

Building transdisciplinary infrastructure for natural history material samples with the Internet of Samples (iSamples)

10 June 2020

SPNHC Virtual General Symposium 1

Ramona Walls, University of Arizona



Neil Davies, Gump Research Station, UC Berkeley



Sarah Wicheter Kansa, Alexandria Archive Institute,
Open Context



John Kunze, California Digital Library



Kerstin Lehnert, Lemont Doherty Earth Observatory,
Columbia University



David Vieglaiss, University of Kansas



Raymond Yee, Open Context

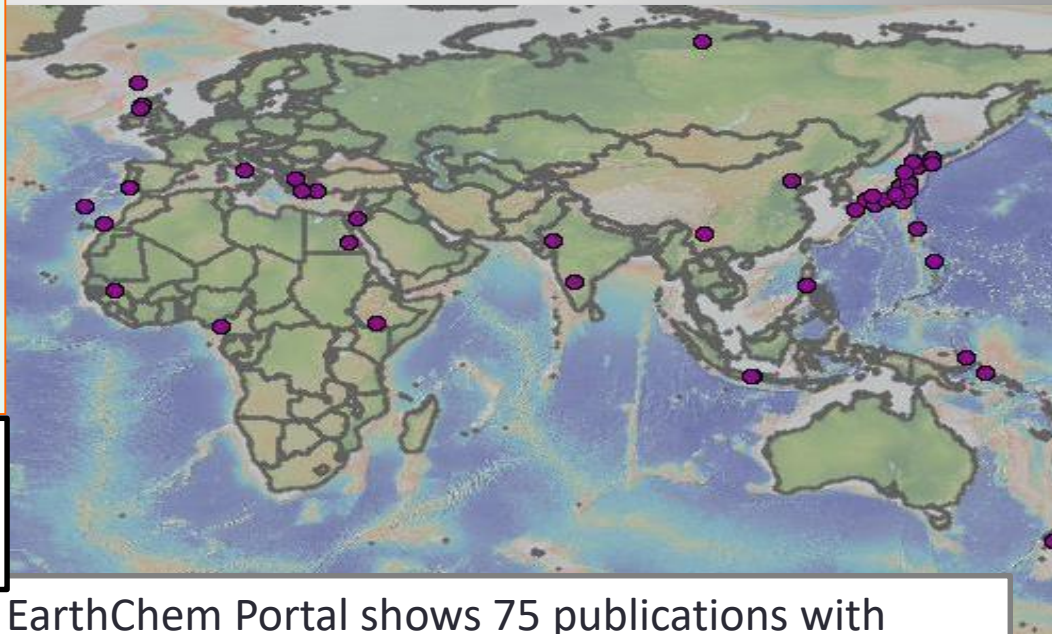
Why a global system for specimen identifiers and data? The Problem of Sample Names

PetDB Identifier: ARGAMPH-003

IGSN: N/A

AMPH D-3(SUN, 1980)
D3(ENGEL, 1964)
PD3(TATSUMOTO, 1965)
PD3(TATSUMOTO, 1966)
AMPH-D3(MACDOUGALL, 1986)
AMPH D-3(SCHILLING, 1975)
S-10(SUBBARAO, 1972)
Other Names: PV D-3(ENGEL, 1965)
AMPH-3D(PINEAU, 1983)
AMPH 3-PD3(HART, 1971)
PD-3(HEDGE, 1970)
PD-3(MUEHLENBACH, 1972)
AMPH3D(PINEAU, 1976)
D-3(SCHILLING, 1971)
D-3(SCHEIDEGGER, 1981)

Names of dredge sample 3 of
the Amphitrite cruise
(PetDB database, www.petdb.org)



The EarthChem Portal shows 75 publications with geochemical data referenced to a sample with the name M1 (or M-1).

