The INCF Digital Atlasing Infrastructure

Ilya Zaslavsky (UCSD)
and INCF Digital Atlasing Infrastructure Task Force

Joint MRC/INCF/SICSA Workshop on Atlas Informatics
Edinburgh, 15-16 May 2012

Oversight Committee

> Chair: **Robert W. Williams**, University of Tennessee, Memphis, USA
> Richard Baldock, Medical Research Council and University of Edinburgh, United Kingdom
> **Jan Bjaalie**, INCF and University of Oslo, Norway
> Gregor Eichele, Max Planck Institute, Göttingen, Germany
> **Mark Ellisman**, INCF and University of California, San Diego, USA
> Tei-ichi Furuichi, RIKEN Brain Science Institute, Wako, Japan
> **Michael Hawrylycz**, Allen Institute for Brain Research, Seattle, USA
> **G. Allan Johnson**, Duke University, Durham, USA
> **Maryann Martone**, University of California, San Diego, USA
> **Alan Ruttenberg**, Creative Commons and University of Buffalo, USA
> Arthur Toga, University of California, Los Angeles, USA

Waxholm Space Task Force

> Lead: **Michael Hawrylycz**, Allen Institute for Brain Science, Seattle, USA
> Alexandra Badea, Duke University, Durham, USA
> **Albert Burger**, MRC Human Genetics Unit and Heriot-Watt University, Edinburgh, United Kingdom
> **James Gee**, University of Pennsylvania, Philadelphia, USA
> Christian Haselgrove, University of Massachusetts Medical Center, Worcester, USA
> **Luis Ibáñez**, Kitware Inc., Clifton Park, NY, USA
> **G. Allan Johnson**, Duke Center for In Vivo Microscopy, Durham, USA
> Stephen Larson, University of California, San Diego, USA
> Piotr Majka, Nencki Institute of Experimental Biology, Warsaw, Poland
> **Lydia Ng**, Allen Institute for Brain Science, Seattle, USA
> Yuko Okamura-Oho, RIKEN Advanced Science Institute, Wako, Japan
> Seth Ruffins, Biological Imaging Center, California Institute of Technology, Pasadena, USA
> Marina Sergejeva, Freidrich-Alexander Universität Erlangen-Nürnberg, Germany

Digital Atlasing Infrastructure Task Force

> Lead: **Ilya Zaslavsky**, San Diego Supercomputer Center, University of California, San Diego, USA
> **Rembrandt Bakker**, Donders Institute for Brain, Cognition and Behavior, Radboud UMC Nijmegen, Netherlands
> **Albert Burger**, MRC Human Genetics Unit and Heriot-Watt University, Edinburgh, United Kingdom
> Christian Haselgrove, University of Massachusetts Medical Center, Worcester, USA
> **Michael Hawrylycz**, Allen Institute for Brain Science, Seattle, USA
> **Andreas Hess**, Freidrich-Alexander Universität Erlangen-Nürnberg, Germany
> Daren Lee, Laboratory for Neuro-Imaging (LONI), University of California, Los Angeles, USA
> Stephen Larson, University of California, San Diego, USA
> **Lydia Ng**, Informatics Group, Allen Institute for Brain Science, Seattle, USA
> Seth Ruffins, Biological Imaging Center, California Institute of Technology, Pasadena, USA
> **Fons Verbeek**, Leiden Institute of Advanced Computer Science, Leiden University, Netherlands

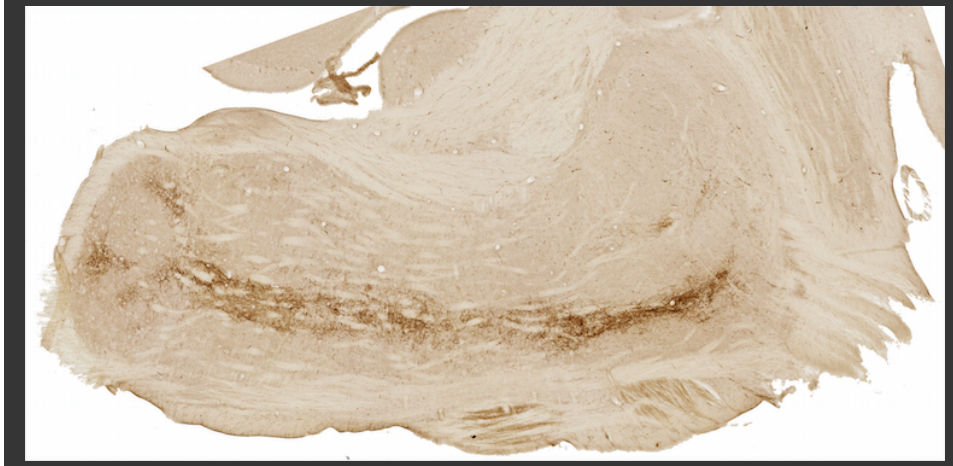
Glenn Rosen
Ken Sugino
Yoonsuck Choe

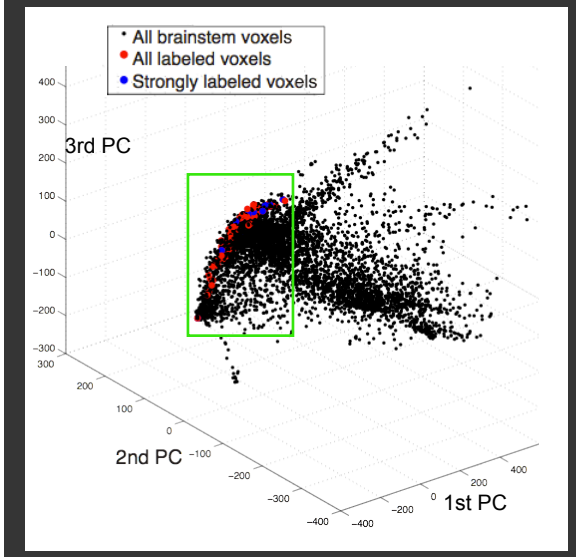
Eszter Papp
Raphael Ritz
Daniel Wojcik

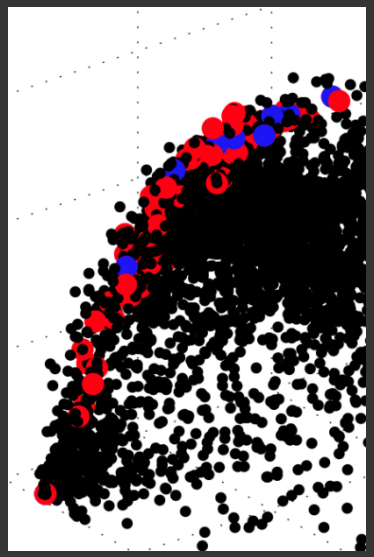
Asif Memon
Kenneth McLeod
Gang Song
Brian Avants
Steve Lamont

incf
Program on Digital Atlasing

The genetic organization of mouse trigeminal nuclei







A) Brainstem slice. B,C) Example of innervated voxels (red) sitting on a small manifold in hyper-gene space.

Does morphology or does innervation determine genetic similarity in brainstem?

We inject anatomical tracers to reveal brainstem inputs, co-register the images with AGEA, and use dimensionality reduction techniques to assess similarity.

Results:

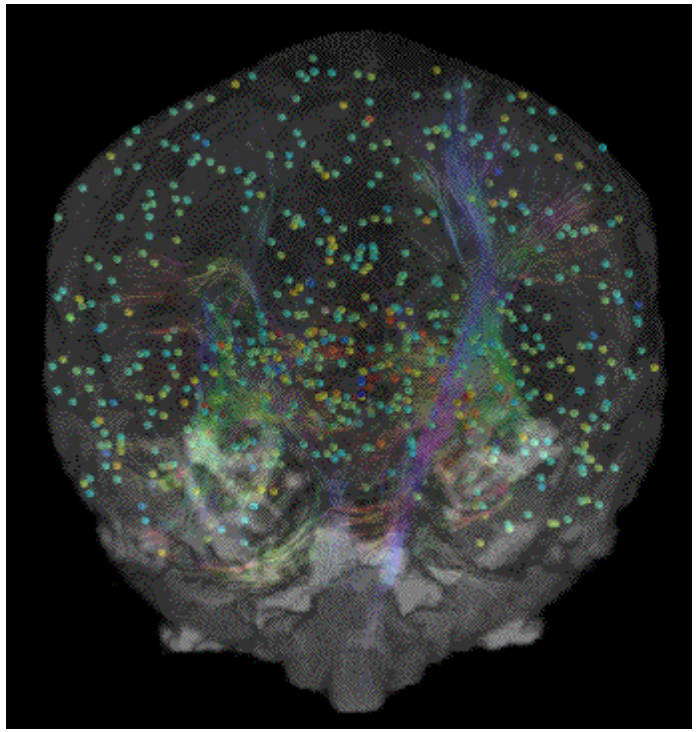
1. Subnuclei with similar inputs are more genetically similar than subnuclei determined histologically.

2. Many genes essential to brainstem organization have not been studied developmentally, making this a powerful informatics tool for gene discovery.

David W. Matthews, UCSD Computational Neurobiology, dm@ucsd.edu

What is an atlas?

