Managing Digital Assets at SI: Current, Planned, and Aspirational Approaches

...

Ian Owens, Deputy Director, NMNH
Rebecca Snyder, Acting Informatics Branch Chief, NMNH

Smithsonian at a Glance

19 museums

9 research centers

1 zoo

155.5M specimens and objects *(32M with digital record, 5M with image)*

2.2M library holdings *(1.5 with digital record, 50K with image)*

Category	Units	Objects and Specimens	Archives Cubic Feet	Library Volumes
Grand Total		155,491,416	163,261	2,214,710
Science	Total	146,232,429	19,108	
	NMNH - Natural History	146,221,207	18,213	
	SG - Gardens	9,297	895	
	NZP - National Zoo	1,925		
History and Culture	Total	8,913,179	86,823	
	NPM - Postal	6,066,255		
	NMAH - American History	1,883,570	16,680	
	NMAI - American Indian	848,584	3,173	
	NASM - Air and Space	71,581	17,953	
	NMAAHC - African American	36,253	611	
	SIB - SI Castle	3,524		
	ACM - Anacostia	3,412	1,523	
	CFCH - Folklife		4,981	
	SIA - SI Archives		41,902	
Art	Total	345,808	46,896	
	CHM - Cooper Hewitt	215,769		
	SAAM - American Art	44,021	19,970	
	FSGA - Freer | Sackler	42,025	2,123	
	NPG - Portrait Gallery	21,369	1,374	
	HMSG - Hirshhorn	12,222		
	NMAFA - African Art	10,402	1,168	
	AAA - Archives American Art		22,261	
Libraries	Total		10,434	2,214,710
	SIL - SI Libraries		10,434	2,214,710

Primary Data Types

SI has 5 main data types:

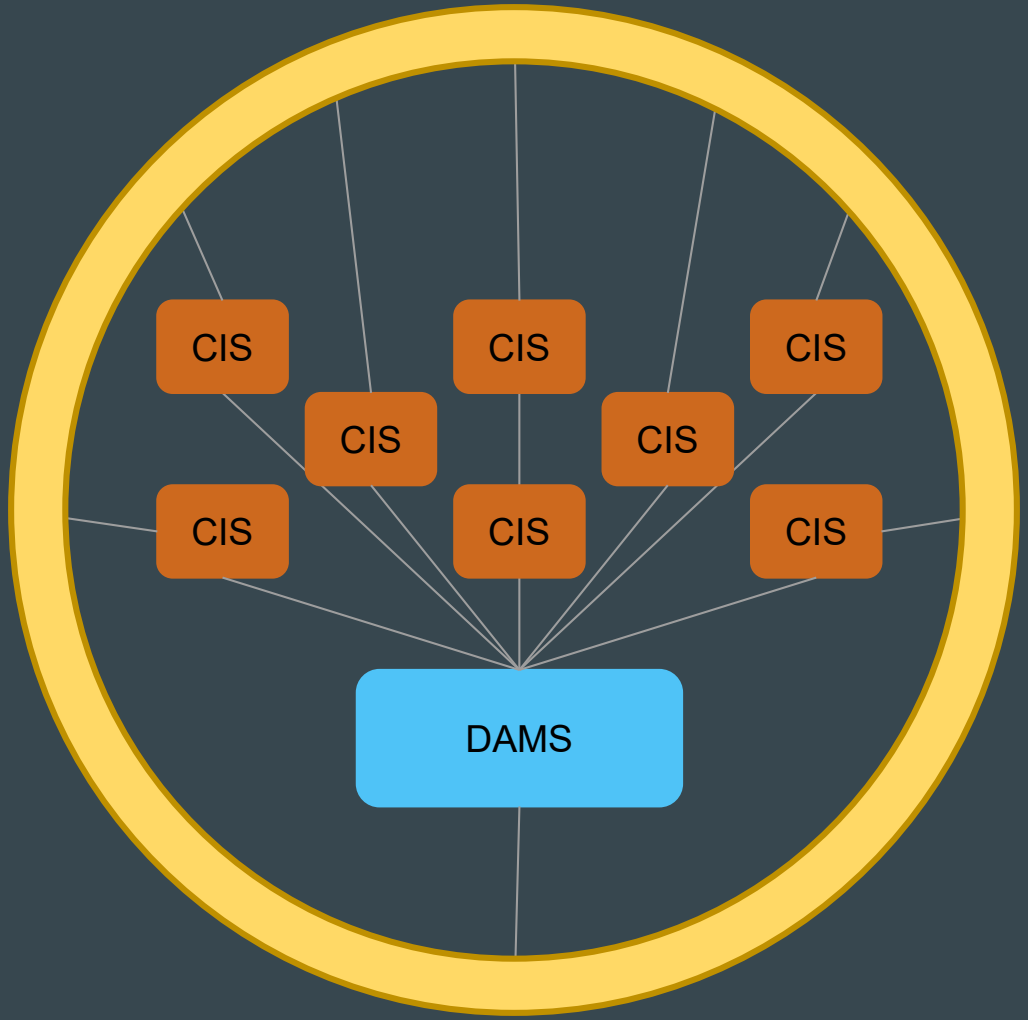
- Descriptive records (metadata of collections, library holding, archival finding aids, etc.)
- Standard multimedia (binary files of images, video, audio, text, etc.)
- Complex data (assets with complex file structures and interrelationships, like 3D, CT, instrument data, etc.)
- Genomic (sequences, eDNA, etc)
- Astronomical (telescope, satellite)

