



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/244558/>

Version: Published Version

Article:

Buckley, Y.M., Finn, A., Molloy, A. et al. (2026) Quality assurance framework for the design, collection and curation of standard data products from a distributed ecology network. Ecology and Evolution, 16 (8). e74000. ISSN: 2045-7758

<https://doi.org/10.1002/ece3.74000>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

RESEARCH ARTICLE OPEN ACCESS

Quality Assurance Framework for the Design, Collection and Curation of Standard Data Products From a Distributed Ecology Network

Yvonne M. Buckley¹  | Alain Finn²  | Aoife Molloy¹  | Anna Mária Csergő³  | Elizabeth Crone⁴  | Annabel L. Smith⁵  | Bret D. Elderd⁶  | Caroline McKeon¹  | Deborah Roach⁷  | Dylan Z. Childs⁸  | Glenda M. Wardle⁹  | María B. García¹⁰  | Maude Baudraz¹¹  | Rachel M. Penczykowski¹²  | Roberto Salguero-Gómez¹³  | Ruth Kelly¹⁴  | Satu Ramula¹⁵  | Anna Liisa Laine¹⁶  | Simone Blomberg⁵  | Johan Ehrlén¹⁷  | Jesús Villellas¹⁸  | Aveliina Helm¹⁹  | Christiane Roscher^{20,21}  | Krista Raveala¹⁶ | Liv Norunn Hamre²²  | Lotte Korell^{21,23}  | Francis Q. Brearley²⁴  | Siri Lie Olsen^{25,26}  | Astrid Wingler²⁷  | Joslin L. Moore^{28,29}  | Ayco J. M. Tack³⁰  | Anna Bucharova³¹  | Olav Skarpaas³²  | Judit Bodis³³  | Aldo Compagnoni^{21,34}  | Sergi Munne-Bosch^{35,36}  | Zuzana Munzbergova^{37,38}  | Joachim Töpper³⁹  | Peter Vesk⁴⁰  | Melanie Morales^{35,36,41}  | Lizzie Wandrag⁴²  | Marjo Saastamoinen¹⁶ | Martin Andrzejak⁴³  | Pil Rasmussen⁴⁴  | Christina Caruso⁴⁵  | Richard Duncan⁴⁶  | John Dwyer⁴⁷  | Matthew Coghill⁴⁸  | Ronny Groenteman⁴⁹  | Meelis Pärtel¹⁹  | William K. Petry⁵⁰  | Michele Lonati⁵¹  | Javier Puy⁵²  | Jane Catford^{53,54}  | Trevor H. Drees⁵⁵  | Christoph Rosche^{21,56}  | Katriona Shea⁵⁷  | Simone Ravetto Enri⁵¹  | Jennifer R. Gremer⁵⁷  | Balázs Deák⁵⁸  | Lauchlan Fraser⁴⁸  | Lauri Laanisto⁵⁹  | Orsolya Valkó⁵⁸  | Robert Drew Siegel⁶⁰  | Benedicte Bachelot⁶¹  | Jennifer Finn⁶² 

Correspondence: Yvonne M. Buckley (buckleyy@tcd.ie)

Received: 4 March 2026 | **Revised:** 24 June 2026 | **Accepted:** 30 June 2026

Keywords: big data | collaborative | coordinated distributed network | data management | data product | demography | macroecology | team science

ABSTRACT

1. Ecological data are increasingly collected by networks of collaborators using replicated designs and methods, which can significantly improve the quality and quantity of data throughout the ecological niche and geographic range of species or communities. The coordinated generation and management of data is critical for producing datasets (Standard Data Products) across multiple sites that can be used by different researchers, over extended time periods and for multiple purposes.
2. We describe and use a Quality Assurance framework for the design, collection and production of reproducible Standard Data Products for distributed ecology projects. We identified six critical project elements of a Quality Assurance framework (QA1–6) to produce ecological Standard Data Products with high immediate and future value.
3. We applied the Quality Assurance framework to the Plantpopnet project as a case-study. Plantpopnet is a coordinated distributed system for population macroecology using the model species *Plantago lanceolata*. We mapped Plantpopnet activities to the Quality Assurance Framework as follows: (QA1) *Measurable objectives*: research project objectives with data requirements, (QA2) *Process control*: governance policies, (QA3) *Project specific procedures*: model organism selection and data collection protocol, (QA4) *Supporting production of high quality data*: recruitment, retention and engagement of participants, (QA5) *Data management*: data management plan and reproducible data cleaning workflow, (QA6) *Production and management of outputs*: Standard Data Products and papers.
4. Explicit use of Quality Assurance, project and data management tools together with standardised ecological methods facilitated the design, collection, maintenance and sustainability of high-quality data products. We provide a Quality Assurance

For affiliations refer to page 11.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Ecology and Evolution* published by British Ecological Society and John Wiley & Sons Ltd.

framework together with governance documents, code and data for a reproducible Standard Data Product. This framework supports a distributed funding model which can be sustainably applied to facilitate future research and applications of coordinated distributed ecology projects.

1 | Introduction

Understanding and predicting processes at large spatial extents and over time underpins ecology, evolution and conservation. Research in ecology is increasingly international, across a species' range (Buckley and Puy 2022) or ecosystem (Borer et al. 2014, Mirtl et al. 2018). Spatially extensive datasets enable the understanding of context dependency of ecological processes across sites and systems (Catford et al. 2022). However, the intensity of field work required for spatially extensive data collection has historically restricted its scope (Buckley et al. 2010; Coutts et al. 2016). One solution to this challenge is distributed ecology projects where networks of researchers address questions at large spatial extents, reducing the intensity of field work for contributors e.g., NutNet (Borer et al. 2014), GrENE-net (Czech et al. 2022), HerbVar (Herbivory Variability Network et al. 2023), iCONNECT (Lucas et al. 2024), GLUE (Santangelo et al. 2022) and Plantpopnet, a distributed population ecology project focussed on the model species *Plantago lanceolata* L. (Figure 1).

Spatially extensive, replicable and reproducible systems can eliminate or minimise apparent context dependency arising from differences in statistical inference or methodological differences and offer an opportunity for learning about replicability and generalisability in ecology (Catford et al. 2022). We contend that Quality Assurance (QA) processes can further minimise apparent context dependency, enabling ecologically meaningful context to be more



FIGURE 1 | A population of *Plantago lanceolata* L. (ribwort plantain) the focal species for the Plantpopnet distributed population ecology project, at Portrane Co Dublin Ireland.

clearly discerned. QA refers to an ongoing, continuous process of monitoring, evaluating, maintaining and enhancing a product or service (QQI 2021), enabling the consistent production of a product or service that minimises defects and increases utility. QA management frameworks are commonly used across industries including environmental science (CEN 2015a; Goldberg et al. 2016; Kim et al. 2022; Sutter et al. 2015). Here we use QA principles throughout the full life-cycle of a coordinated distributed ecology project to define objectives, ensure good governance, develop appropriate and widely applicable protocols, support participation and curate data through reproducible data cleaning, auditing, versioning and the production of Standard Data Products that can be widely used. Standard Data Products (SDP) are validated and quality assured data sets that are well documented and shareable. Research objectives such as FAIR data management (Wilkinson et al. 2016) or the creation and management of large networks such as the International Long Term Ecological Research system can be nested within a Quality Assurance system.

QA increases confidence that quality requirements will be fulfilled through a set of established procedures or processes that ensure: (QA1) *Measurable objectives* are supported by the data, where the proposed data is fit for the objectives of the project, (QA2) *Process controls* are in place. Process in this instance refers to interacting activities that use inputs to deliver an intended result. Process controls are methods and protocols implemented to ensure that the operation and control of these processes are effective, (QA3) *Project specific procedures* are used for accurate data collection, defined as specified ways to carry out activities (or processes), (QA4) *Supports for data collectors* are in place to enable collection of high quality data, (QA5) *Data management* is used to control errors and ensure data integrity during transcription, transfer and storage and (QA6) *Production and dissemination of high quality outputs*, with outputs defined as the result of a process (CEN 2015a). Quality Control (QC) includes technical checks to ensure the product (in this case, data) fulfils quality requirements. Previous work on quality aspects of ecological data collection have tended to focus on individual components of a QA system, such as data management (e.g., Borer et al. 2009; Kim et al. 2022) or the development of modelling standards (e.g., Araújo et al. (2019)), protocols, or networks; but see Borer et al. (2014) for an integrated approach.

Distributed projects often involve a common protocol for data collection, standard methods and consistent reporting of data using templates. These are useful elements of a more comprehensive QA process, but are not sufficient. Distributed collaborative research (team science) in ecology typically uses a common experimental or observational protocol deployed across multiple locations by many researchers (Alexander et al. 2011; Borer et al. 2014; Henry and Molau 1997) run in parallel across geographic scales to test ecological questions (Pastor et al. 2020). Data, samples or both, are collected by participants from multiple sites and used to answer research questions at both the fine grain needed to understand local ecological processes and the large spatial extent needed for contextualisation

and generalisation (Wiens 1989). Team science can widen participation in ecological research, enable sharing of knowledge, expertise and costs among contributors, reduce greenhouse gas emissions and other costs associated with international research (Diagne et al. 2021), enable simultaneous sampling across large areas, and increase the resilience of research to global shocks. However, there are significant challenges in ensuring that the resulting data have sufficient quality, longevity, reproducibility, accessibility and utility to answer the targeted research questions, as well as future value.

The compilation of data collected by collaborating researchers at large spatial and temporal scales has a long history (Heberling et al. 2021; Henry and Molau 1997). However, managing effective distributed networks goes beyond implementing data quality checking. Networks must ensure quality is achieved across project management, team engagement, data curation and the reproducibility of data products used as inputs to research projects (Borer et al. 2014; Maund et al. 2022). Replicability in coordinated distributed networks is built into the observational or experimental protocols, standardised across sites. However, methods for curating and generating reproducible data products from multiple contributions are rarely presented (Hampton et al. (2013); Buckley et al. (2025); but see Herbivory Variability Network (2023) and Lind (2016)).

