Appendix: Example primary persona: Sun

(Primary persona)

Source: Data Conservancy Sun persona: comments from Lynn Rogers. Based in part on a Data Conservancy persona written by Anne Thessen. Revised by Kevin Crowston with some details based on William I. Boarman, USGS.

Tags: non-academic, government, early career, single discipline, field, human and machine-collected data, novice data management, biology

See also: Dr Yolanda Suarez DataONE Scenario

Photo credit: U.S. Army Environmental Command https://www.flickr.com/photos /armyenvironmental/2650014187

Note: the persons pictured and referenced above are not affiliated with DataONE and the use of this image or their names does not constitute endorsement of DataONE.

Background

Name, age, and education

Sun is a biologist specializing in desert tortoises. She did her masters and Ph.D. at California State University San Marcos. She has spent her career studying tortoises in their natural habitat.

Life or career goals, fears, hopes, and attitudes

Sun recently started working for the USGS Western Ecological Research Center, "one of 18 Centers of the Biological Resources Discipline of the U.S. Geological Survey," (http://www.werc.usgs.gov/who.aspx). Her broad interest is how human activity and climate change will affect tortoise populations. Her research needs to inform decisions by land managers in various state and federal agencies. She works with NGOs on conservation issues and speaks to the public on tortoises and conservation issues. For example, she collaborates with biologists at the Wildlife Research Institute (http://www.wildlife-research.org/page10.html) on a project tracking desert tortoises relocated from the expanding Fort Irwin Army Base. She writes technical reports and publishes peer-reviewed journal articles (e.g., http://www.conservation-science.com/Products.html; http://www.werc.usgs.gov/person.aspx?personID=52).

A day in her life

Sun and other members of the research team go into the field with a notebook, camera, simple instruments, and sample containers. They capture and tag tortoises before collecting data about individuals such as age, weight, and sex. They also collect data about entire tortoise populations by taking a census, collecting feces, and monitoring carcasses. Much of these data are recorded in a notebook and later copied onto a spreadsheet for analysis with desktop statistics software. A number of her research subjects are radio tagged, giving her a latitude/longitude position as often as every 10 minutes.

Reasons for using DataONE to share and to reuse data

Needs and expectations of DataONE tools

Sun feels that she cannot easily share her own data for fear of disclosing sensitive information because of the work location and the fact that she works on endangered species. Even an embargoed dataset could be problematic, as tortoises keep the same home range and the lifespan of a tortoise vastly exceeds the duration of any reasonable embargo. However, she might be able to share derivative datasets, if these could be easily created, or a subset of less sensitive data, such as life history, demographic, or behavioral data (e.g., home range size, daily and seasonal activity, diet, social biology, or thermo-regulatory behavior).

DataONE might also be useful in improving Sun's overall data management capabilities, e.g., educating her on best practices for data quality and metadata development. If DataONE provided tools for cataloguing and managing locally-stored data, these could be very useful. She might be willing to deposit data at a member node for limited sharing, preservation, and for ensuring long-term preservation of data (e.g., migration of data formats), though only if its privacy can be assured and doing so were as easy as (or at least, not much harder than) maintaining local backups.

Sun is interested in finding additional data that correspond to the location of tortoise populations, and additional tortoise data so she can put her current study into perspective and perhaps find collaborators. For example, data on invasive species in the area she studies could help explain changes observed in the populations. She does not have much technical support, so she needs easy-to-use tools. Given that her research is motivated by both scientific interests and policy concerns, she is extremely wary of using data of unknown origin or quality, so discoverability and validation of datasets are key issues.

Intellectual and physical skills that can be applied

As a trained research scientist, there should be no overt challenges to dealing with data *per se*. However, though Sun strives to follow established data-collection protocols, the realities of field

research mean that her methods are often adjusted on the fly and her data needs secondary analysis and clean up. If DataONE provides tools to aid in the integration of similar, yet not identical, datasets, and can help her to troubleshoot data-entry and other errors in her own data, her own use and possible subsequent deposition of her data into a DataONE member node would be simple.

