



Brown Dog Data Tilling Service

Luigi Marini

July, 22nd, 2002



National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign

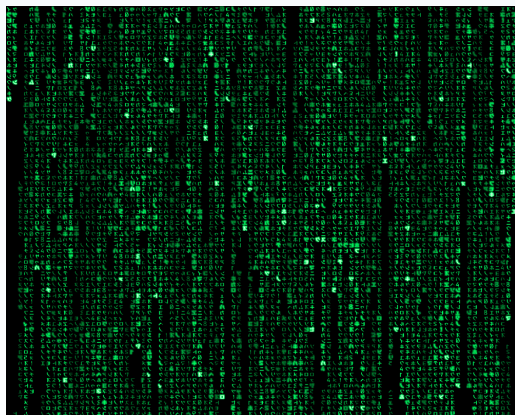
Outline

- DTS?
- Brief background on building blocks
 - Medici
 - Versus
- Sample use cases
 - Census
 - LSVA
 - Great Lakes Monitoring / IML-CZO
- A few more details about DTS
- Adding tools
 - Extractors
 - Metrics

Data Tilling Service

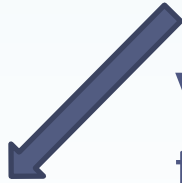
- Till (v): “to prepare and cultivate (land) for crops” (google)
- Data Tilling (v): To prepare and cultivate (*data*) for *analysis*
- Tillage (n): “is the agricultural preparation of soil by mechanical agitation of various types, such as digging, stirring, and overturning.” (wikipedia)
- Data Tillage (n): Is the *computational* preparation of *data* by *algorithmic* agitation of various types, such as digging, stirring, and overturning

Data Tilling Service



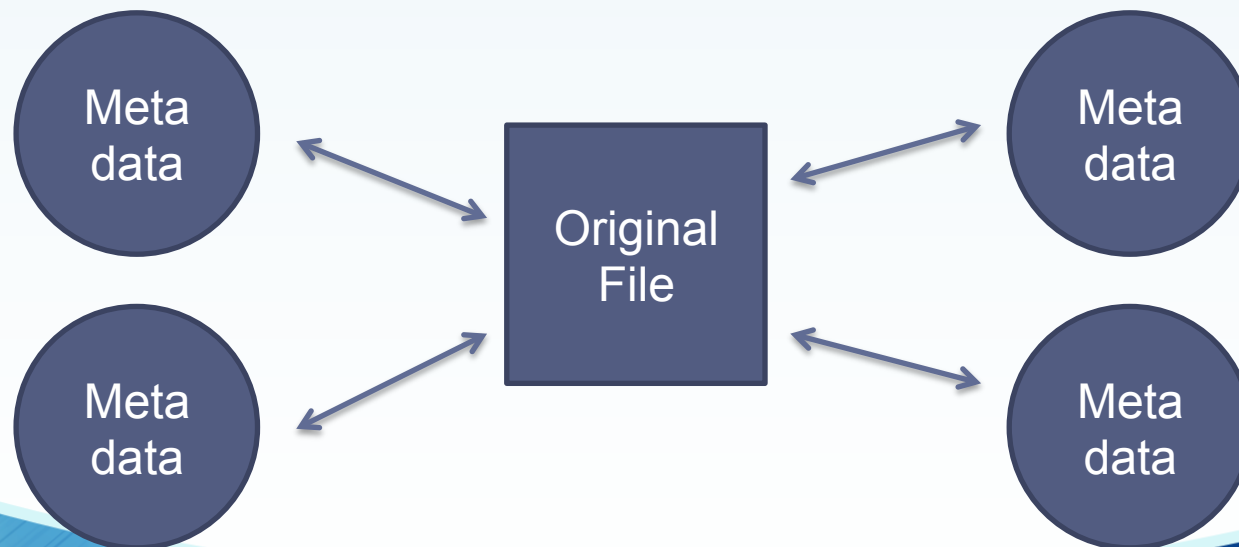
Data
+
Metadata

What falls out
View into the raw data

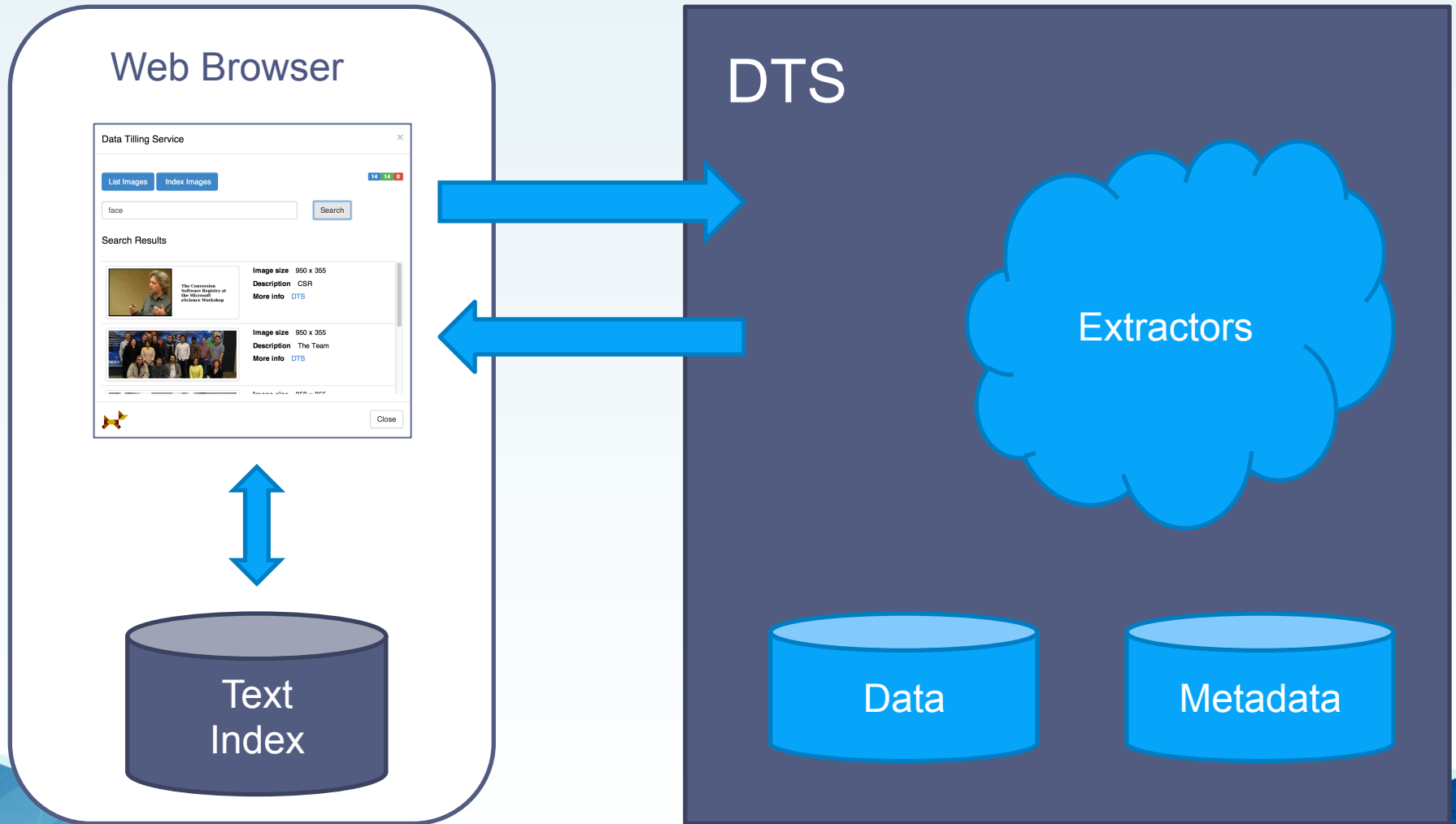


(Pre) Data Analysis

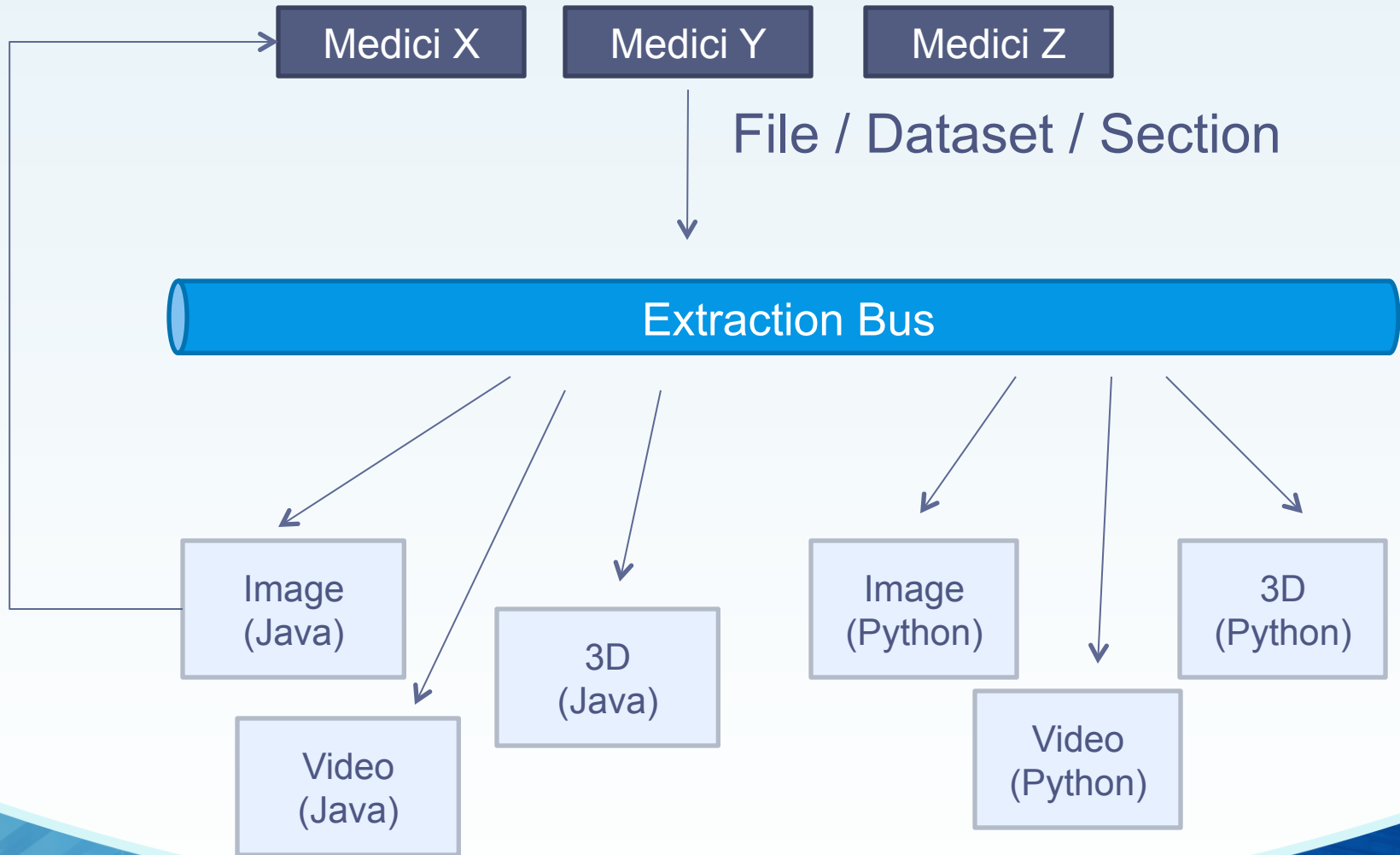
- Not necessary *data cleaning*
- More like metadata extraction
- Not full analysis / Not perfect results
- Apply as many methods as possible
- Support the user in finding the metadata they need



Bookmarklet



Extraction Bus



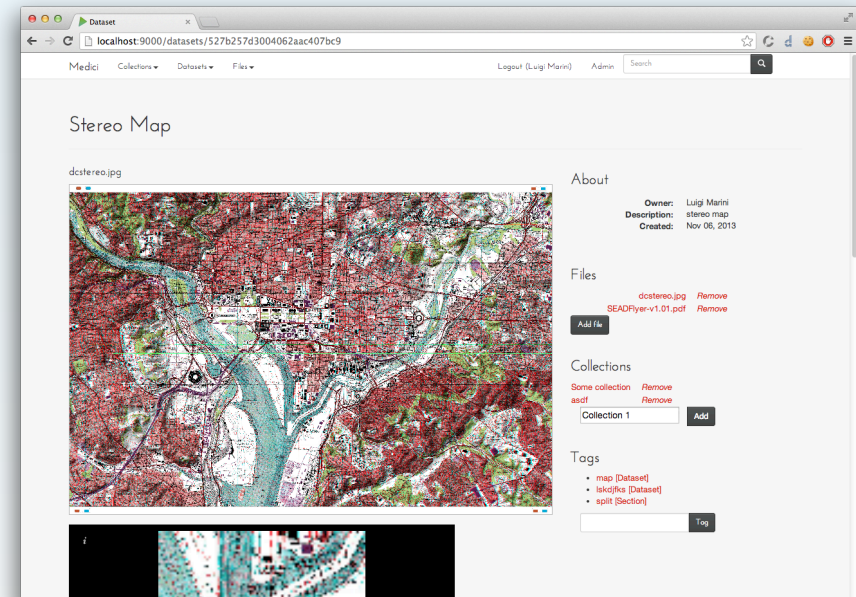
2 building blocks

MEDICI & VERSUS



Medici - Data Management for Research Data

- Any file type
 - Like a file system or cloud storage
- Upload first, organize later
 - Social curation
 - Tags, collections, datasets, spaces, generic metadata
- Extensible
 - Creating new data extractors
 - Creating new data previewers



Organize, Search, Analyze

Open source

- Install service on your own resources
- We host instances at NCSA
 - 18 instances for SEAD Datatnet
 - 15 other instances
- Brown dog DTS maintained by NCSA
 - All code available can be installed and maintained somewhere else

Timeline

- v.1 Development started November 2009
- v.1 First public release June 2010
- v.2 Development started November 2012

Drivers

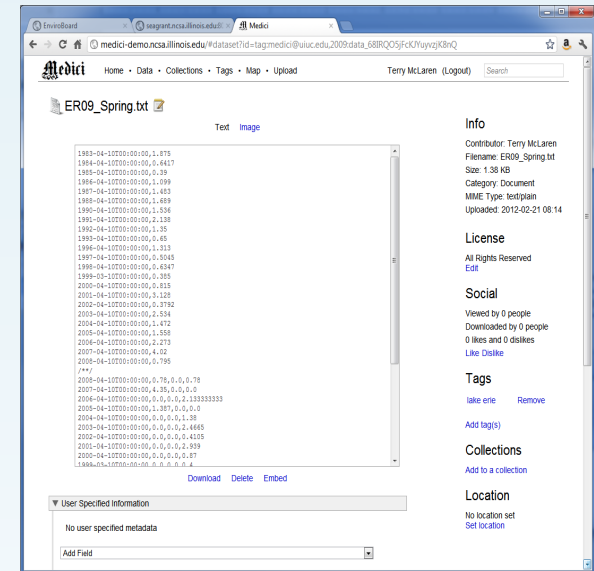
- Cyberinfrastructure that works well across disciplines
 - physical sciences, biology, medicine, humanities, arts, and social sciences
- Large-scale community collections of heterogeneous data and metadata
 - documents, images, video, 3d, sensor, gis, etc.
 - metadata fields based on specific uses cases
- Rapid growth from
 - High throughput instruments
 - Digitization efforts
 - More sources (cameras, phones, tablets)

Funding Sources

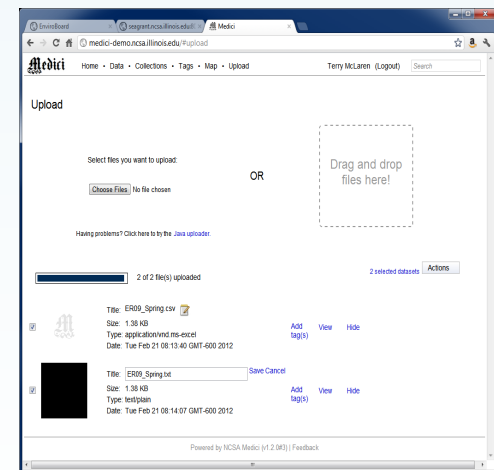
- **ONR** - *Technology Research, Education and Commercialization Center*
- **NARA/NSF OCI** – *Understanding Data Intensive and CPU Intensive Services to Support Preservation and Reconstruction of Electronic Records*
- **NSF CDI** – *Groupscape: Instrumenting Research on Interaction Networks in Complex Social Contexts*
- **NSF DataNet** - *Sustainable Environments-Actionable Data (SEAD)*
- **NSF EAR** – *Critical Zone Observatory Network for Intensively Managed Landscapes (IML-CZO)*
- **NSF/NEH/JISC** – *Digging into Image Data to Answer Authorship Related Questions*
- **NIH** – *Immunomodulatory and Regenerative Effects of Mesenchymal Stem Cells on Allografts*
- **Sea Grant** – *Great Lakes Monitoring*
- **European Commission** – *Linking Scientific Computing in Europe and the Eastern Mediterranean*
- **XSEDE** – *Large Scale Video Analytics*
- **NSF ACI** – *CIF21 DIBBs: Brown Dog*

Medici - Features

- Upload / download
- Automated data extraction and analytic services
- Data type previews
- Indexing / Search / browse
- Tag / comment
- Create collections
- Geo-locate data (map view)
- Define a specific taxonomy
- Access statistics, data provenance
- Citable persistent URLs
- Author defines copyright and license attributes
 - View only, prevent download
- Create relationships between datasets



Archive View



Upload View

Dataset Page

Data Preview

Data provenance

User defined information

Extracted Metadata

The screenshot shows a web browser window displaying a dataset page for 'dataset_9_image1.png'. The page is titled 'Active Content Repository' and includes a navigation menu with 'SEAD', 'Home', 'Data', 'Collections', 'Tags', 'Map', and 'Upload'. The user is identified as 'Terry McLaren (Logout)'. The main content area features a world map with a 'Zoom' button. Below the map are 'Download' and 'Embed' buttons. A 'User Specified Information' table lists the creator as 'James Myers' and 'Luigi Marini'. The 'Info' section provides details such as 'Contributor: James Myers', 'Creator: James Myers', 'Filename: dataset_9_image1.png', 'Size: 107.19 KB', 'Category: Image', 'MIME Type: image/png', and 'Uploaded: 2012-08-09 12:59'. The 'License' is 'All Rights Reserved'. The 'Social' section shows 'Viewed by 8 people', 'Downloaded by 1 people', and '0 likes and 0 dislikes'. The 'Tags' section includes 'nced' and a 'Remove' button. The 'Collections' section shows two collections: 'NCED Abstract images' and 'ACR Demo', each with a 'Remove' button.

