

The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/2059-5816.htm

Excavating archival description: from collection to data level

Excavating
archival
description

AQ:au

Lesley L. Parilla

Smithsonian Libraries, Smithsonian Institution, Washington, District of Columbia, USA

Rebecca Morgan

*Special Collections Library, American Museum of Natural History,
New York, New York, USA, and*

Christina Fidler

*The Museum of Vertebrate Zoology at Berkeley, University of California,
Berkeley, California, USA*

1

Received 14 November 2016
Revised 16 December 2016
Accepted 16 December 2016

Abstract

Purpose – The purpose of this paper is to discuss three projects from three institutions that are dealing with challenges with natural sciences field documentation. Each is working to create the collection, item and data-level description required so that researchers can fully use the data to study how biodiversity has changed over time and space. Libraries, archives and museums recognize the need to make content searchable across material type. To create online catalogs that would make this possible, ideally, all records would describe one item. Museums and libraries describe their materials at the item level; however, archives must balance the need to describe the collection as a whole alongside needs of collection materials that may require more description to reconnect with library and museum items. There is a growing determination inside of archives to increase this flow of data, particularly for the natural sciences, by creating workflows that provide additional description to make these data discoverable. This process is a bit like drilling into the earth: each level must be described before the next can be dealt with.

Design/methodology/approach – The piece describes challenges, approaches and workflows of three institutions developing deeper levels of description for archival materials that will be made available online to a specialized audience. It also describes the methods developed so that the material's data can eventually be accessed at a more granular level and linked to related resources.

Findings – Current systems, schema and standards are adapted as necessary, and the natural sciences archival community is still working to develop best practices. However, they are getting much closer through the collaboration made possible through grants in the recent years.

Originality/value – The work described in this paper is ongoing, and best practices resulting from the work are still under development.

Keywords Digital libraries, Museums, GLAM, Metadata, Digitization, Archival description

Paper type Case study

Introduction

Libraries, archives and museums recognize the need to make content discoverable across material types. To create searchable online catalogs, all records would ideally describe individual items. Museums and libraries often describe their materials at the item level; however, archives most frequently describe materials at the collection level. Archives must balance the need to describe the collection as a whole alongside individual collection materials that may require further description to reconnect with library and museum items. This is particularly true of materials such as natural sciences field notes, which often exist



Digital Library Perspectives
Vol. 33 No. 3, 2017
p. 000
© Emerald Publishing Limited
2059-5816
DOI 10.1108/DLP-11-2016-0043

DLP
33,3

2

as collections of manuscripts in archives. Increasingly, scientists in natural history museums have been expecting more item-level description from archives. Natural sciences archivists believe additional description is paramount to supporting scientists' research. The scientists who create the field notes have long recognized their importance as well. Joseph Grinnell, Founding Director of the Museum of Vertebrate Zoology at University of California at Berkeley and developer of the Grinnell Method for taking field notes, stated:

Remember, that the value of our manuscripts increases as the years go by and faunal changes take place. Some of our earlier notebooks describe conditions now vanished in the localities they dealt with (Grinnell, 1938).

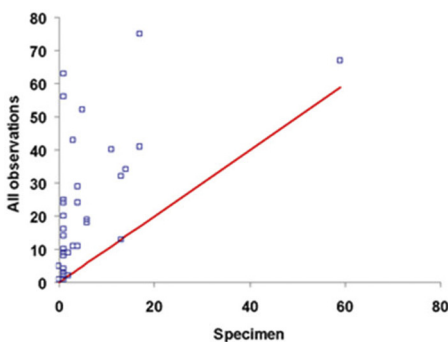
The need is clear but goes against trends in the archival community where "More Product, Less Process" (Greene and Meissner, 2005) enables archivists to manage archival backlogs and make collections accessible to the public by using methods that reduce the amount of time spent per collection.

There is a growing determination inside the archival community to increase this flow of data, particularly for the natural sciences, by creating workflows that provide additional description to make these data discoverable. This process is bit excavating; each level must be described before the next level can be dealt with. This article will discuss three archival projects from three institutions that are dealing with the challenges of describing natural sciences field documentation: the Smithsonian Institution, a federally funded research institution; the University of California at Berkeley, a state university; and the American Museum of Natural History, a private museum. Each is working to create collection, item and data-level descriptions that are required for researchers to fully use data to study how biodiversity has changed over time and space. The institutions' varying sizes, funding streams, audiences and structures also affect how they complete this work.

The three projects focus on managing field books, which are an indispensable reference for natural sciences specimens. Grinnell (1908) said, "The field records will be perhaps the most valuable of all our results". The field book (also known as field notes) holds the most accurate and detailed account of the collecting event: the place and time where collecting occurred. Field notes also describe the habitat, appearance of specimens when collected, etc. These data are usually not fully transferred to the specimen descriptive record in a museum's collections information system (Figure 1).

