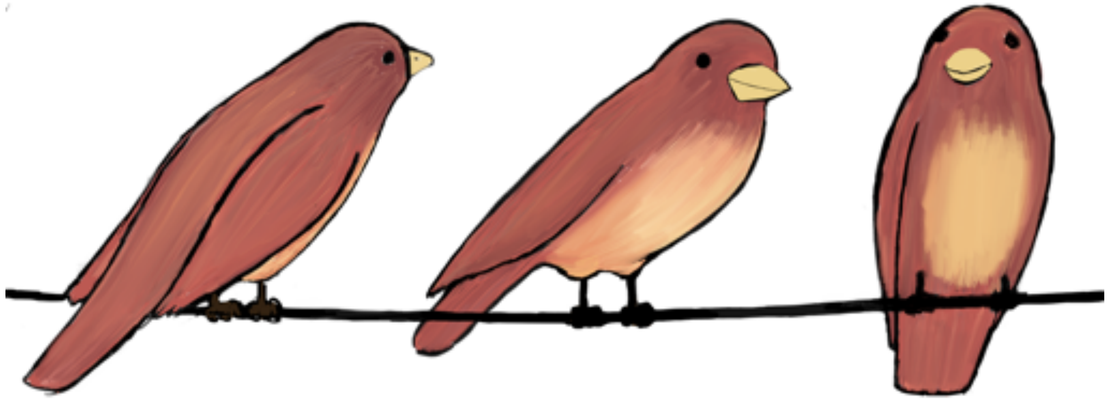
Sparrow: a data management system and cyberinfrastructure component targeted at geochronology laboratories

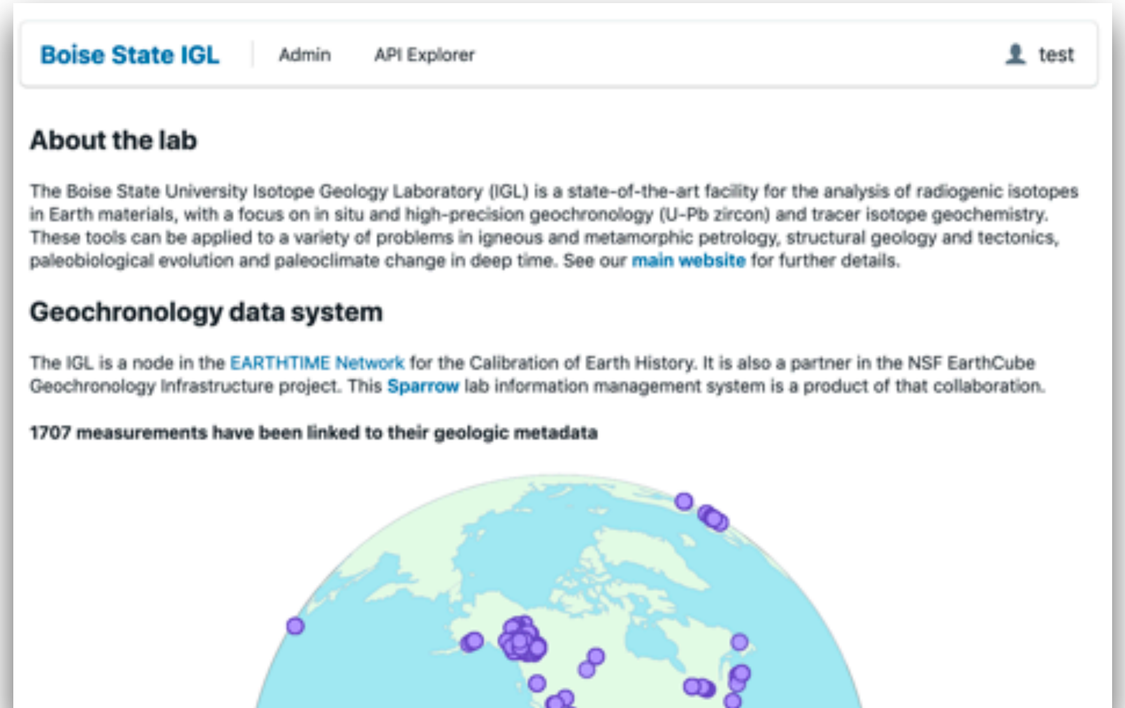
Daven P. Quinn[1], A.J. Schaen[2], K. Sundell[2], G. Gehrels[2], S. Goring[1], S.A. Marcott[1], S.R. Meyers[1], S.E. Peters[1], J.I. Ross[3], M. Schmitz[4], B.S. Singer[1], and J.W. Williams[1]

1. *University of Wisconsin – Madison*
2. *University of Arizona*
3. *New Mexico Bureau of Geology & Mineral Resources*
4. *Boise State University*



GSA 2019, Phoenix, Arizona

Supported by NSF EarthCube



https://github.com/EarthCubeGeochron/Sparrow
https://sparrow-data.org

Use cases for interoperable geochronology data

Reproducibility

- Adjustment and intercalibration
- Archival for re-analysis
- Audit by peer facilities

Audience

Expert workers within the geochronology community

Data requirements

Detailed analytical data





Aggregation and synthesis

- Discovery of relevant geochronology measurements
- Integration with other types of data (stratigraphic, ecological, etc.) for synthesis studies

Audience

Geoscientists in other domains

Data requirements

Interpreted ages, simplified uncertainties

Reconstructions of geologic time underpin synthesis studies across Earth science domains

...always ultimately anchored to detailed geochronology data

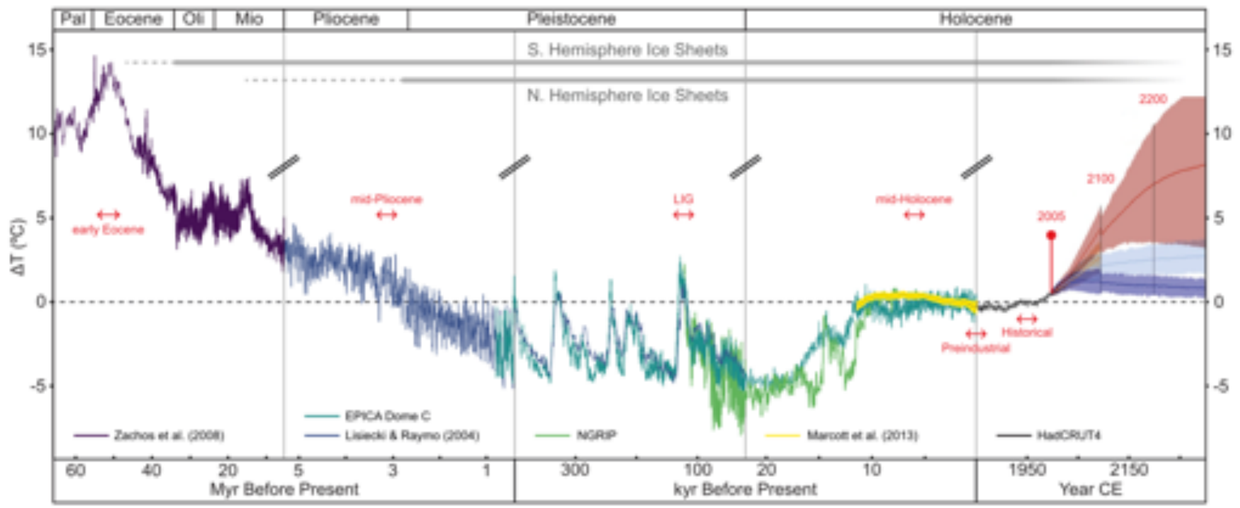
PNAS

Pliocene and Eocene provide best analogs for near-future climates

K. D. Burke[a,1], J. W. Williams[b], M. A. Chandler[c,d], A. M. Haywood[e], D. J. Lunt[f], and B. L. Otto-Bliesner[g]

[a]Nelson Institute for Environmental Studies, University of Wisconsin–Madison, Madison, WI 53706; [b]Department of Geography and Center for Climatic Research, University of Wisconsin–Madison, Madison, WI 53706; [c]Center for Climate Systems Research, Columbia University, New York, NY 10025; [d]Goddard Institute for Space Studies, National Aeronautics and Space Administration (NASA), New York, NY 10025; [e]School of Earth and Environment, University of Leeds, LS2 9JT Leeds, United Kingdom; [f]School of Geographical Sciences, University of Bristol, BS8 1SS Bristol, United Kingdom; and [g]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO 80305

Edited by Noah S. Diffenbaugh, Stanford University, Stanford, CA, and accepted by Editorial Board Member Robert E. Dickinson November 6, 2018 (received for review June 29, 2018)



Toward a Neoproterozoic composite carbon-isotope record

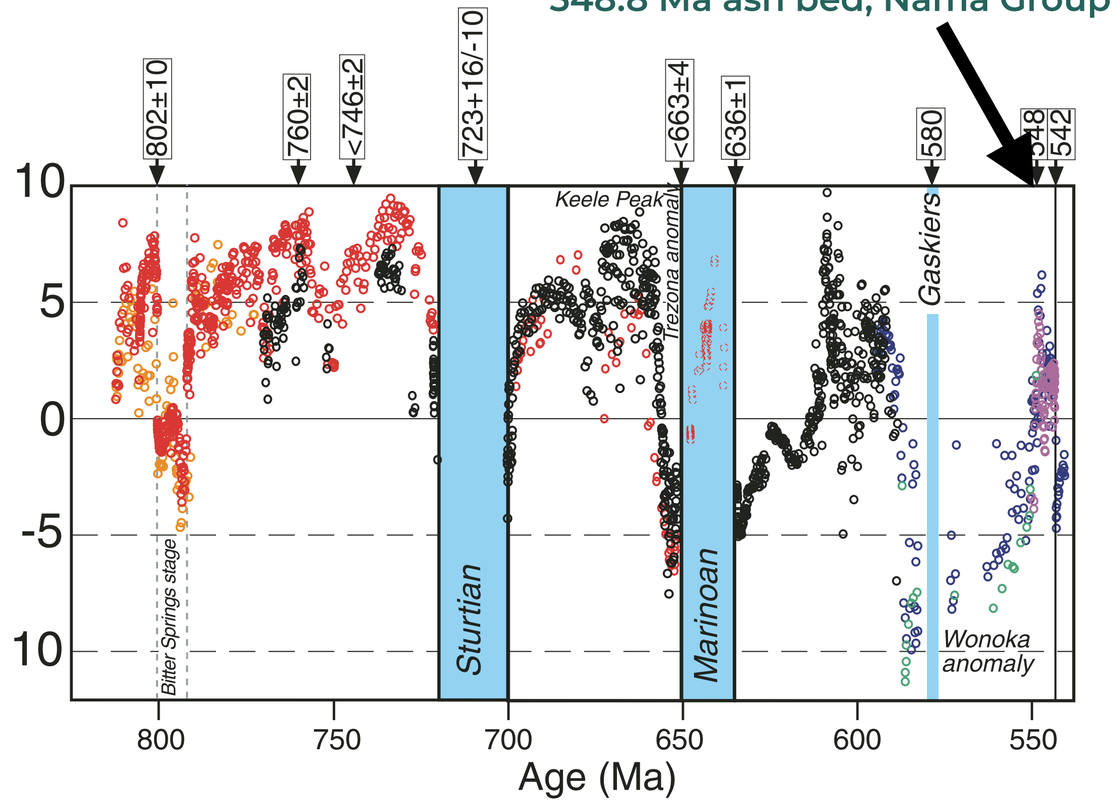
Galen P. Halverson[†]
Department of Earth and Planetary Sciences, Harvard Uni
and Department of Earth, Atmospheric, and Planetary Sci
bridge, Massachusetts 02139, USA

