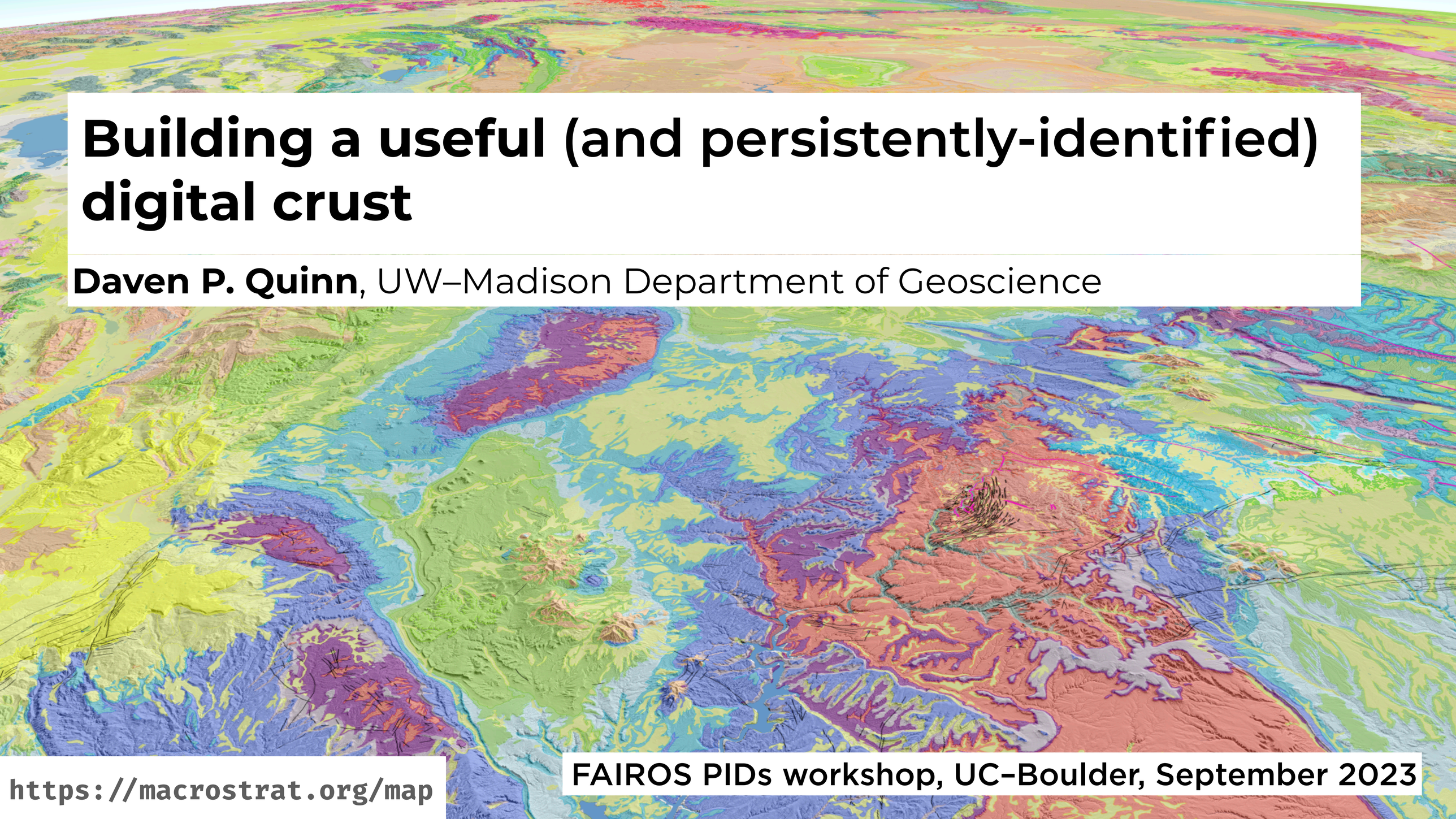
Building a useful (and persistently-identified) digital crust

Daven P. Quinn, UW–Madison Department of Geoscience

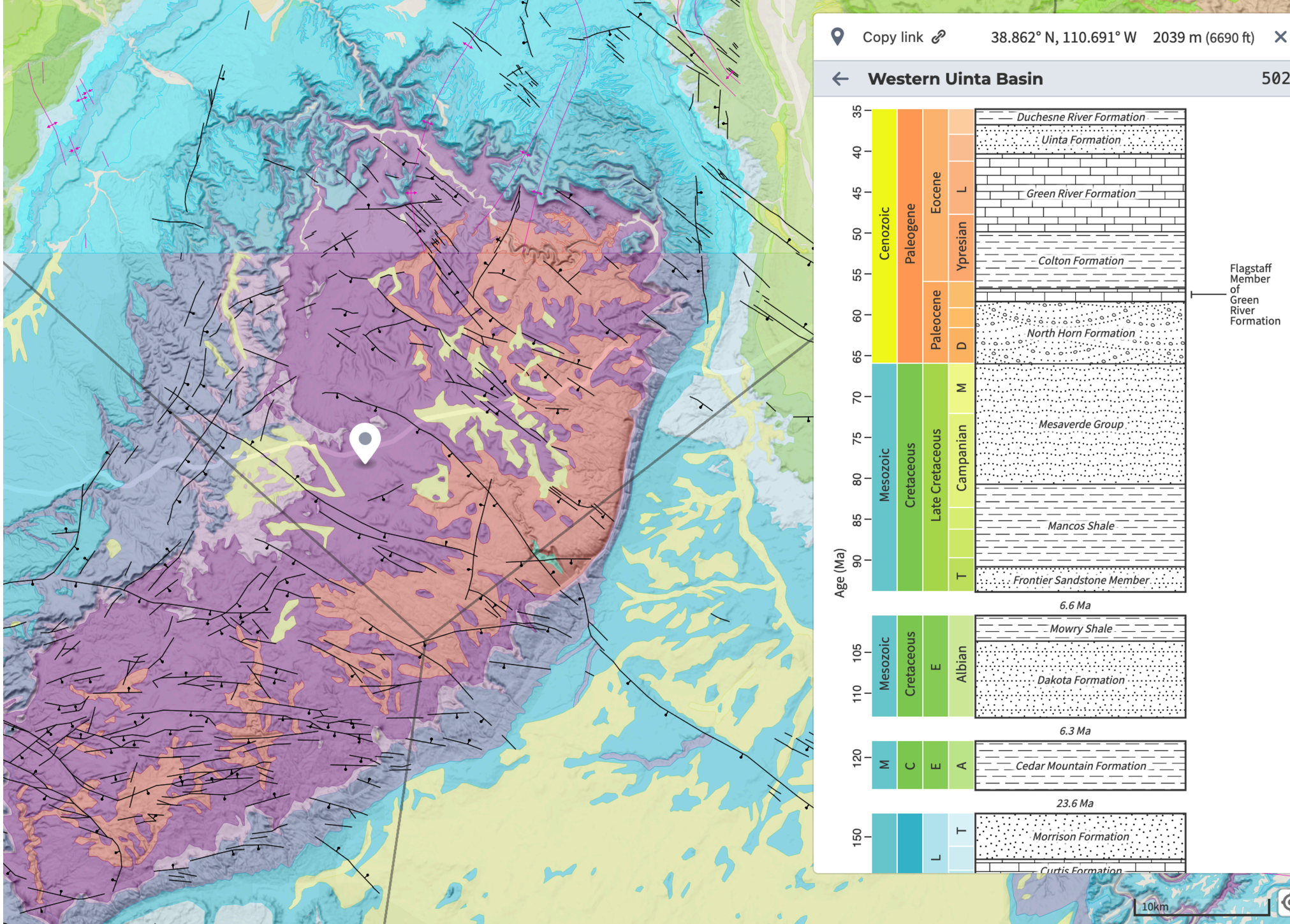


https://macrostrat.org/map

FAIROS PIDs workshop, UC–Boulder, September 2023

MACROSTRAT

A quantitative, descriptive data system for geological information



Shallow-marine sedimentary rock area through geologic time

LETTER

doi:10.1038/nature10969

Formation of the 'Great Unconformity' as a trigger for the Cambrian explosion

Shanan E. Peters[1] & Robert R. Gaines[2]

The transition between the Proterozoic and Phanerozoic eons, beginning 542 million years (Myr) ago, is distinguished by the diversification of multicellular animals and by their acquisition of mineralized skeletons during the Cambrian period[1]. Considerable progress has been made in documenting and more precisely correlating biotic patterns in the Neoproterozoic–Cambrian fossil record with geochemical and physical environmental perturbations[2–5], but the mechanisms responsible for those perturbations remain uncertain[1,2]. Here we use new stratigraphic and geochemical data to show that early Palaeozoic marine sediments deposited approximately 540–480 Myr ago record both an expansion in the area of shallow epicontinental seas and anomalous patterns of chemical sedimentation that are indicative of increased oceanic alkalinity and enhanced chemical weathering of continental crust. These geochemical conditions were caused by a protracted period of widespread

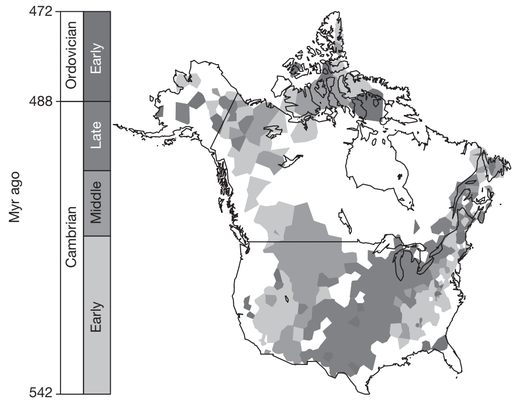
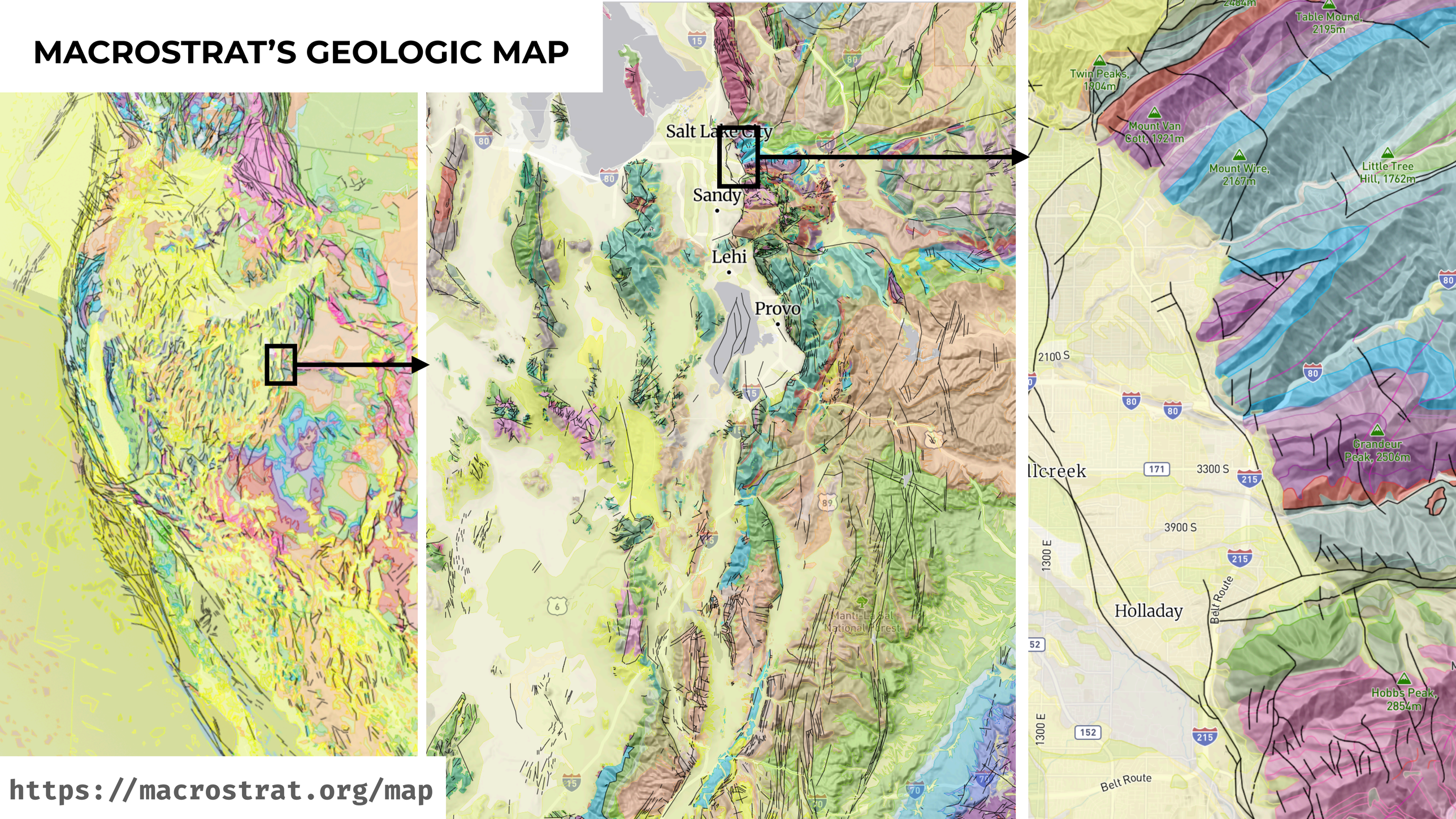


