

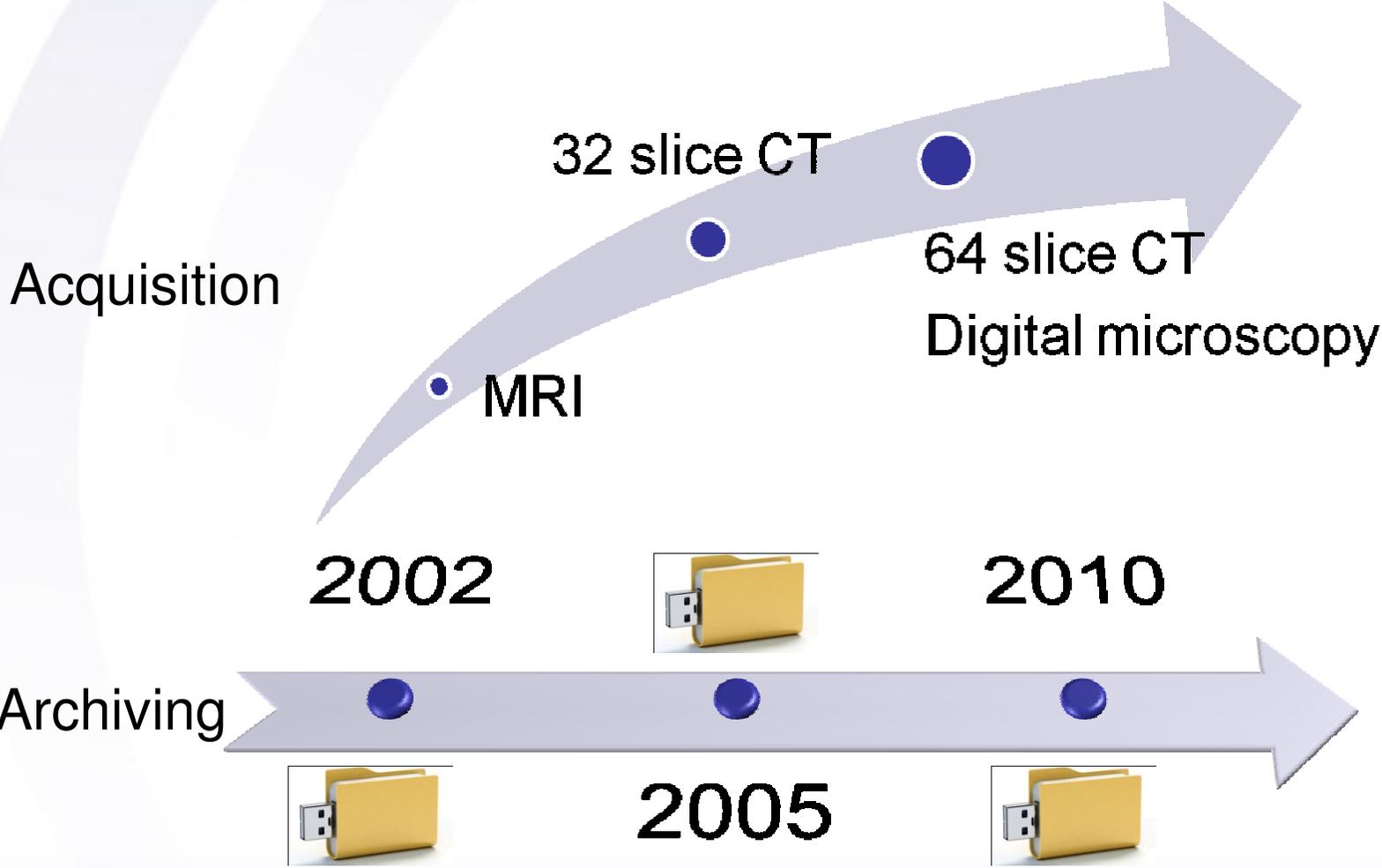


An Open-Source Digital Archiving System for Medical and Scientific Research

