Smithsonian
National Museum of Natural History

Undertaking digitization in the NMNH Paleobiology collection and digitization strategies at the Smithsonian Institution

Kathy Hollis
Collections Manager
NMNH Department of Paleobiology
hollisk@si.edu

Conceptualizing Paleobiology

You can't manage what you don't measure

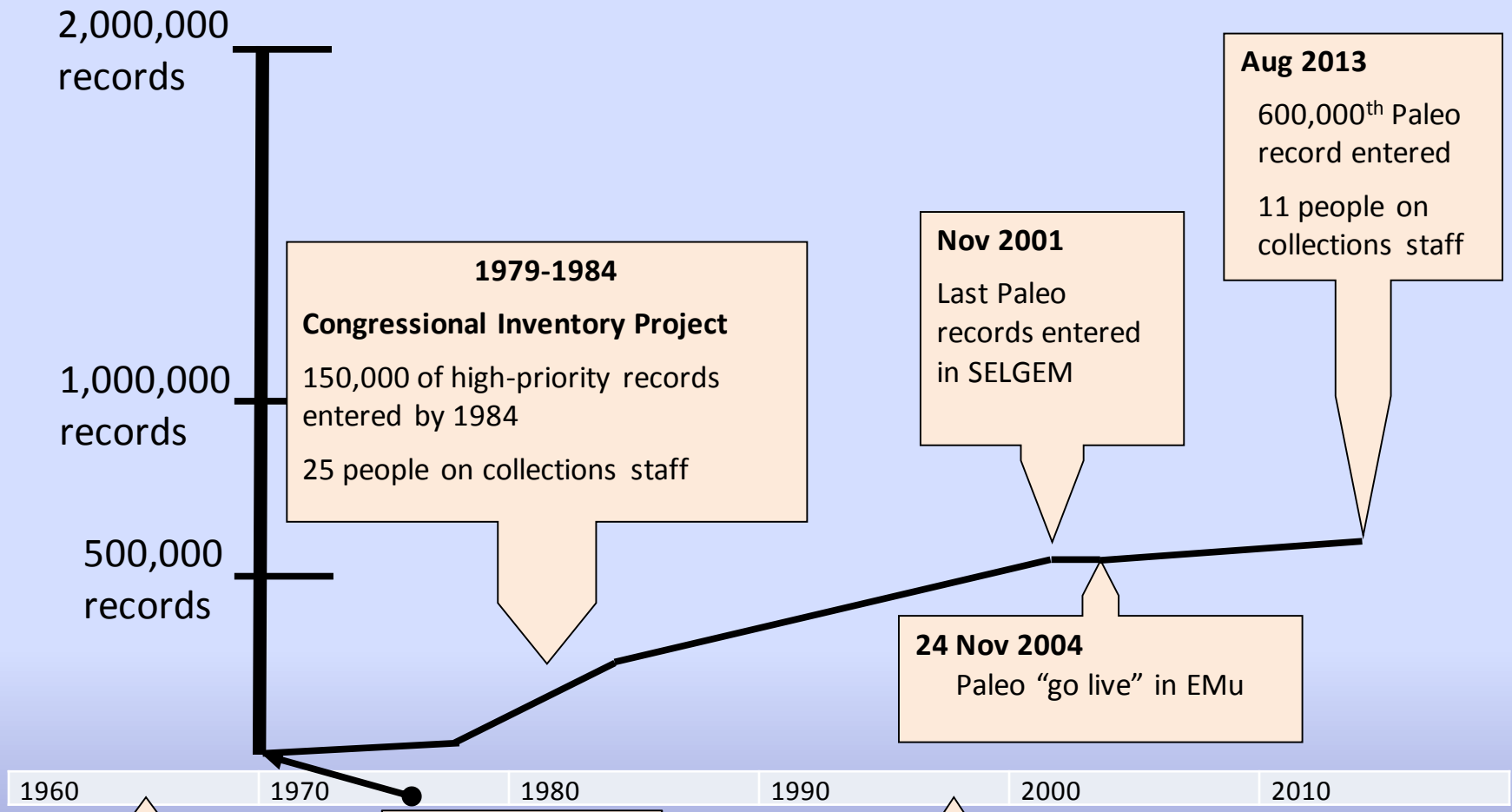
Estimated 40 million individual items

Estimated 2 million Records