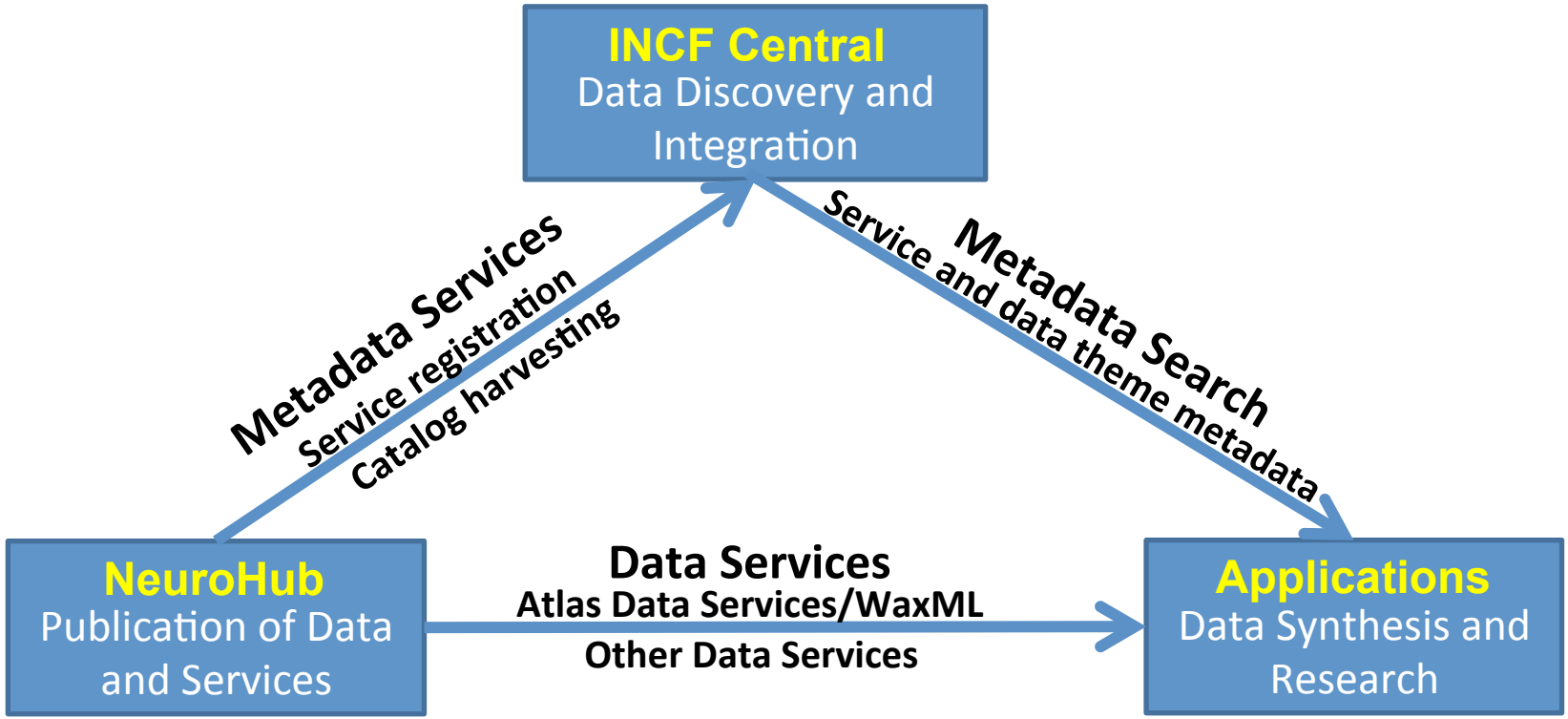
- A collection of 2D images or a 3D volume, possibly with anatomic feature delineations and a set of additional annotations



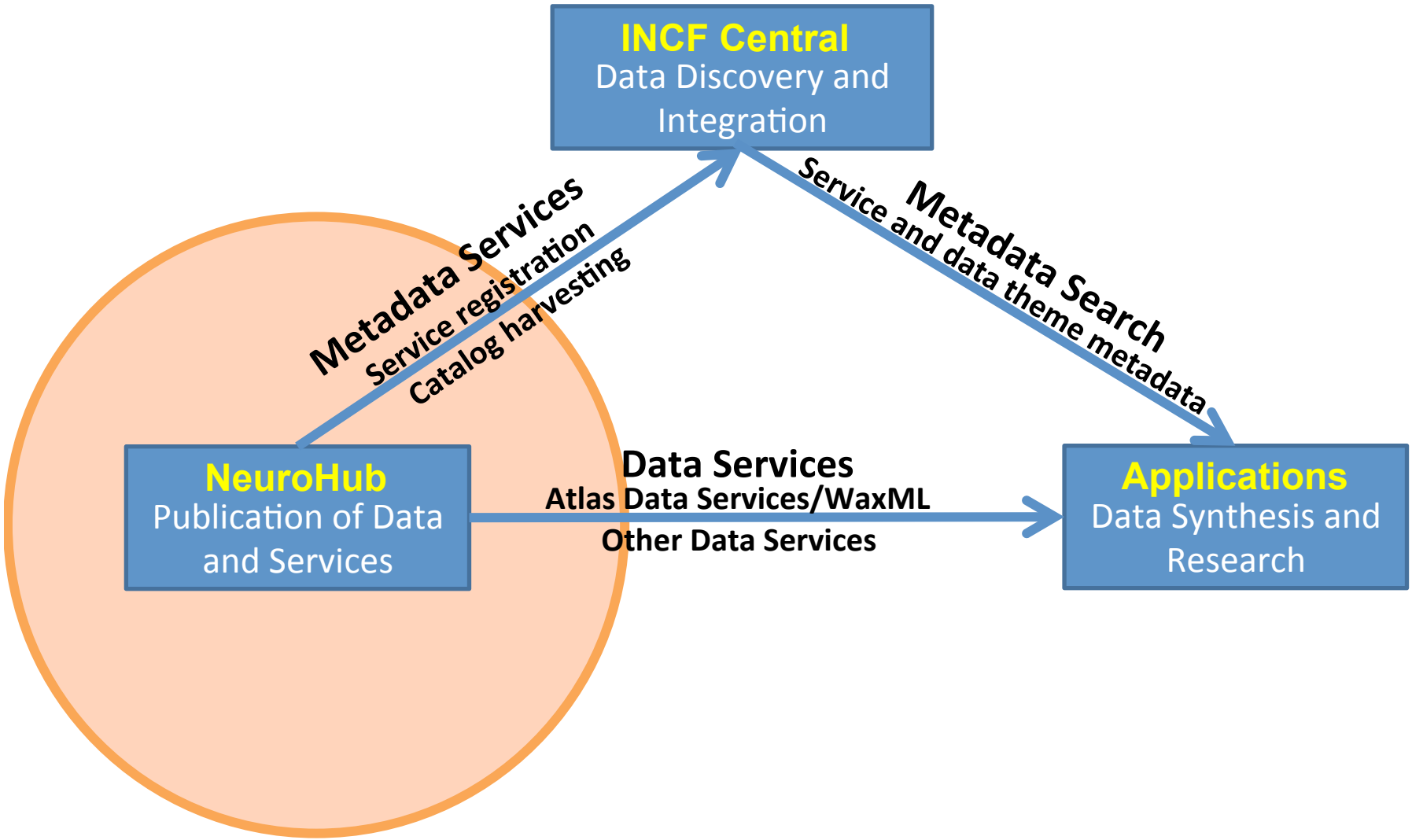
What we need is more than an atlas:

- A gateway to large distributed databases of images, volumes, segmentations, gene expression data, electrophysiology, behavioral, connectivity, other spatially-registered data
  - Ability to ask questions such as “which atlases have images for this part of the brain”, “what genes are expressed here in atlas A”, “compare spatial patterns of protein distribution across atlases”, etc.
- Collection of atlases organized as spatial data sources
- Collection of spatial data registries, service APIs and workflows:
  - image registration, segmentation, spatial selection, spatial analysis, integration of spatial data
- Collection of viewers , integration and annotation tools
- Ability to integrate user-supplied images, 3D volumes, and other resources

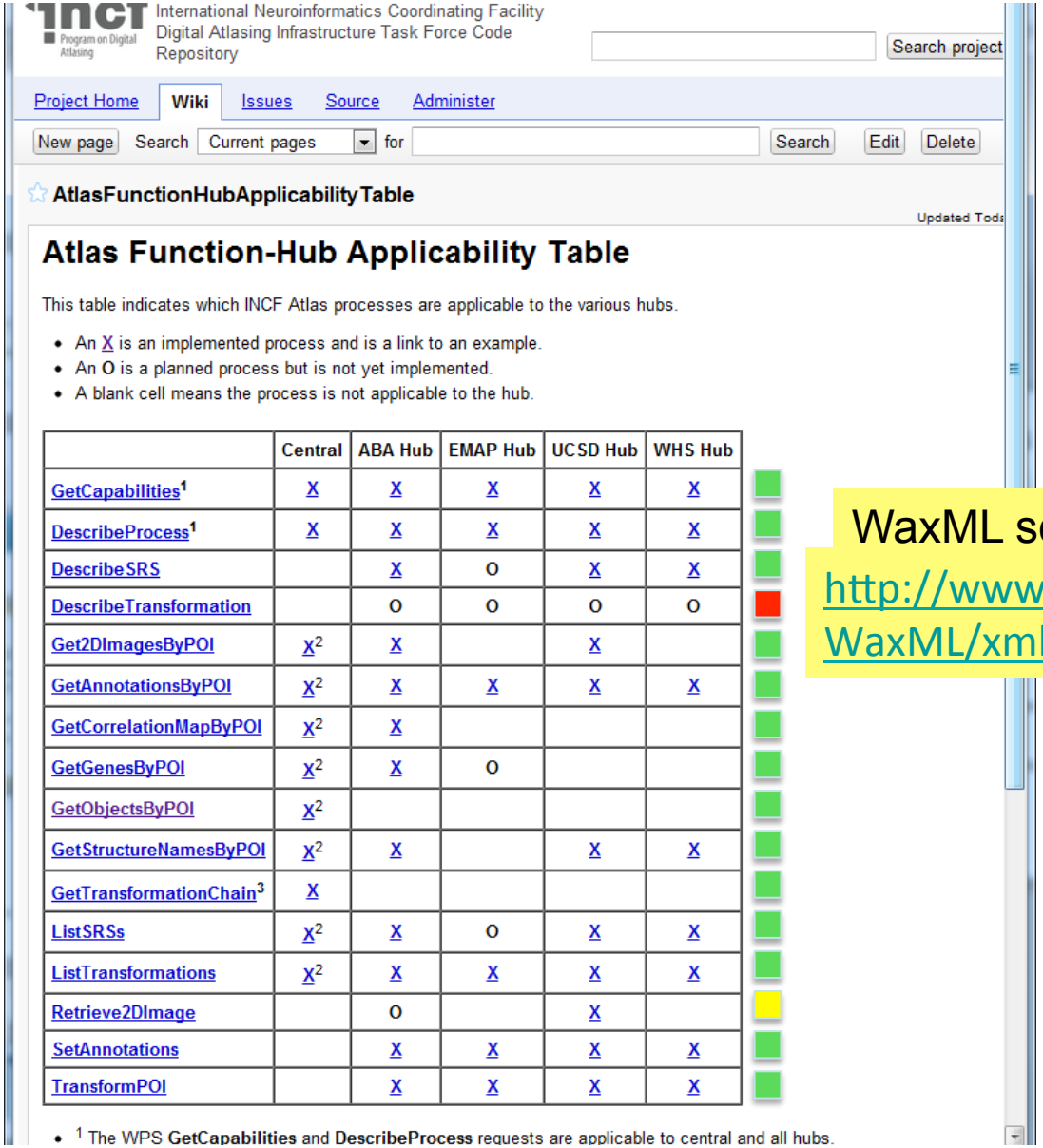
High-level INCF-DAI Design



Publishing Atlas Services



Waxholm Markup Language (WaxML) and standard atlas services



International Neuroinformatics Coordinating Facility
Digital Atlasing Infrastructure Task Force Code Repository

Search project

Project Home | Wiki | Issues | Source | Administer

New page | Search | Current pages | for | Search | Edit | Delete

AtlasFunctionHubApplicabilityTable

Updated Tod

Atlas Function-Hub Applicability Table

This table indicates which INCF Atlas processes are applicable to the various hubs.

