RESEARCH DATA MANAGEMENT IN ACADEMIC INSTITUTIONS: A SCOPING REVIEW

Leanne Trimble[1], Dylanne Dearborn[1], Erik Blondal[2], A. Patricia Ayala[1], Tim Kenny[3], David Lightfoot[4], Roger Reka[5], Mindy Thuna[1], Heather MacDonald[6]

[1] University of Toronto Libraries; [2] Institute of Health Policy, Management and Evaluation, University of Toronto; [3] Gibson D. Lewis Health Science Library, UNT Health Science Center; [4] St. Michael's Hospital Library, St. Michael's Hospital; [5] Faculty of Information, University of Toronto; [6] MacOdrum Library, Carleton University

OBJECTIVE

The goal of this scoping review was to describe the volume, topics, and methodological nature of the existing research literature on research data management in academic institutions.

MEDLINE SEARCH STRATEGY

Database: Ovid MEDLINE(R) <1946 to March Week 4 2016>, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <April 01, 2016>

```
1    Data Collection/ and Information Management/
2        Records as Topic/
3        Datasets as Topic/
4    "research data management".tw.
5    (accessib$ adj2 data$).tw.
6    (accessib$ adj2 research).tw.
7    (shar$ adj2 data$).tw.
8    (shar$ adj2 research).tw.
9    (transparen$ adj2 data).tw.
10     (transparen$ adj2 research).tw.
11     (cyber adj infrastructure).tw.
12     (reus$ adj2 data$).tw.
13       (reus$ adj2 research).tw.
14     (re-us$ adj2 data$).tw.
15       (re-us$ adj2 research).tw.
16   escience.tw.
17   e-science.tw.
18   (esocial adj science).tw.
19   (e-social adj science).tw.
20   eresearch.tw. (6)
21   e-research.tw. (61)
22   (research adj repository).tw. (19)
23   (research adj repositories).tw. (8)
24   (data$ adj repository).tw. (753)
25   (data$ adj repositories).tw. (467)
26   (data adj stewardship).tw. (20)
27   (data adj curation).tw. (89)
28   (data adj preservation).tw. (18)
29   "open research data".tw. (0)
30   cyberscholarship.tw. (0)
31   cyber-scholarship.tw. (0)
32   or/1-31 (14512)
33   "Academies and Institutes"/ (14962)
34   exp Libraries/ (9452)
35   Library Services/ (1063)
36   Universities/ (30455)
37   exp Faculty/ (29513)
38   exp Education, Graduate/ (37595)
39   exp Academic Medical Centers/ (77807)
40   academic.tw. (79057)
41   academia.tw. (4356)
42   library.tw. (98181)
43   libraries.tw. (31048)
44   university.tw. (252326)
45   universities.tw. (13260)
46     faculty.tw. (33305)
47   professor?.tw. (16525)
48   researcher.tw. (10203)
49   researchers.tw. (97773)
50   investigator?.tw. (66641)
51   scientist?.tw. (37274)
52   (graduate adj student?).tw. (2485)
53   (master? adj student?).tw. (117)
54   (PhD adj student?).tw. (268)
55   post-doc$.tw. (353)
56   (research adj fellow$).tw. (443)
57   or/33-56 (778405)
58     32 and 57 (2551)
59   exp Animals/ not Humans/ (4208134)
60     58 not 59 (2402)
```

METHODOLOGY

LITERATURE SEARCH

- Searched 40 databases from all disciplines, from inception to Apr 2016
- Also searched conference proceedings and gray literature
- Scanned references of included studies to identify other articles
- 15,228 articles identified and imported into Covidence

STUDY SELECTION

Criteria for inclusion:

- Topic of research data management
- Include activities or researchers at academic institutions
- Research study with documented qualitative or quantitative methodology (exclude descriptive papers)

Two stage process of assessing articles:

1. Review the title/abstract results using the inclusion criteria
2. Review the full-text for articles that passed the first review

Articles were assessed independently at each stage by two investigators. After assessment, 301 articles were included in the study.

DATA ABSTRACTION

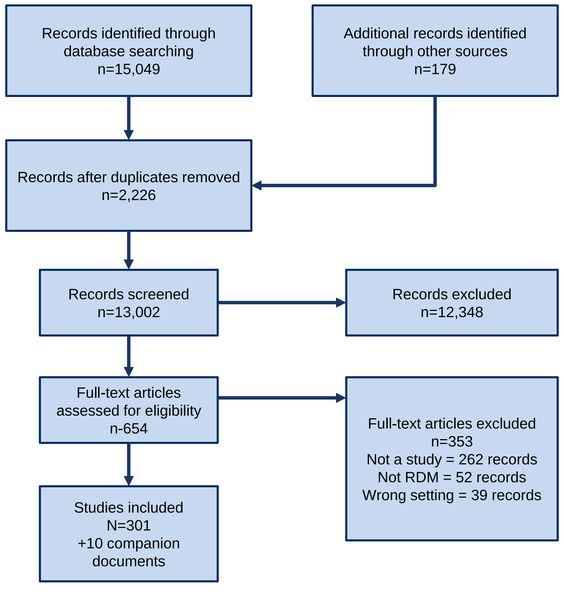
Two investigators independently read each article and extracted relevant data from the 301 articles.

