Bridging data silos to holistically model plant macrophenology

Lizbeth G. Amador[1,2], Tadeo H. Ramirez-Parada[3], Isaac W. Park[3, 4], Susan J. Mazer[3], Aaron M. Ellison[5,6], Margaret O'Brien[7], Eric R. Sokol[8], Colin A. Smith[9], Charles C. Davis[5], Sydne Record[1,2]

[1]Department of Wildlife, Fisheries, and Conservation Biology, University of Maine, Orono, ME 04469, USA
[2]Maine Agricultural and Forest Experiment Station, Orono, ME, 04469, USA
[3]Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA, 93106 USA
[4]Department of Biology, Georgia Southern University, Statesboro, GA, USA
[5]Department of Organismic and Evolutionary Biology, Harvard University Herbaria, 22 Divinity Avenue, Cambridge, MA 02138, USA
[6]Sound Solutions for Sustainable Science, Boston, MA 02135, USA
[7]Marine Science Institute, University of California, Santa Barbara, CA 93111, USA
[8]National Ecological Observatory Network, Battelle, Boulder, CO 80301, USA
[9]Center for Limnology, University of Wisconsin, Madison, WI 53706, USA

Corresponding author: Amador, L.G. (lizbeth.amador@maine.edu)

Summary

- Phenological response to global climate change can impact ecosystem functions. There are various data sources from which spatiotemporal, and taxonomic phenological data may be obtained: mobilized herbaria, community-science initiatives, observatory networks, and remote-sensing. However, analyses conducted to date have generally relied on single sources of these data.
- Siloed treatment of data in analyses may be due to the lack of harmonization across different data sources that offer partially non-overlapping information and often complementary. Such treatment precludes a deeper understanding of phenological responses at varying macroecological scales. Here, we describe a detailed vision for the harmonization of phenological data, including the direct integration of disparate sources of phenological data using a common schema.
- Specifically, we highlight existing methods for data harmonization that can be applied to phenological data: data-design patterns, metadata standards, and ontologies. We describe how harmonized data from multiple sources can be integrated into analyses using existing methods and discuss the use of automated extraction techniques.
- Data harmonization is not a new concept in ecology but the harmonization of phenological data is overdue. We aim to highlight the need for better data harmonization providing a roadmap for how harmonized phenological data may fill gaps while simultaneously integrated into analyses.

Keywords: Data harmonization, Data management, Ontologies, Scales, SDMs

Introduction

Many biological interactions depend on phenological patterns that reflect ecological and evolutionary responses to climatic conditions (e.g., Chmura *et al.*, 2019). For example, plant phenology—the recurring seasonal timing of leaf-out, flowering, fruiting, and leaf senescence—is a key set of genetically- and environmentally-controlled traits that are central to plant reproduction, plant-pollinator interactions, and availability of resources to herbivores. Plant phenology is also linked directly to ecosystem processes and services relevant to human society, such as carbon sequestration, seasonal allergies, and food security (e.g., Fatima *et al.*, 2020; Gray & Ewers, 2021; Cope *et al.*, 2022). Because plant phenology is very sensitive to ongoing rapid environmental change, there is an urgent need to better quantify and predict plant phenological dynamics (e.g., Gallinat *et al.*, 2021; Park *et al.*, 2021), including how they pertain to species range changes (Peng *et al.*, 2024; Ramirez-Parada *et al.*, 2025a).

A major challenge to quantifying and predicting phenology lies in its scale-dependence (D. S. Park *et al.*, 2021). Like many other ecological phenomena, inferences made about phenology depend on how data are combined across space, time, or taxa (Levin, 1992; D. S. Park *et al.*, 2021). Furthermore, environmental drivers of phenology (e.g., temperature, precipitation, insolation) can vary and interact differently across space and time (e.g., Peters *et al.,* 2007; Chamberlain & Wolkovich, 2023), and plastic organismal responses to these drivers can differ among individuals, populations, species, and communities (Inouye *et al.,* 2019; Ramirez-Parada *et al.*, 2024). Studying the effects of phenology on ecological processes at global scales requires data that span scales of time, taxonomy, and levels of biological organization. Understanding and analyzing scale-dependencies of phenological responses is a growing field known as macrophenology, which can inform processes at larger spatial extents (Doi *et al.,* 2017; Gallinat *et al.,* 2021).

Data sources across scales of space, time, taxonomy, and levels of biological organization do exist, although they have rarely been analyzed simultaneously. For plant phenological data, these include herbarium specimens, community science initiatives, observatory networks, and remote-sensing platforms (e.g., Gray & Ewers, 2021; Reyes-González *et al.*, 2021; Davis *et al.*, 2022; Richardson *et al.,* 2018). Different data types capture disparate ecological levels and spatiotemporal scales as a result of their sampling design and effort. For instance, remote sensing may provide continuous landscape-level monitoring over a long period of time, whereas observatory networks may provide periodic sampling with field surveys of individuals and populations that vary in their temporal extent (Fig. 1). These scale mismatches often hinder data harmonization—the direct integration of disparate data types under a common schema.

The lack of harmonization across datasets limits our ability to assess phenological responses to climate at various scales. All phenological data sources have blind spots on the ecological levels, generating observed variation across these levels. For example, remotely sensed data cannot resolve species; population-level metrics (e.g., peak flowering dates) do not resolve individuals, and data from herbarium specimens and community-science platforms (e.g., iNaturalist) are rarely repeated within individuals and populations. Integration of different phenological data sources can capture greater variation across scales. For instance, Iwanycki Ahlstrand *et al.* (2022) found that observatory network, herbarium, and remotely sensed data detected different temporal and spatial variation in phenology, such that combining these datasets captured greater variation. In some instances, data integration that captures greater variation across space leads to discoveries about regional differences in phenological responses. Everingham *et al.* (2023) used historic field data, herbarium records, and contemporary field data across New South Wales, Australia, and detected a delay in flowering phenology through time in the Southern Hemisphere compared to the advancement of flowering phenology in the Northern Hemisphere. Furthermore, single-source analyses cannot directly assess how variability at one ecological level scales to determine patterns at the next level, making them less ecologically informative. For example, detecting a longer flowering season within a community is not enough to assert that the seasonal availability of flowering species is increasing: such lengthening could occur either because populations are flowering longer, or due to a greater spread of flowering onsets among populations that may in fact decrease the average diversity of flowering species available throughout the season (Ramirez-Parada *et al.,* 2025b). Finally, although we emphasize here the value of cross-scale analyses with data integration, it is also important to note that local extent studies provide invaluable data for local management and conservation.

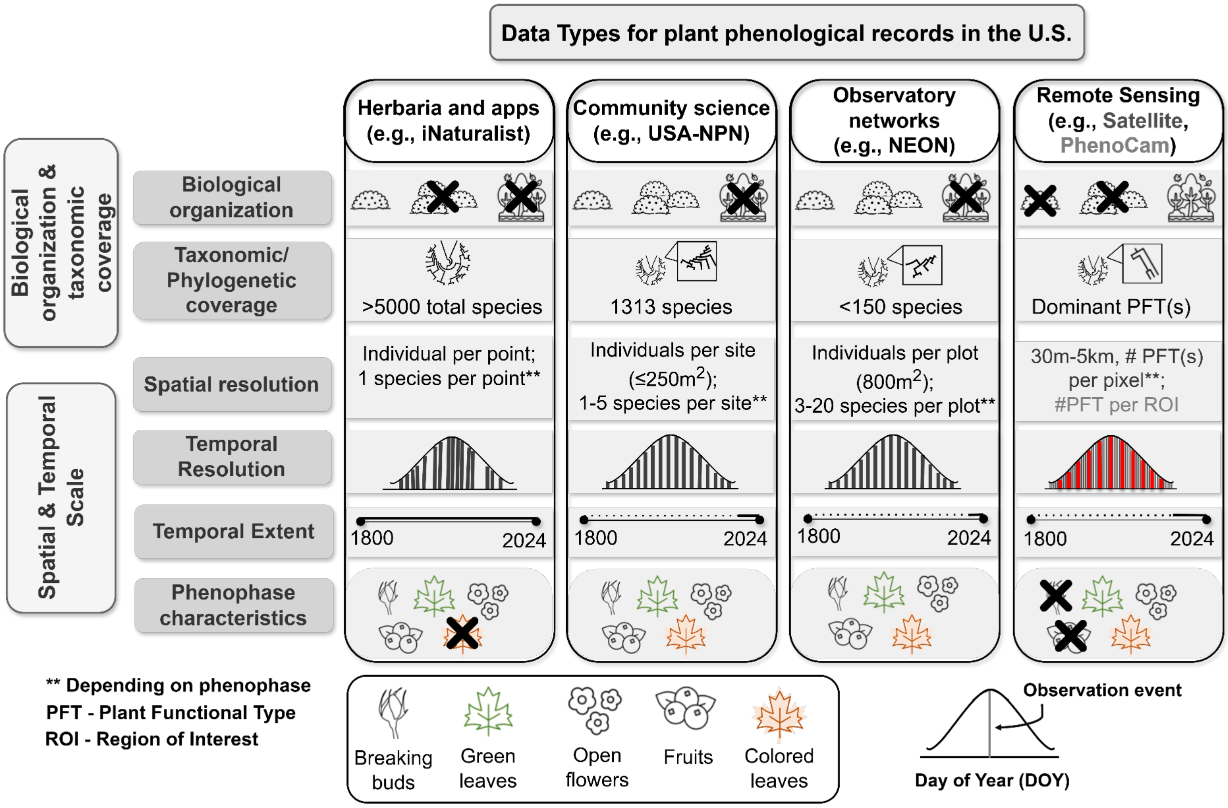


Figure 1. Data aspects across phenological data types with example datasets. Each data type offers a distinct level of biological, taxonomic, spatial, temporal, and phenophase information. Each level contains example-specific information comparing the similarities and differences across data types. Differences between data types (across columns) highlight gaps where data harmonization would benefit and maximize coverage within each level. We use a bell-curve to exemplify the sampling frequency within the duration of a phenophase. The bell-curve in the remote sensing column exemplifies both satellite (bright red) and PhenoCam (grey). The color of their text also distinguishes the spatial resolution for satellites (dark grey) and PhenoCam (light grey). The *young leaves* symbol also include fully opened green leaves for remote sensing and Phenocam, and # PFT per Region of Interest (ROI) varies for Phenocams as the cameras capture different fields of view depending on landscape characteristics (e.g., topography; Liu *et al.,* 2024). Note that there is overlap between the National Ecological Observatory Network (NEON), community science, and remote sensing because NEON uses the data collection protocols of the National Phenology Network and the USA Phenocam Network, and the NEON Airborne Observation Platform collects remotely sensed hyperspectral and red-green-blue imagery.

