APPENDIX 7 - PASTPERFECT CASE STUDY

PastPerfect 5 & Online Case Study

An Assessment Summary and Cleanup Plan for Deschutes County Historical Society

v. July 29, 2023

This project was made possible with funding from the Oregon Cultural Trust.

Created by Rachael Cristine Woody, Founder + Director of Relicura™ LLC

Table of Contents

Introduction

The data in your Collections Management System (CMS) is important. The quality of it impacts how well you can manage your collection and provide access to it. Any cultural heritage organization that's existed longer than a decade knows the challenges that can exist with changing technologies, shifting best practices, reliance on volunteers and interns, and staff transitions. Any one of these items can contribute to inconsistent data creation in a Collections Management System (CMS) and it's safe to say 99% of heritage organizations each have a part of their collection catalog that needs data refinement (aka data cleanup) in order to meet current needs and best practices. It's important to known and get comfortable in the knowledge that messy data exists and we all have messes we feel embarrassed about regardless of whether we're the ones that created the mess. Another truth to get comfortable with is that attempting data cleanup can absolutely feel overwhelming. Where do you start? How much is there? How long will it take? And, How do you even do it? These are big questions and this isn't a topic typically covered in graduate school or post-grad workshops. So, if you're reading this and you don't know the answer to any of the posed questions in this section: great. You're in the right place.

> **Any Data is Better than No Data:** *This section focuses on data cleanup and enhancement, but is in no way implying that data should be 100% correct and complete before being published in the catalog. Any data (any point of access) is better than no data. No data means the collection is completely inaccessible and that's worse than messy data.*

The Case Study

The following is a case study-style report of how Relicura™ LLC worked with Deschutes County Historical Society (DCHS) to perform an assessment of their PP 5 and PP Online data, and create a data cleanup plan. This case study is intended to be paired with the Oregon Heritage Commission produced *A Guide to Collection Data Cleanup*. As we make our way through the case study sections of the Guide will be referenced where you can find more information on a particular topic. While this study is intended to provide you with specific examples, some of the non-essential detail may be omitted for clarity or out of deference to DCHS.

Desired Outcome

The most obvious answer is: Clean and complete data! But, it's more nuanced than that. The goal is data cleanup performed in an effective way and at a sustainable pace. Yes, we'd love to have 100% of our data clean at the end of this, but that's the ultimate "future" goal. Not our "now" goal. We need to make sure we make best use of limited staff time and that the pace is sustainable so as not to overburden staff or have the work done at the expense of other priority items.

> **Good Data:** *For more information on the importance and helpfulness of "good data" please see the "The Power of Good Data" section in A Guide to Collection Data Cleanup.*

Cleanup Limitations

Every CMS will have data cleanup tools and limitations. As both software and user expectations evolve, data cleanup *should* become easier and easier for us, however, no tool is the same. Nor, is any collection or data mess the same. It's important to understand both the possibilities and limitations of your CMS in order to determine the cleanup strategies available to you. The PP product is limited in its support of data cleanup tools or functions. As a result, the majority of data cleanup has to take place at the item-level record and progress item by item. There is an option for bulk data conversion, but it must go through PP and is (currently) billed at $100/hour. Unfortunately, the impact of this limitation disempowers the PP user and can prevent meaningful cleanup from taking place.

Capacity and Realistic Expectations

The museum, archives, and cultural heritage organization sector suffers from chronic under-funding and low staffing levels. As such, the staff in place are doing their best to perform all required duties for collection care, management, and access—and they are already at capacity. With that in mind, the only way data cleanup can happen is if: 1. Capacity is added via additional staff, or grant-funded or otherwise temporary staff; and/or 2. The current slate of priorities are changed and something is taken off the to-do list—even temporarily—in order to accommodate the time needed for data cleanup.

In addition, know that <u>data cleanup is likely a long-term, multi-year effort</u>. In order to proceed it's important to break the cleanup into achievable, short-term chunks. This approach requires performing data assessment up front and the intentional partitioning of cleanup activities for future work. There are additional benefits to this approach: 1. It supports easy "grab and go" cleanup for a window of staff availability or a trusted intern; and 2. It can serve as a framework for a potential grant or donor-funded project.

Finally, sometimes we just need an easy win. With the pre-identification of cleanup areas in your data, you'll have a sense of how pervasive a particular data mess may be. In many instances there are small and confined data messes that require only 5-15 minutes to correct. For example: Updating (or consolidating duplicates of) a creator authority record. Or, perhaps block out 1-hour a week to tackle any data cleanup sets that can be performed in that amount of time. Employing one or more of these tactics will greatly support your overall data cleanup effort and help keep the work moving at a sustainable pace.

Limited Capacity and Realistic Expectations Requires Us to:
- Perform an assessment of your data and identify areas of cleanup.[1]

---

[1] Information and guidance on how to perform data assessment please see *A Guide to Collection Data Cleanup*, section titled "A Review of Data Cleanup Strategies."

- Carve up the areas of cleanup by what needs to be done, by field, and (if there's a lot) by record set. A record set can be an arbitrarily decided number or another narrowing criteria factor.[2]
- Pair each cleanup batch with guidance on what's needed and how the cleanup can be performed.[3]
- Identify the quantity of records per cleanup batch, estimate the time it would take to perform the data cleanup (factor in research if the data is missing), and assign a priority level.[4]
- Finally, indicate if the work can be performed by a volunteer, intern, or staff member. This will depend on the knowledge and capability of the person as well as who is trained to use your CMS.

It is entirely possible to have non-CMS users assist with data cleanup by performing necessary research, imaging the collection, and providing data for review in a spreadsheet format for a staff member to copy/paste or import into the CMS at a later date. This case study and many of the cleanup strategies recommended in *A Guide to Collection Cleanup* rely on spreadsheets for data analysis, data preparation, and (if possible) data import. The *Data Cleanup* section in this case study provides further examples on how to incorporate other team members in the cleanup process.

The Assessment Summary

The Deschutes County Historical Society (DCHS) uses PastPerfect 5 (PP 5) as the primary CMS. Currently, there are approximately 75,000 records in the database with 559 of those records provided through PastPerfect Online (PP Online). This study will offer a summary assessment for both (PP 5 and Online) and the cleanup specified for each.

The best way to evaluate data is as an aggregate and when using PP, the only way to effectively do so is with generating a query and exporting the results as a spreadsheet. A spreadsheet view will display each entry and all of the data in each field. This makes it easier to spot missing data, incorrect data formats, and any data inconsistencies.

> **Data Review via Spreadsheet:** *For more information on the importance and helpfulness of "good data" please see "A Review of Data Cleanup Strategies" section in A Guide to Collection Data Cleanup.*

In PP it's currently impossible to perform an export of more than a few hundred records at a time. In order to review data in the aggregate for DCHS, queries were ran in order to generate spreadsheet reports that could provide an aggregate view of the selected records. Given that DCHS

---

[2] See the examples in this case study for how we divided the data cleanup work.

