

Pilot:

Cambridge Historical Commission Building Files

Digitize > Catalog > Expose > Explore



May 18, 2019, Paul Cote

Resources

These references provide some background on how to organize a digitizing project.

- Digital Public Library of America
Digital Reformatting and File Management
<http://dp.la/info/about/projects/public-library-partnerships/digital-reformatting-and-file-management/>
- National Archives:
**Technical Guidelines for Digitizing Archival Materials for Electronic Access:
Creation of Production Master Files – Raster Images**
<https://www.archives.gov/files/preservation/technical/guidelines.pdf>
- Smithsonian Institution
Digitizing Collections
<https://siarchives.si.edu/what-we-do/digital-curation/digitizing-collections>
- Open Archive Initiative
Standards for Web Content Interoperability
<https://openarchives.org/>

C-Dash Architecture: Documents Linked to Knowledge about People Places, things Events

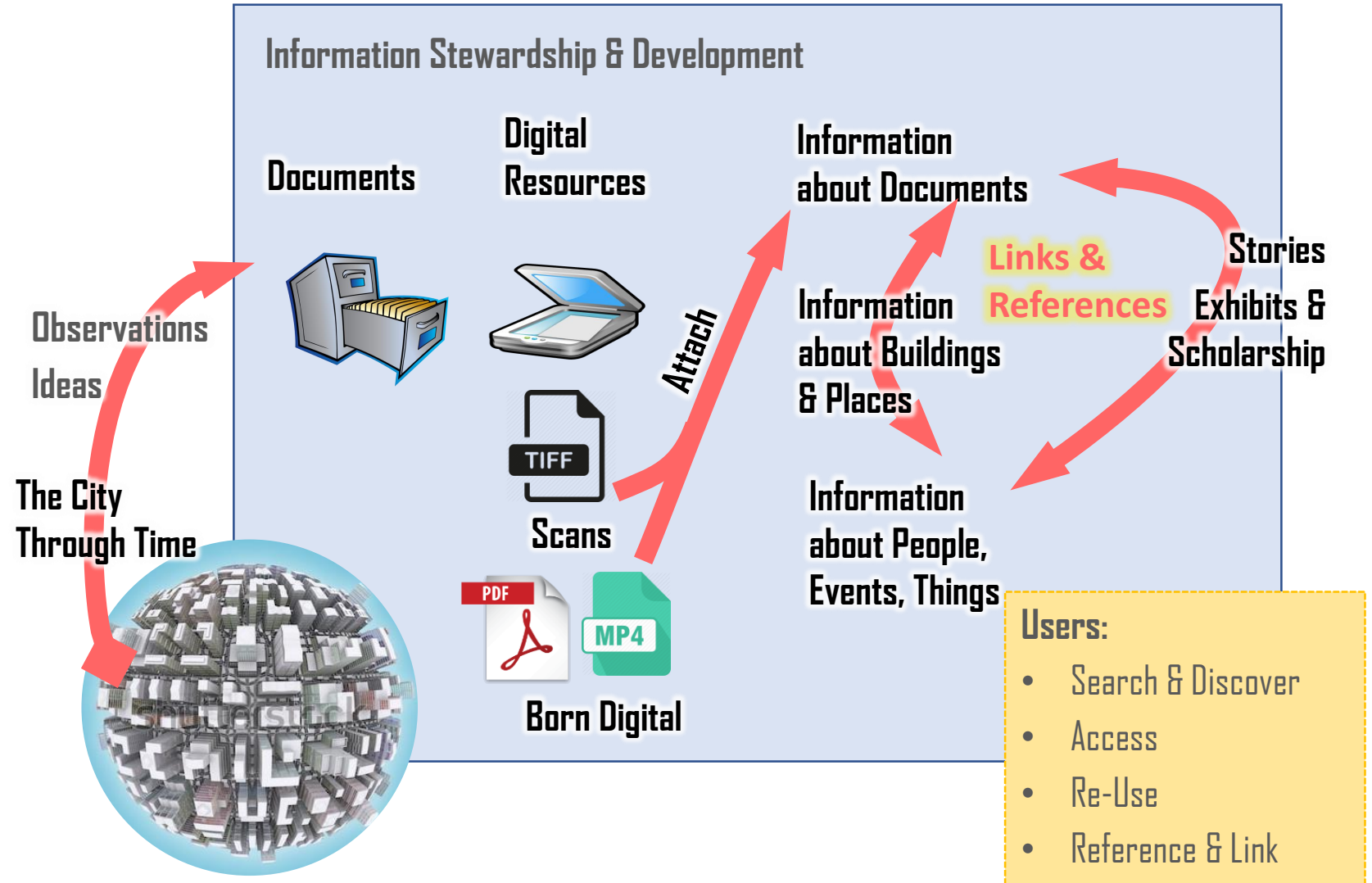
A digital archive for the building files entails a conception of a more holistic system for keeping and developing information about architectural objects in the city.

This project begins with a system for cataloging, annotating and linking **Documents**.

These documents will be used in descriptions of **Buildings and other Named Places**,

In the future, our information system can be used to organize **Accounts of Events** and **Biographical information**, and information about other sorts of things.

This holistic conceptualization will remind us that we will have a container for information about Buildings that is independent from information about documents.