Here we describe a general QA framework which provides a template for the design of distributed ecology projects. We illustrate this QA framework for a distributed population ecology network, Plantpopnet [www.plantpopnet.com], and demonstrate its utility from inception to the production of a Standard Data Product (SDP) ready for use by data consumers (researchers in this case). The production of SDPs facilitates reproducibility of analyses and enables citation of specific instances of evolving data sources. To illustrate our QA approach, we mapped Plantpopnet activities to the QA framework. We provide a QA workflow for reproducible SDPs and assess error rates across sites.

1.1 | Plantpopnet Case-Study

Plantpopnet uses the perennial short-lived herb *Plantago lanceolata* L. (Plantaginaceae) as a model organism to understand range-wide biotic and abiotic drivers of population change (Penczykowski and Sieg 2021). Since 2014, 75 site coordinators in 17 countries and 3 continents have submitted data from 65 populations (Figure 2). The Plantpopnet project enables the development and testing of methods for a biogeographical approach to population ecology based on a unique, spatiotemporally replicated demographic dataset of plants and animals that can be applied to other systems (Alexander et al. 2012; Baudraz et al. 2025; Buckley and Puy 2022; Kuiper and Bos 1992; Penczykowski and Sieg 2021; Sagar and Harper 1961; Smith et al. 2020; Villellas et al. 2021).

1.2 | Quality Assurance Framework

A QA framework for coordinated, distributed ecology projects was adapted from fundamental principles and requirements

of ISO Quality Management Systems (CEN 2015a, 2015b). We outline the application of each QA objective to Plantpopnet and explain how the framework evolved in line with QA guidelines for adaptability and flexibility (an overview is provided in Table 1).

1.2.1 | QA1: Measurable Objectives Supported by Data

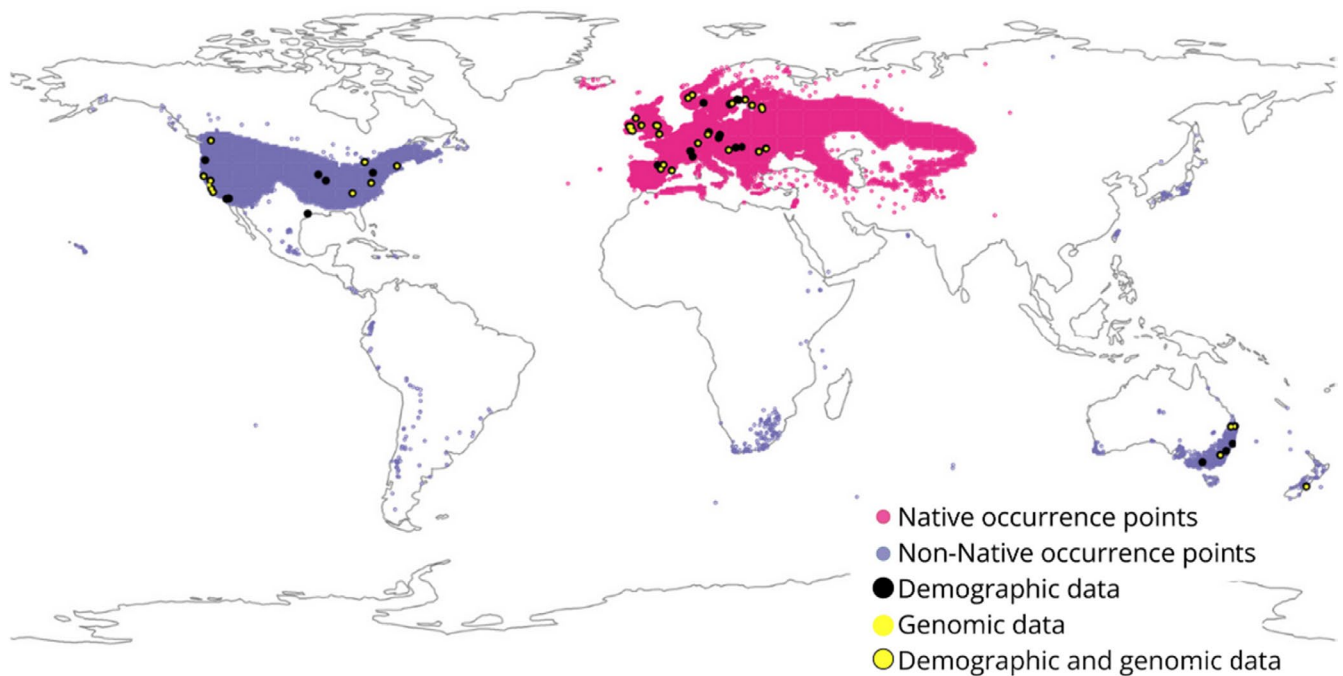
Clear definition of research questions and appropriate research design is imperative (Hayward et al. 2015). Co-design and collaborative processes are particularly useful in distributed ecological networks for defining research objectives (Table 1, PPN1). Engagement of participants ensures that objectives and procedures are robust and that the commitment of participants is secured (McCord et al. 2022). A core set of common research objectives can be supplemented with 'add-on' research objectives that may require additional data collection and/or a more limited number of sites for specific observations or experiments (Borer and Seabloom 2025). For example, Villellas et al. (2021) used core data collected through the standard Plantpopnet protocol (46 populations) with add-on protocols at a subset of sites for seed collection (15 populations) subsequently used for an experiment. Add-on studies have many benefits, including allowing participants to enhance their own research by defining research questions, building on existing data and reducing costs. Add-on studies increase data collection, test new research questions, improve efficiency through utilising existing protocols and methodologies, and enhance collaboration and knowledge exchange.

Importantly, research objectives must be explicitly linked to sampling design, spatial coverage, temporal duration and planned analytical approaches. The ability of a network to address broad-scale ecological questions depends on whether sampled sites adequately represent the environmental, geographic and demographic gradients relevant to the study objectives. Under-representation of particular regions, habitat types or environmental conditions can limit the generalisability of findings and constrain inference regarding species responses across their full range (Borer and Seabloom 2025). Similarly, temporal coverage must be sufficient to capture the ecological processes of interest (Fraser et al. 2013) and support the statistical or modelling approaches proposed. During network design, objectives should therefore be evaluated against expected sampling coverage and data structure to ensure that the resulting dataset is capable of supporting robust inference. Where gaps in coverage remain unavoidable, these limitations should be explicitly recognised, and research questions, analytical methods or recruitment efforts adapted accordingly.

1.2.2 | QA2: Process Controls

Clear governance is needed to ensure transparent management of a coordinated distributed collaborative network (e.g., top-down versus bottom-up management (Aubin et al. 2020)) and should be developed early in the project and revised regularly. Decision making, funding, authorship and data availability processes ensure that the responsibilities and benefits of participation are clear to participants. Decision making may be

(a)



(b)

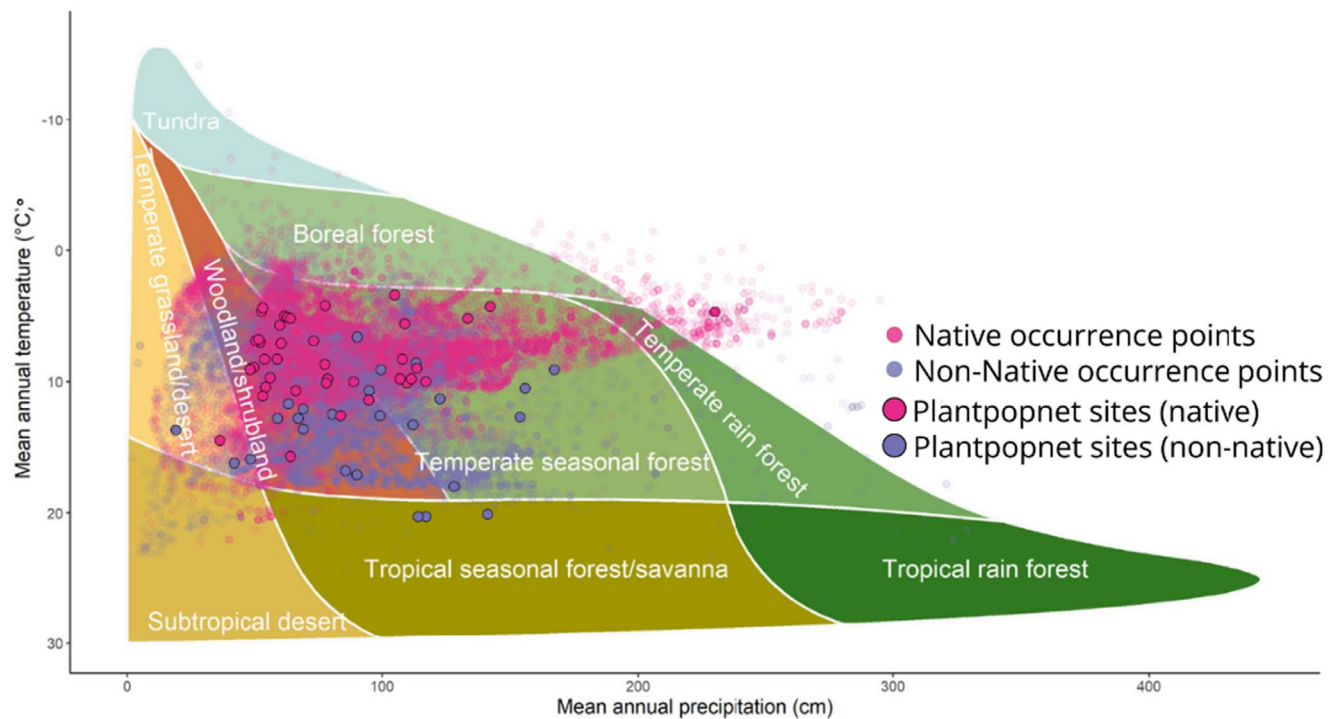


FIGURE 2 | (a) Map of 64 demographic data collection sites with at least one year of demographic data and 37 sites for both genomic and demographic data (Smith et al. 2020) throughout the native (small pink points) and non-native (small purple points) range of *Plantago lanceolata* (as recorded on GBIF (GBIF 2022) & BIEN (Enquist et al. 2016)). (b) an indicative position of the data collection sites in Whittaker Biomes. Each pink (native) and purple (non-native) point represents a data collection site and its corresponding biome. One population in Norway lies outside the definition of Whittaker's biome due to high local precipitation relative to its mean annual temperature. Climatic data in (b), shown as transparent points for *P. lanceolata* observations in GBIF and BIEN, is from WorldClim 2 (Fick and Hijmans 2017) and is produced using 'plotbiomes' (Stefan and Levin 2019).

undertaken by a subset of participants. However, careful consideration should be paid to geographic distribution across regions of research interest, a diverse suite of professional networks,

gender balance, under-represented groups and countries, early career participants and multidisciplinary expertise (Supporting Information S1: 1.1).