Paul F. Hoffman
Daniel P. Schrag
Adam C. Maloof[‡]
Department of Earth and Planetary Sciences, Harvard Univ

A. Hugh N. Rice
Department of Geological Sciences, University of Vienna, A



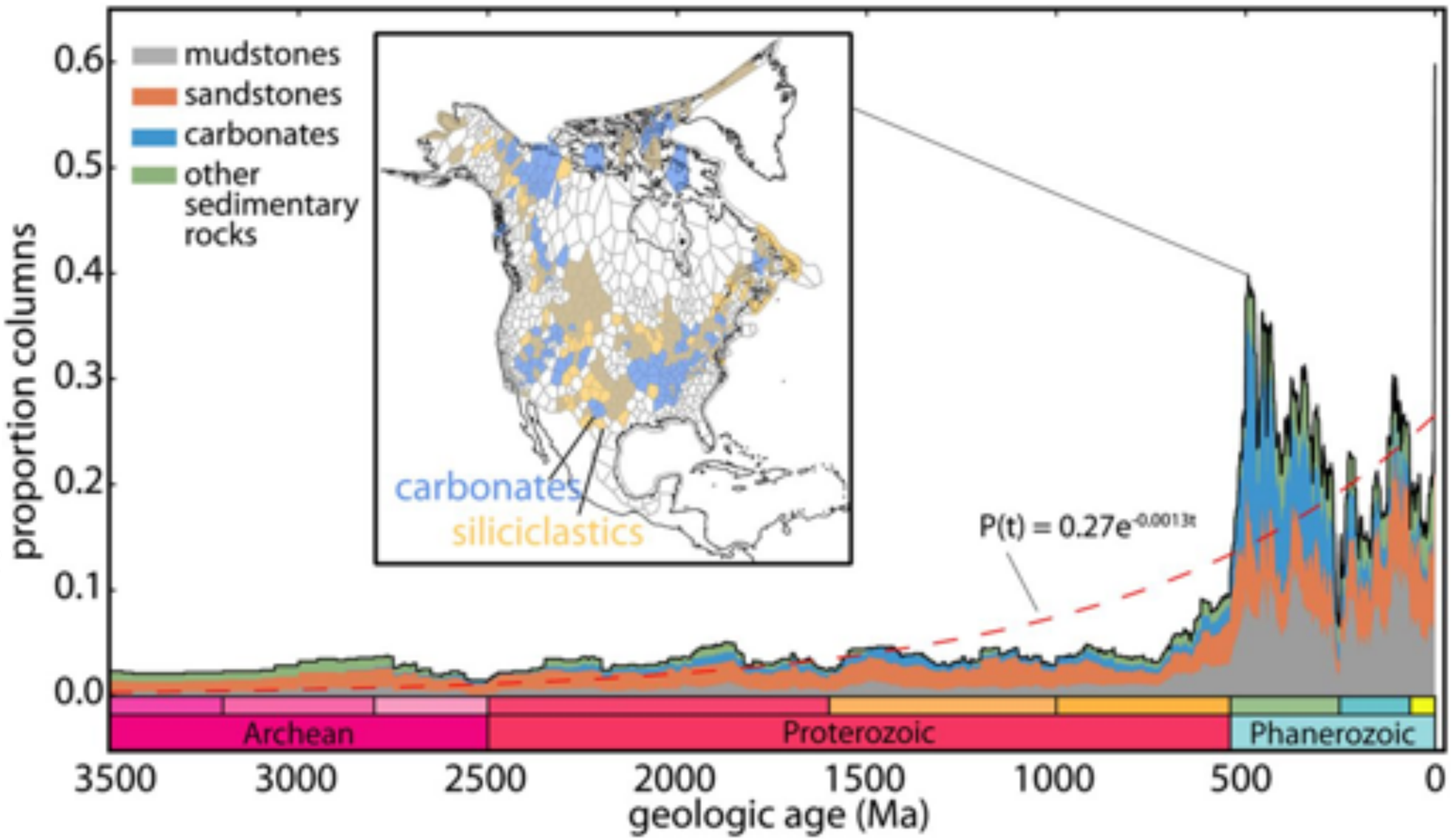
548.8 Ma ash bed, Nama Group



Digital Earth models — new targets for geochronology synthesis

Macrostrat

- Global time-stratigraphic model
- Incorporates a model of *relative* geologic time
- Allows synthesis across the entire stratigraphic record



Husson and Peters, 2018

Can we calibrate this model en-masse against geochronology data?

Digital Earth models — new targets for geochronology synthesis

Same goal in multiple communities!

Macrostrat

Global stratigraphy

NEOTOMADB.ORG

Paleoecology

PaleoBioDB

Paleobiology

...and others

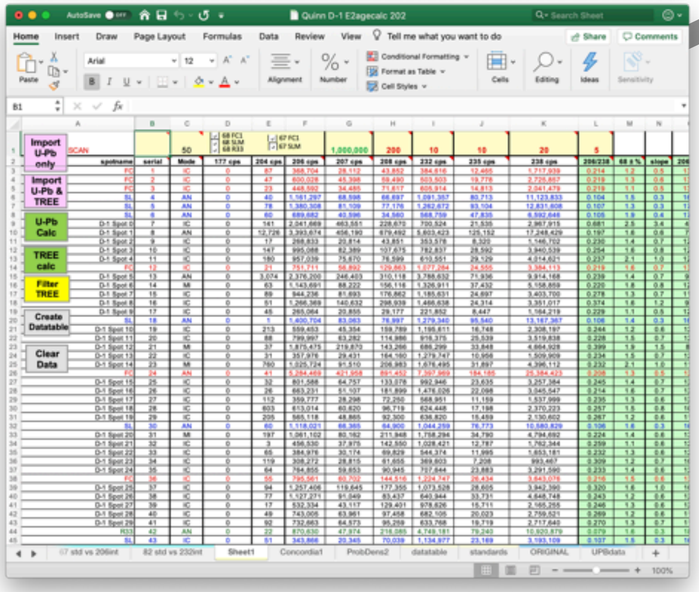
Can we calibrate this model en-masse against geochronology data?

Requirements for digital synthesis

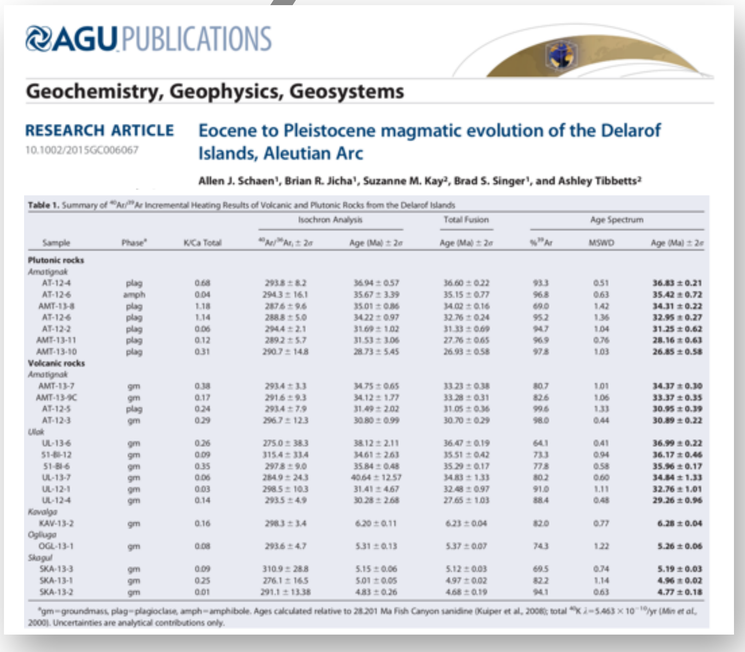


Geological context

Machine-readable
Cross-referenced
Available for synthesis



Analytical data



AGU PUBLICATIONS

Geochemistry, Geophysics, Geosystems

RESEARCH ARTICLE
10.1002/2015GC006067

Eocene to Pleistocene magmatic evolution of the Delarof Islands, Aleutian Arc

Allen J. Schaen[1], Brian R. Jicha[1], Suzanne M. Kay[2], Brad S. Singer[1], and Ashley Tibbetts[2]

Table 1. Summary of $^{40}Ar/^{39}Ar$ Incremental Heating Results of Volcanic and Plutonic Rocks from the Delarof Islands