<http://www.earthchem.org>

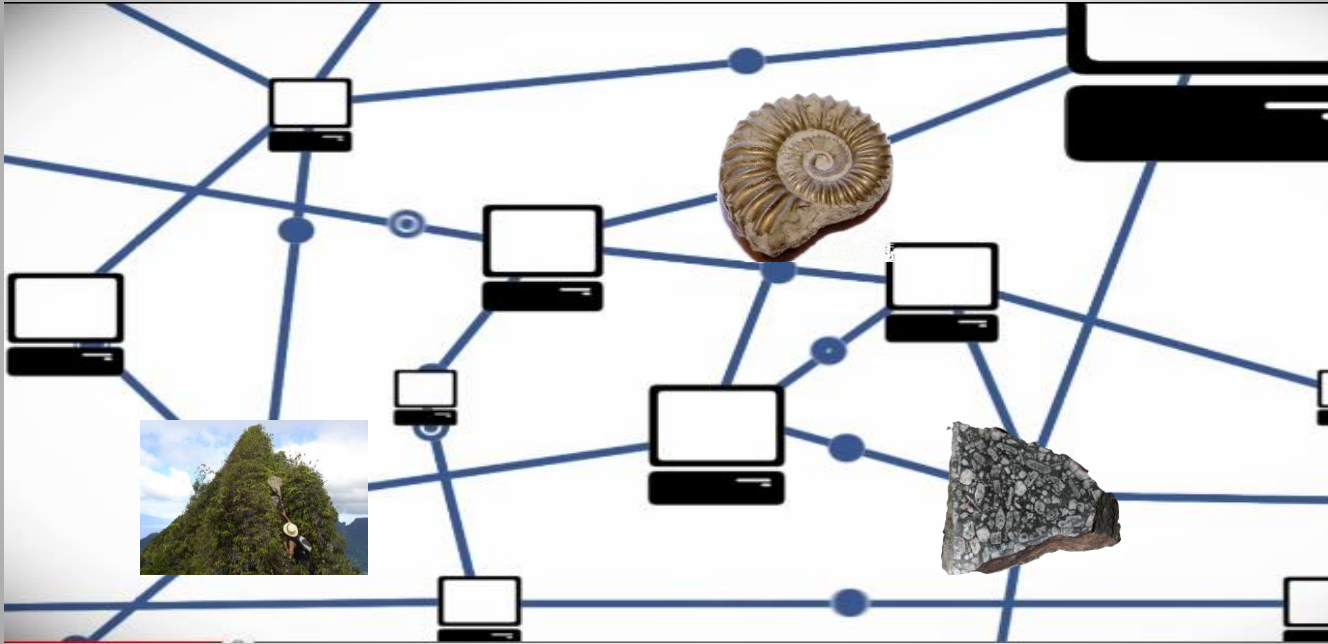
Why a global system for specimen identifiers
and data?

Specimen data are not FAIR

Why a global system for specimen identifiers?



Why a global system for specimen identifiers and data?



Building a “Web of Samples” requires that specimens have permanent, unique, resolvable identifiers, and that those identifiers be use.

The state of specimen IDs in natural history collections

- Solutions needed for both existing and new specimens/collections
 - born digital specimen records are now possible
 - but no standard exists for identifiers and metadata
- Many types of identifiers in use
 - Local identifiers minted by many collections, some can be made globally unique
 - CETAF IDs
 - IGSNs being used for soil samples, plants
 - DOIs
 - ARKs