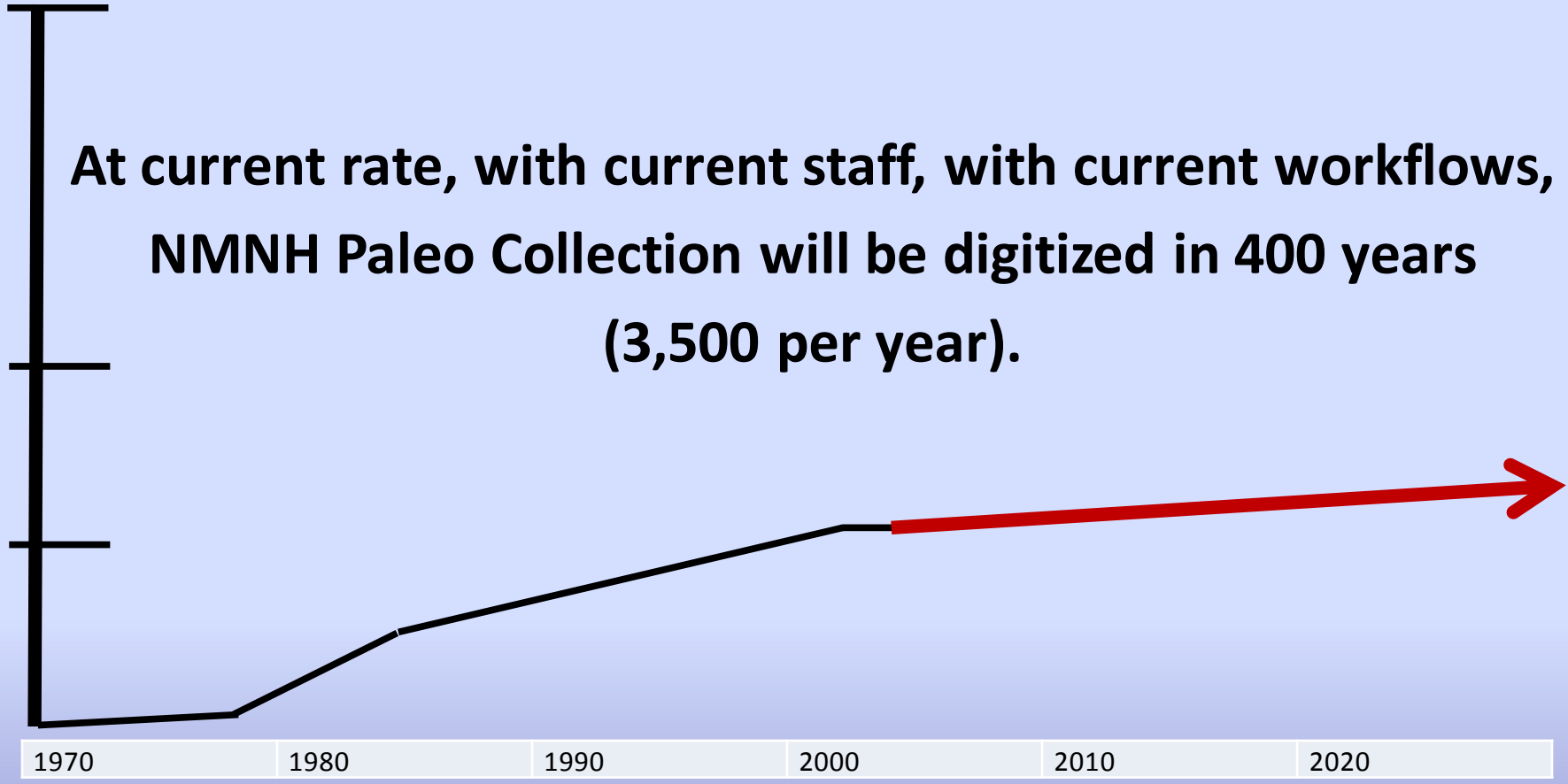
490 Collections

58 Collection Groups

History of digitization at NMNH Paleo



History of digitization at NMNH Paleo

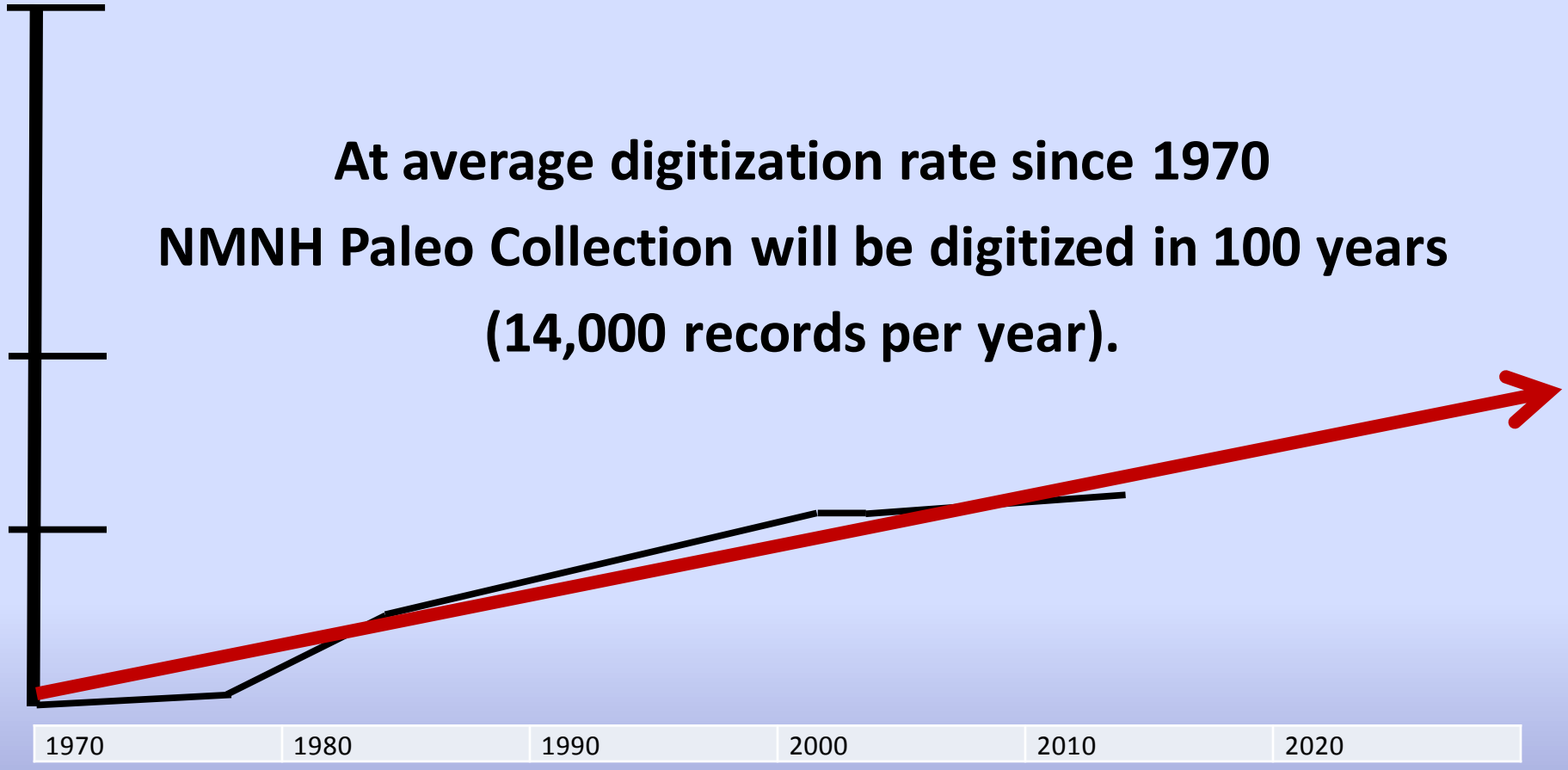


History of digitization at NMNH Paleo

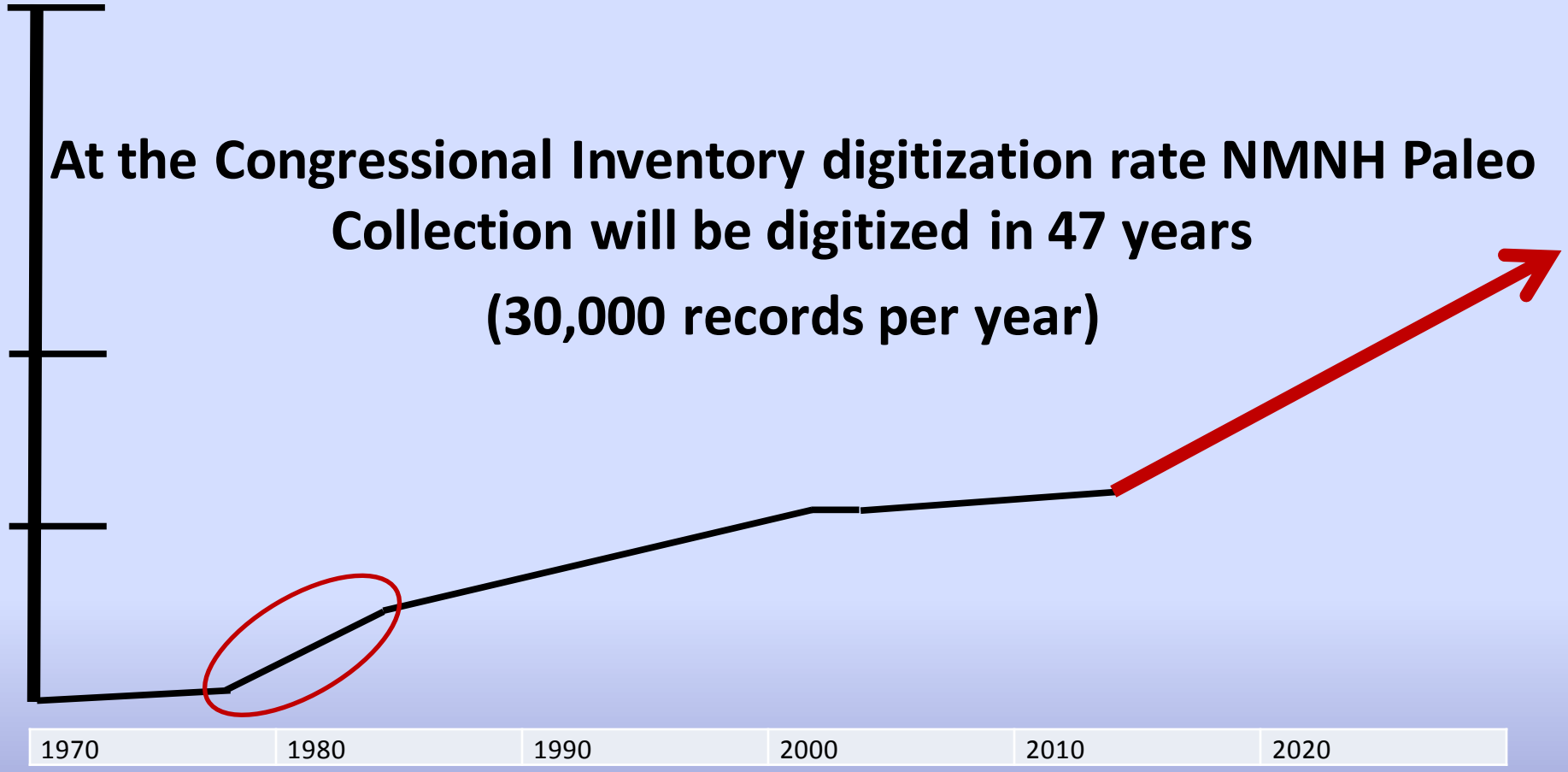
At average digitization rate since 1970

NMNH Paleo Collection will be digitized in 100 years

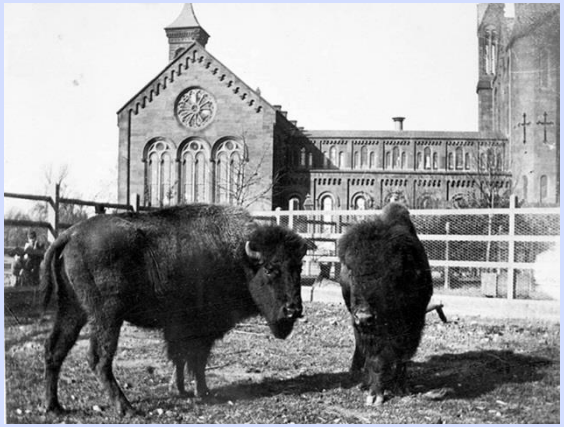
(14,000 records per year).



History of digitization at NMNH Paleo



History of digitization at NMNH Paleo



The Digitization Frontier

Digital records of variable quality

No strategic plan to digitization

The Age of Standards and Planning

Digital Records meet or exceed data quality standards

Digitization projects must align with strategic priorities

As of 2011, SI is requiring a systematic, priority-driven, centrally managed approach.

Contemporary Digitization at SI

Step 1. Define standards for record quality

Emu Record Quality
Rank 1 **NOT DIGITIZED**
Rank 2 **One or more required fields are missing data**
Rank 3 **All required fields are complete**
Rank 4 **All Rank 3 plus if type, HVCI, or dinosaur then has an image**
Rank 5 **All Rank 4 plus properly georeferenced**

Does not meet NMNH Paleo Standards and needs more data to be scientifically useful

