

Evolution of a database: from Excel to Specify and the era of modern data publishing



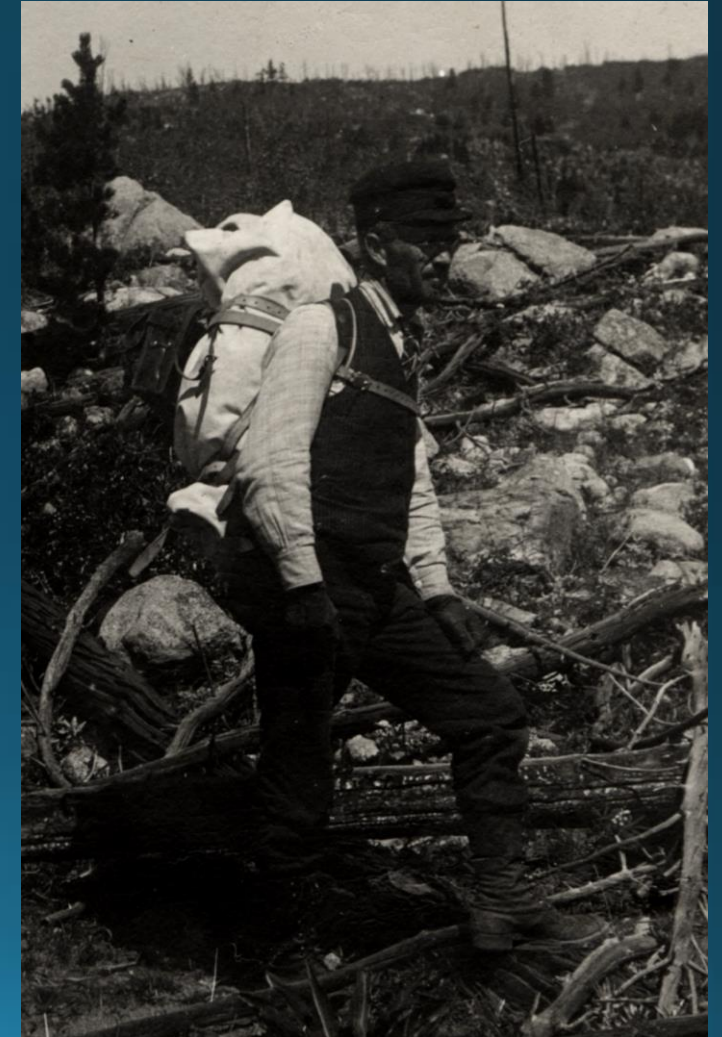
Talia S. Karim¹ and Dena Smith²

¹University of Colorado Museum of Natural History, Boulder, CO

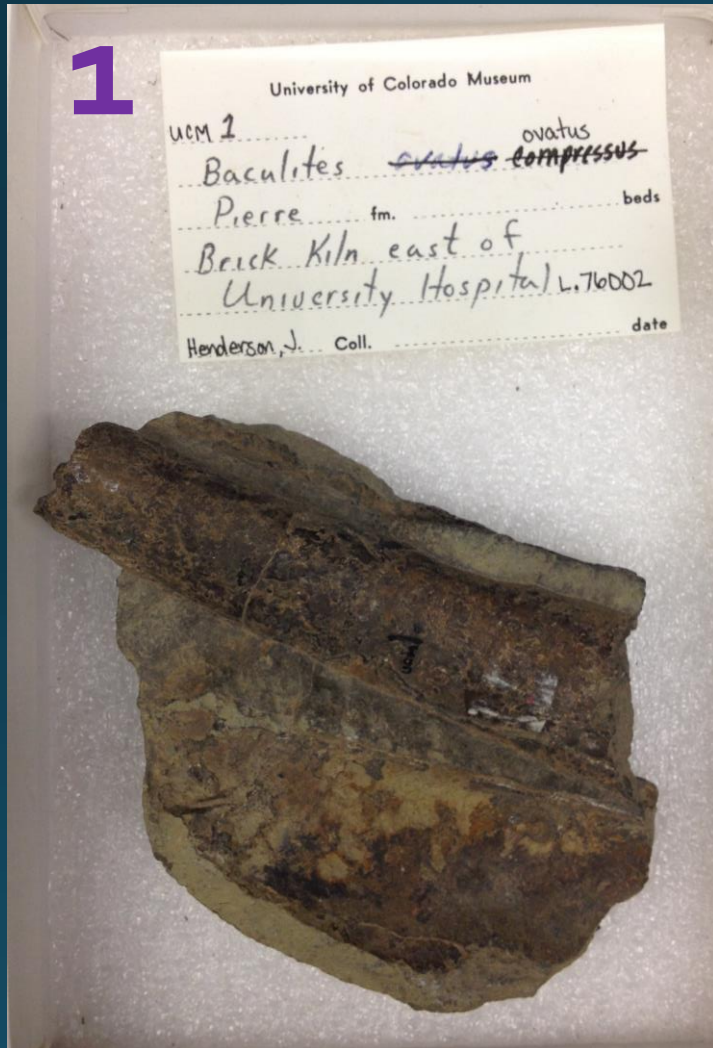
²Sedimentary Geology and Paleobiology, National Science Foundation, Arlington, VA

