USGS
science for a changing world

Development and Release of Phenological Data Products—A Case Study in Compliance with Federal Open-Data Policy

By Alyssa Rosemartin, Madison L. Langseth, Theresa M. Crimmins, and Jake F. Weltzin

Open-File Report 2018–1007

U.S. Department of the Interior
RYAN K. ZINKE, Secretary

U.S. Geological Survey
William H. Werkheiser, Deputy Director
exercising the authority of the Director

U.S. Geological Survey, Reston, Virginia: 2018

Contents

Figures

Tables

Abbreviations

CDI	Community for Data Integration
DMP	Data management plan
DOI	Digital Object Identifier
DPC	Data Product Catalog
FGDC	Federal Geospatial Data Committee
FSP	Fundamental Science Practice
IPDS	Information Product Data System
OFR	Open-File Report
OGC	Open Geospatial Consortium
OMB	Office of Management and Budget
OSTP	Office of Science and Technology Policy
PRISM	Parameter-elevation Relationships on Independent Slopes Model
SI-x	Extended Spring Indices
SPN	Science Publishing Network
TDR	Trusted Digital Repository
URL	Uniform Resource Locator
USA-NPN	USA National Phenology Network
USGS	U.S. Geological Survey
WCS	Web Coverage Service
WMS	Web Map Service

Development and Release of Phenological Data Products—A Case Study in Compliance with Federal Open Data Policy

By Alyssa Rosemartin[1,2], Madison L. Langseth[3], Theresa M. Crimmins[1,2], and Jake F. Weltzin[2,3]

Abstract

In Autumn 2015, USA National Phenology Network (USA-NPN) staff implemented new U.S. Geological Survey (USGS) data-management policies intended to ensure that the results of Federally funded research are made available to the public. The effort aimed both to improve USA-NPN data releases and to provide a model for similar programs within the USGS. This report provides an overview of the steps taken to ensure compliance, following the USGS Science Data Lifecycle, and provides lessons learned about the data-release process for USGS program leaders and data managers.

Introduction

As Federal scientists respond to the call for open science in the digital age, new opportunities and challenges are emerging. Open data enables research on new temporal and spatial scales and greatly facilitates replication of results. However, full documentation, peer review, and public release of datasets is costly and not yet common in government, which limits data reuse and the integration of complementary datasets (Sullivan and others, 2016).

The U.S. Geological Survey (USGS) is a leading provider of valuable geospatial data, although historically not all of these data have been made available (that is, discoverable, accessible, and usable) to other programs within or outside of the USGS, limiting the potential to integrate complementary datasets, replicate results, or develop new applications for existing data.

Recent open-data policies from the Office of Science and Technology Policy (OSTP) and the Office of Management and Budget (OMB) require that Federally funded information products (for example, publications) must be made freely available to the public along with the underlying data on which the conclusions are based (Holdren, 2013). In response, the USGS has developed Instructional Memoranda describing requirements for information management that were later codified in the USGS Survey Manual (U.S. Geological Survey, 2017a), the Public Access Plan (U.S. Geological Survey, 2016), and USGS Fundamental Science Practices (U.S. Geological Survey, 2017b). Collectively these documents describe how USGS data should be managed, reviewed, and shared with the public.

---

[1]University of Arizona,
[2]USA National Phenology Network
[3]U.S. Geological Survey

Purpose and Scope

The new USGS policies seek to improve discovery, accessibility (including machine readability), integration, and use of data for novel applications in support of the broader USGS mission. However, the actual effort required to reach compliance across scales from projects to the USGS as a whole was unknown. In this report we describe how the USA National Phenology Network (USA-NPN, www.usanpn.org) exercised and tested new USGS data-management policies in the production and delivery of national-scale biodiversity data products, including, from a first-person perspective, the process that we followed and the lessons that we learned in the production of policy-compliant data products. We describe (1) our objectives in creating and implementing data products, (2) the formulation of a data-management plan, and (3) the establishment of workflows to ensure that data releases are compliant with current data-management policy. We offer an "on the ground" perspective that focuses on lessons learned about the data-release process. Specific details concerning USGS data-management policies and software tools are presented elsewhere (for example, Chase and others, 2016; Kirk, 2016,) and are outside the scope of this report. We use the USGS Science Data Lifecycle (Faundeen and others, 2013), along with the associated chapters in the USGS Survey Manual (U.S. Geological Survey, 2017a), as a framework for our case study (table 1).

We hope this report will provide inspiration and guidance to other projects and programs as they produce and deliver similar types of data and data products. This purpose is particularly relevant in light of a recent USGS emphasis on creating an "integrated predictive scientific capacity" to produce "powerful new products and services"—including "vulnerability detection and assessment, prediction and forecasting, early warning, and decision support at the scale of decisions"—across scientific boundaries and disciplines (Werkheiser, 2017). Realization of the promise of integrated and predictive science requires a commitment to sound and unified data-management practices.

Table 1. U.S. Geological Survey (USGS) policy requirements based on the USGS Science Data Lifecycle stages (see fig. 2).