In this viewpoint, we assert that data harmonization is critical for improving our understanding of the impacts of climate change on macrophenology (Gallinat *et al.*, 2021; see Melaas *et al.*, 2016; Taylor *et al.*, 2019; Ramirez-Parada *et al.*, 2025a; Everingham *et al.*, 2023; Iwanycki Ahlstrand *et al.*, 2022). We focus on plant phenological datasets collected predominantly in the United States, although similar data have been collected at sites worldwide (see Nagai *et al.,* 2010, Tsuchida *et al.,* 2005; Domingo-Marimon *et al.,* 2022; Iwanycki Ahlstrand *et al.,* 2022; Mariani *et al.,* 2013; Cook *et al.,* 2012; Davis *et al.,* 2022; D.S. Park *et al.,* 2023; Appendix Table S1). We explore the characteristics of these data and reveal the unintentional data silos that limit our ability to answer a range of important ecological and evolutionary questions about phenology. Additionally, we identify how bringing together multiple data sources will enable us to answer new questions. To move toward a common goal of phenological data harmonization, we provide a road map describing methods for harmonization, how harmonization can help to fill gaps in phenological data across space and time, and methods for integrating harmonized data into analyses. We end with a call for harmonization of phenological data to rapidly advance phenological research.

A multiplicity of data sources with different strengths and weaknesses

Phenological data have provided invaluable insights into the varying effects of changing climate on plants (D. Li *et al.*, 2019; Zohner *et al.*, 2023), and recent papers highlight potential new insights to be made from each independent data type (Davis *et al.*, 2022; Dronova & Taddeo, 2022; Binley & Bennett, 2023; Zhu & Song, 2023). We cannot get a complete picture of phenology without integrating across data types due to gaps within any single data type. In this section, we draw attention to the strengths, weaknesses, and gaps of each data type with respect to space, time, taxonomy, life history, and levels of biological organization.

Herbarium Specimens

Each herbarium specimen provides phenological information observed at a specific historical point in time and at a specific location, and therefore reflects an individual plant's phenological response to recent or to long-term climatic conditions. Collectively, herbarium specimens have been mobilized to study many species' and communities' phenological sensitivity to local climatic conditions and to climatic change (e.g., Davis, 2023) at broad spatial scales (e.g., D. S. Park *et al.*, 2018; I. W. Park *et al.,* 2021; Willis *et al.*, 2017; Zhu & Song, 2023; Ramirez-Parada *et al.,* 2025b). Many of the large herbarium collections in the U.S. have increased the accessibility of phenological data contained in these specimens through massive efforts to digitize millions of physical specimens and the information contained in their labels with the centralization of data into repositories (e.g., the Global Biodiversity Information Facility [GBIF], Southwestern Environmental Information Network [SEINet]) (GBIF; SEINet Portal Network, 2023; Hedrick *et al.*, 2020; also see Phang *et al.*, 2022). However, specimens in countries with less digital infrastructure in place are less accessible or less frequently digitized, leading to biogeographical biases in spatial coverage (Daru *et al.,* 2018; Davis *et al.,* 2022). Despite the large taxonomic coverage at the species level, herbarium specimens provide relatively coarse phenophase information, represent single "snapshots" of phenology in space and time, and may exhibit sampling biases that make it unclear whether a specimen represents an early, median, or late observation relative to its source population (Ramirez-Parada *et al.*, 2022; I. W. Park *et al.*, 2024; Schmidt *et al.,* 2025; Fig. 1). Moreover, as specimens represent single observations of individuals distributed widely in space and time, variation in phenology among specimens represents both within- and among-population differences. Thus, identifying the level of ecological organization associated with relationships between phenology and environmental variables–and the mechanisms underlying such relationships–requires careful statistical design and interpretation of results (Davis *et al.,* 2015; Ramirez-Parada *et al.,* 2024; Pearse *et al.,* 2017). Another limitation of herbarium-derived data for use in phenological studies is the difficulty in identifying dates of occurrence for phenological phases other than flowering and fruiting for most species. Additionally, their patchy temporal and spatial coverage can generate sampling biases that may limit their use at global or local scales (Daru *et al.,* 2018; Schmidt *et al.,* 2025).

Community-science initiatives

Community-science initiatives harness the power of volunteers to record phenological data across broad spatial extents while providing high-resolution phenophase information from a variety of taxa (Reyes-González *et al.*, 2021; Domingo-Marimon *et al.*, 2022). Such initiatives vary in the degree of standardization used in data collection. For instance, image contributors for community-sourced app-based records (e.g., iNaturalist) do not follow specific protocols for capturing phenophases. In contrast, the USA National Phenology Network (USA-NPN) is an example of a community-science initiative with volunteer engagement across the country through their Nature's Notebook platform (Crimmins *et al.*, 2017; https://www.usanpn.org/). It has a standardized protocol to facilitate repeated observations of specific individuals (or patches) at a chosen site (Crimmins *et al.*, 2017; Fig. 1). These data provide estimates of date of onset, termination, and duration of multiple phenophases at high temporal resolution and national coverage (Fig. 1). However, while these data encompass observations for thousands of species, most correspond to a narrow set of indicator species for which specific observational protocols have been developed. Furthermore, investment by volunteers leads to large variations in taxonomic coverage and duration of observations (e.g., a single versus multiple years). These data often contain observation bias and inconsistencies in protocol implementation that can limit their application (Reyes-González *et al.*, 2021; Domingo-Marimon *et al.*, 2022). In some cases, participants might record the phenological status of only one individual at one site many times per year (sometimes less), but they might not sample multiple individuals at a given site, thus greatly limiting population-level inferences.

Observatory Networks

Observatory networks provide systematic, long-term field data that follow individuals throughout their phenological cycle (Gallinat *et al.*, 2021), thus providing opportunities to quantify inter- and intraspecific variation in phenology across ecoregions. The U.S. National Science Foundation's National Ecological Observatory Network (NEON) is one such long-term, ecological monitoring network designed to collect data through 2049 (Elmendorf *et al.*, 2016; https://www.neonscience.org/). NEON works closely with the USA-NPN and USA Phenocam networks to collect data and has adopted their standards and protocols (Richardson *et al.*, 2007; https://phenocam.nau.edu/webcam/). This offers an exciting example of how different phenological monitoring systems can coordinate efforts for standardized observations and facilitate data harmonization (Richardson *et al.*, 2007). Similar to NPN, NEON provides high-resolution phenological information through repeated measures in their field sampling design (Fig. 1). NEON also records co-located information on a suite of other biological and physical variables relevant to phenology (e.g., beetle pollinator abundance, climatic variables, carbon dioxide flux) (Nagy *et al.*, 2021). Despite the continental scale and projected 30-year lifespan scale of NEON, their NPN-style field observations are limited in i.) taxonomic coverage due to resource constraints; ii.) spatial coverage as data are collected only at the several dozen established NEON sites; and iii.) current temporal range due to the relatively recent establishment of the network in 2019 (Fig. 1). The Long-term Ecological Research Network has some sites that collect phenological information, but these data are not collected with a standardized protocol and synthesis of them is challenging (Mulder *et al.*, 2021; Schulze, 2023; but see Keenan et al., 2014). Moving forward, NEON promises to be an irreplaceable long-term reference for fine-resolution phenological data compatible with all data amassed by the USA-NPN through their active partnership and delivering NEON data through the USA-NPN portal.

Remote Sensing

Remotely-sensed data are capable of capturing continental and interannual changes in phenology, yet the spatial resolution of the data is often too coarse to discern phenological changes at the species level (Gallinat *et al.*, 2021; Reyes-González *et al.*, 2021). Satellites, near-Earth imagery, and PhenoCams have allowed for spatially continuous observations with increasing temporal resolution (Zarnetske *et al.*, 2019; Lechner *et al.*, 2020; Latifi *et al.*, 2023; Dranova and Taddeo, 2022). Remote-sensing techniques related to phenology have been applied successfully to detect the start and end of the growing season (i.e., leaf out and leaf off) for either dominant tree taxa or functional types, and primarily in temperate deciduous and tropical dry forests (Dronova & Taddeo, 2022). Recent applications of deep-learning algorithms to high-resolution hyperspectral and red-green-blue images (1-m and 0.25-m, respectively) from near-Earth (i.e., airborne) remote sensing of NEON sites enable the segmentation and identification of individual tree crowns (Weinstein *et al.*, 2024), paving the way for the detection of individual tree-crown phenology from frequent near-Earth image acquisition (e.g., by drones). Furthermore, the National Aeronautics and Space Administration's Surface Biology and Geology High Frequency Time Series (SHIFT) near-Earth remote-sensing campaign in 2023 enabled the detection of superblooms in the grasslands of coastal California from weekly flyovers (Angel *et al.*, 2025). Although hyperspectral sensors have broadened the possibilities of remotely-sensed phenological monitoring, such efforts remain limited to specific sites (e.g., NEON sites) or campaigns with high spatial and temporal resolution. Another limitation is that the earliest remote sensing data are limited to the 1970's (i.e. Landsat 1 products) and do not provide substantial pre-global warming information comparable to point-based herbarium data.

