Decentralized Approach to RDM deployment

Andrew White
Director, Library Information Services
Rensselaer Polytechnic Institute

December 5, 2019
Institutional Research Objectives

- Enhance institutional ability to pursue major partnerships
  - corporate, federal, philanthropic, and state partnerships

- Grow / obtain research funding
  - facilities, specialized equipment
  - maintenance / operating capital

- Strengthen basic research endeavors across the Institute
  - Recruit and support leading faculty in identified priority areas

- Expand research reach and impact

- Measure key outcomes related to impact and leadership
Trending RDM role for Libraries in Service to Institution

- Establish institutional repositories
  - for open access research articles
  - for raw research data sets.

“...a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members.”

Feasibility and Operational Library Considerations for Smaller Institutions

- Relatively small number of library staff

- Highly decentralized and federated scholarly communications ecosystem
  - various repositories, like PubMed or arXiv, have their own set of data and publication storing options.

- Service offerings from scholarly publishers
  - deposit and reference research data sets in conjunction with article publication.

https://techcrunch.com/2015/01/10/decentralize-all-the-things/
The New Research Library: Services and ROI proposition

- Enable the archival of research data
  - Develop / assign appropriate ontologies / metadata

- Foster unique cross-disciplinary collaborations
  - Create linkages to facilitate discoverability for researchers across discipline and data silos

- Contribute to cyber-security strategies
  - Provide a check that data is stored in a secure location (but NOT BE that location)
Data Inventory Service

- Assist Rensselaer Libraries to keep track of research data availability and its location.

- The research data availability for each article is first checked automatically via Scholix and Mendeley Data.

- If no data is automatically found, researchers get emails and asked to provide information about data location via an interactive survey.

- The interactive survey is integrated with DataCite.

- All collected information is presented via the Data Monitor dashboards
Congratulations on the publication of your article ‘COMPASS - New modeling and simulation approach to PWR in-vessel accident progression’ by Podowski M.Z., Podowski R.M., Kim D.H., Bae J.H., Son D.G. in Nuclear Engineering and Technology.

Rensselaer Libraries encourages the depositing of research data associated with publication output. We are interested in knowing, when possible, where research data sets might be stored, both to help increase the awareness of your research and to spur collaboration across various disciplines within the Institute. While making research data available online can enhance the discovery and impact of research activities, many federal funding agencies are increasingly requiring that data from supported projects be placed in an online repository as part of grant compliance.

Could you please provide us with some information about this article’s associated research data so that we may better understand and promote the Institute’s research output?

Do you have research data associated with the article ‘COMPASS - New modeling and simulation approach to PWR in-vessel accident progression’?

[ ] Yes [ ] No

Thank you in advance for your contribution.

Sincerely,
Andrew White, PhD
Director of Library Information Services

You can find more information about this project [here](#). For assistance, [please contact us](#).
Confirmation step before the survey is loaded

If dataset has a DOI, metadata will be automatically pre-populated, including the name of repository
Data Location Details

for the article: A 0.4% PAE 194-GHz Signal Source with 1.5-mW Output Power in 65-nm Bulk CMOS Technology

- The research data has been deposited to a repository.

Repository name:

Persistent Identifier (e.g. 10.3517/ccdc.csd.ccl1m1sq5, GSE39549, 4hhb, not sure)

Next

Submit

Data Location Details

for the article: A 0.4% PAE 194-GHz Signal Source with 1.5-mW Output Power in 65-nm Bulk CMOS Technology

- The research data has been deposited to a repository.
- The research data is private.
- The research data is stored locally.

Please provide details of the storage and file formats (e.g. Excel files stored on a lab server, 10MB):

Please provide name and email address of the contact person to reach out for more information:

- I don't know how to deposit the data.
- I cannot deposit the data.

Submit
Results of the third test email campaign as they appear on the Data Monitor/Lighthouse dashboard
### Research articles checked - 670

**Data inventory**

<table>
<thead>
<tr>
<th>First name</th>
<th>Last name</th>
<th>Article DOI</th>
<th>Selected data location option</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luigi</td>
<td>Vanhetti</td>
<td>2.10.73/2015.11252641</td>
<td>The research data has been deposited.</td>
<td>mrp files on a lab server, 100GB (contact details: Elena, <a href="mailto:mpreperts23@gmail.com">mpreperts23@gmail.com</a>)</td>
</tr>
<tr>
<td>Ravishankar</td>
<td>Sundaraman</td>
<td>2.10.73/2015.11252684</td>
<td>The research data is stored locally.</td>
<td>mrp files on a lab server, 100GB (contact details: Elena, <a href="mailto:mpreperts23@gmail.com">mpreperts23@gmail.com</a>)</td>
</tr>
</tbody>
</table>