CUMNH Paleontology Collection

- 1902- Junius Henderson

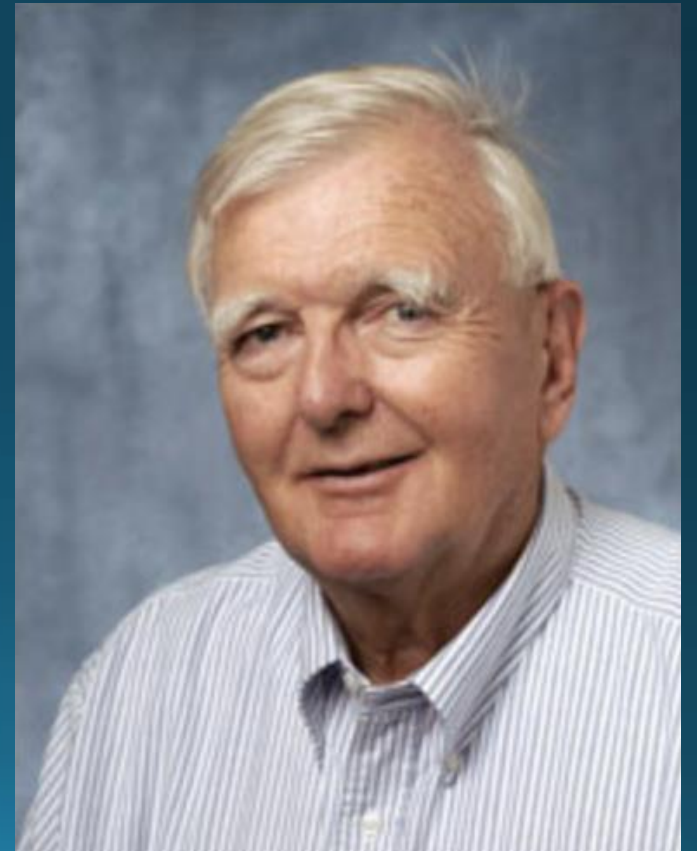


CUMNH Paleontology Collection



CUMNH Paleontology Database

- Advent of electronic databasing... Peter Robinson
- Director: 1971-1982
- Curator of Geology: 1961-2002
- Entering VP records into Excel: 1980's

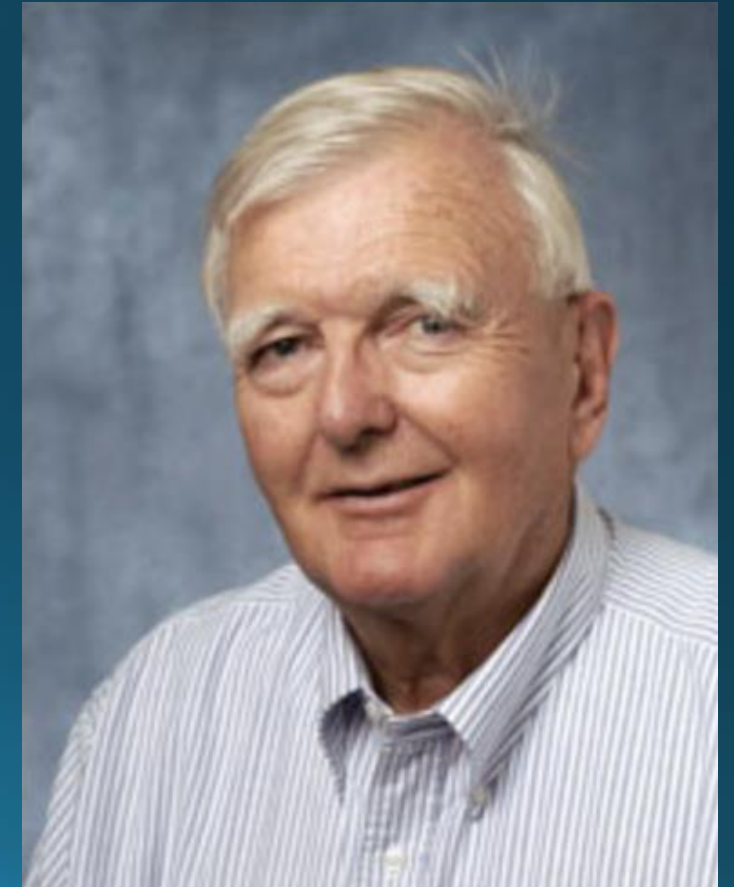


CUMNH Paleontology Collection

100 records per day

IMN	SUFFIX (PCP)	ACCESSION #	COLLECTOR	COLLECTION DATE	DESCRIPTION (LEAF, WING, ETC.)	TAXON	TAXON LEVEL	IDENTIFIER	ID DATE	SPECIAL REFERENCE	LOCALITY #	COMMENTS	ASSIGNED BY	ADDITIONAL INFO
79476		770	D. Kohls	July 16, 2009	body	Auchenorrhyncha	suborder	T. Karim	8/13/14		2009063		K. Schreiber	CP 090716
79477		"	"	"	wing + partial body	"	"	"	"		"		"	"
79478		"	"	"	body	Heteroptera	"	"	"		"		"	"
79479		"	"	"	"	"	"	"	"		"		"	"
79480		"	"	"	"	"	"	"	"		"		"	"
79481	atb	"	"	"	"	"	"	"	"		"		"	"
79482	atb	"	"	"	winged body	"	"	"	"		"		"	"
79483		"	"	"	body	Chrysomelidae	family	"	"		"		"	"
79484		"	"	"	partial body	Coleoptera	order	"	"		"	Assoc. UCM #s: 79484-79485	"	"
79485		"	"	"	body	Mordellidae	family	"	"		"	"	"	"
79486		"	"	"	"	"	"	"	"		"		"	"
79487	atb	"	"	"	"	Carabidae	"	"	"		"		"	"
79488	atb	"	"	"	"	Curculionidae	"	"	"		"		"	"
79489		"	"	"	"	"	"	"	"		"		"	"
79490	atb	"	"	"	elytra	Coleoptera	order	"	"		"		"	"
79491		"	"	"	body	"	"	"	"		"		"	"
79492	atb	"	"	"	"	"	"	"	"		"		"	"
79493	atb	"	"	"	"	"	"	"	"		"		"	"
79494	atb	"	"	"	"	Hymenoptera	"	"	"		"		"	"
79495	atb	"	"	"	winged body	"	"	"	"		"		"	"
79496	atb	"	"	"	body	Muscomorpha	suborder	"	"		"		"	"
79497	atb	"	"	"	winged body	Ephydroidea	family	E. Anderson	"		"		"	"
79498	atb	"	"	"	"	Tipulidae	"	"	"		"		"	"
79499	atb	"	"	"	body	Chironomidae	"	"	"		"		"	"
79500	atb	"	"	"	winged body	Tipulidae	"	"	"		"		"	"