- An X is an implemented process and is a link to an example.
- An O is a planned process but is not yet implemented.
- A blank cell means the process is not applicable to the hub.

	Central	ABA Hub	EMAP Hub	UCSD Hub	WHS Hub
GetCapabilities[1]	X	X	X	X	X
DescribeProcess[1]	X	X	X	X	X
DescribeSRS		X	O	X	X
DescribeTransformation		O	O	O	O
Get2DImagesByPOI	X[2]	X		X	
GetAnnotationsByPOI	X[2]	X	X	X	X
GetCorrelationMapByPOI	X[2]	X			
GetGenesByPOI	X[2]	X	O		
GetObjectsByPOI	X[2]				
GetStructureNamesByPOI	X[2]	X		X	X
GetTransformationChain[3]	X				
ListSRSs	X[2]	X	O	X	X
ListTransformations	X[2]	X	X	X	X
Retrieve2DImage		O		X	
SetAnnotations		X	X	X	X
TransformPOI		X	X	X	X

- [1] The WPS **GetCapabilities** and **DescribeProcess** requests are applicable to central and all hubs.

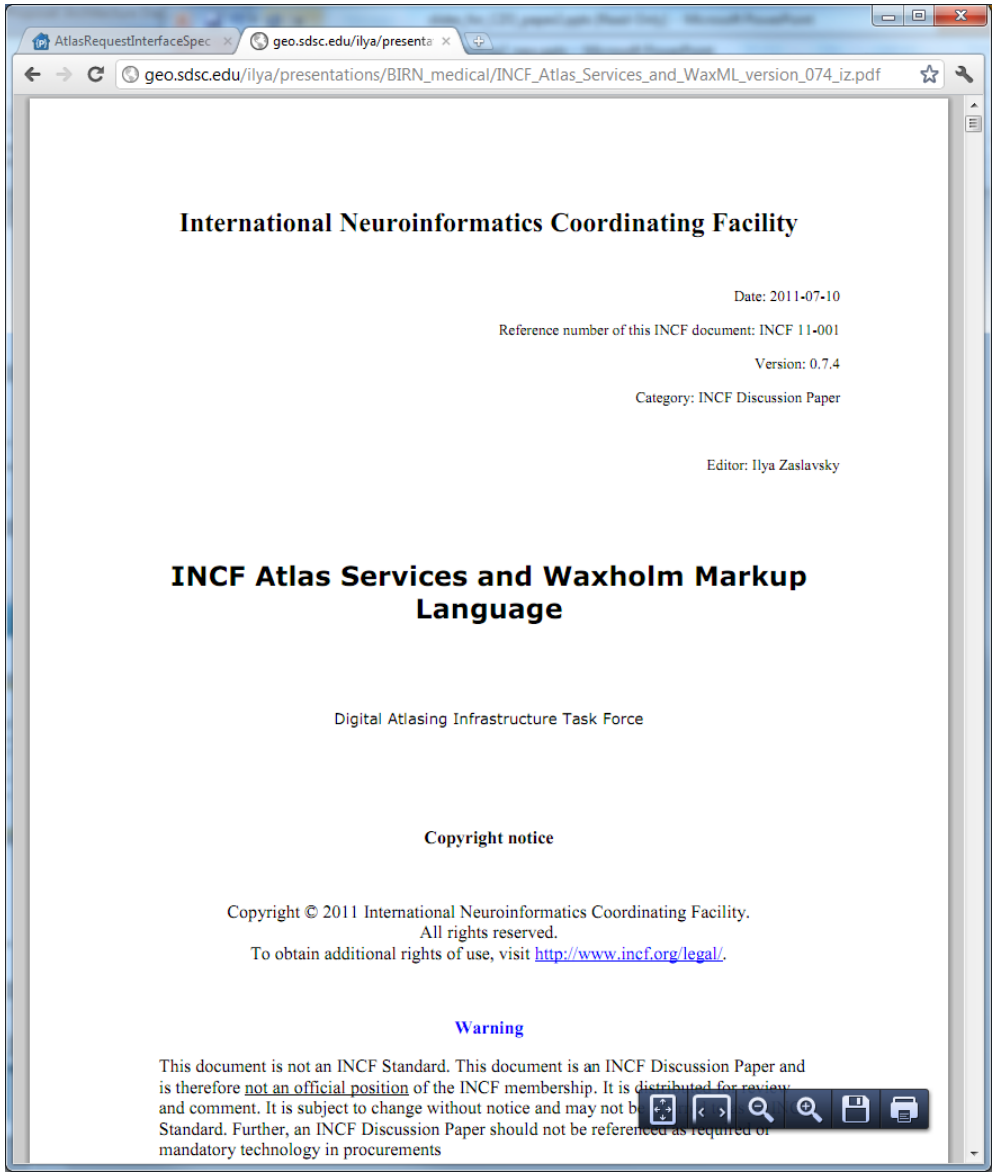
WaxML schema:
http://www.incf.org/WaxML/xmlschema

Waxholm Markup Language (WaxML): XML schema that provides standard constructs for atlas services.

The services are modeled after the **OGC WPS** interface specification

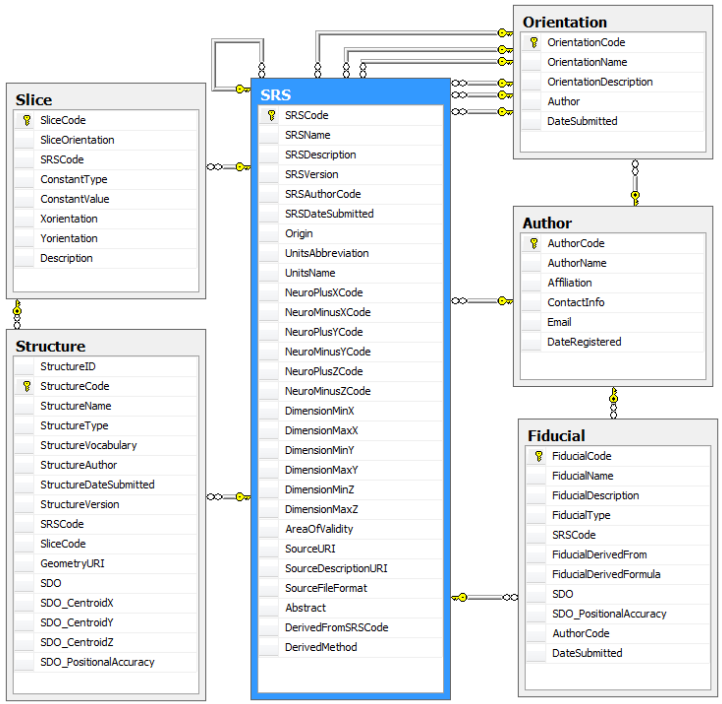
Specification

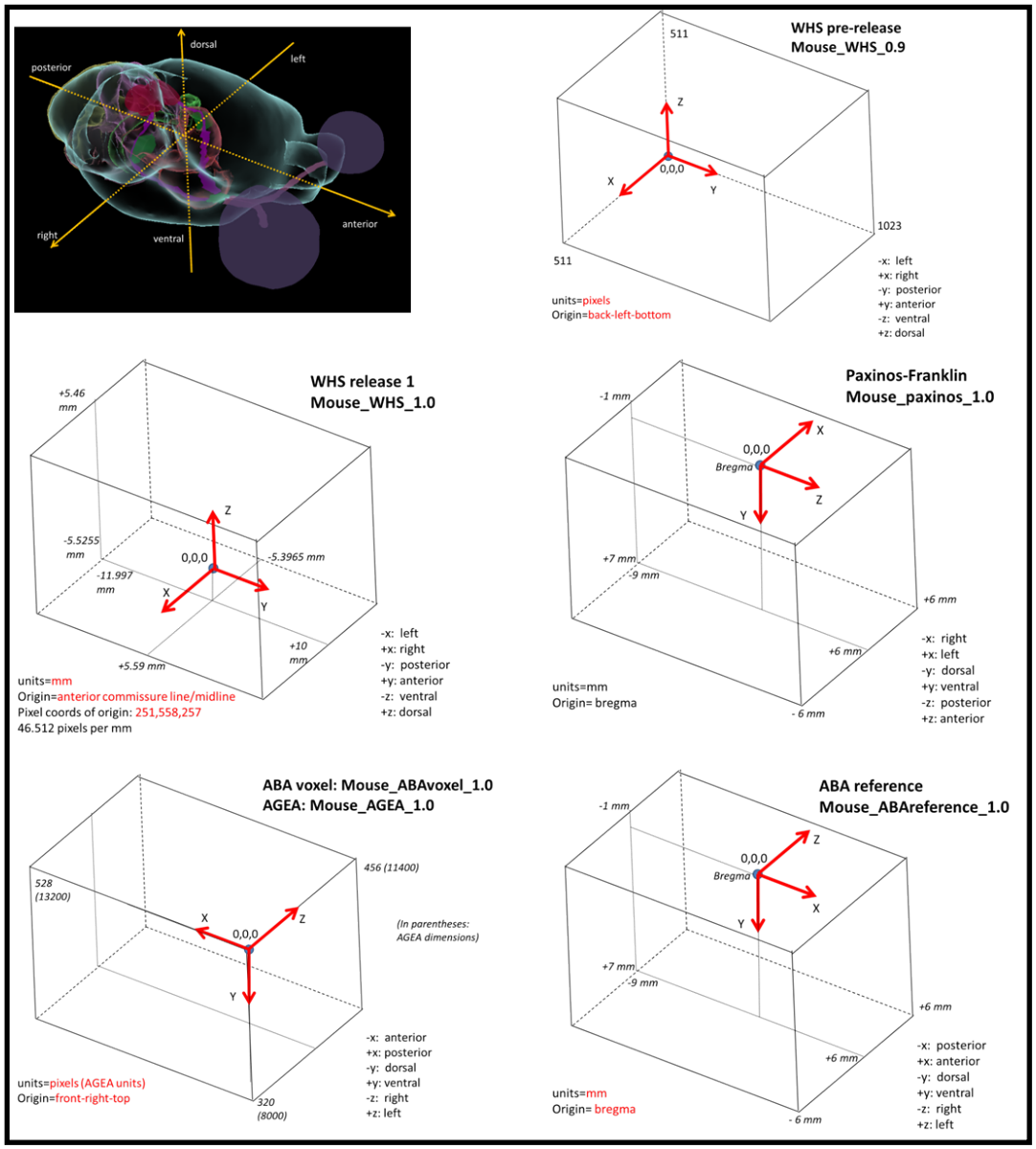
- Version 0.7.4: http://code.google.com/p/incf-dai/wiki/INCFProjectSpecification



Spatial reference systems and SRS registry

- Published by atlas hubs, summarized at
- INCF acts as a naming authority for SRS and transformations
- Accessed via *ListSRSs* and *DescribeSRS*

