

Connecting people with data

Gracen Brilmyer

Ashwini Sriram

Advisor: Robert Glushko

Acknowledgement:

A special thanks to Daniel Chen and Hadrien Renold for their integral involvement in the development of Bridge.



Abstract

People want to share data. Bridge aims to bridge the gap between two user groups: collection managers (CMs), those who produce the datasets and want to make them public, and end users, those who want to search for data in a familiar or new domain. Each user group has particular needs which we are attempting to address through user interface design and research. By providing a web application to assist collection managers build a custom search interface for their dataset, Bridge integrates broad domains of user research to create an optimal user experience for both CM and end user.



Table of contents

1. Collection Managers
 - 1.1. Why Share?
 - 1.2. Sharing Data: Existing Search Interfaces of small and large collections
 - 1.3. Collection Manager Use cases
 - 1.4. Pain points - What trouble do Collection Managers have with sharing data?
 - 1.5. Existing Tools
2. End Users
 - 2.1. Finding Data
 - 2.2. How do end users look for something specific
 - 2.3. How do end users explore
 - 2.4. Existing Tools
3. Solution: Bridge
 - 3.1. Collection Manager User Interface Design
 - 3.2. End User Interface Design
 - 3.3. Product Architecture
 - 3.4. Design Process
 - 3.5. Lessons from Design Process
4. Future Design Questions



1. Collection Managers

1.1 Why share?

Institutions and individuals want to share their data for many reasons -- not only is it a great way to show off the hard work of organizing a collection but also it can connect researchers across the world and increase visibility of such research or institutions to the general public ("The Case for Digitization"). Digitization is the digital recording of an object that can span from databasing the metadata about a collection specimen to digitally imaging to 3D scanning.

“These efforts create lasting records of the natural world that might otherwise be inaccessible, unless you know a friendly curator willing to take you behind the scenes (Roston).

Digitization is a valuable way not only to preserve the content of a collection but also to increase the shareability of a collection.

“For centuries, scientists who wanted to study a particular type specimen had to visit the museum where it is kept or have the specimen sent to them. Either way, the potential for damage was high: fragile body parts would sometimes fall off during inspection or transport, causing irreparable damage (Olsen).”

Museums, in addition to the many individuals who own personal collections of scientific material, music collections, and other “inaccessible” datasets, benefit through digitization because of increased visibility which can:

Increase awareness around preservation of a collection¹, increase foot traffic and act as advertisement for a museum, increase collaboration between people working on similar data. Science stands to benefit not just from the advances in knowledge gained from worldwide access to these specimens, but also by encouraging students to pursue careers in biology or environmental work (Olsen).

Large museums often have the resources to not only digitize their collection but also purchase complex data storage methods and hire web developers to share data. However, small institutions and individuals lack such resources, although they equally have interest in and can benefit from sharing the contents of their collections. Through initial user research we discovered that collection managers (CMs), who either own their own data collection or manage a collection for an organization, often don't have the budget to pay for expensive data storage technologies as well as designed searchable interfaces (Mulji, Oboyski). A large barrier to entry for collections managers is cost and technological know-how. Many people who have small collections or who are trying to aggregate data from different resources cannot afford to pay for a sophisticated collection database as well as a web developer and designer. In addition, many databases and data collection/storage methods (such as KeEmu, Access, Google Forms) often output data in spreadsheet or csv formats.

Though a survey completed by 67 individuals, we found that the most common method of data storage was Google Drive (Docs and Sheets)(56%) and Microsoft Excel sheets (33%) since organizing, categorizing, and storing individual collection records is relative simple with low cost.

¹ “This is a chance to make everyone aware of how valuable the data are in natural history collections, how much they tell us about the natural world,” Dr. Larry Page qtd Olsen

There was also popularity in also storing data in other delimited formats such as csv, tsv, and json. Through this survey, we also learned that people utilize many cloud-based storage methods, such as Google Drive (72%), Dropbox, and Airdrop to not only store but also share data. Using Excel sheets (33%) is also a popular way to share data with collaborators. We also saw an overwhelming majority of people claim to be very comfortable using Microsoft Excel (58%) and Google Sheets (46%), while not being comfortable using more complicated data management tools such as SQL Databases (22%) and Microsoft Access (6%).

Through our user research, we discovered that CM's want to make their datasets public for many reasons. By making data publicly available, they often want their data to contribute to larger research as well as to show off the work that they have done in collecting it (Oboyski). In the case of the privately own entomological collection, the CM frequently wanted to share a specific specimen record with another researcher or connect the specimen record to a taxon publication. In other cases, organizations want to involve the public in their process and expose their data for increased user engagement with the organization through exposure of their data (Mulji). We also discovered that CM's, although they want public exposure and use of their data, they generally don't want to publish and entire dataset for download -- one CM cited that they felt uncomfortable allowing users to "copy and paste" their whole dataset, however they were comfortable with end users downloading a subset that was pertinent to their research (Oboyski). The GLBT museum, in addition, is interested in sharing its archives publicly, to spread awareness about GLBT people and communities, along with their history. The museum has a big community presence and the archivists are passionate about making their collections available to everyone who has even the slightest interest in them although they have concerns about sharing all of the images of their collection (Black).

Because of the financial and technological limitations small institutions and individuals who own collections need a tool to help them share data catering to the existing digital storage methods. Although some CMs choose to learn how to build their own public facing search interface (or find a volunteer to build one (Black)) they experience high learning curves to technology and result in a less-than-satisfactory interface that relies on a single individual to update (Oboyski, Black).

1.2 Sharing Data: Existing Search Interfaces of small and large collections

This research explored the entry point into different types of collections, developed by collection managers (CM) for either novice or expert end users. Collection managers, as established earlier, are intimately familiar with their data -- ways it can be searched, what fields one might search, and domain specific standards. In biological collections, most scientists are familiar with the DarwinCore metadata model used by most CMs to to store, share, and search their data. In other, more obscure domains, formatting norms are often upheld that pertain for the format of the data: most visual collections have a visual entry point and text additions (such as tags or keywords) and audio collections often have strong visual elements (such as album covers), although this research shows that CMs vary in the domain-specific language they use.

Many user interfaces imply that the end user has some familiarity with the collection's domain. Text heavy collections, such as books or scientific collections, offer a search interface with domain specific categories for the end use to search through (Figs 1 & 2). Smaller collections, text-, audio-, and image-based, often present an alphabetic list for their specific domain and

collection, such as the Oscar Wilde and the Fin de siècle Archive (Fig 3) and Numero Group Online Catalog (Fig 4 & 5). Having such features imply that CMs assume their users are familiar with the domain and will be looking for something in particular. In contrast, many smaller collections take advantage of visual elements as points of entry to their material, such as images of album covers (Fig 6), audio recordings (Fig 7), or visual collections (Fig 8 & 9). Although they limit the amount of data that can be shown or browsed at once, image-centered entry points hold the potential for users to browse without knowing a specific domain. Creating, locating, or formatting visual elements is a costly decision for CMs to invest in, although they provide for a much more engaging interface.