Field	Value	Applies To	Action
Creator	James Myers	Document	Edit Remove
	Luigi Marini	Document	Edit Remove

Social Annotations

Data is part of multiple collections

Shapefile Previewer

The screenshot shows a web browser window with the URL `sead.ncsa.illinois.edu/acr/#dataset?id=tag:medici@uiuc.edu,2009:data_AV8LUuDppz5MPHi0b5q5yA`. The page is titled "Medici" and has a navigation menu with links for Home, Data, Collections, Tags, Map, Upload, and Administration. A search bar is located in the top right. The user is logged in as "Luigi Marini (Logout)".

The main content area displays a file named "angelo_basins.zip" with a document icon. Below the filename, it is categorized as "Geospatial Data". A map is shown with a grey-shaded area representing the basins. The map includes a scale bar at the bottom right with coordinates `-13809033.36704, 4814974.96988`. Below the map are several action buttons: "Download", "Delete", "Rerun Extraction", "Embed", and "Upload Derived Data".

On the right side, there is an "Info" section with the following details:

- Creator(s):
- Filename: `angelo_basins.zip`
- Size: 618.22 KB
- Category: Other
- MIME Type: `application/zip`
- Uploaded By: Medici Admin
- Uploaded: 2012-11-28 08:10

Below the info section is the "Data Maturity" section, which shows a dropdown menu for "Current level" set to "Raw".

The "License" section indicates "All Rights Reserved" with an "Edit" link.

The "Social" section shows statistics: "Viewed by 11 people", "Downloaded by 4 people", and "0 likes and 0 dislikes". There are "Like" and "Dislike" buttons.


At the bottom, there are "Tags" for "angelo" and a "Remove" button.

At the bottom of the page, there are two expandable sections: "Collection Context" and "User Specified Metadata".

3D Previewer (x3d, ply)

ICACH Collections ▾ Datasets ▾ Files ▾ Newsletter ▾ Tags ▾ API Login Admin ▾ Search

Z Remove measurement
J Lighting on/off
V Single-sided/double-sided lighting
C drag Change lighting direction
Shift + Q Toggle fullscreen view



[Show/edit file user-generated metadata](#)

Comments Metadata Notes

Streaming Video

The screenshot shows a web browser window with the address bar displaying the URL: `ion-21-15.sdsc.edu:9000/files/53c05c051038bd0822f9ad5d`. The browser's address bar also shows several icons, including a star, a refresh button, and a search icon. The browser's toolbar includes various application icons such as ISDA, Data, Medici, DTS, Delicious, Grainger, Confluence, JIRA, Stash, and Illinois Clearinghouse. The page content includes a navigation menu with items like 'Medici', 'Collections', 'Datasets', 'Files', and 'API', along with 'Login' and 'Admin' links. The main content area displays the filename 'saturday-marathon_512kb.mp4'. Below the filename, the following metadata is shown:

- Author:** Anonymous User
- Filename:** saturday-marathon_512kb.mp4
- Type:** video/mp4
- Uploaded on:** Fri Jul 11 14:49:57 PDT 2014

Below the metadata, there are two buttons: 'Video (popcorn player)' and 'Video'. The 'Video (popcorn player)' button is highlighted in red. Below these buttons is a video player showing a scene of a marathon on a city street. The video player has a progress bar at the bottom, a play/pause button, a volume icon, and a full-screen icon. The current time displayed is 1:07.

Extraction Service - Images with GPS

- Extract GPS location and displays location in Google map.

Date/Time Original	2010-03-05 08:51:14
Exit Image Height	2592 pixels
Exit Image Width	1944 pixels
Exit Version	2.20
FlashPix Version	1.00
Make	google
Model	Nexus One
Resolution Unit	Inch
Thumbnail Data	[61846 bytes of thumbnail data]
Thumbnail Length	61846 bytes
Thumbnail Offset	678 bytes
X Resolution	72 dots per inch
Y Resolution	72 dots per inch
YCbCr Positioning	Center of pixel array
Extractor	
Extractor created	2012-03-28T15:15:23.237Z
Extraction stage 0 start	2012-03-28T15:15:23.246Z
Extraction stage 0 stop	2012-03-28T15:15:23.255Z
Extraction stage 1 start	2012-03-28T15:15:23.484Z
Extraction stage 1 stop	2012-03-28T15:15:25.880Z
Extraction stage 2 start	2012-03-28T15:15:26.037Z
Extraction stage 2 stop	2012-03-28T15:15:36.894Z
Extraction stage 3 start	2012-03-28T15:15:37.086Z
Extraction stage 3 stop	2012-03-28T15:15:37.096Z
Image Height	2592
Image Size	1944x2592
Image Width	1944
GPS	
GPS Altitude	207 meters
GPS Altitude Ref	Sea level
GPS Latitude	40°6'16.57"
GPS Latitude Ref	N
GPS Longitude	88°19'32.6"
GPS Longitude Ref	W
GPS Map Datum	WGS-84
GPS Time-Stamp	8.51.14 UTC
GPS Version ID	2.2.0
Interoperability	

Active Content Repository

sead.ncsa.illinois.edu/#dataset?id=tag:medici@uiuc.edu,2009:data_luLwKoMK9-hHGMyPJkTQ

SEAD Home • Data • Collections • Tags • Map • Upload Terry McLaren (Logout) Search

morrowPlot-withGPSLocation.jpg Image Zoom

Download Delete Embed

▼ User Specified Information

No user specified metadata

Add Field

▶ Extracted Information

▶ User Views

Info

Contributor: Terry McLaren
Creator:
Filename: morrowPlot-withGPSLocation.jpg
Size: 1.26 MB
Category: Image
MIME Type: image/jpeg
Uploaded: 2012-03-28 10:15
Image Size : 1944x2592

License

All Rights Reserved
[Edit](#)

Social

Viewed by 8 people
Downloaded by 0 people
0 likes and 0 dislikes
[Like](#) [Dislike](#)

Tags

[Add tag\(s\)](#)

Collections

[ACR Demo](#) [Remove](#)
[Add to a collection](#)

Location

Embedding

- Embed data in external sites

The screenshot shows a web browser window with the URL `medici-demo.ncsa.illinois.edu/#dataset?id=tag:medici@uiuc.edu,2009:data_XX_c1Q3L3uGAsuMVrsg-DQ`. The main content area displays a historical black and white photograph of a multi-story building on fire. A large crowd of people is gathered in front of the building, and a ladder is leaning against the facade. Below the image, there is a caption: "Shows the rapidity with which the ladders and cage may be sent to the highest point required. The time occupied from the escape arrives until the first delivery is rescued is less than two minutes." Below the caption are buttons for "Download", "Delete", "Rerun Extraction", and "Embed".

On the right side of the browser window, there is a metadata section showing "Uploaded: 2011-10-17 15:16" and "Image Size : 1136x1600".

A "Preview Embedded Dataset" dialog box is open in the foreground. It has a title bar and three tabs: "Image", "Zoom", and "Metadata". The "Image" tab is selected, showing a smaller version of the building fire image. Below the image, there is a caption: "Shows the rapidity with which the ladders and cage may be sent to the highest point required. The time occupied from the escape arrives until the first delivery is rescued is less than two minutes." Below the caption, there is a text field containing the URL "MainSt_Upr-TownHall..." and the Medici logo. A "Cancel" button is located at the bottom of the dialog box.

Below the main image, there is a text area containing the following HTML embed code:

```
<iframe width="320px" height="500px" src="http://medici-demo.ncsa.illinois.edu/embed.html#?id=tag:medici@uiuc.edu,2009:data_XX_c1Q3L3uGAsuMVrsg-DQ" frameborder="0">
```

Below the text area, there are input fields for "Width: 320 px" and "Height: 500 px", and a "Preview" button.

At the bottom of the browser window, there is a section titled "User Specified Information" with a dropdown arrow.

Collection View

The screenshot displays the Active Content Repository interface. At the top, the browser address bar shows the URL: `sead.ncsa.illinois.edu/#collection?uri=tag:cet.ncsa.uiuc.edu,2008/bean/Collection/37ECE16A-1179-4E80-A63C-2A4058886FA9`. The page title is "ACR Demo" by James Myers, dated Thursday, 2012 June 28, with 18 datasets. The interface includes navigation links (Previous, Next), sorting options (Sort by: Date: newest first), and view options (Grid). A grid of 18 dataset thumbnails is displayed, each with a checkbox for selection. A blue circle highlights the "Actions" button above the grid, which is linked to a detailed "Actions" menu on the right. This menu lists various actions: Delete, Download, Add tag(s), Remove tag(s), Add metadata, Add relationships, Change license, Create collection, Add to collection, Remove from collection, Select all on page, and Unselect all. Another blue circle highlights the "Comments" section at the bottom left, showing "0 comments" and a "Write a Comment" field. A third blue circle highlights the "Tags" section at the bottom right, showing "Add tag(s)".

- Select object(s) in collection and apply an action.
- Add comments or tag a Collection

Relationships

The screenshot shows the Medici web application interface. The browser address bar displays `medici-demo.ncsa.illinois.edu/#viewSelected`. The navigation menu includes [Home](#), [Data](#), [Collections](#), [Tags](#), [Map](#), [Upload](#), and [Administration](#). The user is logged in as Luigi Marini, with a [Logout](#) link and a search box.

Selected Datasets

Showing 5 selected datasets

Thumbnail	Dataset Name	Author	Date	Size	Type
	Ex Ante versus Ex Po...	admin	2012 February 17	3.1 MB	Document
	globe_east_2048.tif	Terry McLaren	2012 May 14	7.52 MB	Image
	globe_west_2048.tif	Terry McLaren	2012 May 14	10.39 MB	Image
	Image04.jpg	admin	2012 February 17	418.55 KB	Image

Create Relationship

The 'Create Relationship' tool allows users to link datasets. It features two image thumbnails: `globe_east_2048...` and `globe_west_2048...`. A dropdown menu between them is set to `relates to`. A `Submit` button is located below the thumbnails.

RDF Support (Medici v1)

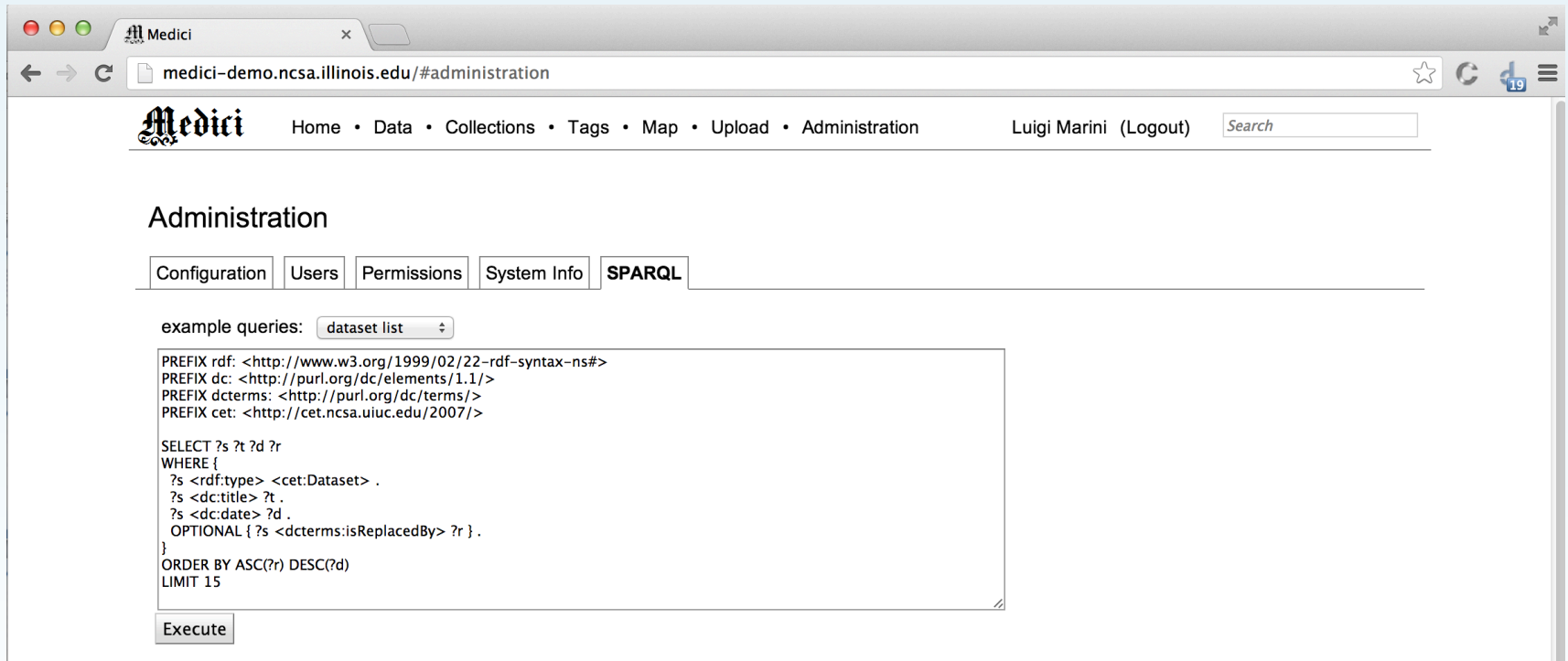
Everything is stored as semantic RDF triples and blobs of binary data

- Open and portable data
- Semantic web



<http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>

SPARQL



The screenshot shows a web browser window with the URL `medici-demo.ncsa.illinois.edu/#administration`. The page title is "Medici" and the navigation menu includes "Home", "Data", "Collections", "Tags", "Map", "Upload", and "Administration". The user is logged in as "Luigi Marini (Logout)".

The "Administration" section is active, with sub-tabs for "Configuration", "Users", "Permissions", "System Info", and "SPARQL".