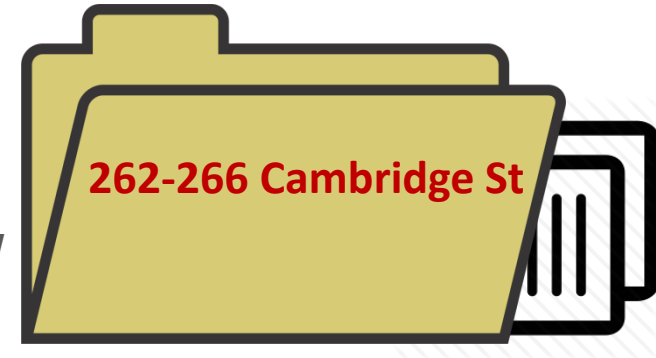


Folders and Items (Existing)

38 File drawers of documentation about Cambridge Buildings



16,000 Folders, each more or less related to a property or building



Items:
000,
001

Step 1. Put Items and Pages in Order

Prepare batches of folders for scanning.

Cover Sheet for Each folder

Multi-Page Documents,
Photo Front and Reverse



Scanning Workflow & Naming Scheme






















Ordered File Names:

264-266_Cambridge_St_0001.tif

264-266_Cambridge_St_0002.tif

264-266_Cambridge_St_0003.tif

 Cambridge_Street	 256_Cambridge_St	 262-266_Cambridge_St_000-PE.jpg JPG File 84.2 KB	 262-266_Cambridge_St_001-AI.jpg JPG File 598 KB
 Fourth_Street	 262-266_Cambridge_St	 262-266_Cambridge_St_002p002-RF.jpg JPG File	 262-266_Cambridge_St_004p001-RN.jpg JPG File
 Otis_Street	 276_Cambridge_St	 262-266_Cambridge_St_004p003-RN.jpg JPG File	 262-266_Cambridge_St_006-CS.jpg JPG File 973 KB
	 292_Cambridge_St	 262-266_Cambridge_St_008-PL.jpg JPG File 965 KB	 262-266_Cambridge_St_009p001-RN.jpg JPG File
	 304-306_Cambridge_St		
	 308_Cambridge_St		
	 314-316_Cambridge_St		
	 320-324_Cambridge_St		

Rough Sort: Identify Common Document Types

The digital collection system relies on titles, and tags to help users find information of interest.

Before uploading bulk-scanned images we can sort them according to rough types that can be evaluated quickly by just looking at the previews for files.

95 percent of all documents in the Building Files collection fit into one of these categories.

These categories allow us to set many useful keywords and generate meaningful titles for each document upon our initial upload into the system.

The rough sort also allows us to identify and associate scans that are parts of multi-page documents, as shown on the next slide.

Text

- AI - Architectural Inventory Form**
- HS - Historic Building Survey**
- RF - Research Form**
- RN- Research Notes**
- CD - Correspondence**
- OB - Official Business**
- AM - Monograph (Article, Paper, Book)**
- EP - Ephemera (Pamphlet, Program, Ads)**
- UC - Unclassified**

Photograph

- PI - Interior**
- PE - Exterior**
- PB - Birds-Eye**
- PD - Detail**
- PC - Contact Sheet**

PP - People Pics (Photo or Drawing)

Illustration

- II - Interior**
- IE - Exterior**
- IB - Birds-Eye**
- ID - Detail**
- IP - Plan/Map**

Rough Sort and Application of Logical File Names

The rough sort and pagination is accomplished by a renaming tool

The user selects the pages of a document from one file explorer window. Then drag and drops the pages onto the appropriate renaming tool.

The result is that the files are given logical file names that identify the Address, Item identifier, Page Identifier and Rough Document Type.

These renamed files are now ready for geo-coding and application of automatically assigned titles and tags.

The image shows a Windows File Explorer window titled 'item_sort_icons_20190522'. It displays a grid of icons representing different document types, each with a two-letter code and a description:

AI Architectural Inventory	RF Research Form	RN Research Notes	HS Historic Bldg Survey	CB CHC Business	AM Article Monograph	CD Correspondence
PE Exterior Photo	PI Interior Photo	PD Detail Photo	PB Birds-Eye Photo	PC Photo Contact Sheet	EP Epitaph	UC Unclassified
IE Exterior Illustration	II Interior Illustration	ID Detail Illustration	IB Bird's Eye Illustration	IP Plan Illustration	Shift	

Annotations on the image include:

- A red dashed box around the 'PE' icon with the text 'Drag and Drop On naming tool' and an arrow pointing to the bottom window.
- A yellow dashed box around the 'PE' icon with the text 'New Logical File Names:' and a list:
 - 264-266_Cambridge_St_001p001-PE.tif
 - 264-266_Cambridge_St_001p002-PE.tif
- A red dashed box around the 'PE' icon with the text 'Select Pages' and an arrow pointing to the bottom window.
- A yellow dashed box around the bottom window showing a list of files with the text 'Initial Scan Names' and a list:
 - 264-266_Cambridge_St_023.tif
 - 264-266_Cambridge_St_024.tif

The bottom window shows a list of files with names like '336-672_Cambridge_St' and '336_Cambridge_St_000 4p0001-VE.tif'. The files are arranged in a grid, and some are highlighted with a red dashed box.