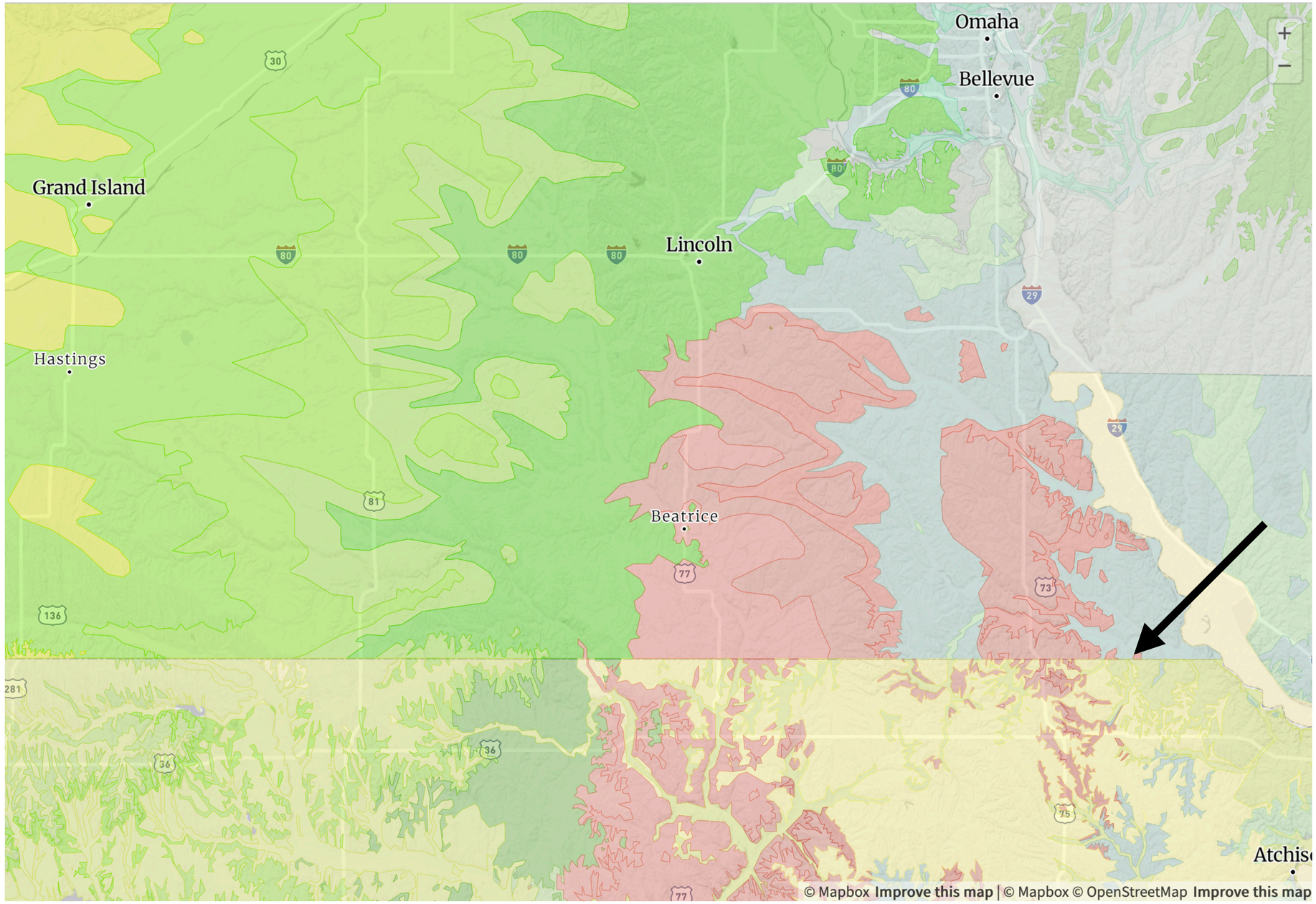
Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

MACROSTRAT'S GEOLOGIC MAP



https://macrostrat.org/map

Macrostrat's geologic map punts on hard problems!



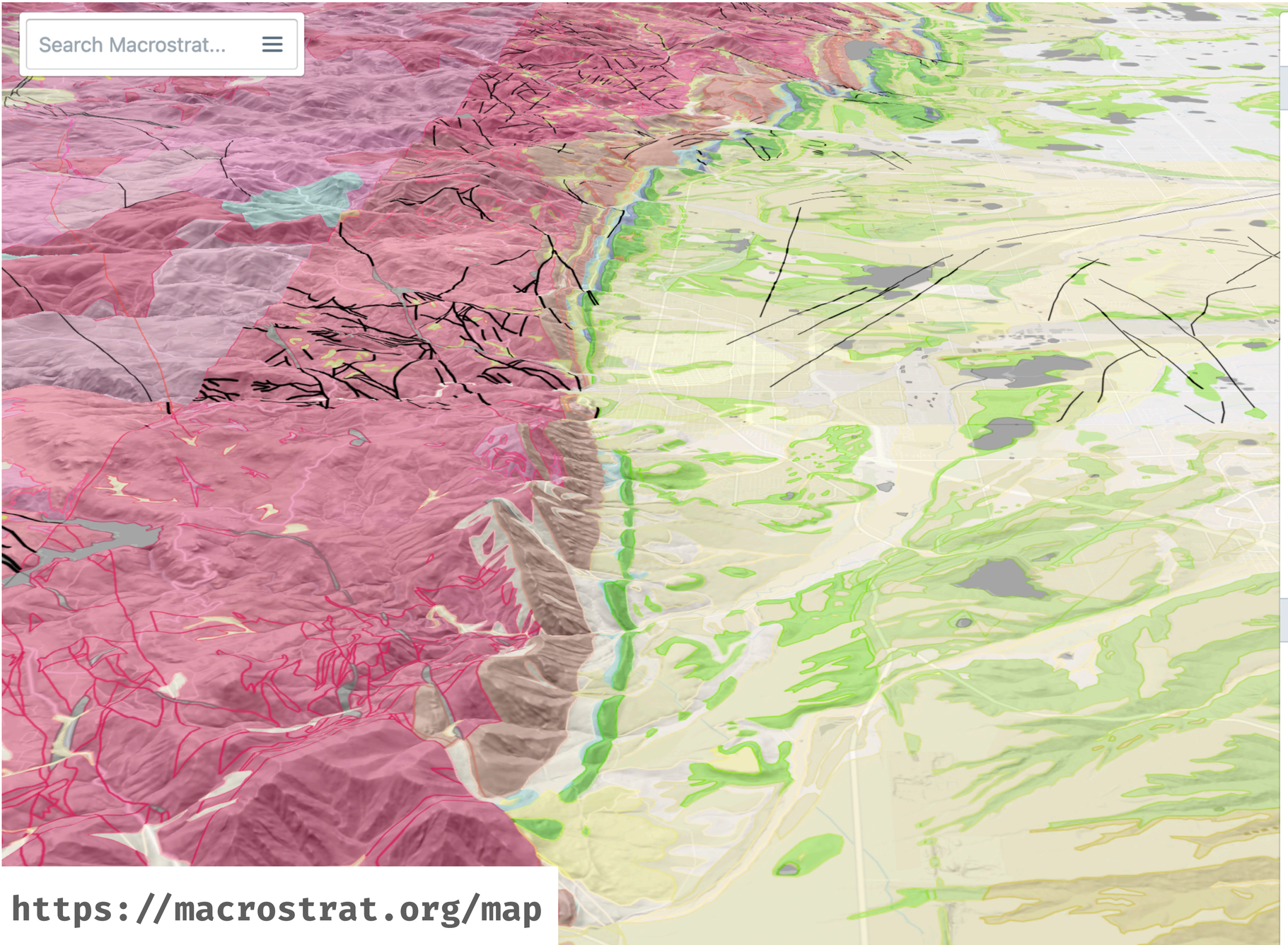
- **NOT seamless map units**
- Source boundaries are often quite apparent

Not a single map…a best-effort harmonization across many maps with a variety of core assumptions

<u>This is controversial among data curators!</u>

Macrostrat – A platform for geological exploration

Web interface



-110.6935, 48.2710 881m | 2890ft

Geologic map

via providers, Macrostrat

Name: Claggett Formation

Age: Campanian 83.6Ma - 72.1Ma

Stratigraphic name(s): Claggett Formation

Lithology: Major:{shale}, Minor:{sandstone}

Description: Dark gray to gray shale that weathers brown, with thin, gray sandstone laminae and beds in upper or middle part and calcareous concretions in lower part. Marine. Thickness as much as 170 m (558 ft).

Original map source: Vuke, S.M., Porter, K.W., Lonn, J.D., and Lopez, D.A., 2007, Geologic Map of Montana - Compact Disc: Montana Bureau of Mines and Geology: Geologic Map 62-C, 73 p., 2 sheets, scale 1:500,000. This map was digitized in 2012 as a r...

Source: Horton, J.D., C.A. San Juan, and D.B. Stoeser, 2017, The State Geologic Map Compilation (SGMC) geodatabase of the conterminous United States: U.S. Geological Survey Data Series 1052, doi: 10.3133/ds1052.133 / 2984391

Macrostrat-linked data

via Macrostrat

Claggett Formation *Matched stratigraphic unit*

Age: **Campanian** **80.725**Ma - **74.975**Ma *Refined using the Macrostrat age model.*

