



HAL
open science

Managing data locally to answer questions globally: The role of collaborative science in ecology

Isabelle Aubin, Françoise Cardou, Laura Boisvert-marsh, Eric Garnier,
Manuella Strukelj, Alison Munson

► To cite this version:

Isabelle Aubin, Françoise Cardou, Laura Boisvert-marsh, Eric Garnier, Manuella Strukelj, et al..
Managing data locally to answer questions globally: The role of collaborative science in ecology.
Journal of Vegetation Science, 2020, 31 (3), pp.509-517. 10.1111/jvs.12864 . hal-03033479

HAL Id: hal-03033479

<https://hal.science/hal-03033479>

Submitted on 23 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Managing data locally to answer questions globally: The role of collaborative science in ecology

Isabelle Aubin¹  | Françoise Cardou^{1,2}  | Laura Boisvert-Marsh¹  |
Eric Garnier³  | Manuella Strukelj⁴  | Alison D. Munson⁴ 

¹Great Lakes Forestry Centre, Canadian Forest Service, Natural Resources Canada, Sault Ste Marie, ON, Canada

²Département de Biologie, Université de Sherbrooke, Sherbrooke, QC, Canada

³CEFE, Univ Montpellier, CNRS, Univ Paul Valéry Montpellier 3, EPHE, IRD, Montpellier Cedex 5, France

⁴Centre for Forest Research, Département des Sciences du bois et de la forêt, Université Laval, Quebec, QC, Canada

Correspondence

Isabelle Aubin, Great Lakes Forestry Centre, Canadian Forest Service, Natural Resources Canada, Sault Ste Marie, ON P6A 2E5, Canada.
Email: isabelle.aubin@canada.ca

Funding information

Natural Resources Canada

Co-ordinating Editor: Francesco de Bello

Abstract

Ecologists are increasingly asking large-scale and/or broad-scope questions that require vast datasets. In response, various top-down efforts and incentives have been implemented to encourage data sharing and integration. However, despite general consensus on the critical need for more open ecological data, several roadblocks still discourage compliance and participation in these projects; as a result, ecological data remain largely unavailable. Grassroots initiatives (i.e. efforts initiated and led by cohesive groups of scientists focused on specific goals) have thus far been overlooked as a powerful means to meet these challenges. These bottom-up collaborative data integration projects can play a crucial role in making high quality datasets available because they tackle the heterogeneity of ecological data at a scale where it is still manageable, all the while offering the support and structure to do so. These initiatives foster best practices in data management and provide tangible rewards to researchers who choose to invest time in sound data stewardship. By maintaining proximity between data generators and data users, grassroots initiatives improve data interpretation and ensure high-quality data integration while providing fair acknowledgement to data generators. We encourage researchers to formalize existing collaborations and to engage in local activities that improve the availability and distribution of ecological data. By fostering communication and interaction among scientists, we are convinced that grassroots initiatives can significantly support the development of global-scale data repositories. In doing so, these projects help address important ecological questions and support policy decisions.

KEYWORDS

data integration, data management, data sharing, ecoinformatics, ecological data

Aubin and Cardou should be considered as joint first authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 Her Majesty the Queen in Right of Canada. International Association for Vegetation Science published by John Wiley & Sons Ltd. Reproduced with the permission of the Minister of Natural Resources Canada.

1 | INTRODUCTION

In a context of data-intensive science, data collection and management are, more than ever, a central focus of ecology (Hampton et al., 2013). However, poor ecological data availability and accessibility are currently a major hurdle in addressing 21st-century challenges, such as more refined forecasts of global change impacts on ecosystems (Aubin et al., 2016; Reichman, Jones, & Schildhauer, 2011; Urban et al., 2016). In response to numerous calls for more open ecological data, global-scale databases, repositories and warehouses have grown rapidly and are now well established. These global repositories have “opened” vast datasets from governmental agencies, universities and observation networks (e.g. Global Biodiversity Information Facility, 2018; iDigBio, 2018). Recognizing the need for improved availability, traceability and citability of data, several funding agencies now consider data accessibility as a criterion for project funding. Increasingly, journals also require authors to either archive raw datasets on open platforms (e.g. Dryad, Zenodo), or offer the possibility of publishing a data paper. Despite this significant top-down pressure for improved data curation, availability of ecological data is still proportionally low compared to other fields (Cheruvellil & Soranno, 2018; Roche, Kruuk, Lanfear, & Binning, 2015). In a paper published in 2011, Reichman, Jones & Schildhauer estimated that less than 1% of ecological data is made accessible after publication. Although these numbers have likely increased somewhat since then, ecological data remain largely scattered, undiscoverable and poorly integrated.

Grounded in the same awareness of the importance of sharing ecological data, many collaborative data integration initiatives have developed among small and/or more cohesive groups of researchers. Scientists organize themselves in a collaborative structure to develop solutions that meet their specific needs at a basic level (e.g. promote data availability and sharing). Because these efforts are initiated and led by cohesive groups of scientists focused on specific goals, they can be defined as grassroots (Seyfang & Smith, 2007). As ecologists tackle important questions that defy the capabilities of individual laboratories and even disciplines (Barlow et al., 2018; Ramirez et al., 2018; Schwartz et al., 2017), such bottom-up initiatives have been found by their initiators to be one powerful way of addressing important scientific and technical challenges, as well as more global data availability objectives. While the impact of some of the larger efforts has been recognized (e.g. TRY, Kattge et al., 2020; sPlot, Bruehlheide et al., 2019), most grassroots initiatives have received far less attention.

