

USA NATIONAL PHENOLOGY NETWORK - DATA MANAGEMENT PLAN

Project Title: USA National Phenology Network

Lead PI: At USGS: Jake Weltzin and at University of Arizona: Theresa Crimmins

Data Point(s) of Contact (if different):

- Kathy Gerst: Overall lead for data product development at USA-NPN.
- Lee Marsh: Technical lead for data and software development, lead for gridded product metadata.
- Jeff Switzer: Programmer, and database support, particularly for gridded products.
- Ellen Denny: Scientific Data Manager, Lead for in-situ protocol development.
- Theresa Crimmins: Lead for data management compliance, including updates of Data Management Plan
- Alyssa Rosemartin: Support for data product and tool development, and compliance.

Plans for Update: Each year, as part of annual planning process, looking at future development, and updating past products.

Plan for DOIs: We created DOIs by landing page for our dynamic datasets, and encourage users to follow the subsetting/date accessed citation [guidelines](#).

Project Description:

The USA National Phenology Network serves science and society by promoting broad understanding of plant and animal phenology and its relationship with environmental change. The Network is a consortium of individuals and organizations that collect, share, and use phenology data, models, and related information.

Data Management Resources: The estimated annual cost in salary and contracts for data management is \$315,000, which includes maintaining server software infrastructure (costs for physical server management, power and conditioning are covered free of charge through our partnership with University of Arizona), ingesting existing collections, programming for data input applications, data processing, output and visualization, quality control and metadata development, engaging with stakeholders around needed data products and project management. This figure does not include outreach staff time spent supporting observers in data collection.

DATA INPUTS – EXISTING COLLECTIONS

Existing collections include data that will be used for the purposes of creating the final data products and/or project deliverables. The collection(s) have already been obtained and have not been collected during the course of the project. Examples: PRISM data, USGS water data, remote sensing, etc.

1	PRISM Daily Temperature Maximum and Minimum Values
Description:	Gridded, daily temperature maximum and minimum across the continental U.S. at a 4km resolution, 1981 to present, used to generate the gridded product suite.
Source:	PRISM Climate Group, Northwest Alliance for Computational Science & Engineering http://prism.oregonstate.edu

Restrictions:	Data may be freely reproduced and distributed for non-commercial purposes. When referring to the data, the source must be clearly and prominently stated and include, at a minimum, the name, URL, and the date of data creation. For example: PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu , created 4 Feb 2004.
Format:	Zipped .BIL files
Fees:	No fees apply.
Quality Checks:	Data used as provided by source.
Data Processing & Scientific Workflows:	Scientific workflow described under Data Products below.
Backup & Storage:	Data are catalogued in a PostGIS database hosted by the USA-NPN. The entirety of the server's file system is backed up weekly to Amazon Cloud Storage as a VM image. Source .BIL files are also maintained alongside PostGIS database for posterity.
Volume Estimate:	50GB
Citation:	PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu , accessed, September 2015.

2 RTMA/URMA Real-Time, Hourly Temperature Maximums and Minimums	
Description:	Gridded, hourly temperature maximums and minimums across the continental U.S. at a 2.5km resolution, 2015-present, and 6 day forecasts, used to generate the gridded product suite.
Source:	NOAA, National Centers for Environmental Prediction (NCEP) RTMA Documentation: http://nomads.ncdc.noaa.gov/docs/rtma-product-description-document-pdd-20030108-final.pdf
Restrictions:	Unrestricted.
Format:	.BIN Files
Fees:	No fees apply.
Quality Checks:	Data used as provided by source.
Data Processing & Scientific Workflows:	Scientific workflow described under Data Products below.
Backup & Storage:	Data are catalogued in a PostGIS database hosted by the USA-NPN. The entirety of the server's file system is backed up weekly to Amazon Cloud Storage as a VM image.
Volume Estimate:	120GB

Citation:	<p>De Pondeva, Manuel SFV, Geoffrey S. Manikin, Geoff DiMego, Stanley G. Benjamin, David F. Parrish, R. James Purser, Wan-Shu Wu et al. "The real-time mesoscale analysis at NOAA's national centers for environmental prediction: Current status and development." <i>Weather and Forecasting</i> 26, no. 5 (2011): 593-612.</p> <p>NOAA National Operational Model Archive & Distribution System, Unrestricted Meso-scale Analysis, accessed September 2015.</p> <p>Both datasets available at: http://www.nco.ncep.noaa.gov/pmb/products/rtma/</p> <p>DOIs are not available for these resources.</p>
-----------	--

