-Olarte.

documenting ecological networks or genotype-phenotype association maps across the tree of life (Deans et al. 2013). By becoming portals of centralized biodiversity data across linked digital resources, monographs will also become far more accessible and useful to scientists outside the systematics community (e.g., Maclean et al. 2019; Lagomarsino and Frost 2020; Nakahama 2020). Further, modern monographs will become effective tools to train the next generation of systematists in state-of-the-art biodiversity informatics and data science. These re-envisioned data products are part of what is needed to reinvigorate long-term funding for taxonomy (Britz et al. 2020). Lastly, the use of PIDs in monography can easily foster international collaboration by integrating data across organizations globally, which can promote inclusivity in taxonomy (Gorneau et al. 2022).

1.3 Challenges and Solutions

Current practices in the use of PIDs have led to a general feeling that two parallel universes exist: one where biodiversity informatics specialists propose data standards and build data archives and another in which taxonomists practice, either not aware of these resources or lacking the funds and/or knowledge and tools to make use of them. Monographs are a clear missing link between these cultures of practice since taxonomists and organismal biologists are already familiar with monographs as a resource when beginning a project on an organismal group or looking for species descriptions and associated data.

Whereas the advocacy of PIDs to link biodiversity knowledge is not new (Guralnick et al. 2015; Page 2016, 2019), the slow adoption of these practices further indicates that the importance and usefulness of PIDs has only reached a portion of the bio-

diversity community. What will PID-enabled monographs look like in practice, especially when it seems that new repositories for storing specialized datasets are developed nearly every day? While the answer to this question may still be unclear—especially when considering both social and technical challenges—as a community we can start to take steps to set up a future in which PIDs are seamlessly interwoven with monographs (Figure 1).

Social challenges

One of the largest hurdles to overcome in adoption of PIDs in monography is the social challenge of changing the ways in which data are presented and archived in monographic projects. The traditional focus of monographs has been on taxonomic groups with measurement and trait data summarized at the species level and individual specimens merely listed in appendices. This has led to a situation in which subsequent researchers often need to "start from scratch" in accumulating data if they want to build on past taxonomic or monographic work (Sterner et al. 2020). The transition from taxon-based to specimen-based monographs implies that taxonomists will need to record and assign a PID to each specimen and its associated data and metadata (Figure 1). Whereas a shift to building "extended specimens" requires careful, intentional thought focused on organizing and linking associated information for each specimen at the outset; it will allow taxonomic works to build upon one another more directly, thus reducing the need for duplication of effort. This added level of granularity will also facilitate the application of explicit taxon concepts and the development of logical tools to evaluate specimen membership as well as to determine changes of circumscription in taxon concepts ("taxonomic intelligence";

Sterner et al. 2020). For example, when specimen membership changes, PIDs could dynamically 'carry over' all linked biodiversity data to the new taxonomic concept. This will make monography more transparent, visual, interoperable, and accessible, and will create exciting opportunities for monographs to become living documents with long lasting impact. Our proposal echoes ideas on the development of semantic annotations to describe phenotypes attached to individual specimens rather than taxonomic concepts (Deans et al. 2012). We look forward to bioinformatics methods used to evaluate semantic phenomic data (Thessen et al. 2020) becoming mainstream in monography.

As a first step, doing a better job of using and referring to existing specimen and data identifiers will aid in increasing the longevity and importance of our collections and data (Table 1; Table 2). While natural history collections may have historically operated as individual entities, the onset of digitization has revealed that linking data across institutions can provide resources that are of great use to the research community. By leveraging the interconnectivity between collections, researchers can gain previously hidden insight about the world. The COVID-19 pandemic perhaps best highlights this necessity by demonstrating the power of connecting bat and coronavirus collections (Upham et al. 2021).

Technical challenges

Although monographs may ultimately be created in an entirely electronic, updatable format that is connected directly into an extended specimen network via PIDs, we recognize that technical challenges exist to implementing this vision. Indeed, many tools exist for researchers (e.g., see Girón et al. 2022), but such resources are currently not centralized nor integrated

Table 1. Types of data typically found in monographs with sources of current identifiers (persistent identifiers in bold). Ideally, all data will be referable using PIDs in the future.