Science Data Lifecycle Stage	USGS Policy Requirements	USGS Survey Manual Reference
Plan	Data management plans must be created before initiating a project and maintained throughout the project lifecycle	Chapter 502.6
Acquire	Methods and techniques for acquiring research data must be planned and documented to ensure that USGS scientific findings are verifiable. Agreements with collaborators/partners and vendors must be created to clarify data ownership responsibilities	Chapter 502.6
Process and Analyze	Processing methods and data analysis techniques must be documented	Chapter 502.6
Describe*	FGDC-compliant metadata must accompany all released data	Chapter 502.7
Manage Quality*	Quality assurance checks must be made throughout the science data lifecycle	Chapter 502.6
Preserve	Data must be preserved in accordance with Federal archiving policy	Chapter 502.9
Publish/Share	Data supporting scholarly conclusions must be reviewed, approved, and released free of charge to the public	Chapter 502.8
Backup and Secure*	Backup copies of data must be maintained to allow recovery from loss	Chapter 502.6

*These stages are treated in the USGS Science Data Lifecycle as continual processes, alongside the six linear steps in the cycle. In this table, they are inserted in the order where they fit in our process of compiling and delivering data for this case study.

USA-NPN: Data Products for Science and Decisionmaking

The mission of the USA-NPN—which was established by USGS in 2007 and has since then operated in collaboration with University of Arizona—is to collect, store, and share phenological data, data products, and information in order to advance science and support natural-resource decisionmaking. The USA-NPN provides an excellent case study for examining issues related to data release, given its unique status as a Government-academic partnership that because of major Federal funding is required to conform to USGS data-management policies.

Data management and delivery at the USA-NPN is accomplished by six staff members (3.5 full-time equivalents) who coordinate data-product delivery, manage databases, write code, test new data products, write documentation, communicate with stakeholders, and ensure compliance with internal and external policies. The USA-NPN maintains a database containing more than 12 million records of organismal plant and animal phenological data from across the United States spanning the period from 1956 to the present. These data—collected according to standardized protocols (Denny and others, 2014) by various contributors, including scientists, resource managers, and citizen-scientists—are contributed to the USA-NPN by way of an online program and user interface called Nature's Notebook (www.naturesnotebook.org). In turn, these data are used to create various value-added data products, or synthetic datasets with defined applications, in support of scientific discovery and decisionmaking at local to national scales (USA National Phenology Network, 2016b).

Here, we outline our planning process and the workflows we established for the production and release of dynamic gridded data products, particularly historical, contemporary, and short-term forecasted predictions of leafout and flowering for several plant species (Crimmins and others, 2017). These products are operational (that is, continuously produced, improved, and documented) and designed to support phenological research and natural-resource-management decisionmaking under changing environments (Ault and others, 2015; McCabe and others, 2013; Monahan and others, 2016). The USA-NPN publishes a total of 10 such products, including 56 subproducts. For the purpose of this report, we track the creation of a single data product entitled "Historical Annual Spring Indices", which is described in more detail as part of the case study below.

The Data-Management Planning Process

Because planning is a foundational step in the USGS Science Data Lifecycle (Faundeen and others, 2013) and USGS policy requires that data-management plans be created and maintained throughout the project lifecycle (table 1), in this section we describe our experience with data-management planning as a context for the case study described in the next section.

We take a broad approach to data management planning, including (1) a Data Management Plan (DMP) that contains an inventory of all source data, software, and models, referencing (2) our Data Product Catalog (DPC; USA-National Phenology Network, 2016b) for an inventory of our data products; and (3) a Data Lifecycle Best Practices document that details knowledge of data-management processes. Together, the DMP and DPC enable us to review all data-related components and ensure compliance. The DPC is appropriate for our stakeholders to evaluate available and forthcoming data products relevant to their interests. The Data Lifecycle Best Practices document serves our current and future staff in managing data effectively.

In January 2016, as part of a proposal to the Community for Data Integration (CDI) competitive grants program, we created a DMP for our gridded data products. Later that year, we created a separate DMP to include all USA-NPN data products and software tools (USA-National Phenology Network 2016a), which brought us into compliance with USGS policy (Survey Manual, chap. 502.6). The USA-

NPN DPC outlines plans for the production and release of observational, modeled, and integrated data products. This DPC is a guiding document to focus our efforts and to inform stakeholders about our plans for the release of various data products. We maintain an updated version of the DPC on our website (https://usanpn.org/pubs/reports#USA-NPN_Technical_Series).

We also created an internal document entitled "Data Lifecycle Best Practices" that details our current thinking about compliance, issues, and decisions related to data management and policy compliance. Whereas the DMP and DPC describe data products and software tools, the Data Lifecycle Best Practices document centers on process and orients staff to the required steps in creating new data products.

