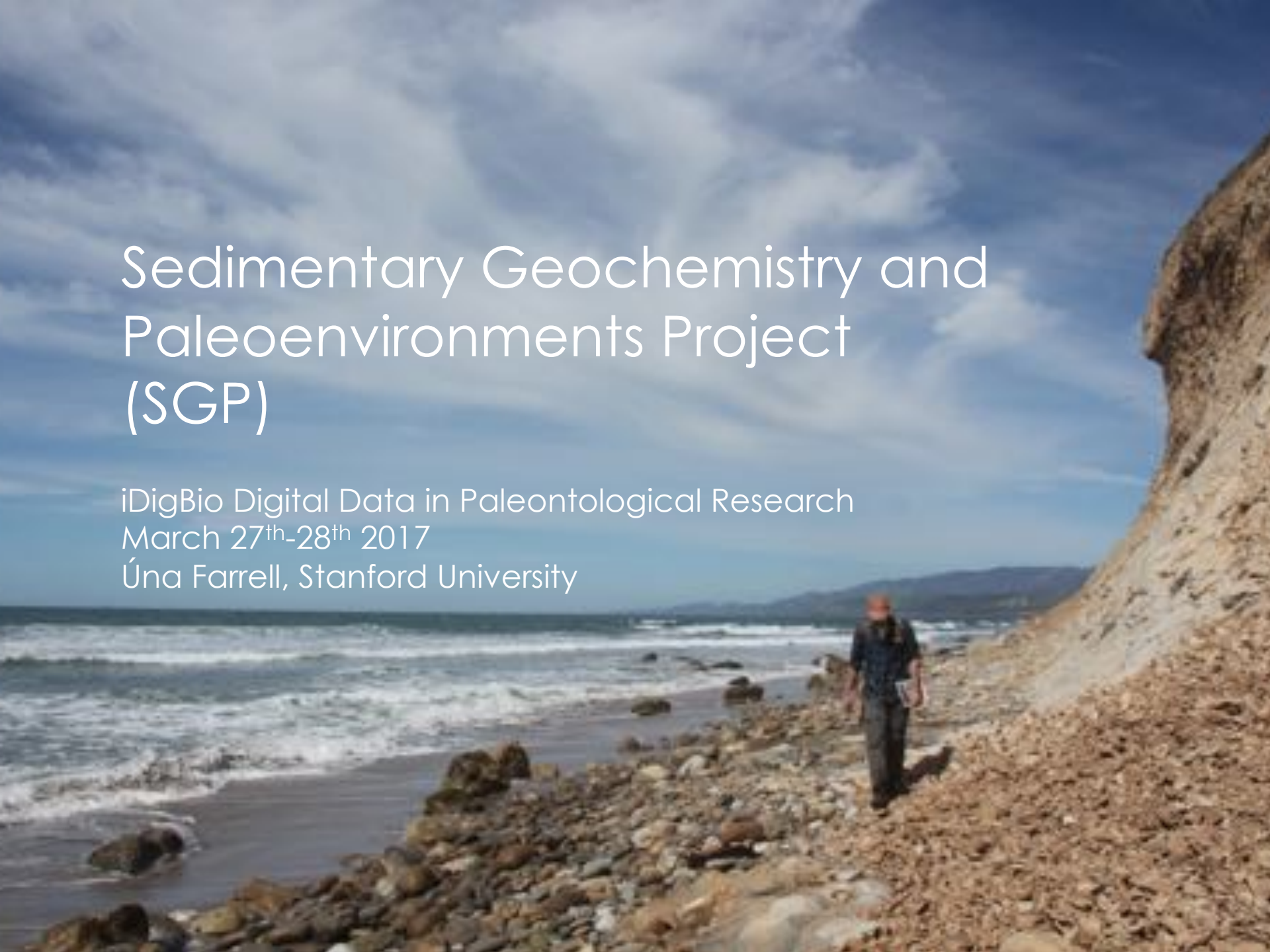
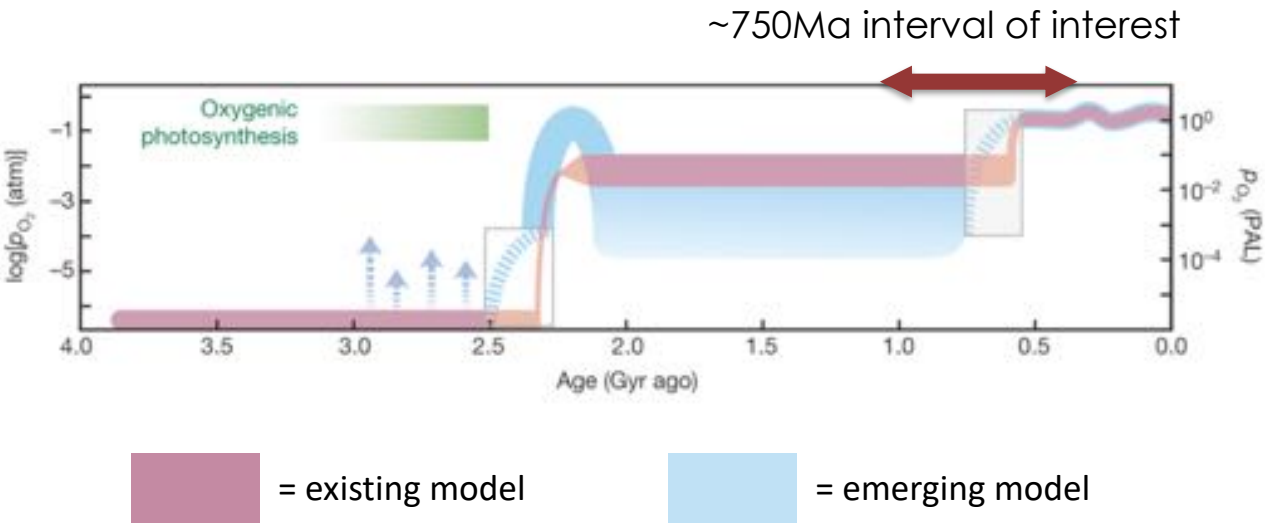


Sedimentary Geochemistry and Paleoenvironments Project (SGP)

iDigBio Digital Data in Paleontological Research
March 27th-28th 2017
Úna Farrell, Stanford University