Moving forward: bridging data silos in macrophenology

Data Harmonization

Data harmonization is not a new concept in ecology. For decades there has been tremendous interest across the scientific community in pooling and harmonizing plant trait data (Keune *et al.,* 1991; Tarboton *et al.,* 2008; Reichman *et al.,* 2011; Wieczorek *et al.,* 2012; Boyle *et al.,* 2013; Pollet *et al.,* 2015; Stucky *et al.,* 2018; Record *et al.,* 2021; Flantua *et al.,* 2023). For instance, the TRY database has excelled in aggregating trait data and supported extensive advances in trait-based plant ecology, but lacks a common format that limits compatibility between data sets (Kattge *et al.*, 2011, 2020). More recently, ecologists have recognized the importance of considering intraspecific trait variation, emphasizing the coordination of open science efforts around individual-level trait information (Violle *et al.*, 2012; Cope *et al.*, 2022). Except by remote-sensing networks, phenological data are collected from individual organisms and allow for exploration of intraspecific trait variation. This makes phenological data an excellent test-bed for developing and testing approaches for data harmonization of individual-level traits. Various approaches exist for harmonizing ecological data that could be applied to phenological data. Many of these approaches incorporate common terminology and structures (i.e., design patterns) representing relational tables tracking organismal information (e.g., taxonomy, measurement [i.e., trait, number of individuals]) and other important metadata (e.g, geographic locations and differences in sampling methodologies (O'Brien *et al.*, 2021; Keller *et al.*, 2023)(Fig. 2).

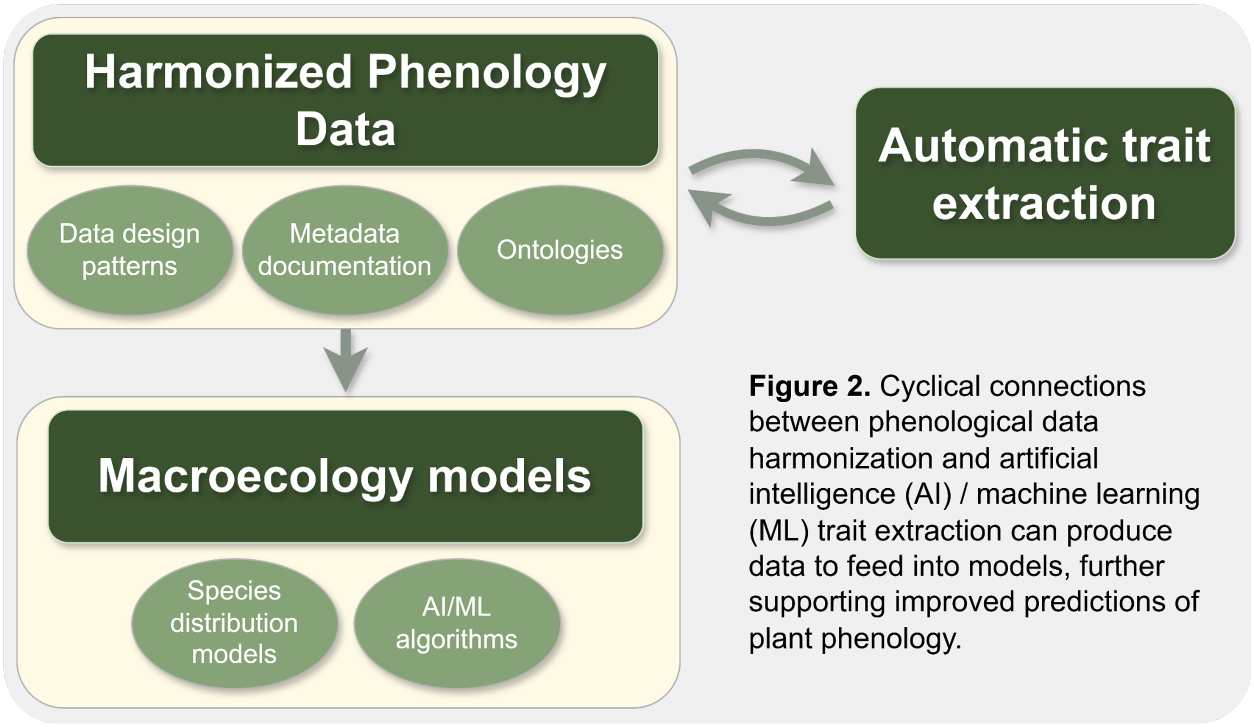


Figure 2. Cyclical connections between phenological data harmonization and artificial intelligence (AI) / machine learning (ML) trait extraction can produce data to feed into models, further supporting improved predictions of plant phenology.

Figure 2. Cyclical connections between phenological data harmonization and artificial intelligence (AI) / machine learning (ML) trait extraction can produce data to feed into models, further supporting improved predictions of plant phenology.

Using common terminology and notation is a key aspect of harmonizing phenological data to make it easier for researchers (and algorithms) to discover and use data. Ontologies provide a structured, formal language for the standardization of terminology and concepts related to data management (see Stucky *et al.*, 2018; Schneider *et al.*, 2019; O'Brien *et al.*, 2021; Lenters *et al.*, 2021; Dumschott *et al.*, 2023; Keller *et al.*, 2023). To our knowledge, the most well-developed ontology of phenology terms is the Plant Phenology Ontology (PPO, Stucky *et al.*,2018), which assembled a robust aggregated vocabulary from global phenological records. Contributions towards such efforts are crucial for dispelling uncertainties in naming conventions for phenophases. For example, intensity-based vocabulary may require a minimum percentage of reproductive organs to be displayed as "open flowers" for an individual's phenophase to be identified as "flowering", whereas qualitative assessments of flowering status may simply require the presence of a single open flower. PPO uses a framework that allows integration with vocabularies that capture other ecological traits and important information (e.g., the location of observation). The Ecological Trait Data Standard (ETS; Schneider *et al.,* 2019) is a common vocabulary to facilitate trait data harmonization and is implemented into some trait data integration networks (Open Traits Network; Gallagher *et al.*, 2020). Notably, ETS incorporates terms used by the Darwin Core Standard (DwC), a glossary of terms to facilitate sharing information about biodiversity maintained by the Biodiversity Information Standards Taxonomic Databases Working Group. Common vocabularies propagate a shared understanding of phenological phenomena that lay the groundwork for the harmonization of data into a common structure. Robust multilingual vocabularies already exist; ENVO (Buttigieg *et al.*, 2016) and EnvThes (FAIRsharing.org, 2025) both describe ecosystem-level concepts that dovetail with phenology and could be integrated at some level. PPO, ENVO, and EnvThes are all capable of handling languages coded by ISO 639-3 (which includes Indigenous languages), although for the most part, translations are limited to major European languages and English. Work to reconcile subtle differences in meaning and cultural context is complex (Vanderbilt *et al.*, 2010), and use of Indigenous languages will require additional extensive work in areas of governance and data sovereignty (Jennings *et al.,* 2025).

Common terminology may also make it easier to consistently document differences in protocols for observing the phenophase across data types, which is important for contextualizing inferences from the data. For instance, when combining community-science and observatory-network data in analyses, it is incumbent to know the difference between observation methods because each has different levels of observation uncertainty (Binley & Bennett, 2023). Accessible information about sampling design would help compare levels of uncertainty between data types and assess their degree of interoperability for specific research objectives (Fig. 3). Sharing reproducible methods in open platforms such as protocol.io (https://protocols.io; Accessed June 3, 2024) will contribute towards the development of standard disciplinary formats that are useful in metadata curation. Shared protocols between NEON and NPN are a notable example, but even their metadata is found only on site-specific publications or their websites. Furthermore, any protocols used in the collection or curation of herbarium specimens (including their digital records) that are relevant to plant phenology should be included in the metadata or the specimen label. This fits perfectly with the concept of the Global Metaherbarium and the Extended Specimen concept (Davis, 2023). Protocol standards with accessible metadata information and sampling disclosures will be key in supporting data integration and harmonization among data sources while highlighting commonalities and differences in observations (Schneider *et al.*, 2019; Dumschott *et al.*, 2023; Keller *et al.*, 2023).