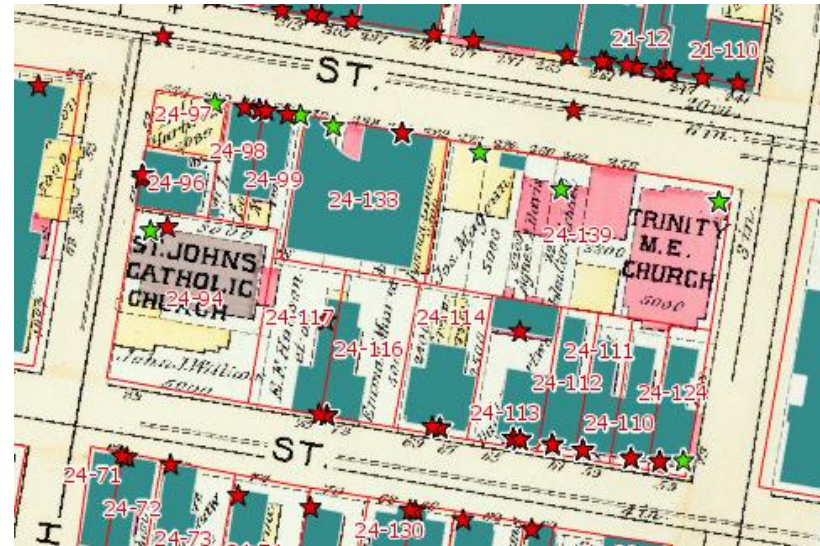
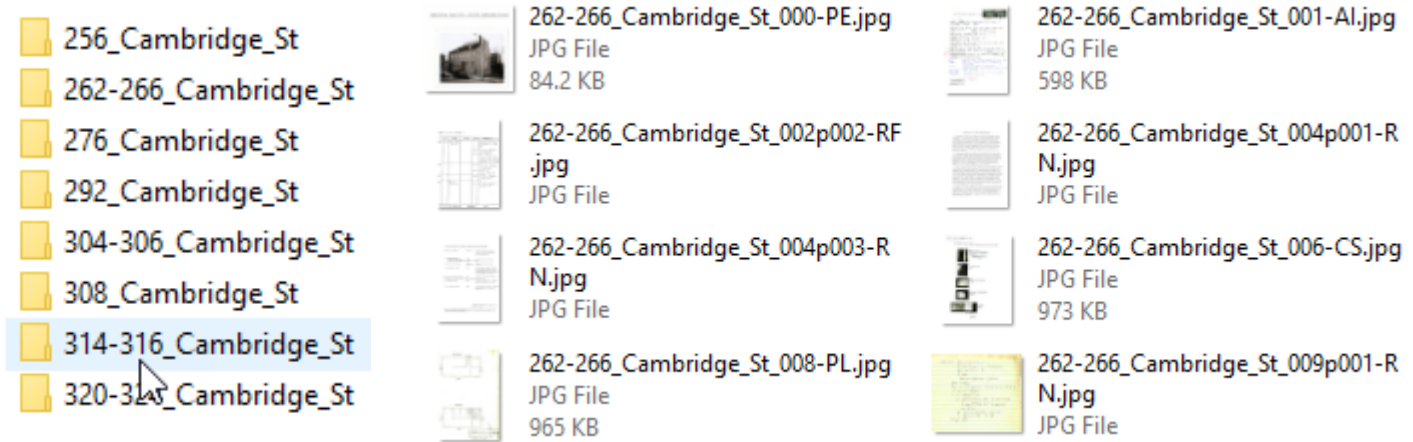
Geocode Addresses and Supplement Master Address List (Using Cambridge Master Address File and Assessing Property Addresses)

Python scripts match CDASH addresses with known address locations from Master Address File and Assessing Database

Many addresses match perfectly. A sequence of steps deals with range addresses and interpolates non-matched house numbers, finally assigning a default location for addresses that find no match.

Result is a dictionary of all CDASH Addresses with:

- Latitude
- Longitude,
- Map-Lot
- Building_ID
- Match Score
- Match Note



Shape	Full_Addr	BldgID	ml	POINT_X	POINT_Y
Point	39 Fourth St	416-31	23-36	-71.081493	42.371644
Point	113-A Fourth St	530-10	26-81	-71.08247	42.368326
Point	111 Fourth St	530-6	26-80	-71.082266	42.368371
Point	113 Fourth St	530-12	26-81	-71.082312	42.368311
Point	119-1/2 Fourth St	530-31	26-151	-71.08254	42.368102
Point	117 Fourth St	530-24	26-150	-71.082433	42.36817
Point	105 Fourth St	507-34	26-59	-71.082241	42.368683
Point	106 Fourth St	514-19	24-2	-71.081914	42.36875
Point	104 Fourth St	514-16	24-3	-71.081913	42.368833
Point	118-R Fourth St	533-14	17-22	-71.081987	42.368169
Point	106-R Fourth St	514-19	24-2	-71.081821	42.368741
Point	23 Fourth St		22-142	-71.081455	42.372479

Correction and Augmentation of Address and Street Layers Using GIS

(Extending Cambridge Master Address File and Street Centerlines)

The CDASH Address Dictionary is opened in ArcMap

Addresses with low scores are corrected.

