

# Cultural Heritage and Research Information: Case 3

Kevin C. Holston  
Swedish Natural History Museum

2014.06.04

**DCH-RP, EUDAT**  
**Digital Preservation: Workshop 2**

## **Natural History Collections – Objects as Samples.**

Representative specimens from (the natural world)  
environmental and evolutionary processes

- from a fauna or flora, geographic region
- from a population of individuals
- from a geological formation

Reference material, vouchers

Increased sampling, better collections/datasets

# **Historical Perspective**

## **Natural History Collections – Objects as Samples.**

Representative specimens from (the natural world)  
environmental and evolutionary processes

Efforts from individual researchers, scientific communities,  
and institutions to improve knowledge on these processes

Unique perspective from public natural history collections  
Historical  
Reassessment (detailed assessment)

# **Historical Perspective**

## **Conflict between curation and research.**

### Collection Inventories

Physical placement, condition – static nomenclature

Focus on summaries on collections

### Biodiversity Informatics Datasets

Scientific observations – dynamic nomenclature

Focus on summaries on trans-collection record sets

### Enrichment of datasets

Secondary data (voucher associations, locality and specimen characterizations, recognition of collectors)

# Historical Perspective

## **Conflicts resolved by new procedures and technological advances.**

Persistent identifiers, (globally) unique

Specimen-level (collection object) registration

Industrial techniques rather than new technology

Data management systems (not-so-big data)

User interfaces to accommodate multiple, novel workflows

Standards and exchange schemas

# **Historical Perspective**

## Data + “Metadata”

Collection Objects

Collection Metadata

Taxonomic  
Names

Storage



Taxonomy and physical placement in the collection is handled using the same mechanisms (schema)

# Dataset Integration – Data Sharing

## **Common ground between curation and research.**

Collection Inventories and Biodiversity Informatics Datasets

Primary data – from analog to digital form

Persistent identifiers, (globally) unique, normalization

Secondary data

Limited resources, difficult to stabilize criteria used to prioritize data capture (which collection objects, which characterizations?)

# **Historical Perspective**

# **Classifications are organizing principles.**

Collection Inventories and Biodiversity Informatics Datasets

Authority files

- Determinations (taxonomic names)

- Georeferences (place names)

- Geological age, lithography, stratigraphy

Access and application rather than local management

## **Useful Assumptions**



## Database records are observations.

Primary data – from analog to digital form

Collector, locality (site), date

*Determinations (taxonomic names), type status*  
*Images (figures)*

Secondary data

Descriptions – do not change over time (e.g, specimen measurements, life stage and sexual maturity)

Voucher associations (e.g., DNA (consensus) sequences)

# Useful Assumptions

# Observation Categories and Data Overlap

## RECORD CONTENT, CIRCUMSCRIPTION

preparation, details  
storage details (transactions)

counts (total within storage unit,  
"actual" collection object totals)

sets: treatment, project, lab  
experiment or sample preparation

collection object, biological attributes  
locality characterization (attributes)

counts (total in field [sample],  
population survey totals)

sets: expedition, field experiment  
(e.g., sampling along transect)

### DATABASE OBJECT DATA

### BIOLOGICAL OBSERVATION DATA

inst_1	storage_1	<b>OBS_0001</b>	<b>Taxon Det.</b>	uniq_obs_1	collevent_1	locality_1
inst_1	storage_1	<b>OBS_0002</b>	<b>Taxon Det.</b>	uniq_obs_2	collevent_2	locality_2
inst_1	storage_1	<b>OBS_0003</b>	<b>Taxon Det.</b>	uniq_obs_3	collevent_3	locality_3
inst_1	storage_1	<b>PAL_0001</b>	<b>Taxon Det.</b>	uniq_obs_4	paleoev_1	paleoev_1
inst_1	storage_1	<b>MUS_0001</b>	<b>Taxon Det.</b>	uniq_obs_5	collevent_4	locality_1
inst_1	storage_1	<b>CITA_0001</b>	<b>Taxon Det.</b>	uniq_obs_5	collevent_4	locality_1
inst_1	storage_1	<b>CITA_0002</b>	<b>Taxon Det.</b>	uniq_obs_5	collevent_4	locality_1
inst_1	storage_1	<b>CITA_0003</b>	<b>Taxon Det.</b>	uniq_obs_5	collevent_4	locality_1
inst_1	storage_2	<b>LAB_0001</b>	<b>Taxon Det.</b>	uniq_obs_5	collevent_4	locality_1
inst_2	storage_3	<b>LAB_0002</b>	<b>Taxon Det.</b>	uniq_obs_5	collevent_4	locality_1
inst_2	storage_4	<b>LAB_0003</b>	<b>Taxon Det.</b>	uniq_obs_5	collevent_4	locality_1