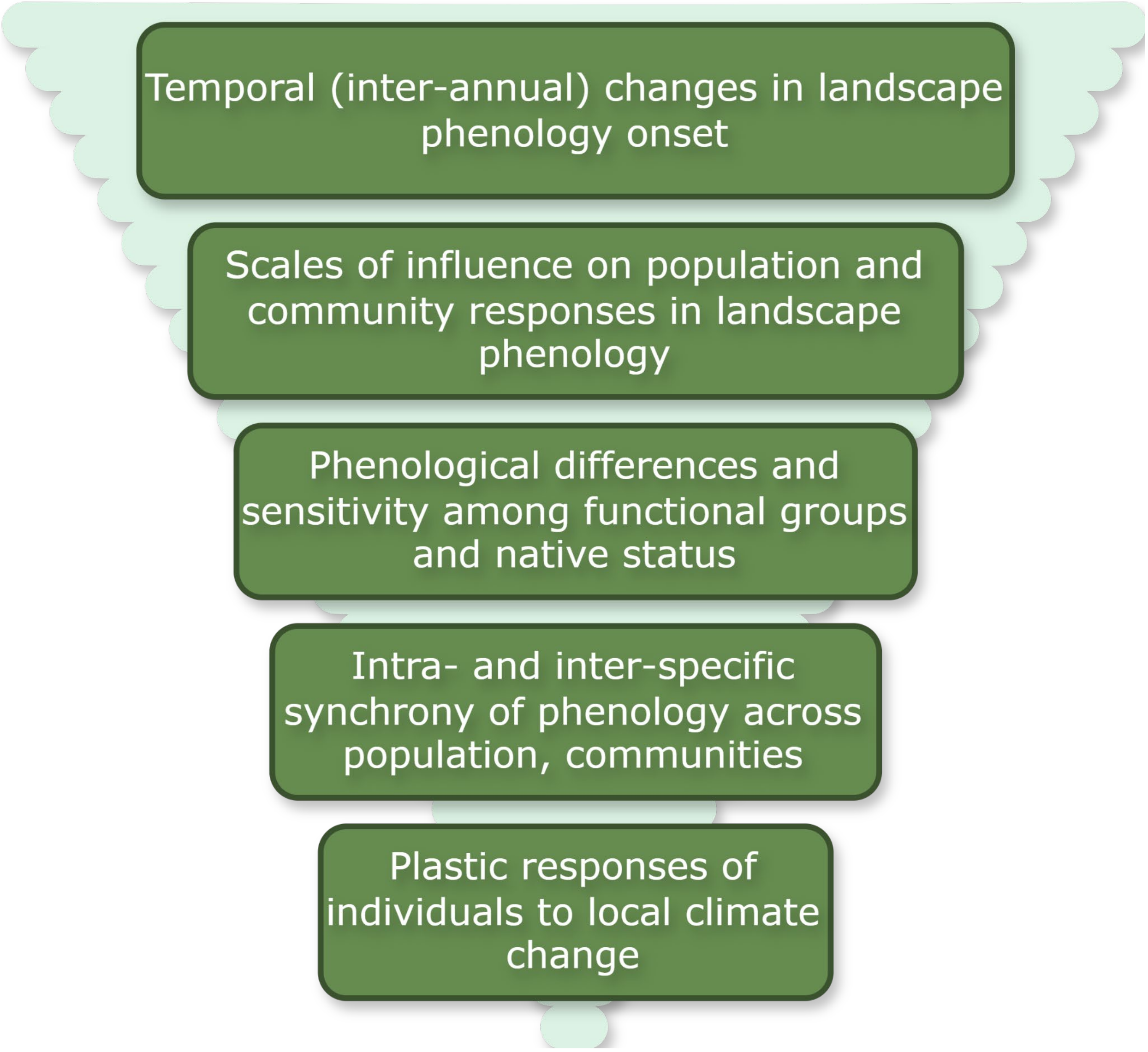


Figure 3. Hierarchy of general objectives in plant macrophenology, from broad to specific. The phenophase event of interest may determine the data types appropriate for achieving the objective.

Along with common terminology and well-documented protocols for phenological data collection, a robust data design pattern (i.e., relational structure) will empower the increased integration of harmonized data into derived data products that may make it easier to account for differences in sampling effort or biases in downstream analyses. Although there are various data design patterns for plant traits (e.g., structural traits of palms incorporating the ETS [Lenters *et al.,* 2021]), there is not a well-adopted data design pattern for plant phenological traits. We propose a phenological trait extension of the Ecological Community Data Design Pattern (ecocomDP), which was developed for harmonizing community ecology biodiversity data (O'Brien *et al.*, 2021). The original ecocomDP model is extended with two features: 1) reconfiguring the table for mapping variables to external dictionaries to allow any variable attribute (e.g., a trait) to be recorded and linked to an external dictionary of concepts, such as the ontologies mentioned above, and 2) adding additional descriptive fields to the dataset summary table (Fig. 4, red boxes). Because ecocomDP already accommodates community-level analyses, this extension would enable researchers to ask questions across levels of biological organization (e.g., from individuals to populations to communities). Another advantage of incorporating trait data into ecocomDP is that existing NEON and Long Term Ecological Research (LTER) Network data from various taxa are already harmonized with ecocomDP, making it a good candidate for the future incorporation of phenological traits as well as additional individual-level traits. ecocomDP also employs concepts used by ETS and DwC, making data harmonized into its structure easily convertible to the DwC-Archive. Finally, ecocomDP strongly emphasizes metadata, which is essential to ensure that downstream users can determine the relevance of the data for their study objectives through filtering.

Although ecocomDP is designed to harmonize in-situ phenological observations, this initial step towards standardization across observations has the potential to increase the compatibility of in-situ and remotely sensed phenological data. When harmonizing in-situ and satellite phenological data, one must consider the spatial resolutions at which they are recorded (Angel *et al.*, 2025). Many in-situ phenological data are point observations or recorded on a designated plot. Remotely sensed phenological satellite data are gridded or rasterized information and may be recorded at different spatial, temporal, and spectral resolutions. Robust data design patterns like ecocomDP ensure that spatial metadata (e.g. latitude/longitude and geographic coordinate systems) for in-situ observations are included in an intermediate format that is readily interoperable and can be further derived into gridded information for alignment with rasterized satellite images (O'Brien *et al.,* 2021). The ecocomDP model is based on when-where-taxon-what_was_measured. The location table is based on a point, which works for small plot data. For satellite data, a pixel is best represented with its center point as the locus, and its size, extent and dimensions recorded in the location_ancillary table. Taxa are often inexact, and this is particularly true for satellite data. Exact identification is not required by the model. Indiscernible taxa can be listed as “taxon 1”, taxon 2”, etc., with finer identification and taxonomic reference added later. A key-value table accommodates any measurement. For satellite data the simplest measurement might be color intensity, or some other measurement obtained via an RS algorithm, linked to an identifier with more information.

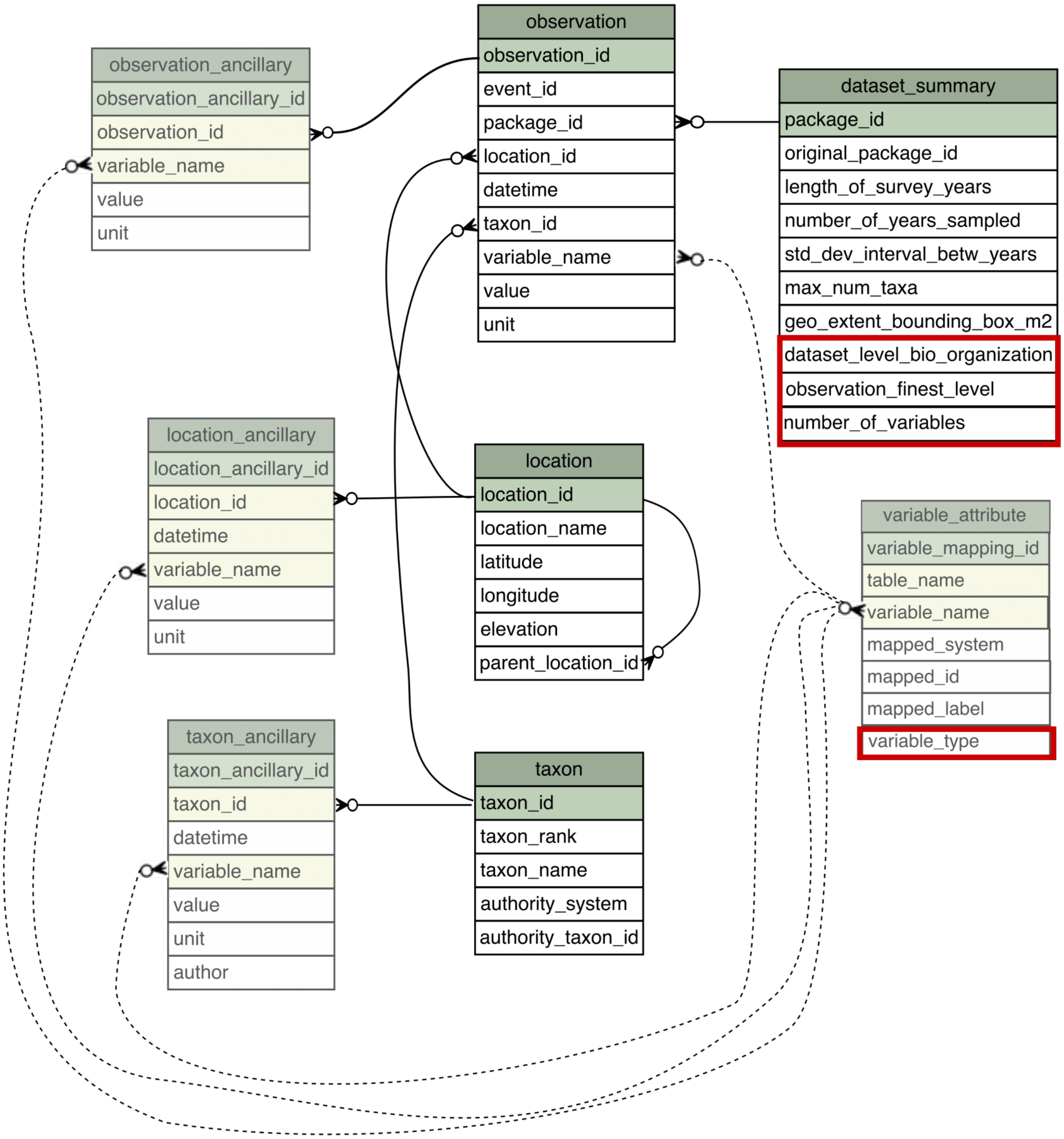


Figure 4. Schema of the updated ecocomDP data design pattern with the extension to accommodate traits (O'Brien *et al.*, 2021). Added fields in the dataset summary table (boxed in red) allow users to include the level of biological organization, level of observation, and the number of variables associated with the trait. The variable attribute (previously variable_mappling) now includes variable type to indicate the type of trait measured (e.g. start

and end dates).

For instance, formatting phenological data from different in-situ sources (e.g., herbaria, phenocams, field observations) into a common intermediate format complete with spatial metadata (e.g., spatial point or bounding box coordinates with geographic coordinate system information) facilitates the integration of in-situ phenological observations and satellite-derived phenological data that may be recorded at different spatial resolutions. Phenological estimates derived from USA-NPN or herbarium point observations can be summarized within the grid cells of a raster that is of the same resolution and extent as the rasterized image of phenological satellite data. For example, USA-NPN creates spring indices to map the onset of spring based on observations submitted by community observers (Crimmins *et al.,* 2017). Furthermore, climate data can be used to create anomaly indices of events or near-term forecasts such as the date of the first appearance of leaves or flowers, similar to the Start of Season (SOS) and End of Season (EOS) satellite-derived products (Crimmins *et al.,* 2017; Schwartz & Hanes, 2010; Wheeler *et al.,* 2024). Derived products from data design patterns such as ecocomDP can streamline reformatting tasks between point and rasterized phenological data, facilitating compatibility between the two data sources.