Sample	Phase[a]	K/Ca Total	Isochron Analysis		Total Fusion	Age Spectrum		
			$^{40}Ar/^{36}Ar, \pm 2\sigma$	Age (Ma) $\pm 2\sigma$	Age (Ma) $\pm 2\sigma$	$\%^{39}Ar$	MSWD	Age (Ma) $\pm 2\sigma$
Plutonic rocks								
Amatignak								
AT-12-4	plag	0.68	293.8 ± 8.2	36.94 ± 0.57	36.60 ± 0.22	93.3	0.51	**36.83 ± 0.21**
AT-12-6	amph	0.04	294.3 ± 16.1	35.67 ± 3.39	35.15 ± 0.77	96.8	0.63	**35.42 ± 0.72**
AMT-13-8	plag	1.18	287.6 ± 9.6	35.01 ± 0.86	34.02 ± 0.16	69.0	1.42	**34.31 ± 0.22**
AT-12-6	plag	1.14	288.8 ± 5.0	34.22 ± 0.97	32.76 ± 0.24	95.2	1.36	**32.95 ± 0.27**
AT-12-2	plag	0.06	294.4 ± 2.1	31.69 ± 1.02	31.33 ± 0.69	94.7	1.04	**31.25 ± 0.62**
AMT-13-11	plag	0.12	289.2 ± 5.7	31.53 ± 3.06	27.76 ± 0.65	96.9	0.76	**28.16 ± 0.63**
AMT-13-10	plag	0.31	290.7 ± 14.8	28.73 ± 5.45	26.93 ± 0.58	97.8	1.03	**26.85 ± 0.58**
Volcanic rocks								
Amatignak								
AMT-13-7	gm	0.38	293.4 ± 3.3	34.75 ± 0.65	33.23 ± 0.38	80.7	1.01	**34.37 ± 0.30**
AMT-13-9C	gm	0.17	291.6 ± 9.3	34.12 ± 1.77	33.28 ± 0.31	82.6	1.06	**33.37 ± 0.35**
AT-12-5	plag	0.24	293.4 ± 7.9	31.49 ± 2.02	31.05 ± 0.36	99.6	1.33	**30.95 ± 0.39**
AT-12-3	gm	0.29	296.7 ± 12.3	30.80 ± 0.99	30.70 ± 0.29	98.0	0.44	**30.89 ± 0.22**
Ulak								
UL-13-6	gm	0.26	275.0 ± 38.3	38.12 ± 2.11	36.47 ± 0.19	64.1	0.41	**36.99 ± 0.22**
51-BI-12	gm	0.09	315.4 ± 33.4	34.61 ± 2.63	35.51 ± 0.42	73.3	0.94	**36.17 ± 0.46**
51-BI-6	gm	0.35	297.8 ± 9.0	35.84 ± 0.48	35.29 ± 0.17	77.8	0.58	**35.96 ± 0.17**
UL-13-7	gm	0.06	284.9 ± 24.3	40.64 ± 12.57	34.83 ± 1.33	80.2	0.60	**34.84 ± 1.33**
UL-12-1	gm	0.03	298.5 ± 10.3	31.41 ± 4.67	32.48 ± 0.97	91.0	1.11	**32.76 ± 1.01**
UL-12-4	gm	0.14	293.5 ± 4.9	30.28 ± 2.68	27.65 ± 1.03	88.4	0.48	**29.26 ± 0.96**
Kovalga								
KAV-13-2	gm	0.16	298.3 ± 3.4	6.20 ± 0.11	6.23 ± 0.04	82.0	0.77	**6.28 ± 0.04**
Ogliuga								
OGL-13-1	gm	0.08	293.6 ± 4.7	5.31 ± 0.13	5.37 ± 0.07	74.3	1.22	**5.26 ± 0.06**
Skagul								
SKA-13-3	gm	0.09	310.9 ± 28.8	5.15 ± 0.06	5.12 ± 0.03	69.5	0.74	**5.19 ± 0.03**
SKA-13-1	gm	0.25	276.1 ± 16.5	5.01 ± 0.05	4.97 ± 0.02	82.2	1.14	**4.96 ± 0.02**
SKA-13-2	gm	0.01	291.1 ± 13.38	4.83 ± 0.26	4.68 ± 0.19	94.1	0.63	**4.77 ± 0.18**

[a]gm = groundmass, plag = plagioclase, amph = amphibole. Ages calculated relative to 28.201 Ma Fish Canyon sanidine (Kuiper et al., 2008); total ^{40}K $\lambda = 5.463 \times 10^{-10}$/yr (Min et al., 2000). Uncertainties are analytical contributions only.

Published science result

https://sparrow-data.org

Geochronology data workflow
and standardization challenges

Rich contextual metadata

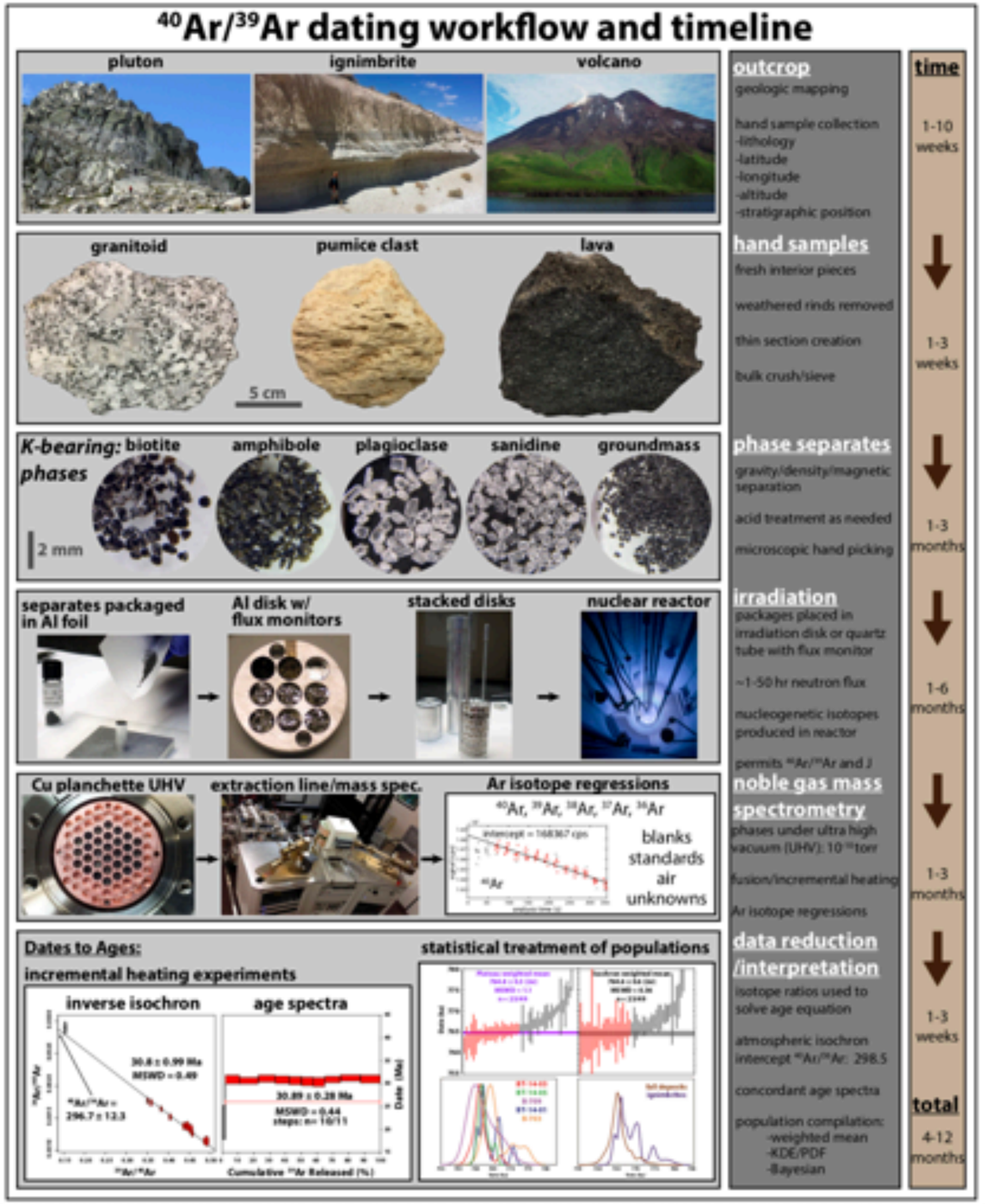
Often not tracked systematically

Analytical process data

Data structure differs across **isotope systems**, **analytical techniques**, *and* **instruments**

