

# Research Data Storage & Management at CHRF, Part II

Michael Wagner

Andy Rupert, Nick Felicelli, Alex Spivak

Biomedical Informatics

Home — Research IT Services - Microsoft Internet Explorer

Address <http://bmi.cchmc.org/>

# The most important slide

Accessibility Contact

Search Site Search

Home Resources Services Help About Us

Log in

You are here: Home

- <http://bmi.cchmc.org>
  - Research IT Services homepage
  - Description of Resources, Services, Software, etc.
  - Statistics, utilization
  - Help/self-help pages
- [HELP@BMI.CCHMC.ORG](mailto:HELP@BMI.CCHMC.ORG)
  - BMI helpdesk, available to all CHRF staff for ANY IT question
  - Monitored 8x5, triaged within 2 hours

The Research IT Services group in the Division of Biomedical Informatics provides computational resources, services and support to investigators at Cincinnati Children's Research Foundation and the University of Cincinnati College of Medicine.

### Hot topics

Getting started

- System access requests
- Usage policies
- Mission

### New resources

- Application for use of Cincinnati Genomic Center for data/samples 14 Dec 09
- Mac OS X Snow Leopard (v. 10.6) 28 Aug 09
- Direct network link to UC 19 Jun 09
- Vector NTI desktop sequence analysis software 09 Mar 09
- Group Wise email from SOAP-compatible clients 26 Jul 08

### InfoPath/SharePoint

Get more information about these tools including access instructions, training

### Computational cluster

Linux Cluster Utilization

last 24 hours

Active CPUs 232 Total CPUs 282 Current Utilization: 82%

Detailed stats, policies, testimonials, user group, more...

### Help

**For desktop support with Windows and Mac,** call the Cincinnati Children's Service Desk, 513-636-4100, or the UCit Help Desk, 513-556-HELP.

**For help with BMI-supported hardware and software,** email the BMI Help Desk.

**For password changes,** go to the following do-it-yourself sites. For all other systems and applications, email the BMI Help Desk.

- Cluster, calendar, wiki, web servers
- BMI Linux desktops, bullpen, internal Citrix servers
- Cincinnati Children's network and email

Local intranet

# [Research] Data Storage

- Unstructured data
  - RDS
  - Xythos
- Structured data
  - Primary data portals
  - Electronic Lab Notebook
  - Research Papers

} Last open house

# Data Storage: Best Practices

- Avoid duplication – share!
- Avoid external hard drives
  - better, more reliable hardware in institutionally provided storage (RDS, Xyθος, ...)
- Allow access only for people that need access (HIPAA)
- Differentiate between access level (Read-Only and Read-Write)
- Make sure backup procedures conform to data retention requirements

# Recap from last time

- [RDS permissions tool](#): self-service web page to request access to G: drives, protected folders and project drives
- <http://ibmi.cchmc.org>: self-service quota increase requests, new project drive requests, etc.
- <http://bmixythos.cchmc.org>: web-based storage, good for sharing with external collaborators.

# Structured data

- Have many different data types, can we do better than just to store them as “anonymous” files on a file server?
- What information besides the file name and timestamp might be useful when trying to
  - search for data?
  - process data?
- Examples:
  - Flow cytometry data (fcs files)
  - Microarray data (CEL files)
  - Affy genotyping data (CEL files)
  - Images (TIFF, jpg, gif, etc.), movies
  - Word documents, pdf documents
  - .....

# Flow Cytometry Data

- Fcs file headers contain information on
  - Acquisition date
  - Cytometer used
  - Experiment name
  - Parameters/Antibodies used
  - Tube name
  - Etc....
- How to make this information useful and searchable?

# FCS Data web portal

- <http://research.cchmc.org/FCSData>
- Accessible to research staff from anywhere on the Internet
- Accounts can be requested by emailing [help@bmi.cchmc.org](mailto:help@bmi.cchmc.org)
- No quotas/limits on data
- Data can be shared with other registered users, self-service user management
- Flexible search interface
- Can tag fcs files with keywords, free text
- Can upload and associate pdf, xls, xml files
- Can define links to search results
- Extensions planned (incl. quality control measures, analysis support)



- (FCSDData demo)

# Portals for Affymetrix expression and genotype data

- Affy expression microarray data portal
  - Accessible from anywhere at <http://research.cchmc.org/microarray>
  - Managed storage for CEL files
  - Similar features as FCSDData portal PLUS
    - RMA processing on BMI Linux cluster enabled
- Affy genotyping portal
  - Accessible from anywhere at <http://research.cchmc.org/genotyping>
  - Can submit Birdseed & CRLMM calling algorithms

- (Microarray and Genotyping Portal Demo)

# Other portals planned

- Imaging data
  - fMRI (in process)
  - General imaging (confocal, gels, other images)
  - Sequence data
  - Your data type
- Ideas?

