International Journal of Research in Library Science (IJRLS)

ISSN: 2455-104X

DOI: 10.26761/IJRLS.9.4.2023.1716

Volume 9, Issue 4 (Oct-Dec.) 2023, Page: 210-219, Paper ID: IJRLS-1716

Received: 29 Sept. 2023 ; Accepted: 14 Dec. 2023 ; Published: 22 December. 2023

Data Curation Tools a Drastic Drift in Information Industries: An Overview and Analysis of Industry Wise Contribution of Data Curation Platforms

Uttam K. Malavi[1]; Dr. Pradnya H. Kshirsagar[2]

Librarian & SPOC NPTEL, Pune Institute of Business Management, Pune[1]; Librarian, Savitribai Phule Mahila Mahavidyalaya, Washim, Maharashtra India[2]

ABSTRACT

It's really, technology-occupied all fields of knowledge and information and data play vital roles and it became inevitable to use data curation technologies. This paper deals with cutting-edge technologies used to curate data to make information available to the right person at the right time.The research initially makes clear the concept of data curation and asserts the importance of describingits needs. The lists of existing data curation tools (platforms) are displayed. The data life cycle is also in the glance. The issues and challenges are exposed with their proper reasons. Selected tools arestudied as a sample as they contribute their services to various industries and data is analyzed for findings. Findings about their services, provided in various information industries are provided in the concluding part.

KEYWORDS: Information Integration, Data Management, Information Systems.

1. INTRODUCTION

Aristotle described human being as "rational animal", considering 'intelligence' as a maindistinguishing feature **G, Keil and N, Kreft** (2019) (1). Man has ability to learn, to assimilate information, to organize and process data and solving complex problems applying knowledge. Inthis modern electronic epoch data curation process has significance to make data available and reusable. So electronic machines like computer, laptops, mobile, tabs are used for quick access ofinformation. Data curation process includes selection of data, data accessing, data storing, data retrieving and preserving the data. Science now became more data centric and collaborative so researchers are using larger and more complex data to find out solutions for research questions **Line, Pouchard (2016)** (2) Extremely rapid increase of big data is sign of significant in the cloudbased management system and new data science epoch. Considering this flow will remain constantthis is clear that man and machine creating large scale of data. It was predicted by IDC study thatdigital universe will house 40,000 exabyte of data still 2020 **Sulayman K., Sowe and Koji Zettsu (2014)** (3). To make information and knowledge field, man started to

invent machines and remain efficacious. The man realized that information is a major factor in getting knowledge and developing every field. Information technology developed rapidly and occupied every field and brought a revolutionary change in human life.

2. LITERATURE REVIEW

In the big data study writer, G, Keil and N, Kreft (2019) demonstrates the design and architectureof cloud-based platforms and big data curation model which describes the picture of use of these platforms to curate data by the community and scientists.

In the research study, Line Pouchard (2016) discussed the characteristics of big data. He has also studied the data lifecycle model along with an application of the big data lifecycle model.

Alqasab, Mariam S, (2018) stated that data need to be curated regularly and curators need to handlehuge amounts of data, which makes it a time-consuming task. Agro-curation workflow and semantic-based architecture discussed in the study. In the study on data curation an opportunity for librarieswriter Amit Tiwari described data curation models and approaches. He has emphasized on importance of data, saying it is the fifth factor of production after land, labor, capital, and entrepreneurship. Anderson, Clifford B. (2017), writes about the data first manifesto, where priority is given to the data curation and interface design in the project of digital scholarship. Further, he discussed data serialization over databases and interfaces of application programming. Jagtap, Urmila L. (2020), in her survey research, stated the importance of information management. In the research, Abrams, S et al., (2014) studied on research data empowering to study data share portals, data preservation, control, discovery, data share architecture,

A research study conducted on the use of information, conducted by Mandal, Baishali B, focuses on recent technologies used for information resource centers.

3. OBJECTIVES

The objectives of this research work is

- ✓ To find out existing data curation tools.
- ✓ To study the importance and need for Data Curation Tools.
- ✓ To examine problems and issues in data curation management.
- ✓ To find out their contribution to various information industries.