Current

8 Collection Information Systems
(30+ implementations)

1 DAMS

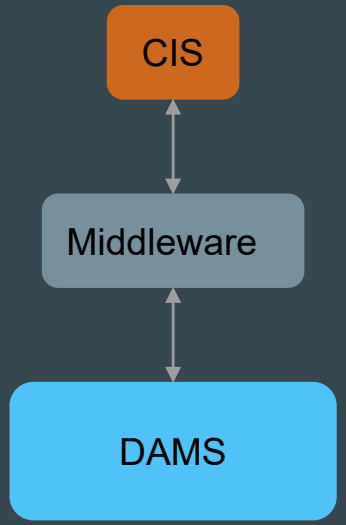
…and one SI Index "EDAN" to connect them all



Collections Data and Standard Multimedia

Most SI Units have a dedicated CIS

- primary focus on collections data
- Most link to central SI DAMS
- DAMS is primary system of record for standard multimedia (images, video, audio, text)
- Middleware for ease of managing +30 CISs connecting to DAMS



Complex Data

Medical CT data on HD and SAN

MicroCT on SAN

Most instruments give data directly to the researcher and SI systems never see it

As of Dec 2019, SI has Figshare for sharing static research data

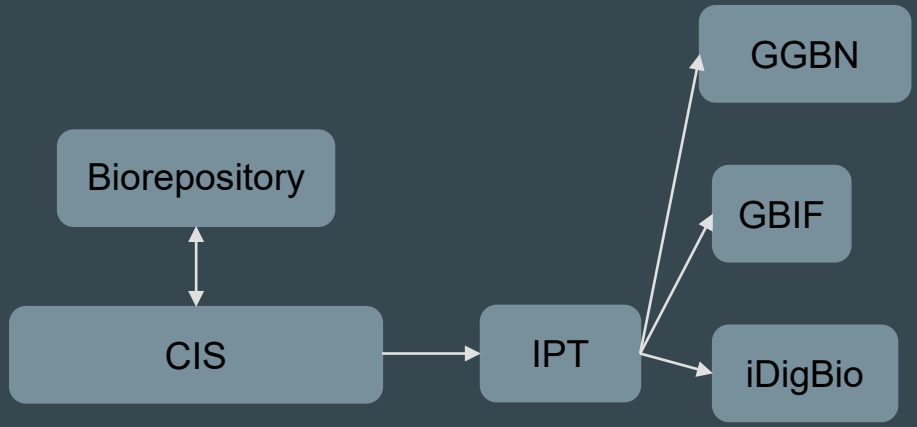
No official mandate to manage, no official system of record