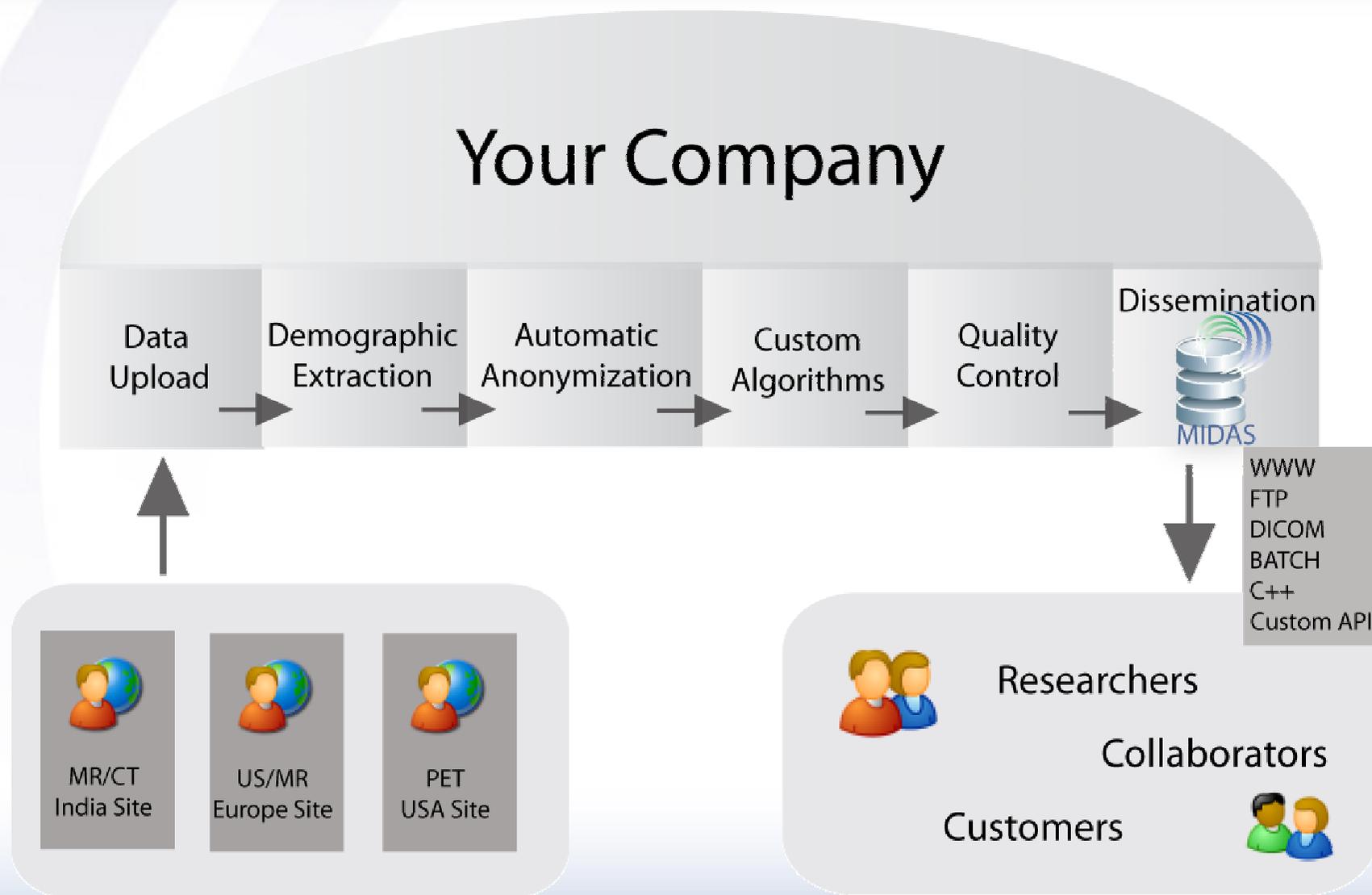
Open Repositories 2010

Julien Jomier
Kitware Inc.