TABLE 1 | Framework for Quality Assurance of distributed ecology projects. We outline general Quality Assurance objectives (QA1–QA6), measurable indicators for Quality Control, specific examples of implementation of objectives for Plantpopnet (PPN1–PPN10, PPN referring to a Plantpopnet activity) and Quality Control (QC1–QC9) evidence is detailed for Plantpopnet activities. More information is available in the [Supporting Information](#).

Quality Assurance objectives	Measurable indicators for Quality Control	Plantpopnet activities (timeline of implementation)	Quality Control evidence
QA1: <i>Measurable objectives supported by data</i>	Number of research questions/research objectives % of research questions/objectives supported by data collected	PPN1: Development of research project abstracts with data requirements (2014)	QC1. Development & refinement of questions at workshops (2014, 2017, 2021)
QA2: <i>Process controls</i>	Number of policies supporting governance, recruitment and retention of participants, funding. % of active projects following network governance policy Number of steering committee meetings	PPN2: Policies for Network Governance (Supporting Information S1: 1.1), Authorship (Supporting Information S1: 1.1) (2014) PPN3: Policy for funding (Supporting Information S1: 1.3) (2019)	QC2. Review & updates at steering committee meetings
QA3: <i>Project-specific procedures</i>	Observational/experimental system in place (pass/fail) Protocol in place (pass/fail) % of sites applying protocol correctly Number of protocol reviews	PPN4: Model organism selection (2014) PPN5: Data collection protocol (2014)	QC3. Selection framework developed (2014) QC4. Network review of protocol (2014, 2019)
QA4: <i>Supports for data collectors to produce quality data</i>	Number of participants and sites recruited % participant and site retention Number of communications with participants Number of training events and resources provided % of participants engaged in training	PPN6: Recruitment, retention and engagement of participants for long term research (2014, 2015, 2016)	QC5. Analysis of participant & site data, including exit data. Review at participants' workshop (2021).
QA5: <i>Data management to control errors and ensure data integrity during transcription, transfer and storage</i>	Data Management Plan in place (pass/fail)	PPN7: General data management and storage processes (2017) PPN8: Reproducible data cleaning workflow & SDPs	QC6. Steering Committee reviews of data management plan QC7. Internal peer review of workflow & scripts (2019, 2020)
QA6: <i>Production & dissemination of high quality outputs</i>	Number of Standard Data Products (SDPs) Number of citations to SDPs Number of publications or other outputs using SDPs Number of citations to outputs using SDPs Number of SDP downloads	PPN9: Publication, dissemination and versioning of Standard Data Products (SDPs) PPN10: Publication of papers & reports (Baudraz et al. 2025 ; Buckley et al. 2019 ; Morales et al. 2021 ; Smith et al. 2020 ; Villellas et al. 2021)	QC8. Data audit of SDPs QC9. Internal & external manuscript peer review

TABLE 2 | Key considerations for protocol development.

Considerations	Justification for Plantpopnet
Provision of a clear methodology for individual plant level demographic monitoring over time	1. Persistence in plots for 3–5 years across a broad environmental gradient 2. Adaptable to different population densities and site conditions 3. Ease of finding transects, plots and individual plants from year to year 4. Minimally invasive monitoring to reduce the impact on population processes 5. Ensure minimum amounts of permits required, specifically consideration of permits required to work in sensitive areas
Accessibility of methods to maximise participation of voluntary participants	1. Flexible choice of site location 2. Inexpensive methodology 3. Clear and easy to follow methodology with data collection template spreadsheets 4. Up to 3 person/days time commitment per year for annual census 5. Modular design of core data collection protocol with optional 'add-on' opportunities for additional data

Process controls include management of funding. A variety of funding mechanisms can be used such as grants for: collaborative workshops, research coordination, data collection at individual sites and individual research questions. This distributed funding model enables flexible research and encourages wider participation by researchers. For Plantpopnet, a distributed funding model was adopted with individual sites largely self-funded. We conservatively estimated a total of 1.2 person years per year of operation has been invested in the collection and curation of core demographic data (equivalent to 8.35 years of one full-time person over 2014–2021, Supporting Information S2).

1.2.3 | QA3: Project-Specific Procedures

Outlining and confirming project procedures enhances data quality (Brown and Williams 2019). Strong project design with clear, shared project objectives and methodology enables the production of long-term, reproducible data which can often address multiple novel research questions (Borer et al. 2014; Fraser et al. 2013). Choice of the system of study is fundamental to the achievement of objectives, and the potential to leverage core data for other kinds of questions both in situ (local sites) and *ex situ* (growth chambers, common gardens) should be considered.

1.2.3.1 | PPN4: Model Organism Selection. Candidate species were assessed against indicators of key ecological and evolutionary attributes using expert knowledge and published literature before *P. lanceolata* was selected (Penczykowski and Sieg 2021) (Table 2).

1.2.3.2 | PPN5: Data Collection Protocol. The first version of the protocol was tested in 2014, revised, and provided to participants in 2015 (V1.01). Second and third revisions of the protocol were provided on FigShare for transparent version control, long-term archiving and citation purposes (Buckley et al. 2019).

Key considerations for protocol development included a clear methodology for demographic monitoring over time and accessibility of methods to maximise sustained participation (Table 2). Trade-offs arose as data for answering some demographic questions was deemed too challenging and/or time-consuming to collect. By prioritising data requirements, we developed a core protocol that every site must complete with additional data collection encouraged through optional 'add-on' studies. For example, the annual demographic census for Plantpopnet is obligatory, but collection of leaves for genetic analysis or seed for experiments was chosen by participants as 'add-on' elements depending on their research interests and capacity. This approach ensures a comprehensive core dataset combined with the agility to ask new research questions according to participant interests.

1.2.4 | QA4: Supporting Data Collectors to Produce Quality Data

The success of a spatially distributed, long-term team science project depends strongly on the network of participants; however, the methodology for achieving group engagement and retention is rarely described. Having clear roles and structures, workshops, training and support for members, goals that all members can contribute to, and having members at different career stages has been identified as contributing to successful team science projects (Cheruvelil et al. 2014; Love et al. 2021). Large-scale spatial projects necessarily require a diverse range of scientists from various geographic locations.

1.2.4.1 | PPN6: Recruitment, Retention and Engagement of Participants for Long Term Research. Several dimensions of sustainability were considered in the design of the research network including: long-term data collection, environmental sustainability, resources and long-term capacity to maintain data collection. Some objectives were achievable within a single or few years of data collection whereas other objectives required long-term (decadal) timescales. The

need for sustainability influenced the recruitment strategy. To encourage diverse participation the steering group consisted of researchers of different career stages, genders and geographic locations who approached potential participants, and promoted the network on social media and at conferences. We recruited participants of different career stages to ensure both short- and longer-term objectives were achievable. We encouraged participants to choose sites close to their place of work or home for ease of sustained data collection. Most participants were recruited in 2015 and 2016 with small numbers of sites joining after this time (Supporting Information S3). The measures/strategies that were taken to maximise participation resulted in 34% of 65 original sites still actively collecting data as of 2024.

Reasons for site abandonment ($n = 16$) included: change in site coordinator circumstance (38%), mass mortality of censused plants (19%), landowner's request (13%) and destruction of the site (accidental or otherwise) (25%). The highest rate of abandonment of sites occurred after 2 years of data collection, where 14% of sites with 2 years of data were abandoned (Supporting Information S4).

In 2018 site participants ($n = 41$) were surveyed online to understand the composition of the network and participants' motivations for joining a team science project (as in García et al. (2019)). 90% of participants completed the survey (see Supporting Information S5 for full 10 question list). Highly ranked reasons for joining included interest in demography (49% ranked it as most important), collaborative research and scientific advancement (41%), personal learning (22%) and network interactions (46%) achievable for relatively low commitment of time, cost and travel (14%) (Figure 3A).

Participants perceived the greatest risks to continuation to be changes in their employment (14%) and geographic location (19%), which is in accordance with the primary observed reason for site abandonment (change in participant circumstances) (Figure 3B; Supporting Information S5).

1.2.5 | QA5: Data Management to Control Errors and Ensure Data Integrity

Units of measurement, file naming and location of stored data should be consistent and agreed prior to commencing (Michener 2018). Detailed metadata with a README file, clearly and consistently labelled files in appropriate folders, clear steps for data processing and utilising centralised, community-maintained controlled-access repositories for metadata and code all contribute to effective data management (Jenkins et al. 2023). To control errors in the data cleaning and analysis processes, code should be written in a manner that enables clear communication, and a formatting style that is easily understandable and consistent (Buckley et al. 2025; Filazzola and Lortie 2022). Specialised apps can be used to control data collection errors, e.g., HerbVar has an app to train herbivory estimation (www.zaxherbivorytrainer.com) and PhenObs (phenological network using botanical gardens, <https://www.idiv.de/en/phenobs>) uses an app to standardise data collection, ensuring consistency in data recording.

1.2.5.1 | PPN7: General Data Management and Storage Processes. The primary investigator at each site (i.e., site coordinator) is responsible for adhering to the protocol (Buckley

et al. 2019) and providing digital data. Data files can be accompanied by images, physical samples, maps and notes.

In 2018, a data management plan for Plantpopnet was developed (Supporting Information S1: 1.4) based on the data management guidelines developed by Borer et al. (2009), detailing file storage, naming, data sharing and access procedures and a checklist for the data management workflow (Figure 4). A key challenge is recording changes made to the raw data before analysis. We took a scripting approach to recording data changes, to ensure transparency, to enable changes to be reversed, and to enable a record of edits to be generated. This approach ensured that the raw data could be converted to an SDP in a reproducible way, following recent data curation reproducibility guidelines (Buckley et al. 2025). Raw data files were not altered, thus any errors arising in compiled datasets could be traced back and queried against the original submission.