Genomics Workflows

Some genomic data is loaded into Biorepository that has limited connection to CIS, GUIDs are assigned

This data is shared via IPT

Other data is loaded directly by researchers into external repositories, not always with SI GUIDs - difficult to link back

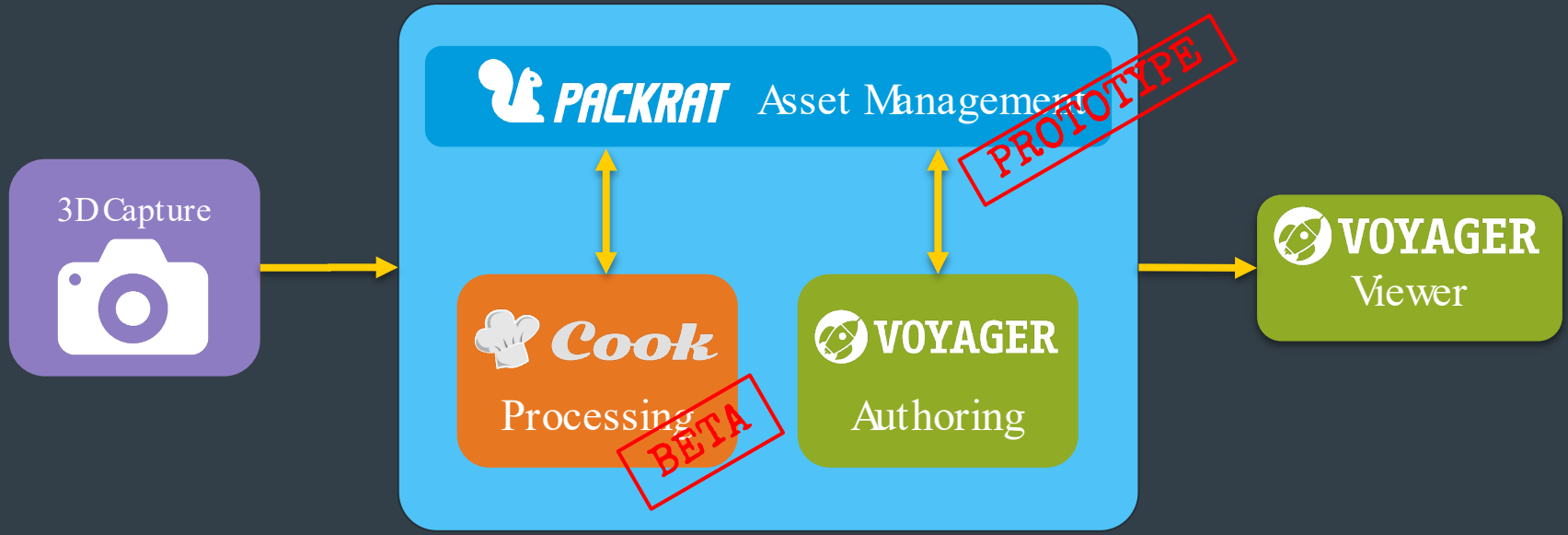


Planned (aka: funded and in progress)

There is a plan for complex data!