4. RESEARCH METHODOLOGY

The descriptive method is used for this research to describe the concept of data curation, to describe the importance and need for data curation tools along with problems and issues in data curation management. Partly quantitative method used to find out industrywide active data curation tools.

4.1 Target Population & Sampling Method

Data curation platforms (software), found on the internet were finalized as the population for this research. Total 107 data curation platforms were found. The target population is too large to analyze data so systematic sampling (which comes under probability sampling) is used to reduce the target population. A total of 10 data curation tools were selected under this sample method.

4.2 Data Collection Tools & Data Analysis

Qualitative and quantitative data are used for this research This research is regarding data curation tools; hence data is collected through the Internet only. Data is collected online from the websites of selected Data curation tools. Data was analyzed with the help of a data analysis tool. Findings are presented with graphs and charts.

5. Research Problem and Importance of Topic

5.1 Research Problem

Due to the information explosion information centers are facing problems in finding, selecting, managing, distributing, and keeping secure data. Many information technology departments of universities have no resources to support bespoke digital Project (**Clifford B. Anderson 2017**). Modern user accepts information within seconds of their fingertips touch. Hence it is need of time to curate data to access with the use of information and communication technology and engage users for data discovery. Collecting data from diverse sources and integrating it into repositories is the major task of data curators and information centers.

While thinking on data curation following basic questions makes us inquisitive F

1. What does data curation mean and what is its process?
2. Which data curation tools are in existence now?
3. What are the benefits of data curation?
4. What is the industry wise contribution to data curation tools?
5. What are the issues in data curation?

To find out above questions this research study is conducted on both qualitative and quantitative data. This research paper covered all the above aspects of data curation.

The study of this research work is regarding data curation tools. It covers only data curation tools availability, service, challenges, and benefits. So the scope of this study will be within the data curation management field. Study is limited with data curation management tools used in various industries. Data curation platforms (software), found on the internet finalized as the population for this research. A total of 107 data curation platforms were found. This research is regarding data curation tools; hence data is collected through the Internet only. So computers and the internet are used as data collection tools for this research

5.2 Data Curation

Data curation is the procedure of discovering, integrating, and cleaning data to use for downstream analytic tasks to extract business value of any organization **Sarvanan Thirumuruganathan et al., 2020**). Information or data in all fields is searched, stored, manipulated, retrieved, and preserved in electronic form by organizations' libraries and information centers.

5.2.1. Definitions of Data Curation

Graduate school of library and information science, University of Illinois defines data curation as: *"The active and ongoing management of data through its lifecycle of interest and usefulness to scholarship, science, and education. Data curation activities enable data discovery and retrieval, maintain its quality, add value, and provide for reuse over time, and this new field includes authentication, archiving, management, preservation, retrieval, and representation."*

According to Cragin et al. (2007), *"Data curation is the active and ongoing management of data through its lifecycle of interest and usefulness; … curation activities enable data discovery and retrieval, maintain quality, add value, and provide for re-use over time*

5.2.2. Data Curation Process (Figure 1)

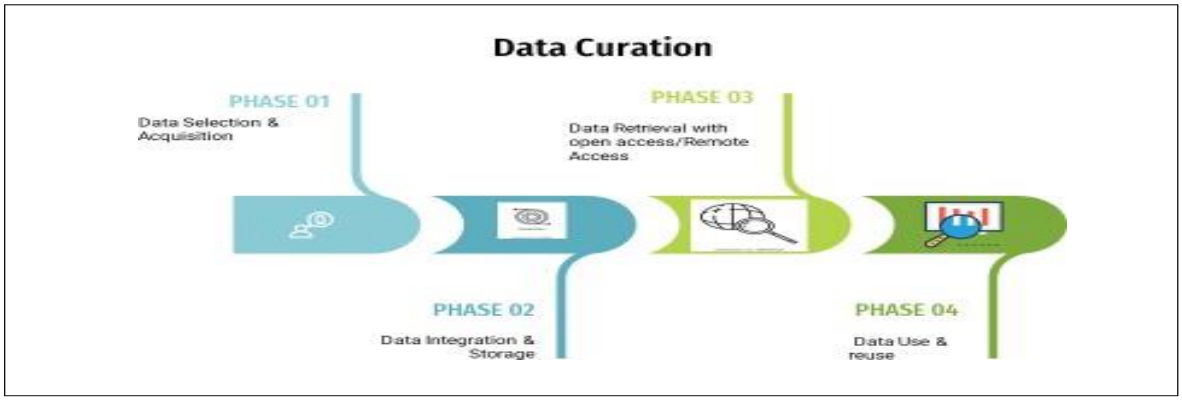


Figure 1 exposes the picture of proper curation of data which is a need of time for today's information centers. In this process, technologies are used skillfully to transform data for its reusefrom its selection to the end user in an easy way.