Throughout the data-management planning process, we learned that a "data management plan" can consist of any number of documents or resources. The most important consideration for formulating a data-management plan is ensuring that it is aligned with a project team's structure and workflows. Our planning process resulted in multiple documents rather than a single document because we have multiple audiences and levels of detail as required for different purposes. All data-management documents are considered living and are updated once or twice annually as plans evolve. To meet the requirements for planning in the USGS Survey Manual (chap. 502.6), only a DMP is needed. We developed the DPC and the Data Lifecycle Best Practices document to best meet the needs of external data users and staff, respectively.

We found the data-management planning process a bit cumbersome at first. Because our team is small, we have ready access to the software codebases and an awareness of the source data underlying our various products. Furthermore, we are experienced data-management professionals who know what is involved in initiating a new project. Therefore, we did not spend much time writing down readily available information. Our Information Technology Coordinator found the CDI DMP template's tabular format with standard categories (for example, quality checks, restrictions) awkward because the scale of the categories did not align with our data. In some cases, the category did not apply; and in others, a more comprehensive response was needed. As a potential alternative, our Information Technology Coordinator proposed writing a paragraph for each product to include all applicable categories. Our Data Product Coordinator noted that the planning process was somewhat complicated by the number of documents describing data products (for example, the DPC and various DMPs); however, once we organized all our documentation, we realized unforeseen efficiencies. For example, when the final progress report on the CDI grant was due, we simply extracted the required details from our DMP.

Ultimately, we found that although our data-management planning documents seemed unnecessary at first, they were useful as communication tools and provided a framework for organizing and prioritizing compliance. For example, all staff members involved in the project could readily determine whether or not metadata were available for all products and could assess at a glance the parties responsible for each aspect of our data-management activities. USA-NPN staff who were not involved in the creation of the products found the documents useful for understanding the suite of new software tools and for finding the context for the various acronyms used in meetings. The staff also see value in the planning documents for handling instances of turnover; new staff will have important information about our infrastructure and processes. Although having a suite of related documents can be cumbersome, we are finding that each document is suited to its purpose and audience, and with the combined set we are able to communicate with our stakeholders and achieve compliance.

Data-management planning provides the broad context for USA-NPN data assets and their current status. The details of compliance with the next steps in the data lifecycle are illustrated by using a case study in the following section (see table 1; Faundeen and others, 2013).

Case Study – Historical Annual Spring Indices

The USA-NPN produces and delivers various climatologic and phenological gridded (raster) data products, including modeled leafout and flowering for several plant species, at a daily time step and 4-kilometer (km) resolution for the Continental United States and 1° resolution for Alaska (Crimmins and others, 2017). For the purpose of this report, we describe the production of a single data product, the Historical Annual Spring Indices (fig. 1), within the context of the data lifecycle (fig. 2).
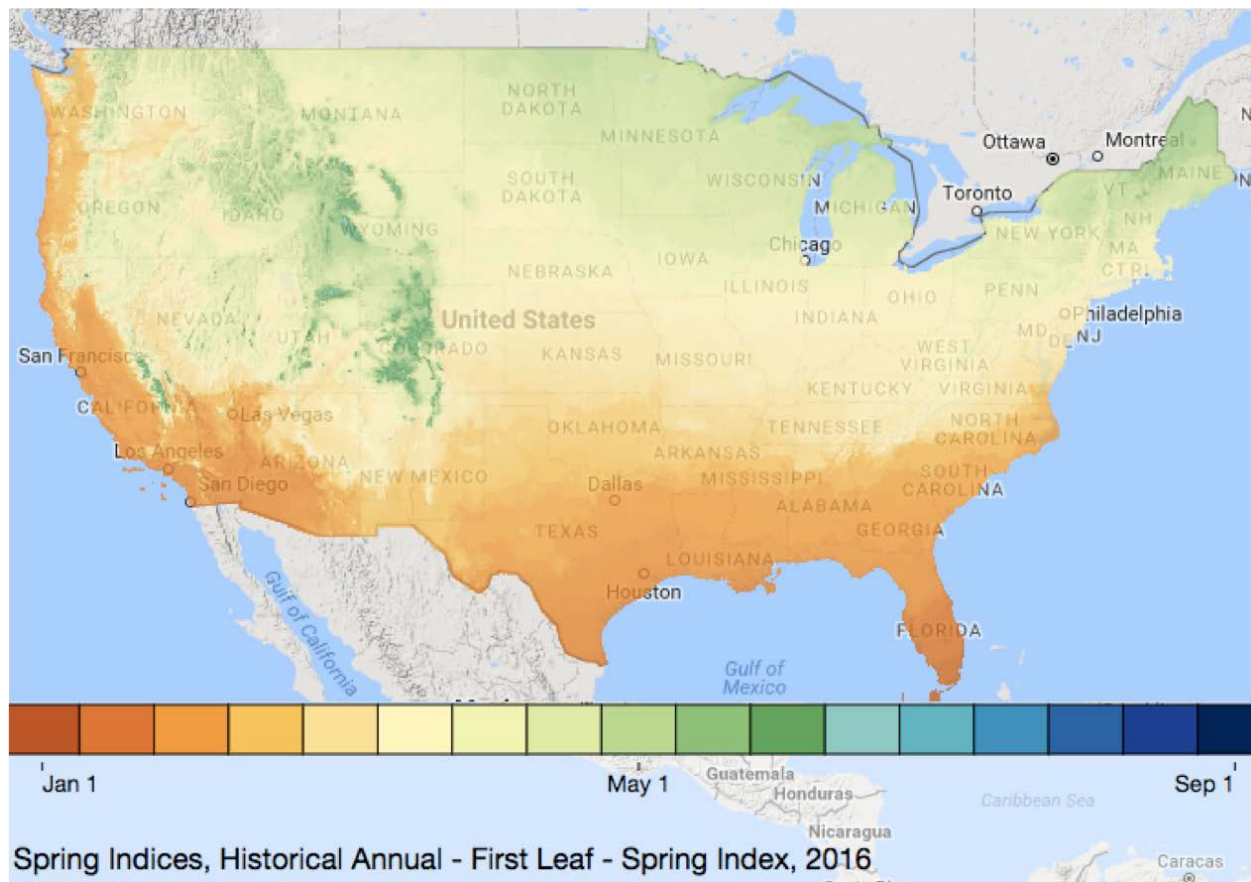
Figure 1. Map showing Historical Annual Spring Leaf Index for 2016, accessed using the USA-NPN Visualization Tool. Each 4-kilometer grid cell on the map is shaded to represent the predicted day of leafout, based on the Spring Leaf Index (Crimmins and others, 2017). During the current year, the map is updated daily, whereas the historical data product (shown here) shows final model output for the year.