Meets or exceeds NMNH Paleo Standards and are minimally scientifically useful

Fields are mapped in Emu to Darwin Core Fields and are auto populated

Contemporary Digitization at SI

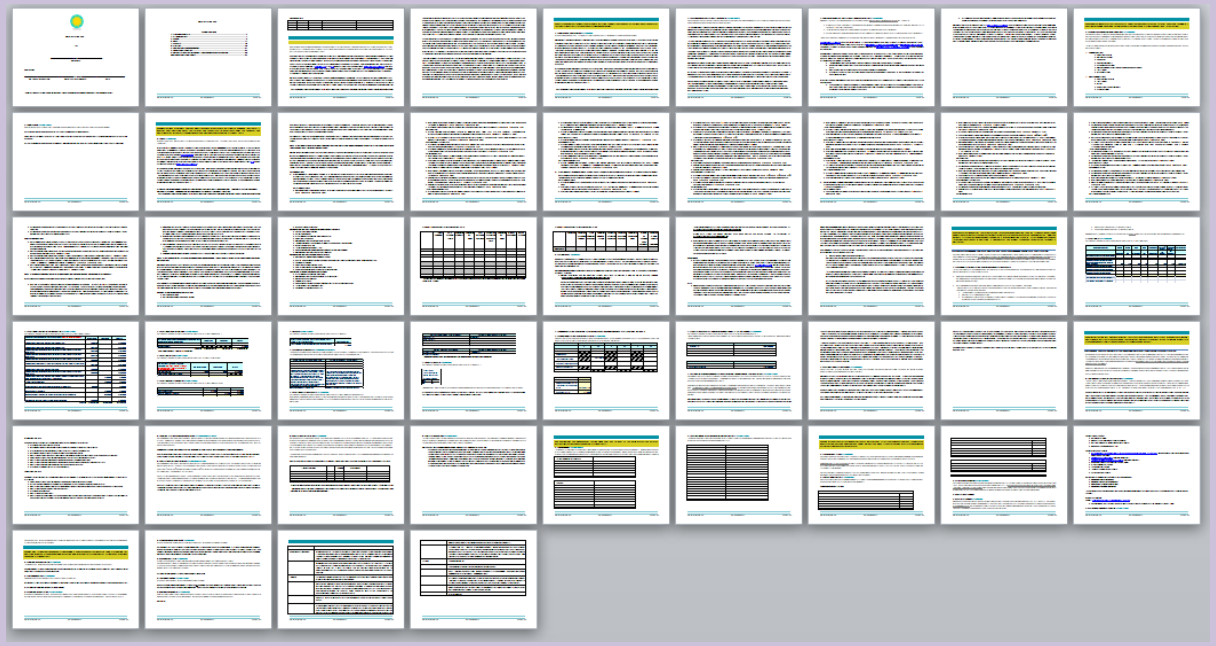
Step 2. Gather Digitization Statistics

Emu Record Quality	Total Emu Records as of 2013	Total Item Count (somewhat meaningless)
Rank 1 **NOT DIGITIZED**	1,280,578	36,508,960
Rank 2 **One or more required fields are missing data**	156,229	392,135
Rank 3 **All required fields are complete**	443,773	7,185,593
Rank 4 **All Rank 3 plus if type, HVCI, or dinosaur then has an image**	8,277	20,775
Rank 5 **All Rank 4 plus properly georeferenced**	0	0
	1,888,857	44,107,463

Contemporary Digitization at SI

Step 3. Set Priorities and Goals

At the Smithsonian, priorities and goals are formalized in Director-approved 3-Year Digitization Plans and are managed through the Digitization Program Office.



Contemporary Digitization at SI

Step 3. Set Priorities and Goals

At the Smithsonian, priorities and goals are formalized in Director-approved 3-Year Digitization Plans and are managed through the Digitization Program Office.

A written digitization plan provides:

1. **Unified vision for digitization activities**
2. **Digitization history**
3. **Alignment with Strategic Plan**
4. **Digitization scope, goals, and priorities**
5. **Capacity and resources**
6. **Collaborators and dependencies (e.g. funding sources)**
7. **Policies, standards, and legal restrictions (e.g. copyright)**
8. **Roles and responsibilities.**

Contemporary Digitization at SI

Step 3. Set Priorities and Goals - **Paleobiology's 2015 goals:**

Goal 1: Create new specimen and locality records through single-record entry or migration of data sets; all will be at least to Rank 3 by 2015

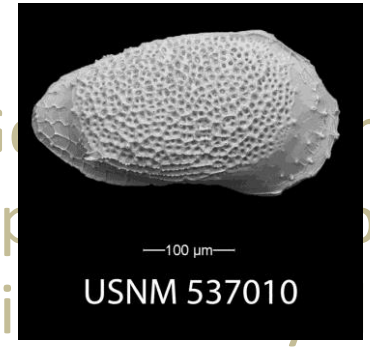
Goal 2: Create new or migrate existing specimen images and attach them to specimen records, all will be at least to Rank 3 by 2015 *See Lightning Slide for details*

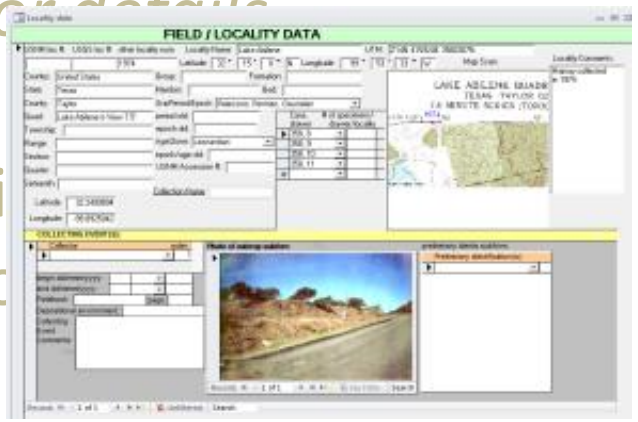
Goal 3: EMu maintenance, standardization, and data clean up in order to share data in online collections or biodiversity data portals by 2015

Contemporary Digitization at SI

Step 3. Set Priorities and Goals - **Paleobiology's 2015 goals:**

Goal 1: Create new specimen and locality records through single-record entry or migration of data sets; all will be at least to Rank 3 by 2015

