Research Data Literacy and Management Skills of Pakistani Researchers

Prof. Kanwal Ameen

Department of Information Management, University of the Punjab, Lahore, Pakistan. kanwal.im@pu.edu.pk

Dr. Muhammad Rafiq

Department of Information Management, University of the Punjab, Lahore, Pakistan. rafiq.im@pu.edu.pk

Introduction

- The challenge of handling long term RDM and its reuse has not been addressed as significant area of research in Pakistan
- A PhD study under my supervision is going on the role of libraries in RDM

Aim

- To conduct a quantitative investigation of the research data literacy and management skills of the academics and researchers of Pakistani, being part of 'Data Literacy Research' Project,

Methodology

- Online survey
- **Sample** – Academic staff and research students of:
  - University of the Punjab, Lahore
  - University of Engineering and Technology, Lahore
  - GC University, Lahore;
  - National University of Science & Technology, Islamabad.

FINDINGS

Respondents Characteristics

	Freq.	%
Gender		
Male	109	52.4
Female	92	44.2
Don't want to disclose	7	3.4
Role		
Academic staff	90	43.3
Research student	118	56.7

Respondents Characteristics

	Freq.	%
Institutional Representation		
GC University, Lahore	81	38.9
NUST, Islamabad	63	30.3
UET Lahore	10	4.8
Punjab University, Lahore	54	26.0

Table 1:
File Types of Data Used (N=208)

File Types	*Freq.*	*%*
Standard office documents (text, spreadsheets, presentations, etc.)	176	85
Source code (scripting, Java, C, C++, etc.)	32	15
Configuration data (parameter settings, logs, library files, etc.)	31	15
Encoded text (XML, SGML, etc.)	27	13

Sources of Getting Data (N=208)

	Freq.	*%*
Always from multiple known sources	124	60
Create new data	99	48
Always from one known source	**36**	**17**

Use of Data Acquired from Others/Outside Sources (N=208)

Statements	*Freq.*	*%*
With a bit of effort for some cleaning and/or modifications	100	48
After spending a lot of time and efforts to make it usable for the project	97	47
I do not use data from others/outside sources	38	18
As it is without any problems	24	12

File Types of Data Produced (N=208)

File Types	*Freq.*	*%*
Standard office documents (text, spreadsheets, presentations, etc.)	155	74
Images (JPEG, GIF, TIFF, PNG, etc.)	100	48
Structured scientific and statistical data (e.g. SPSS, GIS, etc.)	97	47
Encoded text (XML, SGML, etc.)	21	10
Source code (scripting, Java, C, C++, etc.)	19	9
Audio files	14	7

Storage of Produced Data (N=208)

	Freq.	%
Your own devices (your computer, your tablet, external drive, etc.)	189	91
Cloud	49	24
Central servers/repositories of the university	33	16
Outside repositories	13	6

Metadata Assigned (N=208)

	Freq.	*%*
Administrative information (e.g. creator, date of creation, file name, access terms/restrictions, etc)	93	45
Discovery information (e.g. creator, funding body, project title, project ID, keywords, etc.)	69	33
Technical information (e.g. file format, file size, software/hardware needed to use the data, etc)	72	35
Descriptions of the data file (e.g. file/data structure, field tags/descriptions, application rules, etc.)	56	27

RDM Awareness and Skills *(N=208)*

Statements	**Yes** ***Freq.(%)***	**Uncertain** ***Freq. (%)***	**No** ***Freq. (%)***
Does your institution have a Data Management Plan (DMP)?	**51(25)**	96(46)	61(29)
Have you ever used a DMP for your research?	**61(29)**	40 (19)	107(51)
Are you familiar with the term metadata?	**109(52)**	29(14)	70(34)
Do you think a formal training on metadata would be useful for managing research data?	**137(66)**	43(21)	28(13)
Does your university have a prescribed metadata set for uploading data to a repository?	47(23)	**105(51)**	**55(27)**

RDM Awareness and Skills *(N=208)*

Statements	*Yes Freq.(%)*	*Uncertain Freq. (%)*	*No Freq. (%)*
Does your research community use/recommend any standard file naming system?	**67(32)**	**84(40)**	**57(27)**
Are you familiar with the concept of Digital Object Identifier (DOI)?	**106(51)**	**39(19)**	**63(30)**
Does your university recommend any specific guideline for citing data (e.g. APA, Harvard, etc.)?	**146(70)**	**34(16)**	**28(13)**
Have you got any unique researcher identification (like ORCID=Open Researcher and Contributor ID)?	**64(31)**	**50(24)**	**94(45)**
Does your university actively encourage you to share data on open access (OA) mode?	**68(33)**	**81(39)**	**59(28)**
Are you familiar with your university and/or funding body's requirements with regard to data storage?	**58(28)**	**79(38)**	**71(34)**

RDM Skills

How often do you practice the following statements?	***Mean***	***S.D.***
Citing research data	2.30	1.411
Working with data that are generally in the public domain	2.83	1.337
Using file naming convention or standard	3.19	1.376
Working with data that have restricted access	3.19	1.303
Having different versions of the same dataset(s)	3.29	1.298

Note: 1= Almost Always, 2= Often, 3= Sometimes, 4= Rarely, 5=Never

RDM Skills

How often do you practice the following statements?	*Mean*	*S.D.*
Using systems/techniques for version control to easily recognize a specific version	3.39	1.329
Using your own/in-house (your research team) tags and metadata	3.54	1.361
Using metadata standard for tagging your data	3.75	1.314
Using datasets that are tagged with standard metadata	3.76	1.308

Note: 1= Almost Always, 2= Often, 3= Sometimes, 4= Rarely, 5=Never

Preferred Location of Data Storage (N=208)

Location	*Freq.*	*%*
At your university	141	67.8
With the funding body	50	24
At external storage (unpaid)	83	39.9
At external storage (paid)	40	19.2

Formal Trainings Attained (N=208)

Statements	*Freq.*	*%*
Data citation styles	66	31.7
Metadata	**38**	**18.3**
Data management Plan	**36**	**17.3**
Consistent file naming	**21**	**10.1**
Version control of data sets	**14**	**6.7**

Formal Training Needs (N=208)

Statements	Freq.	%
Data management Plan	149	71.6
Metadata	112	53.8
Data citation styles	110	52.9
Consistent file naming	99	47.6
Version control of datasets	93	44.7
No, I am not interested	**30**	**14.4**

Conclusions

- The respondents mostly use basic standard office documents for their research work.
- However, images file formats, Internet and web-based data were also in use.
- Respondents get the data for their research from multiple known sources.
- In terms of creating or producing data from their research, majority of them produce standard office documents.
- a majority researchers store data in their self-owned devices.

……Continued

- They assign some sort of metadata to their data, but needs training in this regard.
- They lack skills in developing a digital data management plan
- Their RDM skills are not enough to manage their data in a desired manner.
- The study establishes a need for researchers' formal trainings on metadata and data management skills.
- It is a baseline study and we recommend further studies focusing on researchers of various disciplines.


Merci!