

DATA QUALITY ASSURANCE & QUALITY CONTROL FOR NATURE'S NOTEBOOK

V1.0 – October 6, 2011

The primary source of observational plant and animal data for the USA-NPN is a national pool of observers ranging from high school students and retirees to professional researchers who participate in *Nature's Notebook*. Observers are not paid and are not typically field-trained by the USA-NPN, and a threshold skill or experience level is not required (or enforceable) for participation in data contribution. In addition, the nature of phenological observation is potentially more subject to observer interpretation than that for other data collection efforts, such as water quality monitoring or precipitation gauging.

To maximize data quality and utility, the USA-NPN has established a suite of quality assurance (QA; before data enters database) and quality control (QC; post-processing) measures for *Nature's Notebook*. Through the full implementation of QA/QC measures, data end users will be able to select observers by skill level as well as track the revision history of a data set, know when observations were made, distinguish between data collected by different observers at a site, and investigate inconsistencies or outliers in the data set. QA/QC measures completed to date and proposed are summarized in Table 1.

Table 1. QA/QC measures for Nature's Notebook completed to date (black text) and proposed (gray text), grouped by potential source of error.

QUALITY ASSURANCE MEASURES	QUALITY CONTROL MEASURES
Species Identification Errors	
<ul style="list-style-type: none"> ● “How to observe” monitoring instructions and Frequently Asked Questions (FAQs) emphasize the importance of accurate species identification and direct observers to general identification resources ● Species profile pages include a photo, range map, written description of the species, and lead the user to other websites with more identification information 	<ul style="list-style-type: none"> ● Site and plant level metadata (e.g., land cover type for sites, watered status for plants) enables data end users to explore outliers ● In a preliminary test of species identification errors, 3.7% of species were registered in states outside of their known range (n of 4857 registered plants and animals) ● Species identification is confirmed via submission of photo with observation (with crowd sourced review of images and expert confirmation on an image subset)
Phenophase Status Evaluation Errors	
<ul style="list-style-type: none"> ● Language in phenophase definitions is carefully chosen for precision and accessibility ● Phenophase definitions are generalized and identical across similar species (within 	<ul style="list-style-type: none"> ● Detection bias in animal phenology reporting is exposed via observer reports of the time spent observing animals and their selection of an animal survey method from a pick list

<ul style="list-style-type: none"> phenological functional types) for consistency • Phenophase definitions are changed as infrequently as possible to simplify observing and to ease the interpretation burden on data-end users • Species-specific additions to the general definitions more completely describe how the phenophase appears in a particular species • Observers are given an ‘uncertain’ option to reduce false positives and false negatives • Observers are not asked to infer the date of a ‘first’; dates of all visits are known explicitly • FAQs address tricky issues in phenophase status evaluation (across species) • Photos or illustrations for each phenophase in each species are provided to observers • Photographic primer teaches phenophase evaluation skills • Online photographic quiz tests and hones observers’ skill in phenophase evaluation 	<ul style="list-style-type: none"> • Spatial interpolation to identify outliers as data density allows • Flagging of phenophases reported out of expected order and logically inconsistent • Comparisons of observation data from expert and non-expert (or trained and untrained) observers at the same site • Assessments in which observers are asked questions about their observations targeted at identifying mischaracterizations of phenophases • Phenophase evaluation is confirmed via submission of photo with observation (with crowd sourced review of images and expert confirmation on an image subset)
--	---

Data Entry Errors

<ul style="list-style-type: none"> • Training and FAQs address data entry issues • Species names and abundance/intensity measures are presented as pick lists • Datasheets mirror the online data entry form • Phenophase and intensity definitions appear on roll-overs in the data entry form • Site location can be entered by Google map or address input; elevation is calculated from USGS digital elevation model, but can be hand-corrected • Observers can review previously submitted observations in user interface (UI) or a downloaded Excel file, and can edit their previously submitted observations in UI • Usability testing has been conducted on user interface to increase intuitiveness and reduce transcription errors • User interface validation on observation methods: <ul style="list-style-type: none"> ○ Users must provide both a measurement and a metric to input data regarding the amount of 	<ul style="list-style-type: none"> • Collect and cross check a sample of observer datasheets with database
---	---

