



Kerstin A. Lehnert

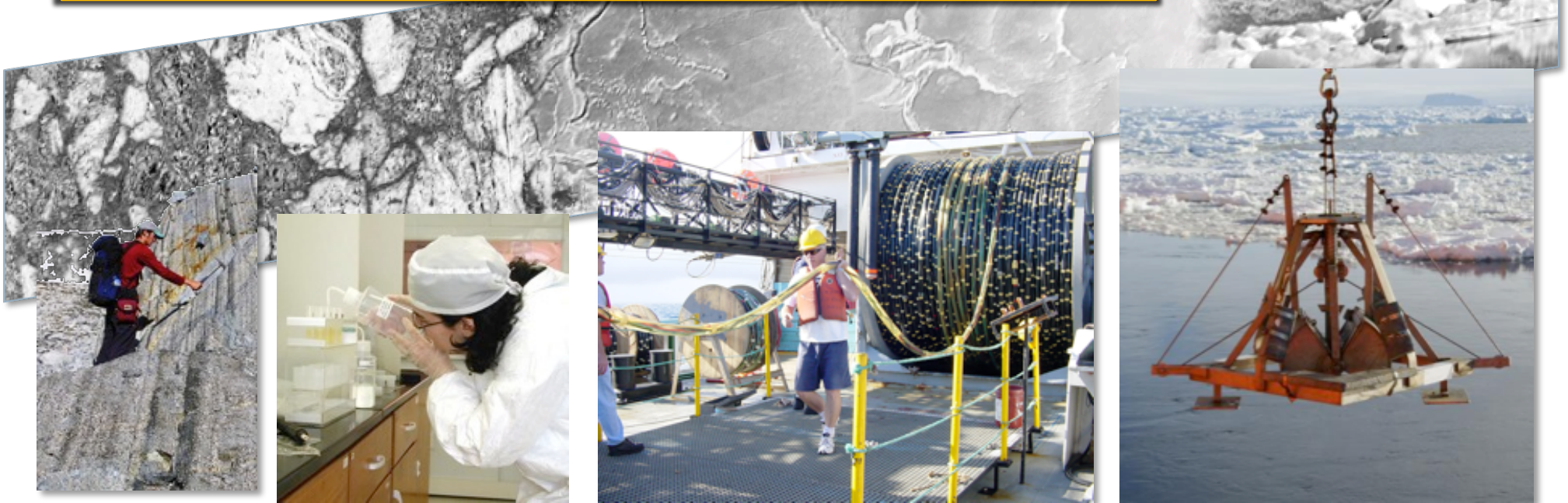
*Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA*

# Data Publication at IEDA: Making Data Fit for Re-use

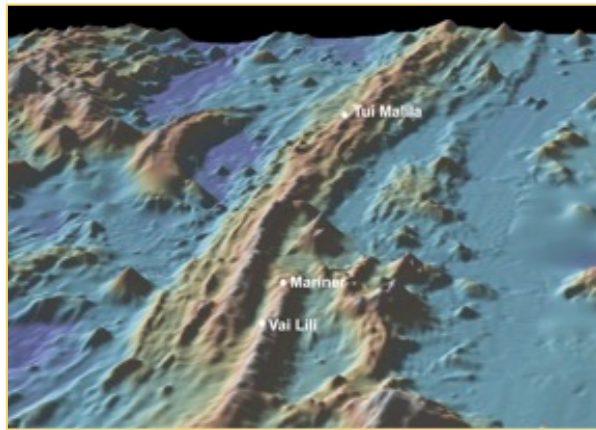


“... a community-based facility that serves to support, sustain, and advance the geosciences by providing a centralized location for the registry of and access to data essential for research in the solid-earth and polar sciences.”

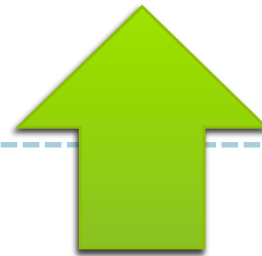
[www.iedadata.org](http://www.iedadata.org)



# IEDA Scope



Derived Data



Field Data



Sensor-based



Sample-based

# IEDA Objectives

- increase productivity and capability of the solid-earth and polar geoscience communities
- provide public access to important solid-earth and polar geoscience datasets

- provide a means for enhancing the interoperability of data sets in research
- provide management

- re-use
- verification

# IEDA Systems

## ➤ Repositories & registries

- ◆ Marine Geoscience Data System, EarthChem Library
- ◆ System for Earth Sample Registration

## ➤ Data syntheses & products

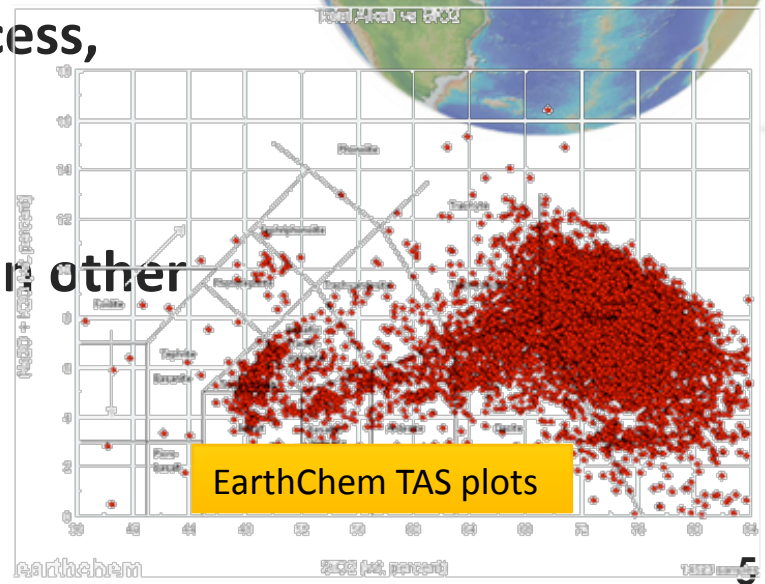
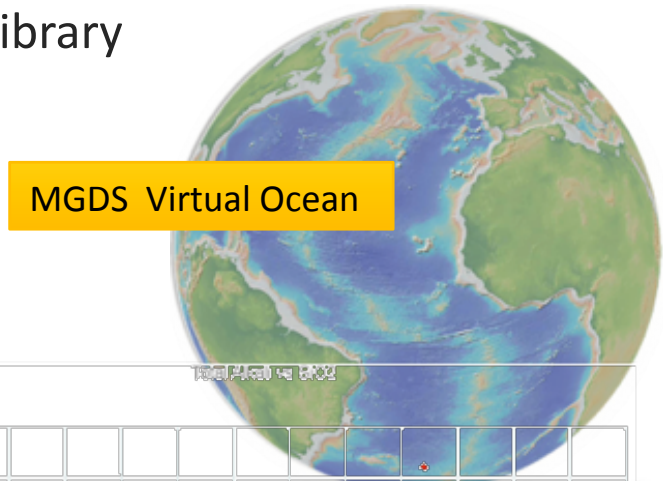
- ◆ GMRT, PetDB, SedDB, Geochron

