

Community Data Annotation/Curation

Dashboard - Wed Jan 17 12:02:54 EST 2007 - Wednesday, January 17, 2007 - Microsoft Internet Explorer

Address: http://www.itk.org/Testing/Dashboard/20070117-0100-Nightly/Dashboard.html

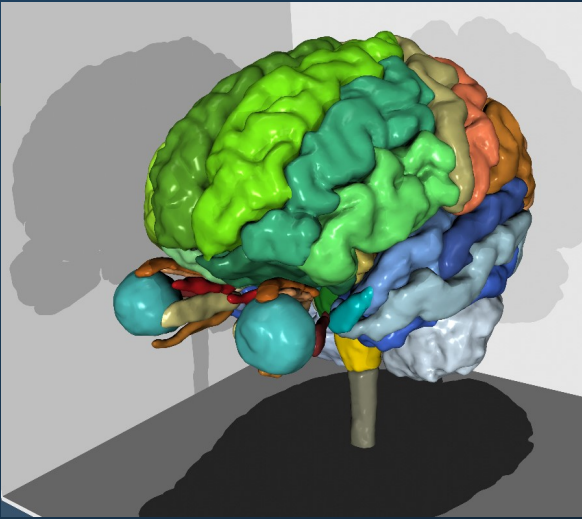
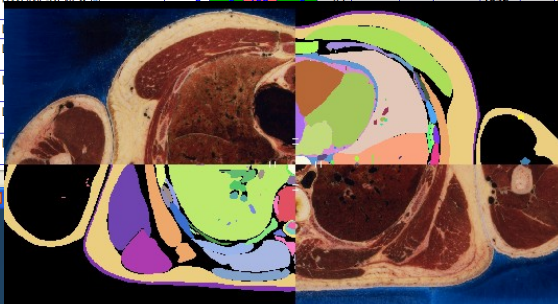
Dashboard - Wed Jan 17 12:02:54 EST 2007
Wednesday, January 17 2007

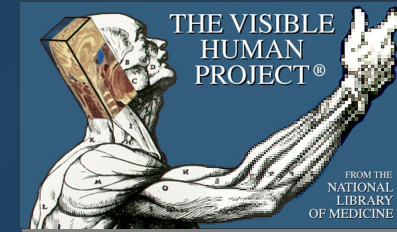
Dashboard | Date | Updates | Tests | Build | CVS | Doxygen | Bugs | Home | Rollup

64 Files Changed by 5 Authors as of 2007-01-17 01:00 GMT

Nightly Builds

| Site | Build Name | Update | Cfg | Build | | | Test | | | | Build Date | Submit Date | |
|-------------------------|---------------------|--------|-----|-------|------|-------|--------|------|------|-----|------------|-----------------------|------------------------------|
| | | | | Error | Warn | Min | NotRun | Fail | Pass | NA | | | Min |
| eno | | | | | | | | | | | | No submission | |
| iMac_intel.SINTEF | Darwin-c++ | | | | | | | | | | | No submission | |
| G4Cube.SINTEF | Darwin-c++ | | | | | | | | | | | No submission | |
| iMac_intel.SINTEF | Darwin-c++ | 77 | 0 | 0 | 24 | 43.6 | 0 | 11 | 1073 | 258 | 25.2 | Jan 17 08:10 CET | Wed Jan 17 03:32:53 EST 2007 |
| midworld.kitware | DarwinG5-g++ | 0 | 0 | 0 | 60 | 397.5 | 0 | 16 | 1107 | 219 | 162.9 | Jan 17 01:31 EST | Wed Jan 17 10:52:23 EST 2007 |
| eno | Darwin-Xcode2.4-rel | 77 | 0 | 0 | 60 | 395 | 0 | 0 | 1084 | 258 | 33.2 | Jan 17 01:59 EST | Wed Jan 17 09:07:32 EST 2007 |
| muse.sci.utah.edu | IRIX64-CC-7.3 | | | | | | | | | | | | No submission |
| kraepelin.uiowa | IRIX64-MPro-7.4-dbg | 77 | 0 | 0 | 1 | 52.9 | 0 | 5 | 1088 | 249 | 383.2 | Jan 16 22:14 CST | Wed Jan 17 06:30:40 EST 2007 |
| kraepelin.uiowa | IRIX64-MPro-7.4-opt | 74 | 0 | 0 | 0 | 61.5 | 0 | 4 | 1089 | 249 | 75.5 | Jan 17 05:39 CST | Wed Jan 17 08:57:01 EST 2007 |
| insight.journal.kitware | KWSbtle_Review | 0 | 0 | 105 | 0 | 0.2 | | | | | | Jan 17 02:09:52 -0500 | Wed Jan 17 02:12:49 EST 2007 |
| pragmatic.crd | | | | | | | | | | | | | No submission |
| chiarugi.uiowa | | | | | | | | | | | | Jan 16 21:54 CST | Wed Jan 17 00:08:49 EST 2007 |
| chiarugi.uiowa | | | | | | | | | | | | Jan 16 23:41 CST | Wed Jan 17 02:14:39 EST 2007 |
| crunch1.isi.nl | | | | | | | | | | | | Jan 17 03:16 CET | Wed Jan 17 02:45:06 EST 2007 |
| camelot.kitware | | | | | | | | | | | | Jan 17 00:15 EST | Wed Jan 17 01:25:37 EST 2007 |