With this paper, we underscore the important but overlooked role of grassroots data integration initiatives in making ecological data available. To show the variety of formats that these initiatives can take, we present three collaborative structures that, from our own experience, each fulfill specific needs and illustrate them with three regional initiatives. Although we consider global-scale bottom-up initiatives to be an important form of grassroots initiatives, we deliberately present smaller, local efforts in which ecologists can readily get involved. We demonstrate how grassroots initiatives can foster greater data openness and interoperability while offering

Box 1. The complexity of ecological data

Ecological data: a definition

Representation of information about the natural world presented in a structured format suitable for interpretation or processing, that could be reinterpreted for use in a different field of study or context (Borgman, Wallis, & Enyedy, 2007; Zimmerman, 2008). Zimmerman (2008) argues that a definition of ecological data without reference to time and place is incomplete. The context provided by metadata, or “data about data,” is crucial to sound interpretation and reuse in ecological studies.

The complexity of ecological data

The data required to make inferences about ecological processes are complex and heterogeneous by nature. Ecological data are everywhere, extremely diversified and can be measured in a variety of ways. Ecological research can range from single, short duration observations to longitudinal studies at spatial scales ranging from sublocal to global (Michener & Jones, 2012). Given the number of taxa to study, and the potential for interactions among them, the breadth of methodologies to address ecological questions is large (Rüegg et al., 2014). By comparison, despite considerable volumes of data (LaDeau, Han, Rosi-Marshall, & Weathers, 2017), genetics generates datasets that are relatively easier to manage (four letters comprise the base of all genetic analyses).

Ecology can be considered a “big data” science (Hampton et al., 2013) because of the “four Vs” (LaDeau et al., 2017): *volume* of data generated; the complexity of ecological data residing mainly in its *variety*; the potential for high *velocity* of change and, occasionally, a requirement for data *veracity* (e.g. validating LiDAR-derived outputs from the field). Several of these “Vs” are reflected in the *heterogeneity of ecological data*, a major impediment to integrated ecological research.

researchers greater recognition and data management support. Overall, we aim to encourage researchers to formalize existing collaborations and to engage in bottom-up approaches to improve ecological data availability.

2 | WHY TOP-DOWN INITIATIVES FIND LIMITED SUCCESS WITH ECOLOGISTS

Ecological data have typically been (and in most cases, still are) collected at relatively small geographical and temporal scales, using ad hoc designs and project specific methods. Ecological data are

more often than not isolated in highly specialized knowledge silos or scattered throughout the grey literature where they are both inaccessible and undiscoverable (Heffernan et al., 2014; Reichman et al., 2011). This limits diffusion to platforms where they could have a significant impact (Rüegg et al., 2014). For instance, while geophysical data are generally available, well described and predicted, we lack fundamental knowledge for a vast majority of biodiversity components, such as precise maps of species' current geographic range. Long-term data are also lacking, precluding our ability to formulate clear trends on species loss or introduction (Proença et al., 2017). The wide scope of ecological research, the spatial and temporal heterogeneity of ecological data and the breadth of methodologies (see Box 1; Michener & Jones, 2012; Rüegg et al., 2014) are some of the obstacles that may explain the difficulties that exist in sharing and integrating data within ecology and with other disciplines.

Top-down initiatives are generally built on generic frameworks inspired by initiatives that have had demonstrated success in other fields (e.g. GenBank from the National Center for Biotechnology Information). Such frameworks have found limited success in ecology because they are rarely customized to take into account the specific sociocultural challenges around sharing ecological data. Ecological data may consist of hard-earned data points representing years of work. Some data generators may be reluctant to contribute data to large anonymous repositories if sharing offers only low probability of tangible rewards or could undermine a competitive advantage (Cheruvilil & Soranno, 2018; Roche et al., 2014). Having invested significant resources into their long-term projects, some ecologists may deliberately avoid funding sources or journals with mandatory public data archiving requirements.

In an effort to increase participation in data archival, large repositories have abandoned many data management practices that were seen as creating a technological barrier for data providers. While this has had the desired effect (data archival is up), it has also had an impact on data discoverability. Large repositories archive ecological datasets in a wide variety of formats, resolutions and levels of supporting documentation (i.e. metadata), making their comparability, discoverability and reuse a challenge (Costello, Michener, Gahegan, Zhang, & Bourne, 2013). The lack of agreement on the use of recognized data standards (e.g. metadata and controlled vocabularies) required to upload datasets in these large repositories also increases the risk of data being misused. Hence, only a small portion of these datasets finds new life in ecological research.

With open-science principles and data-intensive science opening up new exciting areas of research, grassroots data integration initiatives taking a variety of forms are flourishing in academia, governmental and non-governmental research. These initiatives leverage resources across collaborators with different strengths and available infrastructure, improving scientific resource allocation. By maintaining closer ties among data generators and data users, these initiatives may represent an important piece of the puzzle in mitigating inertia at lower levels and alleviate some of the scaling problems at global levels. With no communication plans, outreach

policies or much media buzz, bottom-up initiatives are nevertheless making open science in ecology happen; it is time they get due recognition.

3 | GRASSROOTS INITIATIVES: BUILDING BLOCKS FOR GLOBAL RESEARCH

Ecology has a long history of collaborative research, including multi-author papers and large research teams. More recently however, the development of "team science", a branch of organizational psychology that studies how different team processes and structures can affect research outcomes, has provided a wealth of tools for researchers already involved in various types of collaborations (Börner et al., 2010; Cheruvilil & Soranno, 2018). By integrating these advances, grassroots initiatives have become efficient data-sharing structures. As groups of close-knit researchers within a same region, and/or with common questions and complementary approaches (Wiser, 2016), grassroots initiatives have a high level of relevance within the culture of ecology, increasing buy-in from ecologists. Thus, reframed within collaborative research, together with the support structure that this brings, incentives to share data become more apparent.