5.3 Why Data Curation?

It is expected that machine intelligence will be able to validate, repair, and annotate data within seconds, which might take hours for humans to perform **(Kong et al. 2011).**[1] There is no doubt all people are experiencing this in the world. Let's discuss a few points on 'Why Data Curation'

1) For easy acquisition and dissemination of data
2) Main purpose of data curation is to make data reusable for the future.
3) To minimize the risk of data loss.

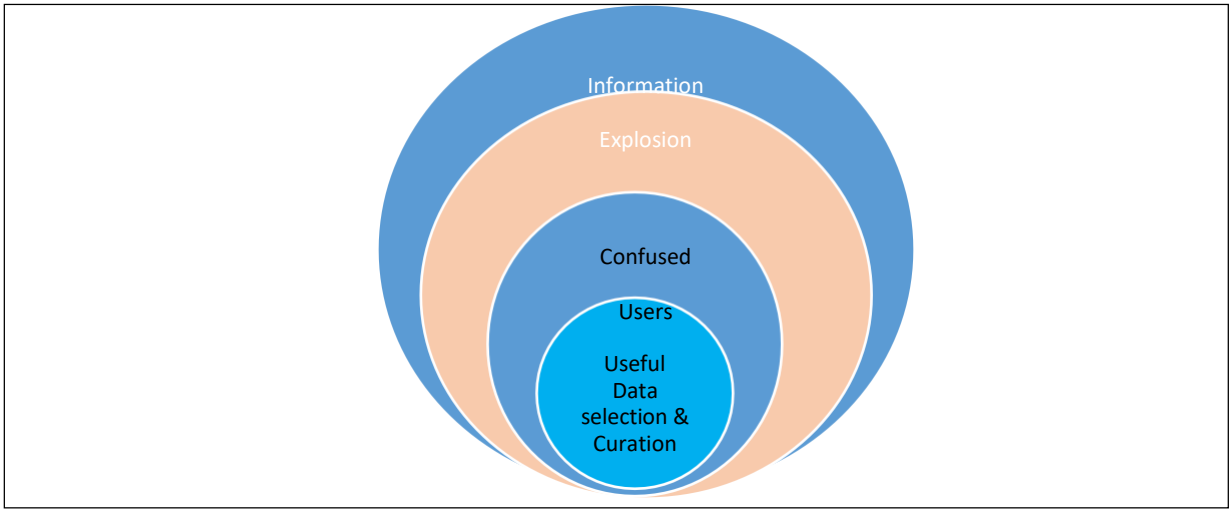
Almost all information centers, research, and academic libraries of higher education have to provide such services efficiently. They have a huge amount of data under their subscriptions and have to provide it to all users in a proper way. Data curation tools are playing a vital role in solving this problem. Many data curation tools organize and take care of data. So it is necessary to conduct a study on data curation tools for all industries of information, information centers, research, and academic libraries. This research study is on 'Data Curation meaning, its importance and need,and their contribution to various information industries. This study will remain beneficial to librarians or information managers to know about data curation tools and it will help to manage their data management system.

Due to the massive creation of e-resources, redundant and partly related documents caused information to explode. Hence information providers and users are confused about how to provide and getthe particular data they want. The option of data curation is best to control this problem. So many private companies that are active in the information technology field are providing this type of services, and proved these platforms as the best tool to satisfy their users. Of course, using theseplatforms is inevitable in the information industry

---

[1] See Jose´ Marı´a Cavanillas (2016), New Horizons for a Data-Driven Economy (Springer International Publishing AG Switzerland)

.5.3.1. Data Curation: Need of Data Curation (Figure 2)