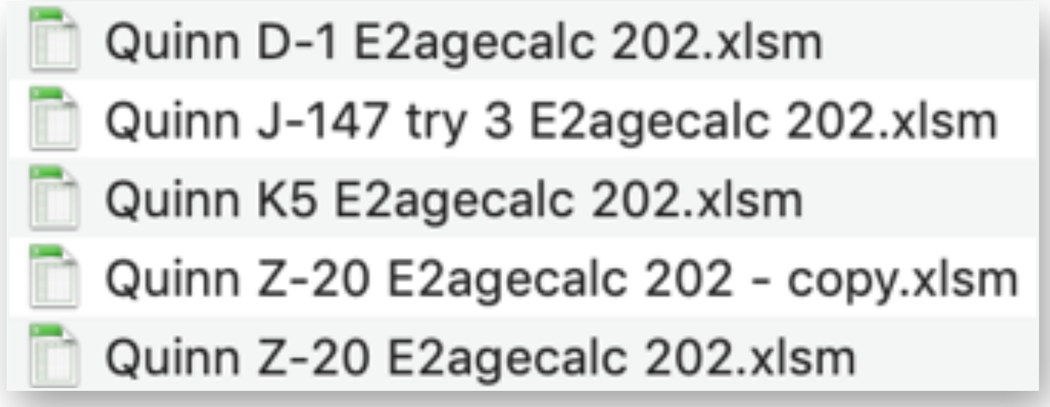
Interpreted age data

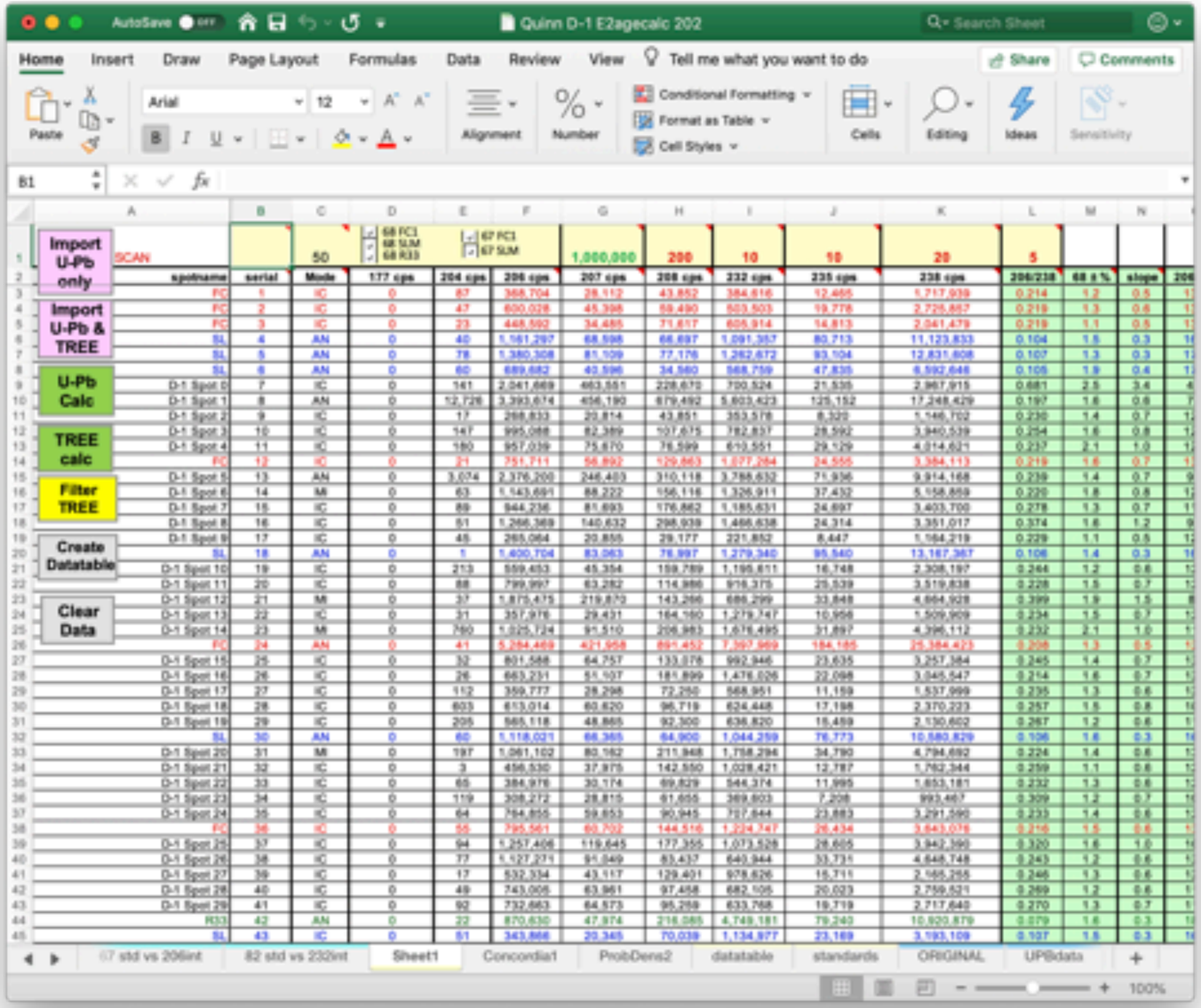
Encodes **expert intervention** *that can be challenging to justify and reproduce*



Geochronology data
management practices

- Analytical data files (e.g. spreadsheets, XML files)
- Ad-hoc file and revision management
- Lab- or method-specific formats
- Not tied to contextual information





A typical E2AgeCalc workbook for detrital zircon geochronology

Community archives — manual process limits usage



Forward-looking focus

- Community standards
- Reproducibility

Centralized model

- Labs must actively upload data
- Dependent on ongoing support

Usability challenges

- Time-consuming manual uploads
- Preference for specific data-reduction software
- Difficult for non-geochronologists to navigate

EAR-1740694: Geochronology Frontier at the Laboratory–Cyberinformatics Interface

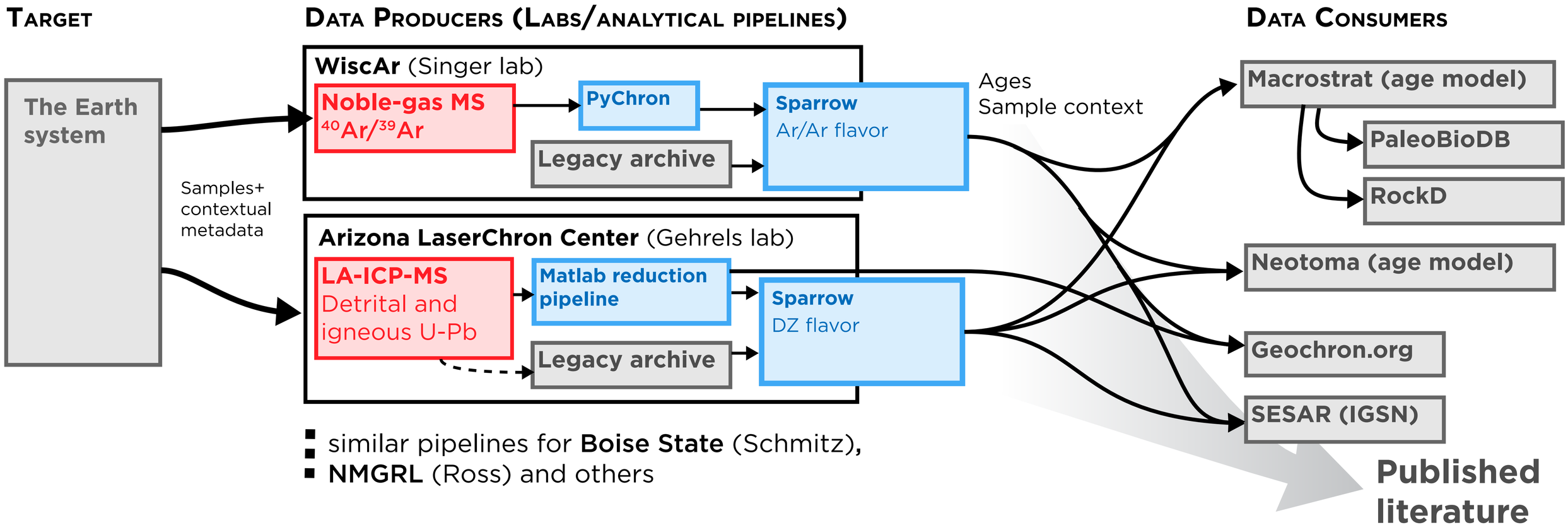
Goal: build a community data infrastructure, emphasizing automation and within-lab workflow needs.

Consortium of high-throughput geochronology laboratories

https://sparrow-data.org

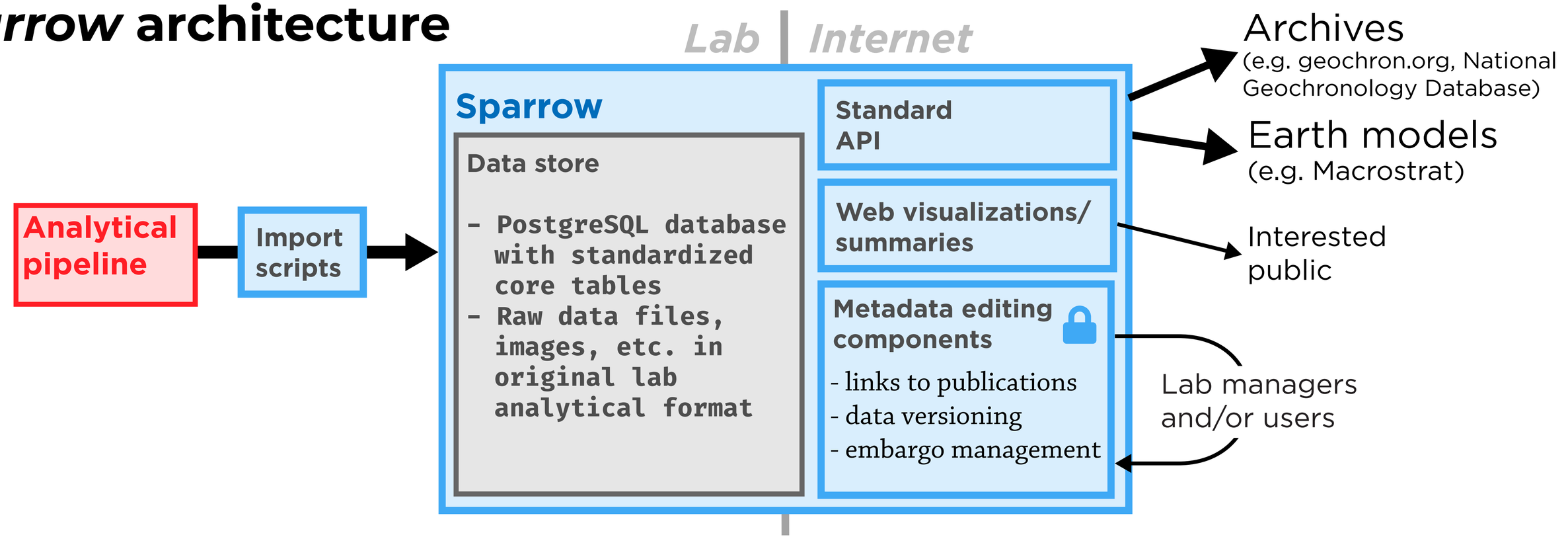
Sparrow: a software interface between a lab and the world

- **Each lab maintains control of its own internal process**
- **Harmonized endpoint enables easy data sharing**



https://sparrow-data.org

Sparrow architecture



Structured data store

- Analytical data in original file format
- Structured subset of data
- Geologic context and project metadata
- Change history/audit trail

Standardized interface to lab data

- Data-driven website
- Web management application
- Extensible set of data types

Sparrow: design principles

Extensible

- A framework with applications across geochemistry
- Customizable at every level (database, API, frontend)

Modular

- Based on standard, widely used tools
- Decoupled API, frontend, and database

Lightweight

- Small footprint at implementing labs
- Adds to (does not replace) current lab processes
- *Does not include or dictate* analytical pipeline
- Abstracts complexity of underlying software (e.g. Docker, PostgreSQL)