[3] What's needed will depend on the field and the perceived data gap or discrepancy. Best practices can help inform what's needed and other items might require research into the collection or item.

[4] How you determine the priority level is up to you. Use this case study for examples in what was prioritized and why.

has 75,000 records, this review could be performed on the 500 PP Online records and using a sample size of PP 5 records from each of the main collecting areas in PP: Archive, Library, Object, and Photograph. The gaps in data outlined below were present in the data reviewed and are likely indicative of missing data across the collections.

Data Standards

The following data standards were used to help conduct the assessment[5]: Dublin Core™ (DC) Metadata Element Set[6] and Cataloging Cultural Objects[7] (CCO). DC is the most universally adopted with its straightforward approach to capturing core data across all collection types. It's also the de-facto standard schema to support Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)—the ability to have collection data "harvested" and placed in a collaborative portal. Each schema has designated fields that are considered the minimum required fields. In addition to what fields to use as a schema, there are also descriptive standards (and subsequent fields) to consider. For many museums and cultural heritage organizations the standard typically used is CCO or (and especially if archives are present) DACS.

> **Data Standards:** *For more information on data standards, please see the "An Overview of Data Standards" section in A Guide to Collection Data Cleanup. For an introduction to the language used while discussing collection data, please see the "Establishing a Common Language" section.*

CCO minimum fields to use:
- Work Type
- Title
- Creator; Role (controlled vocabulary- can be local, an authority, or both)
- Creation Date
- Subject (use of Nomenclature for a controlled vocabulary)
- Current Location
- Measurements
- Materials and Techniques (use of Nomenclature for a controlled vocabulary)

The Cross-Walk from Data Schema, to Descriptive Standards, to PP Fields

The following grid is a cross-walk[8] among the data schema fields (per Dublin Core™), the two descriptive standards (DACS and CCO), and the corresponding field in PP. The numbers found in the DACS and CCO columns refer to the chapter and section of where you can find more

---

[5] For more information on data standards, please see *A Guide to Collection Data Cleanup,* "An Overview of Data Standards."
[6] DCMI: Dublin Core™ Metadata Element Set, Version 1.1, accessed February 15, 2023, via https://www.dublincore.org/specifications/dublin-core/dces/.
[7] Cataloging Cultural Objects (CCO), *A Guide to Describing Cultural Works and Their Images*, accessed February 28, 2023, via https://vraweb.org/wp-content/uploads/2020/04/CatalogingCulturalObjectsFullv2.pdf.
[8] Cross-walk in this case is used in reference to the equating of standard when there are multiple standards or best practices in play. For example, one field title in Dublin Core may be called something else in PastPerfect. Another example is that some standards may require specific fields not included in other standards. By viewing the standards side-by-side in this grid, you will be able to see which fields are considered required and what that field name is across the different standards, and in PastPerfect's case, the name of the field in the database.

6

information on the field via the standards reference guide. Some boxes are absent of content when there's not a corresponding field to use or standard to reference.

Dublin Core Fields	DACS Content Standard	CCO Content Standard	PastPerfect
Identifier			Object ID; Other Number
Publisher	Reference Code Element (2.1); Name and Location of Repository Element (2.2)	Current Location (5.2.1)	Collection; Home Location
Title	Title Element (2.3)	Title (3.1)	Object Name; Title
Creator	Name of Creator(s) Element (2.6); Administrative/Biographical History Element (2.7 added value)	Creator (2.2.1); Creator Role (2.2.2)	Artist, Author, Creator, Photographer
Date	Date Element (2.4)	Creation Date (4.2.3)	Date
Type	Extent Element (2.5)	Work Type (1.2)	Type
Description	Scope and Content Element (3.1)	Description (8.2.1)	Description
	Conditions Governing Access Element (4.1)		
Language	Languages and Scripts of the Material Element (4.5)		n/a
Subject	Access points (See Overview of Archival Description)	Subject (4.1); Classification (7.2.1)	People, Subjects, Classification Search terms
Rights Management	Conditions Governing Reproductions and Use (4.4 added value)		Notes & Legal > Legal Status or Web Rights
Relation	Related Archival Materials (6.3 added value)		Relations
Format		Measurement (3.2.1); Materials and Techniques (3.2.2)	Dimensions; Material; Technique
Current Location		Current Location (5.2.1)	Location > Temporary location

Overview of Data Gaps in PP Online Data

All 559 PastPerfect Online records are from the Photograph collection module in PP. These records are fairly modest in data capture with the consistent use of the following four (4) PastPerfect fields:

- Object ID (OBJECTID)
- Collection
- Type
- Description (DESCRIP)

When comparing the current PP Online fields used to current data schema standards (Dublin Core™) we find the following fields absent:

- Identifier (this is the individual item number)
- Title (of the item being cataloged)
- Creator (controlled vocabulary- can be local, an authority, or both)
- Date (creation date)
- Format (use of Nomenclature for a controlled vocabulary)
- Language (when applicable)
- Subject (use of Nomenclature for a controlled vocabulary)
- Rights (any copyright or other retained rights)
- Relation (when applicable, indicate a relation to other known items in the DCHS or peer museum collection)

Overview of Data Gaps in PP 5 Data

Across all four collections the following data gaps were found (see below). Many of them are part of the data elements (fields) that should be present in a complete record.

Arch Records
- CatBy field infrequently used
- Date field infrequently used
- EarlyDate and LateDate columns usually have a "0"
- HomeLoc has at least one row without an entry
- PubPlace and Publisher columns are inconsistently used – either both, one or the other, or neither are filled out
- Row 6 doesn't have an entry in the Title column
- Sterms [Search Terms] field barely used
- Subject field is barely used (probably less than 10 entries
- Rows 72-76 are missing info in the ImageFile column (possibly more)

Lib Records
- Creator field infrequently used
- CallNo field infrequently used
- CatDate field infrequently used
- CatBy field infrequently used
- Date field infrequently used
- ImageFile field infrequently used
- Language field infrequently used
- PhysDesc field infrequently used
- PubPlace field infrequently used
- Publisher field infrequently used

Object Records
- AccessNo field infrequently used
- Creater field infrequently used
- Row 1 and 2 of CatDate field are missing info
- CatBy field infrequently used
- Date field infrequently used
- Row 4 HomeLoc missing info (possibly more)
- ImageFile field infrequently used
- People field infrequently used

Photo Records
- CatBy field infrequently used
- Date field infrequently used
- Descrip field infrequently used
- HomeLoc field infrequently used
- ImageFile field infrequently used

The Data Cleanup Summary

The PP product is limited in its support of data cleanup tools or functions. As a result, the majority of data cleanup has to take place at the item-level record and progress record by record. With this in mind, it's critically important to spend time analyzing the data, identifying areas of cleanup, prioritizing cleanup tasks, and creating a strategy before any major cleanup effort is made in PP.