Monograph Data Types	Source(s) of Identifiers (hyperlinked where possible)
Physical Specimens	Collection catalog number (including institution & collection codes); Specify; Arctos; Symbiota
Digital Specimen Records	iDigBio; GBIF
Collectors/Identifiers	Bionomia; ORCID, Wikidata QID
Trait data	Dryad; FuTRES; MorphoBank
Genetic data	NCBI; Dryad
Phylogeny	Open Tree of Life; TreeBASE; Dryad
Ecological observations	iNaturalist; NEON
Media files	CalPhotos; MorphoSource; Macaulay Sound Library
Distributions	iDigBio; GBIF
Traditional/Cultural knowledge	BioCultural label
Protective status	Technical reports on biodiversity surveys or conservation assessments
Publications	Arpha Writing Tool in Pensoft

Table 2. Examples of identifiers (IDs) associated with a single specimen of the frog *Arthroleptis tanneri* housed at the California Academy of Sciences and first published by Blackburn (2008). Including more information such as the name of the repository and a URL to the identifier helps to reduce the guesswork and ensure that future researchers will be able to easily link to the data. *The Occurrence ID in this example features the Darwin Core Triplet for this specimen, but the format of the Occurrence ID can vary widely across institutional recordsets. Other examples include "http://n2t.net/ark:/65665/3e4a52adc-b389-4399-a7b9-7b4420af474f" and "06e449b6-c79d-4e26-9404-93c85c3036d6".

Type of Identifier	Effective format	Formats to avoid
Publication DOI	https://doi.org/10.1016/j.ympev.2008.08.015	10.1016/j.ympev.2008.08.015
Museum Catalog Number	CAS:Herp:168823	CAS 168823
Occurrence ID *	urn:catalog:CAS:HERP:168823	n/a
iDigBio	https://www.idigbio.org/portal/records/cceec639-0307-48eb-8834-6a6d491e4c3b	"Data on iDigBio"
NCBI Accession	https://www.ncbi.nlm.nih.gov/nuccore/FJ151056	FJ151056
CalPhotos	https://calphotos.berkeley.edu/cgi/img_query?enlarge=0000+0000+1004+0869	"Image in CalPhotos" or "0000 0000 1004 0869"
MorphoSource ARK	http://n2t.net/ark:/87602/m4/M25583	M25583
Dryad Data Package	https://datadryad.org/stash/dataset/doi:10.5061/dryad.9071	"Data deposited in Dryad"

into traditional publishing pipelines. For researchers preferring a traditional style of publication, a useful starting place for digital linking is to provide a downloadable single, well-structured text file that uses standard headers (e.g., DarwinCore), lists the relevant PIDs as well as traditional identifiers for each specimen (e.g., catalog numbers) and data (e.g., GenBank accession numbers) used in the monograph (see example in Table 3), and includes PIDs for the relevant people involved such as ORCID IDs or Wikidata QIDs. Inclusion of such a machine-readable file as an appendix to a monograph would require little additional effort on the part of a monographer but would hugely increase the degree to which the content of a monograph could be linked to the rest of the digital world. Even a traditionally published monograph that is available online could form a hub of a digital extended specimen network if PIDs are included in such a file.

A significant technical challenge for a researcher new to PIDs may be how to go about getting a PID for specimens or associated data in the first place. For most types of associated data, such as gene sequences or distribution data, PIDs are automatically provided when the data are submitted or uploaded to a repository. Getting PIDs for museum specimens may be more difficult, but many collections have moved to cataloging systems and software (e.g., Specify, Arctos, Symbiota) that automatically generates PIDs at the time of accessioning/cataloging and can provide them for older specimens upon request. For digital records, repositories such as iDigBio (https://www.idigbio.org) generate PIDs that could be cited (e.g., idigbio:uuid; see Table 3). We urge that monographers discuss PIDs with the collections staff in charge of any specimens that they are either using or repositing during the course of producing a monograph.

