# The Plant Phenology Ontology

## Integrating plant phenology data across diverse spatial and temporal scales

Brian J. Stucky
Florida Museum of Natural History
University of Florida
stuckyb@flmnh.ufl.edu

Ellen G. Denny
USA-National Phenology Network
Tucson, AZ, USA

Ramona L. Walls
CyVerse
University of Arizona
Tucson, AZ, USA

John Deck
University of California Berkeley
Berkeley, CA, USA

Robert P. Guralnick
Florida Museum of Natural History
University of Florida
Gainesville, FL, USA

Jennifer Yost
Hoover Herbarium
Cal Poly University
San Luis Obispo, CA, USA

***Abstract*—Plant phenology — the timing of life-cycle events, such as flowering or leafing-out — has cascading effects on multiple levels of biological organization, from individuals to ecosystems. Despite the importance of understanding phenology for managing biodiversity and ecosystem services, we are not currently able to address continent-scale phenological responses to anticipated climatic changes. This is not because we lack relevant data. Rather, the problem is that the disparate organizations producing large-scale phenology data are using non-standardized terminologies and metrics during data collection and data processing. Here, we preview the Plant Phenology Ontology, which will provide the standardized vocabulary necessary for annotation of phenological data. We are aggregating, annotating, and analyzing the most significant phenological data sets in the USA and Europe for broad temporal, geographic, and taxonomic analyses of how phenology is changing in relation to climate change.**

***Keywords—plant phenology; USA-NPN; PEP725; crowd sourcing; PhenoCam; MODIS; Plant Ontology***

### I. Introduction

Plant phenology — the timing of life-cycle events, such as flowering or leafing-out — has cascading effects on multiple levels of biological organization from individuals to ecosystems. Phenology not only affects individual fitness but also the fitness of mutualistic and antagonistic organisms that depend on plants. Changes in phenological responses of plants and animals are known to be highly responsive to environmental drivers and thus strongly influenced by climate change [1-4]. Temporal mismatches between plants and animals can quickly drive populations extinct, cause rapid evolutionary shifts, and can result in billions of dollars of agricultural losses [5-7]. Advancing our understanding of the drivers of phenological response can provide insight into future states of species distributions [8,9], biogeochemistry [10], and other ecosystem services like pollination [11]. Therefore, increasing scientific understanding of relationships between phenology and the structure and function of ecosystems can inform adaptive management of natural resources [12-14].

Phenological research to date has shown that most species flower earlier in response to elevated winter temperatures [15,16]. Even though there are exceptions to this "rule", the advancement of flowering dates in response to a warming climate may be the most widely and best understood general biological response to climate change, familiar to both scientists and the public alike. Nevertheless, species and families differ in the magnitude and direction of their response to inter-annual variation in monthly temperatures, and different phenological phases within species can be sensitive to different environmental cues [15]. The availability of modeled monthly and annual climatic data for any location in the U.S., extending back to 1895 [17], provides new opportunities to measure plant phenological responses to climate change at the continental level and across an unparalleled number of taxa. Access to large-scale phenological data is now the major limitation to such analyses.

Despite its importance for managing biodiversity and ecosystem services, we are not currently able to address continent-scale phenological responses to anticipated climatic changes. This is largely because disparate groups are using non-standardized terminologies and metrics during data collection and processing [16], which hampers data integration and the ability to build trusted large-scale data products. The end result is tremendous inefficiency as data and knowledge producers build similar but non-interoperable end products.

### II. Methods

#### *A. Data Sources*

Most phenological data come from four sources: ground-based observations, satellite remote sensing, digital repeat photography, and historical plant specimens in museum collections (Table 1). Ground based observations are collected through ecological monitoring programs, field notes, agricultural records, and citizen scientist programs. Since the 1970s phenology has also been recorded by capturing images of the earth's surface via remote imaging with satellites (e.g. MODIS data[1]). Sensors that measure leaf reflectance are used

[1] http://modis.gsfc.nasa.gov/data/

This work has been supported by the National Ecological Observatory Network (NEON), CyVerse (NSF-DBI-0735191 and DBI-1265383), the John Wesley Powell Center and the United States Geological Survey, Integrated Digitized Biocollections (iDigBio), the USA National Phenology Network, and the Phenotype Ontology RNC (NSF-DEB-0956049).