Hawaiian Insect - Plant Host Database

	matching	Insect	matching	Plant
Order	equals <input type="button" value="v"/>	<input type="text"/>	equals <input type="button" value="v"/>	<input type="text"/>
Family	equals <input type="button" value="v"/>	<input type="text"/>	equals <input type="button" value="v"/>	<input type="text"/>
Genus	equals <input type="button" value="v"/>	<input type="text"/>	equals <input type="button" value="v"/>	<input type="text"/>
Species	equals <input type="button" value="v"/>	<input type="text"/>	equals <input type="button" value="v"/>	<input type="text"/>



Sort results by Insect
 Sort results by Plant

[About this database](#)

About this database

This searchable database is derived primarily from "Forest Entomology in Hawaii" by O.H. Swezey, 1954. Species names have been updated where information was available. The database will ultimately contain insect and plant names (searchable at the Order, Family, Genus, and Species level) as well as notes on habits (leaf-feeder, seed borer, sap-sucker, etc.) and distribution (from literature sources). The database will continue to grow and develop focusing primarily on literature searches and my personal collections. The "Insects of Hawaii" series by E.C. Zimmerman is being (very) slowly databased beginning with the Lepidoptera.

For a full checklist of the arthropods of Hawaii, including literature citations for select groups (eg. Megalagrion), please refer to the [Bishop Museum On-line Arthropod Checklist](#). If you have questions, comments, corrections, or would like to contribute data, please contact [Pete Oboyski](#).

Fig 1 Hawaiian Insect-Plant Host Database, P. Oboyski

Enter search term(s)...

Scientific Name [e.g., *Chyromya flava*]
Tribe
Subfamily
Family [e.g., Chyromyidae]
Order
Region
Country [specimen data only]
Province/State/Territory [specimen data only]
Types Present
Date Collected [e.g., 1978 or > 2000]
Records per page

List items only with images

Fig 2 Field Museum of Natural History Collections Search Interface

> Home > Contributing Institution > UCLA Share / Save

Clark (William Andrews) Memorial Library, UCLA

2520 Cimarron Street
 Los Angeles, California 90018 Physical collections: 140
 Phone: (323) 731-8529
 Email: rffenning@humnet.ucla.edu
<http://www.humnet.ucla.edu/humnet/clarklib/>



Browse the Collections

Find a collection at this institution

Hit escape key to turn off autocomplete feature.

From A to Z

0-9 **A B C D E F G H I J K L M**
N O P Q R S T U V W X Y Z

: | |

A

- Amalgamated Printers' Association Ephemera Collection
- American Society for Eighteenth-Century Studies Records
- Archer (H. Richard) and the Hippogryph Press Collection
- Armitage (Merle) Collection
- [back to top](#)

B

- Battledore Press Collection
- Baxter tracts.
- Beerbohm (Max), Majl Ewing Collection on
- Bentley (Wilder) Papers
- Book Fair Poster Collection
- Browne (Clyde) and the Abbey Press Collection
- Browning, Oscar Letters to Joseph Leftwich, 1919-1921.
- Burne-Jones (Edward) Letters to Violet Maxse
- Burns (Phil) Papers
- [back to top](#)

C

Cambridge collections

Fig 3 Oscar Wilde and the Fin de siècle Archive (Through the Online Archive of California)

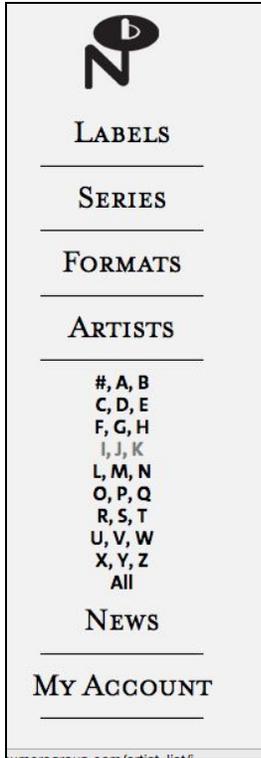


Fig 4 Numero Group Artists Alphabetized Menu

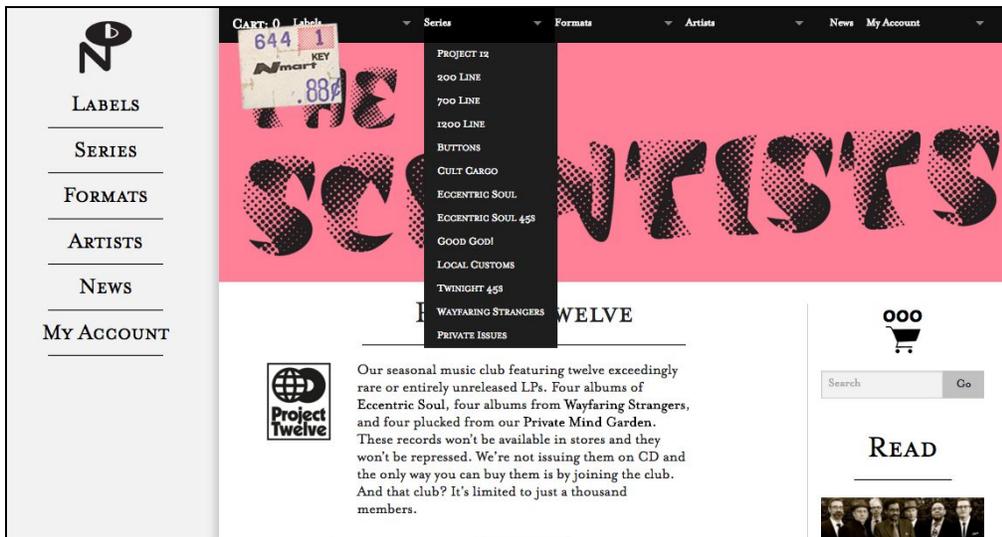


Fig 5 Numero Group Series Menu

CART: 0 Labels Series Formats Artists News My Account



Our archival main line discography features the most lavish packaging allowed by history, taking in the flagship Eccentric Soul series, the Good God!, Cult Cargo, Wayfaring Strangers, Local Customs, and Buttons series records, plus the 200 and 700 series. We've also given our coveted three digits to host of albums reimagined, imagined here for the first time, and currently still beyond imagining.



ECCENTRIC SOUL
THE CAPSOUL LABEL
2xLP, CD, MP3



ANTENA
CAMINO DEL SOL
2xLP, CD, MP3, Cassette



ECCENTRIC SOUL
THE BANDIT LABEL
CD, 3xLP, MP3



BUTTONS
STARTER KIT
2xLP, MP3






Fig 6 Numero Group Featured Albums

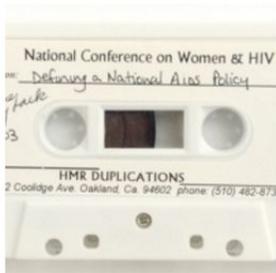
Browse All Browse by Tag Search Items



Linda Meredith's Talk at final CDC hearing

Presentation at the final CDC hearing by Linda Meredith, followed by physicians, about the definition of AIDS activists, especially the inclusion of lesbians, difference from advocates, and inaccurate characterizations.

Tags: [AIDS](#), [Center for Disease Control \(CDC\)](#), [Education](#), [Health Care](#), [Women's](#)



Defining a National Aids Policy

NIH-moderated panel about defining a national AIDS policy, research health services, and prevention in four talks from Maxine Wolfe (ACT UP), Dr. Wanda Jones (CDC), Corola Marta (Yale physician), and community activist Michelle Lopez.

Tags: [AIDS](#), [Center for Disease Control](#)



Meeting at FDA Re accelerated approval

Meeting between ACT UP and other HIV/AIDS activists and employees of the Federal Drug Administration regarding recent FDA activities in accelerated approval and expanded access to drugs.

