

Connecting Content: A Collaboration to Link Field Notes to Specimens and Published Literature

RICHARD T. FISCHER, MLS

California Academy of Sciences, Information Connections Research Intern

Abstract: The California Academy of Sciences Library, located in San Francisco, CA, was recently awarded a three-year Institute of Museum and Library Services (IMLS) National Leadership Grant entitled “Connecting Content: A Collaboration to Link Field Notes to Specimens and Published Literature.” Connecting Content is a collaborative effort involving the California Academy of Sciences, Missouri Botanical Garden, Academy of Natural Sciences of Philadelphia, Harvard University Botany Libraries, Harvard Museum of Comparative Zoology Library, the New York Botanical Garden and the Smithsonian National Museum of Natural History. This cooperative project aims to improve access to biodiversity research materials in archives, libraries, and museums through identifying connections and creating linkages between collections that are physically and geographically dispersed. Further, it seeks to establish connections between collections that are also intellectually dispersed in terms of collection management practices. Substantiating linkages between primary source materials, museum specimen collections, and digitized published literature was the focus of the summer 2011 research project hosted at partner site the Smithsonian National Museum of Natural History.

Introduction

Expedition field books document original scientific observation, and are the documentary forerunner to research publications and refinement of taxonomic description. Across many biodiversity research institutions, field books are housed within individual departments whose researchers created them. Their contents are typically available only to a small body of local researchers. “Connecting Content” seeks to document field book data connections to publications and biological specimen collections, in order to bring the primary description contained in field books to a broader public of researchers. The research questions that come to the fore are the following:

- How may matches be determined between primary source archival materials, museum specimen collections, and published scientific literature?
- Once a potential match is established, how can the certainty of the match be measured?
- How can rules be defined for identifying matches based on common practices of creators of the primary source research documents, the data included on specimen labels, and metadata for literature in a database?

Methodology

Archival scientific field books primarily from 1890–1930, sourced from the California Academy of Sciences (CAS), the Smithsonian National Museum of Natural History Botany Department (NMNH), and the Smithsonian Institution Archives (SIA), were used during the summer 2011 Information Connections Research Internship. Digitized scientific publications in the Biodiversity Heritage Library (BHL) and JSTOR were used as the test portals for citation matches. Institutional biological specimen collection databases from CAS and NMNH were used for specimen collections test results. The research investigated analog and digitized field books for content, format, and style. Collector or Curator name searches tested publication authorship results in BHL and JSTOR, and publication searches appraised specimen numbers and date and location verification. Cross-referencing tested collecting data with



numbers, dates and locations from CAS and NMNH databases. The research is qualitative and represents a random sampling of data sourced in order to glean an impression of the common trends in information relationships between field books, publications, and specimen collections.

Field books containing specimen data and observations, publications resulting from formalized post-expedition research, and natural history specimen collection databases comprise an information relationship with multiple points of entry. The connective thread may be followed in any number of directions depending upon how the sources are cross-referenced. For example, a specimen number (“CAS 3156”) in a California Academy of Sciences Collection Database could be searched in JSTOR to see if it has been cited in an available publication. Assuming it has been cited, one could proceed to search the collector’s field notebooks, to see if the same specimen is recorded in the field.

Beginning with the field book itself, surveying its format and contents for geographical location, dates, and presence of specimen numbers, followed by searching the relevant author or curator in JSTOR and the BHL quickly narrows down whether there is a direct link between a scientific publication and an expedition field book. If such a link exists, then searching the relevant natural history specimen collection database for holdings which can be verified as the same specimens described in the original field book is the next step. Of the different types of matches between these sources that arise through this research methodology, the three of greatest interest to the research goals of the project are the direct three-way match, the indirect three-way match, and the ambiguous possible three-way match. A direct three-way match describes an information relationship in which collected specimens are recorded with numbers in a field book, those same numbers, along with the same locations and dates, are cited in a digitized publication, and an institutional specimen collection database includes the same specimens, citing the original field book number.