Community Annotation/Curation

Demo Project

- Open atlas
 - Individuals
 - Populations (??)

Success criteria

- Acceptance and participation by anatomy community
- Portability of tools to other projects
- At least one “good” atlas

Project cycles

- Identify customers (anatomists) and customer’s customers (radiology, surgery, algorithm developers, educators)
- “Extreme” approach, “release early, release often”

Feasibility studies

- Pick two anatomical areas (thorax, brain)

Deliverables

- Infrastructure/process
- Distributed atlas

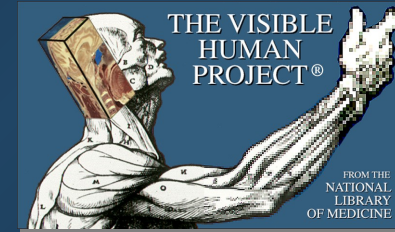
Integration needs

- Visualization
- Federated database
- Ontologies

Issues

- Intellectual property
- Business model

Open Atlas: Requirements



Open data and open process

Customer GUI application

Software Toolkit

Methods for curation

Mechanism for consensus building

Mechanisms for quality control

Continuous process feedback

Provenance

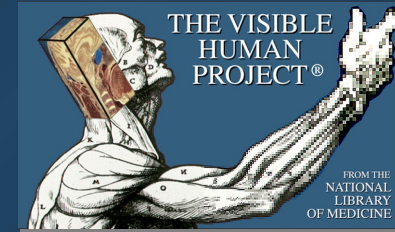
Soup to nuts software

- Reference implementation
 - Visualization
 - Editor
 - Registration, model extraction, etc.
- Query application

Outreach to customer's customer

Local and web based

Open Atlas: Components



User interface

Segmentation tools + manual correction

Interface to multiple ontologies

Revision control

Automated quality assurance

Dashboards

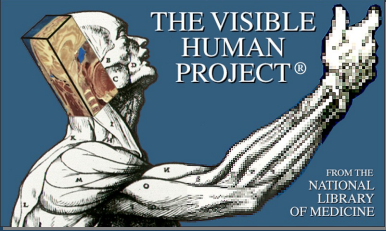
Packaging/delivery

Data repository

API for programmatic access to data/annotations/tools

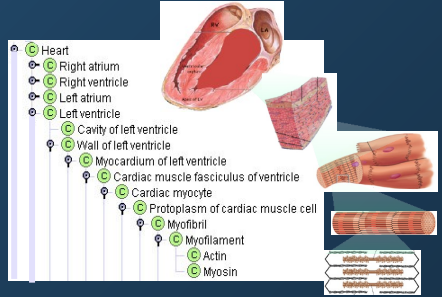
Core team

- Anatomists/Radiologists(Domain expert)
- Database design
- Ontology support
- Image analysis
- Image/Geometry editor
- Process support tools



