

A decorative graphic on a black background. It features a large purple L-shaped block in the top-left corner. Below it is a blue L-shaped block. A horizontal bar with four colored segments (purple, yellow, yellow, red) spans across the middle. On the right side, there is a red rounded rectangle above a yellow rectangle.

# **Liaisons and Data Management**

Data Sets for Librarians, and  
Engaging our Faculty

Brenda Dingley and Scott Curtis

# What's all this talk about data?

When we think about research papers, we often think about original data generated by the researcher through experiment, survey, etc..

However, in an increasingly popular research method, researchers use “raw” data sets from other original research sources.

# “Raw” data set research

This method is used across a wide variety of disciplines, including

- STEM fields
- Economics
- Public Policy
- Business
- Education
- Linguistics
- ....and on and on!

# What to do with data sets?

- Data visualization / “infographics”
- Data comparison across data sets and domains
- Longitudinal studies
- Research on secondary / tertiary analysis methods
- *....and on and on!*

# Where can we find data sets?

Open Data - “Open data is data that can be freely used, reused, and distributed by anyone - subject only, at most, to the requirement to attribute and share-alike.”

- Open Access Directory [list of open data repositories](#)
- [Databib](#) - a directory to 975 open data repositories across disciplines (Purdue and Penn State)
- [re3data](#) - the REgistry of REsearch data REpositories (German Research Foundation)
- [data.gov](#) - “The home of the US Government’s open data”
- Of course Google is getting into the act with its [Google Public](#) data explorer

# Evolution of Federal Data Archiving Policy

## NSF Data Sharing Policy

Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing.“

**January 18, 2011:** NSF requires proposals to include a two page data management plan.

**February 2013:** OSTP memorandum to heads of federal agencies: “Increasing Access to the Results of Federally Funded Scientific Research.” Science agencies with research budgets of \$100 million + annually directed to create a plan to make the results of their funded research freely available to the public. Unlike NIH directive, research results includes underlying data supporting publication results.

# Why Open Data?

1. Reproducibility
2. Detect error or malfeasance
3. Lessen duplicative and expensive data-gathering
4. Create economic opportunity

## Evolution of Federal Data Archiving Policy, cont.

### May 9, 2013: Executive Order: “Making Open and Machine Readable the New Default for Government Information”

“To promote continued job growth, Government efficiency, and the social good that can be gained from opening Government data to the public, the default state of new and modernized Government information resources shall be open and machine readable. Government information shall be managed . . . to promote interoperability and openness, and, . . . to ensure that data are released to the public in ways that make the data easy to find, accessible, and usable.”

See [www.data.gov](http://www.data.gov)



## Evolution of Federal Data Archiving Policy, cont

March 20, 2014: OSTP report on progress in response to OSTP open access memorandum.

May 9, 2014: “US Open Data Action Plan”

Endorses CC0 license for data sets.

# Open Data Principles go global

June 2013: G8 leaders endorse the Open Data Charter

Principle 1: Open Data by Default

Principle 2: Quality and Quantity

Principle 3: Usable by All

Principle 4: Releasing Data for Improved Governance

Principle 5: Releasing Data for Innovation

# Data management challenges

Managing data through its life cycle presents greater challenges than managing digitized text.

- Size of files
- Variety of file types
- Usability of files

# UMKC Data Pilot project

## Goals of project:

1. Respond to funder mandates
2. Test concept of MOspace as data repository
3. Determine researchers' desired repository attributes
4. Develop policy and method to credit data contributor(s)
5. Create appropriate metadata schema for data
6. Generate reports showing demand for data sets

# UMKC Data Pilot project, con't.

University of Missouri System  
COLUMBIA | KANSAS CITY | ROLLA | ST. LOUIS

Login

# MOspace

Preserving the intellectual output & resources of the University of Missouri.

[search](#) | [browse](#) | [add to MOspace](#) | [about](#) | [help](#)

[MOspace Home](#) > [University of Missouri-Kansas City](#) >

## University of Missouri-Kansas City

**Browse by**

- [By Issue Date](#)
- [Authors](#)
- [Titles](#)
- [Subjects](#)
- [Theses Department](#)
- [Theses Semester](#)
- [Theses Advisor](#)

Search within this community and its collections:



## Shifting Gears - Engaging Faculty / Scholarly Communication

To date, faculty usage of our *institutional repository* (MOspace) has been spotty. At least in my STEM fields faculty are not utilizing disciplinary repositories effectively.

There are now mandates for making government grant-sponsored research *publicly accessible*, and for *data management plans* to allow greater sharing of research data between research teams.

Also, faculty are becoming aware of *authors' rights issues* “through the backdoor,” due to finding that UMKC Libraries cannot subscribe to every journal they would publish in or consult in their research.