6. Issues and Challenges of Data Curation:

Very few databases are appropriately managed and preserved with proper documentation and made available for retrieval for users. this is due to a lack of awareness among the researchers regarding the benefits of good practices of data curation and the lack of tools for data curation which areeasy to use (**Stephen, Abrams., et al 2014**). Using data curation tool platforms is a recent trend andnew for both information providers and end users. The data curation process is not affordable for each database and it is a very expensive and time-consuming task in the performance of experts in curation (**Mariam S. Alquasab 2018)** This process needs both human and machine intellect for proper integration of information and retrieval of information, where human intelligence is too little to fulfill this demand. Need of funding is also a remarkable problem. Less awareness of information technology among users and information providers. Only electronic components like computers, laptops, tab, and mobile can be used. Considering all this following issues or challenges are coming into focus.

1) Data curator software is needed for this process which are very costly.
2) Skilled data curators are rare or expensive
3) Sufficient funding is required for this because this is a lengthy process.
4) Long-term preservation of data is a difficult task.
5) Need to update by adding new modules for retrieval.
6) Intellectual property and Licensing are major issues
7) Proper selection of data is also a challenging task.

7. DATA CURATION PLATFORMS

There are several existing data curation platforms providing their services for different industries. 107 data curation tools found after the search which are displayed in the following table.

Table 1 - List of Existing Data Curation Tools

Existing Data Curation Tools							
1	Addict-o-matic	28	Delve	55	Meddle	82	sprinkir
2	Aggregage	29	Dizkover	56	MODX	83	Spundge
3	Alation	30	dot cms	57	Newsle	84	Squarespace
4	Alteryx	31	Drumup	58	OneSpot	85	Stitch Data
5	AnalytiX DS	32	Drupal	59	papaly	86	Storify
6	Ataccama ONE	33	elink	60	Paper.li	87	story stream
7	BagTheWeb	34	Equetia	61	Pearltrees	88	Storyfy
8	BlogBridge	35	eXo Platform	62	Pinterest	89	super desk
9	bold	36	Feedly	63	Pixpa	90	Sutory
10	Bundlepost	37	Flashissue	64	Pluggio	91	Symbaloo
11	Bundlepust	38	Flipboard	65	Pocket	92	TagBoard
12	Buzzsumo	39	Flockler	66	popular	93	Talend
13	Canvas	40	Follozo	67	postplanner	94	Textpattern
14	Catalog Automation	41	Google site	68	Pressjack	95	Trap!t
15	Channelkit	42	HeadSlinger	69	Pulse	96	Triberr
16	CIThread	43	Hubspot	70	Quuu	97	Tweeted Times
17	cognitiveSEO	44	Huzzaz	71	RebelMouse	98	Upcontent
18	Contentful	45	Informatica	72	Reiberr	99	Vidinterest
19	ContentGems	46	Joomla	73	Rock the deadline	100	Waywire
20	Crowdynews	47	Juxtapost	74	Roojoom	101	Wix
21	Crownpeak	48	Kbucket	75	Scoop.it	102	WordPress
22	Curata	49	Kinetico	76	Shareist	103	Wrike
23	Curation Traffic	50	Kweeper	77	ShareIt	104	Yahoo Pipes
24	CurationSoft	51	LinkHub	78	Sharpr	105	Zeef
25	CurationStation	52	List.ly	79	Site Finety	106	Zephry
26	Curatur	53	list.ly	80	SocialPilot	107	Zimilate
27	DayLife	54	Magento	81	Spredfast		

8. DATA ANALYSIS

The target population is too large to analyze data and assert findings. So sample method had to be to reduce the target population. Sequential data selected (10^{th}, 20^{th}, 30^{th} like this) under this used. Hence for this research systematic sampling (which comes under probability sampling) used the sample method

