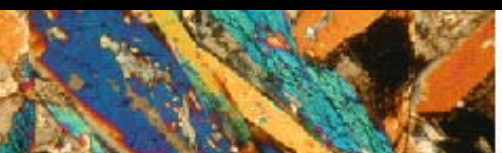
Sample Curation as Part of the Geoscience Cyberinfrastructure

K. Lehnert
Lamont-Doherty Earth Observatory
Of Columbia University

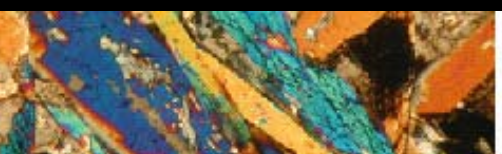


earthchem

ADVANCED DATA MANAGEMENT
IN IGNEOUS GEOCHEMISTRY

Sample Curation as Part of the Geoscience Cyberinfrastructure

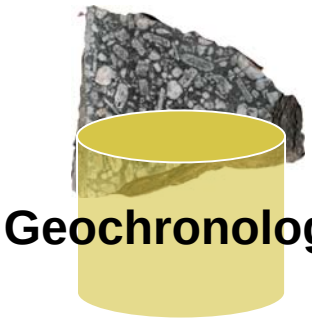
“Catherine C. and Janne B-T contacted me recently about a sample with the fairly useless name AD3-3 that I analyzed so long ago I can't remember. They wanted to reanalyze the Hf isotopes, because they are very high. Having spent several hours looking for the sample without success, I then decided to check the database to see who else might have it (I knew it was a well circulated sample). Turns out the list of people who have analyzed this sample reads like a who's who in isotope geochemistry. Most recently, Vincent S. analyzed it, who confirmed Patchett's high Hf ratio.”



Sample-Based Data

ISES-CI

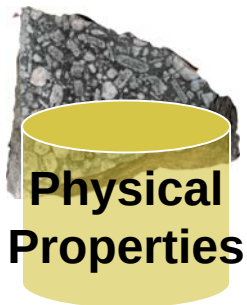
Cyberinfrastructure for Integrated Solid Earth Science



CHRONOS (EarthTime)

EarthChem (SedDB)

MagIC



(MetPetDB)

(PPRM-DB)

PaleoStrat

ISES-CI WGs

Samples as Part of a Geoscience Cyberinfrastructure

Objective:

Link physical samples to digital sample-based data implementations

- **Web-based ‘curation information system’**

Samples as Part of a Geoscience Cyberinfrastructure

The vision:

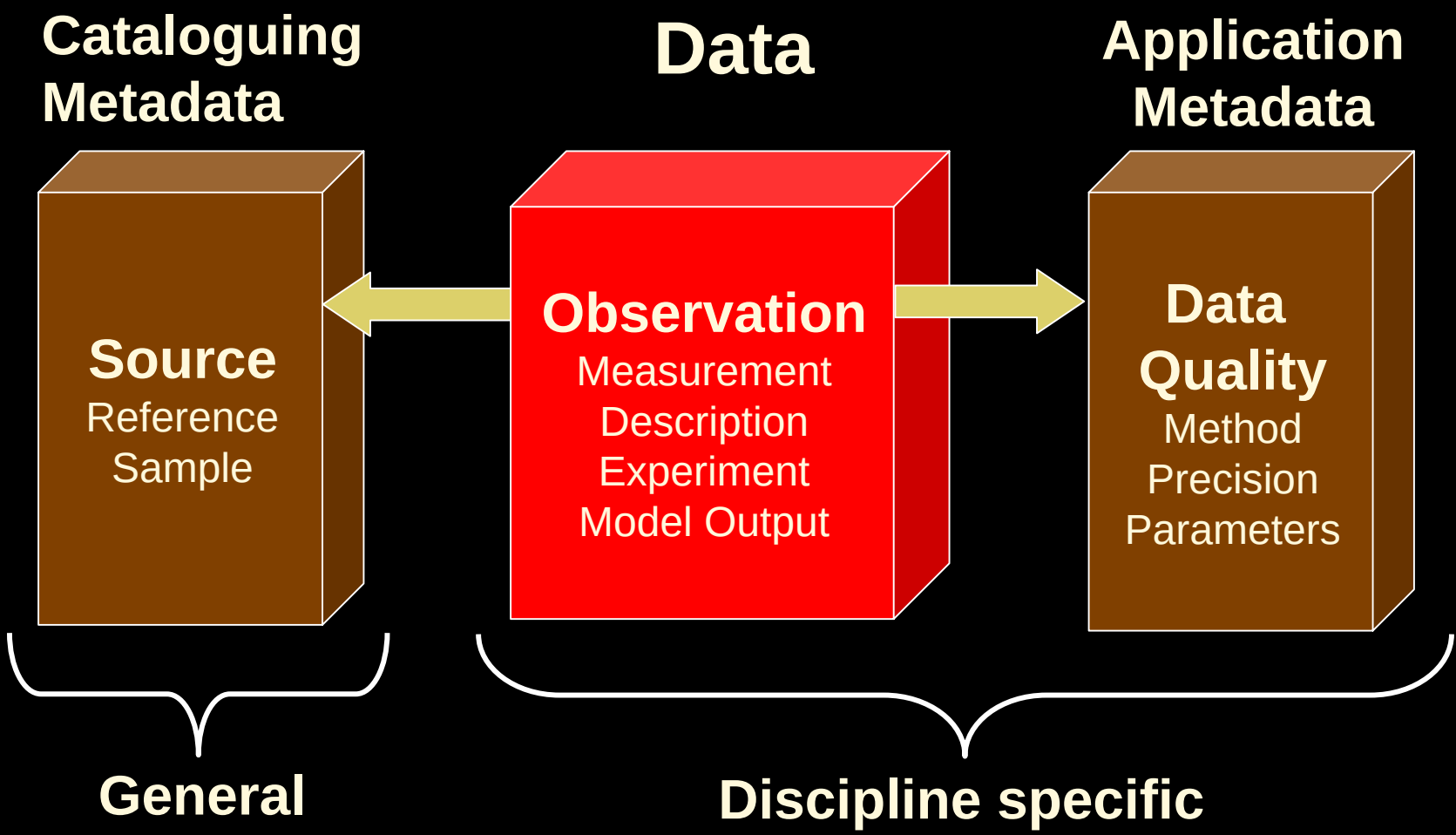
Web-based sample catalogues

- **that are integrated with other sample catalogues**
- **that are integrated with all sample-based data & information systems**

Information System for Sample Curation

- **Complementary**
  - Deliver data that ‘discipline-specific’ systems do not
  - Leave ‘discipline-specific’ data to disciplines
- **Compatible**
  - Metadata standards
  - Controlled Vocabularies
  - Unique identification of samples
- **Interoperable**
  - Access protocols

Content of Databases: Data and Metadata



Metadata: Geochemistry Databases

Red: mandatory metadata
White: essential metadata
Blue: desirable metadata

Cataloguing Metadata

- Rock type
- Location (lat/long
- Location names
- Depth
- Precision of location
- Sampling technique
- Expedition
- Sampling date
- Tectonic setting
- Age / Eruption date
- Rock class
- Texture
- Alteration
- Modal composition
- Archive