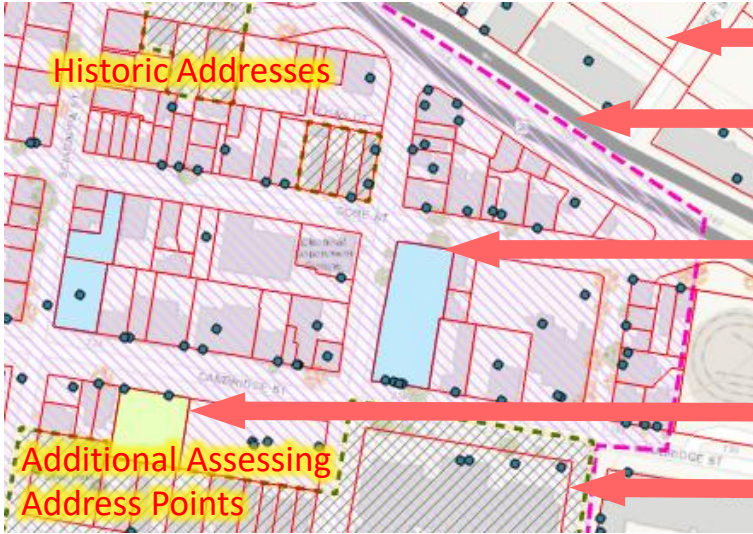
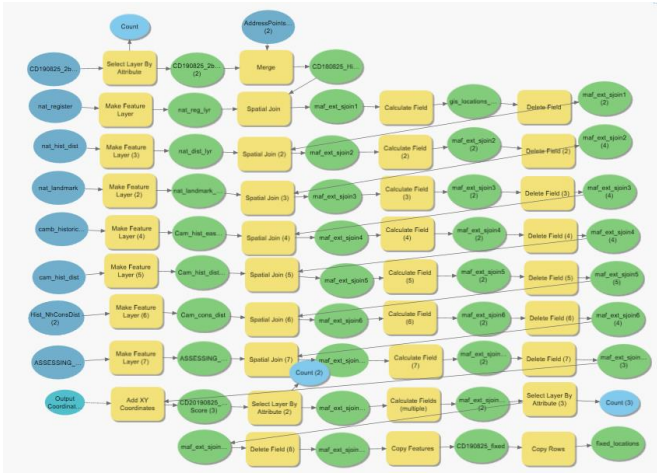
Non-Address locations, like water bodies, parks, etc. have names applied.

GIS Overlay techniques are used to assign official district and historic landmark designations to addresses that fall within GIS layers of historic building information, Titles and names for these locations are assigned.

Assessing information is assigned to re-located address points.

Historical street names are added to a CHC Street Centerlines layer.

At the end we have additions to the Extended Address File that includes historic address locations.



- Assessing Parcels
- Cambridge Neighborhood conservation districts
- Cambridge Landmark Easements
- National Historic Landmarks (not shown)
- National Register
- National Historic District

Prepare Batch of Scans with Rough Catalog

For each image file in the scans folder:

Use the item-type dictionary to apply meaningful titles and keywords

Use the Address Points dictionary to look up the coordinates, building-ID and Map-Lot.

Write a comma-delimited table formatted for the Omeka CSV import tool

Future version could:

- Embed the metadata and coordinates into the image files themselves.
- Import the collection into any content management system that permits systematic import.

Address, Geo-Location & Building and Parcel Links

Shape	Full_Addr	BldgID	ml	POINT_X	POINT_Y
Point	39 Fourth St	416-31	23-36	-71.081493	42.371644
Point	113-A Fourth St	530-10	26-81	-71.08247	42.368326
Point	111 Fourth St	530-6	26-80	-71.082266	42.368371
Point	113 Fourth St	530-12	26-81	-71.082312	42.368311
Point	119- 1/2 Fourth St	530-31	26-151	-71.08254	42.368102
Point	117 Fourth St	530-24	26-150	-71.082433	42.36817
Point	105 Fourth St	507-34	26-59	-71.082241	42.368683
Point	106 Fourth St	514-19	24-2	-71.081914	42.36875
Point	104 Fourth St	514-16	24-3	-71.081913	42.368833
Point	118-R Fourth St	533-14	17-22	-71.081987	42.368169
Point	106-R Fourth St	514-19	24-2	-71.081821	42.368741
Point	23 Fourth St		22-142	-71.08145	42.372479

Rough Item-Type Metadata

- **AI: Architectural Inventory Form**
- **RF: Research Form**
- **PE: Exterior View**
- **PI: Interior Detail**
- **AM: Article or Monograph**
- **RN: Research Notes**



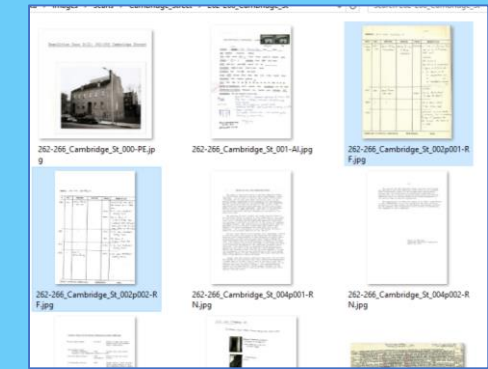
Python and ArcGIS Procedure

Batch
Rough Archive

Batch 20180212

Batch Catalog

Re-Named Images



Introducing Omeka-S

Omeka-S is an open source Content Management System designed for the Library, Archive and Museum Community.

Omeka has many advanced functions for attaching metadata to items, and linking information together across collections and systems – e.g. building photos to building records of other types.

