THE 4TH SOLID SYMPOSIUM · LONDON · 2026

THE NEXT CHAPTER FOR DATA ON THE WEB

# *Halcyon*

## Digital Pathology Whole-Slide Imaging
## *Research Life-cycle Management*

**Erich Bremer**

*Director for Applied Informatics · Stony Brook University*

SPEAKER

# Erich Bremer

## ROLE

Director for Applied Informatics Biomedical Informatics, Stony Brook University

W3C AC Representative, Stony Brook

Co-editor — W3C Linked Web Storage Protocol

## CREDENTIALS

PhD - Biomedical Informatics (in progress) Stony Brook University

M.Sc. - Computer & Systems Engineering, Rensselaer Polytechnic Institute

B.S. - Computer & Systems Engineering, Rensselaer Polytechnic Institute

30+ years - programming, databases, networking, OS

## INTERESTS

AI and medical imaging

Scalable linked-data systems

Data visualization

Standards-based, open research infrastructure

THE PROBLEM

# A single pathology slide is huge.

**40 GB**

uncompressed size of a single whole-slide image

CONTEXT

# Research life cycle of a pathology image

01

**Acquire**

*Scan slide → WSI*

02

**Annotate**

*Expert ground truth*

03

**Train**

*ML model training*

04

**Infer**

*Generate predictions*

05

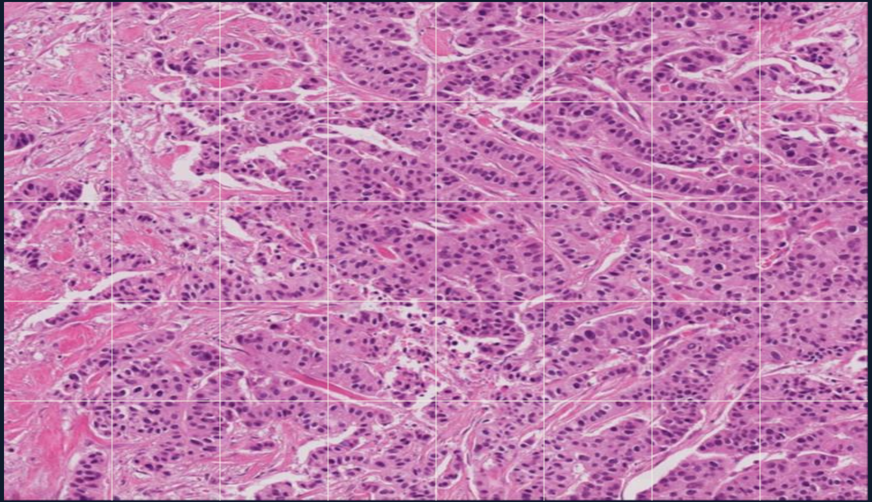
**Validate**

*Compare & curate*

*Every stage generates data that must be linked back to the source image.*

DELIVERY · TILING

# You can't send the whole image.
# Tile it.



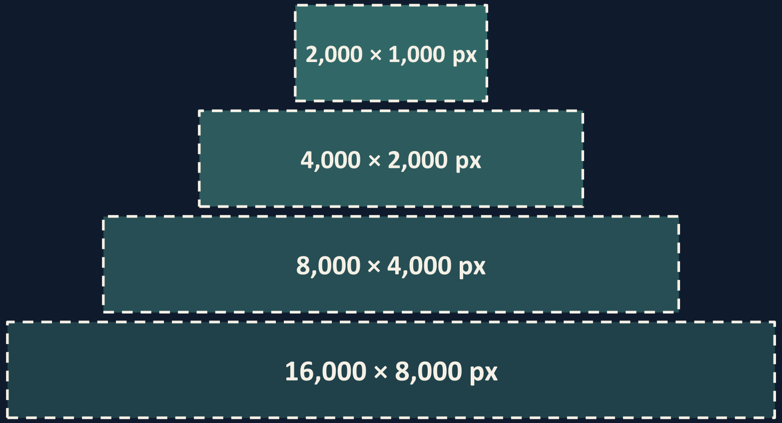
The browser requests only the tiles it needs for the current viewport and zoom level.

**IIIF Protocol**

- International Image Interoperability Framework, the standard used across digital libraries and pathology.

DELIVERY · PYRAMIDS

# Image pyramids

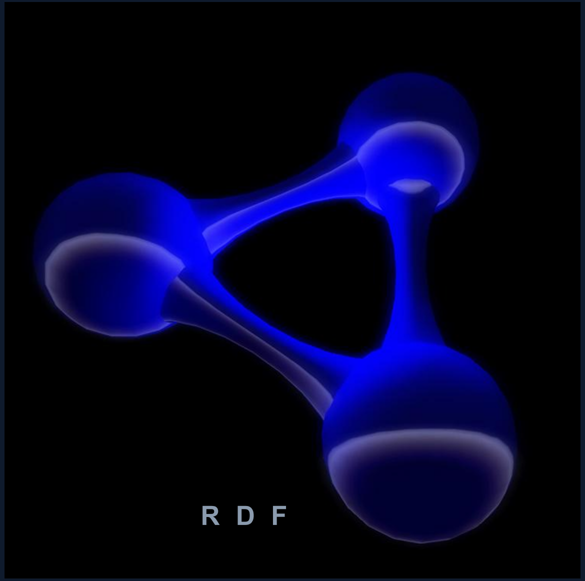


- ● Level 0 · full resolution
  *16,000 × 8,000 px*
- ● Level 1
  *8,000 × 4,000 px*
- ● Level 2
  *4,000 × 2,000 px*
- ● Level 3 · overview
  *2,000 × 1,000 px*

*Each zoom level is a pre-computed, independently addressable layer.*

APPROACH

# So how do you build this?



The answer

## *RDF all the way down.*

Every tile, every annotation, every model output -
*addressable, linkable, queryable.*

ARCHITECTURE

# Halcyon · components

HALCYON

**Apache Jena TDB2**
*RDF triple store*

**Apache Fuseki**
*SPARQL endpoint*

**Eclipse Jetty**
*HTTP server*

**Image Tiling (IIIF)**
*Pyramid delivery*

**Feature Tiling (IIIF)**
*Annotation delivery*

**MCP Interface**
*Agent tool surface*

**Feature / Image Management · OpenSeadragon · WebGL · LWS/Solid**

SPATIAL INDEXING

# Many ways to index 2-D and 3-D spaces

### SPACE-FILLING CURVES

*Linear index, strong locality*

- Geohash
- HHCode
- Z-order curve
- **Hilbert curve**

### TREE-BASED

*Recursive partition of space*

- Quadtree
- Octree
- kd-tree
- BSP-tree
- m-tree

### R-TREE FAMILY

*Bounding-box overlap queries*

- R-tree
- R+ tree
- R* tree
- Hilbert R-tree
- X-tree
- UB-tree

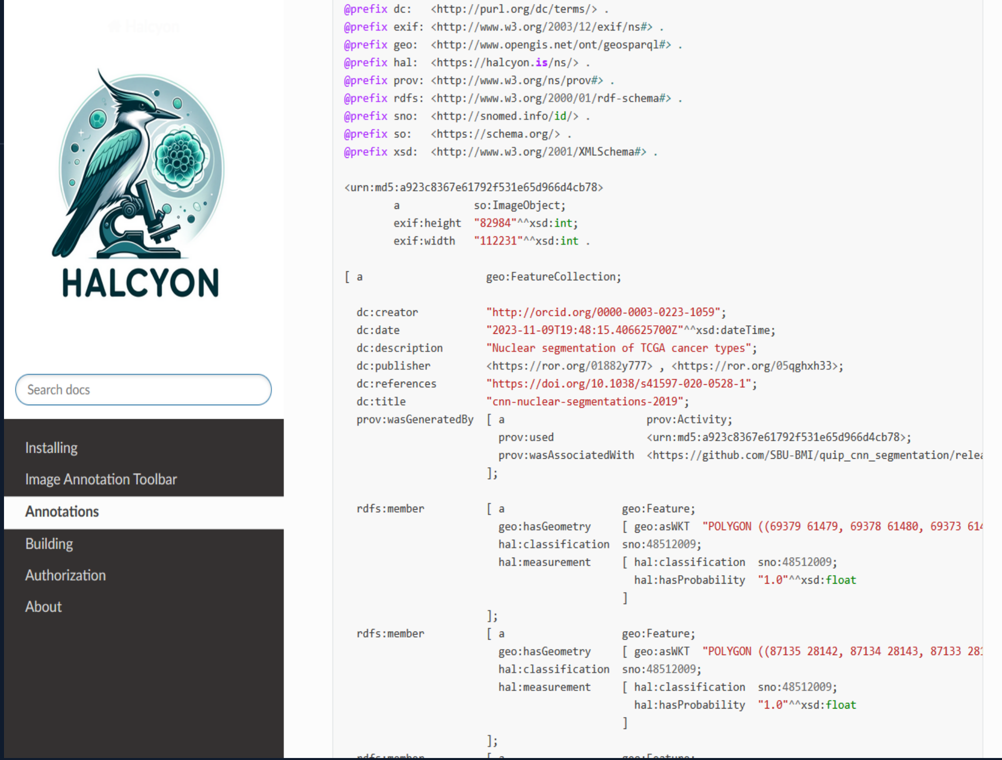
QUERY LAYER

# GeoSPARQL — spatial queries over RDF

The OGC standard that lets you write SPARQL queries like

**EXAMPLE QUERY**

```
SELECT ?feature WHERE {
  ?feature geo:asWKT ?wkt .
  FILTER(
    geof:sfWithin(?wkt, $region)
  )
}
```



```
@prefix dc:   <http://purl.org/dc/terms/> .
@prefix exif: <http://www.w3.org/2003/12/exif/ns#> .
@prefix geo:  <http://www.opengis.net/ont/geosparql#> .
@prefix hal:  <https://halcyon.is/ns/> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix sno:  <http://snomed.info/id/> .
@prefix so:   <https://schema.org/> .
@prefix xsd:  <http://www.w3.org/2001/XMLSchema#> .

<urn:md5:a923c8367e61792f531e65d966d4cb78>
        a            so:ImageObject;
        exif:height  "82984"^^xsd:int;
        exif:width   "112231"^^xsd:int .

[ a                    geo:FeatureCollection;

  dc:creator           "http://orcid.org/0000-0003-0223-1059";
  dc:date              "2023-11-09T19:48:15.406625700Z"^^xsd:dateTime;
  dc:description       "Nuclear segmentation of TCGA cancer types";
  dc:publisher         <https://ror.org/01882y777> , <https://ror.org/05qghxh33>;
  dc:references        "https://doi.org/10.1038/s41597-020-0528-1";
  dc:title             "cnn-nuclear-segmentations-2019";
  prov:wasGeneratedBy  [ a                     prov:Activity;
                         prov:used             <urn:md5:a923c8367e61792f531e65d966d4cb78>;
                         prov:wasAssociatedWith  <https://github.com/SBU-BMI/quip_cnn_segmentation/relea
                       ];

  rdfs:member          [ a                   geo:Feature;
                         geo:hasGeometry     [ geo:asWKT  "POLYGON ((69379 61479, 69378 61480, 69373 61
                         hal:classification  sno:48512009;
                         hal:measurement     [ hal:classification  sno:48512009;
                                               hal:hasProbability  "1.0"^^xsd:float
                                             ]
                       ];
  rdfs:member          [ a                   geo:Feature;
                         geo:hasGeometry     [ geo:asWKT  "POLYGON ((87135 28142, 87134 28143, 87133 28
                         hal:classification  sno:48512009;
                         hal:measurement     [ hal:classification  sno:48512009;
                                               hal:hasProbability  "1.0"^^xsd:float
                                             ]
                       ];
```

*Geometry becomes a queryable property of any resource — slides, annotations, model outputs.*