Methods for integrating harmonized phenological data into analyses

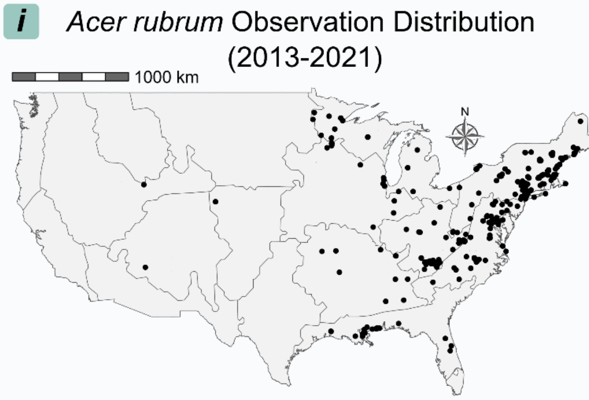
When modeling phenology we need to expand our perspective on where we can apply observations beyond classic phenological models (i.e., location-specific growing degree day models; Chamberlain & Wolkovich, 2023). With the integration of different data types into analyses, models must account for underlying biases from different data containing information on phenology and its drivers across spatial and temporal scales. Here we discuss macroecological approaches to solving two challenges: 1) differences in how phenophases are recorded and 2) spatial and temporal mismatches between phenological, geographic, and climatic data. These hurdles greatly impede efforts of macroscale phenology studies as they limit the geographic scope and questions that may be explored (Gallinat *et al.*, 2021).

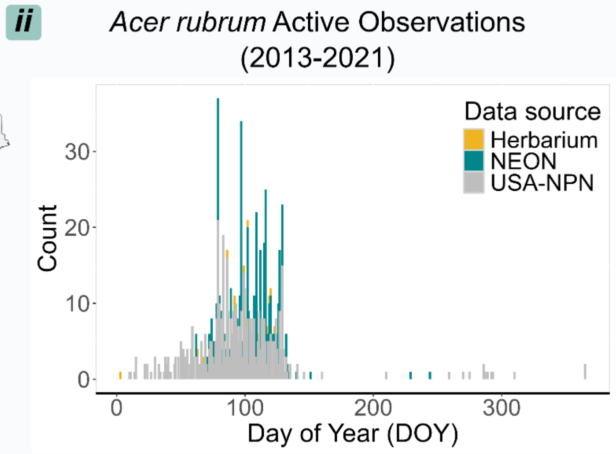
To address the first challenge of differences in how phenophases are recorded, there are many methods phenologists could adapt from species distribution modeling (SDM). Instead of modeling a response of species occurrence or abundance, we can model the probability of occurrence of a phenophase throughout the year. For instance, such a model could be used to create rasterized forecasts of species-level phenological point observations from herbarium, community science, or observatory network data into a spatially gridded dataset that is compatible with remote sensing data (Peng *et al.*, 2024). For example, Yoder *et al.* (2024) used herbarium and community-science data to create gridded predictions of whether Joshua trees are expected to have masted or flowered for each year and location. These rasterized predictions of mast events can then be compared to remotely-sensed gridded data on leaf phenology (i.e., peak greenness). Notably, most airborne or satellite remotely sensed phenological data cannot discern information below the community level (e.g., plant functional types), but combining such remotely sensed data with harmonized in-situ phenological observations (e.g., point data for dominant genera or species from site visits or PhenoCams) can enable inference of higher taxonomic resolutions (see Domingo-Marimon *et al.,* 2022; Angel *et al.,* 2025; Browning *et al.,* 2017; Chandra *et al.,* 2022; Shao *et al.,* 2023).

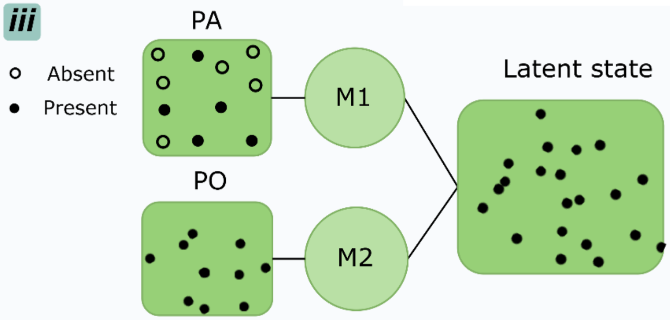
Another approach to account for differences in how phenophases are recorded is occupancy modeling (OM), which is widely used in the field of macroecology to model species distributions and provides a rich methodology by addressing imperfect detection and incorporating geographic location error from specimens for analyses of harmonized phenological data with differences in sampling effort (Erickson & Smith, 2021). One difference in sampling effort presented by phenological data is that some data types only record presence-only (PO) information on phenophases (e.g., herbarium records; iNaturalist), whereas others record both presences and absences (PA) of phenophases (e.g., NPN, NEON). With an OM framework, differences in sampling effort can be accounted for by treating each data type as a designated survey or method to account for differences in detection. Recent advances with integrated SDMs that model PO and PA responses provide a powerful approach for combining data types with differences in sampling effort (Miller *et al.*, 2019; Isaac *et al.*, 2020; Mäkinen *et al.*, 2024). Integrated SDMs could be a powerful way to combine disparate phenological data types (Box 1) and overcome the challenge of accounting for differences in how phenophases are recorded.

Box 1. ISDM roadmap for harmonized models

To illustrate how ISDMs could be used to simultaneously model different types of phenological data, we present a roadmap to constructing a harmonized model using open flowers of red maples (*Acer rubrum*) using open-source data types that differ in detection of flowering (herbarium [PO] and field observations from NEON and USA-NPN [PA]). To better understand the heterogeneity across data types, it is important to first understand where observations occur in space in time to guide model parameterization (*panel i, ii*). Visualizations of observations for each data source depict disparities in sampling effort across time. The main assumption for an ISDM is that the data observed are modeling the same ecological state where the true distribution is unknown (i.e. latent state; *panel iii*). The common parameters shared between each model address the assumption of observations pertaining to the same ecological state (*panel iv*). The ISDM incorporates shared spatial biases and any known sampling biases into the species-specific models to predict the probability of flowering occurring across space at a point in time (*panel iv*) to produce a probabilistic map of flowering across space at that time. Examples of available software for running ISDMs include the pointedSDM or intSDM R packages (Mostert and O'Hara, 2025; Mostert *et al.,* 2025).







iv

M1 $y \sim Binomial(p_i)$
$log(p_i) = \beta_0 + \beta1 * x_i + \zeta_{shared}$

M2 $y \sim Poisson(\omega(s) * e^{\lambda\_p(s)})$
$\lambda_p(s) = \beta_0 + \beta1 * x_i + \zeta_{shared} + \zeta_{bias}$

Where p_i the probability of presence in location i, x_i is a coefficient value at location i, ζ_{shared} is the shared spatial bias, w(s) is a thinning parameter, ζ_{bias} is the spatial random field to account for sampling bias, A second challenge is that phenological data and drivers of phenological responses are often measured at different temporal scales and spatial scales (and therefore, levels of biological organization) (Fig. 1). Automated integration and synthesis tools have begun to be developed to facilitate cross-scale phenological studies. For instance, the Pheno-Synthesis Software Suite (PS3) summarizes ground-based phenological observations into gridded climate and phenological indices (Morisette *et al.*, 2021). One consideration in using such software is understanding what spatial and temporal resolutions and extents have the greatest influence on phenology; to explore the optimal spatial scale between phenological data and its drivers (e.g., climate, land use topography), different grains (e.g., varying radii around a central phenological observation point or pixel) and extents (e.g., continental, ecoregion, site for NPN or NEON) that are then compared in analyses (Zarnetske *et al.*, 2019; Read *et al.*, 2020; Z. Li, 2022).

A promising approach for exploring scales of space and time simultaneously is through interpretable machine learning (ML). Interpretable ML aims to understand what input data characteristics are most important in driving predictions of output data. Local interpretation with ML—wherein the prediction of a model for a single observation in space and time is considered, as opposed to trying to understand the overall predictive behavior of the model across the entire dataset—is especially relevant for exploring spatio-temporal drivers of geo-referenced phenological data. This allows for the visualization and estimation of interactions between location features (i.e., spatial coordinates of phenological data points or grid cells) and other model features (e.g., temperature data represented by different spatial resolutions or temporal lags; Z. Li, 2022). An example of a local interpretation method comes from an extension of the Shapley value in game theory (Shapley 1953), which evaluates how contributions of different players collectively result in a contest's outcome. SHapley Additive exPlanations (SHAP), a recent ML offshoot of Shapley values, quantifies how much each feature collectively contributes to averaged model predictions (Štrumbelj & Kononenko, 2014). Historical phenological data could be used as features in such a model to predict contemporary or future phenological responses. The Shapley value and other local interpretability methods (e.g., Local Interpretable Model-agnostic Explanation or LIME; Ribeiro *et al.*, 2016) offer an exciting new opportunity to simultaneously explore spatiotemporal effects of drivers of harmonized phenological data. Overall, existing modeling approaches from macrosystems ecology and data informatics pose unique solutions to challenges by the integration of phenological data simultaneously into analyses.