Technology stack



docker

Backend

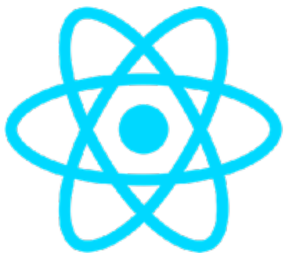


PostgreSQL



python

Frontend



React

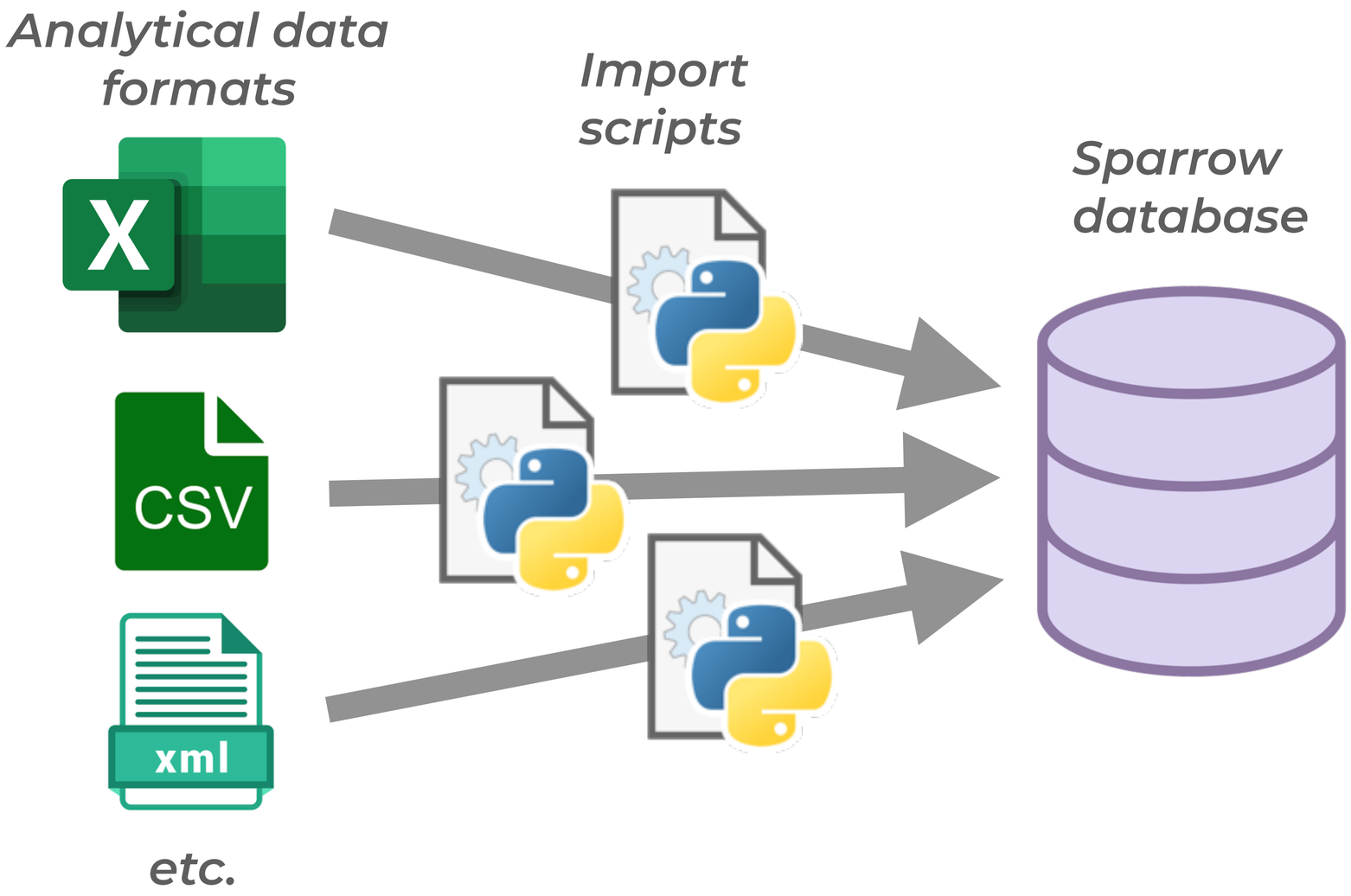


BlueprintJS

Connecting lab data archives to *Sparrow*

Import scripts track data files and build a structured representation.

These scripts will vary by data format and lab requirements.

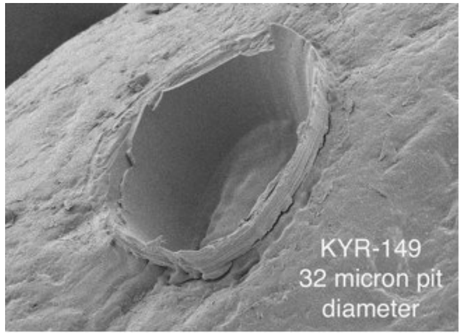


Hierarchical model for analytical measurements

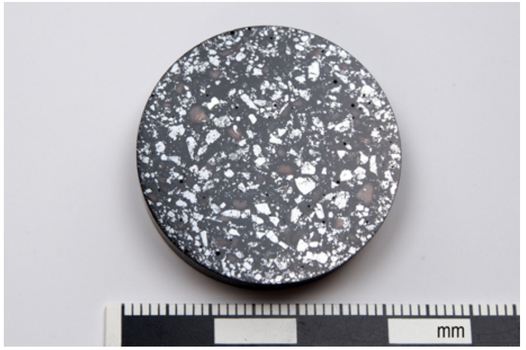
Single measured values

H2O content:	5.34 ±0.534(1s) %
De:	21.6 ±0.900(1s) Gy
OD:	24.0 %
luminescence age:	5.36 ±0.342(1s) ka
total dose rate:	4.03 ±0.195(1s) Gy/ka

Single instance of space- or time-resolved data collection



Grouped measurements in a single instrument run



The geologic material that was originally collected for analysis



Datum → **Analysis** → **Session** → **Sample**

Location

Entity (e.g. geologic unit)

Additional models for measurement context

Project

Publication

Researcher

Not just for geochronology — adaptable to all types of geochemical data!

https://sparrow-data.org

Lab internal data types → Community structured vocabularies

Fields extracted directly from data files

Ratio of 206Pb*/238U	Arizona LaserChron Center
Ratio of 207Pb*/235U*	Arizona LaserChron Center
206Pb*/207Pb* age	Arizona LaserChron Center
206Pb*/238U* age	Arizona LaserChron Center
207Pb*/235U* age	Arizona LaserChron Center
Concordance between U/Pb age systems	Arizona LaserChron Center
Error correlation	Arizona LaserChron Center

Analytical vocabulary

ratio of uranium isotope U238 to lead isotope Pb204	EarthChem
ratio of uranium isotope U238 to lead isotope Pb206	EarthChem
ratio of uranium isotope U238 to thorium isotope Th230	EarthChem
ratio of uranium isotope U238 to thorium isotope Th230, re...	EarthChem
ratio of uranium isotope U238 to thorium isotope Th232	EarthChem
ratio of lead isotope Pb207 to uranium isotope U235	EarthChem
ulvospinel	EarthChem

EarthChem

Internal structure of measurement (for reproducibility and reanalysis by geochronologists)

Interpretive vocabulary

detrital grain best age	Sparrow
igneous formation age	Sparrow
cooling age	Sparrow
exposure age	Sparrow

Interpreted age, meaning, and context (for use outside of geochronology)

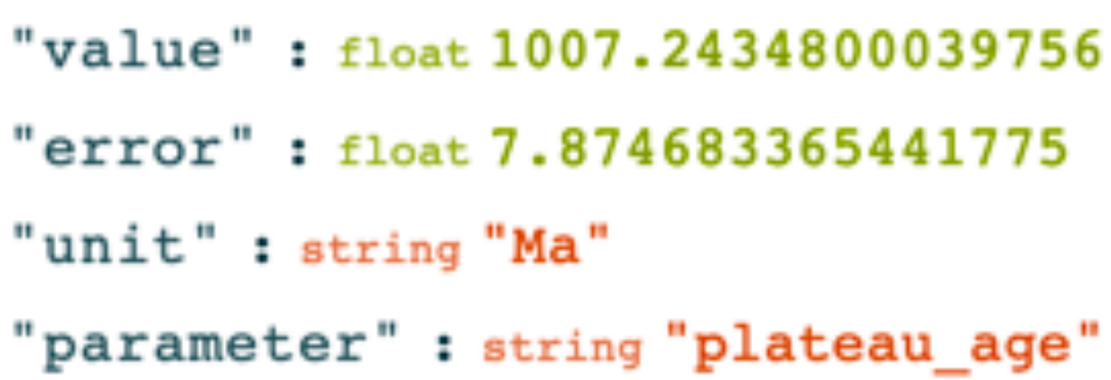
- Labs start with **any format they want**
- Multiple outputs based on requester requirements (e.g. peer labs vs. users)
- Formats can evolve independently based on lab and community needs
- Balances ***flexibility*** and ***interoperability***

Public-facing API

“application programming interface”

https://sparrow.boisestate.edu/api/v1/datum

Machine-readable data



*Consumers **pull** data from labs*

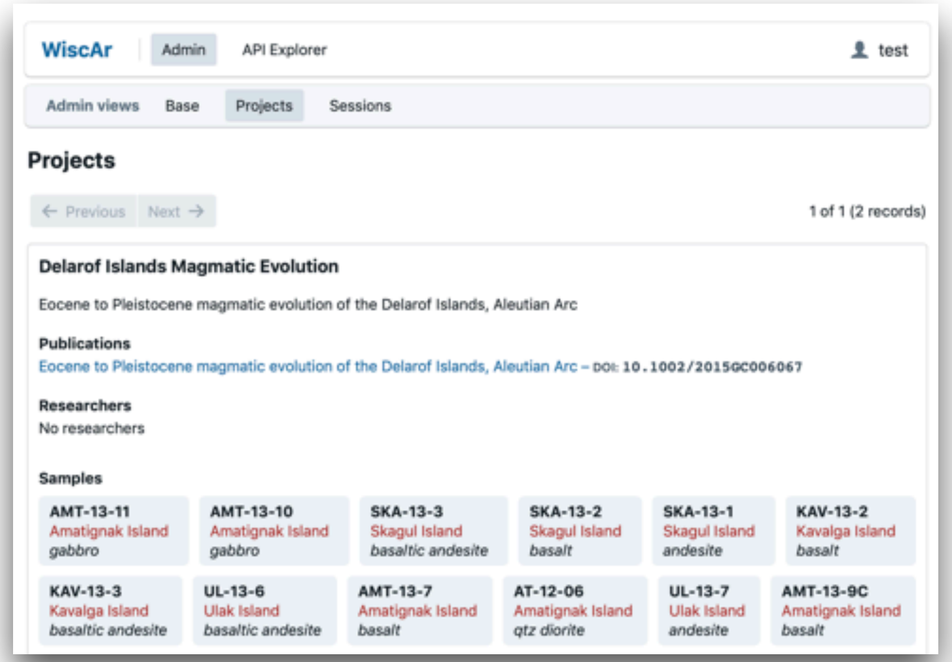
Geochron

Macrostrat

etc.