Tags: [AIDS](#), [Clinical trials](#), [Communities](#), [Consciousness Raising](#), [drugs](#)

Fig 7 Lesbian Herstory Archive Audio Recording Collection

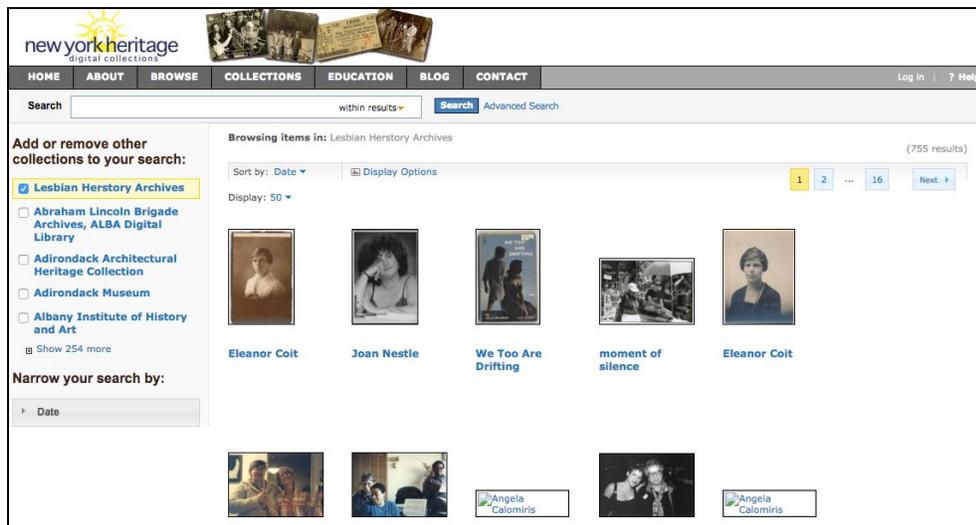


Fig 8 Lesbian Herstory Archive Photograph collection (hosted on New York Heritage Digital Collection)

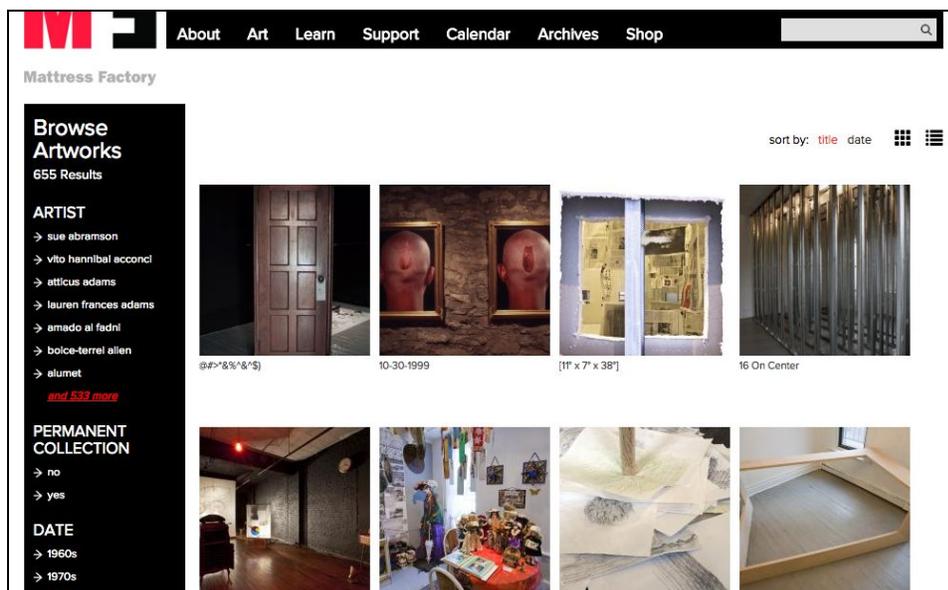


Fig 8 Mattress Factory Collection Page

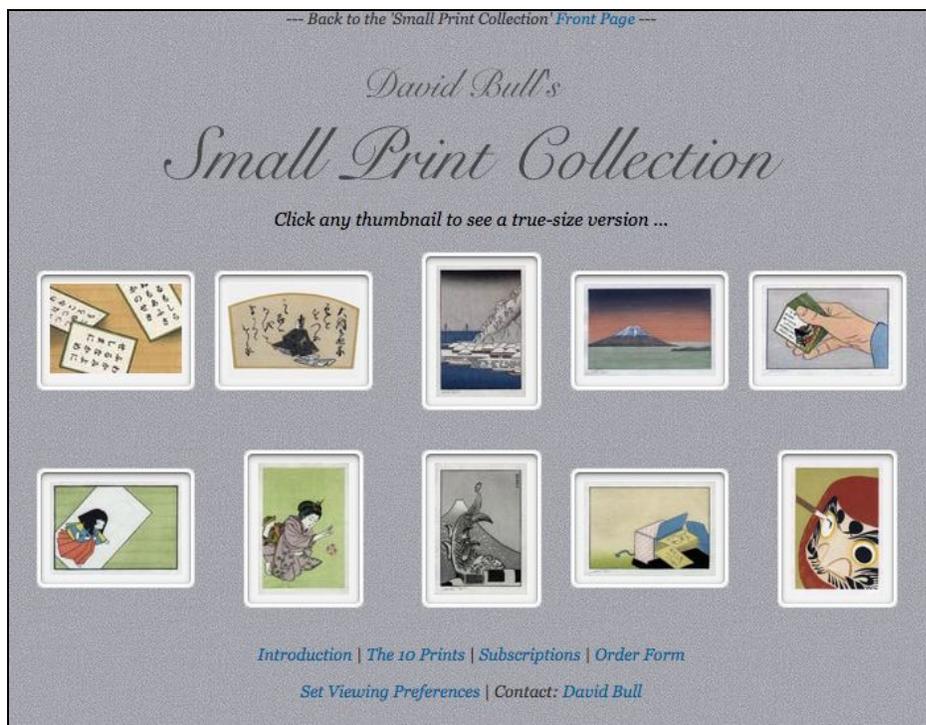


Fig 9 David Bull's Small Print Collection

1.3 Collection Manager Needs and Roles:

a . Collection Manager of Museum Collection

Duties include:

Supervising staff, overseeing the collection, making sure things are protected, properly catalogued, making sure the public has access to literature. There are reference rooms at the office and they are open two days a week. The CM helps people by matching the collections to their research. Allowing them to use the collections that are not their own. She has tools and

mechanisms to understand material but she does not know every bit of information that is out there. She also works with donors.

Main needs:

A volunteer built the current search interface, which isn't exactly what the archive needs: An end user can't use standard search syntax (for example, last name, first name or exempt search) in the search fields and there is no advanced search. No other staff members have the technical skills to update the current website. Right now, people interested in the archive have to come to the physical museum to see both the collection and the digitized records.

Technologies currently used:

- Wordpress for interface and museum website
- Museum database

B. Non-Profit Organizer

Duties include:

Working to improve the democratic process to decide how to spend public funds. Sharing data about public funds with different groups in the public and city staff to get ~25 project ideas of how to spend public money. The projects then get chosen by the public and are implemented by the public institution. They provide technical support to institutions that are doing participatory budgeting and provide materials, and training for implementation.

Main needs:

Make projects and associated data available to the public. Different interest groups have different areas that they want to learn about and they should be able to find that data easily. It would be nice to share different data with different groups as well as allow a novice user to easily browse all of the project proposals. They want to associate people with ideas in a meaningful way.