Concluding remarks

Predicting plant phenological responses to global change is important given its close ties to ecosystem processes and human health. However, given the scale-dependence of plant phenology, it is difficult to make informed predictions in the absence of data that spans space, time, taxa, and levels of biological organization. Fortunately, such data are at our fingertips through various efforts in recording plant phenology at different scales and with different methods of observation, but the distinct types of phenological data need to be harmonized to unlock their full potential. Efforts to bridge phenological data silos can benefit from successful examples from other subdisciplines in ecology. Approaches to harmonize data can be adopted from existing ecological data design patterns, metadata standards, and ontologies. Biogeographic and macroecological studies offer many solutions for integrating disparate data with unique sampling biases into models. They provide a rich methodology for tackling imperfect detection and incorporating geographic location error from specimens (Erickson & Smith, 2021). Data informatics approaches are another promising tool to automate data extraction and harmonization while improving predictions of plant phenology through pattern detection. Data interoperability is not a new concept in ecology, and phenological data harmonization is long overdue.

Glossary

Data design pattern: A blueprint that captures the essential data characteristics so that a centralized workflow can access, reformat, and structure data (O'Brien *et al.*, 2021).

Data harmonization: Direct integration of different plant phenological data categories (e.g., community science, herbarium, remote sensing) under a common schema

Data integration (or data interoperability): Disparate data sources that may be used in tandem and are readably applicable in modeling or management frameworks (Brenskelle *et al.*, 2019; Stucky *et al.*, 2018; Wilkinson *et al.*, 2016).

Data management: The organization and handling of data that supports its continuous discovery, evaluation, and reuse (Wilkinson *et al.*, 2016).

Machine learning (ML): A subset of methodologies that use algorithms to automate learning predictions about data (e.g., Deep Learning, random forest; “Artificial Intelligence (AI) vs. Machine Learning,” n.d.; Pearson *et al.*, 2020).

Occupancy model (OM): A spatially explicit model that determines the occupation of an organism using presence and absence information.

Ontology: Standardized vocabulary and a language framework using formal logic that relates terms to concepts and allows for the integration of different data (Madin *et al.*, 2008; Stucky *et al.*, 2018).

Plant phenology: The timing of recurring life stages (reproductive or growth) of a plant; with a focus on angiosperms.

Phenophase: The phenological stage of a plant or animal, that details a particular life cycle stage (e.g., leaf emergence, migration, breeding).

Species Distribution Model (SDM): A form of occupancy modeling that predicts species distributions over space based on the attributes of the locations where they are currently known to occur.

Acknowledgments

This work was supported by the National Science Foundation through NSF DEB-1556768 (S.J. Mazer and I.W. Park), NSF DEB-2105932 (S.J.M. and I.W.P.), NSF DEB-2105907 (S.Record) and NSF DEB-2105903 (C.C. Davis). Additionally, for S.R. and L.G. Amador the project was supported by the USDA National Institute of Food and Agriculture, Hatch Project Number ME0-22425 through the Maine Agricultural and Forest Experiment Station. The National Ecological Observatory Network is a program sponsored by the U.S. National Science Foundation and operated under cooperative agreement by Battelle. This material is based in part upon work supported by the U.S. National Science Foundation through the NEON Program.

Conflict of Interest Statement

L. G. A. and S. R. are in a working group with Daijiang Li, Kai Zhu, and Tong Qui who may appear as potential reviewers. The authors have no other conflicts of interest to disclose.

Author Contributions

L.G.A., A.M.E, C.C.D. conceived the initial ideas which were further developed and refined with S.J.M., I.W.P., T.R-P., and S.R. L.G.A. designed and developed outlines which were further refined by S.R., C.C.D., A.M.E., I.W.P., S.J.M., and T.R-P. L.G.A. and M.O'B drafted and compiled figures which were further refined by S.R., C.C.D., A.M.E., I.W.P., S.J.M., T.R-P, C.A.S, and E.R.S. L.G.A. led the writing of the manuscript and S.R., C.C.D., A.M.E., I.W.P., S.J.M, M.O'B, C.A.S., and E.R.S. contributed significantly to the subsequent revisions. S.R. served as Ph.D. advisor for L.G.A. All authors contributed critically to the drafts and gave final approval for publication.

Data Availability

The data used to create graphs from Box 1 are openly available in Environmental Data Initiative (EDI) at http://doi.org/[***doi in progress***], reference number [***reference number in progress***]. Additionally, the data derived in this article are available from USA-National Phenology Network at http://doi.org/10.5066/F78S4N1V, National Ecological Observatory Network at https://www.neonscience.org/data, Dryad at https://datadryad.org/stash, and EDI at https://edirepository.org/. These data were derived from the following resources available in the public domain:

Switzer J, Chamberlain S, Marsh L, Wong K (2024). _rnpn: Interface to the National 'Phenology' Network 'API'_. R package version 1.2.8.0, <https://CRAN.R-project.org/package=rnpn>.

NEON (National Ecological Observatory Network). 2024. Plant phenology observations, DP1.10055.001 (RELEASE-2023). 2013-2021 for Region: Contiguous United States. Dataset accessed May 2023 via the USA National Phenology Network at http://doi.org/10.5066/F78S4N1V.

Park, Isaac et al. (2023). Herbarium-Derived Phenological Data in North America [Dataset]. Dryad. https://doi.org/10.25349/D9WP6S. Accessed Feb 2023

Park, D., A. Williams, E. Law, A. Ellison, and C. Davis. 2023. Assessing Plant Phenological Patterns in the Eastern United States Over the Last 120 Years ver 5. Environmental Data Initiative. https://doi.org/10.6073/pasta/bfb70a1701ef23f686fcc73840e6ae17 (Accessed 2023-08).

References

Angel Y, Raiho A, Kathuria D, Chadwick KD, Brodrick PG, Lang E, Ochoa F, Shiklomanov AN. **2025**. Deciphering the spectra of flowers to map landscape-scale blooming dynamics. *Ecosphere* **16**: e70127.

Artificial Intelligence (AI) vs. Machine Learning. *CU-CAI.*

Binley AD, Bennett JR. **2023**. The data double standard. *Methods in Ecology and Evolution* **14**: 1389–1397.

Brenskelle L, Stucky BJ, Deck J, Walls R, Guralnick RP. **2019**. Integrating herbarium specimen observations into global phenology data systems. *Applications in Plant Sciences* **7**: e01231.

Boyle B, Hopkins N, Lu Z, Raygoza Garay JA, Mozzherin D, Rees T, Matasci N, Narro ML, Piel WH, Mckay SJ, *et al.* **2013**. The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics* **14**: 16.

Browning DM, Karl JW, Morin D, Richardson AD, Tweedie CE. **2017**. Phenocams Bridge the Gap between Field and Satellite Observations in an Arid Grassland Ecosystem. *Remote Sensing* **9**: 1071.

Buttigieg PL, Pafilis E, Lewis SE, Schildhauer MP, Walls RL, Mungall CJ. **2016**. The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. *Journal of Biomedical Semantics* **7**: 57.

Chamberlain CJ, Wolkovich EM. **2023**. Variation across space, species and methods in models of spring phenology. *Climate Change Ecology* **5**: 100071.

Chandra S, Singh A, Mathew JR, Singh CP, Pandya MR, Bhattacharya BK, Solanki H, Nautiyal MC, Joshi R. **2022**. Phenocam observed flowering anomaly of Rhododendron arboreum Sm. in Himalaya: a climate change impact perspective. *Environmental Monitoring and Assessment* **194**.

Chmura HE, Kharouba HM, Ashander J, Ehlman SM, Rivest EB, Yang LH. **2019**. The mechanisms of phenology: the patterns and processes of phenological shifts. *Ecological Monographs* **89**: e01337.

Cook BI, Wolkovich EM, Davies TJ, Ault TR, Betancourt JL, Allen JM, Bolmgren K, Cleland EE, Crimmins TM, Kraft NJB, *et al.* 2012. Sensitivity of Spring Phenology to Warming Across Temporal and Spatial Climate Gradients in Two Independent Databases. Ecosystems 15: 1283–1294.

Cope OL, Burkle LA, Croy JR, Mooney KA, Yang LH, Wetzel WC. **2022**. The role of timing in intraspecific trait ecology. *Trends in Ecology & Evolution* **37**: 997–1005.

Crimmins TM, Crimmins MA, Gerst KL, Rosemartin AH, Weltzin JF. **2017**. USA National Phenology Network's volunteer-contributed observations yield predictive models of phenological transitions. *Plos One* **12**: e0182919.

Daru BH, Park DS, Primack RB, Willis CG, Barrington DS, Whitfeld TJS, Seidler TG, Sweeney PW, Foster DR, Ellison AM, *et al.* 2018. Widespread sampling biases in herbaria revealed from large-scale digitization. *The New Phytologist* **217**: 939–955.

Davis CC. **2023**. The herbarium of the future. *Trends in Ecology & Evolution* **38**: 412–423.

Davis CC, Champ J, Park DS, Breckheimer I, Lyra GM, Xie J, Joly A, Tarapore D, Ellison AM, Bonnet P. **2020**. A New Method for Counting Reproductive Structures in Digitized Herbarium Specimens Using Mask R-CNN. *Frontiers in Plant Science* **11**: 1129.

Davis CC, Lyra GM, Park DS, Asprino R, Maruyama R, Torquato D, Cook BI, Ellison AM. **2022**. New directions in tropical phenology. *Trends in Ecology & Evolution* **37**: 683–693.

Davis CC, Willis CG, Connolly B, Kelly C, Ellison AM. **2015**. Herbarium records are reliable sources of phenological change driven by climate and provide novel insights into species' phenological cueing mechanisms. *American Journal of Botany* **102**: 1599–1609.

Doi H, Gordo O, Mori T, Kubo MT. **2017**. A macroecological perspective for phenological research under climate change. *Ecological Research* **32**: 633–641.

Domingo-Marimon C, Masó J, Ester P, Zabala A, Serral I, Batalla M, Ninyerola M, Cristóbal J. 2022. Aligning citizen science and remote sensing phenology observations to characterize climate change impact on vegetation. Environmental Research Letters 17: 085007.