Entered by N. Ridgwell Date 8-28-2014



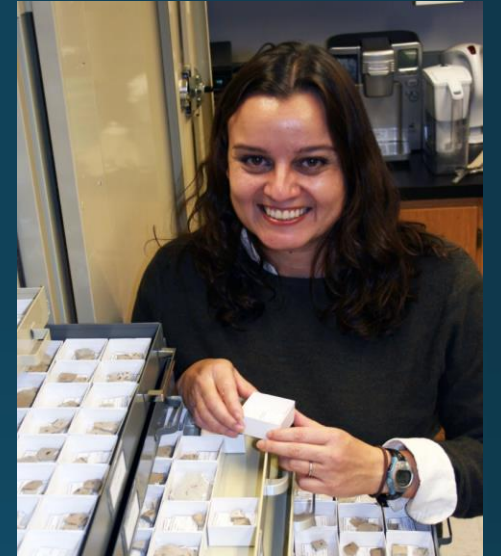
CUMNH Paleontology Database

- VP records migrated into Access... 1990's
- IP records entered straight into Access- flat files
- Data entry mistakes
 - QC time consuming



Specify 5: fossil invertebrates and plants

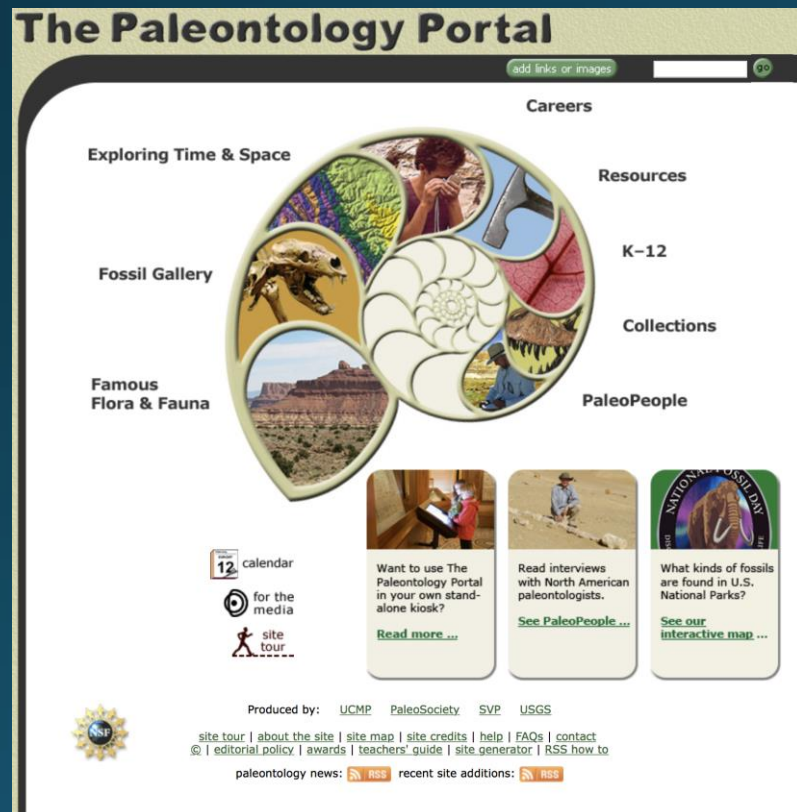
- Migration into Specify 5- ca. 2004
 - Part of NSF Collection Improvement Grant
- Invertebrate and Plant records into Specify 5



POST MORTEM,
WE SORT 'EM.

Data Publishing Take 1

- Data publishing with DiGIR
 - PaleoPortal



PaleoPortal Collections Search

Please read the Paleontology Portal [Data Use Agreements](#) before starting your search.

1. Build Your Query ?

Concept	Comparator	Value
<input type="text" value="Date last modified"/>	<input type="text" value="equals"/>	<input type="text"/>

2. Choose Options more options

Return field set:

Select data providers:

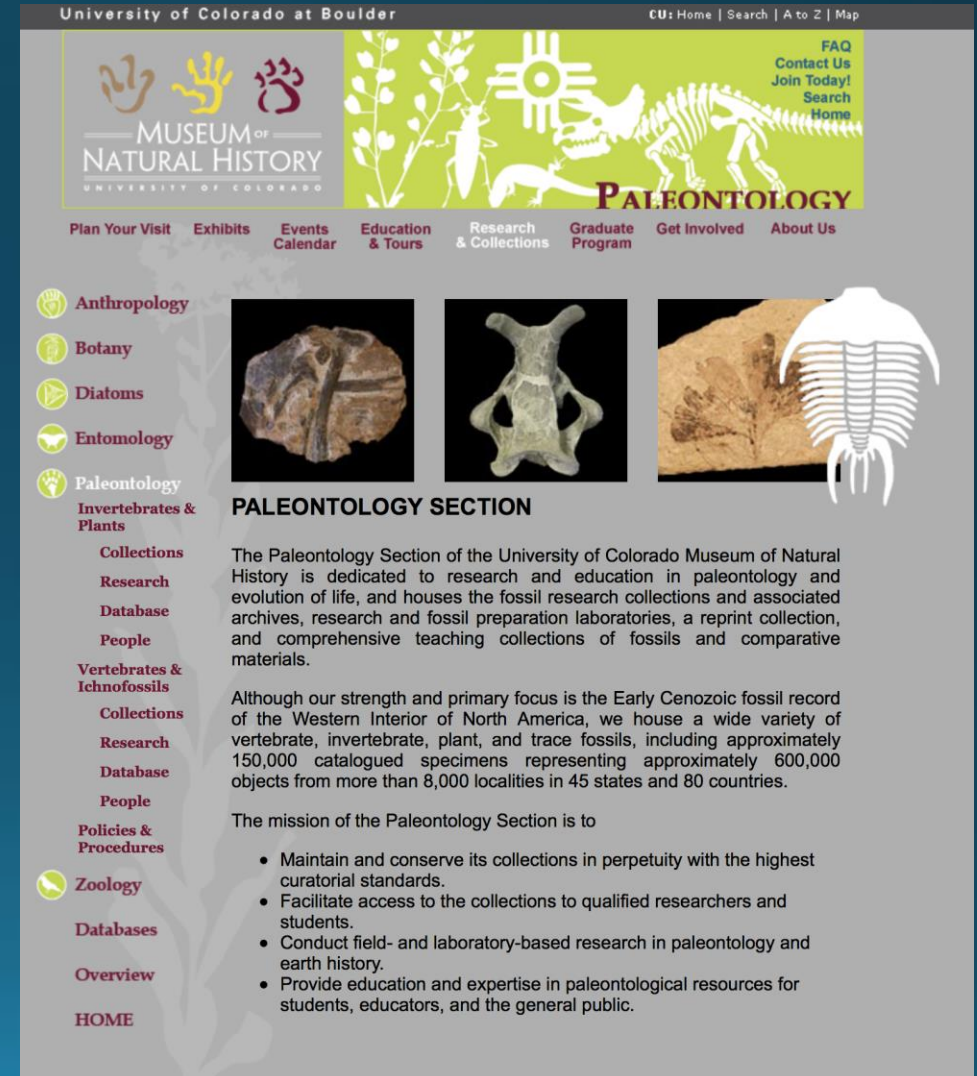
- University of Iowa Paleontology Repository
- University of Colorado Museum of Natural History Paleontology Collection
- American Museum of Natural History
- MorphoBank.org