Technologies currently used:

- Google Sheets, Forms, Docs
- Excel
- Shareabouts
- Formdesk
- Citizinvestor
- Ushahidi

C. Naturalist with personal collection

Duties include:

Working in a natural history museum as a Collection Manager, where he works managing the digitization of a large collection, organizing incoming and outgoing specimen loans, and maintaining collection updates. He also has his own research practice on insects, which he collects for his personal collection. He learned Perl and web programming to build his current interface, which took a lot of effort.

Main needs:

Sharing data with collaborators and fellow researchers, he wants both his interface and data to be easily updatable and accessible from anywhere. He needs the ability to easily create custom search fields since he doesn't want to learn new programming. He also keeps a separate species database with publications, which he wants to be able to associate with his specimen records.

Technologies currently used:

- Microsoft Access
- Microsoft Excel
- DropBox
- Perl, html, css

1.4 Pain points - What trouble do Collection Managers have with sharing data?

There are several tools that partially solve the needs of collection managers, however, there is no complete end-to-end solution for creating search interfaces for collections, as is clear from our user interviews. Collection managers already use tools like Shareabouts, Google Spreadsheets and other Google docs, however, there are several pain points they face while using these tools. We are restricting our analysis to the tools that were mentioned by our interviewees. Literature review and Google searches of our own did not bring up any relevant tools that can be used specifically for sharing collections.

1.5 Existing Tools: Collection managers

All of the collection managers that we interviewed cited similar needs for data sharing. They illustrated that an ideal tool would support easy and efficient data entry, should support popular formats like Excel, csv and Google Spreadsheets (according to the survey, these are the most popular data formats), is intuitive to create custom search fields, and has the ability to be published and integrated into their existing sites.

CM Existing Tool Feature Matrix:

	Shareabouts	Google Spreadsheets	Excel	Wordpress
Data entry is easy and efficient	NO	YES	YES	MAYBE
CM is able to easily create a usable public search interface that is intuitive to use	NO	NO	NO	YES
Collections are easily shareable	YES	YES	NO	YES
Requires very little technical know-how	NO	YES	YES	NO
Affordable	YES	YES	YES	YES
Sophisticated search functionality	YES	NO	NO	MAYBE
Supports images	YES	NO	NO	YES

CM Existing Tools Description:

“**Shareabouts** is an online mapping tool to gather crowdsourced public input in a social and engaging process.” Through this tool, collection managers can drop pins on digital maps and provide ideas and comments for planning and design issues. A quick glance at the documentation tells us that we have to install git, pip and virtualenv, to get started, in addition to requesting access from api.shareabouts.org. Users have to have an understanding of APIs and git to get started. Terms like ‘virtual machine’ and ‘Linux distribution’ are casually thrown around the documentation.

“[The] Hard part is getting all the inputs from in-person events uploaded. Time for data entry is a huge sink. [In] Shareabouts, you have to manually enter each of the ideas into a form. It is buggy. Map based tool should allow you to upload text easily but edit the location later.” (Mulji)

Microsoft Excel is one of the most common ways to organize data. Records can be entered, organized, and stored in this flat database system. Although spreadsheets are a commonly used and understood format, they can be overwhelming for an end user to understand, sort, and make meaning from. In addition, there is a subscription fee for using Microsoft Excel beyond an initial trial period, which varies depending on who uses it (enterprise, small business, non-profit etc.).

Google Sheets are an easy way to organize and share data with different users, and by far one of the most popular tools. Data can also be securely generated using Google Forms, where users can customize fields for end users to populate. Although sharing through Google Drive is

extremely user friendly, the spreadsheet format, as stated earlier, isn't always easy to use for understanding large datasets.

“We can't sort and filter and break off the results. It is always just a massive database.”

(Mulji)

On further research, we found that filters can be created on Google Spreadsheets and on Excel spreadsheets, but it is not obvious how to use them. But Google Spreadsheets are widely used among collection managers to store and share data, and Bridge aims to build on top of the existing affordances of spreadsheets.

Although **Wordpress** is not a software that can be directly used as a database, it came up several times in our interviews as a platform that museums use to build interfaces for their collections.

“We just switched our website. It was little more robust earlier, then **we moved to wordpress platform. It is too limiting.** If you search for “last name, first name”, it doesn't bring up search results. It is unfortunate. There is no exempt search, it is limiting. If we had advanced search function, if I am looking for something, I can't do a lot on my own for in depth research. If people are just looking for a flavor for what we have, then it's great.” (Black).

The value derived from using a platform like Wordpress, while substantial, could be limited by technical know-how of the developer. In this case, a volunteer at the GLBT museum

implemented the search functionality, which turned out to be insufficient and failed to satisfy basic expectations of a search interface. The application needs to be hosted publicly, which requires knowledge of web hosting, which is not always straightforward.

Ushahidi is a data crowdsourcing tool, which allows users to gather posts about a particular topic, from across the world and across devices. Its website says, *“Ushahidi, which translates to “testimony” in Swahili, was developed to map reports of violence in Kenya after the post-election violence in 2008. Since then, thousands have used our crowdsourcing tools to raise their voice. We’re a technology leader in Africa, headquartered in Nairobi, with a global team.”* Their solution seems more tailored towards crowdsourcing events and happenings across the world and technology platforms, rather than on facilitating collection sharing. However, they customize solutions to suit the needs of their customers, and charge a fee for services which include data management, crowdsourcing data, visualization, filtering, sorting and searching of data. This seems to be a promising solution for organizations like Participatory Budgeting, that want to gather data from multiple sources.



2. End Users

2.1 Finding Data

End users, we discovered, may want to access such datasets for research purposes in a domain they are familiar with, or out of curiosity for a new domain. The people we interviewed mentioned several different ways in which they search for things online. On the one hand there were situations when people knew exactly what wanted and searched specifically, using keywords, and on the other, there were situations that called for more exploratory approaches to searching. People interested in specific areas, knew the organizations that were authorities on those topics, and specifically looked those organizations up. Generally, people used the Google search engine to look for anything. Most of our interviewees were unaware of curated collections platforms. One person mentioned data.gov as a source of interesting datasets.

2.2 How end users look for something specific

We found that one of the biggest barriers to successfully finding specific data in a collection is the time taken to negotiate search commands. Ideally, the application has to maximize the time the user searches the collection to get useful information.

"Typically, users can adapt to different database command structures. However, the question is not whether they can adapt, but rather whether they want to spend the time doing so. The less time a user needs to spend negotiating search commands, the more time that user will have to spend searching your collections." (Blackaby and Sandore)

Our research shows that a key problem is the semantic vocabulary gap between experts created descriptions and the implicit and often not domain related preferences of end users. (Wang et al). Using controlled vocabulary and simple, intuitive terms and descriptions could help balance this gap.

