



International Federation of  
Library Associations and Institutions

## Who is the data curator? Defining a vocabulary

Anna Maria Tammara, Vittore Casarosa

Italian Research Conference Digital Library 2018, Udine 25-26 January 2018



# Purpose

IFLA Section Library Theory and Research started the research project in 2016 with the following objectives:

- To identify core vocabulary defining data curation and RDM profession.
- To explore the practice of research data management from the perspective of the professionals working in the field



# Research questions

R1: How is data curation defined by practitioners / professional working in the field?

R2: What terms are used to describe the roles for professionals in data curation area?

R3: What are primary roles and responsibilities of data curators?

R4: What are educational qualifications and competencies required of data curators?



# Stage 1 Literature review

The team conducted a review of library literature to identify possible new and emerging roles for librarians in assisting researchers



## Stage 2 Identification of current curation services

To discover niche services currently provided by subject librarians in the campus libraries, the team used:

- Quantitative content analysis of 441 job announcements for RDM professionals from 34 countries
- Qualitative interviews with professionals working as data librarians, data curators, or research data managers

# Six corpora

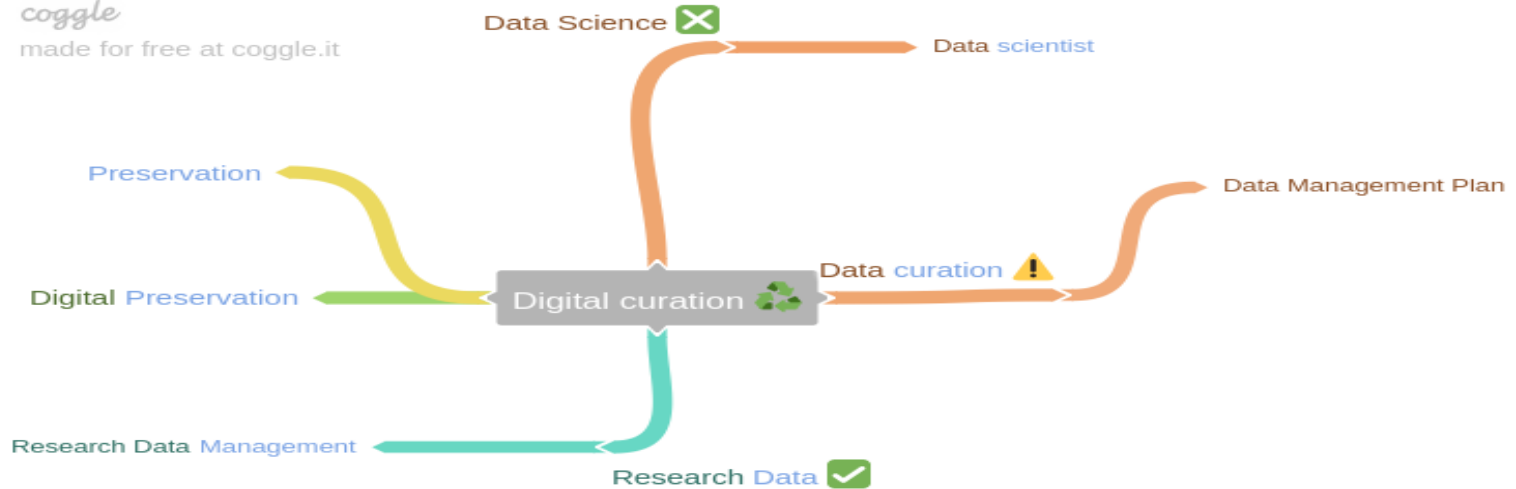
---

- Questionnaire. Text extracted from a set of questionnaires distributed to professionals in the field.
- Bibliography Updated. Text extracted from abstracts and keyword of papers related to Digital Curation and published in 2016 and 2017.
- Bibliography Old. Text extracted from abstracts and keyword of papers related to Digital Curation and published up to 2015.
- Positions/Job offers. Text extracted from job offers and positions, searching for “digital curators”, mostly from academia.
- Interviews. Text extracted from the transcript of interviews with selected respondents of the questionnaires.
- Edison project. Text extracted from deliverables of the Edison project, a European project aiming at defining the skills and the roles of the new profession of Data Scientist.
- Used KD (Keyphrase Digger) from FBK (Fondazione Bruno Kessler) to extract keyphrases