## ➤ Software tools for data discovery, access, visualization and analysis

- ◆ GeoMapApp, Virtual Ocean, EarthChem

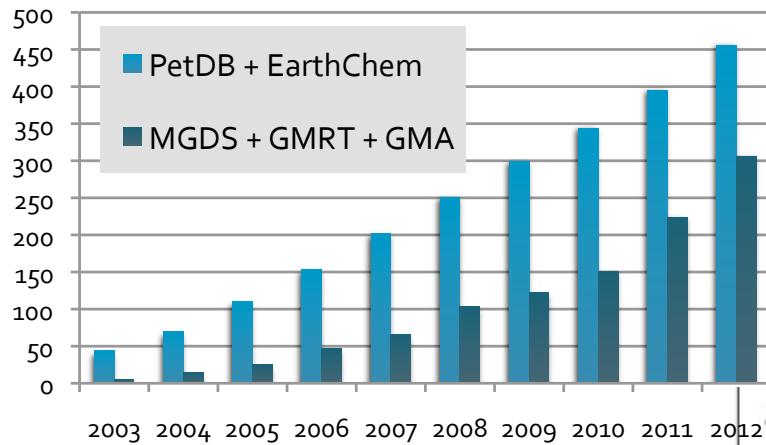
## ➤ Portals to complementary data held in other repositories

- ◆ ASP, EarthChem, USAP-DCC



# Record of Re-use

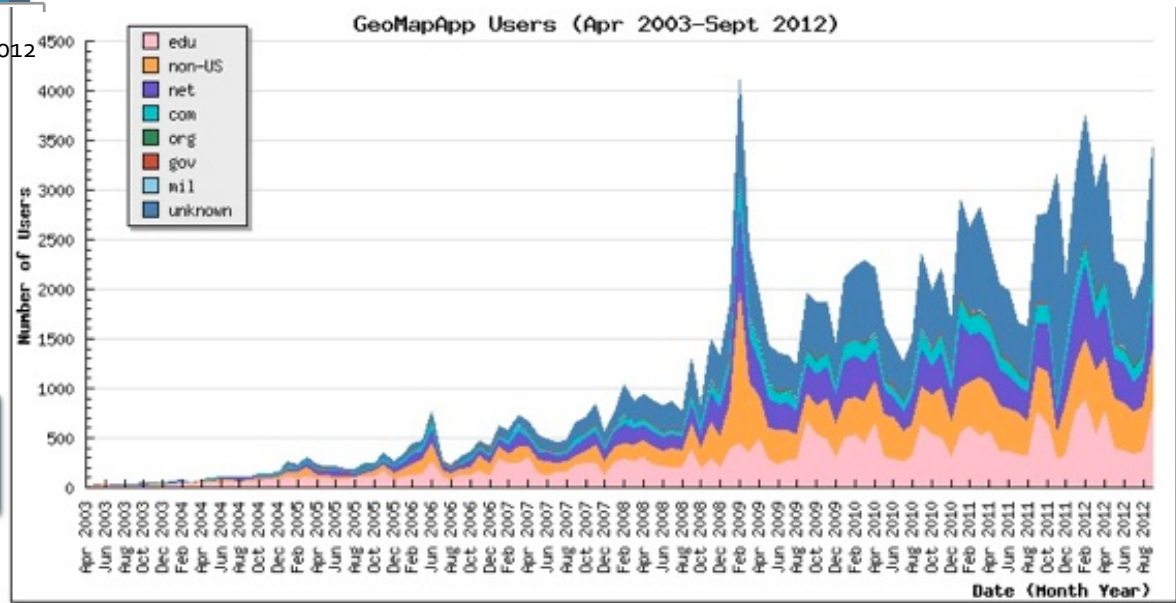
## Downloads



	Unique Users Downloading Data	Unique Download Requests	Total Volume (TB)	Total # Files
MGDS (yr 1)	2,783	5,808	68	521,180
MGDS (yr 2)	2,632	5,049	58	248,442

## Citations

## Visitors



# Making Data Fit for Reuse

- **Document provenance of data**
  - ◆ collection strategies, procedures and underlying assumptions
- **Comply with standards for data collection and representation (formats, semantics, etc.)**
- **Document data precision, errors, missing data, and verify/validate workflows for data quality assurance**