For ecology, the paradigm shift around open science has been momentous. Understandably, the new questions that have opened up with the advent of global-scale data repositories, together with increasing urgency around these questions, has created much enthusiasm. In comparison, local grassroots data integration initiatives represent less a paradigm shift and more the natural evolution of everyday practice. Nonetheless, such structures operating at lower levels are laying the necessary groundwork that feeds the success of many large-scale (and more publicized) initiatives. In fact, they represent the intuitive response of a diverse discipline to the need to make data more available. In spite of this, or perhaps because of, their collaborative structures and best practices are seldom considered as an integral part of the research process, and brief descriptions in the methodology sections are usually all the attention they might receive.

Grassroots data integration initiatives may take different forms, each adapted to suit various teams and their specific goals. In Figure 1, we summarize three types of structures these initiatives may take by drawing on local examples from our own experience. The first collaborative structure aims to improve data availability through the integration of discipline-specific datasets (Figure 1a). The TOPIC database (Box 2) was built expressly for this purpose. This regional initiative provides support to integrate recently acquired and legacy datasets according to international standards and make data available locally to other researchers, and internationally via a contribution to the TRY database (Kattge et al., 2020). Grassroots initiatives can also provide the structure necessary for researchers to ask broad questions that exceed the capabilities of individual laboratories (Figure 1b). In Box 3, we present the example of the Co-VITAS project, which