TABLE I. DATA SOURCES TO BE INTEGRATED USING THE PPO

| **Attribute** | **Data Source** | | | | |
|---|---|---|---|---|---|
| | *USA Natl. Phenology Network/NEON* | *Pan-European Phenology Network* | *PhenoCam* | *MODIS MOD12Q2* | *Herbarium Specimens / iDigBio* |
| *Size/Coverage* | 7.5 million records on 30,000 individuals | 11 million obs. on 121 species | 300 cameras across the US | many gridded images | 5 million records |
| *Data format* | tabular | tabular | image + tabular | raster image | tabular |
| *Aggregation* | direct single observation | direct single observation | direct single obs. + modeled product | modeled product | direct single observation |
| *Spatial grain* | site point radius | site point radius | field of view + ROI | 500 +/- m resolution | site point radius |
| *Temporal grain* | day | day | minute (sec) | year | day |
| *Frequency* | variable but known | unknown | 30 minute +/- | derived yearly from 1-2 day | single occurrence |
| *Biological unit* | individual / species / patch | individual / species / patch | community | community | species / individual |
| *Data standardization* | USA-NPN definitions | BBCH definitions | RGB values | % greenness | currently unstandardized metrics |
| *Inventory /Observation parameters* | date + location (lat/long) | date + location (admin. units) | date + location of camera | date + pixel location | date + location fine-scale |

to determine "greenness," and repeat measures can determine greenness onset and maximum over time. While these images can provide decadal information across large geographic areas, they usually ignore other vegetative phenological measures, such as flowering time, and it can be challenging to reconcile "greenness" with ground-based observations [18]. Digital repeat photography (PhenoCam [10] being perhaps the best example) bridges some of the data gaps between ground based observations and satellite images. This method involves taking photographs of the same place repeatedly through time. RGB (Red-Green-Blue) values are extracted from the images to make inferences about the timing of "greening up" or flowering. Finally, the best large-scale historical record of phenology is housed within herbaria, and the recent accessibility of over 5 million digital records from U.S. herbaria through iDigBio makes broad-scale historical analyses possible.

### B. *Ontology Development*

Initial development of the Plant Phenology Ontology (PPO) began at a January 2016 workshop with participants representing the organizations that collect many of the major sources of phenological data. The goals of this workshop were to produce a draft ontology and use it to annotate and query real phenological data.

Post-workshop development has focused on expanding and refining the ontology design patterns prototyped during the workshop, including completing the logic of leaf phenophases and extending it to flowering and fruiting phenophases. Our current development goals are to:

- Establish the terms, text definitions, and axiomatic definitions needed for the ontology to be most useful for both machine computation and human understanding.
- Establish a development workflow that incorporates best practices for a community ontology, including importing Plant Ontology (PO) [19,20] terms and requesting new terms in the Plant Trait Ontology (TO) [21]. We intend the PPO to be highly compatible with and to complement the PO and the TO, but given the PPO's specific goal of integrating phenological data, we decided it was best implemented as a separate ontology rather than as an expansion of existing ontologies.
- Develop proofs of concept that the PPO supports interoperability and new analyses by annotating key data sources across broad geographic scales.

An initial stable release of the PPO is planned for late 2016.

### C. *Data Mapping*

A major goal of our project is to put the PPO to immediate, practical use by annotating phenological source data with relevant ontology terms. Our data annotation process begins with converting existing tabular data (spreadsheets and relational databases) to RDF graphs using D2RQ[2] and tools developed by the NSF-funded BiSciCol project[3], while applying relevant ontology terms from the Plant Phenology Ontology (PPO), the Environment Ontology (ENVO; used to describe environmental classes such as biomes, environmental features, and environmental materials) [22], and the Biological Collections Ontology (BCO; used to describe biological specimen collection and observation processes plus their inputs and outputs) [23]. The tools for this process are encapsulated by the Biocode FIMS open-source software package[4]. Once data are converted to explicit, model-ready RDF graphs,