Motivation



Data Collection, Management and Distribution



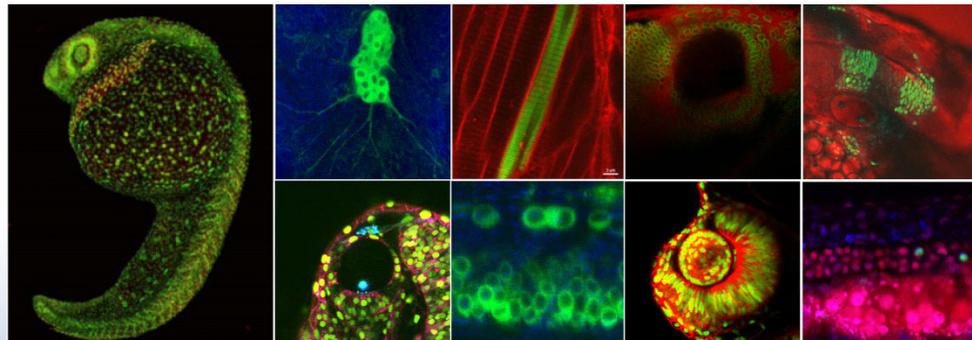
Outline

- **Challenges** in archiving scientific and medical datasets
- **Solutions** for digital archiving of scientific datasets
- The **MIDAS system**
- **Open repositories** using MIDAS
- **Future work**



Archiving Challenges

- Datasets are massive
- Datasets are heterogenous
- Metadata are embedded in the image format
- Metadata are not standardized
- Data provenance is equally important
- Uploading and Downloading data
- Processing
- Visualization



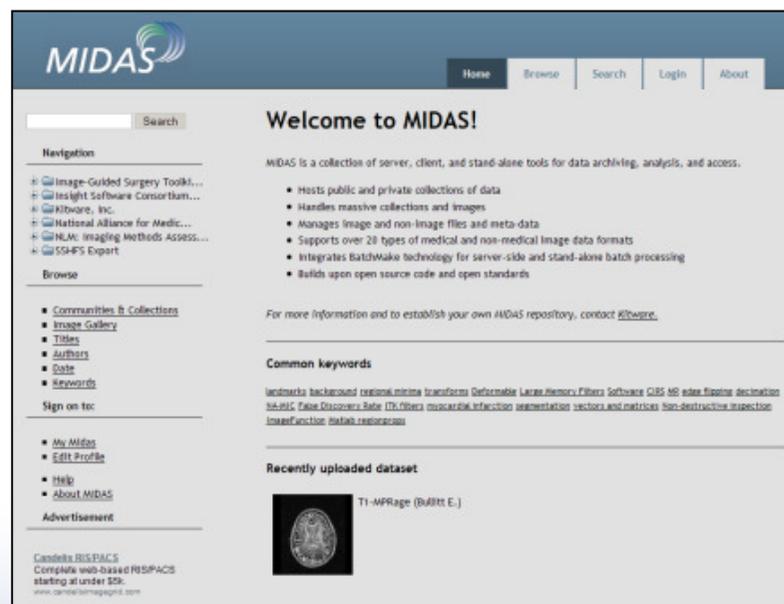
Scientific Archiving Solutions

- Hard drive/Tape Backup
- Revision control systems
- Picture Archiving and Communication System (PACS)
- XNAT: Medical (DICOM) images only
- NBIA: NCI digital archive
- DSpace/EPrint: Publications
- Others...