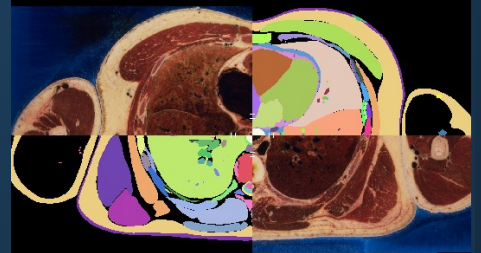
Starting Points

U Wash FMA

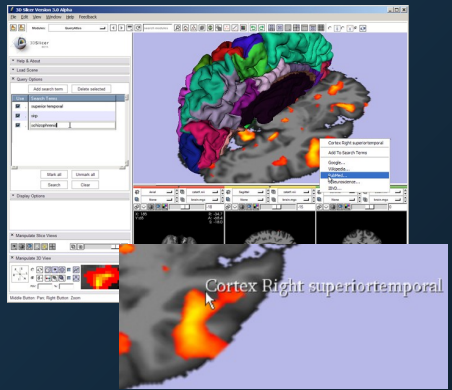


NLM Visible Human Thorax

- Original from EAI
- Enhanced by Virtual Soldier Project



Brigham and Women's Brain Atlas/Slicer



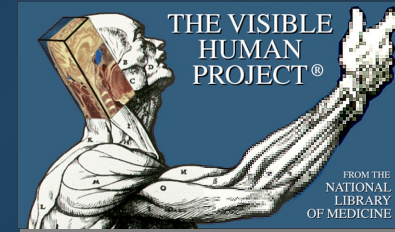
Community Data Annotation/Curation

Background Slides:

Open, Distributed and Collaborative Data Annotation

Bill Lorensen
Insight Software Consorti

Motivation



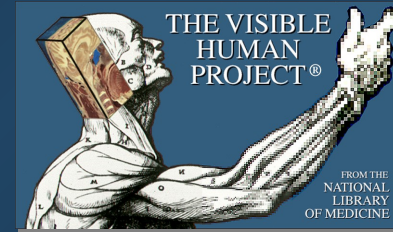
Many imaging communities are data starved

- Algorithm developers
- End users

Lots of raw data, but very little annotated data

- LIDC
- Notre Dame Biometrics Data Distribution

Forms of Annotation



Anatomy labels

Contours

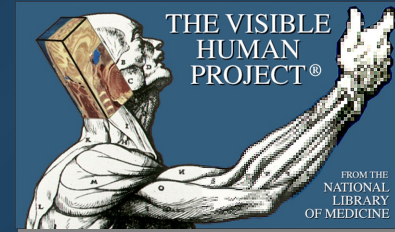
Statistical

Anatomical landmarks

Templates

Ground truth

Problem Statement



Sensors are producing large amounts of data

Annotation adds value

Annotation of large data collections is expensive and error prone

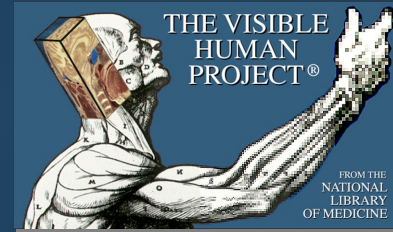
Customers

Algorithm developers

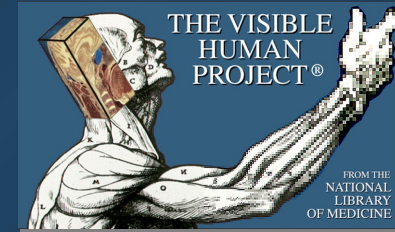
Anatomists

Teachers

Sensor manufacturers



Solution

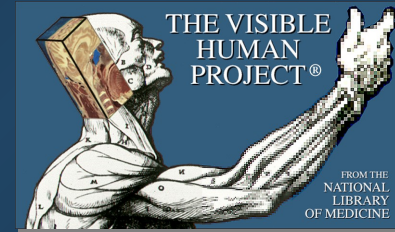


A distributed, coordinated community can efficiently and economically annotate large sets of data