Readily available data and information on phenologic patterns support resource managers in timing such actions as treating invasive species, operating visitor facilities, and scheduling seasonal events (for example, flower festivals, autumn leaf-viewing; Monahan and others, 2016). The Extended Spring Indices (SI-x) are models predicting the "start of spring" at a given locality (Schwartz and others, 2006, 2013) and are curated and run by the USA-NPN. The SI-x models were constructed from historical observations of the timing of first leaf and first bloom in cloned lilacs and honeysuckles, as well as from daily observations from nearby meteorologic stations, to determine the weather conditions (growing degree days and warm periods, generated from daily minimum and maximum temperatures)

5

that precede spring leafout and bloom for a wide range of plants (Schwartz, 1997; Schwartz and others, 2006, 2013). The Historical Annual Spring Indices consists of gridded layers of the SI-x leafout and bloom dates at 4-km resolution across the Contiguous United States, from 1981 to the previous year. (We update this product when the underlying gridded temperature data become available, after the close of each calendar year.) Additional details for all aspects of production and delivery of this and associated data products were provided by Crimmins and others (2017).
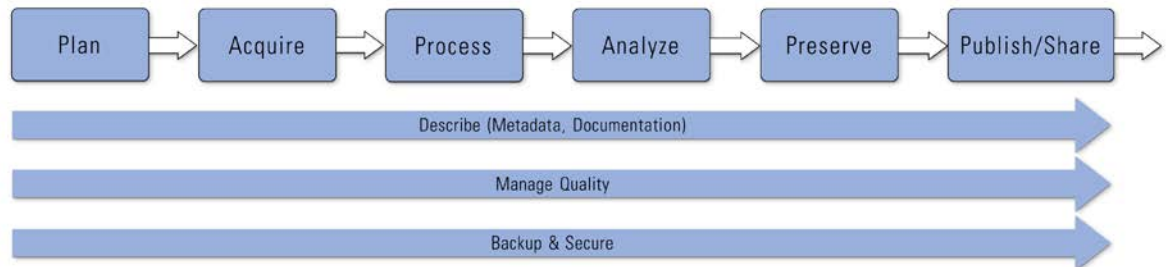

Figure 2.   Diagram of the U.S. Geological Survey Science Data Lifecycle (Faundeen and others, 2013).

Data-Product Development Following the Data Lifecycle

Acquire

The USGS requires that the methods and techniques for acquiring research data be planned and documented to ensure that USGS scientific findings are verifiable. Agreements with collaborators, partners, and vendors must be created to clarify data-ownership responsibilities (table 1). The data required for producing the Historical Annual Spring Indices are daily minimum and maximum temperature for the United States. We selected the Parameter-elevation Relationships on Independent Slopes Model (PRISM) as the data source, given the high-resolution gridded products available from 1981 to the present (PRISM Climate Group, Oregon State University, 2015). These datasets are freely available and no data-sharing agreements were required.

Process and Analyze

USGS policy requires that processing methods and data-analysis techniques be documented (table 1). We established a workflow, implemented in the Python programming language, which downloads minimum and maximum temperature data from the PRISM Climate group (PRISM Climate Group, Oregon State University, 2015) and generates SI-x values for the Historical Annual Spring Indices. The raster data generated by these scripts are loaded in a PostGreSQL database with PostGIS support. Geoserver, an Open Geospatial Consortium (OGC)-compliant web application, is then used to access the database and present these data using Web Map Service (WMS) or Web Coverage Service (WCS) endpoints. The source codes for these tools are available on both GitHub (https://github.com/usa-npn/gridded_models) and Bitbucket (https://my.usgs.gov/bitbucket/projects/CDI/repos/phenology-maps-workflow/browse) repositories.