PROJECT BACKGROUND

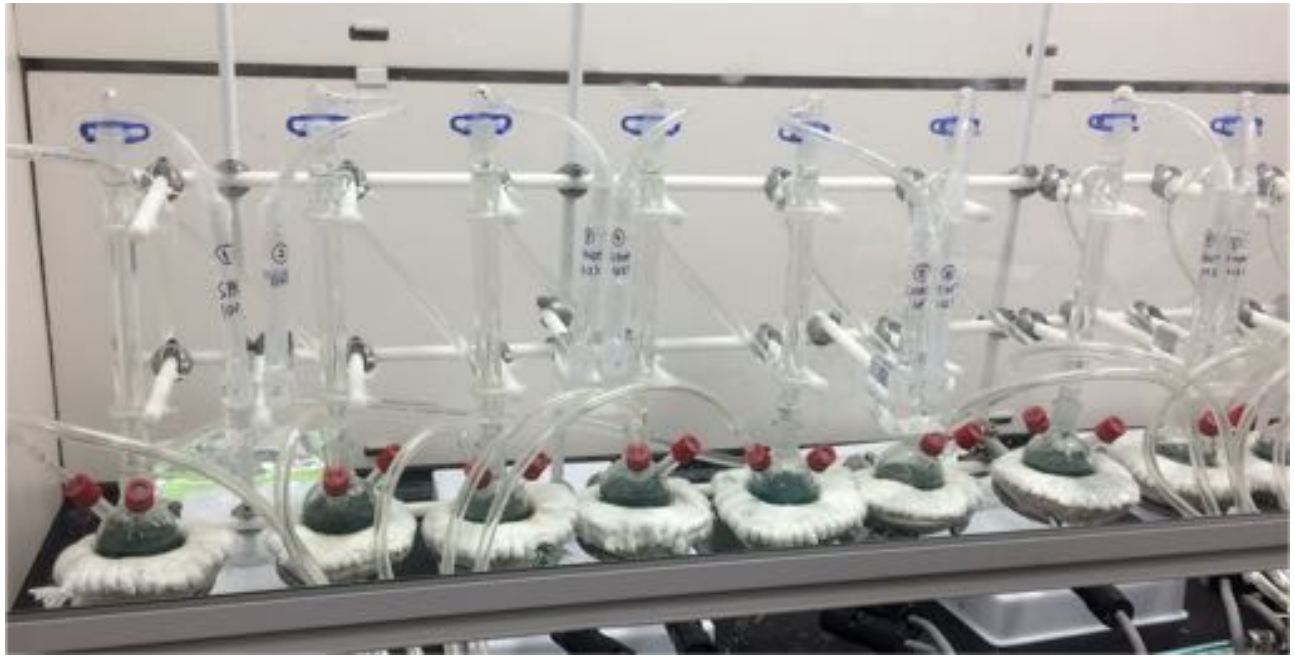
Role of environmental change in the evolution of early animal ecosystems.



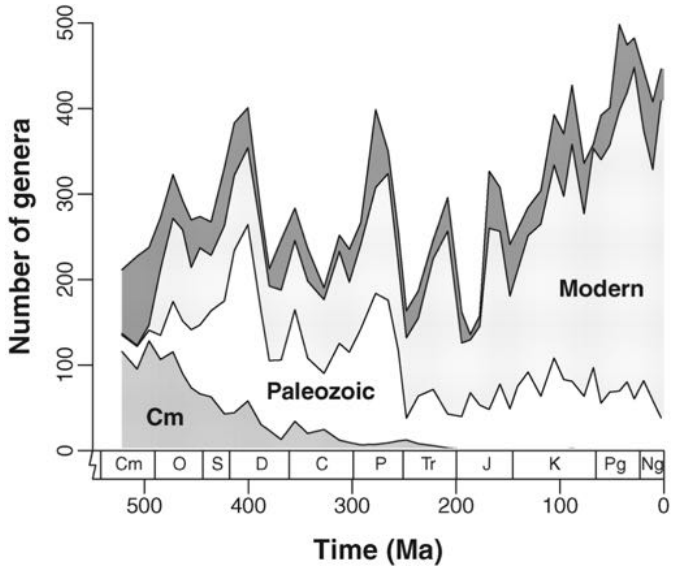
Lyons et al. 2014: Atmospheric oxygen through time

Focus: Neoproterozoic through Paleozoic, where questions remain about the magnitude and timing of environmental change.

Approach: Sedimentary Geochemistry



Iron speciation
Redox-sensitive trace elements
Organic carbon
Isotopes



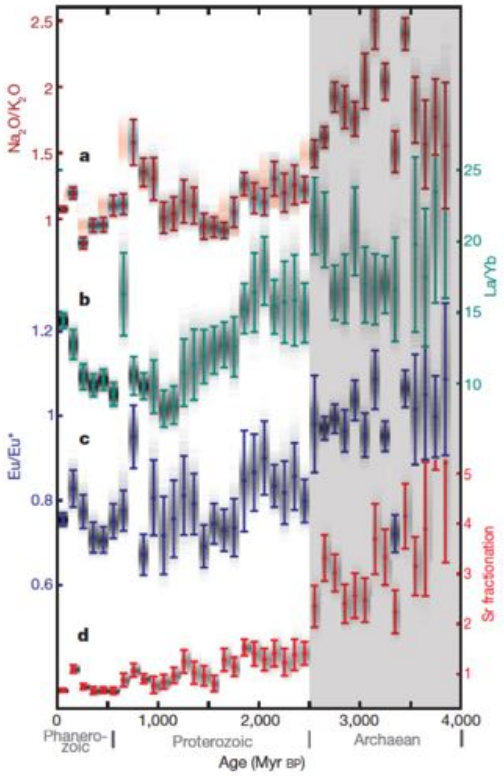
Alroy 2010: Sampling-standardized Phanerozoic diversity curve for the three marine evolutionary faunas.

Modelled after paleobiological studies

and work by other geochemical groups

e.g. Keller and Schoene, 2012

~70,000 samples from existing sources (including EarthChem portal)