- wikipedia
- wikimapia

Extreme programming techniques can be applied to the data annotation process

Examples

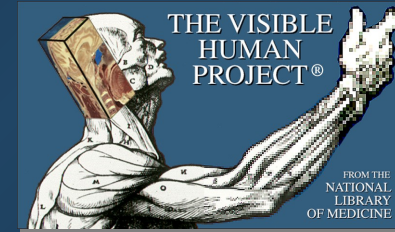


Anatomical atlases

Face recognition

- 2D photos
- 3D range data

Example – FBI Facial Reconstruction

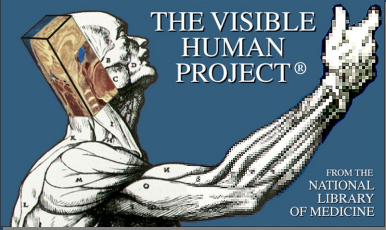


Two data collections

- 300 CT datasets of heads
- 1000 photo and range data of faces

Challenge

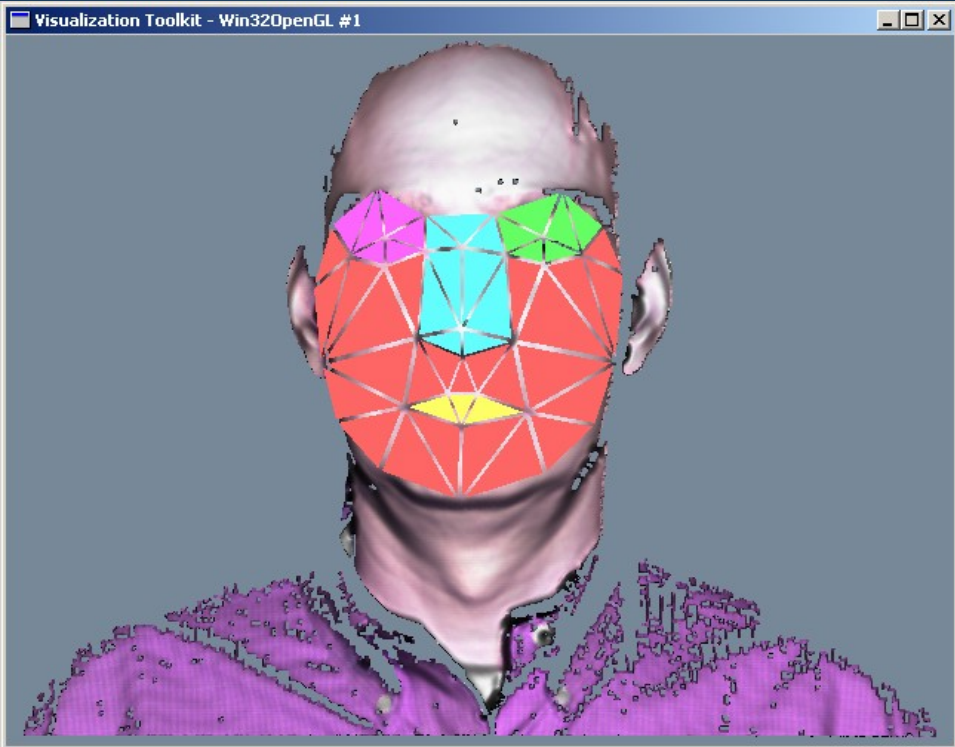
- Extract models of eyes, noses and mouths from range data
- Replace eyes, noses and mouths in CT data with range data models