Thickness: 27 - 408m

Fossil collections: 3

Fossil occurrences: 3

Lithology: siliciclastic

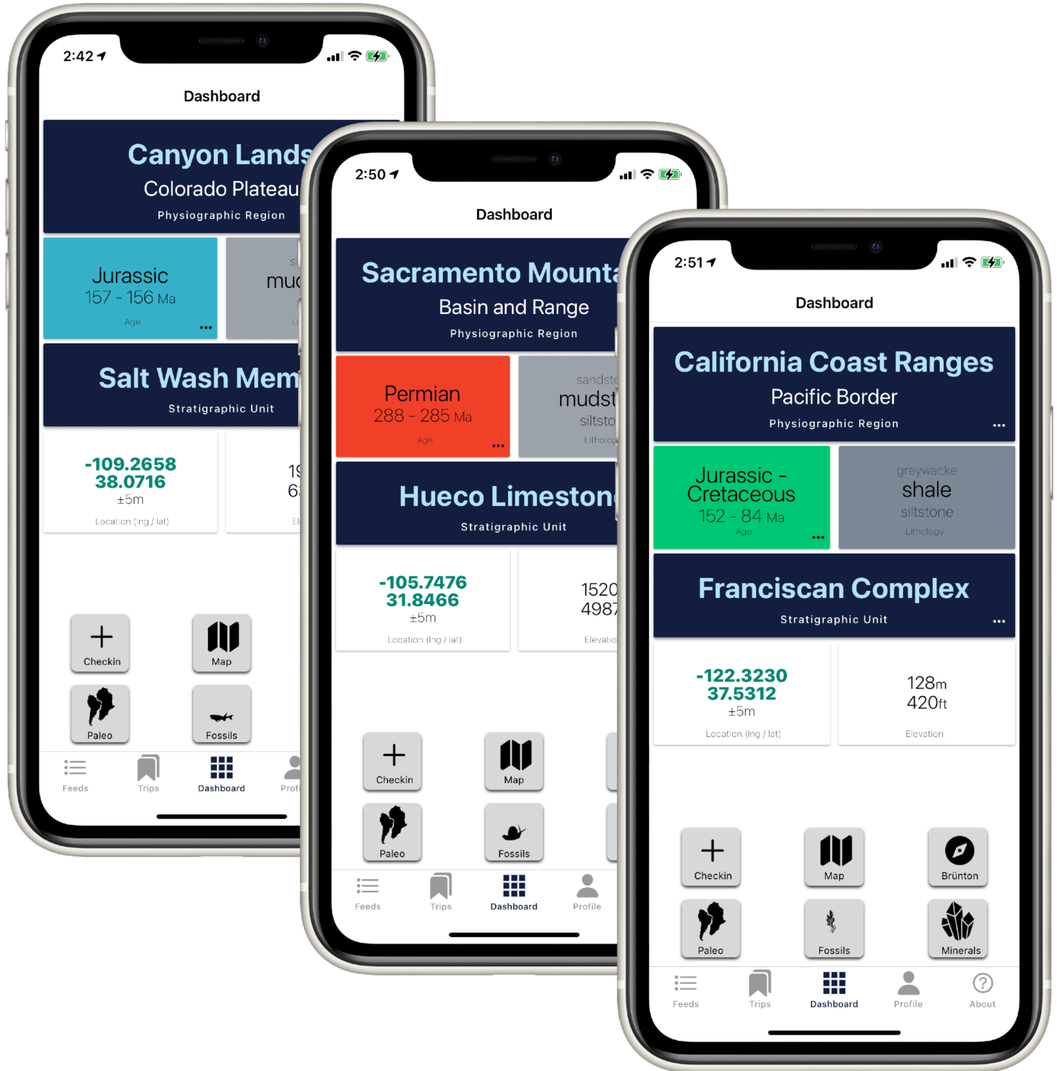
Environment: other

https://macrostrat.org/map

Rockd

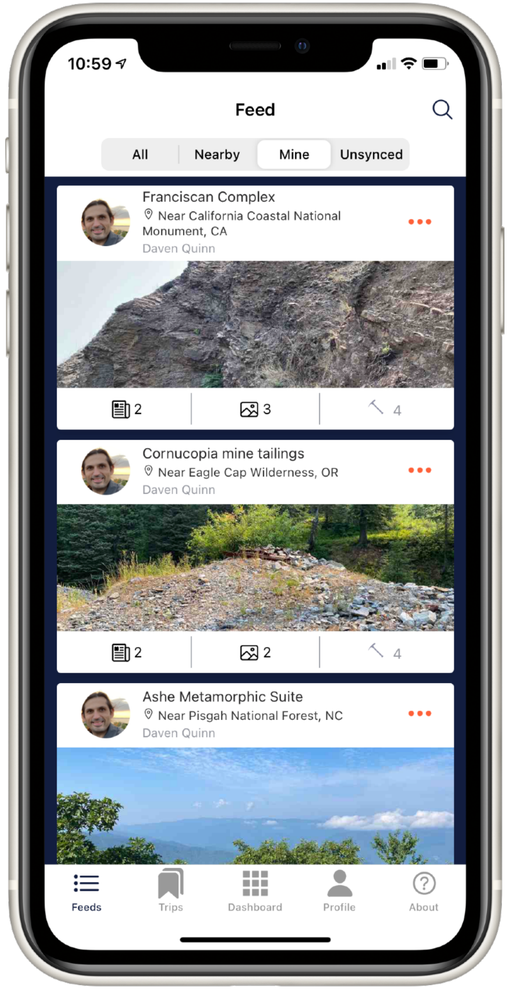
https://rockd.org

Understand your geological surroundings



Explore Macrostrat's map

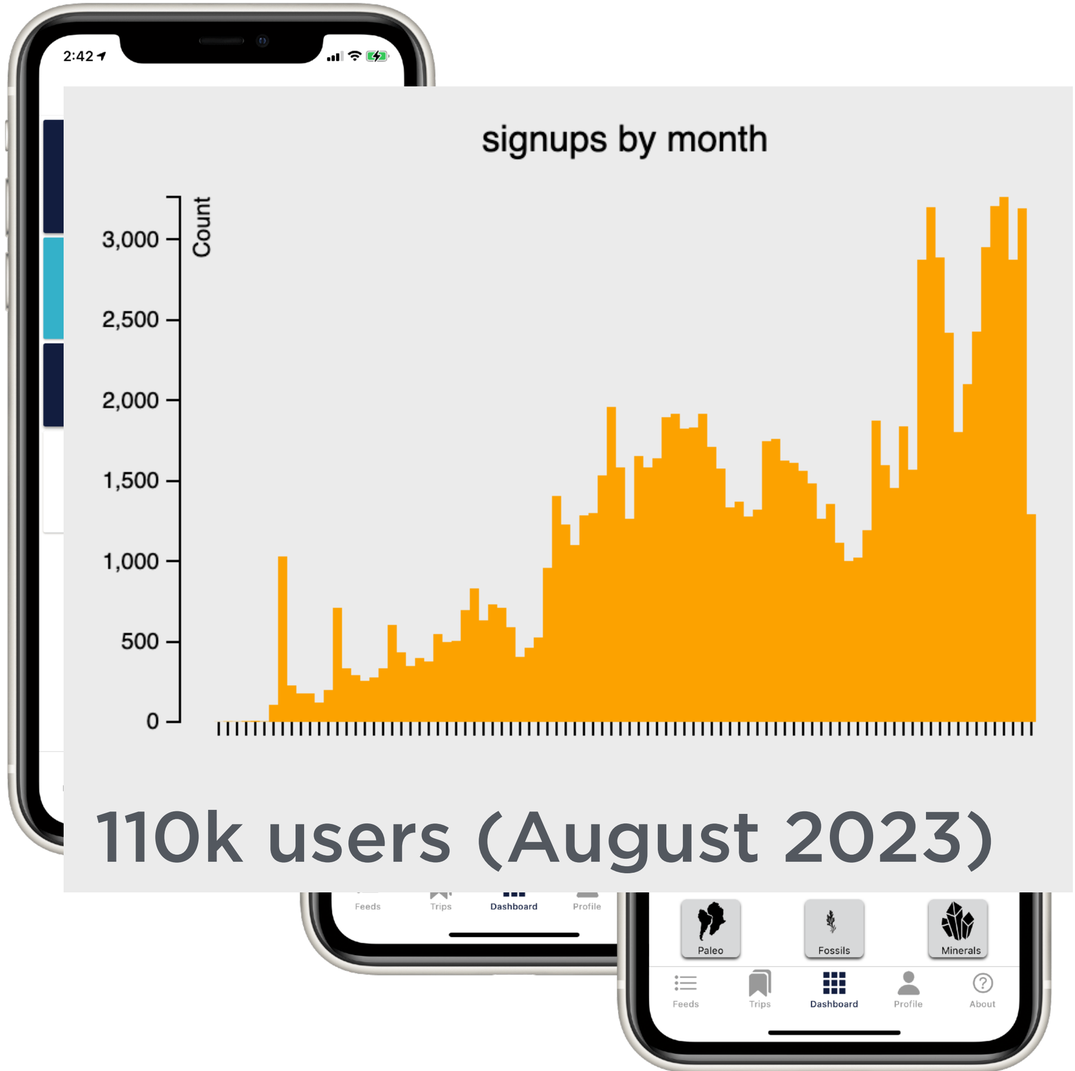
View and contribute local observations



Rockd

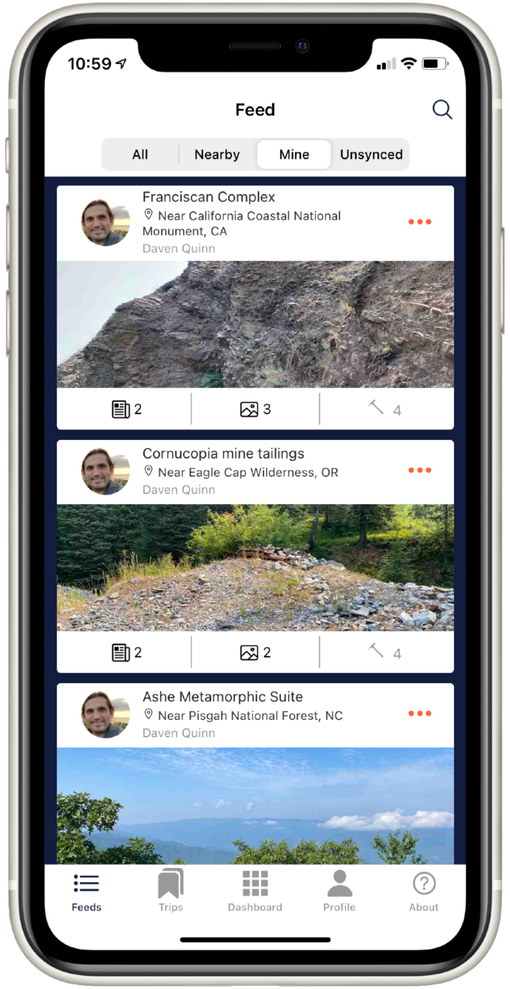
https://rockd.org

Understand your geological surroundings



Explore Macrostrat's map

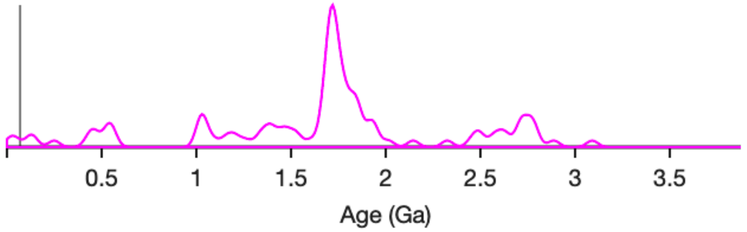
View and contribute local observations



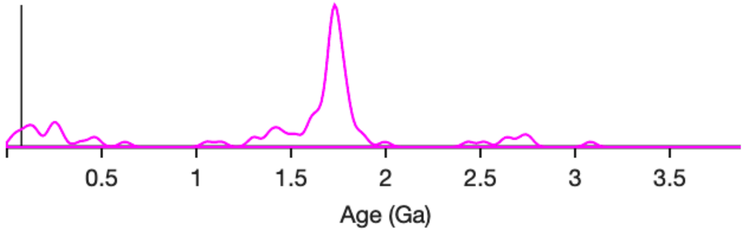
GETTING MORE GEOLOGICAL INFORMATION INTO THIS FRAMEWORK

Detrital-zircon geochronology

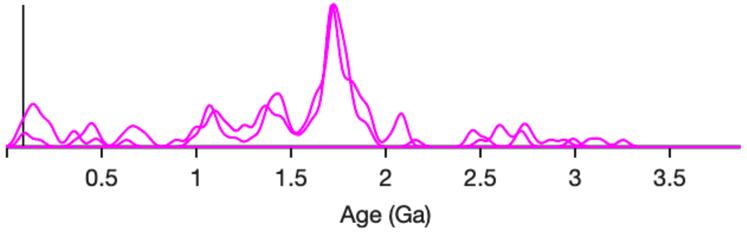
Lance Formation



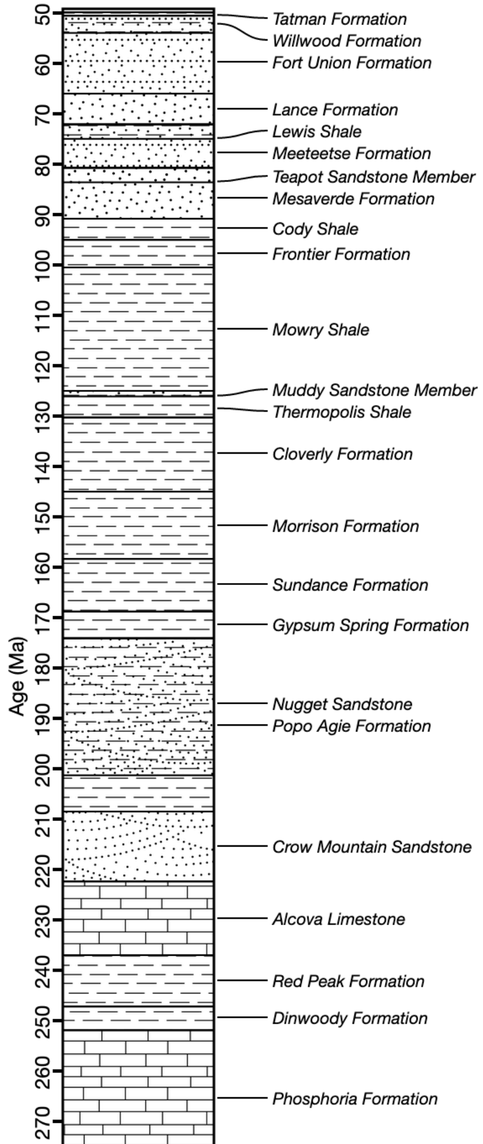
Meeteetse Formation

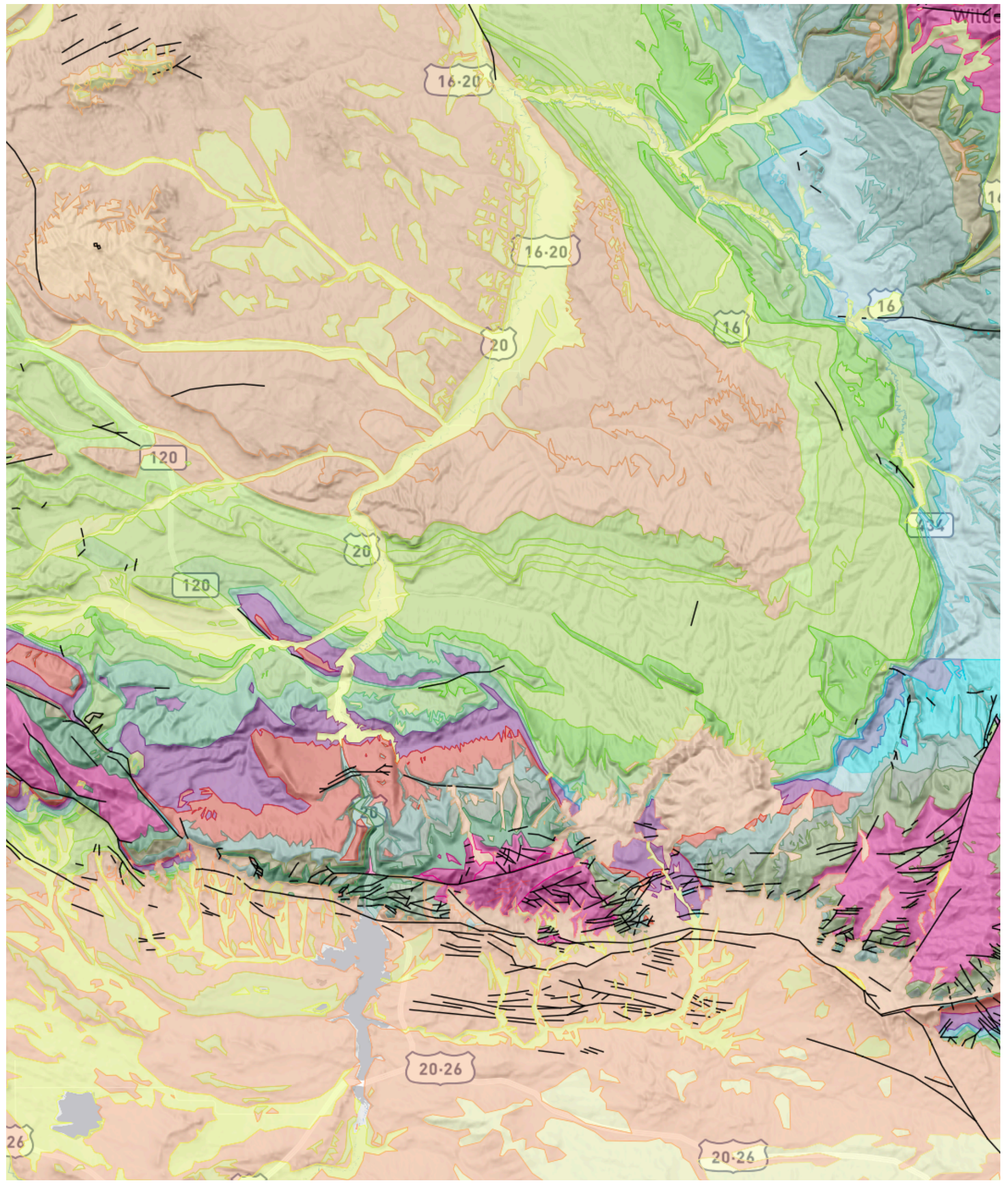


Mesaverde Formation



Bighorn Basin

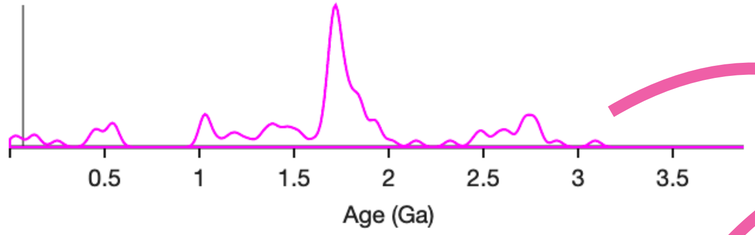




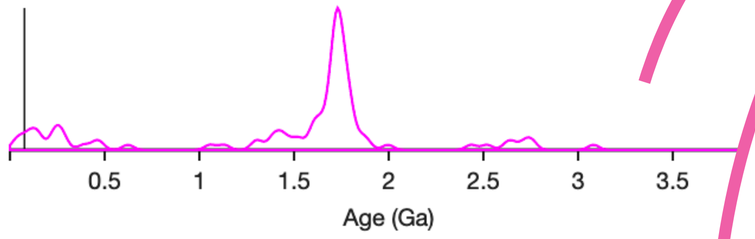
