

Implementing community best practice data and metadata capture into laboratory workflows using Sparrow



Steve Kuehn¹, Daven Quinn², Casey Idzikowski², Victor Atasie¹, Tsega Samson¹

¹Concord University, Athens, WV, sckuehn@concord.edu

²University of Wisconsin, Madison, WI, daven.quinn@wisc.edu cidzikowski@wisc.edu

EarthCube Annual Meeting

Poster presentations 1

14-June-2022, 4:00pm - 5:30pm

Background

International tephra data workshops held in 2014, 2017, 2019, and 2022 focused on improving tephra studies through standardization and encouraging FAIR data practices. These efforts led to comprehensive recommendations for data collection and reporting in tephra studies (<https://zenodo.org/record/3866266>), initial implementations in the StraboSpot field app and database (<https://strabospot.org/files/StraboSpotTephraHelp.pdf>) and at SESAR-EarthChem (<https://earthchem.org/communities/tephra/>) along with other workshop output and archives (<https://tephrochronology.org/cot/Tephra2022/>).

These efforts also revealed a need for additional digital tools, including a bridge between the field (e.g. Strabo) and data repository (e.g. EarthChem) with support for laboratory workflows. From this, the concept of Sparrow-Tephra was born.

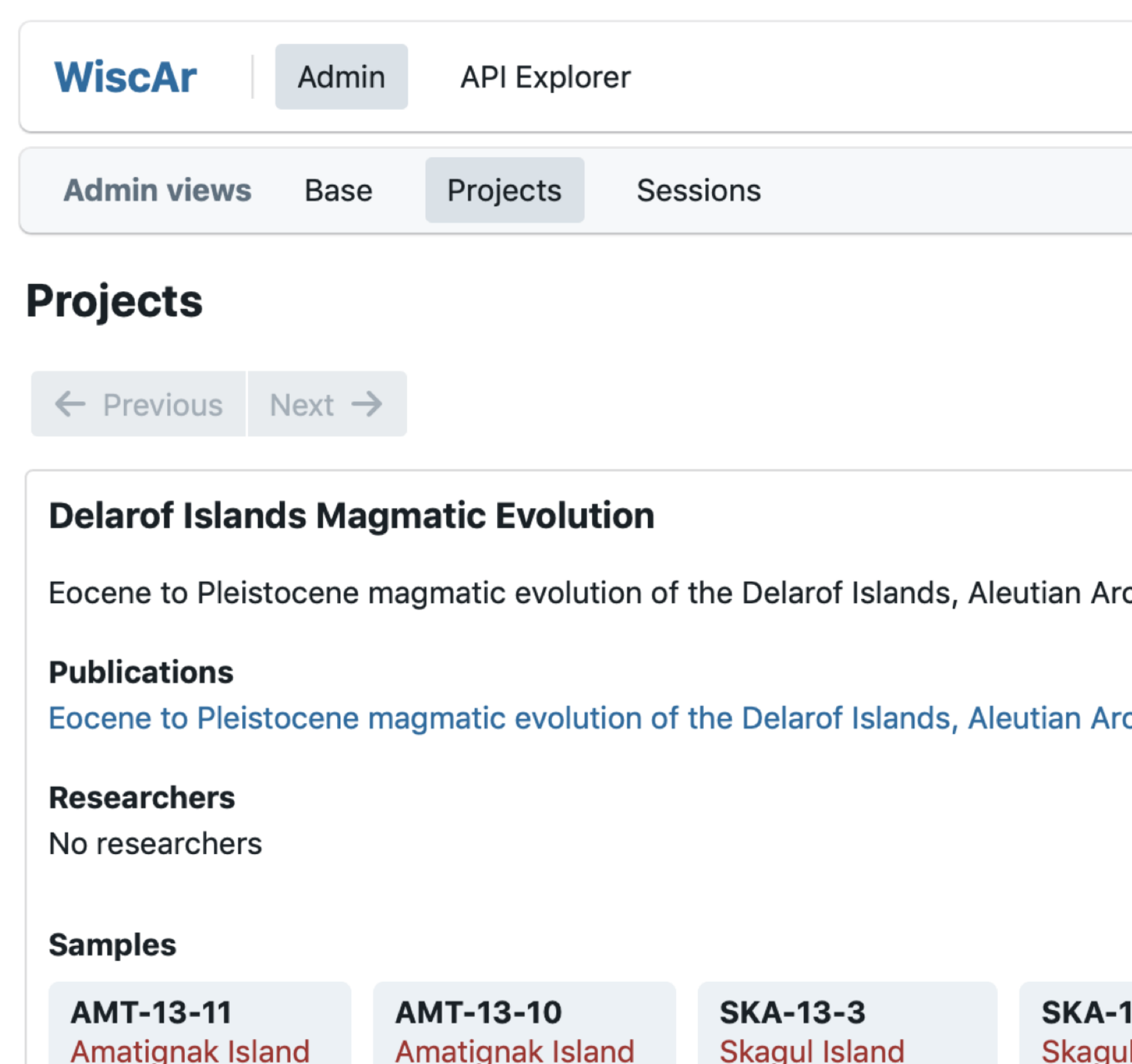
See also the Wallace et al. companion poster on tephra community best practice developments

What is Sparrow?