# 'Data Publication' at IEDA

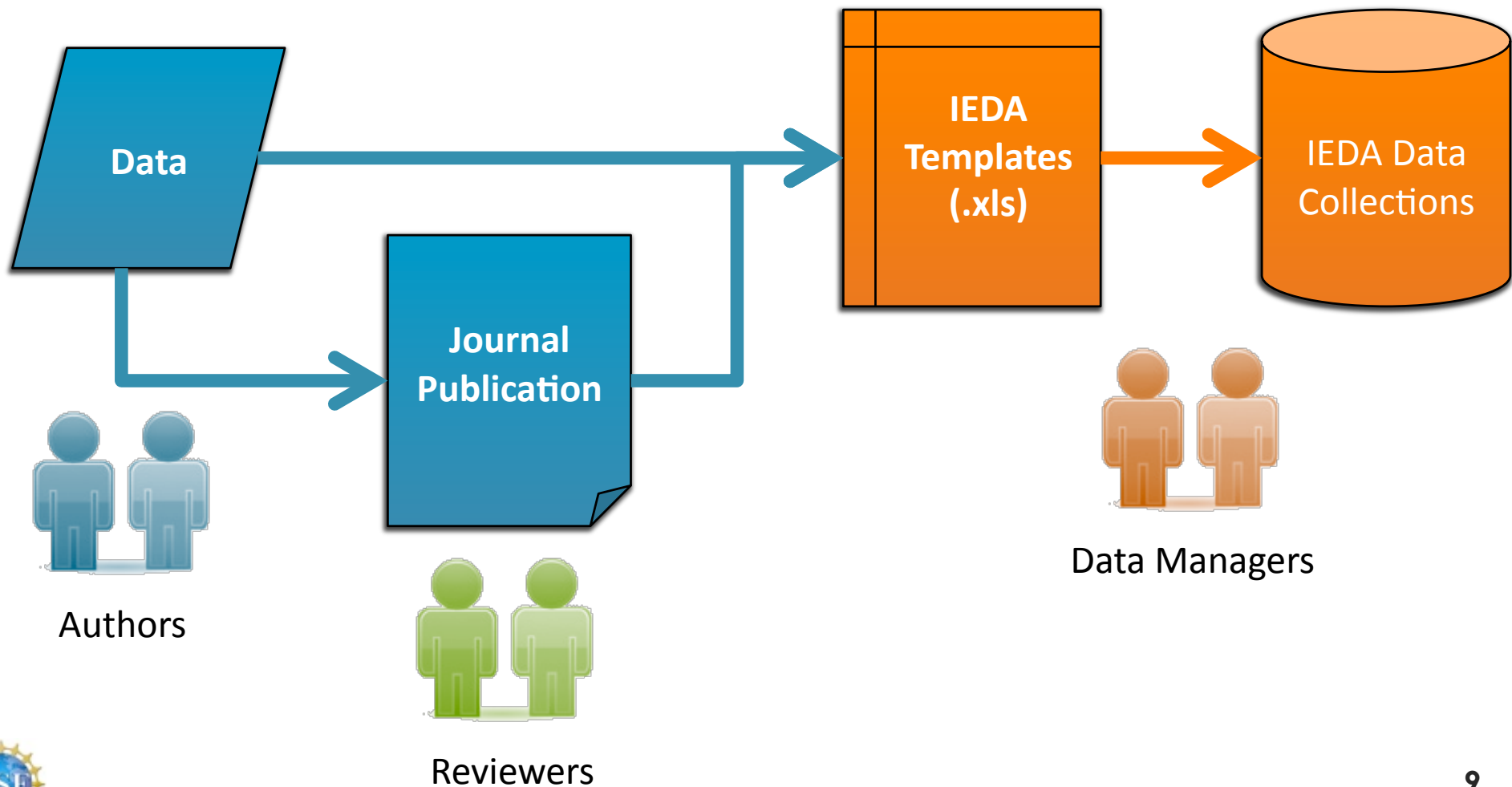
➤ **Past**

➤ **Present**

➤ **Future**



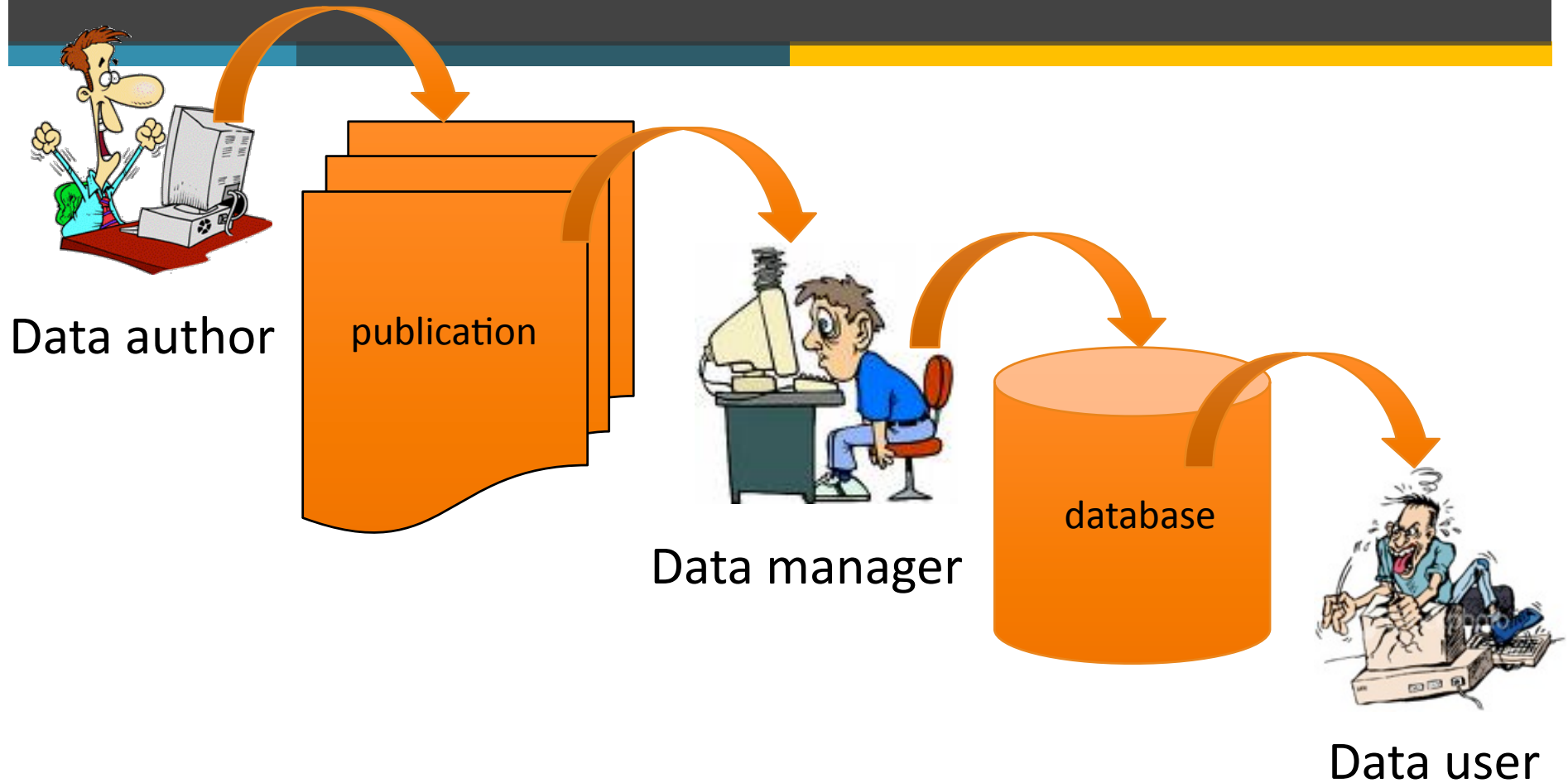
# Past Flow of Data into IEDA



# QC/Review by Data Managers

- **development of metadata for new data sets**
  - ◆ extract from publications
  - ◆ extract from secondary literature
  - ◆ contact authors
- **continued development of metadata schemas and vocabularies to align with evolving community standards**