This data cleanup plan will outline how DCHS staff can effectively conduct data remediation of the 559 online records and will include coverage of the cleanup process, strategies for approach, suggested prioritization, possible tools or strategies to use, and offer any additional advice to help ensure a smooth and supportive cleanup process.

9

The Cleanup Strategy

The record-by-record cleanup strategy is the strategy we selected to inform cleanup actions. This strategy was chosen because PP doesn't support data changes via import of a spreadsheet, nor support any data cleanup tools. Record-by-record cleanup can be a slow and sometimes frustrating process, so it's incredibly important to develop a data cleanup plan that breaks the data cleaning tasks into achievable sections. In order for this work to be sustainable, a doable pace will need to be established.

Cleanup Strategies: *For a refresher on data cleanup strategies please see the "A Review of Data Cleanup Strategies," and the "Practical Application of Data Cleanup Strategies" sections of A Guide to Collection Data Cleanup.*

Leverage a Spreadsheet for Data Work Pre-PastPerfect

While data must be directly entered into PP, it's often easier to gather the information, review it, and create the data ahead of time. This allows time for any research or file pulling that might need to happen, and it can help with data quality control. By entering or editing the data into the spreadsheet first you'll be able to see patterns more easily and benefit from data already created.

For example, many collection items share affinities such as the same creator (Creator field), descriptive details (Description field), and subjects (Subject field). Using a spreadsheet as your template will allow you to easily copy and paste relevant data and help to ensure you're providing a consistent level of detail for each related entry. Additionally, spellcheck can be ran across the entirety of entries, and a secondary review can be conducted efficiently with the spreadsheet before any (potentially erroneous) data enters PP.

Essentially, using a spreadsheet as your draft template helps to save you time and supports quality control of the data before it ever touches the database. Finally, it speeds up the eventual data entry time as the data can be transposed (copy/pasted) into each record with full confidence.

Spreadsheet Template Summary of Benefits:
- Allows you time to gather the information you need.
- Supports the easy comparison of multiple records and makes patterns easier to spot.
- Facilitates the copy/paste function of shared data elements across records, lessening the need to recreate repeat data and helps to ensure consistency.
- Helps to maintain correct field use with each field present as a column header in the spreadsheet, and reenforced with each entry correctly distributing data into each field.
- Provides further quality control mechanisms such as spell check and review of entries before data ever enters the system.
- Speeds up the eventual data entry process into PP.

Approaches to Cleanup

When approaching the spreadsheet for pre-PP work or once ready to enter data into PP, the following two approaches can be selected and used:

Approach A: Fill out each record and strive for record completeness. This looks like working in each record entry and filling out or editing data in each of the required fields. This approach is for brains that like to fully complete one item (record) before they move onto another.

Approach B: Pick a field and work on that field exclusively across all of the records. Then pick the next field and repeat. This approach is for brains that like to see quick progress and the ability to focus on one piece of data at a time.

Either approach is effective and with limited downsides so it really is up to you and how your brain likes to work!

Prioritization and Workflow

Approach A Prioritization

Prioritization can look like selecting a batch of records that are the most interesting and/or most important records. By selecting a prioritizing criterion, you can focus your efforts on records that are immediately beneficial to the heritage organization. This also helps to keep the scope small and achievable, which is an important practice when working with hundreds (and sometimes thousands) of records. Given the content of the PP Online records, we recommend that the priority batches are created based on the shared location or event depicted in the photographs. This will allow the cataloger an opportunity to build consistent, quality descriptions.

Approach A Workflow Steps

1. Select your criteria. Potential areas of focus could be a portion of the collection that is highly request, could aid in an upcoming project, or is "hidden" and could benefit from improved data.
2. Narrow the scope of the records you intend to look at and break it into batches of no more than 100 records (aka items) at a time.
3. Run a query in the system that will pull up the records that match your criteria and select the first 100 item records to export into a spreadsheet file format.
4. Review the spreadsheet and remove any entries that meet best practices and <u>don't</u> need data cleanup. Then highlight each field (column) where every record entry remaining could use further review and data cleanup or enhancement.
5. Make your way through each record (represented as a row in the spreadsheet) and cleanup or fill-in missing data. As you complete an entry, fill the row in a light gray color so that you can easily track which entry you're working on and what's been completed thus far.
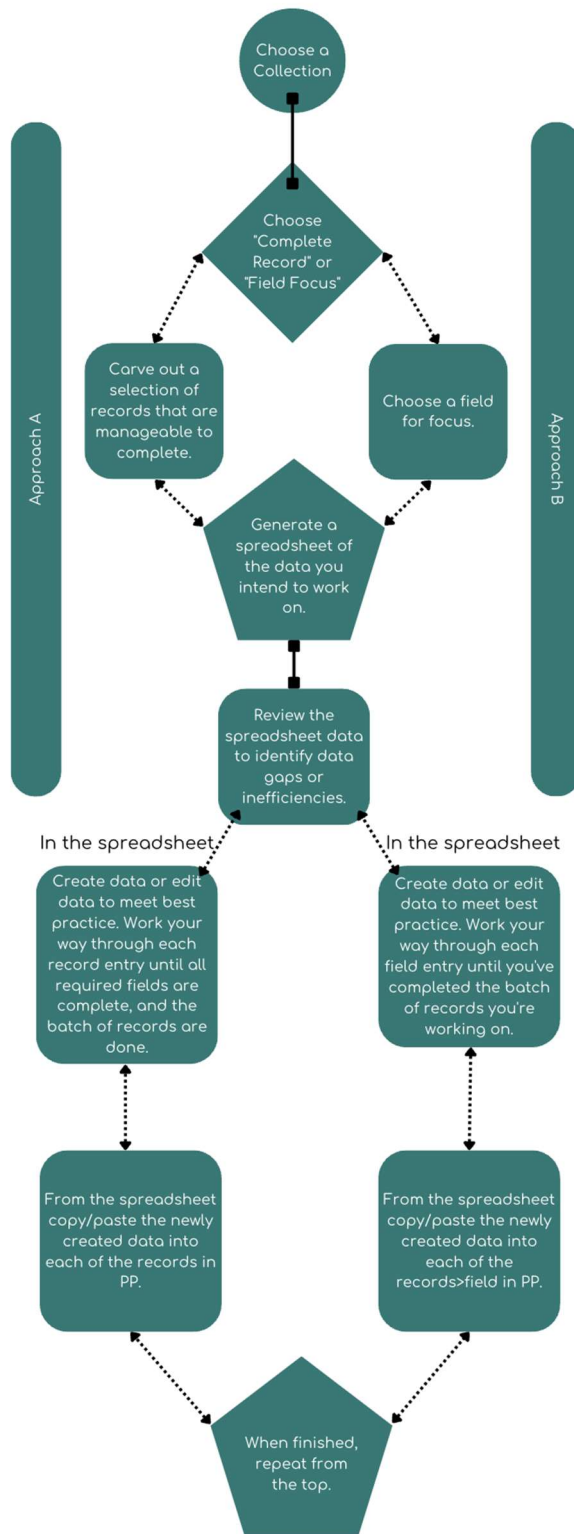
11

6. Once you've completed your work on this set, transfer the cleaned or enhanced data into PP by pulling up each record and copy/pasting the corrected data into the system. (Unfortunately, PP doesn't support data import for cleanup purposes).

