TECHNISCHE UNIVERSITÄT
CHEMNITZ

Department of Computer Science

Distributed and Self-organizing Systems Group

Comparison of existing decentralized RDM solutions

VSR Technical Report

André Langer, Dang Vu Nguyen Hai

1. Categories of existing decentralized RDM solutions

This section categorizes and describes approaches for decentralized research data management. As this thesis focuses on storing, publishing, and sharing aspects, the resulting groups are: (1) Peer-to-peer based approach, (2) Blockchain-based approach, and (3) Data repository (Git) based approach. Each group will be described in detail in the following.

Peer-to-peer based approach

Peer-to-peer (P2P) network is a distributed application architecture, which connects each computer as a node (peer) to a network, with equal privileged. Peers play both roles as suppliers and consumers of resources, which leads to the elimination of a centralized server. One of the typical characteristics of the P2P is distributed data storage so that research data might not only stored in one node but also can replicate in some or all nodes of the network. Besides, one node can directly exchange the data to another node in the network. Although P2P can apply for any data type, in this thesis, we only analyze these implementations which concentrate on research data, i.e., DAT-Project[1] and Academic Torrent[2].

Dat-Project, shown in Figure 1, provides a new protocol for sharing data between computers. Dat-Project is free and open-source software, support researchers, analysts, libraries, and university to achieve and distributed scientific data. By using the peer-to-peer network in behind, Dat works by linking two computers directly, without needing a third-party server, to share from small to large datasets between researchers and institutions.

[1] https://dat.foundation/

[2] http://academictorrents.com/

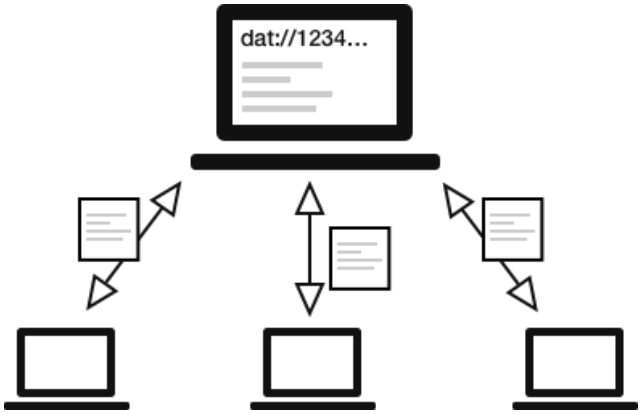

Figure 1 The DAT protocol of Dat Project [22]

Academic Torrent, which its concept is shown in Figure 0, is a distributed system for sharing enormous research datasets. The Academic Torrent network is built for researchers and by researchers. It has two main components: a site where researchers can search for datasets, and a BitTorrent backbone (peer-to-peer network) which makes sharing data scalable and fast.

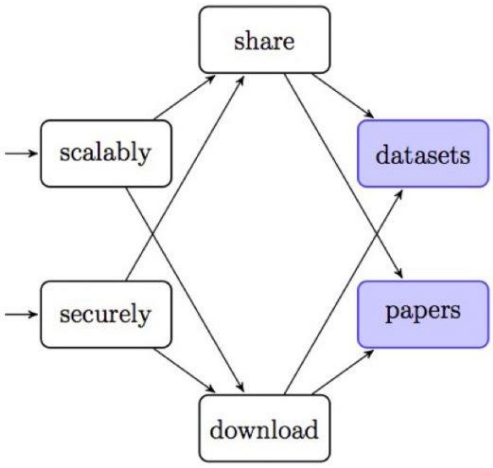

Figure 0 The concept of AcademicTorrents.com

Blockchain-based approach

Blockchain is a decentralized, distributed and public digital ledger that is used to record transactions across many computers so that any involved record cannot be altered

retroactively, without the alteration of all subsequent blocks. The blockchain is built on top of the P2P network and added some new elements, i.e., cryptography, consensus algorithm so that user can store data on dozens of individual nodes, intelligently distributed across the network. Consequently, no central entity is needed to control access to a user's files, which leads to improving security and decreasing costs via decentralized file storage. Datum[3] is one of the well-known implementations of using blockchain technology to store and share research data.