GETTING MORE GEOLOGICAL INFORMATION INTO THIS FRAMEWORK

Detrital-zircon geochronology

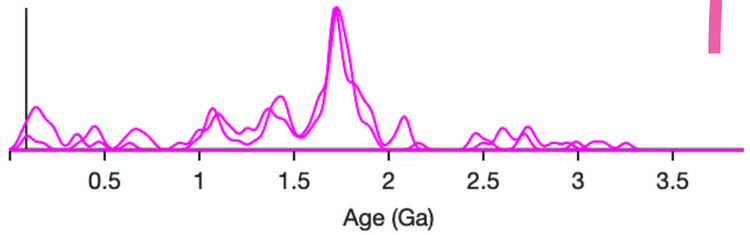
Lance Formation



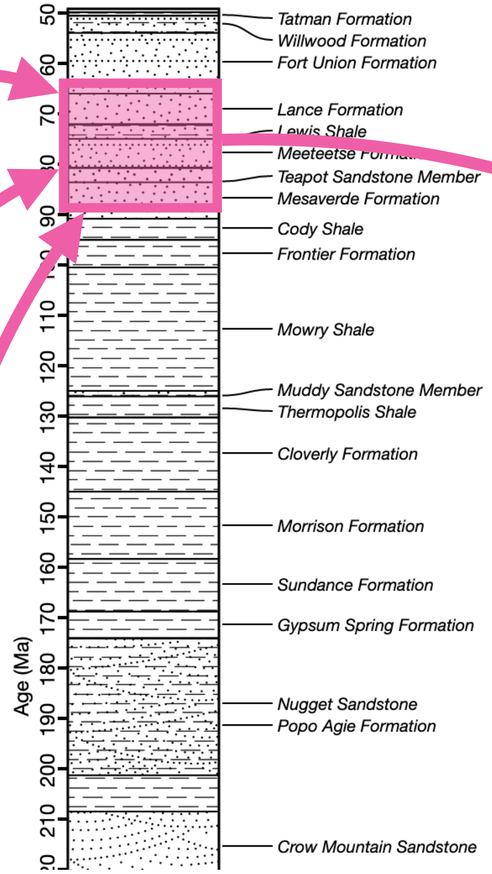
Meeteetse Formation

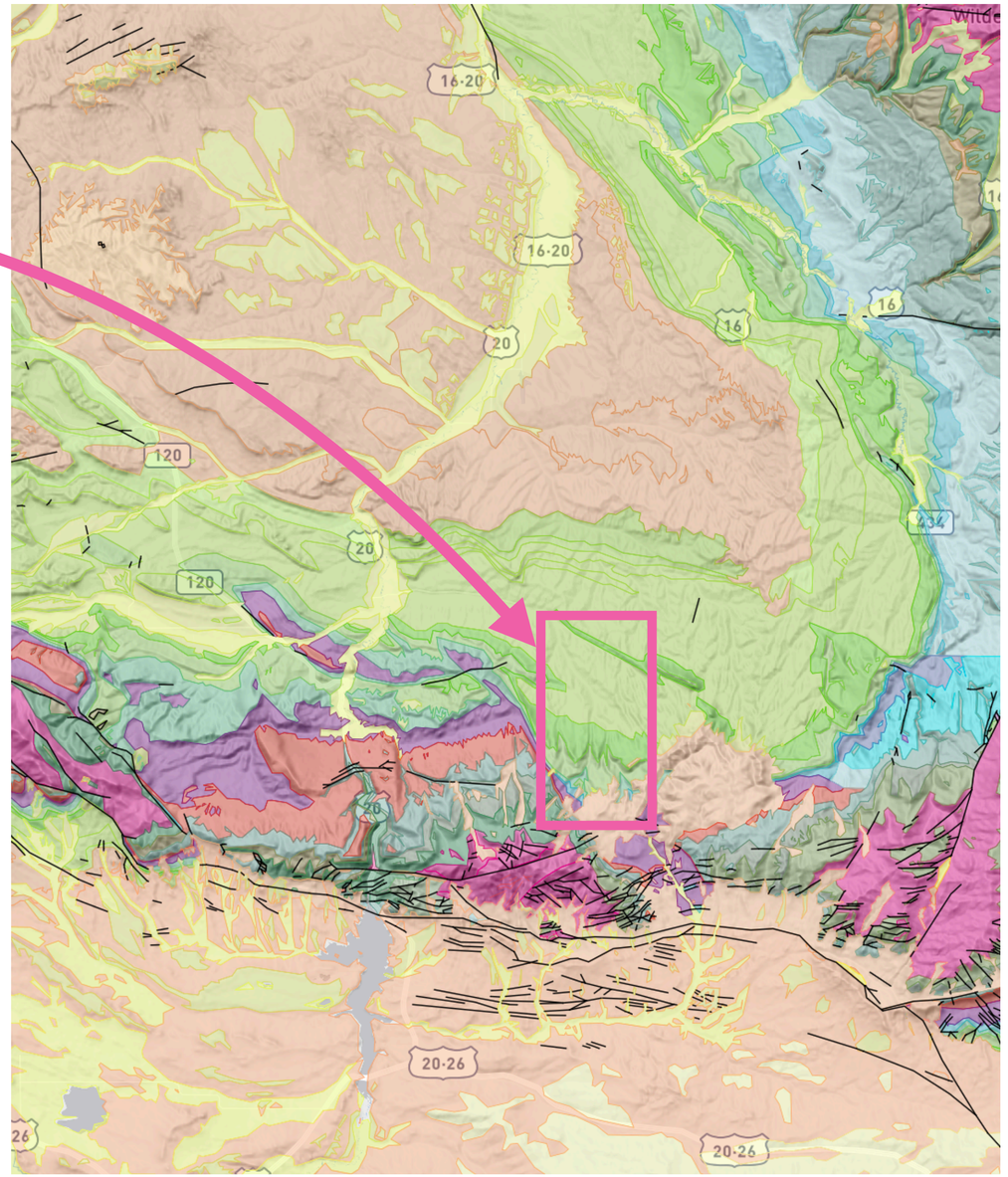


Mesaverde Formation



Bighorn Basin





- Geologic unit name forms a “primary key” that defines the relationship!
- Of course, labs often don’t track this data 😕

SOLUTION: METADATA TRACKING!

Need for good metadata is recognized in geoscience

USGS has a metadata **mandate**, supported by workflows, data curation staff, and tools

USGS science for a changing world

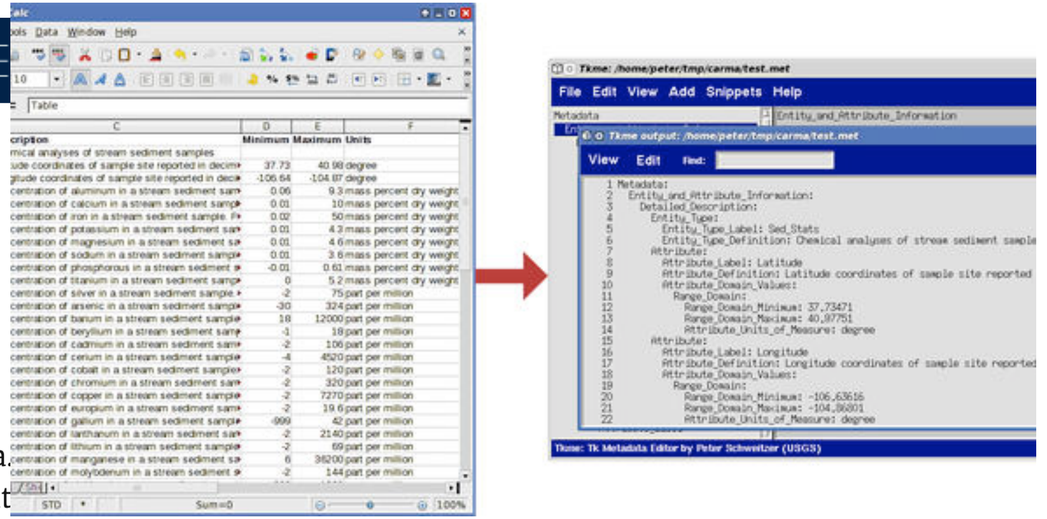
DATA MANAGEMENT

Describe (Metadata/Documentation)

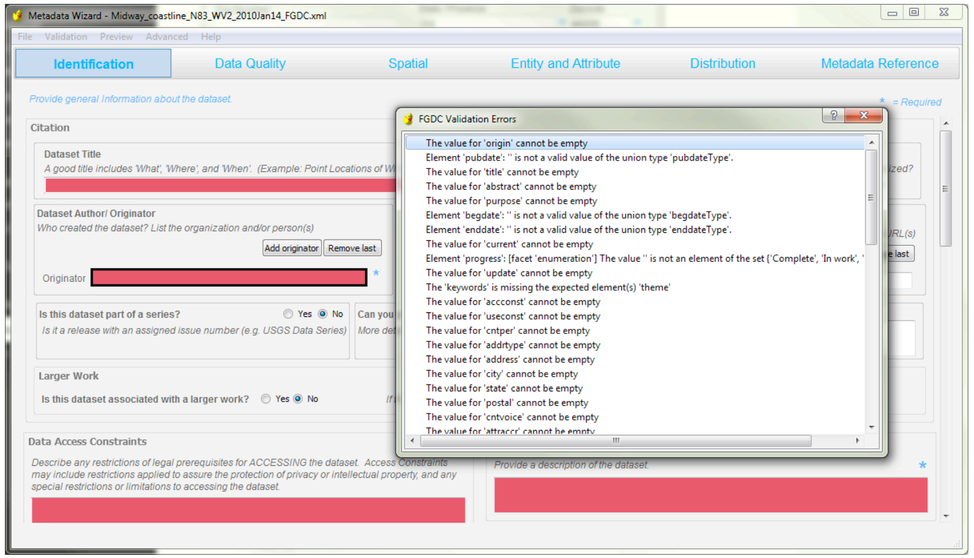
By Data Management

Throughout the data lifecycle process, documentation must be updated to reflect actions taken upon the data. This includes acquisition, processing, and analysis, but may touch upon any stage of the lifecycle. Updated and complete metadata are critical to maintaining data quality.