Findings

Types of Matches of Primary Interest:

Direct Three-Way Match

- Collected specimens are recorded with numbers in field book.
- Same numbers, along with same locations and dates, cited in digitized publication.
- Institutional specimen collection database includes same specimens, citing original field book number.

Indirect Three-Way Match

- Collected specimens are recorded with numbers in field book.
- Same locations and dates, cited in digitized publication.
- Institutional specimen collection database includes same specimens, citing original field book number.

Ambiguous Possible Three-Way Match

- Collected specimens are recorded with numbers in field book.
- General locations and general dates, cited in digitized publication.
- Institutional specimen collection database includes broad date ranges or inconclusive locations.

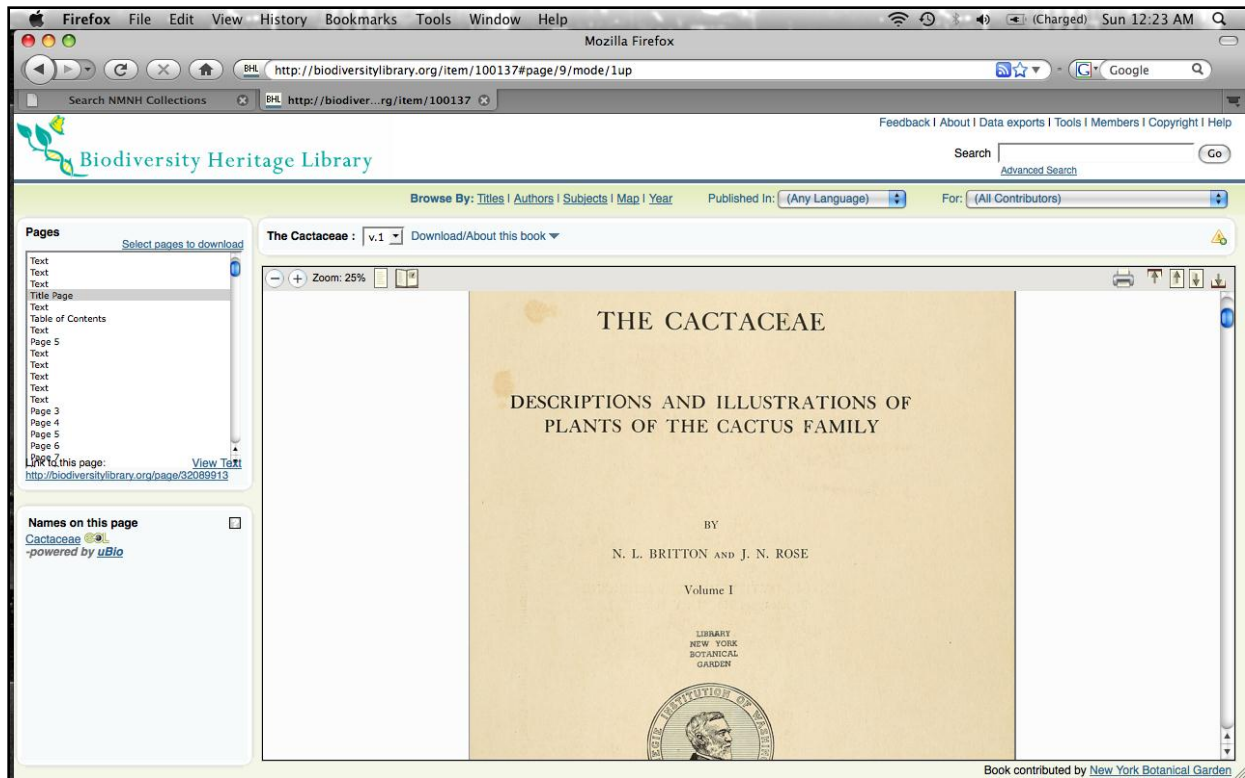


Figure 1. *The Cactaceae*. Biodiversity Heritage Library Screen Shot.