Initially, raw data files were manually verified to ensure correct protocols were followed, that data were consistent and complete, and the established criteria for QC results were met. Each verified raw file was then stored as a 'master' copy. Standardised master data for all sites and all years were combined, and the resulting dataset was error checked using scripts (Figure 4). Finally, the standardised dataset was used to produce SDPs corresponding to cumulative census years and distributed for analysis.

R scripts, R Markdown files detailing processes and changes, and data products associated with the cleaning process were maintained using version control software Git and online repository Github. Github stores the online working repository and version history of the cleaning scripts, whereas the scripts available on Zenodo under the Data Availability Statement are archived for release and reproducibility with this manuscript.

1.2.5.2 | PPN8: Reproducible Data Cleaning Workflow and Standard Data Products (SDPs). As datasets increase in size, so does the complexity of the curation process. A working group was formed to assemble scripts to identify and solve errors in the data. Each script tackled specific variables and was incorporated into an annotated markdown file. Project management was through Github, which was used to track errors and solutions and included internal peer review of scripts.

When passing large datasets through a series of scripts, it is possible to introduce new errors; therefore, a check of the standardised data is a vital step. A random subset of sites per year from the draft SDP was cross checked against the uncleaned master data using the R package 'compareDF' (Joseph 2021), enabling identification of differences between two similarly structured datasets. Sites were randomly selected using R's random number generator, with a fixed starting value (or 'seed') to ensure reproducibility of the sampling process.

All ten sites audited from the Y0 (Year 0, first year of data collection) SDP had some level of modification to the raw data: affecting 2 to 199 rows per site, representing 1.2% to 98% of rows in a site with at least one cell edited. Most were minor formatting or continuity errors, e.g., cells with 'NA' were converted to blank cells. However, six sites had substantial changes, with three of

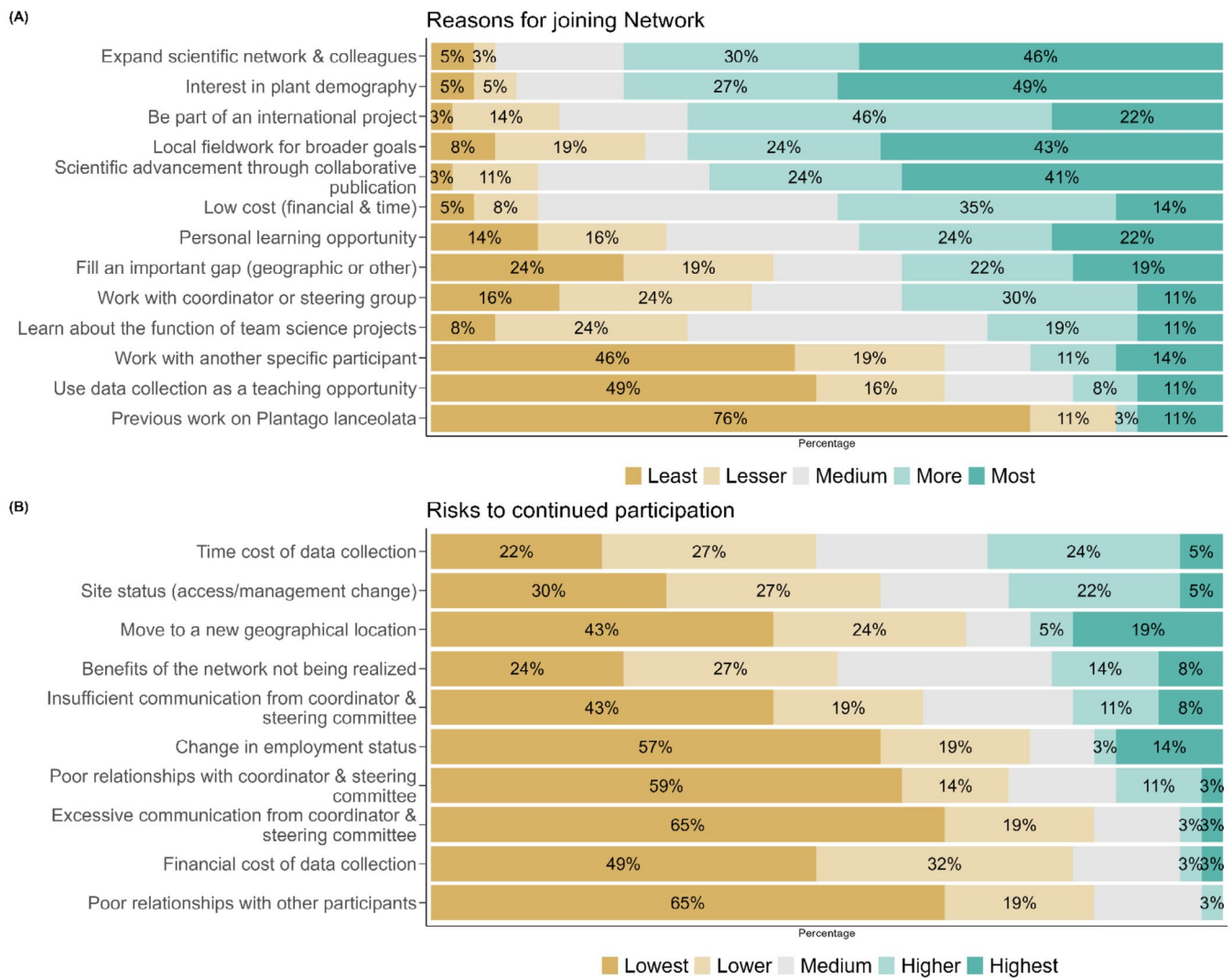


FIGURE 3 | (A) Reasons for joining Plantpopnet ranked from least important to most important & (B) Risks to continued participation in the Plantpopnet network ranked from lowest risk to highest risk. Responses are ordered by most important (using 'more' and 'most' categories) and highest risk (using 'higher' and 'highest' categories) respectively.

these sites requiring removal of some data. Substantial changes were defined as changes to data values, including removal of data not included in the protocol, correcting units and protocol deviations, such as recording multiple values in a single cell. The most common errors included columns which were removed or re-labelled from the protocol template data file, values outside defined ranges or incomplete records. Corrections in these six sites affected 2–68 rows (representing 1.2%–32% of rows in a site with at least one cell edited). Corrections were recorded as annotations in the data checking script. There were no instances of errors introduced through the data cleaning process itself. The data check also incorporated a review of each site and their data submissions, highlighting sites missing from the Y0 (v1.0) dataset.

1.2.6 | QA6: Production & Dissemination of High Quality Outputs

Sharing data outputs that support publications allows for replication, and sharing data within a network and beyond allows

for increased standardisation as the data is re-used (Parr and Cummings 2005). High-quality outputs are produced through rigorous peer review processes and data audits to ensure QC (Ferreira et al. 2016).

1.2.6.1 | PPN9: Publication, Dissemination and Versioning of SDPs. Each SDP combines data from all sites. As new data become available, or new cleaning procedures are implemented, new versions of each SDP are produced, named and old versions archived. For example, the Y0 SDP contains data for all sites from the first year of data collection, the Y1 SDP contains data for 2 years, enabling one demographic transition to be analysed. A record of changes from version to version is maintained for comparison. The first Plantpopnet SDP (Y0v1.3) is provided with this paper, including a summary of the data at site level.

1.2.6.2 | PPN10: Publication of Papers, Reports. To avoid publication conflicts and enhance collaboration, all uses of the (unpublished) core demographic data across multiple sites require approval from the steering group. A participant

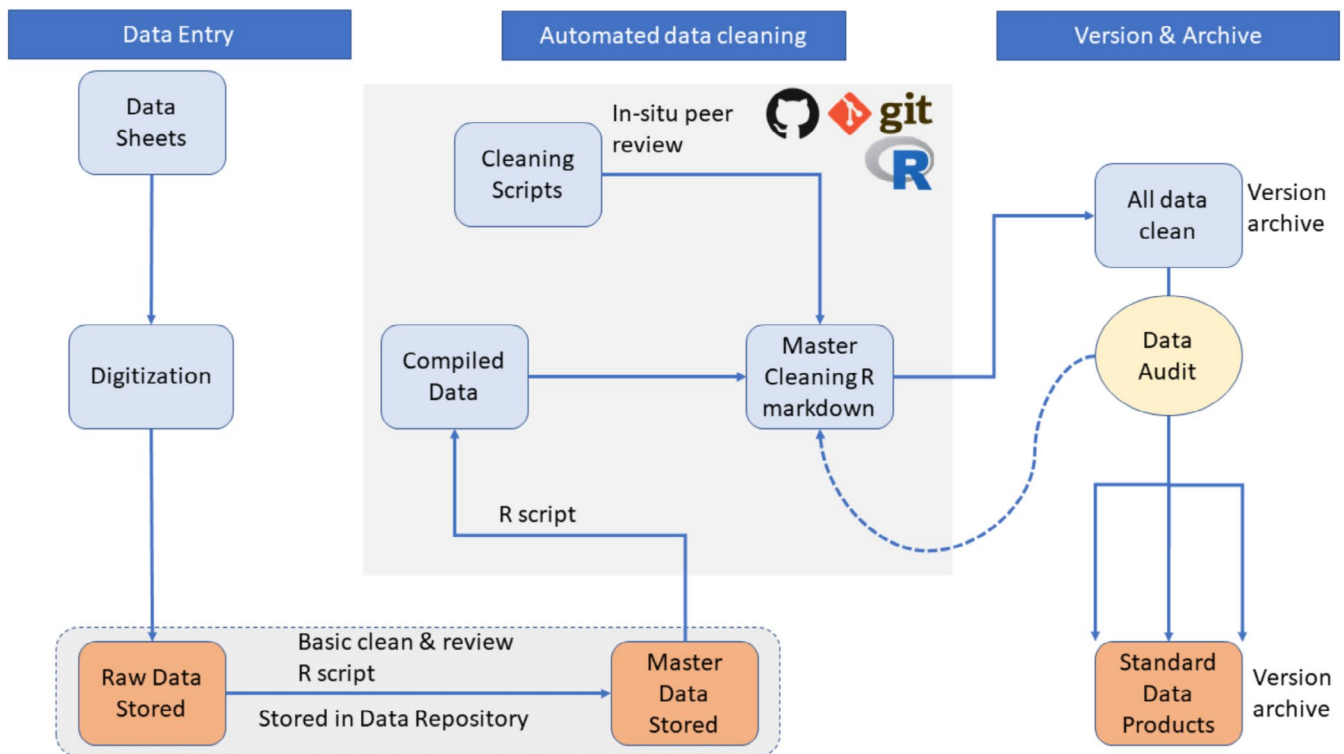


FIGURE 4 | Data management workflow from point of creation by site coordinators to production of Standard Data Products (SDPs). Raw data is converted to master data, which is cleaned via a semi-automated system of R scripts and maintained and stored using Git and the online repository Github (icons under Automated data cleaning represent software tools). ‘Cleaned’ data is subjected to an audit before final export to SDPs. A data audit determines if errors were introduced through the cleaning scripts.

submits an abstract for approval (using the abstract template provided, Supporting Information S1: 1.5) outlining the research question, intended use of data products and any additional data to be collected. Potential authorship is identified and normally includes all data contributors as well as other collaborators with key expertise. An authorship guidelines document is used to determine authorship processes (Supporting Information S1: 1.2). To ensure quality, all manuscripts produced using Plantpopnet data are internally reviewed by participants who have contributed data. A record is kept of manuscripts and their stage towards completion and updated every 12 months.