Exhortative explainer for agency staff



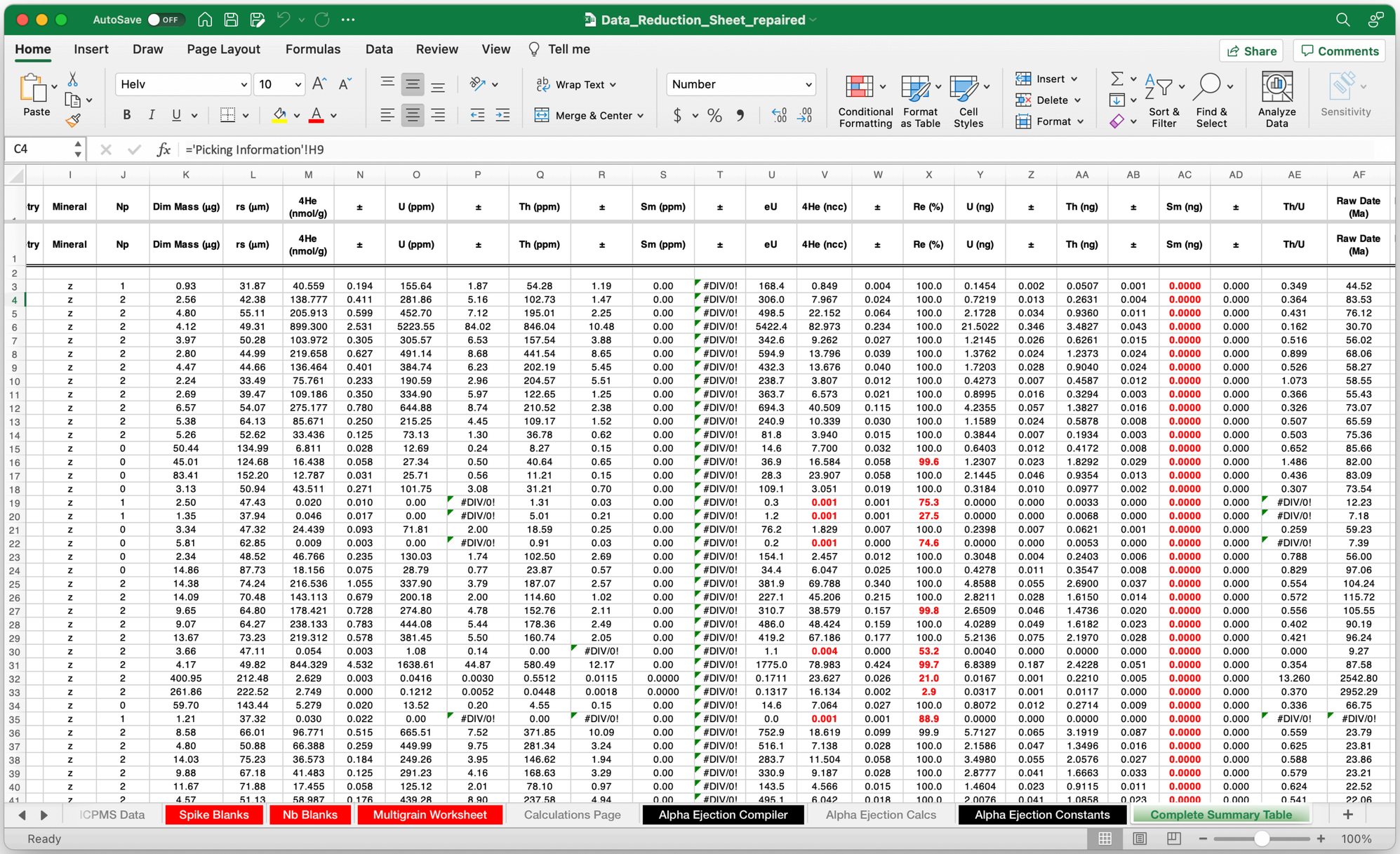
Vocabulary management interface



Scary (and required) metadata wizard

https://www.usgs.gov/data-management/describe-metadatadocumentation

Academic labs live in a world of Excel

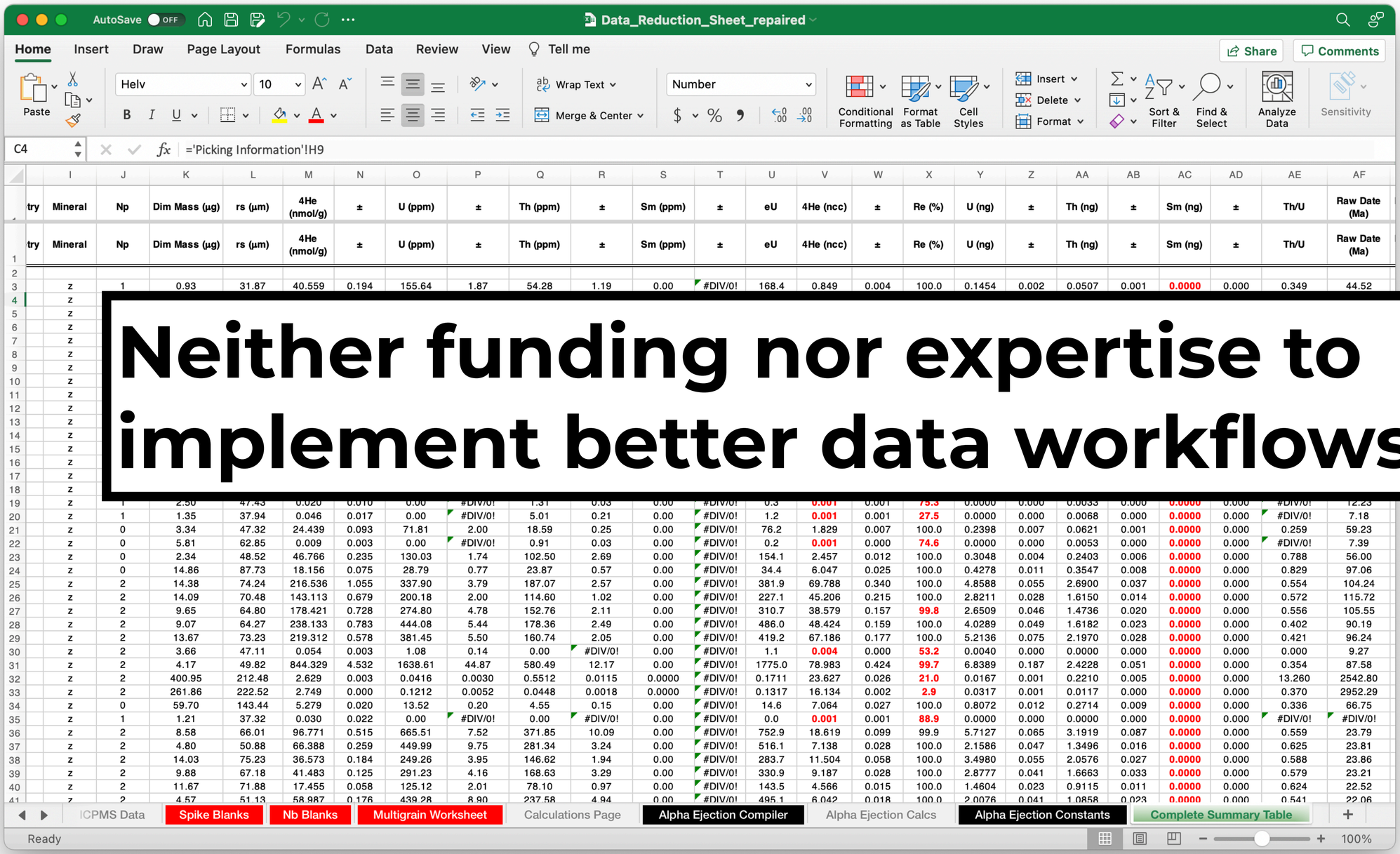


try	Mineral	Np	Dim Mass (µg)	rs (µm)	4He (nmol/g)	±	U (ppm)	±	Th (ppm)	±	Sm (ppm)	±	eU	4He (ncc)	±	Re (%)	U (ng)	±	Th (ng)	±	Sm (ng)	±	Th/U	Raw Date (Ma)
try	Mineral	Np	Dim Mass (µg)	rs (µm)	4He (nmol/g)	±	U (ppm)	±	Th (ppm)	±	Sm (ppm)	±	eU	4He (ncc)	±	Re (%)	U (ng)	±	Th (ng)	±	Sm (ng)	±	Th/U	Raw Date (Ma)
z	1	0.93	31.87	40.559	0.194	155.64	1.87	54.28	1.19	0.00	#DIV/0!	168.4	0.849	0.004	100.0	0.1454	0.002	0.0507	0.001	0.0000	0.000	0.349	44.52
z	2	2.56	42.38	138.777	0.411	281.86	5.16	102.73	1.47	0.00	#DIV/0!	306.0	7.967	0.024	100.0	0.7219	0.013	0.2631	0.004	0.0000	0.000	0.364	83.53
z	2	4.80	55.11	205.913	0.599	452.70	7.12	195.01	2.25	0.00	#DIV/0!	498.5	22.152	0.064	100.0	2.1728	0.034	0.9360	0.011	0.0000	0.000	0.431	76.12
z	2	4.12	49.31	899.300	2.531	5223.55	84.02	846.04	10.48	0.00	#DIV/0!	5422.4	82.973	0.234	100.0	21.5022	0.346	3.4827	0.043	0.0000	0.000	0.162	30.70
z	2	3.97	50.28	103.972	0.305	305.57	6.53	157.54	3.88	0.00	#DIV/0!	342.6	9.262	0.027	100.0	1.2145	0.026	0.6261	0.015	0.0000	0.000	0.516	56.02
z	2	2.80	44.99	219.658	0.627	491.14	8.68	441.54	8.65	0.00	#DIV/0!	594.9	13.796	0.039	100.0	1.3762	0.024	1.2373	0.024	0.0000	0.000	0.899	68.06
z	2	4.47	44.66	136.464	0.401	384.74	6.23	202.19	5.45	0.00	#DIV/0!	432.3	13.676	0.040	100.0	1.7203	0.028	0.9040	0.024	0.0000	0.000	0.526	58.27
z	2	2.24	33.49	75.761	0.233	190.59	2.96	204.57	5.51	0.00	#DIV/0!	238.7	3.807	0.012	100.0	0.4273	0.007	0.4587	0.012	0.0000	0.000	1.073	58.55
z	2	2.69	39.47	109.186	0.350	334.90	5.97	122.65	1.25	0.00	#DIV/0!	363.7	6.573	0.021	100.0	0.8995	0.016	0.3294	0.003	0.0000	0.000	0.366	55.43
z	2	6.57	54.07	275.177	0.780	644.88	8.74	210.52	2.38	0.00	#DIV/0!	694.3	40.509	0.115	100.0	4.2355	0.057	1.3827	0.016	0.0000	0.000	0.326	73.07
z	2	5.38	64.13	85.671	0.250	215.25	4.45	109.17	1.52	0.00	#DIV/0!	240.9	10.339	0.030	100.0	1.1589	0.024	0.5878	0.008	0.0000	0.000	0.507	65.59
z	2	5.26	52.62	33.436	0.125	73.13	1.30	36.78	0.62	0.00	#DIV/0!	81.8	3.940	0.015	100.0	0.3844	0.007	0.1934	0.003	0.0000	0.000	0.503	75.36
z	0	50.44	134.99	6.811	0.028	12.69	0.24	8.27	0.15	0.00	#DIV/0!	14.6	7.700	0.032	100.0	0.6403	0.012	0.4172	0.008	0.0000	0.000	0.652	85.66
z	0	45.01	124.68	16.438	0.058	27.34	0.50	40.64	0.65	0.00	#DIV/0!	36.9	16.584	0.058	99.6	1.2307	0.023	1.8292	0.029	0.0000	0.000	1.486	82.00
z	0	83.41	152.20	12.787	0.031	25.71	0.56	11.21	0.15	0.00	#DIV/0!	28.3	23.907	0.058	100.0	2.1445	0.046	0.9354	0.013	0.0000	0.000	0.436	83.09
z	0	3.13	50.94	43.511	0.271	101.75	3.08	31.21	0.70	0.00	#DIV/0!	109.1	3.051	0.019	100.0	0.3184	0.010	0.0977	0.002	0.0000	0.000	0.307	73.54
z	1	2.50	47.43	0.020	0.010	0.00	#DIV/0!	1.31	0.03	0.00	#DIV/0!	0.3	0.001	0.001	75.3	0.0000	0.000	0.0033	0.000	0.0000	0.000	#DIV/0!	12.23
z	1	1.35	37.94	0.046	0.017	0.00	#DIV/0!	5.01	0.21	0.00	#DIV/0!	1.2	0.001	0.001	27.5	0.0000	0.000	0.0068	0.000	0.0000	0.000	#DIV/0!	7.18
z	0	3.34	47.32	24.439	0.093	71.81	2.00	18.59	0.25	0.00	#DIV/0!	76.2	1.829	0.007	100.0	0.2398	0.007	0.0621	0.001	0.0000	0.000	0.259	59.23