Under "example queries:", there is a dropdown menu showing "dataset list". Below it is a text area containing the following SPARQL query:

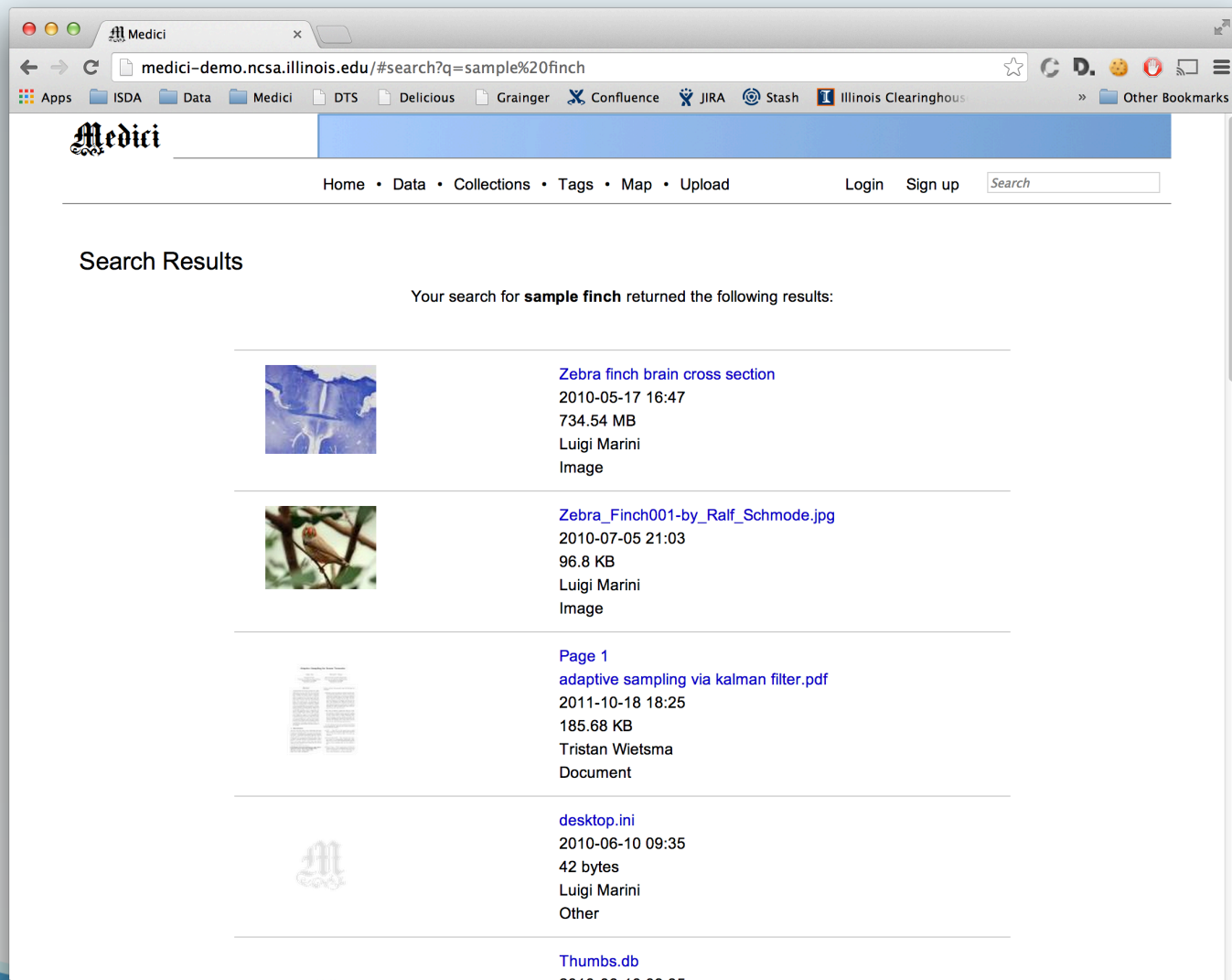
```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX cet: <http://cet.ncsa.uiuc.edu/2007/>

SELECT ?s ?t ?d ?r
WHERE {
  ?s <rdf:type> <cet:Dataset> .
  ?s <dc:title> ?t .
  ?s <dc:date> ?d .
  OPTIONAL { ?s <dcterms:isReplacedBy> ?r } .
}
ORDER BY ASC(?r) DESC(?d)
LIMIT 15
```





Below the query editor is an "Execute" button.

Available as a REST service and through web form (seen above)

Text Based Search



The screenshot shows a web browser window with the URL `medici-demo.ncsa.illinois.edu/#search?q=sample%20finch`. The browser's address bar and tabs are visible at the top. The page header features the Medici logo and navigation links: Home, Data, Collections, Tags, Map, Upload, Login, Sign up, and a search input field. The main content area is titled "Search Results" and displays the message: "Your search for **sample finch** returned the following results:". Below this, four search results are listed, each with a thumbnail image and associated metadata:

Thumbnail	File Name	Date	Size	Author	Type
	Zebra finch brain cross section	2010-05-17 16:47	734.54 MB	Luigi Marini	Image
	Zebra_Finch001-by_Raif_Schmode.jpg	2010-07-05 21:03	96.8 KB	Luigi Marini	Image
	Page 1 adaptive sampling via kalman filter.pdf	2011-10-18 18:25	185.68 KB	Tristan Wietsma	Document
	desktop.ini	2010-06-10 09:35	42 bytes	Luigi Marini	Other

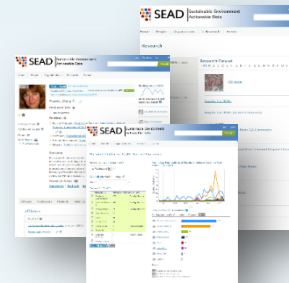
Below the fourth result, a fifth entry is partially visible: [Thumbs.db](#) with a date of 2010-06-10 09:35.

Sustainable Environment | Actionable Data

- NSF DataNet Partnership
- Web services for managing heterogeneous data for sustainability research
- Supporting the long-tail of research data
- “Active and Social Curation”
 - Making curation a natural part of producing and using data
 - Active: on-going, less formal than long term preservation
 - Social: any community member can take part



SEAD Components



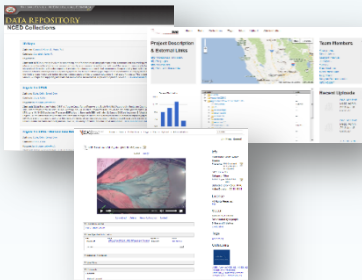
SEAD VIVO:
Browse Through People , Projects,
Publications, Data Citations , and
Organizations, Visualize Networks
and Community Dynamics

SPARQL Queries
HTTP People/Publication links

SPARQL Queries
HTTP Data links

SPARQL Queries
HTTP People/Org links

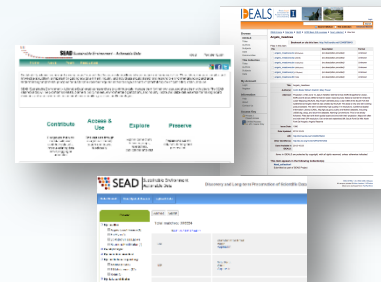
VIVO Service call
HTTP Data/DOI links



Active Content Repository:
Branded Public Access
Active Project Spaces
Individual Data Pages

SPARQL Queries
HTTP Data/DOI links

BAGIT Data/
Metadata
Transfer

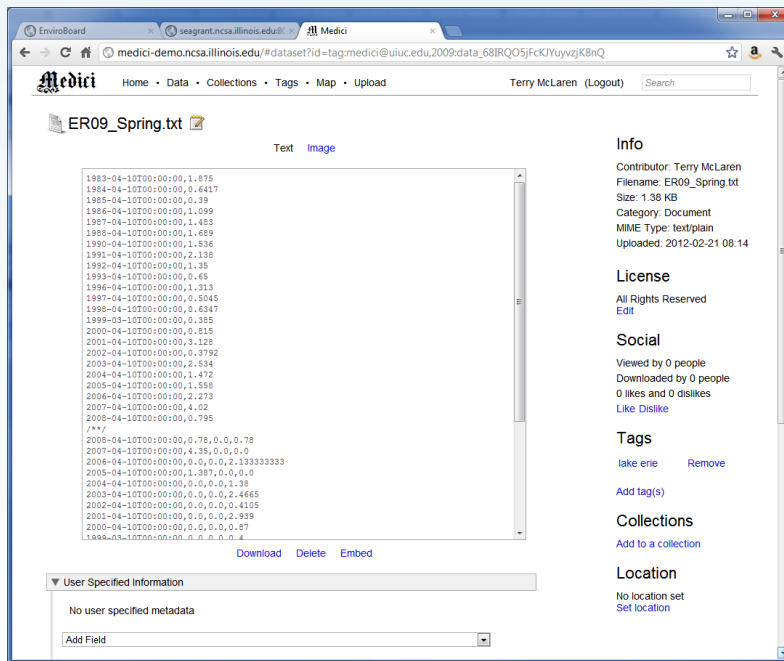


SEAD Virtual Archive:
Policy Driven Curation
Institutional/Cloud/Grid
Storage
Faceted Search for
Reference Data

Versions

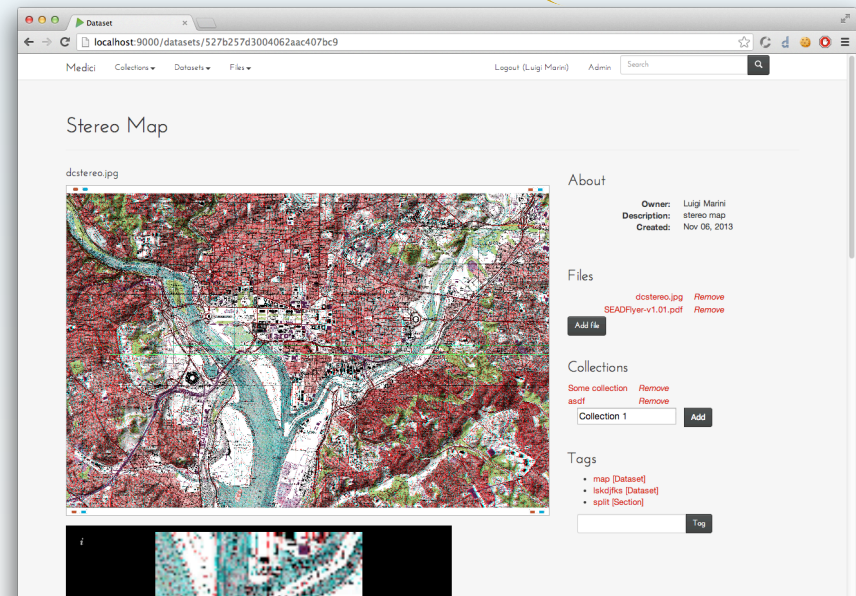
- Scale up and out
- Robust API – easier to write clients
- User experience improvements

1.3



Stable

2.0



In progress

1.0 - What parts were successful?

- Decentralized cloud storage
 - Install your own instance and maintain for your community
- Flexible metadata support
 - No predefined schemas/ontologies
- Framework for adding domain specific features
 - Extractors
 - Dig into the files for information
 - Previewers
 - Visualize information on the web
- Support for discovery of new content
 - Text-based search, social annotation

2.0 Priorities

- Scalable
 - Both horizontally and vertically
- Maintainable
 - Decoupled
 - Higher level languages
 - Build on proven technologies
- Extensible
 - Add extractors and previewers without recompiling
 - Full fledged API
- Usable
 - Modernize the user interface

Version 2.0

- Web scale
 - Across the stack
- Introducing projects
 - Group based access control
- File versioning
 - Keeping generic provenance trail
- Multiple files in a dataset
 - Explicit instead of implicit zip files
- Multimedia search
 - Find similar images, videos, audio
- Recommender system
 - Recommend datasets based on metadata and user activity

Improved UI

The screenshot shows a web browser window with the URL `greatlakesmonitoring.org/medici/datasets`. The page title is "Datasets". The navigation bar includes "Medici", "Collections", "Datasets", "Files", and "API", along with "Login" and "Admin" links. The main content area displays a grid of dataset cards. Each card shows a dataset ID (e.g., ON55Mcnvbin), the description "SeaBird ingested data", and a small icon set (1 document, 1 heart, 0 list, 0 bars). Below the grid is a table with the following data:

Name	Created	Description	
ON55Mcnvbin	May 27, 2014	SeaBird ingested data	No

Support for Themes

The screenshot shows a web browser window displaying the 'Medici Datasets' page. The browser's address bar shows the URL `greatlakesmonitoring.org/medici/datasets`. The page has a red navigation bar with the 'Medici' logo and menu items like 'Collections', 'Datasets', 'Files', and 'API'. The main content area is teal and titled 'Datasets'. It contains a grid of dataset cards, each with a yellow bookmark icon in the top right corner. The cards are arranged in three rows: the first two rows have four cards each, and the third row has two cards. Each card displays a dataset ID in orange text, the description 'SeaBird ingested data', and a set of icons (file, download, list) at the bottom. At the bottom of the page, a table header is visible with columns for 'Name', 'Created', and 'Description'.

Name	Created	Description
------	---------	-------------

Multiple Files per Dataset

Dataset

greatlakesmonitoring.org/medici/datasets/53843eace4b06896ae463ba3

Medici Collections Datasets Files API Logout (Luigi Marini) Admin

HU320221

HU320221.BTL
[Show/edit file community-generated metadata](#)

HU320221.DAT
[Show/edit file community-generated metadata](#)

HU320221.CNV
[Show/edit file community-generated metadata](#)

HU320221.ROS
[Show/edit file community-generated metadata](#)

HU320221.HDR
[Show/edit file community-generated metadata](#)

HU320221.CON
[Show/edit file community-generated metadata](#)

HU320221.BL
[Show/edit file community-generated metadata](#)

Comments Metadata Notes

About

Owner: Anonymous User
Description: SeaBird ingested data
Created: May 27, 2014

Files

HU320221.CNV	Detach	Delete
HU320221.BTL	Detach	Delete
HU320221.DAT	Detach	Delete
HU320221.ROS	Detach	Delete
HU320221.CON	Detach	Delete
HU320221.BL	Detach	Delete
HU320221.HDR	Detach	Delete

Add file

ER090211.BL
Add existing file

Collections

SeaBird Remove

LEC Add

Still Flexible Metadata (but not as native RDF)

HU320221.BL

Show/edit file community-generated metadata

Comments

Metadata

Notes

Auto-generated metadata:

- **HU32:**
 - **sensor:** <http://greatlakesmonitoring.org/medici/api/geostreams/sensors/795>
 - **dashboard:** <http://greatlakesmonitoring.org/#detail/location/HU32/>
 - **stream:** <http://greatlakesmonitoring.org/medici/api/geostreams/streams/636>
 - **datapoints:** http://greatlakesmonitoring.org/medici/api/geostreams/datapoints?stream_id=636

Versus

- Execution and dissemination of customizable content-based file comparison methods
- Generalized content-based comparison and retrieval of files
- Ability to plugin new methods and reuse existing ones
- Consists of
 - Java API + Engine
 - Library of methods
 - Web Service
 - Variety of web and desktop clients

Content-Based Comparison

- Goal: Comparing digital data
- Given two or more digital objects establish their proximity
- Arbitrary?
- Not really, comparing two files, videos, documents, etc. has many applications, for example:

Some Applications

- Information loss
 - Information loss when applying file format conversion
 - Polyglot
- Content-based retrieval
 - Given a multimedia file (image) find the closest ones in a large collection
- Find duplicates
 - Across formats

Census Information Retrieval

Query:

Illinois

Collection:

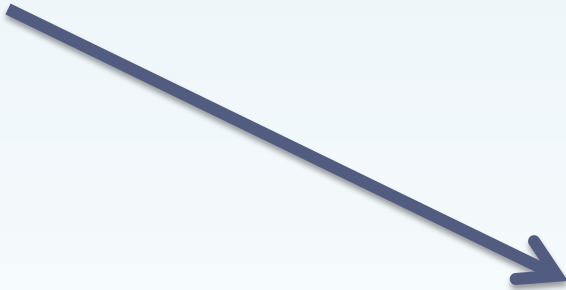
Missouri

Penn

Illinois

None

Kansas



History



- Funding by National Archives and Records Administration (NARA)
- Research and development started in 2010
- Originally focused on pairwise comparison
- Adding support for the creation of indexes over past two years
- Current version is 0.6
 - Usable but still in flux
 - Particular important with APIs
 - Lots of exploratory work over the years

Two Main Components

Core

- A set of Java interfaces
- Multithreaded Execution Engine
- Registry to register and query for methods

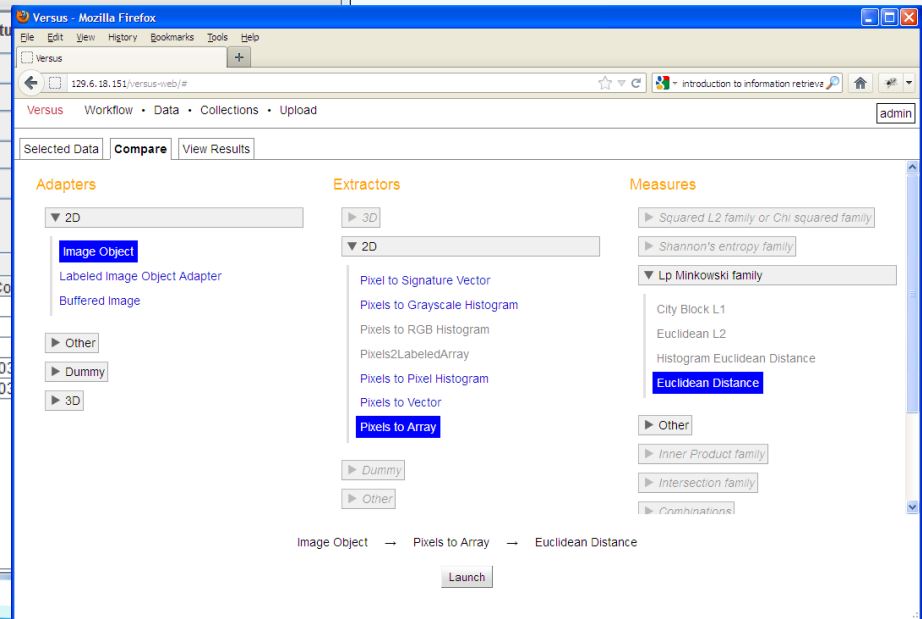
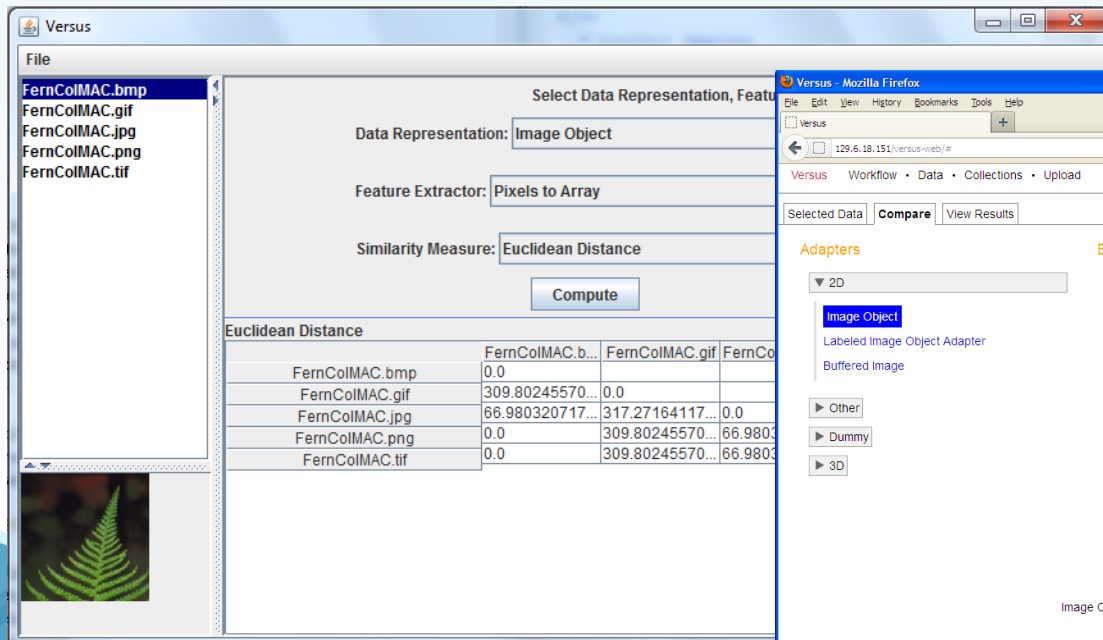
Web Service