# A possible strategy...

In order to engage faculty in a larger discussion of scholarly communication, I've started the following program:

1. Who are the tenure-track faculty in a department? Find them through department web pages.
2. Use database tools to determine number of articles produced by tenure-track faculty in a given period.
3. Where are they publishing? What journals? Which conferences? Are there a wide-variety of publishers, or only a few?
4. Find out if the publishers they have been using have a self-archiving policy. (How? Look at [SHERPA/RoMEO](#))
5. Communicate with department to let them know that UMKC Libraries will take care of archiving; tell them to NOT sign standard author agreements but to request self-archiving clause, and to see if publishers will apply self-archiving clause retroactively.


# Strategy Example: CSEE, Step 1

## Step 1: Who are they?

School of Computing and Engineering f t in

Home Our School Degree Programs Continuing Education Research Careers Scholarships Contact Us

Our School > CSEE Department



- ABOUT US
- CHAIR'S WELCOME
- ADVISORY BOARD
- CSEE FACULTY & STAFF**  
COMPUTER SCIENCE ELECTRICAL ENGINEERING

**Academic Administrators**

- + Place, Jerry, Ph.D.  
Associate Dean for Academic Affairs  
Associate Professor
- + Chaudhry, Ghulam, Ph.D.  
Chair of Computer Science Electrical Engineering Dept  
Professor

**Faculty**

- + Beard, Cory, Ph.D.  
Associate Professor
- + Burris, Eddie, M.S.  
Associate Teaching Professor

14) NOTE - This Directory Includes REGULAR TENURE TRACK Faculty Only - no Teaching P

First Name	Last Name	Rank	e-mail	Website?
Jerry	Place	Associate	<a href="mailto:placej@umkc.edu">placej@umkc.edu</a>	
Ghulam	Chaudhry	Professor	<a href="mailto:chaudhryg@umkc.edu">chaudhryg@umkc.edu</a>	
Cory	Beard	Associate	<a href="mailto:beardc@umkc.edu">beardc@umkc.edu</a>	<a href="http://sce.umkc.edu">http://sce.umkc.edu</a>
Deb	Chatterjee	Associate	<a href="mailto:chatd@umkc.edu">chatd@umkc.edu</a>	
Baek-Young	Choi	Associate	<a href="mailto:choiby@umkc.edu">choiby@umkc.edu</a>	<a href="http://sce.umkc.edu">http://sce.umkc.edu</a>
Masud	Chowdhury	Associate	<a href="mailto:masud@umkc.edu">masud@umkc.edu</a>	<a href="http://sce.umkc.edu">http://sce.umkc.edu</a>
Reza	Derakshani	Associate	<a href="mailto:reza@umkc.edu">reza@umkc.edu</a>	<a href="http://sce.umkc.edu">http://sce.umkc.edu</a>
Deendayal	Dinakarpanian	Associate	<a href="mailto:dinakard@umkc.edu">dinakard@umkc.edu</a>	<a href="http://sce.umkc.edu">http://sce.umkc.edu</a>
Yijie	Han	Associate	<a href="mailto:hanyij@umkc.edu">hanyij@umkc.edu</a>	<a href="http://h.web.umkc.edu">http://h.web.umkc.edu</a>
Lein	Harn	Professor	<a href="mailto:harnl@umkc.edu">harnl@umkc.edu</a>	<a href="http://h.web.umkc.edu">http://h.web.umkc.edu</a>
Jerome	Knopp	Associate	<a href="mailto:knoppj@umkc.edu">knoppj@umkc.edu</a>	



# Strategy Example: CSEE, Step 2

Step 2: What have they published? I did a “Scopus Scrape” for 2010-2014

The screenshot shows the Scopus search results for the author "Kumar, Vijay S." with the search criteria: "AU-ID ('Kumar, Vijay S.' 26661845000) AND (LIMIT-TO (PUBYEAR, 2013)) OR LIMIT-TO (PUBYEAR, 2012)) OR LIMIT-TO (PUBYEAR, 2011)) OR LIMIT-TO (PUBYEAR, 2010))". The results are filtered to 12 document results. The left sidebar shows refinement options for Year (2010-2013), Author Name (Kumar, V., Kurc, T., Khedkar, A., Acharya, D., Kong, J.), and Subject Area (Computer Science, Engineering, Mathematics, Business, Management and Accounting, Decision Sciences). The main results list includes:

Refine	Document Title	Author(s)	Year	Journal/Conference
<input type="checkbox"/> SAVEUS: Saving victims in earthquakes through unified systems 1	Beard, C., Chen, Z.Q., Kumar, V., (...), Leon-Salas, W.D., Rao, P.	2013	International Journal of Communication Networks and Distributed Systems	
<input type="checkbox"/> Pairwise key generation scheme for cellular mobile communication 2	Jaiswal, C., Kumar, V.	2012	Proceedings of the IEEE Symposium on Reliable Distributed Systems	
<input type="checkbox"/> Performance evaluation of data intensive mobile healthcare test-bed in a 4G environment 3	Acharya, D., Kumar, V., Han, H.-J.	2012	Proceedings of the International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)	
<input type="checkbox"/> High performance computing techniques for scaling image analysis workflows 4	Widener, P.M., Kurc, T., Chen, W., (...), Saltz, J.H., Foran, D.J.	2012	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	
<input type="checkbox"/> Self synchronization based air traffic control and collision avoidance system 5	Khedkar, A., Kumar, V.	2011	AIAA/IEEE Digital Avionics Systems Conference - Proceedings	
<input type="checkbox"/> Flash-based logging for database updates	Khedkar, A., Kumar, V.	2011	Proceedings of the 2011	

# Strategy Example: CSEE, Step 3-1

Step 3: Where are they publishing? What journals? What conferences?



Scopus

Search | Alerts | My list | Settings

Live Chat | Help and Contact | Tutorials | MERLIN Library Catalog

Register | Login

Brought to you by UMKC University Lib

AU-ID ("Kumar, Vijay S." 26661845000) AND (LIMIT-TO (PUBYEAR, 2013)) OR LIMIT-TO (PUBYEAR, 2012)) OR LIMIT-TO (PUBYEAR, 2011)) OR LIMIT-TO (PUBYEAR, 2010))

12 document results

Export | Download | View citation overview | View Cited by | More...

Refine	Document	Author	Year	Journal/Conference
<input type="checkbox"/> SAVE AS: Saving victims in earthquakes through unified systems	1	Beard, C., Chen, Z.Q., Kumar, V., (...), Leon-Salas, W.D., Rao, P.	2013	International Journal of Communication Networks and Distributed Systems
<input type="checkbox"/> Pairwise key generation scheme for cellular mobile communication	2	Jaiswal, C., Kumar, V.	2012	Proceedings of the IEEE Symposium on Reliable Distributed Systems
<input type="checkbox"/> Performance evaluation of data intensive mobile healthcare test-bed in a 4G environment	3	Acharya, D., Kumar, V., Han, H.-J.	2012	Proceedings of the International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)
<input type="checkbox"/> High performance computing techniques for scaling image analysis workflows	4	Widener, P.M., Kurc, T., Chen, W., (...), Saltz, J.H., Foran, D.J.	2012	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)
<input type="checkbox"/> Self synchronization based air traffic control and collision avoidance system	5	Khedkar, A., Kumar, V.	2011	AIAA/IEEE Digital Avionics Systems Conference - Proceedings
<input type="checkbox"/> Flash-based logging for database updates		Khedkar, A., Kumar, V.	2011	Proceedings of the 2011

Document Type



# Strategy Example: CSEE, Step 3-2

I exported to Excel and made a spreadsheet for all tenured and tenure-track CSEE

The screenshot shows the Scopus search results interface. The search criteria are: AU-D ("Kumar, Vijay S." 26661845000) AND (LIMIT-TO (PUBYEAR, 2013) OR LIMIT-TO (PUBYEAR, 2012) OR LIMIT-TO (PUBYEAR, 2011) OR LIMIT-TO (PUBYEAR, 2010)). There are 12 document results. A dialog box is open, asking to choose a default reference manager or file type. The dialog box has two main sections: "Choose your default reference manager or file type:" and "Choose the information to export:". The first section has radio buttons for "Save to Mendeley", "RIS Format", "CSV", "BibTeX", and "Text". The second section has a dropdown menu set to "Citation information only" and a list of selected fields: "Author(s), document title, year, source title, volume, issue, pages, citation count, source and document Type".

The screenshot shows an Excel spreadsheet with the following columns: Authors, Title, Year, Source title, Volume, Issue, and Art. N. The data is as follows:

Authors	Title	Year	Source title	Volume	Issue	Art. N
Hossain, N	Analysis of the properties of ZnO nanoparticle	2013	Midwest Symposium on Circuits and Systems			667
Hossain, N	Thermal aware Graphene based Through Silicon	2013	2013 IEEE International 3D Systems Integration Conference, 3DIC 2013			670
Subash, S	A new spatially rearranged bundle of mixed ci	2013	IEEE Transactions on Nanotechnology	12	1	587
Humayun,	On-chip voltage regulator without physical in	2012	2012 IEEE 10th International New Circuits and Systems Conference, NEM			632
Rahaman,	Exploiting negative quantum capacitance of ca	2012	ISCAS 2012 - 2012 IEEE International Symposium on Circuits and Systems			627
Gjanci, J.,	A hybrid scheme for on-chip voltage regulatio	2011	IEEE Transactions on Very Large Scale Integration (V	19	11	559
Vora, P.,	C A circuit implementation for dynamic thermal	2011	Proceedings - IEEE International Symposium on Circuits and Systems			593
Hassan, M	Addressing crossstalk issue in on-chip carbon n	2011	Proceedings - IEEE International Symposium on Circuits and Systems			593
Rahaman,	Spatial- and temporal-reliability aware design	2011	Proceedings - IEEE International Symposium on Circuits and Systems			593
Xu, J.,	Chc Fast waveform estimation (FWE) for timing an	2011	IEEE Transactions on Very Large Scale Integration (V	19	5	541
Chowdhui	An innovative power-gating technique for lea	2011	Circuits, Systems, and Signal Processing	30	1	
Chowdhui	Electrical properties of grain boundaries in po	2011	Journal of Physics D: Applied Physics	44	1	1
Xu, J.,	Chc Circuit level issues of interconnect pipelining	2010	Micro and Nanosystems	2	2	
Roy, A.,	Xi Analysis of the impacts of signal slew and skew	2010	IEEE Transactions on Very Large Scale Integration (V	18	2	528
Rahaman,	Energy efficiency of error control coding in int	2010	Microelectronics Journal	41	1	
Beard, C.,	SAVEUS: Saving victims in earthquakes through	2013	International Journal of Communication Networks	10	4	

## Strategy Example: CSEE, Step 3-3

Where are they publishing? Look at the raw data and find out!

For January 2010 - March 2014, Scopus indicates that UMKC CSEE Department tenured and tenure-track faculty were authors/co-authors of 121 conference papers, 81 journal articles, 8 books or articles in book series....***and only 1 of those was an open-access publication!***

However, faculty published a **LOT** ...like 109 out of 210 records...at IEEE conferences and in IEEE journals. (IEEE = Institute of Electrical and Electronics Engineers). Most of these faculty list IEEE membership and committees on their personal web pages. ***I wonder what IEEE's default author publishing agreement looks like?***

# Strategy Example: CSEE, Step 4

What are the publishers' policies for author self-archiving?  
SHERPA/RoMEO to the rescue!



SHERPA/RoMEO ... opening access to research

Home • Search • Journals • Publishers • FAQ • Suggest •

## Search - Publisher copyright policies & self-archiving English | Español | Magyar | Po

One publisher found when searched for: Publisher: **ieee**

Publisher:	<a href="#">Institute of Electrical and Electronics Engineers (IEEE)</a> , United States
Journals:	<a href="#">158 journals (including journals with special policies)</a> - involving 27 other organisations
RoMEO:	This is a <b>RoMEO green</b> publisher
Copyright:	<a href="#">Electronic information dissemination policy - copyright form</a> (pdf) - <a href="#">Copyright Policy</a>
Updated:	19-Jun-2013

Although this publisher has default policies, individual journals may have special permissions, especially if they involve other organisations or have paid open access options. **Always run a journal title or ISSN search to check.**

### Search again?

Journal titles or ISSNs  **Publisher names**

Exact phrase only  **All keywords**  Any

RoMEO Colour	Archiving policy
<b>Green</b>	Can archive pre-print and post-print or publisher's version/PDF
<b>Blue</b>	Can archive post-print (ie final draft post-refereeing) or publisher's version/PDF
<b>Yellow</b>	Can archive pre-print (ie pre-refereeing)
<b>White</b>	Archiving not formally supported

[More on colours and restrictions](#)

# Strategy Example: CSEE Step 5

Step # 5: Now I'm ready to talk with the CSEE Department!

1. I have data on their publication records in a major database
2. I can ask about whether the author agreements that they've signed with IEEE have included self-archiving clauses
3. I can point out that IEEE allows authors to self-archive both pre-print and publisher's version of articles (Thanks [SHERPA/RoMEO!](#))
4. I can ask them to include self-archiving clauses explicitly in future author agreements, and to contact IEEE about retroactive self-archiving rights for their papers
5. Once IEEE grants retrospective self-archiving rights (should be more than willing to do this), UMKC Libraries will then do the self-archiving work for the faculty, getting the IEEE works in published version into MOspace, our institutional repository

## A useful list of blog articles, web pages

A short overview of using secondary raw data sets at [Write.com](#).

A post from John Kratz at the California Digital Library blog [Data Pub](#) analyzing data repository coverage between DataBib and re3data. {*This just in 5/14/14 - [DataCite, DataBib, and re3data have announced a collaboration and merger of repositories](#)*}

The California Digital Library also has this nice page on [properly citing data sets](#) according to the Joint Declaration of Data Citation Principles.

A post from Romy Misra at [visual.ly](#) with 30 links to sources of Open Data on the web.

[The Open Data Handbook](#) can help researchers negotiate the legal, technical, and social aspects of opening up their data.

## More resources...

The [Data Management Research Guide](#) at UMKC Libraries