Maximum records per provider:

List results per page (leave blank for single page of results)

Data Publishing Take 1

- Custom built websearch on museum website
- Challenges:
 - Updating data cache
 - Change in personnel



The screenshot shows the website for the University of Colorado at Boulder Museum of Natural History, specifically the Paleontology section. The header includes the university name and navigation links like 'CU: Home | Search | A to Z | Map'. The main banner features the museum logo and the word 'PALEONTOLOGY' in large letters. Below the banner is a navigation menu with links such as 'Plan Your Visit', 'Exhibits', 'Events Calendar', 'Education & Tours', 'Research & Collections', 'Graduate Program', 'Get Involved', and 'About Us'. The left sidebar contains a list of categories: Anthropology, Botany, Diatoms, Entomology, Paleontology (with sub-links for Invertebrates & Plants, Collections, Research, Database, People, Vertebrates & Ichnofossils, Collections, Research, Database, People, Policies & Procedures), Zoology, Databases, Overview, and HOME. The main content area is titled 'PALEONTOLOGY SECTION' and includes three images of fossils: a fossilized skull, a fossilized vertebra, and a fossilized trilobite. The text describes the section's dedication to research and education in paleontology and evolution, and lists its mission goals.

University of Colorado at Boulder CU: Home | Search | A to Z | Map

MUSEUM OF NATURAL HISTORY UNIVERSITY OF COLORADO

PALEONTOLOGY

FAQ Contact Us Join Today! Search Home

Plan Your Visit Exhibits Events Calendar Education & Tours Research & Collections Graduate Program Get Involved About Us

Anthropology

Botany

Diatoms

Entomology

Paleontology

Invertebrates & Plants

Collections

Research

Database

People

Vertebrates & Ichnofossils

Collections

Research

Database

People

Policies & Procedures

Zoology

Databases

Overview

HOME

PALEONTOLOGY SECTION

The Paleontology Section of the University of Colorado Museum of Natural History is dedicated to research and education in paleontology and evolution of life, and houses the fossil research collections and associated archives, research and fossil preparation laboratories, a reprint collection, and comprehensive teaching collections of fossils and comparative materials.

Although our strength and primary focus is the Early Cenozoic fossil record of the Western Interior of North America, we house a wide variety of vertebrate, invertebrate, plant, and trace fossils, including approximately 150,000 catalogued specimens representing approximately 600,000 objects from more than 8,000 localities in 45 states and 80 countries.

The mission of the Paleontology Section is to

- Maintain and conserve its collections in perpetuity with the highest curatorial standards.
- Facilitate access to the collections to qualified researchers and students.
- Conduct field- and laboratory-based research in paleontology and earth history.
- Provide education and expertise in paleontological resources for students, educators, and the general public.

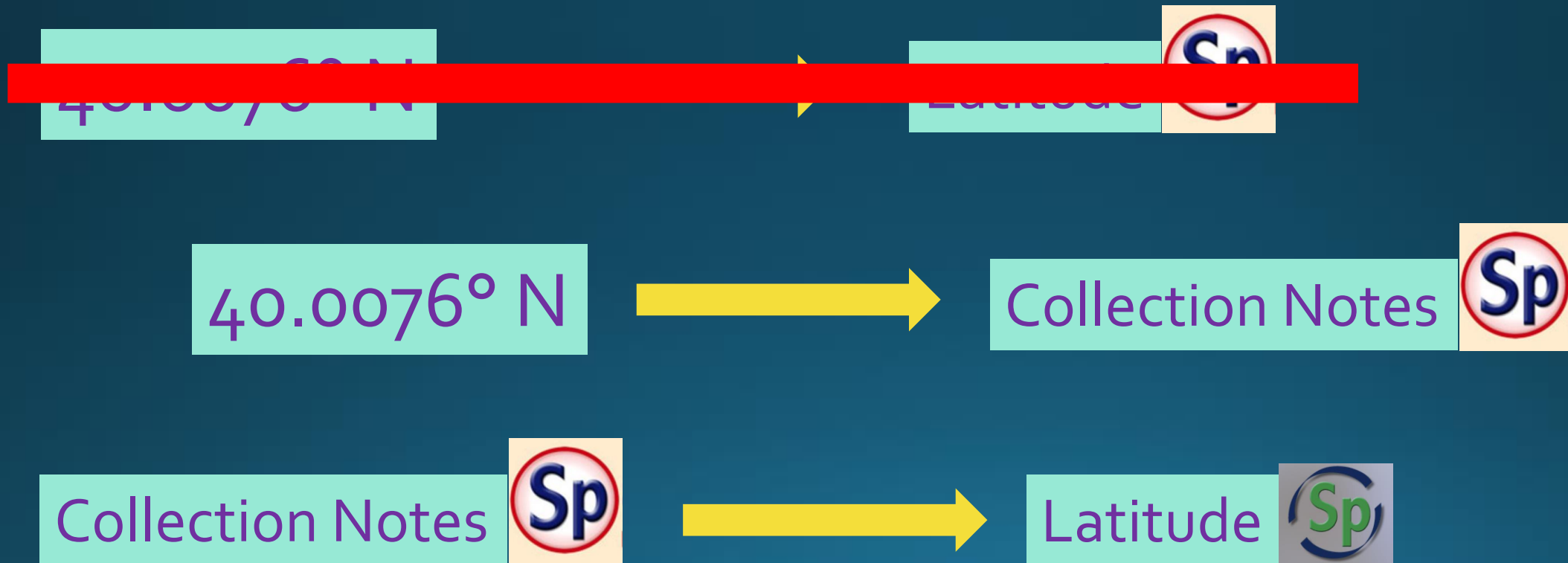
Specify 6: fossil invertebrates and plants