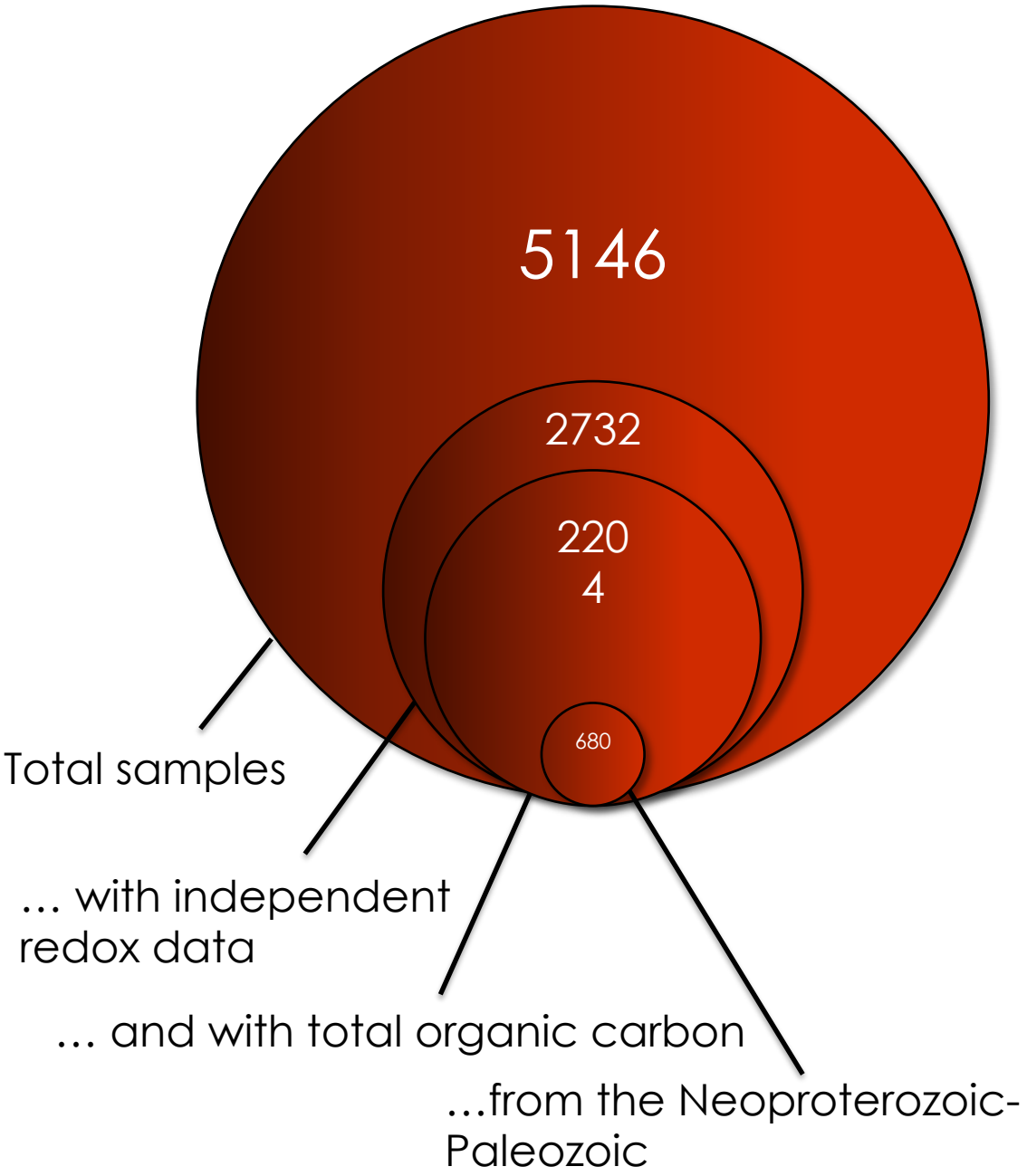
Keller and Schoene 2012: Secular compositional evolution of felsic lithologies.

Not only need large amounts of data but also

comprehensive suite of data per sample

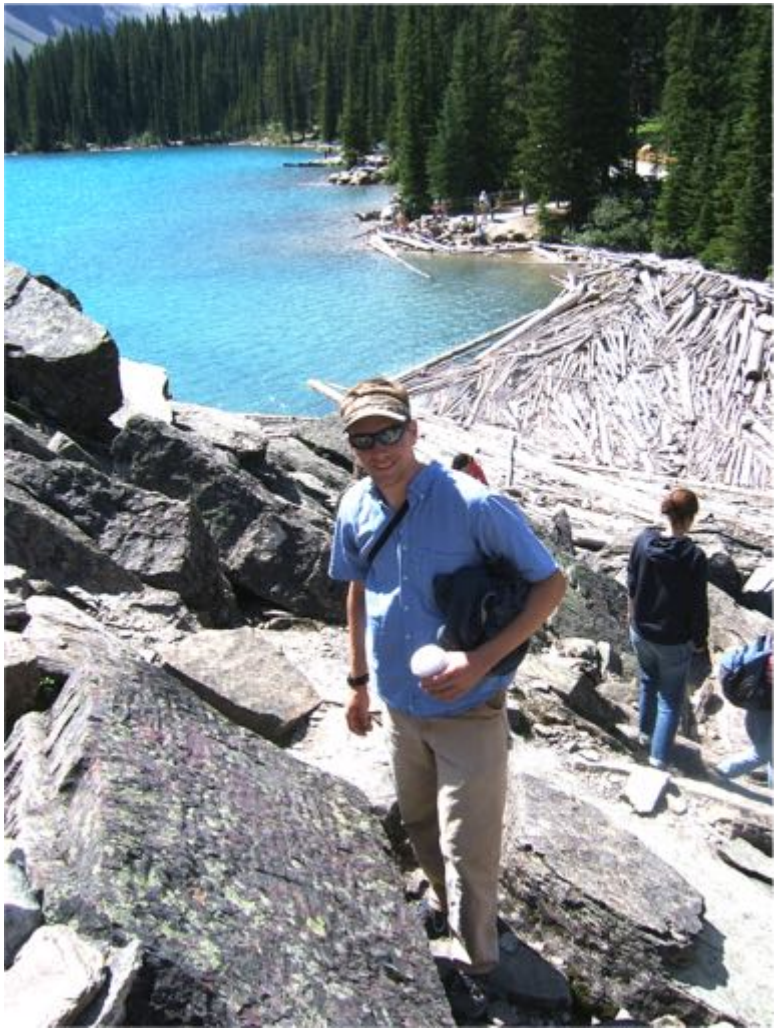
good quality geological and geographical context

e.g. Study of Partin et al. 2013, looked at all published uranium (key element of interest) through time….



Historical Geobiology Lab

August 2015



Dr. Erik Sperling

David Johnston

Noah Planavsky

... et al.

REQUIREMENTS

Establish and maintain good data standards

Easily query e.g. I would like to know the **total iron content** of all **shale** samples from the **Cambrian.**

Upload new datasets as they are acquired, either from our lab or from collaborators.

Easily update existing records e.g. I originally analyzed these 100 samples for iron geochemistry, now I would like to add this trace element data.

Export CSV/excel files e.g. for analysis in R.

Accommodate multiple users

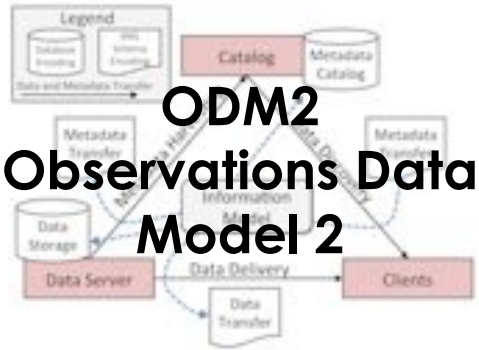
EXISTING SYSTEMS

Earthchem: portal to data from PetDB, SedDB, GEOROC, NavDat, USGS, and GANSEKI (Lehnert et al. 2000)

Geobiodiversity DB: section-based online database system, adopted by the International Commission on Stratigraphy

ODM: "An information model for spatially-discrete, feature based earth observations"

Hsu, L. et al., (2017). Enhancing Interoperability and Capabilities of Earth Science Data using the Observations Data Model 2 (ODM2). Data Science Journal. 16(1), p.4. DOI: http://doi.org/10.5334/dsj-2017-004



Open source

Good documentation

PostGIS for geographic data

Used by the Stanford Paleobiology Lab (Noel Heim) for body-size database

Data model inspired by:

ODM2

EarthChem

Specify 6

Arctos
Collaborative Collection Management Solution

Controlled vocabularies:

EarthChem

ODM2

Macrostrat
A collaborative platform for geological data exploration and integ
1,474 regional columns, 33,903 rock units, 1,750,044 geologic m