3	DAYMET Gridded Climate Variables
Description:	Gridded daily minimum and maximum temperatures, precipitation and daylength.
Source:	Oak Ridge National Laboratory, Distributed Active Archive Center http://daac.ornl.gov
Restrictions:	Unrestricted.
Format:	TXT
Fees:	No fees apply.
Quality Checks:	Data used as provided by source.
Data Processing & Scientific Workflows:	Data is requested and cached on an as-needed basis. Incoming data are used to calculate a series of other aggregated values including: seasonal average minimum and maximum temperatures, seasonal average precipitation, daily accumulated precipitation, and daily, base 0C, accumulated growing degree days, represented in both Celsius and Fahrenheit.
Backup & Storage:	Data are catalogued in a MySQL database and hosted by the USA-NPN. The entirety of the server's file system is backed up weekly to Amazon Cloud Storage as a VM image.
Volume Estimate:	In raw text format, current data is ~ 2.5 GB, growing by ~ 500MB/year.
Citation:	Thornton, P.E., M.M. Thornton, B.W. Mayer, N. Wilhelm, Y. Wei, R. Devarakonda, and R.B. Cook. 2014. Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 2. Data set. Available on-line [http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA. Date accessed: 2015/06/01-Present. Temporal range: 2008/01/01-present. Spatial range: N=52.00, S=14.53, E=52.95, W=131.10. http://dx.doi.org/10.3334/ORNLDAAC/1219 .

4	Lilac and Honeysuckle Data 1956-2008
----------	---

Description:	This dataset represents the legacy portion of the lilac and honeysuckle phenology data collected by the Western and Eastern Networks beginning in the 1950's. It was digitized and curated by Dr. Mark D. Schwartz at the University of Wisconsin until the founding of the USA-NPN in 2008. The dataset was recently combined with analogous data from the modern program, additional quality control measures were included, and it was published in Nature Scientific data (described as a data product, below.)
Source:	Mark D. Schwartz
Restrictions:	Unrestricted.
Format:	CSV
Fees:	No fees apply.
Quality Checks:	As described in Rosemartin et al 2015 for the data product.
Data Processing & Scientific Workflows:	Changes to data structure were made to conform to the National Phenology Database (CSV files split into tables for species, sites, and observation data reformatted from Date of First Leaf, to First Leaf: 0 or 1, to facilitate storage alongside status data).
Backup & Storage:	Server VM images are backed up weekly to Amazon Cloud Storage.
Volume Estimate:	3 MB
Citation:	Schwartz, M.D. and J.M. Caprio, 2003, North American First Leaf and First Bloom Lilac Phenology Data, IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series # 2003-078. NOAA/NGDC Paleoclimatology Program, Boulder CO, USA.

DATA INPUTS – NEW COLLECTIONS

Data that do not currently exist and will be collected or generated during the course of the project for the purposes of creating the final data products and/or project deliverables, for example, a new field data collection. New data collections must be delivered as a project deliverable at project completion and do not need to be added as a data product in the DMP.

1	Plant and Animal Phenology Data for the United States
Description:	<p>Dataset currently consists of 8 million phenology observation records on 1,137 species of plants and animals (as of October 2016), collected by participants in <i>Nature's Notebook</i> in accordance with our Terms of Use (to which participants agree upon registration), since 2009, across the United States. Data collection is ongoing. Updated summary statistics available at https://www.usanpn.org/data/dashboard and all data is available 48 hours after submission.</p> <p>A small number of these records are directly imported into the National Phenology Database from partner efforts which use exact or approximate USA-NPN protocols (e.g. New York Botanical Garden 2009-13: 42,130 records and Arbor Day Foundation, 2010-2012: 2,301 records). This dataset also includes the legacy lilac and honeysuckle observations and ongoing collection of these data.</p>