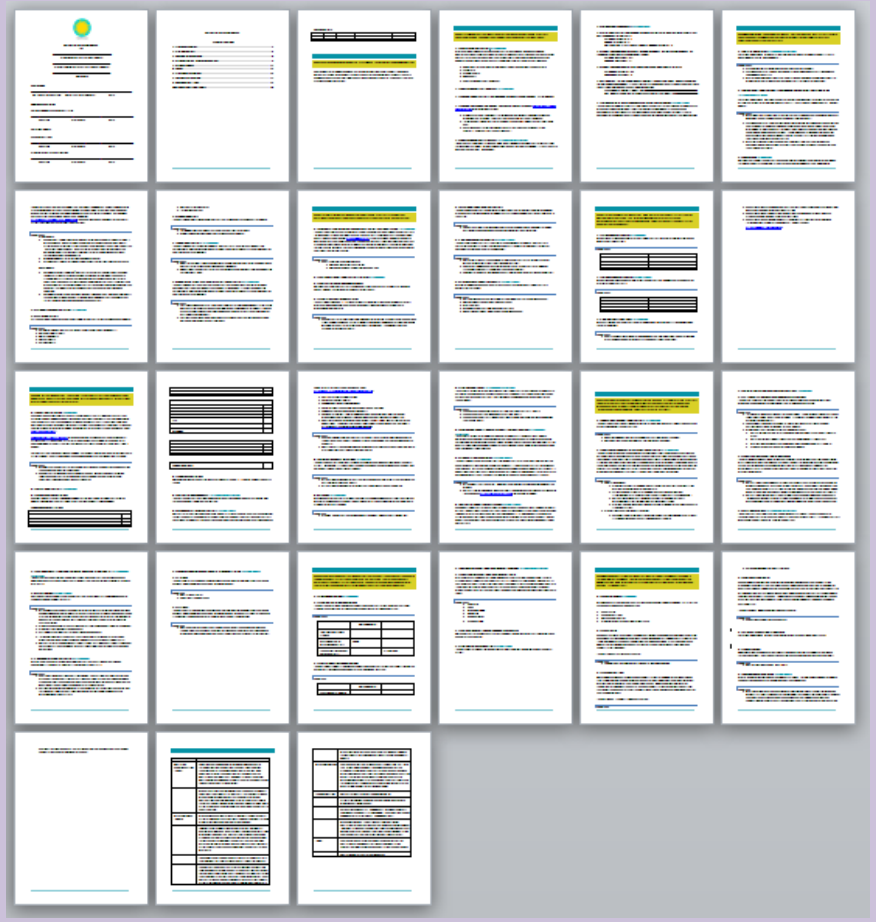




Contemporary Digitization at SI

Step 4. Figure out logistics and workflows

At the Smithsonian, each digitization project should be managed through an approved Digital Asset Management Plan, which are filed at the Digitization Program Office.