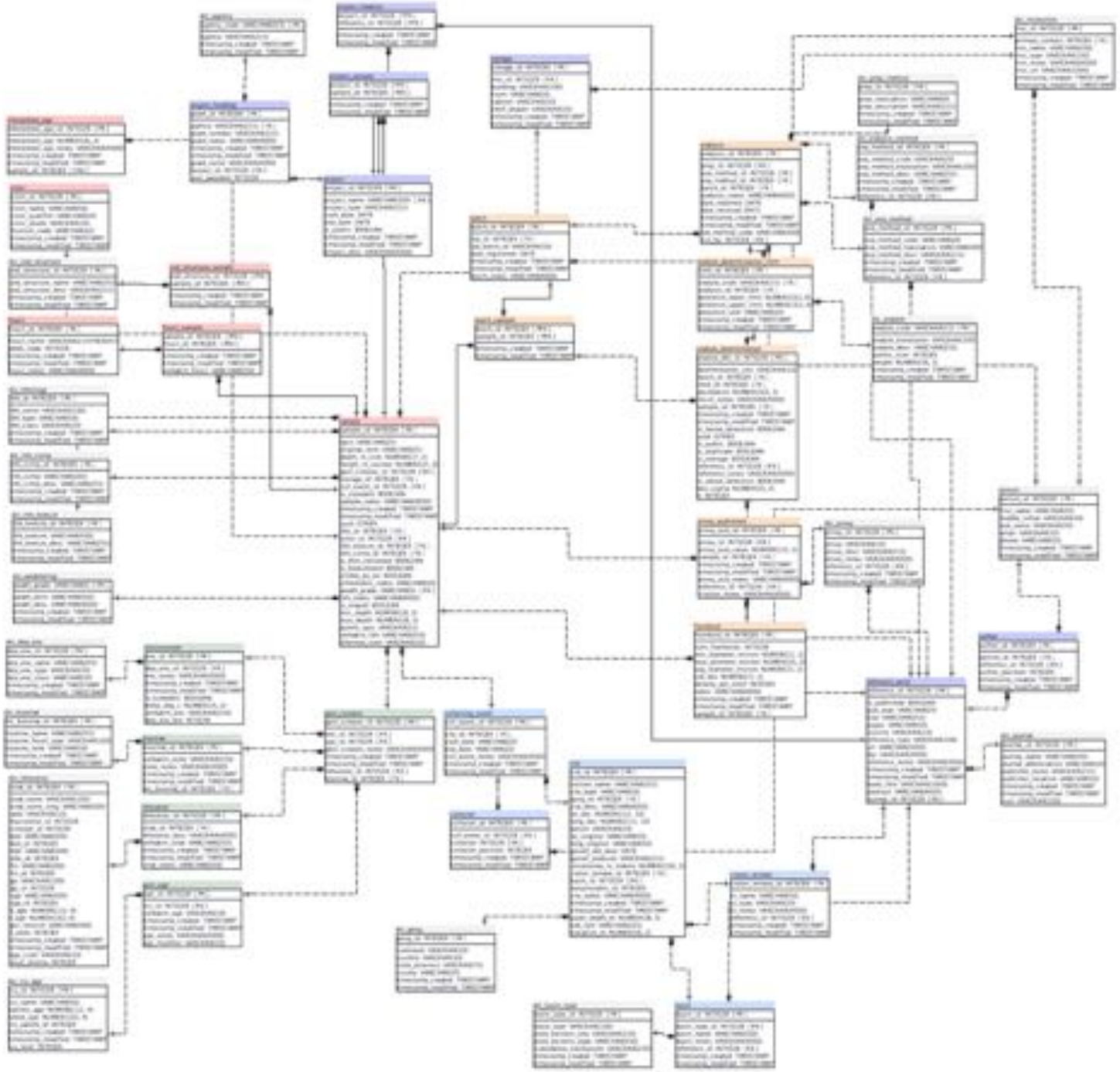
ICS

AAPG, and many more.

Database IDs stored:

Macrostrat
A collaborative platform for geological data exploration and integration.
1,474 regional columns, 33,903 rock units, 1,750,044 geologic map polygons