z	0	5.81	62.85	0.009	0.003	0.00	#DIV/0!	0.91	0.03	0.00	#DIV/0!	0.2	0.001	0.000	74.6	0.0000	0.000	0.0053	0.000	0.0000	0.000	#DIV/0!	7.39
z	0	2.34	48.52	46.766	0.235	130.03	1.74	102.50	2.69	0.00	#DIV/0!	154.1	2.457	0.012	100.0	0.3048	0.004	0.2403	0.006	0.0000	0.000	0.788	56.00
z	0	14.86	87.73	18.156	0.075	28.79	0.77	23.87	0.57	0.00	#DIV/0!	34.4	6.047	0.025	100.0	0.4278	0.011	0.3547	0.008	0.0000	0.000	0.829	97.06
z	2	14.38	74.24	216.536	1.055	337.90	3.79	187.07	2.57	0.00	#DIV/0!	381.9	69.788	0.340	100.0	4.8588	0.055	2.6900	0.037	0.0000	0.000	0.554	104.24
z	2	14.09	70.48	143.113	0.679	200.18	2.00	114.60	1.02	0.00	#DIV/0!	227.1	45.206	0.215	100.0	2.8211	0.028	1.6150	0.014	0.0000	0.000	0.572	115.72
z	2	9.65	64.80	178.421	0.728	274.80	4.78	152.76	2.11	0.00	#DIV/0!	310.7	38.579	0.157	99.8	2.6509	0.046	1.4736	0.020	0.0000	0.000	0.556	105.55
z	2	9.07	64.27	238.133	0.783	444.08	5.44	178.36	2.49	0.00	#DIV/0!	486.0	48.424	0.159	100.0	4.0289	0.049	1.6182	0.023	0.0000	0.000	0.402	90.19
z	2	13.67	73.23	219.312	0.578	381.45	5.50	160.74	2.05	0.00	#DIV/0!	419.2	67.186	0.177	100.0	5.2136	0.075	2.1970	0.028	0.0000	0.000	0.421	96.24
z	2	3.66	47.11	0.054	0.003	1.08	0.14	0.00	#DIV/0!	0.00	#DIV/0!	1.1	0.004	0.000	53.2	0.0040	0.000	0.0000	0.000	0.0000	0.000	0.000	9.27
z	2	4.17	49.82	844.329	4.532	1638.61	44.87	580.49	12.17	0.00	#DIV/0!	1775.0	78.983	0.424	99.7	6.8389	0.187	2.4228	0.051	0.0000	0.000	0.354	87.58
z	2	400.95	212.48	2.629	0.003	0.0416	0.0030	0.5512	0.0115	0.0000	#DIV/0!	0.1711	23.627	0.026	21.0	0.0167	0.001	0.2210	0.005	0.0000	0.000	13.260	2542.80
z	2	261.86	222.52	2.749	0.000	0.1212	0.0052	0.0448	0.0018	0.0000	#DIV/0!	0.1317	16.134	0.002	2.9	0.0317	0.001	0.0117	0.000	0.0000	0.000	0.370	2952.29
z	0	59.70	143.44	5.279	0.020	13.52	0.20	4.55	0.15	0.00	#DIV/0!	14.6	7.064	0.027	100.0	0.8072	0.012	0.2714	0.009	0.0000	0.000	0.336	66.75
z	1	1.21	37.32	0.030	0.022	0.00	#DIV/0!	0.00	#DIV/0!	0.00	#DIV/0!	0.0	0.001	0.001	88.9	0.0000	0.000	0.0000	0.000	0.0000	0.000	#DIV/0!	#DIV/0!
z	2	8.58	66.01	96.771	0.515	665.51	7.52	371.85	10.09	0.00	#DIV/0!	752.9	18.619	0.099	99.9	5.7127	0.065	3.1919	0.087	0.0000	0.000	0.559	23.79
z	2	4.80	50.88	66.388	0.259	449.99	9.75	281.34	3.24	0.00	#DIV/0!	516.1	7.138	0.028	100.0	2.1586	0.047	1.3496	0.016	0.0000	0.000	0.625	23.81
z	2	14.03	75.23	36.573	0.184	249.26	3.95	146.62	1.94	0.00	#DIV/0!	283.7	11.504	0.058	100.0	3.4980	0.055	2.0576	0.027	0.0000	0.000	0.588	23.86
z	2	9.88	67.18	41.483	0.125	291.23	4.16	168.63	3.29	0.00	#DIV/0!	330.9	9.187	0.028	100.0	2.8777	0.041	1.6663	0.033	0.0000	0.000	0.579	23.21
z	2	11.67	71.88	17.455	0.058	125.12	2.01	78.10	0.97	0.00	#DIV/0!	143.5	4.566	0.015	100.0	1.4604	0.023	0.9115	0.011	0.0000	0.000	0.624	22.52
z	2	4.57	51.13	58.987	0.176	439.28	8.90	237.58	4.94	0.00	#DIV/0!	495.1	6.042	0.018	100.0	2.0076	0.041	1.0858	0.023	0.0000	0.000	0.541	22.06

1000s of files spanning 10+ years of operation

Each lab has different file formats, data standards, etc.

Academic labs live in a world of Excel



1000s of files spanning 10+ years of operation

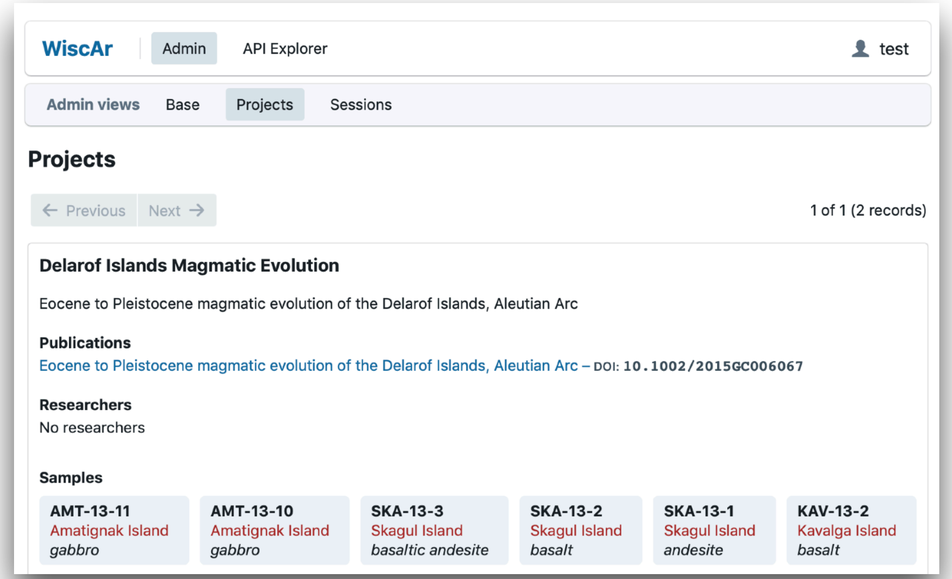
Neither funding nor expertise to implement better data workflows

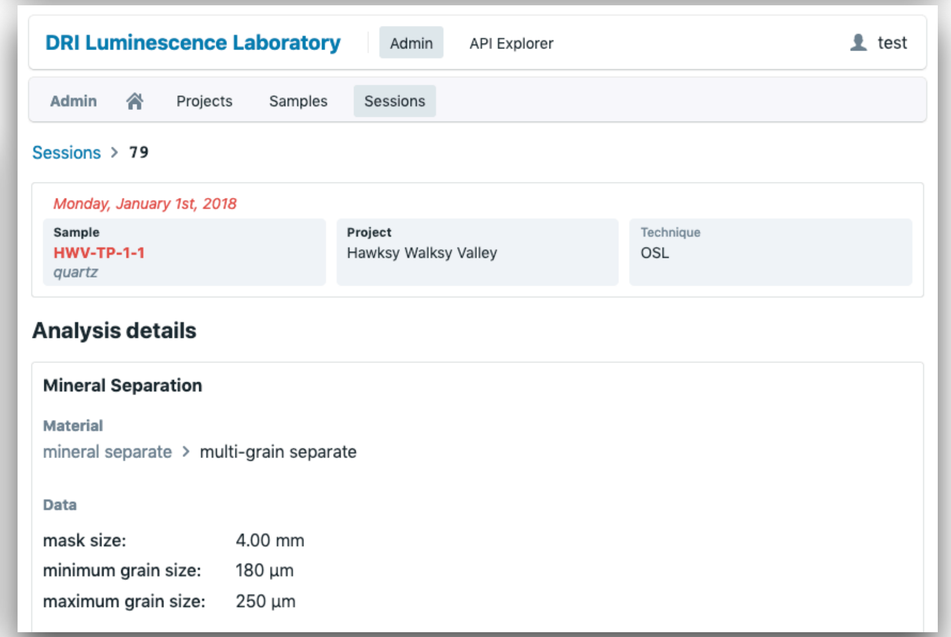
Each lab has different file formats, data standards, etc.