STORAGE BACKEND

# BeakGraph

*Apache Jena graph implementation · RDF HDT packed in HDF5*

BeakGraph · HDF5

**APACHE JENA**

## Drop-in graph

*Implements the Jena Graph interface — use it anywhere Jena runs.*

**RDF HDT**

## Compressed triples

*Header · Dictionary · Triples. Small on disk, queryable without expansion.*

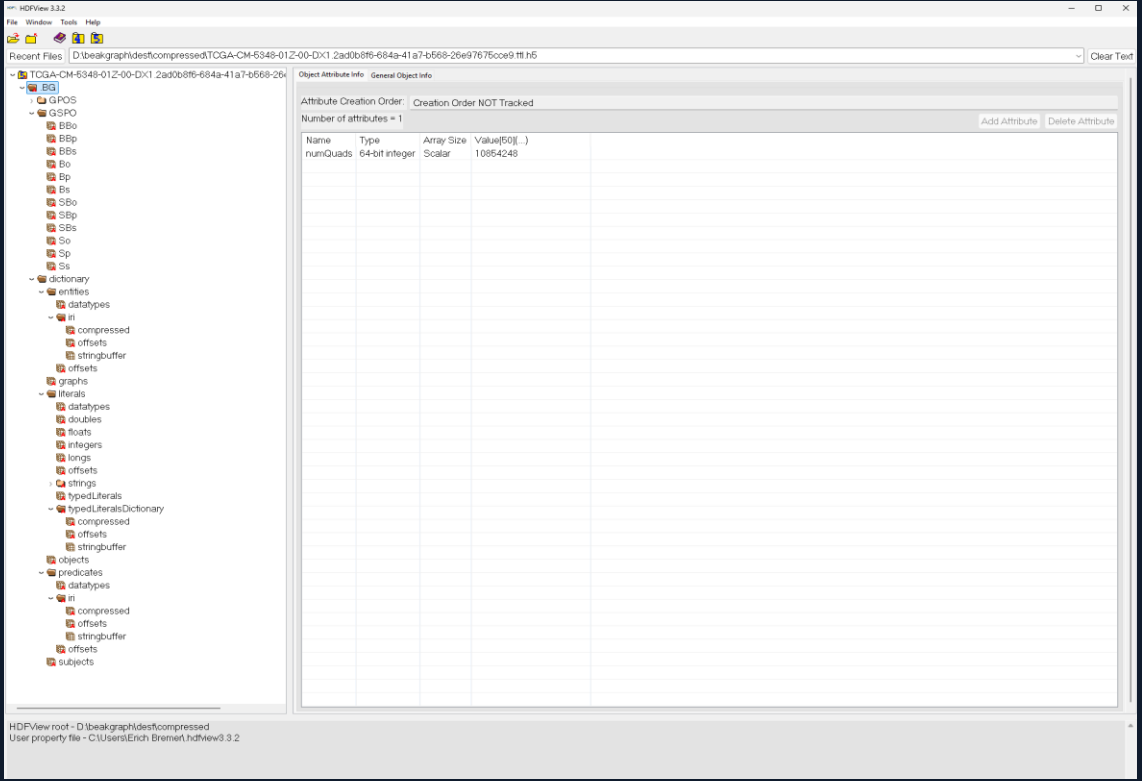
**HDF5 CONTAINER**

## Chunked, portable

*One self-describing file. Random access, version-able, widely supported.*

RUNNING SYSTEM · INTERNALS

# Inside a BeakGraph HDF5 file



D:\beakgraph\dest\compressed\TCGA-CM-5348-01Z-00-DX1.2ad0b8f6-684a-41a7-b568-26e97675cce9.ttl.h5

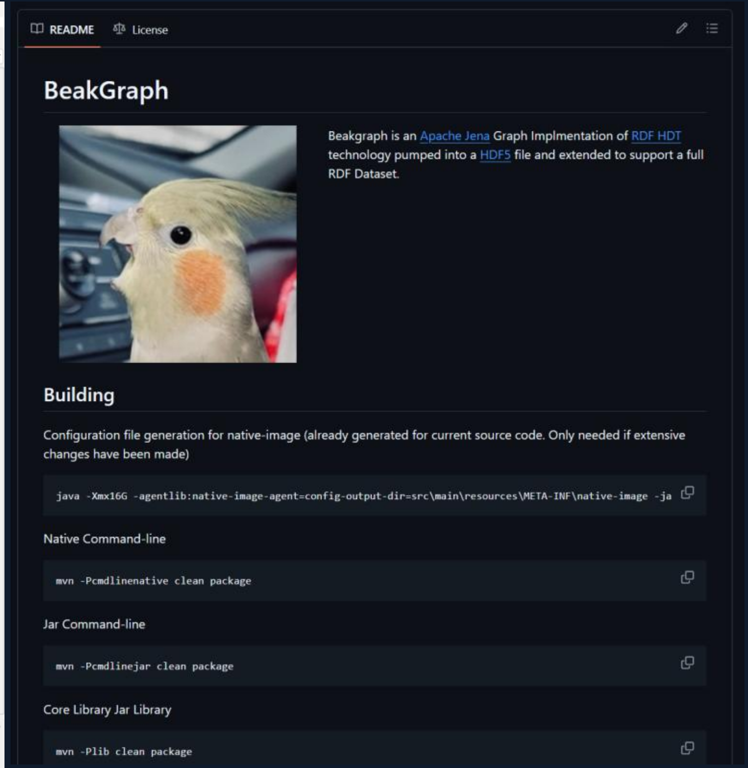
- TCGA-CM-5348-01Z-00-DX1.2ad0b8f6-684a-41a7-b568-26
  - .BG
    - GPOS
    - GSPO
      - BBo
      - BBp
      - BBs
      - Bo
      - Bp
      - Bs
      - SBo
      - SBp
      - SBs
      - So
      - Sp
      - Ss
    - dictionary
      - entities
        - datatypes
        - iri
          - compressed
          - offsets
          - stringbuffer
        - offsets
      - graphs
      - literals
        - datatypes
        - doubles
        - floats
        - integers
        - longs
        - offsets
        - strings
        - typedLiterals
        - typedLiteralsDictionary
          - compressed
          - offsets
          - stringbuffer
      - objects
      - predicates
        - datatypes
        - iri
          - compressed
          - offsets
          - stringbuffer
        - offsets
      - subjects

Attribute Creation Order: Creation Order NOT Tracked

Number of attributes = 1

| Name | Type | Array Size | Value[50](...) |
|---|---|---|---|
| numQuads | 64-bit integer | Scalar | 10854248 |

HDFView root - D:\beakgraph\dest\compressed
User property file - C:\Users\Erich Bremer\.hdfview3.3.2



## BeakGraph

Beakgraph is an Apache Jena Graph Implmentation of RDF HDT technology pumped into a HDF5 file and extended to support a full RDF Dataset.

## Building

Configuration file generation for native-image (already generated for current source code. Only needed if extensive changes have been made)

```
java -Xmx16G -agentlib:native-image-agent=config-output-dir=src\main\resources\META-INF\native-image -ja
```

Native Command-line

```
mvn -Pcmdlinenative clean package
```

Jar Command-line

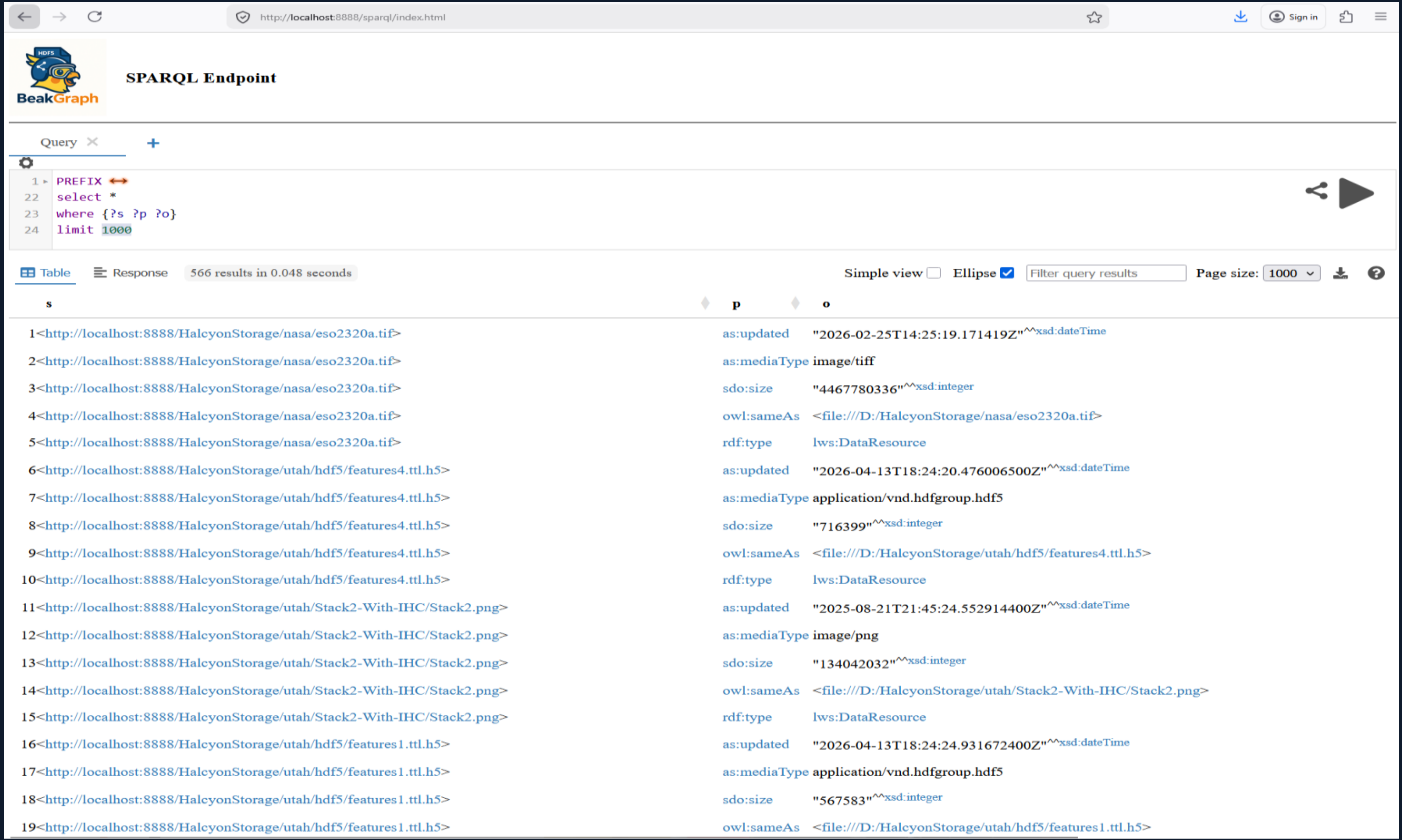
```
mvn -Pcmdlinejar clean package
```

Core Library Jar Library

```
mvn -Plib clean package
```

*Dictionary, triples, and spatial indexes — all as HDF5 datasets, browsable with standard tooling.*

