

AI in the Paleogeosciences: Progress, Challenges, and Opportunities

Deborah Khider
Information Sciences Institute

USC Viterbi

School of Engineering

Information Sciences Institute

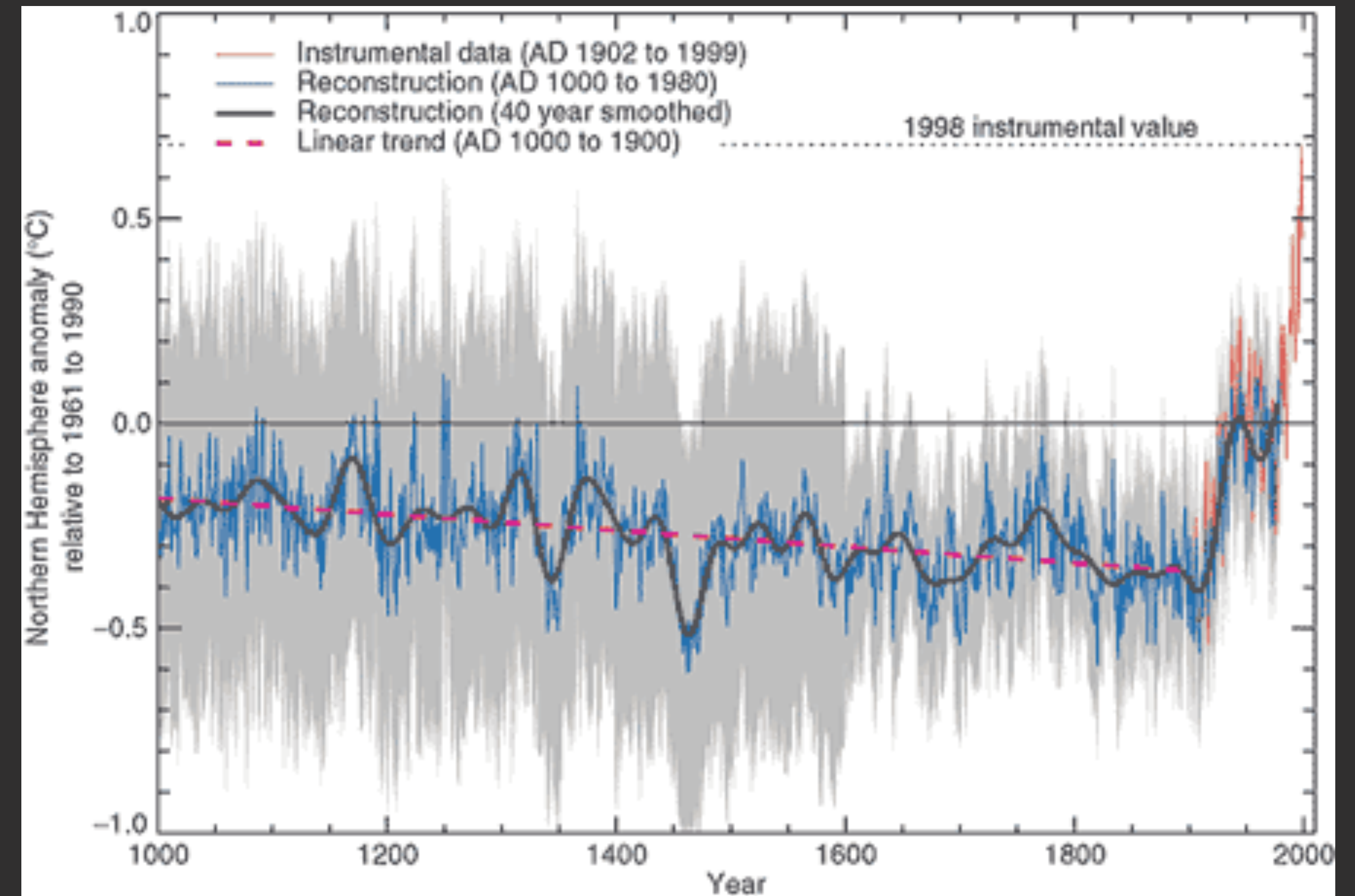
Why study paleoclimatology?

To address questions such as:

- Has climate changed in the past?
- How much?
- How often?
- How fast?
- Why?

To provide test datasets for the verification of climate models

The Hockey Stick



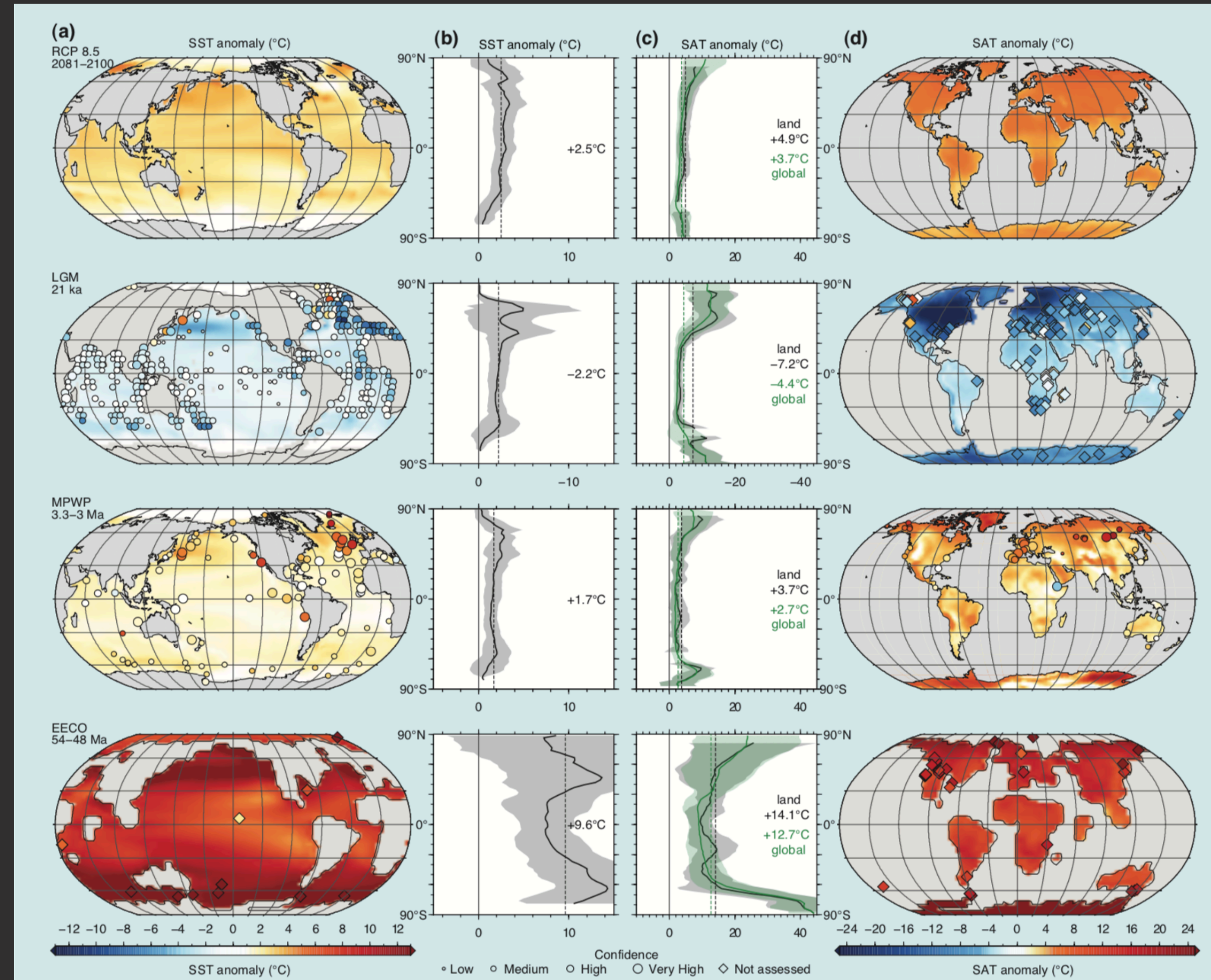
Mann et al. , 1999, Geophysical Research Letters

Why study paleoclimatology?

To address questions such as:

- Has climate changed in the past?
- How much?
- How often?
- How fast?
- Why?

To provide test datasets for the verification of climate models



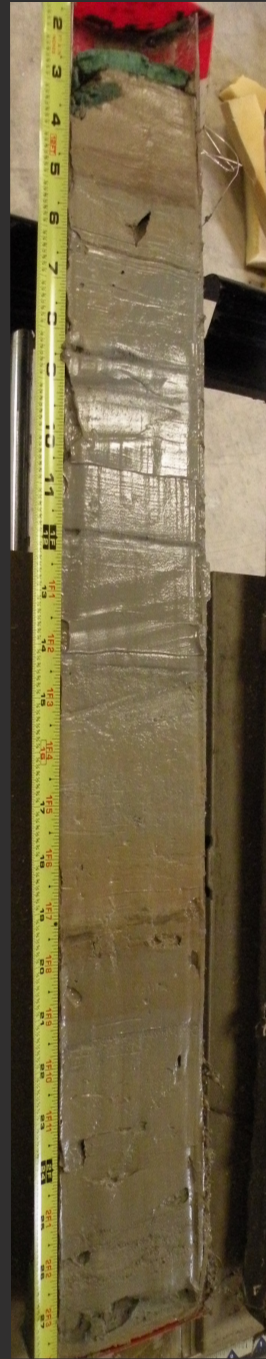
How do we answer these questions?

Past climate quantities are not measured directly

How do we answer these questions?

Past climate quantities are not measured directly

Proxy Archive



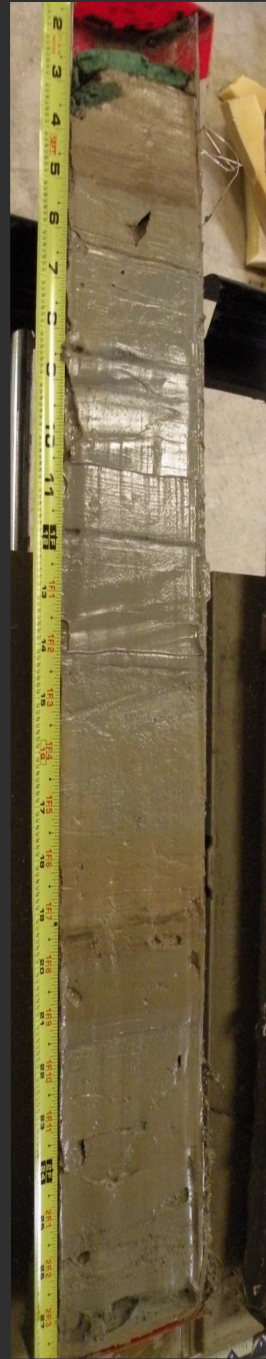
Marine Sediment

How do we answer these questions?

Past climate quantities are not measured directly

Proxy Archive

Proxy Observation



Marine Sediment

Mg/Ca

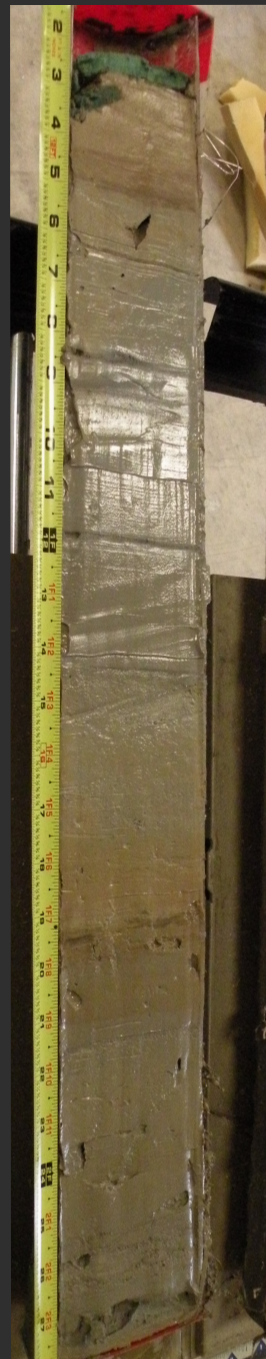
How do we answer these questions?

Past climate quantities are not measured directly

Proxy Archive

Proxy Observation

Environment



Marine Sediment

Mg/Ca

Surface Ocean

How do we answer these questions?

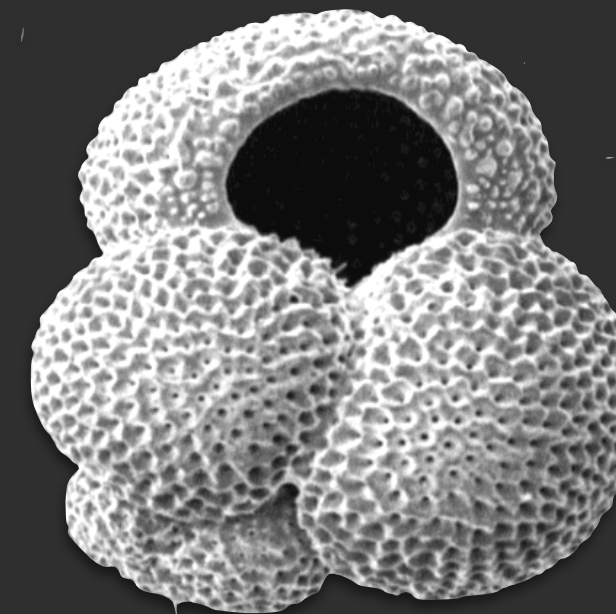
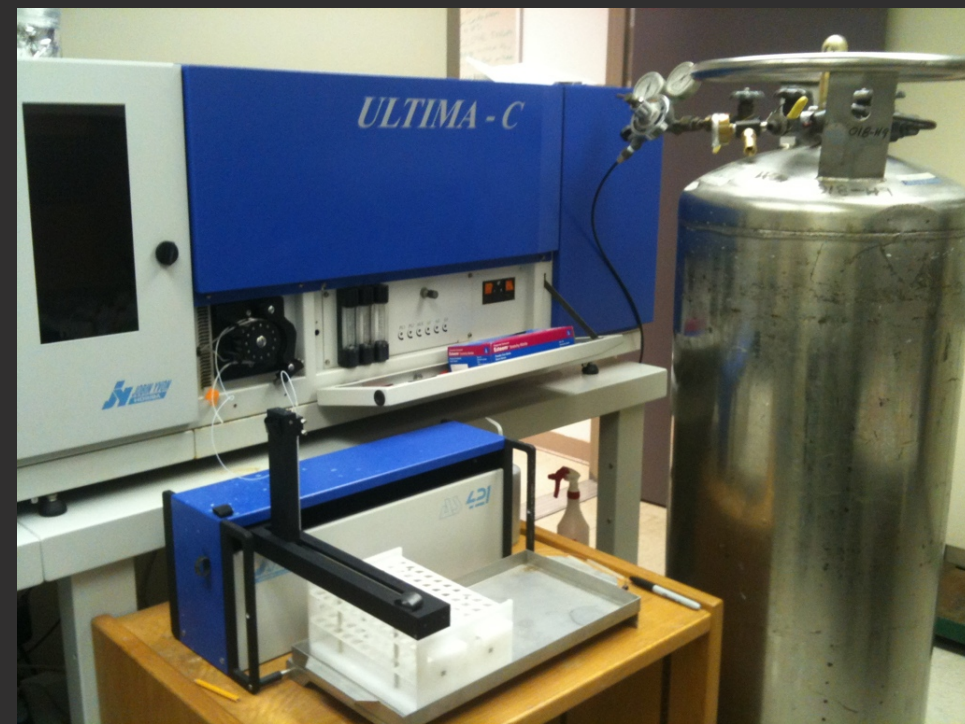
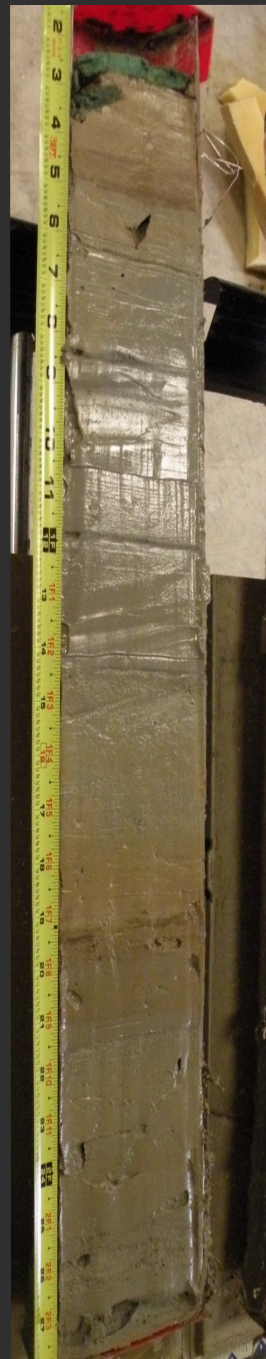
Past climate quantities are not measured directly

Proxy Archive

Proxy Observation

Proxy Sensor

Environment



Marine Sediment

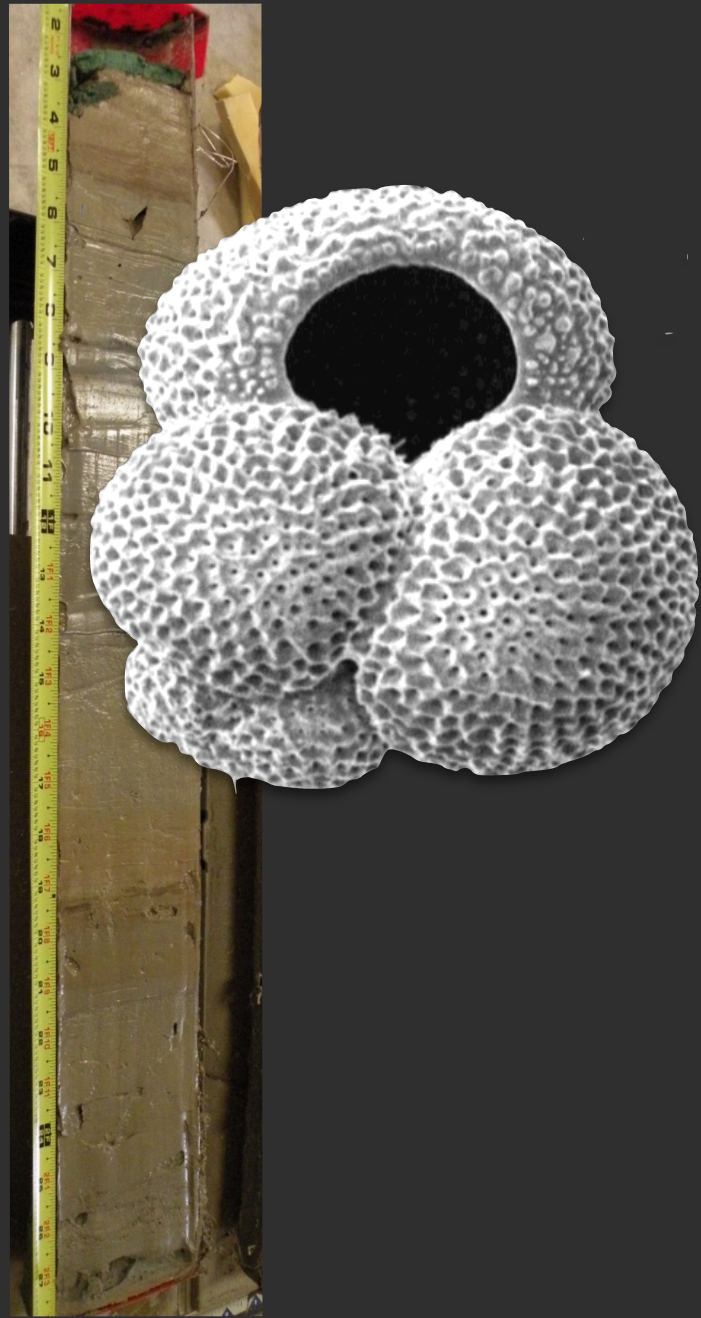
Mg/Ca

Foraminifera
Globigerinoides ruber

Surface Ocean

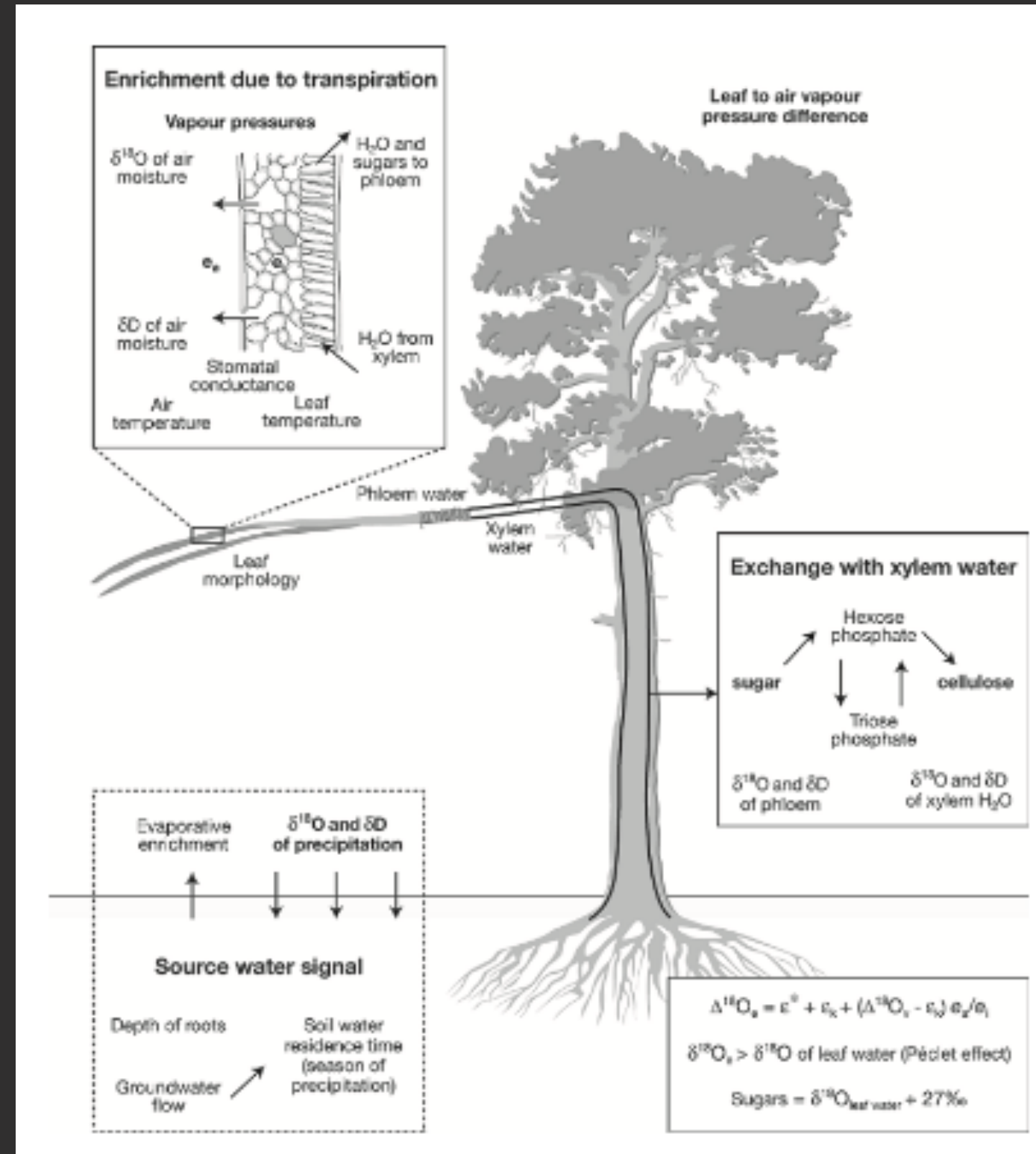
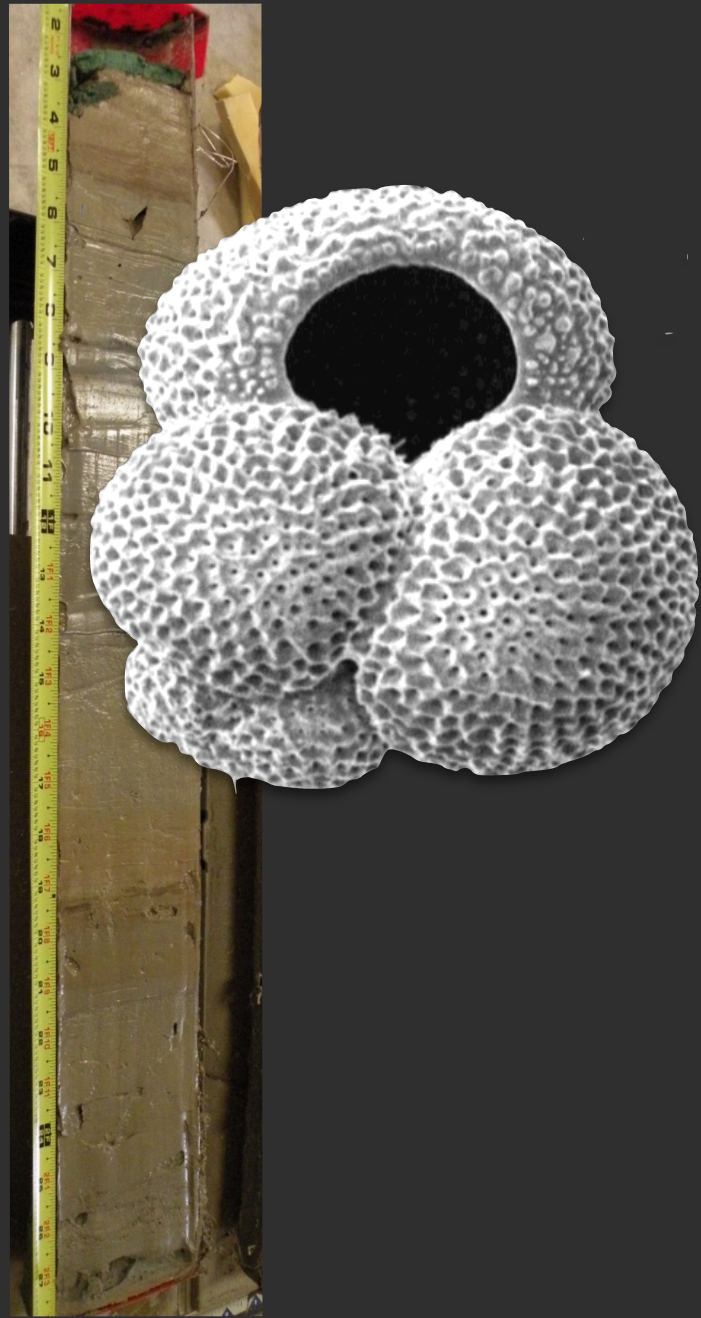
We have multitude of proxy systems!

We have multitude of proxy systems!



Archive: Marine Sediment
Sensor: Foraminifera
Observation: Mg/Ca

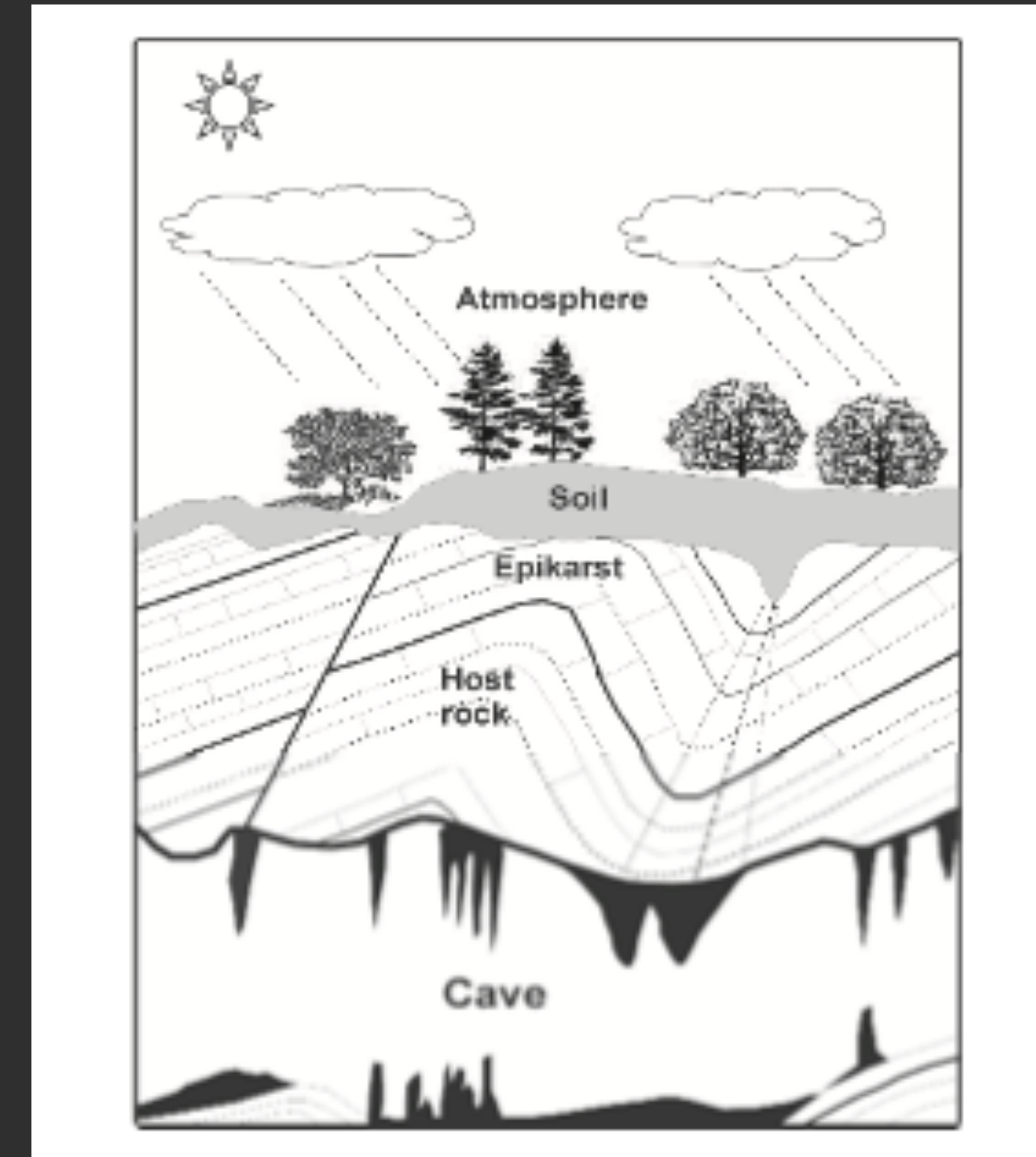
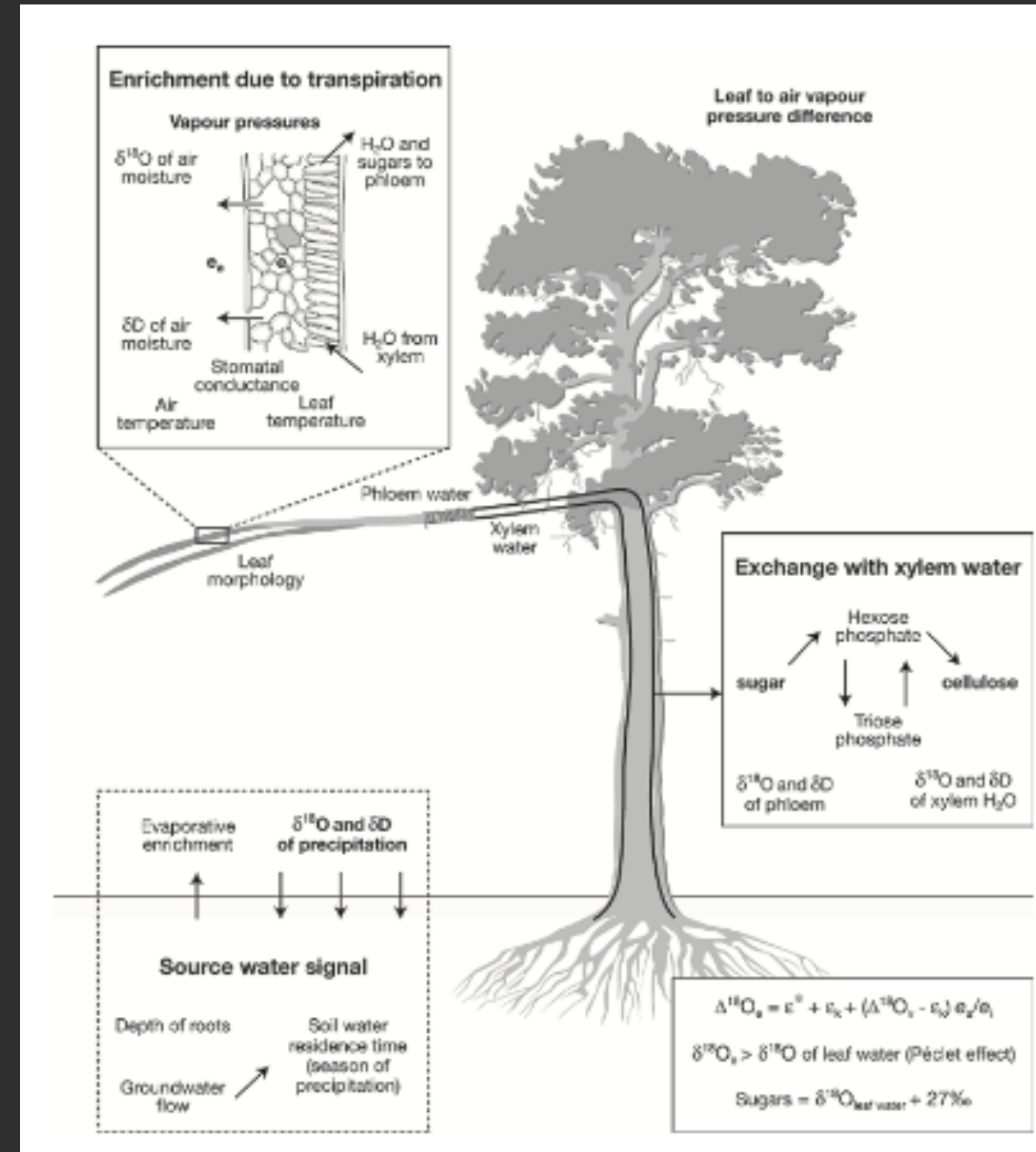
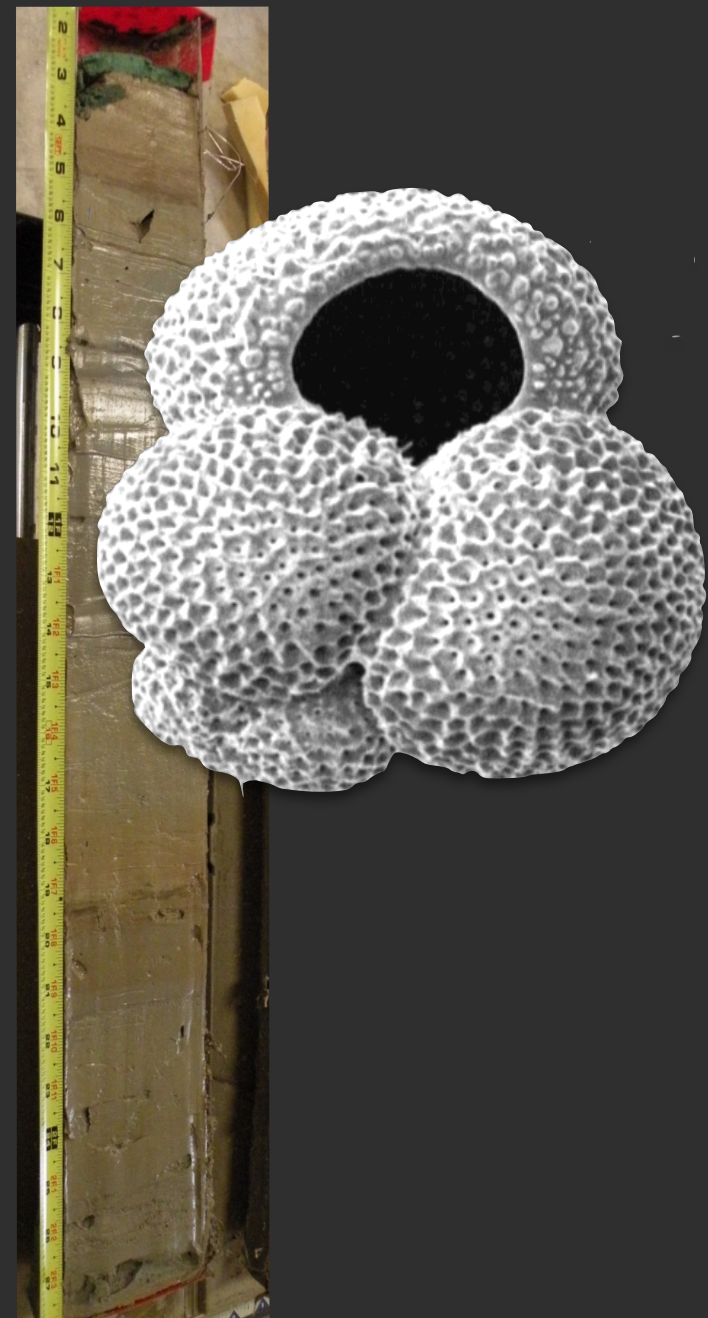
We have multitude of proxy systems!



Archive: Marine Sediment
 Sensor: Foraminifera
 Observation: Mg/Ca

Archive: Wood
 Sensor: Tree
 Observation: $\delta^{18}\text{O}$

We have multitude of proxy systems!

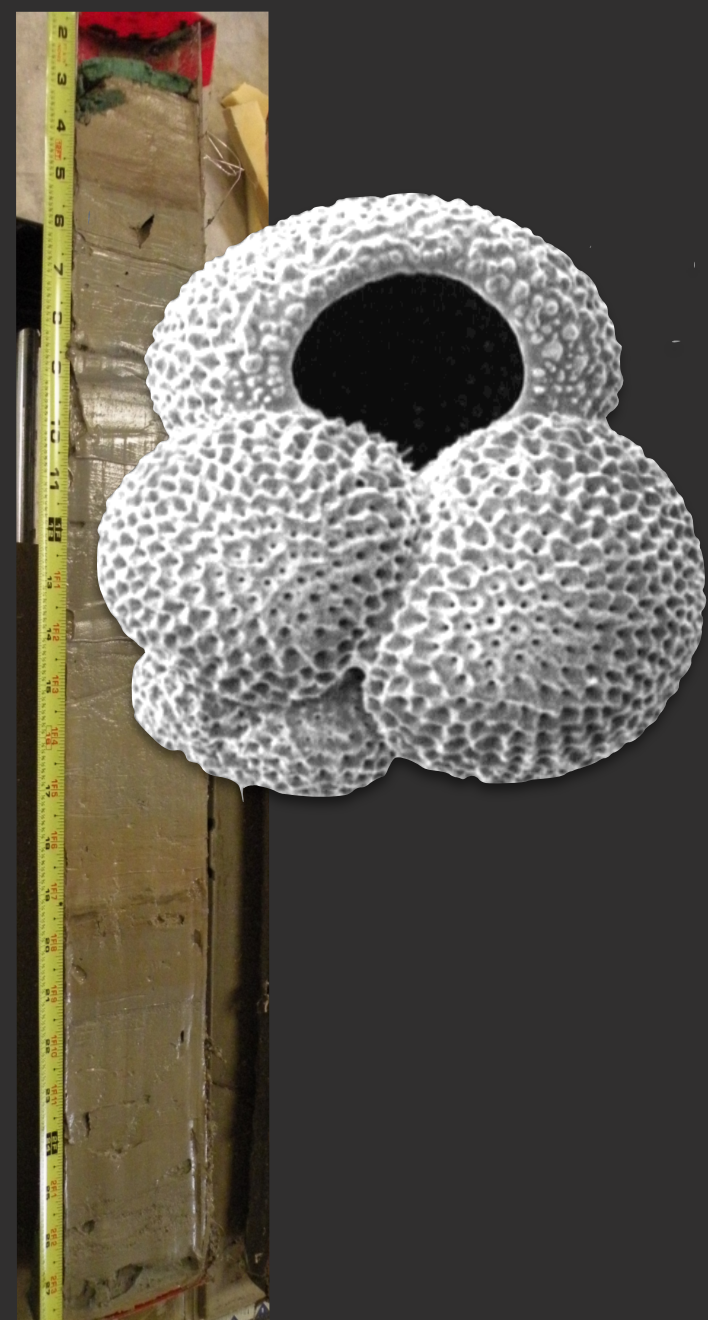


Archive: Marine Sediment
 Sensor: Foraminifera
 Observation: Mg/Ca

Archive: Wood
 Sensor: Tree
 Observation: $\delta^{18}\text{O}$

Archive: Speleothem
 Sensor: Karst
 Observation: $\delta^{18}\text{O}$

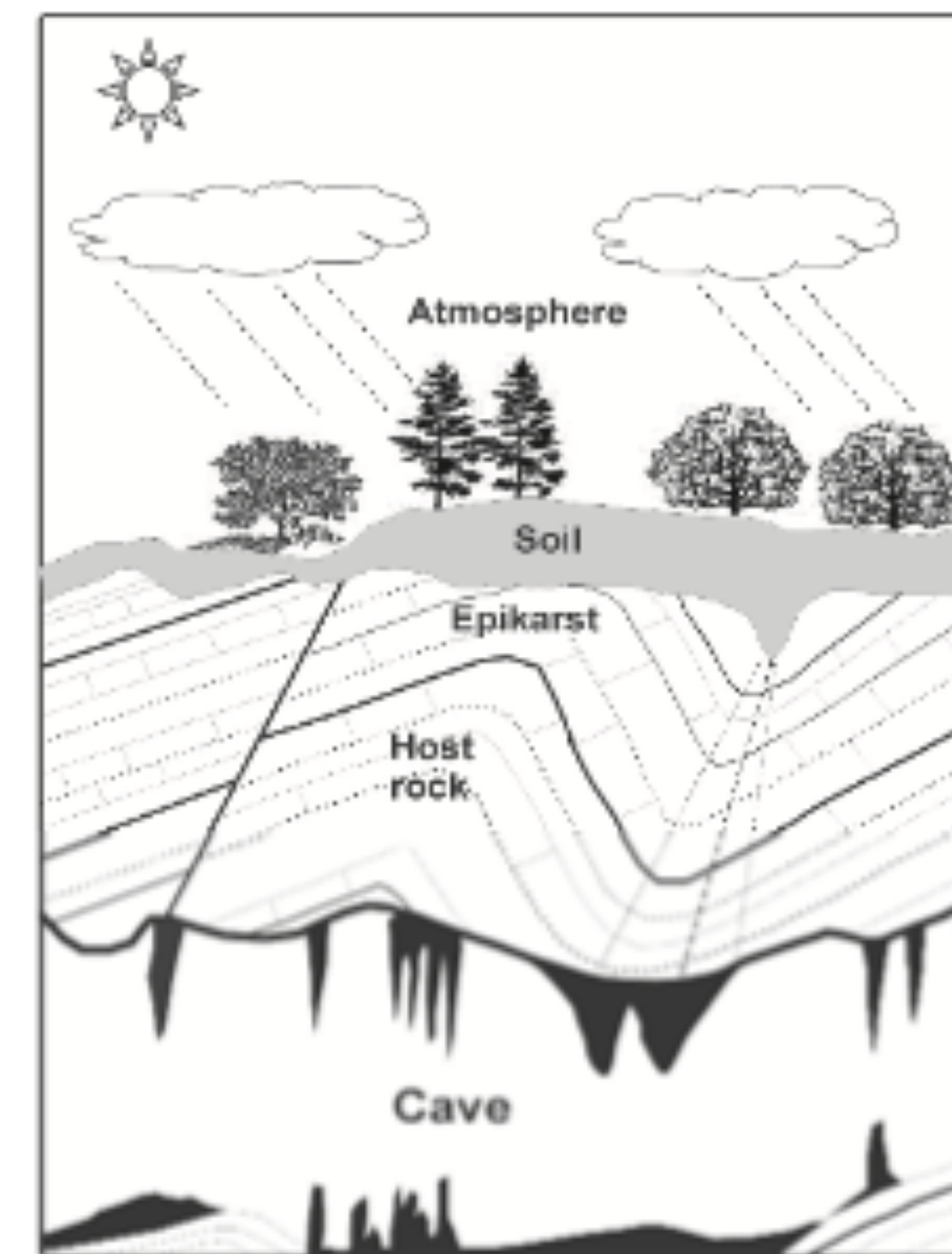
We have multitude of proxy systems!