For identifiers to "persist" over time they need human care and a robust infrastructure. Without a solid understanding of how identifiers need to be managed, sometimes minted identifiers like UUIDs get thrown away. We stress that best practice is to curate identifiers including PIDs to ensure they persist. Once PIDs such as the idigbio:uuid or the dwc:occurrenceID are assigned, there needs to be active intent to not change these identifiers. If they must change, there are ways to share both an old occurenceID and a new one (e.g., the Darwin Core Extension Resource Relationship), but this extension is not yet well-used or understood by our community. We suggest an investment in infrastructure that includes professional capacity development, better data management tools such as registries to mint and store identifiers, support mechanisms for advice, and a responsive infrastructure. Management of PIDs should become more sustainable as these changes are made throughout the user community. Stable identifiers make services and data like those offered through Bionomia (https://bionomia.net/) possible. Through Bionomia one can see metrics for specimen data use, the human network responsible, the knowledge gained, and the museums that care for these objects and support the researchers. These types of functions may serve to help the broader community better understand what these identifiers facilitate.

The adoption of PIDs in monography implies that we may need to carefully reengineer traditional approaches to linking biodiversity data via taxon concepts represented as scientific names (Kennedy et al. 2005). Indeed, it is in monographs where taxon concepts are put into practice and taxon names are proposed (McDade 1995). Monographers study and sort specimens into taxon concepts, and, therefore, such

dwc:scientificName	dwc:institutionCode	dwc:collectionCode	dwc:catalogNumber	dwc:recordedBy	dwc:identifiedByID	idigbio:uuid	dwc:occurrenceID
Cryptantha ptero-carya f. pterocarya	SDSU	-	20360	Mabry, Makenzie | M.G. Simpson	https://orcid.org/0000-0002-6139-9559 | https://orcid.org/0000-0002-6197-2132	447f823d-6efd-441b-ae56-91b1cd-o56ece82628a	21f348c5-4e5a-40c9-8cca-
Escallonia alpina	RSA	RSA	RSA080132	Felipe Zapata	https://orcid.org/0000-0002-9386-0573	e7697ca0-f585-4707-bofo-8dc06d	4924add2-c126-4517-8dcf-d9be8f-bo59306e55ca
Tragopogon castellanus	FLAS	-	261824	Douglas Soltis | Pamela Soltis	https://orcid.org/0000-0001-8638-4137 | https://orcid.org/0000-0001-9310-8659	b9d562db-1486-42d1-b6b8-aabf-aa8482f9cfd	ae32be05-b054-410a-ab70-
Plestiodon fasciatus	UF	Herp	188722	Blackburn, David	https://orcid.org/0000-0002-1810-9886	ba56a38f-fce8-4688-ab3c-df69963278e5e	799262243-c46c-47a6-a89a-
Artibeus lituratus	AMNH	Mammals	M-278081	N.B. Simmons et al.	https://orcid.org/0000-0001-8807-7499	c7046doc-93b0-4a8a-817a-757125753cc5b	urn:catalog:AMNH:Mammals:M-278081

Table 3: Example of Mapped Darwin Core archive with iDigBio UUID. Identifiers minted at the source (e.g., dwc:catalogNumber) vs those by the aggregators (e.g., dwc:occurrenceID) both need to be well taken care of. Additional terms can be found at https://dwc.tdwg.org/list/ such as dwc:associatedSequences for including DNA sequences used for phylogenetic inference.

concepts necessarily may often change based on specimen data and analysis of variation. Specimens and their associated data should move fluidly across taxon concepts if such data support alternative membership, disjointness, or subsumption. Owing to the use of PIDs, such fluidity and interconnectedness of specimen-level data should become fundamental principles in modern monography. Powerful biodiversity data management systems such as Scratchpads (Smith et al. 2012) and TaxonWorks (Dmitriev 2018) hold great potential to accommodate this technical need and transform the future of monography. The Paleobiology Database provides a forward-looking implementation of the fluidity of taxonomic nomenclature linked to specimens via unique identifiers when the underlying data (or taxonomic opinion) change (Peters and McCelennen 2016).