Dronova I, Taddeo S. **2022**. Remote sensing of phenology: Towards the comprehensive indicators of plant community dynamics from species to regional scales. *Journal of Ecology* **110**: 1460–1484.

Dumschott K, Dörpholz H, Laporte M-A, Brilhaus D, Schrader A, Usadel B, Neumann S, Arnaud E, Kranz A. **2023**. Ontologies for increasing the FAIRness of plant research data. *Frontiers in Plant Science* **14**.

Elmendorf SC, Jones KD, Cook BI, Diez JM, Enquist CAF, Hufft RA, Jones MO, Mazer SJ, Miller-Rushing AJ, Moore DJP, *et al.* 2016. The plant phenology monitoring design for The National Ecological Observatory Network. *Ecosphere* **7**: e01303.

Erickson KD, Smith AB. **2021**. Accounting for imperfect detection in data from museums and herbaria when modeling species distributions: combining and contrasting data-level versus model-level bias correction. *Ecography* **44**: 1341–1352.

Everingham, S. E., Blick, R. A. J., Sabot, M. E. B., Slavich, E., & Moles, A. T. 2023. Southern hemisphere plants show more delays than advances in flowering phenology. Journal of Ecology, 111, 380–390. https://doi.org/10.1111/1365-2745.13828.

FAIRsharing.org: EnvThes; Environmental Thesaurus, DOI: 10.25504/FAIRsharing.dS2o69, Last Edited: Wednesday, February 19th **2025**, 3:49, Last Accessed: Thursday, April 17th 2025, 12:41, Last Reviewed: Wednesday, February 19th 2025, 3:49

Fatima Z, Ahmed M, Hussain M, Abbas G, Ul-Allah S, Ahmad S, Ahmed N, Ali MA, Sarwar G, Haque EU, *et al.* 2020. The fingerprints of climate warming on cereal crops phenology and adaptation options. *Scientific Reports* **10**: 18013.

Flantua SGA, Mottl O, Felde VA, Bhatta KP, Birks HH, Grytnes J-A, Seddon AWR, Birks HJB. **2023**. A guide to the processing and standardization of global palaeoecological data for large-scale syntheses using fossil pollen. *Global Ecology and Biogeography* **32**: 1377–1394.

Gallagher RV, Falster DS, Maitner BS, Salguero-Gómez R, Vandvik V, Pearse WD, Schneider FD, Kattge J, Poelen JH, Madin JS, *et al.* 2020. Open Science principles for accelerating trait-based science across the Tree of Life. *Nature Ecology & Evolution* **4**: 294–303.

Gallinat AS, Ellwood ER, Heberling JM, Miller-Rushing AJ, Pearse WD, Primack RB.

2021. Macrophenology: insights into the broad-scale patterns, drivers, and consequences of phenology. *American Journal of Botany* **108**: 2112–2126.

GBIF.

Gray REJ, Ewers RM. **2021**. Monitoring Forest Phenology in a Changing World. *Forests* **12**: 297.

Hedrick BP, Heberling JM, Meineke EK, Turner KG, Grassa CJ, Park DS, Kennedy J, Clarke JA, Cook JA, Blackburn DC, *et al.* 2020. Digitization and the Future of Natural History Collections. *BioScience* **70**: 243–251.

Inouye BD, Ehrlén J, Underwood N. **2019**. Phenology as a process rather than an event: from individual reaction norms to community metrics. *Ecological Monographs* **89**: e01352.

Isaac NJB, Jarzyna MA, Keil P, Dambly LI, Boersch-Supan PH, Browning E, Freeman SN, Golding N, Guillera-Arroita G, Henrys PA, *et al.* 2020. Data Integration for Large-Scale Models of Species Distributions. *Trends in Ecology & Evolution* **35**: 56–67.

Iwanycki Ahlstrand, N., Primack, R.B. & Tøttrup, A.P. 2022. A comparison of herbarium and citizen science phenology datasets for detecting response of flowering time to climate change in Denmark. Int J Biometeorol 66, 849–862. https://doi.org/10.1007/s00484-022-02238-w.

Jennings L, Jones K, Taitingfong R, Martinez A, David-Chavez D, Alegado R 'Anolani, Tofighi-Niaki A, Maldonado J, Thomas B, Dye D, et al. 2025. Governance of Indigenous data in open earth systems science. Nature Communications 16: 572.

Kattge J, Bönisch G, Díaz S, Lavorel S, Prentice IC, Leadley P, Tautenhahn S, Werner GDA, Aakala T, Abedi M, *et al.* 2020. TRY plant trait database – enhanced coverage and open access. *Global Change Biology* **26**: 119–188.

Kattge J, Díaz S, Lavorel S, Prentice IC, Leadley P, Bönisch G, Garnier E, Westoby M, Reich PB, Wright IJ, *et al.* 2011. TRY – a global database of plant traits. *Global Change Biology* **17**: 2905–2935.

Keenan TF, Darby B, Felts E, Sonnentag O, Friedl MA, Hufkens K, O'Keefe J, Klosterman S, Munger JW, Toomey M, et al. 2014. Tracking forest phenology and seasonal physiology using digital repeat photography: a critical assessment. Ecological Applications 24: 1478–1489.

Keller A, Ankenbrand MJ, Bruelheide H, Dekeyzer S, Enquist BJ, Erfanian MB, Falster

DS, Gallagher RV, Hammock J, Kattge J, *et al.* 2023. Ten (mostly) simple rules to future-proof trait data in ecological and evolutionary sciences. *Methods in Ecology and Evolution* **14**: 444–458.

Keune H, Murray AB, Benking H. **1991**. Harmonization of environmental measurement. *GeoJournal* **23**.

Latifi H, Valbuena R, Silva CA. **2023**. Towards complex applications of active remote sensing for ecology and conservation. *Methods in Ecology and Evolution* **14**: 1578–1586.

Lechner AM, Foody GM, Boyd DS. **2020**. Applications in Remote Sensing to Forest Ecology and Management. *One Earth* **2**: 405–412.

Lenters TP, Henderson A, Dracxler CM, Elias GA, Kamga SM, Couvreur TLP, Kissling WD. **2021**. Integration and harmonization of trait data from plant individuals across heterogeneous sources. *Ecological Informatics* **62**: 101206.

Levin SA. **1992**. The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. *Ecology* **73**: 1943–1967.

Li Z. **2022**. Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. *Computers, Environment and Urban Systems* **96**: 101845.

Li D, Stucky BJ, Deck J, Benjamin B, Guralnick RP. **2019**. The effect of urbanization on plant phenology depends on regional temperature. *Nature Ecology & Evolution* **3**: 1661–1667.

Liu Y, Zhang X, Shen Y, Ye Y, Gao S, Tran KH. **2024**. Evaluation of PlanetScope-detected plant-specific phenology using infrared-enabled PhenoCam observations in semi-arid ecosystems. *ISPRS Journal of Photogrammetry and Remote Sensing* **210**: 242–259.

López-Guillén E, Herrera I, Bensid B, Gómez-Bellver C, Ibáñez N, Jiménez-Mejías P, Mairal M, Mena-García L, Nualart N, Utjés-Mascó M, *et al.* 2024. Strengths and Challenges of Using iNaturalist in Plant Research with Focus on Data Quality. *Diversity* **16**: 42.

Madin JS, Bowers S, Schildhauer MP, Jones MB. **2008**. Advancing ecological research with ontologies. *Trends in Ecology & Evolution* **23**: 159–168.

Mäkinen J, Merow C, Jetz W. **2024**. Integrated species distribution models to account for sampling biases and improve range-wide occurrence predictions. *Global Ecology and Biogeography* **33**: 356–370.

Mariani L, Alilla R, Cola G, Monte G, Epifani C, Puppi G, Failla O. 2013. Erratum to: IPHEN—a real-time network for phenological monitoring and modelling in Italy. International journal of biometeorology 57.

Miller DAW, Pacifici K, Sanderlin JS, Reich BJ. **2019**. The recent past and promising future for data integration methods to estimate species' distributions. *Methods in Ecology and Evolution* **10**: 22–37.

Morisette JT, Duffy KA, Weltzin JF, Browning DM, Marsh RL, Friesz AM, Zachmann LJ, Enns KD, Landau VA, Gerst KL, *et al.* 2021. PS3: The Pheno-Synthesis software suite for integration and analysis of multi-scale, multi-platform phenological data. *Ecological Informatics* **65**: 101400.

Mostert P, Bruls A, Ragnhild {Bjørkås}, Koch W, Martin E. **2025**. intSDM: Reproducible Integrated Species Distribution Models Across Norway using 'INLA', *v.2.1.1*. https://cran.r-project.org/package=intSDM

Mostert P, O'Hara B. **2025**. PointedSDMs: Fit Models Derived from Point Processes to Species Distributions using 'inlabru', *v.2.1.3*. https://cran.r-project.org/package=PointedSDMs

Mulder C, Spellman KV, Bonanza Creek LTER. **2021**. Flower and Leaf Phenology of Interior Alaska Forbs and Shrubs as Observed Near Fairbanks Alaska from 2013-2015.

Nagai S, Nasahara KN, Muraoka H, Akiyama T, Tsuchida S. 2010. Field experiments to test the use of the normalized-difference vegetation index for phenology detection. Agricultural and Forest Meteorology 150: 152–160.

Nagy RC, Balch JK, Bissell EK, Cattau ME, Glenn NF, Halpern BS, Ilangakoon N, Johnson B, Joseph MB, Marconi S, *et al.* 2021. Harnessing the NEON data revolution to advance open environmental science with a diverse and data-capable community. *Ecosphere* **12**: e03833.

O'Brien M, Smith CA, Sokol ER, Gries C, Lany N, Record S, Castorani MCN. **2021**. ecocomDP: A flexible data design pattern for ecological community survey data. *Ecological Informatics* **64**: 101374.