- Migration into Specify 6- 2009-2011
 - Conversion downtime- nearly 2 years
 - Completed early 2011
- Challenges:
 - No in house IT support
 - Issue with initial conversion in Specify 5
 - Enter catalog sheet backlog



Specify 6: fossil invertebrates and plants

- Challenges



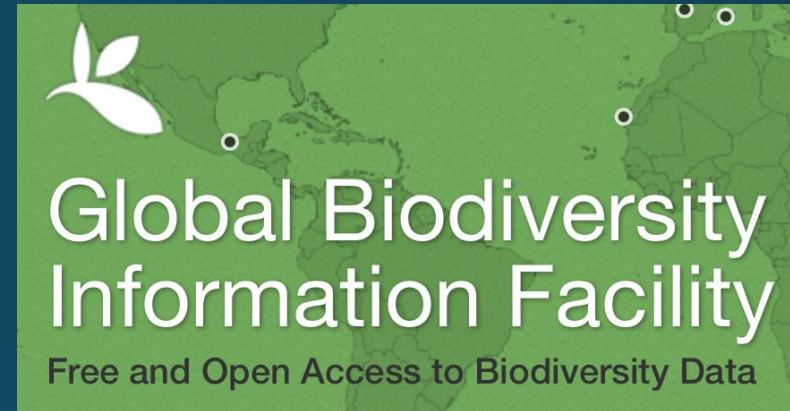
Specify 6: fossil invertebrates and plants

- Migration into Specify 6- 2009-2011
 - Conversion downtime- nearly 2 years
 - Completed early 2011
- Challenges:
 - No in house IT support
 - Issue with initial conversion in Specify 5
 - Enter catalog sheet backlog



Specify 6: fossil invertebrates and plants

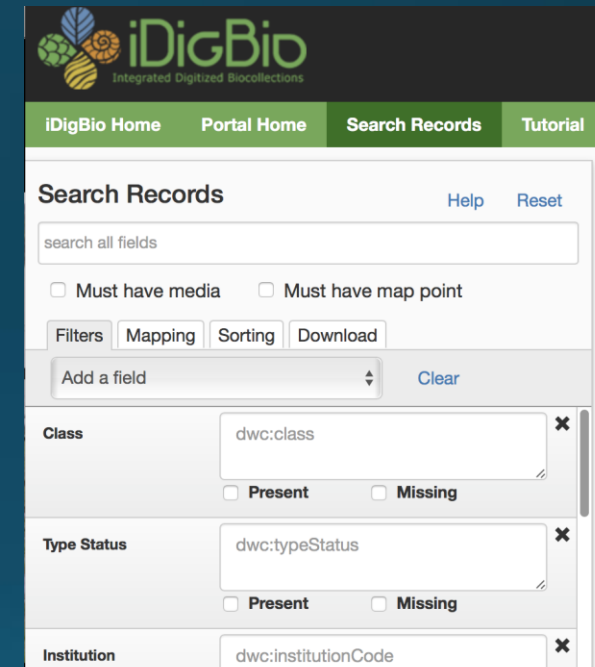
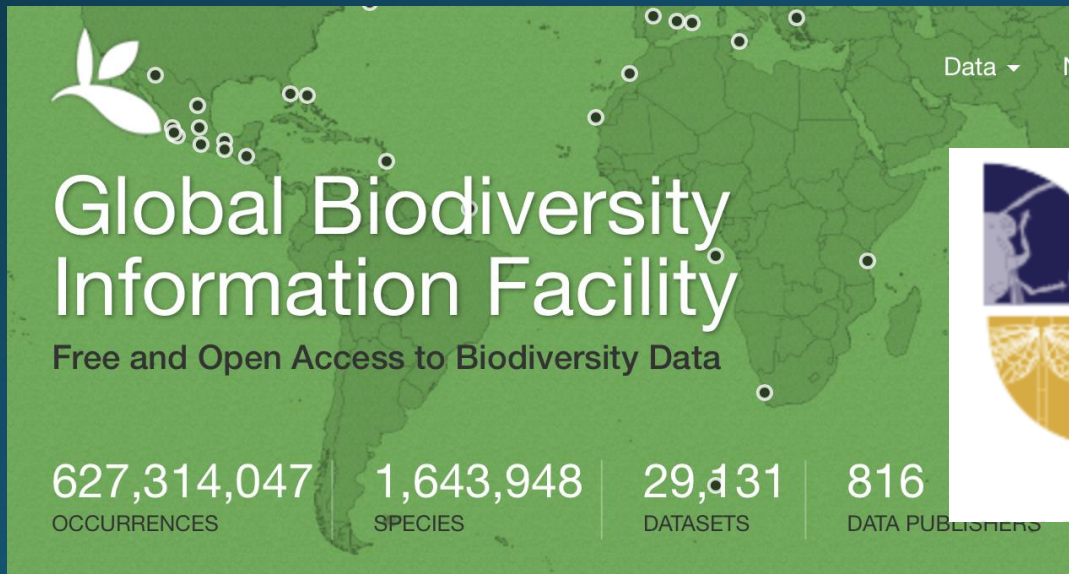
- Data publishing using IPT
- Challenges:
 - Who sets up the IPT server?
 - How do we setup the IPT server?
 - Who manages the IPT server?