Sparrow is a laboratory data management system initially developed for geochronology, but extensible and customizable for other research foci (<https://sparrow-data.org>

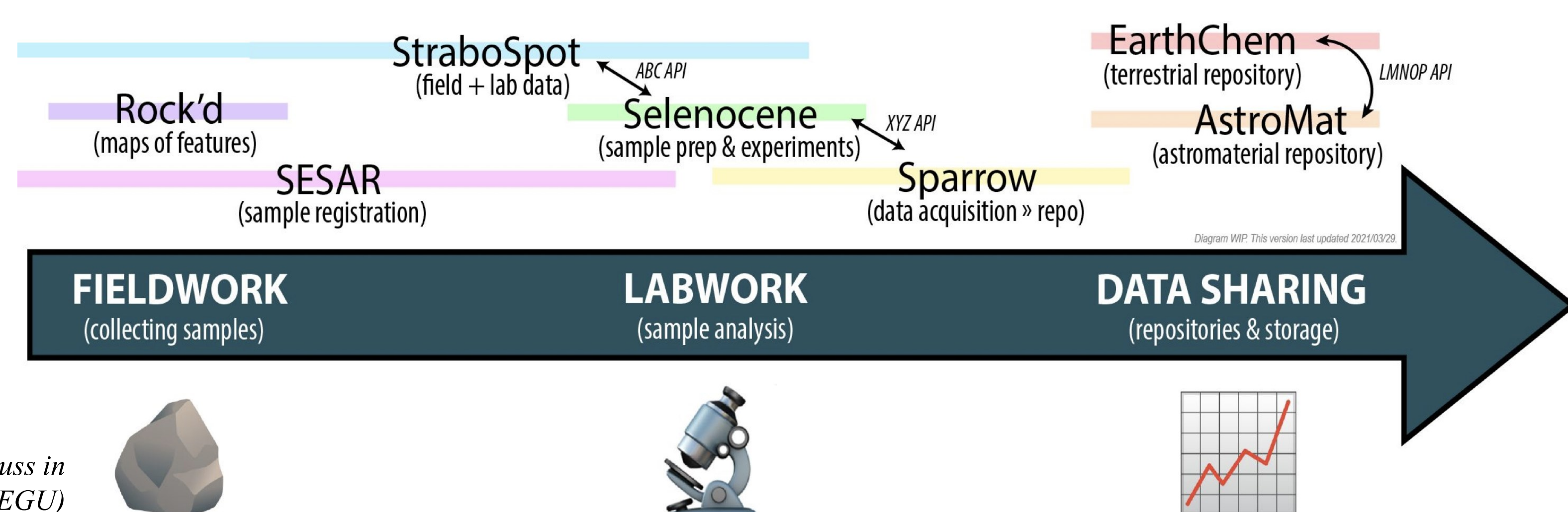
<https://github.com/EarthCubeGeochron/Sparrow>). Sparrow provides the following:

- PostgreSQL database backend
- Python-based API server
- React-based administration interface
- Web front-end and user interface
- Platform for data import, export, embargo, public access, visualization, versioning, & more
- Package containerized using Docker for easy installation on UNIX-type systems
- Lab-managed, lab-controlled archive that can be run locally or in the cloud
- Detailed Sparrow presentation: <https://sparrow-documentation-assets.s3.amazonaws.com/media/2022/14-april-2022-sparrow-kbsi.pdf>



Sparrow-Tephra Goals

- Support for most major types of tephra data
- Alignment to community best-practice recommendations, standards, and vocabularies
- Live capture of laboratory data and metadata with customized support for laboratory workflows, including sample preparation and method details
- Ingest of legacy data and integration of new and legacy data
- Import of sample information from StraboSpot
- Output of standardized data and metadata to e.g. SESAR-EarthChem