Technical support available

Sun has very little computer support within her research group and institution, but she does have experience with field equipment and general computer competencies. Thus far, complex visualizations and data-handling algorithms have not been a factor in her work, so any system that did not offer the option to work with simple datasets using easy tools would probably intimidate her.

Personal biases about data sharing and reuse (and data management more generally)

Sun is interested in reviewing data that might inform her studies, but does not depend on it and it is not yet an important part of her work. On the other hand, she does not have the technical skills to prepare her data for sharing nor does she have large quantities of data that she thinks would be of interest to others. Furthermore, she is hesitant to share her geo-located data because she works with a threatened species. So far, she has only shared raw data with close colleagues.

Currently, Sun only collects data for her personal use. She validates her data and describes it, though not following any broadly-used data quality or metadata standards. Deposit is in the form of publications based on summaries and analyses; the raw data themselves are not shared. These data are then analyzed and used to drive further data collection.

The main motivation for Sun to use DataONE is to improve her data management practices and discover potentially useful data created by other researchers to integrate into her own analyses. Sun could use DataONE tools (and the training in their use) to improve her capabilities for data assurance and description. Under the right conditions, she could use DataONE tools for preparing data for deposit and preservation, and potentially even for reuse of appropriately redacted data by other researchers.
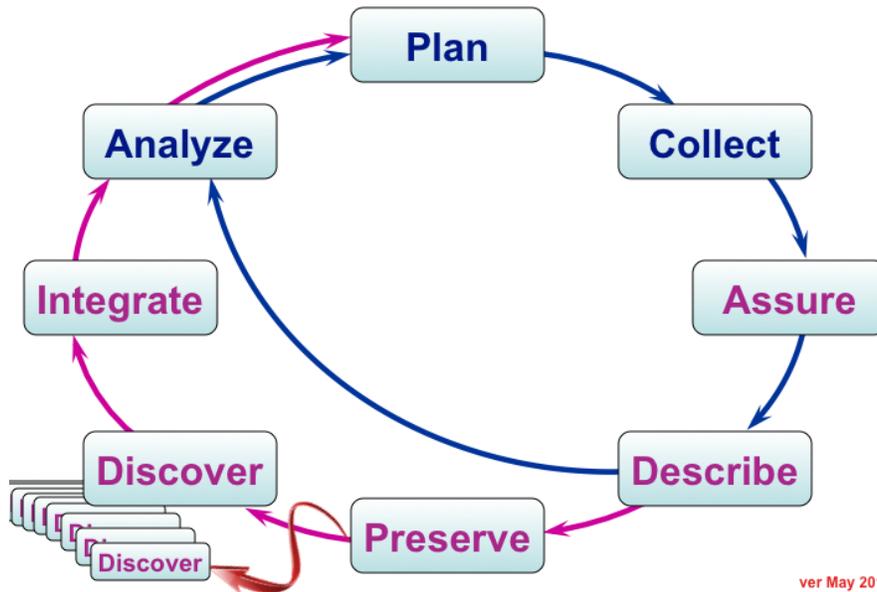
Sun, Desert Tortoise Researcher   **Current Practice**