Multi-level description is an effort to balance the access needs of the primary field book users with descriptive practices in museums, libraries and archives. Primary users of field books (e.g. natural sciences researchers) often need a high level of access to specific data in the field notes, particularly when studying specimens in museum collections. Ideally, field

Figure 1.
During the Museum of Vertebrate Zoology's Grinnell Resurvey Project, researchers found that there were three times more bird species per site observed than those collected.



F1

books would be digitized, fully transcribed and indexed so that they could be linked to specimen records. This need for full access only grows, as the research life cycle moves towards transparency and specimens are reinterpreted through new tools of science.

Increasingly, grant awarders require recipients to make grant-funded data available over the long term. Meanwhile, the US Government has called on federally funded research to be increasingly available to the public, not just reports and publications but also information such as primary datasets. This builds on the research community's own call for open access to research data. As a result, users now have greater access to current research, which offers scholars more opportunities to study contemporary research alongside historical field data.

Natural scientists' increased expectations of access do not fit well with the standard archival description for field note collections, and archivists have developed a system of layered description to meet this new expectation. Below is an overview of the multilevel description used (in varying combinations) by the institutions covered in this article. These institutions have developed ways to provide item- and page-level description along with the standard level of description expected in museums, archives and libraries to ensure materials are searchable across platforms. Each level of description provides content and access to different audiences.

Description levels

Collection description

Collection-level description is still the primary level of description created by archives and is particularly important for field notes. Collection-level description enables users to understand the content as a whole, as in the past, field notes were frequently created to be used and understood by their creator. Most collections are encoded using encoded archival description (EAD), which is a "non-proprietary standard for the encoding of finding aids for use in a networked (online) environment"[1]. Some archives use natural collections description (NCD)[2], which offers specific tags to delineate the taxonomic description not available in EAD.

Creator description

Creator records offer important access points and context for field research. Encoded archival context – corporate bodies, persons and families (EAC-CPF) is "a standard for encoding contextual information about persons, corporate bodies and families related to archival materials using extensible markup language (XML)"[3]. It is also emerging as a standard schema used by the archival community. These records not only inform the field book content but also link to online consortiums such as Social Networks and Archival Context (SNAC)[4] that are primarily used by librarians, archivists and scholars. EAC-CPF may enable the archival community to create an online cooperative that serves a similar purpose as the Name Authority Cooperative Program managed by the Library of Congress.

Item/folder-level description

For many field book users, item- or folder-level description (e.g. one note book or a folder of field correspondence) is close to the ideal. Item-level description also supports interoperability and searchability with records from other museums and libraries that are typically described at the item level. There are important challenges with machine-readable cataloging (MARC)[5], which has long been the cataloging standard of libraries: it requires a lot of time and training and lacks the structure to support the multiple levels of archival description found in a finding aid. Other schemas such as the metadata object description

DLP
33,3

schema (MODS)[6] have also been selected by institutions because they are more flexible, map to the MARC fields and can be learned more quickly by the staff.

Page-level description

Page-level description comes into play when an item is digitized. Page metadata often overlap with image standards. International Press Telecommunications Council[7], a metadata standard commonly used for photography, is used by some institutions in cases where metadata standards follow the existing best practices taken from digital photography. At a minimum, page-level description includes information about pagination and page content (e.g. illustration or text) and may include item-level description such as title and creator. This level of description is necessary for navigating and citing digitized field books. It can be particularly challenging depending on staff availability and how closely an institution seeks to match a field book creator's method of pagination. Many institutions skip this level when creating their digital objects by compiling the digital object as a PDF.

This level of description can be made possible through consortium digital libraries such as the Biodiversity Heritage Library (BHL)[8]. BHL requires page-level metadata for all digital objects and assures description consistency by making available an online interface for metadata creation. Institutions upload individual images and add page-level metadata in the online interface. The system then merges images and descriptions into a digital object.

Data-level description

For many institutions, data-level description is still very much in development. Natural sciences institutions are looking into ways to link to types of data in field notes such as specimen numbers or taxonomic names. This first requires that the field notes be fully transcribed; several institutions are using volunteer crowd-sourced transcription.

Case 1: University of California, Berkeley, Museum of Vertebrate Zoology

The Museum of Vertebrate Zoology (MVZ) holds 1,825 field books, the smallest of the three collections. The archive has a close relationship with the faculty and students. Field notes are actively used and frequently referenced. They function more actively than traditional, closed collections in an archive. The field books are obtained through two sources: required deposit (students) and voluntary contribution (current and emeritus faculty).