Approach B Prioritization

Prioritization is inherent in this approach as you can determine the fields you work on first as priority fields. These can be fields that are the most impactful to good search results such as Date, Description, or Subject. Selecting fields by this prioritization will help ensure your work is immediately focused on the most impactful areas of the data. The following sections titled *An Assessment of the Data* and *An Assessment on the Completeness of Record*, offers a recommendation and priority level for each of the required fields. They are listed here for reference:

Existing Data Priority Areas:
- Object Name (High)
- Description (High)
- Collection (Low)

Absent Data Priority Areas:
- Identifier (Low)
- Publisher (Low)
- Title (High)
- Creator (Medium)
- Date (High)
- Format (and Dimensions) (Low)
- Language (Low)
- Subject (High)
- Rights (Medium)
- Relation (Medium)
- File Name (Low)

Approach B Workflow Steps

1. Select your priority area of focus aka the priority field.
2. Narrow the scope of the records you intend to look at and break it into batches of no more than 100 records (aka items) at a time. For this approach we can start at the beginning of your accession numbers.
3. Run a query in the system that will pull up the records that match your criteria and select the first 100 item records to export into a spreadsheet file format.
4. Review the spreadsheet and remove any entries that meet best practices for your field of focus and don't need data cleanup. Then highlight the priority field (column) to focus on for review and data cleanup or enhancement.
5. Make your way through each record (represented as a row in the spreadsheet) and cleanup or fill-in missing data for the priority field. As you complete an entry, fill the row

12

in a light gray color so that you can easily track which entry you're working on and what's been completed thus far.

6. Once you've completed your work on this set, transfer the cleaned or enhanced data into PP by pulling up each record and copy/pasting the corrected data into the appropriate field in the system. (Unfortunately, PP doesn't support data import for cleanup purposes).

Working with a team? Regardless of which approach you use, follow the same steps but instead make sure the spreadsheet being worked on is a shared spreadsheet using OneDrive, GDrive, Dropbox or similar document sharing and collaboration tool. Make sure everyone follows step #5 and grays out the rows they've completed so that work isn't duplicated.

Or, you can provide each member of the team (volunteers included) their own spreadsheet with set of 100 to work on.

If the other team members have access to and know how to correctly enter their data in to PP, then they can complete step #6 on their own. If not, then have the team submit their completed spreadsheets to you for PP input.

Choose a Collection

Choose "Complete Record" or "Field Focus"

Approach A

Approach B

Carve out a selection of records that are manageable to complete.

Choose a field for focus.

Generate a spreadsheet of the data you intend to work on.

Review the spreadsheet data to identify data gaps or inefficiencies.

In the spreadsheet

In the spreadsheet

Create data or edit data to meet best practice. Work your way through each record entry until all required fields are complete, and the batch of records are done.

Create data or edit data to meet best practice. Work your way through each field entry until you've completed the batch of records you're working on.

From the spreadsheet copy/paste the newly created data into each of the records in PP.

From the spreadsheet copy/paste the newly created data into each of the records>field in PP.

When finished, repeat from the top.

14

An Assessment of the Data

The following sections are an assessment of the data contained in the four fields used in DCHS's 559 PP Online records: Object ID, Object Name, Collection, and Description. Each of these fields has a corresponding descriptive standard to refer to and has helped to inform this assessment.

> **For Brains That Like Spreadsheets:** *If you prefer to see this information in spreadsheet format, please see PP Online Cleanup Actions 06222023.xlsx document on the Oregon Heritage Commission website* https://www.oregon.gov/oprd/OH/pages/technicalresources.aspx#datacleanup

---

Object ID

For the bulk of the PP Online data the numbering appears in this format: YYYY.NNN.NNNN

Y = Year and N = Number

However, there are variations. Those variations are:

YYYY.NNNX.NNNN (in this case with a literal "X" inserted); example: 1999.029X.0043

YYYY.NNN.NNNNL and YYYY.NNNX.NNNNL (L = letter, and it's added at the end); examples: 2003.000.0226A and 1992.018X.0001A

Letters
Sometimes "A" and "B" is used to refer to the front and back of image. Other times A, B, and other letters of the alphabet are used. The current practice at DCHS is to use letters (in alphabetical order) to indicate multiple parts of one unit.

Example: A letter with an envelope. The letter receives an "A" at the end of its Object ID, and the envelope receives a "B".

A few Object IDs from the 1980s are: YYYY.NNN.NNNN.NNN. These numbers are a combination of "Object ID" number and the "Other Number".

Finally, there's: YYYY.DHS.NNNN; example: 2011.DHS.0141. The "DHS" is used in reference to the Deschutes County Historical Society.

Determination: While the varied numbering system has some quirks, it ultimately doesn't prohibit (to a detrimental extent) the usability of the data and the findability of the item records. With this in mind, it's recommended that this data is left as-is. While it's important to document the different numbering schemas used throughout the years, the data doesn't need to be updated. The only exception to this determination is if there are duplicates of Object ID numbers being used as that can impact the integrity of the data with risk of conflating two separate items.

Action: No action.

Priority: N/A

Object Name

Across the 559 records there are the following terms found in the Object Name (OBJNAME) field:

OBJNAME List
- File, digital (2)
- Image, digital (13)
- Lithograph (1)
- Negative, Film (24)
- Negative, film (22)
- Photocopy (1)
- Photograph, Cabinet (1)
- Postcard (1)
- Postcard, photo (73)
- Print, photographic (119)
- Print, Photographic (301)
- Tintype (1)

There's some consolidation that can occur in the OBJNAME field. For example, the sets listed below are most likely duplicates either due to a capitalization difference or a slight variation on identification.

16

File, digital (2)
Image, digital (13)

Negative, Film (24)
Negative, film (22)

Postcard (1)
Postcard, photo (73)

Print, photographic (119)
Print, Photographic (301)

Determination: This data set is at a manageable-level for cleanup of the overlapping Object Names. In order to improve findability and maintain good data hygiene, it's recommended that these terms are reviewed and consolidated.

Action: Review Object Names that are similar and consider consolidating them into one.

Priority: These cleanup activities will improve search result quality and are an "easy" win regarding data cleanup effort. With both of those in mind, this is a **high priority** item.