Face Template

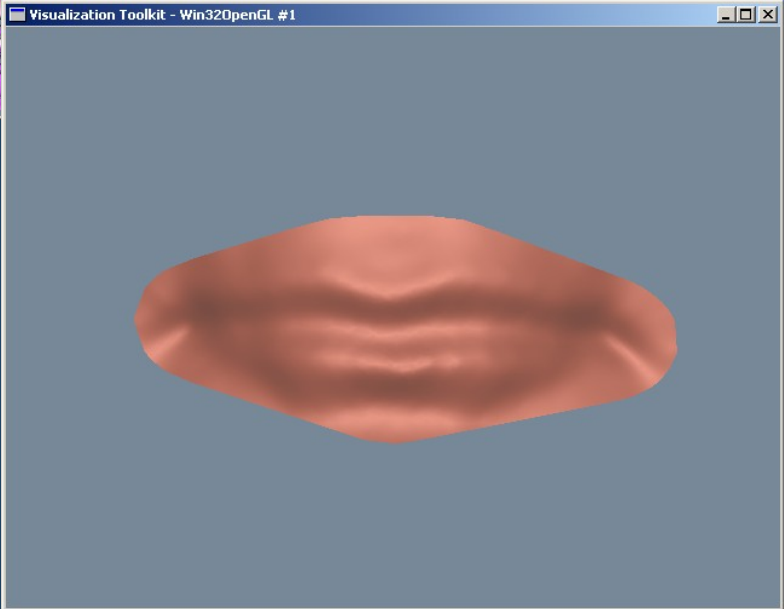
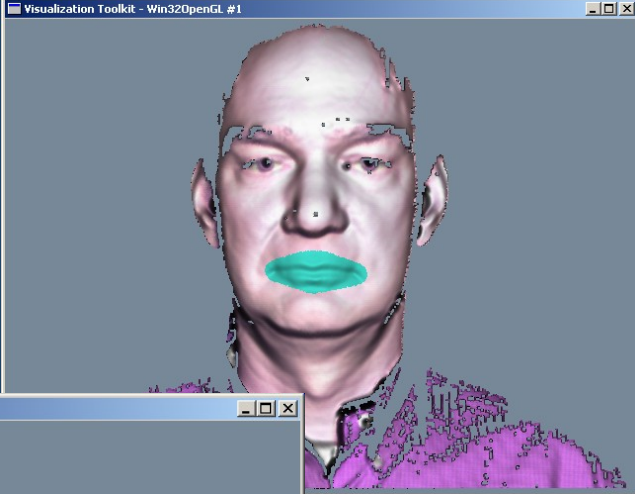
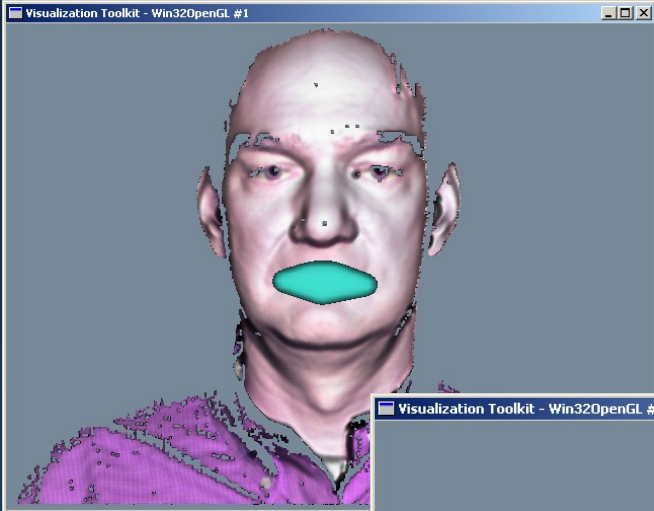


Photo



Range Data

Mouth



Multidisciplinary Project

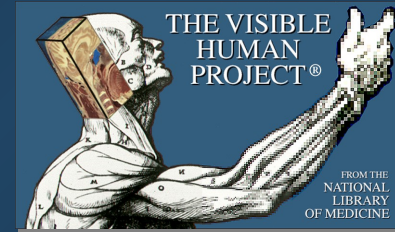


Image Analysis

Anatomy

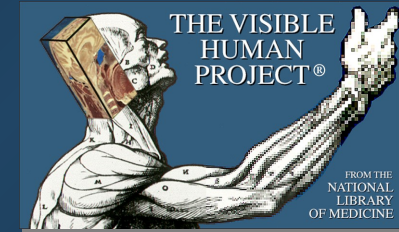
Databases

Ontologies

Software Engineering

Quality Assurance

Visualization



Menu for Success

A Community with a common vision

A pool of talented and motivated developers/scientists

A mix of academic and commercial

An organized, light weight approach to product development

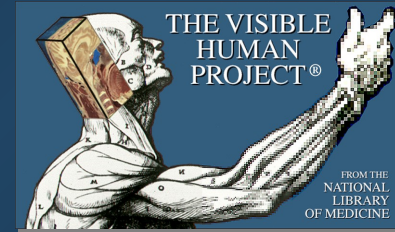
A leadership structure

Communication

A business model



Leadership Structure



Follow NCBC model

Algorithms

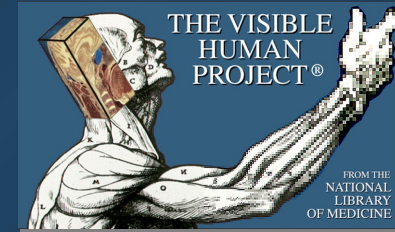
- Ontology creation
- Image analysis

Engineering

Driving Projects

- Open Atlas
- Radiology ground truth

Business Model



All core technology is open, without restriction

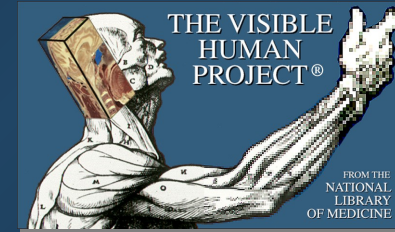
All NLM supported annotation is open, without restriction