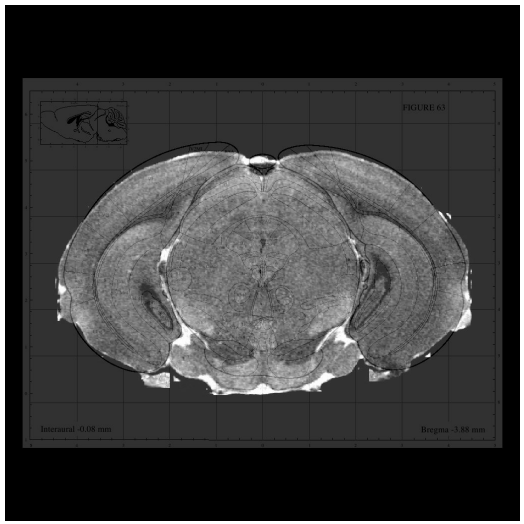


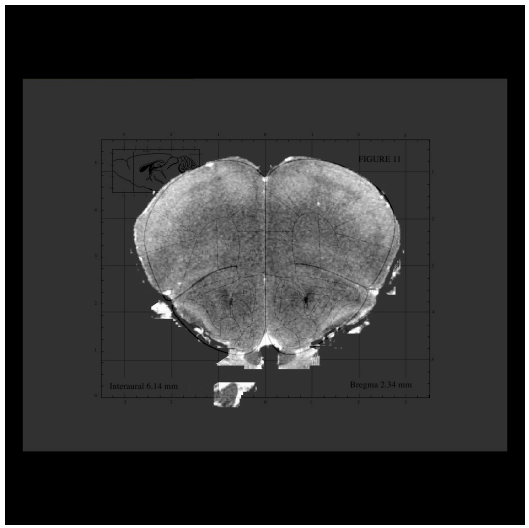


Coordinate Transformations: Different Types

- ABAvoxel to WHS
  - Large volumes for transforms in both directions, between 3D and 3D of different dimensions
  - Then simple lookup in the transformation volume
- ABAvoxel to ABAreference
  - Collection of conversion formulas for individual slices
- ABAvoxel to AGEA
  - Simple scaling
- WHS to Paxinos Mouse Atlas
  - Warping appropriate WHS cuts to match with Paxinos slices; translations per slice and brain region

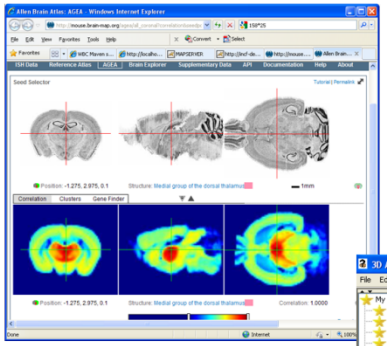
ListTransformations
GetTransformationChain
DescribeTransformation
TransformPOI





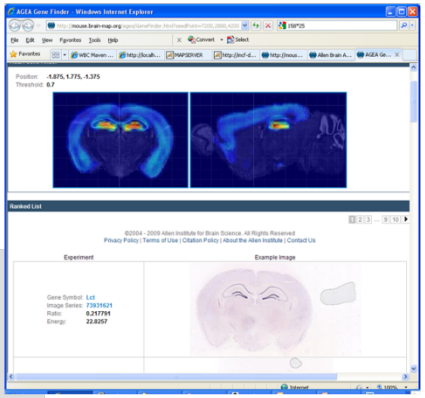
Evaluating and reporting accuracy is critical

Atlas services in action

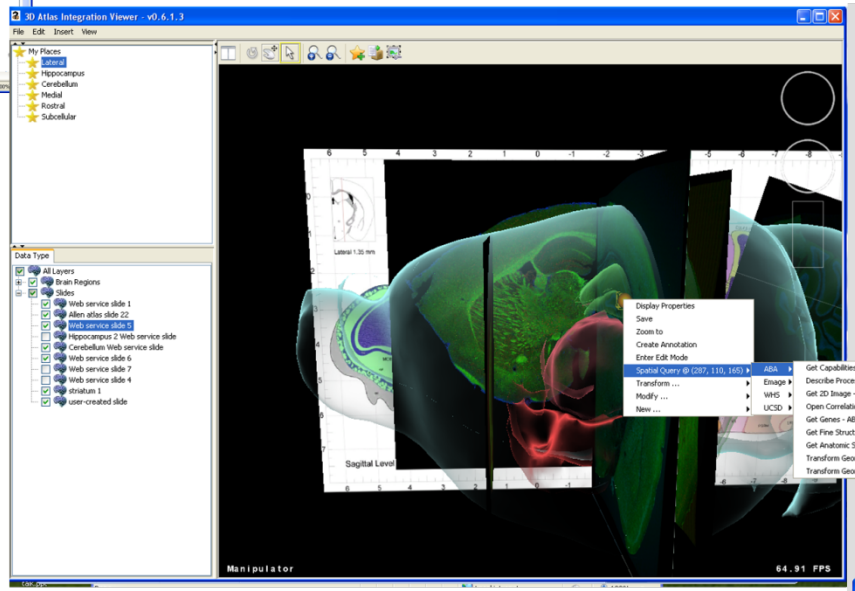


AGEA response to GetCorrelationMapByPOI

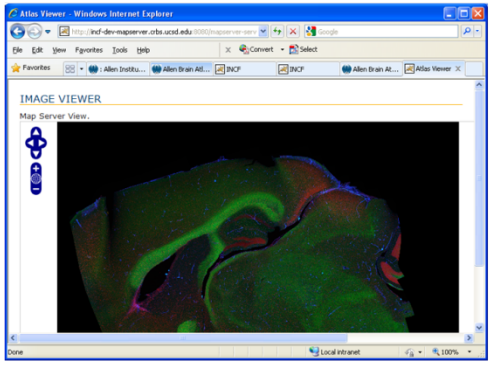
GetGenesByPOI returns gene information from AGEA

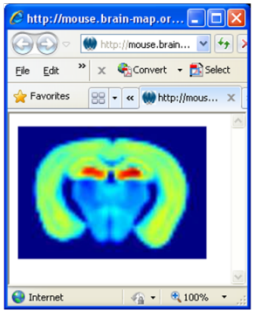


The 3D atlas integration client



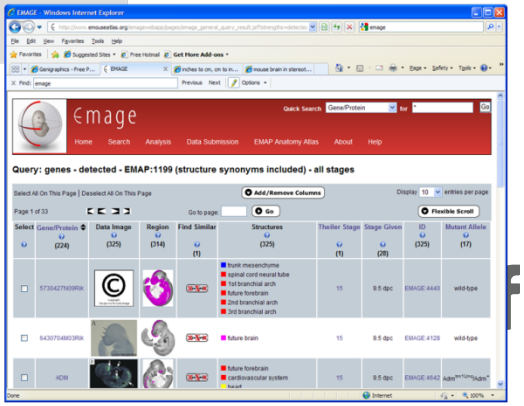
Get2DImageByPOI returns related images from UCSD hub