Reference

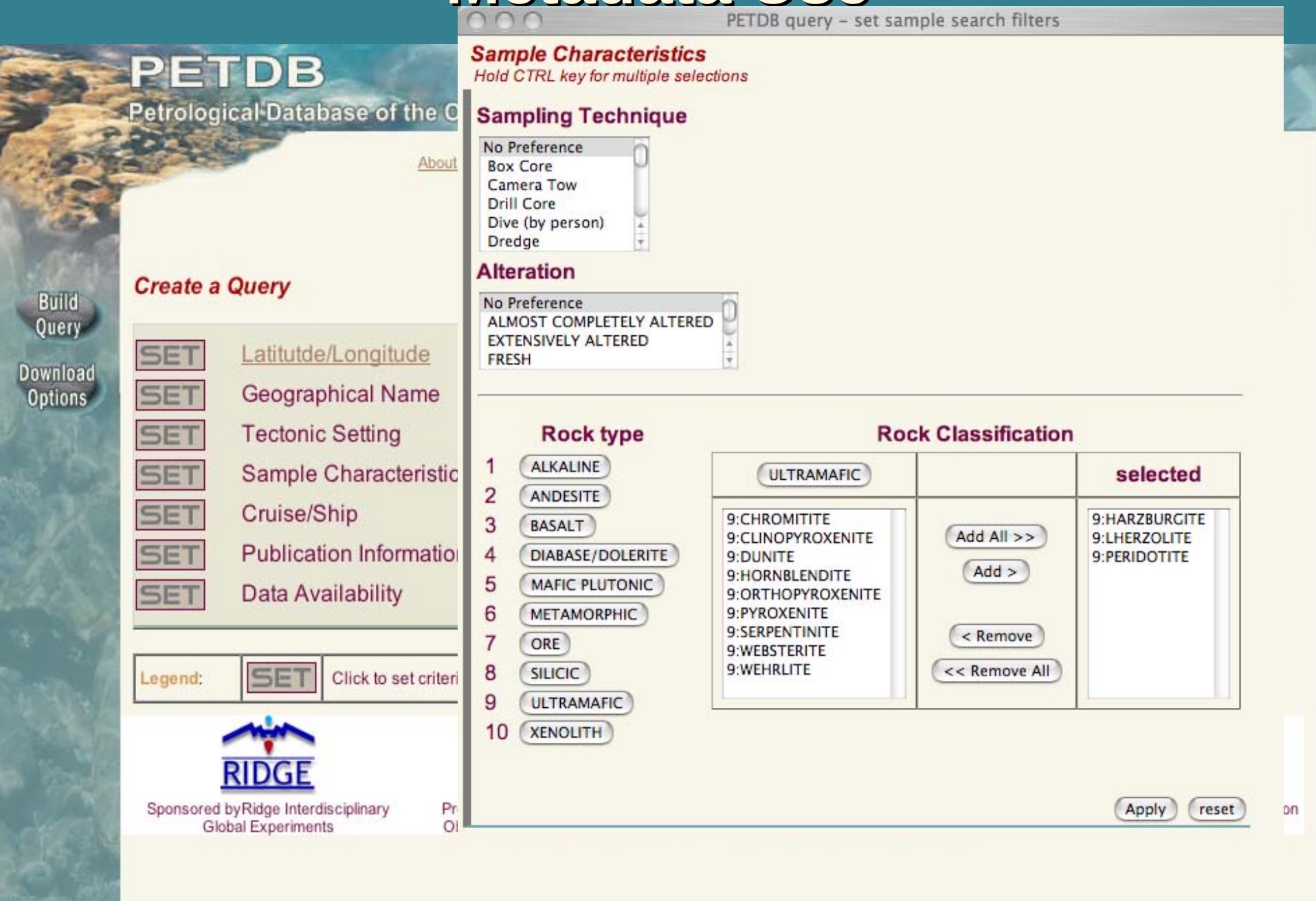
- Authors
- Title
- Journal
- Volume Number
- Publication Year

Application Metadata

- Unit
- Method
- Lab
- Precision
- Standard deviation
- Normalization
- Averaged or individual values
- Standard values
- Material (GL, WR, etc.)
- Mineral metadata: mineral type, grain type, generation,…
- Inclusion metadata: host mineral, heating, …