Weblex Canada
Australian Stratigraphic Units DB

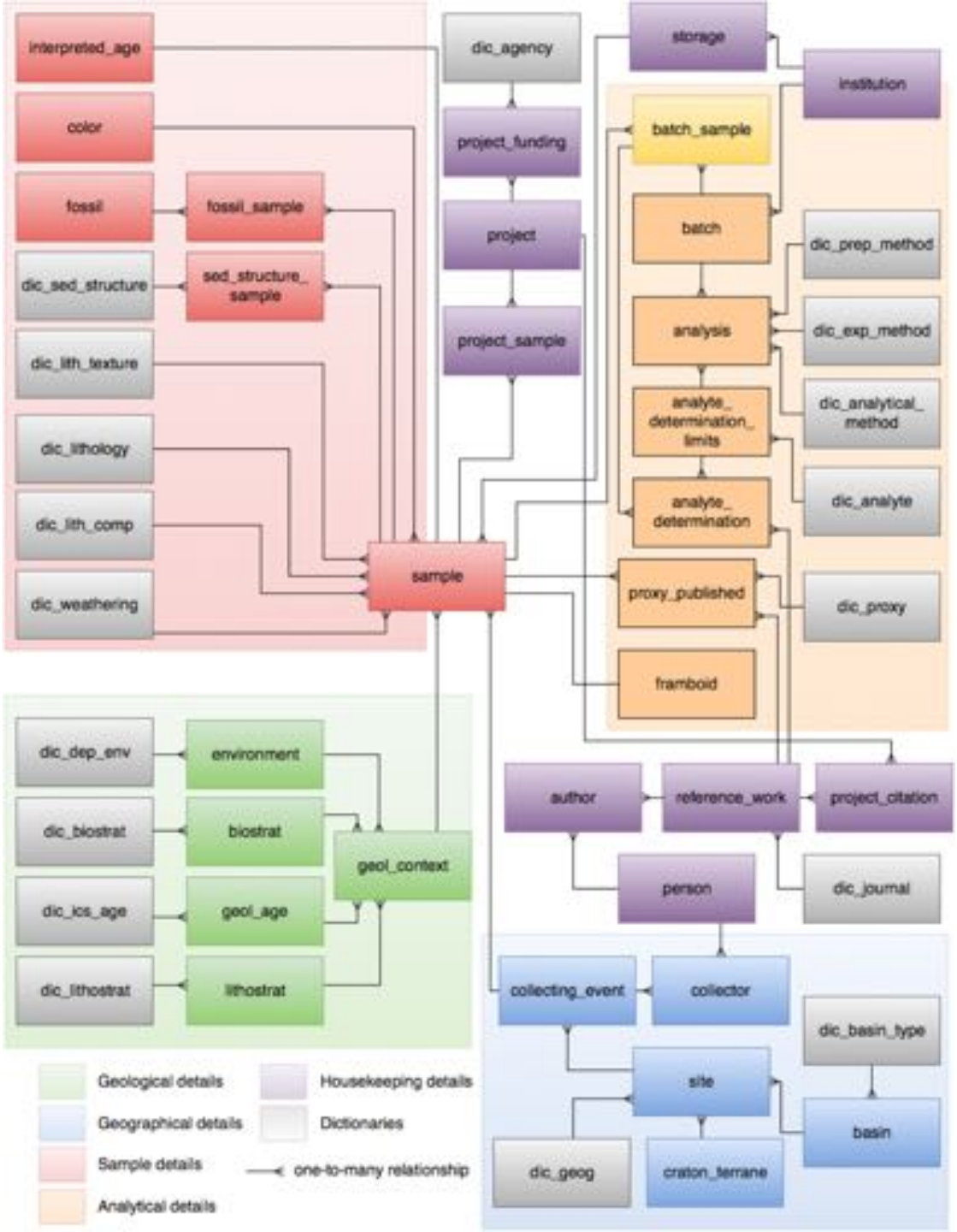


Sample

Analytical

Geological context

Geographical



SAMPLE

Color (Munsell Rock Color chart for new samples)

Lithology (Macrostrat, modified)

- Lithological texture (silty, clayey, sandy etc.)
- Lithological composition (calcareous etc.)

Sedimentary structures

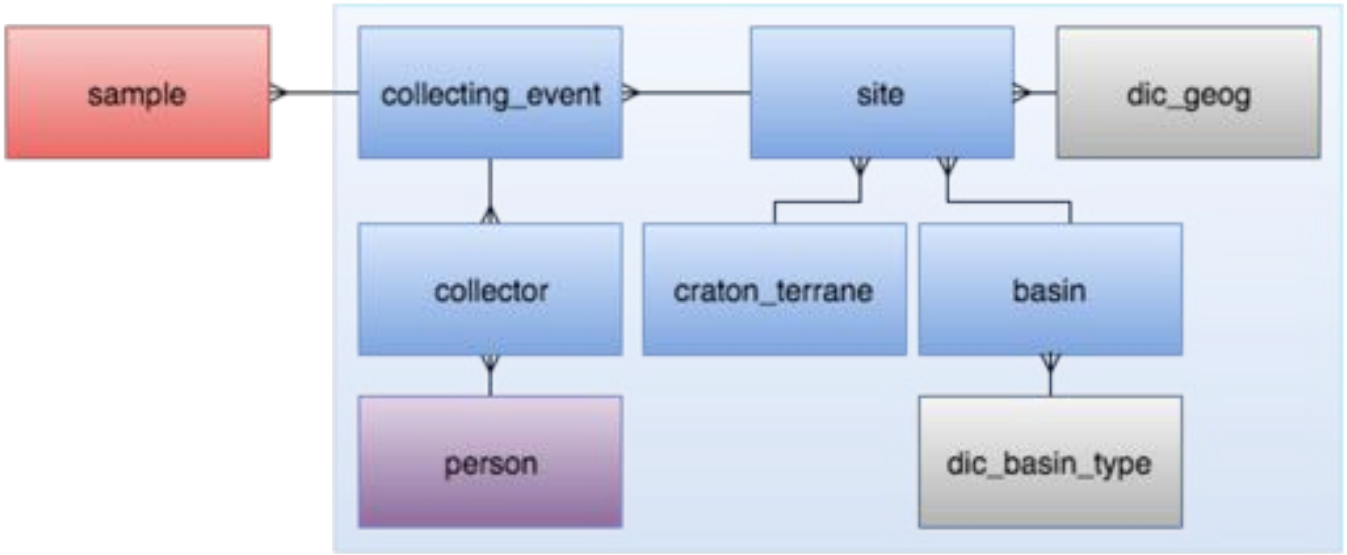
Fossils (with Paleobiology Database ID)

Weathering

Bioturbation

Interpreted age (absolute age)

GEOGRAPHICAL CONTEXT



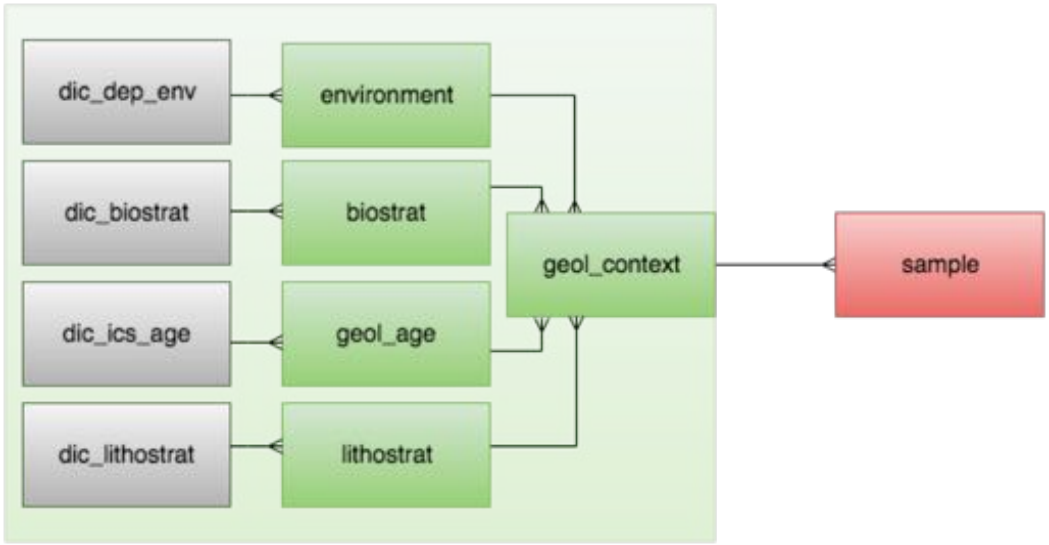
Site – section name, description, lat-long (+ georeferencing details where applicable)

Low-temperature metamorphic bin

Basin (Robertson Tellus Sedimentary Basins)

Higher geography – Country, State/Province, County (NGA GEOnet Names Server (GNS))