**There are benefits from incorporating new procedures at natural history collections.**

Persistent identifiers, normalization, collection object registration

- effective, periodic summaries
- application of data harvesting tools
- engagement of current digitization “personnel”
- data sharing (public) “dissemination”

**Useful Assumptions**

## **Workflows. Imperatives for a Repository or Laboratory.**

**Repository:** Manage/organize the collections!

Summarize holdings, maintain condition (high quality)


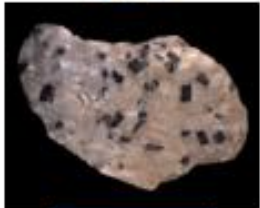




*Taxonomic Determinations*

**Laboratory:** Characterize the collections!

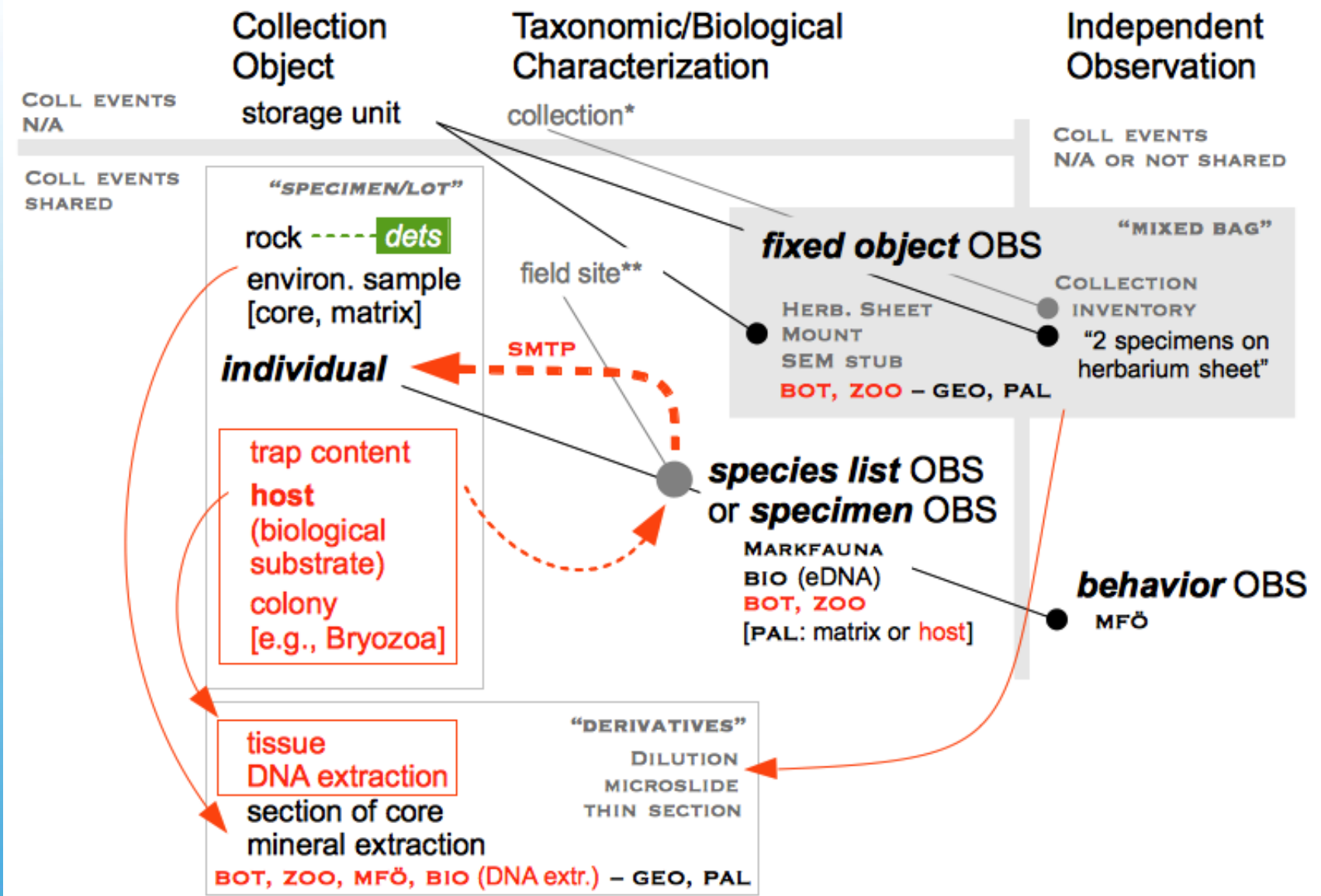
Create derivative objects from collected objects and associate data (as processed datasets)