The field books are managed in the archive, so description is usually at the collection level with finding aids in EAD. Most reference requests focus on geographic location and taxon, seeking specific data within field notes as opposed to seeking individual authors. In 2005, museum staff started creating item-level description to meet these reference needs in an in-house MySQL database, which was intuitive for the curatorial staff, but the data relied heavily on open text fields and could not be easily edited. This was not supportable over the long term, so the MySQL data were migrated to Archivists' Toolkit[9] as digital objects, and in 2015, the MVZ developed a new online portal for accessing field notes called the EcoReader. The EcoReader depends on MODS records created from Archivists' Toolkit. The EcoReader is more dynamic than its predecessor and uses Arctos, MVZ's information management system, as the name authority for geographic terms, allowing researchers to search at the county level for field notes.

MVZ has several qualities that have affected choices in description. Unlike the field notes of many institutions, MVZ's field notes are very similar in structure, frequently following the format established by Grinnell. This has enabled MVZ to produce a high level of description for specimen information and geographic location. This consistency of format

4

and small collection size also allow MVZ to focus on geographic description in a way other institutions might not be able to support.

As a result, there is a highly standardized description that falls to the county level to fit with Arctos. These are cross-walked to Library of Congress Subject Headings when going to bibliographic portals such as BHL. The result has been a demonstrable increase in reference requests, from 8 to 70 over a comparable period.

Case 2: American Museum of Natural History

The American Museum of Natural History (AMNH) has approximately 3,500 field books. Historically, they have been managed not only by the Research Library and Archive but also by the natural sciences departments themselves. Field books are primarily used in support of the specimen collections and were mainly created by the staff or donated to the museum with specimen gifts and purchases. The field book content not only supports the scientific research at the museum but also is used for supporting the broader mission of reaching museum audiences through exhibits, education and other forms of outreach.

Field books in the library are managed as rare books or archival collections, depending on how they were received. Books that are singled out as being of particular importance have been added to the rare book collection. For this reason, MARC has been a predominant method of description. The library staff has significant knowledge of MARC's nuances and has a documented history for using it to describe field books. It continues to be a useful way of describing both the collections and field books at the item level. MARC collection-level description has been a priority because it provides discoverability through online resources such as OCLC's ArchiveGrid[10].

AQ: 1

The library recognizes the importance and benefits of using the EAD schema, which is now an archival standard for providing important context. However, there is a challenge with the potential of overlapping description, as many materials are also described with the MARC collection-level records as well as EAD-encoded finding aids. EAD is being implemented whenever possible to describe the archival collections in ArchivesSpace.

AMNH has been able to take advantage of EAC-CPF, which allows for in-depth description of people and corporate bodies and, thus, can highlight the museum's unique history as part of numerous expeditions. The EAC-CPF schema provides a valuable way of describing important historical context. For this reason, AMNH has been focusing on EAC-CPF record creation for expeditions for several years and is active with organizations such as SNAC that work to aggregate EAC-related content and make it available to archives, libraries and museums.

Case 3: Smithsonian Institution

The Smithsonian has by far the largest of the three collections with over 8,600 field books. Like those of the other institutions, these field books are part of much larger archival holdings. Field books have historically been managed by either natural sciences departments or the Smithsonian Institution Archives. Field books deposited at the Archives are usually voluntarily transferred by the staff or as part of departmental papers. Field books primarily document specimen holdings, but the researcher use varies widely. Field books that are of primary interest to departments are housed in the respective department.

The Smithsonian Libraries and the Smithsonian Archives are two separate units, so when the Archives began to consider what method to pursue for item-level description, they did not have a strong tradition of MARC item-level structure and chose MODS for its adaptability and created an in-house database.

DLP
33,3

6

The Smithsonian Institution has been part of numerous expeditions, so EAC-CPF has been an important tool for highlighting relationships between entities. An expeditionary event contains many relationships between entities, often individuals participate in multiple expeditions and very often expeditions build on one another. In addition, the Smithsonian holds many field books that were created by the staff who worked for affiliated federal agencies. EAC-CPF records are helpful for clarifying these relationships.

Lessons learned

All three institutions described in this article are able to describe field notes both within the context of their collections and as individual bibliographic records. Larger institutions often face the task of reconciling legacy MARC records and workflows with current archival processing. Whether an institution begins with finding aids or item-level records, inevitably, the material will need to be described both ways.

The three institutions seek to create all the levels of description, but factors such as size, historical management of collections and variety of format affect what descriptive standards are feasible. Field notes are most often deposited in archives; their description usually stems from archival descriptive standards. Smaller archives looking to share records and content with library and archive aggregators such as Digital Public Library of America[11] must figure out how to create item-level description but often, do not have access to library catalogs or technical services and are, thus, ill-equipped to create MARC records. Yet, MARC is still the format standard for most library online aggregators.