Park DS, Breckheimer I, Williams AC, Law E, Ellison AM, Davis CC. **2018**. Herbarium specimens reveal substantial and unexpected variation in phenological sensitivity across the eastern United States. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**: 20170394.

Park DS, Newman EA, Breckheimer IK. **2021**. Scale gaps in landscape phenology: challenges and opportunities. *Trends in Ecology & Evolution* **36**: 709–721.

Park DS, Lyra GM, Ellison AM, Maruyama RKB, dos Reis Torquato D, Asprino RC, Cook BI, Davis CC. **2023**. Herbarium records provide reliable phenology estimates in the understudied tropics. *Journal of Ecology* **111**: 327–337.

Park IW, Ramirez-Parada T, Mazer SJ. **2021**. Advancing frost dates have reduced frost risk among most North American angiosperms since 1980. *Global Change Biology* **27**: 165–176.

Park IW, Ramirez-Parada T, Record S, Davis C, Ellison AM, Mazer SJ. 2024. Herbarium data accurately predict the timing and duration of population-level flowering displays. *Ecography*: e06961.

Pearse WD, Davis CC, Inouye DW, Primack RB, Davies TJ. **2017**. A statistical estimator for determining the limits of contemporary and historic phenology. *Nature Ecology & Evolution* **1**: 1876–1882.

Pearson KD, Nelson G, Aronson MFJ, Bonnet P, Brenskelle L, Davis CC, Denny EG, Ellwood ER, Goëau H, Heberling JM, *et al.* 2020. Machine Learning Using Digitized Herbarium Specimens to Advance Phenological Research. *BioScience* **70**: 610–620.

Peng S, Ramirez-Parada TH, Mazer SJ, Record S, Park I, Ellison AM, Davis CC. **2024**. Incorporating plant phenological responses into species distribution models reduces estimates of future species loss and turnover. *New Phytologist* **242**: 2338–2352.

Peters DPC, Bestelmeyer BT, Turner MG. **2007**. Cross–Scale Interactions and Changing Pattern–Process Relationships: Consequences for System Dynamics. *Ecosystems* **10**: 790–796.

Phang, A., Atkins, H. and Wilkie, P. 2022. The effectiveness and limitations of digital images for taxonomic research. TAXON, 71: 1063-1076. https://doi.org/10.1002/tax.12767

Pollet TV, Stulp G, Henzi SP, Barrett L. **2015**. Taking the aggravation out of data aggregation: A conceptual guide to dealing with statistical issues related to the pooling of individual-level observational data. *American Journal of Primatology* **77**: 727–740.

Ramirez-Parada TH, Park IW, Mazer SJ. **2022**. Herbarium specimens provide reliable estimates of phenological responses to climate at unparalleled taxonomic and spatiotemporal scales. *Ecography* **2022**: e06173.

Ramirez-Parada TH, Park IW, Record S, Davis CC, Ellison AM, Mazer SJ. **2024**. Plasticity and not adaptation is the primary source of temperature-mediated variation in flowering phenology in North America. *Nature Ecology & Evolution* **8**: 467–476.

Ramirez-Parada TH, Park IW, Peng S, Nishino M, Kartesz JT, Record S, Davis C, Mazer SJ. **2025a**. Shifts in phenology and species ranges restructure the flowering season across North America.

Ramirez-Parada TH, Park IW, Record S, Davis CC, Mazer SJ. 2025b. Scaling flowering onset and duration responses among species predicts phenological community reassembly under warming. **Ecosphere 16 (3), e70070**

Read QD, Zarnetske PL, Record S, Dahlin KM, Costanza JK, Finley AO, Gaddis KD, Grady JM, Hobi ML, Latimer AM, *et al.* 2020. Beyond counts and averages: Relating geodiversity to dimensions of biodiversity (V Bahn, Ed.). *Global Ecology and Biogeography* **29**: 696–710.

Record S, Voelker NM, Zarnetske PL, Wisnoski NI, Tonkin JD, Swan C, Marazzi L, Lany N, Lamy T, Compagnoni A, *et al.* 2021. Novel Insights to Be Gained From Applying Metacommunity Theory to Long-Term, Spatially Replicated Biodiversity Data. *Frontiers in Ecology and Evolution* **8**.

Reichman OJ, Jones MB, Schildhauer MP. **2011**. Challenges and Opportunities of Open Data in Ecology. *Science* **331**: 703–705.

Reyes-González ER, Gómez-Mendoza L, Barradas VL, Terán-Cuevas ÁR. **2021**. Cross-scale phenological monitoring in forest ecosystems: a content-analysis-based review. *International Journal of Biometeorology* **65**: 2215–2227.

Ribeiro MT, Singh S, Guestrin C. **2016**. 'Why Should I Trust You?': Explaining the Predictions of Any Classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM, 1135–1144.

Richardson A, Jenkins J, Braswell B, Hollinger D, Ollinger S, Smith M-L. **2007**. Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. *Oecologia* **152**: 323–34.

Richardson AD, Hufkens K, Milliman T, Aubrecht DM, Chen M, Gray JM, Johnston MR, Keenan TF, Klosterman ST, Kosmala M, *et al.* 2018. Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. *Scientific Data* **5**: 180028.

Schneider FD, Fichtmueller D, Gossner MM, Güntsch A, Jochum M, König-Ries B, Le Provost G, Manning P, Ostrowski A, Penone C, *et al.* 2019. Towards an ecological trait-data standard. *Methods in Ecology and Evolution* **10**: 2006–2019.

Schulze MD. **2023**. Vegetative Phenology observations at the Andrews Experimental Forest, 2009 - Present.

SEINet Portal Network. **2023**. http//:swbiodiversity.org/seinet/index.php.

Shapley, LS. 1953. A value for n-person games. Princeton University Press. 307-318.

Shao C, Shuai Y, Wu H, Deng X, Zhang X, Xu A. **2023**. Development of a Spectral Index for the Detection of Yellow-Flowering Vegetation. *Remote Sensing* **15**: 1725.

Schmidt RJ, Davis C, Struwe L, Saban KE. **2025**. Identifying the collector practices that shape spatial, temporal, and taxonomic bias in herbaria.

Stevens S, Wu J, Thompson MJ, Campolongo EG, Song CH, Carlyn DE, Dong L, Dahdul WM, Stewart C, Berger-Wolf T, *et al.* 2024. BioCLIP: A Vision Foundation Model for the Tree of Life.

Štrumbelj E, Kononenko I. **2014**. Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems* **41**: 647–665.

Stucky BJ, Guralnick R, Deck J, Denny EG, Bolmgren K, Walls R. **2018**. The Plant Phenology Ontology: A New Informatics Resource for Large-Scale Integration of Plant Phenology Data. *Frontiers in Plant Science* **9**.

Tarboton DG, Horsburgh JS, Maidment DR. **2008**. CUAHSI Community Observations Data Model (ODM) Version 1.1 Design Specifications. *Des Doc*.

Tsuchida S, Nishida K, Iwao K, Kawato W, Oguma H, Iwasaki A. 2005. Phenological Eyes Network for Validation of Remote Sensing Data. 25: 282–288.

Vanderbilt KL, Blankman D, Guo X, He H, Lin C-C, Lu S-S, Ogawa A, Ó Tuama É, Schentz H, Su W. **2010**. A multilingual metadata catalog for the ILTER: Issues and approaches. *Ecological Informatics* **5**: 187–193.

Violle C, Enquist BJ, McGill BJ, Jiang L, Albert CH, Hulshof C, Jung V, Messier J. **2012**. The return of the variance: intraspecific variability in community ecology. *Trends in Ecology & Evolution* **27**: 244–252.

Weaver WN, Smith SA. **2023**. From leaves to labels: Building modular machine learning networks for rapid herbarium specimen analysis with LeafMachine2. *Applications in Plant*

Sciences **11**: e11548.

Weinstein BG, Marconi S, Zare A, Bohlman SA, Singh A, Graves SJ, Magee L, Johnson DJ, Record S, Rubio VE, *et al.* 2024. Individual canopy tree species maps for the National Ecological Observatory Network. *PLOS Biology* **22**: e3002700.

Wheeler KI, Dietze MC, LeBauer D, Peters JA, Richardson AD, Ross AA, Thomas RQ, Zhu K, Bhat U, Munch S, *et al.* 2024. Predicting spring phenology in deciduous broadleaf forests: NEON phenology forecasting community challenge. *Agricultural and Forest Meteorology* **345**: 109810.

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D. **2012**. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLOS ONE* **7**: e29715.

Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, *et al.* 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**: 160018.

Willis CG, Ellwood ER, Primack RB, Davis CC, Pearson KD, Gallinat AS, Yost JM, Nelson G, Mazer SJ, Rossington NL, *et al.* 2017. Old Plants, New Tricks: Phenological Research Using Herbarium Specimens. *Trends in Ecology & Evolution* **32**: 531–546.

Yoder JB, Andrade AK, DeFalco LA, Esque TC, Carlson CJ, Shryock DF, Yeager R, Smith CI. 2024. Reconstructing 120 years of climate change impacts on Joshua tree flowering. Ecology Letters. 27(8):e14478. doi:10.1111/ele.14478. [accessed 2024 Oct 17]. https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.14478

Zarnetske PL, Read QD, Record S, Gaddis KD, Pau S, Hobi ML, Malone SL, Costanza J, M. Dahlin K, Latimer AM, *et al.* 2019. Towards connecting biodiversity and geodiversity across scales with satellite remote sensing. *Global Ecology and Biogeography* **28**: 548–556.

Zhu K, Song Y. **2023**. Harnessing herbaria to advance plant phenology research under global change. *New Phytologist* **n/a**.

Zohner CM, Mirzagholi L, Renner SS, Mo L, Rebindaine D, Bucher R, Palouš D, Vitasse Y, Fu YH, Stocker BD, *et al.* 2023. Effect of climate warming on the timing of autumn leaf senescence reverses after the summer solstice. *Science* **381**: eadf5098.