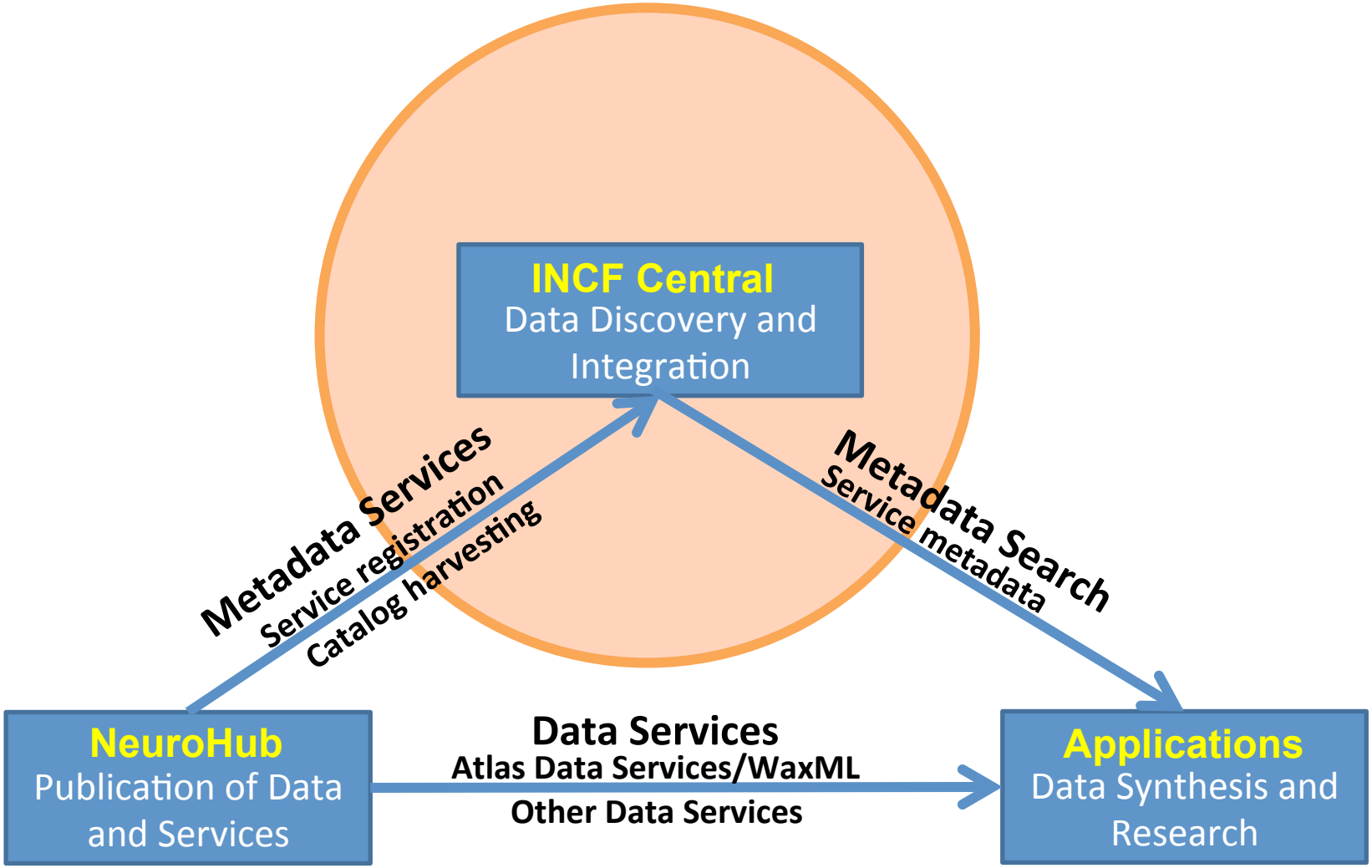


Get2DImageByPOI returns images from AGEA

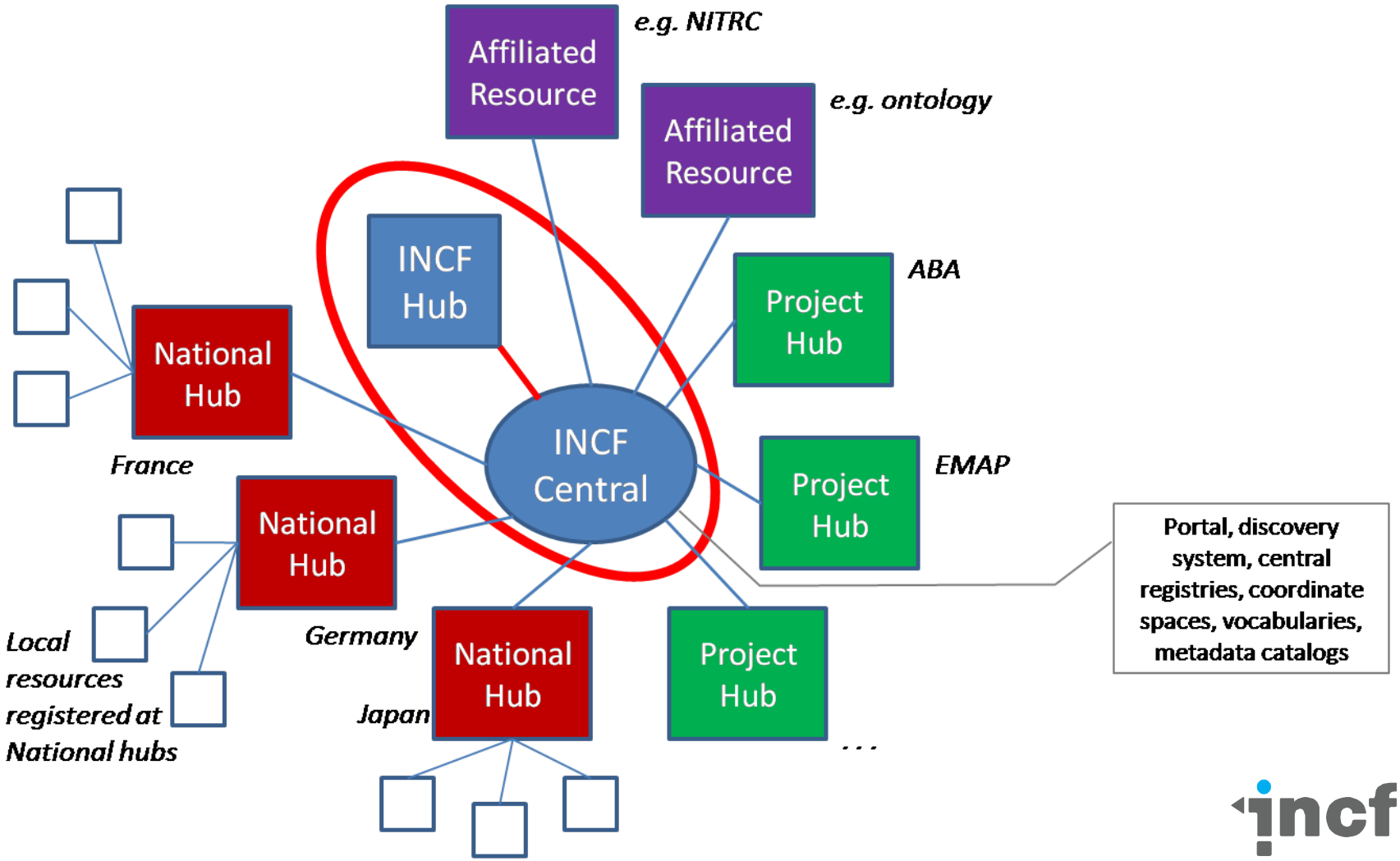
GetGenesByStructureName returns gene information from EMAGE



INCF-DAI Central Portal and Services

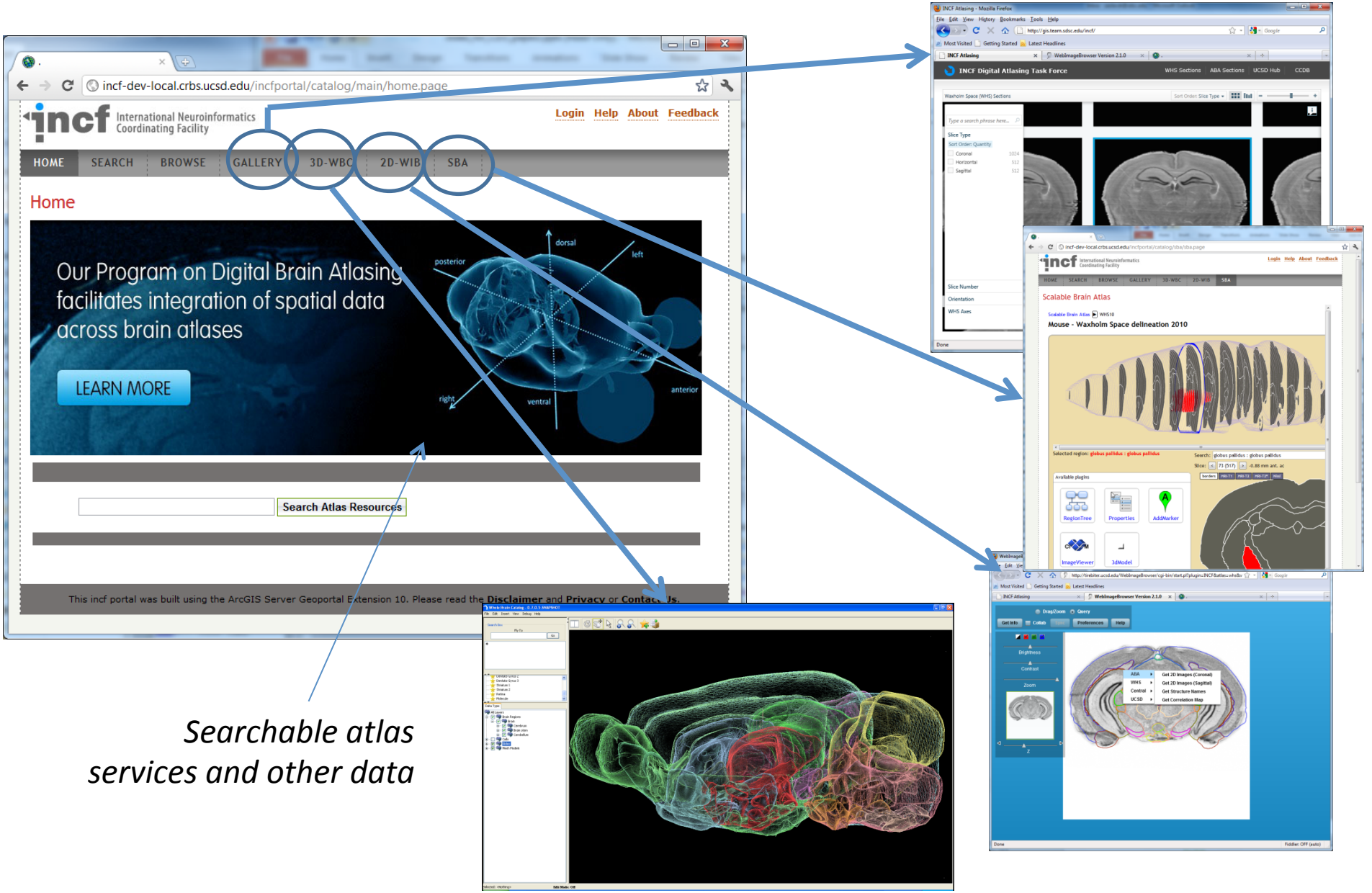


INCF hubs and INCF central



Atlas data catalog portal prototype

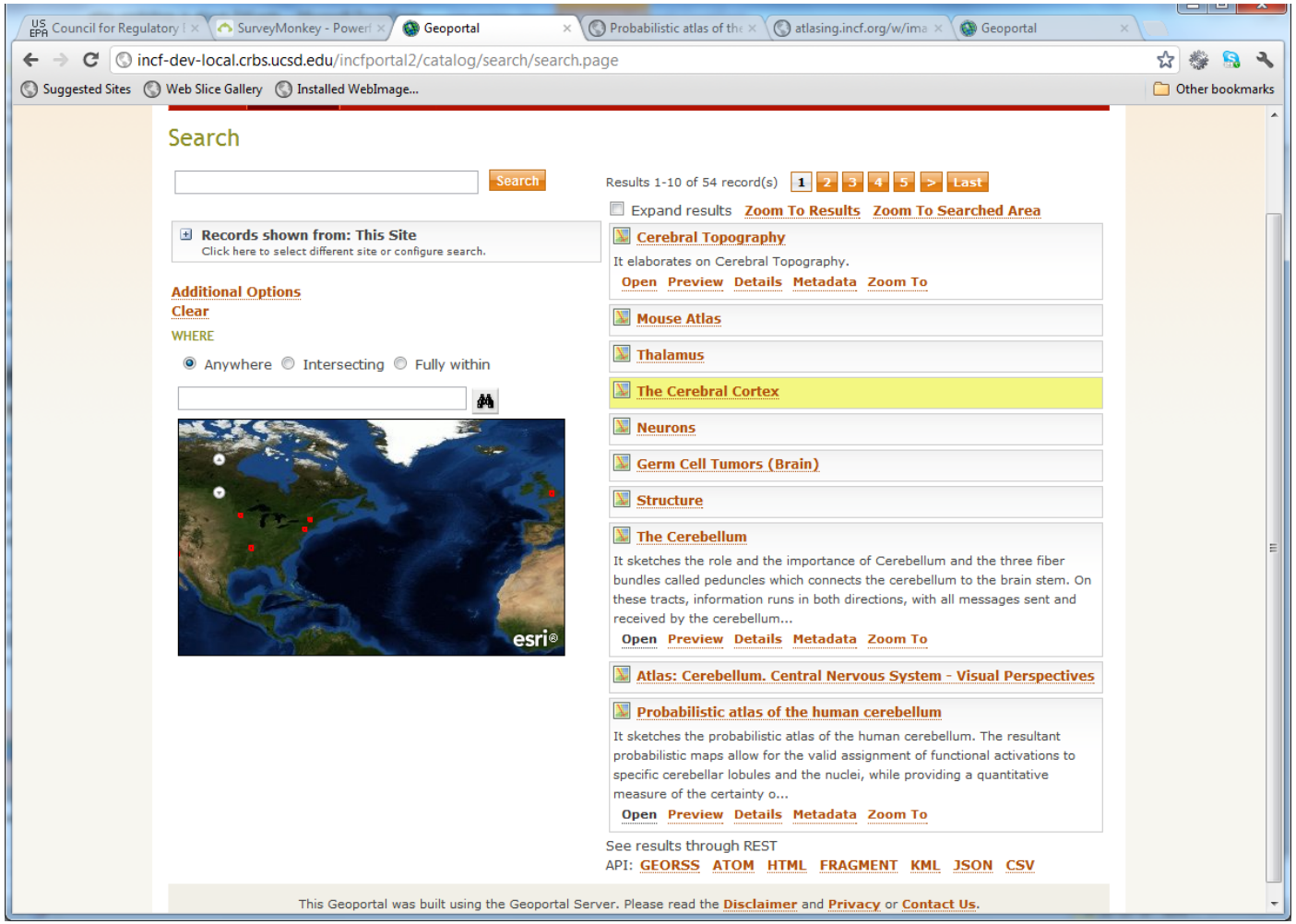
http://incf-dev-local.crbs.ucsd.edu/incfportal/