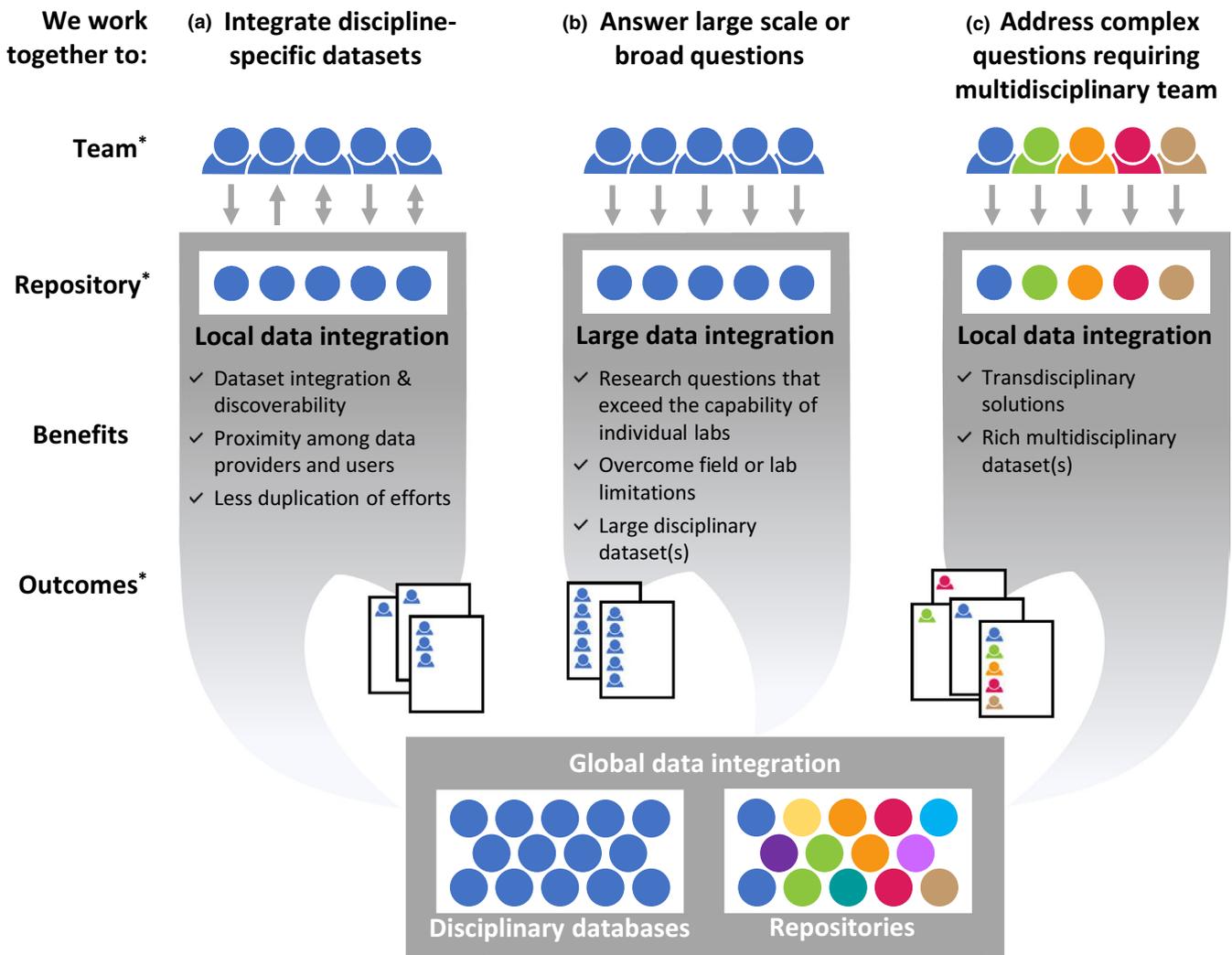


FIGURE 1 Three structures that can be used by grassroots data integration initiatives to address specific goals and which produce various research outcomes. (a) Local initiatives which provide support to integrate recently acquired and legacy datasets according to international standards. Data are made available locally to researchers, and internationally via contribution to global-scale disciplinary databases. See the TOPIC database example in Box 2. (b) Group of researchers focused on a common question or objective that exceeds the capabilities of a single laboratory. The large disciplinary dataset created around a large-scale or broad research question contributes to global-scale disciplinary databases and/or repositories. See the Co-VITAS example in Box 3. (c) Multidisciplinary team of researchers addressing a common problem requiring transdisciplinary solutions. An integrated multidisciplinary dataset is created which contributes to discipline-specific research, transdisciplinary science and global-scale disciplinary databases and/or multidisciplinary repositories. See the Island Lake experiment example in Box 4. All these structures favour interoperability with global-scale databases and repositories. * Colours represent different disciplines [Colour figure can be viewed at wileyonlinelibrary.com]

studies intraspecific trait variability at a subcontinental scale. Intensive coordinated field sampling for plant traits at that scale would represent a massive undertaking for any single laboratory. In this project, logistical field and laboratory limitations were overcome using a collaborative structure. Participants collected data to answer specific questions defined together from the outset, yielding a well-documented, high-quality dataset that can be more easily integrated into disciplinary repositories. Going further, collaborative structures can help generate “rich” datasets stemming from multidisciplinary research (Figure 1c). Breaking down barriers to data sharing across disciplines can be challenging, notably because of differences in scale, ontology and measurement types.

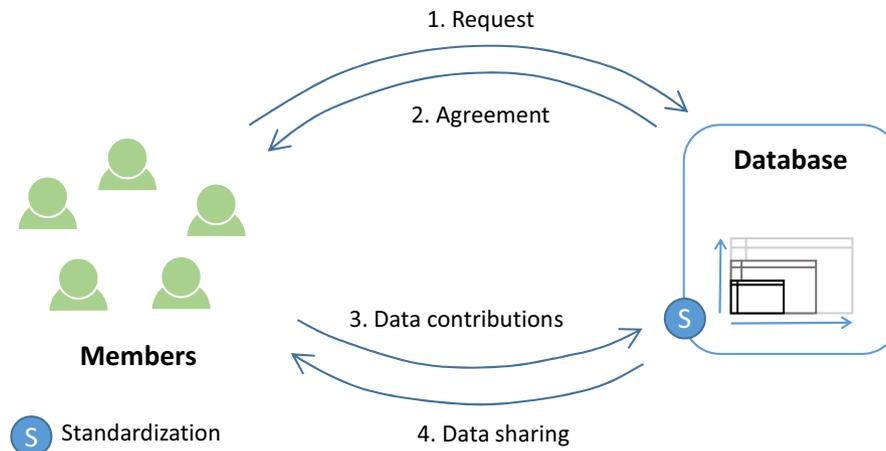
These challenges may preclude data integration entirely and result in missed opportunities for collaboration and new discoveries. By fostering a common language among different branches of ecology, collaborative structures lay the groundwork for the development of tools that favour interoperability and data exchange, ultimately removing these roadblocks. In the Island Lake Biomass Harvesting experiment (Box 4), scientists from different disciplines work on the same experimental design, generating advances in their own disciplines as well as contributing integrative insight into ecosystem functioning in the boreal forest. By developing multidisciplinary teams, data integration initiatives can both bridge knowledge silos and develop strategic data collecting campaigns

Box 2. The TOPIC (Traits of Plants in Canada) database: working together to integrate discipline-specific datasets

Why: Before the creation of TOPIC in 2009, Canadian plant species were severely underrepresented in the international trait database, TRY. Researchers seeking to apply the trait approach to Canadian plant communities had to single-handedly document functional traits for their species of interest. In the absence of a centralized, integrated repository in which to index them, data aggregated in this way had a short lifespan, and efforts were duplicated by independent research teams unaware of work already done elsewhere.

How does it work: Data contributed by members are archived in a structured pool of literature-based data, or in a repository of empirical geolocalized measurements. TOPIC acts as a hub for Canadian researchers to integrate small datasets and make them available locally and to the international scientific community via its collaboration with the international TRY database (Kattge et al., 2020). Support to standardize and document data is also offered.

Website: <http://cfs.cloud.nrcan.gc.ca/ctn/topic.php>



around complex questions requiring transdisciplinary solutions. In these three cases, interoperability with other databases and data sharing among participants and disciplines are facilitated by the use of common data management standards.

4 | BENEFITS OF GRASSROOTS INITIATIVES

4.1 | Maintain proximity for greater involvement, recognition and data openness

An important sociocultural hurdle in convincing ecologists to share their data is the fear that the data will be interpreted or used beyond their applicability, rendering the data provider an involuntary participant in flawed science (Michener, 2015; Mills et al., 2015; Tenopir et al., 2015). To ensure appropriate further use, the nature of ecological data, as well as its limitations, are critical information that must be preserved during data integration. The small-scale collaborative context of grassroots initiatives can alleviate some of these concerns via proximity, dialogue and trust among data generators and users. Not only can help and guidance be offered and sought more directly, errors are also more easily identified locally,

which greatly improves data quality. These initiatives can also set sharing guidelines adapted to their members' needs and ensure that they are respected. Ultimately, this results in more easily interpretable data, higher quality archival, and enhanced discoverability and interoperability. The proximity with the data generator also promotes new collaborations, providing tangible and immediate benefits to the participants. Even a simple recognition of the effort (e.g. thank you email, Serra-Diaz, Enquist, Maitner, Merow, & Svenning, 2017) can go a long way, acknowledging that data acquisition and curation represent a substantial investment of time and money. When these sociocultural challenges are faced, then technical difficulties can be addressed.