Collection

There are six records assigned to the Deschutes Pioneers' Association[9]:

- 1980.050.0059.001
- 1980.050.0059.003
- 1980.050.0106.040
- 1980.050.0106.067
- 1980.050.0106.089
- 1980.050.0108.011

DCHS would like to retain what came over as Deschutes Pioneers' Association.

Exhibit Materials
Additionally, there are 16 entries listed as "Exhibit Materials". DCHS suspects that these materials are exhibit materials such as signs and other non-original object materials. This supposition is based on the known practices of early DCHS staff who used PP to help track non-original object items; such as furniture, supplies, and exhibit materials. Is this an accurate Collection

---

[9] In the early 1940s or 1950s the Deschutes Pioneers Association was established and operated as a small history museum for the area. When DCHS was created in 1975 there was a coalescing of efforts and the Association transferred their collection holdings to DCHS. The Association continued to exist as an organization until January 2022, when the Association dissolved and transferred their remaining assets to DCHS.

identification to use? Was it meant to represent a loan, or items from the DCHS collection used for an exhibit? The answer may help us determine if and how we should approach this data.

Collection Names

In the larger PP 5 data sets there is a variation in Collection name assignment. Some are a variation of DCHS. Others are names from the creator or donor of the collection. Either is appropriate, but should be definitively decided and formalized into DCHS policy.

Determination: The PP Online data set has minimal cleanup needed, but does open the door to policy questions for all DCHS data.

Action: There are a few actions to consider. At a minimum for the PP Online data, it is determined that "Exhibit Material" can be disposed of.

1. Decide on a DCHS policy regarding the assignment of PP "Collection" names. Are they all DCHS? Are they identified by "creator" or "donor" names? What are the criteria for selecting and assigning a collection name? And, depending on the answer, is there another PP field where DCHS can capture either DCHS or the creator/donor collection name—as an example?

2. Review the Collection entries to determine if any need to be revised.

3. Additionally, commit to policy a statement on what is *not* considered a collection item and therefore does *not* go into the collection database.

4. For items such as "exhibit materials" consider "deaccessioning" or otherwise removing those items from the PP database in order to limit confusion regarding collection materials.

Priority: While the accuracy of this data is important, it can be considered a low priority.

Description

The descriptions available across the 559 records vary in length, style, and quality. The description field is typically the largest field and takes the longest time to construct and review. While having any data can meet the minimum data schema requirements, many of the entries conflict with descriptive standards. The following data quality issues were found in these records:

- Abbreviations
- Fragmented sentences
- A list of items in the photograph instead of a narrative description
- Inconsistency in length or quality of content provided across similar image sets
- Informal notes on where the item was found
- Concatenated data

- Referencing item numbers that don't appear to be captured in the PP Online records
- Informal commentary on the quality of the item
- Donor notes that may not be appropriate for inclusion
- Provision of an entire history of the photograph's point of interest versus a focus on what the photograph has captured
- Grammar and spelling errors (throughout)

Abbreviations

Abbreviations should be fully spelled out for both consistent quality and to support effective keyword searching across multiple fields (including the Description field).

Examples:
- "photo" for photograph
- "B&W" for "black and white"
- "Hwy" for highway
- "Lbr Co" presumably lumber company?

Fragmented Sentences

Fragmented sentences make descriptions difficult for general users to read and limit the multi-purpose use on behalf of the museum. Descriptions are used to help inform exhibits, research, and curriculum (to name a few examples) and the quality of description can impact the usability of the data.

Example: *George Bradetich place: color photo of horse drawn equipment, potato harvest. Potatoes on ground, in mesh buckets, in bags.*

Corrected Example: This is the George Bradetich residence during potato harvest. This color photograph depicts horse drawn equipment and potatoes on the ground, in mesh buckets, and in bags.

Lists

Other descriptions appear to be more list-like rather than descriptive. Some entries don't even contain a "." at the end of the sentence. This may not impact searchability, but it does impact user interpretation and experience.

Example: *Pilot Butte Inn: lawn, patio, tables, umbrellas, director's chairs*

Corrected Example: This scene was taken at Pilot Butte Inn. Pictured in the foreground is the patio and surrounding lawn, complete with tables, umbrellas, and director's chairs.

Inconsistent Length or Quality Among Related Image Sets

The majority of the descriptions are fairly modest—which is both fine from a best practices perspective, as well as understandable from a staff capacity perspective. However, keep in mind the balance and consistency of longer descriptions. For example, this longer description is also connected to Pilot Butte and contains a good chunk of information that should be considered for inclusion in all of the Pilot Butte item records. As many of these images are related, it may be helpful to view the descriptions all together and then standardize the information that is provided in each image.

> *SW Corner of Wall and Newport with Pilot Butte Inn seen on right edge of photo. Wall Street in Bend in 1904, showing post office. In 1885, a filing was proposed for a post office to be called Farewell Bend. The post office was established January 18, 1886 with only the name of Bend allowed as there was already a post office of the former name in a community along the Snake River. In 1904 the Bend Post Office was located at the southwest corner of Wall and Newport. The corner of the Pilot Butte Inn porch appears on the right.*

Informal Notes on Where the Item Was Found

This description includes an informal note on how the item was found. In previous cataloging practices, catalogers opted to add the majority of collection information in the Description field instead of using more specific and appropriate fields. This type of note is more appropriate as an internal note as it serves no findability or usability purpose to those external to DCHS.

> *Group photo of children,adults, and dog in woods. Possibly logging camp residents. Backing says, "Rowena M. Hogan 121 Oregon St. Bend, Ore." From The Hogans Photo Studio (JJ and Rowena M Hogan).. portrait and framing. Framed photo was found in back cupboard in Historical Center 05/2003*

Concatenated Data

This and a few other entries appear to have concatenated the data in a way that has duplicated the content.

Referencing Item Numbers that Don't Exist

This example also references a related item with No. 74. As this number doesn't appear in any of the PP Online data, how do we know what this number is referring to? If a connection can be made then it may be worth keeping this data. If not, then this data should be removed to limit future confusion.

> *Wagons; Grain; Harvesting; Horses; Farm Equipment. Farm wagons with horses, grain harvesting, 1919; farm equipment. Harvesting, 1919. Farm equipment in the field and horses with wagons. Grain harvesting, 1919. Farm equipment in the field and horses with wagons. Horses with wagons, farm equipment, involved with grain harvesting, 1919. Harvesting grain, 1919. Horses with wagons shown. Copied by Don Ellis from Roy Newell Collection, 1980. Print 3.5" x 5"*

> *Same as No. 74: Railroad Day, Downtown Redmond Sept 30, 1911.*

Informal Commentary on Quality

This example includes references to a related item and number, and it offers informal (subjective) commentary not consistent with descriptive best practices.