Enrichment due to transpiration

Leaf to air vapour

1. Broad expertise
2. Extremely diverse data

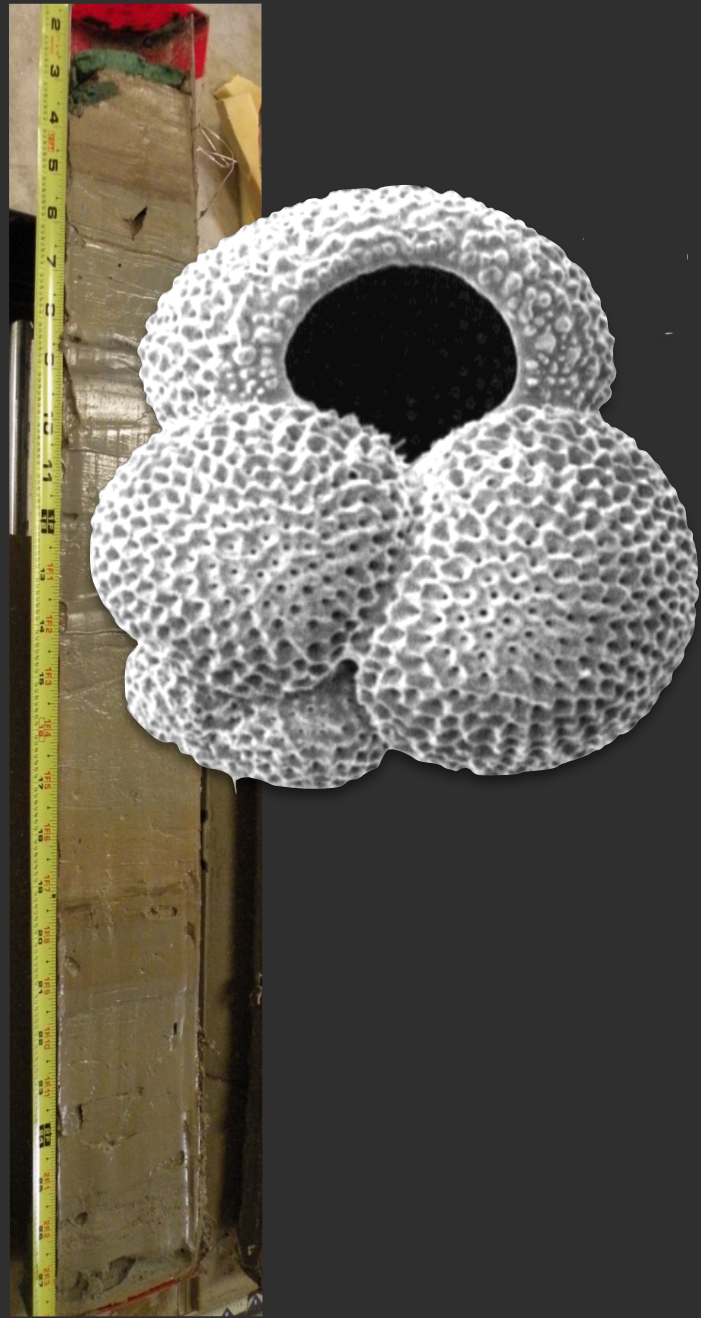


Archive: Marine Sediment
Sensor: Foraminifera
Observation: Mg/Ca

Archive: Wood
Sensor: Tree
Observation: $\delta^{18}\text{O}$

Archive: Speleothem
Sensor: Karst
Observation: $\delta^{18}\text{O}$

We have multitude of proxy systems!

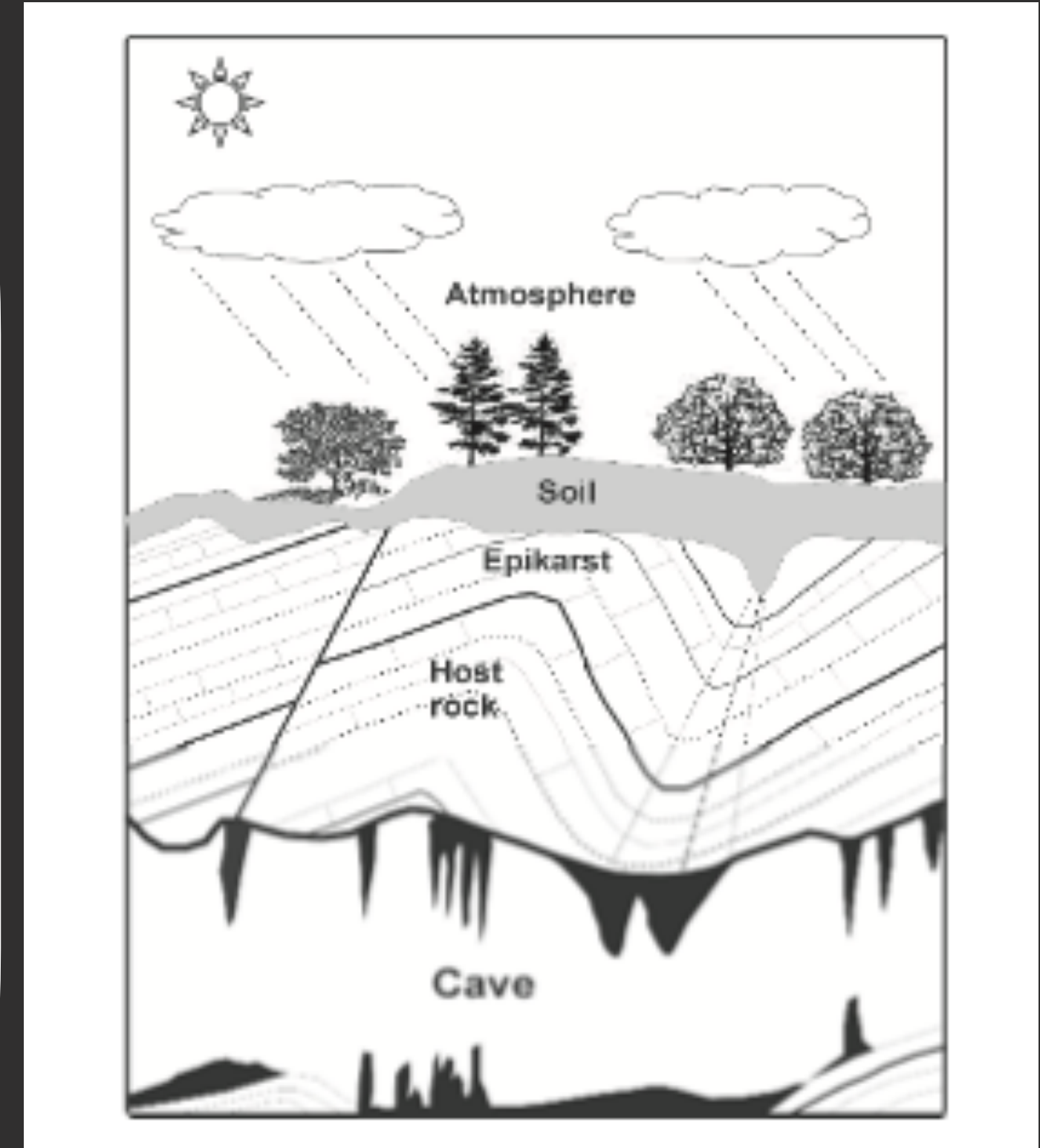


Enrichment due to transpiration

Leaf to air vapour

1. Broad expertise
2. Extremely diverse data

Long-tail approach to data collection:
each record is provided by an
individual investigator.

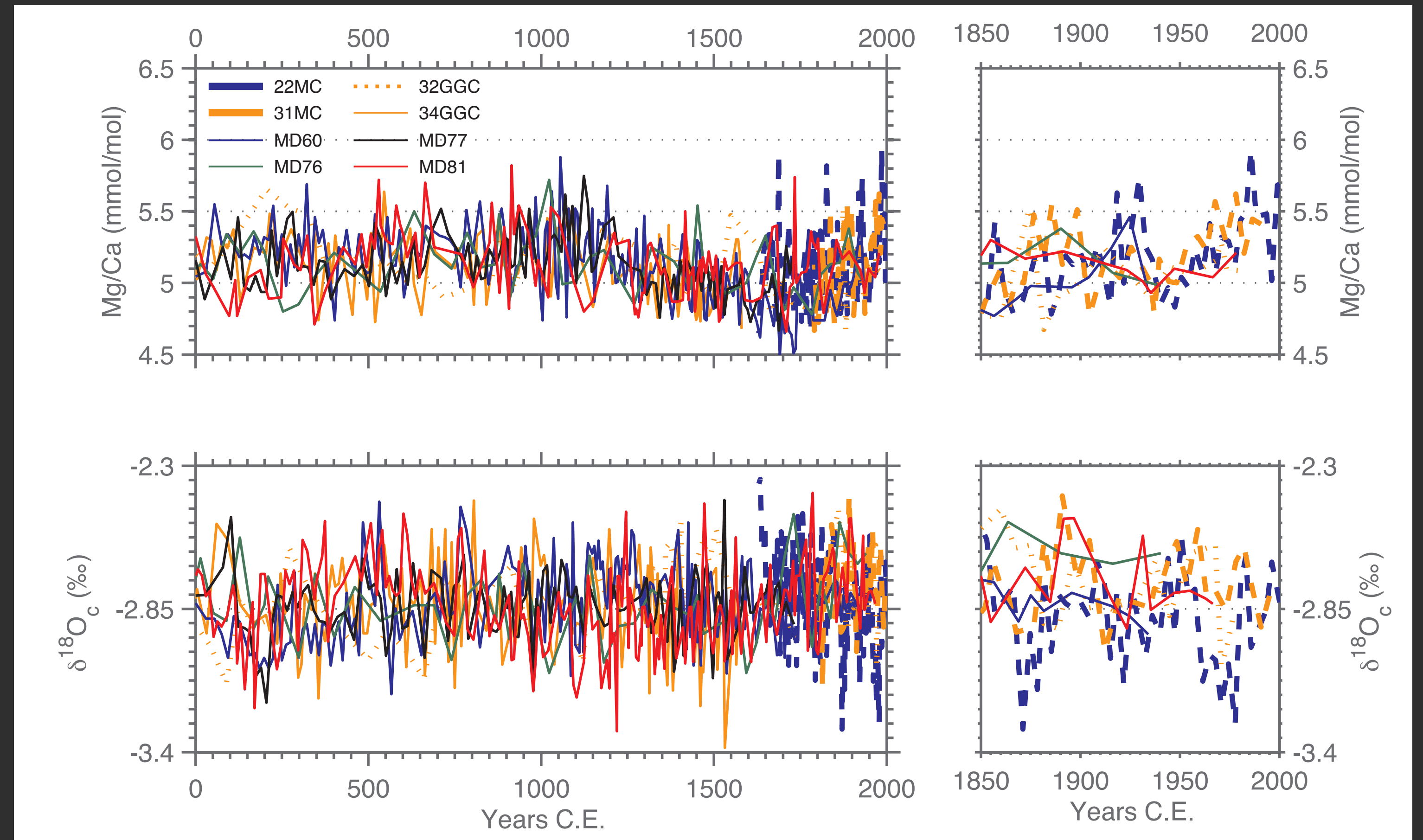
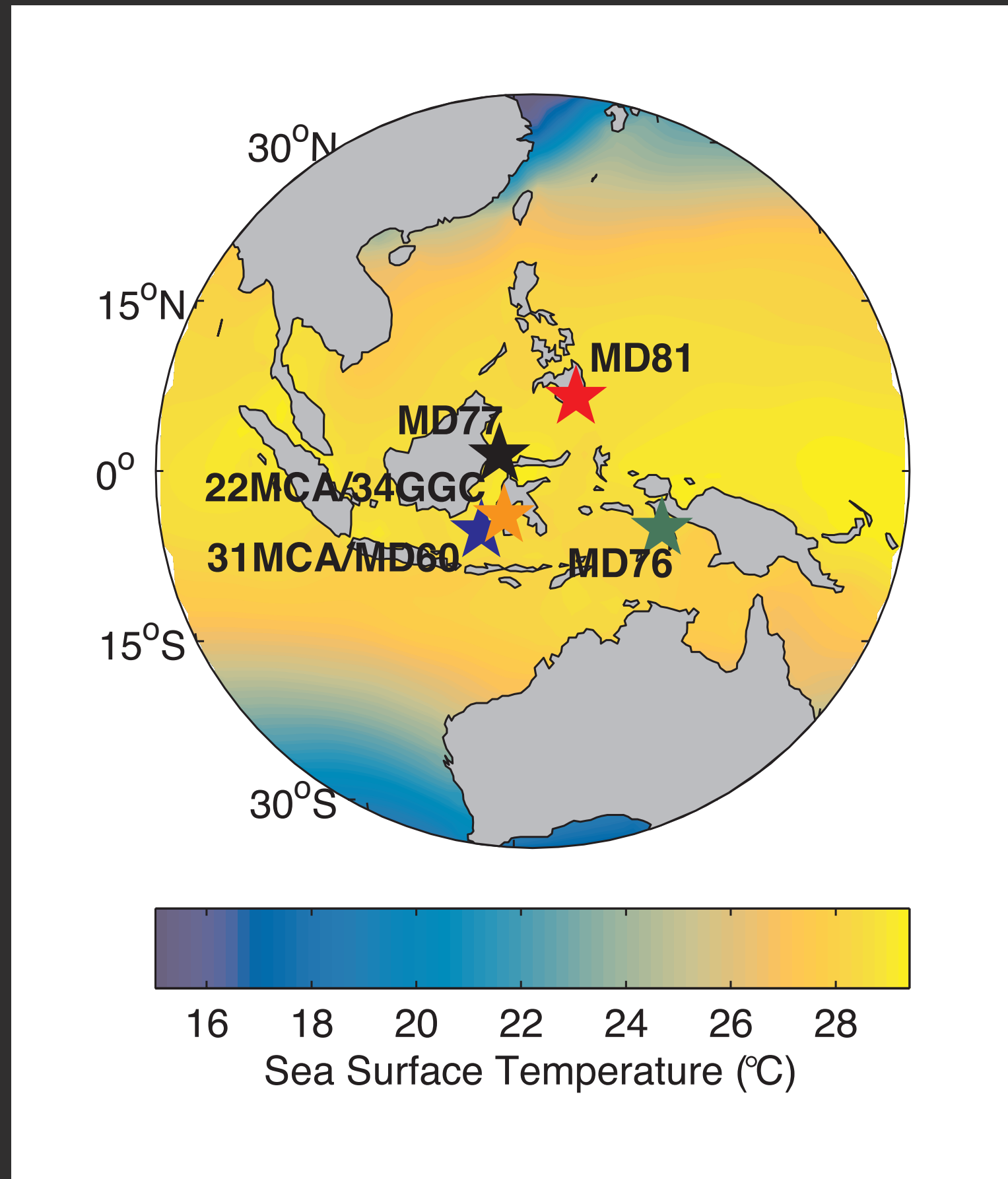


Archive: Marine Sediment
Sensor: Foraminifera
Observation: Mg/Ca

Archive: Wood
Sensor: Tree
Observation: $\delta^{18}\text{O}$

Archive: Speleothem
Sensor: Karst
Observation: $\delta^{18}\text{O}$

They offer a picture of the climate at one location through time



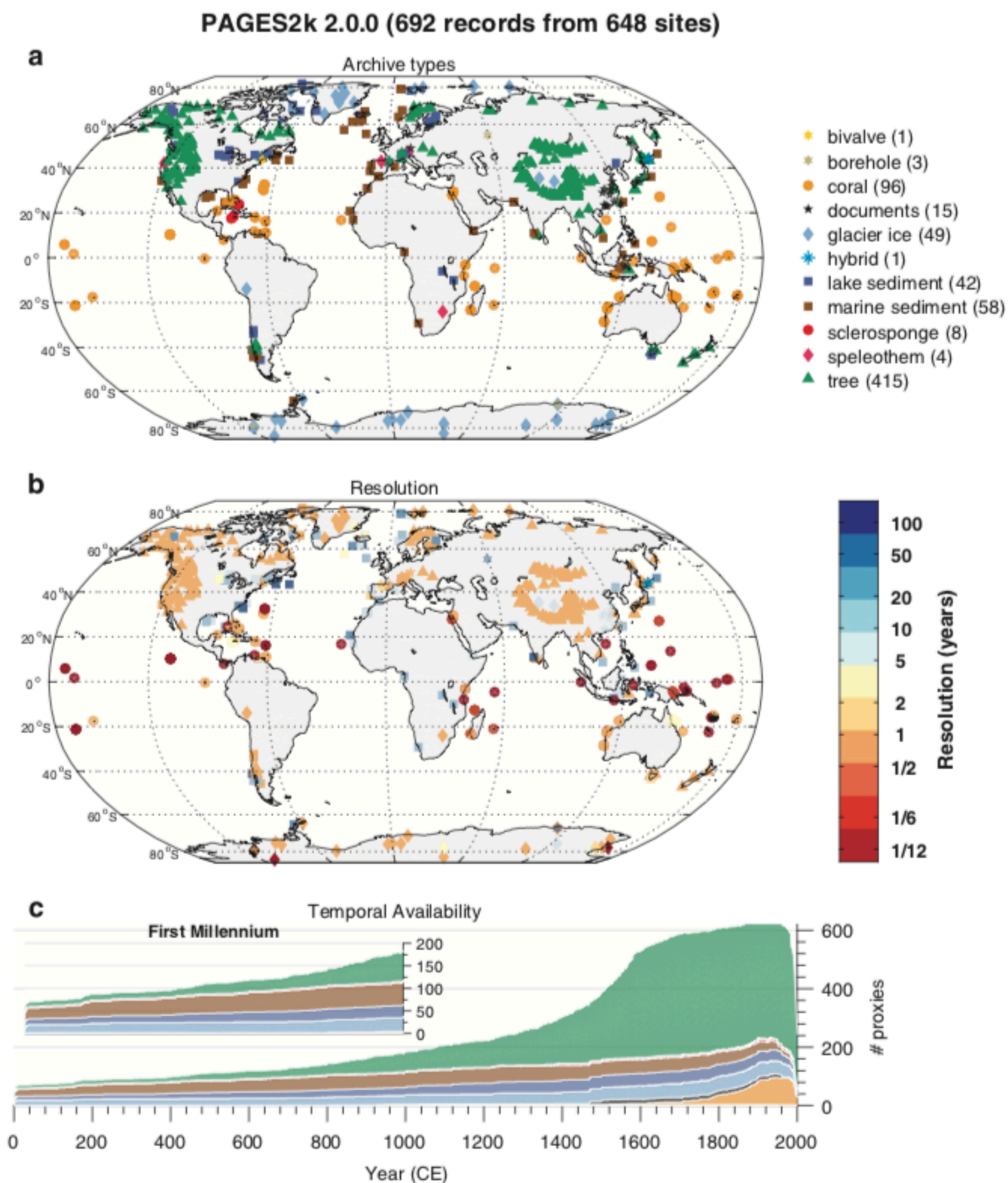


Figure 1. Spatiotemporal data availability in the PAGES2k database. (a) Geographical distribution, by archive type, coded by color and shape. (b) Temporal resolution in the PAGES2k database, defined here as the median of the spacing between consecutive observations. Shapes as in (a), colors encode the resolution in years (see colorbar). (c) Temporal availability, coded by color as in (a).

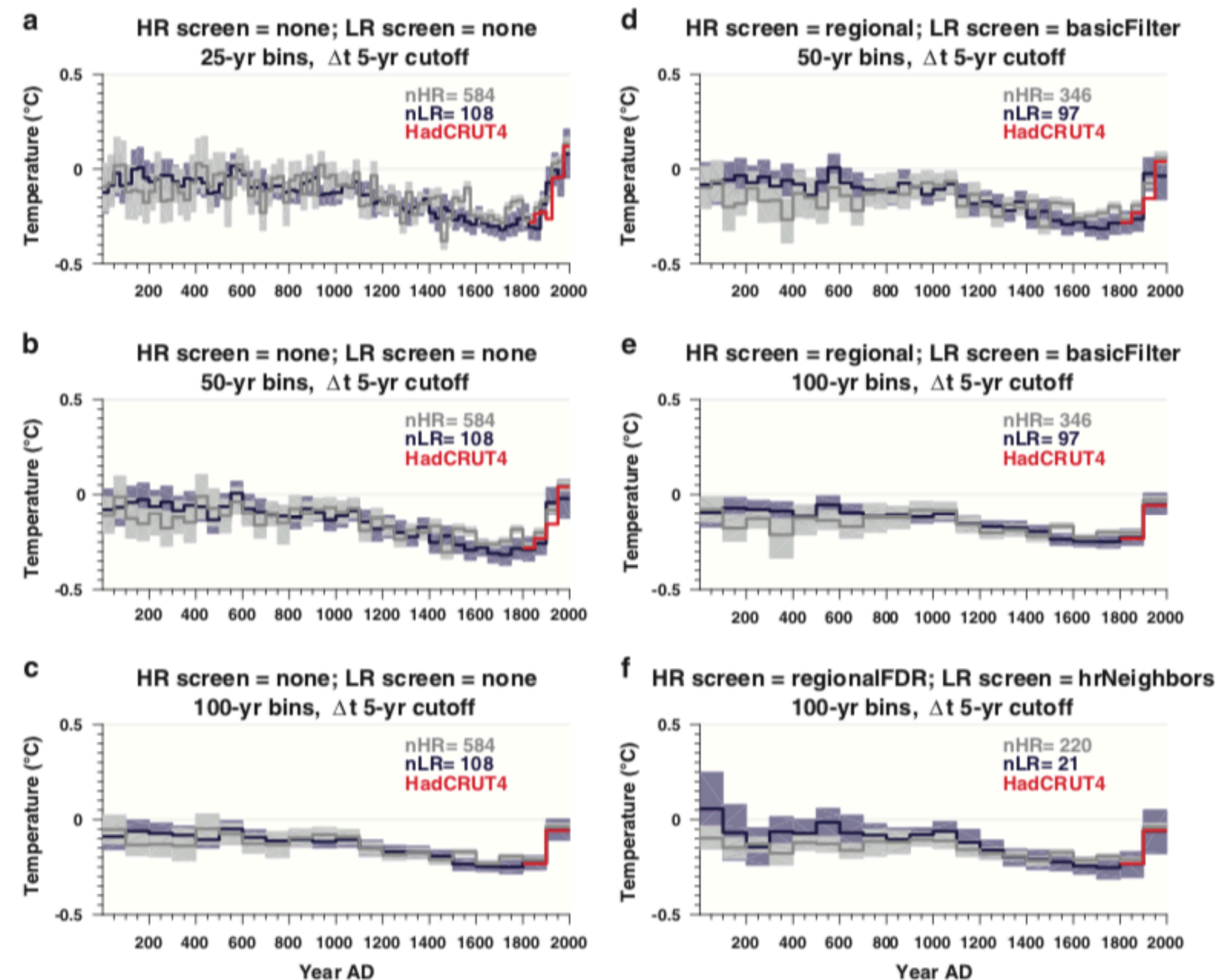


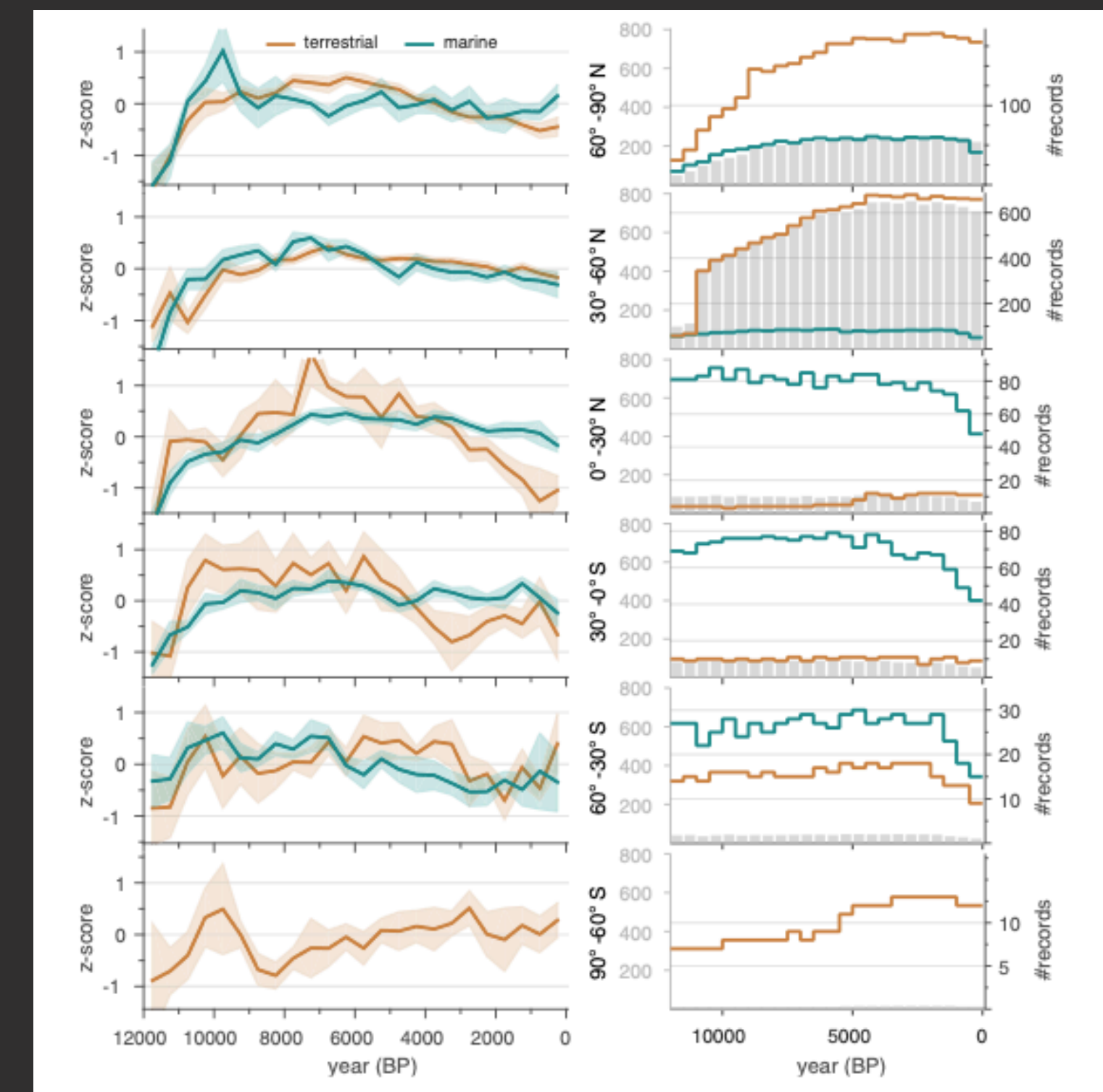
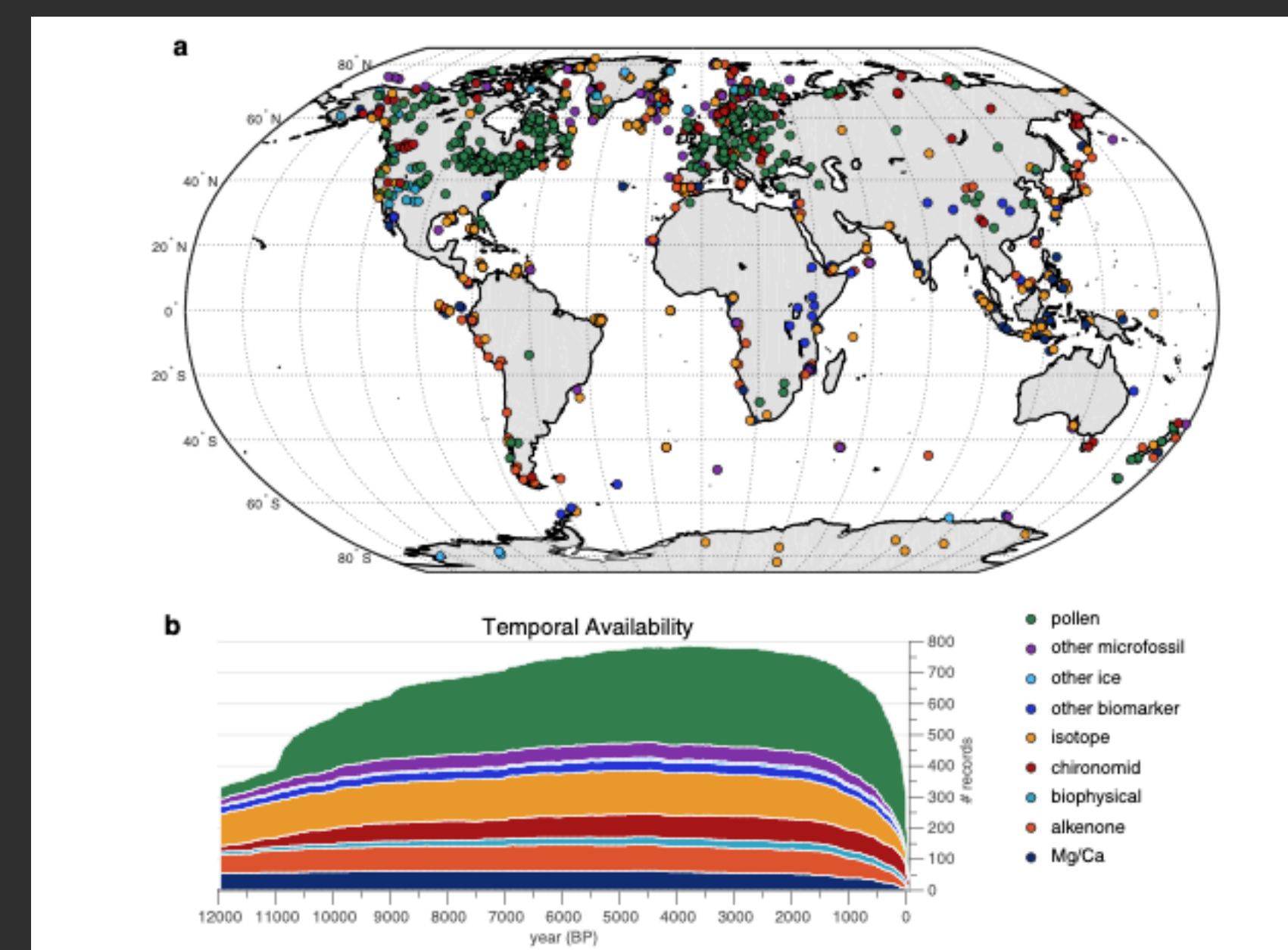
Figure 7. Global composites for various binning intervals and screening criteria, as indicated in subplot titles. The composites are scaled to temperature for comparison, and the shading denotes 95% bootstrap confidence intervals with 500 replicates, to constrain uncertainties. The cutoff between high-resolution (HR) and low-resolution (LR) records is defined as a median resolution of 5 years. Screening options comprise: no screening (*none*), regional temperature screening (*regional*), or regional screening adjusted for the false discovery rate (*regionalFDR*). For low-resolution records, *basicFilter* denotes records that comprise at least 20 values over the Common Era (Supplementary Fig. 2), while *hrNeighbors* denotes records with at least one significantly correlated HR neighbor.

A global database of Holocene paleotemperature records

Darrell Kaufman [✉](#), Nicholas McKay, Cody Routson, Michael Erb, Basil Davis, Oliver Heiri, Samuel Jaccard, Jessica Tierney, Christoph Dätwyler, Yarrow Axford, Thomas Brussel, Olivier Cartapanis, Brian Chase, Andria Dawson, Anne de Vernal, Stefan Engels, Lukas Jonkers, Jeremiah Marsicek, Paola Moffa-Sánchez, Carrie Morrill, Anais Orsi, Kira Rehfeld, Krystyna Saunders, Philipp S. Sommer, Elizabeth Thomas, Marcela Tonello, Mónika Tóth, Richard Vachula, Andrei Andreev, Sebastien Bertrand, Boris Biskaborn, Manuel Bringué, Stephen Brooks, Magaly Caniupán, Manuel Chevalier, Les Cwynar, Julien Emile-Geay, John Fegyveresi, Angelica Feurdean, Walter Finsinger, Marie-Claude Fortin, Louise Foster, Mathew Fox, Konrad Gajewski, Martin Grosjean, Sonja Hausmann, Markus Heinrichs, Naomi Holmes, Boris Ilyashuk, Elena Ilyashuk, Steve Juggins, Deborah Khider, Karin Koinig, Peter Langdon, Isabelle Larocque-Tobler, Jianyong Li, André Lotter, Tomi Luoto, Anson Mackay, Eniko Magyari, Steven Malevich, Bryan Mark, Julieta Massaferrero, Vincent Montade, Larisa Nazarova, Elena Novenko, Petr Pařil, Emma Pearson, Matthew Peros, Reinhard Pienitz, Mateusz Płóciennik, David Porinchu, Aaron Potito, Andrew Rees, Scott Reinemann, Stephen Roberts, Nicolas Rolland, Sakari Salonen, Angela Self, Heikki Seppä, Shyhrete Shala, Jeannine-Marie St-Jacques, Barbara Stenni, Liudmila Syrykh, Pol Tarrats, Karen Taylor, Valerie van den Bos, Gaute Velle, Eugene Wahl, Ian Walker, Janet Wilmshurst, Enlou Zhang & Snezhana Zhilich -Show fewer authors

Scientific Data **7**, Article number: 115 (2020) | [Cite this article](#)

9846 Accesses | **3** Citations | **353** Altmetric | [Metrics](#)



Has paleoclimate fulfilled its promises?

Has paleoclimate fulfilled its promises?

- Mostly yes, especially in the past 10-20 years

Has paleoclimate fulfilled its promises?

- Mostly yes, especially in the past 10-20 years
- Advancements in (geo)informatics have made this possible

Geosciences in the future

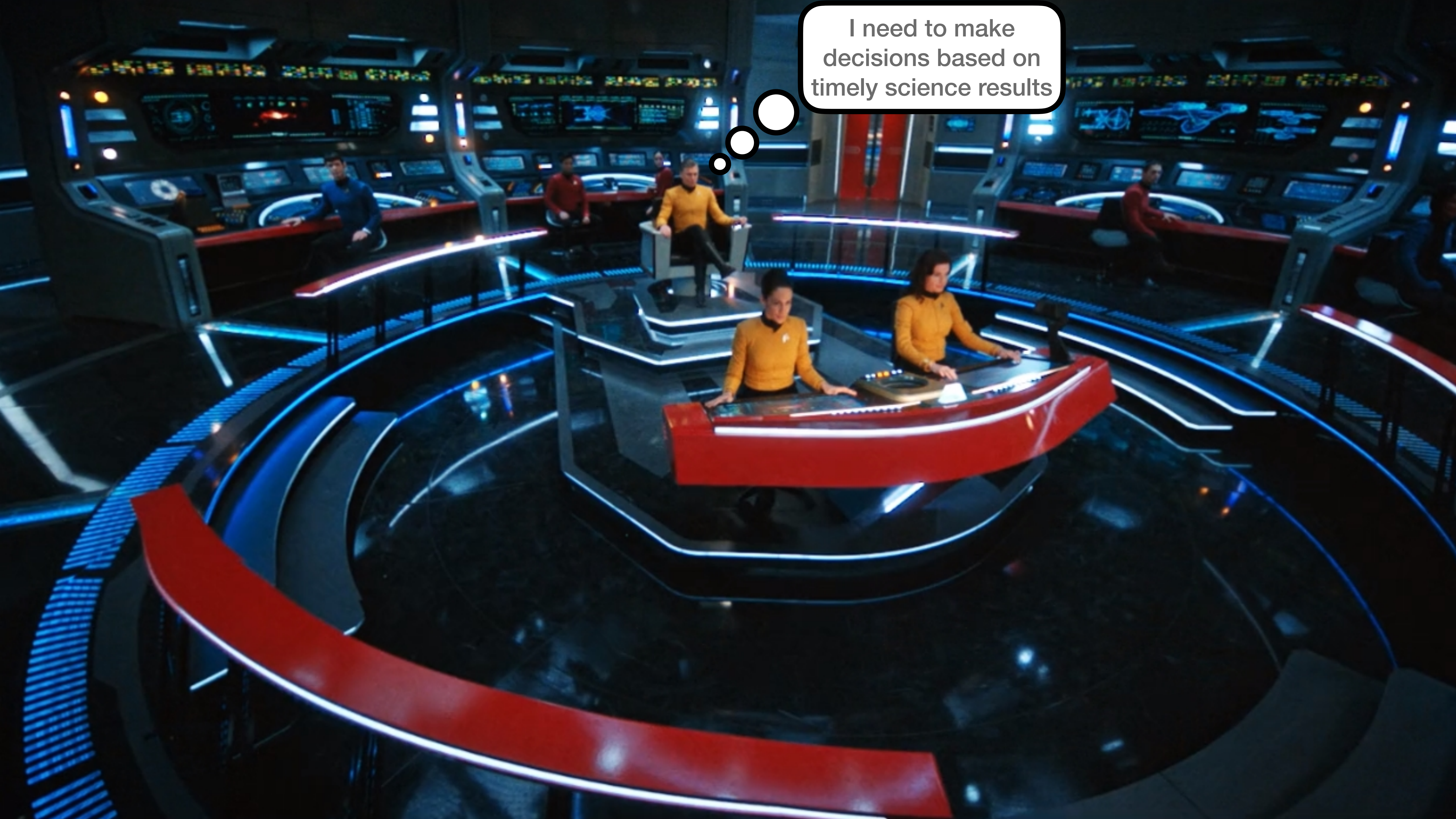


Geosciences in the future






I need to make decisions based on timely science results



A wide-angle shot of a futuristic control room, likely from the Star Trek franchise. The room is filled with various consoles, screens, and control panels. Several crew members are seated at their stations. The lighting is a mix of blue and red, creating a high-tech atmosphere. The room has a circular layout with multiple levels and walkways.

I need to make decisions based on timely science results

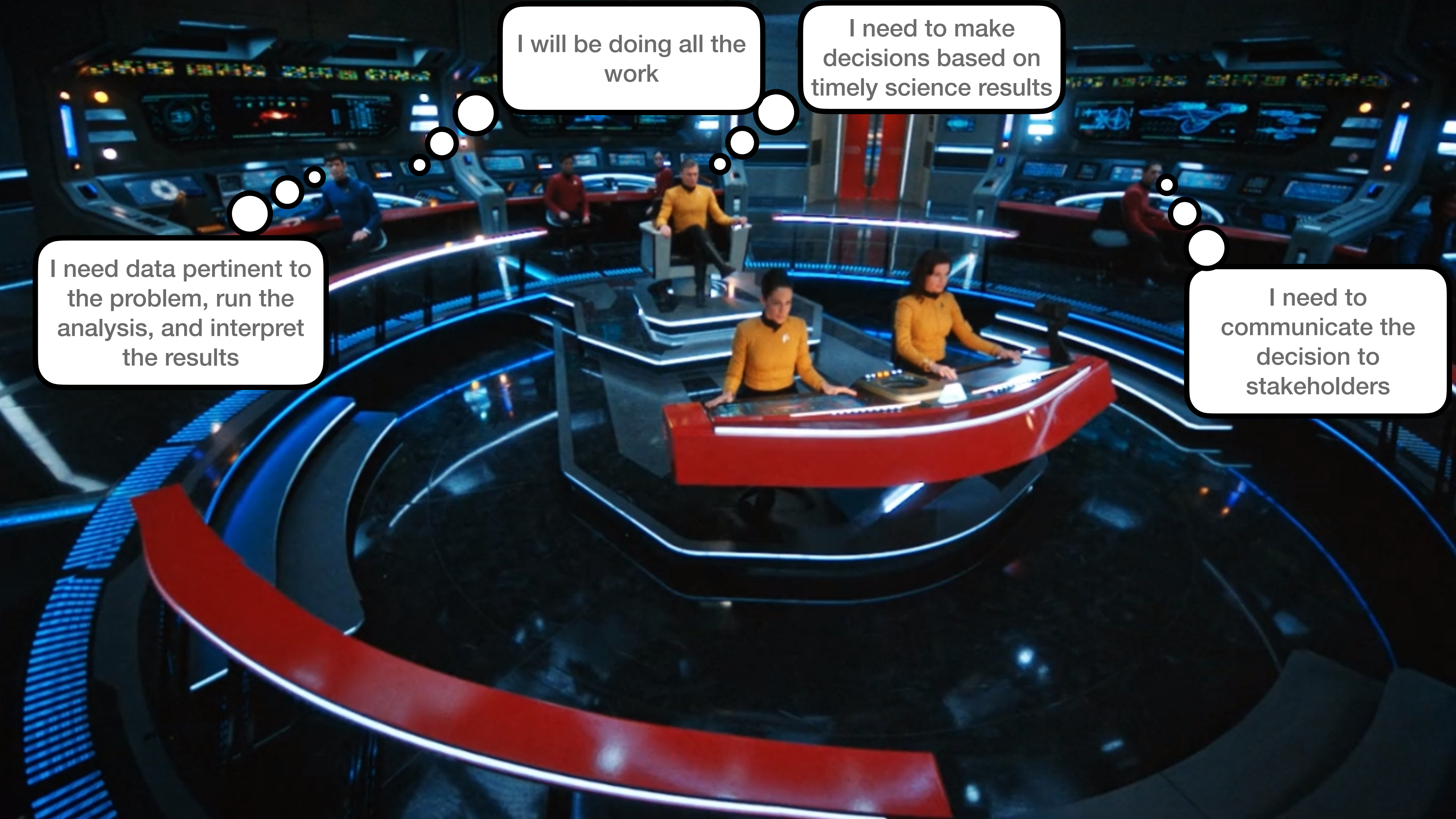
I need data pertinent to the problem, run the analysis, and interpret the results



I need to make decisions based on timely science results

I need data pertinent to the problem, run the analysis, and interpret the results

I need to communicate the decision to stakeholders

A wide-angle shot of the Star Trek: Enterprise bridge. The room is filled with crew members in various uniforms (yellow, red, blue) working at their stations. The bridge has a futuristic design with red and blue lighting. Several thought bubbles are overlaid on the image, each connected to a specific area of the bridge. The bubbles contain text related to data analysis and decision-making. The background shows various control panels and screens displaying data.

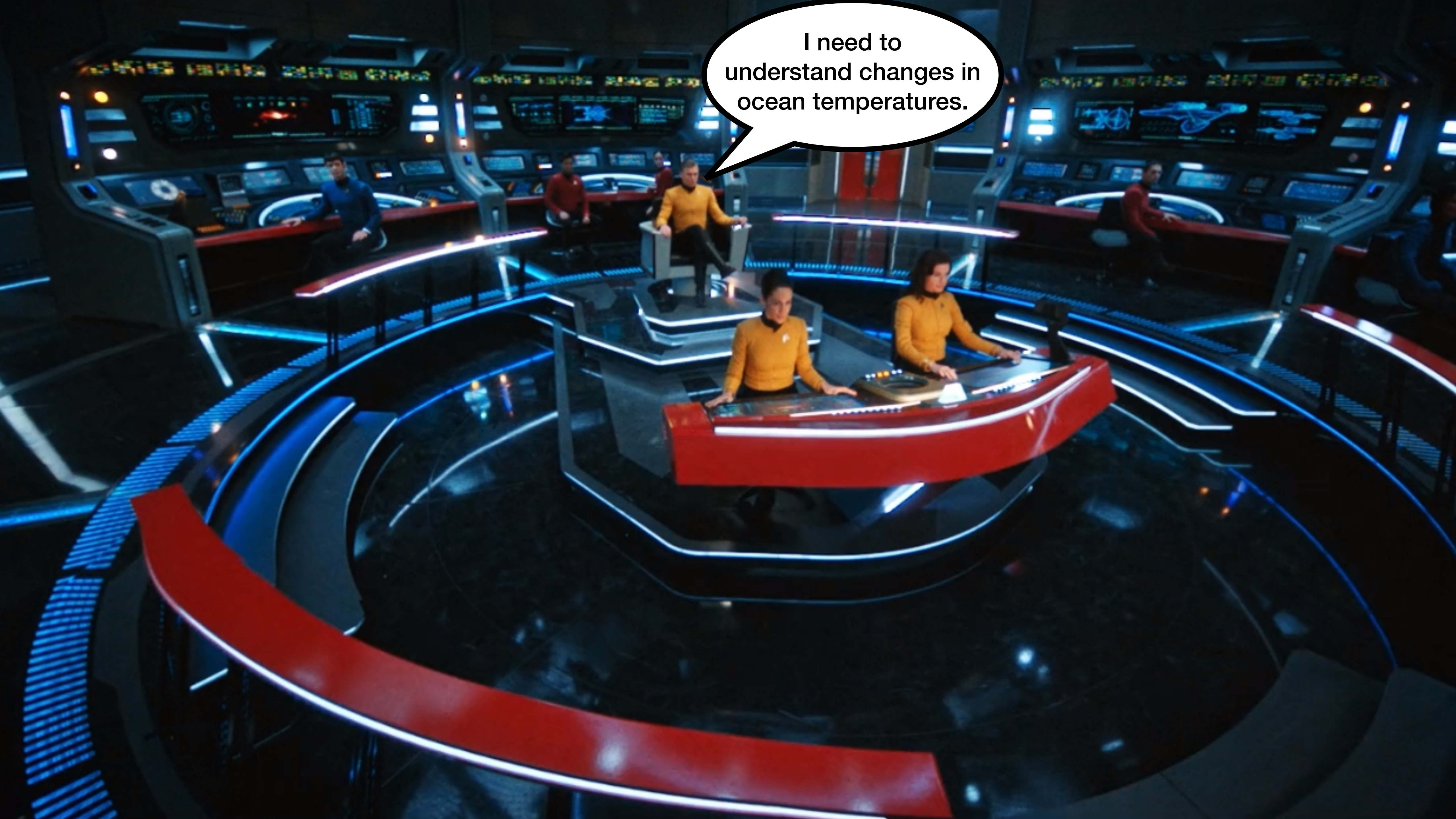
I will be doing all the work

I need to make decisions based on timely science results

I need data pertinent to the problem, run the analysis, and interpret the results

I need to communicate the decision to stakeholders

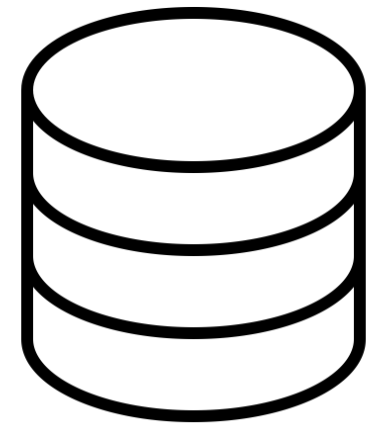
I need to understand changes in ocean temperatures.