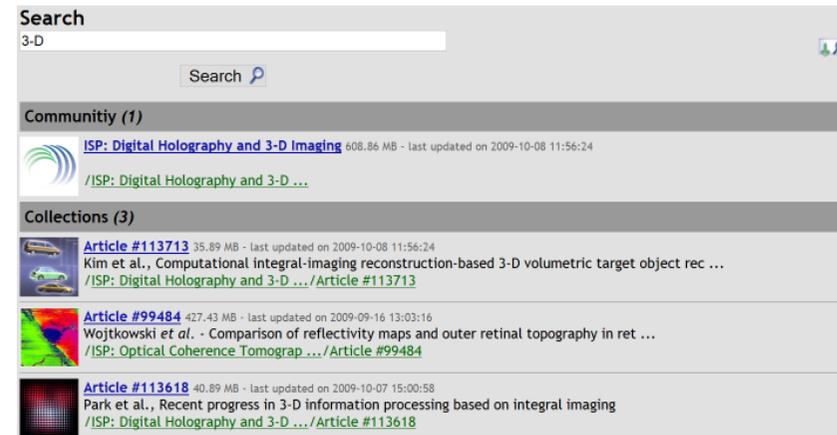
MIDAS

- Web-based Multimedia Digital Archiving System
- Written in PHP with PostGreSQL backend
- Modular and highly customizable framework
- Open Source (BSD)
- Based on DSpace
- API: REST, C++
- Grid computing
- Online visualization



MIDAS Features

- OAI-PMH
- Handle server
- Advanced search
- Upload Filters
- Upload Templates
- WebDAV support (upload/download)
- Comments and Ratings
- MIDAScpp (C++ API)
- MIDASDesktop



The screenshot displays the search results page for '3-D'. At the top, there is a search bar with the text '3-D' and a search icon. Below the search bar, the results are organized into sections: 'Community (1)' and 'Collections (3)'. The 'Community' section shows a single entry for 'ISP: Digital Holography and 3-D Imaging' with a file size of 608.86 MB and a last update date of 2009-10-08 11:56:24. The 'Collections' section lists three articles: 'Article #113713' (35.89 MB, last updated 2009-10-08 11:56:24) by Kim et al., 'Article #99484' (427.43 MB, last updated 2009-09-16 13:03:16) by Wojtkowski et al., and 'Article #113618' (40.89 MB, last updated 2009-10-07 15:00:58) by Park et al. Each article entry includes a small thumbnail image and a link to the full article.



OPEN REPOSITORIES USING MIDAS



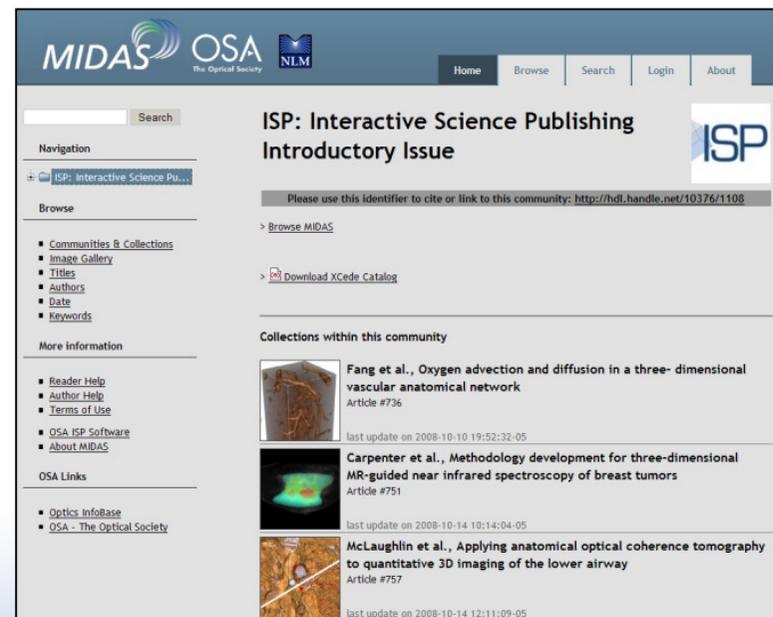
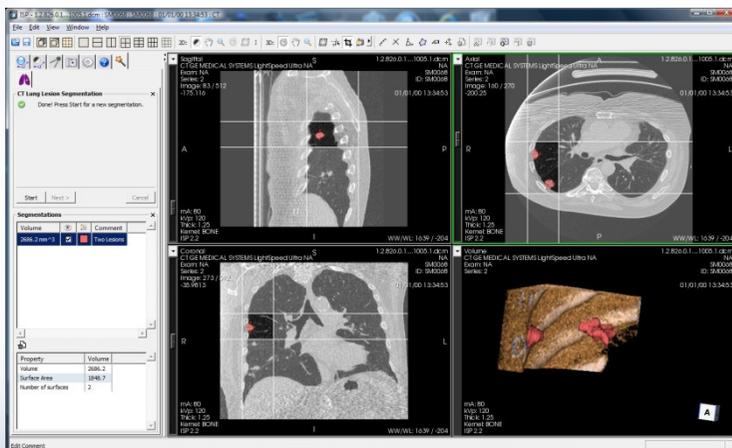
June. 03, 2010