- HTTP API wrapping Core
- Master/slave architecture

Several Clients

- Command Line Interface
- Web Application
- Desktop App
- Medici 2

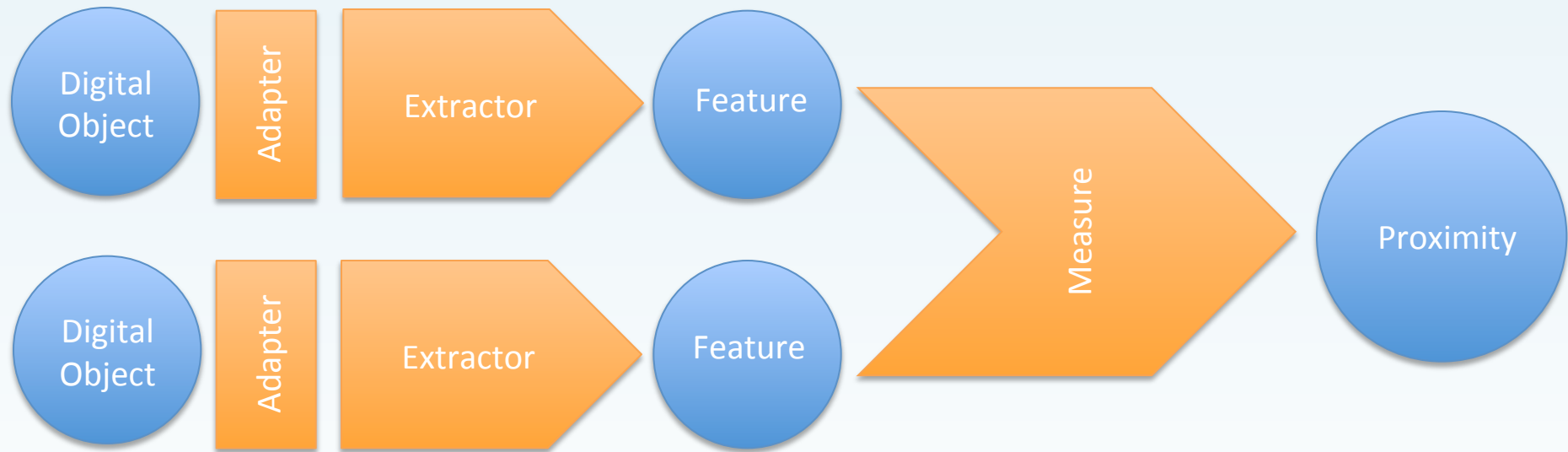
```
Terminal — bash — 115x11
faye:bin lmarini$ ./versus-cli.sh file1.tiff file2.tiff edu.illinois.ncsa.versus.adapter.impl.BytesAdapter edu.illinois.ncsa.versus.extract.impl.MD5Extractor edu.illinois.ncsa.versus.measure.impl.MD5DistanceMeasure
DEBUG [main] (ComputeThread.java:67) - Selected adapter is edu.illinois.ncsa.versus.adapter.impl.BytesAdapter
DEBUG [main] (ComputeThread.java:70) - Selected extractor is edu.illinois.ncsa.versus.extract.impl.MD5Extractor
DEBUG [main] (ComputeThread.java:72) - Selected measure is edu.illinois.ncsa.versus.measure.impl.MD5DistanceMeasure
DEBUG [main] (ExecutionEngine.java:58) - Job submitted
DEBUG [pool-1-thread-1] (ComputeThread.java:130) - Compared file1.tiff with file2.tiff = 0.0
Comparison's result: 0.0
faye:bin lmarini$
```



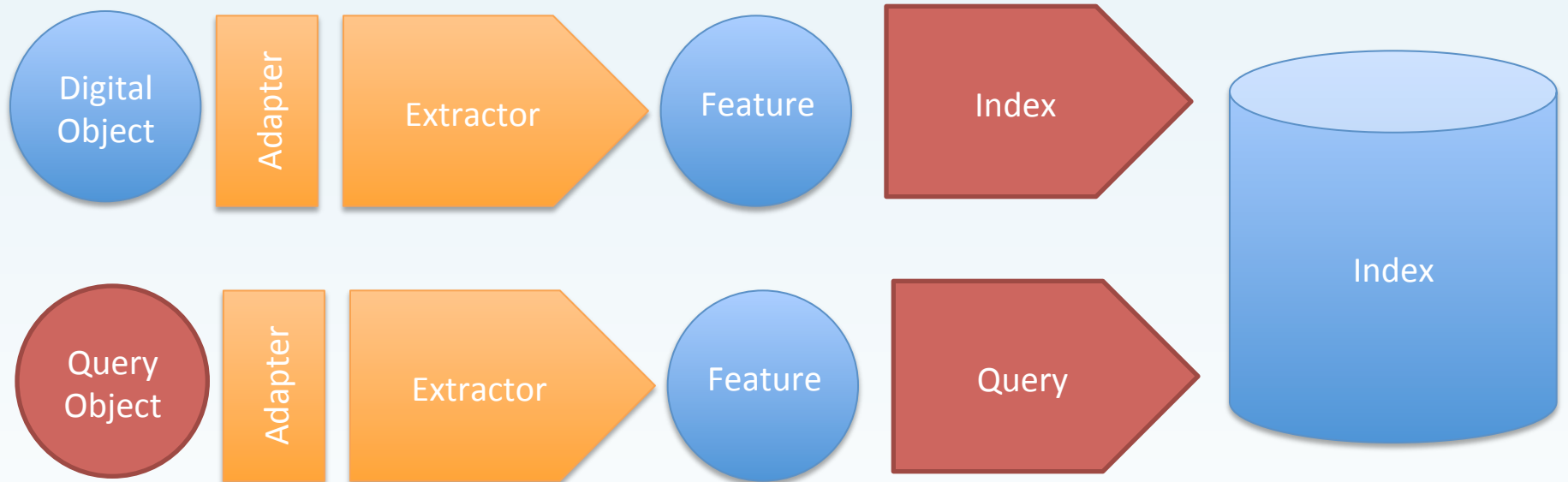
Why would one use Versus instead of writing specific implementations as need be?

- **Reuse** existing methods
- **Share** methods with community
- **Organize** code in clear components
- **Leverage** execution environment and service infrastructure

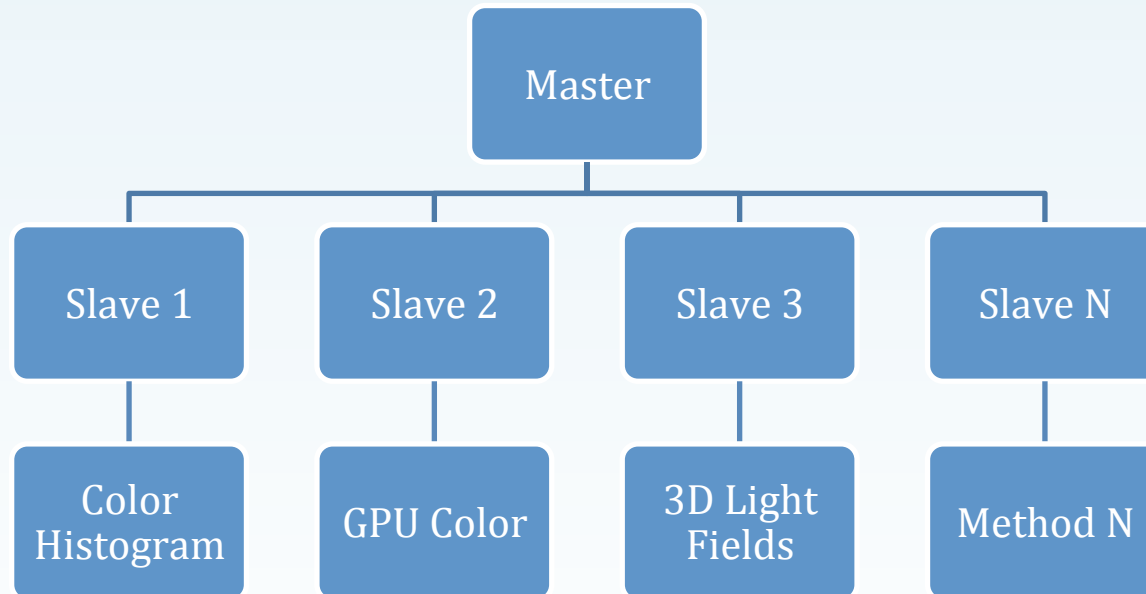
Pairwise Comparison API



Indexing API



Master/Slave



Library of Adapters

Name	Package	Description
Mesh	3D	Load 3D files content into a mesh made up of vertices and polygons connecting those vertices.
Audio	Audio	Encapsulation of audio files.
Bytes	Core	Simplest possible representation of data.
PDF	Doc2Learn	Encapsulation of the Doc2Learn PDF document.
Buffered Image	Image	Standard Java representation of image data.
Image Object	Image	Encapsulation of the Im2Learn Image Object.
SIFT GPU	GPU	Encapsulation of image data for SIFT Gpu specific processing.

Library of Descriptors

Name	Package	Description
Double Array	Core	A single dimensional array containing double values.
MD5 Digest	Core	A data integrity structure generated from the raw data.
Three Dimensional Double Array	Core	A three-dimensional array containing double values.
Vector	Core	A list of generic elements, allows greater storage flexibility.
Label Histogram	Doc2Learn	A histogram of labels obtained through Doc2Learn.
Keypoint	Image	Generic container for invariant feature detectors.
Pixel	Image	Generic type for various image package descriptors.
Color Layout	Image	A two dimensional grid of sub-images over the input image.
Grayscale Histogram	Image	A one-dimensional grayscale image histogram.
RGB Histogram	Image	A three-dimensional RGB color histogram.
Pixel Histogram	Image	A multidimensional histogram for a pixel's intensity and position.
MOPS Features	Fiji	Invariant feature type used for image stitching.
SIFT Features	Fiji	Popular invariant feature type used for image comparison and object matching.
SIFT Gpu	Gpu	Same as SIFT but implemented through Gpu libraries.
Harris Corners	OpenCV	Well-known corner detector used for image inference, tracking, and recognition.
Hough Circles	OpenCV	Circles detected in an image with the Hough

Library of Extractors

Name	Package	Description
Light Field	3D	Surface is represented by silhouettes taken from 3 canonical positions capturing the surface shape minus any concavities (i.e. the convex hull).
Statistics	3D	Ignores the surface and focuses on the vertices of a 3D object returning their mean and standard deviation. Simple, but fast to compute.
Surface Area	3D	The sum of the area occupied by the polygons making up a surface. Considers surface and is still fast to compute.
Audio	Audio	Sampling of audio from existing file for histogram usage and comparison.
MD5	Core	Creation of the MD5 hash from data.
Image Histogram	Doc2Learn	Generates a non-standard color histogram
Line Graphics Histogram	Doc2Learn	Generates a histogram to compare vector graphics found in documents.
Text Histogram	Doc2Learn	Generates a label histogram based on word frequency.
Array Feature	Image	Generates the three-dimensional double array; a generic image container.
Color Average Vector Feature	Image	Generates an average RGB color over 9 regions taken from the image.
Grayscale Histogram	Image	Generates the histogram for grayscale images. Useful for image comparison.
Pixel Histogram	Image	Generates the multidimensional histogram for feature matching.
RGB Histogram	Image	Generates the histogram for color images. Useful for image comparison.
Signature Vector	Image	Feature vector (for an image) containing colorspace

Library of Measures

Name	Package	Description
Chessboard Distance	Core	Also known as Chebyshev; the greatest difference along any coordinate dimension (between two vectors)
Dynamic Time Warping	Core	Similarity metric between two (possibly) varying sequences over time.
Euclidean Distance	Core	Distance between two n-dimensional points in Euclidean space.
Manhattan Distance	Core	Absolute difference of coordinates of points, distance between two points measured along right angled axes.
MD5 Hash	Core	Binary measure; either equal or not.
Bhattacharyya Distance	Image	Measures the overlap between two probability distributions.
Neyman's χ^2	Image	Tests the <i>goodness of fit</i> between two distributions. Variant of the standard χ^2 test.
Czekanowski Distance	Image	Sum of the absolute value of the difference of two distributions divided by the sum of the two distributions.
Histogram Euclidean Distance	Doc2Learn / Image	Bin-by-bin comparison using the standard Euclidean distance. Well known and widely used.
Histogram Intersection	Doc2Learn / Image	Sum of the absolute value of the difference of two distributions, scaled by one-half. Well known and widely used.
KL Divergence	Image	Non-symmetric measure of the difference between two probability distributions. Well known measure of entropy.
Jeffrey Divergence	Image	Symmetric measure of the difference between two probability distributions.
Motyka Distance	Image	Sum of the maximum of two distributions divided by the sum of the two distributions.
Normalized Cross Correlation	Image	Similar to sum of squared differences; invariant to the magnitude of two points.
Ruzicka Similarity	Image	Sum of the minimum of two distributions divided

NARA / XSEDE

CENSUS



The Product of Digitization

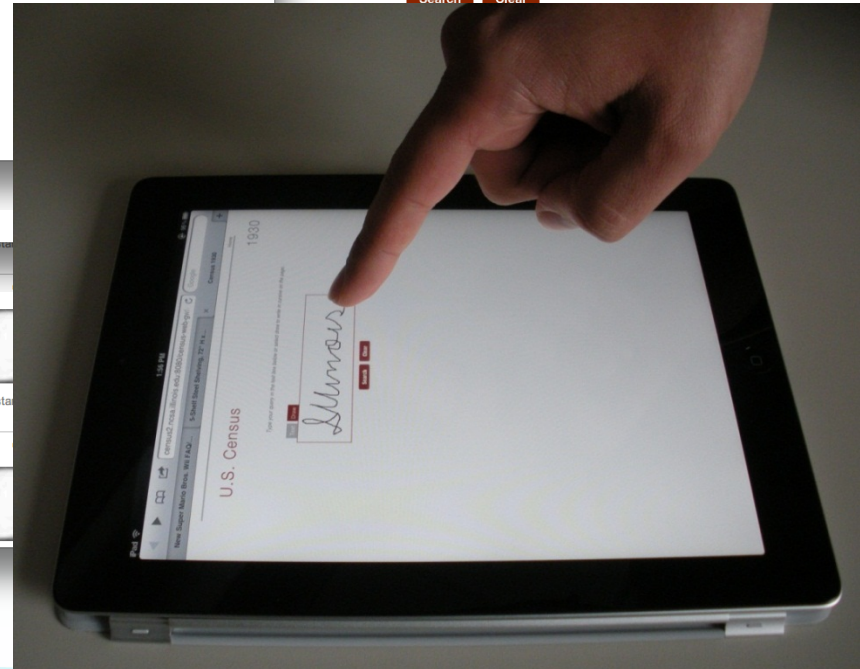
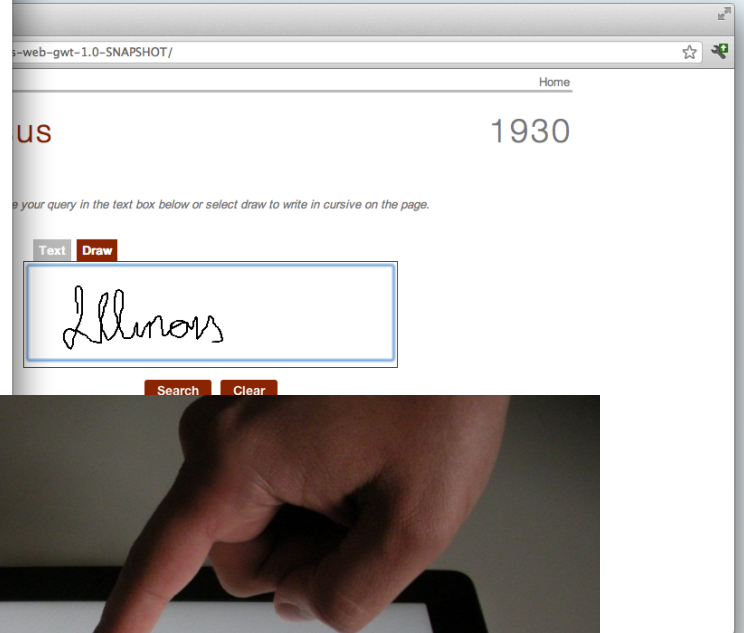
Form 100
 DEPARTMENT OF COMMERCE—BUREAU OF THE CENSUS
 FIFTEENTH CENSUS OF THE UNITED STATES: 1930
 POPULATION SCHEDULE