Extracted data included:

- Study design and details (e.g. purpose, methodology, sample size, data collection)
- Setting and population or sample (e.g. geographic location, discipline)
- Alignment with research data lifecycle phases (e.g. creating data)

Validated guidelines for narrative synthesis were used for the descriptive summary of results, and dominant group and clusters of characteristics were also analysed.

PRISMA FLOW DIAGRAM

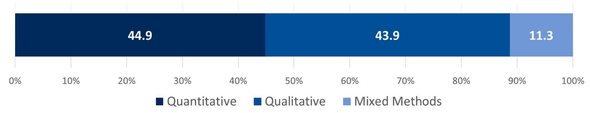


RESULTS

CHARACTERIZING THE LITERATURE

- The number of articles published on this topic has risen dramatically in recent years, with 85% of articles published post-2009
- 35% of all studies were classified as multidisciplinary. Medicine and Information Science also had significant amounts of RDM literature
- No studies used validated data collection instruments; most were either developed by authors or not described
- Fairly equal number of quantitative and qualitative studies

TYPE OF EVIDENCE INCLUDED IN ARTICLES BY PERCENT

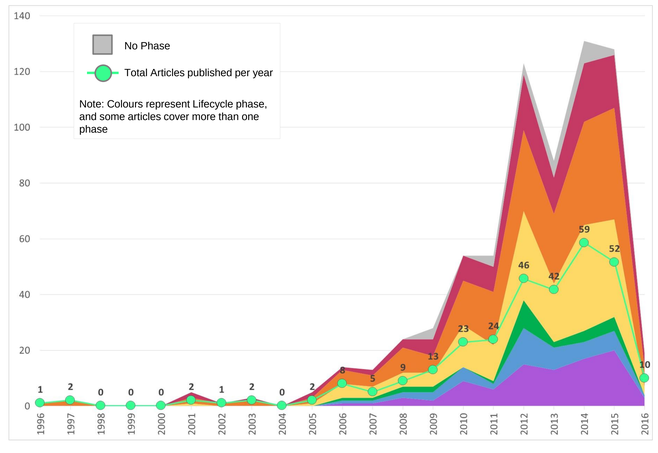


ALIGNMENT WITH RESEARCH DATA LIFECYCLE

- The largest number of articles were aligned with "giving access to data", followed by "preserving data"
- 28.9% were aligned with two phases, while 24.3% were aligned with three phases and 23.3% with one phase
- 9.6% were not aligned with any phase (this included those discussing librarian training and skillsets)

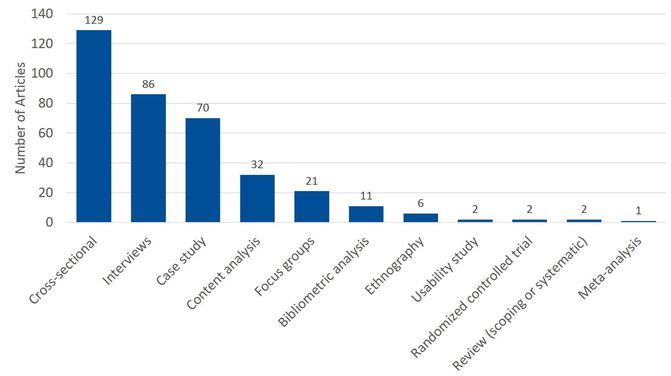
ARTICLES CORRESPONDING TO LIFECYCLE PHASE (UKDA[1])

NUMBER OF ARTICLES PUBLISHED BY YEAR/LIFECYLE PHASE



- The most common study types were cross-sectional surveys, interviews, and case studies

NUMBER OF ARTICLES BY STUDY TYPE



DISCUSSION

This analysis revealed a number of limitations to the research in this area:

- Limited range of study designs used; reliance on self-reports and case studies (nearly 80% of studies)
- Lack of standardized or validated data collection tools
- Lack of transparency in reporting
- Limited studies demonstrating the impact of RDM activities

This analysis suggested that research has emphasized access and preservation, with less emphasis on the beginning of the data lifecycle, such as writing data management plans and preparing to handle data.

Tracing articles back to their original literature database indicates that 86% of the studies were from the science literature, indicating higher interest and knowledge in this area.

See forthcoming paper for deeper analysis into dominant groups and clusters of characteristics found in the results[2].

CONCLUSIONS

Further research could be encouraged in several areas:

- Activities in the beginning of the data lifecycle
- Activities regarding the quality and usefulness of data in repositories, as well as facilitators and barriers for deposit and reuse
- Studies that use empirical evidence to demonstrate the impact of activities and interventions related to research data management

In order to ensure a high quality of publications and datasets investigating research data management, researchers should make efforts to follow best practices in research reporting.

Data available from Zenodo, DOI: 10.5281/zenodo.557043
A scoping review protocol was prepared (available upon request).
Article detailing results available in PLoS ONE[2]

REFERENCES

[1] UK Data Archive. Research Data Lifecycle. Available at: http://www.data-archive.ac.uk/create-manage/life-cycle. Accessed February 27, 2017.

[2] Perrier L, Blondal E, Ayala AP, Dearborn D, Kenny T, Lightfoot D, et al. (2017) Research data management in academic institutions: A scoping review. PLoS ONE 12(5): e0178261. https://doi.org/10.1371/journal.pone.0178261