A final step in realizing the full potential of the ESN—a goal that could be met in this new age of electronic data and publication—would be the assignment of a separate PID for each piece of data associated with every specimen or observation in a monograph (Figure 1). This would mean that not only every specimen would have a PID, but each unique measurement (e.g., plant height, leaf length, fruit width, etc.) or other trait observation would have an associated PID; projects such as FuTRES (J. Deck, pers. comm.; https://futreswebsite.netlify.app/) are now implementing PIDs for specimen-based measurement data. However, before this can be fully realized, we suggest a first step should be the development of a funded and sustainable registry to mint and care for these PIDs long term. Promising efforts to make such a registry include The Global Registry of Biodiversity Repositories (GrBio; Schindel et al. 2016) which is now being shepherded by GBIF and in the process of being re-designed (www.gbif.org/grscicoll; Grosjean et al. 2021). While we are unaware of any examples of such a specimen-based, fully linked, and fully updatable monograph, we can imagine this as part of the future of monography in the digital age.

Finally, publishers play a key role in our vision of the future of monography. One unique example is Pensoft and their ARPHA writing tool (Penev et al. 2017), which currently is being updated to facilitate, at least in part, what we envision here. The publishing workflow envisioned at Pensoft includes managing identifiers that authors have mapped to data standards so that all identifiers in a text can be found by computers. Before fully linked, PID-enabled monographs can become a reality, publishers and distributors of monographs will need to allow and facilitate the changes that we have outlined above to bring monographs fully into the digital age.

3 CONCLUSIONS

While the specimen data from museum collections and other resources central to monographs are increasingly made digitally available, the scarceness of centralized repositories linking these resources is a limitation for the development of the "extended specimen" and "cyber-enabled taxonomy" (Lendemer et al. 2020; Miralles et al. 2020). We argue that monographs can work as effective distributed platforms for PIDs which will facilitate the integration of specimen data and metadata across these disparate data resources. However, for a full realization of the ESN, the development of a funded and sustainable registry must first occur. While currently social and technological challenges prevent this, we suggest that there are several actions that we can take as a

Table 4: Suggested actions for different user groups in implementing PIDs.

User Group	Action Items
Researchers	• Publish a Darwin Core Archive file with Monograph files. • Development of biodiversity informatics tools.
Museum/Collection staff	• When supporting and encouraging researchers to use collections, send guidance on how they expect the specimens to be cited. • Just as specimens need curation, PIDs will need active curation and software that supports best practices for managing and sharing these identifiers. • Include / integrate this information into undergrad and graduate coursework.
Repositories	• Help desk for assistance. • Development of bioinformatic tools. • Implementation of PIDs.
Publishers	• Make PIDs required for publishing. Add tools to the publishing process to support this work.
Funding agencies	• Create a central registry for minting, storing, and linking PIDs.
Professional Organizations and Meetings	• Support training events and sessions on these topics (SPNHC, TDWG, iDigBio, DiSSCo, Botany, Mammalogy, Paleo, Entomology, etc.).

community to achieve PID-enabled monography (see Table 4). Embracing the use and citation of PIDs would facilitate compiling and linking the data and metadata used to build monographs and could increase the scientific impact of these works by making taxonomic data FAIR and improve citation, attribution, and credit tracking. In addition, by adopting PIDs for voucher specimens—the core components of a monograph—it will be possible to increase the transparency of taxonomic decisions presented in monographs, thus making taxonomy more consistent and reproducible. Development of bioinformatic tools, such as R packages, that use API resources from repositories and data aggregators would facilitate easy discovery and aggregation of PIDs for specimens and associated data. End-users of monographs including taxonomists, ecologists, genomicists, policymakers, collection managers, and even the general public would benefit from these changes of practice in monographic work. PID-enabled monography can enable the discoverability and reusability of biodiversity data in novel ways and would allow for institutions—and individuals—to better measure the impact of their specimens and associated data.

Funding

Acknowledgements

References

Blackburn DC. Biogeography and evolution of body size and life history of African frogs: phylogeny of squeakers (*Arthroleptis*) and long-fingered frogs

(*Cardioglossa*) estimated from mitochondrial data. Mol Phylogenet Evol. 2008;49(3):806-826. https://doi.org/10.1016/j.ympev.2008.08.015

Britz R, Hundsdörfer A, Fritz U. Funding, training, permits—the three big challenges of taxonomy. Megataxa. 2020;1(1):49-52. https://doi.org/10.11646/megataxa.1.1.10

Cook JA, Galbreath KE, Bell KC, Campbell ML, Carrière S, Colella JP, Dawson NG, Dunnum JL, Eckerlin RP, Fedorov V, and others. The Beringian Coevolution Project: Holistic collections of mammals and associated parasites reveal novel perspectives on evolutionary and environmental change in the North. Arct Sci. 2016;3(3):585–617. https://doi.org/10.1139/as-2016-0042