A wide-angle shot of the Star Trek: The Next Generation bridge. The room is filled with crew members at their stations, illuminated by blue and red lights. Two speech bubbles are overlaid on the scene. One bubble, located in the upper right, contains the text 'I need to understand changes in ocean temperatures.' The other bubble, located in the lower left, contains the text 'Computer, search for records that represent sea surface temperature over the past 10,000 years and run spectral analysis!'.

I need to understand changes in ocean temperatures.

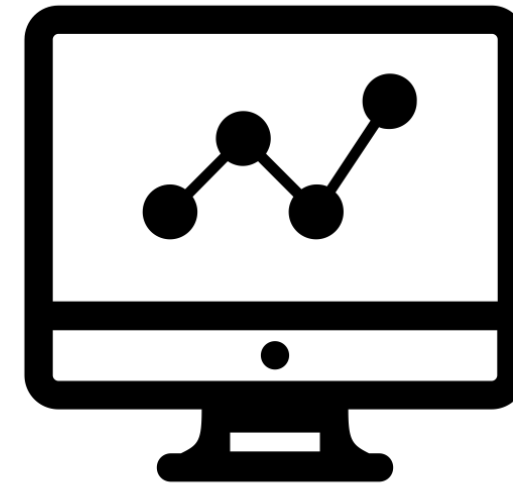
Computer, search for records that represent sea surface temperature over the past 10,000 years and run spectral analysis!



Query Data Catalog



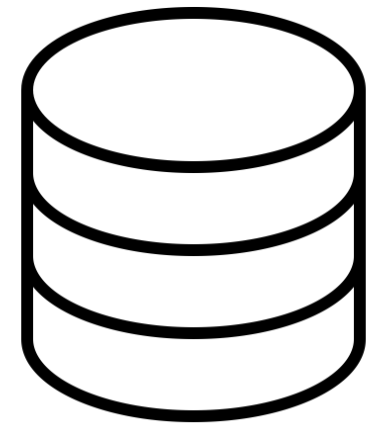
Download datasets



Spectral Analysis



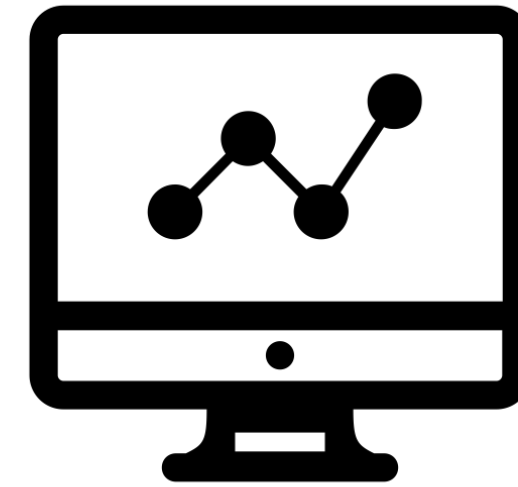
**Visualize and Interpret
Results**



Query Data Catalog



Download datasets



Spectral Analysis



**Visualize and Interpret
Results**

1. Standardization is key to enable automation
2. Create analysis code that makes use of this standard

LinkedEarth: Bringing paleoclimatology into the 21st century



Yolanda Gil
USC



Deborah Khider
USC



Julien Emile-Geay
USC



Nick McKay
NAU



Daniel Garijo
USC



Feng Zhu
USC



Jun Hu
USC

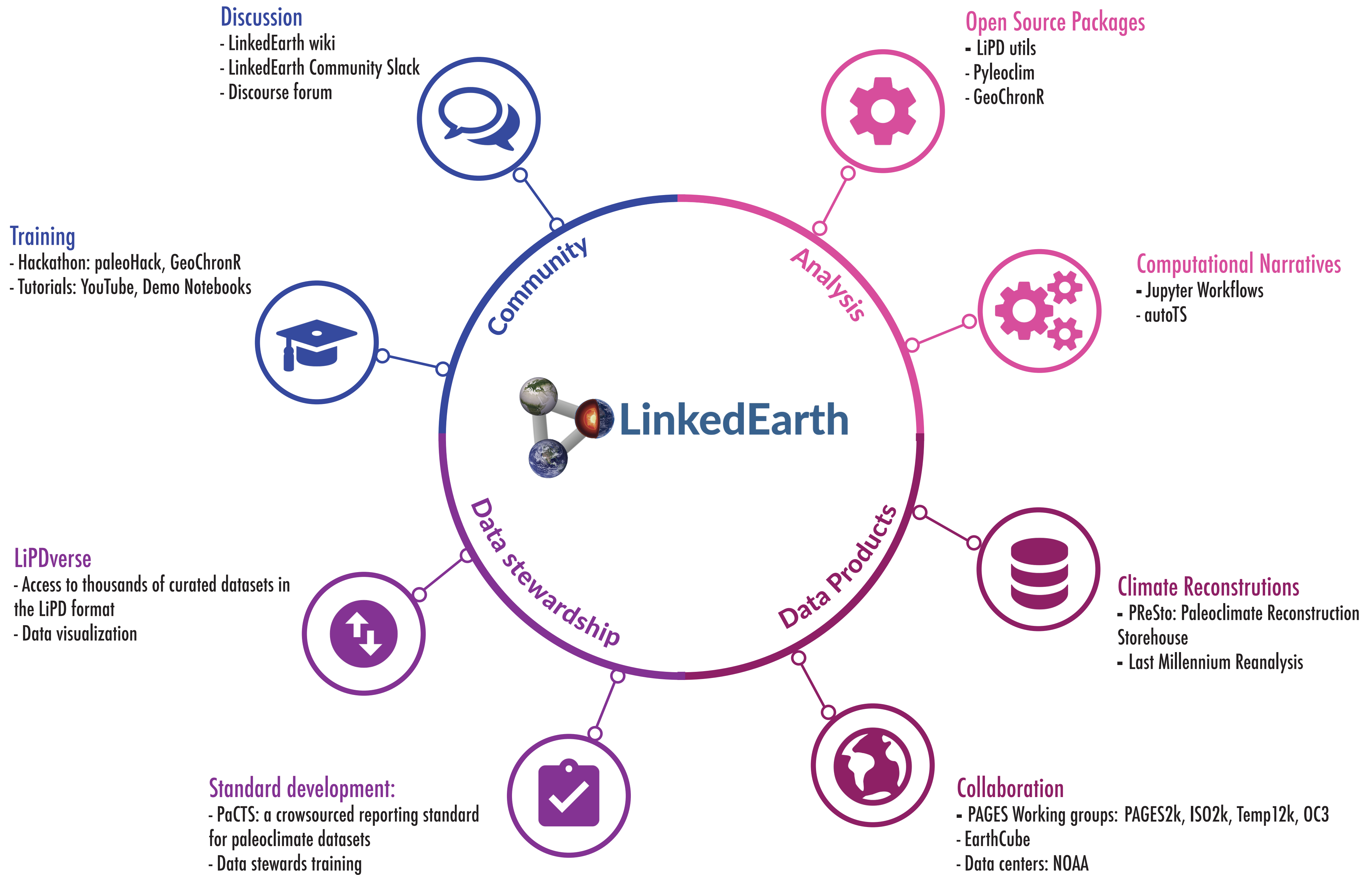


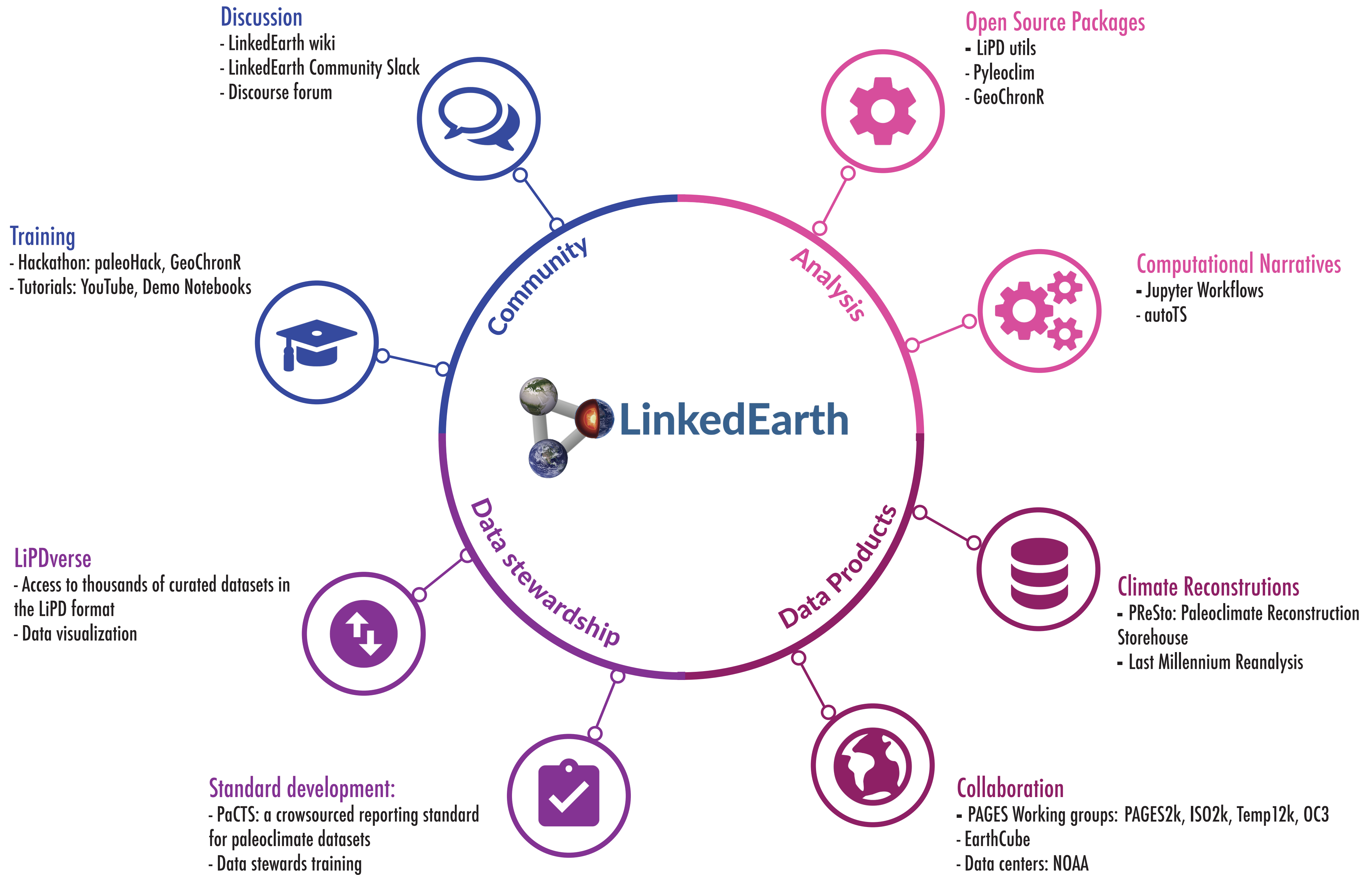
Varun Ratnakar
USC



Chris Heiser
NAU







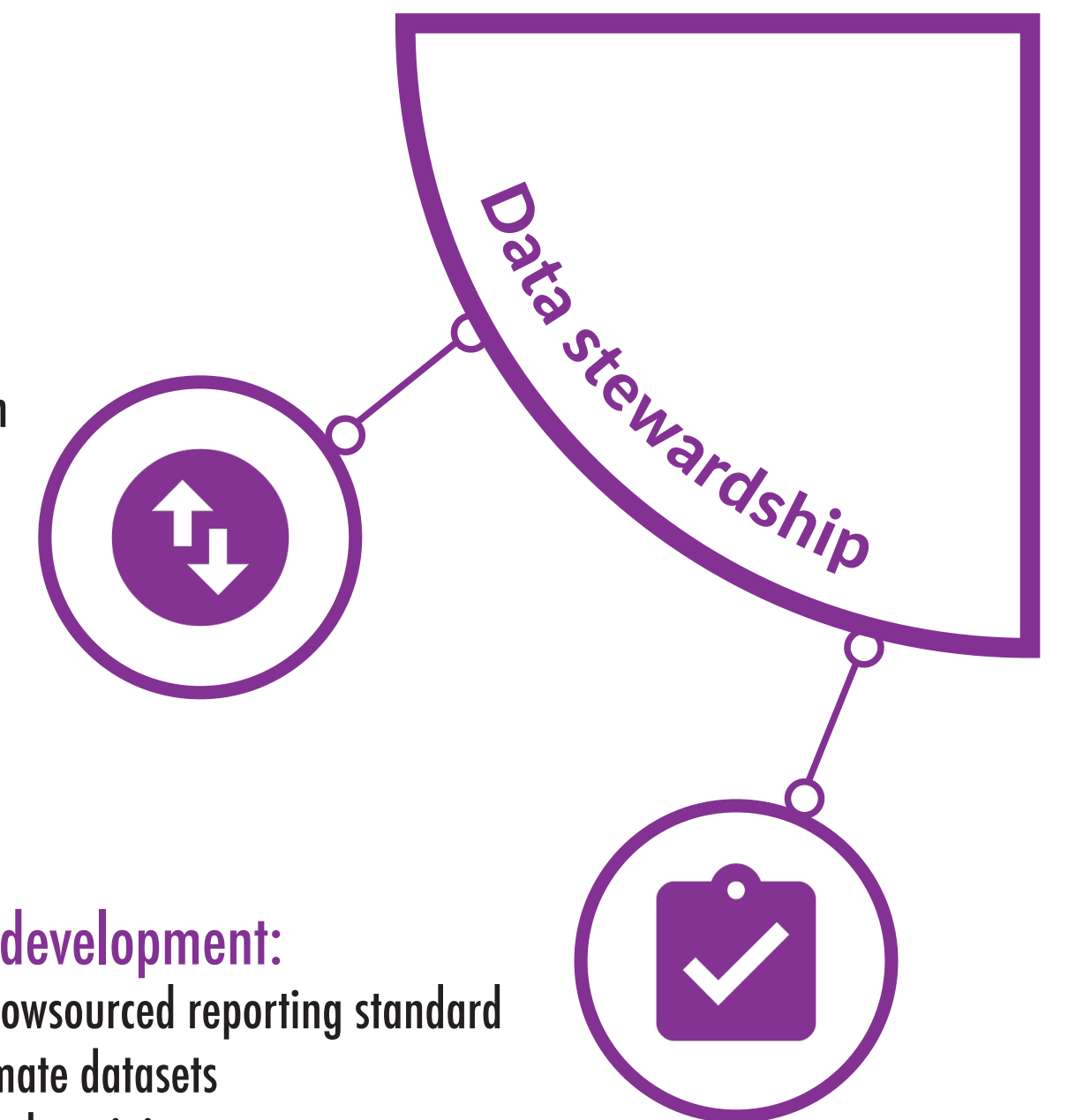
What is a standard?

“A public specification documenting some practice or technology that is adopted and used by a community.”

EarthCube 2015

LiPDverse

- Access to thousands of curated datasets in the LiPD format
- Data visualization



Standard development:

- PaCTS: a crowdsourced reporting standard for paleoclimate datasets
- Data stewards training

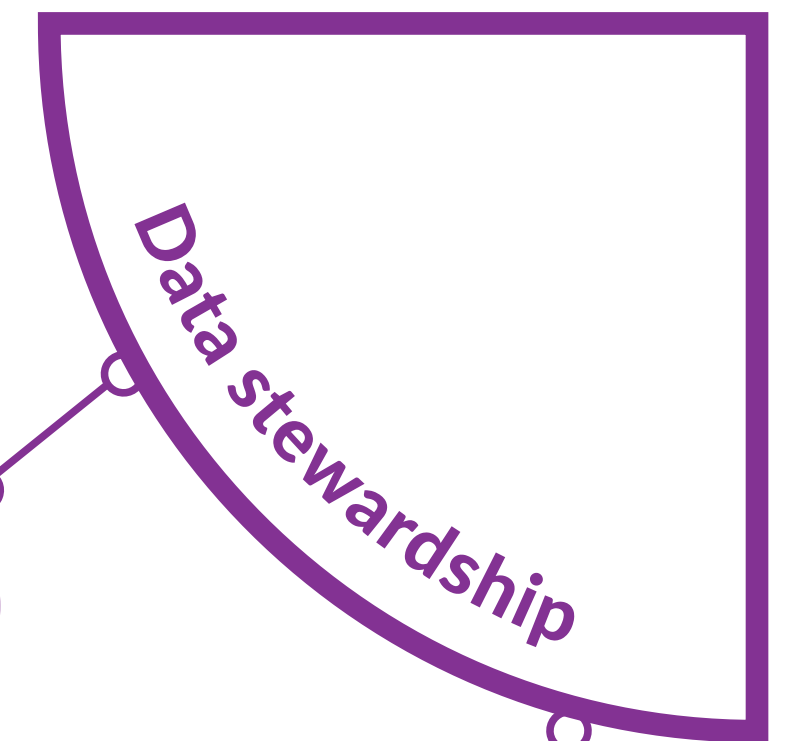
What is a standard?

“A public specification documenting some practice or technology that is adopted and used by a community.”

EarthCube 2015

LiPDverse

- Access to thousands of curated datasets in the LiPD format
- Data visualization



Standard development:

- PaCTS: a crowdsourced reporting standard for paleoclimate datasets
- Data stewards training



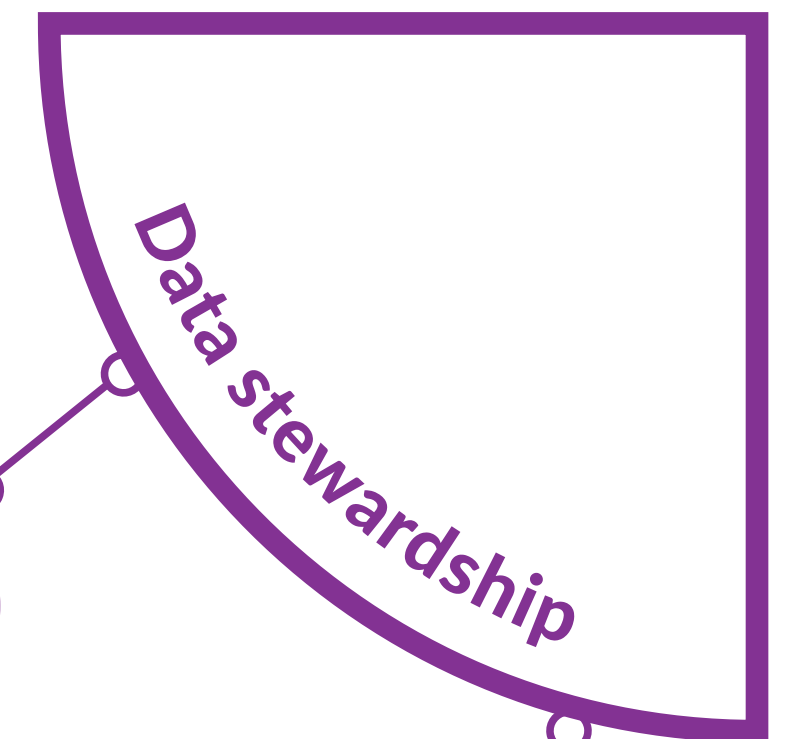
What is a standard?

“A public specification documenting some **practice or technology** that is adopted and used by a **community**.”

EarthCube 2015

LiPDverse

- Access to thousands of curated datasets in the LiPD format
- Data visualization



Standard development:

- PaCTS: a crowdsourced reporting standard for paleoclimate datasets
- Data stewards training



What is a standard?

A standard representation for the data

A standard terminology for the data

Standard guidelines for reporting data

What is a standard?

A standard representation for the data

LiPD/LinkedEarth Ontology

A standard terminology for the data

Standard guidelines for reporting data

What is a standard?

A standard representation for the data

LiPD/LinkedEarth Ontology

A standard terminology for the data

PaST Thesaurus (NOAA)

Standard guidelines for reporting data

What is a standard?

A standard representation for the data

LiPD/LinkedEarth Ontology

A standard terminology for the data

PaST Thesaurus (NOAA)

Standard guidelines for reporting data

PaCTSv1.0

What is a standard?

A standard representation for the data

LiPD/LinkedEarth Ontology

A standard terminology for the data

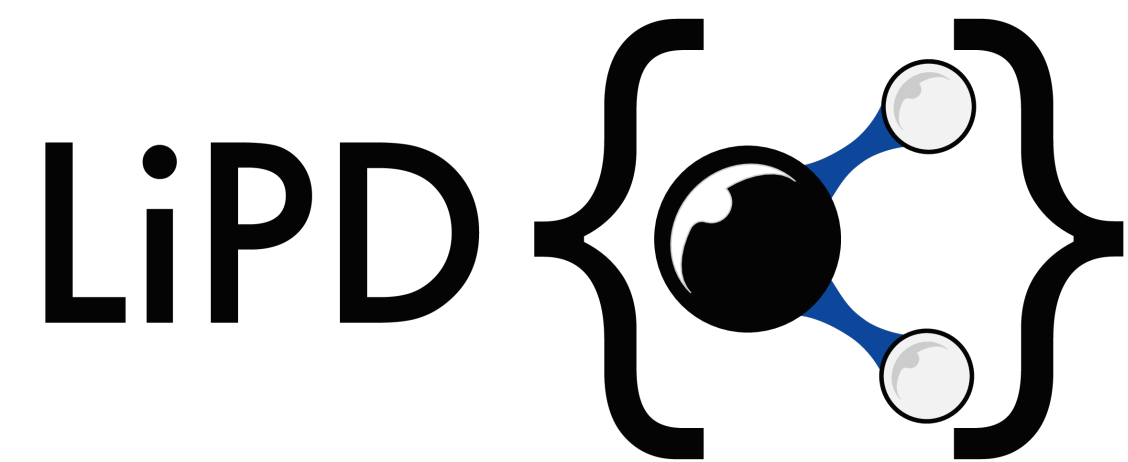
PaST Thesaurus (NOAA)

Standard guidelines for reporting data

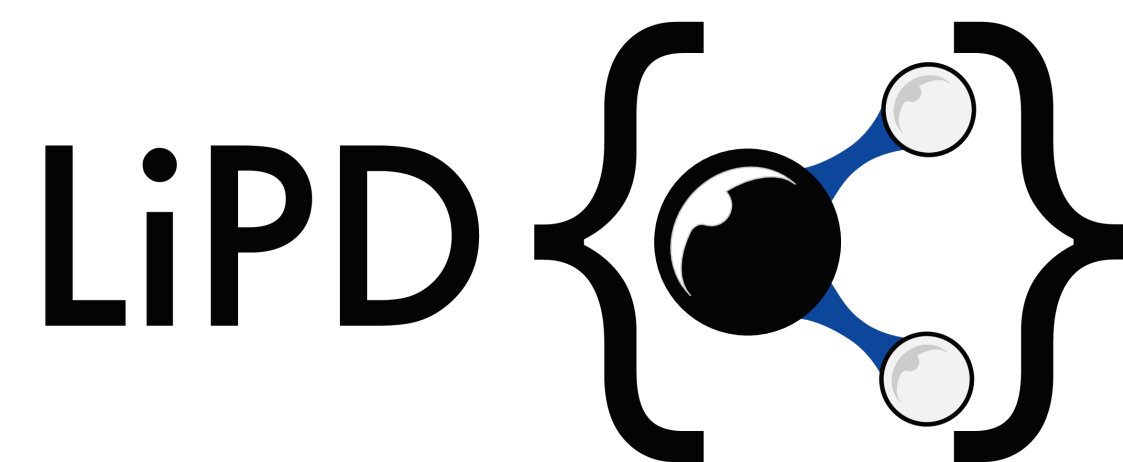
PaCTSv1.0

The Linked Paleo Data Format (LiPD)


The Linked Paleo Data Format (LiPD)



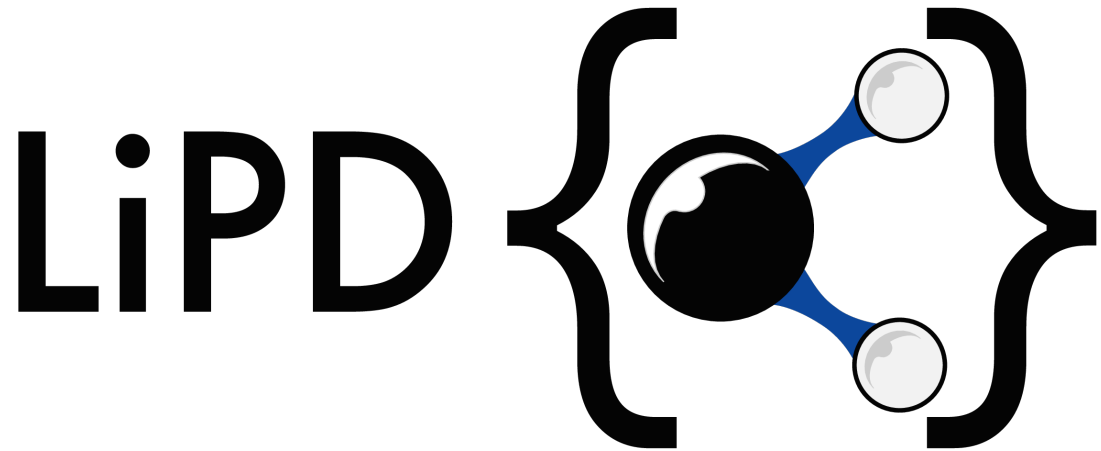
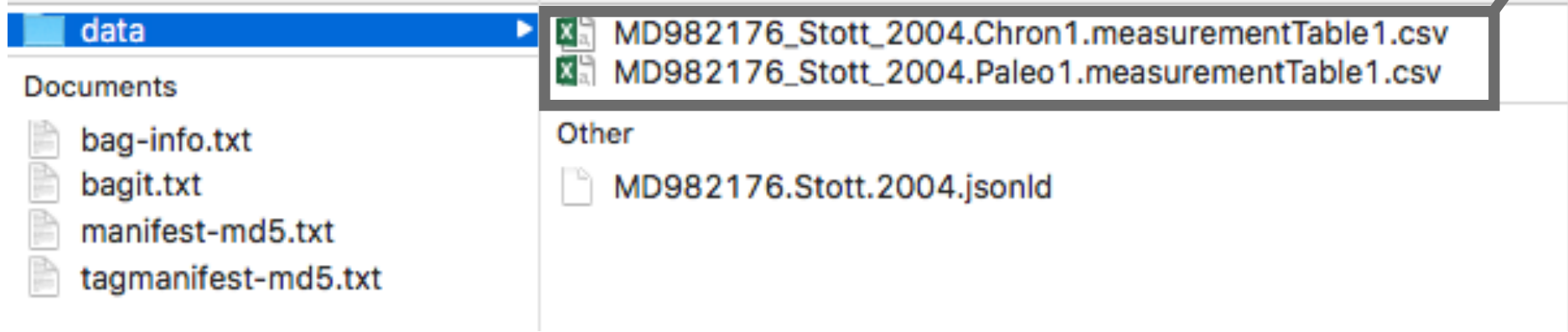
The Linked Paleo Data Format (LiPD)



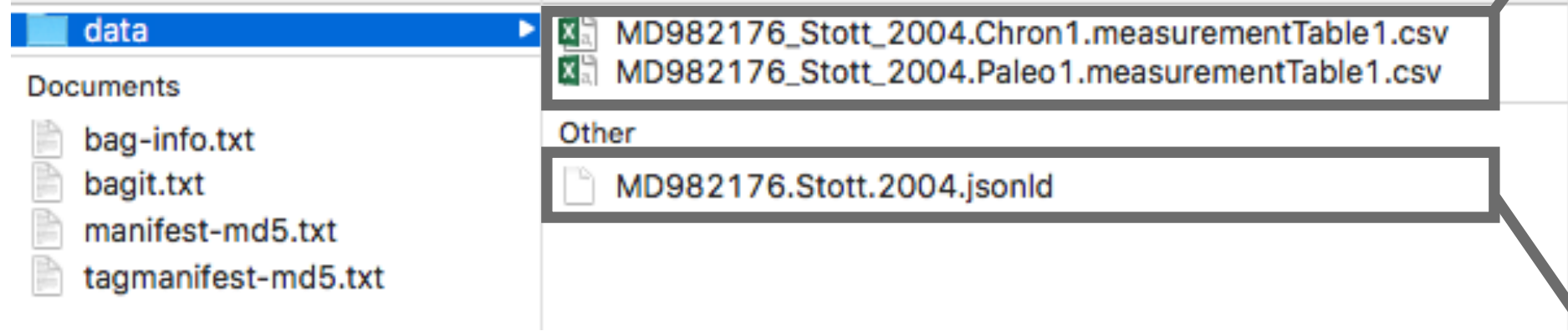

The Linked Paleo Data Format (LiPD)



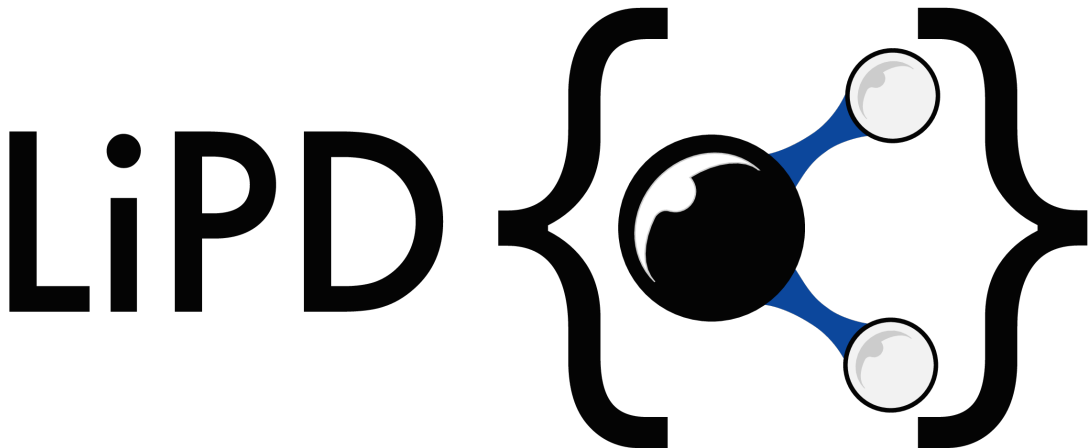

1.0	139.3	-2.561	0.188	4.88	28.4
3.0	161.1	-2.603	0.188	4.97	28.6
5.0	182.8	-2.558	0.358	5.28	29.2
7.0	204.5	-2.44	0.372	5.04	28.7
9.0	226.3	-2.813	-0.001	5.03	28.7
11.0	248.0	-2.668	0.102	4.96	28.5
13.0	269.7	-2.915	-0.27	4.68	27.9
15.0	291.4	-2.676	0.011	4.77	28.1
17.0	313.2	-2.41	0.318	4.87	28.3
19.0	334.9	-2.578	0.088	4.71	28.0
21.0	356.6	-2.821	-0.051	4.93	28.5



The Linked Paleo Data Format (LiPD)

1.0	139.3	-2.561	0.188	4.88	28.4
3.0	161.1	-2.603	0.188	4.97	28.6
5.0	182.8	-2.558	0.358	5.28	29.2
7.0	204.5	-2.44	0.372	5.04	28.7
9.0	226.3	-2.813	-0.001	5.03	28.7
11.0	248.0	-2.668	0.102	4.96	28.5
13.0	269.7	-2.915	-0.27	4.68	27.9
15.0	291.4	-2.676	0.011	4.77	28.1
17.0	313.2	-2.41	0.318	4.87	28.3
19.0	334.9	-2.578	0.088	4.71	28.0
21.0	356.6	-2.821	-0.051	4.93	28.5

```

    "paleoData": [
      {
        "paleoMeasurementTable": [
          {
            "filename": "MD982176_Stott_2004.Paleo1.measurementTable1.csv",
            "columns": [
              {
                "number": "1",
                "hasMaxValue": 887,
                "hasMeanValue": 423.07228915663,
                "hasMedianValue": 419,
                "hasMinValue": 1,
                "resolution": null,
                "units": "cm",
                "TSid": "PYT28ZWS0M8",
                "variableName": "depth",
                "variableType": "measured"
              },
              {
                "number": "2",
                "hasMaxValue": 14831,
                "hasMeanValue": 5836.9088353414,
                "hasMedianValue": 5449.5,
                "hasMinValue": 139.3,
                "resolution": null,
                "units": "yr BP",
                "TSid": "PYTLE0KGBBD",
                "inferredVariableType": "Age",
                "variableName": "yrbp",
                "takenAtDepth": {
                  "foundInTable": "MD982176 Stott 2004.Paleo1.measurementTable1",
                  "hasColumnName": "1",
                  "hasMaxValue": 887,
                  "hasMeanValue": 423.07228915663,
                  "hasMedianValue": 419,
                  "hasMinValue": 1,
                  "hasProxySystem": {
                    "proxyArchiveType": "MarineSediment",
                    "proxySensorType": {
                      "sensorGenus": "Globigerinoides",
                      "sensorSpecies": "ruber"
                    }
                  }
                }
              },
              {
                "hasResolution": {
                  "hasMaxValue": 182,
                  "hasMeanValue": 59.240725806452,
                  "hasMedianValue": 55.8,
                  "hasMinValue": 5.700000000000007
                }
              }
            ]
          }
        ]
      }
    ]
  
```

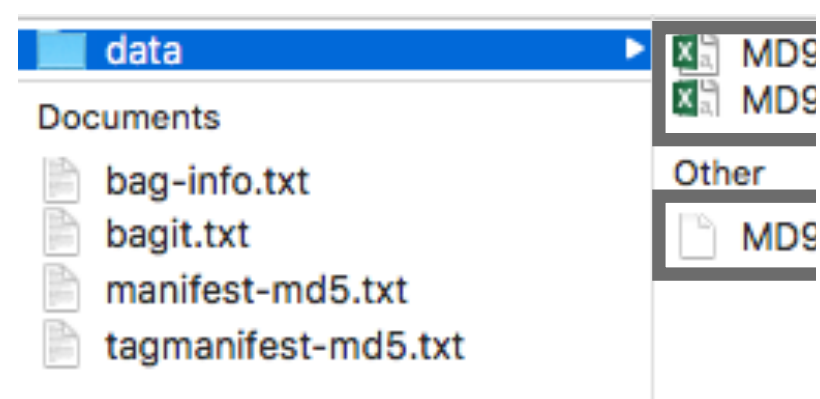


```

"paleoData": [
  {
    "paleoMeasurementTable": [
      {
        "filename": "MD982176_Stott_2004.Paleo1.measurementTable1.csv",
        "columns": [
          {
            "number": "1",
            "hasMaxValue": 887,
            "hasMeanValue": 423.07228915663,
            "hasMedianValue": 419,
            "hasMinValue": 1,
            "resolution": null,
            "units": "cm",
            "TSid": "PYT28ZWSOMB",
            "variableName": "depth",
            "variableType": "measured"
          },
          {
            "number": "2",
            "hasMaxValue": 14831,
            "hasMeanValue": 5836.9088353414,
            "hasMedianValue": 5449.5,
            "hasMinValue": 139.3,
            "resolution": null,
            "units": "yr BP",
            "TSid": "PYTLE0KGBBD",
            "inferredVariableType": "Age",
            "variableName": "yrbp",
            "takenAtDepth": {
              "foundInTable": "MD982176 Stott 2004.Paleo1.measurementTable1",
              "hasColumnName": "1",
              "hasMaxValue": 887,
              "hasMeanValue": 423.07228915663,
              "hasMedianValue": 419,
              "hasMinValue": 1,
              "hasProxySystem": {
                "proxyArchiveType": "MarineSediment",
                "proxySensorType": {
                  "sensorGenus": "Globigerinoides",
                  "sensorSpecies": "ruber"
                }
              }
            },
            "hasResolution": {
              "hasMaxValue": 182,
              "hasMeanValue": 59.240725806452,
              "hasMedianValue": 55.8,
              "hasMinValue": 5,
              "resolution": 10,
              "units": "yr BP"
            }
          }
        ]
      }
    ]
  }
]

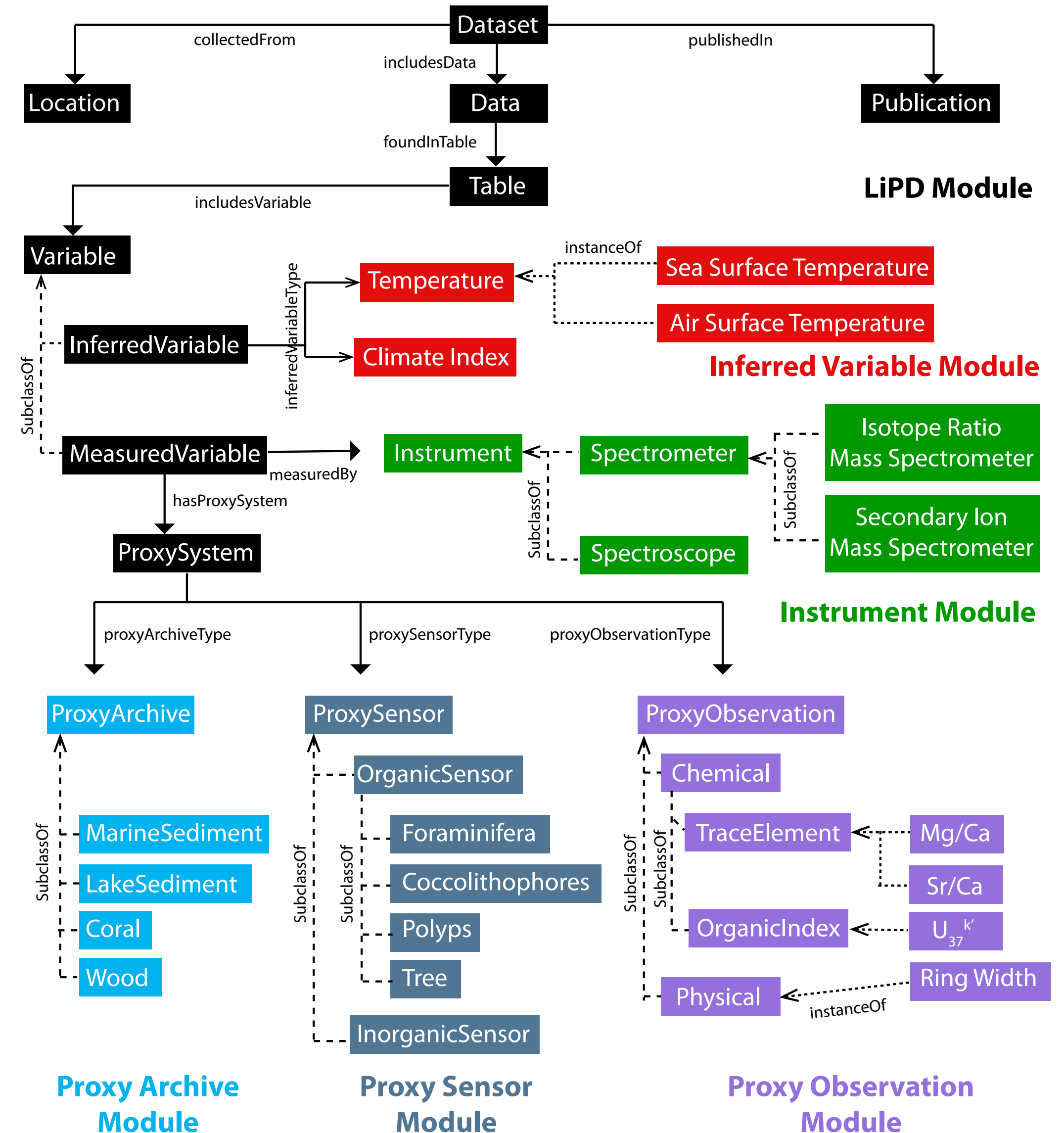
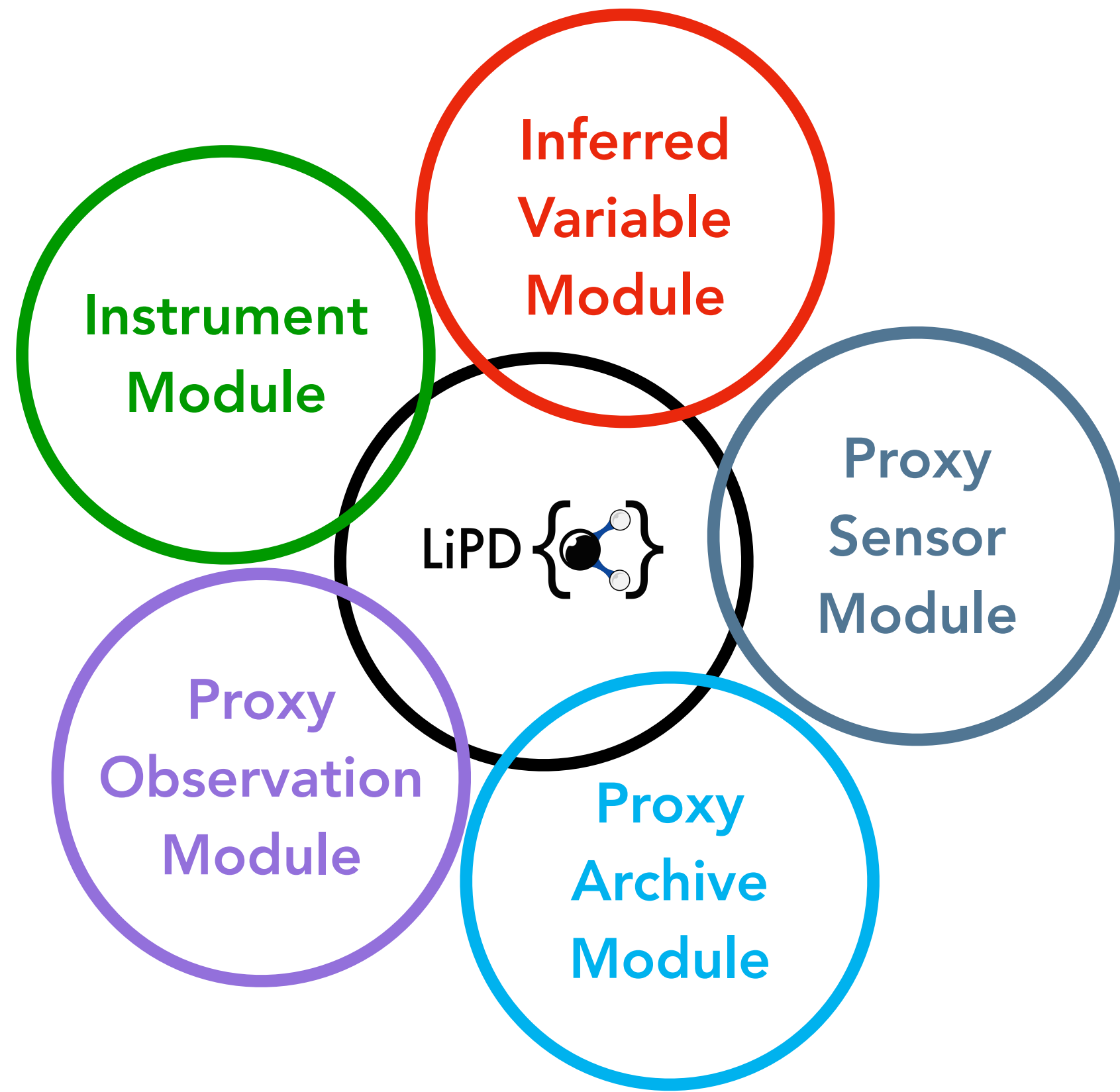
```

18	28.4
17	28.6
18	29.2
14	28.7
13	28.7
16	28.5
18	27.9
17	28.1
17	28.3
11	28.0
13	28.5



LiPD

The LinkedEarth Ontology



Emile-Geay et al., in revisions, Earth Science Informatics
 Gil et al., 2017, ISWC
<http://linked.earth/ontology>

The LinkedEarth Platform

Map-Based Visualizations

MD982181.Khider.2014
ProxyObservation: Mg/Ca

Supported by Queries

```

{{#ask:
[[Category:Dataset_(L)]]
[[IncludesPaleoData_(L).FoundInMeasurementTable_(L).IncludesVariable_(L).HasProxySystem_(L).ProxyArchiveType_(L)::MarineSediment]] OR [[Category:Location_(L)]]
[[CoordinatesFor.ArchiveType::marine sediment]]
|?CollectedFrom_(L).Coordinates
|?Name_(L)
|?IncludesPaleoData_(L).FoundInMeasurementTable_(L).IncludesVariable_(L).ProxyObservationType_(L)
|showtitle=off
|maxzoom=14
|minzoom=1
|limit=500
|template=ByArchiveLocation
|format=leaflet
}}
    
```

Dataset Annotation Interface

PYT6K1XJRVm.mg/ca-g.rub-w
(MeasuredVariable (L))

HasProxySystem (L)	ProxySystem.MarineSediment.Globigerinoides ruber.Mg/Ca	(By Khider)
MeasuredBy (L)	USC.Stott.ICPAES	(By Khider)
MeasuredOn (L)	MD98-2181	(By Khider)
ProxyObservationType (L)	Mg/Ca	(By Khider)
HasUnits (L)	mmol/mol	(By Khider)
Method (L)	Not defined!	
Standard (L)	Standard solution made from solid Mg and reagent grade CaCO3 in an elemental ratio of 5.63 mmol/mol	(By Khider)
FoundInTable (L)	MD982181.Khider.2014.paleo1measurement1	(By Khider)
HasResolution (L)	PYT6K1XJRVm.mg/ca-g.rub-w.Resolution	(By Khider)
HasUncertainty (L)	PYT6K1XJRVm.mg/ca-g.rub-w.Uncertainty1	(By Khider)
InterpretedAs (L)	PYT6K1XJRVm.mg/ca-g.rub-w.Interpretation1	(By Khider)
TakenAtDepth (L)	PYTMU9RVYQL.depth cm	(By Khider)
HasColumnNumber (L)	2	(By Khider)
HasMaxValue (L)	5.86	(By Khider)
HasMeanValue (L)	5.191	(By Khider)
HasMedianValue (L)	5.18	(By Khider)
HasMinValue (L)	4.59	(By Khider)
HasMissingValue (L)	NaN	(By Khider)
HasValue (L)	Not defined!	
HasVariableID (L)	PYT6K1XJRVm	(By Khider)
Notes (L)	Average duplicates from paleo1measurementTable1	(By Khider)

Extra Information

Name (L) mg/ca-g.rub-w (By Khider)

Incoming Properties

- MD982181.Khider.2014.paleo1measurement1 > IncludesVariable (L) > PYT6K1XJRVm.mg/ca-g.rub-w
- PYTES973TGM.sst > InferredFrom (L) > PYT6K1XJRVm.mg/ca-g.rub-w

Author Credit

D. Khider
(Person (L))

Contributions:
Created 299 pages Imported 42 LIPD Edited 578 pages Edited 1 datasets Edited 41 terms Created 29 terms

Working Groups:
Marine Sediment Working Group Chronologies Working Group Uncertainty Working Group

Wiki Privileges:
bureaucrat editorial-board sysop

Email (L)	Not defined!
OrcidNumber (L)	0000-0001-7501-8430
Name (L)	D. Khider
Has User ID	Khider
Expertise	Paleoclimatology Geochemistry
Highest Degree	PhD
University	University of Southern California
Current Affiliation	University of Southern California
Current Position	Postdoctoral Scholar
Twitter ID	@dkhider

Community Discussions

Subcategories

This category has the following 10 subcategories, out of 10 total.