1	Plant and Animal Phenology Data for the United States
Exclusive Use:	No limitation on data use is made; Data Use Policy .
Restrictions:	Data are not restricted; Data Use Policy .
Format:	Data are stored in a MySQL database and made available in CSV format, and by web service as JSON or XML.
Protocols:	Standardized protocols were developed, vetted and documented in: Denny, E. G. et al. Standardized phenology monitoring methods to track plant and animal activity for science and resource management applications. International Journal of Biometerology (2014).
Quality Checks:	Many quality assurance and quality control measures are in place, and documented in a technical information sheet: https://www.usanpn.org/files/shared/files/QA-QC_tech_info_sheet_v2.4.pdf
Data Processing & Scientific Workflows:	Data processing described under data products, below.
Metadata:	FGDC metadata
Volume Estimate:	In text format, the entirety of the raw, underlying data is ~2.4GB in size, as of March 2016, accumulating at a rate of ~25% each year.
Backup & Storage:	Server VM images are backed up weekly to Amazon Cloud Storage.
Repository for Data:	Metadata for this data is or will be stored in the USGS Science Data Catalog, Ecosystems.data.gov and Data.gov, ScienceBase, Dryad and KNB.
Citation:	USA National Phenology Network. Year of download date. Plant and Animal Phenology Data for the United States, Date range of data used. USA-NPN, Tucson, Arizona, USA. Data set accessed YYYY-MM-DD at http://www.usanpn.org/results/data .
Digital Object Identifier (DOI)/Link:	http://dx.doi.org/10.5066/F78S4N1V
Lifespan of Data	50+ years

MODELS

Describe the function and methodology used for any models that are part of the project. Any code developed to execute the model (if any was/will be developed by the project) should be described in the custom software/code section.

1	Extended Spring Indices
Description	The Extended Spring Indices are models that predict the “start of spring” (timing of leaf out or bloom in lilacs and honeysuckles) at a particular location, based on antecedent weather conditions (Schwartz 1997 , Schwartz et al. 2006 , Schwartz et al. 2013 , Ault et al 2015).
Model Version	SI-x, Extended Spring Indices

1	Extended Spring Indices
Source/Link:	The model is described and implemented in MatLab via publication below. We have also made a version in Python available at: https://github.com/usa-npn/gridded_models Ault, Toby R., Raul Zurita-Milla, and Mark D. Schwartz. "A Matlab© toolbox for calculating spring indices from daily meteorological data." Computers & Geosciences 83 (2015): 46-53.
Model Input(s)	The model is calculated from daily minimum and maximum temperatures, which are used to generate growing degree hours and number of high energy synoptic events.
Model Output(s)	Estimated date leaf and bloom, an average of cloned and common lilacs and two species of honeysuckles.
Calibration Details	Calibration and validation approaches (comparisons to climate modes, and native species and crops) are described in the literature cited in the description above.

CUSTOM SOFTWARE/CODE AND WEB TOOLS

Describe any custom software or code used as part of this project. If a web tool (e.g., visualization, decision support, etc.), is a project deliverable that should be included in this section.

In addition to the custom tools described below we leverage Tableau Desktop and Public for data exploration and partner dashboard, as well as four Google services: Maps, Reverse Geocoder, Time Zone Service and Elevation Service.

Maintenance and Support for Web Tools: The USA-NPN is a 30-year project, with relatively secure base support. Maintenance of the tools developed is planned across the tool's lifespan. Tools are upgraded or replaced as standards and best practices for web development change.

1	Web Services
Description:	A series of REST and SOAP based web services providing input/output access to the underlining USA-NPN database. Useful for external collaborators to interface with NPN data for custom built applications and data access.
Source/Link:	Documentation for use: https://docs.google.com/document/d/1yNjupricKOAXn6tY1sI7-EwkcfwdGUZ7lxYv7fcPjO8/edit SVN Repository: https://www-dev.usanpn.org/svn/npn_ws/trunk/npn_portal Access available upon request
Restrictions:	Unrestricted.
Languages:	PHP
Environment:	Linux

2 Nature's Notebook Web Interface	
Description:	<p>Web interface which allows <i>Nature's Notebook</i> participants to register sites, plants and animals to observe and submit phenology observations via on screen controls or a spreadsheet template. Participation in shared sites, a phenology calendar visualization, and digital merit badges are also available through this interface.</p> <p>Accessible by creating an account, begin here: https://www.usanpn.org/nn/become-observer</p>
Source/Link:	Access available upon request for: https://www-dev.usanpn.org/svn/mynpnapp/
Restrictions:	Unrestricted.
Languages:	Java/JavaScript
Environment:	Linux