2.3 How end users explore

Our research has shown that adding visual elements, such as images and data visualizations, can help end users explore a collection. By adding interactive visual elements, an interface can support user exploration. If the data set is large and/or complex, aside from concerns around computing power, designers should split the data into palpable chunks for the end user (Tominski 10). Those chunks then, “become more effective because they are tailored to emphasize a particular aspect of the data, allowing users to concentrate on task-relevant questions (Tominski 10). Jacques Bertin famously said,

“a graphic is not ‘drawn’ once and for all; it is ‘constructed’ and reconstructed until it reveals all the relationships constituted by the interplay of the data. The best graphic operations are those carried out by the decision-maker himself.” (16)

Through abstraction and aesthetic arrangement of data, an end user can engage with a dataset. The figure below is a widely accepted overview of data visualization, showing, “how data is to be transformed through several stages from a data source into data tables, into visual abstractions, and finally into visualization views. Data transformations, including filtering, clustering, error correction, can be found in the early stage of the pipeline.”(Tominski)

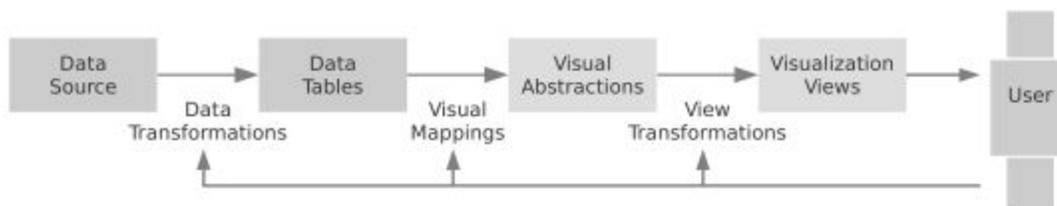


Figure 2.1: The visualization pipeline (adapted from Card et al., 1999).

2.4 Existing Tools: End Users

Our many interviews with end users highlighted key points for searching and exploring data.

Users wanted to easily be able to locate datasets online, intuitively search using keywords and categories, have the ability to filter and sort the collection, and be presented with some interactive element of the data if they were unfamiliar with the domain. Through our survey, we found that Excel is a popular way (67%) along with Google Sheets (45%) to understand a new dataset. Some people also utilize visualization tools such as Tableau to explore new data, while others took a more technical route citing Python and r as primary tools for data exploration.

When exploring a new domain, end users cited feeling overwhelmed if they encountered a large spreadsheet of data without understanding the context or knowing the most important aspects of the data. They also overwhelmingly cited preferring using Google keyword searches to help locate data and to make sense of it through Google's search result summary.



3. Bridge as a Solution

Bridge is a tool that allows collection managers, without being tech- or design-savvy to easily share their data, create visualizations, and a well-designed custom search interface for their collection. Our research identified and illuminated the key pain points for collection managers as well as for expert and novice users, resulting in a tool that bridges the gaps between collection managers, their data, and end users. We based our design off of three use cases inspired by our CM user research:

3.1 Collection Manager User Interface Design

List of Section Screenshots:

- Landing Page

- Log In Page

- Sign Up Page

- Data Upload Page

- Create Search Interface: Page Details

- Create Search Interface: Create full-text search, custom search fields & visualizations

- Create Search Interface: Select custom colors, cover image, and publish

- Profile Page

Bridge's CM-facing interface is clean and easy to understand and works with simple data storage methods, such as Google Drive and Microsoft Excel. A new user can create a new log-in or log in to an existing one (See screenshot: *Log In Page*), which is encrypted in our database to provide a secure environment. Using FileStack, an API for file web integration, a CM can upload their preferred file format, as Bridge accepts .csv, .xls, xlsx, and connects to a

CM's Google Drive to accept Google Sheets (See screenshot: Data Upload Page). The CM can also add media files to individual records by labeling the data column "media" and including the URL to an image file. This column will automatically displayed in the grid-view (discussed below). Once uploaded, the CM can see a preview of their data.

After uploading their chosen dataset, the CM can then populate information about their institution and/or collection. Our CM user research revealed that both CMs and end users prefer and regularly use keyword search (the most common example being Google). Because of the common use and understanding of cross-category search, Bridge provides a full text search option; a CM can select to include a full text search in their interface and it will automatically provide a text entry box on their interface that will search every cell in their collection.

Our user research also illuminated a need for a CM to create custom search fields, where one search field searches multiple database fields. Often collection objects will have unique database fields for storing different levels of information, however, an end user may not know in which database field a piece of information is categorized. With Bridge, the CM can create one search field (e.g "Location") that looks in multiple data fields (e.g. State, Province, City, Township, Town, County) (See screenshots: Create Search Interface).

In addition to creating custom search fields that run SQL queries on their collection, CMs can further customize their interface by selecting primary and secondary accent colors (that style the navigation bar and buttons) along with a cover image(See screenshots: Create Search Interface). Based on our Existing Search Interfaces research (section 1.2), this seemed to be the most important styling features that we could implement in our minimum viable product.

Much of our end user research focused on two user groups: expert and novice users. As expert users often seek out the search fields to be able to retrieve a certain subset of collection records, novice users do not always know what to look for in an unfamiliar collection or domain. Through this research we discovered that data visualizations are a useful tool that both assist CMs and expert users get a better understanding of what's in a collection and help novice users what's in a collection and what they might search for. Bridge provides 3 options of visualizations for a CM to select from: a **bar chart** that caters to categorical data, a **line chart** that can display numerical or categorical data, and a **map** that can plot data that has latitude and longitude coordinates (See screenshots: Create Search Interface). Once the CM has selected what type of visualization they want to display for their collection, they select what columns of data they want the visualization to apply to. The selected visualization is displayed on a clean splash page, styled similarly to the search interface, for the collection that provides a high-level summary of the collection.

In addition to creating a search interface, a CM can also view, edit, and delete their existing interfaces in their profile page (See screenshots: Profile Page). Here, they can also add details about themselves and their organizations.

Create a custom interface to share your data in minutes!

1. Upload

2. Create Search

3. Publish

+ CREATE EXPLORE

About Bridge Contact us

Landing Page

bridge Search Collections Signup

LOGIN

Email:

Password:

LOG IN SIGNUP

Log In Page

The screenshot shows the 'Sign Up' page of the Xbridge website. The header is blue with the 'Xbridge' logo on the left and 'Search Collections' and 'Log in' on the right. The main content area is light gray and contains a white sign-up form. The form has the title 'SIGN UP' and four input fields: 'First name:', 'Last name:', 'Email:', and 'Password:'. Below the form is a red button with the text 'SIGN UP'.

Sign Up Page

The screenshot shows the 'Data Upload' page of the Xbridge website. The header is blue with the 'Xbridge' logo on the left and 'Profile' and 'Log out' on the right. The main content area is white and features a large heading '1. Upload Collection'. Below this is a section titled 'How to format your data' with four numbered instructions: 1. The first row of data will be used as your column headers (You will be able to build customized search fields based on column headers); 2. You will have the option to create a visualization based on the columns; 3. If you would like to add images, please add a column named **media** in your spreadsheet with complete url links to images; 4. The first 5 columns will be displayed on your web interface. Below the instructions is a red button with a file icon and the text 'Pick File'. Underneath the button, it says 'Accepted formats' followed by '.csv, .xls, .xlsx, Google Spreadsheet'. At the bottom of the page, there is a section labeled 'File Preview'.