Describe

USGS policy requires that Federal Geospatial Data Committee (FGDC)-compliant metadata accompany all data releases (table 1). We documented the methods used to generate the gridded products in an Open-File Report (OFR) describing the suite of gridded products, including the Historical

Annual Spring Indices (Crimmins and others, 2017). The report details each data-product layer, with scientific justification, algorithms, technical specifications, and instructions for access, including workflow processing and quality-control measures. To aid in communication, we provided introductory information about the suite of products in a Technical Information Sheet (https://www.usanpn.org/files/reports/USA-NPN_AGDD-and-SiX.pdf) and created a landing page (https://www.usanpn.org/data/spring_indices) on the USA-NPN Website, which describes and provides examples of all gridded products.

We authored FGDC-compliant metadata for the layers by using the USGS Metadata Wizard (https://www.sciencebase.gov/metadatawizard) and reviewed the metadata in desktop text-editing programs, using GitHub to share and compile changes. We relied on an FGDC interactive graphic (https://www.fgdc.gov/csdgmgraphical/index.html) to better understand the FGDC standard. Writing FGDC metadata encouraged us to reexamine our information-management practices. For example, because the FGDC metadata standard calls for information on use constraints and attributions, we examined our terms of use (https://www.usanpn.org/terms), which governs products and information delivered via our website, and found no need for changes.

We created FGDC-compliant, layer-level metadata that can be used by researchers to autopopulate field names and other descriptive information in a GIS environment. Layer-level metadata files exist for historical annual leafout and bloom layers (as well as for other layers not detailed here; see Crimmins and others, 2017). To facilitate product discovery, we also created a metadata file that describes the full suite of SI-x products. This metadata file was submitted to the USGS Science Data Catalog, a comprehensive USGS metadata catalog. From the Science Data Catalog it is harvested by the Ecosystems Website (ecosystems.data.gov). The layer-level metadata files reference this suite-level file, and vice versa; however, the individual metadata files were not submitted to the Science Data Catalog, because they are not targeted at data discovery. Instead, the layer-level metadata files, which are most useful to data end users, are hosted on USA-NPN's Webserver, are referenced by Geoserver, and are readily available from a Geoserver Request Builder (http://usanpn.org/geoserver-request-builder/).

Proper documentation of a dataset includes a suggested citation. Several years ago, we settled on a citation format for dynamic datasets that supports human traceability (see Preserve section below for further discussion of machine-traceable identifiers). Following DataONE (2016) and Earth Science Information Partners (2016) guidelines, we request that researchers include dataset search parameters and date accessed in each citation (see USA-NPN Data Use Policy; https://usanpn.org/terms#DataUse). For gridded products, the citation request also includes a parameterized Uniform Resource Locator (URL) for the Geoserver Request Builder software tool, which opens with the filters set to replicate the download, enabling a researcher to both download the referenced dataset and explore related data. An example requested citation for the SI-x, Bloom Index for 2016 follows:

USA National Phenology Network. 2017. Spring Indices, Historical Annual (1981-Previous Year) - First Bloom - Spring Index, Year: 2016. Region: 49.9375°,-66.4791667° (UR), 24.0625°,-125.0208333° (LL). http://data.usanpn.org/geoserver-request-builder?service=wms&layer=si-x:average_bloom_prism&year=2016&format=image/png&projection=4269&width=1700&height=800&colors=1. USA-NPN, Tucson, Arizona, USA. Data set accessed 2017-8-2. http://doi.org/10.5066/F7XD0ZRK.

Unlike many Federal data-management groups, we opted to use institutional rather than individual authorship. Staff members are funded primarily to compile and deliver operational datasets and accordingly are not rewarded for dataset citations tied to individuals. Although datasets themselves

are not authored by individuals, staff do author publications on the creation, storage, and synthesis of USA-NPN products, allowing for recognition of individual contributions.

Perhaps more importantly, much of the value of the USA-NPN's observational datasets is created by thousands of citizen-scientist volunteers across the country. Because it is difficult to include these individuals as authors, we instead composed a standard acknowledgment for observers in the USA-NPN Data Attribution Policy (https://usanpn.org/terms#DataAttribution), which reads as follows:

> "Data were provided by the USA National Phenology Network and the many participants who contribute to its *Nature's Notebook* program."

The USGS requires all data releases to have associated FGDC-compliant metadata and a recommended citation with a Digital Object Identifier (DOI). Though not required, OFRs are a good option for describing data in addition to metadata records because they provide a more comprehensive and human-readable view of the documentation. Data managers may instead choose to author a peer-reviewed journal article, a USGS Techniques and Methods series report, or a Webpage to describe a data resource. In fact, current guidance encourages service-delivered data to have the descriptive text placed on or linked from the Webpage where the data are delivered.