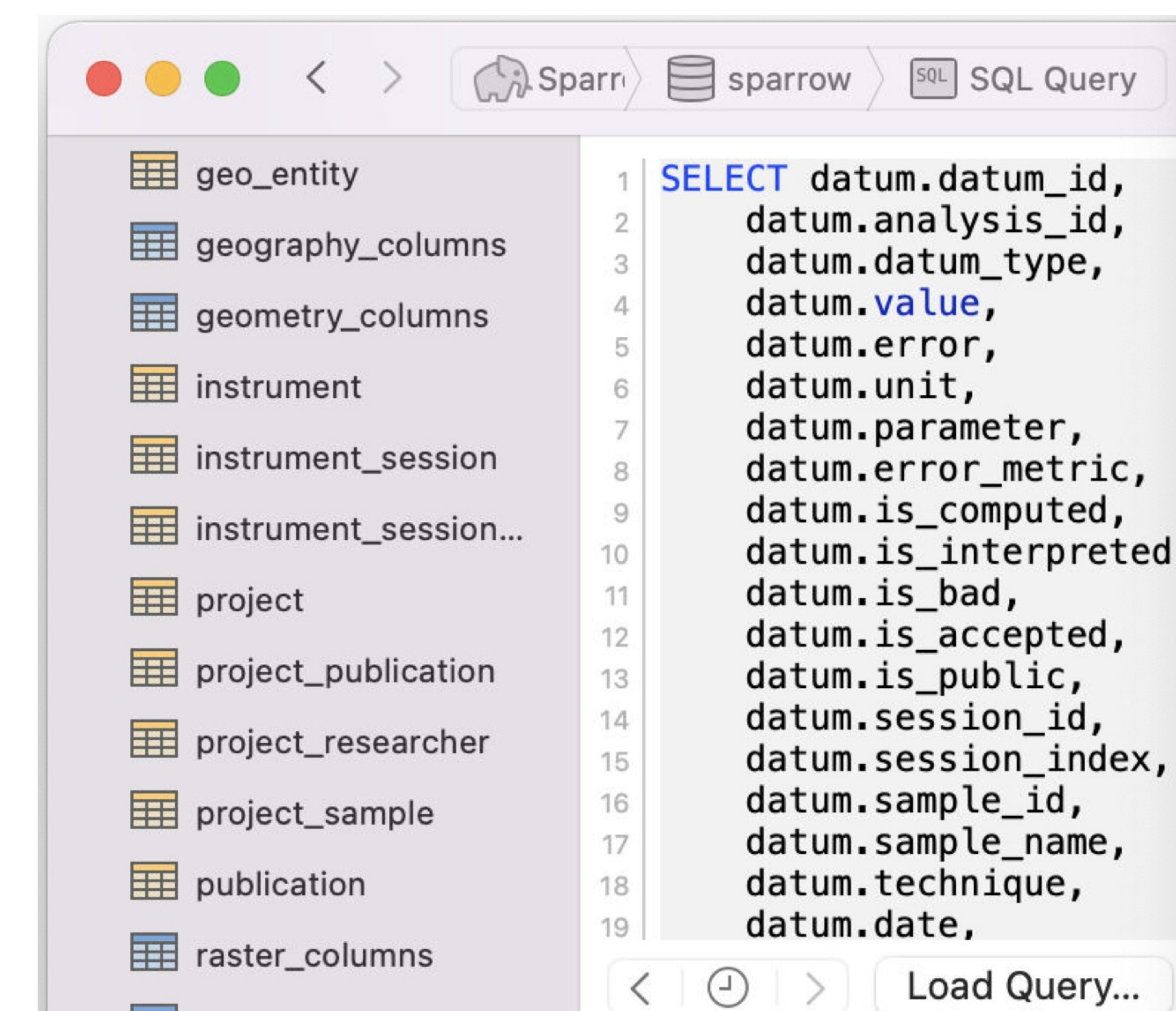
Graphic by Beck Strauss in
Lehnert et al. 2021 (EGU)

Sparrow-Tephra Data Coverage and Draft Schema

Tephra consists of fragmental material ejected from volcanoes. Tephra research has interdisciplinary applications from volcanology to geochronology, and this research involves a wide range of data types which overlap work in sedimentology-stratigraphy and geochemistry-petrology.