> *Big snow, 1919, downtown Redmond, Oregon; same as No. 104, but better.*

Donor Notes

Unless specifically stated in the donor's Deed of Gift that public acknowledgement of the gift should be made in the catalog, donation information is otherwise internal-only information. This is an area DCHS can develop policy around. For example, there can be inclusion of a field to capture a collection name or a credit line.

> *Brooks Scanlon Lumber Co. Logging Camp #1, Wet Weather Springs. Camp families pose in Sunday best, with camp homes, logging office and 1920's automobiles in background. Donated by Delmar Davis. Related to photo#1980.050.0059.001.*

Provision of Entire History

This description captures an entire history which is unusual for an item-level description.

> *St. Charles Hospital, downtown Bend, Shrine donated by Klondike Kate (Kate Rockwell), 1937. Article published in Bend Bulletin December 9, 1937 states that the grotto was constructed by Ray Williams and "the impressive grotto, some 15 feet high and holding a statue of St. Joseph and the Christ child, has been built against the new wing of the hospital". According to the article, Klondike Kate (referred to here as Mrs. John Matson) donated the grotto after making a promise to Sister Louise years before the installation. All stones used were collected by Kate, outside of a few purchased from local rock collectors. In another article from July 1, 1957, it was reported that the St. Joseph statue in the grotto was smashed. "A statue in a niche of gemstones at St. Charles Memorial Hospital was smashed Friday afternoon and police were called on to investigate. The state was that of St. Joseph and was given to the hospital a number of years ago by the late "Klondike Kate", Mrs. W.L. Van Duren. The cemented stones forming thje acover were not damaged, but the plaster of which the statue was made was broken into fragments. The vandilism occurred Friday afternoon, possibly around 3 pm. The statue was in front of the old hospital unit. Two men were seen on the hospital grounds at the time and attendants indicated they may have been intoxicated. Possibility that the vandilism might be associated with the recent entry of St. Francis Catholic church and telephoned threats to the pastor were considered by officers. It was learned that Trinity Episcopal Church was entered at the same time that the Catholic church was burglarized. All candles on the Episcopal Church alter were lighted and burned out. Entry of the Episcopal church was not discovered for several days, the pastor being out of town."*

Determination: Descriptions (by the very nature of the field) are the largest data set to review and revise. Improvement in this field can also be the most rewarding because it yields increased user engagement and can add to staff time-savings in reference and product (exhibition, research, etc.) output.

Action: Identify and prioritize chunks of data for cleanup. Specifics on how to break out this type of data cleanup into achievable chunks will be covered in the next section: *A Plan for PastPerfect Online Data Cleanup.*

1. Prioritize cleanup of the Description field for the 559 Online records.

2. Consider a policy for using a collection name or credit line.

Priority: This area of cleanup will take the longest time as it requires reviewing, editing, and creating narratives. But, it's also one of the most important areas of the record as it describes what's depicted in each photograph. This is considered a **high priority**, but one that should be approached in achievable batches.

An Assessment and Plan on the Completeness of Record

Data Schema Standards

Data schema standards outline which data elements (fields) should be present in a complete record. When comparing the current PP Online fields used to current data schema standards (Dublin Core™) we find the following fields absent:

- Identifier
- Publisher
- Title
- Creator
- Date
- Format (and Dimensions)
- Language
- Subject
- Rights
- Relation

The following section will outline an approach to data enhancement for missing or anemic data. It will describe the potential impact each field has on item management and use, share any relevant PP 5 data notes, and offer a recommendation with priority level assignment.

22

Identifier

Identifier: The identifier is an individual item number. For object-based collections it is typical for the Accession Number to also serve as the Object ID. However, anytime there's a multi-item accession—more typical of Photograph and Archives collections—there's the need to apply individual item numbers as well. These can be based on the Accession Number, but should always be unique.[10]

Example A: a tea set

The tea set is composed of dozens of pieces, but the set as acquired as a whole and accessioned as a whole, receiving just one official accession number: 20120706.11

If item-level or component-level cataloging needs to occur—as is typical for many cultural heritage institutions—then the accession number remains the same, but the component can receive its own object ID: 20120706.11.a.

Example B: a photograph collection

A photograph collection is comprised of 100s of photographs. The collection received an accession number: 20230827.01

If each photograph is cataloged at an item-level, the accession number (20230827.01) remains the same, but each photograph receives its own object ID, starting at 001, 002, and so on.

DHCS: The PP Online data captures the Object ID, but is lacking the Accession Number.

Potential Impact: An absent Accession Number may not hinder discoverability but it could impact collection management especially in situations where staff memory or institutional knowledge is no longer available. The missing Accession Numbers can also hinder data migration as it's often the data that helps tie related records (such as Conservation or Exhibition Loans) together especially if the CMS-assigned identifier (e.g. record ID) is impossible to migrate with the related data.

PP 5 Data: In reviewing sample Photograph collection data in PP 5, it appears both Accession Number and Object ID was used in nearly every entry.

Recommendation: If and when DCHS intends to participate in a collaborative portal this will become a higher priority. Until then, this is a low priority.

---

[10] Unique numbers are required in order to avoid accidental merging or conflating of two separate items.

Publisher

Publisher: The publisher field is intended to capture the person or organization who is publishing these materials to the catalog. In other words, the repository name. While this may seem obvious when cataloging material that all belong to the heritage organization, it's a great practice to get into as it will be necessary when contributing collection content to a larger collaborative portal—such as Northwest Digital Heritage or Digital Public Libraries of America.

Potential Impact: While there's no immediate impact, this piece of data will be required for any future collaborative portal opportunity.

PP 5 Data: PP 5 doesn't have a specific field carved out for capturing a repository name. In the DCHS data the Collection field is sometimes used to indicate the "Deschutes County Historical Society"; however, it's also used for capturing more granular collection names for collections at DCHS. The repository name can also be captured as part of the Home Location field in PP 5, but it will share space with more specific location information—information that may not be desirable for public consumption.

Recommendation: If neither Collection nor Home Location feel like a good fit for the repository name, consider using a customizable field in PP to capture the data. If and when DCHS intends to participate in a collaborative portal this will become a higher priority. Until then, this is a low priority.

Title

Title: The title of the item being cataloged. This field is particularly helpful (and a required field) in Archives and Photograph collections because it can convey key information on the item.

It can be intimidating for some to create a title for an item. To help steer past any roadblocks consider the following prompts:

- What's the most important thing about this photograph?
- Who or what is the focus of the photograph?
- What event is being captured in the photograph?

Examples:

- The Exterior of Pilot Butte Inn Featuring the Patio
- Mrs. Jane Smith with Juniper Elementary School 1st Graders
- A Harvesting Scene with Farm Equipment, Grain, and Horses

Potential Impact: Titles serve to increase discoverability of materials and their use. Titles are one of the first (and sometimes only) fields a user reads before they make an adjudication on the item's usefulness to their search and move on.

