Research Data Storage & Management at CHRF, Part II

Michael Wagner

Andy Rupert, Nick Felicelli, Alex Spivak

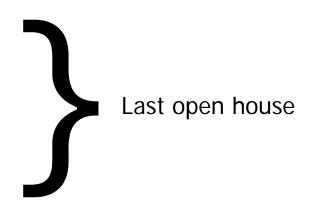
Biomedical Informatics





[Research] Data Storage

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





Data Storage: Best Practices

- Avoid duplication share!
- Avoid external hard drives
 - better, more reliable hardware in institutionally provided storage (RDS, Xythos, ...)
- 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

change the outcome. Word documents, pdf documents

–



Flow Cytometry Data

- Fcs file headers contain information on
 - Acquisition date
 - Cytometer used
 - Experiment name
 - Paramters/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
- 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)



• (FCSData 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 FCSData 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,
 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
- <u>help@bmi.cchmc.org</u> -- BMI help desk
- http://ibmi.cchmc.org -- storage request system
- http://bmirdsdp/DAP -- permissions request system
- http://research.cchmc.org/FCSData
- 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