<p>time spent observing, time spent traveling to observation site, and time spent searching for animals</p> <ul style="list-style-type: none"> ● Reordering of plant and animal lists reorders data entry form and datasheet printout ● When a plant is deleted, rationale for deletion is requested and the deleted plant data is retained ● Comments box provided at the site, plant and observation level ● User interface validation on date/time: <ul style="list-style-type: none"> ○ Date field required; default is to select from a calendar ○ Time field optional; selected from pick list ○ Dates in the future not allowed ○ After the date is entered it appears above the phenophase column for every species ○ Duplicate date/time values not allowed ○ Observations cannot be made about an individual after it has been marked as 'inactive' ○ User is automatically warned by UI of changing phenophases through time ● User interface validation on phenophases: <ul style="list-style-type: none"> ○ User may only enter "Yes," "No" or "Uncertain" on the interface, using mutually exclusive click points; if no response is checked no database record is created ○ User may not enter abundance or intensity measure unless the phenophase is set to "Yes" or "Uncertain" ● Observers see their data re-presented to them in a visualization on their "My Account" page ● Mobile applications for data collection eliminate datasheet to interface transcription errors 	
---	--

Training and Observer Skill Level

<ul style="list-style-type: none"> ● Field observing methods (selecting a site, selecting species, making observations) are accessible as: <ul style="list-style-type: none"> ○ Printable handbook ○ Handouts ○ Powerpoint presentation with script for aspects of data collection and data entry ○ Voiced "training video" version of the 	<ul style="list-style-type: none"> ● Self-reporting of training, skill and experience level by observers ● Record of observer's online quiz scores ● Scores based on comparison of observer's data to expert's data at sites where both are monitoring the same species
--	--

<p>Powerpoint presentations</p> <ul style="list-style-type: none"> • FAQs available on separate page, and as clickable web links from places on website where user questions might arise • In-person and online workshops provide training opportunities for a limited number of observers • A variety of peer-support networks can be implemented, from user forums on the website to power-observers who review other observers' data 	<ul style="list-style-type: none"> • Use rainlog, eBird or another program with more easily interpolated/QC-ready data to determine characteristics of skilled observers; apply the findings to the Nature's Notebook observer pool
--	--

In addition to these error-specific QA/QC measures, the USA-NPN has taken several broader steps to ensure high quality data, outlined below.

Recruitment and Retention

- Biweekly reminders to observers are targeted to increase temporal resolution of the data set
- Retention efforts to maintain observers long term include leaderboards, prizes, demonstrations of utility of data and quarterly email updates
- Recruitment efforts are targeted around priority geographic regions
- Recruitment is targeted at high-school age and above audiences
- Shared sites enable on-the-ground administrators to access and review their member's data

Methodology

- Individual plants are tracked through time, controlling for variation across organisms and in microclimate
- Observers are encouraged to monitor multiple individuals of each species of plant at each site to capture variation
- Monitoring plant and animal taxa at the same site enables analyses of species interactions

Data Management

- Data submitted via alternate interfaces (partner, mobile or facebook apps) are tagged as such with associated metadata about the interface
- Data integrated from other programs are tagged as such with associated metadata about methodology and interface

Options for Data End Users

- Observer contact information enables NCO-mediated follow up from data end users on outliers or biases in the data
- Data end users can select data to fit their criteria:
 - Collected by trained volunteers with a particular program (e.g. Signs of the Seasons)
 - Sites where plants were not watered or fertilized; or where no water or feeding stations were available for animals
 - Collected on a weekly basis, or more frequently