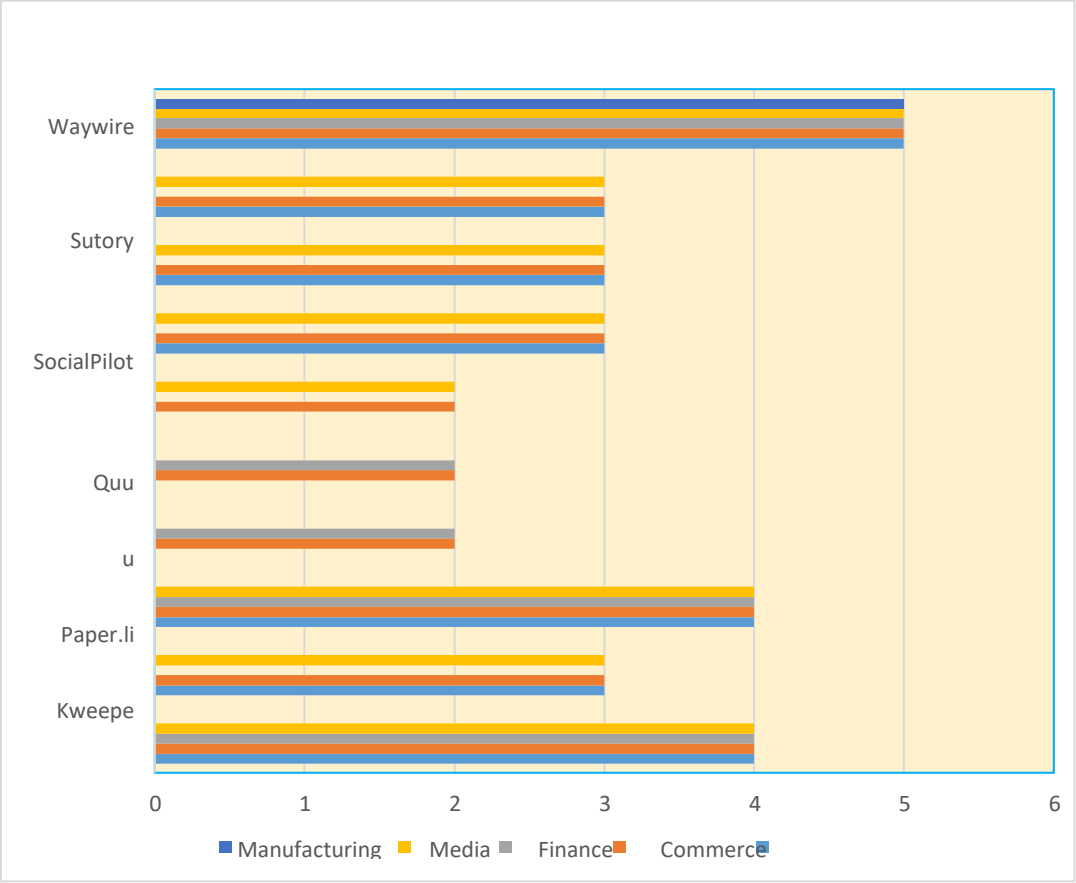
8.1 Table 2 - List of Selected Data Curation Tools (using systematic sampling)

There are many existing data curation tools and it is necessary to reduce the population for study. Sofollowing 10 data curation tools are selected for further study.

Bundlepost	Kweeper (Crunchbase)	Sutory
Crowdynews	Paper.li	Waywire - Crunchbase
dot cms	Quuu	
Follozo	SocialPilot	

8.2 Industry wide Data Curation Contribution

Data curation tools are used in various industries like the education Industry, retail and commerce industry, financial service industry, media and entertainment industry, and Manufacturing industry. Data collected and analyzed according to industry wide contribution in this research. The outcome is presented in the chart bar below.