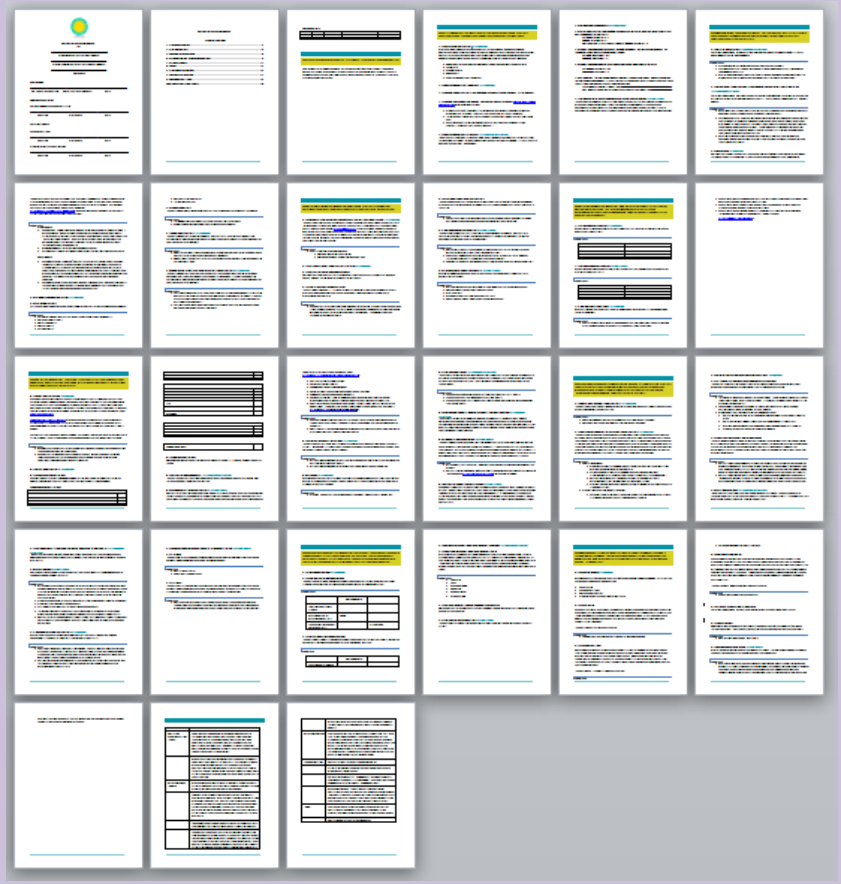


Contemporary Digitization at SI

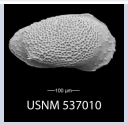
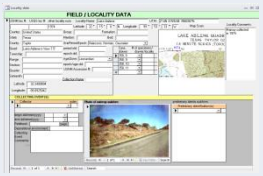
Step 4. Figure out logistics and workflows

A written Project Plan is for:

1. **Stating the significance of the digital assets**
2. **How the project supports museum and Strategic goals**
3. **Stating the scope including type, volume, and owner of digital assets**
4. **Workflow**
5. **Stating audience and usage for digital assets**
6. **Policies and restrictions**
7. **Lifecycle management**
8. **Technical requirements**
9. **Roles and responsibilities**



Logistics and workflows for Goal 1: Create new specimen and locality records; all will be at least to Rank 3 by 2015

Digitization Project *DOES NOT INCLUDE IAMGING*	Records created by end of 2015	Percent of collection databased by 2015	*Workflow*
Bown-Rose Eocene Mammals	9,000	45% 	Pull specimens, type data into Emu from specimen labels, verify data as needed
Ongoing Type specimen cataloging	Records created as needed; "filler" work	?% 	Pull specimens, type data into Emu from specimen labels, verify data as needed
Ostracoda Biologic	2,000	100% 	Pull specimens, type data into Emu from specimen labels, verify data as needed
All Exhibit specimens	2,000	100% 	Create inventory labels for all specimens, update Excel file pre-mapped to Emu fields, migrate final data set to Emu.
Paleobotany (Cleared Leaves and Pz Plants)	6,000	44% 	Pull specimens, type data into pre-mapped Emu migration template rom specimen labels, verify data as needed; Map Access database to Emu, standardize data where necessary, and migrate to EMu.
Paleobotany Localities	1,500	100% 	Map Access database to Emu, standardize data where necessary, and migrate to EMu.
Green River plants and insects	30,000	100% 	Map Filemaker database to Emu, standardize data where necessary, and migrate to Emu.

OLD SCHOOL

NEW SCHOOL

NMNH Paleo tips for undertaking digitization projects

1. Collect data on where you are starting.
2. Write plan and priorities for what you want to digitize in the next few years.
3. Define written standards for record content and quality, especially if workflow varies among digitization projects. Standards should meet interoperability requirements.
4. Centralized, stable storage during data and image processing is essential. No free-standing data on external hard drives!

hollisk@si.edu

NMNH Paleo tips for undertaking digitization projects

1. Collect data on where you are starting.
2. Write plan and priorities for what you want to digitize in the next few years.
3. Define written standards for record content and quality, especially if workflow varies among digitization projects. Standards should meet interoperability requirements.
4. Centralized, stable storage during data and image processing is essential. No free-standing data on external hard drives!

Next steps for NMNH Paleo digitization:

1. Learn more about efficient, time-saving workflows and put them in place.
2. Educate department on digitization and data sharing so it is a universal priority.
3. Start a georeferencing initiative.
4. Set our data free!

hollisk@si.edu