The state of specimen IDs in biodiversity

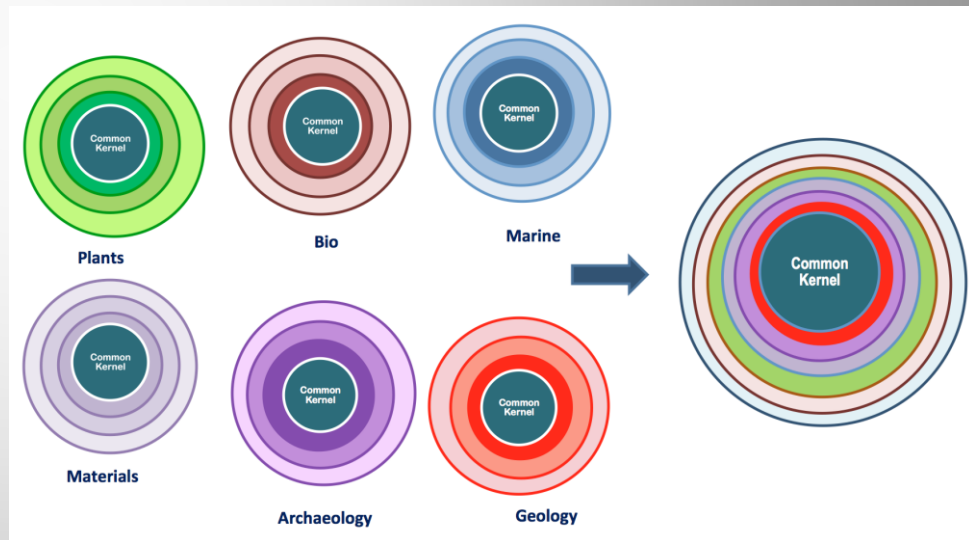
GBIF occurrence	Identifier type	Identifier	Catalog number	Collection
872747863	LSID	urn:lsid:biosci.ohio-state.edu:osuc_occurrences:OSUC__169968	OSUC 169968	C.A. Triplehorn Insect Collection
896421698	URN	urn:occurrence:Arctos:MVZ:Bird:157675:1526959	MVZ 157675	MVZ Bird Collection
784060956	URN	urn:catalog:UMMZ:Mammals:171041	UMMZ 71041	UMMZ Mammal Collection
575336458	HTTP URI	http://data.rbge.org.uk/herb/E00115694	E00115694	Royal Botanic Garden Edinburgh Herbarium
1050474791	HTTP URI	http://arctos.database.museum/guid/UAM:Ento:230092	UAM 230092	UAM Entomology Collection
1050474791	DOI	10.7299/X7VQ32SJ	UAM 230092	UAM Entomology Collection
624211191	UUID	EF0A4D3E-702F-4882-81B8-CA737AEB7B28	UF 161444	UF FLMNH Ichthyology
476850316	Darwin Core Triplet	MCZ:Mamm:8831	MCZ 8831	MCZ, Harvard University

