How to demonstrate the value of research computing and data

Jackie Milhans, Manager Computing and Data Support Services, Northwestern University

EDUCAUSE Research Computing and Data CG
February 25, 2021
Many contributed to the collection, presentation, and visualization of this data:

- Christina Maimone, Lead, Research Data Services
- Arman Pazouki, Lead, Computational Services
- Alper Kinaci, Lead Computational Research Engineer
- Janna Nugent, Senior Bioinformatics Specialist
Agenda

• Audience
• What story are you trying to tell?
• KPI Examples
Who Is Your Audience?

- IT leadership, Institutional leadership, etc.
- Your boss
- Your team
- Your “customers” or researchers
What Do You Need to Show?

SUCCESES

CONSTRAINTS

PROGRESS
What is the story that you need to tell?
OVERALL IMPACT
How Much Research is Supported?

Report RCD value by understanding the number of research dollars you potentially support.

$400+M FY19
$740+M FY20

Based on research dollars associated with PIs using HPC and research storage services. Does not include people who attended workshops and training.

Steps Involved: Requires gathering University IDs and working with the Office for Sponsored Research.
## Percent HPC Researchers by School/College

<table>
<thead>
<tr>
<th>School</th>
<th>FY20 Buy-In</th>
<th>FY20 General Access</th>
<th>FY21 Buy-In</th>
<th>FY21 General Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicine</td>
<td>34%</td>
<td>22%</td>
<td>34%</td>
<td>30%</td>
</tr>
<tr>
<td>Arts and Sciences</td>
<td>22%</td>
<td>24%</td>
<td>22%</td>
<td>24%</td>
</tr>
<tr>
<td>Engineering</td>
<td>18%</td>
<td>24%</td>
<td>18%</td>
<td>26%</td>
</tr>
<tr>
<td>Other (Office for Research, IT, smaller schools, etc.)</td>
<td>8%</td>
<td>4%</td>
<td>8%</td>
<td>4%</td>
</tr>
<tr>
<td>Journalism</td>
<td>2%</td>
<td>0%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>Management</td>
<td>2%</td>
<td>4%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>Law</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Communication</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>2%</td>
</tr>
</tbody>
</table>

- This shows the diversity and breadth of the research supported. Demonstrate skills and staff needed to support researchers.
- For institutions looking for investments from schools or departments, understand where usage is.
- For institutions focusing on specific schools to build out research areas, understand impact and service adoption.
INFRASTRUCTURE CAPACITY
Cluster Wait Times

- General Access is proposal-based
- Buy-In is purchased, dedicated resources
- Demonstrate capacity constraints
- Demonstrate effectiveness of scheduler tuning or other efforts
Proposal Access Research Allocations

Requested

- Demonstrate interest and service adoption
- Demonstrate ability to fulfill proposals due to infrastructure capacity
Jobs Completed per Month

Total Number of Jobs (millions)

- Show impact and infrastructure capacity
- This can be tricky, since it can be difficult to report total capacity of compute available or how it relates to total compute hours
Total Compute Hours

- Tricky, since total available is difficult to compute.
- Usually based on resources available, but can show they are being used.
Demand for Data Storage

Adoption and Capacity Projections (Petabytes)

- Demonstrate total capacity and service adoption.
- Maintenance of storage with extreme growth requires staff to manage platform and infrastructure investment.

Growth of ~3% (~200 TB) of new data per month
SUPPORT AND PEOPLE
New HPC Users per Month

- Several new users need onboarding and support
- This also shows service adoption and demand
HPC Consultations

- Demonstrate support requirements and gaps
- Define ability to provide this level of service
Ticket Support

- Demonstrate effect of an employee out of the office
- Ability to respond in timely fashion
- Ability to resolve and troubleshoot issues
- Improvements could show effects of improved documentation, videos, increased staff, improved user interfaces, etc.
Demand for Data Help

Demand for Data Consultations is Growing
Averaging 22 consults per month in 2020

Workshop Registration Hours More than Doubled in FY20
Workshops Taught or Sponsored by Research Computing Services

Demonstrate demand, support needed, and ability to provide workshops to more researchers by teaching remotely.
Data Services Workshop Hours

Workshop Hours by Year

Note changes to format in remote world:
To better accommodate a remote format, we offered shorter workshops more frequently rather than longer in-person sessions.

Demonstrate total work, instead of number of workshops, since they can vary in length.
Constraints and Approaches - Data

Demand for DataCamp Online Training Continues

- Winter 2021
- Fall 2020
- Summer 2020
- Spring 2020
- Winter 2020
- Fall 2019
- Summer 2019
- Spring 2019
- Winter 2019
- Fall 2018
- Summer 2018
- Spring 2018
- Winter 2018
- Fall 2017
- Summer 2017
- March 2017
- February 2017
- December 2016

BYOD Working Groups
Demand for new program exceeds capacity

- Fall 2020
- Summer 2020

Longer-term guidance and support for a project, by proposal.

Demand for online training. Requires funding and partial staff to manage.
## Service Levels and Staffing

### Research Computing and Data: Evolution of Support Services

<table>
<thead>
<tr>
<th>Staffing Requirements</th>
<th>FY13</th>
<th>FY14</th>
<th>FY15</th>
<th>FY16</th>
<th>FY17</th>
<th>FY18</th>
<th>FY19</th>
<th>FY20</th>
<th>FY21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computational Research</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Operational Support</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Onboarding Consultations</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Research Project Support</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Data and Computing Environments</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Training</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Services and Infrastructure Staff</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Data Science</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Data Science Consultations</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Research Project Support</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Data Visualization</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Training</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Services Staff</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Data Workflow and Implementation</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Staff for Fee</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Info Security/Storage Support</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
</tbody>
</table>

### Service Levels:

- **NONE**: Only documentation provided, service eliminated completely, or service never provided despite demonstrated demand.
- **REDUCED**: Unable to meet a significant portion of demand. Portions of service have been eliminated.
- **LIMITED**: No capacity to develop or expand service to support new research areas or existing service. Provide basic support for existing services and existing research areas. Does not meet full demand of requests for service.
- **FULL**: Capacity to respond to most service requests within service definition. Capacity to take on a limited number of novel projects, such as new research approaches or service improvements.
- **INNOVATING**: Capacity to support all service requests within service definition, including novel requests. Service meets modern research demands and evolves to support new research areas adopting service. Allows Northwestern to be a leader in this area.

- Demonstrate service levels
- Show evolution of number of staff
- Request staffing
Other Metrics

• Number of users
• Number of projects supported
• Hours spent on projects
• Net promoter or survey results
• PIs with allocations or Usage by PI
• Staff time spent in operations, projects, leave, professional development
Questions and Discussion

• Thank you to Northwestern Research Computing Services staff for the collection and visualization of data.

• Additional questions: milhans@northwestern.edu