Data Upload Page

2. Create Search Interface
Customize colors, cover image, and search fields for your web interface

1. Page details

Enter title for collection (ex: GLBT Archives)

About your collection

2. Customize Search Interface

Choose how users will search your collection

Full Text Search
Check this if you want a Google-like full text searchbox for your entire document

Customize Search
Check this if you want create search fields that search specific columns in your spreadsheet

Create Search Interface: Page Details

2. Customize Search Interface

Choose how users will search your collection

Full Text Search
Check this if you want a Google-like full text searchbox for your entire document

Customize Search
Check this if you want create search fields that search specific columns in your spreadsheet

Create Search Fields

Display Name	Name	Description	Dropdown
CommonName, Category	Name	Description	Dropdown
SpecimenID	Specimen Number	Description	Textbox

Create Visualizations (optional) No Yes

Line Graph
Recommended for showing comparisons between categories

Bar Chart
Recommended for comparing two or more categories

Map View
Recommended for location data (named place or latitude/longitude)

Create Search Interface: Create full-text search, custom search fields & visualizations

3. Style your interface

a. Choose primary color
The color of top navigation bar on web interface

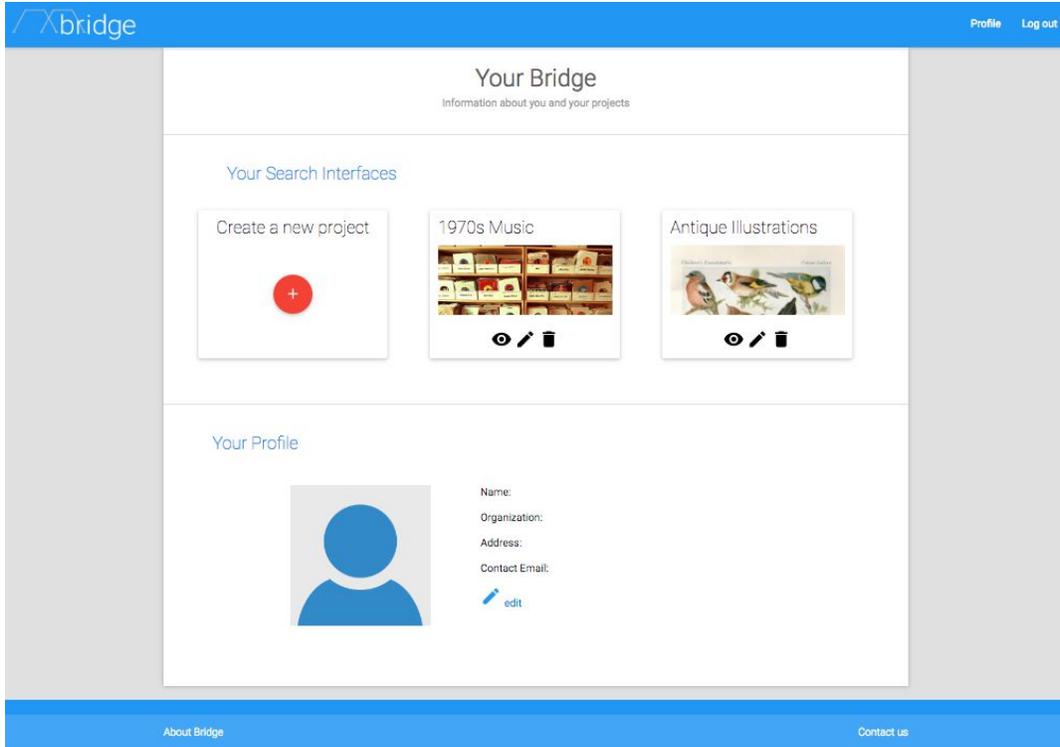
b. Choose accent color
The color of all buttons and highlights on web interface

Upload Cover Photo:
The top banner image on web interface

About Bridge [Contact us](#)

The image shows a configuration screen for styling a web interface. It has a title '3. Style your interface'. There are two color selection options: 'a. Choose primary color' with a color picker set to black and a description 'The color of top navigation bar on web interface'; and 'b. Choose accent color' with a color picker set to light gray and a description 'The color of all buttons and highlights on web interface'. Below these is an 'Upload Cover Photo' section with a description 'The top banner image on web interface' and a red 'FILE' button. At the bottom of the main content area are 'BACK' and 'PUBLISH' buttons. A blue footer bar contains 'About Bridge' and 'Contact us'.

Create Search Interface: Select custom colors, cover image, and publish



Profile Page

3.2 End User Interface Design

List of Section Screenshots:

- Results: List View
- Results: Grid View
- Record Modal View
- Visualization Page

The search page is also designed in such a way where both novice and expert users can see the collection. Both the search fields and entire collection are shown simultaneously, so that the user can see the entire collection in a list view and grid view. Both the list and grid views display the first 5 columns of data, a design choice that both appeases some CMs concerns about exposing their entire collection, and the end users that feel overwhelmed when presented with the entirety of the data fields (See screenshots: Results). An end user can click on any record to see the entirety of the data fields for an individual record. Using Materialize's Modal view, each record, when clicked on, exposes a pop-up window with all record data and displays the record image if there is one (See screenshot: Record Modal View). The grid view and individual record views also show any media links that provided in the CM's uploaded spreadsheet. Through our user research, we found that novice users liked to be presented with some information from the collection since they wouldn't know what terms to search for. The collection visualization is another entry point for users to explore the collection. The visualization is an interactive dashboard so that when a user manipulates a controller, the chart changes to reflect the collection data at that granularity (See screenshot: Visualization Page). The visualization is meant to summarize the collection in novel ways, to engage both novice and expert end users. Not only does it provide novice users with aspects of the collection that they might search for, but also summarizes the collection for expert users who might want a high-level view of the collection contents.

Search ?

SEARCH CLEAR

number

Category

Choose option

SEARCH CLEAR

Search Results

37 Record(s)

LIST VIEW GRID VIEW

SpecimenID	CommonName	Category	CollectorCode	Collector
MUS1078364	Tiger Beetle	Beetle	12345	C. Smith
MUS1078365	Tiger Beetle	Beetle	45829	G. Deleuze
MUS1078366	Tiger Beetle	Beetle	16847	S. Ahmed
MUS1078367	Tiger Beetle	Beetle	98206	S. Ahmed

Results: List View

Search ?

SEARCH CLEAR

number

Category

Choose option

SEARCH CLEAR

Search Results

37 Record(s)

LIST VIEW GRID VIEW

SpecimenID: MUS1078364
CommonName: Tiger Beetle
Category: Beetle
CollectorCode: 12345
Collector: C. Smith



SpecimenID: MUS1078365
CommonName: Tiger Beetle
Category: Beetle
CollectorCode: 45829
Collector: G. Deleuze



SpecimenID: MUS1078366
CommonName: Tiger Beetle
Category: Beetle
CollectorCode: 16847
Collector: S. Ahmed



SpecimenID: MUS1078367
CommonName: Tiger Beetle
Category: Beetle
CollectorCode: 98206
Collector: S. Ahmed



SpecimenID: MUS1078368
CommonName: Scarab Beetle
Category: Beetle
CollectorCode: 27047
Collector: S. Ahmed



SpecimenID: MUS1078369
CommonName: Scarab Beetle
Category: Beetle
CollectorCode: 73057
Collector: S. Ahmed



SpecimenID: MUS1078370
CommonName: Tiger Beetle
Category: Beetle
CollectorCode: 28370
Collector: S. Ahmed



SpecimenID: MUS1078371
CommonName: Tiger Beetle
Category: Beetle
CollectorCode: 45561
Collector: S. Ahmed