[2] http://d2rq.org/

[3] http://biscicol.org

[4] http://github.com/biocodellc/biocode-fims

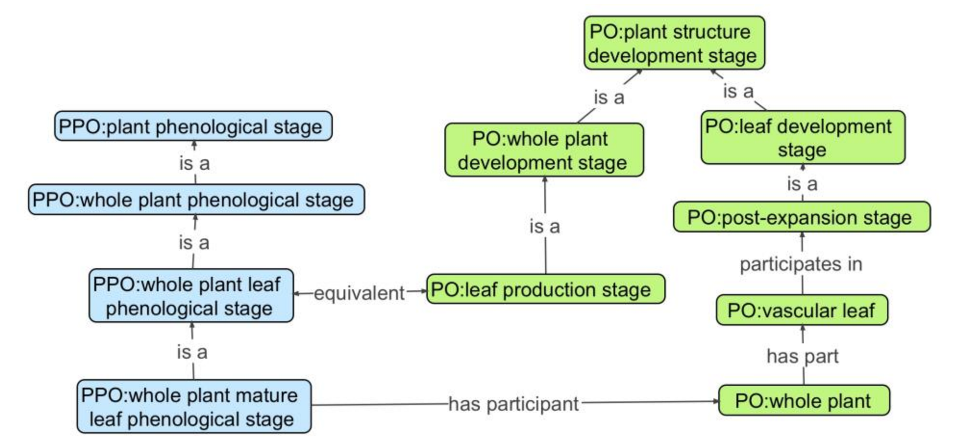


Fig. 1. PPO classes (in blue) map to PO classes (in green) for consistency and interoperability. New PPO classes are logically defined using PO classes and follow the PO design patterns as much as possible. E.g., PPO:whole plant mature leaf phenological stage *is a* PPO:whole plant leaf phenological stage that *has participant* PO:whole plant that *has part* PO:vascular leaf.

queries can be assembled across the set of graphs covering all of our data sources (Table 1).

## III. RESULTS

The initial development workshop in January 2016 produced a draft PPO that covers leaf phenological stages of whole plants and related phenological characteristics. Participants worked together to reach agreement on definitions of key PPO terms, to develop ontology design patterns for phenological stages and characteristics related to the timing of leaf development, and to ultimately produce the draft PPO.

The draft ontology references terms from the PO and follows design patterns similar to those used in the PO to ensure interoperability with PO-annotated datasets (Fig. 1). For example, the hierarchies of leaf phenological stages and characters in the PPO (where, for example, a whole plant leaf phenological stage *is a* whole plant phenological stage which *is a* plant phenological stage; see Fig. 1) mirror the plant structure hierarchy in the PO (where a leaf production stage *is a* whole plant development stage which *is a* plant structure development stage; Fig. 1). The PPO also builds upon the Phenotype and Trait Ontology (PATO) [5]; all phenological characters in the PPO are subclasses of "quality" from PATO (Fig. 2). The working copy of the PPO is available on GitHub [6] and is open source.

Workshop participants also developed a set of use cases for testing the competency of PPO for data integration and its value for answering long-standing questions in the field of plant phenology, such as testing Hopkin's Spring Time Law [24]. Using the draft PPO and focusing on a limited subset of data, workshop participants mapped subsets of data from the USA-National Phenology Network (USA-NPN) [7] and the Pan-European Phenology Network (PEP725) [8] and demonstrated a proof-of-concept in the ability to query and analyze across these data sets. Work is now ongoing on the next stages of developing the PPO, as detailed in the methods.

At the time of writing, the PPO includes 84 native classes and incorporates three ontology design patterns that cover the phenological stages of whole plants, the phenological stages of plant parts, and count-based phenological traits (e.g., the number of young leaves present at a particular time). However, these numbers will change substantially as PPO development