4.2 | Foster individual data stewardship

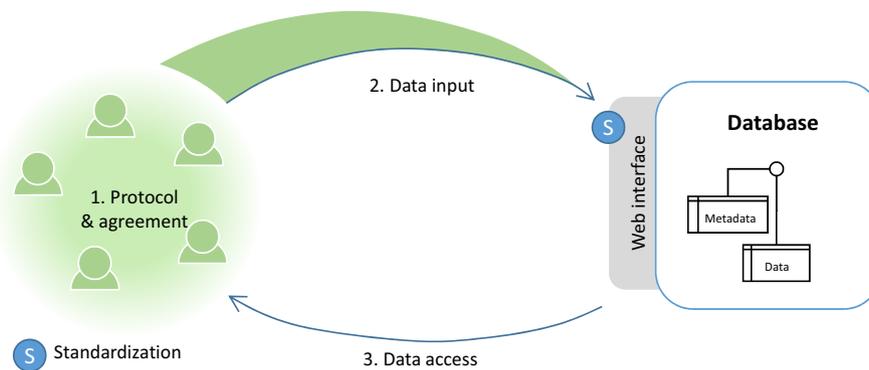
Despite a willingness to share data, lack of experience with data management can be a major deterrent for many ecologists (Hampton et al., 2017; Michener, 2015). Compounding the problem, literacy in new ecoinformatic tools remains a central challenge (Michener & Jones, 2012), and keeping up with best practices can seem like a moving target for most researchers (Wilson et al., 2014). Grassroots initiatives have benefited from advances in ecoinformatics, open

Box 3. The Co-VITAS project (Collaboration on Intraspecific Trait Variability of Above and Belowground Traits): working together to answer large-scale research questions

Why: Ubiquitous North American understory species were identified by a multidisciplinary working group as important for predicting species persistence under global change. Their wide geographical range and ecological breadth as well as the strong climatic gradients to which they are exposed make them perfect candidates for a large-scale study on intraspecific trait variability, a crucial component of their response to global change. This study involved systematic sampling of several traits across the species' range, a task well beyond the capabilities of a single laboratory. Collaborative science was the best way to overcome the prohibitive logistical field and laboratory requirements associated with answering this question.

How does it work: Following a general call for participants, 21 research teams from 18 institutions accepted to collect data for this project at their existing field sites. They received protocols and instructional videos detailing how to collect trait and ancillary environmental data. Field work was conducted at a total of 81 locations over 5,000 km within the same two-week period in July 2014. Sample processing tasks were split between participants for simple and easily standardized measurements, and a few core laboratories for tasks requiring specialized equipment or expertise. Data entry into an online database was performed by each participating team. Following the initial data management plan, archival of the collaborative dataset was achieved after data quality control. Agreement was reached to make the data available for reuse after publication through a disciplinary database and global-scale repository.

Reference: Kumordzi et al., (2019).



science and data-intensive science (Cheruvellil & Soranno, 2018). Several conceptual and applied tools have recently been developed to facilitate dataset aggregation (i.e. collecting together) and integration (i.e. merging into a single dataset) toward the goal of making ecological data more FAIR (Findable, Accessible, Interoperable and Reusable; Wilkinson et al., 2016). For instance, machine-readable metadata formats (e.g. EML, Ecological Metadata Language; Jones et al., 2019) and ontologies (e.g. OBOE for ecology; Madin, Bowers, Schildhauer, & Jones, 2008) help organize ecological information using logical relationships among domain-specific terms with controlled vocabularies and standardized definitions (e.g. T-SITA for soil invertebrate traits; Pey et al., 2014; Thesaurus of Plant Characteristics; Garnier et al., 2017). New technologies like Blockchain (Dai et al., 2018) in combination with artificial intelligence (Gil, Greaves, Hender, & Hirsh, 2014) may open new ways of managing data in the future, but most ecologists will need improved data management literacy levels to implement them. Collaborative settings improve literacy in, and resource allocation to, data management by leveraging resources from the various members and

offering a responsive platform via which data management questions can be addressed.

4.3 | Better handle complexity to facilitate interoperability

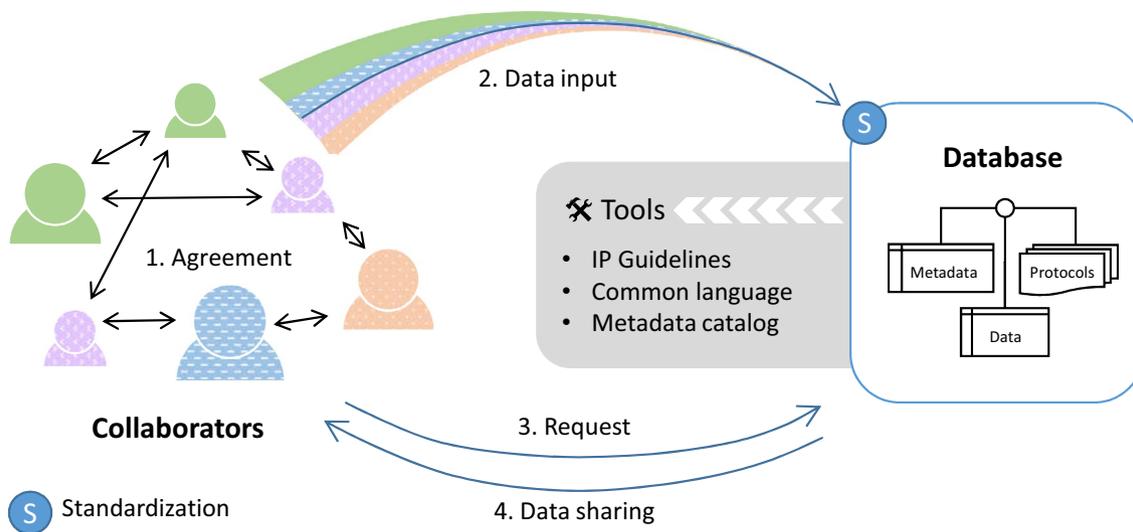
By facilitating data description for smaller groups of researchers with closer disciplinary, cultural or geographical ties, grassroots initiatives can serve as a critical link between individual labs and global-scale repositories. More cohesive, and based on close professional relationships, these groups focus on integration at a level at which the complexity of ecological data can still be tackled. Solutions proposed from the bottom up and developed through consensus ensure a better fit than data standards imposed through top-down mechanisms, thus improving buy-in. In collaborative settings, a structure can be developed that both makes sense to the data provider and supports dataset integration into global initiatives. Grassroots initiatives are therefore in a better position than top-down-driven projects to