- **ongoing evaluation to ensure completeness of metadata for existing data holdings**
- **data verification ensuring that data files are readable**

# Fit for Re-use?



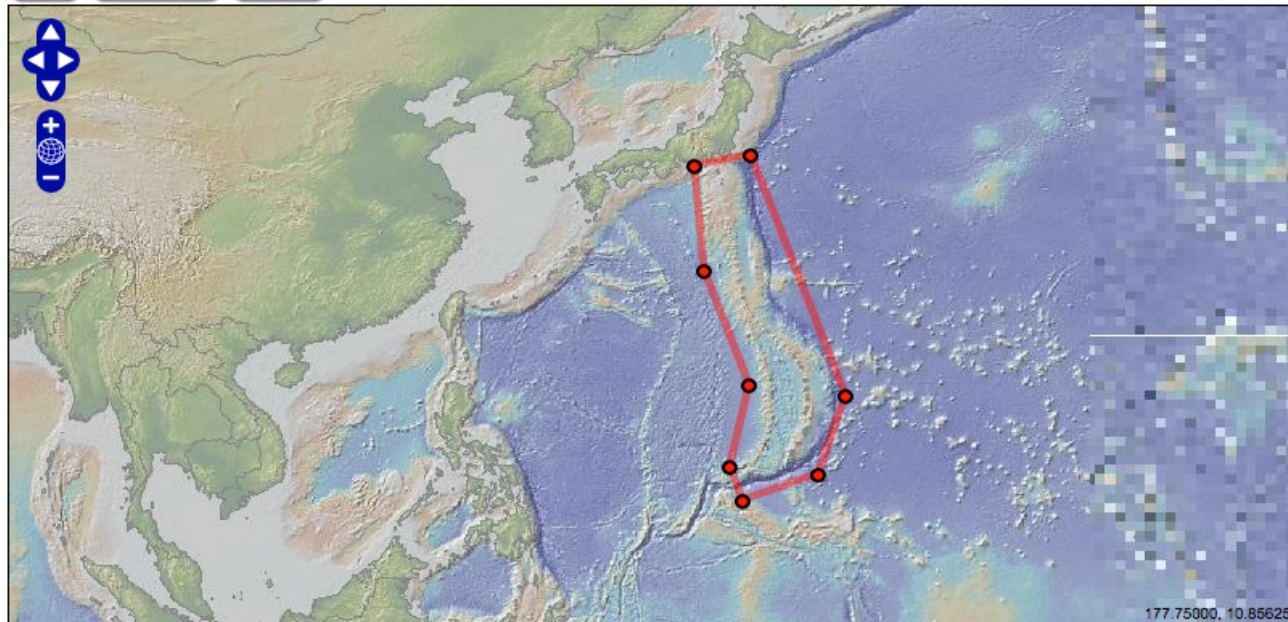
# geospatial metadata

## Earthchem Polygon Map

Click on the map below to define a polygon to search in. You can drag the map dynamically, as well as use the zoom bar to zoom in and out. Use shift-click to create a dynamic zoom range for more detailed zooming. Click 'clear' to start over.