Manage Quality

USGS policy requires quality-assurance checks to be made throughout the Science Data Lifecycle (table 1). As for all gridded data products, our products contain uncertainty or error arising from multiple sources, including error propagated forward from underlying climate data and from models used in the calculations (Crimmins and others, 2017). We manage quality by ensuring the validity of values as part of all data processing. In addition, comparisons of independent-weather-station and grid-cell values for all data products are either available now or planned in the near term (Crimmins and others, 2017). The quality of the SI-x model and underlying data are well described in the literature (Schwartz and others, 2013; Rosemartin and others, 2015). Ault and others (2015) described the quality and reliability of a gridded implementation of the SI-x.

The primary way in which the USGS ensures quality of scientific outputs is through Fundamental Science Practice (FSP) review. In fall 2016, we submitted the data, metadata, and OFR (Crimmins and others, 2017) as a package for FSP review, in compliance with the USGS Survey Manual (chap. 502.8). The USGS Science Publishing Network (SPN) formatted the OFR. We had originally planned to write a Techniques and Methods series report because the content of our report primarily describes data processing; we were instead advised to create an OFR because the content is expected to change as new layers are added to the suite. Additionally, we learned that the data and metadata must be tracked in Information Product Data System (IPDS) as a separate entity from the report, which may not always be ideal, because it separates a report describing data processing from the data itself. We added a note to each IPDS entry to cross-reference the products to one another.

Preserve

USGS policy requires that data be preserved in accordance with National Archives and Records Administration standards. USA-NPN data products are stored on secure servers at the University of Arizona, as described in the Back Up and Secure section below. We are formulating a preservation plan to address data storage at the end of the project's life.

Preserving Dynamic Data

One of the more difficult issues in the preservation of dynamic data is the creation of persistent identifiers (a unique and lasting reference to a dataset, publication, or other file). We explored a solution based on "snapshotting" dynamic databases that was suggested by the Research Data Alliance (Rauber and others, 2016). This approach would require dataset versioning (in other words storing all previous versions of the data) or maintaining a static copy of every dataset downloaded in order to function correctly, which we found to be too costly both programmatically and in terms of storage space. Instead, we opted to create a single DOI through the USGS DOI Tool for the landing page that describes the suite of Spring Index products and serves as a portal for all data access (https://doi.org/10.5066/F7XD0ZRK). Although using a single DOI enables preservation and persistent access to the data, we recognize that this approach limits the data's provenance, so we are seeking lower-cost solutions to this issue.

Publish/Share

USGS policy requires that data supporting scholarly conclusions be reviewed, approved, and available free of charge to the public (table 1). Once we were ready to share these data products, we requested informal review from two external reviewers. We had a strong interest in sharing the provisional data with partners and potential users in spring 2016 and sought guidance on official-release procedures. Because the products were time sensitive, but had yet to receive full FSP review and approval, the USGS Bureau Approving Official advised that it was appropriate to release the new products as "non-emergency provisional data." The pre-review disclaimer that is required in all metadata and in the Geoserver layer descriptions reads as follows:

> These data are preliminary or provisional and are subject to revision. They are being provided to meet the need for timely best science. The data have not received final approval by the U.S. Geological Survey (USGS) and are provided on the condition that neither the USGS nor the U.S. Government shall be held liable for any damages resulting from the authorized or unauthorized use of the data.

In spring 2016, we made the layers available directly from the Geoserver by way of OGC-compliant Web services, in compliance with the requirement to share data in widely used and nonproprietary formats, and facilitated access with Geoserver Request Builder (http://usanpn.org/geoserver-request-builder/), a software tool that creates a URL to download a layer based on a user's specifications through a user-friendly graphical user interface. At the same time, the layers were released through the USA-NPN Visualization Tool (https://www.usanpn.org/data/visualizations), together with USA-NPN observational data and phenologically relevant climate data.

Publishing Dynamic Data

Although a protocol for data versioning is included on the USGS FSP Website, the USA-NPN and similar efforts do not fit neatly within existing standards because our data products are updated daily. Because dynamic datasets are relatively new and unusual within the USGS, the USGS guidance on Documenting Revisions to Data Releases was modified in response to our description of the challenges of versioning for dynamic data, to clarify that versioning is not required for such datasets, when appropriate processes for quality control are in place and documented elsewhere (for example, Crimmins and others, 2017). The procedures for managing the quality of dynamic data were developed

by the FSP Advisory Committee; implementation of these procedures for this case study is described in the Manage Quality section above and Current Status section below.

Backup and Secure

USGS policy requires that backup copies of data be maintained to allow recovery from loss (table 1). All USGS data products are stored on servers at the University of Arizona. USA-NPN data are currently stored on high-quality servers, in a climate-controlled room, with appropriate security and redundancy procedures through Amazon Web Services, including weekly backups of server images, as described by the University of Arizona Information Technology Services. We are currently constructing a system to regularly export USA-NPN data to ScienceBase, so that our data will be stored in Trusted Digital Repositories (TDR), per the USGS Survey Manual (chap. 502.9) preservation requirements for digital scientific data. The data will be accessible only by administrators of the USA-NPN collection on ScienceBase to prevent multiple, potentially outdated datasets from being used.