GEOLOGICAL CONTEXT

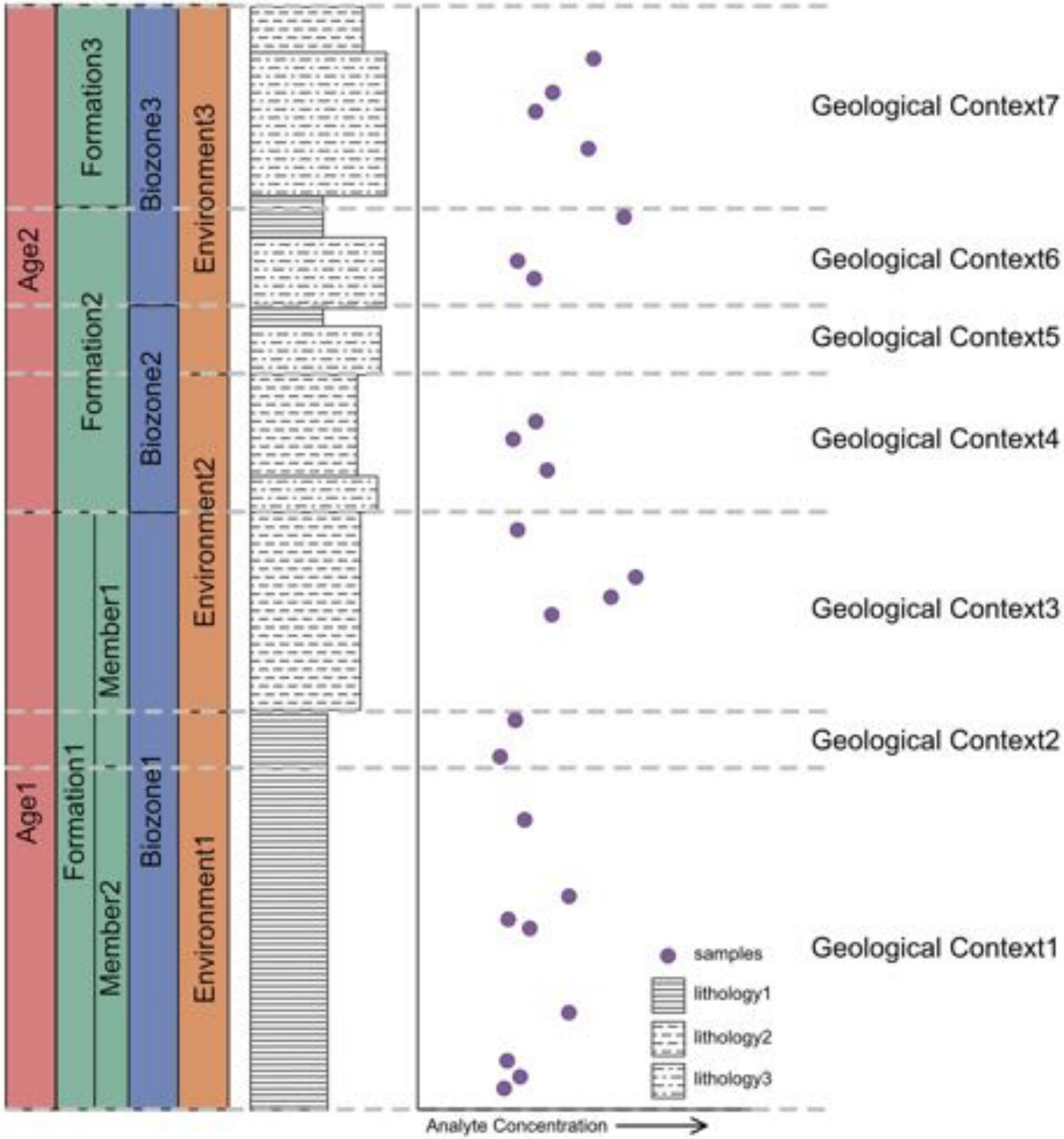


Paleoenvironment (Macrostrat, modified and sensu Sperling et al. 2015)

Biostratigraphy

Geological age (International Commission on Stratigraphy)

Geological Unit name (Macrostrat, with additions) - including Macrostrat ID, British Geological Survey code, Australian Stratigraphic Units code, Canadian Weblex code….


Age1
Age2
Formation1
Formation2
Formation3
Member1
Member2
Biozone1
Biozone2
Biozone3
Environment1
Environment2
Environment3
Geological Context7
Geological Context6
Geological Context5
Geological Context4
Geological Context3
Geological Context2
Geological Context1
samples
lithology1
lithology2
lithology3
Analyte Concentration

ANALYTICAL

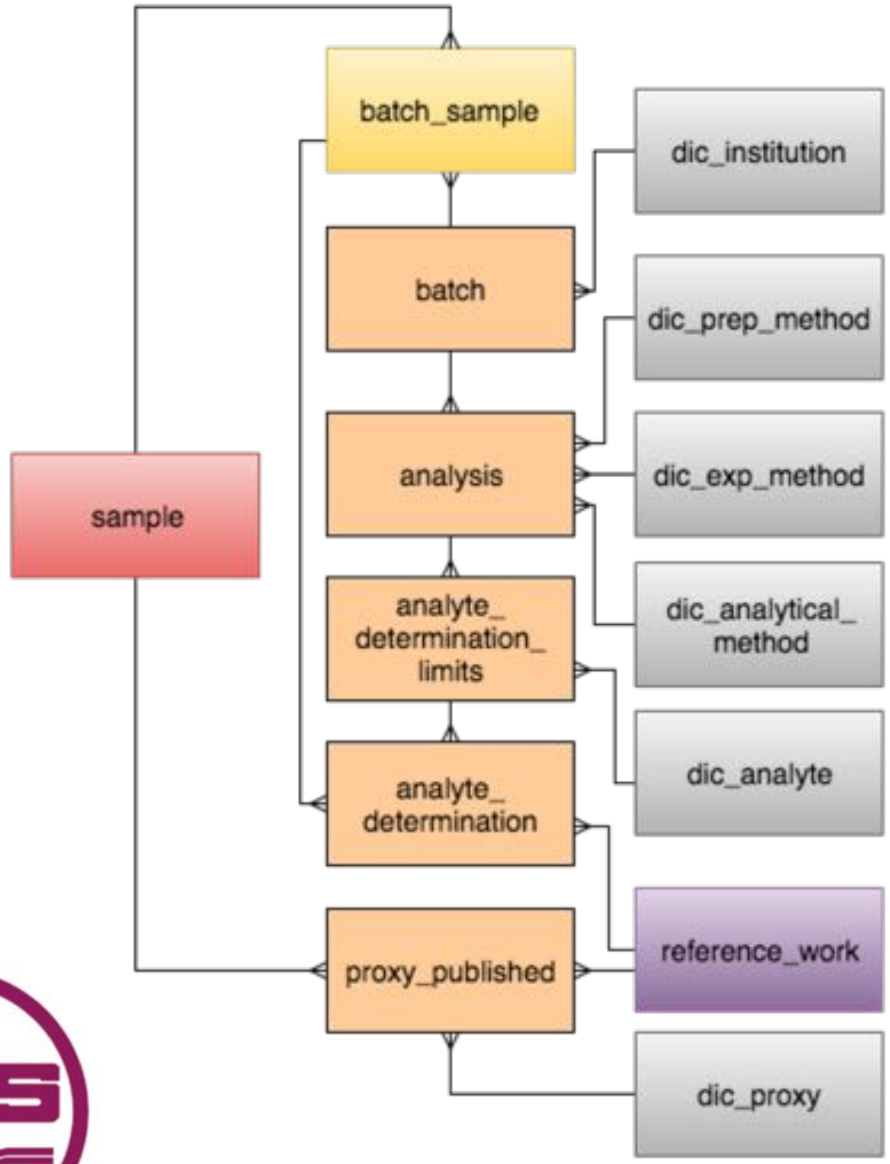
Preparation method (e.g. powdered using tungsten-carbide shatterbox)

Experimental method (e.g. 48hr acetate extraction)

Analytical method (e.g. Ferrozine spectrophometry)

Analytes (e.g. TOC, Fe-carb etc.)