Searchable atlas services and other data

A companion portal prototype: organizes the same sources by locations of labs

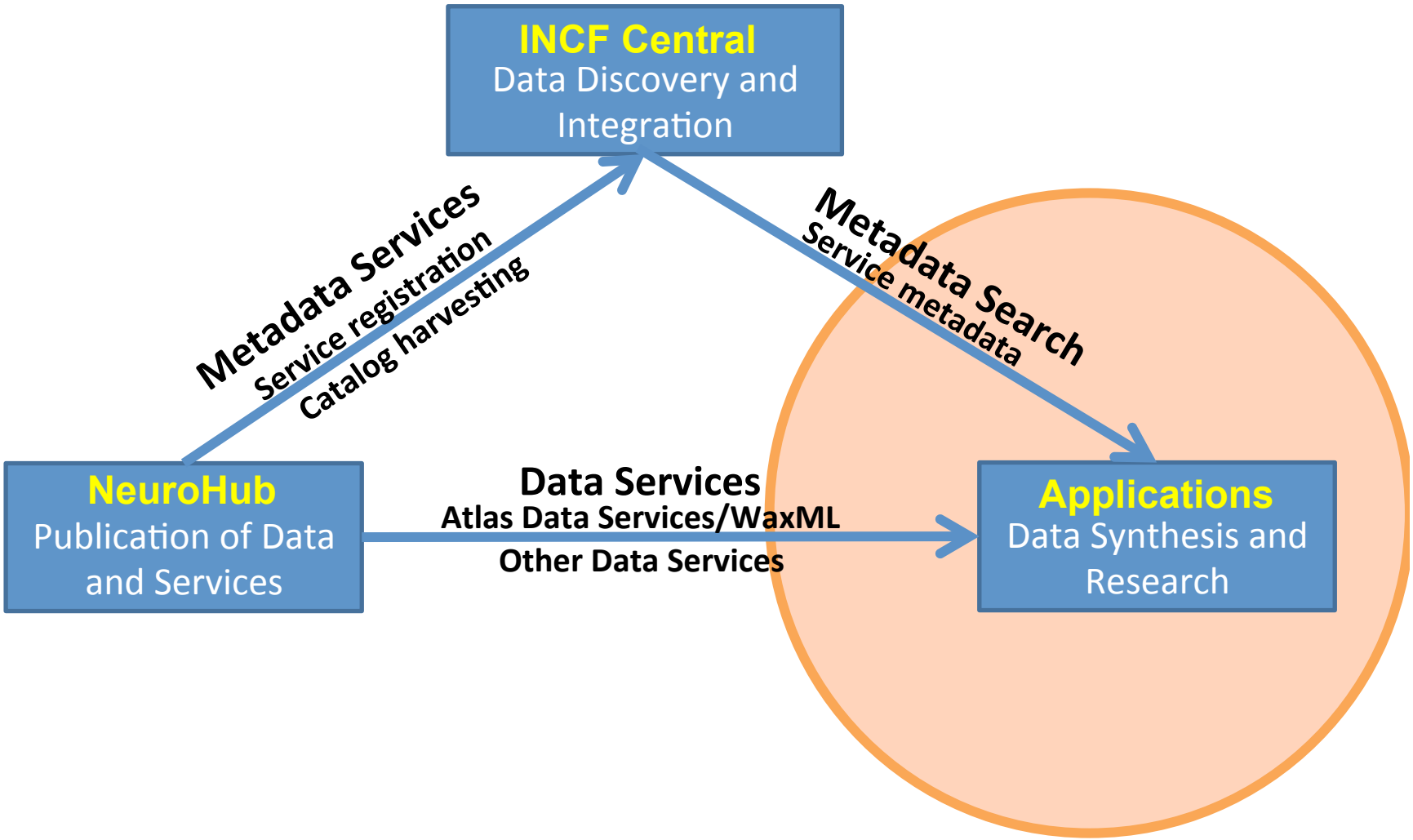
http://incf-dev-local.crbs.ucsd.edu/incfportal2/



Some features of the portals

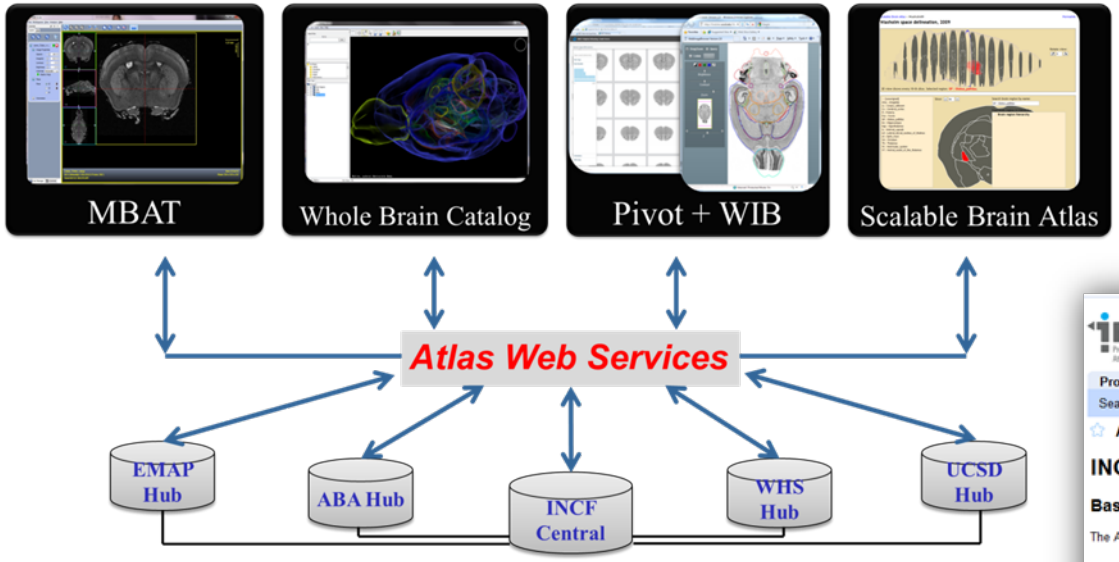
- The underlying catalog is compliant with the Catalog Services for the Web (CSW) standard, which defines how to search/add/update/delete catalog records programmatically.
- Support for several metadata profiles, for both data and services.
- Support for registration and indexing of multiple types of resources, including standard-compliant services, individual downloadable files, file collections, web sites, offline data, or other portals.
- Access control and authentication via Active Directory.
- Support for data publication and management workflow.
- Ability to synchronize/reharvest registered resources on a schedule or manually.
- Can federate with other CSW catalogs: for example catalogs supporting data portals established by other hubs.
- Integrates with an ontology service, and supports semantics-based search.
- The portal is customized for neuroscience/atlasing data types.
- The portal has a built-in viewer for image services, but also includes WIB and WBC visualization clients, and a Pivot-based gallery of neuroscience images.
- Ability to integrate with other (non-spatial) content management systems and portal frameworks, and ease of customization.
- Open source software model (available on sourceforge), with large development community following.

INCF-DAI Client Applications

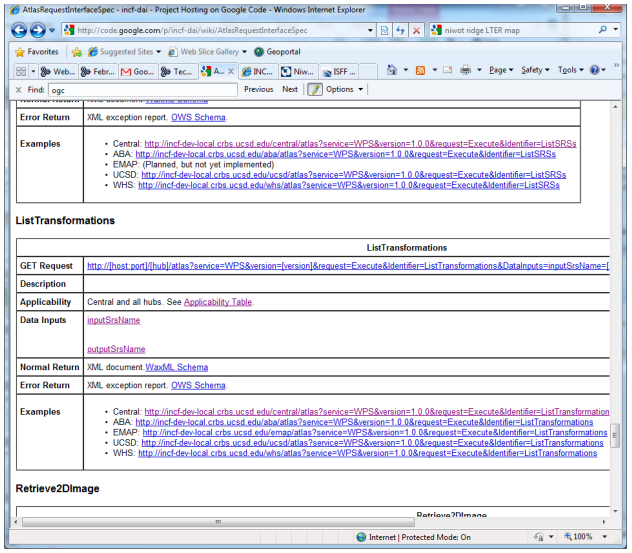


Accessing DAI from a software client

http://waxholm.neurocommons.org/page/Background_for_Developers



+ Python Libraries



AtlasRequestInterfaceSpec - incf-dai - Project Hosting on Google Code - Windows Internet Explorer

Error Return	XML exception report. OWS Schema.
Examples	• Central: http://incf-dev-local.crbs.ucsd.edu/central/atlas?service=WPS&version=1.0.0&request=Execute&Identifier=ListSRSs • ABA: http://incf-dev-local.crbs.ucsd.edu/aba/atlas?service=WPS&version=1.0.0&request=Execute&Identifier=ListSRSs • EMAP: (Planned, but not yet implemented) • UCSD: http://incf-dev-local.crbs.ucsd.edu/ucsd/atlas?service=WPS&version=1.0.0&request=Execute&Identifier=ListSRSs • WHS: http://incf-dev-local.crbs.ucsd.edu/whs/atlas?service=WPS&version=1.0.0&request=Execute&Identifier=ListSRSs

ListTransformations

ListTransformations	
GET Request	http://[host:port]/[hub]/atlas?service=WPS&version=[version]&request=Execute&Identifier=ListTransformations&DataInputs=inputSrsName=[
Description	
Applicability	Central and all hubs. See Applicability Table.
Data Inputs	inputSrsName outputSrsName
Normal Return	XML document.WaxML Schema
Error Return	XML exception report. OWS Schema.
Examples	• Central: http://incf-dev-local.crbs.ucsd.edu/central/atlas?service=WPS&version=1.0.0&request=Execute&Identifier=ListTransformation • ABA: http://incf-dev-local.crbs.ucsd.edu/aba/atlas?service=WPS&version=1.0.0&request=Execute&Identifier=ListTransformations • EMAP: http://incf-dev-local.crbs.ucsd.edu/emap/atlas?service=WPS&version=1.0.0&request=Execute&Identifier=ListTransformations • UCSD: http://incf-dev-local.crbs.ucsd.edu/ucsd/atlas?service=WPS&version=1.0.0&request=Execute&Identifier=ListTransformations • WHS: http://incf-dev-local.crbs.ucsd.edu/whs/atlas?service=WPS&version=1.0.0&request=Execute&Identifier=ListTransformations

Retrieve2DImage

incf incf-dai
International Neuroinformatics Coordinating Facility Digital Atlasing Infrastructure Task Force Code Repository

Project Home | Wiki | Issues | Source

Search Current pages for Search

AtlasRequestInterfaceSpec

INCF Atlas Server Request Interface

Base URIs

The Atlas base URIs identify the server location (URL) and the server application or context.