Current Status

Once the data release and associated OFR were approved through FSP review and the OFR was released (Crimmins and others, 2017), the data became "approved" rather than "provisional." At this point, the staff updated the disclaimers in the metadata on the products themselves, the visualization tool, and metadata. The required "approved" disclaimer reads as follows:

Unless otherwise stated, all data, metadata and related materials are considered to satisfy the quality standards relative to the purpose for which the data were collected. Although these data and associated metadata have been reviewed for accuracy and completeness and approved for release by the U.S. Geological Survey (USGS), no warranty expressed or implied is made regarding the display or utility of the data on any other system or for general or scientific purposes, nor shall the act of distribution constitute any such warranty. (Fundamental Science Practices, 2016)

With further clarification from the Bureau Approving Official, we determined that our Information Technology Coordinator could serve as Data Quality Manager. The Information Technology Coordinator's review is the only review required when a new year of data is added to an existing approved data product, such as the Historical Annual Spring Indices.

Conclusions and Recommendations

Our experience in bringing a suite of new, real-time products into compliance with USGS data-management policy resulted in the following recommendations for other groups going through a similar process. First, it is important to clarify how the various USGS data-management policies apply to the specific data resources being released. For example, at the broadest level, does Federal funding for the USA-NPN mean that an organizational DMP is required? The question of whether policy applies is also relevant at a more granular level, for example, should data be considered "emergency" or "non-emergency"?

Second, it is helpful to prioritize the different aspects of compliance and to define staff roles in achieving compliance. In a dynamic work environment, individual roles shift, policies evolve in response to new cases, and plans for data-product development respond to stakeholder interests. It is important to keep in mind that a renegotiation of priorities and roles will likely be ongoing and that regular communication among project members is critical.

Several challenges were associated with USA-NPN's multiorganizational nature. Because many of our staff members are not USGS employees, we sometimes had difficulty accessing USGS software tools (for example, for creating DOIs and metadata) and USGS internal Websites. To surmount these barriers, the USGS member of our team assigned the DOIs, whereas university staff used the Metadata Wizard (which is publicly accessible) to create and edit metadata.

Although compliance with new policies may seem overwhelming at first, taking it "one step at a time" was an effective approach, and the returns tended to increase after some of the initial implementation costs were paid. For example, writing the first DMP is more labor-intensive than updating it regularly, and over time new staff will benefit from access to the DMP. Through this project, the USA-NPN made considerable progress on data documentation and delivery. In the future, we plan to address creating better traceability for dynamic data.

Compliance can be seen as a conversation (or a negotiation) between USGS leadership—which seeks standardized solutions for open data—and individual projects or programs with unique challenges based on their scientific emphasis and the demands of their stakeholders. Following the framework of the USGS Science Data Lifecycle and relying on the associated policies and software tools, we were able to reach near compliance in a relatively short time.

The dynamic nature of USA-NPN datasets created challenges relative to USGS policy for data preservation and release. USA-NPN's dynamic data products do not need to be versioned, as described above in the Publish/Share section. For dynamic DOIs, we continue to search for a solution to achieve compliance, although this issue will likely affect all other producers of dynamic data within USGS.

The USA-NPN is a somewhat unusual group within USGS, in that its major function is operational, that is, to produce data and data products primarily in support of research and decisionmaking by collaborating partners. A more typical USGS group might be primarily responsible for science, with data management as a secondary priority. Our staffing model of pararesearch professionals enabled us to create a suite of complex, relevant, robust, well-documented, and USGS-compliant datasets. As the commitment to big, open data grows throughout the Federal Government and academia, such models could be considered to improve efficiency and to reach emerging goals and priorities (for example, Werkheiser, 2017).

Acknowledgments

References Cited

Ault, T.R., Schwartz, M.D., Zurita-Milla, R., Weltzin, J.F., and Betancourt, J.L., 2015, Trends and natural variability of North American spring onset as evaluated by a new gridded dataset of spring indices: Journal of Climate, v. 28, no. 21, p. 8363–8378.

Chase, K.J., Bock, A.R., and Sando, R., 2017, Sharing our data—An overview of current (2016) USGS policies and practices for publishing data on ScienceBase and an example interactive mapping application: U.S. Geological Survey Open-File Report 2016–1202, 10 p., https://doi.org/10.3133/ofr20161202.

Crimmins, T.M., Marsh, R.L., Switzer, J.R., Crimmins, M.A., Gerst, K.L., Rosemartin, A.H., and Weltzin, J.F., 2017, USA National Phenology Network gridded products documentation: U.S. Geological Survey Open-File Report 2017–1003, 27 p., https://doi.org/10.3133/ofr20171003.