- Chronologies Working Group
- Ice Cores Working Group
- Lake Sediments Working Group

Pages in category "Working Group"

The following 4 pages are in this category, out of 4 total.

What should be the primary way of reporting depth of samples taken from lake sediments:

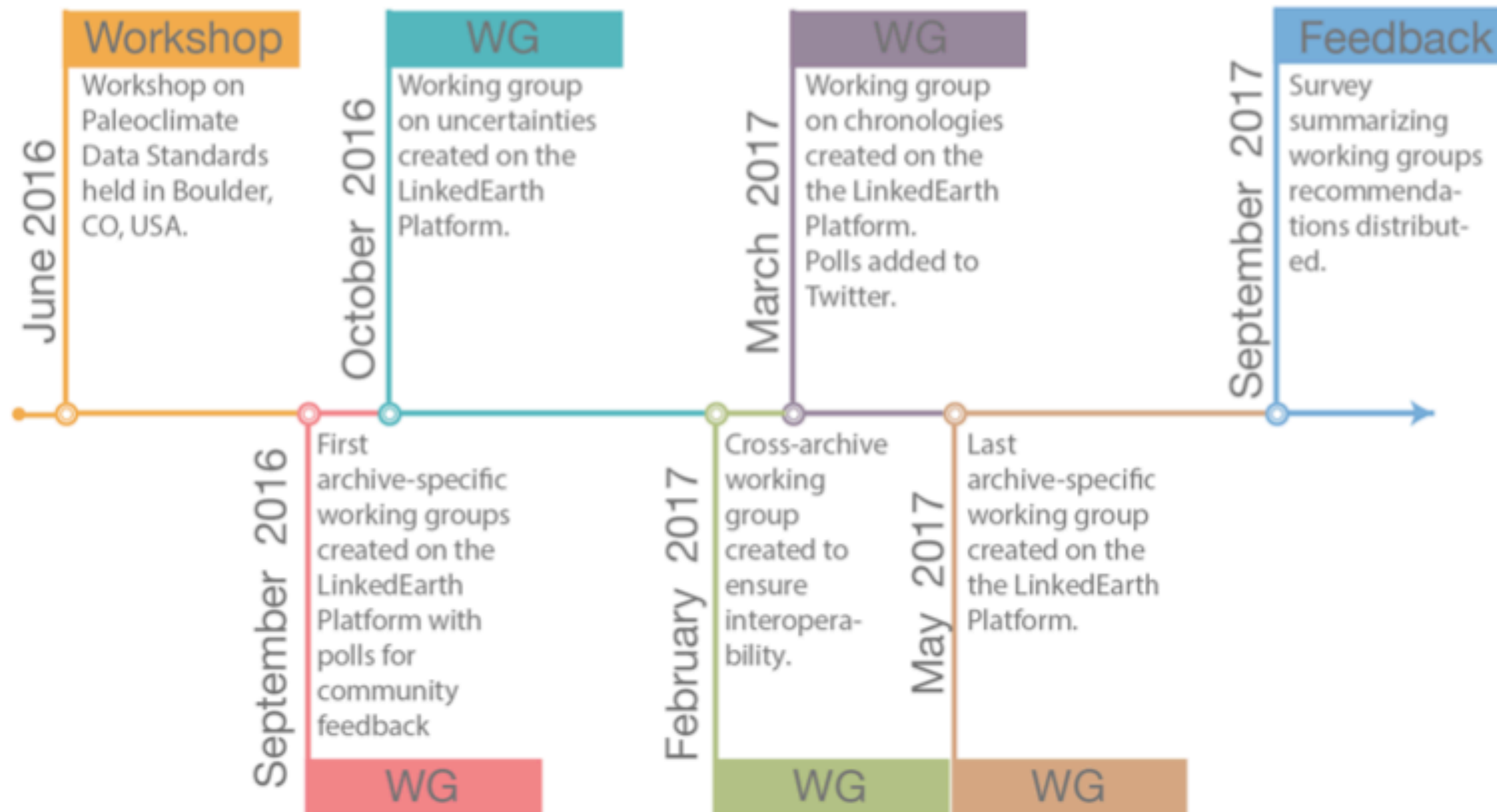
You voted for "Mid-point depth and thickness" on 14 March 2017 at 10:31. You can change your vote by clicking a different answer below.

- Mid-point depth and thickness (5 votes)
- Top and bottom depth (3 votes)
- Something else (1 vote)
- I want to revoke my vote

Poll for Decision Making

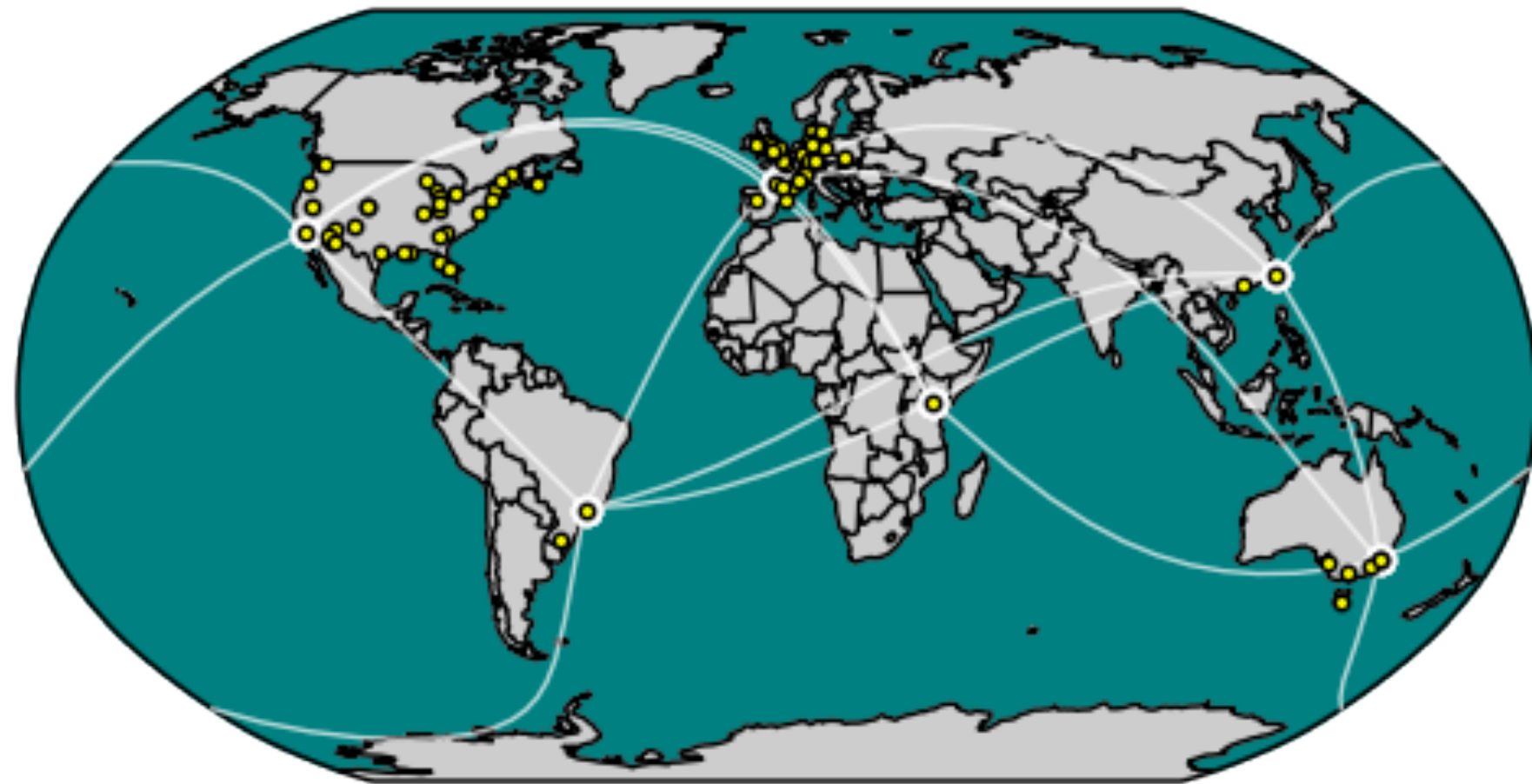
There were 9 votes since the poll was created on 12:18, 21 September 2016.

PaCTS: A crowdsourced reporting standard for paleoclimate data



PaCTS: A crowdsourced reporting standard for paleoclimate data

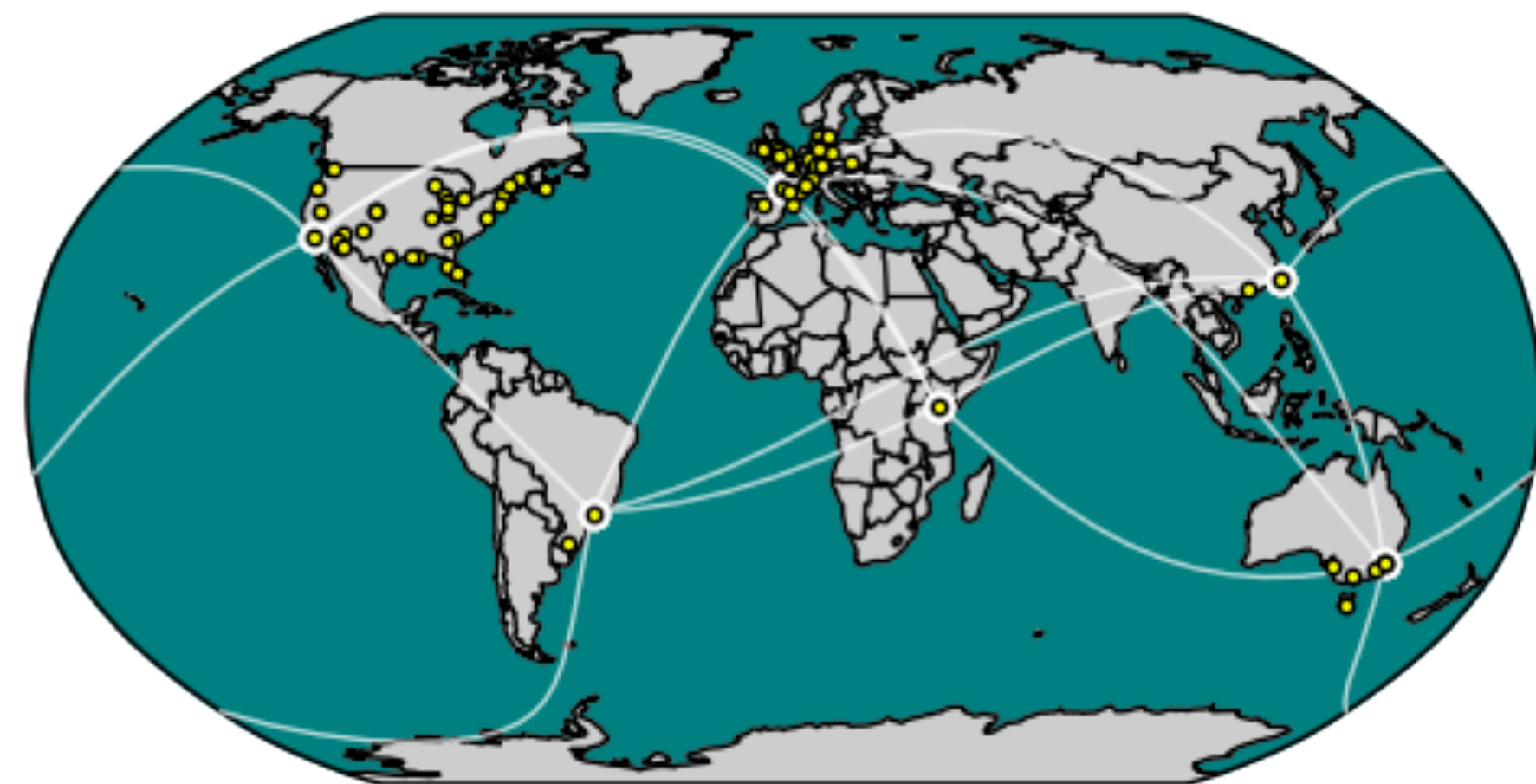
135 researchers...



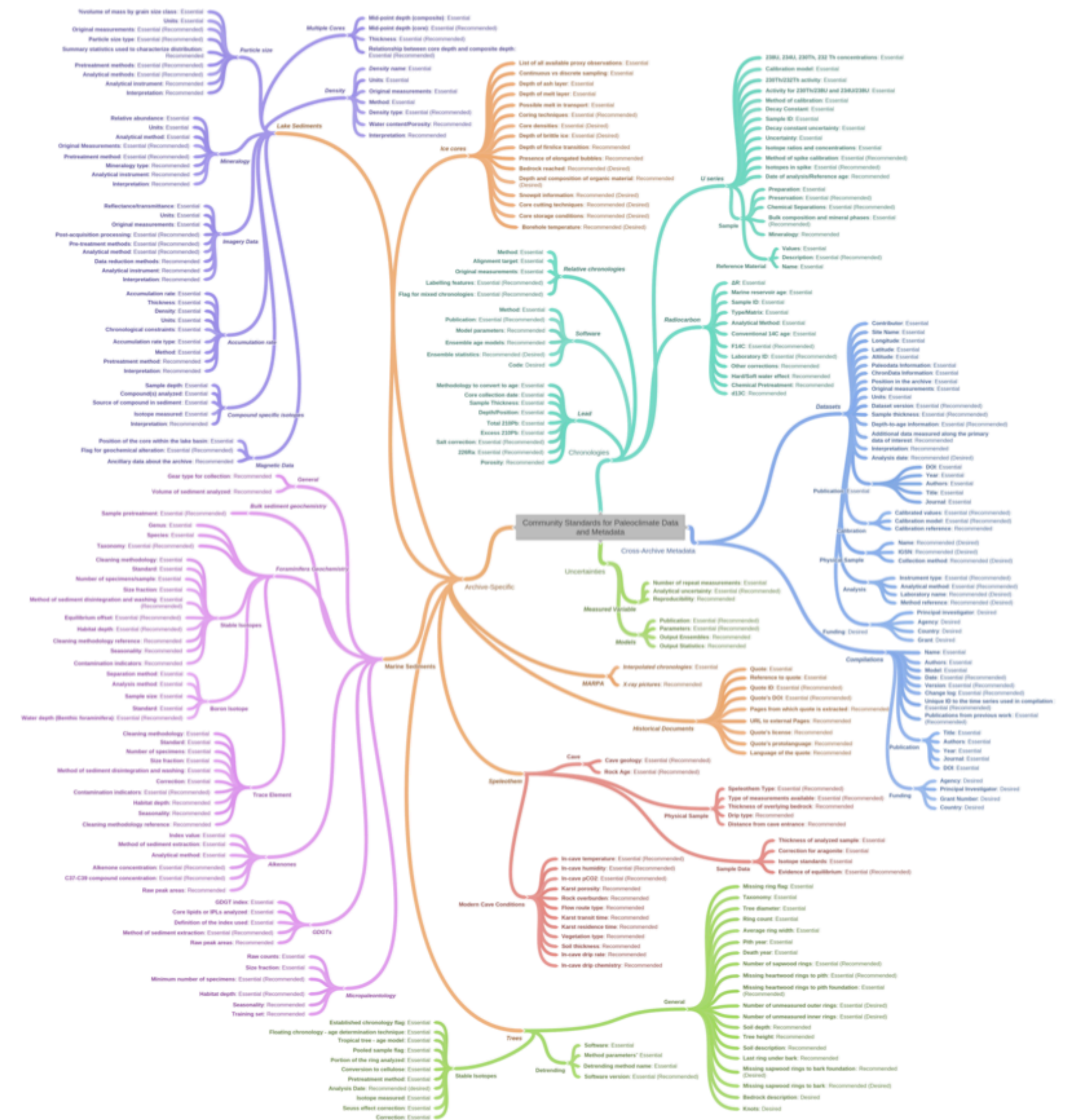
... voted on 603 properties

PaCTS: A crowdsourced reporting standard for paleoclimate data

135 researchers...

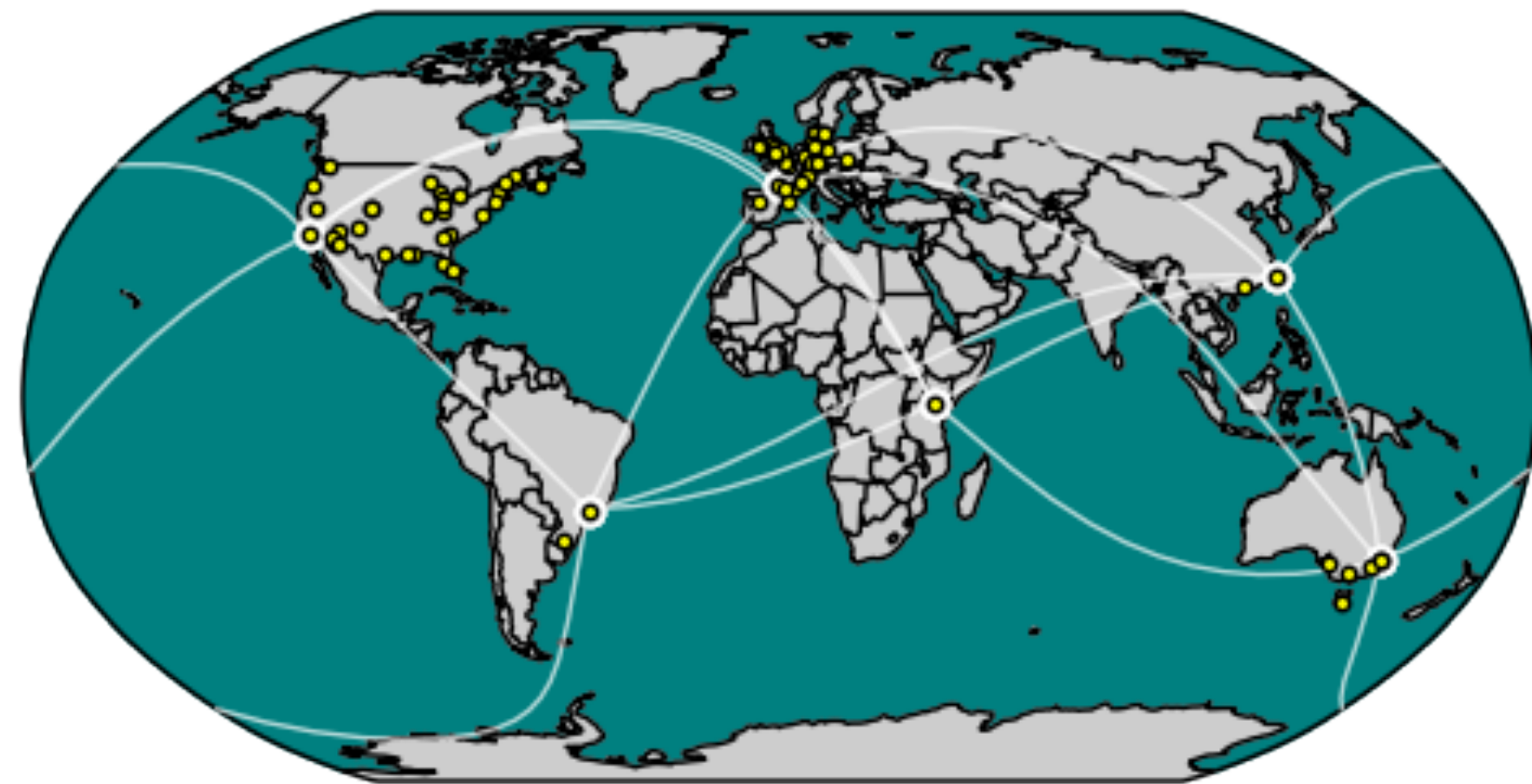


... voted on 603 properties

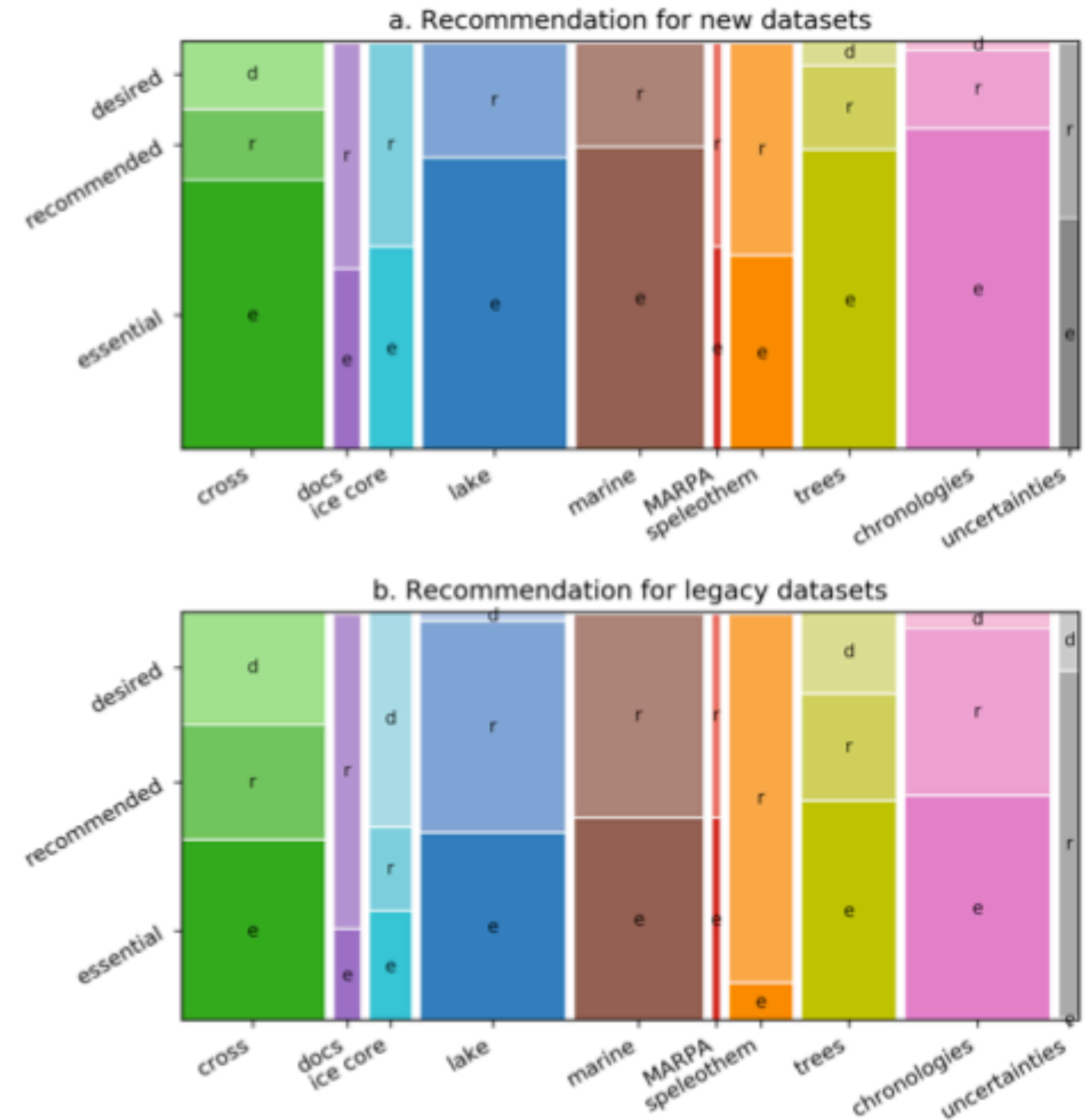


PaCTS: A crowdsourced reporting standard for paleoclimate data

135 researchers...



... voted on 603 properties



Implementing PaCTS

1. Lower the barrier to metadata archiving

2. Change the incentive structure

 Funding agencies

 Editors and publishers

 Professional societies

 Science-enabling code

Implementing PaCTS

1. Lower the barrier to metadata archiving


2. Change the incentive structure

 Funding agencies

 Editors and publishers


 Professional societies

 Science-enabling code



Paleoceanography and Paleoclimatology

FEATURE ARTICLE
10.1029/2019PA003632



PaCTS 1.0: A Crowdsourced Reporting Standard for Paleoclimate Data

D. Khider^{1,2}, J. Emile-Geay², N. P. McKay³, Y. Gil¹, D. Garijo¹, V. Ratnakar¹, M. Alonso-Garcia⁴, S. Bertrand⁵, O. Bothe⁶, P. Brewer⁷, A. Bunn⁸, M. Chevalier⁹, L. Comas-Bru^{10,11}, A. Csank¹², E. Dassié¹³, K. DeLong¹⁴, T. Felis¹⁵, P. Francus¹⁶, A. Frappier¹⁷, W. Gray¹⁸, S. Goring¹⁹, L. Jonkers¹⁵, M. Kahle²⁰, D. Kaufman³, N. M. Kehrwald²¹, B. Martrat^{22,23}, H. McGregor²⁴, J. Richey²⁵, A. Schmittner²⁶, N. Scroxton²⁷, E. Sutherland²⁸, K. Thirumalai²⁹, K. Allen³⁰, F. Arnaud³¹, Y. Axford³², T. Barrows²⁴, L. Bazin¹⁸, S. E. Pilaar Birch³³, E. Bradley³⁴, J. Bregy³⁵, E. Capron³⁶, O. Cartapanis³⁷, H.-W. Chiang³⁸, K. M. Cobb³⁹, M. Debret⁴⁰, R. Dommain⁴¹, J. Du²⁶, K. Dyez⁴², S. Emerick⁴³, M. P. Erb³, G. Falster⁴⁴, W. Finsinger⁴⁵, D. Fortier⁴⁶, Nicolas Gauthier⁴⁷, S. George⁴⁸, E. Grimm⁴⁹, J. Hertzberg⁵⁰, F. Hibbert⁵¹, A. Hillman⁵², W. Hobbs⁵³, M. Huber⁵⁴, A. L. C. Hughes^{55,56}, S. Jaccard³⁷, J. Ruan⁵⁷, M. Kienast⁵⁸, B. Konecky⁵⁹, G. Le Roux⁶⁰, V. Lyubchich⁶¹, V. F. Novello⁴³, L. Olaka⁶², J. W. Partin⁶³, C. Pearce⁶⁴, S. J. Phipps⁶⁵, C. Pignol³¹, N. Piotrowska⁶⁶, M.-S. Poli⁶⁷, A. Prokopenko⁶⁸, F. Schwanck⁶⁹, C. Stepanek⁷⁰, G. E. A. Swann⁷¹, R. Telford⁷², E. Thomas⁷³, Z. Thomas⁷⁴, S. Truebe⁷⁵, L. von Gunten⁷⁶, A. Waite⁷⁷, N. Weitzel⁷⁸, B. Wilhelm⁷⁹, J. Williams⁸⁰, M. Winstруп⁸¹, N. Zhao⁸², and Y. Zhou⁸³

Key Points:

- First version of a crowdsourced reporting standard for paleoclimate data
- The standards arose through collective discussions, both in person and online, and via an innovative social platform
- The standard helps meet the interoperability and reuse criteria of FAIR (Findable, Accessible, Interoperable, and Reusable)

Implementing PaCTS

1. Lower the barrier to metadata archiving
2. Change the incentive structure



\$ Funding agencies

📖 Editors and publishers

👜 Professional societies

🌌 Science-enabling code

Paleoceanography and Paleoclimatology

FEATURE ARTICLE
10.1029/2019PA003632

PaCTS 1.0: A Crowdsourced Reporting Standard for Paleoclimate Data

D. Khider^{1,2}, J. Emile-Geay², N. P. McKay³, Y. Gil¹, D. Garijo¹, V. Ratnakar¹, M. Alonso-Garcia⁴, S. Bertrand⁵, O. Bothe⁶, P. Brewer⁷, A. Bunn⁸, M. Chevalier⁹, L. Comas-Bru^{10,11}, A. Csank¹², E. Dassié¹³, K. DeLong¹⁴, T. Felis¹⁵, P. Francus¹⁶, A. Frappier¹⁷, W. Gray¹⁸, S. Goring¹⁹, L. Jonkers¹⁵, M. Kahle²⁰, D. Kaufman³, N. M. Kehrwald²¹, B. Martrat^{22,23}, H. McGregor²⁴, J. Richey²⁵, A. Schmittner²⁶, N. Scroxton²⁷, E. Sutherland²⁸, K. Thirumalai²⁹, K. Allen³⁰, F. Arnaud³¹, Y. Axford³², T. Barrows²⁴, L. Bazin¹⁸, S. E. Pilaar Birch³³, E. Bradley³⁴, J. Bregy³⁵, E. Capron³⁶, O. Cartapanis³⁷, H.-W. Chiang³⁸, K. M. Cobb³⁹, M. Debret⁴⁰, R. Dommain⁴¹, J. Du²⁶, K. Dyez⁴², S. Emerick⁴³, M. P. Erb³, G. Falster⁴⁴, W. Finsinger⁴⁵, D. Fortier⁴⁶, Nicolas Gauthier⁴⁷, S. George⁴⁸, E. Grimm⁴⁹, J. Hertzberg⁵⁰, F. Hibbert⁵¹, A. Hillman⁵², W. Hobbs⁵³, M. Huber⁵⁴, A. L. C. Hughes^{55,56}, S. Jaccard³⁷, J. Ruan⁵⁷, M. Kienast⁵⁸, B. Konecky⁵⁹, G. Le Roux⁶⁰, V. Lyubchich⁶¹, V. F. Novello⁴³, L. Olaka⁶², J. W. Partin⁶³, C. Pearce⁶⁴, S. J. Phipps⁶⁵, C. Pignol³¹, N. Piotrowska⁶⁶, M.-S. Poli⁶⁷, A. Prokopenko⁶⁸, F. Schwanck⁶⁹, C. Stepanek⁷⁰, G. E. A. Swann⁷¹, R. Telford⁷², E. Thomas⁷³, Z. Thomas⁷⁴, S. Truebe⁷⁵, L. von Gunten⁷⁶, A. Waite⁷⁷, N. Weitzel⁷⁸, B. Wilhelm⁷⁹, J. Williams⁸⁰, M. Winstруп⁸¹, N. Zhao⁸², and Y. Zhou⁸³

Key Points:

- First version of a crowdsourced reporting standard for paleoclimate data
- The standards arose through collective discussions, both in person and online, and via an innovative social platform
- The standard helps meet the interoperability and reuse criteria of FAIR (Findable, Accessible, Interoperable, and Reusable)

Implementing PaCTS

1. Lower the barrier to metadata archiving

2. Change the incentive structure

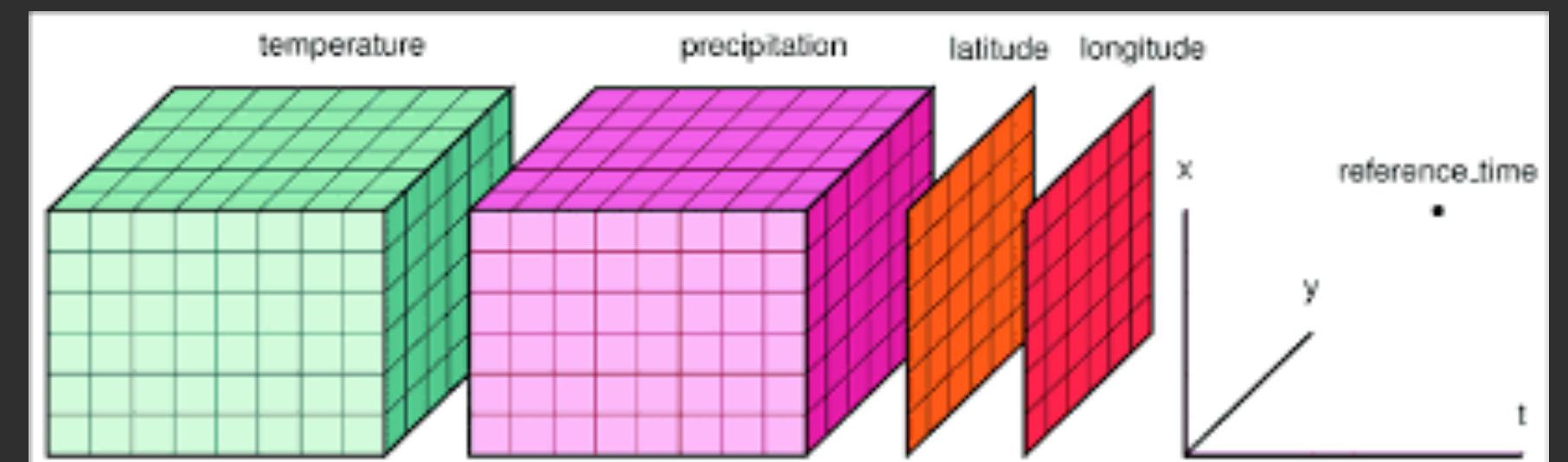
\$ Funding agencies

📖 Editors and publishers

👜 Professional societies

🔬 Science-enabling code

NetCDF-CF



Implementing PaCTS

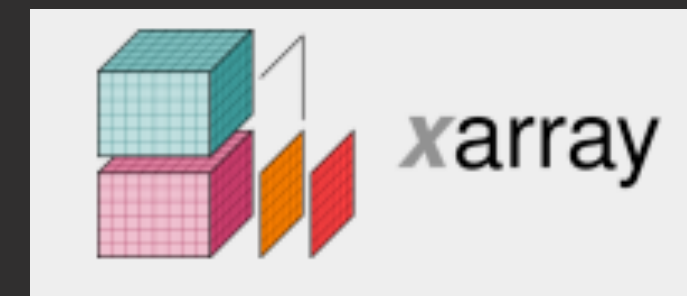
1. Lower the barrier to metadata archiving
2. Change the incentive structure

\$ Funding agencies

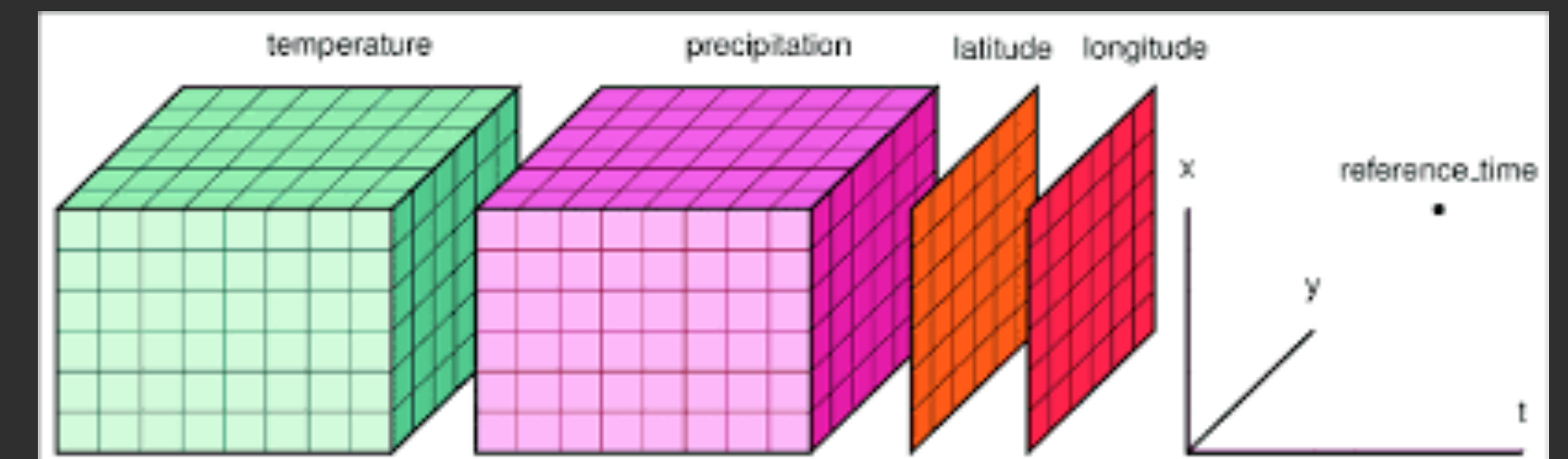
📖 Editors and publishers

👜 Professional societies

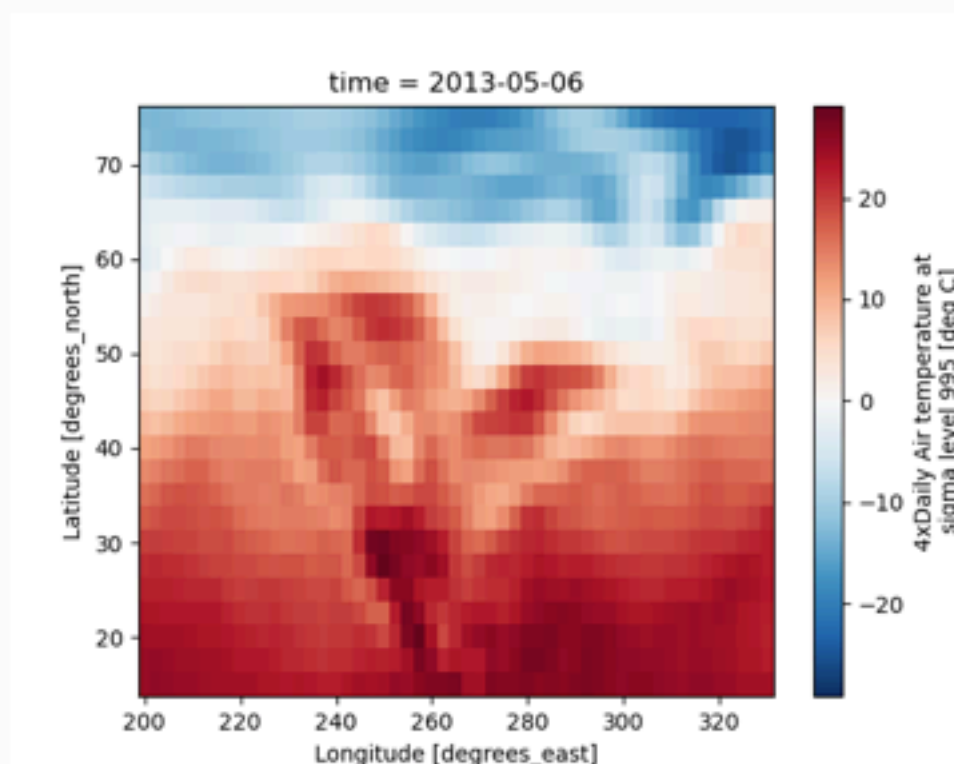
🔬 Science-enabling code

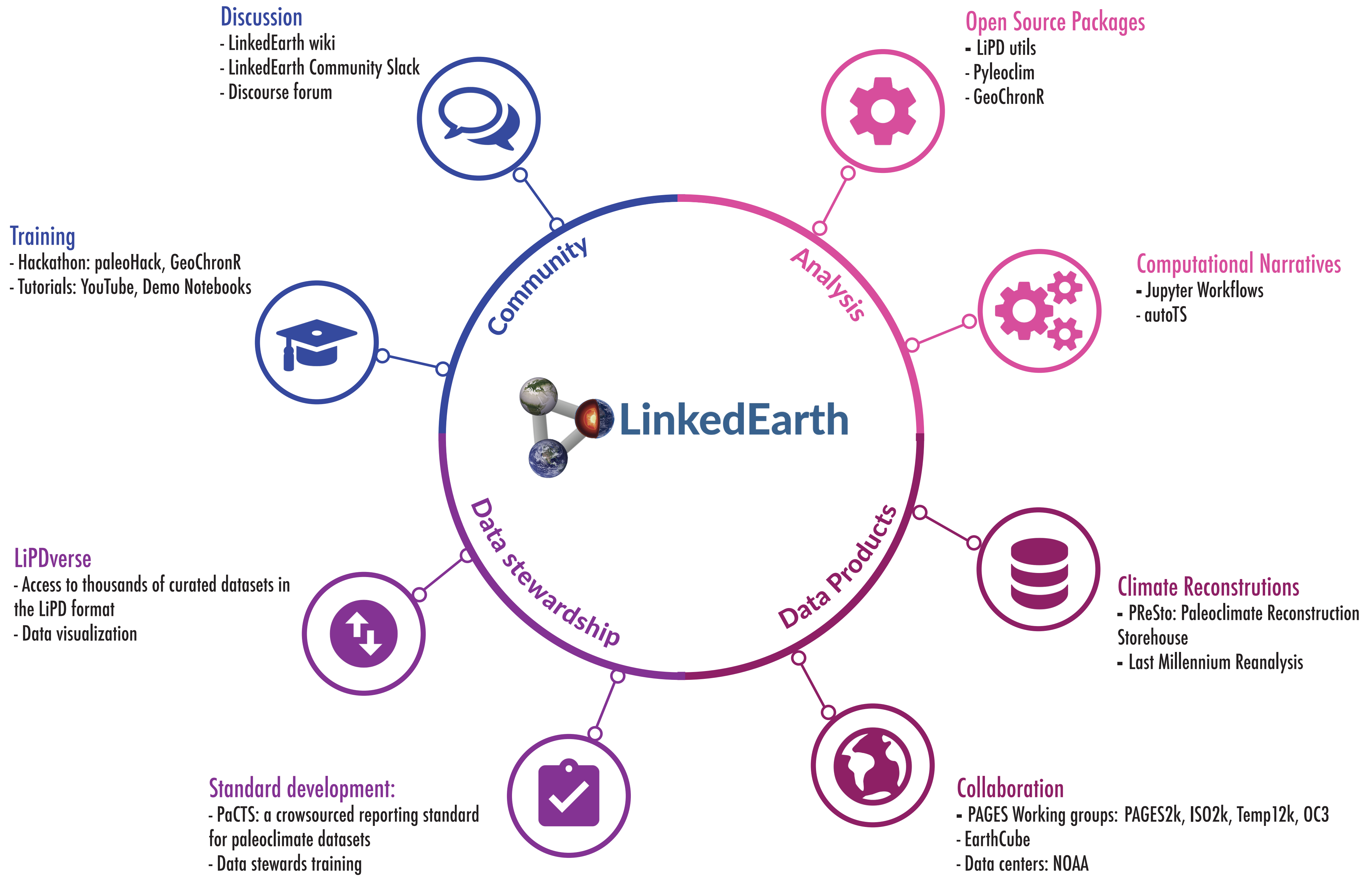


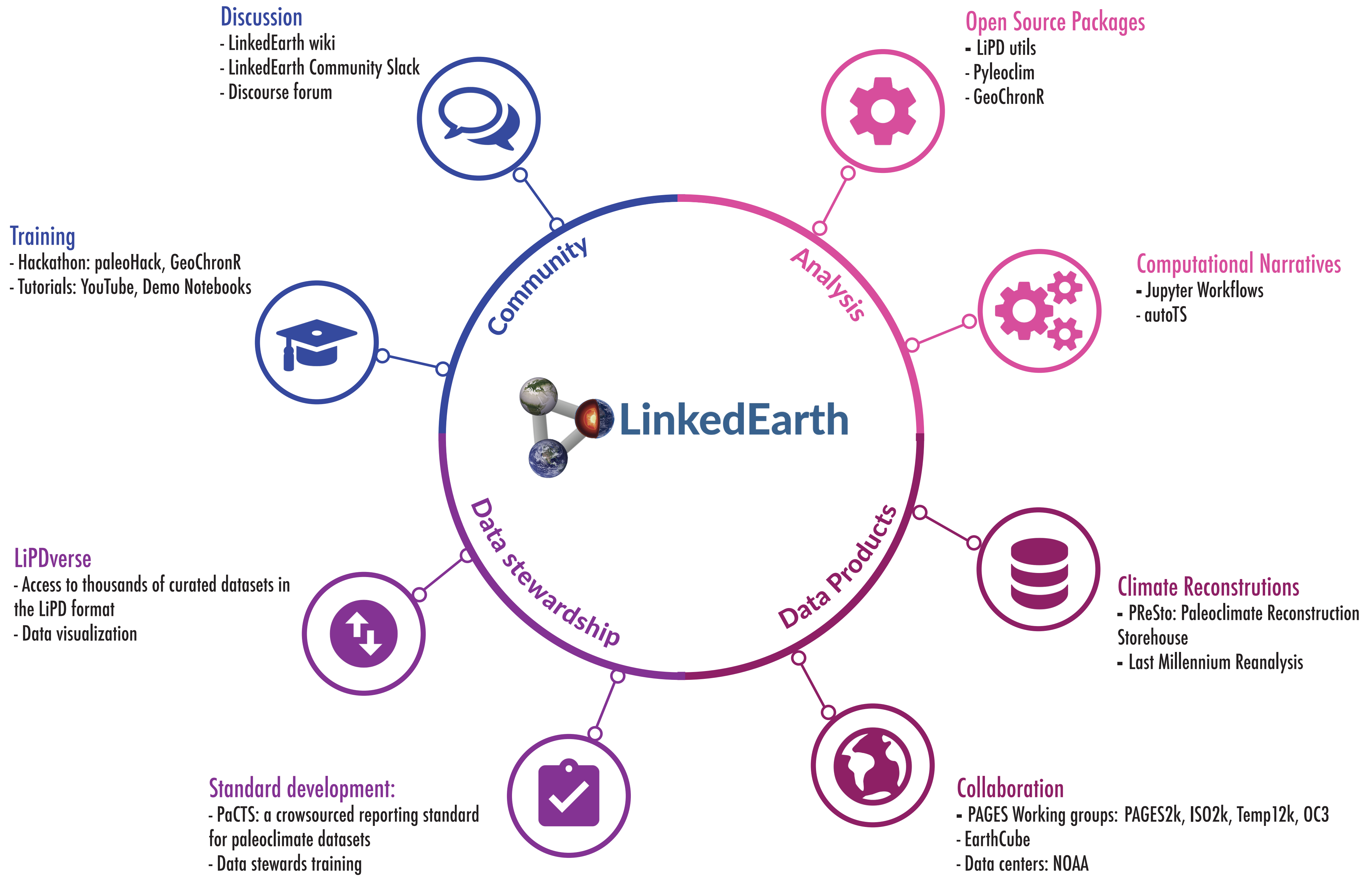
NetCDF-CF



```
In [43]: air2d = air.isel(time=500)
In [44]: air2d.plot()
Out[44]: <matplotlib.collections.QuadMesh at 0x7f4ad45c8400>
```







The diagram features a large, light blue, semi-circular shape on the left side of the slide. The word "Analysis" is written in a dark blue, sans-serif font, following the curve of the top edge of this shape. A thin, dark blue line extends from the top edge of the semi-circle, passing through a small white circle, and then continues to connect to another small white circle located at the bottom-left corner of a red-bordered rounded rectangle. This rectangle contains the "Open Source Packages" section.