State Illinois Incorporated place Madison village Enumeration District No. 10-30 Sheet No. 993
 County Champaign Ward of city _____ Block No. _____ Superintendent's District No. 19 4A 42
 Township or other district of county Madison township Unincorporated place _____ Enumeration by name on April 10, 1930, James E. Walker

PLACE OF BIRTH			RELATION to head of family or to person with whom dwelling is kept	SEX AND AGE		MARITAL STATUS	EDUCATION	PLACE OF BIRTH			MOTHER TONGUE OR NATIVE LANGUAGE OF FOREIGN BORN	CITIZENSHIP, ETC.	OCCUPATION AND INDUSTRY		MILITARY SERVICE				
Male	Female	Total		Male	Female			Foreign born	Native born	Foreign born			Native born	Foreign born		Native born	Occupation	Industry	Year of entry into service
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93
94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113
124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183
194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213
234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253
274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293
314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333
354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373
394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413
434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453
474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493
514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533
554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573
594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613
634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653
674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693
714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733
754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773
794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813
834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853
874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893
914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933
954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973
994	995	996	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013

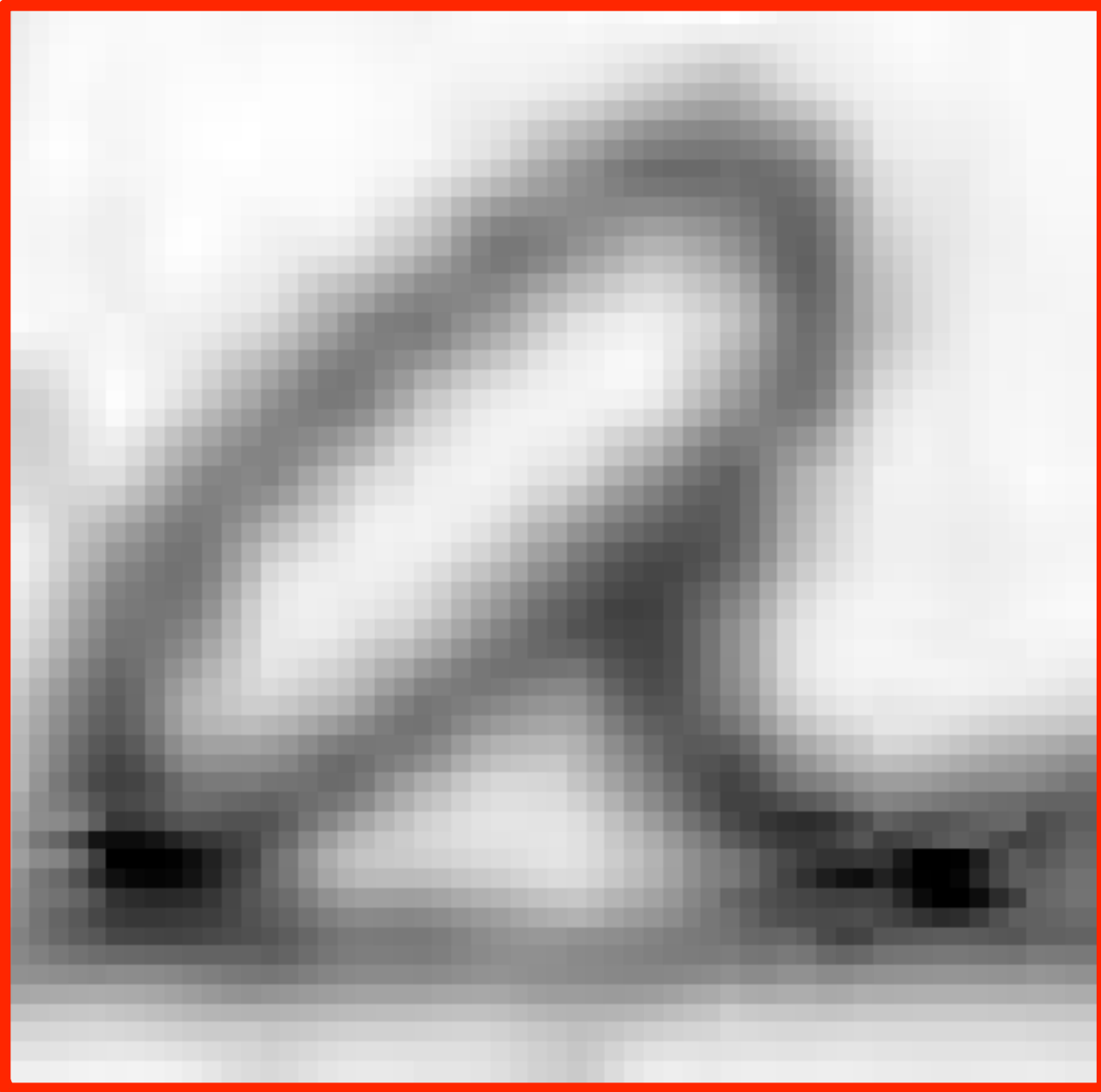
APPROPRIATIONS TO BE MADE IN COLUMN 18 (MILITARY SERVICE)
 INSTRUCTIONS TO ENUMERATORS
 STATE OF ILLINOIS
 STATE OF ILLINOIS
 STATE OF ILLINOIS

User Queries



A Computer Vision Problem

State



A Computer Vision Problem

A 10x10 grid of numerical values, likely representing a feature map or a similarity matrix. The values are arranged in a grid with a red border. A zoomed-in inset of the top-right corner is shown above the main grid.

1.0	1.0	1.0	1.0	1.0	0.3	0.2	0.2	0.3	1.0
1.0	1.0	1.0	1.0	0.3	0.2	0.2	0.3	1.0	1.0
1.0	1.0	1.0	0.3	0.2	0.2	0.3	1.0	1.0	1.0
1.0	1.0	0.3	0.2	0.2	0.3	0.3	1.0	1.0	1.0
1.0	1.0	0.3	0.2	0.3	0.3	1.0	1.0	1.0	1.0
1.0	1.0	0.3	0.2	0.3	1.0	1.0	1.0	1.0	1.0
1.0	0.3	0.2	0.3	0.3	1.0	1.0	1.0	1.0	1.0
1.0	0.3	0.2	0.3	1.0	1.0	1.0	1.0	1.0	1.0
1.0	0.3	0.2	0.3	1.0	1.0	1.0	1.0	1.0	1.0
0.3	0.2	0.2	0.3	1.0	1.0	1.0	1.0	1.0	1.0

Form Segmentation

DEPARTMENT OF COMMERCE - BUREAU OF THE CENSUS
 FIFTEENTH CENSUS OF THE UNITED STATES: 1930
 POPULATION SCHEDULE

Enumeration District No. 229
 Supervisor's District No. A

City of Chicago
 Ward of city

Place of birth of each person whose place of abode on April 1, 1930, was in the United States

PLACE OF BIRTH	NAME	RELATION	HOME DATE	MARITAL STATUS	SEX	AGE	PLACE OF BIRTH		MOTHER'S LANGUAGE	MOTHER'S BIRTH	CITIZENSHIP	OCCUPATION AND INDUSTRY		LAST VISIT	REMARKS
							FATHER	MOTHER				OCCUPATION	INDUSTRY		
Illinois	Thomas J. Jones	Head	1878	M	M	52	Illinois	Illinois	English	Illinois	U.S.	Farmer	Farmer	1928	
Illinois	Thomas J. Jones	Wife	1880	F	F	50	Illinois	Illinois	English	Illinois	U.S.	Farmer	Farmer	1928	
Illinois	Thomas J. Jones	Child	1905	M	M	25	Illinois	Illinois	English	Illinois	U.S.	Farmer	Farmer	1928	
Illinois	Thomas J. Jones	Child	1908	F	F	22	Illinois	Illinois	English	Illinois	U.S.	Farmer	Farmer	1928	
Illinois	Thomas J. Jones	Child	1910	M	M	20	Illinois	Illinois	English	Illinois	U.S.	Farmer	Farmer	1928	
Illinois	Thomas J. Jones	Child	1912	F	F	18	Illinois	Illinois	English	Illinois	U.S.	Farmer	Farmer	1928	
Illinois	Thomas J. Jones	Child	1915	M	M	15	Illinois	Illinois	English	Illinois	U.S.	Farmer	Farmer	1928	
Illinois	Thomas J. Jones	Child	1918	F	F	12	Illinois	Illinois	English	Illinois	U.S.	Farmer	Farmer	1928	
Illinois	Thomas J. Jones	Child	1920	M	M	10	Illinois	Illinois	English	Illinois	U.S.	Farmer	Farmer	1928	
Illinois	Thomas J. Jones	Child	1922	F	F	8	Illinois	Illinois	English	Illinois	U.S.	Farmer	Farmer	1928	
Illinois	Thomas J. Jones	Child	1925	M	M	5	Illinois	Illinois	English	Illinois	U.S.	Farmer	Farmer	1928	
Illinois	Thomas J. Jones	Child	1928	F	F	2	Illinois	Illinois	English	Illinois	U.S.	Farmer	Farmer	1928	
Illinois	Thomas J. Jones	Child	1930	M	M	0	Illinois	Illinois	English	Illinois	U.S.	Farmer	Farmer	1928	

Form Segmentation

Birmingham Francis

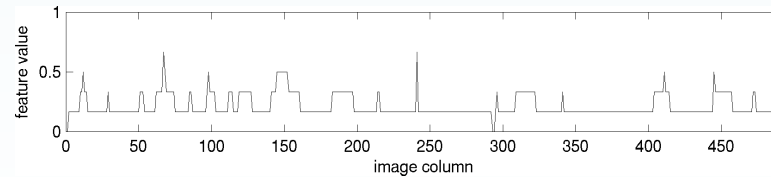
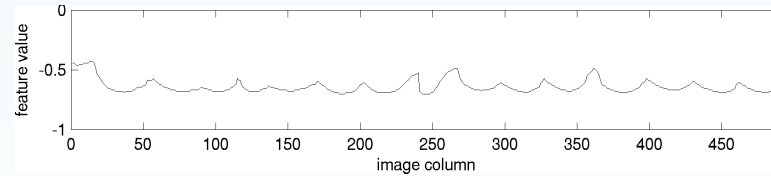
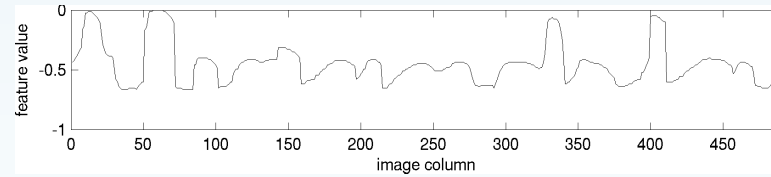
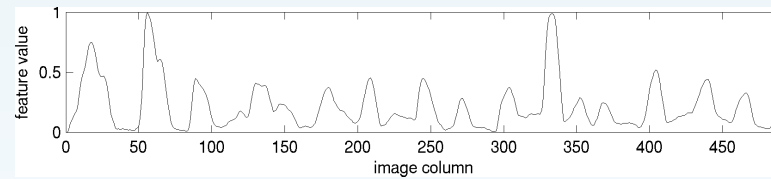
A Les

Thomas Louis

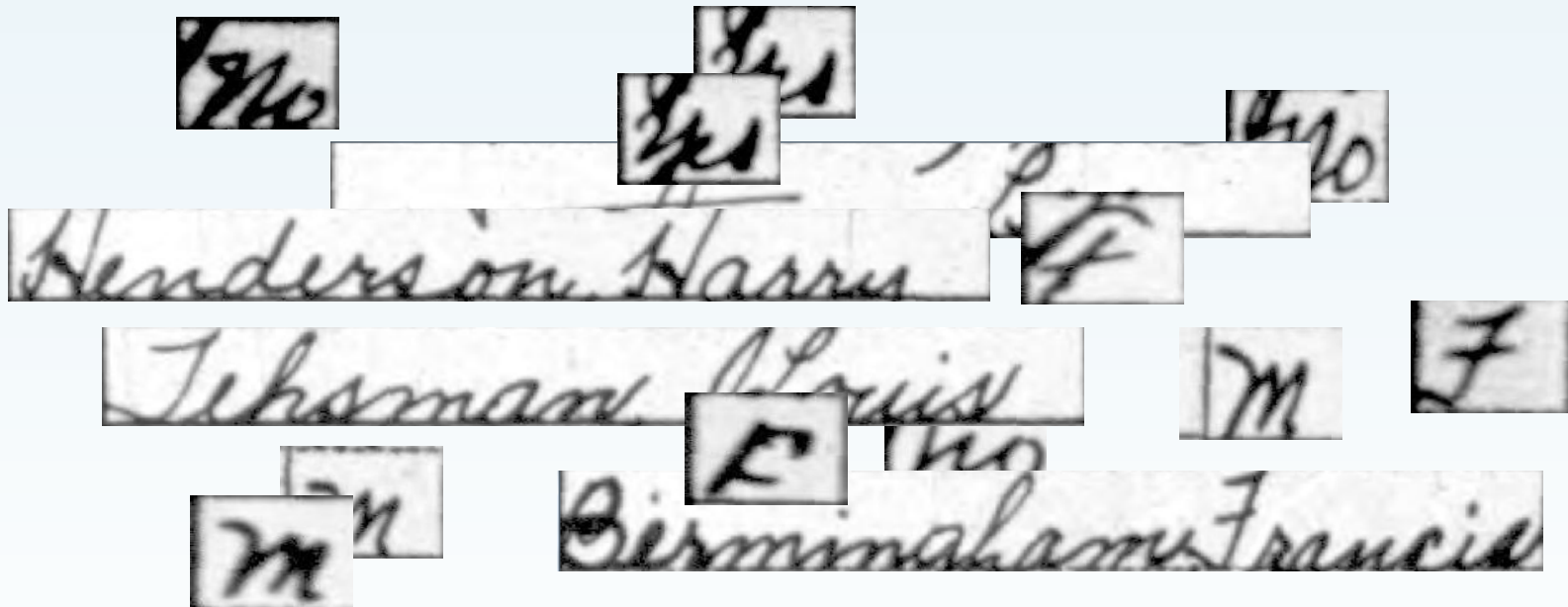
Henderson Harris

Word Spotting

Dynamic Time Warping [T. Rath, 2003]



Searching in Image Databases

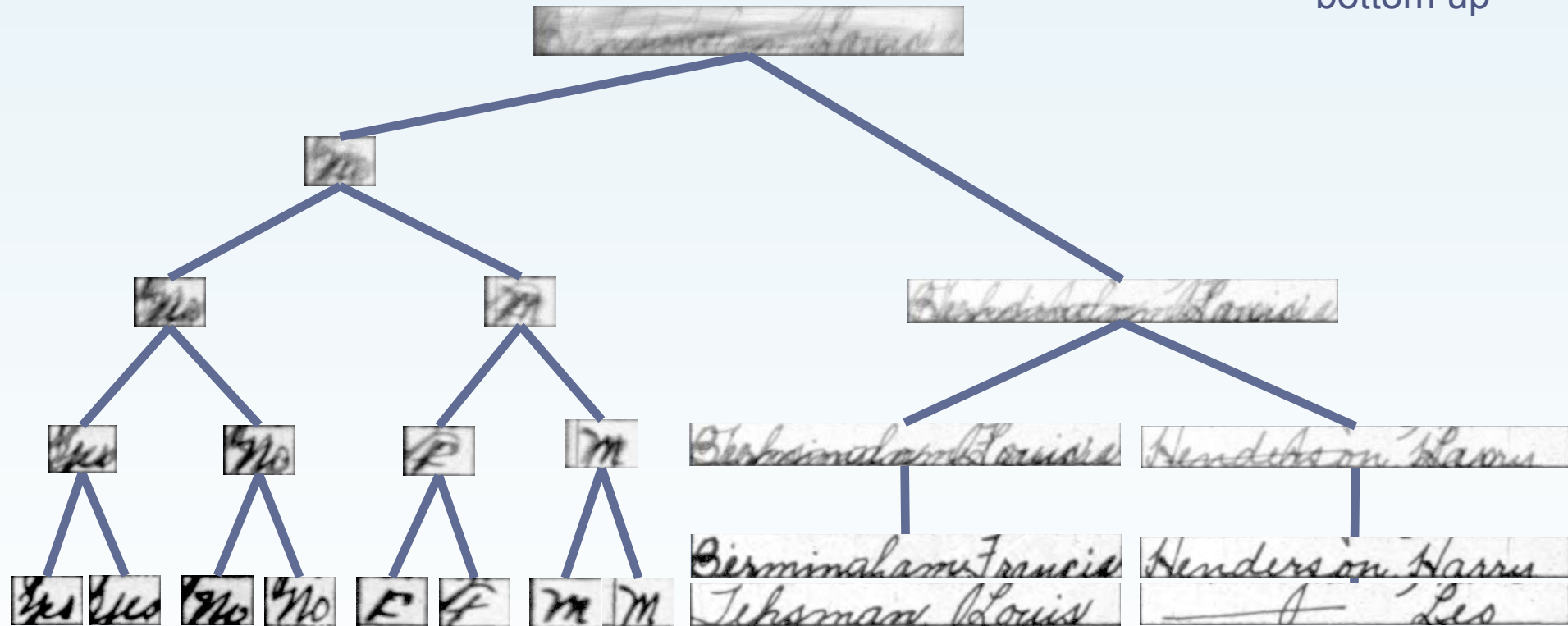


Linear Search

No	Yes	Yes	A	Les	
No	Henderson Harris			F	
Jehsman Louis			M	F	F
No	M	M	Birmingham Francis		