PP 5 Data: In reviewing the PP 5 records there appears to be Titles for approximately 40% of the sample size. A comparison of the PP 5 records with the 559 PP Online records may help with completing any data gaps or offer examples of past titles used.

Recommendation: Given the immediate attention a title can command when reviewing items in search results, this is recommended as a high priority area to focus on.

Creator

Creator: The creator is most often a controlled vocabulary- can be local, an authority, or both.[11] In the case where a creator is not known, DCHS has elected to confirm this with the use of "Unknown" in the Creator field.[12]

Examples:

- Unknown
- Smith, Jane
- The Bend Bulletin

Remember: Use the "People" classification in PP to establish the correct names of Creators so that they can be searched for and applied consistently.

Potential Impact: Capturing creator names help provide contextual information to both collection workers and collection users. Using the Creator field provides a tie across collection items that share a creator and can offer insight into the inter-relationship of the collection.

PP 5 Data: There is Creator field data populated in the PP 5 database for the Photograph collection. A comparison of the PP 5 records with the 559 PP Online records may help with completing any data gaps.

Recommendation: This is a helpful field but not quite as critical as Date or Subject. It's a medium priority field.

---

[11] In PartPerfect 5 this is the "Artist/Author/Creator/Photographer" Authority File though when viewing the data via spreadsheet, the column (field) name is "Creator".

[12] The use of Unknown in the creator field of any database can depend on the descriptive standards followed and individual heritage organization policy. Some prefer to leave the field blank because they consider the use of "Unknown" as junk data. Those who elect to use Unknown typically choose to do so in order to definitively capture that the creator is truly unknown versus an omission of detail as it can lead to future questions and (ultimately) dead end research.

Date

Date: The creation date of the object. Many of the descriptions capture a date or era and can be used to populate the Date field.

In PP there are two date[13] capture fields:

- Date: free text can capture circa, month names, etc.
- Year Range: YYYY-YYYY
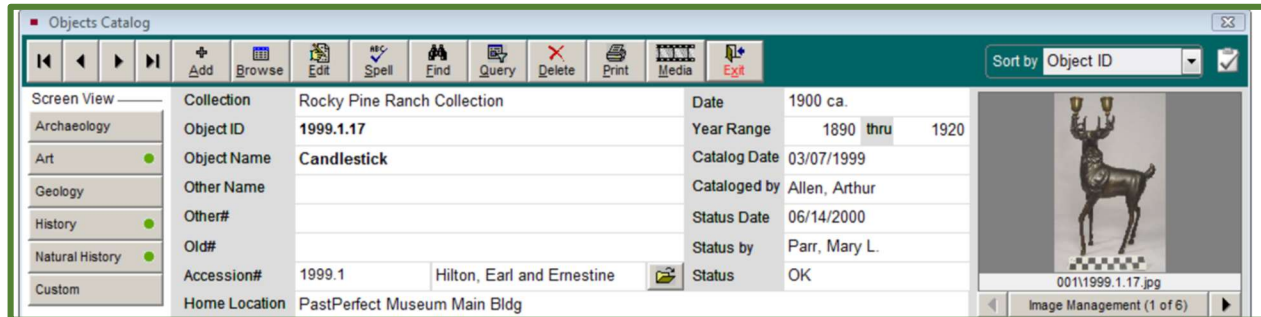
Examples to Use in Date:

- August 15, 2010
- August 2010
- 2010
- Summer 1964
- circa 1990s

Examples to Use in Year Range:

- 1890-1910 (if spanning multiple items; e.g. an archival collection)

Examples Not to Use in Date:

- 1919?
  - Correct: circa 1919
- c. 1941
  - Correct: circa 1941
- ca. 1941
  - Correct: circa 1941

---

[13] PastPerfect Date Style: PastPerfect supports a variety of date formats, including ANSI (2009.04.28), American (04/28/2009), British (28/04/2009), German (28.04.2009), Italian (28-04-2009), Japan (2009/04/28), or USA (04-28-2009). Since dates are stored as Julian dates, changing the date style will reformat the display and printing of all new and previously entered dates. Read more at https://museumsoftware.com/WebHelp/Chapters/PP5-3f.htm, accessed May 26, 2023.

26

- 1947-48
  - Correct: 1947-1948
- 09/00/1915
  - Correct: September 1915
- 1940s
  - Correct: 1940-1949 (if multiple items, then put in Year Range); or circa 1940s
- Circa 1925
  - Correct: circa 1925
- Aug-42
  - Correct: August 1942
- possibly 1937
  - Correct: circa 1937

Potential Impact: Date is one of the top fields we (and our audiences) use to help limit materials we're searching for and intend to work with. With that in mind, any date provided, even an estimation based on contextual information, can be incredibly helpful.

PP 5 Data: There is Date field data populated in the PP 5 database for the Photograph collection. A comparison of the PP 5 records with the 559 PP Online records may help with completing any data gaps.

Recommendation: It's recommended that when using an approximation, pair it with a year range. Not all systems are able to recognize "circa 1910" as an acceptable record for search results when someone wants an item created between 1905 and 1920. Given the potential helpfulness this data lends to item discoverability, this is a high priority area to focus on.

Format

Format (use of Nomenclature for a controlled vocabulary): For Archives and Photograph collections, this is where we would typically expect to see what is currently in Object Name; such as "Print, Photographic". DCHS may also elect to use the Materials vocabulary. Dimensions can also be included as part of the Format for data entry. However, in PP there is a separate Dimensions field to use.

Potential Impact: While format is an important field to understanding the item being viewed, it's doesn't immediately impact management or use of the item.

Recommendation: This is a lower priority item and one to reconsider if and when DCHS wishes to participate in a collaborative portal—as facet searching based on format is a commonly used tool by our audiences.


Language

Language (when applicable): This field is most used in the Archives collection module; however, it can increase accessibility when used to indicate if there is written language present within the image.

Potential Impact: For heavily visual collections this is a "would be nice" requirement. It doesn't immediately impact discovery or use in a significant way and can wait until other high priority data areas have been addressed.

Recommendation: This is a low priority area.