# RUNNING SYSTEM · SPARQL



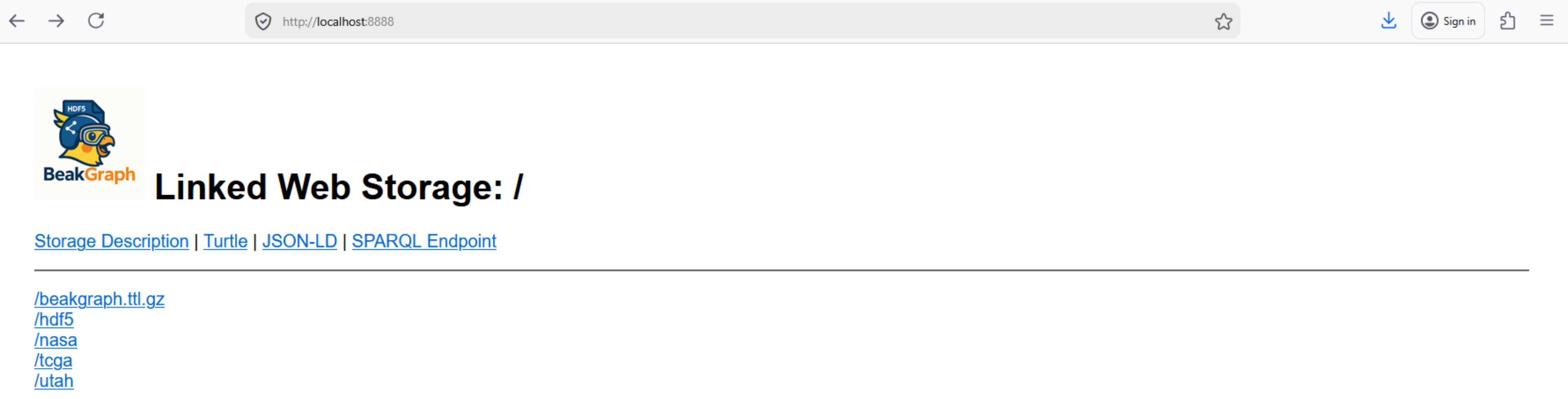
**SPARQL Endpoint**

```
PREFIX ↔
select *
where {?s ?p ?o}
limit 1000
```

566 results in 0.048 seconds

Simple view ☐ Ellipse ☑ Page size: 1000

| | s | p | o |
|---|---|---|---|
| 1 | <http://localhost:8888/HalcyonStorage/nasa/eso2320a.tif> | as:updated | "2026-02-25T14:25:19.171419Z"^^xsd:dateTime |
| 2 | <http://localhost:8888/HalcyonStorage/nasa/eso2320a.tif> | as:mediaType | image/tiff |
| 3 | <http://localhost:8888/HalcyonStorage/nasa/eso2320a.tif> | sdo:size | "4467780336"^^xsd:integer |
| 4 | <http://localhost:8888/HalcyonStorage/nasa/eso2320a.tif> | owl:sameAs | <file:///D:/HalcyonStorage/nasa/eso2320a.tif> |
| 5 | <http://localhost:8888/HalcyonStorage/nasa/eso2320a.tif> | rdf:type | lws:DataResource |
| 6 | <http://localhost:8888/HalcyonStorage/utah/hdf5/features4.ttl.h5> | as:updated | "2026-04-13T18:24:20.476006500Z"^^xsd:dateTime |
| 7 | <http://localhost:8888/HalcyonStorage/utah/hdf5/features4.ttl.h5> | as:mediaType | application/vnd.hdfgroup.hdf5 |
| 8 | <http://localhost:8888/HalcyonStorage/utah/hdf5/features4.ttl.h5> | sdo:size | "716399"^^xsd:integer |
| 9 | <http://localhost:8888/HalcyonStorage/utah/hdf5/features4.ttl.h5> | owl:sameAs | <file:///D:/HalcyonStorage/utah/hdf5/features4.ttl.h5> |
| 10 | <http://localhost:8888/HalcyonStorage/utah/hdf5/features4.ttl.h5> | rdf:type | lws:DataResource |
| 11 | <http://localhost:8888/HalcyonStorage/utah/Stack2-With-IHC/Stack2.png> | as:updated | "2025-08-21T21:45:24.552914400Z"^^xsd:dateTime |
| 12 | <http://localhost:8888/HalcyonStorage/utah/Stack2-With-IHC/Stack2.png> | as:mediaType | image/png |
| 13 | <http://localhost:8888/HalcyonStorage/utah/Stack2-With-IHC/Stack2.png> | sdo:size | "134042032"^^xsd:integer |
| 14 | <http://localhost:8888/HalcyonStorage/utah/Stack2-With-IHC/Stack2.png> | owl:sameAs | <file:///D:/HalcyonStorage/utah/Stack2-With-IHC/Stack2.png> |
| 15 | <http://localhost:8888/HalcyonStorage/utah/Stack2-With-IHC/Stack2.png> | rdf:type | lws:DataResource |
| 16 | <http://localhost:8888/HalcyonStorage/utah/hdf5/features1.ttl.h5> | as:updated | "2026-04-13T18:24:24.931672400Z"^^xsd:dateTime |
| 17 | <http://localhost:8888/HalcyonStorage/utah/hdf5/features1.ttl.h5> | as:mediaType | application/vnd.hdfgroup.hdf5 |
| 18 | <http://localhost:8888/HalcyonStorage/utah/hdf5/features1.ttl.h5> | sdo:size | "567583"^^xsd:integer |
| 19 | <http://localhost:8888/HalcyonStorage/utah/hdf5/features1.ttl.h5> | owl:sameAs | <file:///D:/HalcyonStorage/utah/hdf5/features1.ttl.h5> |



