



# Rensselaer

why not change the world?®

## 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

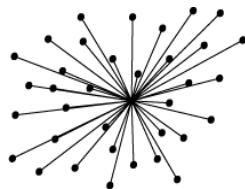
- for open access research articles
- for raw research data sets.

“...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.”

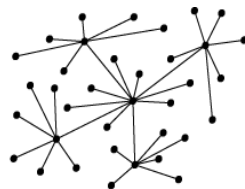
Clifford A. Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," *ARL Bimonthly Report* 226 (February 2003), 1-7.

# Feasibility and Operational Library Considerations for Smaller Institutions

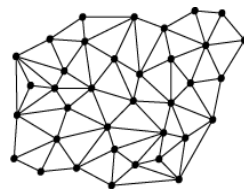
- Relatively small number of library staff
- Highly decentralized and federated scholarly communications ecosystem
  - various repositories, like PubMed or arXiv, have their own set of data and publication storing options.
- Service offerings from scholarly publishers
  - deposit and reference research data sets in conjunction with article publication.



centralised



decentralised



distributed

<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)



- 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

You will now be redirected to the online survey. Would you like to continue?

Yes No



### Data Location Details

for the article: *Waxing and cultivar affect Salmonella enterica persistence on cucumber (Cucumis sativus L.) fruit*

☐ The research data has been deposited to a repository.

☐ The research data is stored as supplementary material with the article.

☐ The research data is private.

☐ The research data is stored locally.

☐ I don't know how to deposit the data.

☐ I cannot deposit the data.

Submit

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

### Data Location Details

for the article: *Waxing and cultivar affect Salmonella enterica persistence on cucumber (Cucumis sativus L.) fruit*

☒ The research data has been deposited to a repository.

Persistent Identifier (e.g. 10.5517/ccdc.csd.cc1m1sq5, GSE39549, 4hhb, not sure)\*

Next

+ Add Dataset

☐ The research data is stored as supplementary material with the article.

☐ The research data is private.

☐ The research data is stored locally.

☐ I don't know how to deposit the data.

☐ I cannot deposit the data.

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.

Repository name\*

Persistent Identifier (e.g. 10.5517/ccdc.csd.cc1m1sq5, GSE39549, 4hhb, not sure)\*

Next

+ Add Dataset

- ☐ The research data is private.
- ☐ The research data is stored locally.
- ☐ I don't know how to deposit the data.
- ☐ I cannot deposit the data.

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

Dashboard

Management

Research articles checked - 246

Data inventory

First name	Last name	Email	Article DOI	Scopus Article ID	Selected data location option	Additional information
Rahmi	Ozisk	ozisk@rpi.edu	10.1016/j.carbon.2019.05.072	2-s2.0-85067040079	The research data is stored locally.	Various files in various formats. (contact details: J.C. Cabanelas, caba@ling.uc3m.es)
Wayne	Gray	grayw@rpi.edu	10.1016/j.cogpsych.2018.11.003	2-s2.0-85057981570	I don't know how to deposit the data.	grayw@rpi.edu
Steven	Roecker	roecs@rpi.edu	10.1029/2019GL082780	2-s2.0-85068208407	The research data has been deposited.	
Joel	Giedt	giedtj@rpi.edu	10.1103/PhysRevD.98.095017	2-s2.0-85058084673	The research data is stored locally.	Stored on my group's repository. File format is ASCII file. (contact details: Joel Giedt, giedtj@rpi.edu)
Mamadou	Diagne	diagnm@rpi.edu	10.1137/1.9781611975758.15	2-s2.0-85073907820	I cannot deposit the data.	
Antoinette	Maniatty	manias@rpi.edu	10.1109/TCPMT.2018.2886813	2-s2.0-85058878359	The research data is stored locally.	ASCII files (contact details: Yuanxiang Zhang, jarson_zhang@hotmail.com)
Antoinette	Maniatty	manias@rpi.edu	10.1109/TCPMT.2019.2940566	2-s2.0-85072523950	The research data is stored locally.	Excel files stored on lab server (contact details: Matt Ring, Matt.Ring@onseml.com)
Antoinette	Maniatty	manias@rpi.edu	10.1007/s00466-017-1496-y	2-s2.0-85032514375	The research data is stored locally.	Excel files stored on lab computer (contact details: Antoinette Maniatty, manias@rpi.edu)