- Format: http://[server].[port]/atlas-[hub]
- Example: http://incf-dev-local:8080/atlas-aba

Full URIs

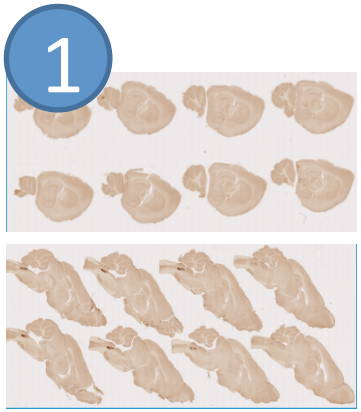
Full Atlas URIs are consistent with, but a subset of, the OGC (Open Geospatial Consortium) WPS (Web Processing Service)

The Base URIs will be appended with a query string which will include these case- and order-sensitive keys:

- service
  - all requests
  - value = WPS
- version
  - all requests except GetCapabilities
  - value be that returned by the GetCapabilities request
- request
  - all requests
  - value = { GetCapabilities | DescribeProcess | Execute }
- Identifier
  - all Execute requests
  - values enumerated below
- DataInputs
  - described below

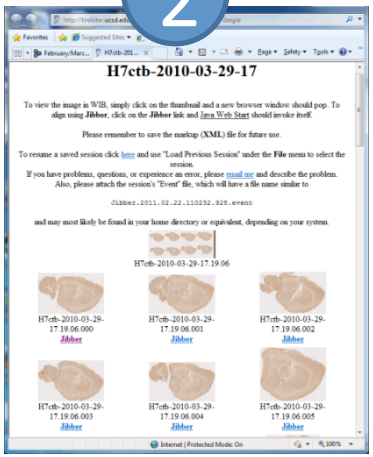
For example:

- http://incf-dev-local.crbs.ucd.edu/atlas-aba?service=WPS&request=GetCapabilities
- http://incf-dev-local.crbs.ucd.edu/atlas-aba?service=WPS&version=1.0.0&request=DescribeProcess
- http://incf-dev-local.crbs.ucd.edu/atlas-aba?service=WPS&version=1.0.0&request=Execute&Identifier=GetGenesByPOI&DataInputs=srsCode=WHS;x=263.7;y=15



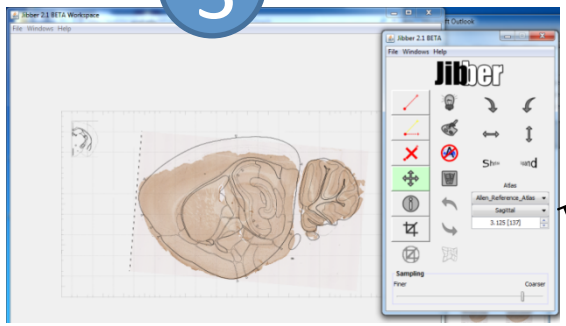
1

Segmenting and packaging for processing



2

Registering based on landmarks, using Jibber



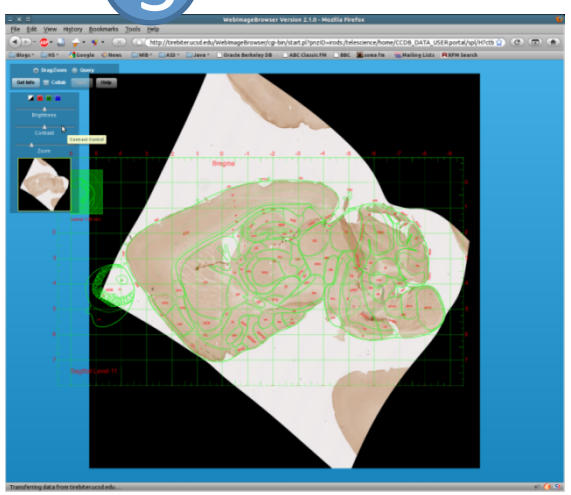
3

Warping, computing transformations

4

Compute cluster – Jetsam

Transformation Matrix



5

Visualize in WIB

Setting up in DAI

6

Atlas Service Update: new SRS: BrainStem, new transformations (from and to BrainStem SRS):
ListSRSs,
DescribeSRS
ListTransformations
TransformPOI

iRODS

Spatial Registration Workflow

(also with ITK/ANTs)

Once atlas services are updated…

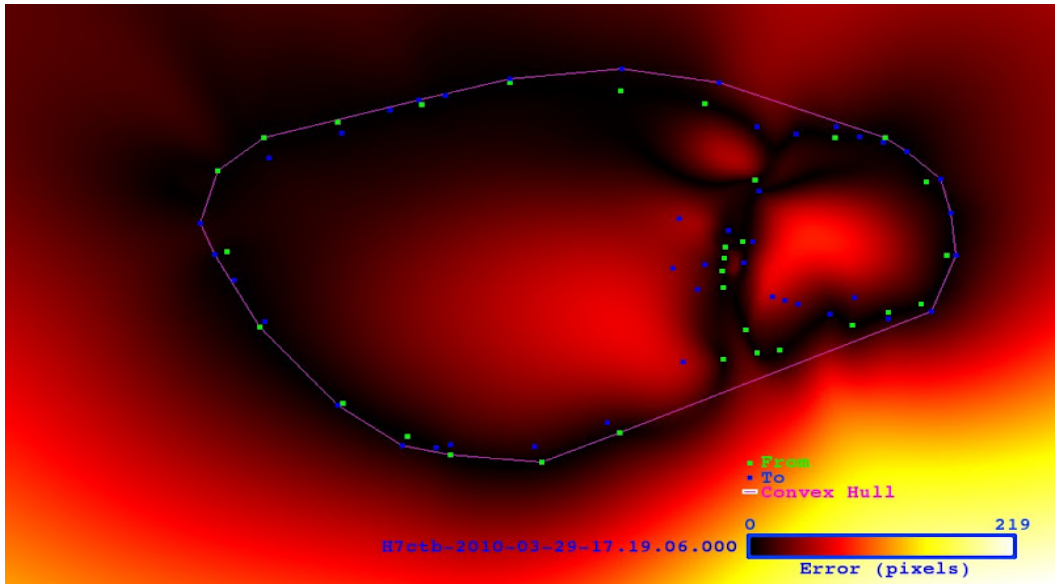
Your image is aligned.
Please go to

http://tirebiter.ucsd.edu/WebImageBrowser/cgi-bin/start.pl?plugin=INCF&pnzID=irods:/telescience/home/CCDB_DATA_USER.portal/CCDB_DATA_USER/acquisition/project_1/microscopy_74540/BrainStem/pnz/H7ctb-2010-03-29-17.19.06.003.pnz

to begin queries.

Additional work:

- Accuracy of registration
- WHS for the rat

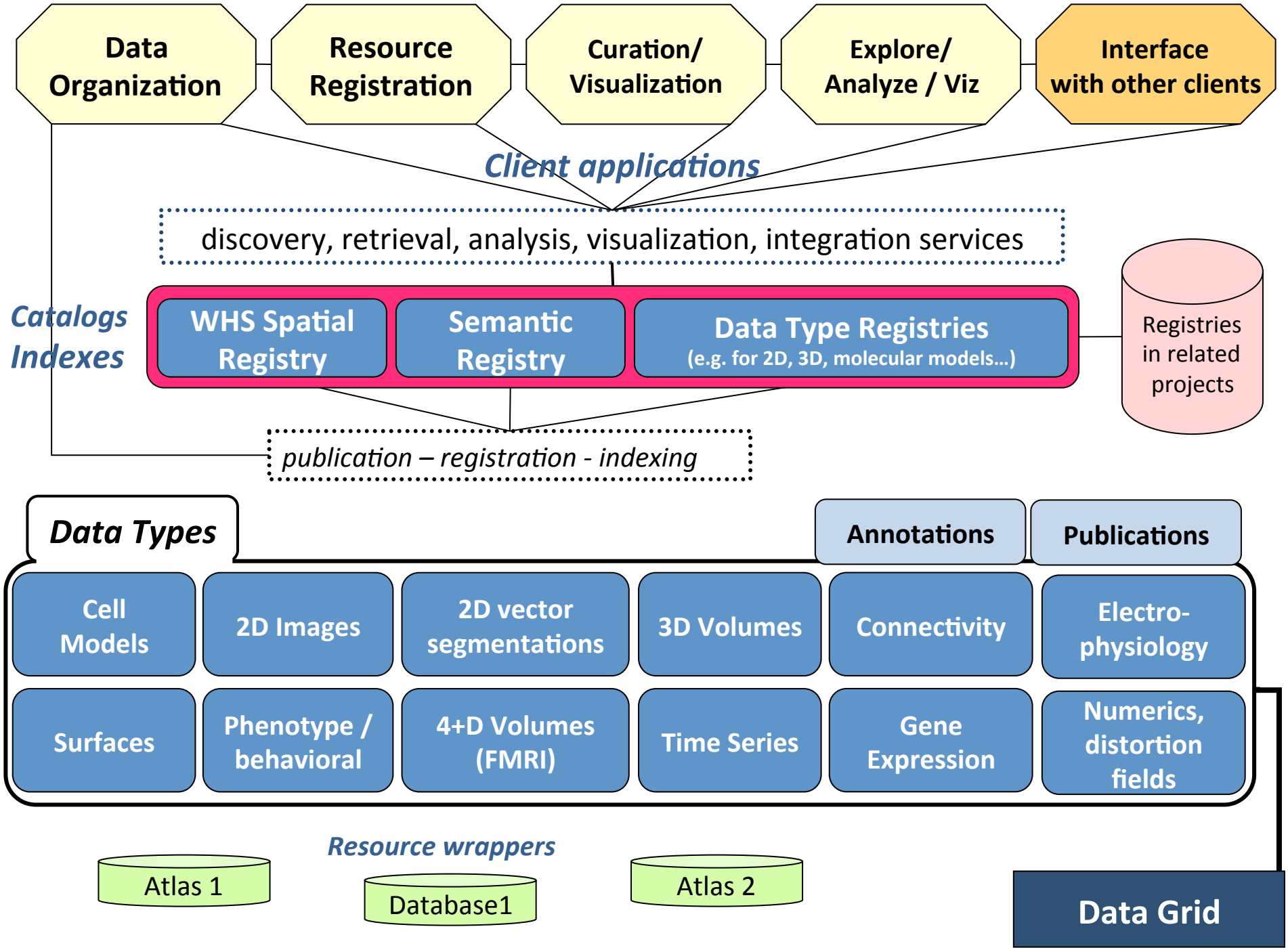


A higher-level view

Towards INCF CI

Initial INCF-CI requirements

- Support for **publishing, cataloging, discovery and access** to key types of neuroscience data
- Enabling straightforward **sharing of large datasets** via common elastic virtual data space (data grid, cloud)
- Enabling **distributed computation** and effective pooling of compute resources (computational grid)
- Common and transparent **information model and API foundation** that developers can use to contribute code, applications, data and resources
- **Scalability** to additional users, larger datasets, new types of data and resources
- Support for **long-term preservation and on-demand availability** of neuroscience resources of different types
- Support for a **range of integration models**: ability to integrate data across scales; species; development stages
- Ability to compose and run complex **workflows**, and ensure **provenance tracking**, **re-usability** and **accuracy** assessment of the results
- Enabling re-use of data by adopting or developing key **interoperability standards and reference frameworks**, including semantic and spatial reference models
- A **governance model** (rather, models) that addresses expectations of all stakeholders