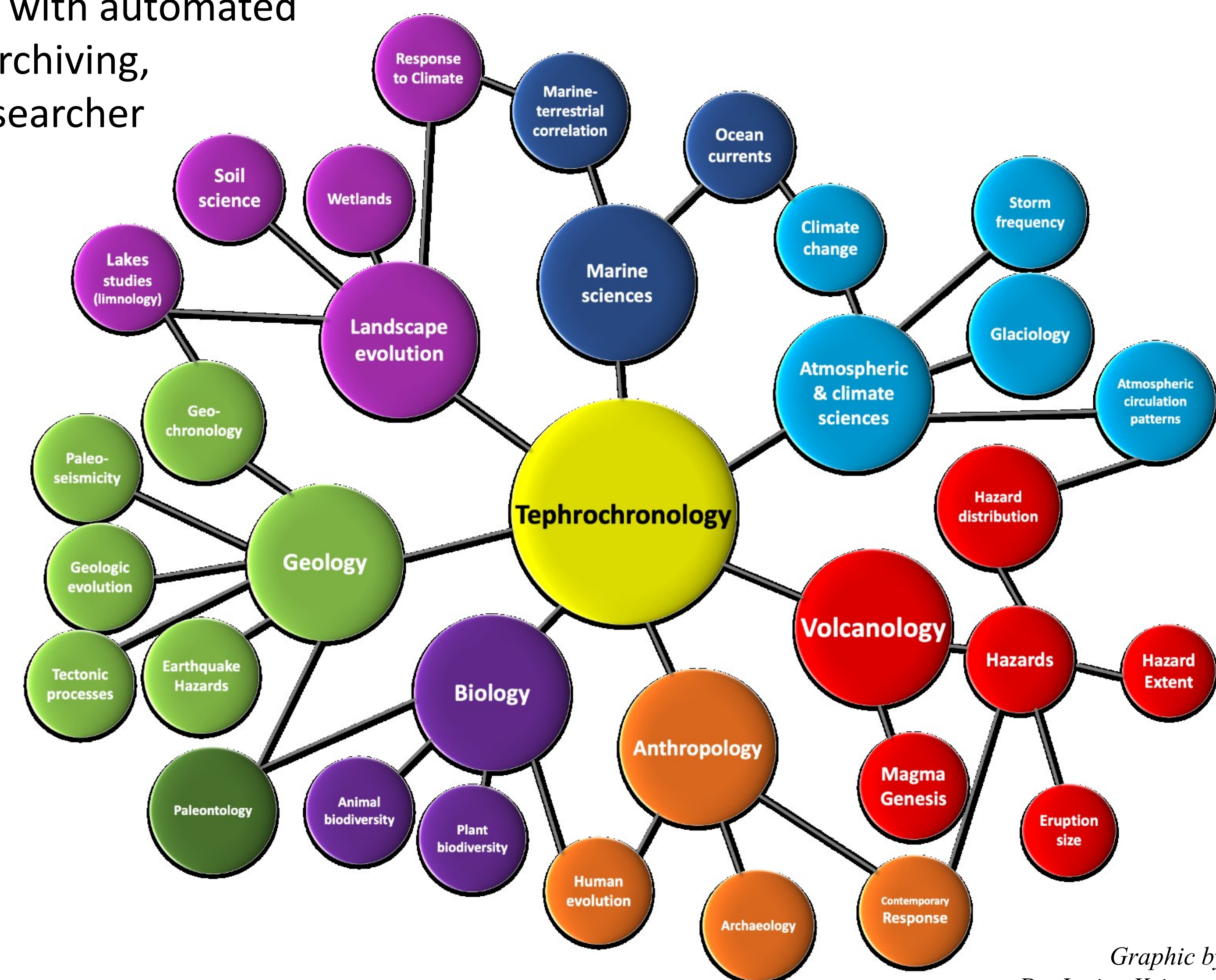
Linked data tables will cover:

- Projects, researchers, funding, and publications
- Field locations, strata, and cores
- Samples, sample curation, and reference materials
- Sample preparation and resulting fractions, grain mounts, and thin sections
- Analytical instruments, methods, and sessions
- Geochemistry with full association of corresponding sample and reference material/calibration data (initial focus EPMA, LA-ICP-MS)
- Optical and electron microscopy images
- Grain size (sieve data and discrete measurements)
- Grain type (componentry, mineralogy) and grain density
- Volcanic source assignment using Smithsonian GVP IDs
- Data access rights and embargos
- ORCIDs, IGSNs, DOIs, grant IDs, etc. captured as available



Anticipated Sparrow-Tephra Applications and Benefits

- Digital-first information capture within laboratory workflows
- Reduced laboratory data fragmentation and improved documentation following research community standards
- Easier data management with automated pathways to repository archiving, thus relieving a major researcher pain point and obstacle to production of FAIR data
- Accelerated data accessibility for secondary users
- Broad coverage of data types addresses applications in sedimentology, mineralogy, petrology, and other disciplines which utilize similar data types



Graphic by
Dr. Janine Krippner

NSF EAR 1740694 Geochronology Frontier at the Laboratory-Cyberinformatics Interface

NSF ICER 1740669 EarthCube Integration: THROUGHPUT: Standards and Services for Community Curated Repositories