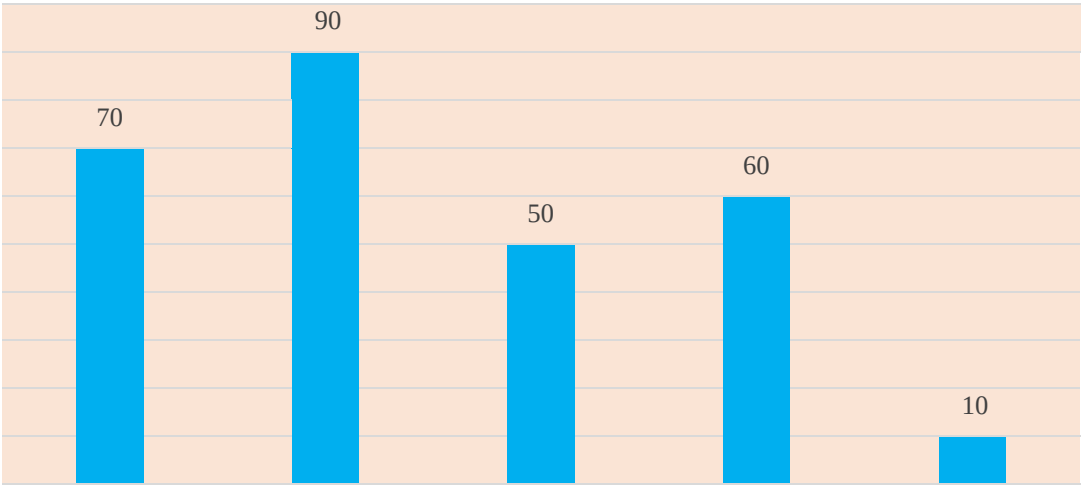


8.2.1 Curation Platforms providing services Different Industries

This clustered column chart shows the picture of the data curation platform services for different industries. After the study of 10 data curation platforms, it was found that almost all data curation platforms are providing multi-industry services. For example, 'dot cms' platform providing services for education, marketing, financial services, media and entertainment and manufacturingindustry.

The above bar chart shows the data curation platforms providing services for different industries where Waywire, Dot CMS and Bundle Post platforms are providing services to maximum industries. On the other hand, Paper li, Kweeper, and Folloze curation platforms are providing services to minimum industries. Waywire has been providing services to five different industries.

8.3 Graph of Data Curation Tools and Their Services (In percentage)



FINDINGS

After the study and observation of the above graph, the following things are found.

1. 7 out of 10 curation platform found means 70% of all data curation tools are providing services in the education field.
2. 90 % Data curation tools are providing services for retail and E-commerce.
3. 50% of Data Curation tools found active in the financial field.
4. 60% are active in the Media and entertainment field, and
5. Only 10% of services are provided for the manufacturing field.
6. After this analysis we realized that, maximum data curation tool platforms providing their services in education, marketing, and media and entertainment industries. Data curation tools are used in every information industry as a need of time.

Scope for Future Research

The scope for future researchers is to study data curation tools in depth with technical architecture and security analysis.

CONCLUSION

There are so many companies and information centers that are using various data curation softwareplatforms. After the other research study, it is found that it is inevitable today to use the data curationprocess for data management in this information explosion. These technologies made it easy to access data for users as fast as they wanted. After a study of selected software, it is found that one platform of data curation is providing services to various industries. Though data there are some issues and challenges to use these technologies in developing countries like India because of funding and lack of skills in information technology. This paper will remain useful to further researchers to get information regarding data curation tools in the future.

.

REFERENCES

[1] Aristotle's Anthropology. (2019). In G. Keil & N. Kreft (Eds.), *Aristotle's Anthropology* (pp. I- Ii). Cambridge: Cambridge University Press

[2] Pouchard, L.C. (2016). Revisiting the Data Lifecycle with Big Data Curation. *International Journal of Digital Curation, 10*, 176-192.

[3] Benítez, J.A., Labra Gayo, J.E., Quiroga, E., Martín, V., García, I., Marqués-Sánchez, P., & Benavides, C. (2017). A Web-Based Tool for Automatic Data Collection, Curation, and Visualization of Complex Healthcare Survey Studies including Social Network Analysis. *Computational and Mathematical Methods in Medicine, 2017*.

[4] Abrams, S.L., Cruse, P., Strasser, C., Willet, P., Boushey, G., Kochi, J., Laurance, M., & Rizk- Jackson, A. (2014). DataShare: Empowering Researcher Data Curation. *Int. J. Digit. Curation, 9*,110-118.

[5] Alesso, H. P. (2006). Thinking on the web: Tim Berners Lee, G Godel and Turing. Alqasab, M. S. (2018). *AMPLIFYING DATA CURATION EFFORTS TO IMPROVE THE*.

[6] Anderson, C. B. (2017). Data-first manifesto : Shifting priorities in scholarly communications.

Information Services & Use., 37(3), 335-342. doi:10.3233/ISU-170852

[7] Bryan Heidorn, P. (2011). The emerging role of libraries in data curation and e-science. *Journal of Library Administration*, 662-672.