The screenshot shows the GBIF Integrated Publishing Toolkit (IPT) interface. At the top, there is a navigation bar with the GBIF logo, the text "GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)", and the tagline "free and open access to biodiversity data". There are also search fields for "email" and "password", and buttons for "login" and "ENGLISH". Below the navigation bar, there are two buttons: "Home" and "About". The main content area is titled "Hosted resources available through this IPT" and features a table with columns for "Logo", "Name", "Organisation", "Type", "Subtype", "Records", "Last modified", "Last publication", and "Next publication". The table contains two rows of data. Below the table, there is a "Filter:" field and a "Showing 1 to 2 of 2 resources" message. At the bottom, there is a footer with the text "The most recently updated resources are also available as an RSS feed." and a navigation bar with links for "IPT Version 2.1.1-r4640", "About the IPT project", "User manual", "Report a bug", and "Request new feature".

Logo	Name	Organisation	Type	Subtype	Records	Last modified	Last publication	Next publication
--	CU Invertebrate Paleontology	University of Colorado Museum of Natural History	Occurrence	Specimen	74,862	2016-09-06	2016-09-06	--
--	UCMC Entomology	University of Colorado Museum of Natural History	Occurrence	Specimen	80,516	2015-12-03	2015-07-17	--

Specify 6: fossil invertebrates and plants

- Data publishing using IPT
- Benefits:
 - Share data with everyone!



iDigBio Home Portal Home Search Records Tutorial

Search Records [Help](#) [Reset](#)

search all fields

Must have media Must have map point

Filters Mapping Sorting Download

Add a field [Clear](#)

Class [x](#)

Present Missing

Type Status [x](#)

Present Missing

Institution [x](#)



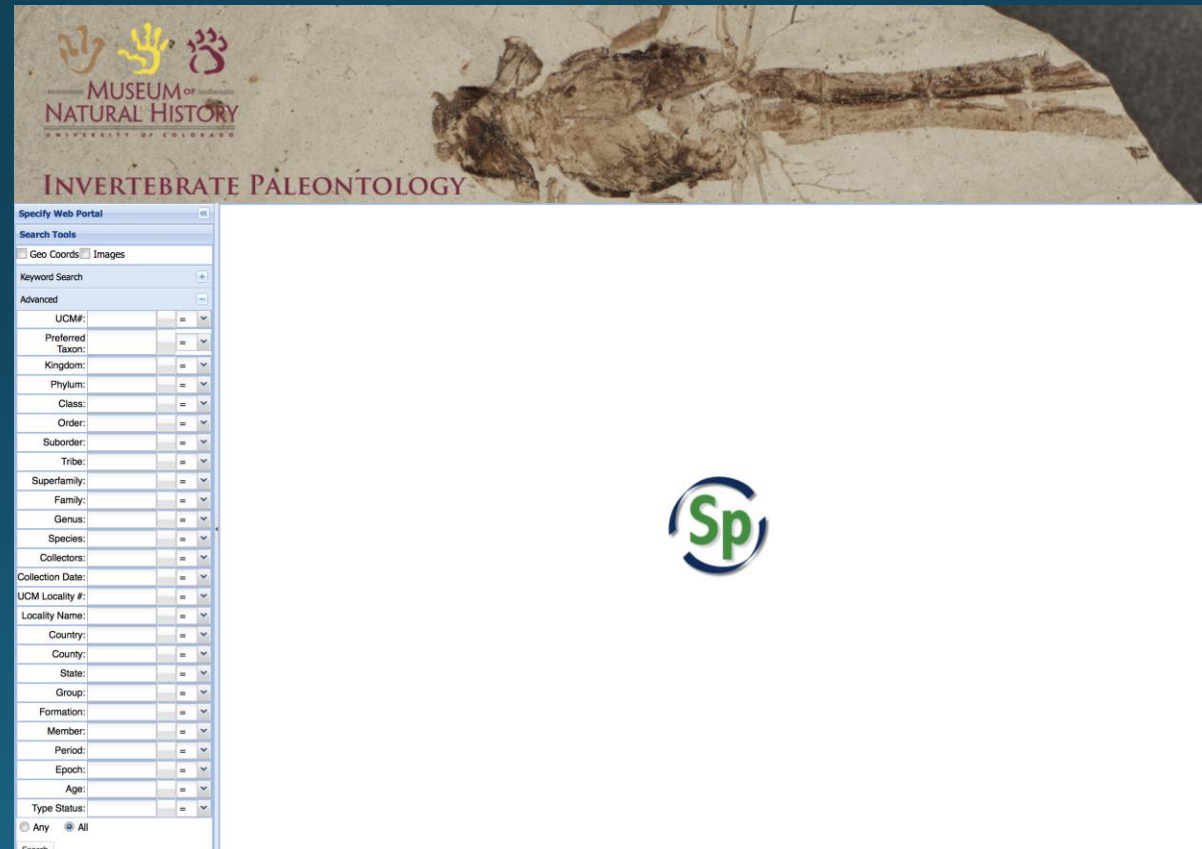
Specify 6: fossil invertebrates and plants

- iDigBio and working with ADBC
- Setup Specify 6 Attachment Server and Websearch