**Datasets & links**

<table>
<thead>
<tr>
<th>First name</th>
<th>Last name</th>
<th>Article DOI</th>
<th>Dataset title</th>
<th>Repository name</th>
<th>Other contributors</th>
<th>Persistent identifier</th>
<th>Linked (Yes/No)</th>
<th>School (Yes/No)</th>
<th>Original Data</th>
<th>Linking Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xing</td>
<td>Wang</td>
<td>10.10.73/2016.11252640</td>
<td>Chromosome fusions triggered by noncoding RNA</td>
<td>Figshare</td>
<td>Grace J. Uttarabai, Derek M. Clay, Xiao Chen, Sierra S. McCloud, Jingmei Wang, Xing Wang, Kaehtli Shetty, Mariliz Nowak, John R. Brattin, Eran C. Callahan, Laura R. Landweber</td>
<td>10.10.73/2015.34252777</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Limited</td>
</tr>
<tr>
<td>Xing</td>
<td>Wang</td>
<td>10.10.73/2016.11252640</td>
<td>Chromosome fusions triggered by noncoding RNA</td>
<td>Figshare</td>
<td>Grace J. Uttarabai, Derek M. Clay, Xiao Chen, Sierra S. McCloud, Jingmei Wang, Xing Wang, Kaehtli Shetty, Mariliz Nowak, John R. Brattin, Eran C. Callahan, Laura R. Landweber</td>
<td>10.10.73/2015.3420283</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Limited</td>
</tr>
<tr>
<td>Xing</td>
<td>Wang</td>
<td>10.10.73/2016.11252640</td>
<td>Chromosome fusions triggered by noncoding RNA</td>
<td>Figshare</td>
<td>Grace J. Uttarabai, Derek M. Clay, Xiao Chen, Sierra S. McCloud, Jingmei Wang, Xing Wang, Kaehtli Shetty, Mariliz Nowak, John R. Brattin, Eran C. Callahan, Laura R. Landweber</td>
<td>10.10.73/2015.34752777</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Limited</td>
</tr>
<tr>
<td>Xing</td>
<td>Wang</td>
<td>10.10.73/2016.11252640</td>
<td>Chromosome fusions triggered by noncoding RNA</td>
<td>Figshare</td>
<td>Grace J. Uttarabai, Derek M. Clay, Xiao Chen, Sierra S. McCloud, Jingmei Wang, Xing Wang, Kaehtli Shetty, Mariliz Nowak, John R. Brattin, Eran C. Callahan, Laura R. Landweber</td>
<td>10.10.73/2015.3264886</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Limited</td>
</tr>
</tbody>
</table>

**Unsubscribed from this email campaign**

<table>
<thead>
<tr>
<th>First name</th>
<th>Last name</th>
<th>Article DOI</th>
<th>Opt-out reason</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ryan</td>
<td>Gilbert</td>
<td>2.10.73/2015.11252640</td>
<td>I cannot share the data.</td>
<td>I don't have the data</td>
</tr>
</tbody>
</table>
Articles with associated research data - trial #1

Articles that have associated research data – 32 (4.8%):

- Deposited & linked data (found automatically) - 14
- Deposited data - 5
- Stored locally data - 9
- Private data - 1
- Data belongs to a different university or person who left - 2
- Needs advice what to do with the data - 1

Datasets related to the articles deposited to the open data repositories

<table>
<thead>
<tr>
<th>Institution</th>
<th>Covered period</th>
<th>Articles checked</th>
<th>Shared datasets</th>
<th>Linked datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rensselaer Polytechnic Institute</td>
<td>2017-01-25 - 2018-05-10</td>
<td>661</td>
<td>35</td>
<td>30</td>
</tr>
</tbody>
</table>
Articles that have associated research data – 11 (4.5%):
- Deposited & linked data (found automatically) - 3
- Deposited data - 1
- Stored locally data - 5
- Private data - 1
- Needs advice what to do with the data - 1

Datasets related to the articles deposited to the open data repositories

<table>
<thead>
<tr>
<th>Institution</th>
<th>Covered period</th>
<th>Articles checked</th>
<th>Shared datasets</th>
<th>Linked datasets</th>
</tr>
</thead>
</table>
Trial 1 Summary

- Test run April 24 –May 10, 2018.
- 406 RPI faculty members were included in the test email campaign.
- The response rate was very high:
  - from 217 RPI researchers who received at least one email 19% responded to emails.
- 661 research articles published by the RPI faculty members were checked for the data availability.
- Based on results of the automatic checks and researcher responses: 5% articles have associated research data.
  - This is inline with the expected outcome.
- Most popular repositories:
  - CCDC
  - Figshare
  - Dryad
  - LTER Network Data Portal
  - Mendeley Data.
- Based on researcher responses, 11% articles don’t have research data.
  - However, manual check of these 74 articles revealed that 34% may have associated research data.
Test run October 30 – November 25, 2019.

100 faculty members were included in the test email campaign.

The response rate was very high:

- 19% responded to emails.

261 research articles and conference papers published by the RPI faculty members were checked for the data availability.

Based on results of the automatic checks and researcher responses: 4.5% articles have associated research data.

This is inline with the expected outcome.

Most popular repositories:

- Cambridge Crystallographic Data Centre
- Figshare
- eLife
- IRIS DMC

Information regarding data availability for 25 articles (10%) was collected and recorded in Data Monitor/Lighthouse during the campaign.
Lessons learned from pilots

- A few researchers actively deposited data after the campaign and shared links
- Need for faculty education about RDM
- Research activities and processes heavily dependent upon graduate assistants
- Many researchers do not yet see the value of DMP
  - Have not experienced “penalties” for not depositing
- Definition of research data varies greatly by
  - research foci
  - research workload
  - personality
  - technologies
Thank You

whitea9@rpi.edu