Example of a Three-Way Match: J.N. Rose's Cacti

From the Introduction:¹

The writers began field, greenhouse, and herbarium studies of the Cactaceae in **1904** and in the years following they made studies and collections over wide areas in the **United States, Mexico, and the West Indies** . . . In **1914 and 1915** Dr. Britton again visited **Porto Rico [sic]** . . . In **1914** Dr. Rose went to the west coast of South America, making short stops at **Jamaica and Panama**. He made extensive collections in central and southern **Peru**, central **Bolivia**, and northern and central **Chile** . . . The types of the new species described in this work **are deposited in the herbaria of the New York Botanical Garden and the United States National Museum**, unless otherwise indicated.

In this and other publication introductions, collecting details and logistical information such as dates and place-names provide a foundation for further inquiry into field books and specimen collections.

- Botanical family of Cactacea.
- Date range given of 1904–1918.
- Locations given of North America, South American and the Caribbean.

¹ N.L. Britton and J.N. Rose, *The Cactaceae: Descriptions and Illustrations of Plants of the Cactus Family* (Washington: Government Printing Office, 1919), 3–4.

NMNH Botany Collections Database search with these fields:

Family	Species	Collector	Number	Date	Country	State
Cactaceae	Opuntia lindheimeri Engelm. va	Rose, J. N.	8222	1905-06-27	United States	Texas
Cactaceae	Opuntia lindheimeri Engelm. va	Rose, J. N.	8223	1905-06-27	United States	Texas
Cactaceae	Echinocereus enneacanthus	Rose, J. N.	8224	1905-06-27	United States	Texas
Cactaceae	Opuntia schottii var. schottii	Rose, J. N.	8225	1905-06-27	United States	Texas
Cactaceae	Grusonia schottii (Engelm.) H. F	Rose, J. N.	8225	1905-06-27	United States	Texas
Cactaceae	Wilcoxia poselgeri	Rose, J. N.	8226	1905-06-27	United States	Texas
Cactaceae	Opuntia lindheimeri Engelm. va	Rose, J. N.	8235	1905-06-27	United States	Texas
Cactaceae	Thelocactus bicolor	Rose, J. N.	8240	1905-06-28	Mexico	Coahuila
Cactaceae	Heliocereus speciosus	Rose, J. N.	8242	1905-06-30	Mexico	Distrito Federal
Cactaceae	Mammillaria magnimamma	Rose, J. N.	8243	1905-06-30	Mexico	Distrito Federal
Cactaceae	Opuntia tomentosa Salm-Dyck	Rose, J. N.	8244	1905-06-30	Mexico	Distrito Federal
Cactaceae	Opuntia imbricata (Haw.) DC.	Rose, J. N.	8262	1905-07-01	Mexico	Distrito Federal
Cactaceae	Opuntia imbricata (Haw.) DC.	Rose, J. N.	8262	1905	Mexico	México
Cactaceae	Echinocereus cinerascens	Rose, J. N.	8263	1905-07-01	Mexico	Distrito Federal
Cactaceae	Pachycereus marginatus	Rose, J. N.	8278	1905-07-03 to 1905	Mexico	Hidalgo
Cactaceae	Pachycereus marginatus	Rose, J. N.	8278	1905-07-03	Mexico	Hidalgo
Cactaceae	Pachycereus marginatus	Rose, J. N.	8278	1904-07-03	Mexico	Hidalgo
Cactaceae	Pachycereus marginatus	Rose, J. N.	8278	1904-07-03	Mexico	Hidalgo
Cactaceae	Mammillaria magnimamma	Rose, J. N.	8281	1905-07-03	Mexico	Hidalgo
Cactaceae	Mammillaria sp.	Rose, J. N.	8281	1905-07-03	Mexico	Hidalgo
Cactaceae	Stenocereus dumortieri	Rose, J. N.	8282	1905-07-03	Mexico	Hidalgo
Cactaceae	Stenocereus dumortieri	Rose, J. N.	8282	1905-07-03	Mexico	Hidalgo
Cactaceae	Stenocereus dumortieri	Rose, J. N.	8282	1905-07-03	Mexico	Hidalgo
Cactaceae	Stenocereus dumortieri	Rose, J. N.	8282	1905-07-03	Mexico	Hidalgo
Cactaceae	Opuntia cantabrigiensis Lynch	Rose, J. N.	8284	1905-07-03	Mexico	Hidalgo
Cactaceae	Opuntia sp.	Rose, J. N.	8284	1905-07-03 to 1905	Mexico	Hidalgo
Cactaceae	Opuntia imbricata (Haw.) DC.	Rose, J. N.	8287	1905-07-03 to 1905	Mexico	Hidalgo
Cactaceae	Opuntia imbricata (Haw.) DC.	Rose, J. N.	8287	1905-07-03	Mexico	Hidalgo
Cactaceae	Myrtillocactus geometrizans (M)	Rose, J. N.	8288	1905-07-03 to 1905	Mexico	Hidalgo
Cactaceae	Myrtillocactus geometrizans (M)	Rose, J. N.	8288	1905-07-03	Mexico	Hidalgo
Cactaceae	Opuntia x "dendroide"	Rose, J. N.	8289	1905-07-03 to 1905	Mexico	Hidalgo
Cactaceae	Opuntia streptacantha	Rose, J. N.	8289	1905-07-03	Mexico	Hidalgo