Box 4. The Island Lake Biomass Harvest Experiment: complex question requiring a multidisciplinary team

Why: To address sustainability issues of harvesting for bioenergy, a multidisciplinary project was established in Ontario (Canada). The research team includes soil scientists, foresters and biodiversity scientists working on taxa ranging from microbes to arthropods and plants, as well as other collaborators from the private and public sector. To facilitate data integration, an ecoinformatic platform and a series of tools were developed at the start of the project, favouring interoperability and data exchange among scientists from different disciplines. The sociocultural aspects of data exchange were also taken in consideration, emphasizing the importance of continued communication and intellectual property guidelines. Well planned data management and continued communication preserve the “proximity” between the data generator and data end-user, ensuring data quality as well as appropriate use.

How does it work: Given the varied nature of the data collected, this project has made a concerted effort to preserve data in a way that future projects can track and build upon. Data management was central at all stages of the project and includes catalogues of protocols and metadata. This effort facilitates data discovery and reuse, but also makes transdisciplinary research possible.

Reference: Kwiaton et al., (2014).



ensure that participants are “speaking the same language” when it comes to nomenclature (Herrando-Pérez, Brook, & Bradshaw, 2014), definitions (i.e. thesaurus, see Garnier et al., 2017), protocols (e.g. Pérez-Harguindeguy et al., 2013) and, ultimately, research outcomes. Integration with global initiatives is easier when datasets are coherent, well described and understandable at the scale at which they were collected. Complexity that is well handled locally sets the foundation for high-quality integration at larger scales.

This groundwork is critical if ecology is to overcome individual dataset singularities to achieve integration without loss of information. However, in order to break the insularity of ecological data, dataset documentation should also take into account the broader context. Here, interoperability is key (Lowndes et al., 2017). These smaller initiatives must be well connected to larger-scale consortia to accurately prepare for larger-scale integration. The ecological community as a whole still lacks mechanisms to adequately link small and large-scale initiatives, a matter which should be considered carefully in the development of cyberinfrastructures for ecology.

5 | CONCLUSION

To work toward predictive ecology and actionable science, ecologists need to embrace data management approaches that also preserve the unique nature of ecological data. One way to achieve this is for researchers to get involved in grassroots initiatives that benefit their research program while participating in the larger data-sharing momentum in ecology. Despite the flexibility of these collaborative structures, some common requirements for success emerge: (a) a data management plan; (b) data archiving and description standards; (c) dedicated resources to project coordination and data provider support; (d) a clear objective/mandate; and (e) an explicit link to global data initiatives.

Beyond the aggregation of datasets, integrative science also requires a new kind of collaboration among researchers. Data sharing and integration is not only a question of data, it is first and foremost a question of individual scientists communicating and interacting among themselves. Substantial changes must be made to the way that we manage data. Based on our experience, we are convinced

that grassroots collaborative initiatives will be an important agent of such changes. Only then can knowledge stemming from these historically discrete fields efficiently address today's ecological questions.

ACKNOWLEDGEMENTS

We would like to thank the members of the TOPIC and CoVITAS initiatives, as well as the scientists involved in the Island Lake Biomass Harvest experiment. This commentary synthesizes hours of constructive discussion with them. Thanks to S. Gachet for comments on an earlier version of the manuscript. This work was supported by the Canadian Forest Service, Natural Resources Canada.

AUTHOR CONTRIBUTIONS

IA and FC led the writing of the manuscript. All authors contributed to development of the ideas, provided critical feedback on the drafts and gave final approval for publication.

DATA AVAILABILITY STATEMENT

No data were used in this paper.