Specify 6: fossil invertebrates and plants

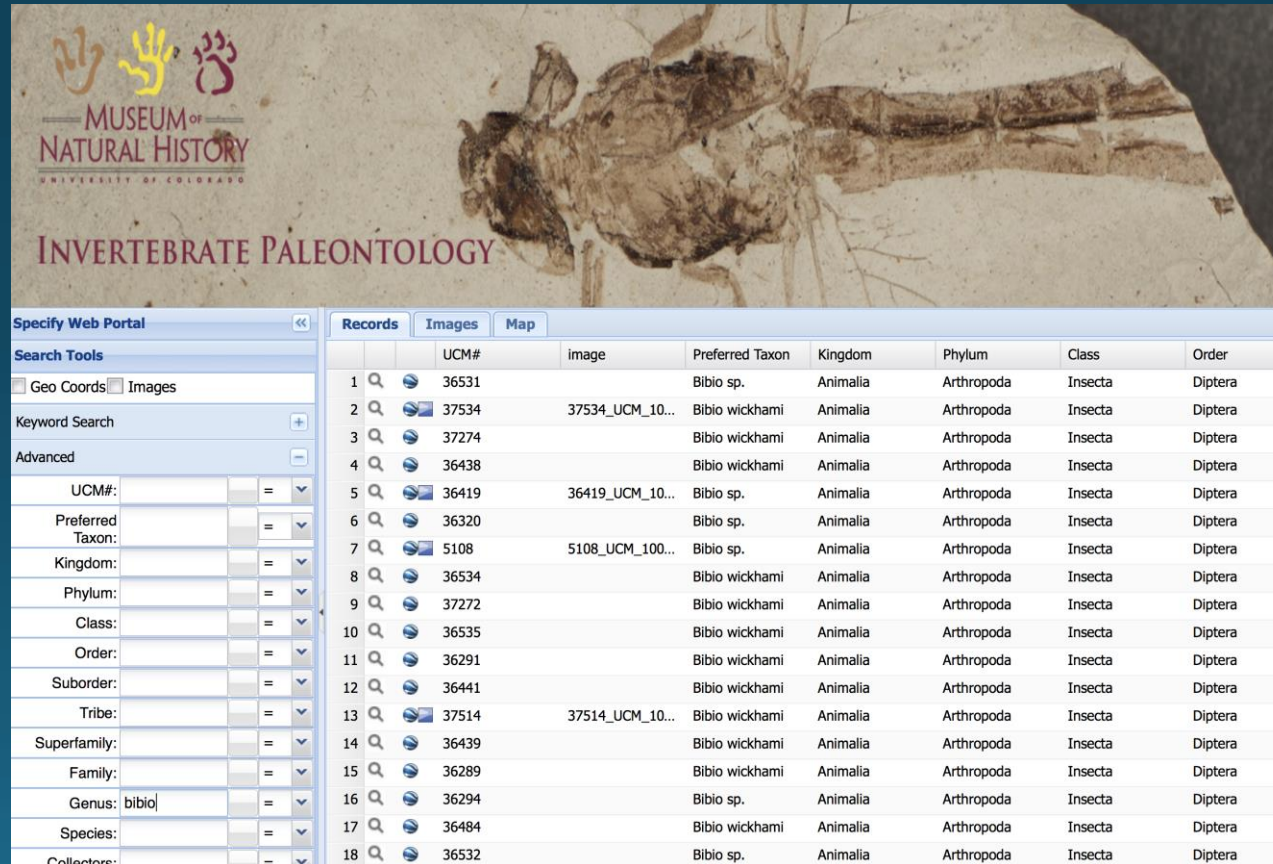
- Specify 6 Websearch
 - <http://invertpaleosearch.colorado.edu>



The image shows a screenshot of the Specify 6 web portal. At the top, there is a banner for the "MUSEUM of NATURAL HISTORY UNIVERSITY OF COLORADO" with the text "INVERTEBRATE PALEONTOLOGY" and a fossil specimen. Below the banner is a search interface titled "Specify Web Portal". The interface includes a "Search Tools" section with checkboxes for "Geo Coords" and "Images". A "Keyword Search" field is present. An "Advanced" search section contains numerous dropdown menus for taxonomic and geographic filtering, including: UCM#, Preferred Taxon, Kingdom, Phylum, Class, Order, Suborder, Tribe, Superfamily, Family, Genus, Species, Collectors, Collection Date, UCM Locality #, Locality Name, Country, State, Group, Formation, Member, Period, Epoch, Age, and Type Status. At the bottom of the advanced search section, there are radio buttons for "Any" and "All". A "Sp" logo is visible in the center of the main content area.