Figure 2. National Museum of Natural History Botany Collections Database.

- Rose’s field books “6801–8500”, “8501–11538”, the NMNH Botany Collection Database returns approximately 300 specimen level matches.

Example of an Indirect Three-Way Match: Standley’s Flora of Glacier National Park

The botanist Paul Standley kept an expedition field book in Montana’s Glacier National Park in July – September, 1919,² documenting the flora present, in work towards a field guide for visitors with an interest in botany. A scientific publication also resulted, in the introduction to which, Standley wrote: “This list of species here presented is the result chiefly of field work conducted by the author during July, August, and September, 1919.”³ This gives an indirect, non-numerically specific indication of what collection dates will correspond to specimen collections in the NMNH Botany Collection Database. The Standley Collector Specimen Numbers, 15019–18314 (not all, approx. 50% randomly dispersed) are reflected in the database.

Example of an Ambiguous Possible Three-Way Match: Walcott’s Cambrians

Charles D. Walcott’s study of cambrian brachiopoda was conducted over several years and in many different geographic locales. Utilizing the archival field books held at the Smithsonian Institution Archives (SIA) and referencing the data with Walcott’s book *Cambrian Brachiopoda* (1912), as well as the National Museum of Natural History Paelobiology Collections Database illustrates the prevalence of ambiguity in the process of connecting the three information points.

² Paul Standley, *Field Book, Glacier National Park* (3 Vols.): *Specimens 14194–18317* (Unpublished, 1919), 1.

³ Paul Standley, *Flora of Glacier National Park, Montana* (Washington: Government Printing Office, 1921), 3.