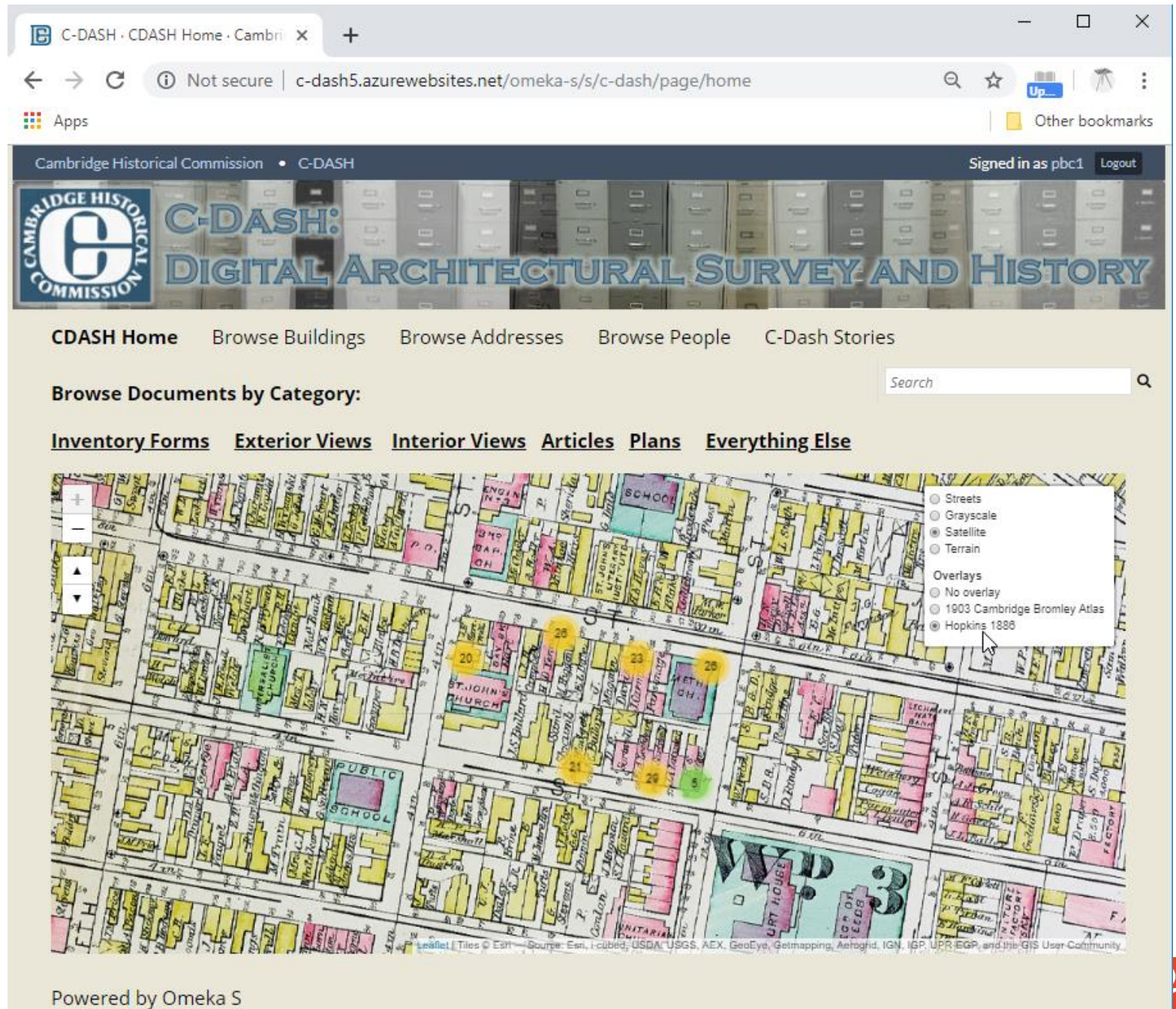
Omeka has been under development since 2008, has a very active community of users and developers.

Many high-profile installations use Omeka-S to develop searchable digital collections and to integrate archival information into web-based exhibits.

Omeka_S is a PHP application similar to the popular content management system, word-press. It is relatively easy to install, configure and modify for someone with basic linux skills.

Visit the C-Dash Omeka-S Pilot Site at the temporary URL:

<http://chcomeka.azurewebsites.net/omeka-s>



The screenshot shows a web browser window displaying the C-DASH website. The browser's address bar shows the URL c-dash5.azurewebsites.net/omeka-s/s/c-dash/page/home. The website header includes the Cambridge Historical Commission logo and the title "C-DASH: DIGITAL ARCHITECTURAL SURVEY AND HISTORY". The navigation menu contains links for "CDASH Home", "Browse Buildings", "Browse Addresses", "Browse People", and "C-Dash Stories". A search bar is located on the right side of the page. Below the navigation, there is a section titled "Browse Documents by Category:" with links for "Inventory Forms", "Exterior Views", "Interior Views", "Articles", "Plans", and "Everything Else". The main content area features a historical map of a city block, overlaid with a legend. The legend includes options for "Streets", "Grayscale", "Satellite", "Terrain", "Overlays", "No overlay", "1903 Cambridge Bromley Atlas", and "Hopkins 1886". The map shows various buildings, streets, and landmarks, with some areas highlighted in yellow and green. The bottom of the page indicates "Powered by Omeka S" and the "bcgis" logo is visible in the bottom right corner.