2 | Discussion

We developed and applied a QA management framework to ensure that the data from coordinated, distributed ecology projects have sufficient quality, longevity, reproducibility, accessibility and utility for current and future use. Applying the QA framework to the Plantpopnet project, we show how coordinated actions can be structured, reviewed and checked. For sustained benefits to be realised, we emphasise the importance of data management processes, including the reproducibility of data-cleaning and the production of SDPs. While other projects have used individual QA or QC components, we have implemented a full project lifecycle approach to QA that supports better reproducibility of the many different components of a successful coordinated distributed network project from inception to data production and sustainable use.

Reproducible methods for SDP production are important given the heterogeneity in the quality of the data submitted, the use of the resulting data by many different researchers, and the expected longevity of the unique datasets produced. QA and QC principles applied to data curation supported consistent production of SDPs to be used for further analysis. The SDPs will be critical inputs into ongoing data analysis projects within the network and as SDPs are progressively publicly released, they will become an important resource for the plant demography and ecological research community. By releasing the Y0 SDP together with the raw data and scripts for data cleaning we hope to inspire wider uptake of this approach to reproducible data cleaning in ecology (Buckley et al. 2025). QA is consistent with the emergence of the Open Science Framework (Foster and Deardorff 2017; O’Dea et al. 2021) and the FAIR (Findability, Accessibility, Interoperability and Reusability) Data Management principles (Wilkinson et al. 2016), which provide a platform and tools needed to radically improve transparency, reproducibility and reliability in science. In adhering to open data principles, it is also important to consider local and jurisdictional-level laws and customs, to ensure local and Indigenous communities maintain data rights (Carroll et al. 2020).

Team science in ecology requires a range of new methods and ways of working to maximise benefits from collaborative, distributed data collection and analysis. Contributors have the opportunity to provide data, a different perspective and further spatial extent (Borer et al. 2023; Czech et al. 2022; Herbivory Variability Network et al. 2023; Lucas et al. 2024; Santangelo et al. 2022). Collecting consistent field data across multiple

network participants is challenging however, and although a detailed protocol and data collection template were provided, checks of the cleaned data identified that logistical issues and differences between ecosystems in the field contributed to deviations from the template. We found that the most frequent discrepancies in the raw data were basic structural and formatting deviations, commonly from modifications to the template for field use. In isolation, these inconsistencies seem trivial, but in a large dataset they can be difficult to manage. The need for post-collection processing can be minimised through careful design of field data-sheets to maximise their utility and minimise the potential for incorrect data entry (Supporting Information S6). Utilising an app for data collection could also help to prevent incorrect data entry, such as transcription errors, e.g., the PhenObs network (Sporbert et al. 2022).

A wide geographic distribution of sites was achieved for Plantpopnet, but important spatial gaps became apparent. The eastern and southern parts of the native range and arid areas are under-sampled (Figure 2). Language capacity, lack of scientific institutions and networks, research culture differences, incomplete occurrence data and resource limitations may have hindered recruitment of participants. Language barriers can seriously limit knowledge sharing through barriers of evidence generation by non-English speakers and the global synthesis of evidence in different languages (Amano and Berdejo-Espinola 2025). Future efforts to identify potential collaborators and funding strategies to support participants in targeted locations may be needed to fulfil the aims of better representing the full extent of biotic and abiotic niche dimensions of this species. Wider availability of automated translation tools will enable future networks to more effectively reach potential collaborators throughout the species' range. Institutional partnerships and the co-design, emergence and recognition of benefits of being part of coordinated distributed networks for individual researchers and their institutions should enable future networks to more effectively recruit and retain members in key locations. Currently benefits include co-authored papers, access to important datasets, access to collaborative networks and international recognition; however, an understanding of different research and institutional cultures may enable additional benefits to be identified or co-created.

2.1 | Funding Network Science

Plantpopnet operates with a distributed funding model where individual sites are funded by network participants. Central funding has primarily been used to operate the digital infrastructure needed for the network. However, once the outputs are published, the data repository and curation systems become public goods, opening the possibilities of freeloading, where anyone can benefit (Neylon 2017). Centralised funding is common across shared infrastructures; however, sustainability models need to adapt and change over the life cycle of an infrastructure, reflecting the changes in the community that uses the service (Neylon 2017). Making the data curation and cleaning tools available to researchers within and external to the network reduces the need for centralised data curation funding, and it may be beneficial to move to an alternative funding model to ensure the longevity of the network. For example, a shift to sources that

fund specific uses of the data rather than funding general data collection and curation. Securely archiving open access data together with strong documentation on protocols under which data were collected and curated can mitigate the risk of data loss on conclusion of funding streams.

2.2 | Serendipitous Benefits

Plantpopnet has displayed an unanticipated robustness to global shocks. In 2020, it was expected that elective projects like Plantpopnet would suffer from the pressures and workloads brought about during the Covid-19 pandemic. Just 6 of 43 active sites were unable to sample in 2020 due to the pandemic. Surprisingly, 6 new sites joined the network in 2020, more than in any of the previous 3 years (Supporting Information S3). The key factors enabling this resilience to shock were the proximity of sites to the researchers' base and the ease of sampling.

Environmental sustainability has been an achievement of the network. Carbon costs of undertaking research are under increasing scrutiny (e.g., ALLEA (2022); Baer (2018, 2019); The Wellcome Trust (2021); Wynes et al. (2019)) and in future, environmental sustainability may need to be addressed for research ethics approval and funding.

3 | Conclusions

An explicit QA framework for management of multi-site, multi-participant collaborative data collection networks can support sustained engagement of participants and robustness of data collection and dissemination. The most important risks to continued data collection were instability in the location and employment of participants. This needs to be considered for the design of future networks where long term datasets are to be collected. One of the strengths of Plantpopnet was the inclusion of participants as independent site coordinators from a wide range of career stages. While changes in personal circumstances were more likely in early career participants, some maintained involvement through collection of different types of data (e.g., genetic samples), setting up of new sites, or through recruitment of collaborators to maintain their original site.

The project management and data curation skills needed to progress spatially distributed ecology projects are not commonly outlined or transferred as part of undergraduate and postgraduate courses or continuing professional development. These skills are highly transferable and in demand within academia and beyond. By making project management transparent and central to the design and implementation of a collaborative, distributed team science project, we hope to inform the structure and design of future projects as well as to stimulate debate and the development of new methods.

Author Contributions

Yvonne M. Buckley: conceptualization (equal), data curation (lead), formal analysis (lead), funding acquisition (lead), investigation (equal), methodology (equal), project administration (lead), resources (lead),

software (equal), supervision (lead), validation (equal), visualization (supporting), writing – original draft (lead), writing – review and editing (equal). **Alain Finn:** data curation (equal), formal analysis (equal), methodology (equal), project administration (equal), writing – original draft (equal), writing – review and editing (equal). **Aoife Molloy:** data curation (equal), formal analysis (equal), investigation (equal), project administration (equal), writing – review and editing (equal). **Anna Mária Csörgő:** conceptualization (equal), data curation (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), writing – review and editing (equal). **Elizabeth Crone:** conceptualization (equal), investigation (equal), methodology (equal), project administration (equal), writing – review and editing (equal). **Annabel L. Smith:** data curation (equal), investigation (equal), methodology (equal), resources (equal), software (equal), writing – review and editing (equal). **Bret D. Elder:** conceptualization (equal), data curation (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), writing – review and editing (equal). **Caroline McKeon:** data curation (equal), investigation (equal), resources (equal), software (equal), writing – review and editing (equal). **Deborah Roach:** conceptualization (equal), data curation (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), writing – review and editing (equal). **Dylan Z. Childs:** conceptualization (equal), data curation (equal), formal analysis (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), writing – review and editing (equal). **Glenda M. Wardle:** conceptualization (equal), data curation (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), writing – original draft (equal), writing – review and editing (equal). **María B. García:** conceptualization (equal), data curation (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), writing – review and editing (equal). **Maude Baudraz:** data curation (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), software (equal), writing – review and editing (equal). **Rachel M. Penczykowski:** conceptualization (equal), data curation (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), writing – review and editing (equal). **Roberto Salguero-Gómez:** conceptualization (equal), data curation (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), writing – review and editing (equal). **Ruth Kelly:** data curation (equal), investigation (equal), writing – review and editing (equal). **Satu Ramula:** conceptualization (equal), data curation (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), writing – review and editing (equal). **Anna Liisa Laine:** conceptualization (equal), data curation (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), writing – review and editing (equal). **Simone Blomberg:** conceptualization (equal), methodology (equal), project administration (equal), writing – review and editing (equal). **Johan Ehrlén:** conceptualization (equal), data curation (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), writing – review and editing (equal). **Jesús Vilellas:** conceptualization (equal), data curation (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), writing – review and editing (equal). **Aveliina Helm:** data curation (equal), investigation (equal), writing – review and editing (equal). **Christiane Roscher:** data curation (equal), investigation (equal), writing – review and editing (equal). **Krista Raveala:** data curation (equal), investigation (equal), writing – review and editing (equal). **Liv Norunn Hamre:** data curation (equal), investigation (equal), writing – review and editing (equal). **Lotte Korell:** data curation (equal), investigation (equal), writing – review and editing (equal). **Francis Q. Brearley:** data curation (equal), investigation (equal), writing – review and editing (equal). **Siri Lie Olsen:** data curation (equal), investigation (equal), writing – review and editing (equal). **Astrid Winkler:** data curation (equal), investigation (equal), writing – review and editing (equal). **Joslin L. Moore:** data curation (equal), investigation (equal), writing – review and editing (equal). **Ayco J. M. Tack:** data curation (equal), investigation (equal), writing – review and editing (equal). **Anna Bucharova:** data curation