Proprietary enhancement of annotated data is allowed

Annotated data can be used in commercial products without restriction

Guiding Principles

Extreme Data Annotation

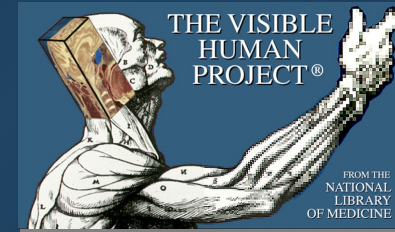


The community owns the data

Although the origin of the data is retained, others are free to correct defects and enhance each other's data

In the end, all of the data should appear as though one person annotated it

Extreme Data Annotation

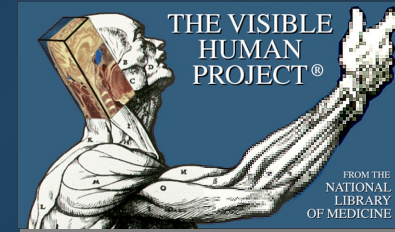


Release early, release often

Although people are tempted to keep their data under wraps until it is perfect, the process encourages them to release their data as soon as it passes some minimum quality control tests

The longer the data is visible to the community, the better integrated it will be

Extreme Data Annotation

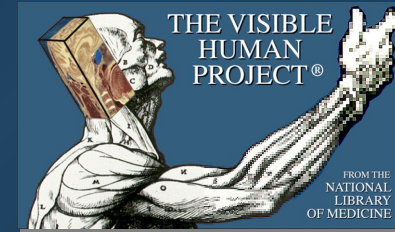


Continuous integration

There is no scheduled porting to databases or model formats

All new data is integrated into supported databases and data formats continuously

Extreme Data Annotation



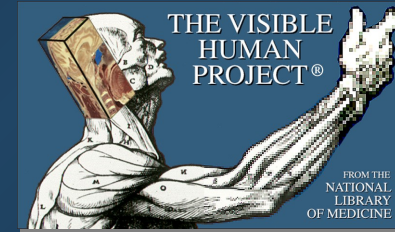
Everyone agrees to keep the data free of defects

Although everyone is encouraged to submit their data early, the data must pass quality tests and integration tests nightly

A continuous QA process sends e-mails to people who check in data that does not meet quality control tests

More effectively, the community enforces the commitment through peer pressure

Software/Data Analogies



Software

Program

Text editor

Compilation error

Compilation

Style

Data

Annotated data

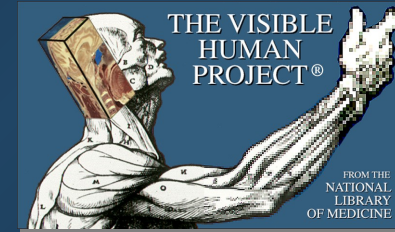
Image editor

Collisions

Model creation

Ontology

Why NLM?



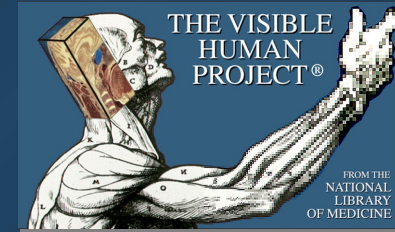
NLM produces, collects, annotates, stores and distributes data

- Medline
- Visible Human Project
- Mayo Data Collection

NLM has managed distributed, collaborative, multidisciplinary projects

- Insight Toolkit
- HPCC Internet 2

What is needed?



Select a pilot project

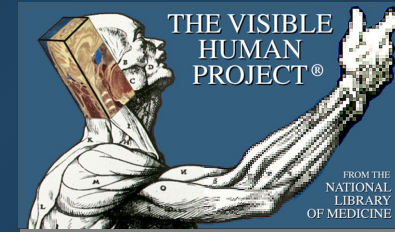
- Open Atlas Project

Select customers

Select core team

- Anatomists
- Database design
- Ontology support
- Image analysis
- Image/Geometry editor
- Process support tools

Open Atlas Project



Create anatomical atlases from cross-sectional image data

Semi-automatic and manual labeling of structures

Engage the anatomy community