Omeka Tour: Delegate Admin and Contributor Privileges

Control look, feel and Branding of site using pre-fab [“Themes”](#)

[Create new Items](#)

[Upload Files](#) (associate them with Items)

[Edit Metadata](#) and [Location Info](#) for Items

[Create New Item Types](#)

[Create new metadata fields](#) and controlled lists (MACRIS-ID)

[Inspect and Edit Tags Globally](#)

[More Omeka Documentation](#)

Admin dashboard · Cambridge H x +

Not secure | c-dash5.azurewebsites.net/omeka-s/admin

Apps | Other bookmarks

Admin dashboard

Welcome to the Cambridge Historical Commission admin dashboard!

Manage resources

- Items (154) +
- Item sets (16)
- Vocabularies (4)
- Resource templates (3)

Manage sites

- C-DASH

Login

Email * paulbcote@gmail.com

Password * [masked]

Log in

Forgot password?

Powered by Omeka 5

Beyond Omeka: Open Archive Strategy

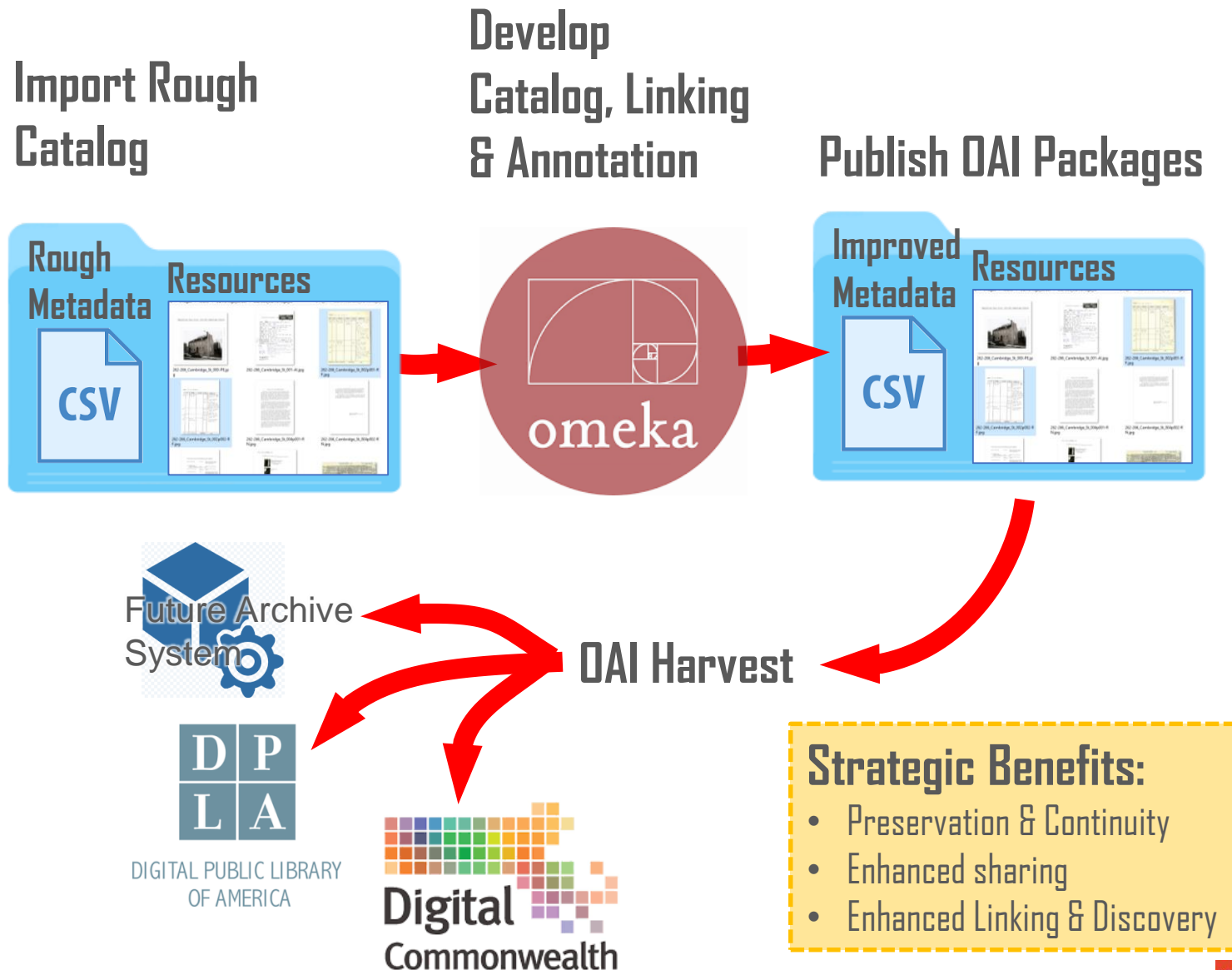
The [Open Archive Initiative](#) (DAI) provides a standard for exchanging resources and metadata between digital archive systems

Following the DAI Protocol for Metadata Harvesting (DAI-PMH) ensures that the collection does not become "locked in" to a specific software platform.

Our metadata template is based on crosswalk of the Omeka Base Resource and the Digital Commonwealth DAI Interchange template.

Our crosswalk template and other useful resources from digital commonwealth are available at www.pbcGIS.com/cdash/crosswalk.

The C-Dash base resource template is used as a starting point for resource templates specialized for our various document types.