(equal), investigation (equal), writing – review and editing (equal). **Olav Skarpaas:** data curation (equal), investigation (equal), writing – review and editing (equal). **Judit Bodis:** data curation (equal), investigation (equal), writing – review and editing (equal). **Aldo Compagnoni:** data curation (equal), investigation (equal), writing – review and editing (equal). **Sergi Munne-Bosch:** data curation (equal), investigation (equal), methodology (equal), writing – review and editing (equal). **Zuzana Munzbergova:** data curation (equal), investigation (equal), writing – review and editing (equal). **Joachim Töpper:** data curation (equal), investigation (equal), writing – review and editing (equal). **Peter Vesk:** data curation (equal), investigation (equal), writing – review and editing (equal). **Melanie Morales:** data curation (equal), investigation (equal), writing – review and editing (equal). **Lizzie Wandrag:** data curation (equal), investigation (equal), writing – review and editing (equal). **Marjo Saastamoinen:** data curation (equal), investigation (equal), writing – review and editing (equal). **Martin Andrzejak:** data curation (equal), investigation (equal), writing – review and editing (equal). **Pil Rasmussen:** data curation (equal), investigation (equal), writing – review and editing (equal). **Christina Caruso:** data curation (equal), investigation (equal), writing – review and editing (equal). **Richard Duncan:** data curation (equal), investigation (equal), writing – review and editing (equal). **John Dwyer:** data curation (equal), investigation (equal), writing – review and editing (equal). **Matthew Coghill:** data curation (equal), investigation (equal), writing – review and editing (equal). **Ronny Groenteman:** data curation (equal), investigation (equal), writing – review and editing (equal). **Meelis Pärtel:** data curation (equal), investigation (equal), writing – review and editing (equal). **William K. Petry:** data curation (equal), investigation (equal), writing – review and editing (equal). **Michele Lonati:** data curation (equal), investigation (equal), writing – review and editing (equal). **Javier Puy:** data curation (equal), investigation (equal), writing – review and editing (equal). **Jane Catford:** data curation (equal), investigation (equal), writing – review and editing (equal). **Trevor H. Drees:** data curation (equal), investigation (equal), writing – review and editing (equal). **Christoph Rosche:** data curation (equal), investigation (equal), writing – review and editing (equal). **Katriona Shea:** data curation (equal), investigation (equal), writing – review and editing (equal). **Simone Ravetto Enri:** data curation (equal), investigation (equal), writing – review and editing (equal). **Jennifer R. Gremer:** data curation (equal), investigation (equal), writing – review and editing (equal). **Balázs Deák:** data curation (equal), investigation (equal), writing – review and editing (equal). **Lauchlan Fraser:** data curation (equal), investigation (equal), writing – review and editing (equal). **Lauri Laanisto:** data curation (equal), investigation (equal), writing – review and editing (equal). **Orsolya Valkó:** data curation (equal), investigation (equal), writing – review and editing (equal). **Robert Drew Sieg:** data curation (equal), investigation (equal), writing – review and editing (equal). **Benedicte Bachelot:** data curation (equal), investigation (equal), writing – review and editing (equal). **Jennifer Finn:** data curation (equal), investigation (equal), writing – review and editing (equal).

Affiliations

¹Co-Centre for Climate + Biodiversity + Water, School of Natural Sciences, Trinity College Dublin, Dublin 2, Ireland | ²Department of Biology, Maynooth University, Maynooth, Co. Kildare, Ireland | ³Department of Botany, Buda Campus, Hungarian University of Agriculture and Life Sciences, Budapest, Hungary | ⁴Department of Evolution and Ecology, University of California, Davis, California, USA | ⁵School of the Environment, University of Queensland, Brisbane, Queensland, Australia | ⁶Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana, USA | ⁷Department of Biology, University of Virginia, Charlottesville, Virginia, USA | ⁸School of Biosciences, University of Sheffield, Sheffield, UK | ⁹School of Life and Environmental Sciences, University of Sydney, Sydney, New South Wales, Australia | ¹⁰Pyrenean Institute of Ecology (CSIC), Zaragoza, Jaca, Spain | ¹¹School of Natural Sciences, Zoology, Trinity College Dublin, Dublin 2, Ireland | ¹²Department of Biology, Washington University in St. Louis, St. Louis, Missouri, USA | ¹³Department of Biology,

University of Oxford, Oxford, UK | ¹⁴Agri-Food and BioSciences Institute, Belfast, UK | ¹⁵Department of Biology, University of Turku, Turku, Finland | ¹⁶Faculty of Biology and Environmental Sciences, University of Helsinki, Helsinki, Finland | ¹⁷Department of Ecology, Environment, and Plant Sciences, Stockholm University, Stockholm, Sweden | ¹⁸Departamento de Ciencias de la Vida, Alcalá de Henares, Universidad de Alcalá, Madrid, Spain | ¹⁹Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia | ²⁰Department for Physiological Diversity, Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany | ²¹German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany | ²²Department of Civil Engineering and Environmental Sciences, Western Norway University of Applied Sciences, Bergen, Norway | ²³Department of Species Interaction Ecology, Helmholtz Centre for Environmental Research, Leipzig, Germany | ²⁴Department of Natural Sciences, Manchester Metropolitan University, Manchester, UK | ²⁵Faculty for Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway | ²⁶Norwegian Institute for Nature Research, Oslo, Norway | ²⁷School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland | ²⁸Department of Energy, Environment and Climate Action, Arthur Rylah Institute for Environment Research, Heidelberg, Victoria, Australia | ²⁹School of Biological Sciences, Monash University, Clayton, Victoria, Australia | ³⁰Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden | ³¹Faculty of Biology, University of Marburg, Marburg, Germany | ³²Natural History Museum, University of Oslo, Oslo, Norway | ³³Institute for Wildlife Management and Nature Conservation, Georgikon Campus, Hungarian University of Agriculture and Life Sciences, Keszthely, Hungary | ³⁴Martin Luther University Halle-Wittenberg, Halle, Germany | ³⁵Department of Evolutionary Biology, Ecology and Environmental Sciences, University of Barcelona, Barcelona, Spain | ³⁶Institute of Research in Biodiversity (IRBio), University of Barcelona, Barcelona, Spain | ³⁷Institute of Botany, Czech Academy of Sciences, Průhonice, Czech Republic | ³⁸Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic | ³⁹Norwegian Institute for Nature Research, Bergen, Norway | ⁴⁰School of Agriculture, Food and Ecosystem Sciences, University of Melbourne, Melbourne, Victoria, Australia | ⁴¹University of Talca, Talca, Chile | ⁴²University of Tasmania, Hobart, Tasmania, Australia | ⁴³Ecology and Genetics, University of Oulu, Oulu, Finland | ⁴⁴The National Research Centre for the Working Environment, Copenhagen, Denmark | ⁴⁵Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada | ⁴⁶Centre for Conservation and Ecology Genetics, University of Canberra, Canberra, Australian Capital Territory, Australia | ⁴⁷School of the Environment, The University of Queensland, Brisbane, Queensland, Australia | ⁴⁸Department of Natural Resource Sciences, Thompson Rivers University, Kamloops, British Columbia, Canada | ⁴⁹Manaaki Whenua—Landcare Research, a group of the Bioeconomy Science Institute, Lincoln, New Zealand | ⁵⁰Department of Plant and Microbial Biology, North Carolina State University, Raleigh, North Carolina, USA | ⁵¹Department of Agricultural, Forest and Food Science, University of Torino, Torino, Italy | ⁵²Departamento de Biología Vegetal y Ecología, Facultad de Biología, Universidad de Sevilla, Seville, Spain | ⁵³Department of Geography, King's College London, London, UK | ⁵⁴Fenner School of Environment & Society, The Australian National University, Canberra, Australian Capital Territory, Australia | ⁵⁵Department of Biology and IGDP in Ecology, The Pennsylvania State University, University Park, Pennsylvania, USA | ⁵⁶Institute of Biology/Geobotany and Botanical Garden, University of Halle, Halle, Germany | ⁵⁷Department of Evolution and Ecology, University of California, Davis, Davis, California, USA | ⁵⁸Lendület Seed Ecology Research Group, Institute of Ecology and Botany, HUN-REN Centre for Ecological Research, Vácrátót, Hungary | ⁵⁹Estonian University of Life Sciences, Institute of Agricultural and Environmental Sciences, University of Tartu, Tartu, Estonia | ⁶⁰Agricultural and Biological Sciences Department, Truman State University, Kirksville, Missouri, USA | ⁶¹Department of Biology, Oklahoma State University, Stillwater, Oklahoma,

USA | ⁶²Biogeosciences, Faculty of Science & Technology, Queensland University of Technology, Brisbane, Queensland, Australia

Acknowledgements

We thank Jan van Groenendael, Cheryl Schultz, Sam Levin, Maria Hohn, Kevin Kožic, Suvi Ikonen, Greg Vose, Anna Roeder, Thomas Panchard, Tiit Kull, Jennifer Firn and Emily Griffoul for their valuable contributions to site coordination and/or data collection. This publication has emanated from research conducted with the financial support of Research Ireland, Northern Ireland's Department of Agriculture, Environment and Rural Affairs (DAERA), UK Research and Innovation (UKRI) via the International Science Partnerships Fund (ISPF) under Grant number [22/CC/11103] at the Co-Centre for Climate + Biodiversity + Water. Funding for development and coordination of Plantpopnet was provided through Australian Research Council DP120101462 (2012, YMB), Marie Curie Career Integration Grant 631105 (2014, YMB), Science Foundation Ireland ERC Development 15/ERC/D2803 (2015, YMB), SFI conferences & workshops grant 17/CW/4943 (2017, YMB) and Irish Research Council Laureate award IRCLA/2017/60 (2018, YMB). The survey was approved by the ethics committee of the School of Natural Sciences, TCD.