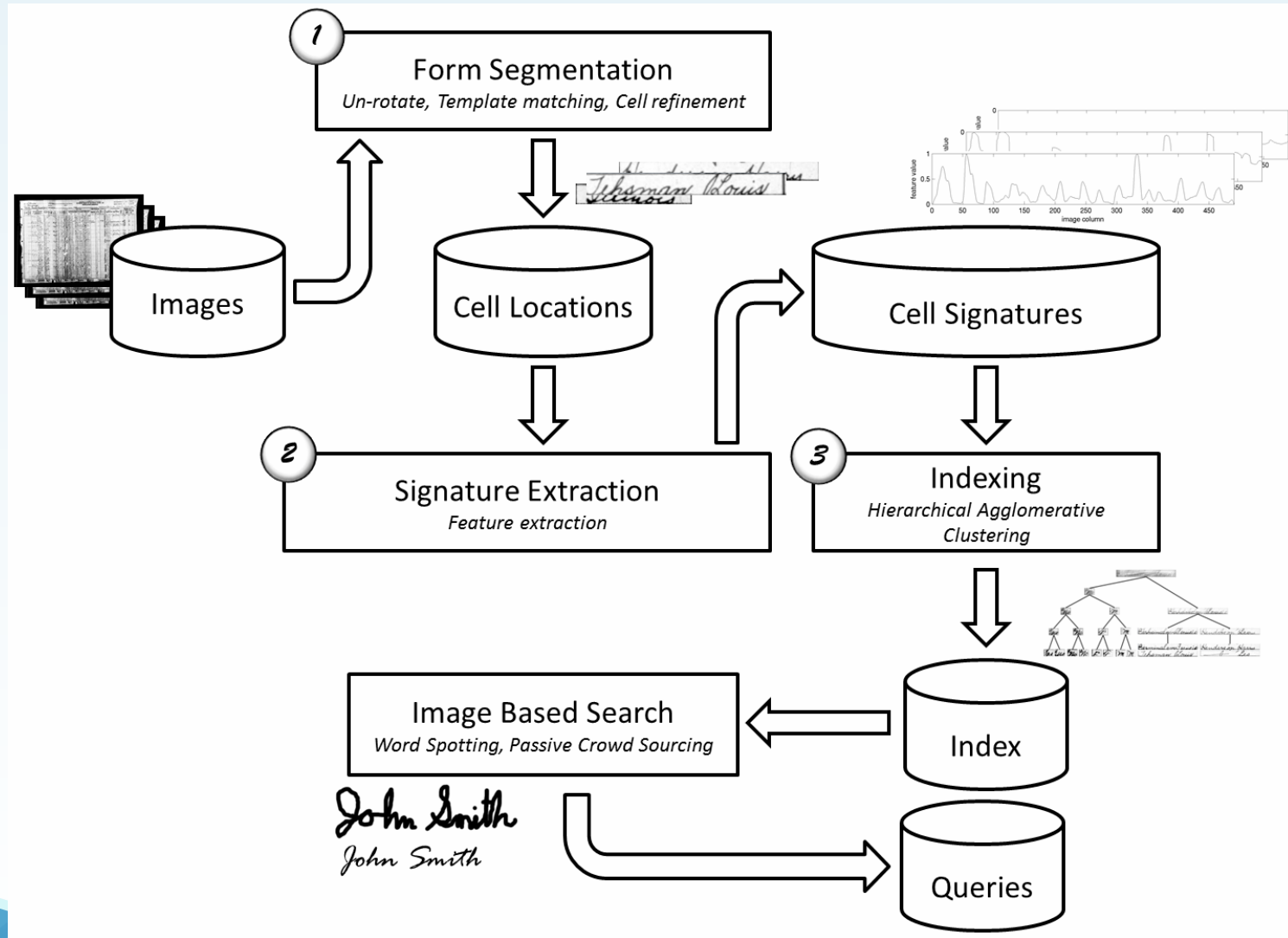
Hierarchical Agglomerative Clustering

“bottom up”

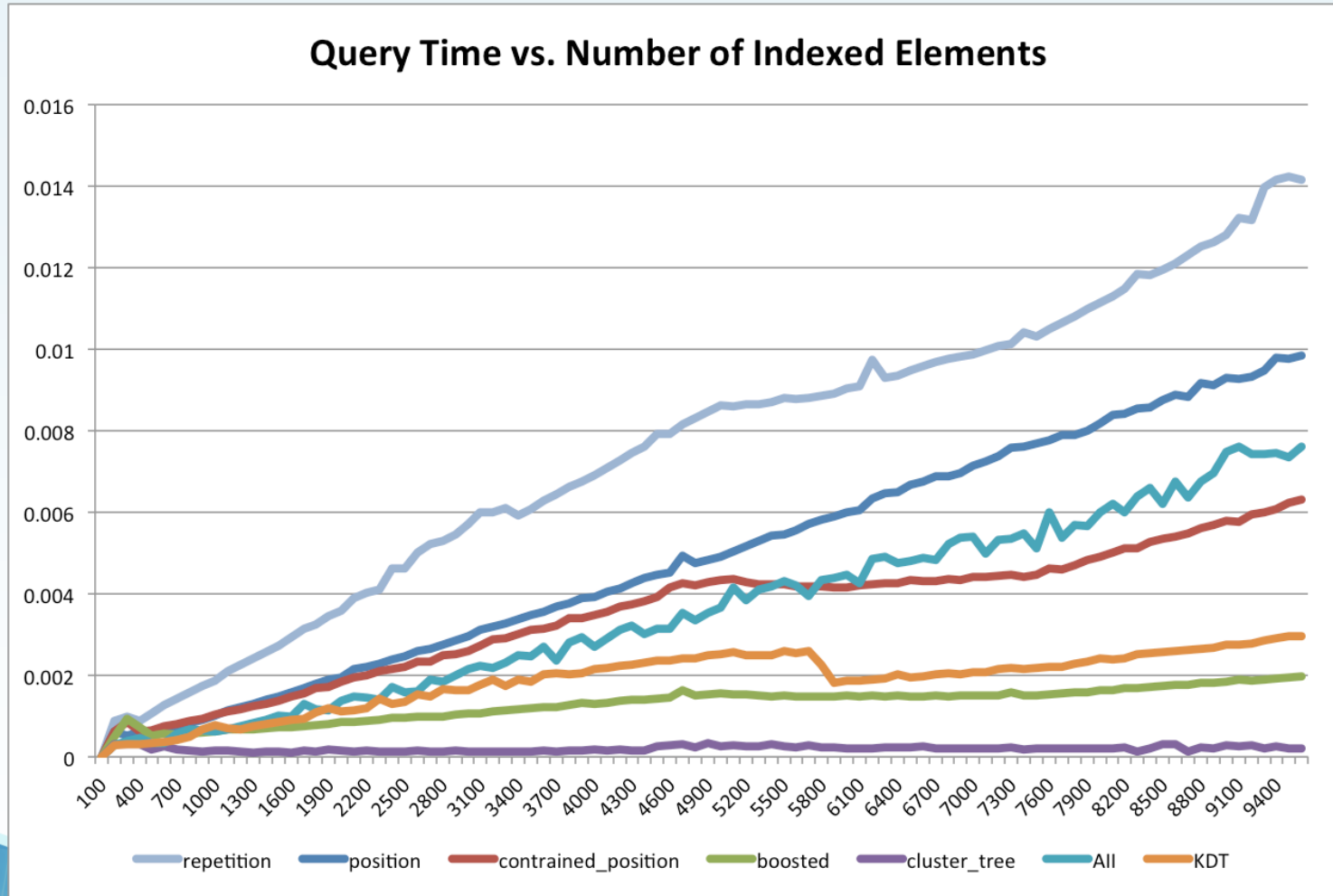


[D. Defays, 1977]

Content-Based Image Retrieval



Index Build Time



XSEDE

LARGE SCALE VIDEO ANALYTICS



Large-Scale Video Analytics (XSEDE)

- Aims to facilitate humanities research on moving images at large scale
- Dedicated nodes on Gordon at SDSC
- Identification of shots in videos
- Multimedia retrieval of similar shots

Large-Scale Video Analytics (XSEDE)

Medici Datasets Files Search Login

localhost:9000/datasets/5139ff634566900401ce0946

- Women artists
- Video Aktiv
- Politics
- Culture
- Eastern Europe

Text input Tag

Shots

#	Thumbnail	Start Time	Find
1		0:48	Similar
2		1:59	Similar
3		2:43	Similar
4		3:10	Similar


Large-Scale Video Analytics (XSEDE)

Search Results


localhost:9000/searchbyfeature?section_id=513a16424566637619e44aec

Medici Datasets Files Search Login


Query



Distance: 14.583333333333348



Distance: 14.583333333333348



Distance: 15.821347736440428

localhost:9000/datasets_by_section/513a0b0b4566637619e4429a

Large-Scale Video Analytics (XSEDE)

Search Results



localhost:9000/searchbyfeature?section_id=513a204345667c4176ad8704

Medici Datasets Files Search Login


Search

EDGE_HISTOGRAM

Query



Distance: 0.0

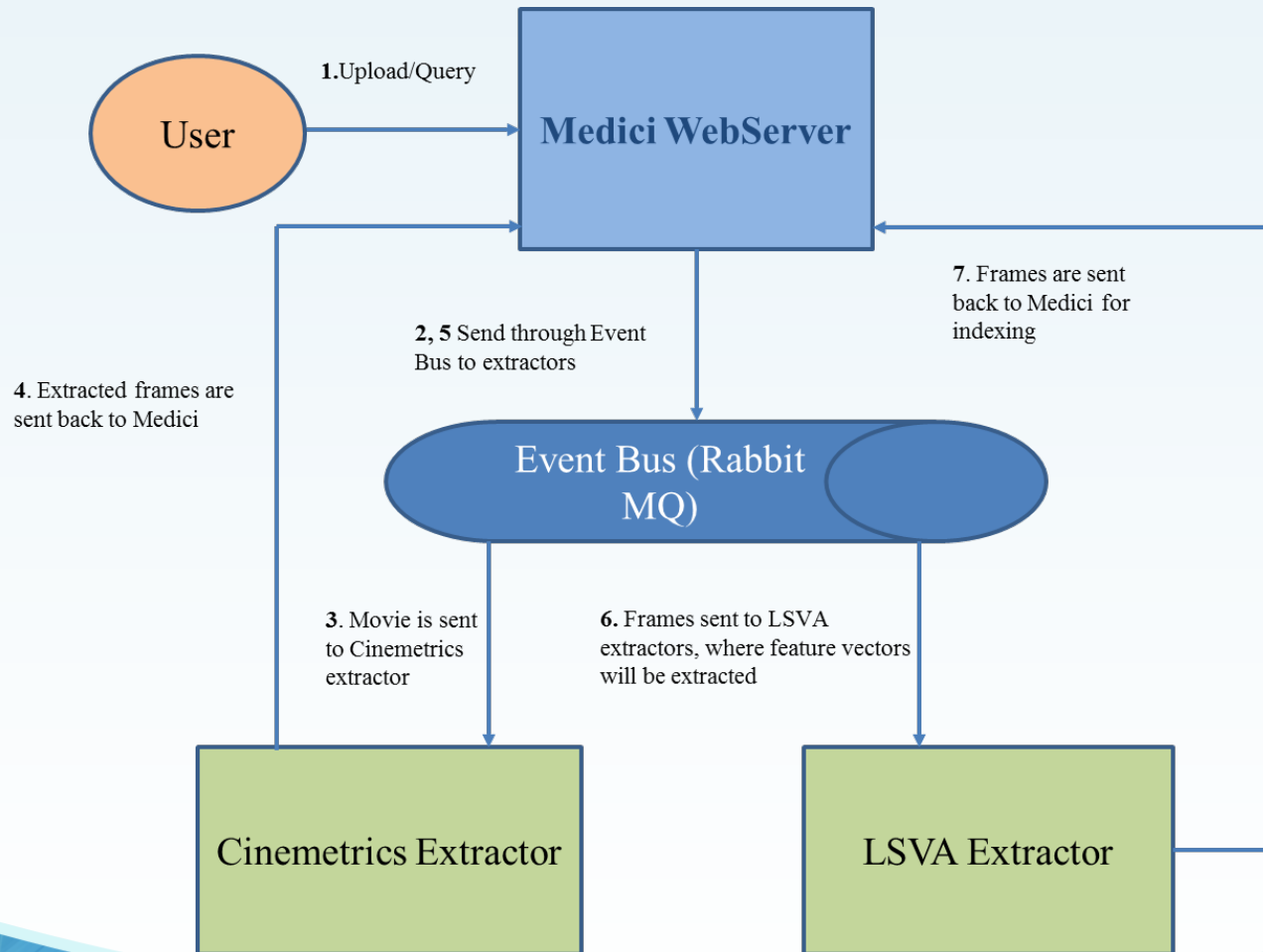


Distance: 0.65056111111111109

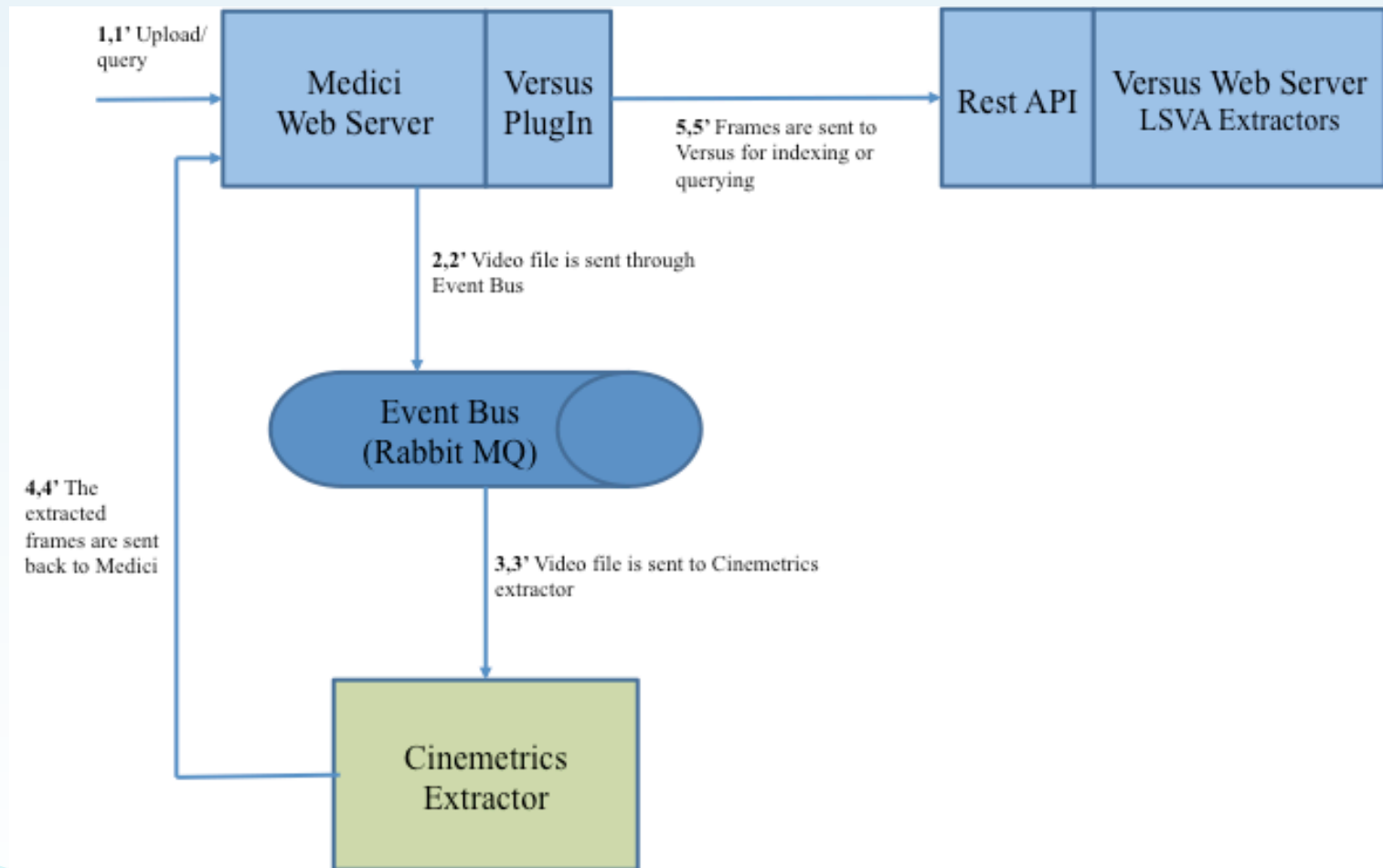
LSVA Versus Methods

Extractor	Descriptor	Measure
ColorLayoutCBExtractor	ColorLayoutCBDescriptor	ColorLayoutL2DistanceMeasure
ColorLayoutCExtractor	ColorLayoutCRDescriptor	ColorLayoutL2DistanceMeasure
ColorLayoutDCEExtractor	ColorLayoutDCTDescriptor	ColorLayoutL2DistanceMeasure
ColorLayoutExtractor	ColorLayout1DDescriptor	ColorLayoutDistanceMeasure
ColorLayoutYExtractor	ColorLayoutYDescriptor	ColorLayoutL2DistanceMeasure
EdgeHistogramExtractor	EdgeHistogramDescriptor	EdgeHistogramDistanceMeasure
GaborTextureExtractor	GaborTextureDescriptor	GaborEnergiesMeasure
GrayHistogramExtractor	GrayHistogramDescriptor	GrayHistogramDistanceMeasure
HSLHistogram1DExtractor	HSLColorHistogram1DDescriptor	HSLColorHistogram1DDistanceMeasure
HSLHistogram3DExtractor	HSLColorHistogram3DDescriptor	HSLColorHistogram3DMeasure
HSVHistogram1DExtractor	HSVColorHistogram1DDescriptor	HSVColorHistogram1DDistanceMeasure
HSVHistogram3DExtractor	HSVColorHistogram3DDescriptor	HSVColorHistogram3DDistanceMeasure
RoughColorLayoutExtractor	RoughColorLayoutDescriptor	RoughColorLayoutDistanceMeasure