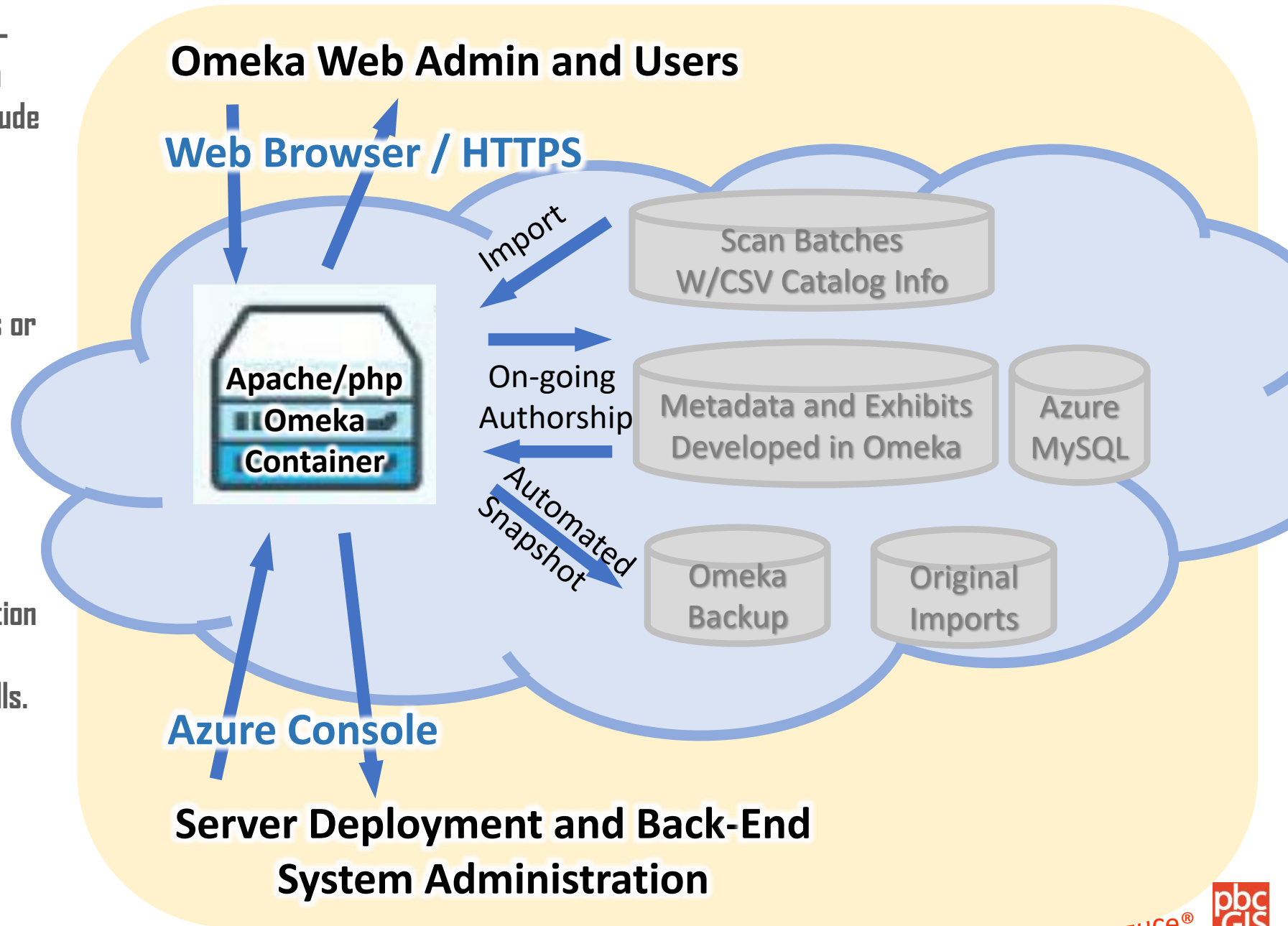
Developing, Using and Preserving Historical Resources with Cloud Application

The next phase of development for the C-Dash involves developing of a production system. A few goals for this system include protecting information assets from:

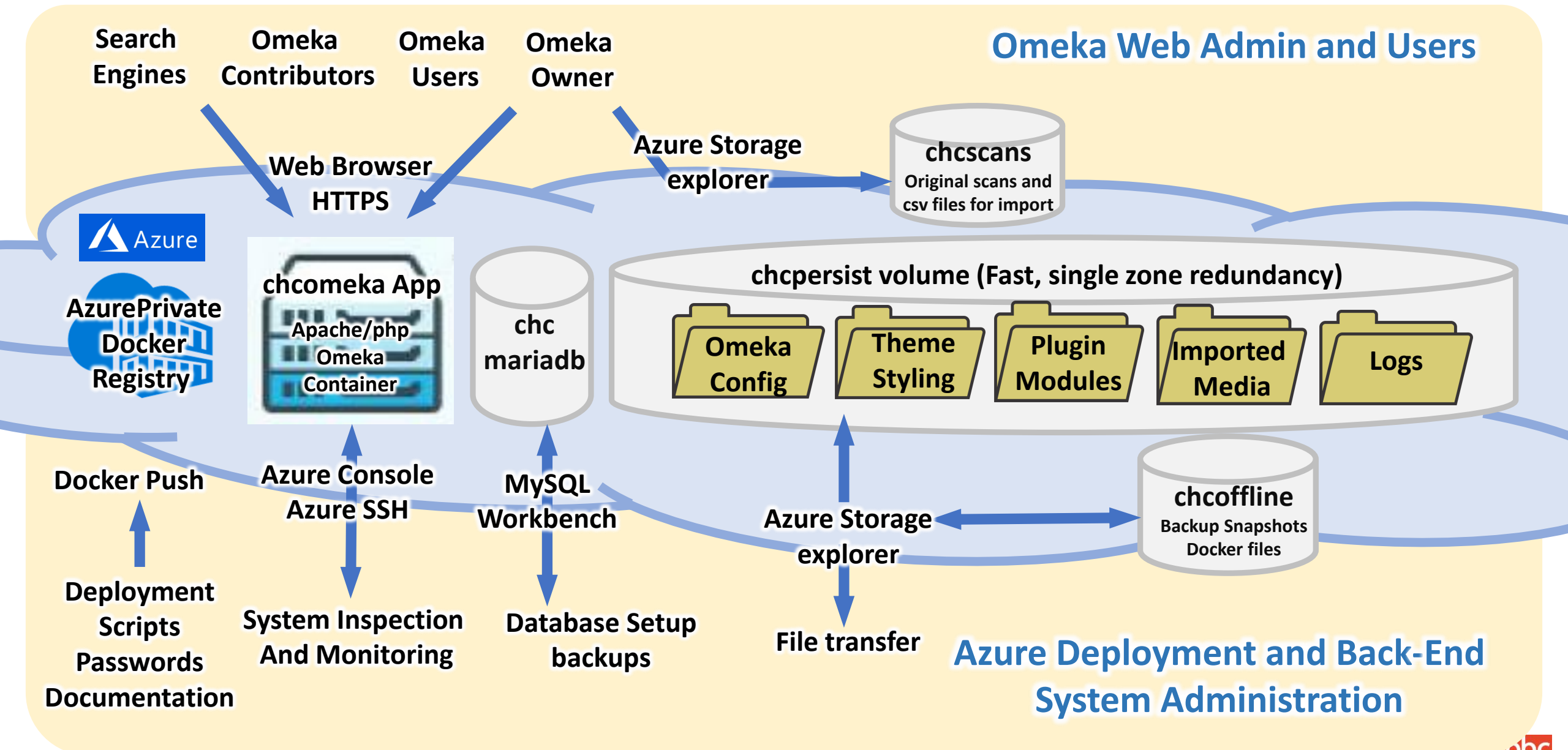
- Disasters
- Mistakes
- Unplanned obsolescence of software
- Existential dependence on Individuals or private firms

Our design addresses these goals in a Microsoft Azure Web Application

- Server maintenance and data redundancy are managed by Azure.
- The application, data and documentation are easily understood by system administrators with off-the-shelf skills.
- Recommend review of Cambridge IT department.
- Consider hosting Cambridge Azure Subscription.



Omeka in Azure Cloud (detail)



Roles & Responsibilities (Draft)

CHC Omeka Owner: (Emily? Charlie?)

Is responsible for the long term viability of the Omeka installation.

Makes and maintains agreements with the IT department.

Understands executive view of file management, backup and recovery.

- Receives reports of routine snapshot activity

CHC CDASH Administrator (Meta)

Directs ongoing development of resources

- Create new Omeka accounts
- Understands how to do bulk imports and Routine item import

Delegates design and development

- metadata templates
- Exhibits
- Create and improve Item Metadata and linking.

Omeka/Azure Developer: (pbc)

Designs and implements C-Dash system

- Development (test) sandbox
- Production environment
- Documentation
- Training
- Handles batch rough-sort, initial catalog and Omeka Imports
- Assists C-Dash Owner with developing look/Feel and Features
- Assists with developing new metadata templates and linking infrastructure

Monitors regular snapshot activities

Tests and verifies recoverability

Maintains all system rebuild scripts and passwords, up-to-date in a secure location that is accessible to Omeka/Azure Owner.

Azure Owner (Cambridge IT)

Pays monthly Azure bills

Holds full rights to delegate anything on Azure resources

Evaluates and has last word on system design.

Understands enough about the architecture, rebuild scripts, passwords and documentation to delegate full system control to new Omeka/Azure Developer in case of emergency.

Omeka/Azure Developer Capabilities

In the event that the Omeka/Azure developer should need to be replaced, a consultant who has experience with commonphp/linux applications such as wordpress should be able to figure out how this system works from the documentation.

Cambridge-Specific Internet Interfaces (Suggested)

CHC Domain

Ideally, once this service is ready for production, it should have a regular URL within the Cambridgema/CHC branch of the web site.

www.cambridgema.gov/chc/c-dash

HTTPS Certificates:

For now the Omeka service is using plain http. Once users start registering we may want to use HTTPS. Would it make sense to use a Cambridge certificate, or create a new one for C-Dash?

(Optional) CHC SendMail Address

Certain aspects of new user registration and notification require the sending of emails. This requires reference to a sendmail server. Optimally, these emails would come from.

(Optional) VPN

Currently all files are transferred through the Azure Management Console. The bulk import workflow involves one stage of unzipping the archives using the SSH console. Eventually, this could be more user friendly to non-linux users if file access was through a VPN connection – which, it seems requires authentication information on the client. This is not urgent, and requires further investigation.