ORCID

Isabelle Aubin  <https://orcid.org/0000-0002-5953-1012>
 Françoise Cardou  <https://orcid.org/0000-0002-6527-9212>
 Laura Boisvert-Marsh  <https://orcid.org/0000-0002-0939-8196>
 Eric Garnier  <https://orcid.org/0000-0002-9392-5154>
 Manuella Strukelj  <https://orcid.org/0000-0002-0391-2051>
 Alison D. Munson  <https://orcid.org/0000-0001-6013-7998>

REFERENCES

- Aubin, I., Munson, A. D., Cardou, F., Burton, P. J., Isabel, N., Pedlar, J. H., ... McKenney, D. (2016). Traits to stay, traits to move: A review of functional traits to assess sensitivity and adaptive capacity of temperate and boreal trees to climate change. *Environmental Reviews*, 24, 164–186. <https://doi.org/10.1139/er-2015-0072>
- Barlow, J., Stephens, P. A., Bode, M., Cadotte, M. W., Lucas, K., Newton, E., ... Pettoelli, N. (2018). On the extinction of the single-authored paper: The causes and consequences of increasingly collaborative applied ecological research. *Journal of Applied Ecology*, 55, 1–4. <https://doi.org/10.1111/1365-2664.13040>
- Borgman, C. L., Wallis, J. C., & Enyedy, N. (2007). Little science confronts the data deluge: Habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries*, 7, 17–30. <https://doi.org/10.1007/s00799-007-0022-9>
- Borner, K., Contractor, N., Falk-Krzesinski, H. J., Fiore, S. M., Hall, K. L., Keyton, J., ... Uzzi, B. (2010). A multi-level systems perspective for the science of team science. *Science Translational Medicine*, 2, 49cm24. <https://doi.org/10.1126/scitranslmed.3001399>
- Bruelheide, H., Dengler, J., Jiménez-Alfaro, B., Purschke, O., Hennekens, S. M., Chytrý, M., ... Zverev, A. (2019). sPlot – a new tool for global vegetation analyses. *Journal of Vegetation Science*, 30, 161–186. <https://doi.org/10.1111/jvs.12710>
- Cheruvilil, K. S., & Soranno, P. A. (2018). Data-intensive ecological research is catalyzed by open science and team science. *BioScience*, 68, 813–822. <https://doi.org/10.1093/biosci/biy097>
- Costello, M. J., Michener, W. K., Gahegan, M., Zhang, Z.-Q., & Bourne, P. E. (2013). Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology & Evolution*, 28, 454–461. <https://doi.org/10.1016/j.tree.2013.05.002>
- Dai, H., Young, H. P., Durant, T. J., Gong, G., Kang, M., Krumholz, H. M., Jiang, L. (2018) TrialChain: A Blockchain-Based Platform to Validate Data Integrity in Large, Biomedical Research Studies. arXiv preprint arXiv. <http://arxiv.org/abs/1807.03662>
- Garnier, E., Stahl, U., Laporte, M.-A., Kattge, J., Mougnot, I., Kühn, I., ... Klotz, S. (2017). Towards a thesaurus of plant characteristics: An ecological contribution. *Journal of Ecology*, 105, 298–309. <https://doi.org/10.1111/1365-2745.12698>
- Gil, Y., Greaves, M., Hendler, J., & Hirsh, H. (2014). Amplify scientific discovery with artificial intelligence. *Science*, 346, 171–172. <https://doi.org/10.1126/science.1259439>
- Global Biodiversity Information Facility (2018). What is GBIF? Retrieved from <https://www.gbif.org/what-is-gbif>
- Hampton, S. E., Jones, M. B., Wasser, L. A., Schildhauer, M. P., Supp, S. R., Brun, J., ... Aukema, J. E. (2017). Skills and knowledge for data-intensive environmental research. *BioScience*, 67, 546–557. <https://doi.org/10.1093/biosci/bix025>
- Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., ... Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, 11, 156–162. <https://doi.org/10.1890/120103>
- Heffernan, J. B., Soranno, P. A., Angilletta, M. J., Buckley, L. B., Gruner, D. S., Keitt, T. H., ... Weathers, K. C. (2014). Macrosystems ecology: Understanding ecological patterns and processes at continental scales. *Frontiers in Ecology and the Environment*, 12, 5–14. <https://doi.org/10.1890/130017>
- Herrando-Pérez, S., Brook, B. W., & Bradshaw, C. J. A. (2014). Ecology needs a convention of nomenclature. *BioScience*, 64, 311–321. <https://doi.org/10.1093/biosci/biu013>
- iDigBio (2018). Welcome to iDigBio. Retrieved from https://www.idigbio.org/wiki/index.php/Welcome_to_iDigBio
- Jones, M. B., O'Brien, M., Mecum, B., Boettiger, C., Schildhauer, M., Mitchell, M., Chong, S. (2019) Ecological Metadata Language version 2.2.0 KNB Data Repository. <https://doi.org/10.5063/F11834T2>
- Kattge, J., Bönsch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., ... Wirth, C. (2020). TRY plant trait database – enhanced coverage and open access. *Global Change Biology*, 26, 119–188. <https://doi.org/10.1111/gcb.14904>
- Kumordzi, B. B., Aubin, I., Cardou, F., Shipley, B., Violle, C., Johnstone, J., ... Munson, A. D. (2019). Geographic scale and disturbance influence intraspecific trait variability in leaves and roots of North American understorey plants. *Functional Ecology*, 33, 1771–1784. <https://doi.org/10.1111/1365-2435.13402>
- Kwiaton, M., Hazlett, P., Morris, D., Fleming, R., Webster, K., Venier, L., & Aubin, I. (2014) *Island Lake Biomass Harvest Research and Demonstration Area: Establishment Report*. Information Report GLC-X-11. Sault Sainte Marie, ON: Natural Resources Canada, Great Lakes Forestry Centre.
- LaDeau, S. L., Han, B. A., Rosi-Marshall, E. J., & Weathers, K. C. (2017). The next decade of Big Data in ecosystem science. *Ecosystems*, 20, 274–283. <https://doi.org/10.1007/s10021-016-0075-y>
- Lowndes, J. S. S., Best, B. D., Scarborough, C., Afflerbach, J. C., Frazier, M. R., O'Hara, C. C., ... Halpern, B. S. (2017). Our path to better science in less time using open data science tools. *Nature Ecology & Evolution*, 1, 160. <https://doi.org/10.1038/s41559-017-0160>
- Madin, J. S., Bowers, S., Schildhauer, M. P., & Jones, M. B. (2008). Advancing ecological research with ontologies. *Trends in Ecology & Evolution*, 23, 159–168. <https://doi.org/10.1016/j.tree.2007.11.007>
- Michener, W. K. (2015). Ecological data sharing. *Ecological Informatics*, 29, 33–44. <https://doi.org/10.1016/j.ecoinf.2015.06.010>