Results: Grid View

Record Detail

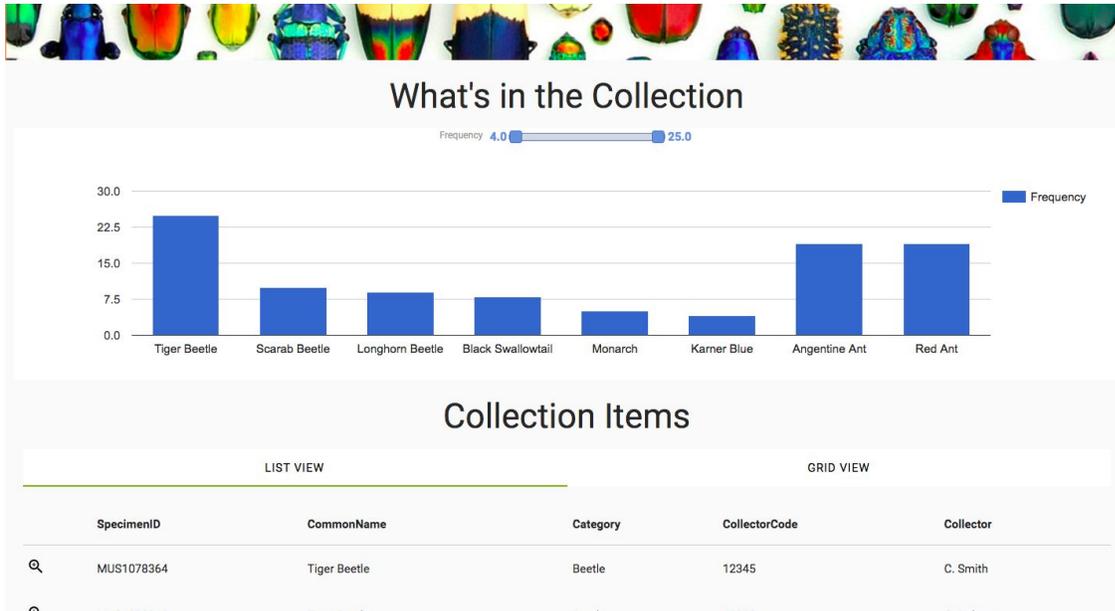
SpecimenID: MUS1078365
CommonName: Tiger Beetle
Category: Beetle
CollectorCode: 45829
Collector: G. Deleuze
State: Texas
City: Marfa
Month: June
Year: 1946
Collection Location: East Wing



CLOSE

MUS1078371	Tiger Beetle	Beetle	45561	S. Ahmed
MUS1078372	Longhorn Beetle	Beetle	86255	C. Smith

Record Modal View

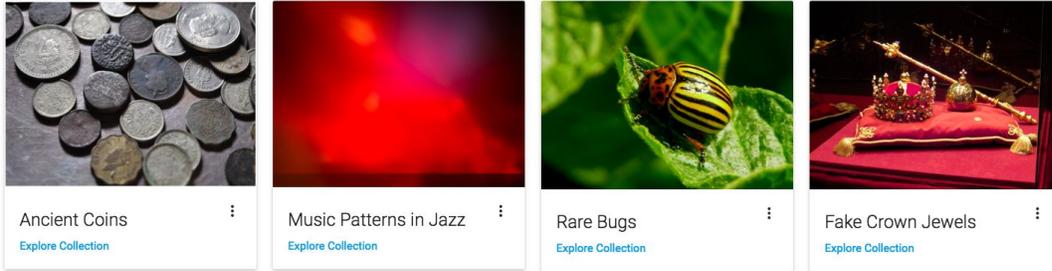


Visualization Page

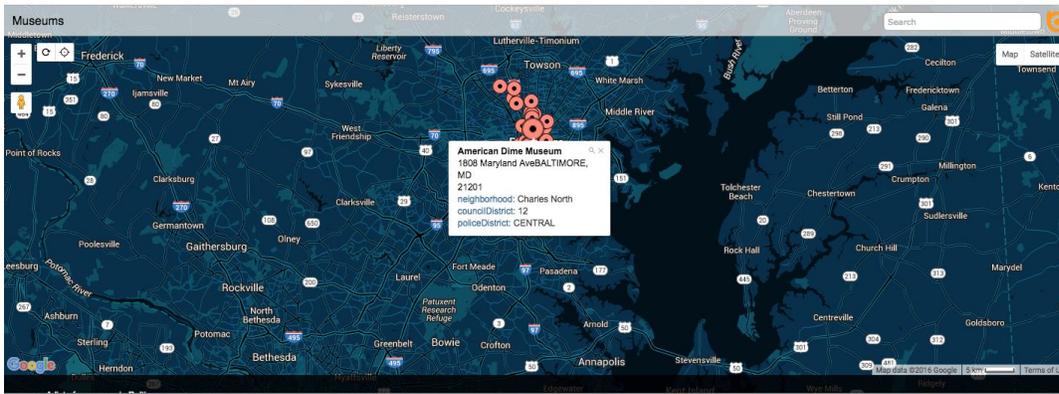
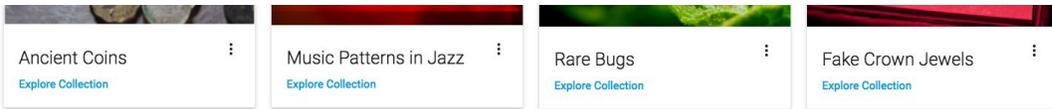
Explore collections around the world