Proxies (e.g. FeHR/FeT) – published and calculated



TRACKING SAMPLES: a major challenge

TRACKING SAMPLES: a major challenge

New data published from same samples

TRACKING SAMPLES: a major challenge

New data published from same samples

Previously published data reported a second time (not always cited)

TRACKING SAMPLES: a major challenge

New data published from same samples

Previously published data reported a second time (not always cited)

Same samples given different sample numbers in different publications

TRACKING SAMPLES: a major challenge

New data published from same samples

Previously published data reported a second time (not always cited)

Same samples given different sample numbers in different publications

Sample identifiers we store:

Sample_id (database-generated)

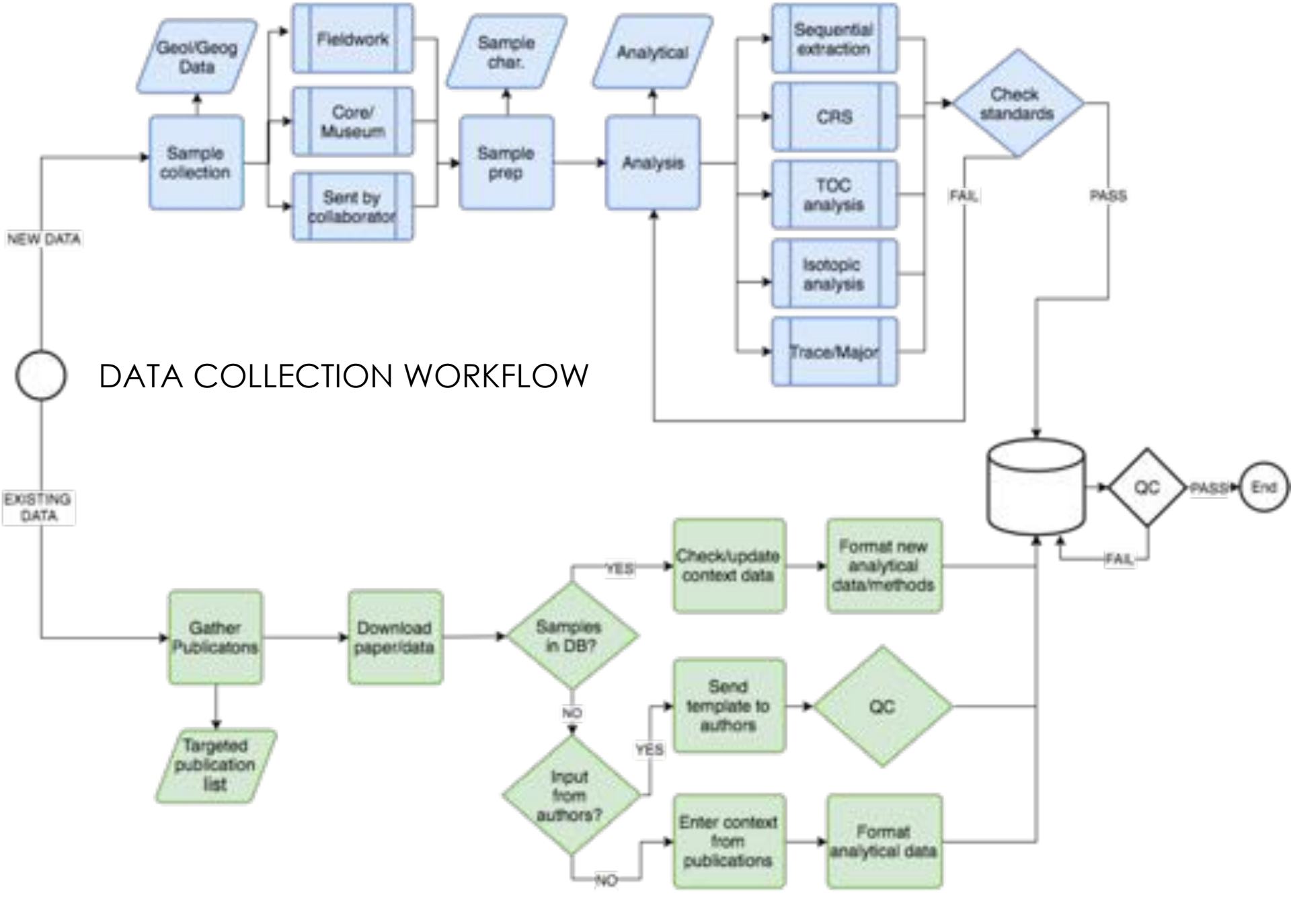
Original sample number (in our case: section name+meter)

Alternate numbers

IGSN (International Geosample Number)

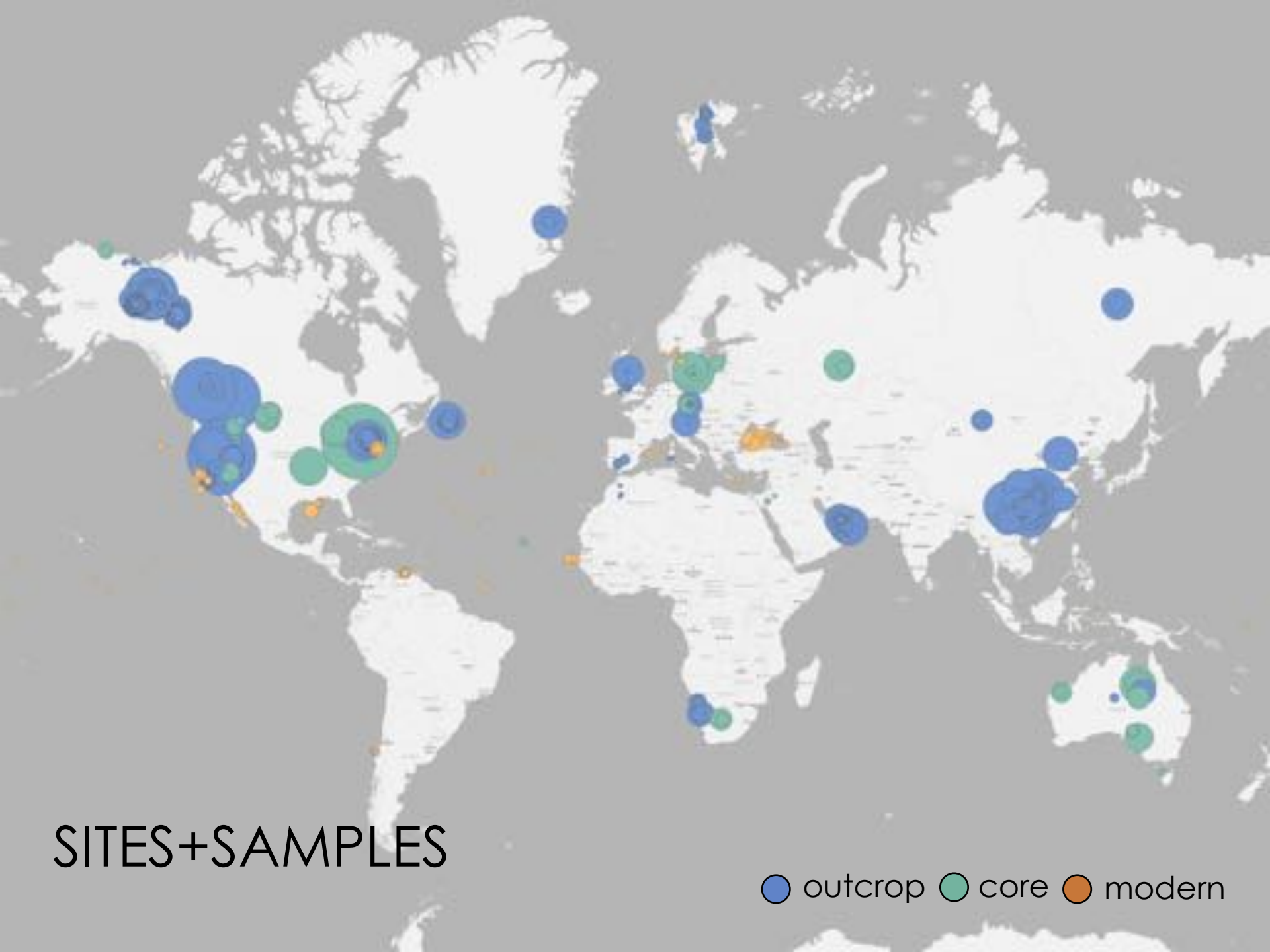
UUID (database-generated)