[8] Katherine, S. (2020). *Designing for Serendipity : Research Data Curation in Topic Spaces.*

Mclure, M. (2014). Data Curation: A study of researcher practices and needs. 139-164.Petersen, K. S. (2019). *From Data Collection To Analysis.*

[9] Tiwari, A. (2018). Data Curation: An Opportunity for the libraries. *Emerging Trends inLibrarianship*, 3-5.

(Hill et al., 2020)Bryan Heidorn, P. (2011). The emerging role of libraries in data curation and e-science. *Journal of*

[10] *Library Administration*, *51*(7–8), 662–672. https://doi.org/10.1080/01930826.2011.601269

Hill, B. R., Stein, C., & Williams, H. (2020). *Internalizing Externalities : Designing Effective Data Policies †*. 49–54.

[11] Katherine, S. (2020). *UC Santa Barbara UC Santa Barbara Electronic Theses and DissertationsDesigning for Serendipity : Research Data Curation in Topic Spaces*.

[12] Laulederkind, S. J. F., Shimoyama, M., Hayman, G. T., Lowry, T. F., Nigam, R., Petri, V., Smith, J. R., Wang, S. J., De Pons, J., Kowalski, G., Liu, W., Rood, W., Munzenmaier, D.H., Dwinell, M. R., Twigger, S. N., & Jacob, H. J. (2011). The Rat Genome Database curation tool suite: A set of optimized software tools enabling efficient acquisition, organization, and presentation of biological data. *Database*, *2011*, 1–8. https://doi.org/10.1093/database/bar002

[13] Lord, P., Macdonald, A., Lyon, L., & Giaretta, D. (n.d.). *From data deluge to data curation*. http://scholar.google.com/citations?view_op=view_citation&hl=en&user=r23WHA8AAAA

J&citation_for_view=r23WHA8AAAAJ:u-x6o8ySG0sC

[14] Mclure, M., Level, A. V., Cranston, C. L., Oehlerts, B., & Culbertson, M. (2014). Data Curation: A study of researcher practices and needs. *Portal*, *14*(2), 139–164. https://doi.org/10.1353/pla.2014.0009

Petersen, K. S., & Technology, H. (2019). *From Data Collection To Analysis*. 1–24.

[15] Ray, J. (2012). The rise of digital curation and cyberinfrastructure: From experimentation toimplementation and maybe integration. *Library Hi Tech*, *30*(4), 604–622. https://doi.org/10.1108/07378831211285086

[16] Rehm, G., Bourgonje, P., Hegele, S., Kintzel, F., Schneider, J. M., Ostendorff, M., Zaczynska, K., Berger, A., Grill, S., Räuchle, S., Rauenbusch, J., Rutenburg, L., Schmidt, A., Wild, M.,Hoffmann, H., Fink, J., Schulz, S., Ševa,

[17] J., Quantz, J., … Heine, F. (2020). QURATOR: Innovative technologies for content and data curation. *CEUR Workshop Proceedings*, *2535*,1–15.

[18] Sowe, S. K., & Zettsu, K. (2014). Curating big data made simple: Perspectives from scientificcommunities. *Big Data*, *2*(1), 23–33. https://doi.org/10.1089/big.2013.0046

[19] Valenza, J. K., Boyer, B. L., & Curtis, D. (2014). Curation Platforms. *Library TechnologyReports*, *50*(7), 2,60-65. https://search.proquest.com/docview/1628649481?accountid=17242

[20] Vassilakopoulou, P. (2019). *Enabling openness of valuable information resources : Curbingdata subtractability and exclusion*. *October 2016*, 768–786. https://doi.org/10.1111/isj.12191

[21] Weiner, K., Will, C., Henwood, F., & Williams, R. (2020). Everyday curation? Attending to data, records and record keeping in the practices of self-monitoring. *Big Data and Society*,*7*(1).
https://doi.org/10.1177/2053951720918275

[22] Xiao, A. (2019). *How did the data extraction business model come to dominate ? Changes in the web use ecosystem before mobiles surpassed personal computers*. *35*(5), 272–285.

[23] Zilinski, L., Scherer, D., Bullock, D., Horton, D., & Matthews, C. (2014). Evolution of data creation, management, publication, and curation in the research process. *Transportation Research Record*, *2414*, 9–19. https://doi.org/10.3141/2414-02