The state of specimen IDs in natural history collections

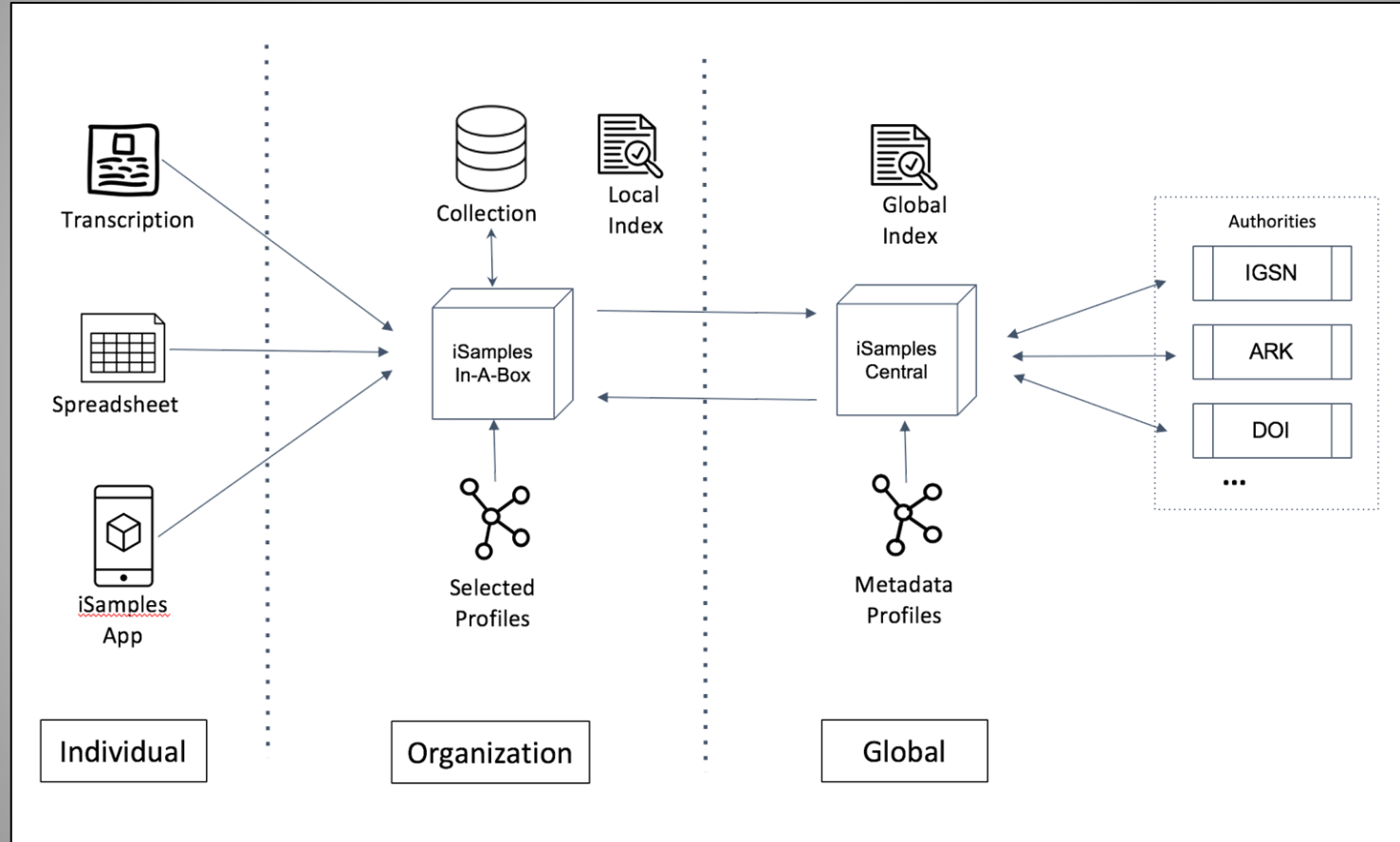
- Biology:
 - DwC Material sample core – never published in IPT, still *Occurrence* based
 - INSDC BioSamples database
 - Allows material sample ID from Darwin Core/MlxS; no DataCite metadata
- Agriculture:
 - FAO issues free DOI for plant genetic resources for food and ag (PGRFA)
 - <http://www.fao.org/plant-treaty/areas-of-work/global-information-system/doi/en/>
- Archeology:
 - Open Context mints stable URIs for each observed artifact, ecofact, sample, and context. ~ 1.6 million indexed specimen records, plan to use IGSN
- Material Science:
 - NIST High Throughput Experimental Materials Science Registry and Repository
 - <https://mgi.nist.gov/hitemc>
- Geology/Earth and Space Science
 - IGSNs used for mineral samples of all sorts

A way forward

- Baseline need for stable, resolvable GUIDs for specimens, collections, locations (sampling features)
- Recognition of existing systems and stakeholder needs
 - A federated approach
- Cross-disciplinary metadata standards
 - Core elements for all specimens
 - Shared elements for closely related disciplines
 - Discipline specific profiles



iSamples Architecture



Initial iSamples implementations

- Open Context
 - Archaeology, IGSNs
- Genomic Observatories Metadata Database (GEOME)
 - Biodiversity, ARKs
- System for Earth Sample Registration (SESAR)
 - GeoSciences, IGSNs
- Smithsonian
 - Biodiversity, ARKs; Geosamples, IGSNs
- USGS
 - GeoSciences, IGSNs

Role of SPNHC and ICOM NATHIST communities

- Input on interdisciplinary standards
- Advice on governance
- Possible new implementations