DataONE, 2016, Provide a citation and document provenance for your dataset: DataONE web page, accessed December 17, 2016, at https://www.dataone.org/best-practices/provide-citation-and-document-provenance-your-dataset.

Earth Science Information Partners, 2016, Interagency data stewardship/citations/provider guidelines: Earth Science Information Partners web page, accessed February 6, 2017, at http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/Citations/provider_guideline.

Faundeen, J.L., Burley, T.E., Carlino, J.A., Govoni, D.L., Henkel, H.S., Holl, S.L., Hutchison, V.B., Martín, E., Montgomery, E.T., Ladino, C.C., Tessler, S., and Zolly, L.S., 2013, The United States Geological Survey science data lifecycle model: U.S. Geological Survey Open-File Report 2013–1265, 4 p., http://dx.doi.org/10.3133/ofr20131265.

Holdren, J.P., 2013, Memorandum for the heads of executive departments and agencies—Increasing access to the results of Federally funded scientific research: Executive Office of the President, Office of Science and Technology Policy dated February 22, 2013, accessed August 8, 2017, at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

Kirk, K., 2016, Use Case—Data release via USGS web site or USGS web service: USGS web page, accessed August 8, 2017, at https://www2.usgs.gov/datamanagement/documents/Data-Release-Web-Use-Case.pdf.

Monahan, W.B., Rosemartin, A., Gerst, K.L., Fisichelli, N.A., Ault, T., Schwartz, M.D., Gross, J.E., and Weltzin, J.F., 2016, Climate change is advancing spring onset across the U.S. national park system: Ecosphere, v. 7, no. 10, http://dx.doi.org/10.1002/ecs2.1465.

PRISM Climate Group, 2017, PRISM climate data: PRISM web page, accessed August 8, 2017, at http://prism.oregonstate.edu.

Rauber, A., Asmi, A., van Uytvanck, D. and Pröll, S., 2016, Identification of reproducible subsets for data citation, sharing and re-use: Bulletin of IEEE Technical Committee on Digital Libraries, v. 12, no. 1, p. 6–15, accessed December 15, 2016 at https://www.rd-alliance.org/system/files/documents/TCDL-RDA-Guidelines_160411.pdf.

Rosemartin, A.H., Denny, E.G., Weltzin, J.F., Marsh, R.L., Wilson, B.E., Mehdipoor, H., Zurita-Milla, R., and Schwartz, M.D., 2015, Lilac and honeysuckle phenology data 1956–2014: Scientific data, v. 2, https://doi.org/10.1038/sdata.2015.38.

Schwartz, M.D., 1997, Spring index models—An approach to connecting satellite and surface phenology: Phenology in Seasonal Climates, p. 23–38.

Schwartz, M.D., Ahas R., and Aasa, A.,2006, Onset of spring starting earlier across the Northern Hemisphere: Global Change Biology, v. 12, no. 2, p. 343–351.

Schwartz, M.D., Ault T.R,. and Betancourt, J.L., 2013, Spring onset variations and trends in the continental United States—Past and regional assessment using temperature-based indices: International Journal of Climatology, v. 33, no. 13, p. 2917–2922.

Sullivan, B.L., Phillips, T., Dayer, A.A., Wood, C.L., Farnsworth, A., Iliff, M.J., Davies, I.J., Wiggins, A., Fink, D., Hochachka, W.M., and Rodewald, A.D., 2016, Using open access observational data for conservation action—A case study for birds: Biological Conservation, v. 208, p. 5–14, http://dx.doi.org/10.1016/j.biocon.2016.04.031.

USA-National Phenology Network National Coordinating Office, 2016a, USA National Phenology Network Data Management Plan: USA-National Phenology Network, also available at https://www.usanpn.org/files/shared/USA-NPN_DMP_12-5-16.pdf.

USA- National Phenology Network National Coordinating Office, 2016b, USA-National Phenology Network Data Product Development Framework and Data Product Catalog, version 1.1: USA-NPN Technical Series 2016–001, also available at https://www.usanpn.org/files/npn/reports/USA-NPN_Data-Product-Catalog1.1-2016Dec.pdf.

U.S. Geological Survey, 2016, Public access to results of Federally funded research at the U.S. Geological Survey—Scholarly publications and digital data: U.S. Geological Survey web page, accessed December 22, 2016, at https://www2.usgs.gov/quality_integrity/open_access/downloads/USGS-PublicAccessPlan-APPROVED-v1.03.pdf.

U.S. Geological Survey, 2017a, U.S. Geological Survey Manual. Accessed August 8 2017 at https://www2.usgs.gov/usgs-manual/.

U.S. Geological Survey, 2017b, Guidance on disclaimer statements allowed in USGS science information products: U.S. Geological Survey, Fundamental Science Practices web page, accessed April 7, 2017, at https://www2.usgs.gov/fsp/fsp_disclaimers.asp.

Werkheiser, W., 2017, USGS Director's Annual Bureau Guidance for Fiscal Year 2018: U.S. Geological Survey Memorandum, dated June 9, 2017.