# Why bother?!

- Free storage
- Good organization
- Much easier retrieval of data
- Easily accessible from anywhere on the web
- Easy sharing with others
- Ongoing in-house development (let us know if you have ideas)
- Everything is backed up, secure.
- Use the BMI cluster for processing (e.g., RMA for microarrays)

# The Electronic Lab Notebook

- Information available at <http://eln.cchmc.org>
- Software is available for CHRF investigators at no charge, supported by BMI
- Secure, centralized storage of research records
- Ca. 25 active users at CHMC currently, more seats available
- Please send inquiries to [help@bmi.cchmc.org](mailto:help@bmi.cchmc.org), we are happy to schedule demonstrations, set you up with test accounts, etc.

- (ELN short demo)

# Mendeley PDF management tool

- Freely available at <http://www.mendeley.com>
- Collect, organize and share pdf libraries
- All fully searchable
- Mendeley provides client software and web interface
- Web browser plugin for quick additions, e.g., from PubMed
- MS Word integration for bibliographies
- Not officially supported by us, but do check it out!



- (Mendeley demo)

# Collection of links from this talk

- <http://bmi.cchmc.org> -- Research IT homepage
- [help@bmi.cchmc.org](mailto:help@bmi.cchmc.org) -- BMI help desk
- <http://ibmi.cchmc.org> -- storage request system
- <http://bmirdsdp/DAP> -- permissions request system
- <http://research.cchmc.org/FCSDData>
- <http://research.cchmc.org/microarray>
- <http://research.cchmc.org/genotyping>
- <http://eln.cchmc.org> -- Electronic Lab Notebook
- <http://www.mendeley.com> -- PDF organizer

# Acknowledgements

- Andy Rupert (BMI)
  - Nick Felicelli (BMI)
  - Alex Spivak (co-op student)
  - Prakash Velayutham (cluster admin)
- 
- Funding from Rheumatology NIH P30 Informatics core and from BMI

Home — Research IT Services - Microsoft Internet Explorer

Address <http://bmi.cchmc.org/>

# The most important slide

Accessibility Contact

Search Site Search

Home Resources Services Help About Us

Log in

You are here: Home

- <http://bmi.cchmc.org>
  - Research IT Services homepage
  - Description of Resources, Services, Software, etc.
  - Statistics, utilization
  - Help/self-help pages
- [HELP@BMI.CCHMC.ORG](mailto:HELP@BMI.CCHMC.ORG)
  - BMI helpdesk, available to all CHRF staff for ANY IT question
  - Monitored 8x5, triaged within 2 hours

The Research IT Services group in the Division of Biomedical Informatics provides computational resources, services and support to investigators at Cincinnati Children's Research Foundation and the University of Cincinnati College of Medicine.

### Resources

- Software
- Collaboration tools
- Research data warehouse
- Computational cluster
- Data storage
- Database servers
- Web services

### Services

- Basic research support
- Clinical research support
- Software development
- Hosting
- Mac, Linux and Solaris support

### Hot topics

#### Getting started

- System access requests
- Usage policies
- Mission

#### New resources

- Application for use of Cincinnati Genomic Center for data/samples 14 Dec 09
- Mac OS X Snow Leopard (v. 10.6) 28 Aug 09
- Direct network link to UC 19 Jun 09
- Vector NTI desktop sequence analysis software 09 Mar 09
- Group Wise email from SOAP-compatible clients 26 Jul 08

#### Research data storage

- Research Data Storage Request System (available on internal network only)
- Permissions Management System (available on internal network only)
- Completed for research

#### InfoPath/SharePoint

Get more information about these tools including access instructions, training

### Computational cluster

#### Linux Cluster Utilization

last 24 hours

Active CPUs 232 Total CPUs 282 Current Utilization: 82%

Detailed stats, policies, testimonials, user group, more...

### Help

**For desktop support with Windows and Mac,** call the Cincinnati Children's Service Desk, 513-636-4100, or the UCit Help Desk, 513-556-HELP.

**For help with BMI-supported hardware and software,** email the BMI Help Desk.

**For password changes,** go to the following do-it-yourself sites. For all other systems and applications, email the BMI Help Desk.

- Cluster, calendar, wiki, web servers
- BMI Linux desktops, bullpen, internal Citrix servers
- Cincinnati Children's network and email

Local intranet