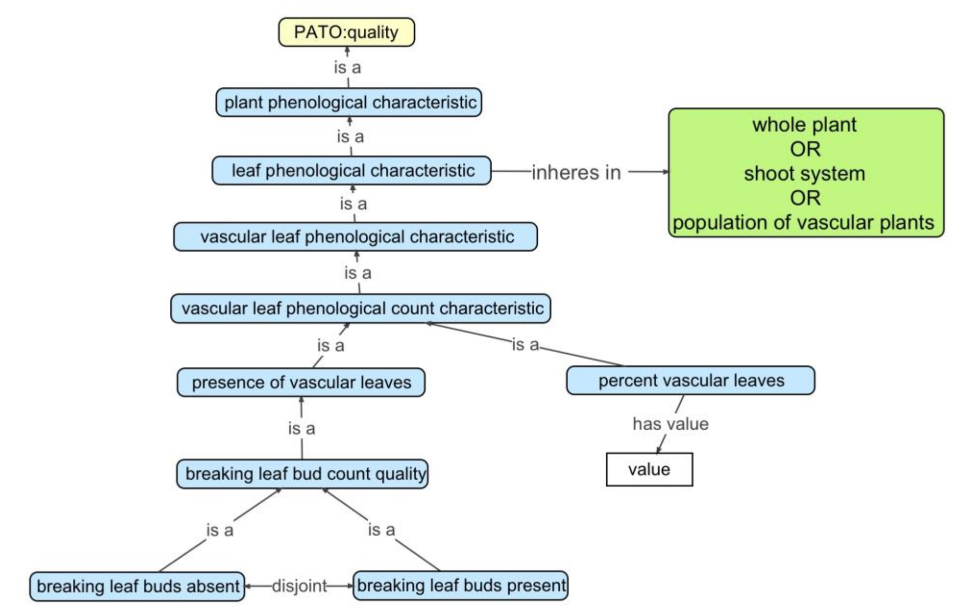


Fig. 2. Example of a partial hierarchy for phenological characters in the PPO, showing that leaf phenological characters are subclasses of "quality" from PATO (the Phenotype and Trait Ontology) that can inhere in a whole plant or shoot system (from the PO) or a population of vascular plants. Phenological characters can be categorical (present/absent) or continuous (e.g., a percentage).

[5] http://www.obofoundry.org/ontology/pato.html

[6] https://github.com/PlantPhenoOntology/PPO

[7] https://www.usanpn.org/

[8] http://www.pep725.eu/

continues.

The PPO is currently able to represent about one half of all USA-NPN and PEP725 data and should be able to represent all of these data by the end of the year. PhenoCam, MODIS, and herbarium data are targets for future development effort.

## IV. Future Directions

The PPO is a community resource, and as such must interoperate well with other resources used by the plant science, biodiversity, and ecological communities. Ongoing work is aimed at improving interoperability with domain-specific ontologies such as the PO and ENVO, but also with more general observation ontologies such as BCO (used for biodiversity collections data) and OBOE (Extensible Ontology for Observations; used in earth and environmental sciences) [25]. Outreach to drive adoption is another important aspect of a community ontology, so we are working with USA-NPN, PEP725, and PhenoCam researchers to train them in the use of PPO and ontologies in general, and to demonstrate the utility of ontologies.

The PPO aims to be a key resource for facilitating phenological data integration for applications ranging from global change research to crop breeding to conservation planning. Providing interoperable data from the four major types of phenological data, assembled into a platform for further discovery and analyses, will provide an unprecedented view of continental and global patterns of phenological change, allowing researchers to address questions such as:

- Which habitats and vegetation types are most phenologically sensitive to changes in precipitation and temperature?
- To which climatic parameters (minimum temperature, maximum temperature, dew point, cumulative growing degree days) is each species/community most sensitive?
- How do temperature and precipitation patterns interact to influence the onset of vegetative growth and flowering?
- Are there predictable responses in plant phenology that can assist managers in planning for future change?

## Acknowledgment

The authors thank the participants of the Plant Phenology workshops at the Powell Center and UC Berkeley.

## References

[1] C Parmesan and G Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37–42. doi:10.1038/nature01286

[2] TL Root, JT Price, KR Hall, SH Schneider, C Rosenzweig and JA Pounds. 2003. Fingerprints of global warming on wild animals and plants. Nature 421:57–60. doi:10.1038/nature01333

[3] EE Cleland, I Chuine, A Menzel, HA Mooney and MD Schwartz. 2007. Shifting plant phenology in response to global change. Trends Ecol Evol 22:357–65. doi:10.1016/j.tree.2007.04.003

[4] IPCC 2007. Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, RK Pachauri and A Reisinger (eds.)]. IPCC, Geneva, Switzerland, 104 pp.