# Linked Web Storage: /

Storage Description | Turtle | JSON-LD | SPARQL Endpoint

/beakgraph.ttl.gz
/hdf5
/nasa
/tcga
/utah

*LWS containers expose Turtle, JSON-LD, and a SPARQL endpoint at every level.*

UI · ARCHITECTURE

# Vandegraph

*How Halcyon renders SHACL-defined forms with Apache Wicket and Jena.*

**APACHE WICKET**

Component-based UI

**APACHE JENA**

RDF model + SHACL

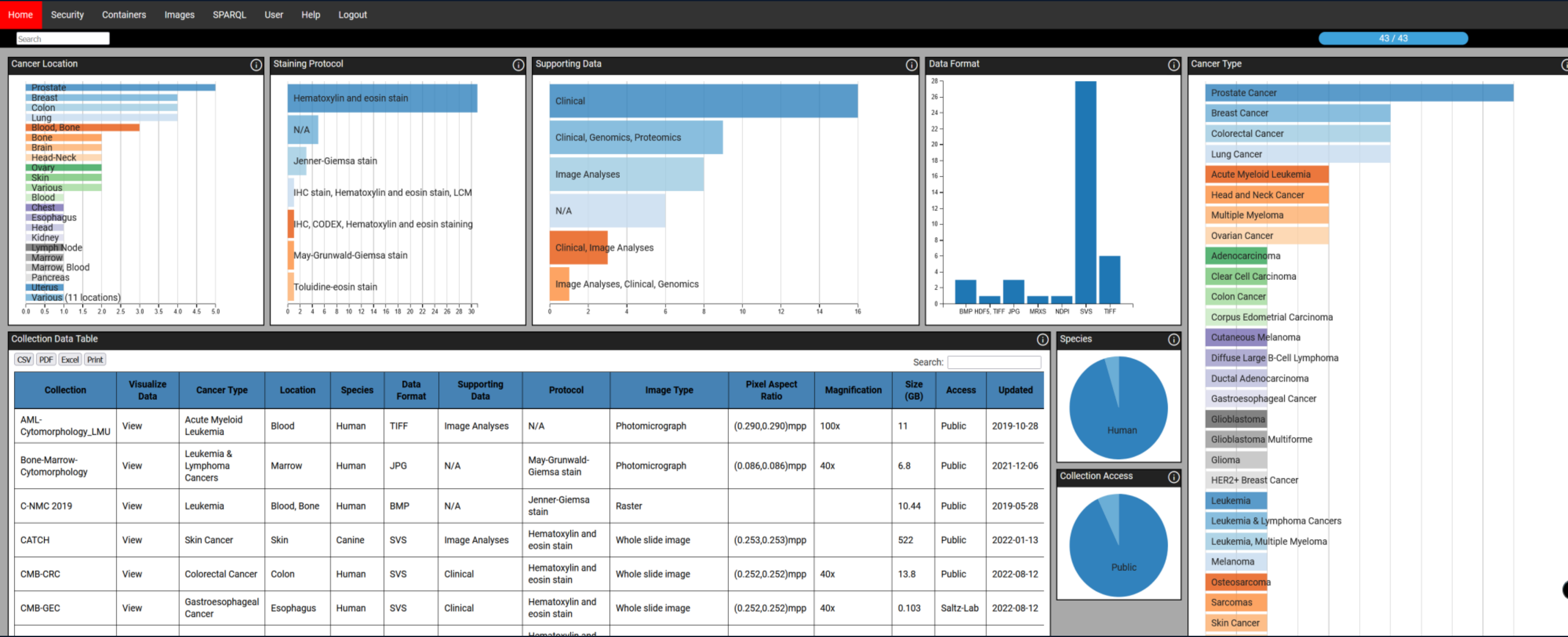
**VANDEGRAPH**

Wicket ⟷ SHACL glue ⟷ Apache Jena

*A SHACL shape becomes a Halcyon form. No hand-coded UI per entity type.*

RUNNING SYSTEM · DASHBOARD

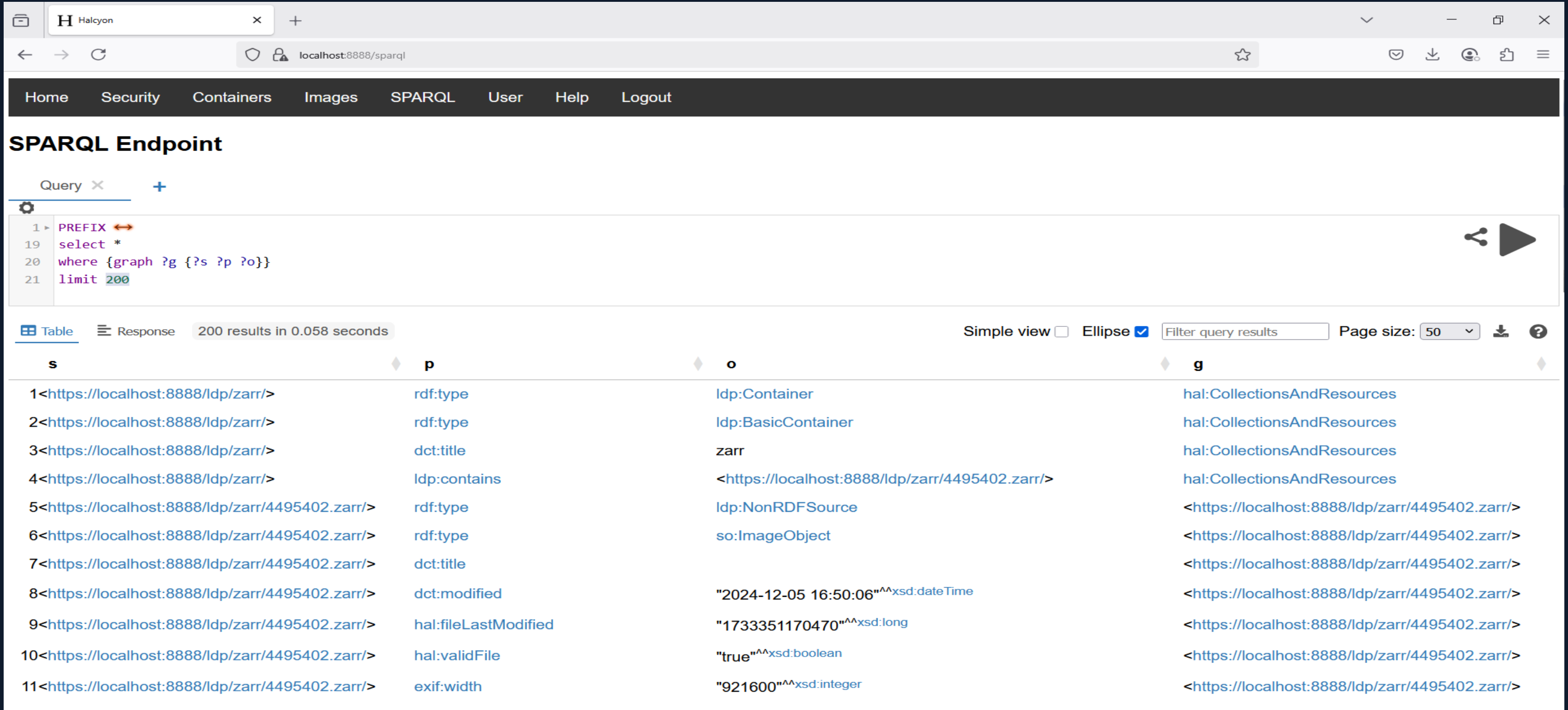
# Collection — summary dashboard



| Collection | Visualize Data | Cancer Type | Location | Species | Data Format | Supporting Data | Protocol | Image Type | Pixel Aspect Ratio | Magnification | Size (GB) | Access | Updated |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AML-Cytomorphology_LMU | View | Acute Myeloid Leukemia | Blood | Human | TIFF | Image Analyses | N/A | Photomicrograph | (0.290,0.290)mpp | 100x | 11 | Public | 2019-10-28 |
| Bone-Marrow-Cytomorphology | View | Leukemia & Lymphoma Cancers | Marrow | Human | JPG | N/A | May-Grunwald-Giemsa stain | Photomicrograph | (0.086,0.086)mpp | 40x | 6.8 | Public | 2021-12-06 |
| C-NMC 2019 | View | Leukemia | Blood, Bone | Human | BMP | N/A | Jenner-Giemsa stain | Raster |  |  | 10.44 | Public | 2019-05-28 |
| CATCH | View | Skin Cancer | Skin | Canine | SVS | Image Analyses | Hematoxylin and eosin stain | Whole slide image | (0.253,0.253)mpp |  | 522 | Public | 2022-01-13 |
| CMB-CRC | View | Colorectal Cancer | Colon | Human | SVS | Clinical | Hematoxylin and eosin stain | Whole slide image | (0.252,0.252)mpp | 40x | 13.8 | Public | 2022-08-12 |
| CMB-GEC | View | Gastroesophageal Cancer | Esophagus | Human | SVS | Clinical | Hematoxylin and eosin stain | Whole slide image | (0.252,0.252)mpp | 40x | 0.103 | Saltz-Lab | 2022-08-12 |

*Facets, histograms, and the slide table — all driven by SPARQL under the hood.*

# SPARQL endpoint — embedded in Halcyon



## SPARQL Endpoint

```
PREFIX ↔
select *
where {graph ?g {?s ?p ?o}}
limit 200
```

200 results in 0.058 seconds

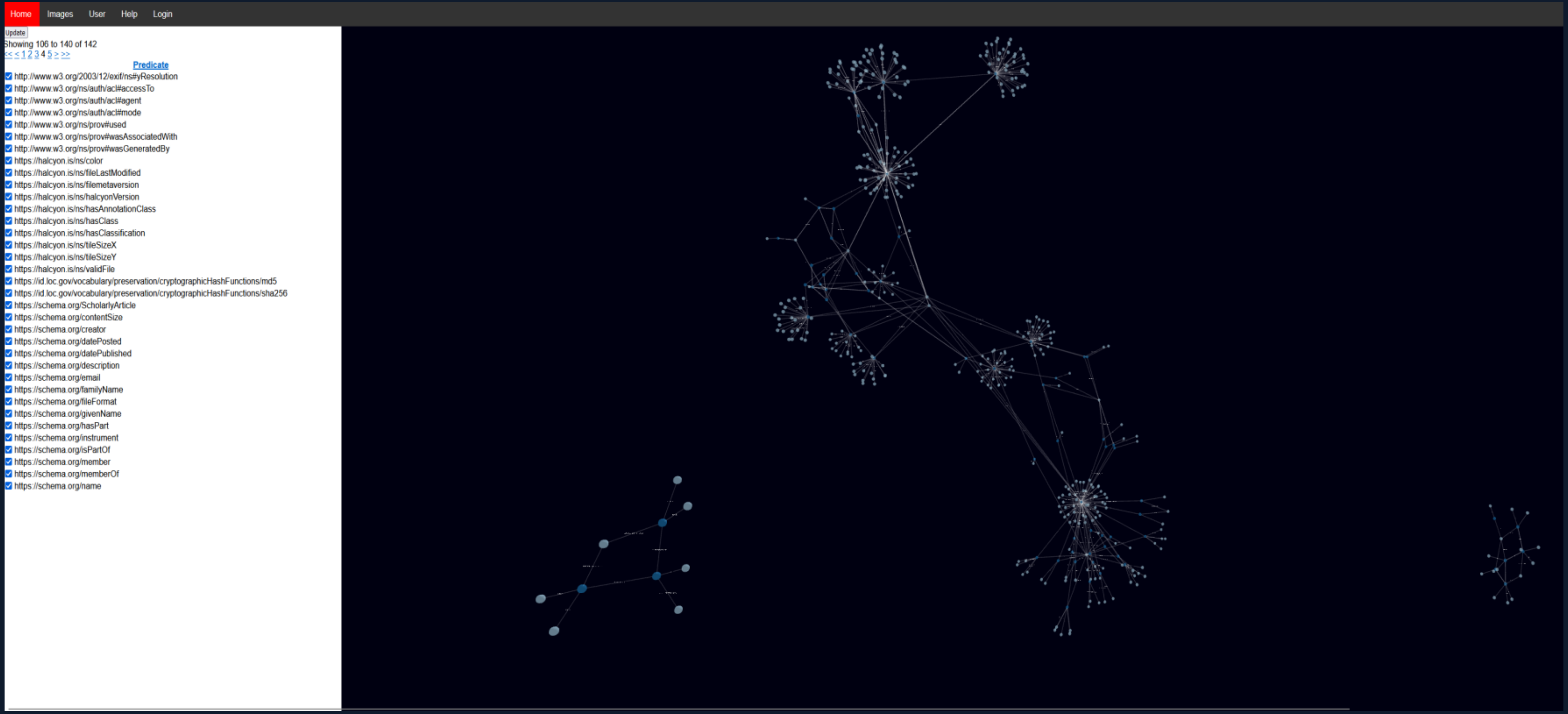
Simple view ☐ Ellipse ☑ Filter query results Page size: 50

| | s | p | o | g |
|---|---|---|---|---|
| 1 | <https://localhost:8888/ldp/zarr/> | rdf:type | ldp:Container | hal:CollectionsAndResources |
| 2 | <https://localhost:8888/ldp/zarr/> | rdf:type | ldp:BasicContainer | hal:CollectionsAndResources |
| 3 | <https://localhost:8888/ldp/zarr/> | dct:title | zarr | hal:CollectionsAndResources |
| 4 | <https://localhost:8888/ldp/zarr/> | ldp:contains | <https://localhost:8888/ldp/zarr/4495402.zarr/> | hal:CollectionsAndResources |
| 5 | <https://localhost:8888/ldp/zarr/4495402.zarr/> | rdf:type | ldp:NonRDFSource | <https://localhost:8888/ldp/zarr/4495402.zarr/> |
| 6 | <https://localhost:8888/ldp/zarr/4495402.zarr/> | rdf:type | so:ImageObject | <https://localhost:8888/ldp/zarr/4495402.zarr/> |
| 7 | <https://localhost:8888/ldp/zarr/4495402.zarr/> | dct:title | | <https://localhost:8888/ldp/zarr/4495402.zarr/> |
| 8 | <https://localhost:8888/ldp/zarr/4495402.zarr/> | dct:modified | "2024-12-05 16:50:06"^^xsd:dateTime | <https://localhost:8888/ldp/zarr/4495402.zarr/> |
| 9 | <https://localhost:8888/ldp/zarr/4495402.zarr/> | hal:fileLastModified | "1733351170470"^^xsd:long | <https://localhost:8888/ldp/zarr/4495402.zarr/> |
| 10 | <https://localhost:8888/ldp/zarr/4495402.zarr/> | hal:validFile | "true"^^xsd:boolean | <https://localhost:8888/ldp/zarr/4495402.zarr/> |
| 11 | <https://localhost:8888/ldp/zarr/4495402.zarr/> | exif:width | "921600"^^xsd:integer | <https://localhost:8888/ldp/zarr/4495402.zarr/> |