Submit

Clear Topography Satellite



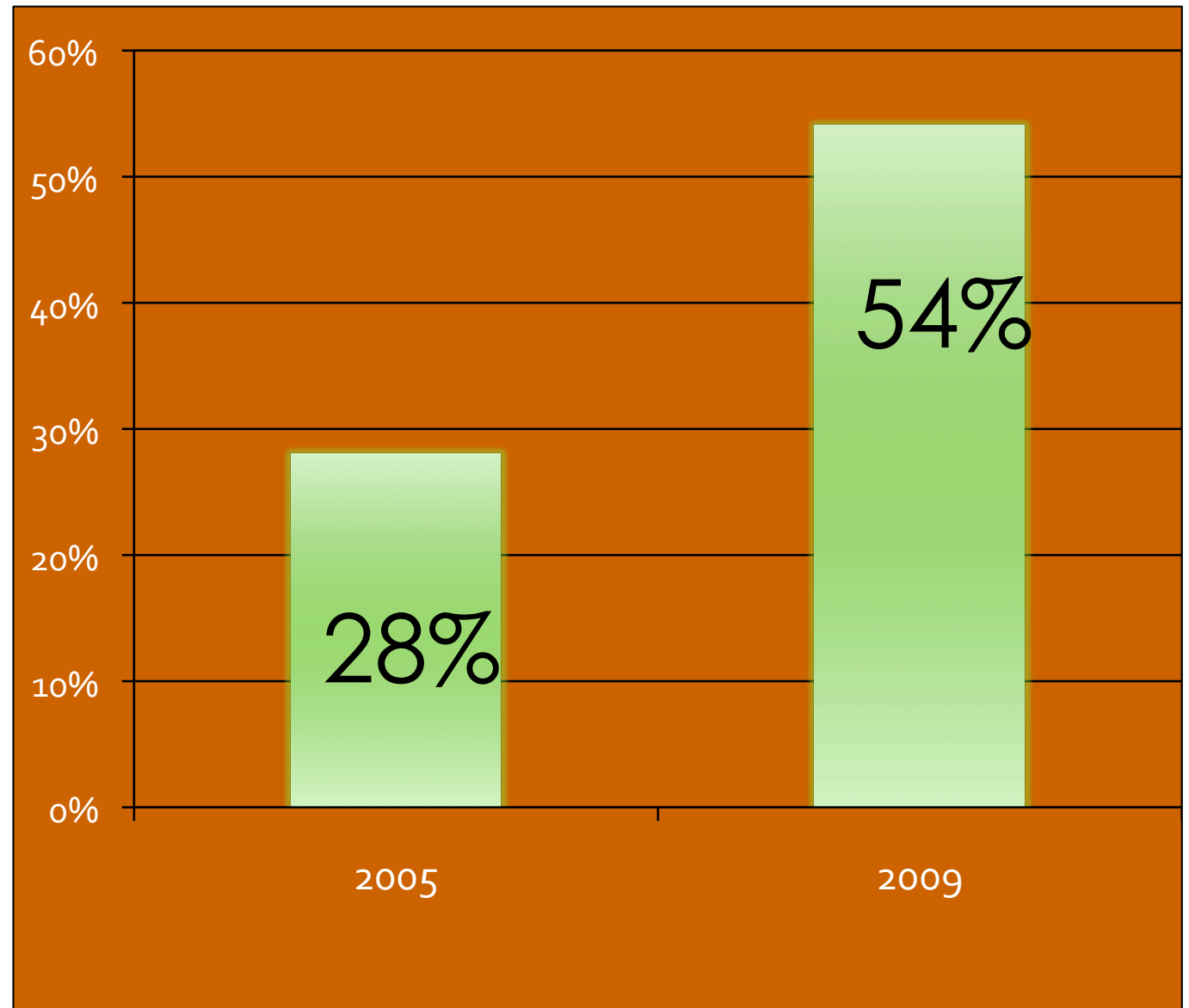
177.75000, 10.85625



missing metadata

**Percentage of publications that lists geospatial coordinates for sample locations**

**Example:**  
*Journal of Petrology*



# database production

magic?



"I CAN ASSURE YOU MR RUMPLESTILSKIN, WEAVING STRAW INTO GOLD IS A SKILL WE CAN CERTAINLY USE ..."

# Standards for Data Reporting

- **Build community consensus what metadata are essential for evaluation of data for re-use.**
- **Ensure that these metadata are included with the data in publications or databases to ensure their long-term utility.**
- **Ensure that data are reported in a consistent way so that they can be easily integrated and compared.**

# Developing Data Standards

- **EarthChem community workshops**
- **Editors Roundtable at Goldschmidt 2008 and AGU 2008 with representation from EPSL, Chem Geology, G-Cubed, Journal Petrology, GCA, Nature, Science, Geology, GSA Publications, Springer**
  - ◆ Joint policy recommendation approved
    - Complete data publication
    - Essential metadata
  - ◆ EPSL implemented in January 2011
  - ◆ Basis for IEDA Data Submission guidelines for geochemistry