Smithsonian 3D Pipeline Overview

Smithsonian
3D DIGITIZATION



Submission Information Package (SIP)

Aspirational (Rebecca's hopes and dreams - aka not funded)

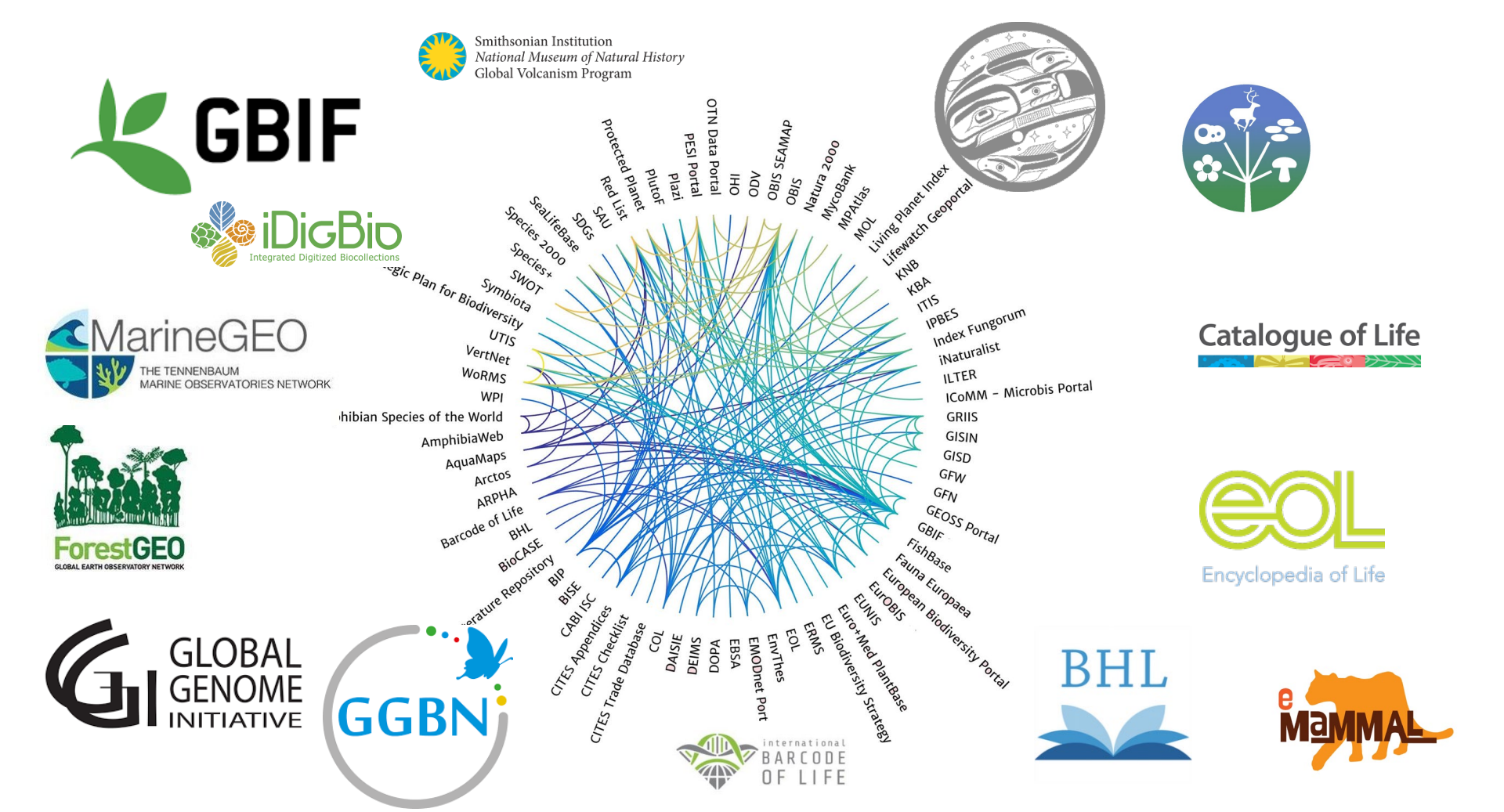
SI goes all in on digital preservation and stewardship

- New central SI office for Digital, akin to our National Collections Program, whose job is to coordinate all digital systems and projects as well as facilitate a cohesive strategy for digital preservation and stewardship
- Packrat goes enterprise and integrates with SI and Unit systems of record and dataflows (DAMS, CISs, EDAN, etc. using our newly implemented SI arks.
- SI attempts for all digital management systems to achieve Trusted Digital Repository certifications