*The same SPARQL surface the UI uses is exposed to the researcher directly.*

RUNNING SYSTEM · VISUALIZATION

# Resource graph — interactive view



Showing 106 to 140 of 142

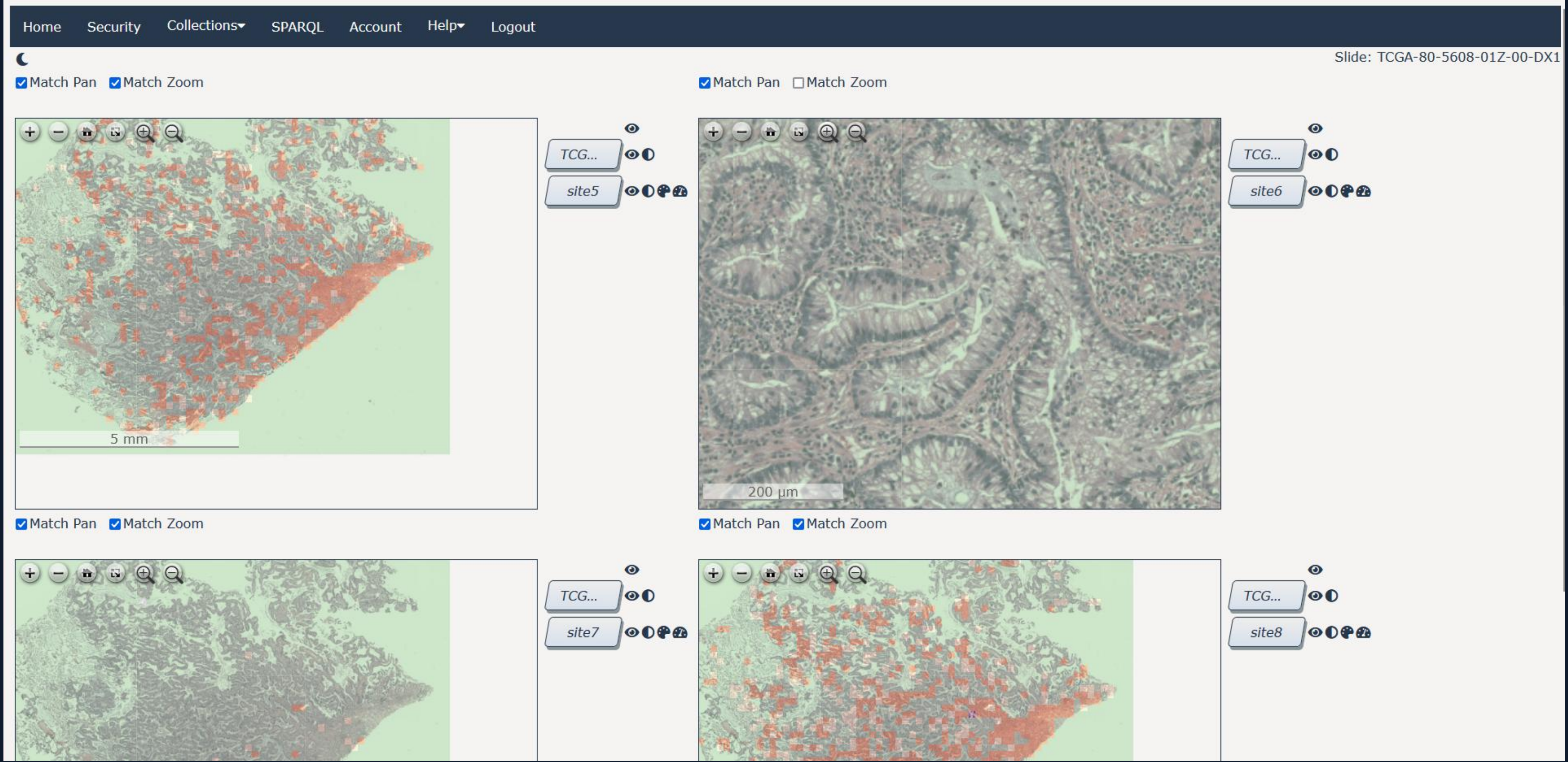
<< < 1 2 3 4 5 > >>

Predicate

☑ http://www.w3.org/2003/12/exif/ns#yResolution
☑ http://www.w3.org/ns/auth/acl#accessTo
☑ http://www.w3.org/ns/auth/acl#agent
☑ http://www.w3.org/ns/auth/acl#mode
☑ http://www.w3.org/ns/prov#used
☑ http://www.w3.org/ns/prov#wasAssociatedWith
☑ http://www.w3.org/ns/prov#wasGeneratedBy
☑ https://halcyon.is/ns/color
☑ https://halcyon.is/ns/fileLastModified
☑ https://halcyon.is/ns/filemetaversion
☑ https://halcyon.is/ns/halcyonVersion
☑ https://halcyon.is/ns/hasAnnotationClass
☑ https://halcyon.is/ns/hasClass
☑ https://halcyon.is/ns/hasClassification
☑ https://halcyon.is/ns/tileSizeX
☑ https://halcyon.is/ns/tileSizeY
☑ https://halcyon.is/ns/validFile
☑ https://id.loc.gov/vocabulary/preservation/cryptographicHashFunctions/md5
☑ https://id.loc.gov/vocabulary/preservation/cryptographicHashFunctions/sha256
☑ https://schema.org/ScholarlyArticle
☑ https://schema.org/contentSize
☑ https://schema.org/creator
☑ https://schema.org/datePosted
☑ https://schema.org/datePublished
☑ https://schema.org/description
☑ https://schema.org/email
☑ https://schema.org/familyName
☑ https://schema.org/fileFormat
☑ https://schema.org/givenName
☑ https://schema.org/hasPart
☑ https://schema.org/instrument
☑ https://schema.org/isPartOf
☑ https://schema.org/member
☑ https://schema.org/memberOf
☑ https://schema.org/name

*Every node is a URI you can click and follow.*

# VIEWER



Slide: TCGA-80-5608-01Z-00-DX1

☑ Match Pan ☑ Match Zoom

☑ Match Pan ☐ Match Zoom

☑ Match Pan ☑ Match Zoom

☑ Match Pan ☑ Match Zoom

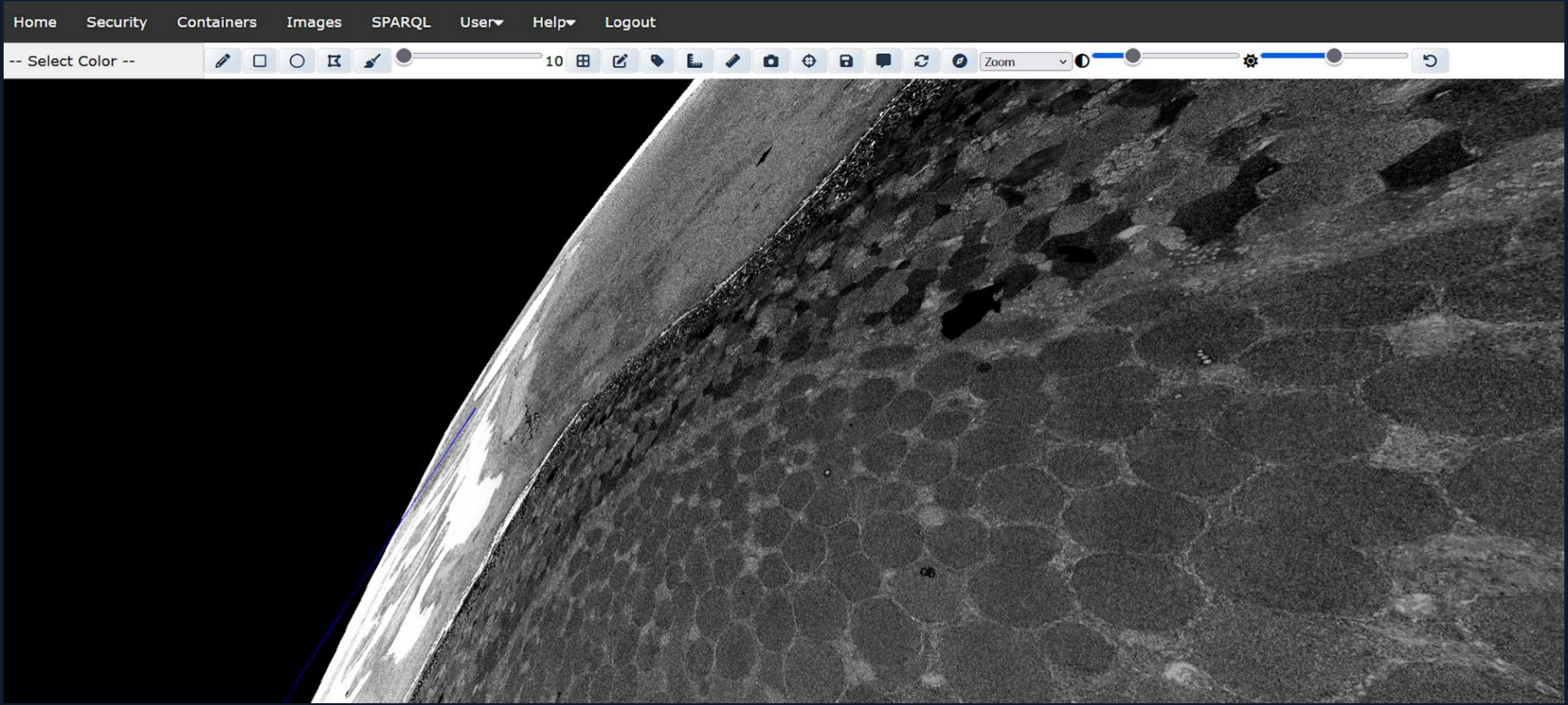
VIEWER

# Zephyr

*A new WebGL image viewer with annotation tools.*

| | |
|---|---|
| **WEBGL RENDERING** | Hardware-accelerated, smooth pan and zoom across gigapixel tiles. |
| **IIIF-NATIVE** | Consumes standard IIIF tile pyramids — no proprietary image server. |
| **ANNOTATION TOOLS** | Freehand, polygon, rectangle, point — every shape is RDF when saved. |
| **SHACL-DRIVEN UI** | Annotation class palette is generated from the active shape graph. |