Digital Archiving with MIDAS



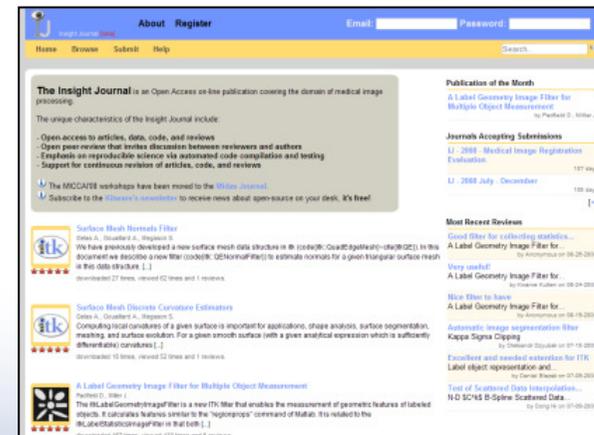
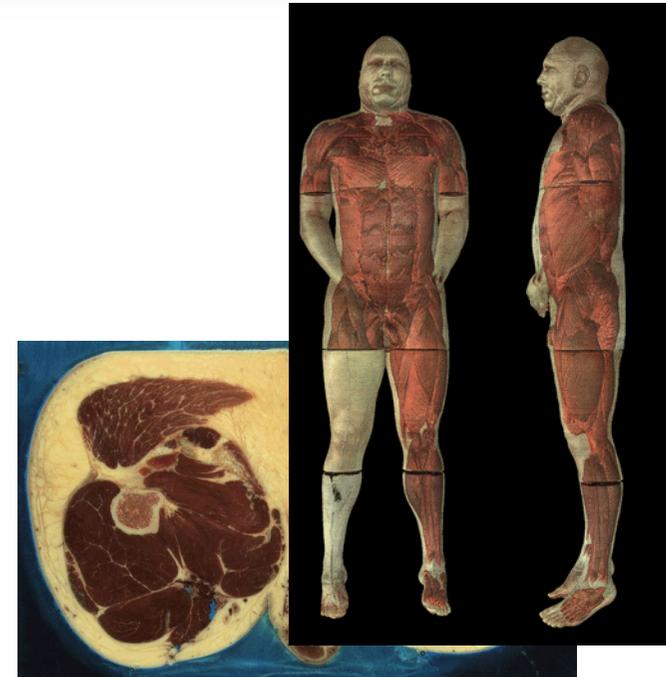
Optical Society of America

- Interactive Science Publishing
- **Enhance traditional publishing with 3D rendering**
- Sponsored by the National Library of Medicine
- Low-resolution < 10s
- High-resolution = background
- Open Datasets
- <http://midas.osa.org>



