CNI 2016 report-back
ECAR/CNI white paper on developing capacity for institutional digital humanities support

- Practical guide
- Capacity-building framework
- “Digital humanities” vs “digital scholarship”
- “Services” vs “partnerships”
ECAR/CNI white paper on developing capacity for institutional digital humanities support

- Getting started
  - Needs assessment
  - Organizational models
- Community engagement
- Communications and outreach
- Funding models
- Governance

- Infrastructure
  - Technology
  - Staffing
  - Facilities
- Acceptance and support (P&T)
### DRAFT - INFRASTRUCTURE

<table>
<thead>
<tr>
<th>Early Stage</th>
<th>Established</th>
<th>High Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researchers working on standard desktops with no support for research applications.</td>
<td>Dedicated support for research projects</td>
<td>Central dedicated DH lab space with dedicated staff</td>
</tr>
<tr>
<td>Research is conducted 'off the side of the desk' with only basic tools</td>
<td>Faculty-wide access to software and hardware</td>
<td>Portal that informs and connects providers and practitioners with available resources/services</td>
</tr>
<tr>
<td>Researchers can take advantage of national high performance computing resources, but may not get campus support for onboarding.</td>
<td>Some physical space where researchers can access research resources, such as a dedicated GIS lab</td>
<td>Infrastructure facilitates seamless collaboration and access; no boundary constraints in networks and systems</td>
</tr>
<tr>
<td>Researcher have to pay for their own specialized software.</td>
<td>HPC resources can be accessed on Campus with dedicated support how understand DH needs.</td>
<td>Data management and preservation practices and systems are operational with clear policies in place</td>
</tr>
<tr>
<td>Little to no guidance on data management or planning</td>
<td>Resources provisioned; DH practitioners know about available resources and how to access them</td>
<td></td>
</tr>
</tbody>
</table>


Cliff Lynch plenary

- Uptake of preprint servers
- Discipline-oriented repositories are more effective for community building
- Bringing museum collections to teaching and research
- Recording and archiving algorithms (e.g. Google search)
Building Tools and Services to Support Research Software Preservation and Sharing

Micah Altman, MIT
Rick Johnson, ARL / Univ. of Notre Dame

Jeffrey Spies, Open Science
Fernando Rios, Johns Hopkins

- Open Science Framework: managing, curating, sharing, and preserving research workflow

OSF integrations make your workflow more efficient

- Dropbox
- GitHub
- Amazon Web Services
- Box
- Google Drive
- figshare
- Dataverse
- Mendeley
Building Tools and Services to Support Research Software Preservation and Sharing

- Software not showing up in institutional repositories
- Lack of progress compared to data
- Force 11 has software citation principles
- Incentives are around publishing, not getting it right
  - Need context for reproducibility, replicability, extensibility
  - Researchers not interested in managing “code”, don’t want to be “coders”
- Preservation environment connected directly to compute environment to archive code
Research IT @ Illinois: Establishing Service Responsive to Investigator Needs

John Towns, Deputy CIO for Research IT, Univ. of Illinois

● “Year of Cyberinfrastructure” initiative
  ○ Create common understanding of resources as part of CI
  ○ Highlight how CI supports research
  ○ Implementation plan for more CI

● 27 focus groups, 130 faculty, 12/14 colleges, 155 p. notes, only 5% faculty

● Key findings:
  ○ 1) Access to expertise
  ○ 2) Communications (“I need X.” “We already have X.” “Never heard of it. Is it what I need?”)
  ○ 3) Data needs
  ○ 4) Tech needs (storage issues, software licensing, access to survey tools, etc.)
Research IT @ Illinois: Establishing Service Responsive to Investigator Needs

- Planned for bold investment in research, then funding didn’t come through
- RIT support
  - Training
  - Communications / marketing
  - RIT portal
  - Research user support
  - Research apps & software development
  - Data viz & analysis service
- Research computing
  - ScienceDMZ
  - IL campus cluster program
  - High throughput computing
  - VM and containers for research
  - Cloud computing for research
Research IT @ Illinois: Establishing Service Responsive to Investigator Needs