Deans AR, Yoder MJ, Balhoff JP. Time to change how we describe biodiversity. Trends Ecol Evol. 2012;27(2):78-84. https://doi.org/10.1016/j.tree.2011.11.007

DiSSCo Tech Blog: Debunking reliability myths of PIDs for Digital Specimens [cited 17 May 2021] https://dissco.tech/2020/10/21/debunking-reliability-myths-for-pids-for-digital-specimens/

Dmitriev DA. TaxonWorks. Biodivers Inf . 2018;2:e25560. https://doi.org/10.3897/biss.2.25560

Elliott MJ, Poelen JH, Fortes JAB. Toward reliable biodiversity dataset references. Ecol Inform. 2020;59:101132. https://doi.org/10.1016/j.ecoinf.2020.101132

Girón JC, Valderrama E, O'Connor PM, Simmons NB, Paul DL, Yoder MJ. Enhanced monography in a collaboratively evolved hub for systematic biology. Bull Soc Syst Biol. 2022;1(1):8340. https://doi.org/10.18061/bssb.v1i1.8340

Gorneau JA, Ausich WI, Bertolino S, Bik H, Daly M, Demissew S, Donoso DA, Folk R, Freire-Fierro A, Ghazanfar SA, and others. Framing the Future for Monography: Improving Recognition, Support, and Access. Bull Soc Syst Biol. 2022;1(1):8328. https://doi.org/10.18061/bssb.v1i1.8328

Grace OM, Pérez-Escobar OA, Lucas EJ, Vorontsova MS, Lewis GP, Walker BE, Lohmann LG, Knapp S, Wilkie P, Sarkinen T, and others. Botanical monography in the Anthropocene. Trends Plant Sci. 2021;26(5):433-441. https://doi.org/10.1016/j.tplants.2020.12.018

Groom Q, Hyam R, Güntsch A. Stable identifiers for collection specimens. Nature. 2017;546(7656):33. https://doi.org/10.1038/546033d

Grosjean M, Høfft M, Gonzalez M, Robertson T, Hahn A. GRSciColl: Registry of Scientific Collections maintained by the community for the community. Biodivers Inf Sci. 2021;5:e74354. https://doi.org/10.3897/biss.5.74354

Güntsch A, Hyam R, Hagedorn G, Chagnoux S, Röpert D, Casino A, Droege G, Glöckler F, Gödderz K, Groom Q, and others. Actionable, long-term stable and semantic web compatible identifiers for access to biological collection objects. Database. 2017;2017:1-9. https://doi.org/10.1093/database/bax003

Guralnick R, Cellinese N, Deck J, Pyle R, Kunze J, Penev L, Walls R, Hagedorn G, Agosti D, Wieczorek J, and others. Community next steps for making globally unique identifiers work for biocollections data. ZooKeys. 2015;494:133. https://doi.org/10.3897/zookeys.494.9352

Guralnick R, Conlin T, Deck J, Stucky BJ, Cellinese N. The trouble with triplets in biodiversity informatics: a data-driven case against current identifier practices. PloS One. 2014;9(12):e114069. https://doi.org/10.1371/journal.pone.0114069

Kennedy JB, Kukla R, Paterson T. Scientific names are ambiguous as identifiers for biological taxa: their context and definition are required for accurate data integration. In International Workshop on Data Integration in the Life Sciences. 2005:80-95. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11530084_8

Lagomarsino LP, Frost LA. (2020) The central role of taxonomy in the study of neotropical biodiversity. Ann Mo Bot Gard. 2020;105(3):405-421. https://doi.org/10.3417/2020601

Lendemer J, Thiers B, Monfils AK, Zaspel J, Ellwood ER, Bentley A, LeVan K, Bates J, Jennings D, Contreras D, and others. The extended specimen network: A strategy to enhance US biodiversity collections, promote research and education. BioScience. 2020;70(1):23-30. https://doi.org/10.1093/biosci/biz140

MacLean HJ, Nielsen ME, Kingsolver JG, Buckley LB.