LAMONT-DOHERTY EARTH OBSERVATORY
THE EARTH INSTITUTE AT COLUMBIA UNIVERSITY

Metadata Use



Data Summary for ARGAMPH-001

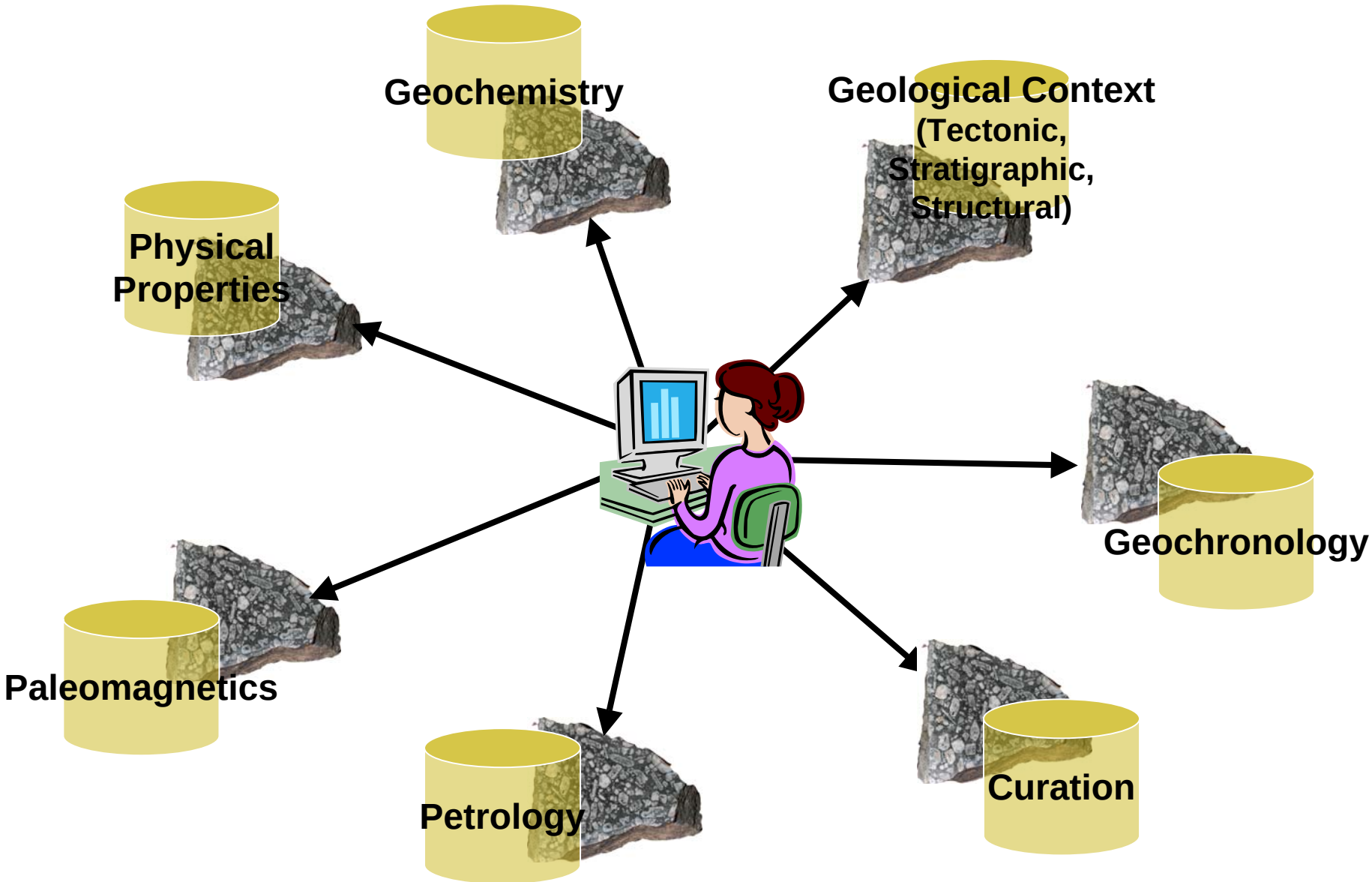
12 publications

analyses_num	material	reference	method and data quality	item_measured
27579	not specified	SUN,1980(201)	MS(707)	Pb206_Pb204, Pb207_Pb204, Pb208_Pb204
27773	whole rock	ENGEL,1964(215)	NN(723)	Al2O3, CaO, Fe2O3, FeO, H2OM, H2OP, K2O, MgO, MnO, Na2O, P2O5, SiO2, TiO2
29056	not specified	HART,1969(350)	XRF(772)	Sr
29057	not specified	HART,1969(350)	MS-ID(770)	Cs, K, Rb
32604	not specified	TATSUMOTO,1966(748)	MS-ID(893)	Pb
32605	not specified	TATSUMOTO,1966(748)	MS(894)	Pb206_Pb204, Pb207_Pb204, Pb208_Pb204, Th232_U238, U238_Pb204
32941	glass	MACDOUGALL,1986(388)	MS(909)	Nd143_Nd144, Sr87_Sr86
32942	glass	MACDOUGALL,1986(388)	MS-ID(910)	Nd, Rb, Sm, Sr
42816	whole rock	KAY,1970(185)	XRF(1047)	Al2O3, CaO, FeOT, K2O, MgO, SiO2, TiO2
42817	whole rock	KAY,1970(185)	AAS(1048)	Na2O
42845	whole rock	KAY,1970(185)	MS-ID(1050)	Ba, Ce, Cs, Er, Eu, K, La, Nd, Rb, Sm, Sr, Yb
63749	whole rock	SUBBARAO,1972(346)	AAS(1626)	K
63750	whole rock	SUBBARAO,1972(346)	XRF(1627)	Sr
63751	whole rock	SUBBARAO,1972(346)	MS-ID(1628)	Rb
63752	whole rock	SUBBARAO,1972(346)	MS(1630)	Sr87_Sr86
65797	whole rock	ENGEL,1965(126)	ES(1727)	Ba, Co, Cr, Cu, Ga, Li, Ni, Sc, Sr, V, Y, Yb, Zr
67732	not specified	PINEAU,1983(157)	MS(1753)	DELTA_O18
69456	whole rock	HART,1971(371)	MS-ID(1841)	Cs, K, Rb