● Data Services
  ○ Sensitive data
  ○ REDCap
  ○ HIPAA compliant

● RIT strategy
  ○ Needs collection
  ○ CI master plan
  ○ UIUC IT architecture

● Deferred
  ○ Mapping (Hadoop, Mpreduce)
  ○ Data intensive computing
  ○ Social media lab
  ○ DB as a service
  ○ Grant proposal support service
  ○ Allocations service (unified allocations process for RIT resources and services)
Research IT @ Illinois: Establishing Service Responsive to Investigator Needs

● Data Services
  ○ Sensitive data
  ○ REDCap
  ○ HIPAA compliant

● RIT strategy
  ○ Needs collection
  ○ CI master plan
  ○ UIUC IT architecture

● Deferred
  ○ Mapping (Hadoop, Mpreduce)
  ○ Data intensive computing
  ○ Social media lab
  ○ DB as a service
  ○ Grant proposal support service
  ○ allocations service (unified allocations process for RIT resources and services)
Expanding Research Data Services

Michelle Claibourn, UVirginia
Bryan Sinclair, Georgia State

- Georgia State
  - Established data management advisory team in 2012-2013
    - Business in DMP less robust now
  - Faculty & students say Excel for biggest training need
  - Gaps in support of the actual research phases
    - Drop-in hours for SPSS help
    - Outreach to courses on relevant methods
  - Partnership with VP for Research Office
    - OSF for long-term preservation, meeting with research computing
    - Outreach for help with proposals about restricted data
  - Expanded program is 4 months old, emphasis on communications
Expanding Research Data Services

Michelle Claibourn, UVirginia  
Bryan Sinclair, Georgia State

- UVirginia
  - Primary activity: direct engagement w/researchers, consultation, collaboration
    - 1800 consultations in first 3 years, 60% data analysis wrangling & statistics
    - 375 more consultations from this fall
    - 800 researchers, 45% met with multiple people, repeatedly
    - 45% grad students, targeted that community explicitly
    - 90 workshops in first 3 years
  - Lots of recent changes, more support for data-oriented research
Expanding Research Data Services

CADRE Computation and Data Resource Exchange

CADRE is a University-wide exchange designed to simplify the process of finding local research technology solutions. The CADRE website provides a comprehensive list of the software, services, training opportunities, storage solutions, and computational hardware available to UVA scholars. CADRE is multi-disciplinary and supports all eleven schools of the University.

Data Analysis & Software
Consulting and tools for data analysis, data wrangling, and statistics

Computation
High-performance computing, big data solutions, Rivanna, project consulting, programming language communities

Visualization Services
Hardware and software tools for visualizing data, medical imagery

Storage
Short-term and long-term storage options for any size dataset
Expanding Research Data Services

CADRE Computation and Data Resource Exchange

Need Help Contact a CADRE Agent

Computation

University of Virginia scholars have access to computing systems that can significantly increase the speed of calculations, analyze and store large data sets, and perform advanced simulations.

Rivanna

Rivanna is a high-performance computing cluster suitable for a range of computational tasks. The major section of the system consists of 240 nodes with 128 GB of RAM and 20 cores each, connected through a high-speed InfiniBand internal network. These nodes can be used for computationally-intensive jobs and parallel programs. Approximately 1200 additional cores are available for serial programs. Four high-memory nodes, each with 16 cores and 1 TB of RAM, can be accessed by request. Two nodes contain NVIDIA K20 GPUs along with 16 cores and 256 GB of RAM, also accessible by request. All nodes have access to 1.4 PB of fast Lustre disk storage for temporary working files. Information on longer-term storage options can be found on the CADRE storage page. A technical description of the cluster is available on the Rivanna specifications page.

To request a startup allocation, complete the request form.

Ivy: Secure Compute Environment
Documenting the Now

http://app.docnow.io

- Affordances for ethical practice
- Notifications into tweet stream “Researcher X is doing data collection for Y reason, go here to opt out.”
- Twitter API requirements: store only tweet IDs for later rehydration
- Data retention policies
- Traditional knowledge labels (inspired by Mukurtu)
Disciplinary repositories

John Howard, Univ. College Dublin  Francis McManamon, ASU

- Consortium of European Social Science Data Archives (CESSDA)
  - Take-away point: easier access to European data sets

- The Digital Archaeological Record (tDAR)
  - Based out of ASU
  - Numerous large government contracts
  - Has per-item fee, meant to cover long-term preservation
    - “Item” = up to 10 MB
  - More focused on archiving than Open Context, better metadata templates