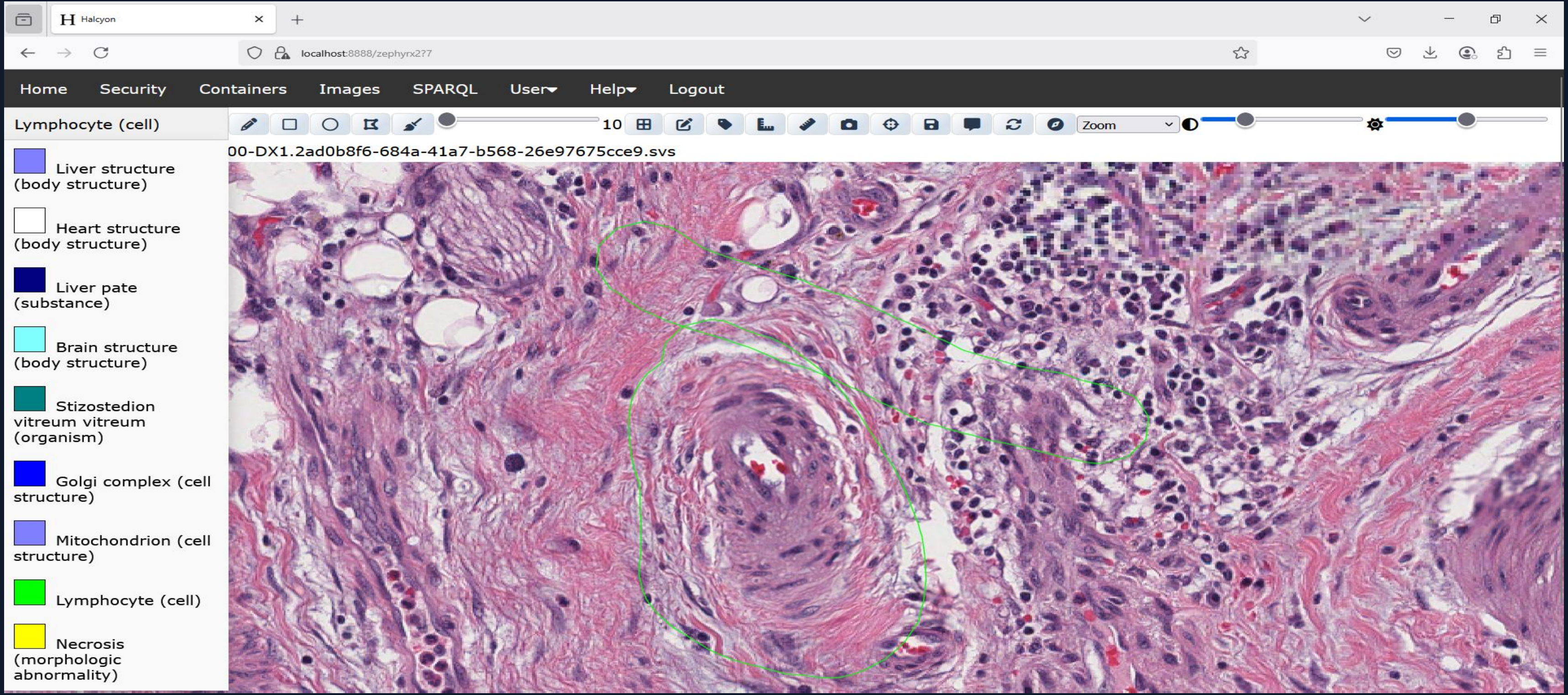
# RUNNING SYSTEM · COMPARISON



*Multiple slides, same viewport, synchronized pan & zoom.*

ZEPHYR · VIEWING

# Tissue at low magnification

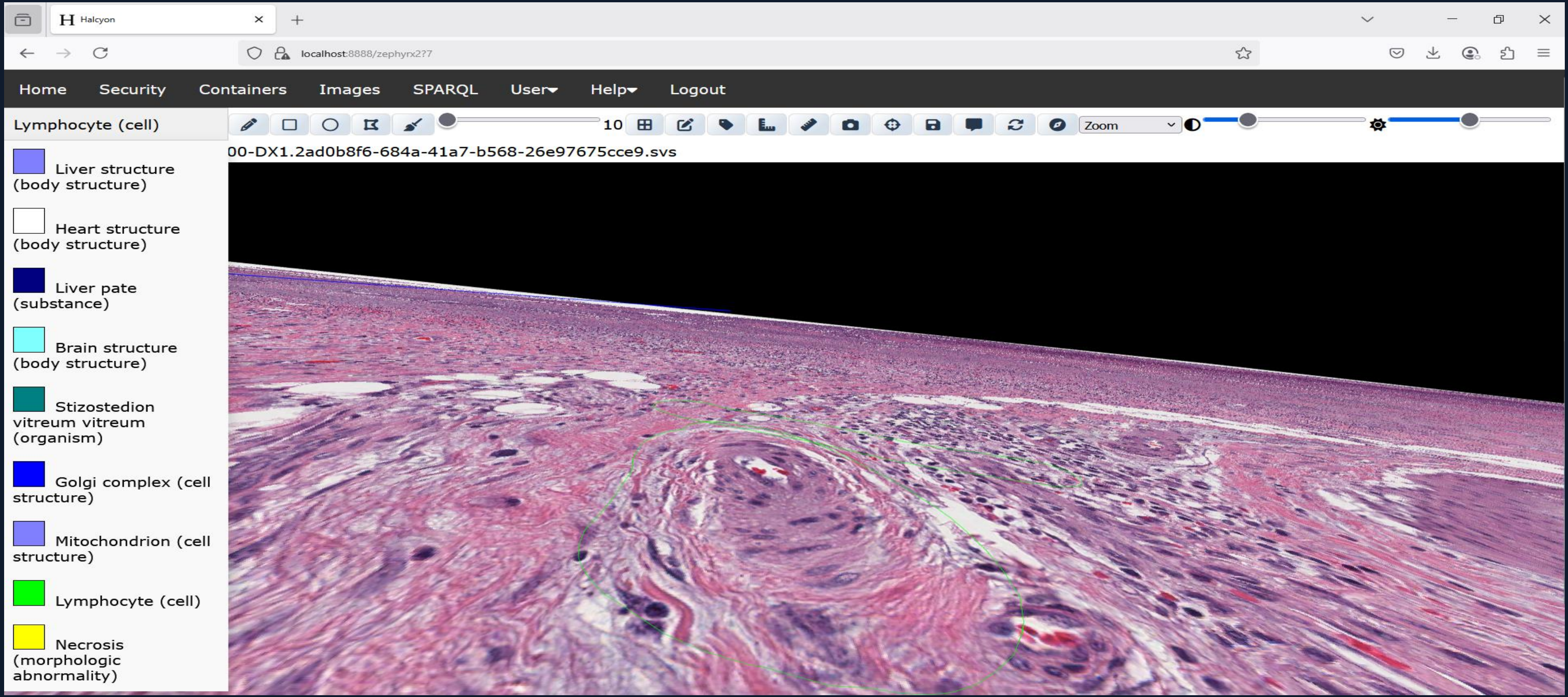


00-DX1.2ad0b8f6-684a-41a7-b568-26e97675cce9.svs

*Annotation class palette on the left is generated directly from SHACL shapes.*

ZEPHYR · ZOOMING

# IIIF — streamed tile pyramid



Lymphocyte (cell)

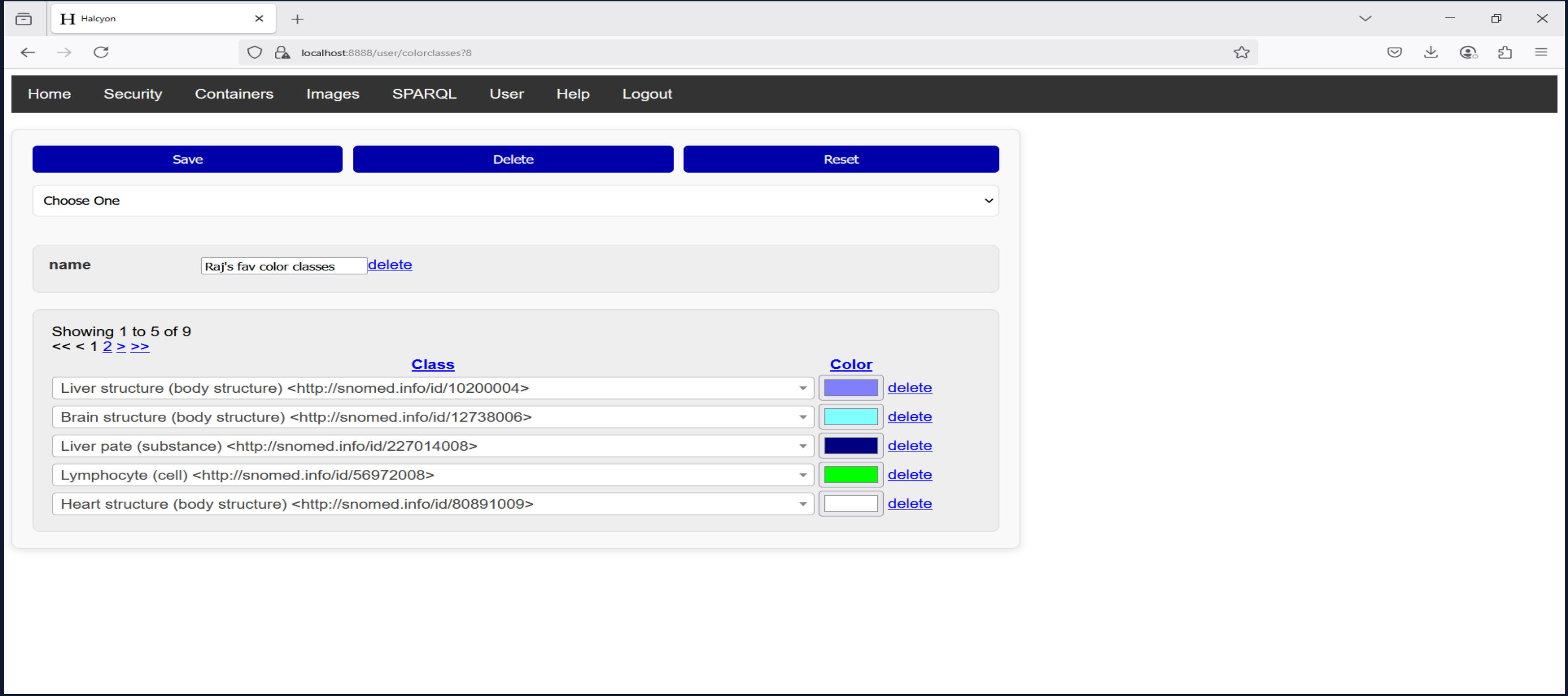
00-DX1.2ad0b8f6-684a-41a7-b568-26e97675cce9.svs

- 🟪 Liver structure (body structure)
- ⬜ Heart structure (body structure)
- 🟦 Liver pate (substance)
- 🟦 Brain structure (body structure)
- 🟦 Stizostedion vitreum vitreum (organism)
- 🟦 Golgi complex (cell structure)
- 🟪 Mitochondrion (cell structure)
- 🟩 Lymphocyte (cell)
- 🟨 Necrosis (morphologic abnormality)

*At 40× the viewer pulls only the tiles covering the visible region.*

ZEPHYR · CONFIGURATION

# Annotation class editor



**name** Raj's fav color classes delete

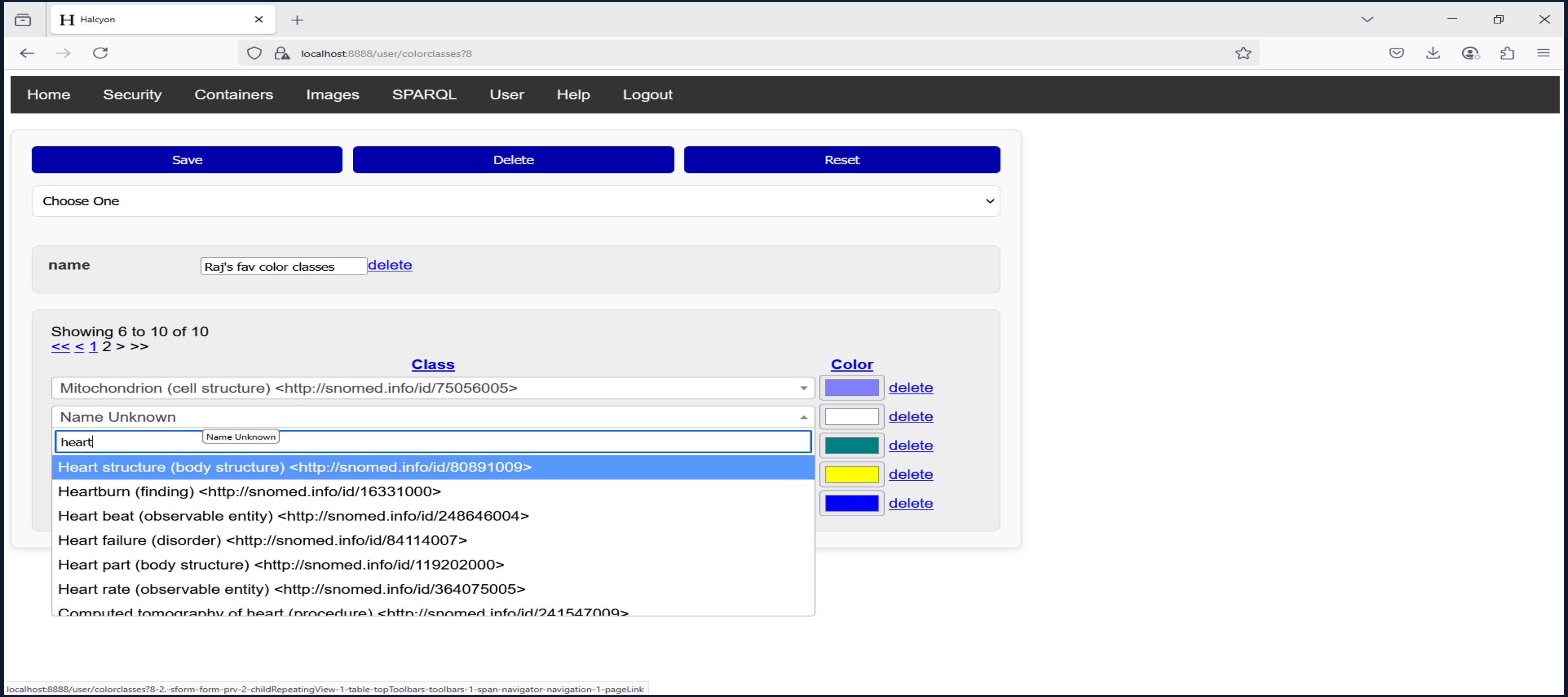
Showing 1 to 5 of 9
<< < 1 2 > >>

| Class | Color | |
|---|---|---|
| Liver structure (body structure) <http://snomed.info/id/10200004> | 🟪 | delete |
| Brain structure (body structure) <http://snomed.info/id/12738006> | 🟦 | delete |
| Liver pate (substance) <http://snomed.info/id/227014008> | ■ | delete |
| Lymphocyte (cell) <http://snomed.info/id/56972008> | 🟩 | delete |
| Heart structure (body structure) <http://snomed.info/id/80891009> | ⬜ | delete |