Aspirational (Rebecca's hopes and dreams - aka not funded)

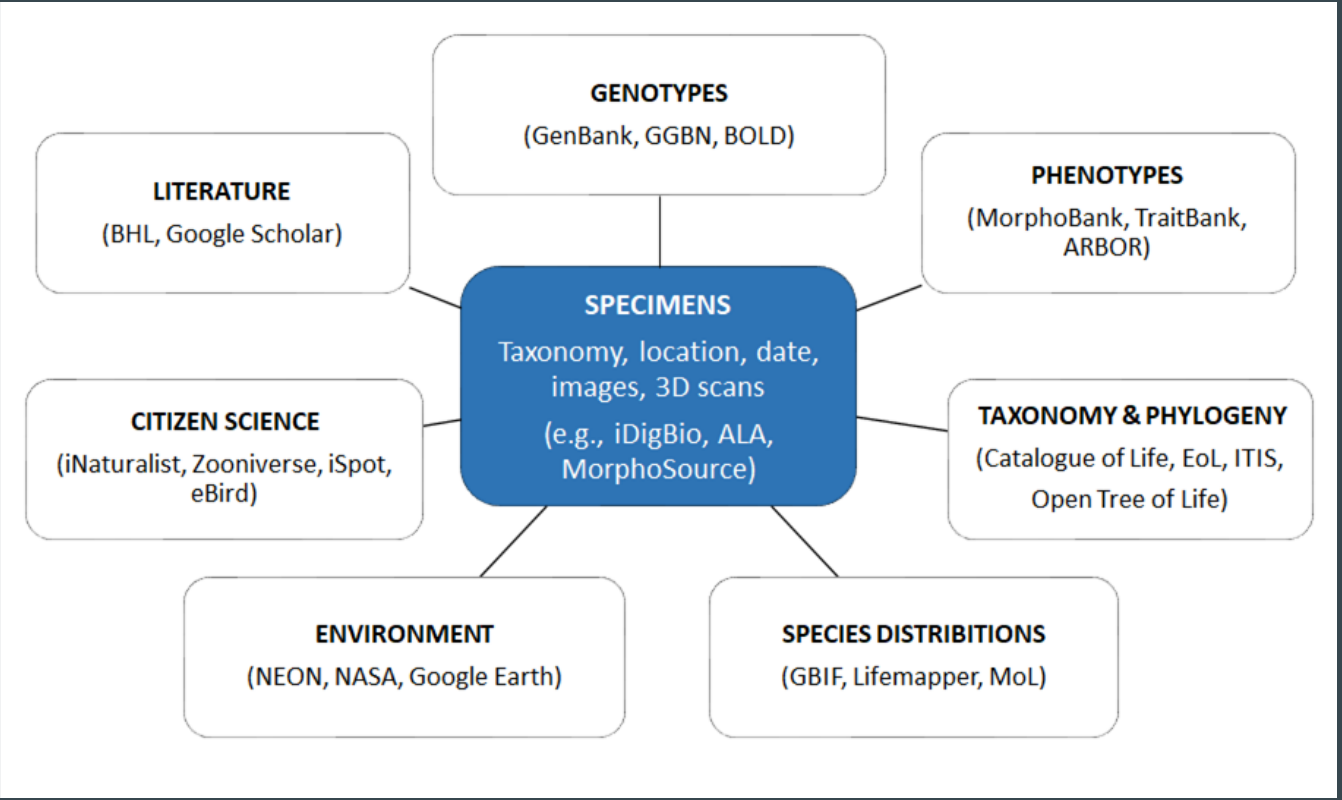
SI leverages our robust infrastructure to build knowledge graphs

- SI shares all data globally in an interconnected and cross referenced way, linking all SI collections with other external knowledge repositories and centralized authorities
- With everyone using persistent and resolvable identifiers and playing FAIR, the potential to connect "ALL THE THINGS" is limitless