Plan

Analyze

Collect

Integrate

Assure

Discover

Describe

Preserve
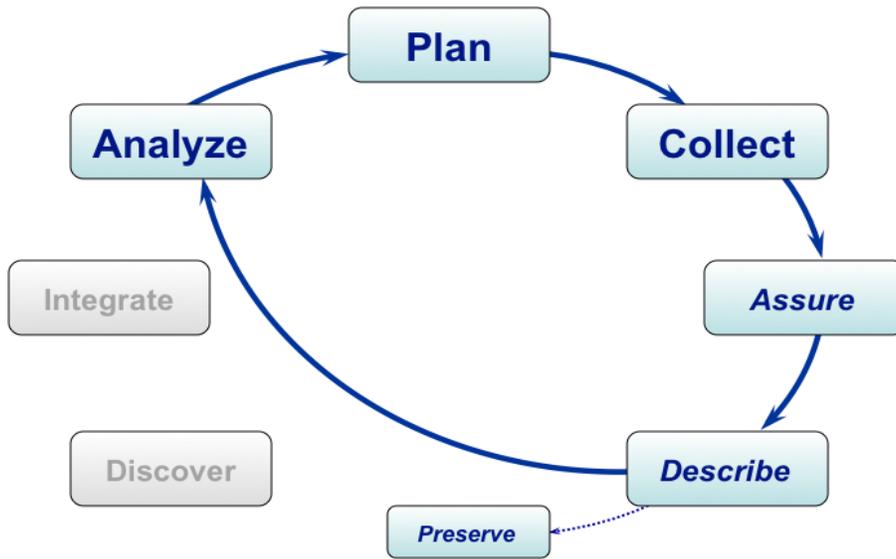
ver May 2012

Sun, Desert Tortoise Researcher   **Current Practice**
DataONE Enables…

Plan

Analyze

Collect

Integrate

Assure

Discover

Describe

Preserve

Discover

ver May 2012

Comparison of current and DataONE-enabled practices:

Current project planning:
No explicit attention to data issues in project planning.

DataONE-enabled project planning:
Management Planning: Develops a project Data Management Plan following examples provided on the DataONE portal.

Current data collection:
Collects tortoise field data.

DataONE-enabled data collection:
No change.

Current data assurance:
Validates data using own standards.

DataONE-enabled data assurance:
Could apply more broadly-used data-quality standards and assurance tools.

Current data description:
Describes data for her own purposes, using her own data description techniques.

DataONE-enabled data description:
- *Training:* Learns how to use *Morpho* (a metadata management editor) based on instructional materials available in the DataONE Best Practices Database and associated downloadable video instructions.
- Creates metadata for datasets following best practices.

Current data preservation:
Sun publishes summary and analysis results but does not deposit data. Data preservation is done only within her lab.

DataONE-enabled data preservation:
Sun might deposit data with a DataONE member node for long-term preservation, with appropriate protections for sensitive data.
- *Data Preservation:* Deposits data and metadata in the USGS data repository with appropriate protections for sensitive data and redaction to create shareable data subsets.
- *Data Preservation:* Submits a research paper to an ecological journal associated with Dryad—a DataONE Member Node. Upon acceptance, she submits the publication-relevant data, metadata, and model to Dryad where they are given a DOI (Digital Object Identifier) and preserved in the Dryad repository.
- *Citation:* Upon publication, she adds the publication reference and the data citation (including DOIs for both; provided by Dryad and the journal) to her CV.

Current data discovery:
Does not use other researchers' data.

DataONE-enabled data discovery:
The possibility of discovering relevant data from other researchers is likely to be a main motivation for Sun's use of DataONE and DataONE tools.

- *Data Discovery, Access, Use, and Dissemination:* Searches for tortoise food web and area meteorological data in the region at the DataONE portal. Searches for land-use histories, especially for former grazing lands. Searches for co-locality data for other animal species as possible signals for other ecological changes in the region.
- *Data Discovery, Access, Use, and Dissemination:* Identifies relevant data and downloads data and metadata from previous LTER studies as well as data collected by state and Federal agency scientists (i.e., non-LTER).
- *Data Discovery, Access, Use, and Dissemination:* Acquires supplemental data from another DataONE Member Node with complete citation information.
- *Citation:* Another scientist working in Mexico on a similar study discovers the new publication and data created by Sun and cites her in his work.

Current data integration: Does not use other researchers' data.	*DataONE-enabled data integration:* Use DataONE tools to integrate her data with data discovered from other researchers.
Current data analysis: Uses standard desktop data analysis tools.	*DataONE-enabled analysis:* *Data Visualization:* Uses data analysis and visualization tools identified through DataONE Tools Database or available as part of the Investigator Toolkit to analyze existing data and develop initial model parameters that she will use in her own research.