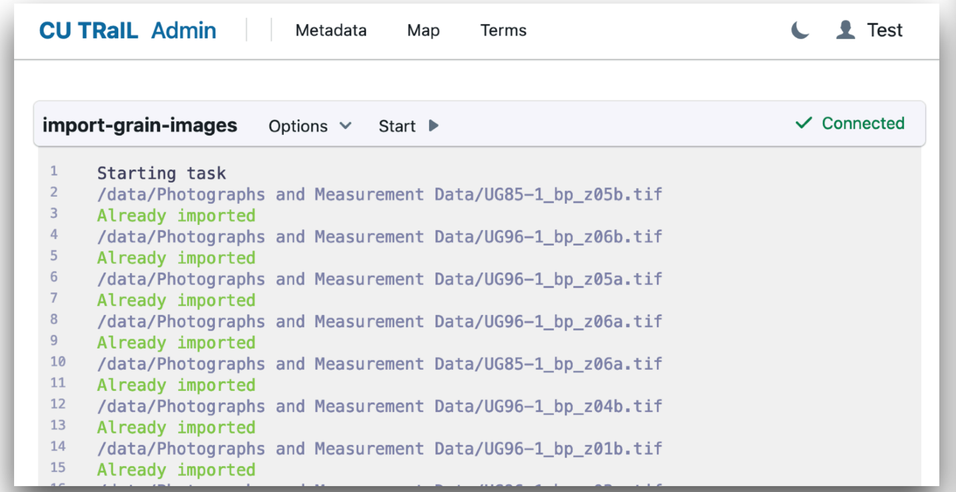
Sparrow

Web-based **metadata management platform** for a geochemical lab's data archive

- **Set embargo for projects and samples**
- **Manage project- and sample-level metadata**
- Search and link publications
- Access analytical data
- Import, export, and track original data files







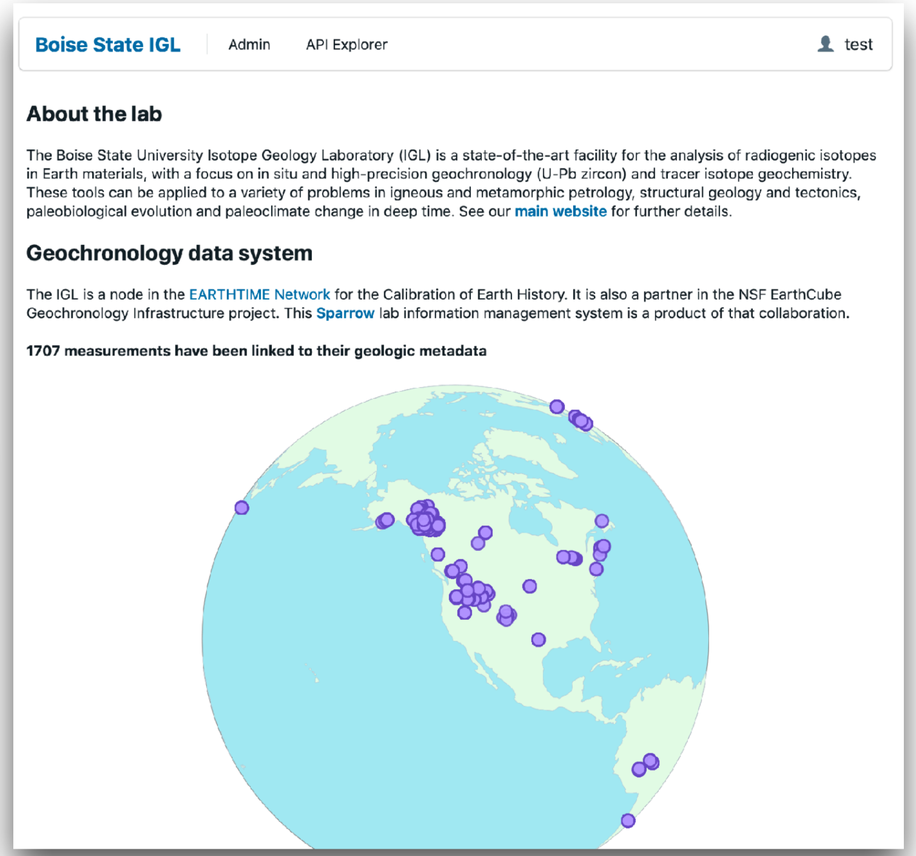
https://sparrow-data.org

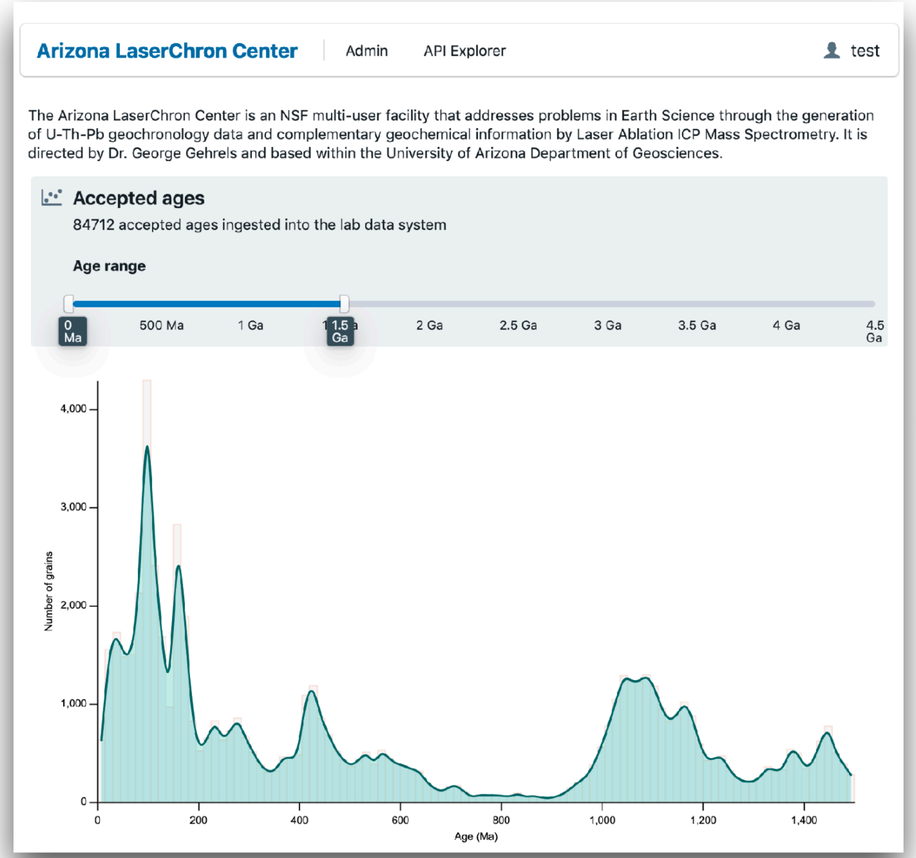
Sparrow's role in labs

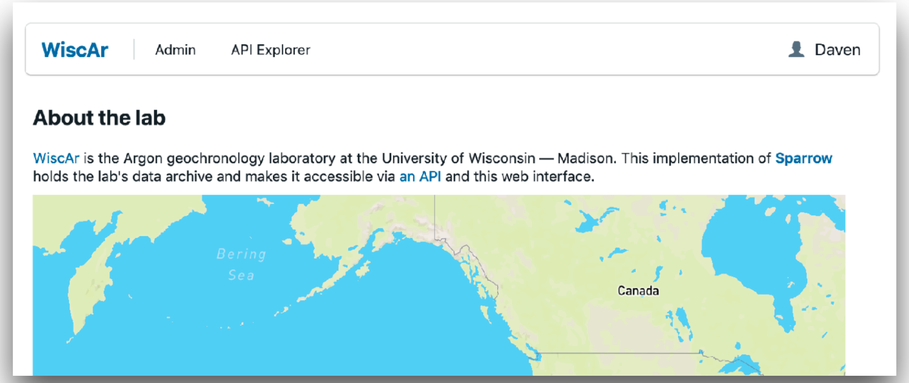
A public data portal

- Summary statistics
- Publications links
- Maps of samples
- Sample-specific information

- Helps fulfill open-data and reporting requirements
- **Extensible** for method-specific needs







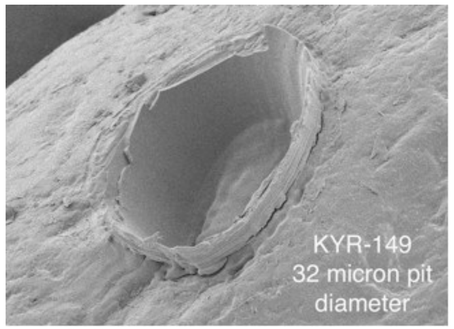
https://sparrow-data.org

Sparrow's database

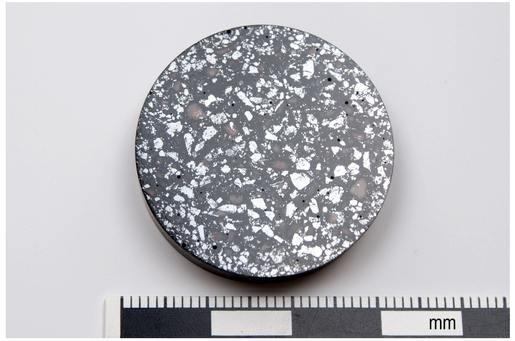
A single measured value

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

An 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**

Instrument session

Project

Location

Entity (e.g. geologic unit)

Publication

Researcher

Metadata tables

Geological context, publication status, embargo…

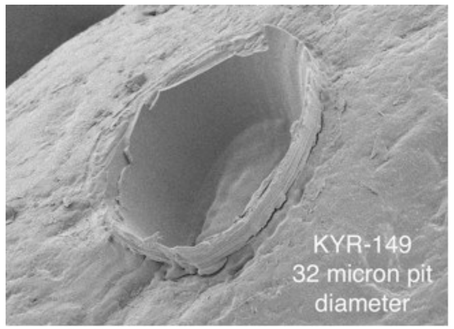
Sparrow's database

A platform for data curation

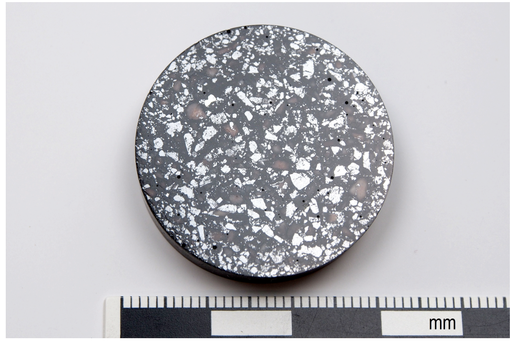
A single measured value

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

An 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** ***IGSN***

Metadata tables

Geological context, publication status, embargo…

Instrument session

Location

Entity (e.g. geologic unit)

Project ***NSF ID***

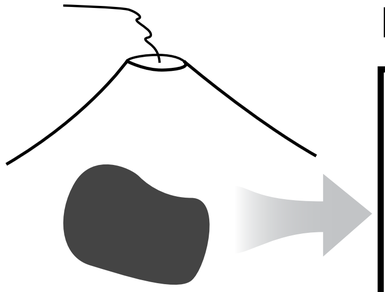
Publication ***DOI***

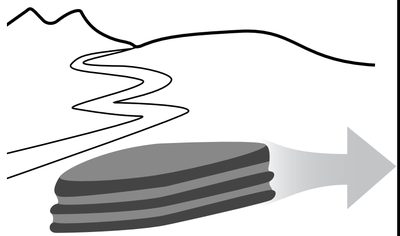
Researcher ***ORCID***

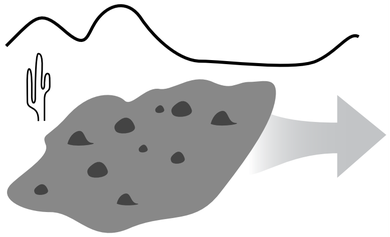
https://sparrow-data.org

Vision: *connecting lab data systems to the community*

...in a way where labs also benefit on their own terms







Samples + geologic context

Data Producers (Labs/analytical pipelines)

WiscAr (Igneous $^{40}Ar/^{39}Ar$)

Noble-gas MS $^{40}Ar/^{39}Ar$ → **PyChron** / **Legacy ArArCalc** → **Sparrow** $^{40}Ar/^{39}Ar$ plugins

Arizona LaserChron Center (Detrital Zircon)

LA-ICP-MS Detrital and igneous U-Pb → **AgeCalcML** / **Legacy AgeCalc** → **Sparrow** DZ plugins

DRI Luminescence Lab (OSL)

Photon Reader → **OSL Analyst**

ICP-AES / **ICP-MS** → **DRAC** → **Excel**

→ **Sparrow** OSL plugins

congruent pipelines for **Boise State IGL, NMGRL, WiscSIMS** and others

Propagation of ***ages + sample context*** into scientific record

new automated connections

manual data transfer and reporting

Geochronology Data Consumers

Earth Models

Domain-Specific Data Compilations

Macrostrat · **Neotoma** · **PaleoBioDB**

Archival Facilities

Geochron.org · **SESAR (IGSN)**

and many others!

Published literature

SO HOW'S THAT GOING?

~10 in-progress laboratory implementations **across domains** of geochronology and geochemistry

- Uranium–Lead
- Argon–Argon
- Cosmogenic nuclides
- (U+Th)/He thermochronology
- Optically-stimulated luminescence
- In-situ stable isotope geochemistry (SIMS)
- Electron Microprobe
- Carbon radioisotopes/AMS *(planned)*

ARIZONA LASERCHRON CENTER
Department of Geosciences
University of Arizona

UW Cosmogenic Nuclide Lab

DRI
Desert Research Institute

CONCORD U
Microprobe lab

NAU Paleoclimate Dynamics Lab

British Geological Survey

https://sparrow-data.org

SO HOW'S THAT GOING? *But really, how is it going? (five years in)*

~10 in-progress laboratory implementations **across domains** of geochronology and geochemistry

- Uranium–Lead
- Argon–Argon
- Cosmogenic nuclides
- (U+Th)/He thermochronology
- Optically-stimulated luminescence
- In-situ stable isotope geochemistry (SIMS)
- Electron Microprobe
- Carbon radioisotopes/AMS *(planned)*

One *operational* pipeline

ARIZONA LASERCHRON CENTER
Department of Geosciences
University of Arizona

New Mexico Geochronology Research Laboratory

WiscAr

ISOTOPE GEOLOGY LABORATORY
BOISE STATE UNIVERSITY

UW Cosmogenic Nuclide Lab

WiscSIMS

CU TRaIL
Thermochronology Research and Instrumentation Laboratory

DRI
Desert Research Institute

CONCORD U
Microprobe lab

NAU Paleoclimate Dynamics Lab

BGS British Geological Survey

https://sparrow-data.org

SO HOW'S THAT GOING? *But really, how is it going? (five years in)*

~10 in-progress laboratory implementations **across domains** of geochronology and geochemistry

- Ura
- Arg
- Cos
- (U+
- Opt
- In-s
- Elec
- Car

Academic labs lack expertise to implement/ manage metadata tracking.