3 Nature's Notebook iPhone App	
Description:	<p>An iPhone app is optimized to allow <i>Nature's Notebook</i> participants to submit phenology observations on registered plants and animals (while it is possible to add new sites, plants and animals on the phone).</p> <p>https://itunes.apple.com/us/app/natures-notebook/id508465801?mt=8</p>
Source/Link:	Access available upon request for: https://www-dev.usanpn.org/svn/NaturesNotebook-Mobile/iOS/NaturesNotebook
Restrictions:	Unrestricted.
Languages:	Objective C
Environment:	iOS 6+

4 Nature's Notebook Android App	
Description:	<p>Our Android app is optimized to allow <i>Nature's Notebook</i> participants to submit phenology observations on registered plants and animals (while it is possible to add new sites, plants and animals on the phone).</p> <p>https://play.google.com/store/apps/details?id=org.usanpn.android.naturesnotebook&hl=en</p>
Source/Link:	Access available upon request for: https://www-dev.usanpn.org/svn/NaturesNotebook-Mobile/Android/NaturesNotebook
Restrictions:	Unrestricted.

4 Nature's Notebook Android App	
Languages:	Java
Environment:	Android Platform 2.3+

5 AT Seasons Android App	
Description:	An Android app, based on the primary <i>Nature's Notebook</i> app, supports observing with the AT Seasons project (providing custom sites, species, and animal vocalizations). https://play.google.com/store/apps/details?id=org.usanpn.android.atseasons&hl=en
Source/Link:	Access available upon request for: https://www-dev.usanpn.org/svn/NaturesNotebook-Mobile/Android/NaturesNotebook/branches/atmobile
Restrictions:	Unrestricted.
Languages:	Java
Environment:	Android Platform 2.3+

6 Visualization Tool	
Description:	A web-based visualization tool supports exploration of <i>in situ</i> phenology data via map interfaces, a calendar and a plot of phenophase onset with predictor variables from Daymet. Additionally, map layers of gridded accumulated growing degree days and Extended Spring Indices for the Continental United States are available. https://www.usanpn.org/data/visualizations
Source/Link:	https://github.com/npnlee85/npn-viz-tool
Restrictions:	Unrestricted.
Languages:	HTML5/Javascript
Environment:	Linux

7 Phenology Observation Portal	
Description:	A web interface which supports the download of customized <i>in situ</i> raw and summarized phenology data, along with Daymet temperature, precipitation and daylength data for the location at which the observations were made. Users can customize the species, locations and phenophases they need, select additional information (e.g., details about sites, quality control flags), and access metadata.

7	Phenology Observation Portal
	https://www.usanpn.org/results/data
Source/Link:	Access available upon request for: https://www-dev.usanpn.org/svn/ddt/trunk
Restrictions:	Unrestricted.
Languages:	PHP
Environment:	Linux

8	Bulk Uploader Tool
Description:	This tool allows users who have set up sites and species in Nature’s Notebook to download a datasheet in Excel format. They may then upload it with their observations. It was developed to support groups with limited internet connection and higher numbers of plants and animals.
Source/Link:	Access available upon request for: https://www-dev.usanpn.org/svn/bulk/trunk
Restrictions:	Unrestricted.
Languages:	Java/Javascript
Environment:	Linux

9	Geoserver Request Builder
Description:	This tool enables access and download Accumulated Growing Degree Day or Extended Spring Index raster phenology map products and images. https://www.usanpn.org/geoserver-request-builder
Source/Link:	https://github.com/usa-npn/geoserver-request-builder
Restrictions:	Unrestricted.
Languages:	Javascript/angular2
Environment:	Linux

DATA PRODUCTS (E.G., DELIVERABLES)

Project deliverables and data products that were developed as a result of the project’s funding are described in Table 3 (pg 15) the [USA-NPN Data Product Development Framework and Catalog](#). Note that all data products are freely and readily available and thus are unrestricted, and for non-exclusive use.