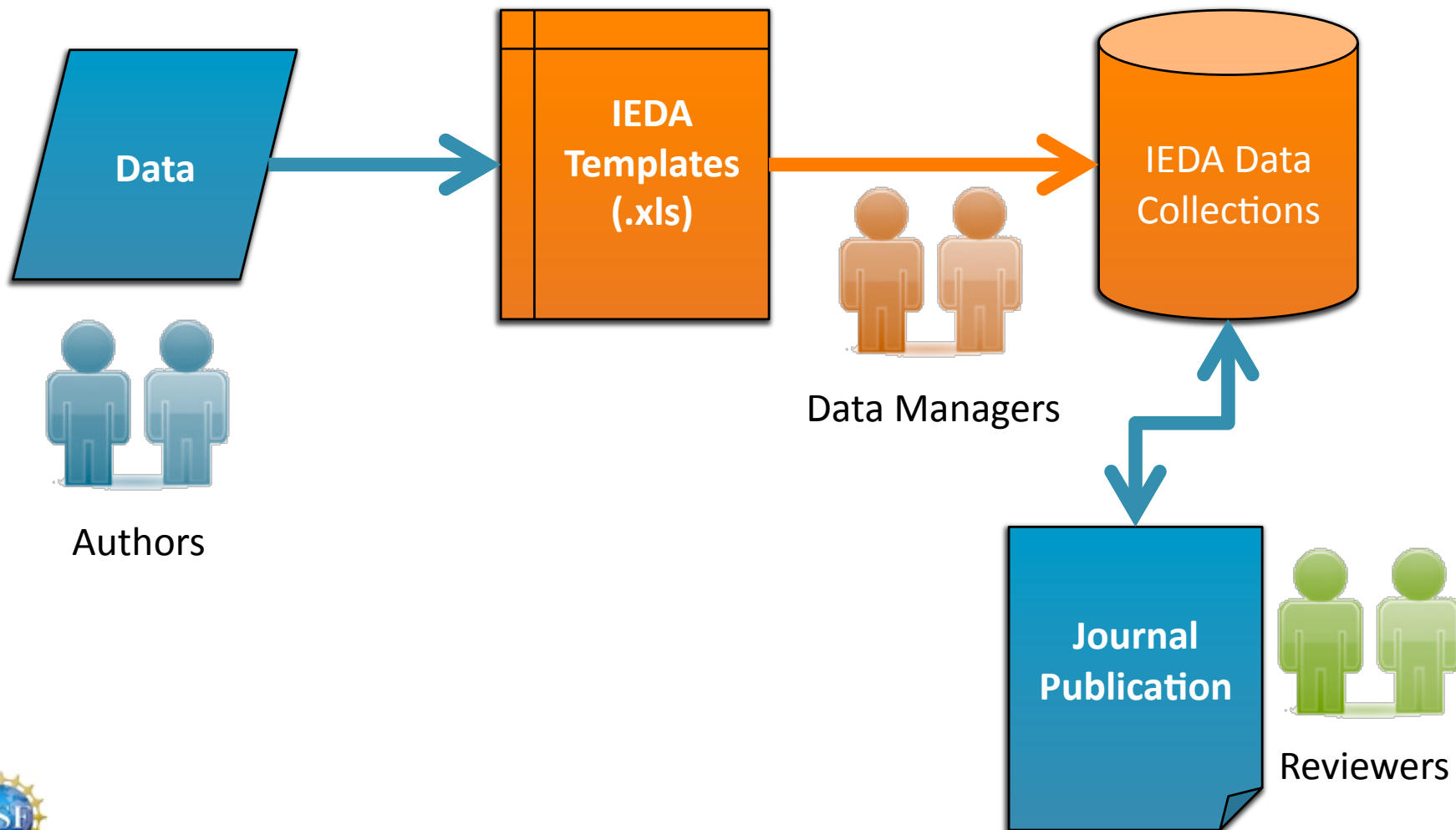


# Editors RT Policy Recommendation

- **Complete data publication**
  - ◆ Require authors to include tables presenting all data used in a publication (“if it is plotted it should be tabulated”).
- **Essential metadata**
  - ◆ Require sample metadata, e.g. location, lithology, whenever relevant
- **Unique identification of samples**
- **Open access**
  - ◆ If a relevant database exists, editors should adopt policies to ensure that the data will be deposited there.
  - ◆ Incorporation into the databases should be as seamless as possible.



# Present Flow of Data into IEDA



# Changes from Past to Present

- **Cooperative Agreement with NSF as facility**
  - ◆ Provides new level of sustainability
  - ◆ Requirement: Community governance & guidance
    - Formal Policy Committee & User Committee
    - Community feedback & acceptance (outreach, demonstrate use, utility, and scientific impact)
  - ◆ Request: Support for compliance with NSF Data Policies
    - Open access
    - Moratorium periods
    - Data Management Plans

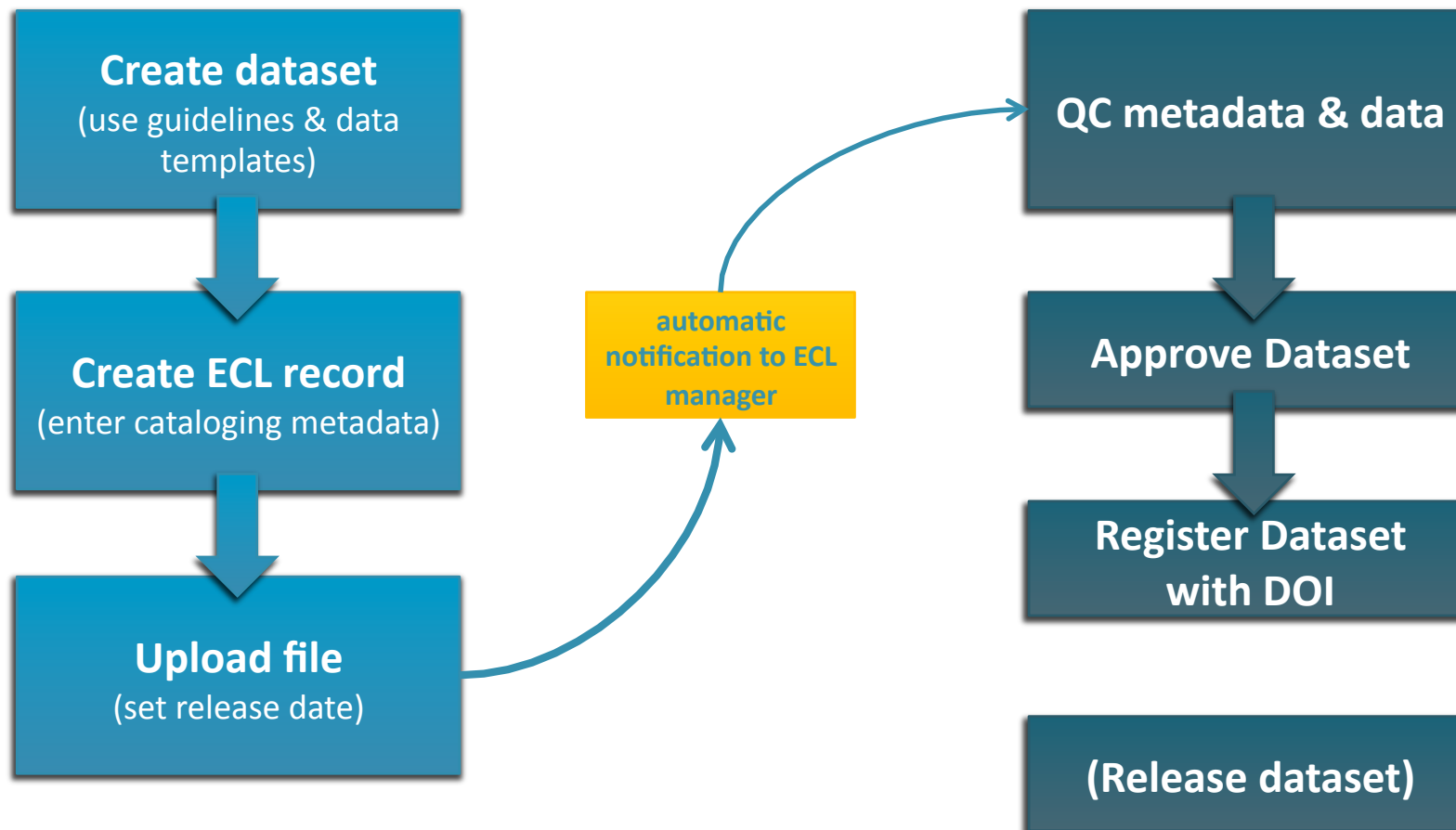
# Changes from Past to Present

- **Major improvements to infrastructure, management, and policies**
  - ◆ Rigorous risk management
  - ◆ Persistent identification of data & samples (DOI, IGSN)
  - ◆ Long-term archiving via agreements with NGDC and Columbia University Libraries
  - ◆ Cross-referencing with publishers, data citation index, etc.
- **Increasing focus on user contributions**
- **Member of World Data System**

# EarthChem Library Data Publication

Investigator

GRL Manager



# Unresolved Issues

- **Requests to accept new data types without community standards**
- **Need to include long narratives describing procedures (data journal publication?)**
- **Migration to synthesis data collections for**
- **Recording feedback from data re-use**

# More Changes from Past to Present

## ➤ **Growing enforcement of data policies at NSF**

- ◆ Data Management Plan requirement
- ◆ Still inconsistent monitoring of compliance, but improving

## ➤ **Culture Change (slow, but moving)**

- ◆ Driven by need for easy access to data within and across domains
- ◆ Supported by an evolving infrastructure of data publication & citation (data DOI, Data Citation Index)

# IEDA Support for Data Compliance

- ➔ Tools for data management
  - ◆ Data Management Plans
  - ◆ Data Compliance Reports
- ➔ Templates for data documentation
- ➔ Tutorials & videos

## Data Compliance Reporting Tool (v1.0)

**Instructions:** The IEDA Data Compliance Report Tool enables the easy preparation of reports to demonstrate compliance with NSF Data Policies. Enter a NSF award and this service will provide a list of related data sets and their release status. Note that data will only be returned for awards that are currently cataloged within IEDA, and data sets returned are based on data and metadata received to date. Please [contact us](#) with comments or questions, or to [submit additional data or metadata](#).