Open Source Packages

- LiPD utils
- Pyleoclim
- GeoChronR

The diagram features a large, light blue, semi-circular shape on the left side of the slide. The word "Analysis" is written in a dark blue, sans-serif font, following the curve of the top edge of this shape. A thin, dark blue line extends from the top edge of the semi-circle, passing through a small white circle, and then continues to connect to another small white circle located at the bottom-left corner of a red-bordered rounded rectangle. This rectangle contains the "Computational Narratives" section.

Computational Narratives

- Jupyter Workflows
- autoTS

Pyleoclim

Pyleoclim

What is it?

- ◆ Python package for the analysis and visualization of paleoclimate data
- ◆ The package works with data stored in the LiPD format.
- ◆ Data transformation is made automatically within the package

Pyleoclim

What is it?

- ✦ Python package for the analysis and visualization of paleoclimate data
- ✦ The package works with data stored in the LiPD format.
- ✦ Data transformation is made automatically within the package

```
In [2]: d_cave = pyleo.Lipd('../data/Crystal.McCabe-Glynn.2013.lpd')
```

Disclaimer: LiPD files may be updated and modified to adhere to standards

```
reading: Crystal.McCabe-Glynn.2013.lpd  
1.42 MB :That's a big file! This may take a while to load...  
Finished read: 1 record
```

Let's load the d18O record.

```
In [3]: ts = d_cave.to_LipdSeries()
```

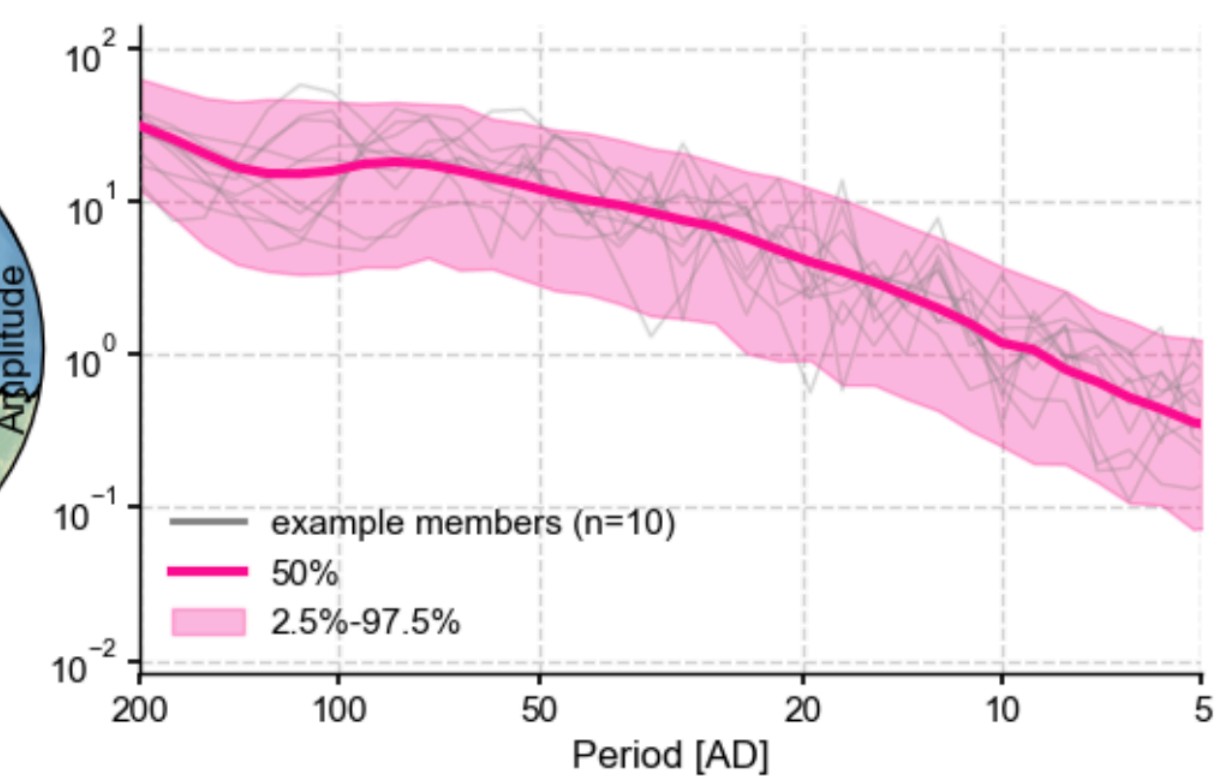
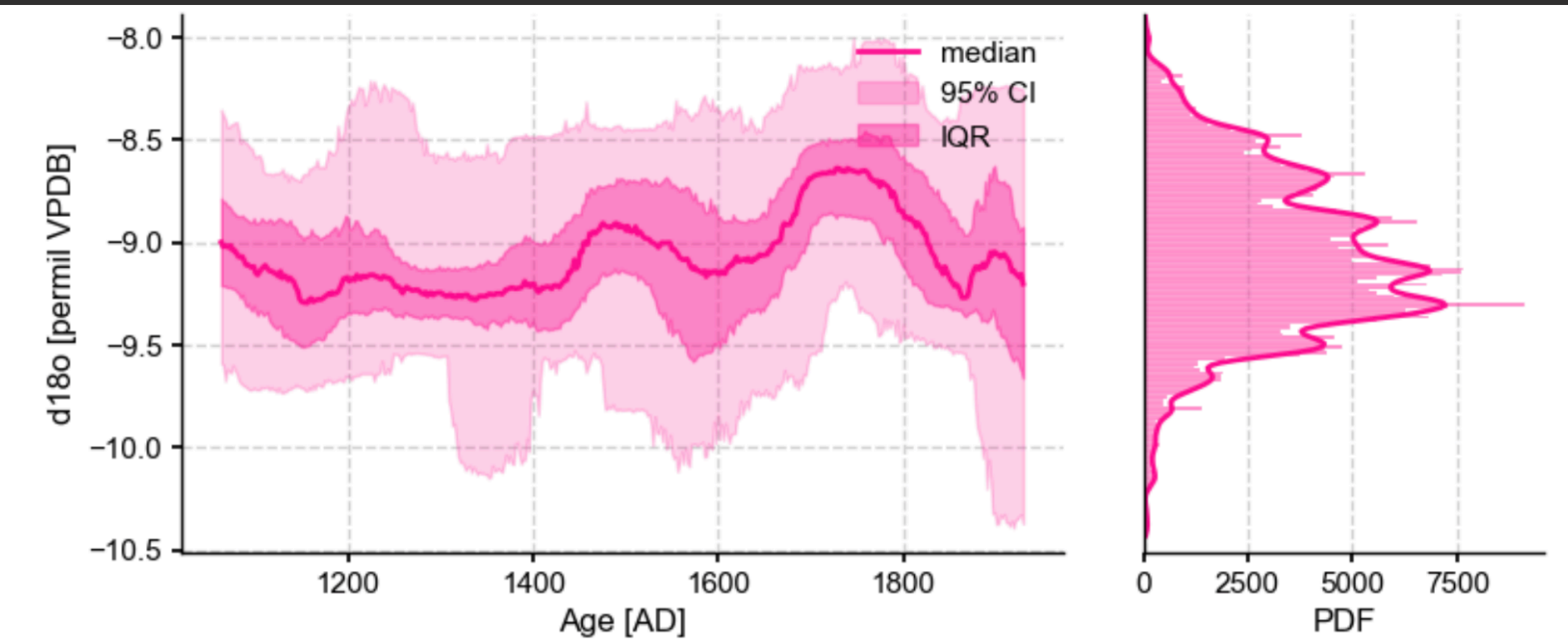
```
extracting paleoData...  
extracting: Crystal.McCabe-Glynn.2013  
Created time series: 3 entries  
0 : Crystal.McCabe-Glynn.2013 : speleothem : depth  
1 : Crystal.McCabe-Glynn.2013 : speleothem : age  
2 : Crystal.McCabe-Glynn.2013 : speleothem : d18o  
Enter the number of the variable you wish to use: 2
```


Pyleoclim

What is it?

- ◆ Python package for the analysis and visualization of paleoclimate data
- ◆ The package works with data stored in the LiPD format.
- ◆ Data transformation is made automatically within the package

```
ts.dashboard(metadata=False, ensemble=True, D=d_cave)
```



Pyleoclim

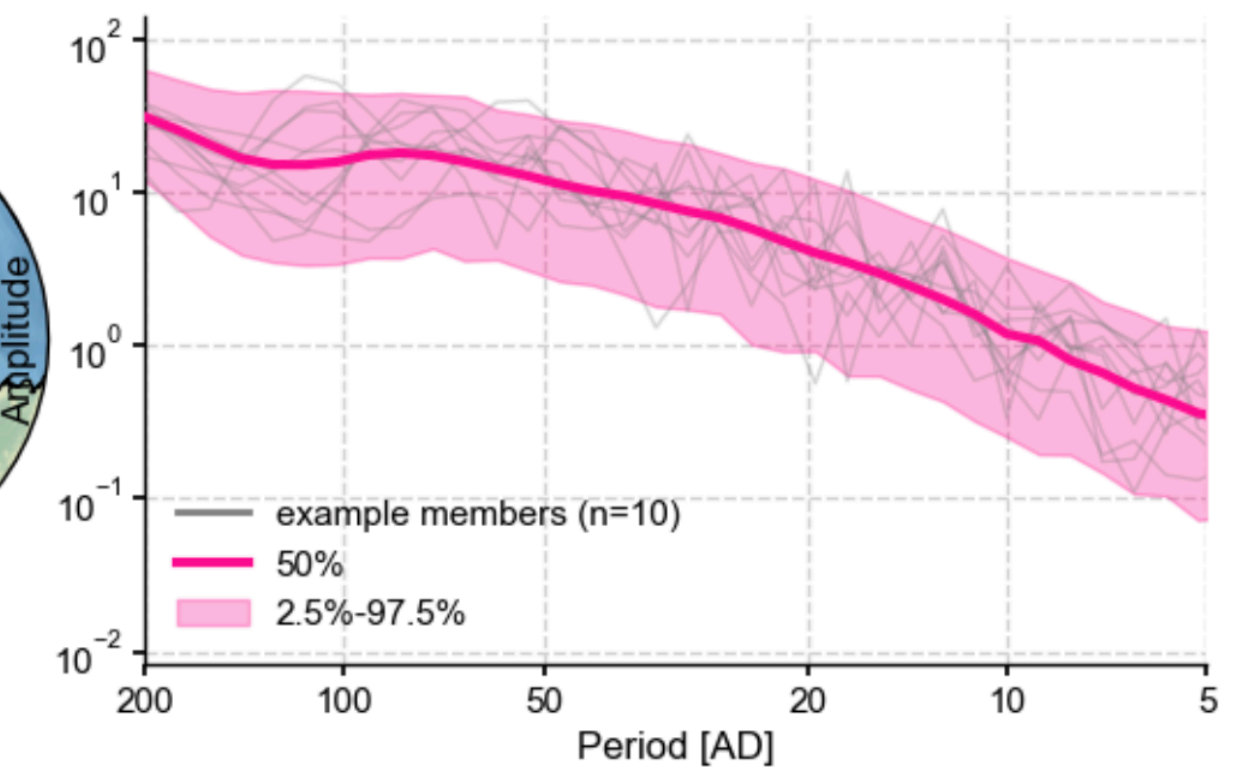
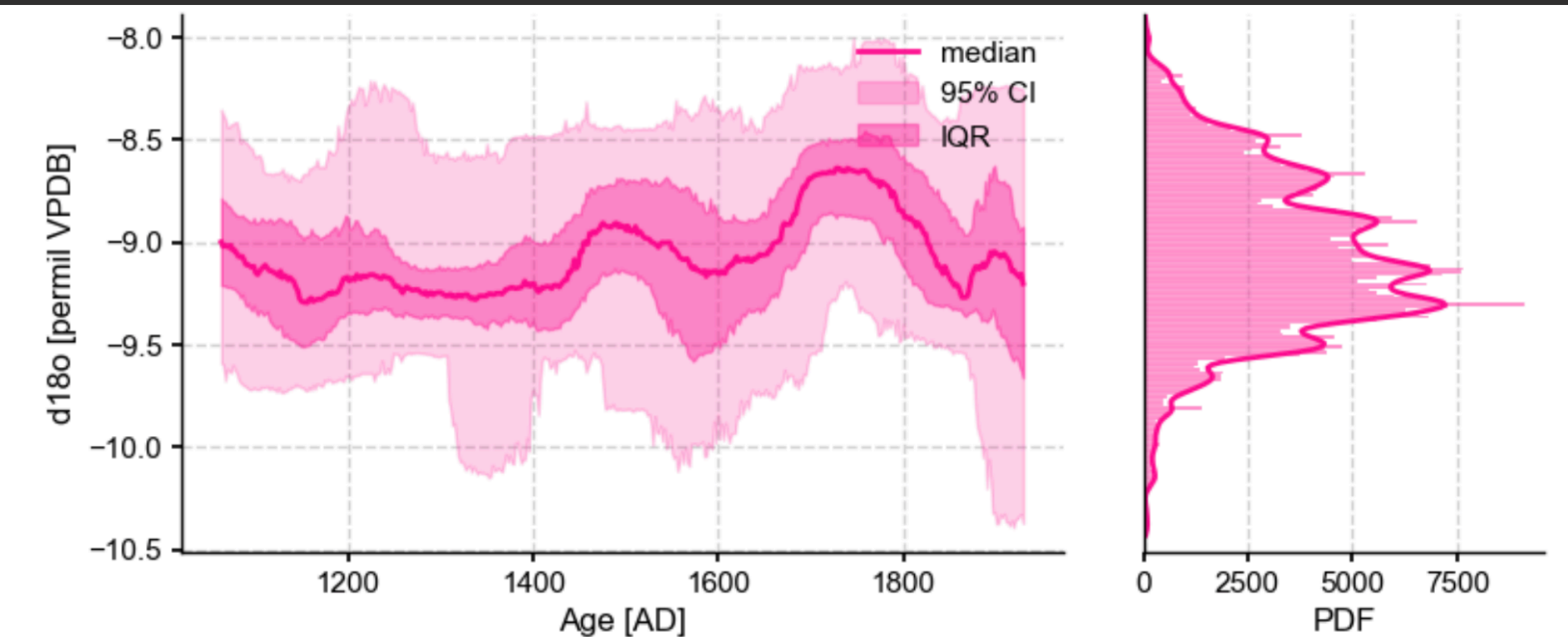
What is it?

- ◆ Python package for the analysis and visualization of paleoclimate data
- ◆ The package works with data stored in the LiPD format.
- ◆ Data transformation is made automatically within the package

What does it do?

- ◆ Mapping
- ◆ Plotting
- ◆ Timeseries analysis
 - ❖ Spectral
 - ❖ Wavelet
 - ❖ Causality
 - ❖ Pre-processing

```
ts.dashboard(metadata=False, ensemble=True, D=d_cave)
```



Pyleoclim

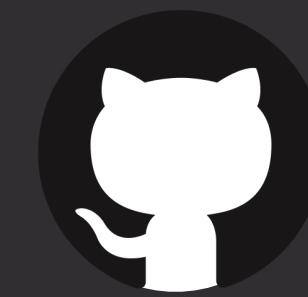
What is it?

- ◆ Python package for the analysis and visualization of paleoclimate data
- ◆ The package works with data stored in the LiPD format.
- ◆ Data transformation is made automatically within the package

What does it do?

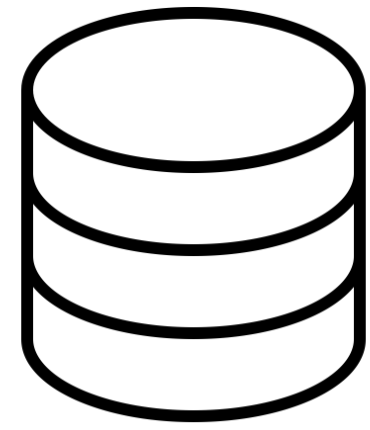
- ◆ Mapping
- ◆ Plotting
- ◆ Timeseries analysis
 - ❖ Spectral
 - ❖ Wavelet
 - ❖ Causality
 - ❖ Pre-processing

Where can I get it?

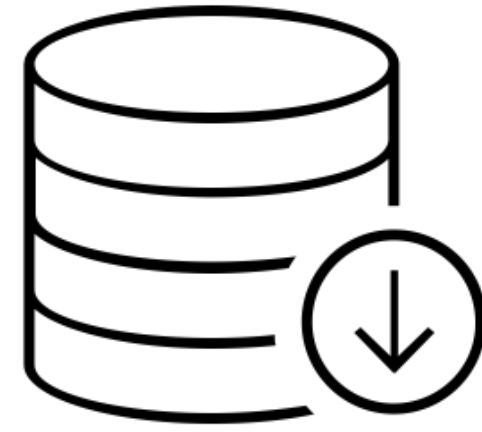


Github: LinkedEarth

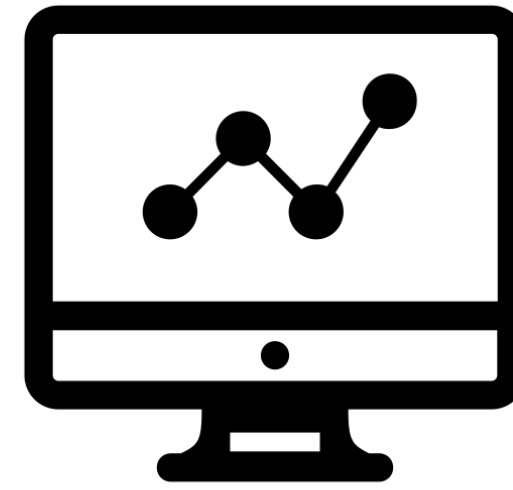
Pypi



Query Data Catalog



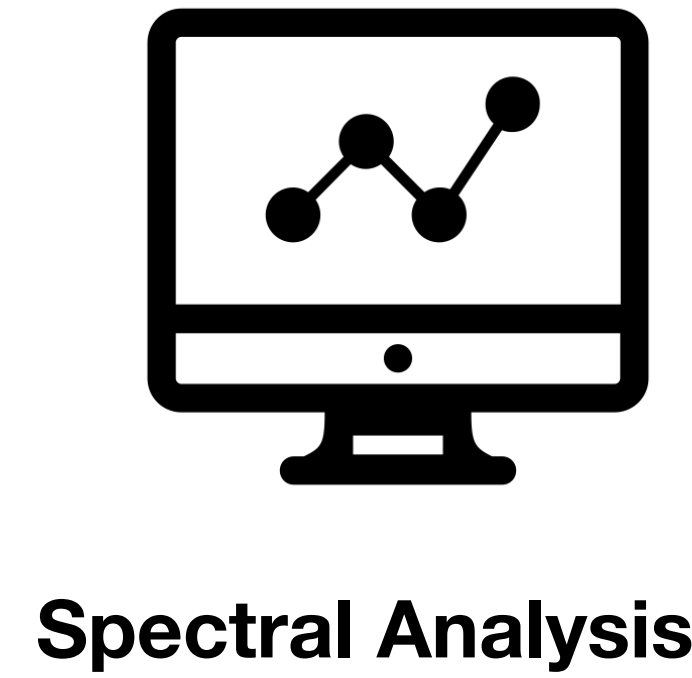
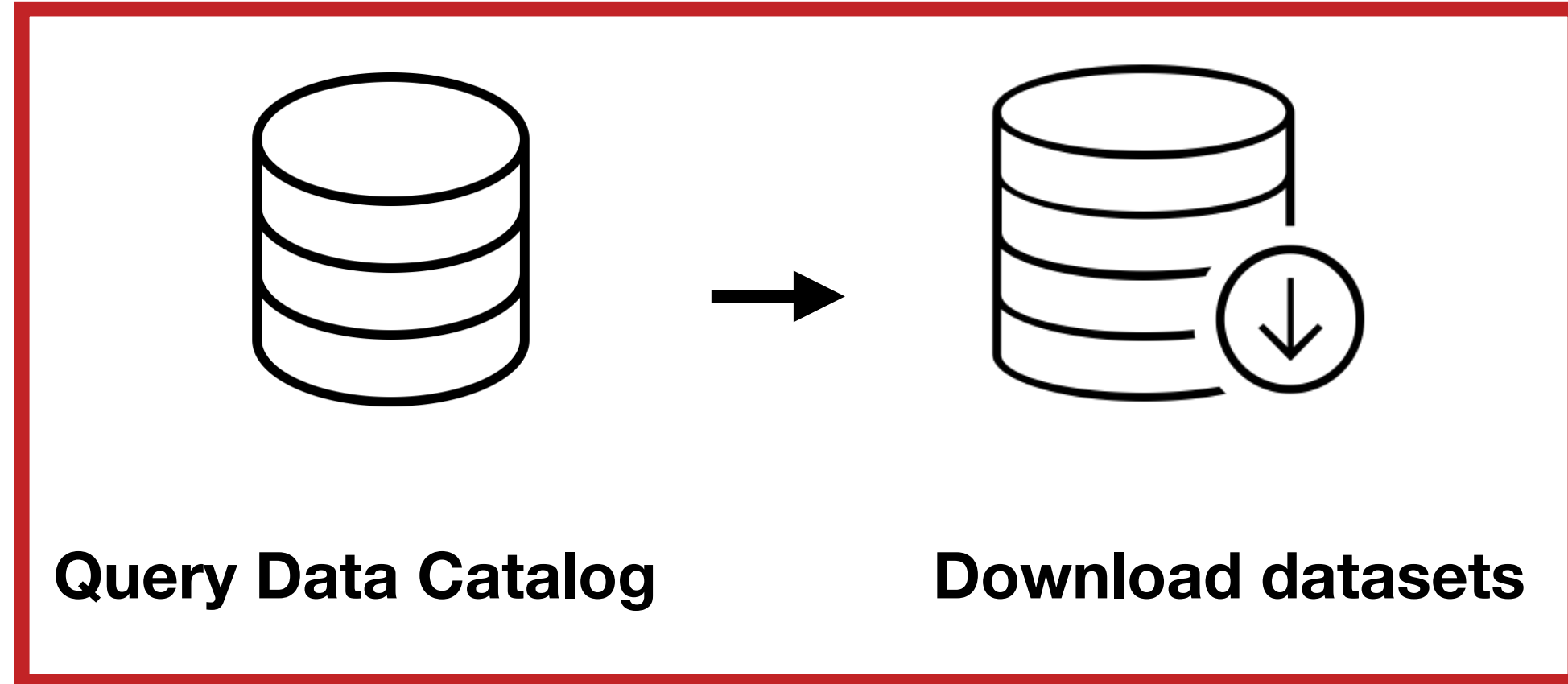
Download datasets



Spectral Analysis



**Visualize and Interpret
Results**

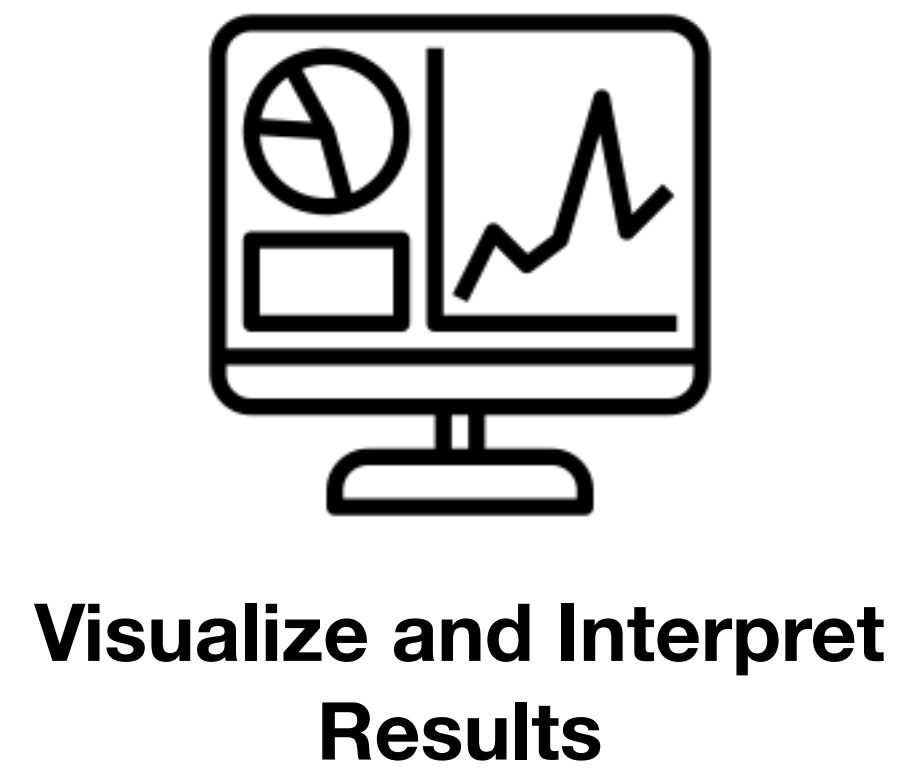
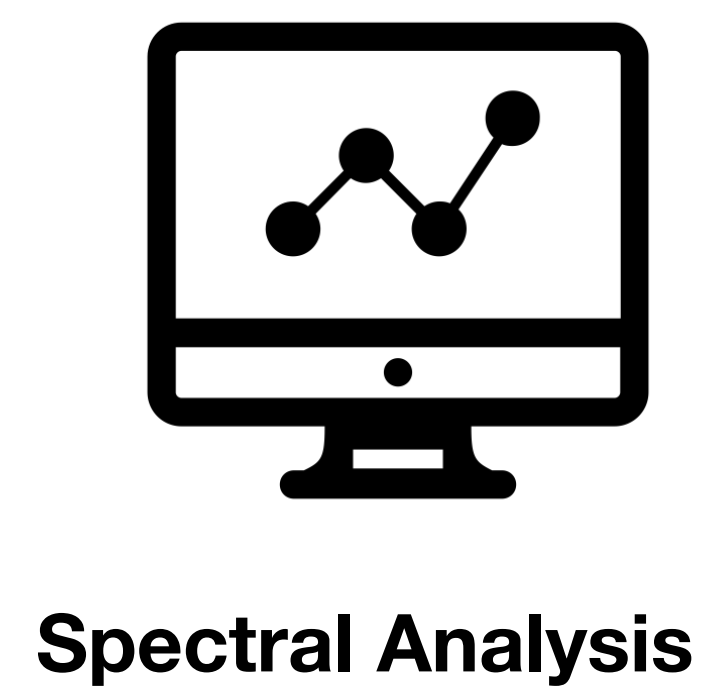
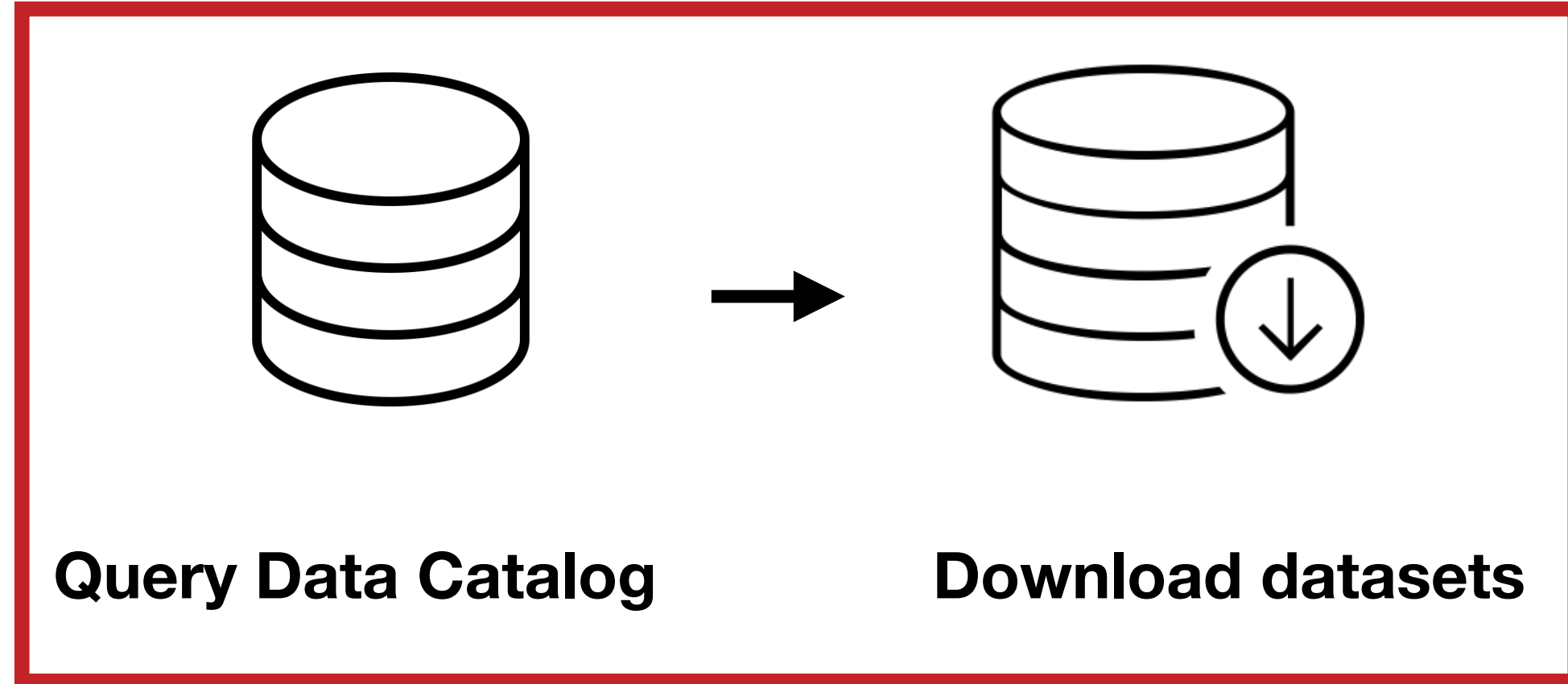


Query

First, we will query the [LinkedEarth database](#) for marine sedimentary records recording sea surface temperature spanning the [Holocene](#) epoch. The database has a [SPARQL](#) endpoint, which is directly accessible through the Pyleoclim package.

```
In [2]: # variables and parameters
archiveType=["marine sediment", "Marine Sediment"]
proxyObsType=[ ]
infVarType=["Sea Surface Temperature"]
sensorGenus=[ ]
sensorSpecies=[ ]
interpName=["temperature", "Temperature"]
interpDetail=["sea surface"]
ageUnits=["kyr BP"]
ageBound=[0,10]
ageBoundType=["any"]
recordLength=[4]
resolution=[ ]
lat=[-14,1.5]
lon=[110,135]
alt=[-10000,0]
#functions
res = pyleo.utils.queryLinkedEarth(archiveType=archiveType, proxyObsType=proxyObsType,
infVarType = infVarType, sensorGenus=sensorGenus,
sensorSpecies=sensorSpecies, interpName=interpName,
interpDetail =interpDetail, ageUnits = ageUnits,
ageBound = ageBound, ageBoundType = ageBoundType,
recordLength = recordLength, resolution = resolution,
lat = lat, lon = lon, alt = alt,
print_response = True, download_lipd = True,
download_folder = './lipd')
#outputs
```

<http://wiki.linked.earth/Special:URIResolver/MD01-2D2378.Xu.2008>
<http://wiki.linked.earth/Special:URIResolver/GeoB10069-2D3.Gibbons.2014>



Query

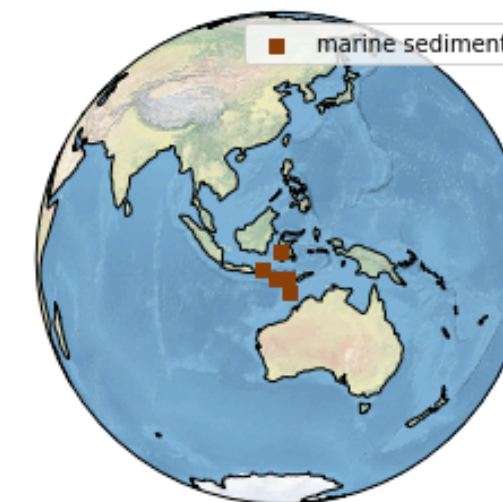
First, we will query the [LinkedEarth database](#) for marine sedimentary records recording sea surface temperature spanning the [Holocene](#) epoch. The database has a [SPARQL](#) endpoint, which is directly accessible through the Pyleoclim package.

```
In [2]: # variables and parameters
archiveType=["marine sediment", "Marine Sediment"]
proxyObsType=[ ]
infVarType=["Sea Surface Temperature"]
sensorGenus=[ ]
sensorSpecies=[ ]
interpName=["temperature", "Temperature"]
interpDetail=["sea surface"]
ageUnits=["kyr BP"]
ageBound=[0,10]
ageBoundType=["any"]
recordLength=[4]
resolution=[ ]
lat=[-14,1.5]
lon=[110,135]
alt=[-10000,0]
#functions
res = pyleo.utils.queryLinkedEarth(archiveType=archiveType, proxyObsType=proxyObsType,
infVarType = infVarType, sensorGenus=sensorGenus,
sensorSpecies=sensorSpecies, interpName=interpName,
interpDetail =interpDetail, ageUnits = ageUnits,
ageBound = ageBound, ageBoundType = ageBoundType,
recordLength = recordLength, resolution = resolution,
lat = lat, lon = lon, alt = alt,
print_response = True, download_lipd = True,
download_folder = './lipd')
#outputs
http://wiki.linked.earth/Special:URIResolver/MD01-2D2378.Xu.2008
http://wiki.linked.earth/Special:URIResolver/GeoB10069-2D3.Gibbons.2014
```

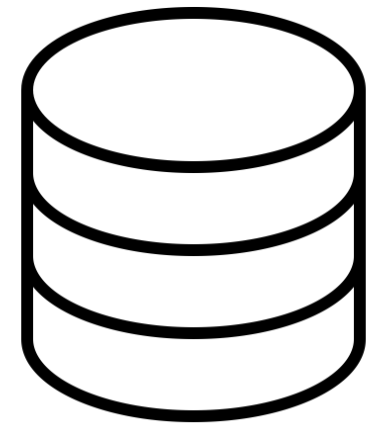
```
In [9]: # variables and parameters
#functions
D = pyleo.Lipd(usr_path=os.getcwd())
D.mapAllArchive(projection='Orthographic', proj_default={'central_longitude':120, '
central_latitude':-5})
#outputs
```

Disclaimer: LiPD files may be updated and modified to adhere to standards

Found: 5 LiPD file(s)
 reading: BJ8-03-13GGC.Linsley.2010.lpd
 reading: GeoB10069-3.Gibbons.2014.lpd
 reading: MD98-2165.Levi.2007.lpd
 reading: BJ8-03-70GGC.Linsley.2010.lpd
 reading: MD01-2378.Xu.2008.lpd
 Finished read: 5 records



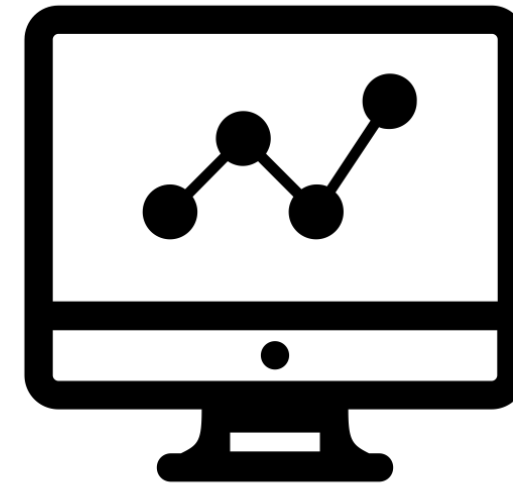
Out[9]: (<Figure size 432x288 with 1 Axes>, <cartopy.mpl.geoaxes.GeoAxesSubplot at 0x1a29230e48>)



Query Data Catalog



Download datasets



Spectral Analysis



Visualize and Interpret Results

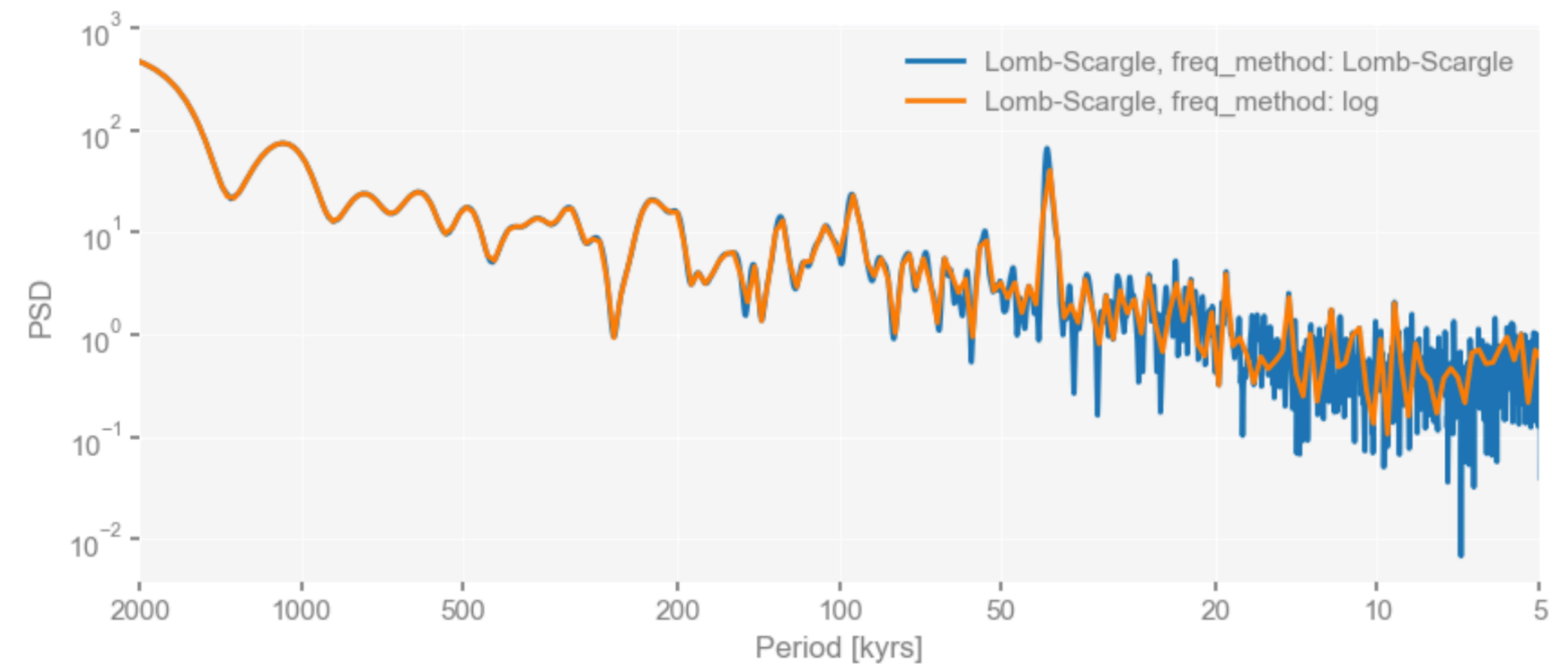
Query

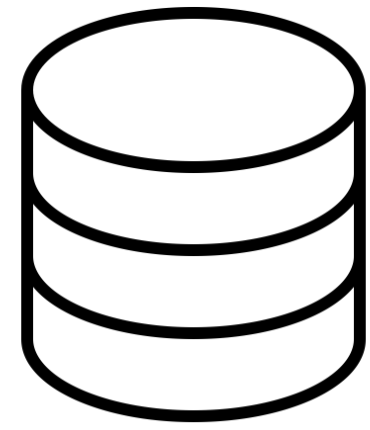
First, we will query the [LinkedEarth database](#) for marine sedimentary records recording sea surface temperature spanning the [Holocene](#) epoch. The database has a [SPARQL](#) endpoint, which is directly accessible through the Pyleoclim package.

```
In [2]: # variables and parameters
archiveType=["marine sediment","Marine Sediment"]
proxyObsType=[ ]
infVarType=["Sea Surface Temperature"]
sensorGenus=[ ]
sensorSpecies=[ ]
interpName=["temperature","Temperature"]
interpDetail=["sea surface"]
ageUnits=["kyr BP"]
ageBound=[0,10]
ageBoundType=["any"]
recordLength=[4]
resolution=[ ]
lat=[-14,1.5]
lon=[110,135]
alt=[-10000,0]
#functions
res = pyleo.utils.queryLinkedEarth(archiveType=archiveType, proxyObsType=proxyObsType,
infVarType = infVarType, sensorGenus=sensorGenus,
sensorSpecies=sensorSpecies, interpName=interpName,
interpDetail =interpDetail, ageUnits = ageUnits,
ageBound = ageBound, ageBoundType = ageBoundType,
recordLength = recordLength, resolution = resolution,
lat = lat, lon = lon, alt = alt,
print_response = True, download_lipd = True,
download_folder = './lipd')
#outputs
```