*Create, edit, and color-code the classes researchers will draw.*

ZEPHYR · CONFIGURATION

# Annotation classes with SNOMED bindings



**name** Raj's fav color classes delete

Showing 6 to 10 of 10
<< < 1 2 > >>

| Class | Color | |
|---|---|---|
| Mitochondrion (cell structure) <http://snomed.info/id/75056005> | 🟪 | delete |
| Name Unknown | ⬜ | delete |
| | 🟩 | delete |
| | 🟨 | delete |
| | 🟦 | delete |

heart

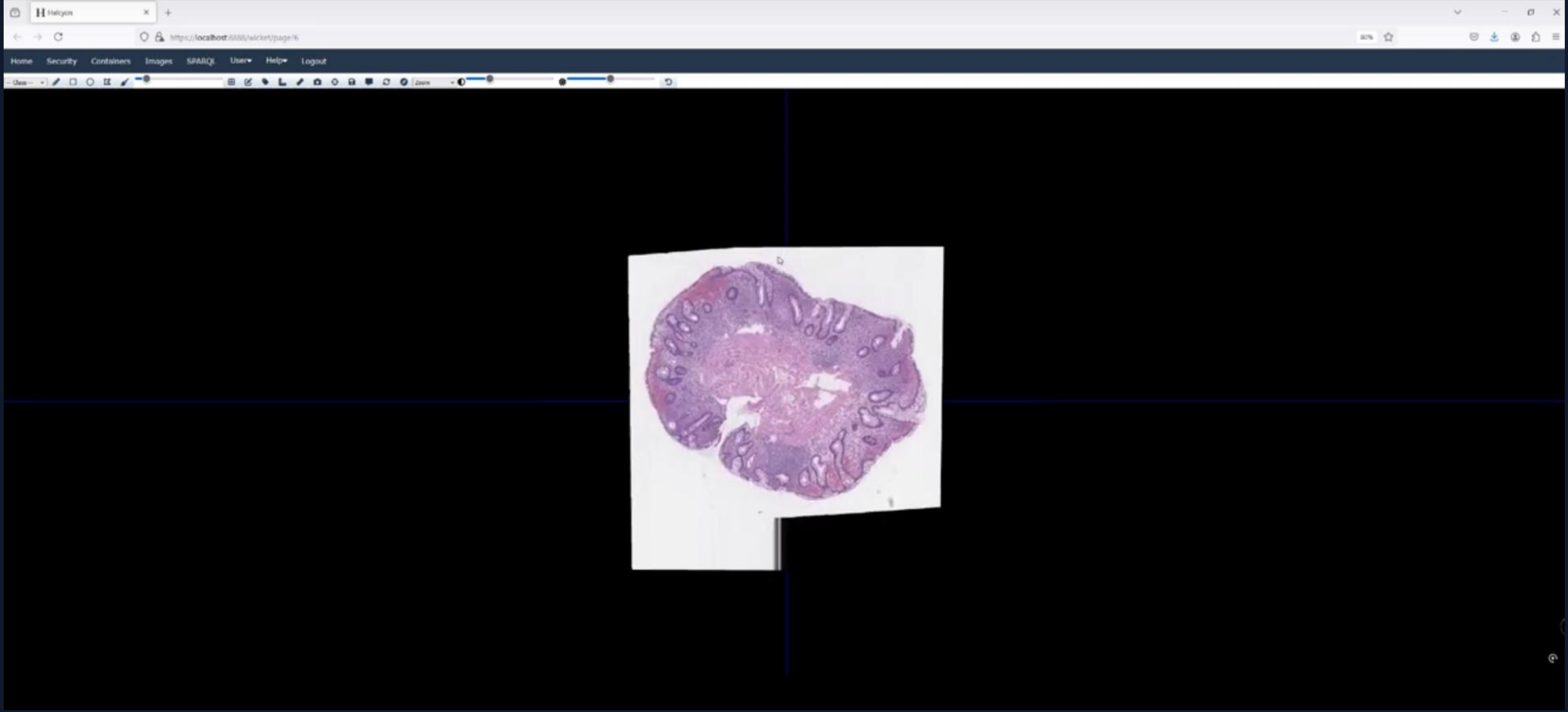
Name Unknown

- Heart structure (body structure) <http://snomed.info/id/80891009>
- Heartburn (finding) <http://snomed.info/id/16331000>
- Heart beat (observable entity) <http://snomed.info/id/248646004>
- Heart failure (disorder) <http://snomed.info/id/84114007>
- Heart part (body structure) <http://snomed.info/id/119202000>
- Heart rate (observable entity) <http://snomed.info/id/364075005>
- Computed tomography of heart (procedure) <http://snomed.info/id/241547009>

localhost:8888/user/colorclasses?8-2.-sform-form-prv-2-childRepeatingView-1-table-topToolbars-toolbars-1-span-navigator-navigation-1-pageLink

*Each class is bound to a SNOMED CT concept — annotations become semantic.*

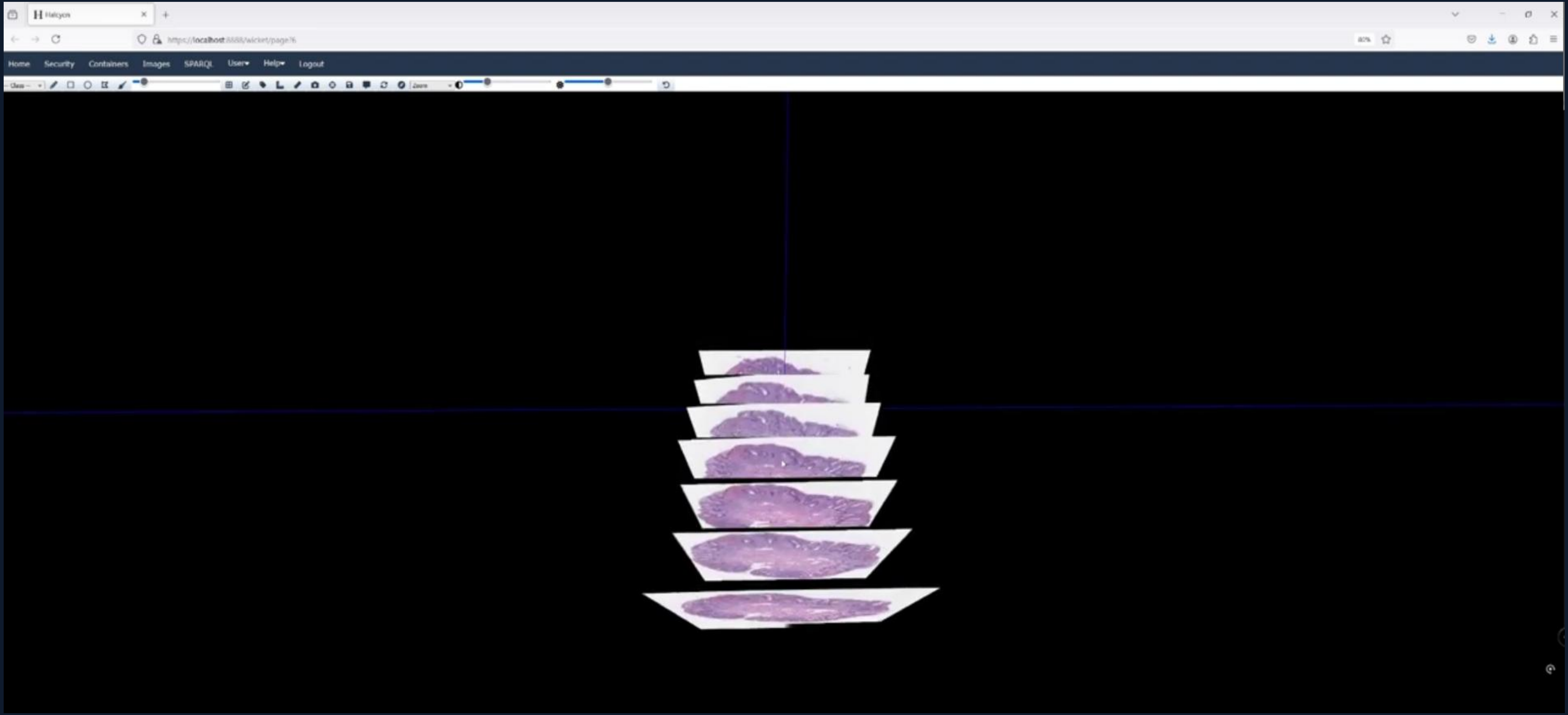
# Slide stack — single tissue level



*Each slide is a 2-D surface in a 3-D scene — enabling serial-section alignment.*