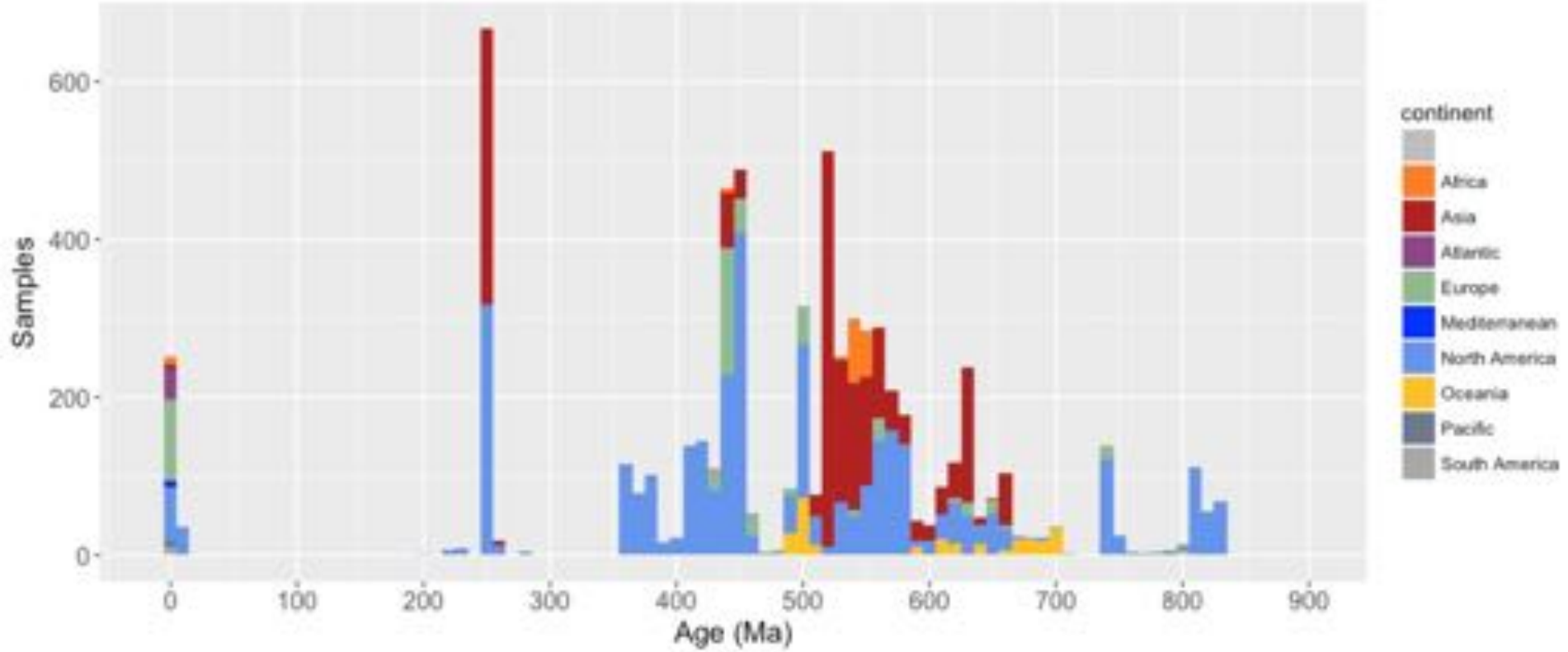
DATA COLLECTION WORKFLOW



Contributed context data and/or new analyses

COLLABORATORS


SITES+SAMPLES
outcrop
core
modern



7731 samples, 150455 unique results

FUTURE WORK

User interface - working with Stanford IT

IGSN

Publication

Sharing data

Suggestions?

Acknowledgements

Erik Sperling

David Johnston (Harvard)
Noah Planavsky (Yale)

Austin Miller (Stanford)

Stanford Historical Geobiology Lab group

Carl Watson (British Geological Survey)

Noel Heim (Stanford)

Collaborators, including:
Alan Rooney (Yale)
Alex Morgan (Cambridge)
Alf Lenz (Western Ontario)
Andy Knoll (Harvard)
Austin Miller (Stanford)
Ben Gill (Virginia Tech)
Brad Sageman (Northwestern)
Chao Li (China University of Geosciences)
Clint Scott (USGS)
David Loydell (Portsmouth)
Devon Cole (Yale)
Don Canfield (Southern Denmark)
Emma Hammarlund (Southern Denmark)
Emmy Smith (Johns Hopkins)
Francis Macdonald (Harvard)
Galen Halverson (McGill)
Inessa Yurchenko (Stanford)
JC Creveling (Oregon State)
Joe Schaffer (Wayfair)
Justin Strauss (Dartmouth)
Lawrence Och (Dr. von Moos AG)
Liam Bhajan (Stanford)
Malcom Hodgskiss (Stanford)
Marcus Kunzmann (CSIRO)
Matthew Clarkson (Otago)
Paulo Linarde Dantas Mascena (Federal U. of Rio Grande do Norte)
Phil Wilby (BGS)
Rachel Wood (Edinburgh)
Richard Stockey (Stanford)
Samantha Ritzer (Stanford)
Stephanie Plaza-Torres (Puerto-Rico)
Tiffani Fraser (Yukon Geological Survey)
Tom Boag (Stanford)
Will Thompson-Butler (Stanford)

Combine new and existing data for increased statistical power

Collaborative Research project

Incentive to contribute: opportunity to address exciting research questions of direct interest to the researchers.