It is just not their core task and is treated accordingly.

Even with purpose-built software, the "last-mile" problem is difficult to solve.

One *operational* pipeline

IRaIL
Thermochronology Research and Instrumentation Laboratory

Desert Research Institute

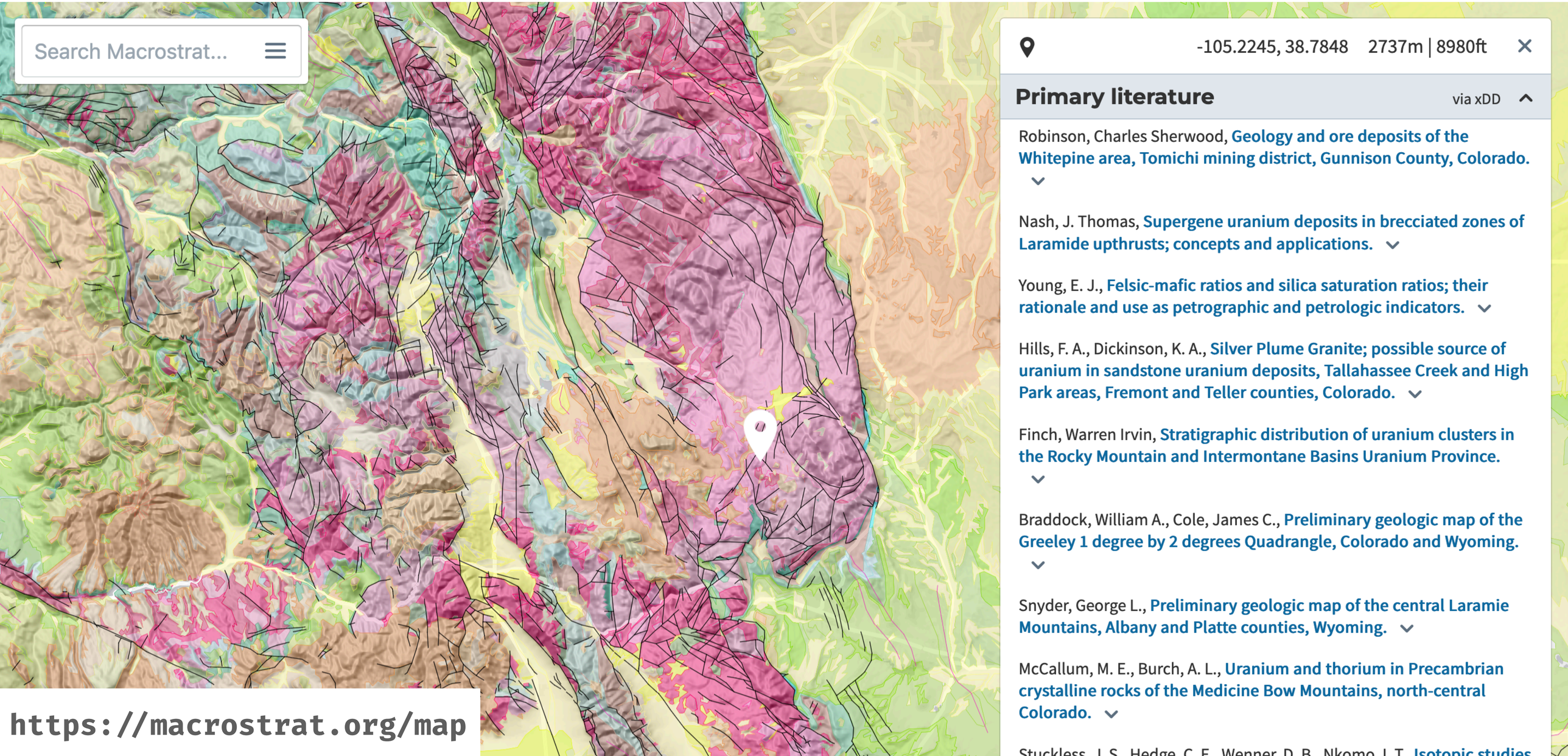
Microprobe lab

NAU Paleoclimate Dynamics Lab

BGS British Geological Survey

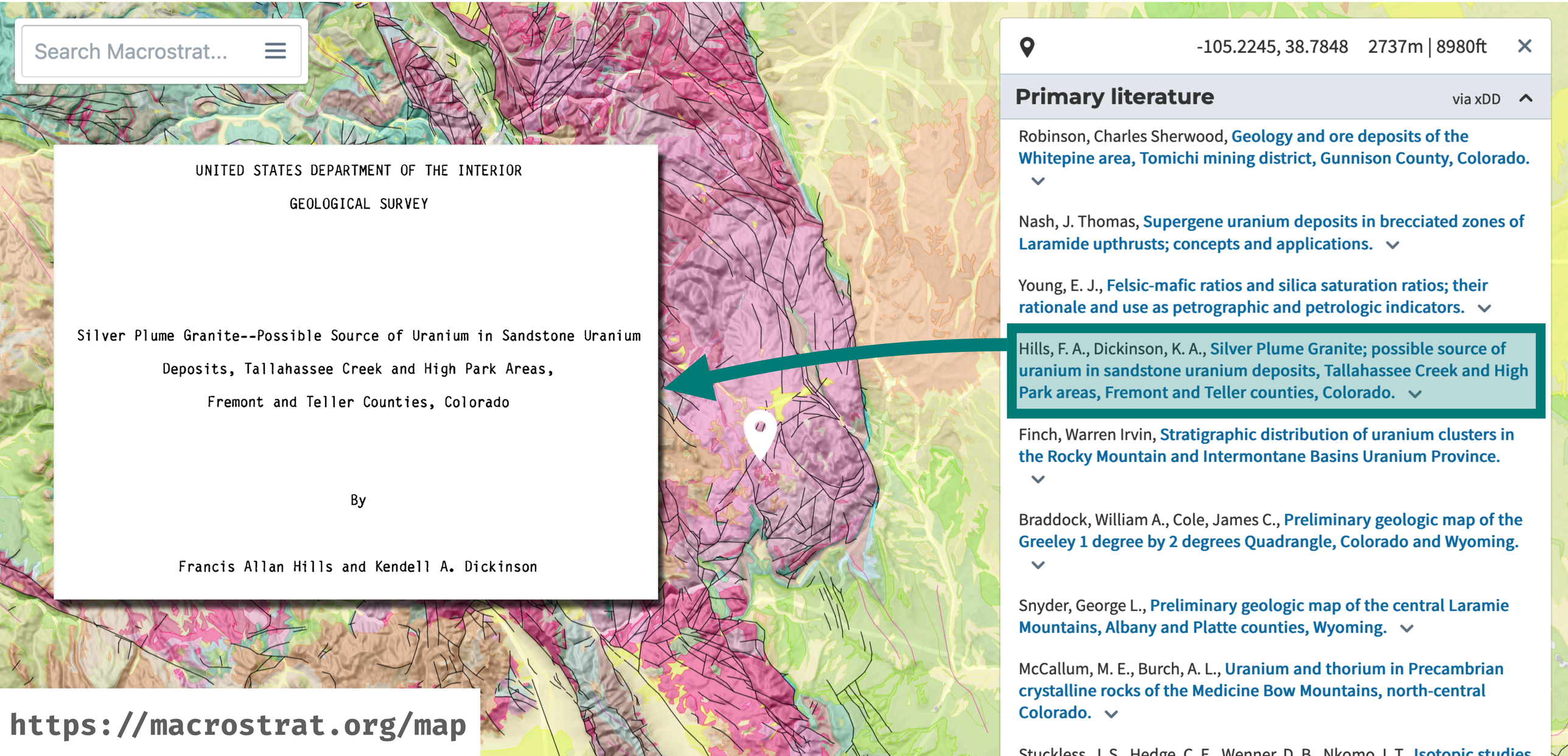
https://sparrow-data.org

Final note: **Automation to the rescue?** 🤖



`https://macrostrat.org/map`

Final note: **Automation to the rescue?** 🤖



https://macrostrat.org/map

Final note: Automation to the rescue? 🤖

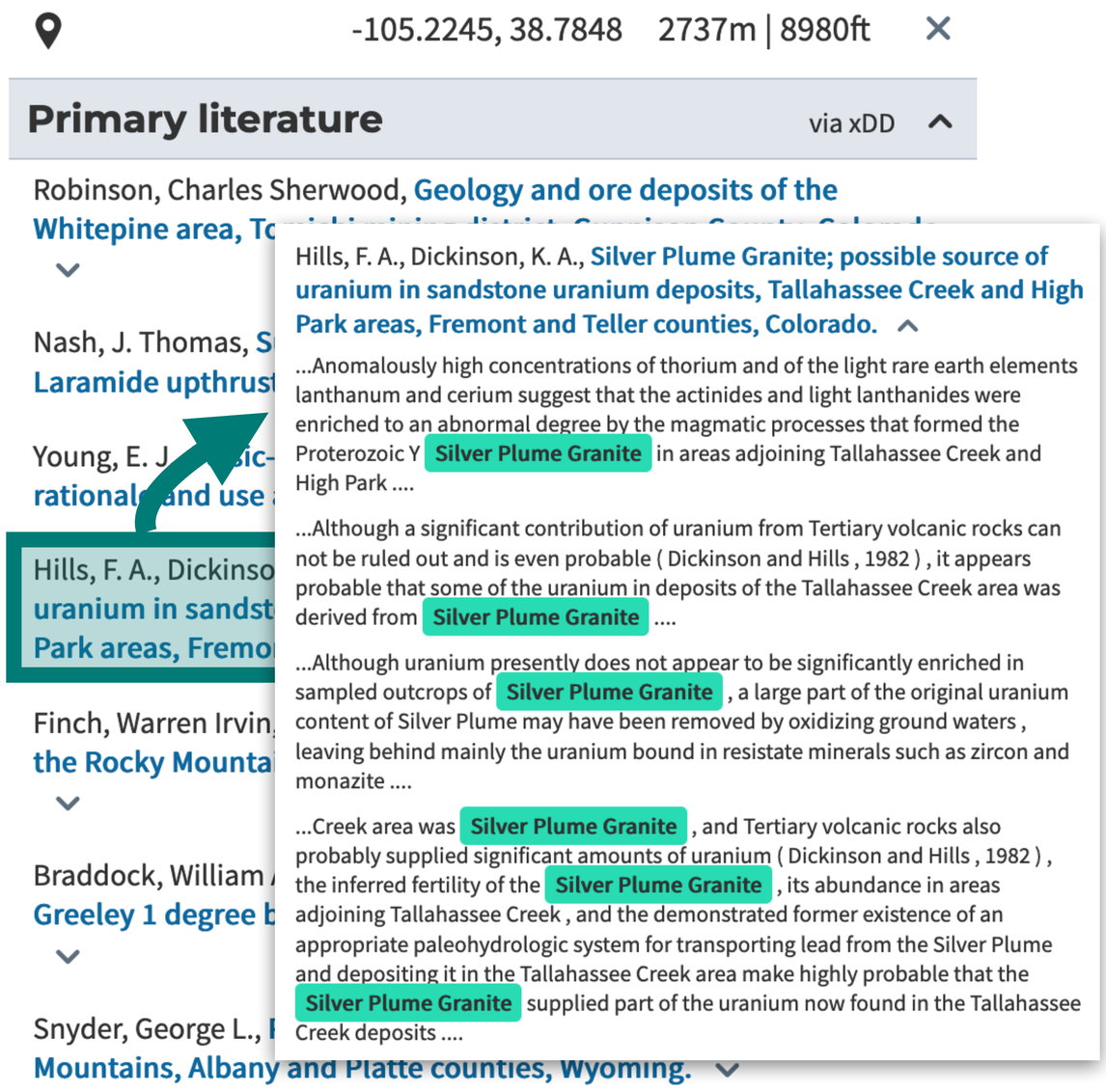
Macrostrat is linked to the **xDD** (formerly, GeoDeepDive) machine reading library, data infrastructure, and API

16,909,371 documents

108,486 added this month

25,112 added this week

3,683 added in the last 24 hours



- Detecting and surfacing references to geologic units in the scientific literature
- These aren't real metadata-level links, but they are pretty useful

https://xdd.wisc.edu