**DINA project case studies**

# Collection object relationships – containers, lots and mixed bags. Batch, composite and derivative objects.

CONTAINER, GEOLOGICAL SAMPLE		FIXED OBJECT		BATCH		COMPOSITE	
storage unit	rock	environ. sample	matrix	trap content	host		
							
<b>CO</b>	<b>CO</b>	<b>CO</b>	<b>CO</b>	<b>CO</b>	<b>CO</b>		
descr.	det	descr.	det	det	det		
<b>OBS</b>	det	<b>OBS</b>	<b>OBS</b>	<b>OBS</b>		<b>OBS</b>	
<b>OBS</b>	det	<b>OBS</b>	<b>OBS</b>	<b>OBS</b>			
<b>OBS</b>	det	<b>OBS</b>	<b>OBS</b>	<b>OBS</b>			

# Collection Object Relationships



# Collection Object Relationships

## Workflows.

**Repository:** Manage/organize the collections!

Summarize holdings, maintain condition (high quality)

*Taxonomic Determinations*

Svenska insektfaunaarkivet (Swedish Malaise Trap Project)

faunal inventory project

2003-2006 field sampling

2007-present, sample sorting

material for researchers to examine and analyze

# Workflow case studies

**Svenska insektfaunaarkivet**



## Workflows.

**Laboratory:** Characterize the collections!

Create derivative objects from collected objects and associate data (as processed datasets)

*(Taxonomic Determinations) Voucher*

Department of Bioinformatics and Genetics, NRM  
(Molecular Systematic Laboratory)

DNA extractions (derived objects), bioinformatics services to researchers (equipment, technical expertise) for analysis results

# Workflow case studies

# Department of Bioinformatics and Genetics

## Workflows.

**Laboratory:** Characterize the collections!

Create derivative objects from collected objects and associate data (as processed datasets)

*Taxonomic Determinations*

Department of Geosciences (Laboratory for Isotope Geology)