Enter NSF Award

0527053

Locate a NSF Award through the [Fastlane Award Search](#)

URL to this dynamic report: [http://www.iedadata.org/compliance/report?award\\_id=0527053](http://www.iedadata.org/compliance/report?award_id=0527053)

[Download as PDF](#)

NSF Award Info

**Award Title:** [Collaborative Research: Anatomy of an Overlapping Spreading Center: Geochemical and Geological Study of the EPR 9 Degrees 30'N OSC](#)  
**Investigator(s):** Kenneth Sims  
**Award ID:** 0527053

Sensor Data Linked to Award

Data System	Expedition/Compilation	Data Type(s)	# of Data Sets	Instrument Info	Investigator(s)	Release Information	Citations
<a href="#">MGDS</a>	<a href="#">AT15-17</a>	<a href="#">Navigation:Primary</a>	3	<i>Atlantis</i> Navigation	Klein, Emily	Released	Not Supplied
<a href="#">MGDS</a>	<a href="#">AT15-17</a>	<a href="#">Photograph</a> <a href="#">Photograph:WebGallery</a>	2	<i>TowCam</i> Camera:Digital	Fornari, Daniel Klein, Emily	Released	Not Supplied
<a href="#">MGDS</a>	<a href="#">AT15-17</a>	<a href="#">Photograph:WebGallery</a>	1	<i>Jason II</i> Camera:Digital	Fornari, Daniel Klein, Emily	Released	Not Supplied
<a href="#">MGDS</a>	<a href="#">AT15-17</a>	<a href="#">Bathymetry:Swath</a>	2	<i>Jason II</i> Sonar:Multibeam	Ferrini, Vicki Fornari, Daniel Soule, S.	Released	<a href="#">Soule et al., 2008</a>
<a href="#">MGDS</a>	<a href="#">AT15-17</a>	<a href="#">Bathymetry:Swath</a>	1	<i>DSL-120A</i> Sonar:Multibeam	Klein, Emily White, Scott	Released	Not Supplied
<a href="#">MGDS</a>	<a href="#">AT15-17</a>	<a href="#">Bathymetry:Phase Sidescan</a>	2	<i>DSL-120A</i> Sonar:Sidescan	Klein, Emily White, Scott	Released	Not Supplied
<a href="#">MediaBank</a>	<a href="#">AT15-17</a>	<a href="#">Educational Image(s) and/or Video(s)</a>		Not Applicable	Fornari, Daniel Klein, Emily		
<a href="#">WHOI:NDSF</a>	<a href="#">AT15-17</a>	<a href="#">Photograph:Video:FrameGrab</a>		<i>Jason II</i> Camera:Video	Klein, Emily		
<a href="#">R2R</a>	<a href="#">AT15-17</a>	<a href="#">Underway Data and Final Navigation</a>					