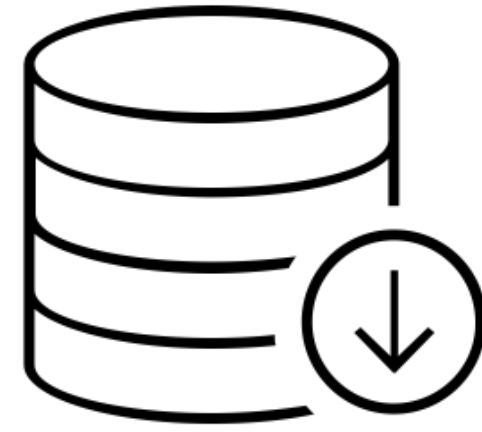
<http://wiki.linked.earth/Special:URIResolver/MD01-2D2378.Xu.2008>
<http://wiki.linked.earth/Special:URIResolver/GeoB10069-2D3.Gibbons.2014>

```
In [8]: psd_ls_ls = ts.standardize().spectral(method='lomb_scargle',freq_method = 'lomb_scargle')
fig, ax = psd_ls_ls.plot(label='Lomb-Scargle, freq_method: Lomb-Scargle', mute = True)
psd_ls.plot(label='Lomb-Scargle, freq_method: log', mute = True, ax=ax)
pyleo.showfig(fig)
```

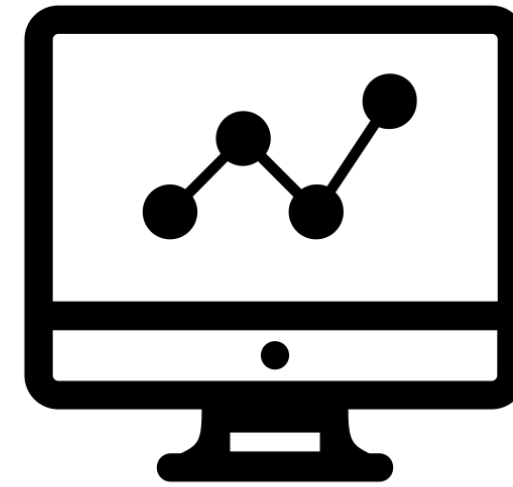




Query Data Catalog



Download datasets

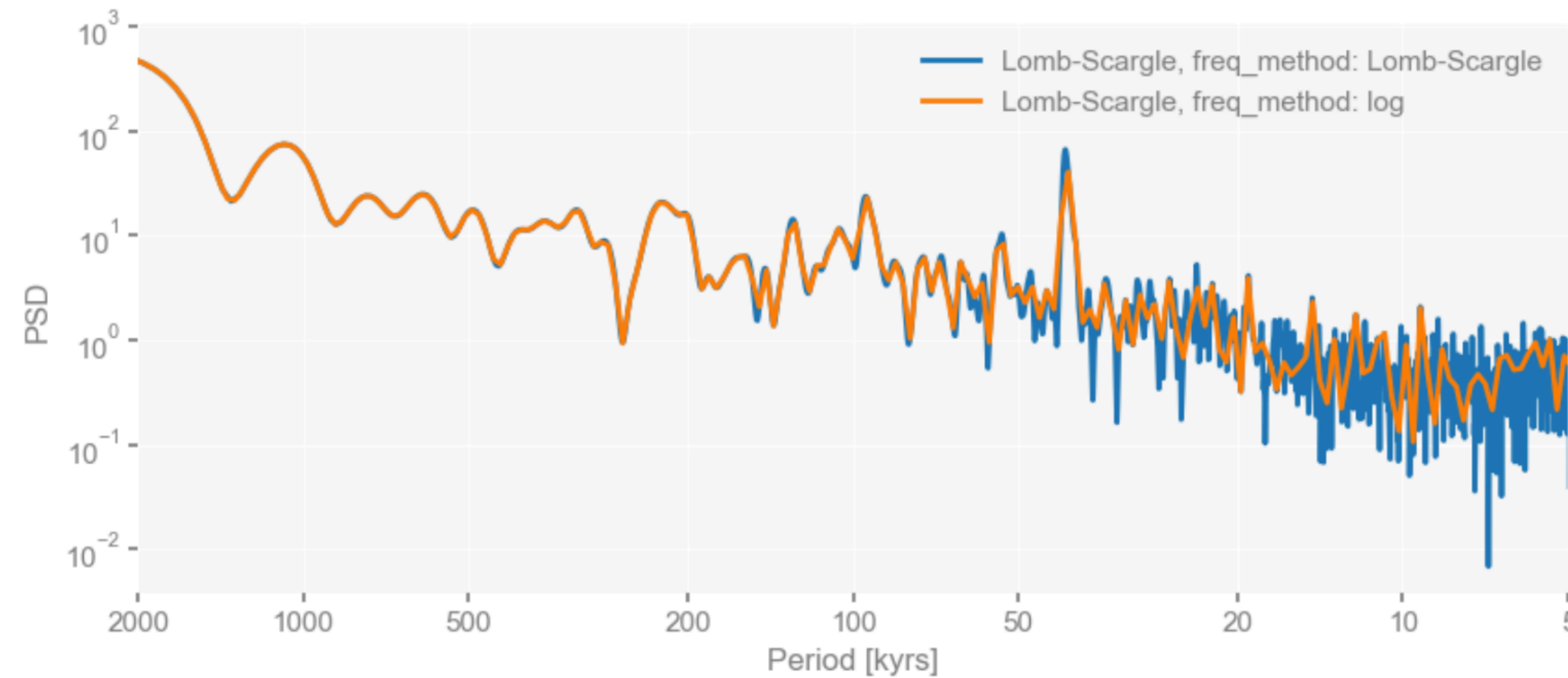


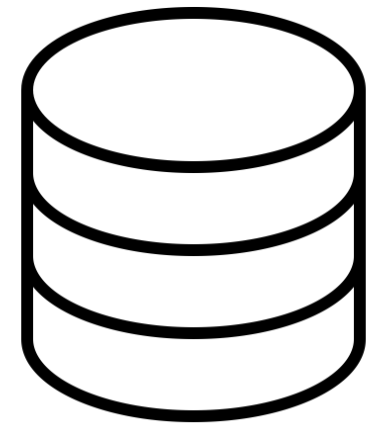
Spectral Analysis



Visualize and Interpret Results

```
In [8]: psd_ls_ls = ts.standardize().spectral(method='lomb_scargle',freq_method = 'lomb_scargle')
fig, ax = psd_ls_ls.plot(label='Lomb-Scargle, freq_method: Lomb-Scargle', mute = True)
psd_ls.plot(label='Lomb-Scargle, freq_method: log', mute = True, ax=ax)
pyleo.showfig(fig)
```

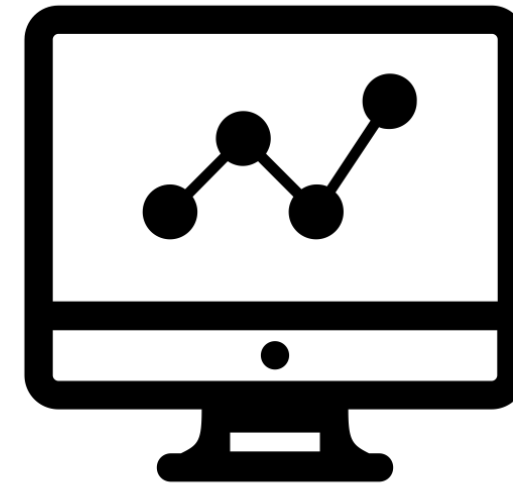




Query Data Catalog



Download datasets

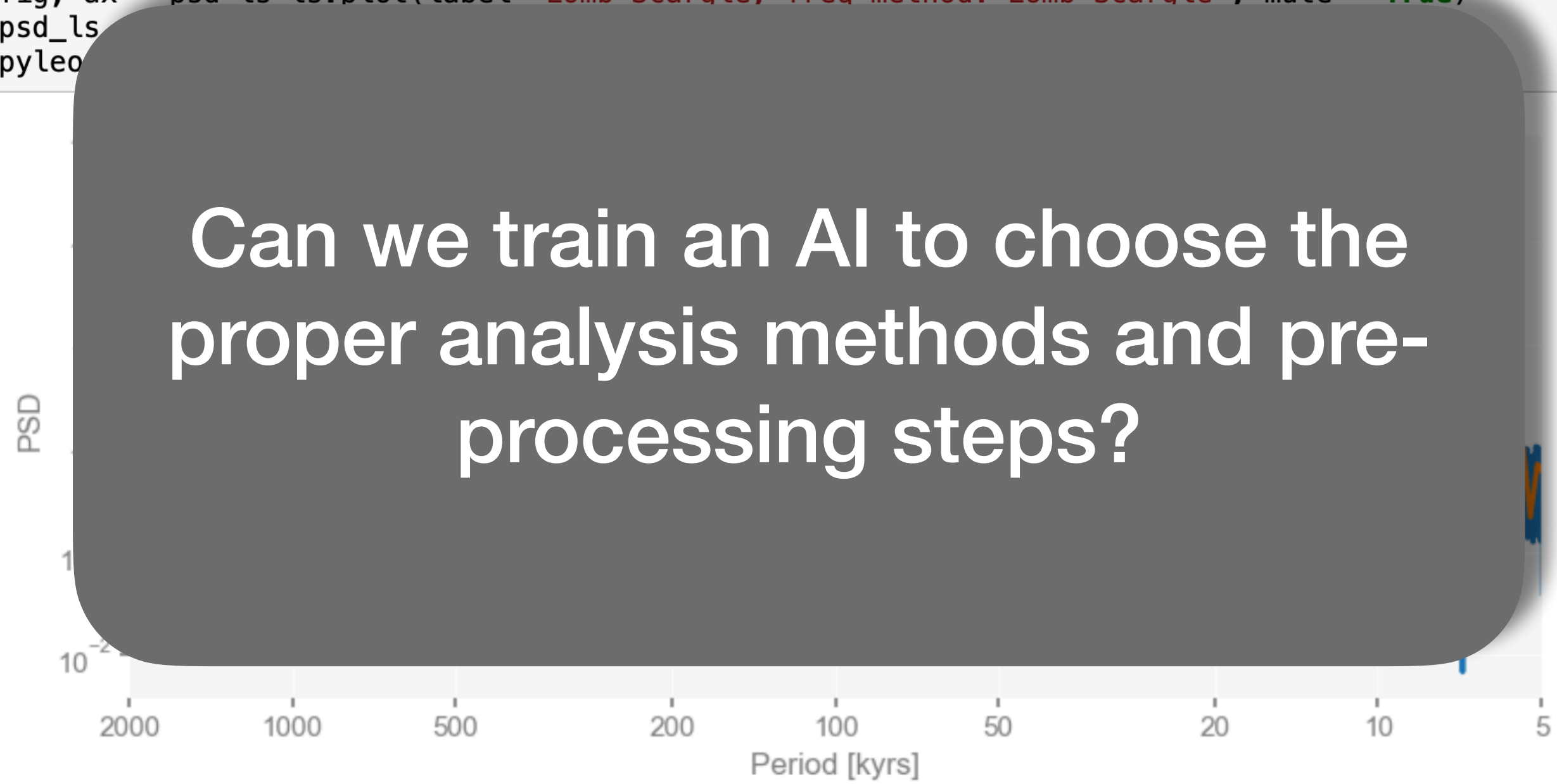


Spectral Analysis



Visualize and Interpret Results

```
In [8]: psd_ls_ls = ts.standardize().spectral(method='lomb_scargle', freq_method = 'lomb_scargle')
fig, ax = psd_ls_ls.plot(label='Lomb-Scargle, freq method: Lomb-Scargle', mute = True)
psd_ls
pyleo
```



Can we train an AI to choose the proper analysis methods and pre-processing steps?

autoTS: Automated Machine Learning for Time Series Analysis



Yolanda Gil
USC



Deborah Khider
USC



Feng Zhu
USC



Pratheek Atherya
USC



Myron Kwan
USC



Varun Ratnakar
USC



Analysis



Open Source Packages

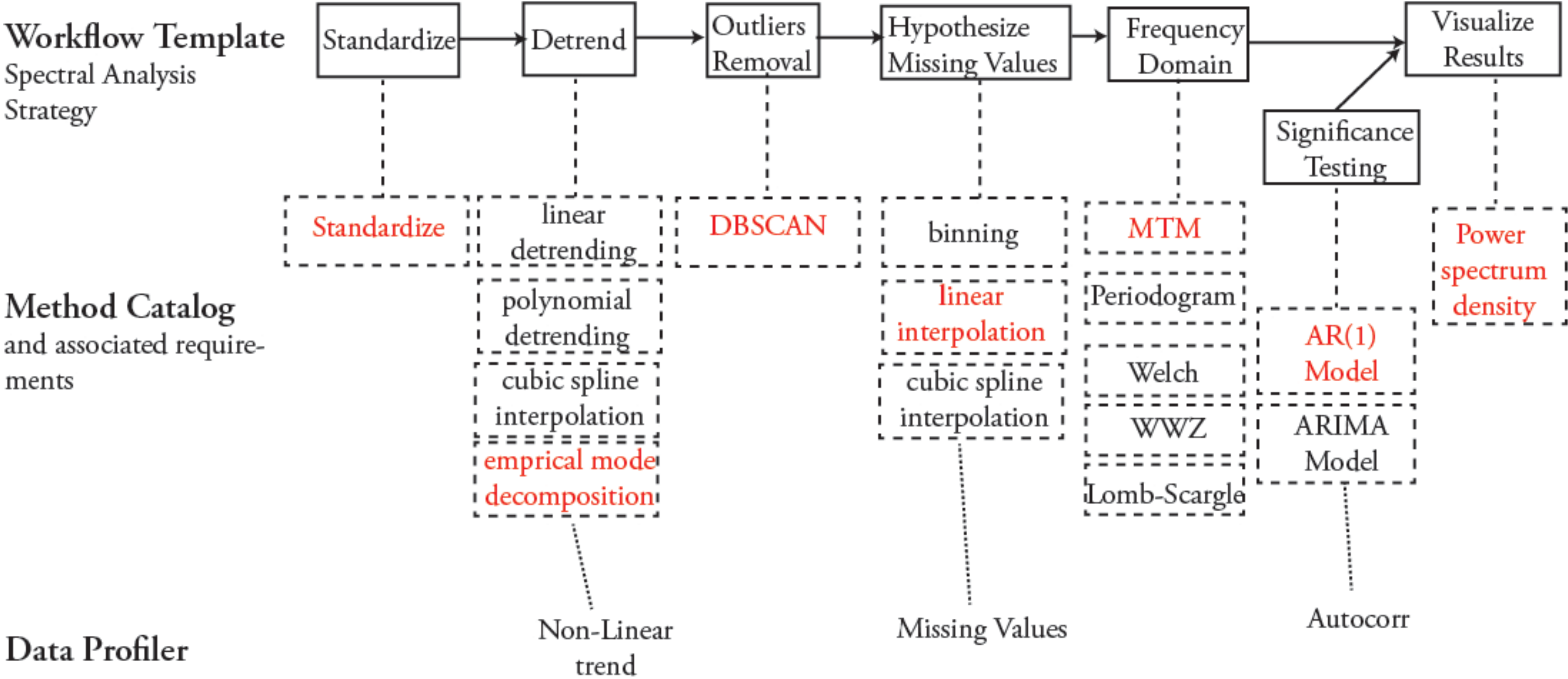
- LiPD utils
- Pyleoclim
- GeoChronR



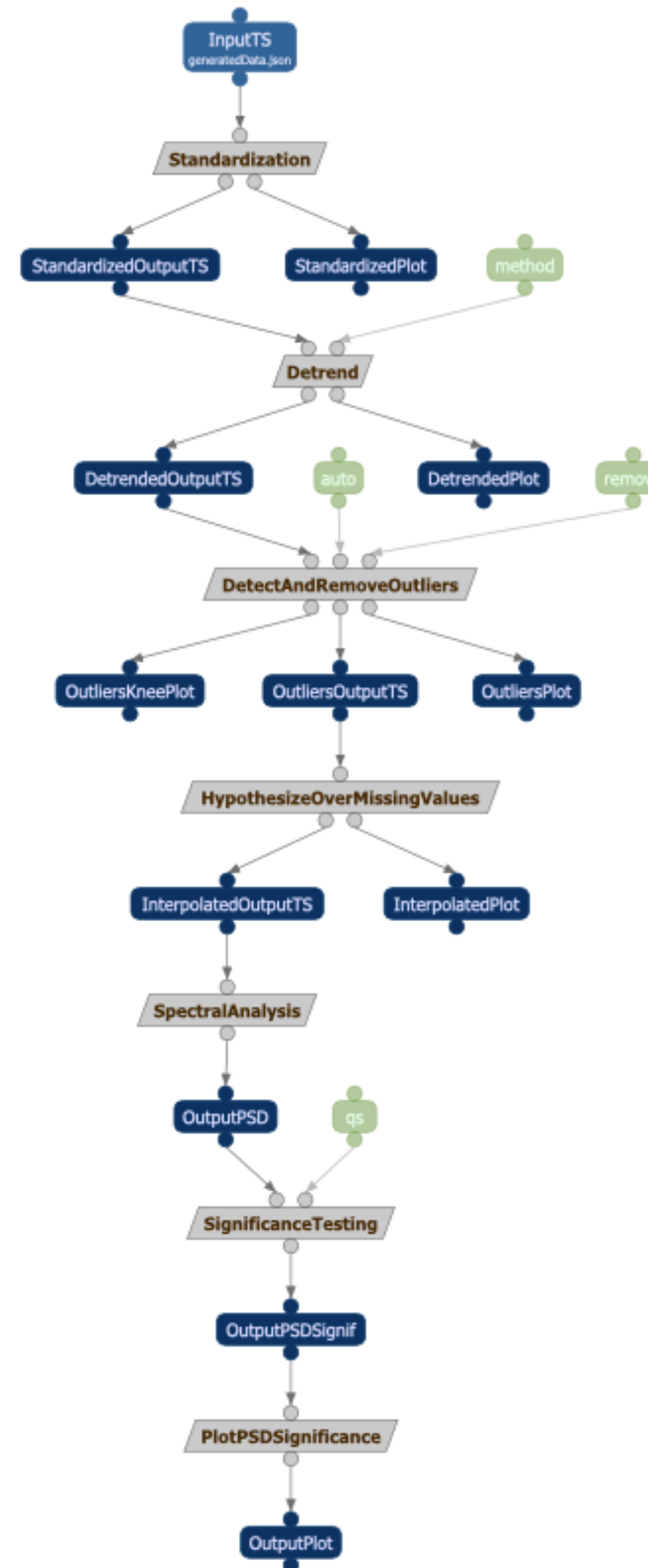
Computational Narratives

- Jupyter Workflows
- autoTS

Strategy

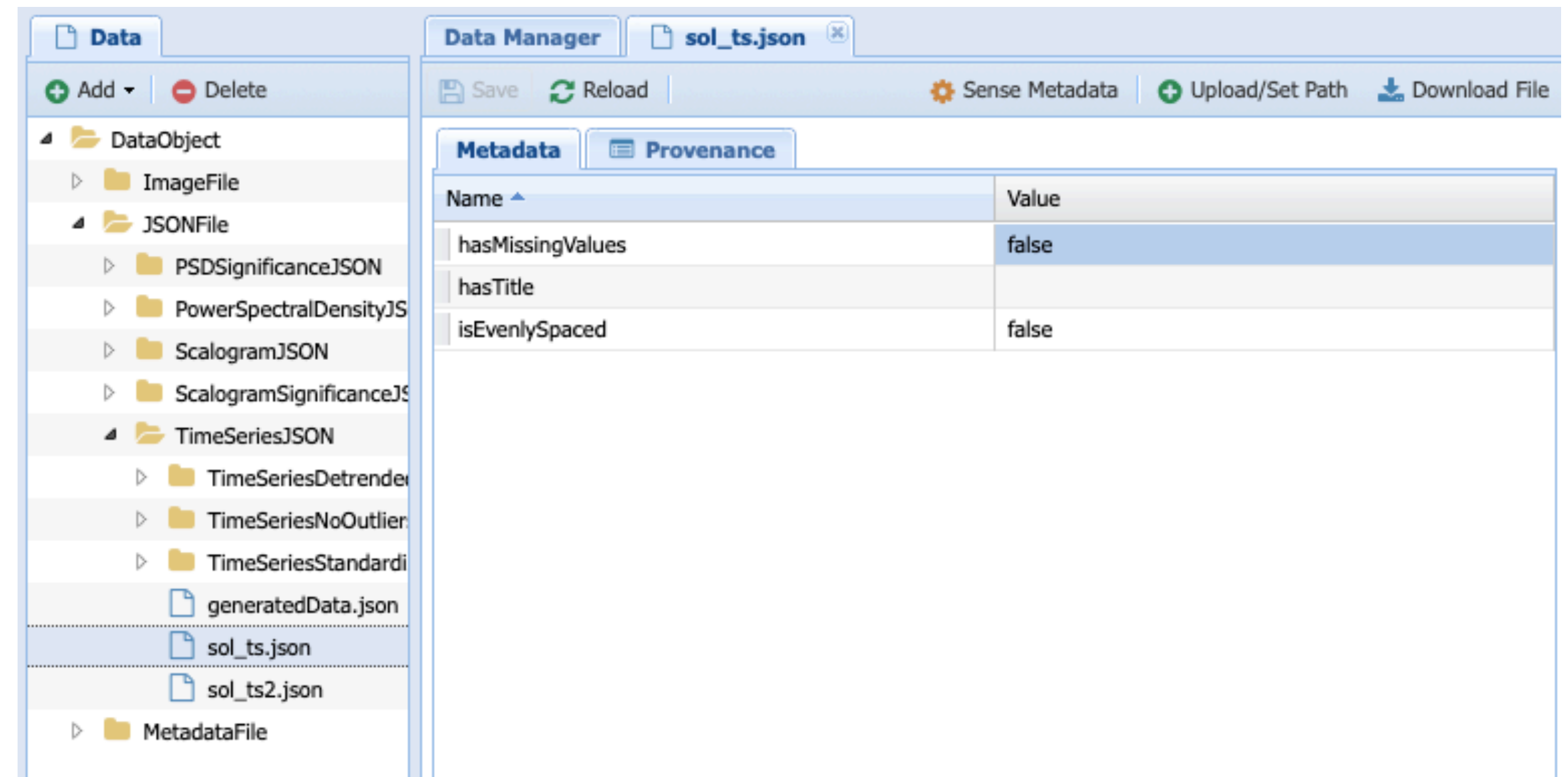


WINGS can capture all the steps involved in spectral analysis in a abstract workflow template



- Analysis
 - SignificanceTesting
 - SignificanceTestingDefault
 - SpectralAnalysis
 - LombScargle
 - MTM
 - Periodogram
 - WWZ
 - Welch
 - WaveletAnalysis
 - WWZ-wavelet
 - WaveletSignificanceTesting
 - WaveletAR1sim
- MetadataExtraction
 - TimeSeriesMetadataExtraction
 - TSMetadataExtraction
- PreProcessing
 - DetectAndRemoveOutliers
 - RemoveOutliers
 - Detrend
 - DetrendDefault
 - HypothesizeOverMissingValues
 - LinearInterpolation
 - Standardization
 - StandardizationDefault
- Visualization
 - PlotPSDSignificance
 - PlotPSDSignificanceDefault
 - PlotScalogramSignificance
 - PlotScalogramSignificanceDefault

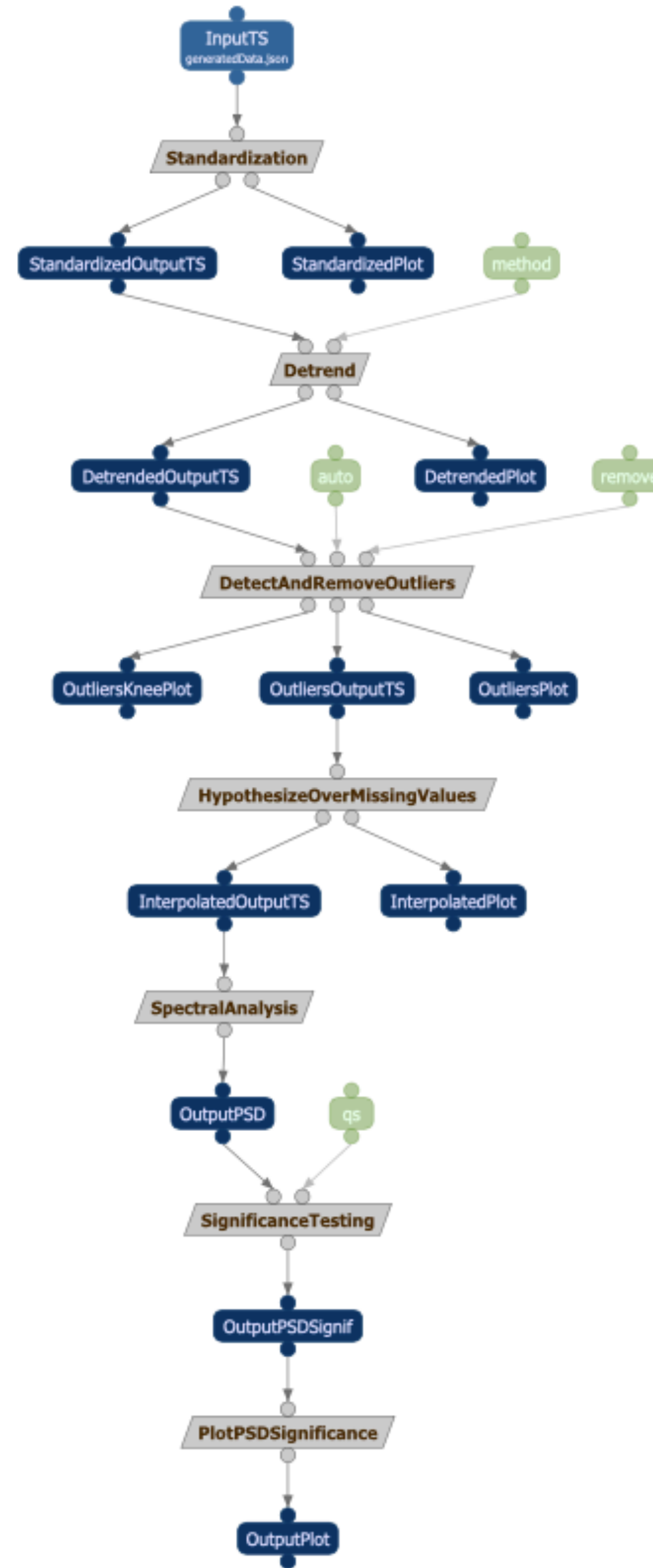
Each dataset has metadata properties that reflect its characteristics, and are used to express constraints that result in the generation of the most effective workflows



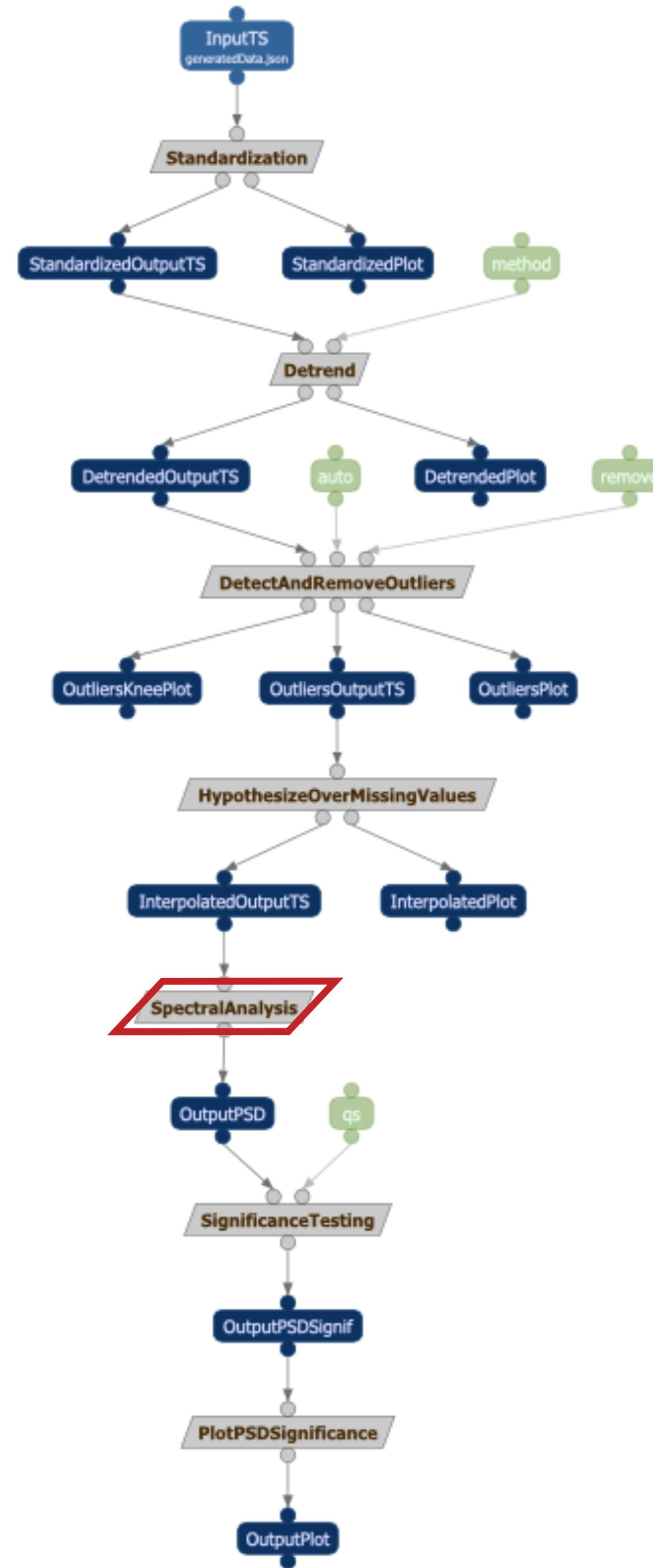
The screenshot displays a software interface for managing data. On the left, a 'Data' pane shows a hierarchical tree structure under 'DataObject'. The tree includes folders for 'ImageFile', 'JSONFile', 'PSDSignificanceJSON', 'PowerSpectralDensityJS', 'ScalogramJSON', 'ScalogramSignificanceJS', 'TimeSeriesJSON', 'TimeSeriesDetrende', 'TimeSeriesNoOutlier', and 'TimeSeriesStandardi'. Below these are files 'generatedData.json', 'sol_ts.json', and 'sol_ts2.json', and a 'MetadataFile' folder. The 'sol_ts.json' file is selected. On the right, the 'Data Manager' window shows the selected file 'sol_ts.json'. It has a toolbar with 'Save', 'Reload', 'Sense Metadata', 'Upload/Set Path', and 'Download File'. Below the toolbar are two tabs: 'Metadata' (active) and 'Provenance'. The 'Metadata' tab displays a table with the following data:

Name	Value
hasMissingValues	false
hasTitle	
isEvenlySpaced	false

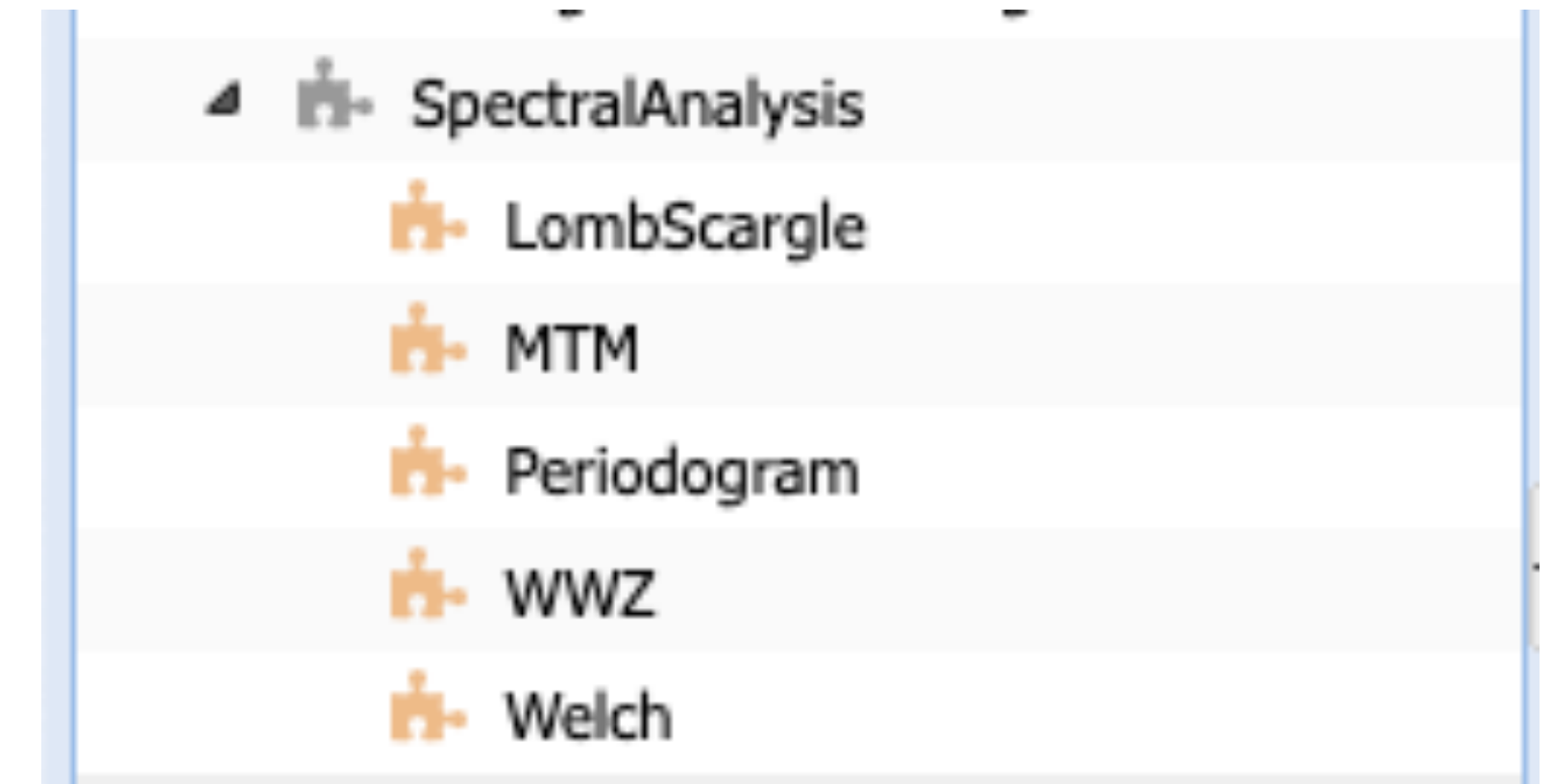
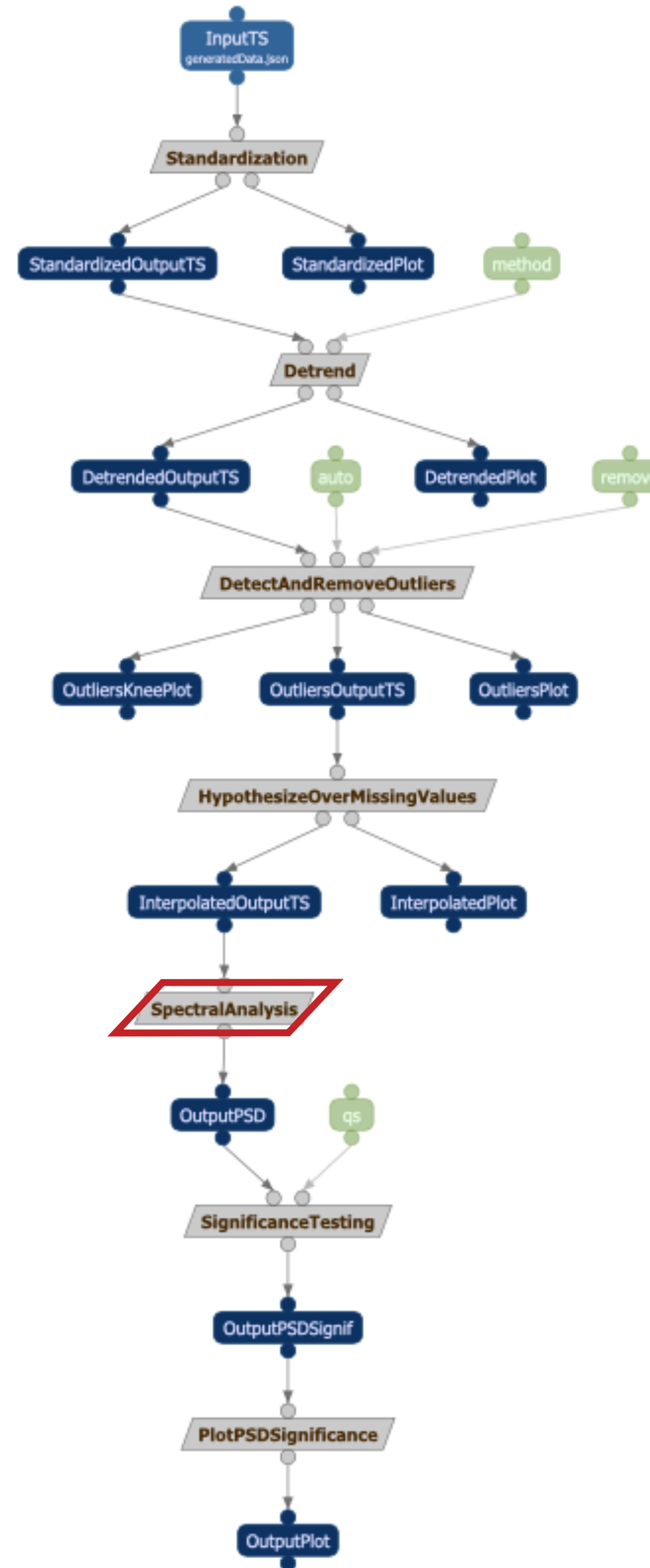
WINGS was extended to allow for a type of constraints that skips an abstract workflow step



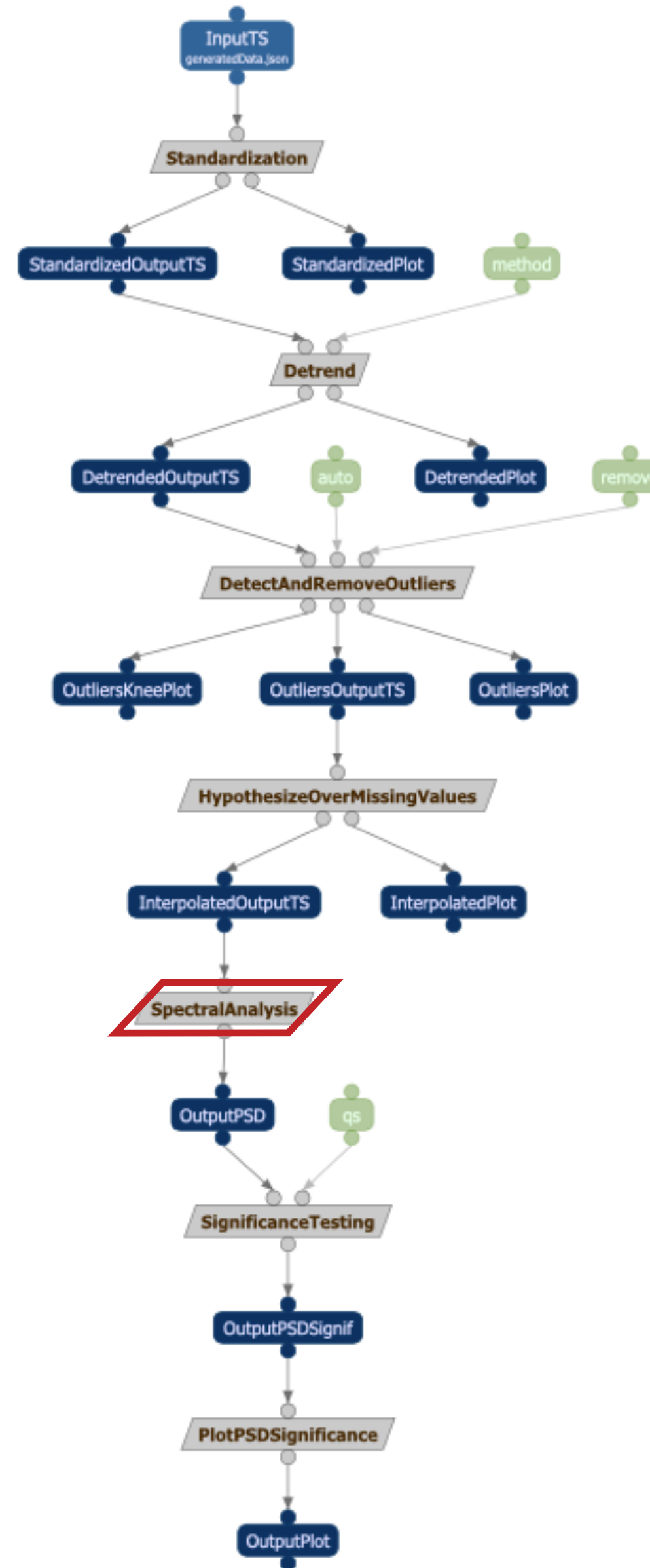
WINGS was extended to allow for a type of constraints that skips an abstract workflow step



WINGS was extended to allow for a type of constraints that skips an abstract workflow step



WINGS was extended to allow for a type of constraints that skips an abstract workflow step



SpectralAnalysis

- LombScargle
- MTM
- Periodogram
- WWZ
- Welch

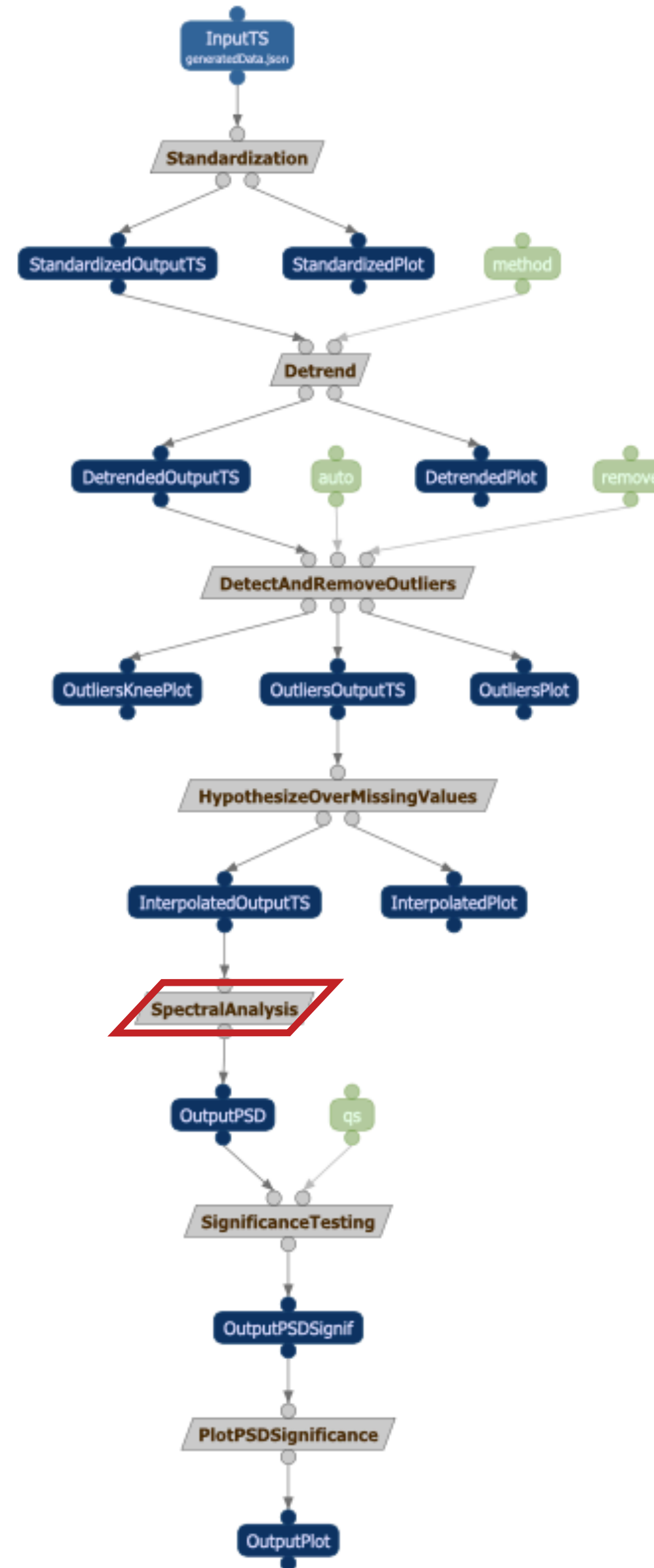
Add Rule ▾

```

1 [ RequirementOfInputToBeEvenlySpaced:
2   (?c rdf:type acdom:MTMClass)
3   (?c ac:hasInput ?InputTS)
4   (?InputTS ac:hasArgumentID 'InputTS')
5   -> (?InputTS dcdom:isEvenlySpaced 'true'^^xsd:boolean)
6   print(?c 'MTM requires evenly spaced data')
7 ]

```


WINGS was extended to allow for a type of constraints that skips an abstract workflow step