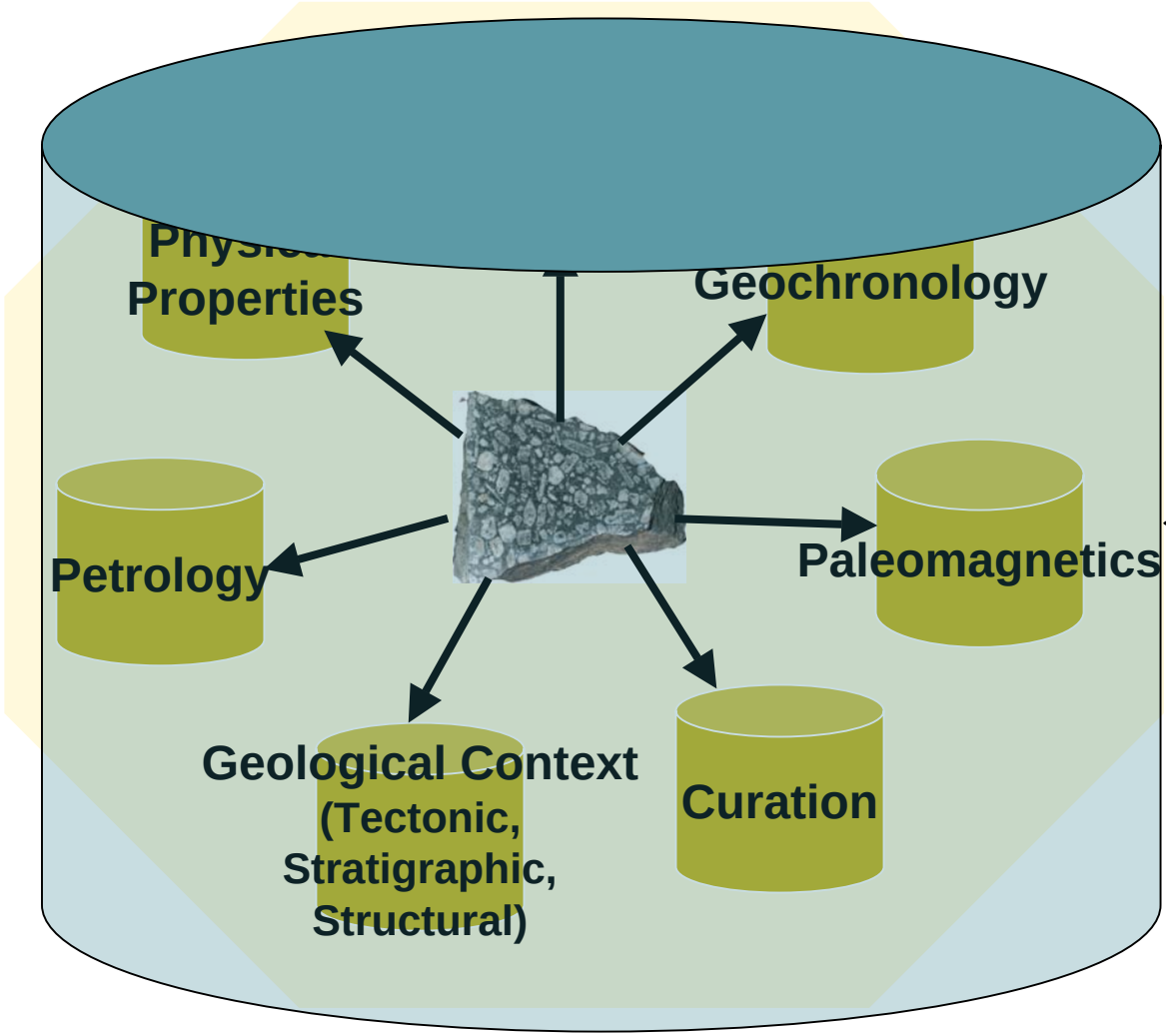
Curation-specific Data

- **Descriptive information**
  - Petrography (texture, alteration, etc.)
  - Visual material (hand specimen & thin section photos
- **Material**
  - Size of sample / weight
  - Subsamples
    - processed rock splits (powders)
    - mineral/fossil separates
  - Material Processing
- **Sampling History**
  - Who sampled?
  - What was the purpose?

Sample-based data: The current situation



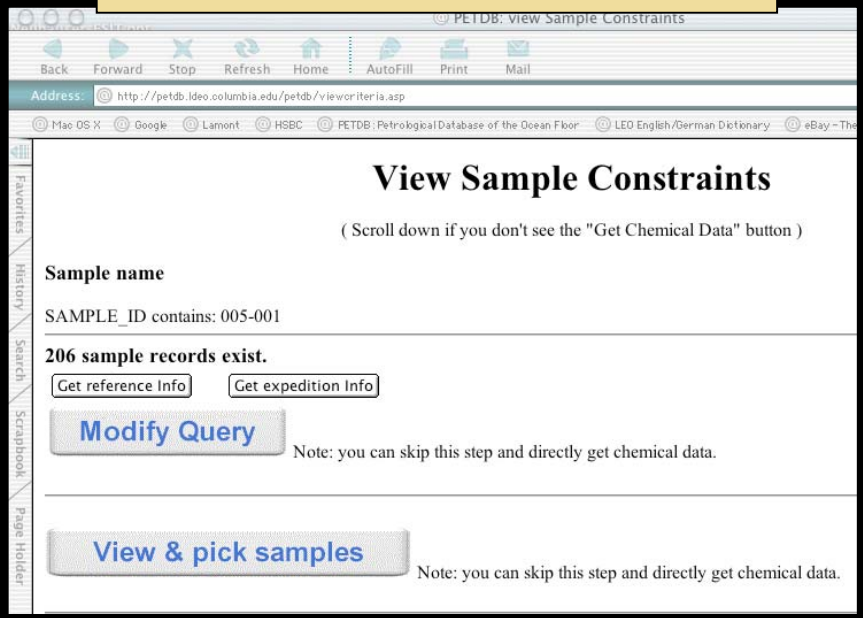
Integrated Sample-Based Information System



Users can retrieve and integrate data from distributed sources.

The Problem of Samples Names

101 different samples in PetDB have the name 5-1



Aliases

Alias	reference
PV D-1	ENGEL, A E J; 1965
AMPH-1D	PINEAU, F; 1983
AMPH D1	KAY, R; 1970
AMPH D-1	SUN, S S; 1980
D1	ENGEL, A E J; 1964
S-9	SUBBARAO, K V; 1972
PD-1	HEDGE, C E; 1970
AMPH PD1-P	HART, S R; 1969
AMPH 1-PDIP	HART, S R; 1971
AMPH1D	PINEAU, F; 1976
AMPH-D1	MACDOUGALL, J D; 1986
PD1	TATSUMOTO, M; 1966

Different names for sample ARGAMPH-001 (Dredge 1 from the cruise AMPHITRITE on R/V Argo)

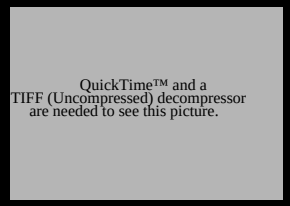
The Need for Unique Sample Identification

- **Integration of data in a distributed system requires unique identification of samples.**
- **Currently, naming of samples is ambiguous.**
  - Different samples have identical names.
  - Samples are renamed.
- **Metadata that allow unique identification are often missing for terrestrial samples.**
- **Institutions have their own naming protocols, no assurance that names are unique on a global scale.**