# Vocabulary

*coggle*

made for free at coggle.it



# Overlapping between result sets

	Questionnaire	Bibliography Updated	Bibliography Old	Positions/Job offers	Interviews	Edison ALL
Questionnaire (56 key phrases)		curation data curation data management data management planning data quality digital curation digital preservation information management preservation repositories research data research data management research data services	curation data curation data management data quality data sharing data stewardship digital curation digital preservation preservation repositories research data research data management	curation data curation data management data management planning digital curation digital preservation preservation research data research data management	curation data curation data management digital curation research data research data management	curation data curation data management data quality
Bibliography Updated (69 key phrases)	curation data curation data management data management planning data quality digital curation digital preservation information management preservation repositories research data research data management research data services		curation data curation data curation education data curation profiles data management data quality digital curation digital preservation digital preservation escience professionals large scale information management problems metadata plurality of curation roles preservation repositories research data research data curation research data management	curation data curation data management data management planning digital curation digital preservation preservation research data research data curation research data management	curation curators data curation data management digital curation metadata research data research data management	curation data curation data management data quality data science



# Overlapping between result sets

	Questionnaire	Bibliography Updated	Bibliography Old	Positions/Job offers	Interviews	Edison ALL
Bibliography Old (60 key phrases)	curation data curation data management data quality data sharing data stewardship digital curation digital preservation preservation repositories research data research data management	curation data curation data curation education data curation profiles data management data quality digital curation digital preservation escience professionals large scale information management problems metadata plurality of curation roles preservation repositories research data research data curation research data management		curation data curation data management digital curation digital preservation preservation research data research data curation research data management	area of data curation data curation data management data management plan digital curation metadata research data research data management	curation data curation data management data quality
Positions (40 key phrases)	curation data curation data management data management planning digital curation digital preservation preservation research data research data curation research data management	curation data curation data management data management planning digital curation digital preservation digital preservation preservation research data research data curation research data management	curation data curation data management data management planning digital curation digital preservation preservation preservation research data research data curation research data management		curation data curation data management digital curation research data research data management scholarly communication	curation data curation data management

# Overlapping between result sets



	Questionnaire	Bibliography Updated	Bibliography Old	Positions/Job offers	Interviews	Edison ALL
Interviews (34 key phrases)	curation data curation data management digital curation research data research data management	curation curators data curation data management digital curation metadata research data research data management	area of data curation data curation data management data management plan digital curation metadata research data research data management	curation data curation data management digital curation research data research data management scholarly communication		curation data curation data management
Edison ALL (25 key phrases)	curation data curation data management data quality	curation data curation data management data quality data science	curation data curation data management data quality	curation data curation data management	curation data curation data management	



## Preferred terms (WIKIDATA definitions)

### **Research Data Management (RDM)**

Activities around the life cycle of research-related data

Research Data: collection of facts produced through systematic inquiry (Q15809982)  
(Q30089794)

### **Data curation**

Work performed to ensure meaningful and enduring access to data

Digital curation: selection, preservation, maintenance, collection and archiving of digital assets (Q5276060)  
(Q15088675)

### **Data management**

All disciplines related to managing data as a valuable resource

Data Management Plan (Q1149776)  
(Q17085509)



# **RDA Task Force: Four professional areas**

1. research librarians;
2. research administrators;
3. research infrastructure managers / operators;
4. researchers



# Conclusions

This study demonstrates a significant variation in terminology, level of policy development, models of RDM services, and technical infrastructure across institutions and countries

Common themes in social aspects of RDM