SpectralAnalysis

- LombScargle
- MTM
- Periodogram
- WWZ
- Welch

Add Rule ▾

```

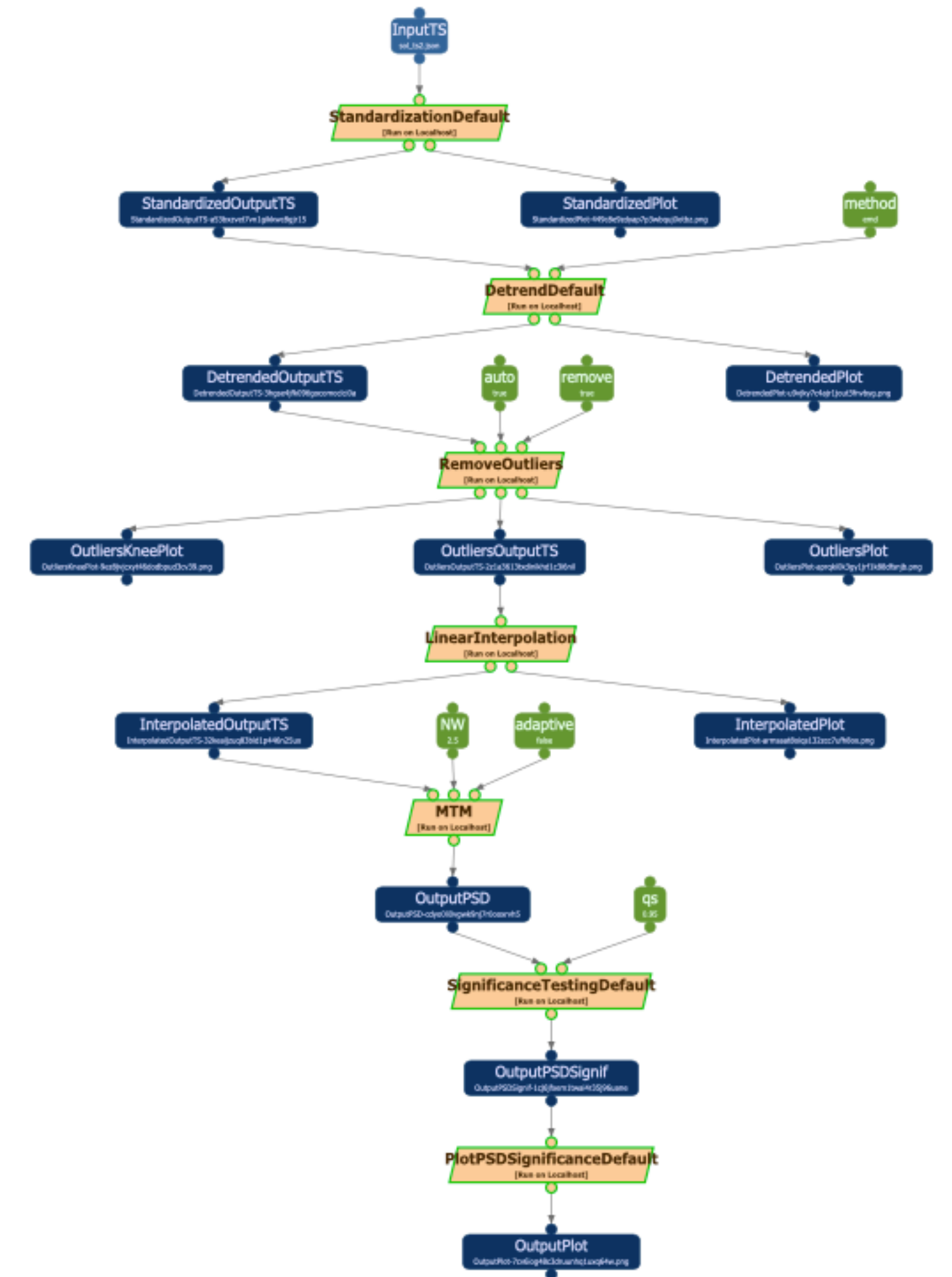
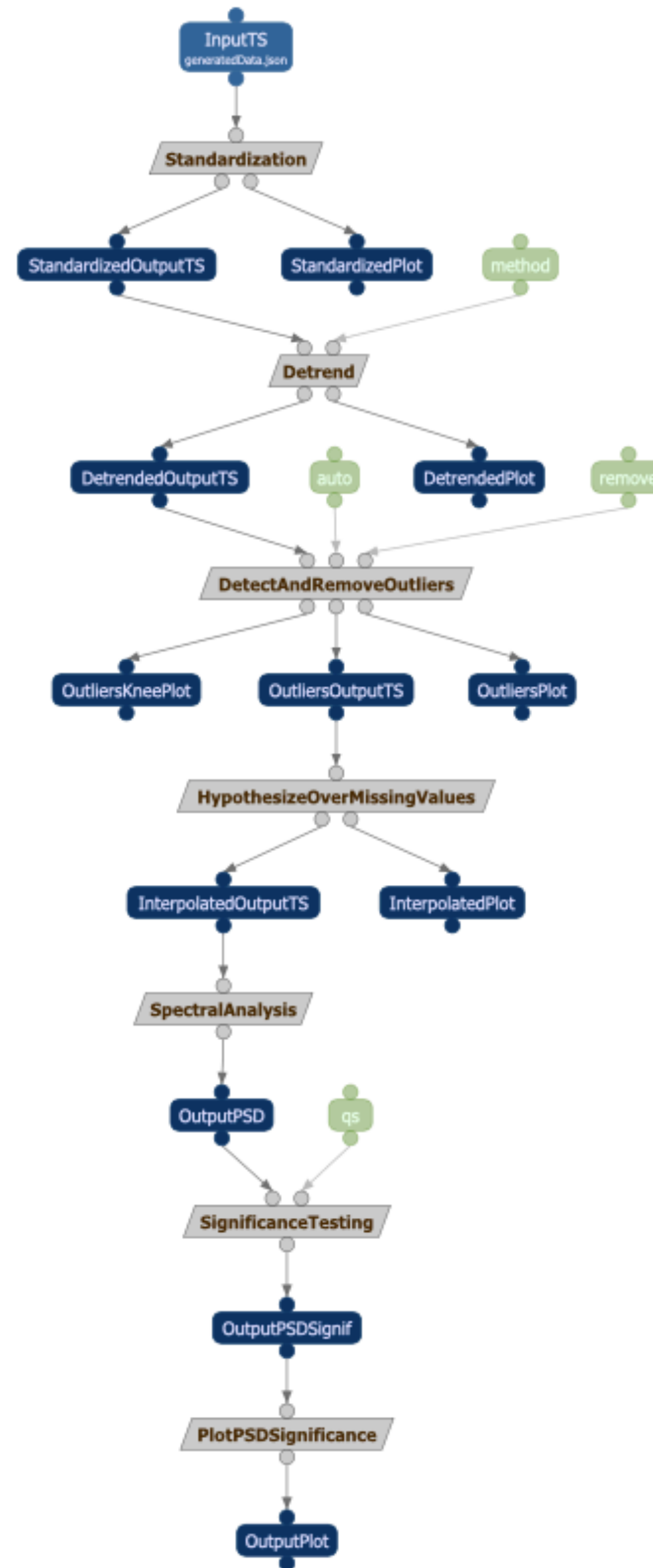
1 [ RequirementOfInputToBeEvenlySpaced:
2   (?c rdf:type acdom:MTMClass)
3   (?c ac:hasInput ?InputTS)
4   (?InputTS ac:hasArgumentID 'InputTS')
5   -> (?InputTS dcdom:isEvenlySpaced 'true'^^xsd:boolean)
6   print(?c 'MTM requires evenly spaced data')
7 ]
  
```

Add Rule ▾

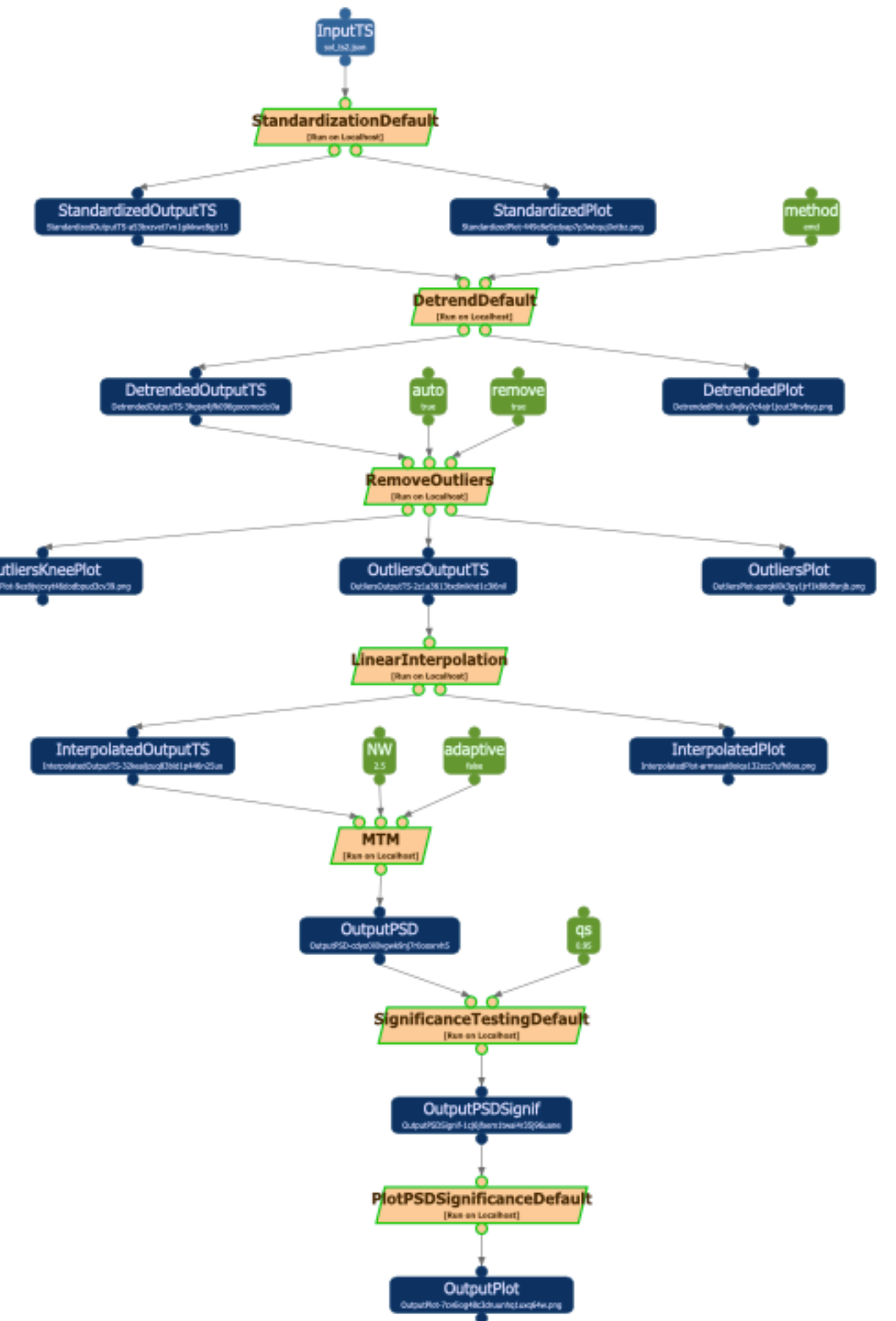
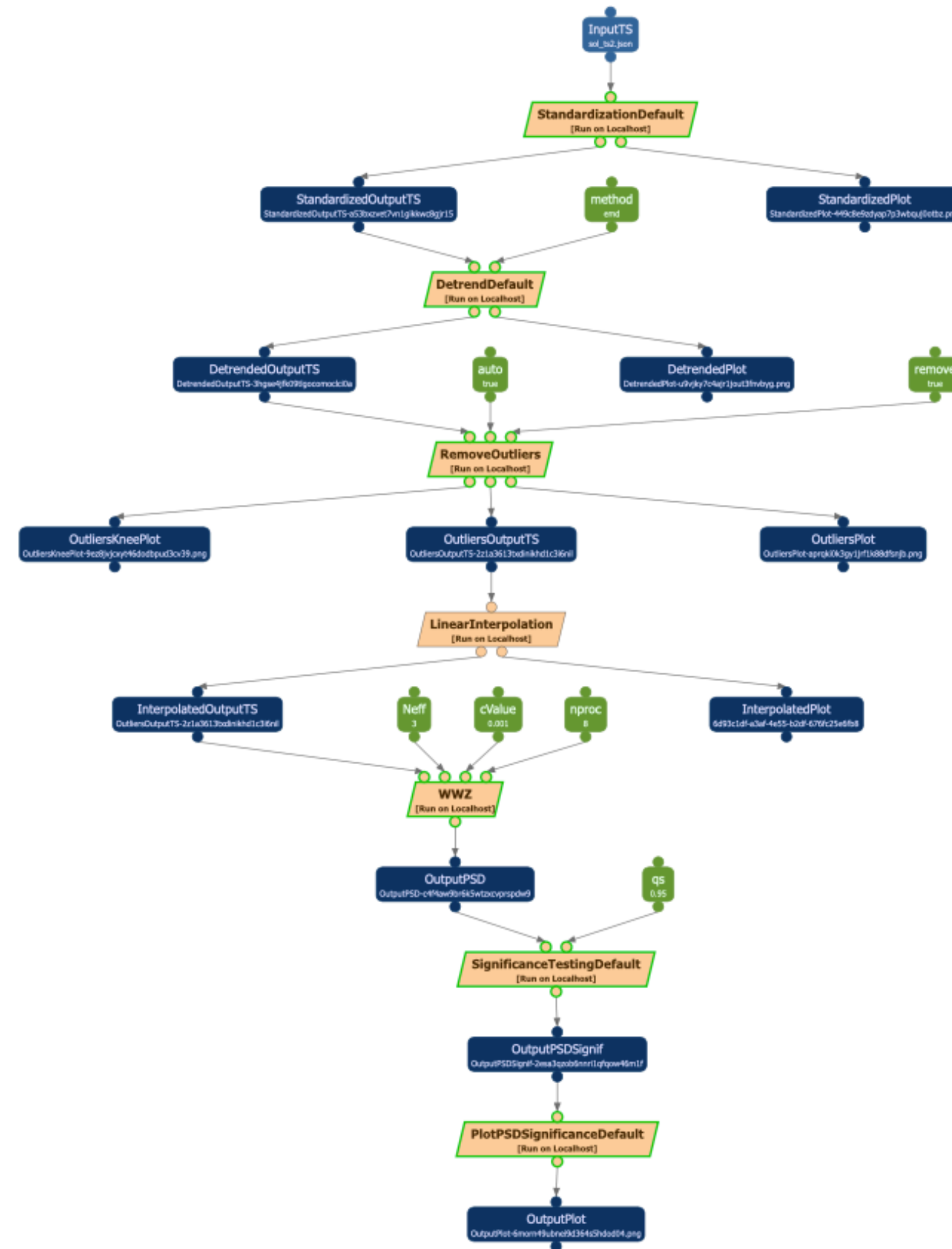
```

1 [ NoRequirementOfInputToBeEvenlySpaced:
2   (?c rdf:type acdom:WWZClass)
3   (?c ac:hasInput ?InputTS)
4   (?InputTS ac:hasArgumentID 'InputTS')
5   -> (?InputTS dcdom:isEvenlySpaced 'false'^^xsd:boolean)
6   print(?c 'WWZ does not require evenly spaced data')
7 ]
8
  
```

WINGS was extended to allow for a type of constraints that skips an abstract workflow step



WINGS was extended to allow for a type of constraints that skips an abstract workflow step



Welcome to the Wings Portal

This portal provides access to the **Wings semantic workflow system** by allowing you to browse, set up, and run workflows. Some unique features of Wings that can assist a user include:

- Choosing datasets that are valid for a workflow you have selected, based on the metadata properties of the datasets and the semantic constraints of that workflow
- Setting parameter values that are appropriate for the datasets you have selected, based on the semantic constraints of the workflow steps
- Handling collections of data in a compact way, based on compact collection reasoning in the workflow templates
- Specializing abstract workflow steps, based on the constraints imposed by both datasets and other workflow steps

You can walk through some [Wings Tutorials and Videos](#) that guide you through the basics of creating and using workflows.

About Wings

Workflow systems can automate many aspects of workflow creation and execution. The WINGS workflow system developed at USC/ISI illustrates these capabilities, with WINGS providing user assistance and automatic workflow validation and generation. The WINGS portal is setup by default to run locally, but it can use different execution engines such as [OODT](#) and [Pegasus](#)

WINGS (Workflow Instance Generation and Specialization) uses **AI planning and semantic reasoners** to assist users in creating workflows while validating that the workflows comply with the requirements of the software components and datasets. WINGS can reason about the constraints of the components and the characteristics of the data and propagate them through the workflow structure. WINGS reasons over *semantic workflow representations* that consist of both a traditional dataflow graph as well as a network of constraints on the data and components of the workflow.

Wings workflow repositories are available through other portal sites installed for use of specific research groups. These collections include workflows for population genomics, for educational student assessment, and for social network analysis. Each of these sites contains datasets and the executable codes for all workflow steps in their collection, so the workflows can be submitted for execution at the location where that particular Wings portal installation is set up.

The WINGS workflow system has an open modular design and can be easily integrated with other existing workflow systems and execution frameworks to extend them with semantic reasoning capabilities. We have integrated the Wings semantic workflow system with other user interfaces, and submitted workflows with a variety of execution engines. WINGS is built on open web standards from the World Wide Web Consortium (W3C) such as the Web Ontology Language (OWL), the Resource Description Framework (RDF), and the SPARQL query language for RDF.

If you are interested in using the Wings semantic workflow reasoners with your own workflow framework, or in setting up your own installation of a Wings workflow portal, please [contact us](#).

Publications providing overviews and technical descriptions of how Wings works are available in the [Wings web site](#).

Welcome to the Wings Portal

This portal provides access to the **Wings semantic workflow system** by allowing you to browse, set up, and run workflows. Some unique features of Wings that can assist a user include:

- Choosing datasets that are valid for a workflow you have selected, based on the metadata properties of the datasets and the semantic constraints of that workflow
- Setting parameter values that are appropriate for the datasets you have selected, based on the semantic constraints of the workflow steps
- Handling collections of data in a compact way, based on compact collection reasoning in the workflow templates
- Specializing abstract workflow steps, based on the constraints imposed by both datasets and other workflow steps

You can walk through some [Wings Tutorials and Videos](#) that guide you through the basics of creating and using workflows.

About Wings

Workflow systems can automate many aspects of workflow creation and execution. The WINGS workflow system developed at USC/ISI illustrates these capabilities, with WINGS providing user assistance and automatic workflow validation and generation. The WINGS portal is setup by default to run locally, but it can use different execution engines such as [OODT](#) and [Pegasus](#)

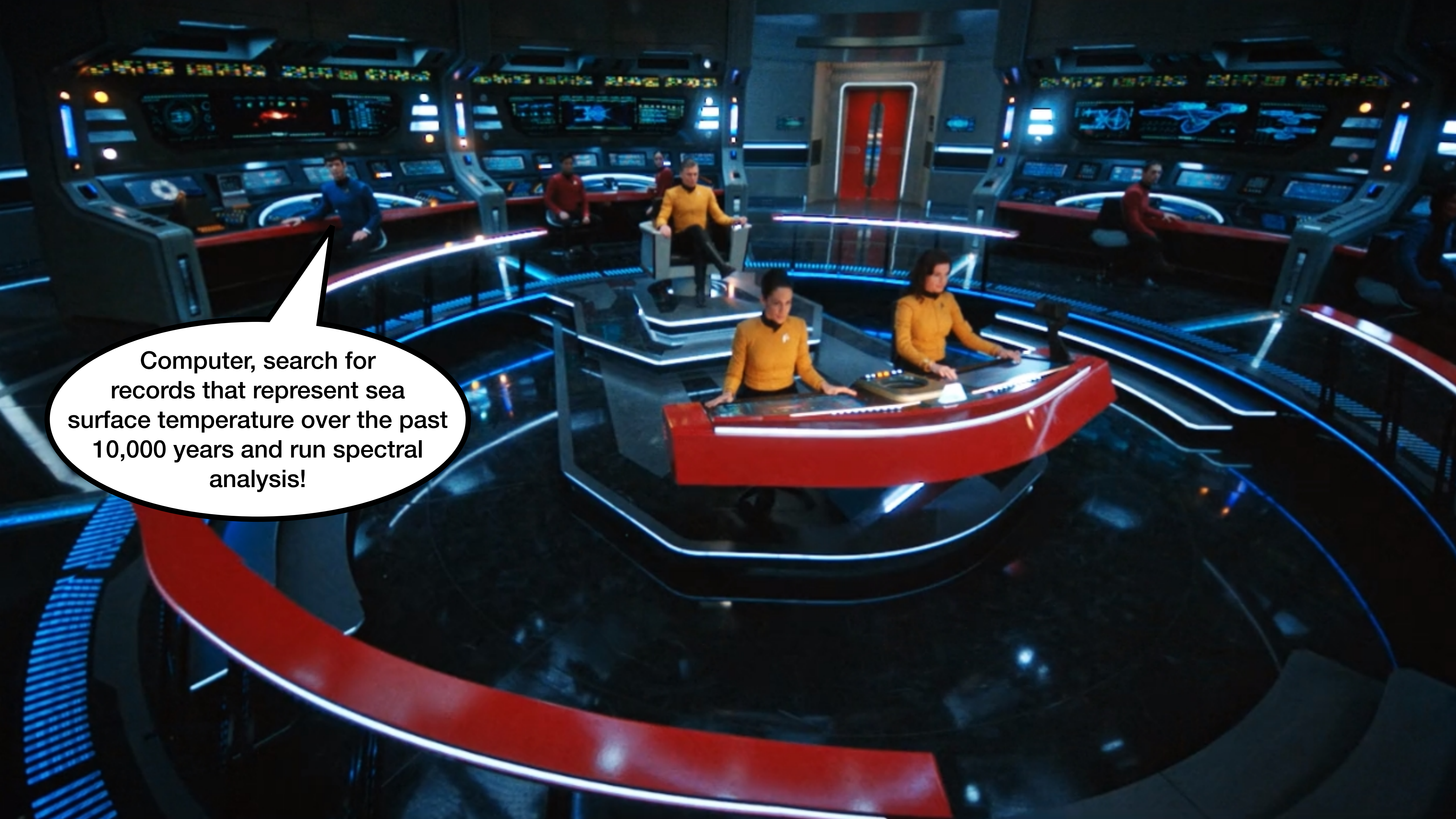
WINGS (Workflow Instance Generation and Specialization) uses AI planning and semantic reasoners to assist users in creating workflows while validating that the workflows comply with the requirements of the software components and datasets. WINGS can reason about the constraints of the components and the characteristics of the data and propagate them through the workflow structure. WINGS reasons over *semantic workflow representations* that consist of both a traditional dataflow graph as well as a network of constraints on the data and components of the workflow.

Wings workflow repositories are available through other portal sites installed for use of specific research groups. These collections include workflows for population genomics, for educational student assessment, and for social network analysis. Each of these sites contains datasets and the executable codes for all workflow steps in their collection, so the workflows can be submitted for execution at the location where that particular Wings portal installation is set up.

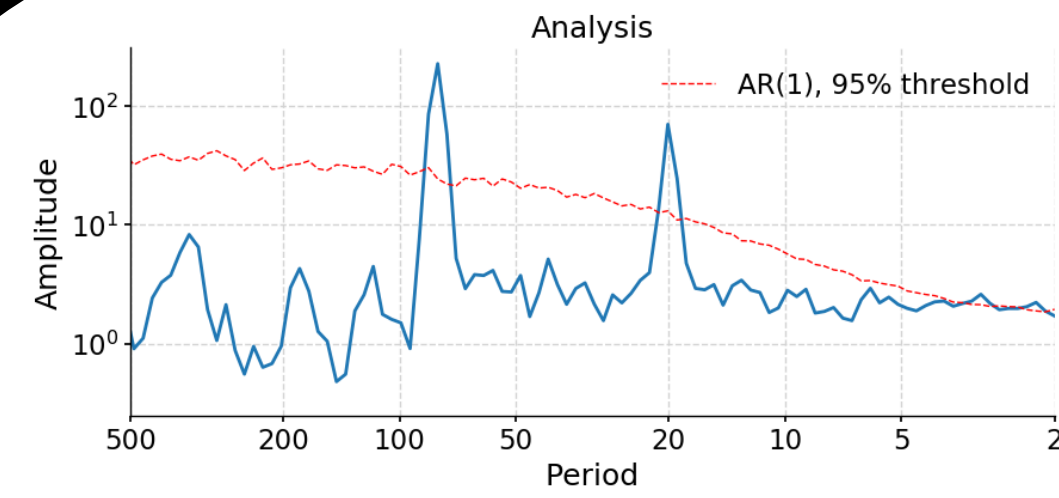
The WINGS workflow system has an open modular design and can be easily integrated with other existing workflow systems and execution frameworks to extend them with semantic reasoning capabilities. We have integrated the Wings semantic workflow system with other user interfaces, and submitted workflows with a variety of execution engines. WINGS is built on open web standards from the World Wide Web Consortium (W3C) such as the Web Ontology Language (OWL), the Resource Description Framework (RDF), and the SPARQL query language for RDF.

If you are interested in using the Wings semantic workflow reasoners with your own workflow framework, or in setting up your own installation of a Wings workflow portal, please [contact us](#).

Publications providing overviews and technical descriptions of how Wings works are available in the [Wings web site](#).

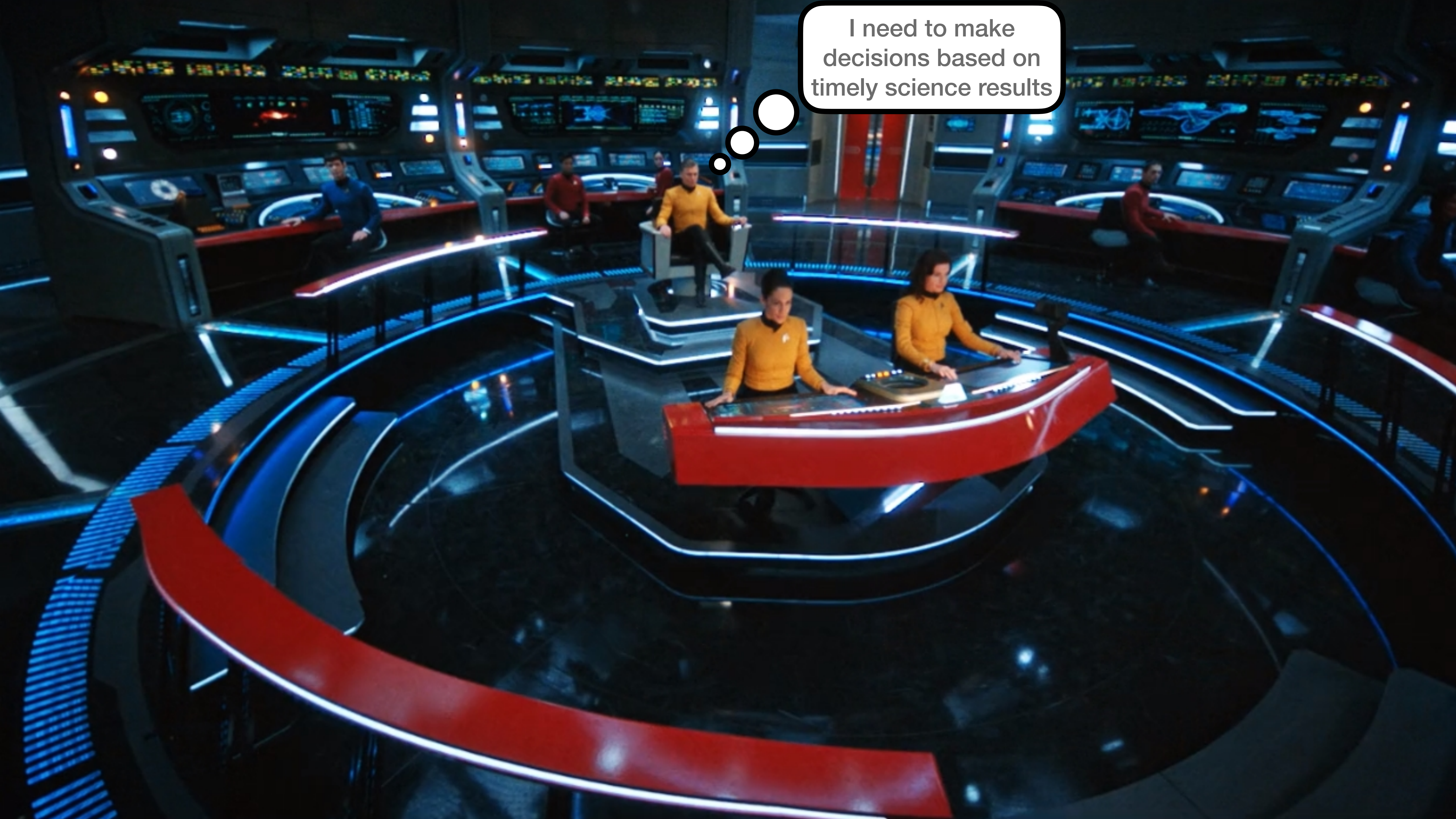
A wide-angle shot of the Star Trek: Voyager bridge. The room is filled with crew members in their respective uniforms (blue, red, and yellow) seated at various consoles. The bridge is illuminated with blue and red lights, and the background features a large red door. A white speech bubble with a black border is overlaid on the left side of the image, containing text.

Computer, search for records that represent sea surface temperature over the past 10,000 years and run spectral analysis!



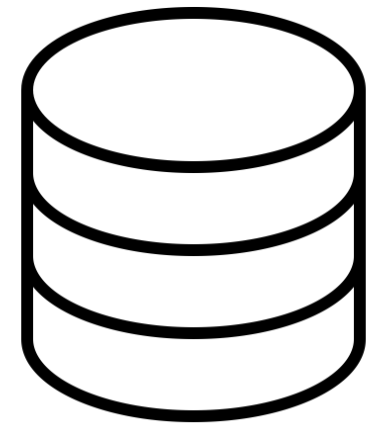
Computer, search for records that represent sea surface temperature over the past 10,000 years and run spectral analysis!

I need to make decisions based on timely science results



I need to make decisions based on **timely** science results

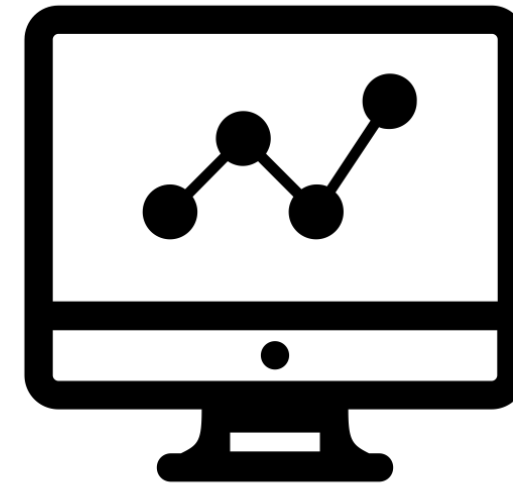




Query Data Catalog



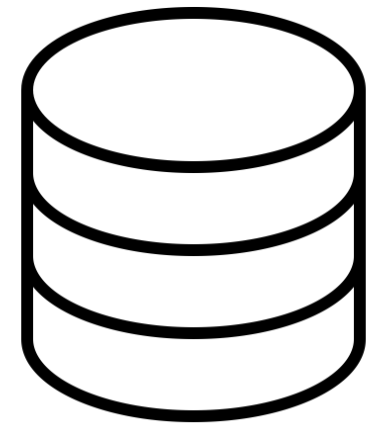
Download datasets



Spectral Analysis



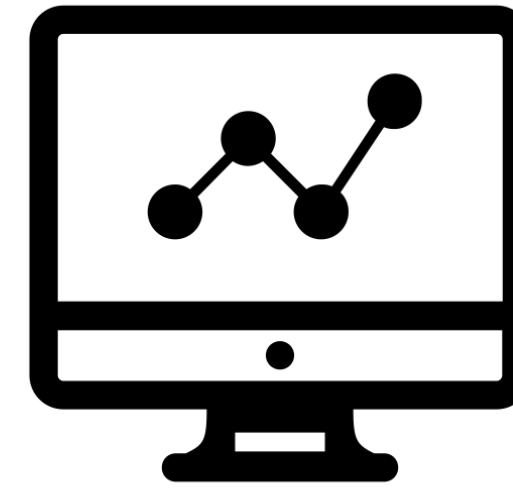
Visualize and Interpret Results



Query Data Catalog



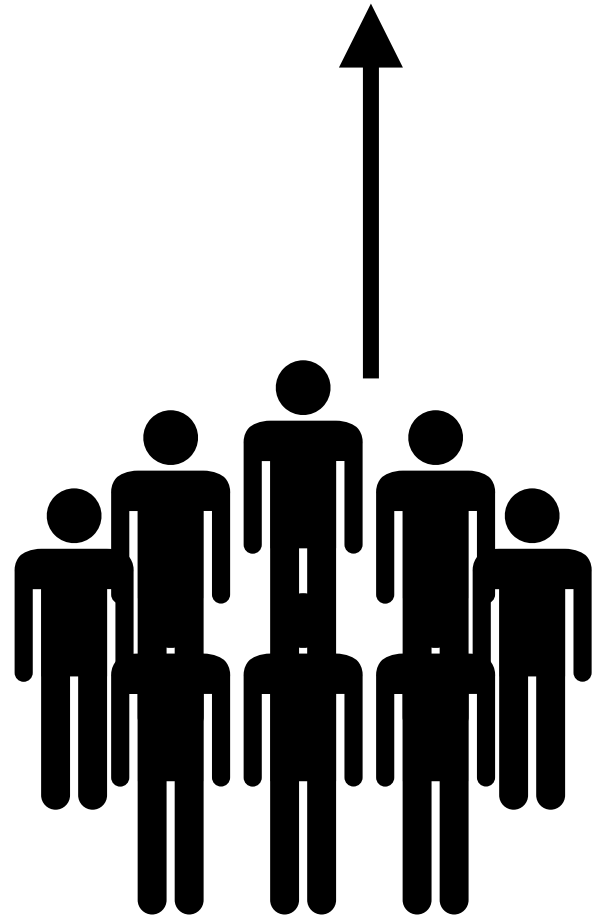
Download datasets

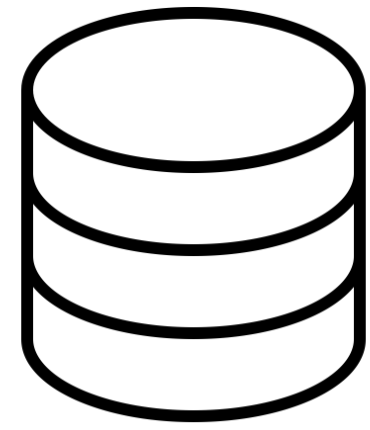


Spectral Analysis

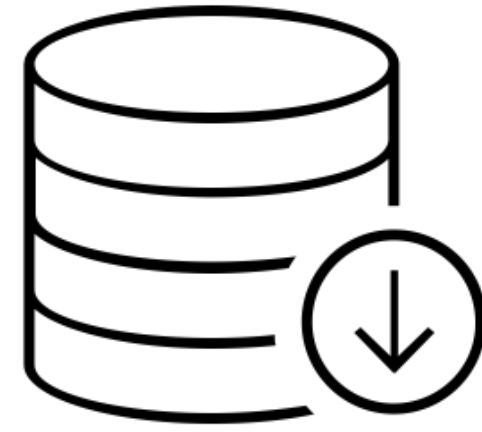


Visualize and Interpret Results

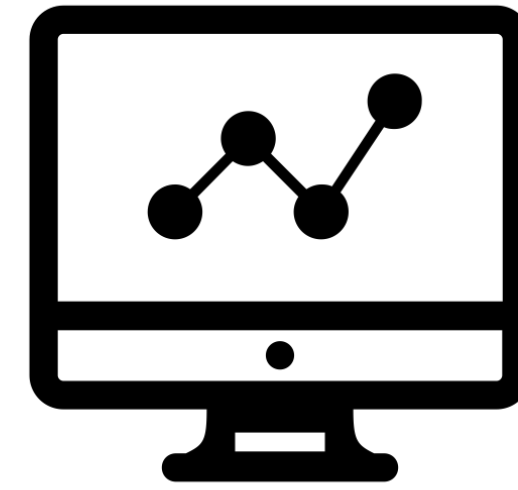




Query Data Catalog



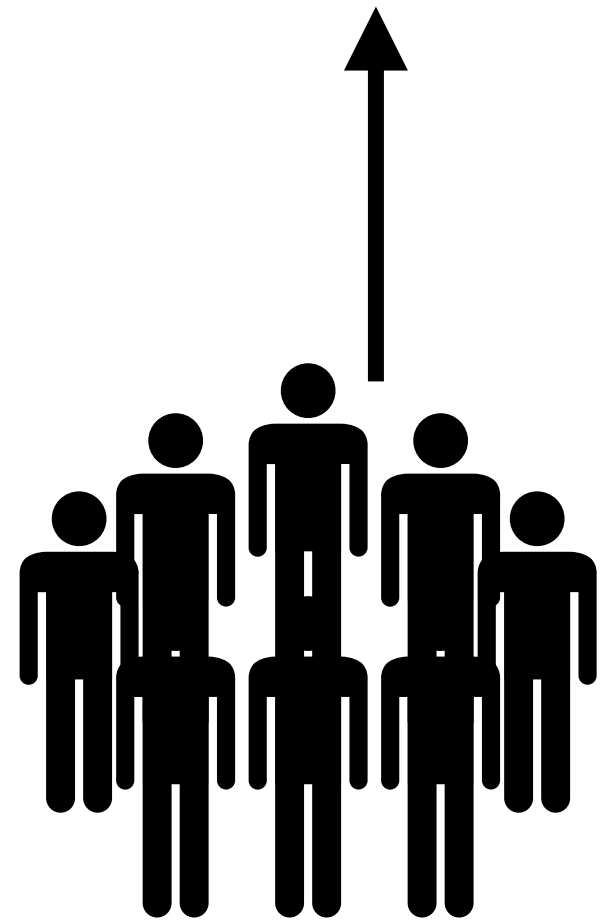
Download datasets



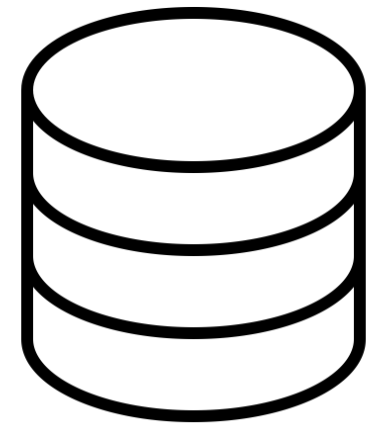
Spectral Analysis



Visualize and Interpret Results



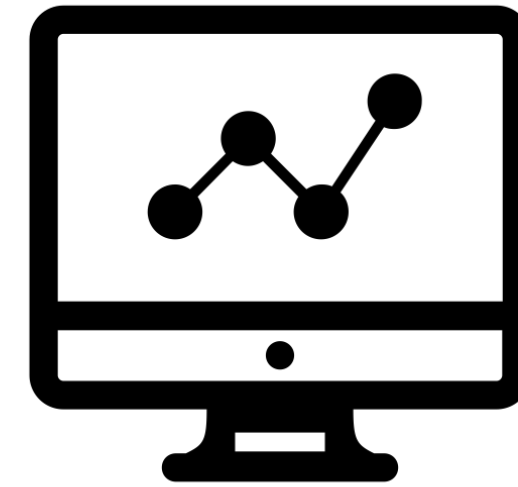
Can we leverage machine learning to help with metadata annotation?



Query Data Catalog



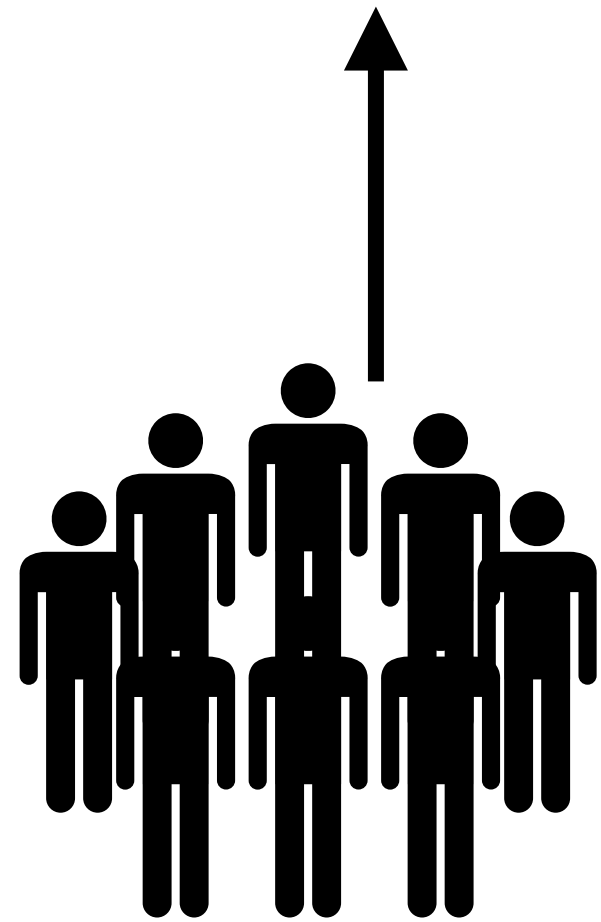
Download datasets



Spectral Analysis



Visualize and Interpret Results



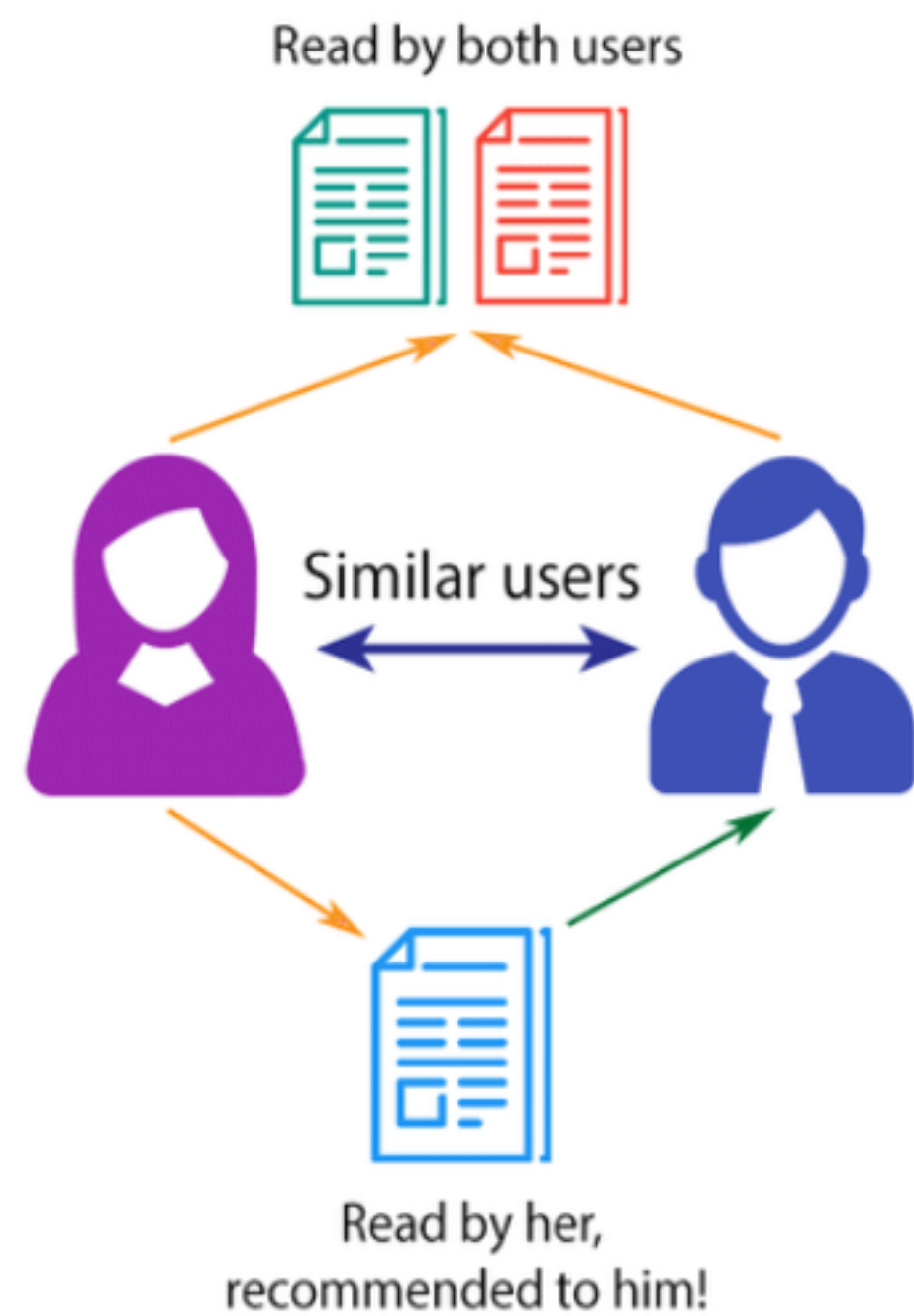
Can we leverage machine learning to help with metadata annotation?