Unique Sample Identification

- **Unique naming requires clearinghouse that provides a centralized registry of sample identifiers on a global basis.**
  - **To avoid ambiguity**
  - **To systematize sample designation**
  - **To ensure that all information associated with a sample will in fact be unique.**

An Approach to Unique Sample Identification



IGSN

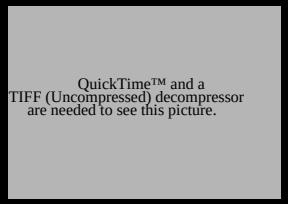
International Geo Sample Number

Solid Earth Sample Registry SESAR

"Establish a system that provides identifiers for solid earth samples to global sharing, linking, and information and data about these

LAMONT-DOHERTY EARTH OBSERVATORY
THE EARTH INSTITUTE AT COLUMBIA UNIVERSITY

An Approach to Unique Sample Identification



IGSN

International Geo Sample Number

Solid Earth Sample Registry SESAR

- **Easy access for registration of samples with minimum set of required metadata**
- **Generation of bar codes**
- **Robust and secure IT environment**
- **Full integration with sample-based data systems**

The Challenge

Development and implementation of a *"Cyberinfrastructure for Solid Earth Science"* requires a change in the investigators' mindset.

"This is community's data for which I am responsible."

The Challenge

Development and implementation of a *"Cyberinfrastructure for Solid Earth Science"* requires a change in the investigators' mindset.

"This is community's samples for which I am responsible."

Metadata in Sample-based Data Management

- Level 0: Mandatory
  - Location: Latitude, longitude, elevation
  - Rock type
- Level I: Essential
  - Age
  - Geological context (Tectonic setting, stratigraphic context, etc.)
  - (Geographical / political location)
  - Sampling Process
- Level II: Desirable
  - Petrographic descriptions
  - Components

Metadata

Metadata is

1. used for discovering existence of data by searching a metadata catalogue =
   ⇨cataloguing metadata
2. documentary information describing the content, context, quality, structure, accessibility of a specific data set =
   ⇨application metadata

Any particular parameter may be used in either way depending on the user's purpose.

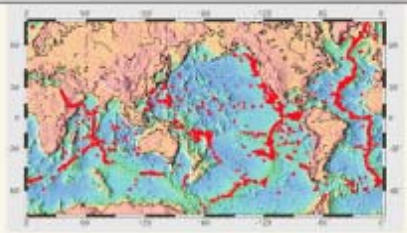
Helly, J. et al.: Scalable models of data sharing in Earth sciences. G³, 2003

Use of ‘Cataloguing’ Metadata

PETDB: Single sample info

Sample info

Sample ID:All0107-6-040-013



Location:

Latitude: 54.422°W
Longitude: 1.528°N
Elevation: -2724 to -3240
Tectonic setting: FRACTURE_ZONE
Location: SOUTHWEST INDIAN RIDGE:SPREADING_CENTER
Location Comment: NORTHWEST WALL

Sampling Information

Cruise: All107-6 Chief Scientist: DICK
(ATLANTIS II, 1980) Technique: Dredge
Date: Station: All0107-6-040

Sample Description:

Rock type: U
Classification: PERIDOTITE
Description:
Alteration: EXTENSIVELY ALTERED
Age:
Archived at:

Mineral Data

Reference	Alias	Method	Mineral	SiO2	TiO2	Al2O3	Cr2O3	FeO	MnO	MgO	CaO	Na2O	Ce	Nd	Sm	Eu	Dy	Er	Yb	Cr	Sc	Sr	Ti	V	Zr
JOHNSON, 1990	All107:40-13	EMP	CPX	51.38	0.1	4.27	1.23	2.35	0.1	16.89	23.99	0.09													
JOHNSON, 1990	All107:40-13	EMP	OPX	55.49	0.07	3.19	0.67	6.02	0.12	33.57	1.27	0.01													
JOHNSON, 1990	All107:40-13	IMP	CPX										0.01	0.01	0.05	0.03	0.38	0.43	0.46	8519.0	53.0	1.2	368.0	291.0	0.1