The Virtual Notebook: a "TurboTax®" approach to improving research metadata

Central Arizona-Phoenix Long-Term Ecological Research
CAP LTER

Global Institute of Sustainability

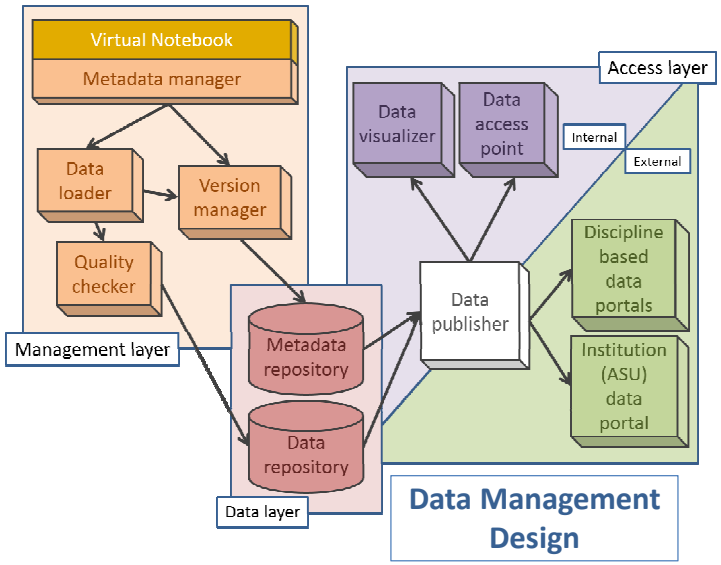
Philip Tarrant, David Julian, Ryan Raub. Arizona State University, Tempe, AZ.

ASU

Problem statement:

A major challenge for all researchers is the long term preservation of their research data. This complex challenge covers not only the collection and storage of multiple data formats, but also the creation of relevant (and sufficient) metadata to describe the data for future re-use. A key issue is that the creation of dataset metadata can be an afterthought, prepared by an investigator who is focused on new research activities rather than old projects.

But let's face it, documenting your research data is like completing your tax return – you don't want to do it, but you have to, so you appreciate your tax preparation software, which helps you manage this onerous task. Enter the Virtual Notebook, the TurboTax® for your research data!



Data Management Design

Proposed solution:

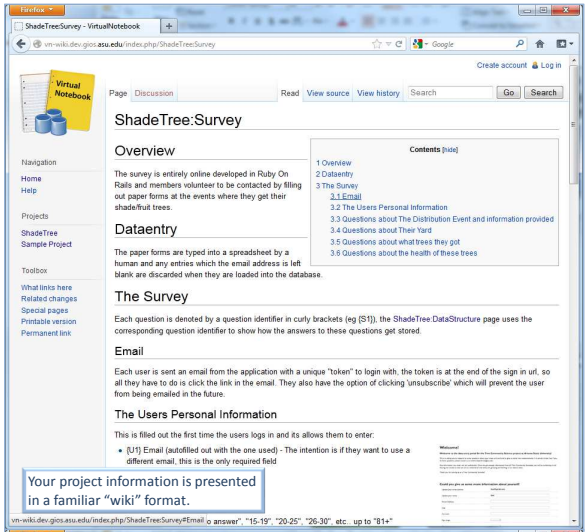
The Informatics team at the Global Institute of Sustainability is developing the Virtual Notebook, a knowledge management system, to help our investigators manage research related information and prepare their data for curation and publication.

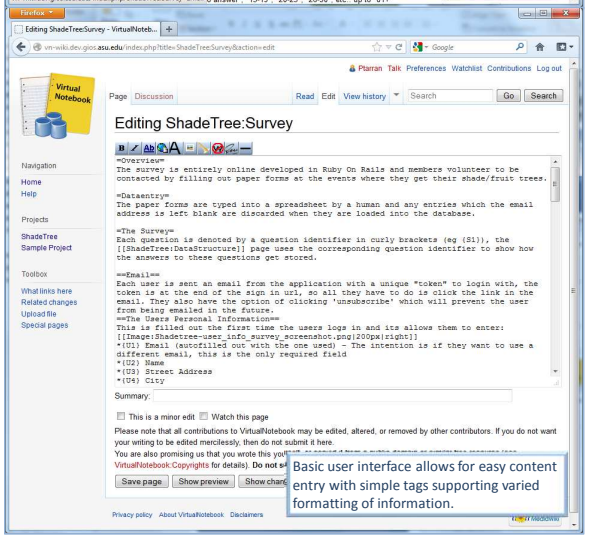
This system will address a major root cause of poor metadata quality (the temporal disconnect between data creation and metadata preparation), by encouraging the incremental development of project information, including project abstracts, methods, protocols and dataset metadata. Programming interfaces will eventually support dataset ingestion by the LTER Network Information System as well as other data repositories.

Project goals:

1. Create a single source for research information, supporting relationships between projects, people, methods, references, media, and dataset metadata
2. Must be multipurpose, supporting the presentation of organization data (people, projects, publications) on websites as well as facilitate the preparation of private and public datasets for publication, access and download

User Interface





Phase One – proof of concept:

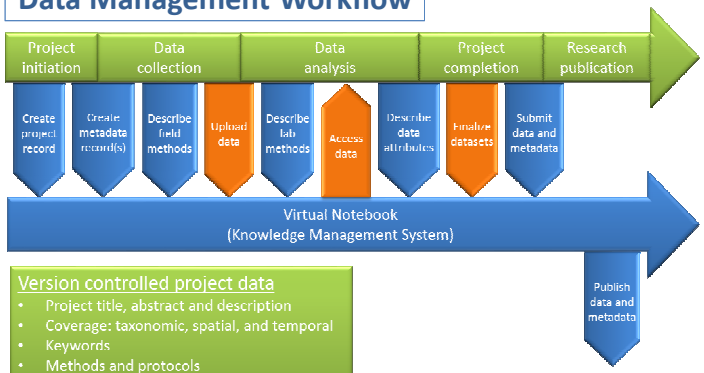
The first phase of the Virtual Notebook development is to create a simple wiki based proof of concept that will allow us to test the potential value of the notebook functions without committing to a specific design or investing large amounts of development effort without researcher feedback.

Feedback by early adopters will help us to ensure that the final product is "fit for purpose". If you want to help us develop the Notebook by putting your research information into this prototype, please discuss your needs with us.

Ultimately, we want the Notebook to:

- Accommodate interdisciplinary projects
- Provide help and guidance throughout a project
- Help you manage your research data submissions
- Be shared with other interested parties
- Be capable of producing specific outputs for different purposes

Data Management Workflow



Version controlled project data

- Project title, abstract and description
- Coverage: taxonomic, spatial, and temporal
- Keywords
- Methods and protocols
- Dataset descriptions
- Attribute descriptions

Effort is required to input project and dataset information, but:

- The "Pay As You Go" model reduces the back-end workload, so you can plan for the future
- You have a single, resilient place to store your research project information – accessible by all team members
- The latest "version" is always available and accessible
- Data backup, replication and preservation are managed for you
- Notebook content is available as input to manuscripts
- Federal data management requirements are automatically met
- Writing proposals gets easier because you can reference existing published datasets (assuming you used the system for your previous project, of course!)