Funding

This work was supported by Research Ireland, 22/CC/11103, UK Research and Innovation, 22/CC/11103, Department of Agriculture, Environment and Rural Affairs, UK Government, 22/CC/11103, FP7 People: Marie-Curie Actions, 631105 and Australian Research Council, DP120101462.

Disclosure

Statement on inclusion: The data were collected by participants in the distributed ecology network Plantpopnet. All data collectors were invited to contribute further to the manuscript and represent a wide range of countries worldwide where the research was carried out. We provide further information in the main text of the paper on the recruitment and management of authorship for this inclusive approach, including a link to the authorship guidelines for Plantpopnet.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Standard Data Product: SDP Y0.1.3 and associated summary table of Plantpopnet site characteristics provided on Dryad DOI: <https://doi.org/10.5061/dryad.m37pvmjdc>. An archived version of the code and raw data for the data cleaning process and production of Standard Data Product (SDP) is available on Zenodo: <https://doi.org/10.5281/zenodo.16608913>.

References

- Alexander, J. M., C. Kueffer, C. C. Daehler, et al. 2011. "Assembly of Nonnative Floras Along Elevational Gradients Explained by Directional Ecological Filtering." *Proceedings of the National Academy of Sciences of the United States of America* 108: 656–661. <https://doi.org/10.1073/pnas.1013136108>.
- Alexander, J. M., M. Van Kleunen, R. Ghezzi, and P. J. Edwards. 2012. "Different Genetic Clines in Response to Temperature Across the Native and Introduced Ranges of a Global Plant Invader." *Journal of Ecology* 100: 771–781. <https://doi.org/10.1111/j.1365-2745.2011.01951.x>.
- ALLEA. 2022. *Towards Climate Sustainability of the Academic System in Europe and Beyond*. All European Academies.

- Amano, T., and V. Berdejo-Espinola. 2025. "Language Barriers in Conservation: Consequences and Solutions." *Trends in Ecology & Evolution* 40: 273–285. <https://doi.org/10.1016/j.tree.2024.11.003>.
- Araújo, M. B., R. P. Anderson, A. Márcia Barbosa, et al. 2019. "Standards for Distribution Models in Biodiversity Assessments." *Science Advances* 5: eaat4858. <https://doi.org/10.1126/sciadv.aat4858>.
- Aubin, I., F. Cardou, L. Boisvert-Marsh, E. Garnier, M. Strukelj, and A. D. Munson. 2020. "Managing Data Locally to Answer Questions Globally: The Role of Collaborative Science in Ecology." *Journal of Vegetation Science* 31: 509–517. <https://doi.org/10.1111/jvs.12864>.
- Baer, H. A. 2018. "Grappling With Flying as a Driver to Climate Change: Strategies for Critical Scholars Seeking to Contribute to a Socio-Ecological Revolution." *Australian Journal of Anthropology* 29: 298–315. <https://doi.org/10.1111/taja.12291>.
- Baer, H. A. 2019. "The Elephant in the Sky: On How to Grapple With Our Academic Flying in the Age of Climate Change." *Anthropology Today* 35: 21–24. <https://doi.org/10.1111/1467-8322.12518>.
- Baudraz, M. E. A., D. Z. Childs, R. Kelly, et al. 2025. "Several Candidate Size Metrics Explain Vital Rates Across Multiple Populations Throughout a Widespread Species' Range." *Journal of Ecology* 113, no. 11: 3159–3176. <https://doi.org/10.1111/1365-2745.70148>.
- Borer, E. T., W. S. Harpole, P. B. Adler, et al. 2014. "Finding Generality in Ecology: A Model for Globally Distributed Experiments." *Methods in Ecology and Evolution* 5: 65–73. <https://doi.org/10.1111/2041-210X.12125>.
- Borer, E. T., A. S. MacDougall, C. J. Stevens, L. L. Sullivan, P. A. Wilfahrt, and E. W. Seabloom. 2023. "Writing a Massively Multi-Authored Paper: Overcoming Barriers to Meaningful Authorship for All." *Methods in Ecology and Evolution* 14: 1432–1442. <https://doi.org/10.1111/2041-210X.14096>.
- Borer, E. T., and E. W. Seabloom. 2025. "Maximizing Inference From Distributed Experimental Networks via "Add-On" Studies." *Frontiers in Ecology and the Environment* 23: e70007. <https://doi.org/10.1002/fee.70007>.
- Borer, E. T., E. W. Seabloom, M. B. Jones, and M. Schildhauer. 2009. "Some Simple Guidelines for Effective Data Management." *Bulletin of the Ecological Society of America* 90: 205–214. <https://doi.org/10.1890/0012-9623-90.2.205>.
- Brown, E. D., and B. K. Williams. 2019. "The Potential for Citizen Science to Produce Reliable and Useful Information in Ecology." *Conservation Biology* 33: 561–569. <https://doi.org/10.1111/cobi.13223>.
- Buckley, Y. M., R. Bardgett, R. Gordon, et al. 2025. "Using Dynamic Documents to Mend Cracks in the Reproducible Research Pipeline." *Journal of Ecology* 113: 270–274. <https://doi.org/10.1111/1365-2745.14454>.
- Buckley, Y. M., E. E. Crone, A. M. Csörgő, et al. 2019. "Plantpopnet Protocol V1.03 2017 477277 Bytes." <https://doi.org/10.6084/M9.FIGSH.ARE.7982477.V9>.
- Buckley, Y. M., and J. Puy. 2022. "The Macroecology of Plant Populations From Local to Global Scales." *New Phytologist* 233: 1038–1050. <https://doi.org/10.1111/nph.17749>.
- Buckley, Y. M., S. Ramula, S. P. Blomberg, et al. 2010. "Causes and Consequences of Variation in Plant Population Growth Rate: A Synthesis of Matrix Population Models in a Phylogenetic Context." *Ecology Letters* 13: 1182–1197. <https://doi.org/10.1111/j.1461-0248.2010.01506.x>.
- Carroll, S. R., I. Garba, O. L. Figueroa-Rodríguez, et al. 2020. "The CARE Principles for Indigenous Data Governance." *Data Science Journal* 19: 1–12. <https://doi.org/10.5334/dsj-2020-043>.
- Catford, J. A., J. R. U. Wilson, P. Pyšek, P. E. Hulme, and R. P. Duncan. 2022. "Addressing Context Dependence in Ecology." *Trends in Ecology & Evolution* 37: 158–170. <https://doi.org/10.1016/j.tree.2021.09.007>.
- CEN. 2015a. *Quality Management Systems—Fundamentals and Vocabulary (ISO 9000:2015)*. European Committee for Standardization. <https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en>.
- CEN. 2015b. *Quality Management Systems—Requirements (ISO 9001:2015)*. European Committee for Standardization. <https://www.iso.org/standard/62085.html>.
- Cheruvilil, K. S., P. A. Soranno, K. C. Weathers, et al. 2014. "Creating and Maintaining High-Performing Collaborative Research Teams: The Importance of Diversity and Interpersonal Skills." *Frontiers in Ecology and the Environment* 12: 31–38. <https://doi.org/10.1890/130001>.
- Coutts, S. R., R. Salguero-Gómez, A. M. Csörgő, and Y. M. Buckley. 2016. "Extrapolating Demography With Climate, Proximity and Phylogeny: Approach With Caution." *Ecology Letters* 19: 1429–1438. <https://doi.org/10.1111/ele.12691>.
- Czech, L., Y. Peng, J. P. Spence, et al. 2022. "Monitoring Rapid Evolution of Plant Populations at Scale With Pool-Sequencing." *bioRxiv*: 2022.02.02.477408. <https://doi.org/10.1101/2022.02.02.477408>.
- Diagne, C., J. A. Catford, F. Essl, M. A. Nuñez, and F. Courchamp. 2021. "What Are the Economic Costs of Biological Invasions? A Complex Topic Requiring International and Interdisciplinary Expertise." *NeoBiota* 63: 25–37. <https://doi.org/10.3897/neobiota.63.55260>.
- Enquist, B. J., R. Condit, R. K. Peet, M. Schildhauer, and B. M. Thiers. 2016. "Cyberinfrastructure for an Integrated Botanical Information Network to Investigate the Ecological Impacts of Global Climate Change on Plant Biodiversity." *PeerJ Preprints* 4: e2615v2. <https://doi.org/10.7287/peerj.preprints.2615v2>.
- Ferreira, C., G. Bastille-Rousseau, A. M. Bennett, et al. 2016. "The Evolution of Peer Review as a Basis for Scientific Publication: Directional Selection Towards a Robust Discipline?" *Biological Reviews of the Cambridge Philosophical Society* 91: 597–610. <https://doi.org/10.1111/brv.12185>.
- Fick, S. E., and R. J. Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37: 4302–4315. <https://doi.org/10.1002/joc.5086>.
- Filazzola, A., and C. Lortie. 2022. "A Call for Clean Code to Effectively Communicate Science." *Methods in Ecology and Evolution* 13: 2119–2128. <https://doi.org/10.1111/2041-210X.13961>.
- Foster, E. D., and A. Deardorff. 2017. "Open Science Framework (OSF)." *Journal of the Medical Library Association* 105: 203–206. <https://doi.org/10.5195/jmla.2017.88>.
- Fraser, L. H., H. A. Henry, C. N. Carlyle, et al. 2013. "Coordinated Distributed Experiments: An Emerging Tool for Testing Global Hypotheses in Ecology and Environmental Science." *Frontiers in Ecology and the Environment* 11: 147–155. <https://doi.org/10.1890/110279>.
- García, M. B., J. L. Silva, P. Tejero, I. Pardo, and D. Gómez. 2019. "Tracking the Long-Term Dynamics of Plant Diversity in Northeast Spain With a Network of Volunteers and Rangers." *Regional Environmental Change* 19: 391–401. <https://doi.org/10.1007/s10113-018-1350-6>.
- GBIF. 2022. "Plantago lanceolata Occurrence Download." <https://doi.org/10.15468/dl.kq6g5p>.
- Goldberg, C. S., C. R. Turner, K. Deiner, et al. 2016. "Critical Considerations for the Application of Environmental DNA Methods to Detect Aquatic Species." *Methods in Ecology and Evolution* 7: 1299–1307. <https://doi.org/10.1111/2041-210X.12595>.
- Hampton, S. E., C. A. Strasser, J. J. Tewksbury, et al. 2013. "Big Data and the Future of Ecology." *Frontiers in Ecology and the Environment* 11: 156–162. <https://doi.org/10.1890/120103>.
- Hayward, M. W., L. Boitani, N. D. Burrows, et al. 2015. "FORUM: Ecologists Need Robust Survey Designs, Sampling and Analytical Methods." *Journal of Applied Ecology* 52: 286–290. <https://doi.org/10.1111/1365-2664.12408>.