- Michener, W. K., & Jones, M. B. (2012). Ecoinformatics: Supporting ecology as a data-intensive science. *Trends in Ecology & Evolution*, 27, 85–93. <https://doi.org/10.1016/j.tree.2011.11.016>
- Mills, J. A., Teplitsky, C., Arroyo, B., Charmantier, A., Becker, P. H., Birkhead, T. R., ... Zedrosser, A. (2015). Archiving Primary Data: Solutions for Long-Term Studies. *Trends in Ecology & Evolution*, 30, 581–589. <https://doi.org/10.1016/j.tree.2015.07.006>
- Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P., ... Cornelissen, J. H. C. (2013). New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany*, 61, 167–234. <https://doi.org/10.1071/BT12225>
- Pey, B., Laporte, M.-A., Nahmani, J., Auclerc, A., Capowiez, Y., Caro, G., ... Hedde, M. (2014). A thesaurus for soil invertebrate trait-based approaches. *PLoS ONE*, 9, e108985–e108985. <https://doi.org/10.1371/journal.pone.0108985>
- Proença, V., Martin, L. J., Pereira, H. M., Fernandez, M., McRae, L., Belnap, J., ... van Swaay, C. A. M. (2017). Global biodiversity monitoring: From data sources to Essential Biodiversity Variables. *Biological Conservation*, 213, 256–263. <https://doi.org/10.1016/j.biocon.2016.07.014>
- Ramirez, K. S., Berhe, A. A., Burt, J., Gil-Romera, G., Johnson, R. F., Koltz, A. M., ... Tuff, K. (2018). The future of ecology is collaborative, inclusive and deconstructs biases. *Nature Ecology & Evolution*, 2, 200. <https://doi.org/10.1038/s41559-017-0445-7>
- Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *Science*, 331, 703–705. <https://doi.org/10.1126/science.1197962>
- Roche, D. G., Kruuk, L. E. B., Lanfear, R., & Binning, S. A. (2015). Public data archiving in ecology and evolution: How well are we doing? *Plos Biology*, 13, e1002295. <https://doi.org/10.1371/journal.pbio.1002295>
- Roche, D. G., Lanfear, R., Binning, S. A., Haff, T. M., Schwanz, L. E., Cain, K. E., ... Kruuk, L. E. B. (2014). Troubleshooting public data archiving: Suggestions to increase participation. *PLoS Biology*, 12, e1001779. <https://doi.org/10.1371/journal.pbio.1001779>
- Rüegg, J., Gries, C., Bond-Lamberty, B., Bowen, G. J., Felzer, B. S., McIntyre, N. E., ... Weathers, K. C. (2014). Completing the data life cycle: Using information management in macrosystems ecology research. *Frontiers in Ecology and the Environment*, 12, 24–30. <https://doi.org/10.1890/120375>
- Schwartz, M. W., Hiers, J. K., Davis, F. W., Garfin, G. M., Jackson, S. T., Terando, A. J., ... Brunson, M. W. (2017). Developing a translational ecology workforce. *Frontiers in Ecology and the Environment*, 15, 587–596. <https://doi.org/10.1002/fee.1732>
- Serra-Diaz, J. M., Enquist, B. J., Maitner, B., Merow, C., & Svenning, J.-C. (2017). Big data of tree species distributions: How big and how good? *Forest Ecosystems*, 4, 30. <https://doi.org/10.1186/s40663-017-0120-0>
- Seyfang, G., & Smith, A. (2007). Grassroots innovations for sustainable development: towards a new research and policy agenda. *Environmental Politics*, 16, 584–603. <https://doi.org/10.1080/09644010701419121>
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., ... Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE*, 10, e0134826–e0134826. <https://doi.org/10.1371/journal.pone.0134826>
- Urban, M. C., Bocedi, G., Hendry, A. P., Mihoub, J.-B., Peer, G., Singer, A., ... Travis, J. M. J. (2016). Improving the forecast for biodiversity under climate change. *Science*, 353, aad8466–aad8466. <https://doi.org/10.1126/science.aad8466>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Wilson, G., Aruliah, D. A., Brown, C. T., Chue Hong, N. P., Davis, M., Guy, R. T., ... Wilson, P. (2014). Best practices for scientific computing. *Plos Biology*, 12, e1001745–e1001745. <https://doi.org/10.1371/journal.pbio.1001745>
- Wiser, S. K. (2016). Achievements and challenges in the integration, reuse and synthesis of vegetation plot data. *Journal of Vegetation Science*, 27, 868–879. <https://doi.org/10.1111/jvs.12419>
- Zimmerman, A. S. (2008). New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology, & Human Values*, 33, 631–652. <https://doi.org/10.1177/0162243907306704>

How to cite this article: Aubin I, Cardou F, Boisvert-Marsh L, Garnier E, Strukelj M, Munson AD. Managing data locally to answer questions globally: The role of collaborative science in ecology. *J Veg Sci*. 2020;31:509–517. <https://doi.org/10.1111/jvs.12864>