Search by category, keyword or location

MOST POPULAR

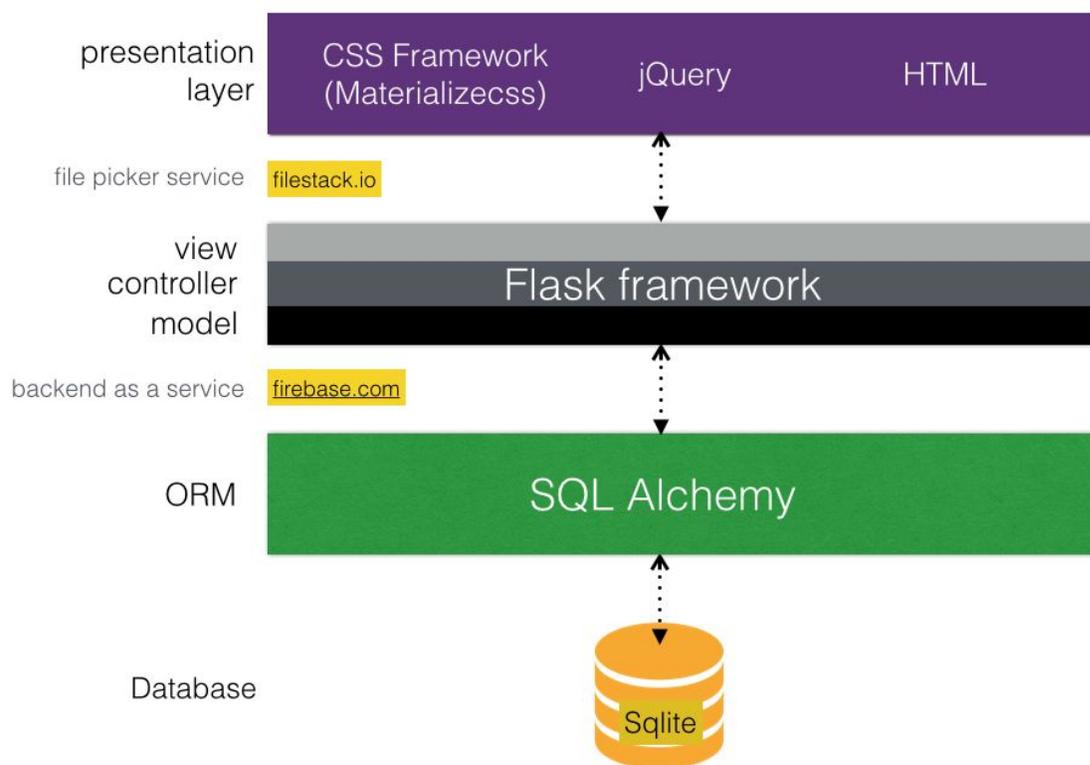


Explore Page: Most Popular



Explore Page: Map

3.3 Product Architecture

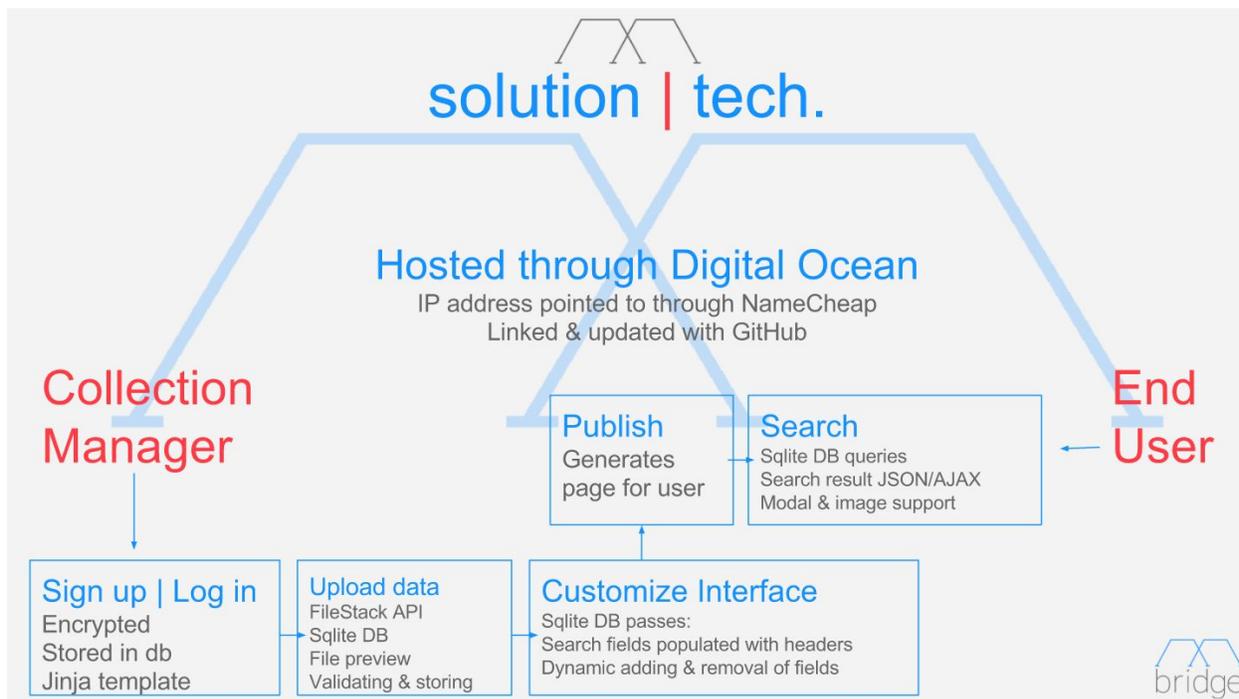


Bridge Product Architecture

The signup/signin page integrates user data and pulls from the database on all pages to see if the user is logged in. The user data is encrypted and stored in a SQLite database. On createUpload.html the user picks a file (from local machine or Google Drive) to upload using FileStack API. A Javascript ajax POST request is done to the /saveFile endpoint on the backend, which processes the file and populates the database from it (search_interface, document, headers) and creates a new table on the fly to store the file in. The front end preview is also updated with this file.

On createSearch.html, the page template is passed the names of the headers from the uploaded dataset. The headers are integrated into the drop down options for creating custom search fields. The CM can dynamically create, edit and delete custom search fields, which is done through JavaScript and then passed to the database. On search.html the search interface is served for users to search and view results. The front-end performs an AJAX call which retrieves a cache of the database and displays the first five columns of data. If any of the columns are named "media", the image is inserted into the gridview. In addition, using JQuery, Materialize modal popups, windows that hover above the page, are dynamically added with buttons for every record that display all columns for the record as well as and the image.

The customized search fields are displayed in the right side bar on the interface. When an end user enters search criteria into the text boxes or drop down menus, the criteria are passed to the database using Python as sql queries. The search results are retrieved through another AJAX call which updates the results table in the interface.



Product Architecture

3.4 Lessons from the Design Process:

- Use a robust front-end framework
- Test application with large collections
- Think about edge cases beforehand
- Front end and back end code are not entirely abstracted away from each other
- Deeper end user research to assess how people with different search goals search for things online
- Explore the concept of a peer-based data-validation system to validate collections



4. Future Design Questions

Explore Page: Some of our user research, both on the collection manager and end-user side, illuminated an interest in indexing collections. One end user cited appreciating websites that aggregated resources where they could search through a range of information. The explore page is a first design draft into how Bridge might connect different collections that use the product: by allowing users to search for collection by keyword to find a particular domain, through mapping collections, where end users can find collections by locality, and by displaying most popular and most recent collections on the Bridge Explore page. A map serves as an aid for filtering collections by location- the idea being that if museums use Bridge to share collections, the end-user can simply search for and click on that particular museum on the map, to view its collections. Moreover, users can get an overview of a collection, simply by looking at a square card, and on further interest, can click on it to view the entire collection.

Mapping real spaces to data: Our target users include archivists at museums, who want to share digital collections that pertain to their physical collections. **How can we allow archivists to map physical museum spaces to digital data?** One way to do this would be to allow collection managers to upload a map of their museum (could be 3D), and attach digital collections to spaces where these collections can be found in the museum.

Analytics: We would like to use the affordances of the internet to gauge and present the interest in collections as helpful statistics to the collection manager.

Interpreting collections: How can we equip collection managers with tools to visually present their data? We have taken a step in that direction with our visualization tool, but it is limited to three types of visualizations. Additionally, we now rely on the collection manager's own interpretation of their data. We would like to head towards a more exploratory design where the collection is automatically and intelligently presented to the users in meaningful ways.

Efficient Media Presentation: Many collections go beyond text, to include images and videos. How can collection managers easily associate and present media with their collections? This is a question we struggled to answer this time, since we don't have any control over the volume and type of media that collection managers would want to present. Moreover we have little foresight on how collection managers would want to associate media with their collections.



Bibliography

Bertin, Jacques. *Graphics and Graphic Information-Processing*. Berlin: Walter De Gruyter, 1981.

Print.

Black, Joanna. Personal Interview. 3 Mar 2016.

Blackaby, Jim, and Beth Sandore. "Building integrated museum information retrieval systems: practical approaches to data organization and access." *Archives and museum informatics* 11.2 (1997): 117-146.

Oboyski, Peter. Personal interview. 17 Dec 2015.

Olsen, Erik. "Museum Specimens Find New Life Online." *The New York Times*. The New York Times, 19 Oct. 2015. Web. 3 Apr. 2016.

Mulji, Aseem.. Personal interview. 3 Dec 2015.

Roston, Michael. "A Guide to Digitized Natural History Collections." *The New York Times*. The New York Times, 19 Oct. 2015. Web. 12 Jan. 2016.

"The Case for Digitization." *Why Digitize Collections? Digital Library Program*, UCLA. Web. 22 Feb. 2016.

Tominski, Christian. *Interaction in Visualization*. San Rafael: Morgan & Claypool, 2015. Print.

Wang, Yiwen, Lora M. Aroyo, Natalia Stash, and Lloyd Rutledge. "Interactive User Modeling for Personalized Access to Museum Collections: The Rijksmuseum Case Study." *User Modeling 2007 Lecture Notes in Computer Science*: 385-89. Web.