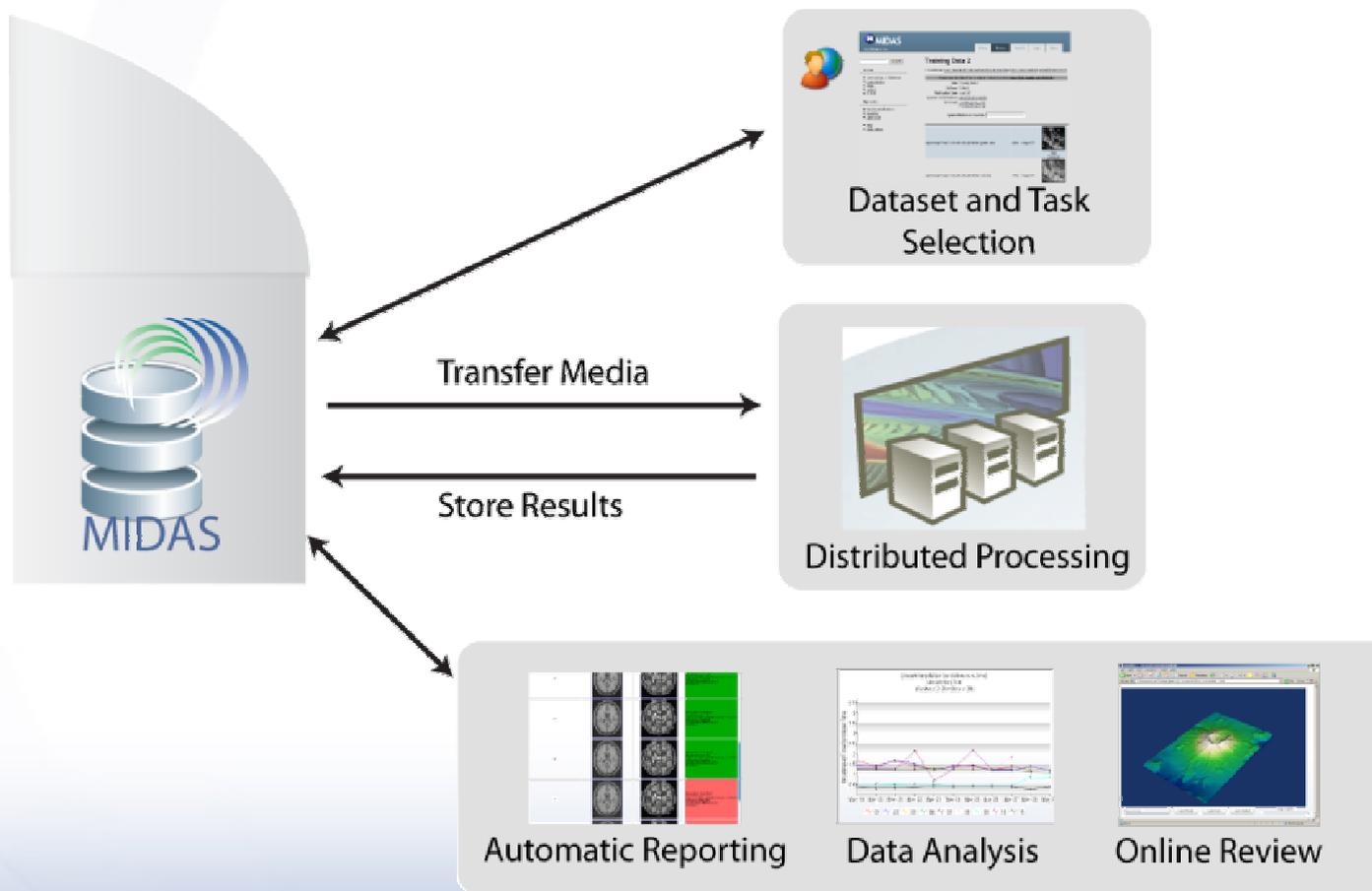
National Library of Medicine

- **Insight Journal**
 - Open-Access Journal
 - 1600+ registered users
 - 775 reviews
 - 400 Publications
- www.insight-journal.org
- Disseminating datasets:
 - **Visible Human**
 - Army research labs