subsamples of rocks and water

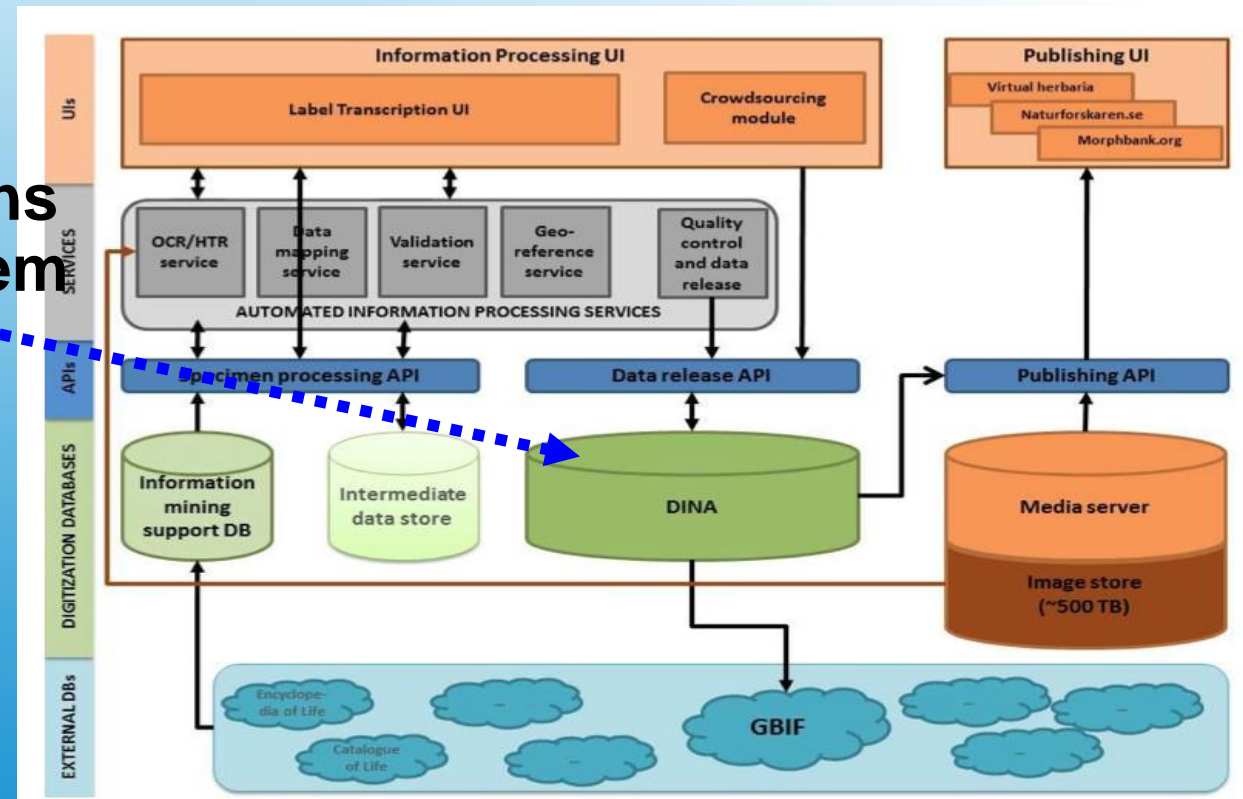
services to researchers (equipment, technical expertise) for analysis results

# Workflow case studies

**Department of Geosciences**

- Further workflow complexity for systems interoperability and data accessibility

## National Collections Management System



# Workflow case studies

## **Workflows. Participation of Diverse Agents.**

Curators, technicians, researchers (students, guests, staff), volunteers

The Crowd. (*quality issues*)

Programmers, database administrators, business managers. (*control issues*) – *The Cloud*

# **Workflow case studies**

## Summary

Collection objects as samples, collections as the result of research-oriented sampling of biodiversity and the environment.

*cultural heritage historical archive for research*

Workflows intended to support further characterization of collection objects. *repository – laboratory*

Participation of collaborating groups of agents as co-creators of digitized objects and records (metadata).

*understanding evolutionary and environmental processes*

**Thank you for your attention.**

- DINA, for infrastructure and development

**DINA (digital information system for natural history collections) is a project aiming to develop and implement a national database system for collection management.**

Museum of Evolution, Uppsala University

Herbarium GB, Gothenburg University

Gothenburg Natural History Museum

Svenska insektfaunaarkivet, Station Linné, Ölands Skogsby

Zoological Museum, Lund University

Swedish Museum of Natural History

# Historical Perspective



- DINA, for infrastructure and development

**DINA (digital information system for natural history collections) is a project aiming to develop and implement a national database system for collection management.**

**The Swedish Species Service (ArtDatabanken)  
2013-2016**

<http://specifysoftware.org>

**Historical Perspective**

- **DINA project**

Per Ericson	Principal Investigator
Fredrik Ronquist	Principal Investigator
Karin Karlsson	Project Leader
Ida Li	Software Developer
Marcus Englund	Systems Analyst
Kevin Holston	Systems Analyst, Database Admin

**International Consortium** – open source development  
Denmark, Germany (MfN), Estonia (Tartu, Pluto-F),  
Agriculture and Agri-food Canada, Ottawa  
Harvard University Herbaria, Boston  
Kansas and Royal Botanical Garden, Edinburgh.

## Historical Perspective