Datasets & links

< 1 >

First name	Last name	Email	Article DOI	Scopus Article ID	Dataset title	Repository name	Dataset URL	Other contributors	Persistent identifier	Linked (Yes/No)	Scholix (Yes/No)	Original Data	Linking Status
Chunyu	Wang	wangc5@rpi.edu	10.1038/s41422-019-0153-8	2-s2.0-85062881756	Abstract	eLife Sciences Publications, Ltd			10.7554/elife.12813.001	Yes	Yes	Yes	Linked
Cynthia	Collins	collic3@rpi.edu	10.1186/s12934-019-1187-9	2-s2.0-85070571049	MOESM1 of Metabolic engineering of Bacillus megaterium for heparosan biosynthesis using Pasteurella multocida heparosan synthase, PmHS2	figshare	https://springernatur.../articles/1/MOESM1_of_Meta.../9565850	Yanlei Yu, Dordick, Jonathan, Mattheos Koffas, Williams, Asher, Vaidyanathan, Deepika, Gedeon, Kamil, Collins, Cynthia, Linhardt, Robert	10.6084/m9.figshare.9565850	Yes	Yes	Yes	Linked
Cynthia	Collins	collic3@rpi.edu	10.1186/s12934-019-1187-9	2-s2.0-85070571049	MOESM1 of Metabolic engineering of Bacillus megaterium for heparosan biosynthesis using Pasteurella multocida heparosan synthase, PmHS2	figshare	https://springernatur.../articles/1/MOESM1_of_Meta.../9565850/1	Yanlei Yu, Dordick, Jonathan, Mattheos Koffas, Williams, Asher, Vaidyanathan, Deepika, Gedeon, Kamil, Collins, Cynthia, Linhardt, Robert	10.6084/m9.figshare.9565850.v1	Yes	Yes	Yes	Linked
Cynthia	Collins	collic3@rpi.edu	10.1186/s12934-019-1187-9	2-s2.0-85070571049	Metabolic engineering of Bacillus megaterium for heparosan biosynthesis using Pasteurella multocida heparosan synthase, PmHS2	figshare	https://springernatur.../collections/1/Metabolic_engineeri.../4615439/1	Yanlei Yu, Dordick, Jonathan, Mattheos Koffas, Williams, Asher, Vaidyanathan, Deepika, Gedeon, Kamil, Collins, Cynthia, Linhardt, Robert	10.6084/m9.figshare.c4615439.v1	Yes	Yes	Yes	Linked
K. V.	Lakshmi	lakshk@rpi.edu	10.1021/acs.organomet.9b00134	2-s2.0-85063089966	CCDC 1899152: Experimental Crystal Structure Determination	Cambridge Crystallographic	http://www.ccdc.cam.../services	Cai, Irene C., Nicolay, Amélie, Ziegler, Micah S., Tilley, T. Don,	10.5517/ccdc.csd.cc21r6z0	Yes	Yes	Yes	Linked

Properties Email campaign dashboard

## Research articles checked - 670

### Data inventory

First name	Last name	Article DOI	Selected data location option	Additional information
Luigi	Vanfretti	2-s2.0-85021256941	The research data has been deposited.	
Ravishankar	Sundararaman	2-s2.0-85014359684	The research data is stored locally.	mp4 files on a lab server, 100GB (contact details: Elena, myreports23@gmail.com)

YES

### Datasets & links

< 1 2 >

First name	Last name	Article DOI	Dataset title	Repository name	Dataset URL	Other contributors	Persistent identifier	Linked (Yes/No)	Scholar (Yes/No)	Original Data	Linking Status
Xing	Wang	10.1080/15476286.2016.1195940	Chromosome fusions triggered by noncoding RNA	Figshare		Grace J. Uttarotai, Derek M. Clay, Xiao Chen, Sierra S. McCloud, Jingmei Wang, Xing Wang, Keerthi Shetty, Mariusz Nowacki, John R. Bracht, Evan C. Callihan, Laura F. Landweber	10.6084/m9.figshare.3475277.v1	Yes	Yes	Yes	Linked
Xing	Wang	10.1080/15476286.2016.1195940	Chromosome fusions triggered by noncoding RNA	Figshare		Grace J. Uttarotai, Derek M. Clay, Xiao Chen, Sierra S. McCloud, Jingmei Wang, Xing Wang, Keerthi Shetty, Mariusz Nowacki, John R. Bracht, Evan C. Callihan, Laura F. Landweber	10.6084/m9.figshare.3422083	Yes	Yes	Yes	Linked
Xing	Wang	10.1080/15476286.2016.1195940	Chromosome fusions triggered by noncoding RNA	Figshare		Grace J. Uttarotai, Derek M. Clay, Xiao Chen, Sierra S. McCloud, Jingmei Wang, Xing Wang, Keerthi Shetty, Mariusz Nowacki, John R. Bracht, Evan C. Callihan, Laura F. Landweber	10.6084/m9.figshare.3475277	Yes	Yes	Yes	Linked
Xing	Wang	10.1080/15476286.2016.1195940	Chromosome fusions triggered by noncoding RNA	Figshare		Grace J. Uttarotai, Derek M. Clay, Xiao Chen, Sierra S. McCloud, Jingmei Wang, Xing Wang, Keerthi Shetty, Mariusz Nowacki, John R. Bracht, Evan C. Callihan, Laura F. Landweber	10.6084/m9.figshare.c.3266887.v1	Yes	Yes	Yes	Linked