Current Setup



Future Setup with Versus



Sea Grant / EPA - Great Lakes Monitoring

- Nutrients
 - Total Phosphorus
 - Chlorophyll a
 - Silica
 - Nitrogen Nitrite



	A	B	C	D	E	F	G	H	I	J	K	L
1	YEAR	MONTH	LAKE	STATION	SAMPLING_DATE	SMPL_DEPTH, m	DEPTH_CODE	QC_ID	SAMPLE_ID	Chlorophyll-a, ug/l	TP, ug/l	NNN, mg/l
2	1983	Apr	Huron	HU06	1983/04/24 06:20 EDT	0.91	EP	RFS	83GB01S01	1.5	3.5	0.27
3	1983	Apr	Huron	HU06	1983/04/24 06:20 EDT	23.77	EP	RFS	83GB01S02	2.5	4.5	0.27
4	1983	Apr	Huron	HU06	1983/04/24 06:20 EDT	37.8	EP	RFS	83GB01S03	2.3	4.5	0.27
5	1983	Apr	Huron	HU06	1983/04/24 06:20 EDT	45.72	EP	RFS	83GB01S04	1.9	3.2	0.27
6	1983	Apr	Huron	HU09	1983/04/24 03:55 EDT	29.26	EP	RFS	83GB06S02	1.9	4.5	0.26
7	1983	Apr	Huron	HU09	1983/04/24 03:55 EDT	0.91	EP	RFS	83GB06S01	2.5	3.5	0.26
8	1983	Apr	Huron	HU09	1983/04/24 03:55 EDT	44.2	EP	RFS	83GB06S03	1.9	4.5	0.26
9	1983	Apr	Huron	HU09	1983/04/24 03:55 EDT	53.34	EP	RFS	83GB06S04	2	4.5	0.26
10	1983	Apr	Huron	HU12	1983/04/24 00:45 EDT	0.91	EP	RFS	83GB11S01	2	3.6	0.31
11	1983	Apr	Huron	HU12	1983/04/24 00:45 EDT	42.06	EP	RFS	83GB11S02	2.3	3.5	0.32
12	1983	Apr	Huron	HU12	1983/04/24 00:45 EDT	73.46	EP	RFS	83GB11S03	1.9	5.5	0.31
13	1983	Apr	Huron	HU12	1983/04/24 00:45 EDT	82.6	EP	RFS	83GB11S04	NRR	3	0.32
14	1983	Apr	Huron	HU15M	1983/04/23 22:20 EDT	10.06	EP	RFS	83GB16S03	1.9	4	0.31
15	1983	Apr	Huron	HU15M	1983/04/23 22:20 EDT	20.12	EP	RFS	83GB16S04	2	4	0.31
16	1983	Apr	Huron	HU15M	1983/04/23 22:20 EDT	29.87	EP	RFS	83GB16S05	2.2	4	0.31
17	1983	Apr	Huron	HU15M	1983/04/23 22:20 EDT	49.99	EP	RFS	83GB16S07	2.2	4	0.31

The Intensively Managed Landscapes- Critical Zone Observatory (IML-CZO)

Home x

imlcz02.ncsa.illinois.edu/geodashboard

Intensive Managed Landscapes Critical Zone Observatory

Explore the Data Advanced Search About the Data

All

Geo Layers

- USRB Flow Lines
- USRB Boundary
- USRB DEM
- IL Moraines
- IL County Boundaries
- USRB Terraces
- ClearCreek Streams
- ClearCreek Watershed
- ClearCreek Catchment

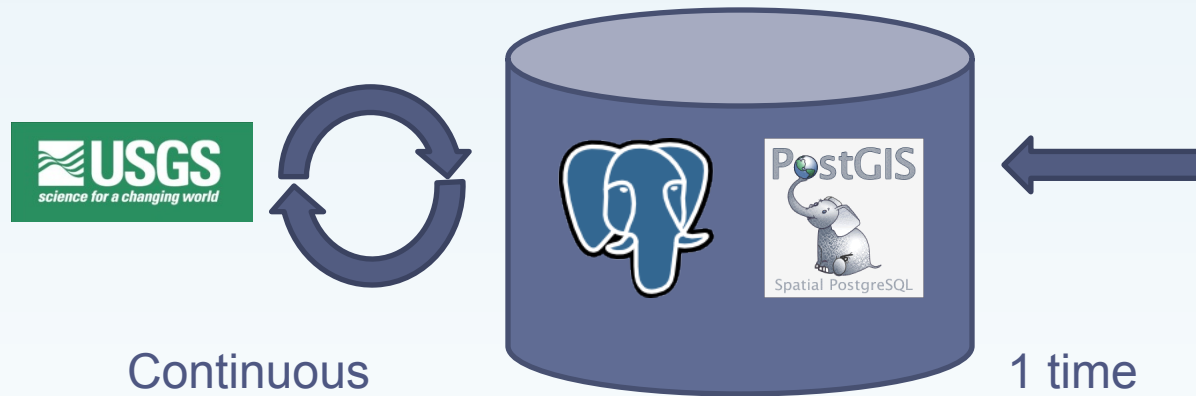
10 km
5 mi

Leaflet | © OpenStreetMap contributors

imlcz02.ncsa.illinois.edu/geodashboard#accordion-nav-collapse-layers

Geostreaming API

Historical and Real Time Data

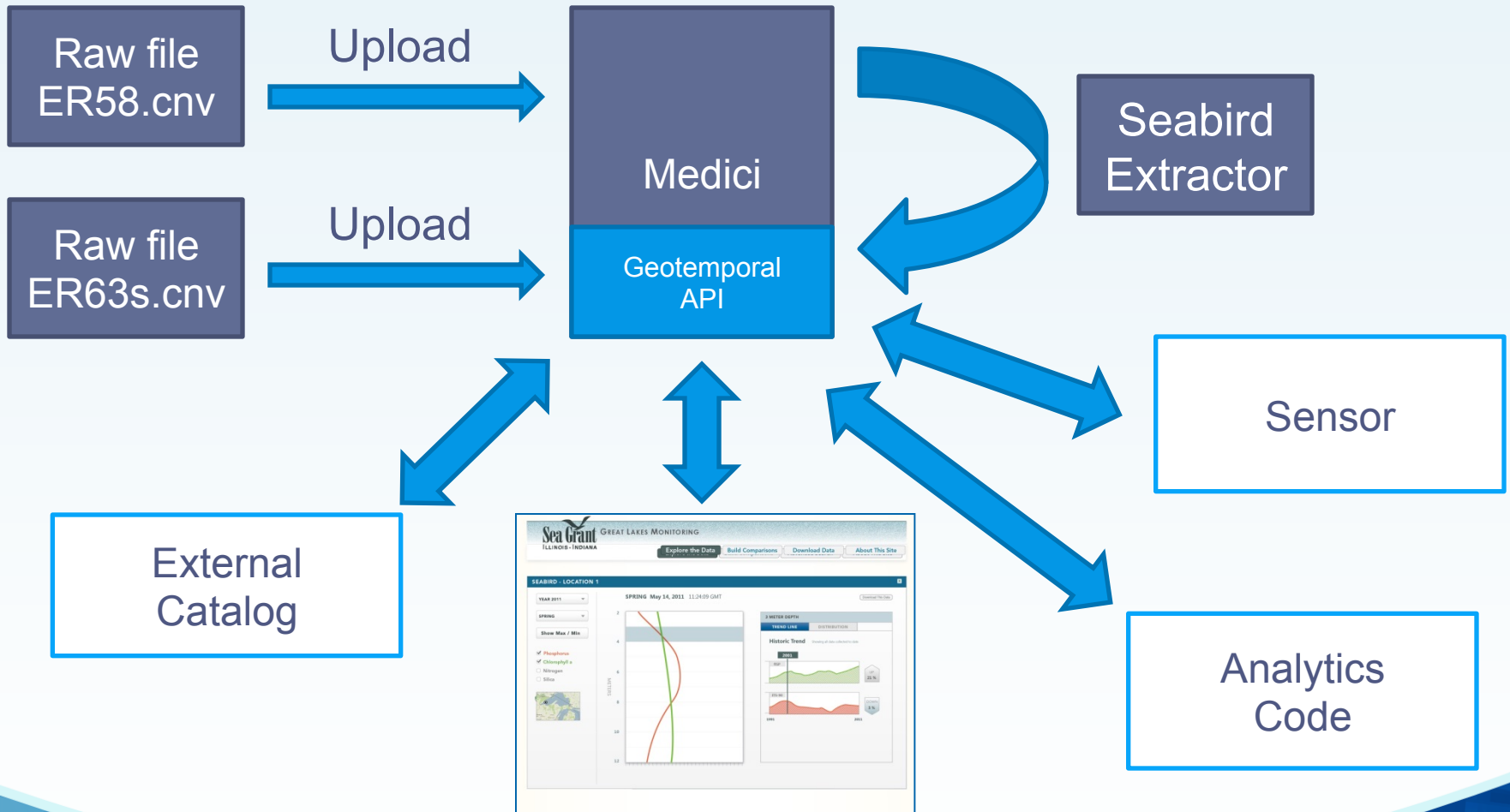


	A	B	C	D	E
1	samID	station	Date	tp	Chl_a
2	11	1	4-May-01	11	
3	12	1	14-May-01	7	0.5
4	13	1	31-May-01	10	0.4
5	14	1	13-Jun-01	9	0.4
6	15	1	25-Jun-01	11	0.9
7	16	1	10-Jul-01	8	0.4
8	17	1	24-Jul-01	5	0.4
9	18	1	6-Aug-01		
10	19	1	21-Aug-01	8	0.8
11	20	1	5-Sep-01	6	0.3
12	21	1	17-Sep-01	9	0.8
13	32	2	4-May-01	7	
14	33	2	14-May-01	5	0.5
15	34	2	31-May-01	7	0.6
16	35	2	13-Jun-01	7	0.8
17	36	2	25-Jun-01	10	1
18	37	2	10-Jul-01	9	0.6
19	38	2	24-Jul-01	6	0.5
20	39	2	6-Aug-01		
21	40	2	21-Aug-01	10	0.4
22	41	2	5-Sep-01	7	1.5
23	42	2	17-Sep-01	6	1
24	43	1	15-May-02	6	1.5
25	44	1	28-May-02	10	2.2
26	45	1	10-Jun-02	8	1
27	46	1	28-Jun-02	5	0.7
28	47	1	8-Jul-02	7	0.5
29	48	1	24-Jul-02	7	0.7
30	49	1	9-Aug-02	7	0.4
31	50	1	20-Aug-02	15	0.7
32	51	1	5-Sep-02	8	0.3
33	52	2	15-May-02	5	1.4
34	53	2	28-May-02	10	1.3
35	54	2	10-Jun-02	8	0.9
36	55	2	28-Jun-02	6	0.6
37	56	2	8-Jul-02	10	0.6
38	57	2	24-Jul-02	9	0.7
39	58	2	9-Aug-02	7	0.6
40	59	2	20-Aug-02	12	0.8
41	60	2	5-Sep-02	10	1.1
42	61	1	6-May-03	20	1.4