Data Organization
Resource Registration
Curation/ Visualization
Explore/ Analyze / Viz
Interface with other clients
Client applications
discovery, retrieval, analysis, visualization, integration services
Catalogs Indexes
WHS Spatial Registry
Semantic Registry
Data Type Registries
(e.g. for 2D, 3D, molecular models…)
Registries in related projects
publication – registration - indexing
Data Types
Annotations
Publications
Cell Models
2D Images
2D vector segmentations
3D Volumes
Connectivity
Electro-physiology
Surfaces
Phenotype / behavioral
4+D Volumes (FMRI)
Time Series
Gene Expression
Numerics, distortion fields
Resource wrappers
Atlas 1
Database1
Atlas 2
Data Grid

What is our Interoperability Target

- Levels of interoperability:
  - Resources follow different information and service models, but can be discovered using Dublin Core metadata (title, abstract, author, when published, etc.)
  - Resources are available as services, whose capabilities can be requested and registered
  - Resources are available via standard services (e.g. Atlas Services)
  - Resources are defined wrt formal spatial, temporal and semantic conventions (vocabularies) managed by a naming authority
  - Resources follow common information models (e.g. WaxML)
- Not all components will reach our target interop level at the same time – but still need to be registered and discoverable → INCF Data Portal

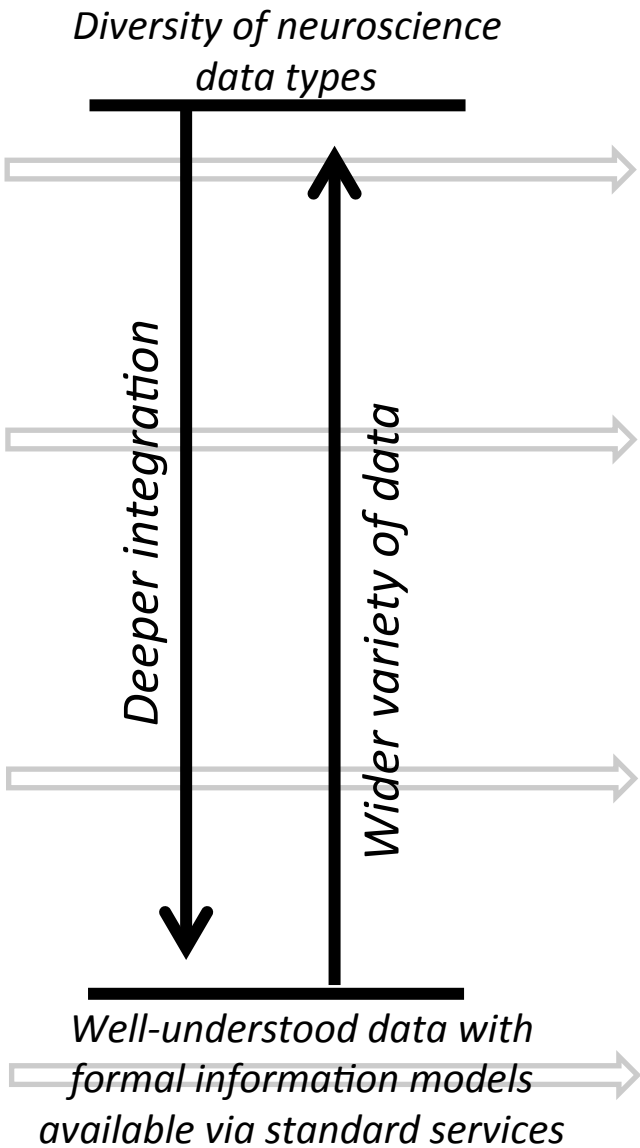
Levels of interoperability

Find and retrieve resources: files and file collections, services, documents – by thematic category and type

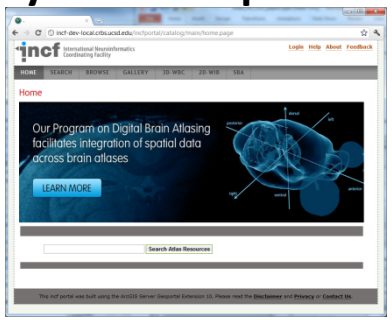
Data available in compatible semantics: ontologies, controlled vocabularies

Data available via standard service interfaces (e.g. WaxML, also OGC services) but different information models

Compatibility at the level of domain information models and databases

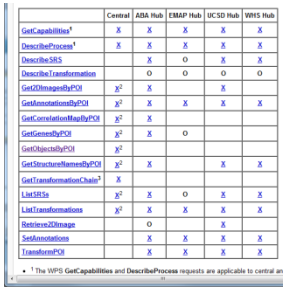


System components

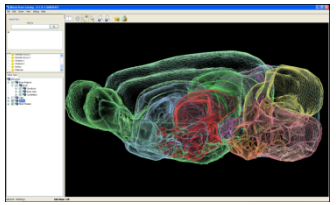


Data discovery portal

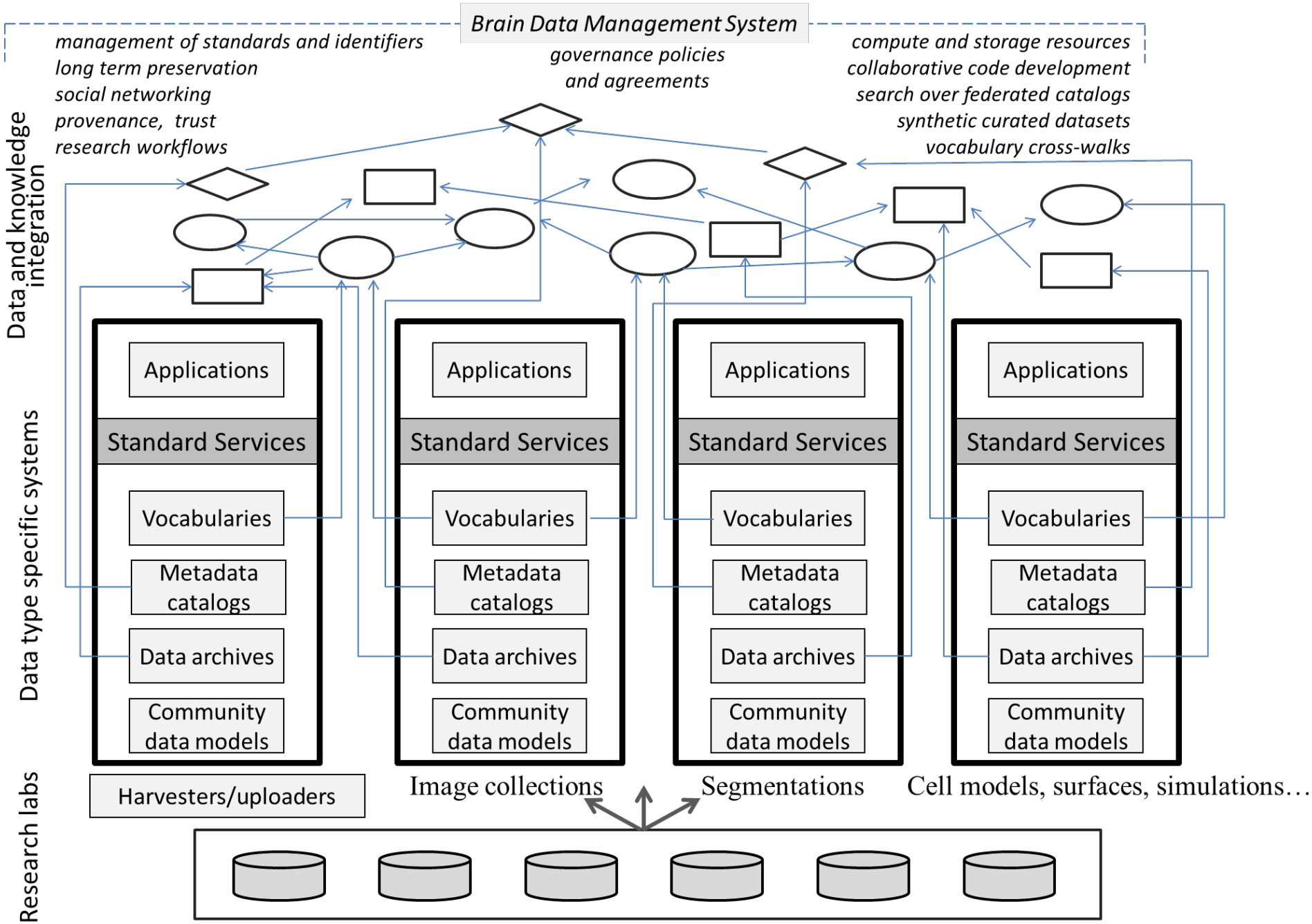
Shared vocabularies and ontology management



Service administration



Web and desktop applications


Brain Data Management System
management of standards and identifiers
long term preservation
social networking
provenance, trust
research workflows
governance policies
and agreements
compute and storage resources
collaborative code development
search over federated catalogs
synthetic curated datasets
vocabulary cross-walks
Data and knowledge integration
Data type specific systems
Research labs
Applications
Standard Services
Vocabularies
Metadata catalogs
Data archives
Community data models
Harvesters/uploaders
Image collections
Segmentations
Cell models, surfaces, simulations…