- Heberling, J. M., J. T. Miller, D. Noesgaard, S. B. Weingart, and D. Schigel. 2021. "Data Integration Enables Global Biodiversity Synthesis." *Proceedings of the National Academy of Sciences of the United States of America* 118: e2018093118. <https://doi.org/10.1073/pnas.2018093118>.
- Henry, G. H. R., and U. Molau. 1997. "Tundra Plants and Climate Change: The International Tundra Experiment (ITEX)." *Global Change Biology* 3: 1–9. <https://doi.org/10.1111/j.1365-2486.1997.gcb132.x>.
- Herbivory Variability Network[†], M. L. Robinson, P. G. Hahn, et al. 2023. "Plant Size, Latitude, and Phylogeny Explain Within-Population Variability in Herbivory." *Science* 382: 679–683. <https://doi.org/10.1126/science.adh8830>.
- Jenkins, G. B., A. P. Beckerman, C. Bellard, et al. 2023. "Reproducibility in Ecology and Evolution: Minimum Standards for Data and Code." *Ecology and Evolution* 13: e9961. <https://doi.org/10.1002/ece3.9961>.
- Joseph, A. 2021. "CompareDF: Do a Git Style Diff of the Rows Between Two Dataframes With Similar Structure (2.3.2) [R]." <https://CRAN.R-project.org/package=compareDF>.
- Kim, A. Y., V. Herrmann, R. Barreto, et al. 2022. "Implementing GitHub Actions Continuous Integration to Reduce Error Rates in Ecological Data Collection." *Methods in Ecology and Evolution* 13: 2572–2585. <https://doi.org/10.1111/2041-210X.13982>.
- Kuiper, P. J. C., and M. Bos. 1992. *Plantago: A Multidisciplinary Study, Ecological Studies*. Springer.
- Lind, E. M. 2016. "Unified Data Management for Distributed Experiments: A Model for Collaborative Grassroots Scientific Networks." *Ecological Informatics* 36: 231–236. <https://doi.org/10.1016/j.ecoinf.2016.08.002>.
- Love, H. B., J. E. Cross, B. Fosdick, K. R. Crooks, S. VandeWoude, and E. R. Fisher. 2021. "Interpersonal Relationships Drive Successful Team Science: An Exemplary Case-Based Study." *Humanities & Social Sciences Communications* 8: 106. <https://doi.org/10.1057/s41599-021-00789-8>.
- Lucas, M. S., I. Hensen, C. D. Barratt, et al. 2024. "Re-Focusing Sampling, Design and Experimental Methods to Assess Rapid Evolution by Non-Native Plant Species." *Biological Invasions* 26: 1327–1343. <https://doi.org/10.1007/s10530-024-03249-x>.
- Maund, P. R., J. W. Bentley, G. E. Austen, et al. 2022. "The Features and Processes Underpinning High-Quality Data Generation in Participatory Research and Engagement Activities." *Methods in Ecology and Evolution* 13: 68–76. <https://doi.org/10.1111/2041-210X.13746>.
- McCord, S. E., J. L. Welty, J. Courtwright, et al. 2022. "Ten Practical Questions to Improve Data Quality." *Rangelands* 44: 17–28. <https://doi.org/10.1016/j.rala.2021.07.006>.
- Michener, W. K. 2018. "Quality Assurance and Quality Control (QA/QC)." In *Ecological Informatics: Data Management and Knowledge Discovery*, edited by F. Recknagel and W. K. Michener, 55–70. Springer International Publishing. https://doi.org/10.1007/978-3-319-59928-1_4.
- Mirtl, M. T., E. Borer, I. Djukic, et al. 2018. "Genesis, Goals and Achievements of Long-Term Ecological Research at the Global Scale: A Critical Review of ILTER and Future Directions." *Science of the Total Environment* 626: 1439–1462. <https://doi.org/10.1016/j.scitotenv.2017.12.001>.
- Morales, M., D. A. Roach, B. M. Quarles, et al. 2021. "Validity of Photo-Oxidative Stress Markers and Stress-Related Phytohormones as Predictive Proxies of Mortality Risk in the Perennial Herb *Plantago lanceolata*." *Environmental and Experimental Botany* 191: 104598. <https://doi.org/10.1016/j.envexpbot.2021.104598>.
- Neylon, C. 2017. "Sustaining Scholarly Infrastructures Through Collective Action: The Lessons That Olson Can Teach Us." *Kula* 1: 3. <https://doi.org/10.5334/kula.7>.
- O'Dea, R. E., T. H. Parker, Y. E. Chee, et al. 2021. "Towards Open, Reliable, and Transparent Ecology and Evolutionary Biology." *BMC Biology* 19: 68. <https://doi.org/10.1186/s12915-021-01006-3>.
- Parr, C. S., and M. P. Cummings. 2005. "Data Sharing in Ecology and Evolution." *Trends in Ecology & Evolution* 20: 362–363.
- Pastor, A., E. Hernández-del Amo, P. Giménez-Grau, et al. 2020. "Early-Career Coordinated Distributed Experiments: Empowerment Through Collaboration." *Frontiers in Education* 5: 13. <https://doi.org/10.3389/educ.2020.00013>.
- Penczykowski, R. M., and R. D. Sieg. 2021. "*Plantago* spp. as Models for Studying the Ecology and Evolution of Species Interactions Across Environmental Gradients." *American Naturalist* 198: 158–176. <https://doi.org/10.1086/714589>.
- QQI. 2021. "Quality Assurance of Education and training|Quality and Qualifications Ireland." [WWW Document]. Accessed April 2, 2025. <https://www.qqi.ie/what-we-do/quality-assurance-of-education-and-training>.
- Sagar, G. R., and J. L. Harper. 1961. "Controlled Interference With Natural Populations of *Plantago lanceolata*, *P. Major* and *P. media*." *Weed Research* 1: 163–176. <https://doi.org/10.1111/j.1365-3180.1961.tb00017.x>.
- Santangelo, J. S., R. W. Ness, B. Cohan, et al. 2022. "Global Urban Environmental Change Drives Adaptation in White Clover." *Science* 375: 1275–1281. <https://doi.org/10.1126/science.abk0989>.
- Smith, A. L., T. R. Hodkinson, J. Villellas, et al. 2020. "Global Gene Flow Releases Invasive Plants From Environmental Constraints on Genetic Diversity." *Proceedings of the National Academy of Sciences* 117: 4218–4227. <https://doi.org/10.1073/pnas.1915848117>.
- Sporbert, M., D. Jakubka, S. F. Bucher, et al. 2022. "Functional Traits Influence Patterns in Vegetative and Reproductive Plant Phenology – A Multi-Botanical Garden Study." *New Phytologist* 235: 2199–2210. <https://doi.org/10.1111/nph.18345>.
- Stefan, V., and S. Levin. 2019. "plotbiomes: Plot Whittaker Biomes With ggplot2." (0.0.0.9001) [R]. <https://valentininitelav.github.io/plotbiomes/>.
- Sutter, R. D., S. B. Wainscott, J. R. Boetsch, C. J. Palmer, and D. J. Rugg. 2015. "Practical Guidance for Integrating Data Management Into Long-Term Ecological Monitoring Projects." *Wildlife Society Bulletin* 39: 451–463. <https://doi.org/10.1002/wsb.548>.
- The Wellcome Trust. 2021. *Carbon Offset Policy for Travel*. Wellcome Trust [WWW Document]. Accessed February 25, 2025. <https://wellcome.org/node/369>.
- Villellas, J., J. Ehrlén, E. E. Crone, et al. 2021. "Phenotypic Plasticity Masks Range-Wide Genetic Differentiation for Vegetative but Not Reproductive Traits in a Short-Lived Plant." *Ecology Letters* 24: 2378–2393. <https://doi.org/10.1111/ele.13858>.
- Wiens, J. A. 1989. "Spatial Scaling in Ecology." *Functional Ecology* 3: 385–397.
- Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3: 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Wynes, S., S. D. Donner, S. Tannason, and N. Nabors. 2019. "Academic Air Travel Has a Limited Influence on Professional Success." *Journal of Cleaner Production* 226: 959–967. <https://doi.org/10.1016/j.jclepro.2019.04.109>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Supporting Information: S1.** The supporting materials can be found at: 1.1. Network governance principles for Plantpopnet (<https://doi.org/10.6084/m9.figshare.24299050>). 1.2. Authorship guidelines for Plantpopnet (<https://doi.org/10.6084/m9>).

[figshare.24299089](https://doi.org/10.6084/m9.figshare.24299089)). 1.3. Funding guidelines for Plantpopnet (<https://doi.org/10.6084/m9.figshare.24299062>). 1.4. Plantpopnet data Management (<https://doi.org/10.6084/m9.figshare.24299116>). 1.5. Abstract template for approval by steering committee (<https://doi.org/10.6084/m9.figshare.24299122>). **Supporting Information: S2.** Estimated person years invested in Plantpopnet core demographic data 2014–2021. We assumed 230 working days per year. Data curation included digital data curation and physical curation of samples (seeds, leaves etc.) as well as dealing with enquiries from participants. **Supporting Information: S3.** Recruitment of participants to the network by year of first data collection. **Supporting Information: S4.** Timing of demographic data collected in 2022 from the ongoing Plantpopnet project and estimate of survey person days for each year of additional data. Survey person days were estimated as 3 person-days per site per year including the field census and initial digitisation of data. Reasons for abandonment are given for 16 of the 65 sites. *Sites where the coordinator was informed. This means abandoned sites that were acknowledge by the site coordinator to the Plantpopnet data curators. **Supporting Information: S5.** Participant survey questions (<https://doi.org/10.6084/m9.figshare.24299278>). **Supporting Information: S6.** Key considerations and challenges to overcome when setting up a network with suggested solutions learned.