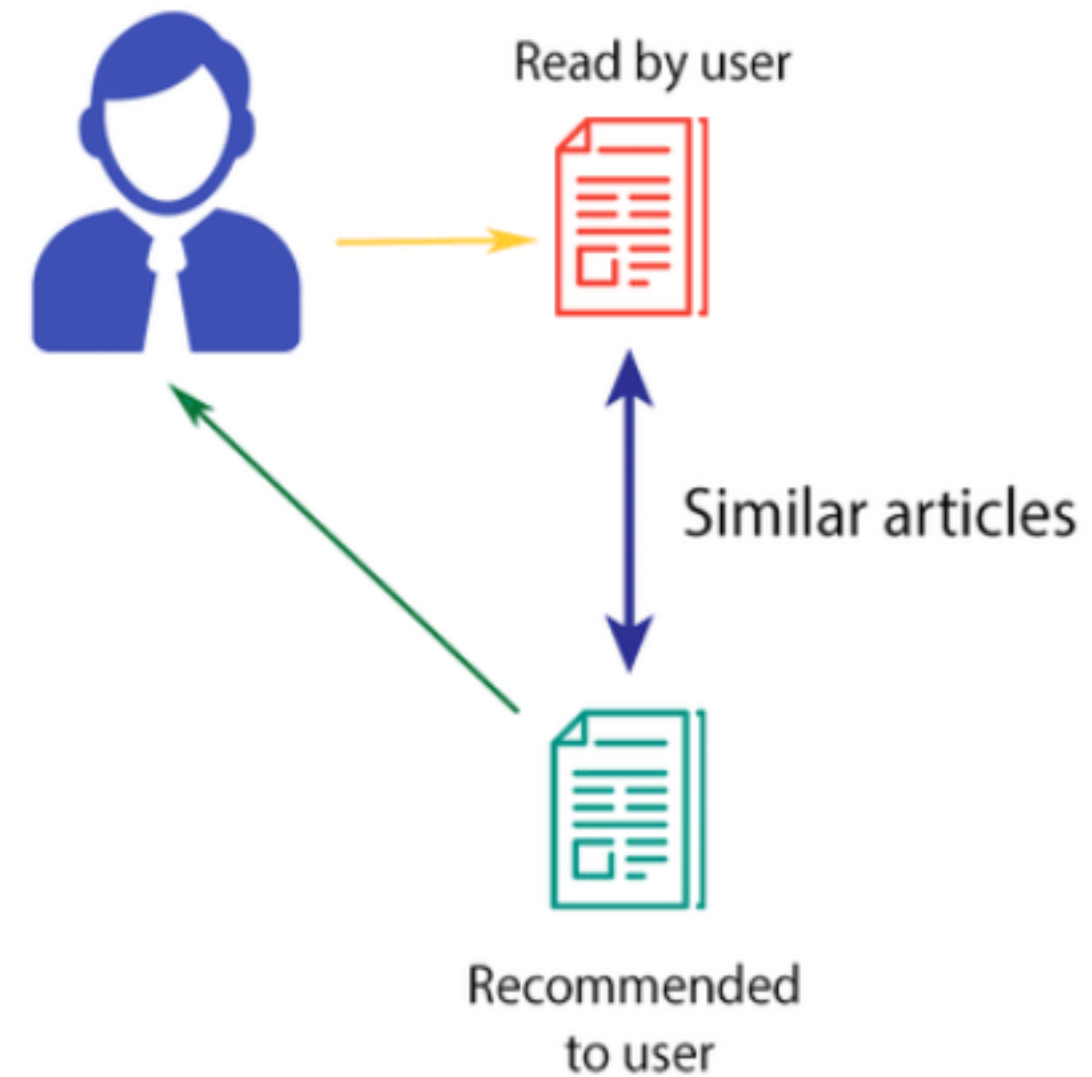
Shravya Manety

Types of recommender systems

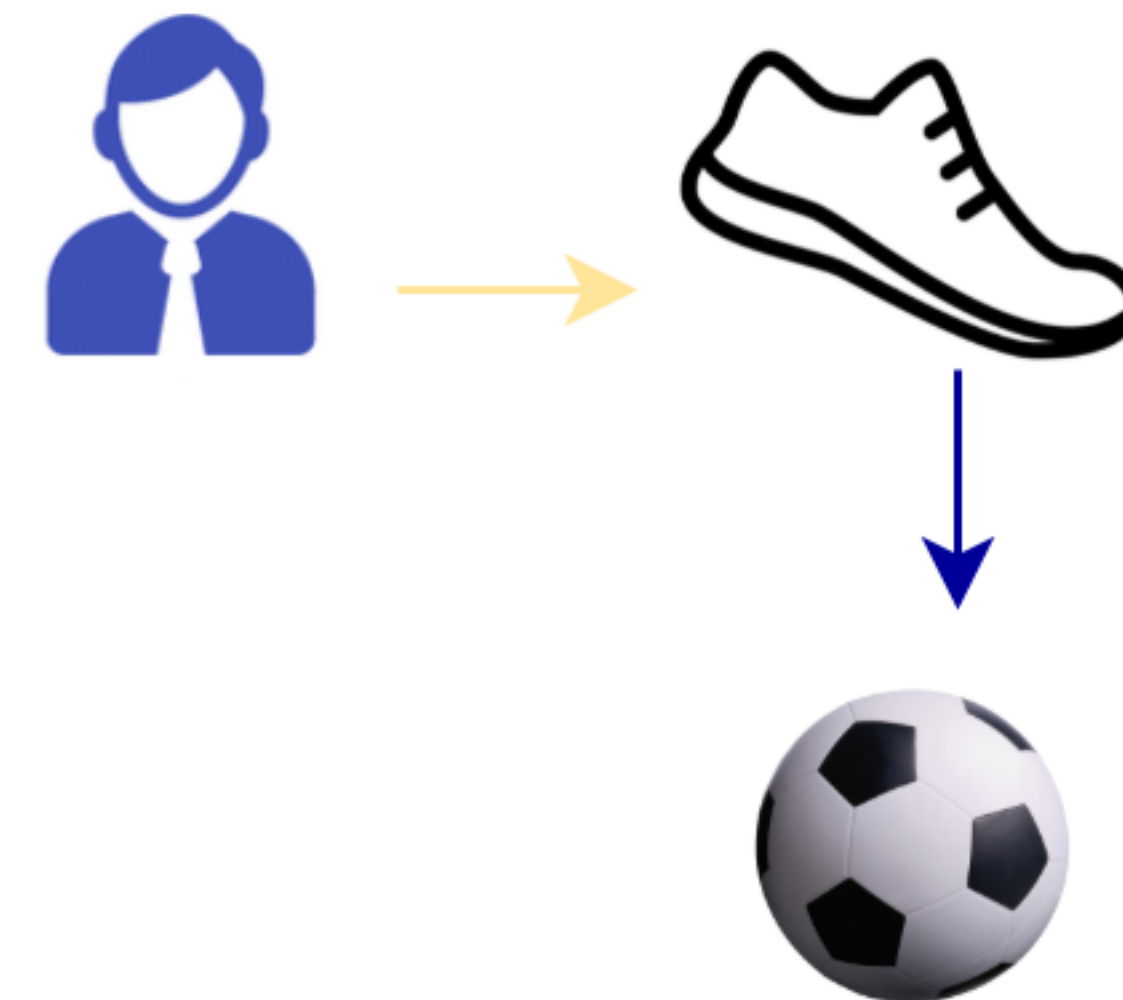
COLLABORATIVE FILTERING



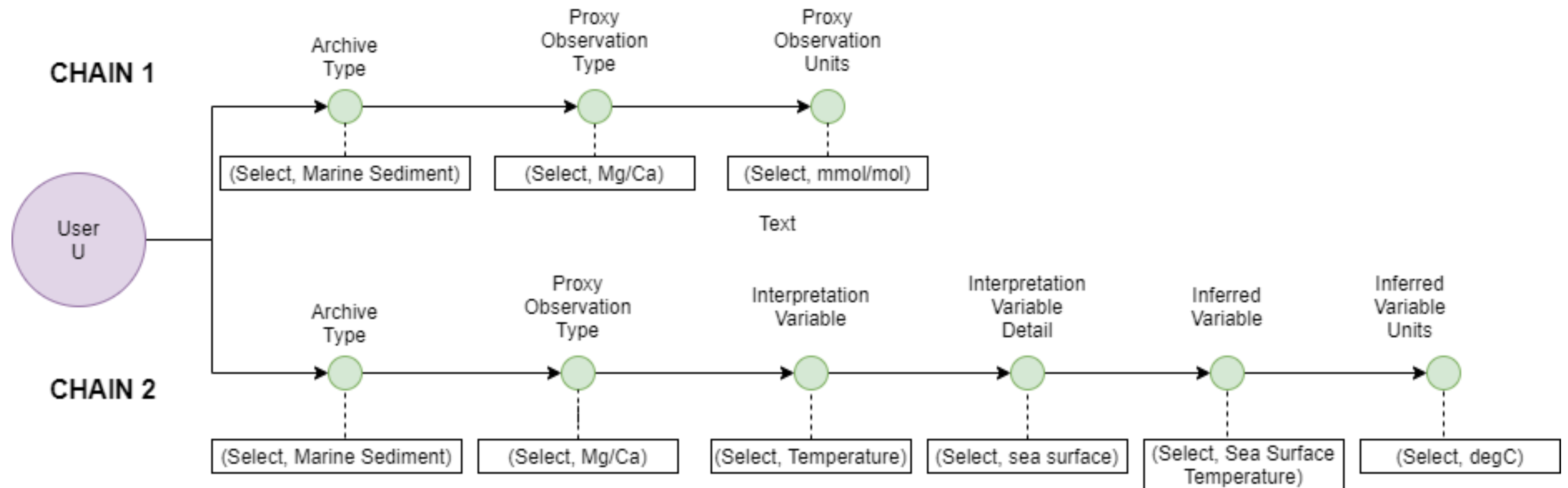
CONTENT-BASED FILTERING



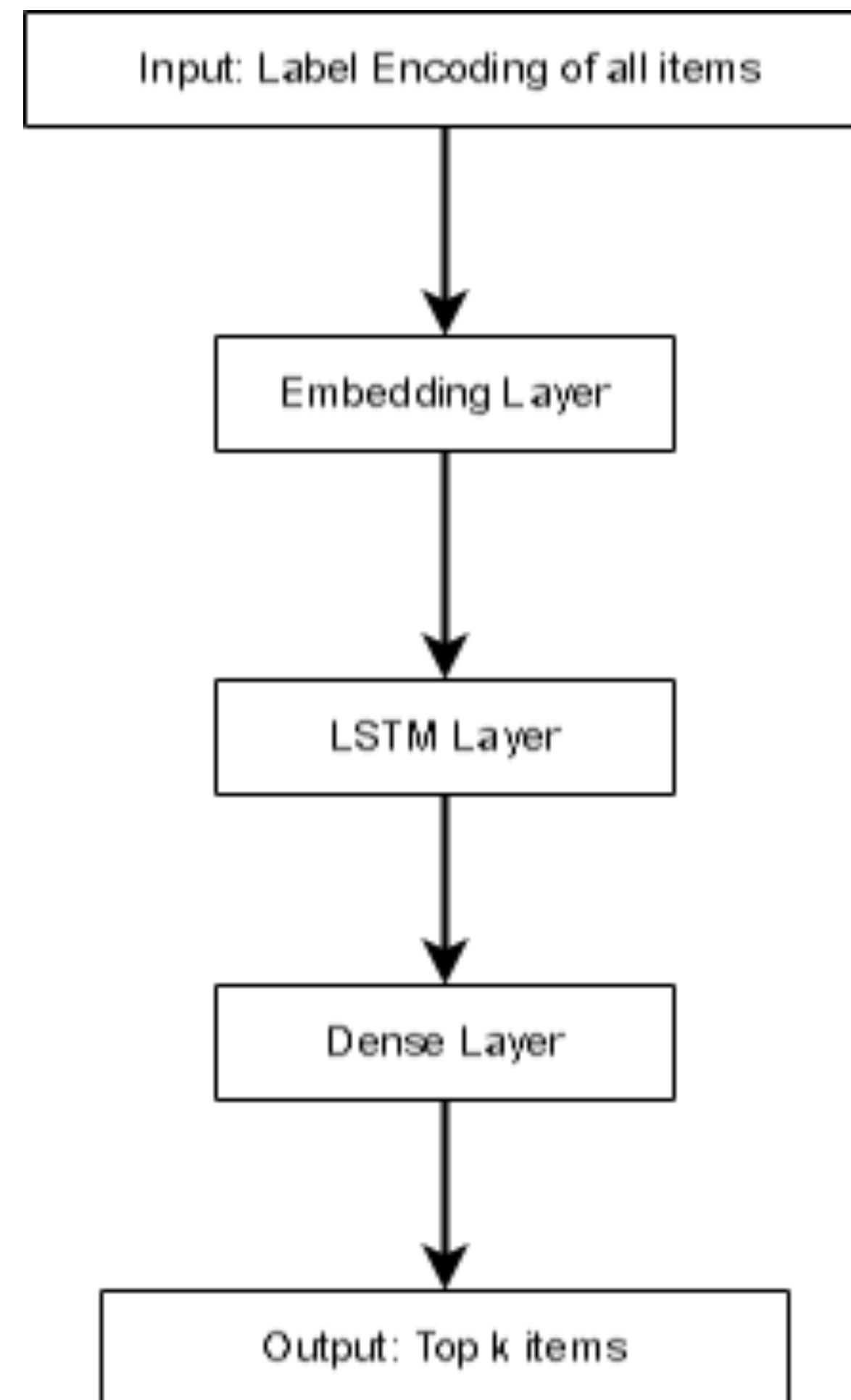
SEQUENTIAL RECOMMENDATION



PaleoRec: A transaction-based sequential recommender system

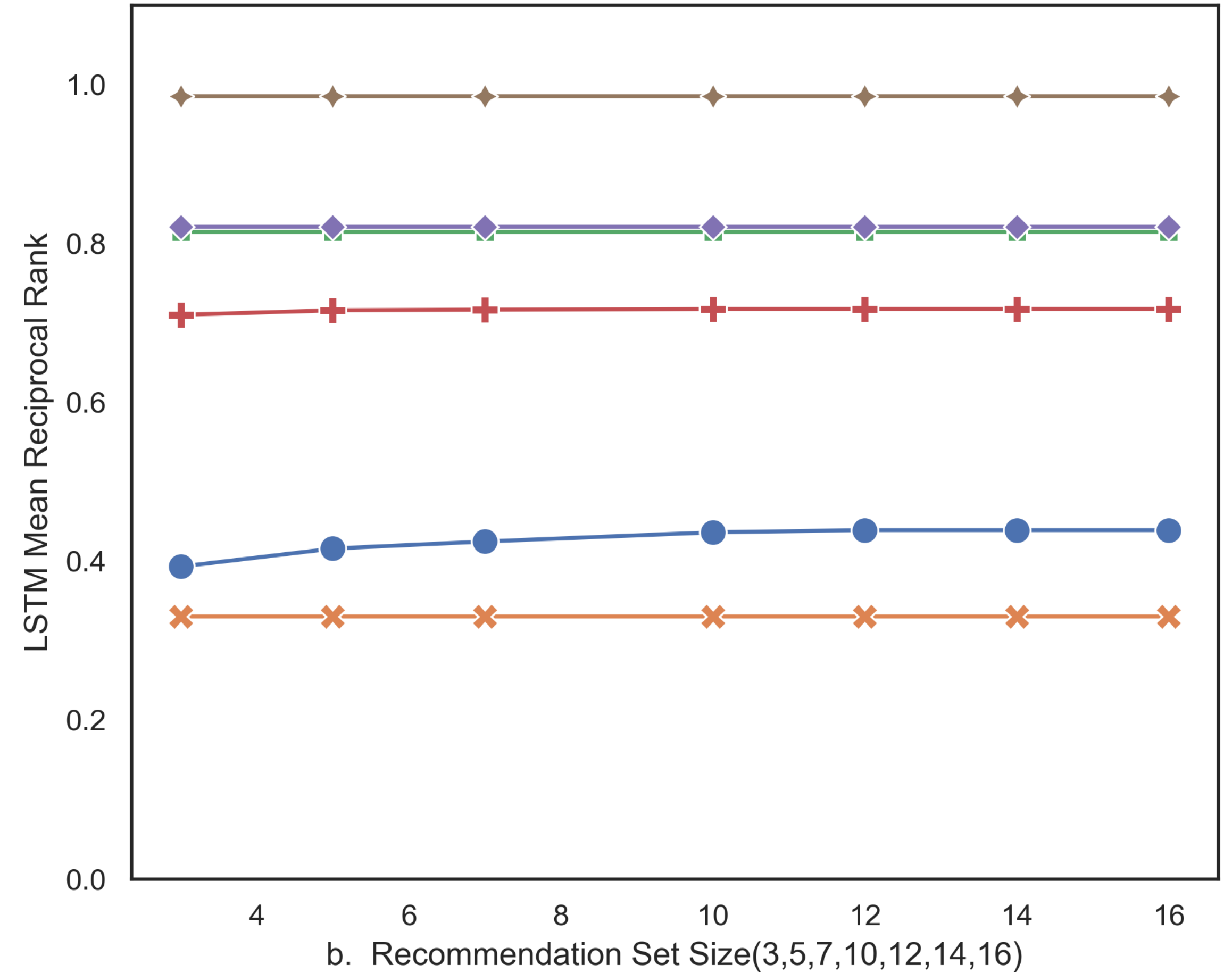
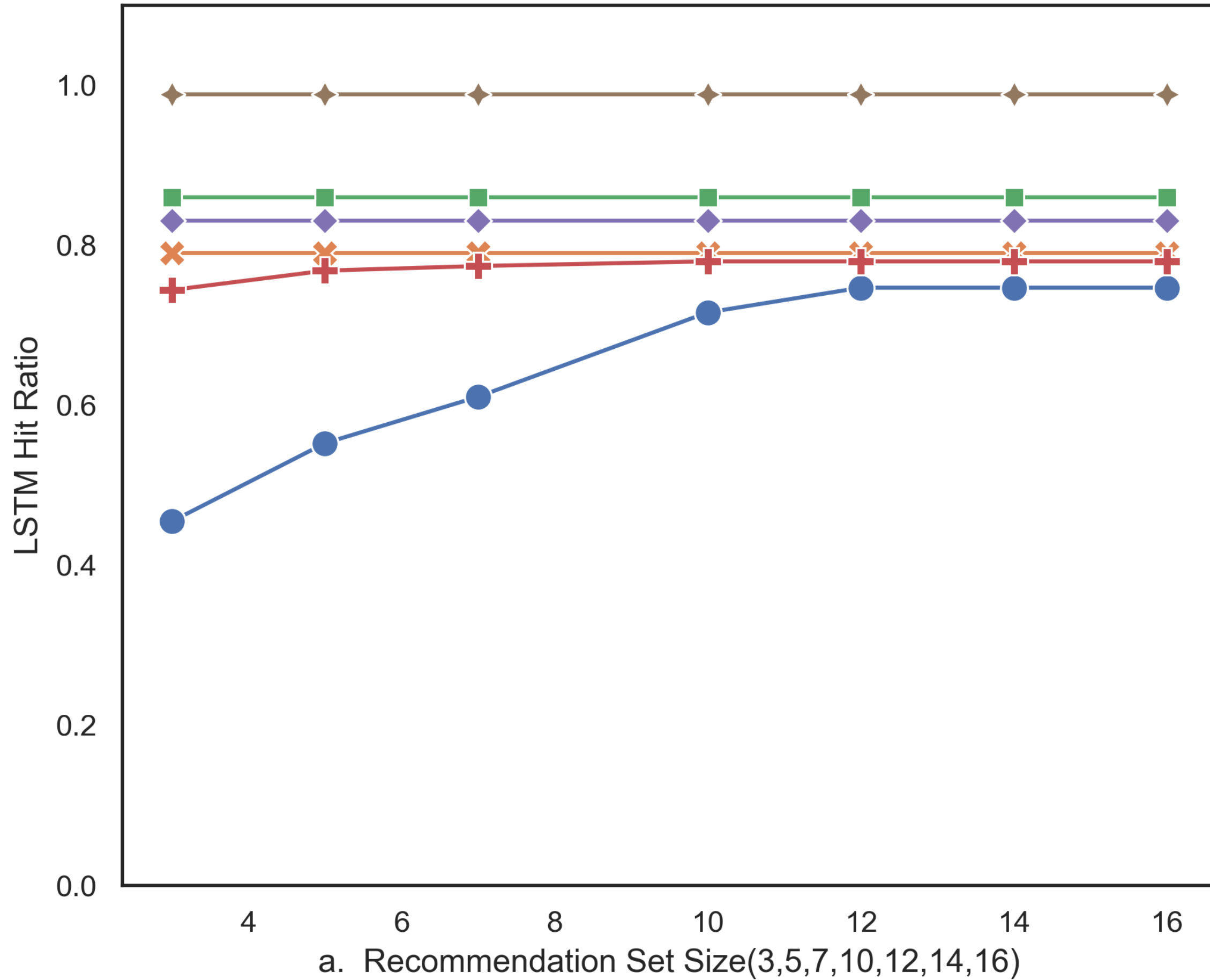


PaleoRec: A transaction-based sequential recommender system



	archiveType	proxyObservationType	units	interpretation/variable	interpretation/variableDetail	inferredVariable	inferredVarUnits
0	Wood	Rbar	NA	NA	NA	NA	NA
1	Wood	Trsgi	NA	Temperature	Air Surface	Air Surface Temperature	deg C
2	MarineSediment	TEX86	NotApplicable	Temperature	Sea Surface	Sea Surface Temperature	deg C
3	MarineSediment	D18O	permil	Temperature	Sea Surface	Sea Surface Temperature	deg C
4	LakeSediment	Sampleid	NotApplicable	NA	NA	NA	NA
5	Wood	Trsgi	NA	Temperature	Air Surface	Air Surface Temperature	deg C
6	MarineSediment	D18O	permil	D18Osw	Bottom Water	Temperature	deg C
7	Wood	Corrs	NA	NA	NA	NA	NA
8	Wood	Ringwidth	cm	NA	NA	NA	NA
9	MarineSediment	Depth	cm	NA	NA	NA	NA

PaleoRec: Metrics



Playground

MAP GRAPH



Are you new here? Take the tour to learn more about how to use this page.

Playground

MAP GRAPH



Are you new here? Take the tour to learn more about how to use this page.

I need to make decisions based on **timely** science results



I need to make
decisions based on
timely science results



Forecasting with paleoclimate data

- From the data
 - Sparse (extremely!)
 - At different resolution
 - Not the variables we are interested in

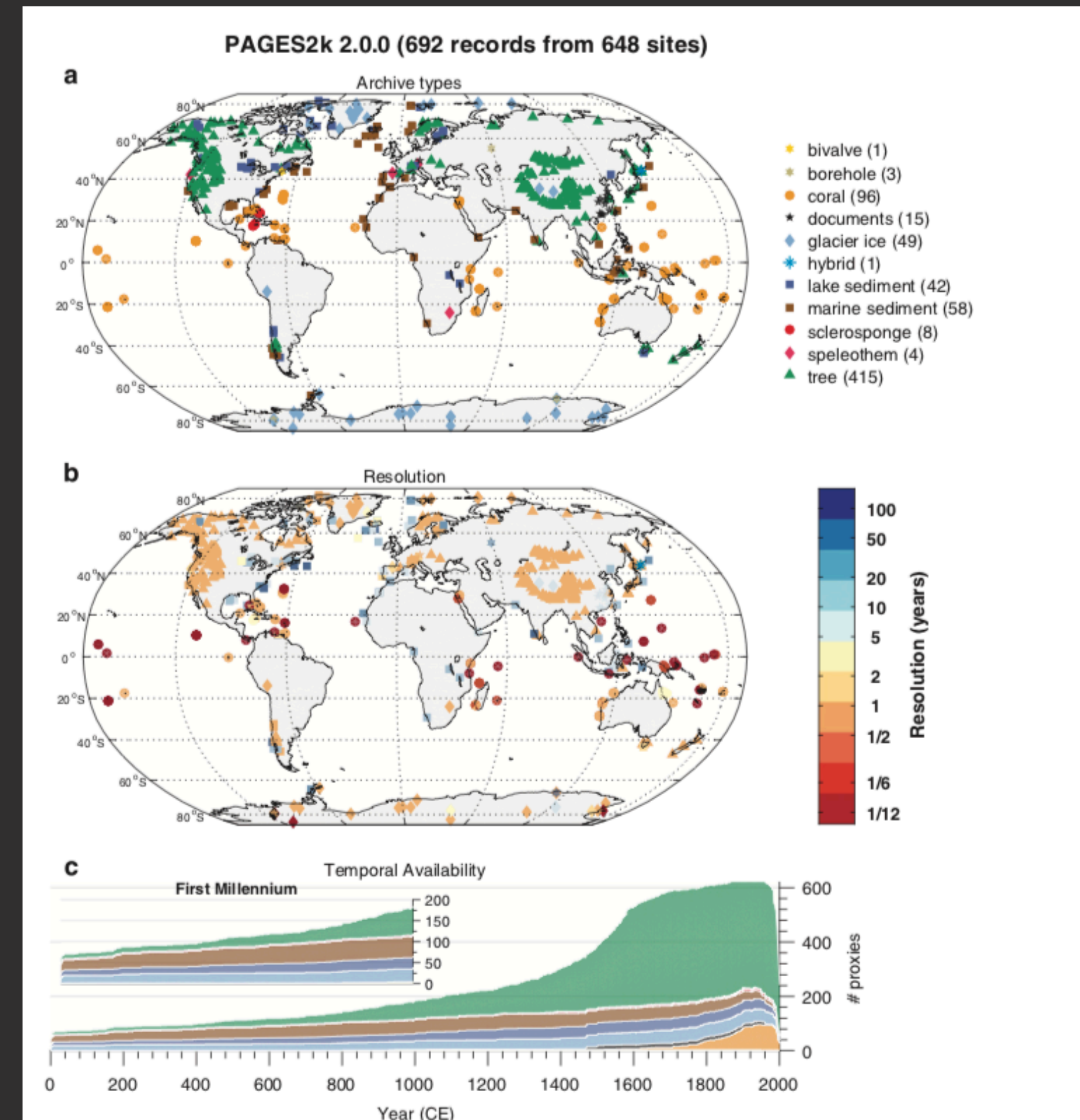
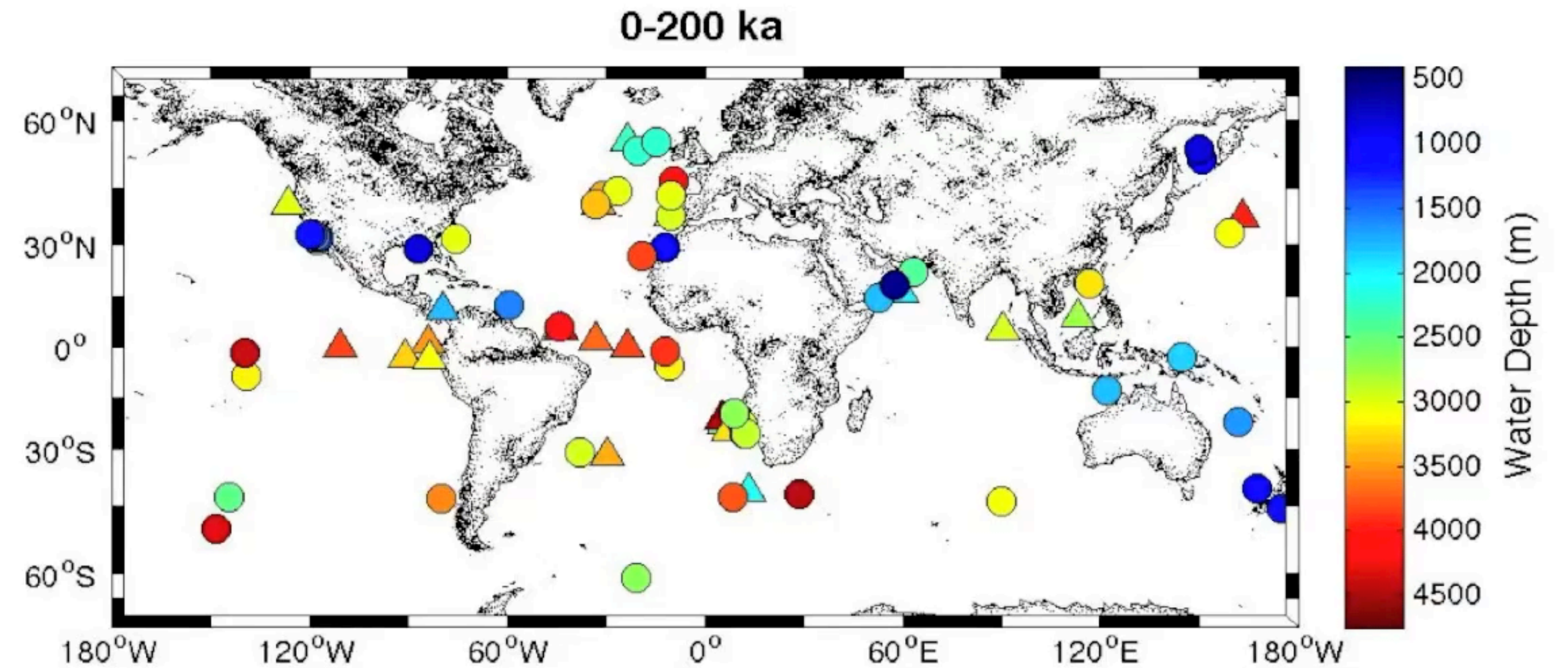


Figure 1. Spatiotemporal data availability in the PAGES2k database. (a) Geographical distribution, by archive type, coded by color and shape. (b) Temporal resolution in the PAGES2k database, defined here as the median of the spacing between consecutive observations. Shapes as in (a), colors encode the resolution in years (see colorbar). (c) Temporal availability, coded by color as in (a).

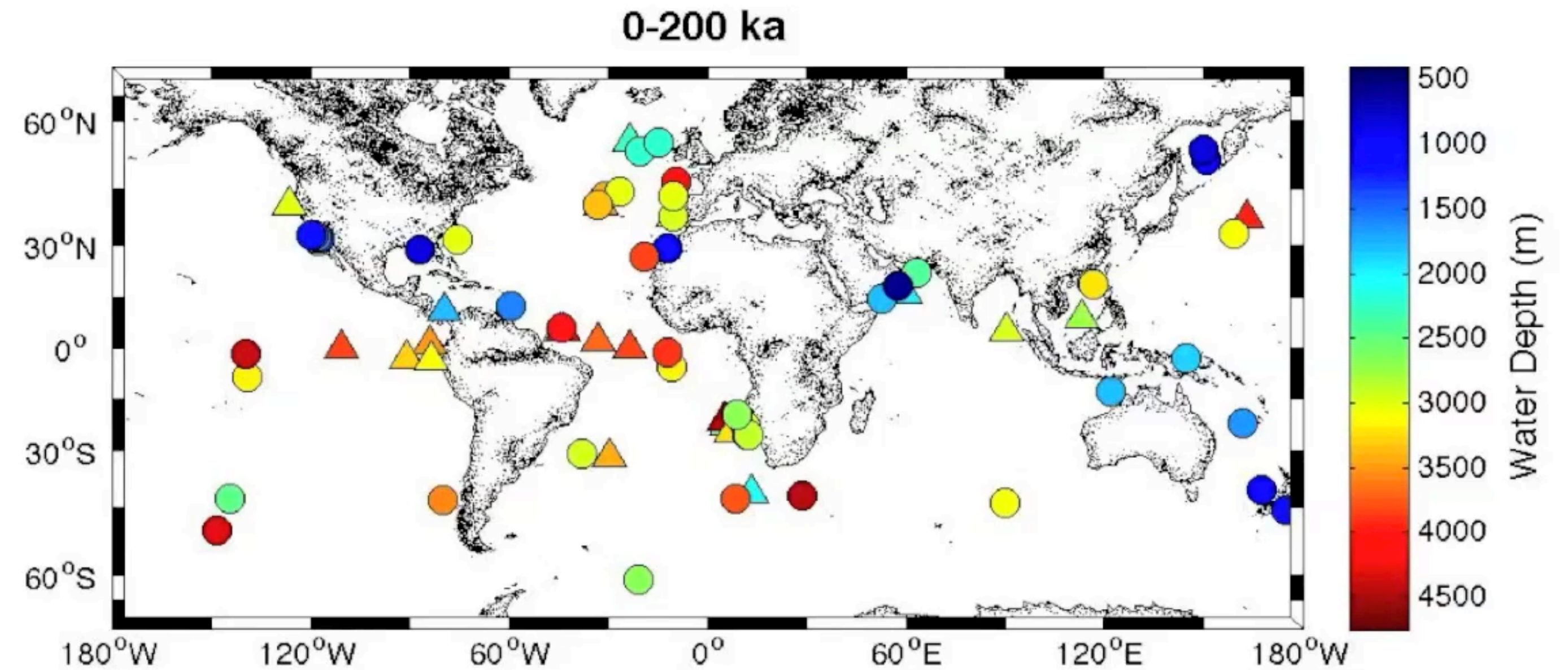
Forecasting with paleoclimate data

- From the data
 - Sparse (extremely!)
 - At different resolution
 - Not the variables we are interested in



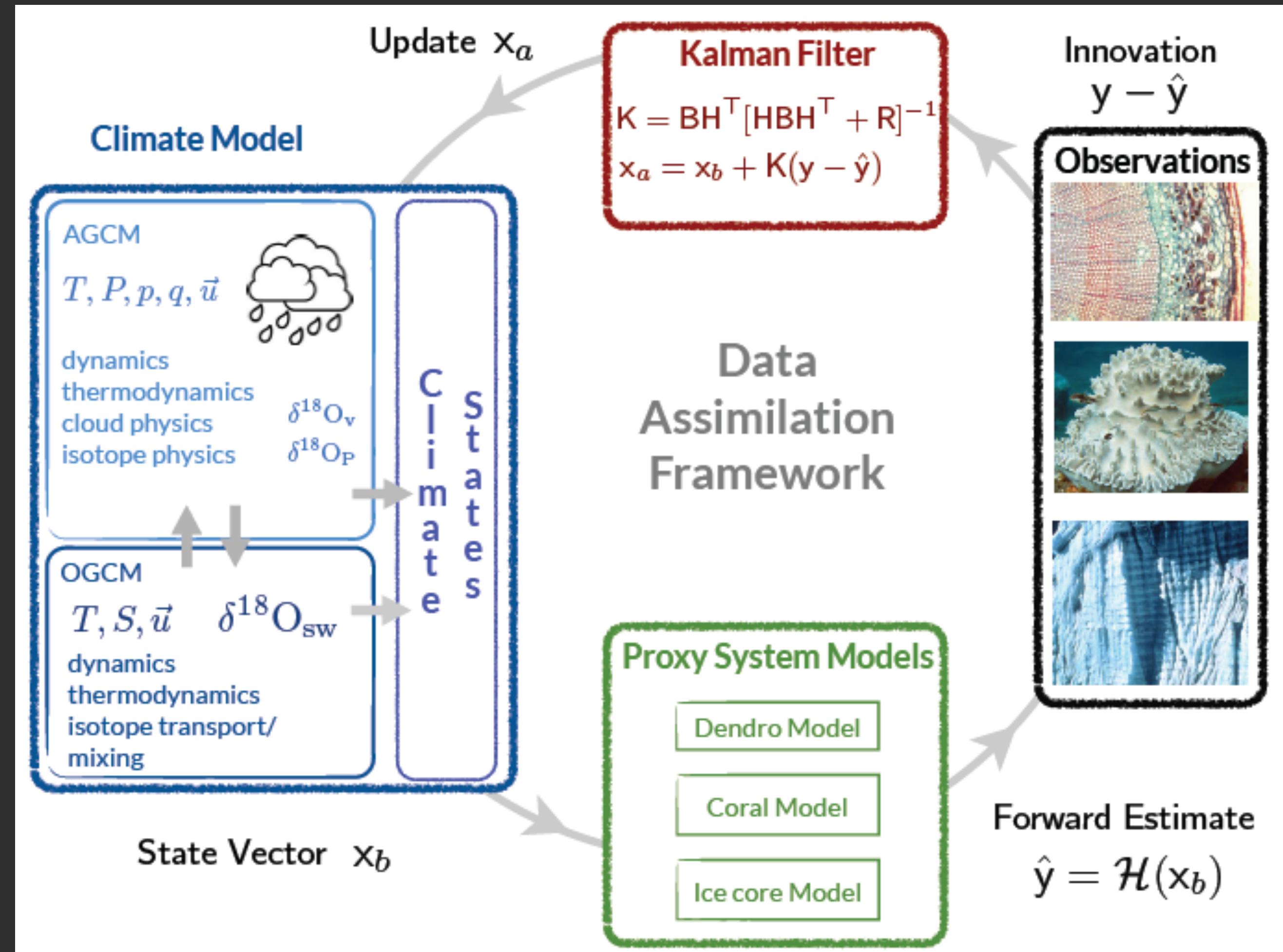
Forecasting with paleoclimate data

- From the data
 - Sparse (extremely!)
 - At different resolution
 - Not the variables we are interested in



Forecasting with paleoclimate data

- From reanalysis:
 - Model Bias
 - PSM representation is simplistic





I need to communicate the decision to stakeholders



I need to communicate the decision to stakeholders

Characterizing the counter-narratives of climate change



Abhilash Pandurangan
USC



Aditya Jajodia
USC



Sushmitha Ravikumar
USC



Vanshika Sridharan
USC



Fred Morstatter
USC



Deborah Khider
USC



**Climate scientists post their findings and views
regularly on social media**

**Climate scientists post their findings and views
regularly on social media**

These views are often met with harassment...

**Climate scientists post their findings and views
regularly on social media**

These views are often met with harassment...

... which is not picked up by already trained classifiers

Harassment of scientist on the Twitter platform

Harassment of scientist on the Twitter platform

- Harassment: attacking someone based on their race, gender, character...

Replying to @KHayhoe @CriticalStress_ and 3 others

You are not a climate scientist. You are a political propagandist pushing fraudulent, incomplete, deceptive data to bolster a false narrative. You've been exposed multiple times and you're only supported by fellow globalist criminals. Joseph Goebbels would be very proud of you.

Harassment of scientist on the Twitter platform

- Harassment: attacking someone based on their race, gender, character...
- **Laymansplaining**: prevalent form of harassment where user condescendingly explains to a scientist their research

Replying to @KHayhoe @CriticalStress_ and 3 others

You are not a climate scientist. You are a political propagandist pushing fraudulent, incomplete, deceptive data to bolster a false narrative. You've been exposed multiple times and you're only supported by fellow globalist criminals. Joseph Goebbels would be very proud of you.

Replying to @GeraldKutney

Actually, I just look at the facts about 'climate change':

- record sea ice extend
- no accelerated sea level rise anywhere
- polar bears are doing more than fine
- no relation between CO2 and global temperature
- etc

Conclusion: climate alarmism is a lost cause.

6:44 AM · Mar 10, 2020 · Twitter for iPhone

Harassment of scientist on the Twitter platform

- Harassment: attacking someone based on their race, gender
- **Laymansp** prevalent for harassment condescen to a scientist

Can we train a classifier to recognize this particular form of harassment/identify content with scientific misinformation?

Replying to @KHayhoe @CriticalStress_ and 3 others

You are not a climate scientist. You are a political propagandist pushing fraudulent, incomplete, deceptive data to bolster a false narrative. You've been exposed multiple times and you're only supported by

- etc

Conclusion: climate alarmism is a lost cause.

6:44 AM · Mar 10, 2020 · Twitter for iPhone

Harassment of scientist on the Twitter platform

- Harassment: attacking someone based on their race, gender, etc.
- **Laymansp** prevalent form of harassment: condescension to a scientist

Replying to @KHayhoe @CriticalStress_ and 3 others

You are not a climate scientist. You are a political propagandist pushing fraudulent, incomplete, deceptive data to bolster a false narrative. You've been exposed multiple times and you're only supported by

A HUMAN

Can we train a classifier to recognize this particular form of harassment/identify content with scientific misinformation?

- etc

Conclusion: climate alarmism is a lost cause.

6:44 AM · Mar 10, 2020 · Twitter for iPhone

Harassment of scientist on the Twitter platform

- Harassment: attacking someone based on their race, gender
- **Laymansp** prevalent for harassment condescen to a scientist

Can we train a classifier to recognize this particular form of harassment/identify content with scientific misinformation?

~60% accuracy

Replying to @KHayhoe @CriticalStress_ and 3 others

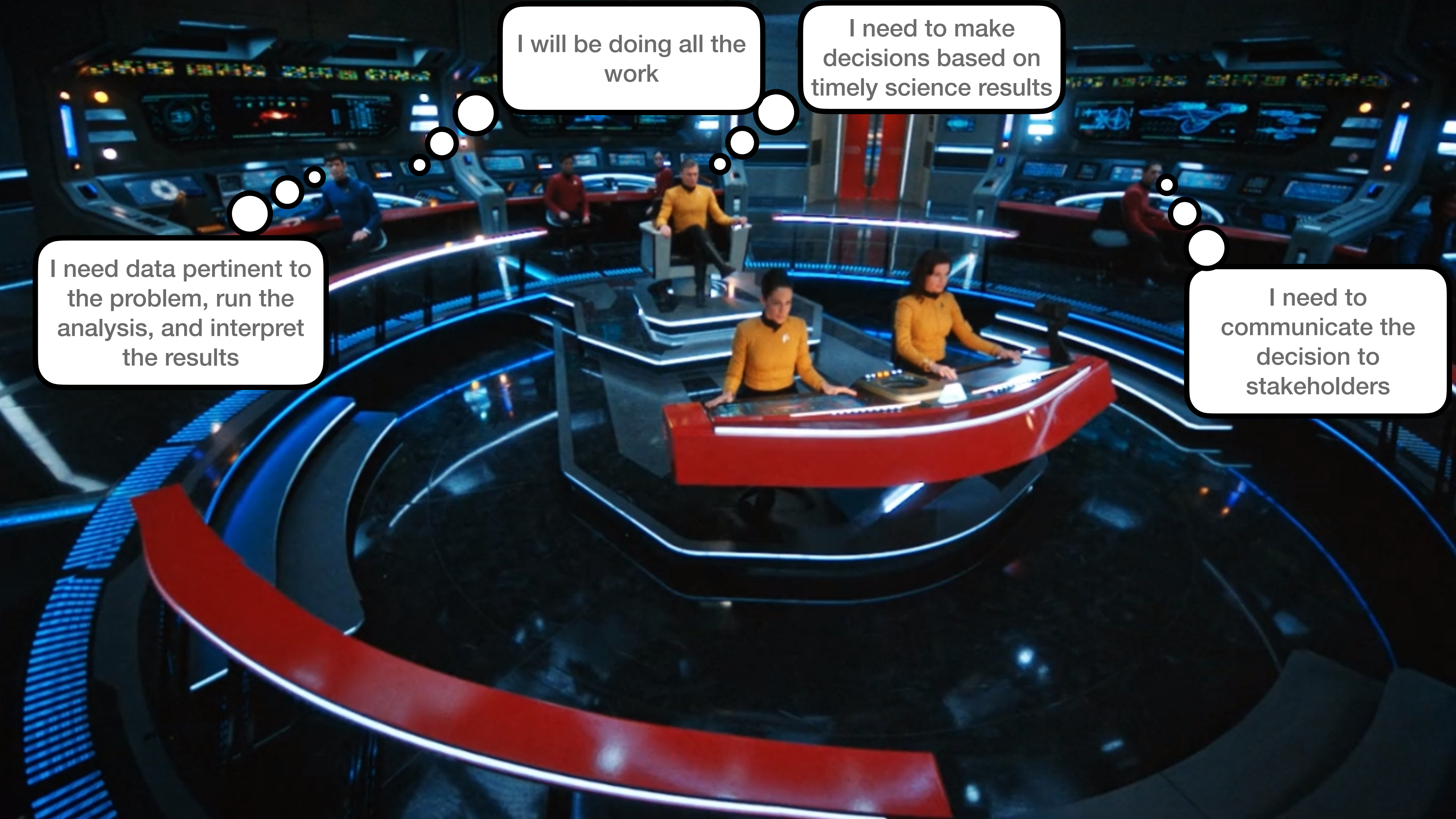
You are not a climate scientist. You are a political propagandist pushing fraudulent, incomplete, deceptive data to bolster a false narrative. You've been exposed multiple times and you're only supported by

ange':

- etc

Conclusion: climate alarmism is a lost cause.

6:44 AM · Mar 10, 2020 · Twitter for iPhone

A wide-angle shot of the Star Trek: Enterprise bridge. The room is filled with red and grey consoles, glowing blue lights, and multiple viewscreens. Several crew members in yellow and red uniforms are seated at their stations. Overlaid on the image are four thought bubbles, each connected to a specific area of the bridge by a series of smaller bubbles. The bubbles contain text that appears to be a list of tasks or needs.

I will be doing all the work

I need to make decisions based on timely science results

I need data pertinent to the problem, run the analysis, and interpret the results

I need to communicate the decision to stakeholders



khider@usc.edu



@dkhider