Web management platform

- **Set embargo policy**
- **Manage project- and sample-level metadata**
- Access detailed analytical data and original data files
- Import and export data files
- Search and link publications
- Request and link IGSNs *(optional; planned)*



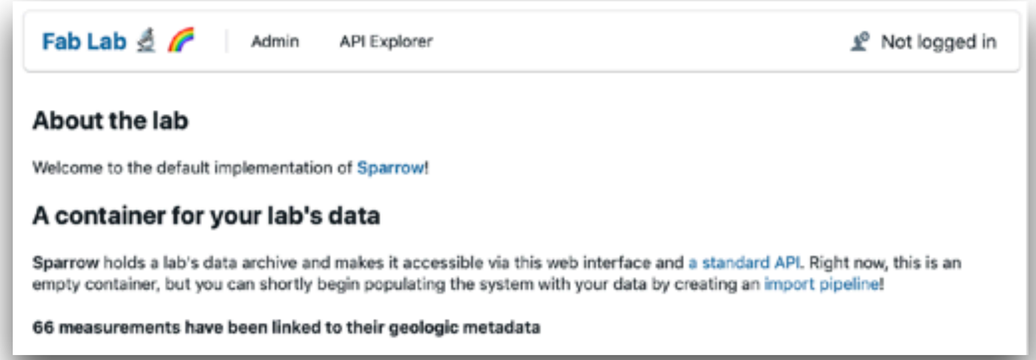


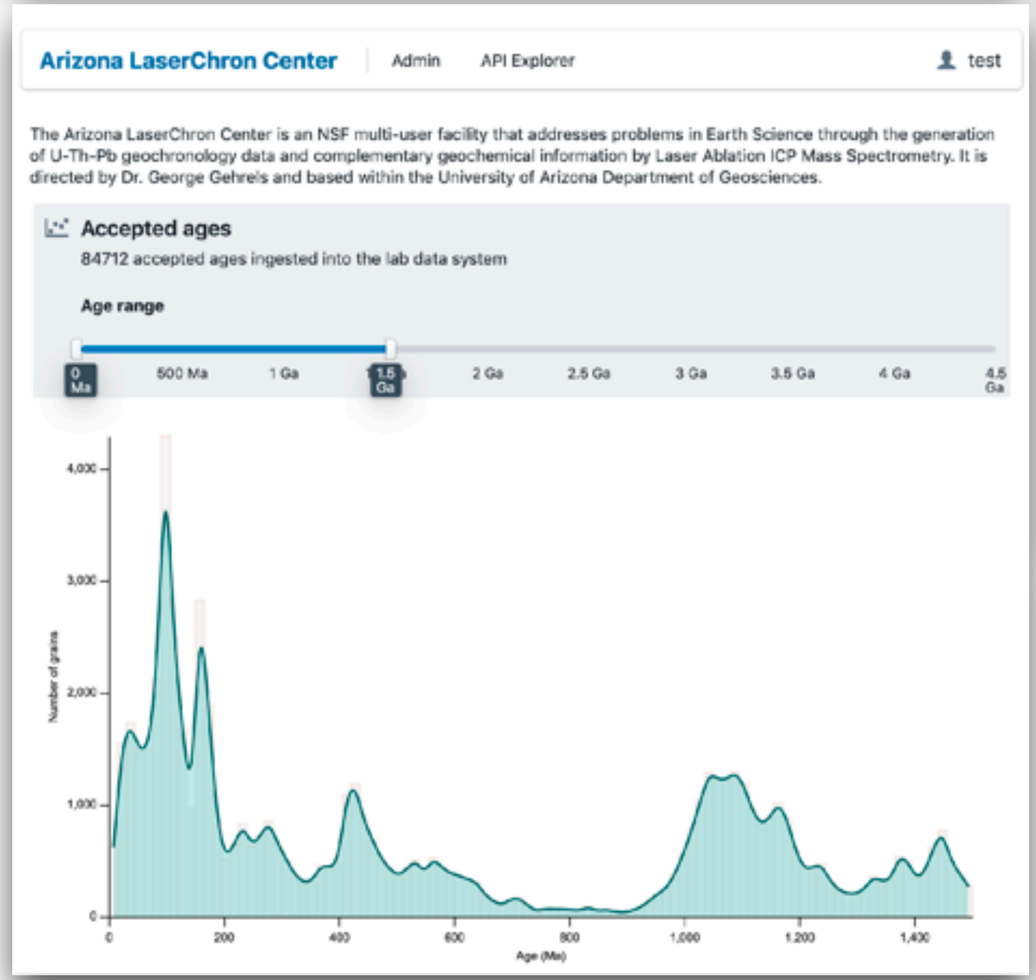


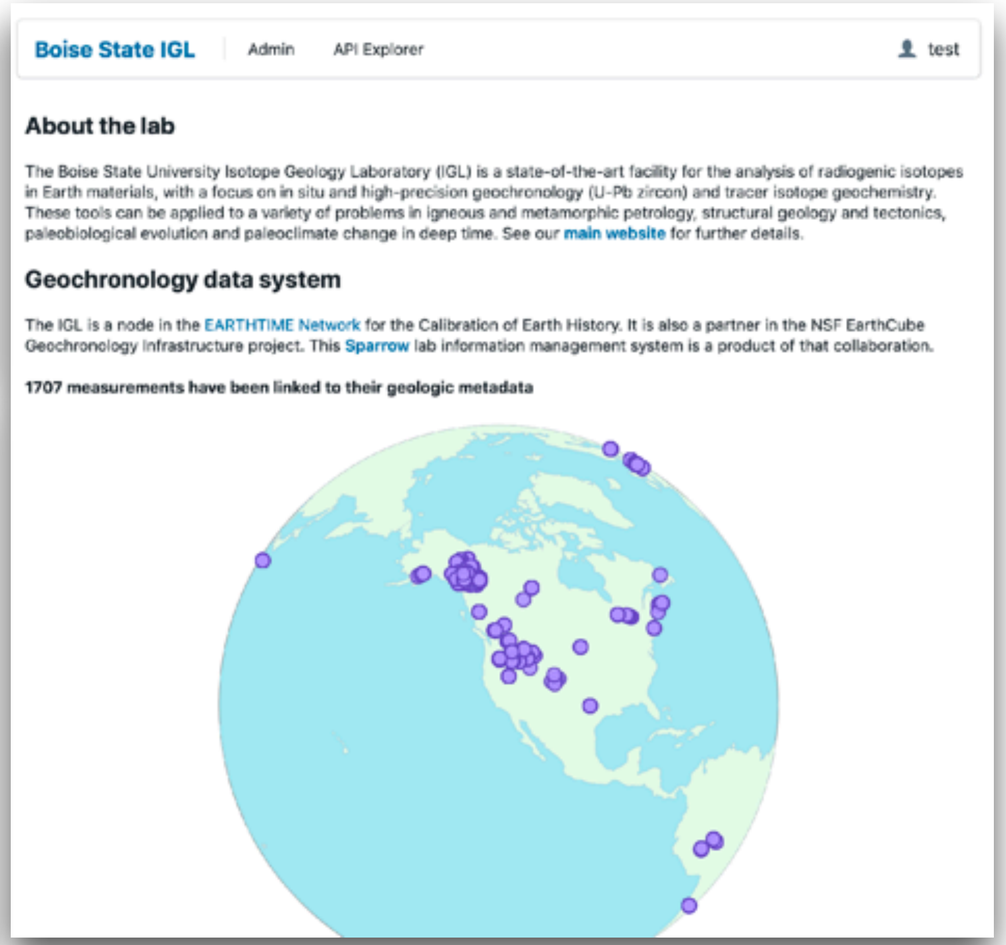
https://sparrow-data.org

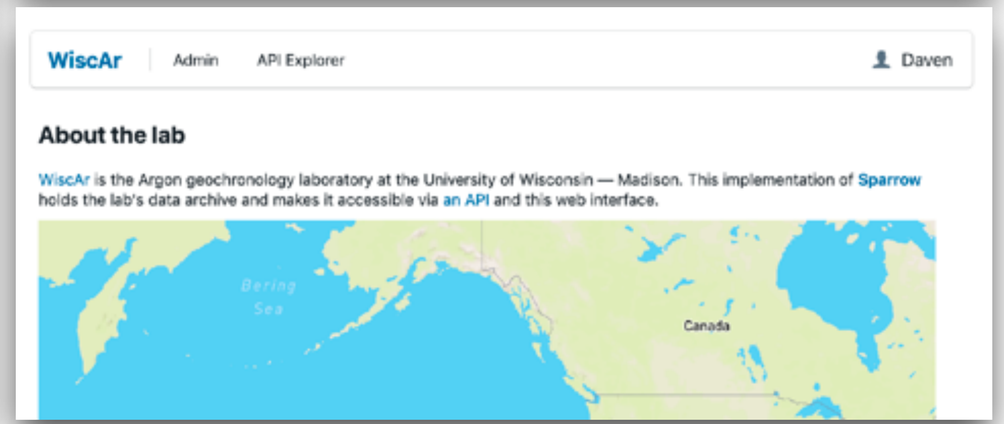
Data-driven website

- Lab information
- Summary statistics
- Publications
- Map of analyses









https://sparrow-data.org

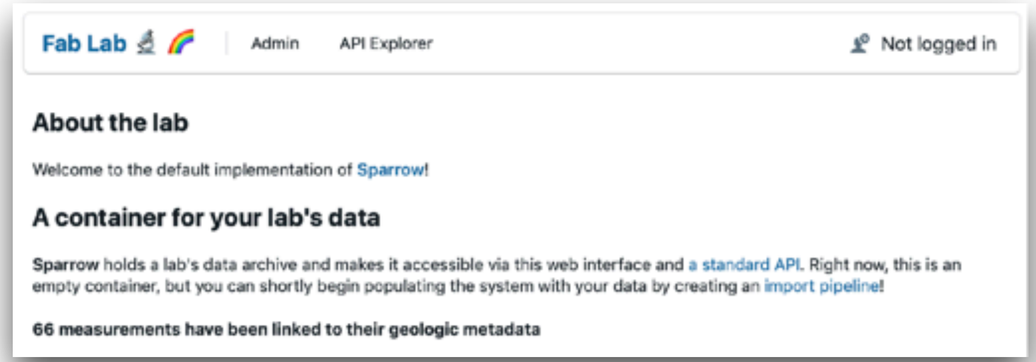
Extensible for different needs

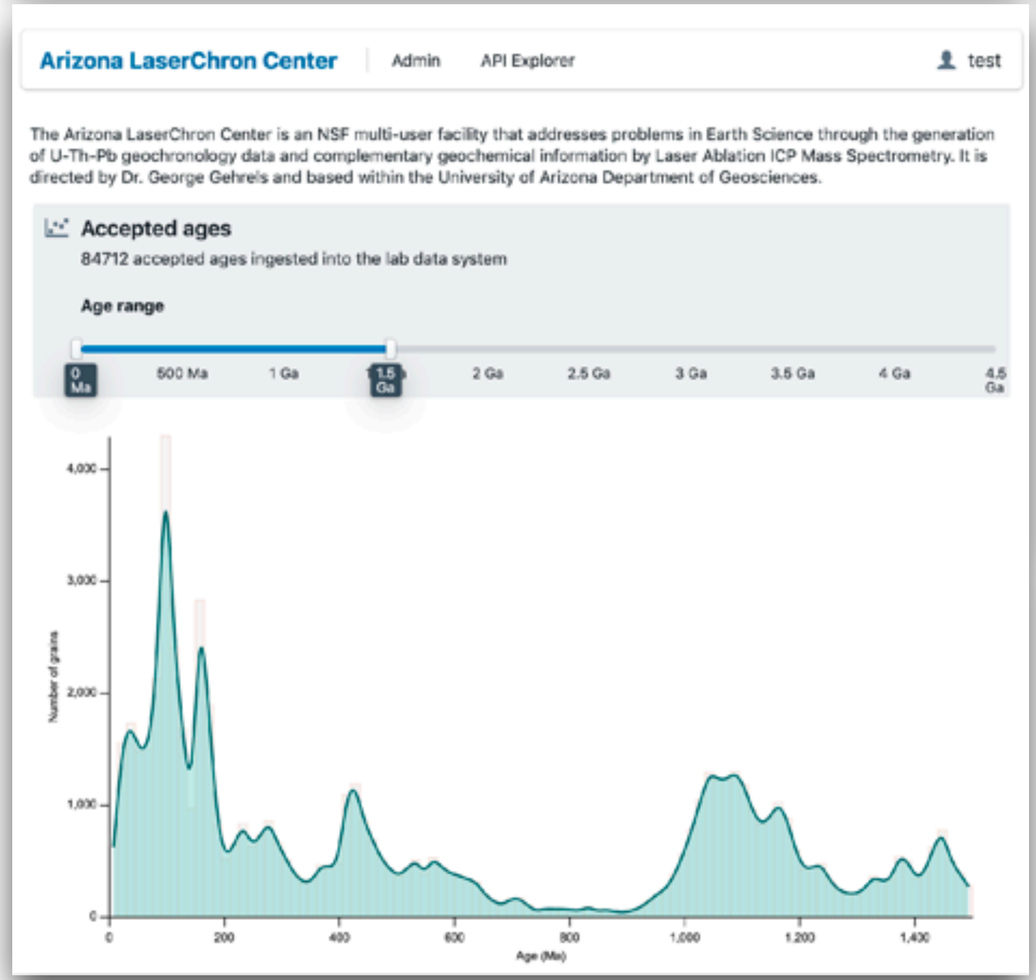
Sparrow provides core functionality that should be useful to most labs.

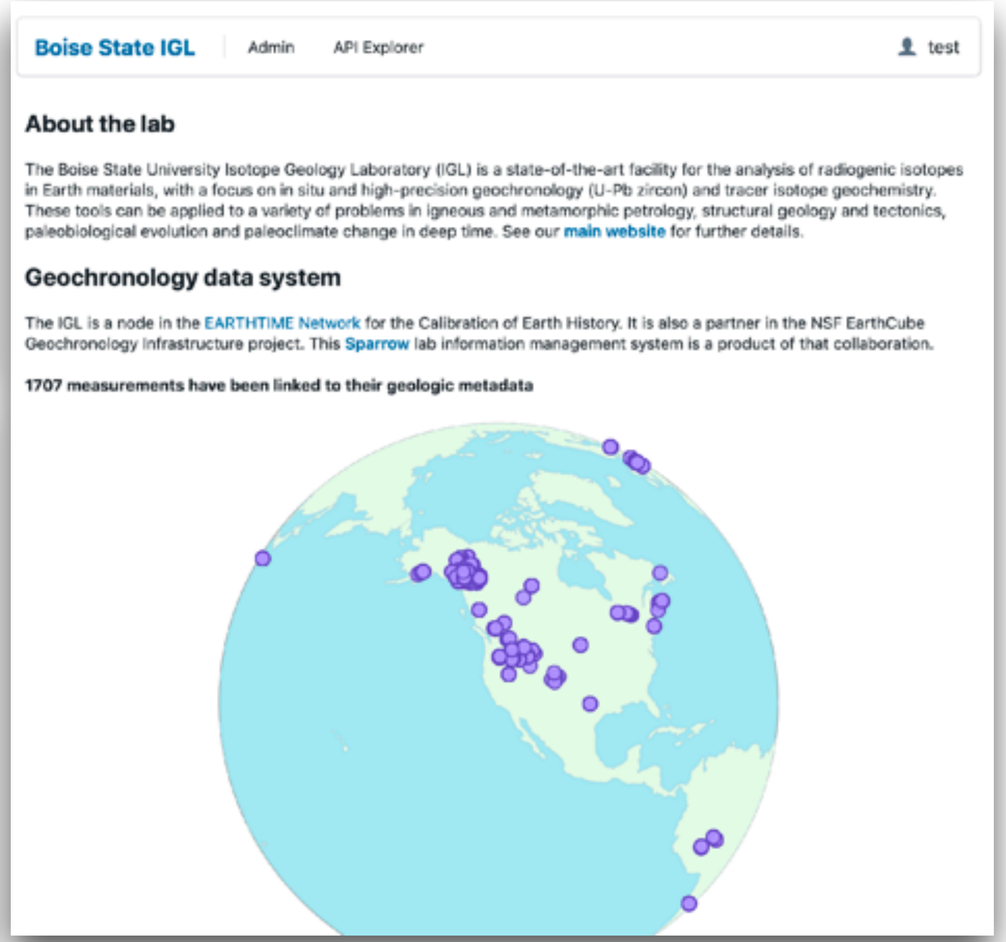
All parts of application are customizable (database, API, and frontend)

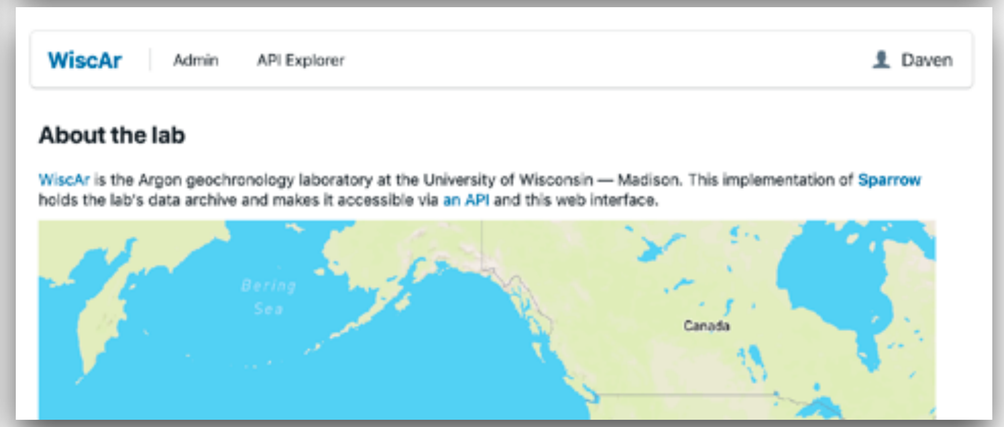
Model:

WordPress *for geochemistry data!*





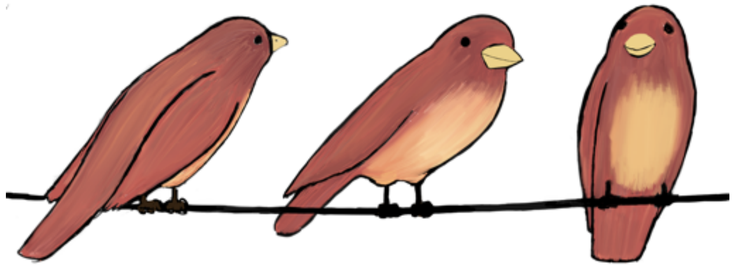




https://sparrow-data.org

Project status

Fall 2018–Sept. 2019



Core software development

- Interactive UI components, documentation, and design evaluation
- Tests for use with chemostratigraphy and optically-stimulated luminescence data

Community feedback

- EarthCube workshop in May 2019
- Beginning to work with other contributors at British Geologic Survey, elsewhere.

Deployments

Arizona LaserChron Center (U-Pb)