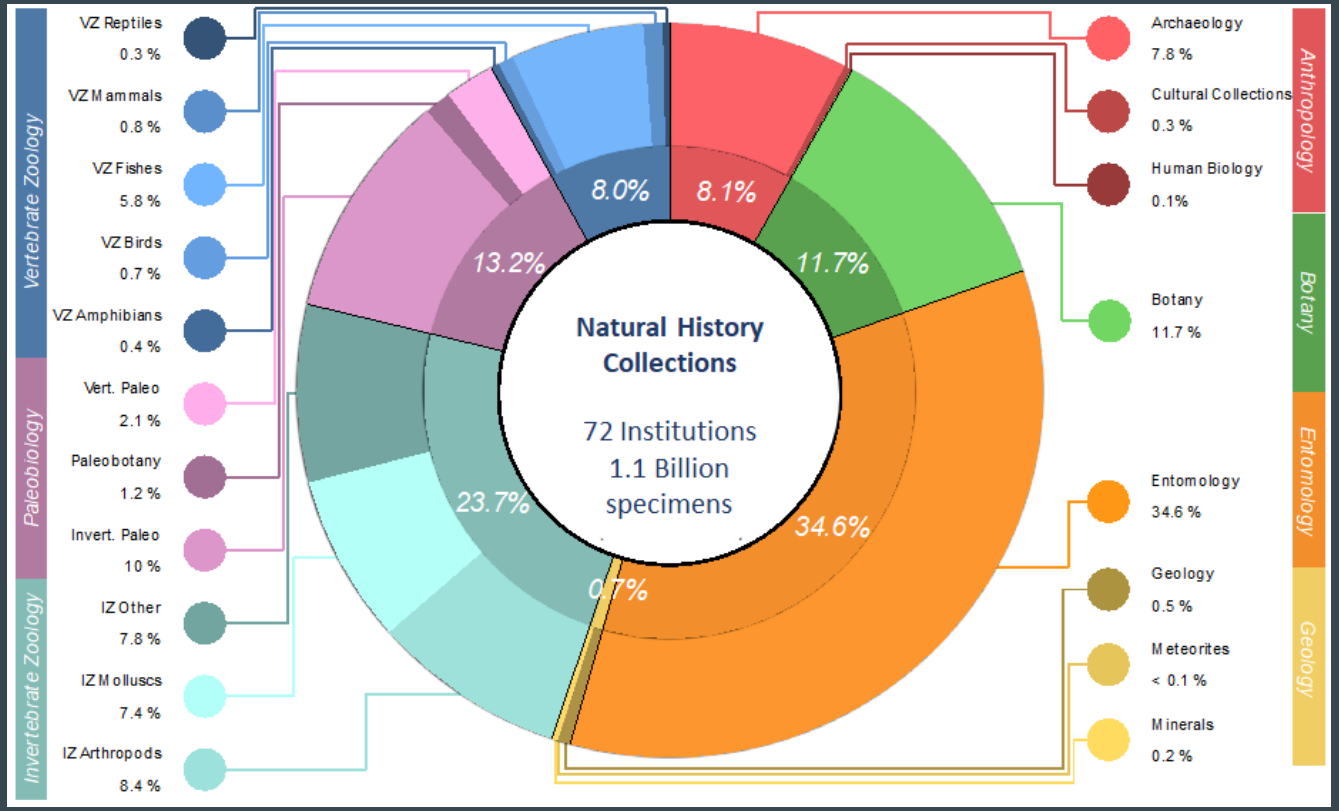
Redrawn from: Bingham H, Et al., 2017. The Biodiversity Informatics Landscape: Elements, Connections and Opportunities. Research Ideas and Outcomes 3: e14059.