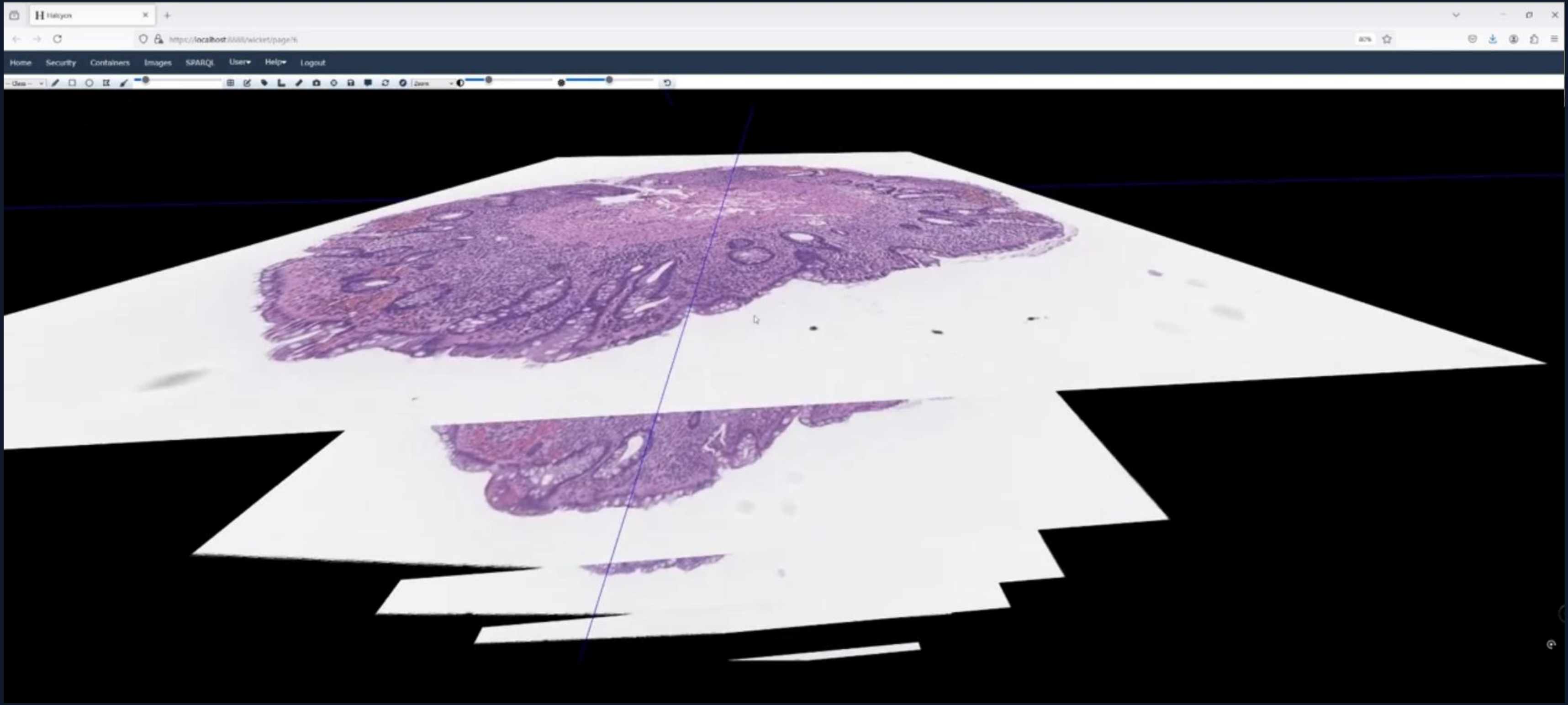
ZEPHYR · 3-D COMPOSITION

# Multi-level stack



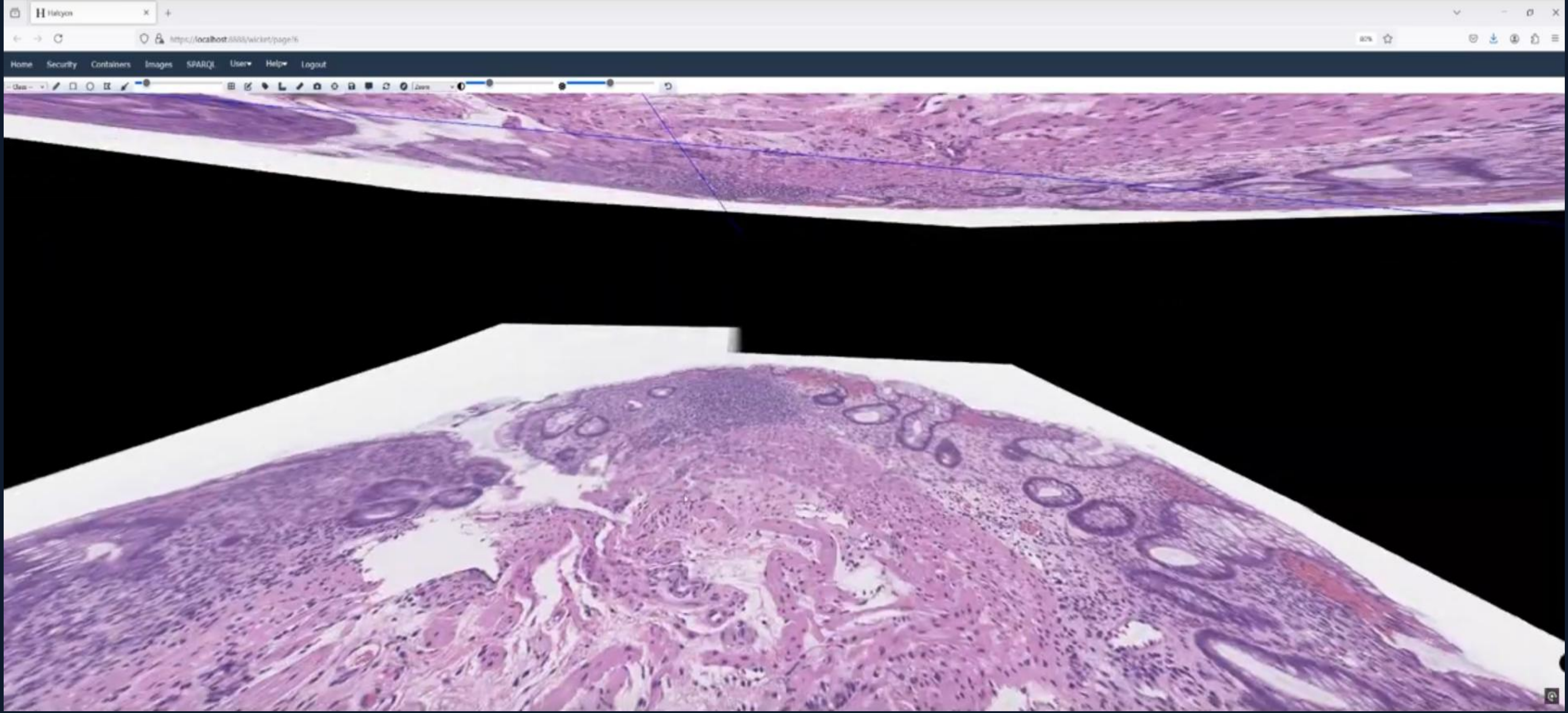
*Browse up and down through the stack — correlate features across sections.*

# Angled view of the stack



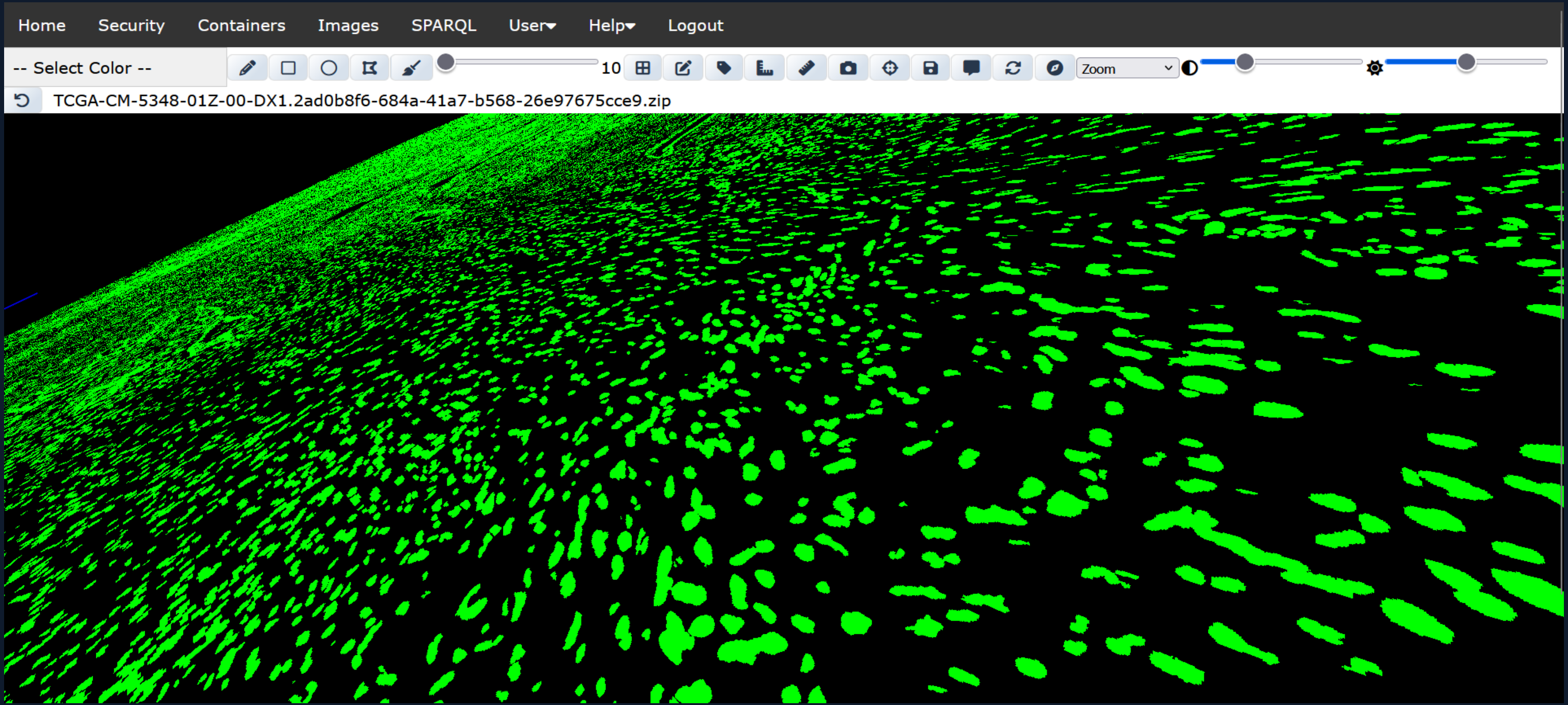
*All native WebGL — no plugins, no downloads.*

# Annotations rendered over the stack



*Every drawn shape is written back to the graph — queryable by SPARQL.*

# 1.5M nuclei rendered as an overlay



TCGA-CM-5348-01Z-00-DX1.2ad0b8f6-684a-41a7-b568-26e97675cce9.zip

*Each green shape is one polygon.*

TOOLING

# dcm2rdf

*Convert DICOM metadata to RDF — losslessly.*

| | |
|---|---|
| ● **LOSSLESS** | Every DICOM tag becomes an RDF property — nothing is discarded. |
| ● **ROUND-TRIPPABLE** | The RDF is being designed faithfully to DICOM spec to promote round-tripping |
| ● **QUERYABLE** | SPARQL across thousands of studies without opening a single DCM file. |
| ● **OPEN SOURCE** | Apache 2.0 licensed, community-developed, runs on any JVM. |

**GITHUB** github.com/ebremer/dcm2rdf

DCM2RDF · MAPPING

# DICOM tags become first-class RDF.

**DICOM TAG**

```
(0008,0018)
SOPInstanceUID
1.3.6.1.4.1.5962.99...
```

**RDF TRIPLE**

```
<urn:oid:...>
  dcm:00080018
    <urn:oid:...> .
```

**QUERY**

```
SELECT ?s WHERE {
  ?s dcm:00080060
     "CT" .
}
```

*Full worked examples and the source code at github.com/ebremer/dcm2rdf*

USE CASE · AN INTRODUCTION

# The Cancer Imaging Archive

**TCIA**

A public, open-access repository of de-identified cancer imaging freely available to researchers worldwide.
Principal Investigator: Fred Prior

**30.9M+**
de-identified images

**37,500+**
subjects across collections

**2011**
launched by the NCI

cancerimagingarchive.net

*Funded by NCI · NIH*

THE WORK

# What TCIA does, and who runs it

WHO THEY ARE

## A public service of the NCI

Funded by the National Cancer Institute (Cancer Imaging Program, DCTD).

Operated by the University of Arkansas for Medical Sciences since 2015.

Managed day-to-day by the Cancer Imaging Informatics Lab at FNLCR.

**DE-IDENTIFY & CURATE**
Rigorously strip PHI from incoming DICOM data to meet HIPAA and NIH sharing policies.

**HOST & ORGANIZE**
Maintain collections of CT, MRI, PET, and digital pathology - grouped by cancer type.

**ENABLE OPEN ACCESS**
Free web and API downloads, with persistent DOIs for exact dataset citation.

**SUPPORT NIH PROGRAMS**

Power QIN, CPTAC, APOLLO, and public imaging-analysis challenges.

BUILDING IT

# A pod-native research layer

*W3C Linked Web Storage / SOLID separates where data lives from who controls it — exactly the split TCIA researchers need.*

CANONICAL SOURCE

### TCIA Archive

Stays the authoritative host of curated images, DOIs, and collection metadata. One place, one truth.

OPEN PROTOCOL

### Linked Web Storage

Standardized identity, access control, and URIs — so images, annotations, and models cite each other by ID.

RESEARCHER OWNS

### Your Pod

Annotations, segmentations, notes, and model outputs live in the researcher's own storage — revocable, portable, theirs.

**Annotate in place** — No download-edit-reupload. Link to exact DICOM URIs.

**Share on your terms** — Grant a lab, a collaborator, or the public — revoke anytime.

**Submit from your pod** — Draft collections privately, let TCIA pull when ready.

AGENTS IN THE LOOP

# Agents that work for researchers

***MCP servers + reusable Skills*** *give AI agents a standard way to reach TCIA and a researcher's pod — without either side giving up control.*

### AI Agent

*Acts on behalf of the researcher*

### Skills + MCP

*Bundles the know-how, wires it in*

### TCIA MCP · Pod MCP

*Read from archive, write to pod*

**Research workflows, one prompt away**

"Fetch the LIDC collection, run my segmentation, save masks to my pod." The agent resolves URIs across both systems.

**Access stays with the owner**

The pod enforces permissions. Grant, scope, or revoke any tool's reach at any time.

**Curation-assisted submissions**

Agents pre-scan DICOM headers for PHI, suggest labels, and draft metadata before TCIA intake.

*TCIA stays the trusted archive. · Researchers keep the keys. · Agents do the legwork.*

# THANK YOU

### Stony Brook University
Tammy DiPrima
Joseph Balsamo
Feiqiao Wang
Raj Gupta
Joel Saltz

### FHIR RDF Group
David Booth
Eric Prud'hommeaux
Jim Balhoff
Gaurav Vaidya
Detlef Grittner (sohard.de)

### TCIA
Fred Prior
Dwayne Dobbins
Justin Whorton
Jeff Tobler
Kirk Smith
Lawrence Tarbox
Michael Rutherford
Natasha Honomichl
Jeremy Jarosz
Tracey Nolan

### Rensselaer Polytechnic Institute
Andrew White
Jonas Braasch

### NIH/NCI
Jonas Almeida

### Emory University
Ryan Birmingham
Tony Pan
Ashish Sharma

THANK YOU

# *Questions?*

**ERICH BREMER**

*Director for Applied Informatics · Stony Brook University*

**EMAIL**
erich.bremer@stonybrook.edu

**GITHUB**
github.com/ebremer