~1500 samples (80,000 zircon grains) imported (~10% of archive)

WiscAr (Ar-Ar)
~1200 samples ingested for MAP spectrometer, sample metadata for ~1400 samples discovered and linked

Boise State Isotope Geology Lab (U-Pb)
Entire metadata archive ingested from IGSN, sample import underway

UW Madison Cosmogenic Nuclide Lab
Archive ingested, ongoing work to create a data management web interface

New Mexico Geochron. Research Lab (Ar-Ar)
Integration with PyChron system complete.

Test instances underway for British Geological Survey, Desert Research Institute, others.

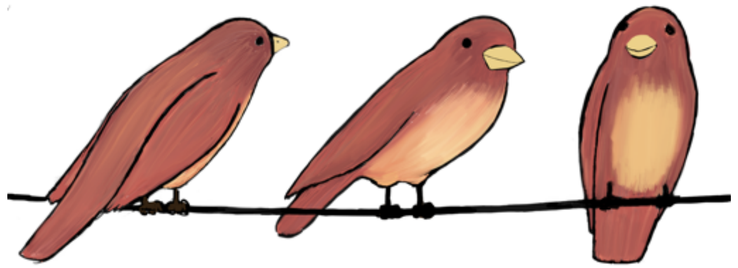
Thank you!

https://sparrow-data.org
Project information, documentation, and examples

https://github.com/EarthCubeGeochron/Sparrow
Source code for core software



Sparrow



EARTHCUBE
TRANSFORMING GEOSCIENCES RESEARCH

Questions?

Data archival software for geochemical labs

- Organizes lab analytical data in a flexible, modern, industry-standard database application
- Augments current lab analytical processes
- Tools for curating and showcasing metadata

Interfaces with data users and community archives through a standard API

Does not dictate specific data-analysis tooling

Keeps labs in control of the canonical version of data

https://sparrow-data.org
Project information, documentation, and examples

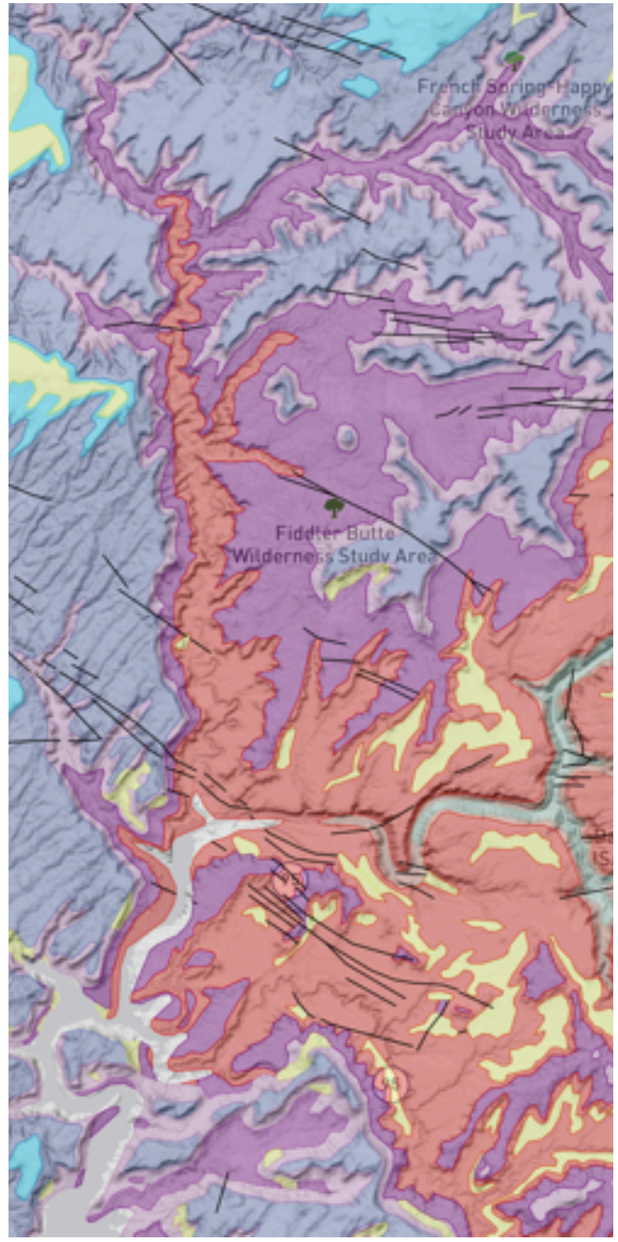
https://github.com/EarthCubeGeochron/Sparrow
Source code for core software