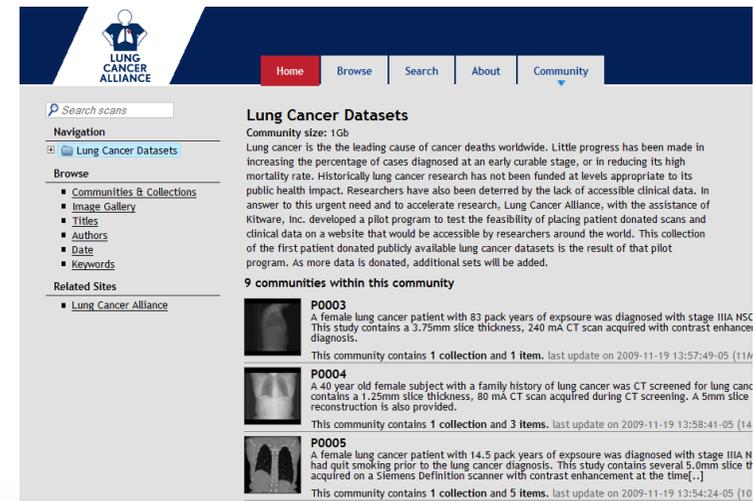
National Cancer Institute

- High Throughput for small animal imaging
- >1TB data every day



Give-A-Scan Project

- Lung Cancer Alliance and Kitware
- First patient-powered open access database
- **10 patients** with advanced lung cancer
 - computer tomography scans
 - # pack years
 - respiratory functions
 - patient information
- <http://www.giveascan.org>

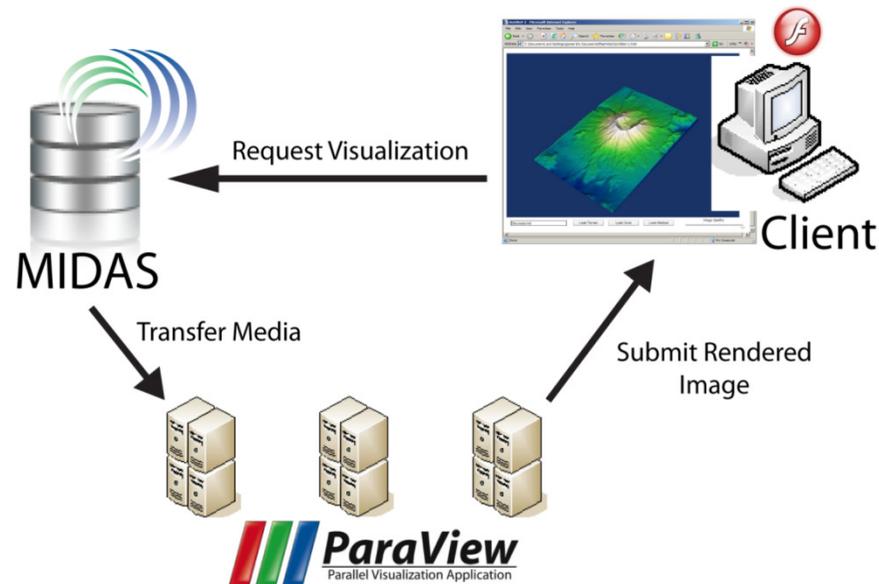


The screenshot shows the 'Lung Cancer Datasets' website. The header includes the Lung Cancer Alliance logo and navigation links: Home, Browse, Search, About, and Community. A search bar is labeled 'Search scans'. The main content area is titled 'Lung Cancer Datasets' and includes a description of the project, a list of navigation options (Communities & Collections, Image Gallery, Titles, Authors, Date, Keywords), and a list of related sites (Lung Cancer Alliance). Below this, there are three patient cases listed with their IDs (P0003, P0004, P0005), brief descriptions, and thumbnail images of CT scans.



Future MIDAS developments

- Current release 2.6 (last week)
- Medical datasets sharing policies
- OSA/Microsoft processing infrastructure
- Visualization on the web



Conclusion

- MIDAS: open-source data archiving/management framework
- Open-source has shown how it can impact Science, Open-data is next
- Data (and associated metadata) is key to scientific Research



Acknowledgments

- The National Institutes of Health
 - STTR 5R41NS059095-02
 - STTR 2R42NS059095-03
- The National Library of Medicine
- The Optical Society of America
- The University of North Carolina at Chapel Hill

- <http://www.kitware.com/midas>





An Open-Source Digital Archiving System for Medical and Scientific Research

Thank you!

julien.jomier@kitware.com