Luigi	Vanfretti	10.1109/TPWRS.2017.2679131	Data for: Intramembranal disulfide cross-linking elucidates the super-quaternary structure of mammalian CatSper	Mendeley Data			10.17632/jkp2rxdy8.1	Yes	No	Yes	
-------	-----------	----------------------------	---	---------------	--	--	----------------------	-----	----	-----	--

< 1 2 >

### Unsubscribed from this email campaign

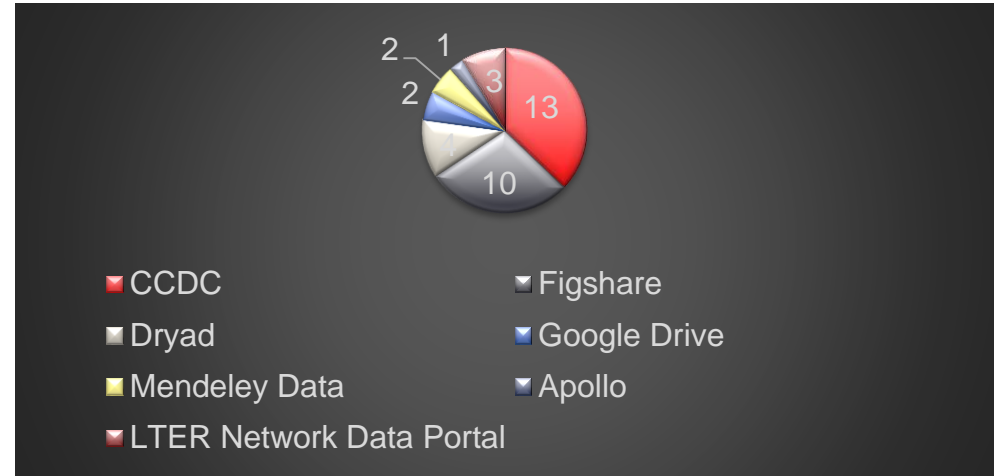
First name	Last name	Article DOI	Opt-out reason	Additional information
Ryan	Gilbert	2-s2.0-85016396572	I cannot share the data.	I don't have the data

NO

# 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



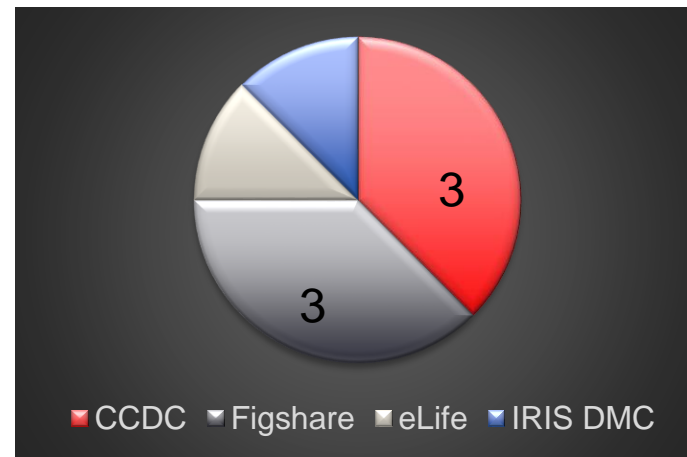
Datasets related to the articles deposited to the open data repositories

Institution	Covered period	Articles checked	Shared datasets	Linked datasets
Rensselaer Polytechnic Institute	2017-01-25 - 2018-05-10	661	35	30

# Articles with associated research data - trial #3

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

Institution	Covered period	Articles checked	Shared datasets	Linked datasets
Rensselaer Polytechnic Institute	2018-06-01 – 2019-11-25	246	8	7

- 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



Q&A



# Rensselaer

**why not change the world?®**

Thank You

[whitea9@rpi.edu](mailto:whitea9@rpi.edu)