Subject

Subject (use of Nomenclature for a controlled vocabulary): Subjects are used as search tools to indicate thematic relationships and meaning among items. There doesn't appear to be any Subject field use in the 559 PP Online records. In this case, it's suggested that DCHS refer to Nomenclature and utilize the People, Subjects, and Classification search terms to improve findability. Below is the corresponding PP subscreen:
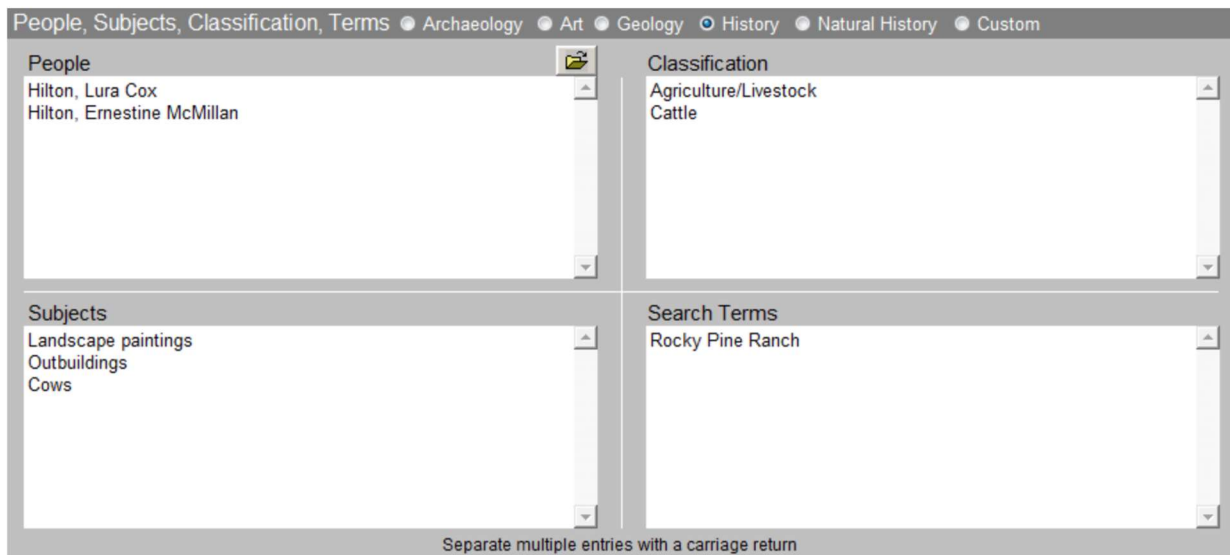
Potential Impact: Subject and Search Terms are incredibly powerful tools to aid in item discovery and use. If there are no subjects or inconsistent subjects in a CMS then it makes the items harder to be found—relying heavily on Title (if it exists) and Description fields. No or inconsistent Subject or Search Term use also hinder the ability to browse the collection as there's no thru-line in the material tied together by subject matter.

PP 5 Data: When viewing a sampling of PP 5 material we can see the Subject and Search Term fields were used at one point, but not with controlled vocabulary.

Recommendation: This is a high priority field that can yield immediate benefits to the discovery and use of the collection.

Rights

Rights (any copyright or other retained rights): This field indicates to the external audience what rights are tied to this item and whether or not the item is eligible for use and what type of use. It also indicates to the user that there should be an appropriate citation if the item is used. This field can be found in the Notes, Legal, and Provenance sub-screen. The Web Rights field will most likely be the most common field to use; however, Legal can also be used if there are other legal or rights considerations beyond a DCHS copyright.

Potential Impact: This statement is a proactive risk mitigation tool to gently educate audiences that there are different rights to using collection content. The inclusion of this statement is of particular importance if the collections are to be published, made discoverable, and shared online. It makes it clear who the photograph belongs, if it's covered under copyright or CreativeCommons[14] license, or can be used with the stated credit line.[15] While the impact of this statement isn't immediate nor does it impact discoverability, it is an important piece of data to include for all collection items online.

Recommendation: Select a copyright, Creative Commons, or other statement approved by DCHS and apply it to all applicable items in the collection. This is a medium priority for collection items that are available online.

Relation

Relation (when applicable, indicate a relation to other known items in the DCHS or peer museum collection). Given that several of these records refer to other related items, it makes sense to use this field to indicate relationships. The Relation field can also be used to capture a relationship to items at a different heritage organization. *Note: There's not necessarily going to be a relationship to record for every item, so this field is indicated as required when applicable.*

---

[14] About Creative Commons Licenses, Creative Commons, accessed March 18, 2023, via https://creativecommons.org/about/cclicenses/.

[15] For example, using a credit line can attribute creatorship to a photographer or *The Bend Bulletin* while also including the repository (DCHS) name.

Relations subscreen

Potential Impact: Taking the time to indicate relationships among items is an investment that can have tremendous benefit. While Creator or Subjects can assist by inferring a relationship, the use of the Relation field can help indicate relationships among items that aren't as obvious as Creator or Subject. Using this field can save future time searching for future exhibits and research.

Recommendation: Brainstorm how relationships may span the collection and which items should indicate the other. This field is a medium priority.

Data Descriptive Standards

The remaining field not already included in the Dublin Core™ minimum data element set and is included in the CCO minimum fields to use set is: Current Location.

Current Location: Unless the photographs move frequently, this field is not helpful to add to the record. This field is most appropriate for the object collection.

Value-Added Fields to Consider

The following field isn't required for minimum best practice, but it does offer improvement in collection management and user engagement: File Name.

File Name (to reference for uploaded digital file): The image file name can serve as a helpful reference to quickly and easily access a digital file for exhibit or researcher use. It can also be immensely helpful in any future database migration when both data and digital files are migrated from one system to another and need to be matched back up.

Potential Impact: This field may not have immediate impact, but it will if the data and related digital files ever need to move to a different CMS platform, or be contributed to a collaborative portal. Knowing which record each digital file belongs to—and the actual file name as it exists outside of PP—can be critical to maintaining the pair.

PP 5 Data: In reviewing the data from PP 5 there does appear to be use of the IMAGEFILE field. A comparison of the PP 5 records with the 559 PP Online records may help with completing any data gaps for this field.

Recommendation: It's not an immediate priority, but it is incredibly important as cataloging continues. Start capturing the file name for attached digital files moving forward. As time allows, retroactively populate the file names in a custom filed that you can call upon later when you need it. This is a low priority for cleanup.

Potential Use of Existing Data

In the DCHS PP 5 data for the Photograph collection, it looks like many of the required fields are already present in the data. What's unclear is if the PP Online records have a carbon copy with a fuller dataset in PP 5. If that's the case it's always easier to use existing data, even if it needs some enhancement or editing.

If there is a fuller set of data in PP 5, then the following is recommended:

1. Run a query in PP5 from the Photograph collection for all photographs that transitioned to PP Online. (The PP Online data is a selection of photographs from 1979-2014).
2. Once the query is executed generate a spreadsheet.
3. Review the data in the spreadsheet to see if each required field is filled in. Highlight any required field that's empty.
4. Review data for quality issues such as incomplete sentences, inconsistencies in details, or factual errors.
5. Select which approach (A or B) you intend to take for data cleanup. Then follow the suggested cleanup prioritization for that approach.
6. When data augmenting is complete: run spellcheck.
7. Using the Object ID, work through each record entry and copy the refined data from your spreadsheet (informed in part from the PP 5) into the PP Online record.

A Guide to Creating Data Content

Please see the "An overview of Data Standards," section in *A Guide to Collection Data Cleanup.*