Geochronology synthesis with digital Earth models

Macrostrat

- Community-sourced model for global stratigraphy
- **Digital model of relative geologic time**

Period /Epoch	Groups Formations & Members			Thick-ness	Rock Type
Neogene	Ox Valley Tuff			12-15	
	Hiko Tuff			0-300	
	Racer Canyon Tuff			0	
	Page Ranch			0-200	
	Rencher			0-200	
	Quichapa			0-500	
	Isom			0-23	
Paleogene	Needles Range			0	
	Needles Range	Escalante Desert		0	
	Claron			0-100	
				0-300	
	Pine Hollow			0-100	
	Canaan Peak			0-200	
K Cretaceous	Kaiparowits			0-1000	
	Iron Springs			0-1200	
Jurassic (J)	San Rafael	Entrada Sandstone		0-100	
		Carmel	Winsor	60-100	
			Paria River	20-50	
			Crystal Creek	0-60	
	Glen Canyon	Navajo Sandstone		0	
	San Rafael	Carmel	Kolob Limestone	70-160	
	Glen Canyon	Navajo Sandstone		460-700	
		Kayenta		100-400	
		Moenave		90-200	
sic	Chinle	Petrified Forest		100-350	

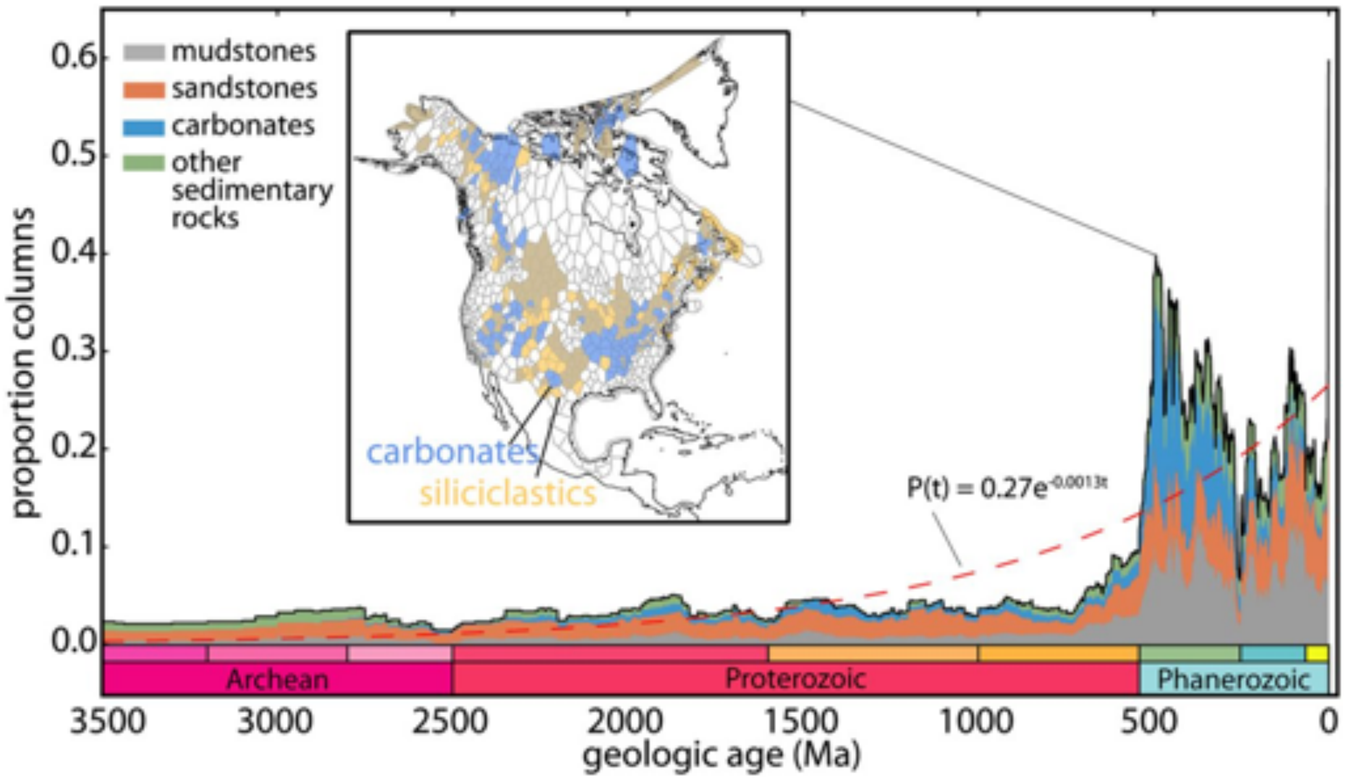


https://sparrow-data.org

Geochronology synthesis with digital Earth models

Macrostrat

- Community-sourced model for global stratigraphy
- Digital model of *relative* geologic time
- **Allows synthesis across the entire stratigraphic record**

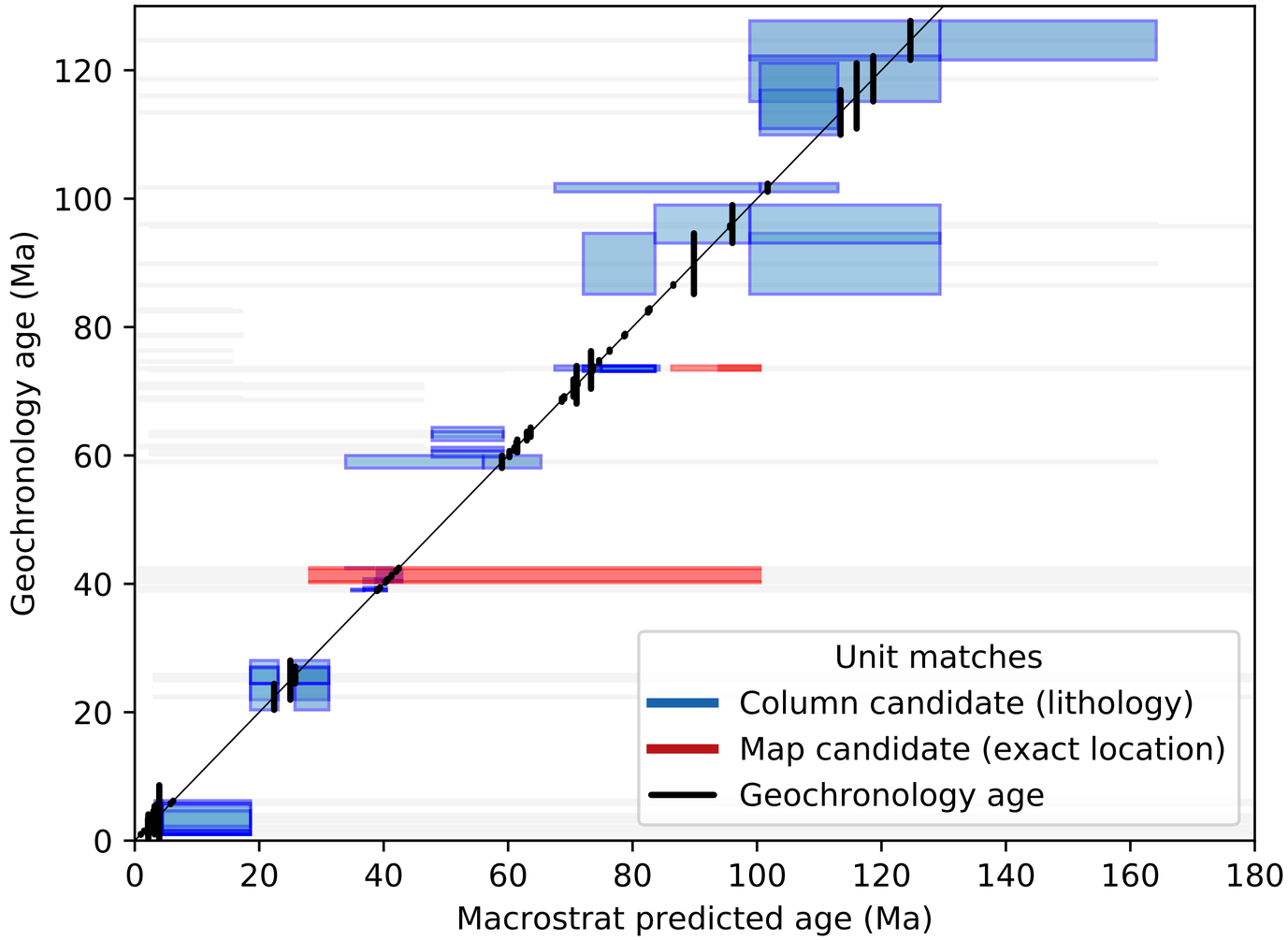


Husson and Peters, 2018

Geochronology synthesis with digital Earth models

Macrostrat

- Community-sourced model for global stratigraphy
- Digital model of *relative* geologic time
- Allows synthesis across the entire stratigraphic record
- **Goal: en-masse calibration against geochronology age data**



Preliminary model matching Macrostrat units from stratigraphic columns and geologic maps to geochronology age data

Geochronology synthesis with digital Earth models

Macrostrat

Global stratigraphy

Time-integrated archives and synthesis models in multiple communities

NEOTOMADB.ORG

Paleoecology

PaleoBioDB

Paleobiology

...and others

All gain from integration with different types of geochronology data