LEC_EPA_Data_Request_2012_Nutri

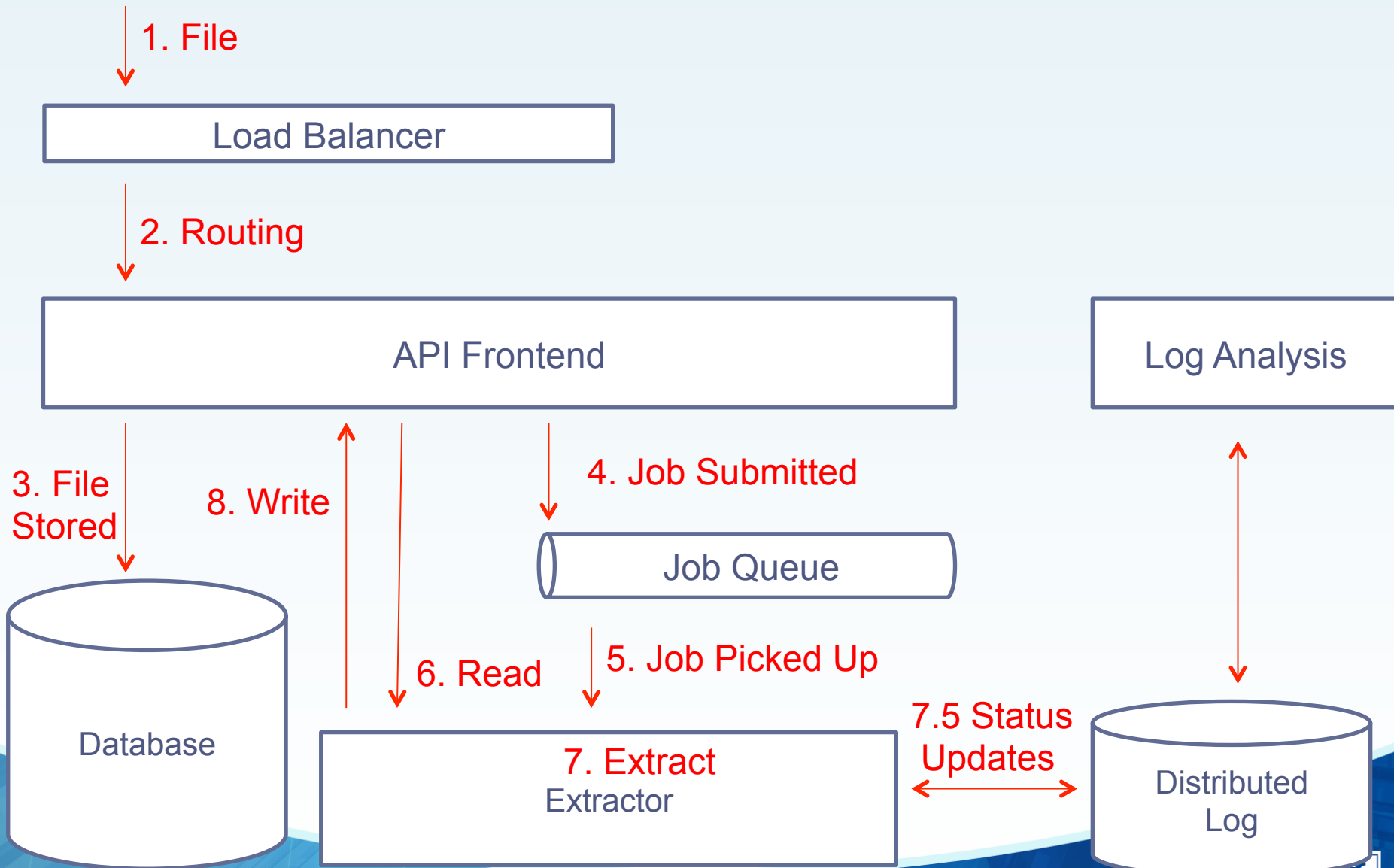
Normal View Ready

Data Lifecycle



EXTRACTION BUS

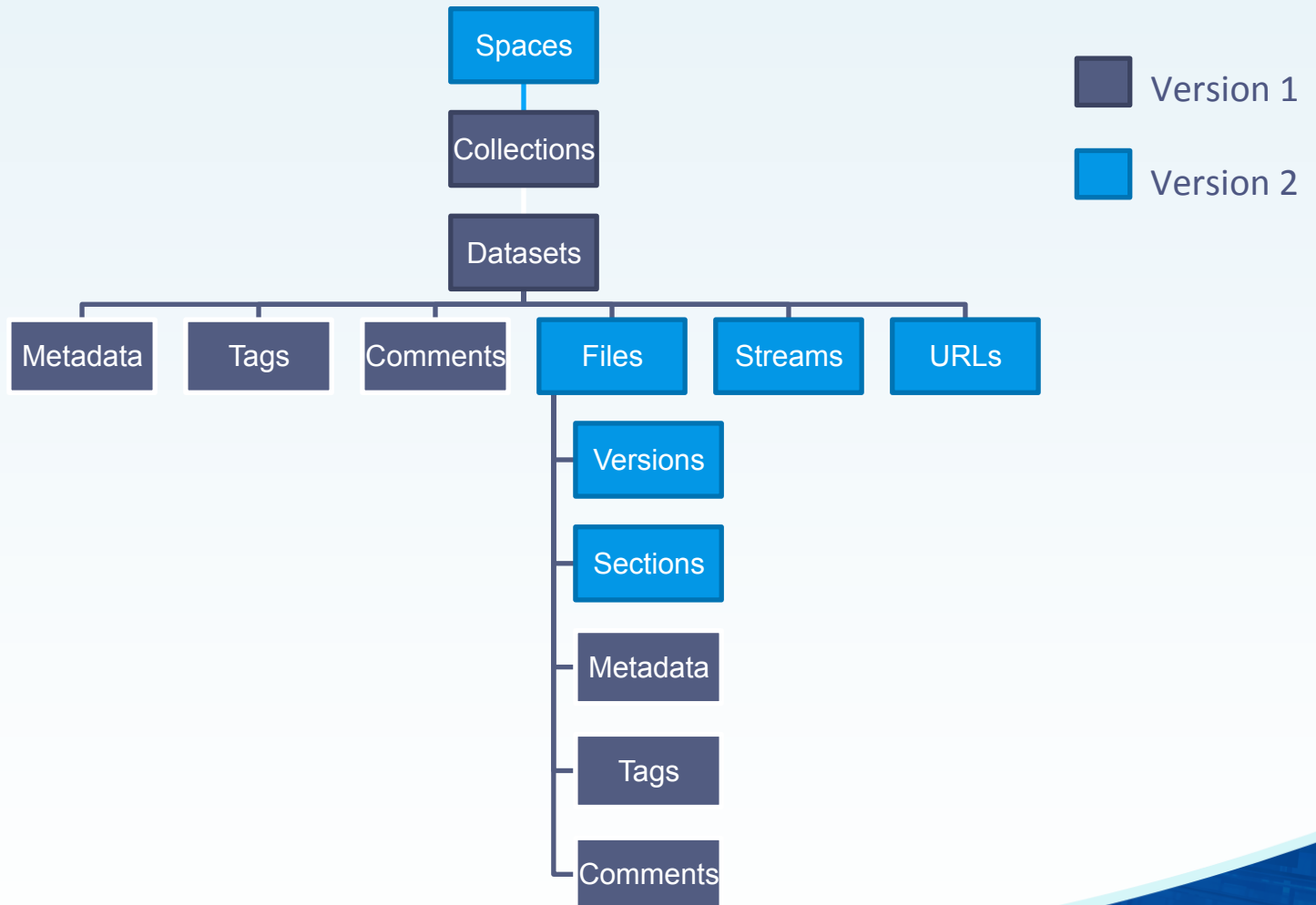
Extraction Architecture w/ Distributed Logging



Extractors Registration

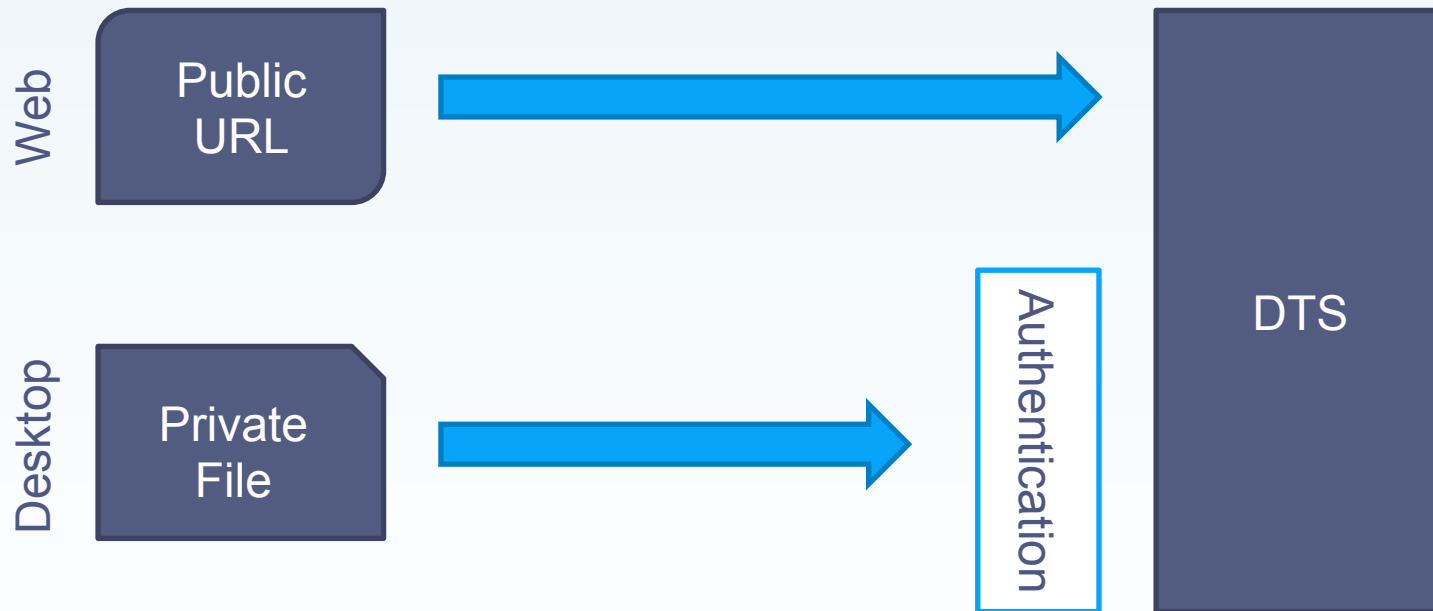
- Extractors register on
 - Internet media type
 - application/pdf
 - image/gif
 - Model/x3d+xml
 - Object type
 - file
 - file section
 - file preview
 - dataset

Medici Data Model

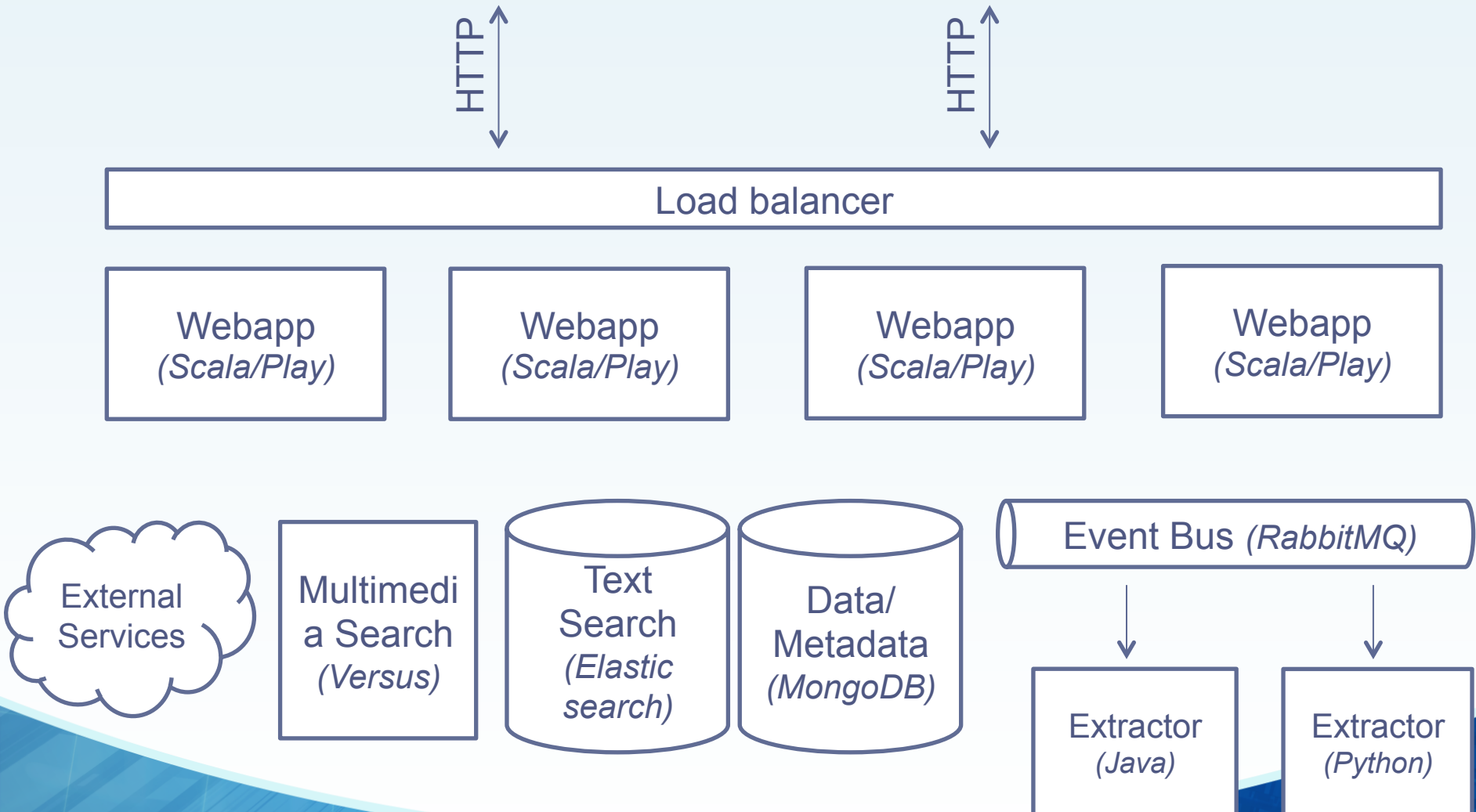


Privacy

- Public Web (public URLs)
- Private Web (private files)

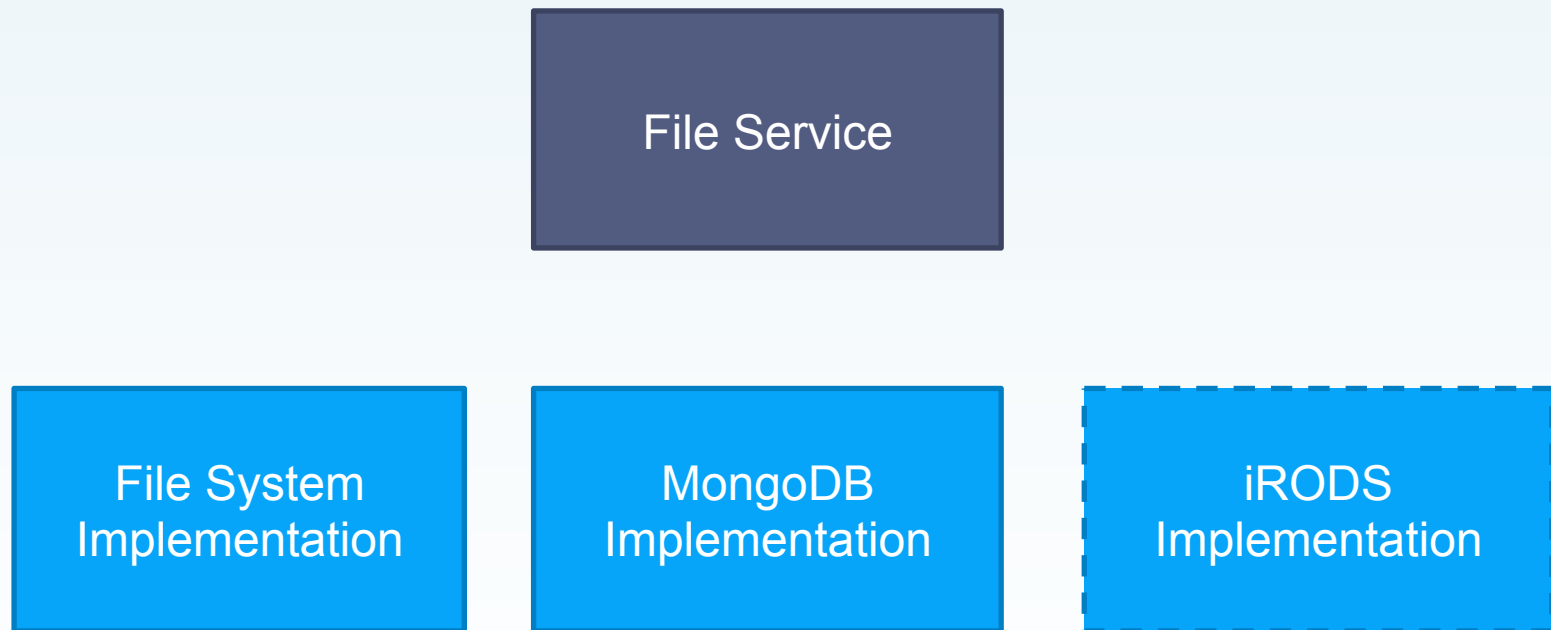


Medici Architecture (v.2)



Plugins / Services

- Ability to swap underlying implementation



Technology



elasticsearch



PostgreSQL



CREATING EXTRACTORS AND MEASURES

Creating New Extractors

- Be able to talk HTTP and parse JSON
- Learn the API
- Use a RabbitMQ client library to connect to broker
- Learn the extraction architecture
 - What messages to listen for and publish
 - How to parse message payload
- Tutorial
 - <https://opensource.ncsa.illinois.edu/confluence/display/BD/Medici+Extractor+in+Python>

Creating Versus Methods

- Write Java Code
- Can execute external code using
 - `Runtime.getRuntime().exec(args)`
 - JNI
- Register them using Java services
 - Add fully qualified class name to respective service file
 - For example add
 - `edu.illinois.ncsa.versus.extract.impl.RGBHistogramExtractor`
 - To
 - `/META-INF/services/edu.illinois.ncsa.versus.extract.Extractor`
 - Restart

Creating Versus Methods

- Adapters

- <https://opensource.ncsa.illinois.edu/stash/projects/VS/repos/versus-core/browse/src/main/java/edu/illinois/ncsa/versus/adapter/Adapter.java>

- Extractors

- <https://opensource.ncsa.illinois.edu/stash/projects/VS/repos/versus-core/browse/src/main/java/edu/illinois/ncsa/versus/extract/Extractor.java>

- Descriptors

- <https://opensource.ncsa.illinois.edu/stash/projects/VS/repos/versus-core/browse/src/main/java/edu/illinois/ncsa/versus/descriptor/Descriptor.java>

Creating Versus Methods

- Proximities

- <https://opensource.ncsa.illinois.edu/stash/projects/VS/repos/versus-core/browse/src/main/java/edu/illinois/ncsa/versus/measure/Proximity.java>

- Indexers

- <https://opensource.ncsa.illinois.edu/stash/projects/VS/repos/versus-core/browse/src/main/java/edu/illinois/ncsa/versus/search/Indexer.java>

- Service Registration

- <https://opensource.ncsa.illinois.edu/stash/projects/VS/repos/versus-image/browse/src/main/resources/META-INF/services/edu.illinois.ncsa.versus.extract.Extractor>

For more information

- Medici demo site
 - <http://medici-demo.ncsa.illinois.edu/>
- Medici documentation, source code, bugs
 - <https://opensource.ncsa.illinois.edu/projects/project.php?key=MMDB>
- Medici manual
 - <http://isda.ncsa.illinois.edu/documentation/medici/>
- Versus documentation, source code, bugs
 - <https://opensource.ncsa.illinois.edu/projects/project.php?key=VS>
- Versus manual
 - http://isda.ncsa.illinois.edu/documentation/versus/manual_strapdown.html

Contributors

- Rob Kooper
- Rui Liu
- Liana Diesendruck
- Ashwini Vaidya
- Nicholas Tenczar
- Sandeep Puthanveetil
- Constantinos Sophocleous
- Michael Simeone
- Theerasit Issaranon
- Smruti Padhy
- Kenton McHenry
- Jong Lee
- Mario Felarca
- Virginia Kuhn
- Alan Craig
- Dave Mattson
- Inna Zharnitsky

If I missed you please come kick me ☹️

Questions

<http://browndog.ncsa.illinois.edu/>

isda@ncsa.illinois.edu