Archives of any size may choose to use a MARC-mapped schema such as MODS that offers additional flexibility to suit local needs but will then need to work with online aggregators to adapt their workflows so that records can be shared successfully. It is the classic case of trying to fit a square peg into a round hole. The natural history museum archival community is growing, and inevitably, it will need to decide whether to continue trying to participate in aggregators that depend on MARC records, or if there are viable alternatives that are more supportive of smaller archival repositories containing fewer resources.

Presently, without access to online public access catalogs, some natural history museums rely on specimen information management systems, such as KE EMu[12] and Collection Space[13] to catalog field notes. The others use spreadsheets or only rely on finding aids, describing field notebooks in the context of personal papers. Access to library professionals is not universal in natural history museums, but the need to access the data from field notes is increasing. In developing this paper, the authors recognize the need for a more accessible platform to describe and make field notebook data available to researchers. Further, the efforts of the Integrated Digitized Biocollections (iDigBio)[14] have brought together library professionals and museum curators. The natural history community has a unique opportunity to collaboratively create a viable alternative and is working towards that goal.

In an ideal world, programs such as ArchivesSpace[15] would have more robust methods for describing individual items, and this will likely evolve. ArchivesSpace supports both item-level description and digital object record output, but the resulting record currently does not have the granularity needed to optimize the discoverability needed for field books in online aggregators. For now, it is necessary to leverage the data standard that best describes individual items, which is MARC, MARCxml, or at least a MARC equivalent. As libraries prepare for life after MARC as Bibframe[16] continues to be developed, institutions may find that the lifespan of MARC will cause item description to rely more heavily on other schema.

The institutions considered in the case study have begun to develop the answers to these questions by trying out methods founded in the information science professional standards and seeing what works and what does not. It is also important to recognize and capitalize on what is unique to an institution and its collections. AMNH and the Smithsonian Institution have both been avid early adopters of EAC-CPF, which provides a powerful tool for describing their rich expedition histories. There are presently no widely adopted best practices for describing field notes, but each has taken calculated risks to meet the needs of researchers who overwhelmingly require item-level description of field notes.

Grants such as the Council of Library and Information Resources Hidden Collections are also an important tool for discovering solutions; they permit institutions to dedicate staff time to trying out new methods to answer these questions. They empower archivists to take on the challenge of item-level description of materials that would benefit from the additional attention. Collaboration through these grants enables these institutions to develop models for describing field notes with varying resources. While locally tailored, other institutions can look to these use cases to determine what may work best for them.

By working closely with museum curators, library professionals have an opportunity to create an innovative agnostic tool for field notes. The scientific community has many such examples. While working with different information management systems, museums still contribute to aggregators such as the National Science Foundation-funded VertNet[17] portal or the collections sponsored by iDigBio. With the same models, field notebooks can also be made available to the scientific community in a more agnostic approach, divorced from MARC and realistic of the resource challenges facing natural history museums.

While analyzing these three use cases, the authors recognize shared challenges and overlap. However, the differences in their description requirements are not insignificant. For the natural history archival community to approach the description and access of the [breath](#) of field notes across the country, we will need to collaborate with the natural history curator community to develop an aggregator that relies on records beyond MARC.

Notes

1. www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-description-ead/encoded-archival-description-ead#.V02qO1UrK_4
2. www.tdwg.org/activities/ncd/
3. www2.archivists.org/groups/technical-subcommittee-on-eac-cpf/encoded-archival-context-corporate-bodies-persons-and-families-eac-cpf#.V02pCVUrK_4
4. <http://socialarchive.iath.virginia.edu/>
5. www.loc.gov/marc/
6. www.loc.gov/standards/mods/
7. <https://iptc.org/standards/photo-metadata/iptc-standard/>
8. www.biodiversitylibrary.org/
9. www.archiviststoolkit.org/
10. <https://beta.worldcat.org/archivegrid/>
11. <https://dp.la/>
12. <https://emu.kesoftware.com/>

DLP
33,3

8

13. www.collectionspace.org/
14. www.idigbio.org/
15. <http://archivesspace.org/>
16. www.loc.gov/bibframe/
17. <http://portal.vertnet.org/search>

References

- Greene, M. and Meissner, D. (2005), "More product, less process: revamping traditional archival processing", *The American Archivist*, Vol. 68 No. 2, pp. 208-263.
- Grinnell, J. (1908), *Letter to Annie Alexander*, Museum of Vertebrate Zoology Archives, Historical Correspondence, Berkeley.
- Grinnell, J. (1938), *Guide to Taking Field Notes*, Museum of Vertebrate Zoology Archives, Museum Methodology and Practices, Berkeley.

Corresponding author

Lesley L. Parilla can be contacted at: lparilla@gmail.com

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQau— Please confirm the given-names and surnames are identified properly by the colours.

■=Given-Name, ■= Surname

The colours are for proofing purposes only. The colours will not appear online or in print.

AQ1— Please spell out OCLC.