[5] ME Visser and C Both. 2005. Shifts in phenology due to global climate change: the need for a yardstick. P Roy Soc B-Biol Sci 272: 2561–69. doi:10.1098/rspb.2005.3356

[6] C Korner and D Basler. 2010. Phenology under global warming. Science. 327:1461–62. doi:10.1126/science.1186473

[7] NE Miller-Struttmann, JC Geib, JD Franklin, PG Kevan, RM Holdo, D Ebert-May, et al. 2015. Functional mismatch in a bumble bee pollination mutualism under climate change. Science 349:1541-44. doi:10.1126/science.aab0868

[8] I Chuine. 2010. Why does phenology drive species distribution? Philos T Roy Soc B 365:3149–60. doi:10.1098/rstb.2010.0142

[9] AJ Miller-Rushing, TT Hoye, DW Inouye and E Post. 2010. The effects of phenological mismatches on demography. Philos T Roy Soc B 365:3177–86. doi:10.1098/rstb.2010.0148

[10] AD Richardson, TA Black, P Ciais, N Delbart, MA Friedl, N Gobron, et al. 2010. Influence of spring and autumn phenological transitions on forest ecosystem productivity. Philos T Roy Soc B 365:3227–46. doi:10.1098/rstb.2010.0102

[11] J Memmott, PG Craze, NM Waser and MV Price. 2007. Global warming and the disruption of plant–pollinator interactions. Ecol Lett 10:710–17. doi:10.1111/j.1461-0248.2007.01061.x

[12] GR Walther. 2010. Community and ecosystem responses to recent climate change. Philos T Roy Soc B 365:2019–24. doi:10.1098/rstb.2010.0021

[13] B Young, E Byers, K Gravuer, G Hammerson and A Redder. 2010. Guidelines for using the NatureServe Climate Change Vulnerability Index. NatureServe, Arlington.

[14] CAF Enquist, JL Kellermann, KL Gerst and AJ Miller-Rushing. 2014. Phenology research for natural resource management in the United States. Int J Biometeorol 58:579–89. doi:10.1007/s00484-013-0772-6

[15] SJ Mazer, SE Travers, BI Cook, TJ Davies, K Bolmgren, NJB Kraft, et al. 2013. Flowering date of taxonomic families predicts phenological sensitivity to temperature: implications for forecasting the effects of climate change on unstudied taxa. Am J Bot 100:1381–97. doi:10.3732/ajb.1200455

[16] EM Wolkovich, J Regetz and MI O'Connor. 2012. Advances in global change research require open science by individual researchers. Glob Change Biol 18:2102–10. doi:10.1111/j.1365-2486.2012.02693.x

[17] PRISM Climate Group. 2016. Oregon State University. URL: http://www.prism.oregonstate.edu/ (accessed 06/2016).

[18] MD Schwartz and BC Reed. 1999. Surface phenology and satellite sensor-derived onset of greenness: an initial comparison. Int J Remote Sens 20:3451–57. doi:10.1080/014311699211499

[19] L Cooper, RL Walls, J Elser, MA Gandolfo, DW Stevenson, B Smith, et al. 2013. The Plant Ontology as a tool for comparative plant anatomy and genomic analyses. Plant Cell Physiol 54:e1. doi:10.1093/pcp/pcs163

[20] RL Walls, B Athreya, L Cooper, J Elser, MA Gandolfo, P Jaiswal, et al. 2012. Ontologies as integrative tools for plant science. Am J Bot 99:1263–75. doi:10.3732/ajb.1200222

[21] P Jaiswal. 2011. Gramene Database: A hub for comparative plant genomics. In Plant Reverse Genetics, edited by A Pereira, 678:247–75. Totowa, NJ: Humana Press.

[22] PL Buttigieg, N Morrison, B Smith, CJ Mungall, SE Lewis and ENVO Consortium. 2013. The environment ontology: contextualising biological and biomedical entities. J Biomedical Semantics, 4:43. doi:10.1186/2041-1480-4-43

[23] RL Walls, J Deck, R Guralnick, S Baskauf, R Beaman, S Blum, et al. 2014. Semantics in support of biodiversity knowledge discovery: An introduction to the Biological Collections Ontology and related ontologies. PLoS ONE 9:e89606. doi:10.1371/journal.pone.0089606

[24] AD Hopkins. 1938. Bioclimatics: a science of life and climate relations. USDA Miscellaneous Publications 280:1–188.

[25] J Madin, S Bowers, M Schildhauer, S Krivov, D Pennington and F Villa. 2007. An ontology for describing and synthesizing ecological observation data. Ecol Inform 2:279–96. doi:10.1016/j.ecoinf.2007.05.004