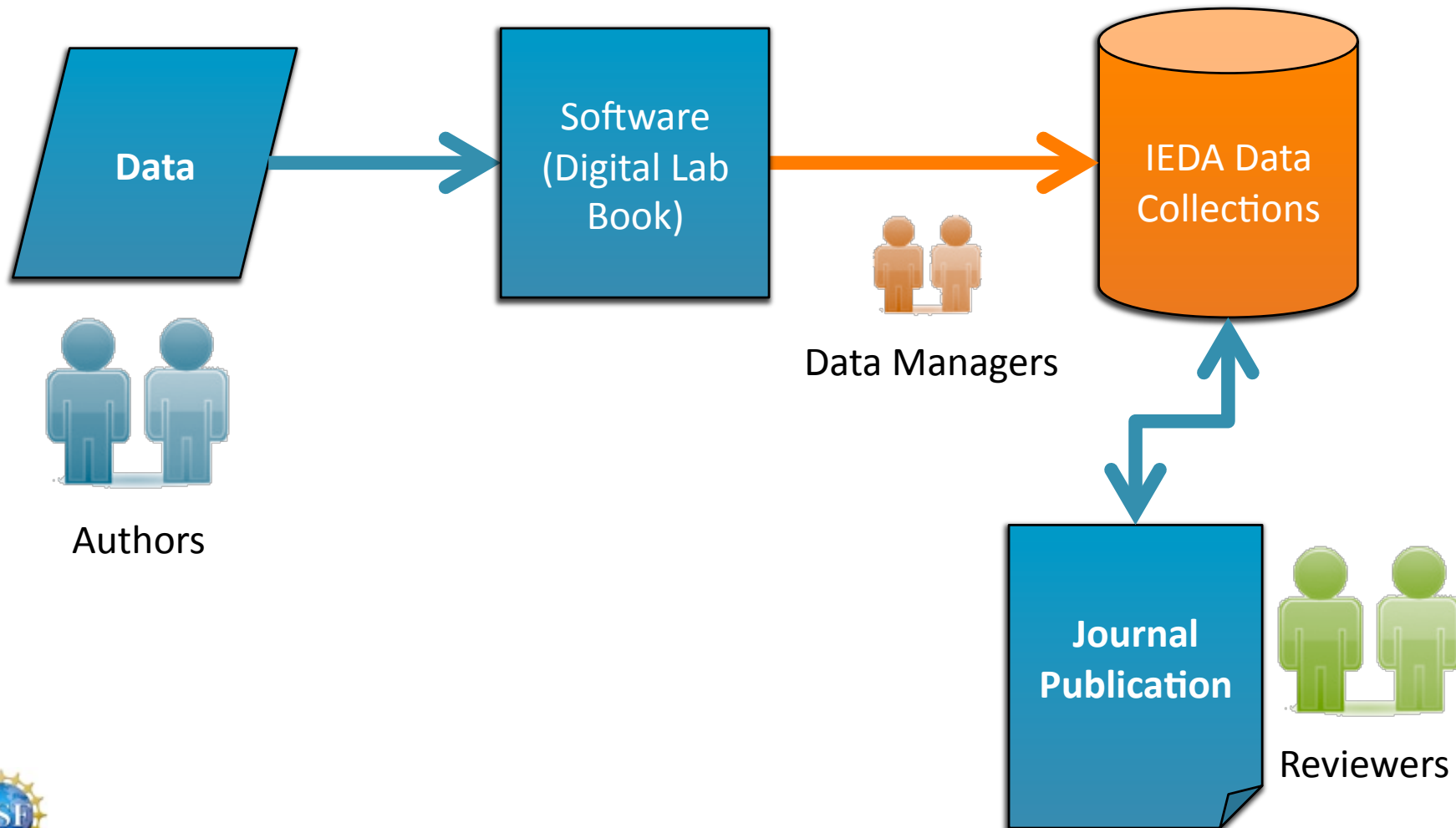
Sample-Based Data Linked to Award

**Data Collection Repository Data Link**  
 EarthChem Library [Data](#)  
 PetDB [Data](#)





# Future Flow of Data into IEDA



# Add Sample Metadata

<<Back to My Data <<Back to Dataset (My Data 2)

## Provide List of Samples for: My Data 2

To register samples, (1) Download a Sample Metadata Template, (2) Fill in the template using Excel, (3) Upload the Sample Metadata Template.

For each sample, please provide (mandatory metadata in **bold**):

- **Sample name**
- **IGSN** (International Geo Sample Number) In order to get IGSNs for your samples, please go to [www.geosamples.org](http://www.geosamples.org) and register the samples.
- **Geospatial Coordinates** (latitude/longitude, elevation or water depth)
- **Material** (e.g. rock, soil, water) and a classification
- Depth in Core (if sample taken from a core)

If possible and applicable, also provide:

- Age
- Keywords for geological context, e.g. geological unit, tectonic setting, physiographic feature
- Field program/cruise, on which the sample was collected
- Name of parent such as a core or dredge

Please, use the EarthChem Sample Metadata Template to submit your sample metadata.

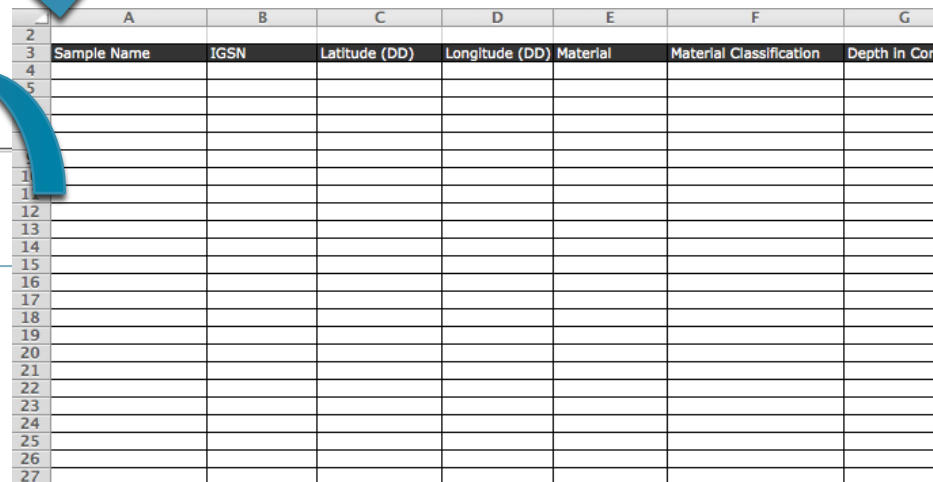
### Download Sample Metadata Template

### Upload Sample Metadata

Metadata File:

Browse...

Upload



	A	B	C	D	E	F	G
2							
3	Sample Name	IGSN	Latitude (DD)	Longitude (DD)	Material	Material Classification	Depth In Core
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							

# 'Digital Lab Book' (under development)

## Analytical Lab Profiles

This section shows your personal Analytical Lab Profiles. Register an Analytical Lab profile to store the Technique, Instrument, Lab Name, Department, Institution, and variables commonly associated with the lab profile. Once stored, you can retrieve the profile when creating a data set template, and will not need to enter the same information again.

Technique	Instrument	Lab	Department	Institution	
FTIR (FOURIER TRANSFORM INFRARED SPECTROSCOPY)		Goldstein MS	Lamont-Doherty Earth	Columbia	<a href="#">Edit</a>

## Data Sets

This section shows your data sets and their current status. Datasets marked "pending" are not yet finalized, while datasets marked "ready" are finalized and can be submitted to EarthChem through the GRL. Registering a dataset requires three steps, (1) providing the sample information via spreadsheet (2) providing the variables measured using a registered lab profile, and (3) entering the data via spreadsheet.

Title	Created On	Status			
Test34	1/03/2013	pending	<a href="#">View/Edit</a>	<a href="#">Delete</a>	
Test3	12/11/2012	pending	<a href="#">View/Edit</a>	<a href="#">Delete</a>	
Test two	10/16/2012	ready	<a href="#">View/Edit</a>	<a href="#">Delete</a>	<a href="#">Submit to GRL</a>
test title	9/10/2012	pending	<a href="#">View/Edit</a>	<a href="#">Delete</a>	
My Data 2	7/24/2012	pending	<a href="#">View/Edit</a>	<a href="#">Delete</a>	
My dataset	6/05/2012	pending	<a href="#">View/Edit</a>	<a href="#">Delete</a>	

[Register New Data Set](#)



# Cross-Domain Re-use

*Project proposed as a “Data Infrastructure Building Block (DIBB)”:*

*“Managing community input to support cross-domain interoperability and fitness-for-use assessment”, NSF Proposal, August 2012*

*Lead-PI: Ilya Zaslavsky, SDSC*