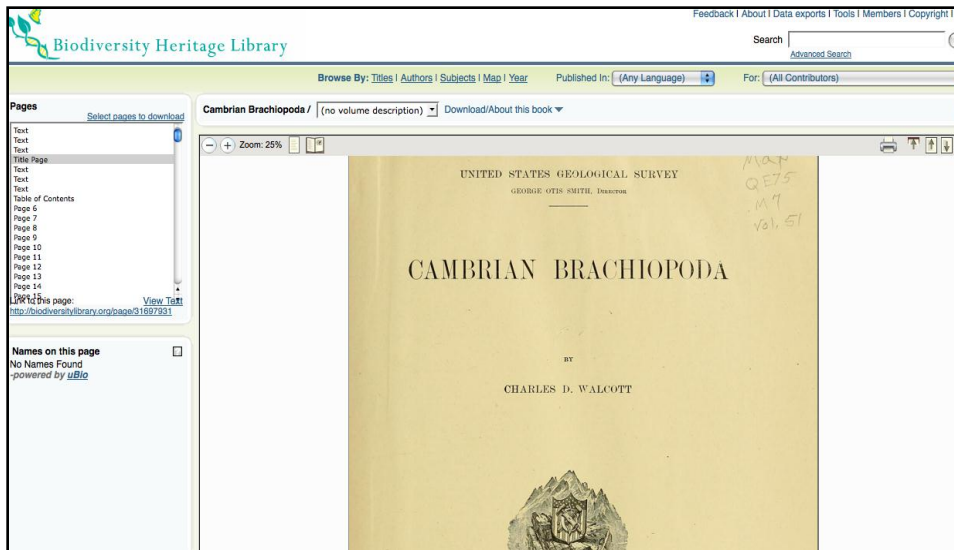


Figure 3. *Cambrian Brachiopoda*. Biodiversity Heritage Library.

- Ambiguity for hundreds of specimen-level records in NMNH Paleobiology Collections, due to broad date range (1910–1917) given in Date field.
- Vice versa, for specimens that do have exact dates in Date field, in many cases their Scientific Name is not given, only the higher level Collection name, and there is no Catalog number.
- Therefore, certain or probable matches are very low proportionally to possible matches.

<input type="checkbox"/>	<input type="checkbox"/>	PAL 83947.A	Burgess Shale Type	Crustacea	<i>Burgessia bella</i> Walcott		Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 83947.F	Burgess Shale Type	Crustacea	<i>Burgessia bella</i> Walcott		Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 83947.K	Burgess Shale Type	Crustacea	<i>Burgessia bella</i> Walcott		Canada	Cambrian
<input checked="" type="checkbox"/>	<input type="checkbox"/>	PAL 83947.B	Burgess Shale Type	Crustacea	<i>Burgessia bella</i> Walcott		Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 83947.L	Burgess Shale Type	Crustacea	<i>Burgessia bella</i> Walcott		Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199715	Burgess Shale Biolog	Worm	<i>Burgessochaeta setigera</i> (Walcott)	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199806	Burgess Shale Biolog	Worm	<i>Burgessochaeta setigera</i> (Walcott)	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199777	Burgess Shale Biolog	Worm	<i>Burgessochaeta setigera</i> (Walcott)	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 14348	Brachiopoda Type		<i>Camarophoria cooperensis</i> (Shumard)		United States	Carboniferous
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199467	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199437	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199448	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199458	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199488	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199438	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199468	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199508	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199498	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199478	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199499	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199459	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199469	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199509	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199489	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199479	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199439	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199449	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199450	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199510	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian
<input type="checkbox"/>	<input type="checkbox"/>	PAL 199470	Burgess Shale Biolog	Worm	<i>Canadia setigera</i>	1910 to 1917	Canada	Cambrian

Figure 4. National Museum of Natural History Paleobiology Collections Database.

Discerning *precise* matches between field books, publications, and specimen collections is uncommon compared to the instances of discerning *probable* or *ambiguous* connections. One or two points of connection still represent valuable supplemental data, however.

Conclusion

The information connections research conducted at the Smithsonian National Museum of Natural History in Washington, D.C. over the summer of 2011 has not yet been implemented on a broader scale. It will be utilized in the final year of the “Connecting Content” IMLS grant project, synthesized with other supporting digitization and analytical work.

References

N.L. Britton and J.N. Rose, *The Cactaceae: Descriptions and Illustrations of Plants of the Cactus Family* (Washington: Government Printing Office, 1919).

Paul Standley, *Field Book, Glacier National Park (3 Vols.): Specimens 14194–18317* (Unpublished, 1919).

Paul Standley, *Flora of Glacier National Park, Montana* (Washington: Government Printing Office, 1921).