Schematic illustration of a common knowledge platform for natural history specimens

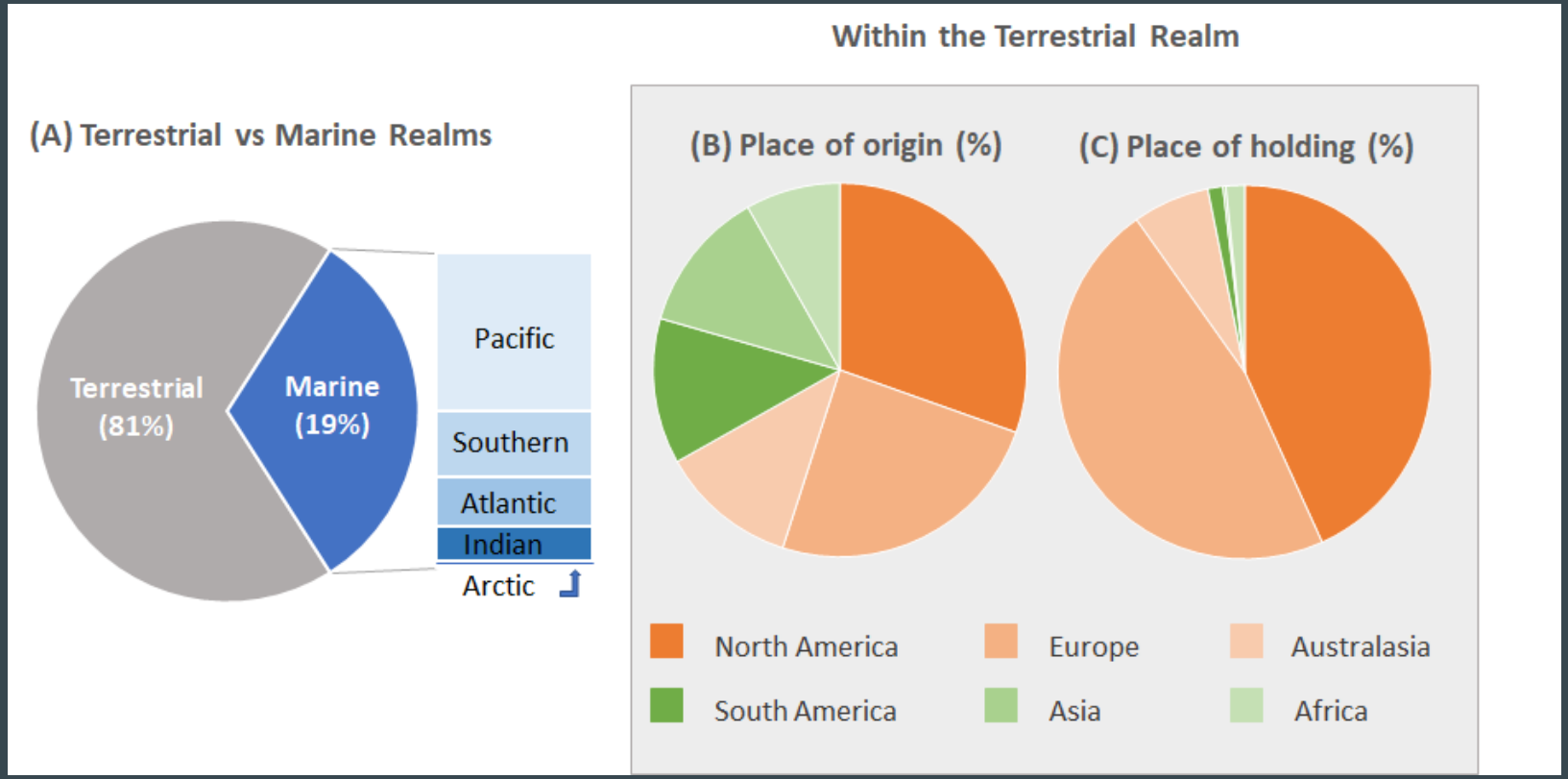


Geographic distribution of participating institutions

Systematic composition of the collections across participating institutions



Geographical distribution of specimens



Questions?

Incorporating Digital Preservation into our Workflows