Specify 6: fossil invertebrates and plants

- Specify 6 Websearch
 - <http://invertpaleosearch.colorado.edu>



MUSEUM OF NATURAL HISTORY
UNIVERSITY OF COLORADO
INVERTEBRATE PALEONTOLOGY

Specify Web Portal

Search Tools

Geo Coords Images

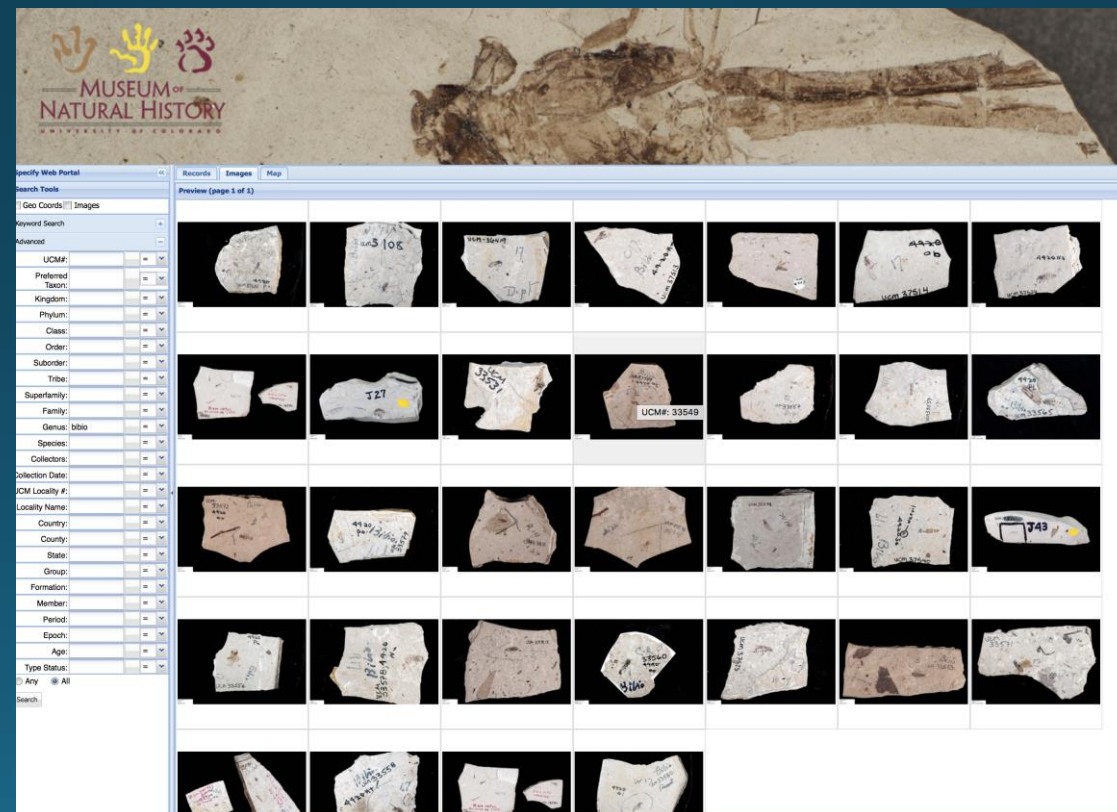
Keyword Search

Advanced

	UCM#	image	Preferred Taxon	Kingdom	Phylum	Class	Order
1	36531		Bibio sp.	Animalia	Arthropoda	Insecta	Diptera
2	37534	37534_UCM_10...	Bibio wickhami	Animalia	Arthropoda	Insecta	Diptera
3	37274		Bibio wickhami	Animalia	Arthropoda	Insecta	Diptera
4	36438		Bibio wickhami	Animalia	Arthropoda	Insecta	Diptera
5	36419	36419_UCM_10...	Bibio sp.	Animalia	Arthropoda	Insecta	Diptera
6	36320		Bibio sp.	Animalia	Arthropoda	Insecta	Diptera
7	5108	5108_UCM_100...	Bibio sp.	Animalia	Arthropoda	Insecta	Diptera
8	36534		Bibio wickhami	Animalia	Arthropoda	Insecta	Diptera
9	37272		Bibio wickhami	Animalia	Arthropoda	Insecta	Diptera
10	36535		Bibio wickhami	Animalia	Arthropoda	Insecta	Diptera
11	36291		Bibio wickhami	Animalia	Arthropoda	Insecta	Diptera
12	36441		Bibio wickhami	Animalia	Arthropoda	Insecta	Diptera
13	37514	37514_UCM_10...	Bibio wickhami	Animalia	Arthropoda	Insecta	Diptera
14	36439		Bibio wickhami	Animalia	Arthropoda	Insecta	Diptera
15	36289		Bibio wickhami	Animalia	Arthropoda	Insecta	Diptera
16	36294		Bibio sp.	Animalia	Arthropoda	Insecta	Diptera
17	36484		Bibio wickhami	Animalia	Arthropoda	Insecta	Diptera
18	36532		Bibio sp.	Animalia	Arthropoda	Insecta	Diptera

Specify 6: fossil invertebrates and plants

- Specify 6 Websearch
 - <http://invertpaleosearch.colorado.edu>



Specify 6: fossil invertebrates and plants

- Specify 6 Websearch
- Challenges:
 - Still need help from a server admin
- Benefits:
 - Easy to update
 - Easy to customize (with some help!)



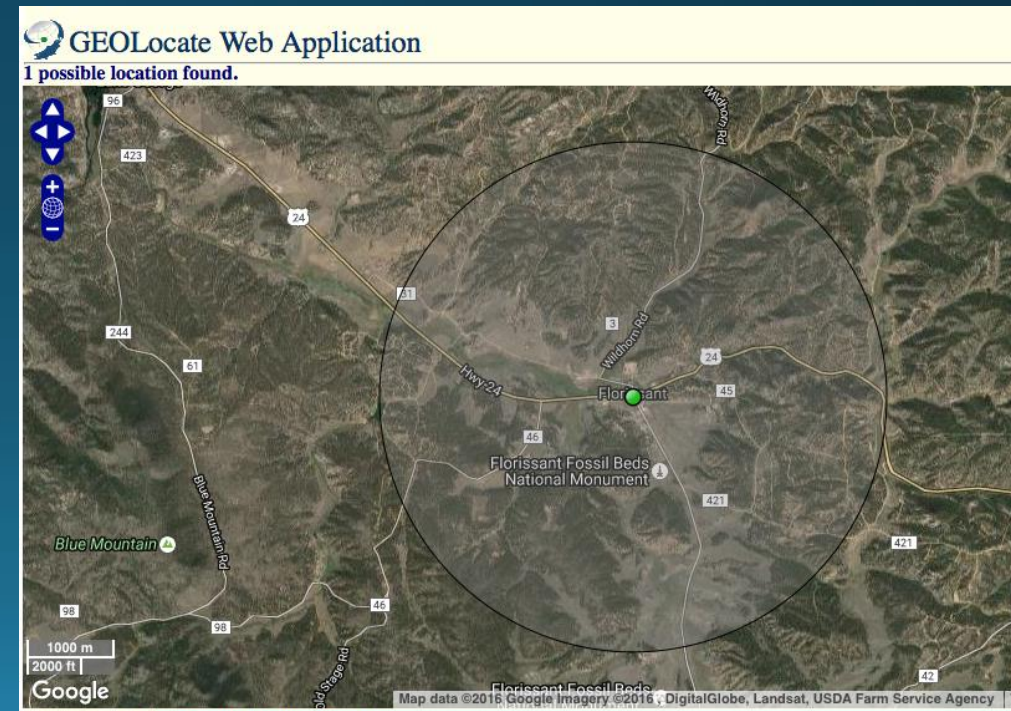
Specify 6: fossil invertebrates and plants

- iDigBio and working with ADBC
- Standardize our data with DarwinCore:
 - Dates– out of open text, into formatted fields
 - Add occurrenceID to Specify forms
 - Lots of future work!



Future Work

- Data cleanup- Access Legacy
 - Collector names- parse out into separate fields
 - Specimen count field
 - Move data in open text/notes fields into separate vocabulary controlled fields
- Do our forms accurately reflect our data?
 - Move data to new fields?
 - Remove unused fields?
- Georeferencing
 - Modify forms to capture data



Acknowledgements

