echinopscis (an extensible notebook for open science on specimens): overview & design principles

TaxonWorks Together (TWT) - 26th October 2023
Nicky Nicolson¹, Eve Lucas²
(1) Digital revolution, RBG Kew, (2) Accelerated taxonomy, RBG Kew
Context

• Transitioned from software development into research
• Interested in how we can use software development practices in research:
  • Automation
  • Version control
  • Dependency management
  • Continuous integration
• Institutional commitment to use digital technologies to accelerate the process of taxonomy
Community aim: digital extended specimen

• Integrate specimens and associated data across multiple research infrastructures
  • allowing the investigation of wider scale research questions

- How to get there? Activities at a range of scales:
  - Large scale: computational
  - Distributed: lightweight tools, link construction in context by researchers
Community aim: digital extended specimen

- Integrate specimens and associated data across multiple research infrastructures
  - allowing the investigation of wider scale research questions

- How to get there? Activities at a range of scales:
  - Large scale: computational
  - Distributed: lightweight tools, link construction in context by researchers
• A personal knowledge manager: for creating & linking research notes
• Emphasises linking
• Data stored locally, using open formats
  • Markdown and optional structured data frontmatter
• Works offline
• Extensible architecture – plugins for data access and citation processing
• Active user and developer community
• A personal knowledge manager: for creating & linking research notes
• Emphasises linking
• Data stored locally, using open formats
  • Markdown and optional structured data frontmatter
• Works offline
• Extensible architecture – plugins for data access and citation processing
• Active user and developer community

...sounds a lot like OpenRefine, which we have adopted with some success

Could this contribute to our management of linked, semi-structured data, as Open Refine has for tabular data?
Extend Obsidian for specimen research

- Access of relevant data
  - Specimens (GBIF)
  - Names (International Plant Names Index)
  - Collections (Global Registry of Scientific Collections)
  - People (Bionomia)
  - Literature (crossref)
- Creation of links, spatial and network exploration
- Citation in new work
- Open science working practices
Start here

Welcome

This is the sample vault for "echinopsris" - an experiment in creating an extensible notebook for open science.

These instructions are a page in Obsidian, formatted in Markdown and located on your local computer. You can edit this page - toggle between view and edit mode using Ctrl + E.

By default these instructions are "pinned" so that they will always be visible as we create new pages. Any new pages that you create or open by navigating a link will open in a separate pane to the right.

About this demonstration

This vault has been configured with some tools which enable easy access, linking and visualisation of biodiversity informatics data.

The rest of this page gives a worked example of how to use these. Occasionally, some extra information about Obsidian and pointers to the Obsidian documentation has been included in this page in “tip” sections. These look like this - if you click on the coloured bar the contents will unroll.

Tips and extra information about Obsidian

Functions are available via the command palette in Obsidian - a searchable list of the functions in the current system.
• Our physical herbarium includes spaces for layout & comparison of large sets of specimens

• Working digitally means we can access specimen images from anywhere

• But how to effect this kind of interaction in digital environment?
Arranging specimens: virtual prototype
Roadmap

1. **Personal research environment** based on Markdown authoring and linking (now working on “seeding” – pre-populating data)

2. **Web publication** using static site generators (conceptual similarities with the [GBIF hosted portal](https://www.gbif.org) and TaxonPages work)

3. **Document production**: with structured bibliographic/specimen references

4. **Dataset production**: mobilisation of content and links into DarwinCore archives for aggregator harvesting
Curation of Camellia
Exploring and testing methods of taxonomic curation and identification of herbarium specimens

Celia Acear
2022

Thesis submitted in partial fulfillment for the MSc in the Biodiversity and Taxonomy of Plants.
Taxonomic Literature (TL-2) – multi-volume work presented in Obsidian

Anton Savchenko
@notgaudi

Release alert 🎉
I present a digital version of Taxonomic Literature ed. 2, commonly known as TL-2. It can be viewed online or downloaded as an Obsidian database and used offline on your PC. See it here: tl2.io
Integration: widescale adoption of Reconciliation API

• “strings to things” – the echinopscis demo shows multiple workflows where a user translates a piece of text (like a specimen reference) into an entity managed by an external authority (like a specimen record)

• The Open Refine Reconciliation API is now documented as a standard and can be implemented:
  • by the data provider, those that maintain an authoritative dataset
    • Bionomia: https://bionomia.net/reconcile
    • Catalogue of Life: https://github.com/CatalogueOfLife/backend/issues/1265
    • Datasette (generic): https://github.com/drkane/datasette-reconcile
  • by data consumers: developers of taxonomic / data manipulation tools:
    • Open Refine: https://openrefine.org/docs/manual/reconciling
    • Python: https://github.com/jvfe/reconciler
    • Could we build an Obsidian plugin?
Integration: navigation of linked data

- “entity explosion” is a browser plugin that shows links from Wikidata based on the currently loaded page.
- Should be possible to develop an Obsidian equivalent, that uses an ID stored in page frontmatter to run a Wikidata query and show useful links.
Explore

Background, demo videos, installation information, roadmap and ideas for contributions:

ewitness.github.io

Source code and documentation:

github.com/echinopscis

n.nicolson@kew.org
n.nicolson
n.nicolson