Using museum specimens to track morphological shifts through climate change. Philos Trans R Soc B. 2019;374(1763): 20170404. https://doi.org/10.1098/rstb.2017.0404

McDade LA. Species concepts and problems in practice: insight from botanical monographs. Syst Bot. 1995;20(4):606-622. https://doi.org/10.2307/2419813

McMurry JA, Juty N, Blomberg N, Burdett T, Conlin T, Conte N, Courtot M, Deck J, Dumontier M, Fellows DK, and others. Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. PLoS Biol. 2017;15(6):e2001414. https://doi.org/10.1371/journal.pbio.2001414

Miralles A, Bruy T, Wolcott K, Scherz MD, Begerow D, Beszteri B, Bonkowski M, Felden J, Gemeinholzer B, Glaw F, and others. Repositories for taxonomic data: where we are and what is missing. Syst Biol. 2020;69(6):1231-1253. https://doi.org/10.1093/sysbio/syaa026

Nakahama N. Museum specimens: An overlooked and valuable material for conservation genetics. Ecol Res. 2021;36(1):13-23. https://doi.org/10.1111/1440-1703.12181

National Academies of Sciences, Engineering, and Medicine. Biological Collections: Ensuring Critical Research and Education for the 21st Century. 2020. Washington, DC: The National Academies Press. https://doi.org/10.17226/25592

Nelson G, Sweeney P, Gilbert E. Use of globally unique identifiers (GUIDs) to link herbarium specimen records to physical specimens. Appl Plant Sci. 2018;6(2):e1027. https://doi.org/10.1002/aps3.1027

Page R. Ozymandias: a biodiversity knowledge graph. PeerJ. 2019;7:e6739 https://doi.org/10.7717/peerj.6739

Page R. Towards a biodiversity knowledge graph. Res Ideas Outcomes. 2016;2:e8767. https://doi.org/10.3897/rio.2.e8767

Parr CS, Guralnick R, Cellinese N, Page RD. Evolutionary informatics: unifying knowledge about the diversity of life. Trends Ecology Evol. 2012;27(2):94-103. https://doi.org/10.1016/j.tree.2011.11.001

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

Peters SE, McClennen M. The Paleobiology Database application programming interface. Paleobiology. 2016;42(1):1-7. https://doi.org/10.1017/pab.2015.39

Schindel DE, Miller SE, Trizna MG, Graham E, Crane AE. The global registry of biodiversity repositories: a call for community curation. Biodivers Data J. 2016;4:e10293. https://doi.org/10.3897/BDJ.4.e10293

Smith VS, Rycroft SD, Brake I, Scott B, Baker E, Livermore L, Blagoderov V, Roberts D. Scratchpads 2.0: a virtual research environment infrastructure for biodiversity data. [cited 2012 Nov 19] Available from: http://scratchpads.eu

Sterner BW, Upham N, Sen A, Franz NM. Avenues into Integration: Communicating taxonomic intelligence from sender to recipient. Biodivers Inf Sci. 2020;4:e59006. https://doi.org/10.3897/biss.4.59006

Thessen AE, Walls RL, Vogt L, Singer J, Warren R, Buttigieg PL, Balhoff JP, Mungall CJ, McGuinness DL, Stucky BJ, and others. (2020) Transforming the study of organisms: Phenomic data models and knowledge bases. PLoS Comput Biol. 2020;16(11):e1008376. https://doi.org/10.1371/journal.pcbi.1008376

Upham N, Poelen JH, Paul DL, Groom QJ, Simmons NB, Vanhove MPM, Bertolino S, Reeder DM, Bastos-Silveira C, Sen A, and others. Liberating host-virus knowledge from biological dark data. Lancet Planet Health. 2021;5(10):E746-E750. https://doi.org/10.1016/S2542-5196(21)00196-0

Webster M, editor. The extended specimen: Emerging frontiers in collections-based ornithological research. Boca Raton, FL: CRC Press; 2017.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, Bonino da Silva Santos L, Bourne PE, and others. Comment: The FAIR Guiding Principles for scien-

tific data management and stewardship. Sci Data. 2016;3:160018. https://doi.org/10.1038/sdata.2016.18

Winston JE. Describing species: practical taxonomic procedure for biologists. Columbia University Press; 1999.