Datum, shown in Figure 3, "is a decentralized data store allowing users to store structured data securely running on a smart contract blockchain" [23]. Datum uses a DAT token for data storage and sharing. It leverages BigChainDB and IPFS to provide a scalable, decentralized data storage backend. The Datum network aims to give researchers efficient and frictionless access to data while respecting the data owner's terms and conditions.

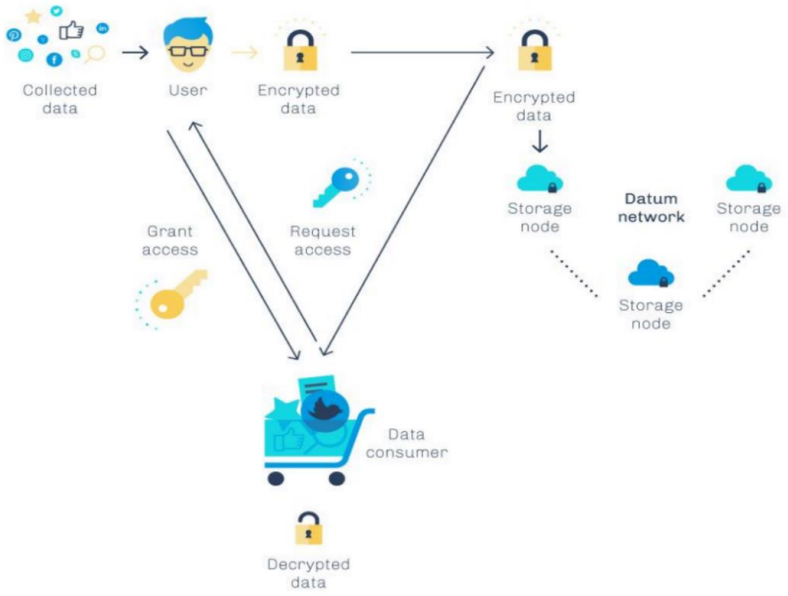

Figure 3 How datum.org works [23]

Data Repository (Git) based approach

Git is a free and open-source distributed version control system (DVCS) and was initially developed to maintain code repositories in the software industry. Git can be ap-

[3] https://datum.org/

plied to manage research data by "providing a lightweight yet robust framework that is ideal for managing the full suite of research outputs, e.g., datasets, statistical code, figures, lab notes, and manuscript" [24]. A Git user can store their data at a local working directory on their computer. Git provides command-line tools or GUI clients for the researcher to manage, track, and version data. Later, the researcher can publish their data by pushing it to Git hosting services (e.g., GitHub, and GitLab) or research data portals (e.g., Datalad). With Git and Git repositories, researchers have great collaborative tools that enable them to share and co-edit data with their peers.

Git hosting services (GitHub, GitLab) "allow researchers to store and share their research data with the appropriate versions of the files. Other researchers with access permissions, can read the data, work asynchronously, and merge their contributions at any time, all the while maintaining a complete authorship trail" [24]. However, there are some limitations to manage research datasets with original tool. Git is only good at dealing with smaller files (e.g., GitHub has a size limitation of 100Mb), and the data are text-based formats (e.g., CSV, XML, JSON). Meanwhile, research datasets are much more massive and have many different forms, e.g., binary, and blobs. In order to overcome these limitations and provide more effective and efficient research data management portal, there are several implementations which are based on Git but extends it, by combining some extensions such as git-anex[4] or git-lfs[5]. One typical example is DataLad platform [25].

DataLad[6], shown in Figure 4, is a data portal, built on top of git-anex and extends it to enables researchers to operate on research data while transparently managing data access and authorization. Moreover, DataLad aims to provide all the tools necessary for creating and publishing data distributions to data sharing and collaborative work.

[4] https://git-annex.branchable.com/
[5] https://git-lfs.github.com/
[6] https://www.datalad.org/

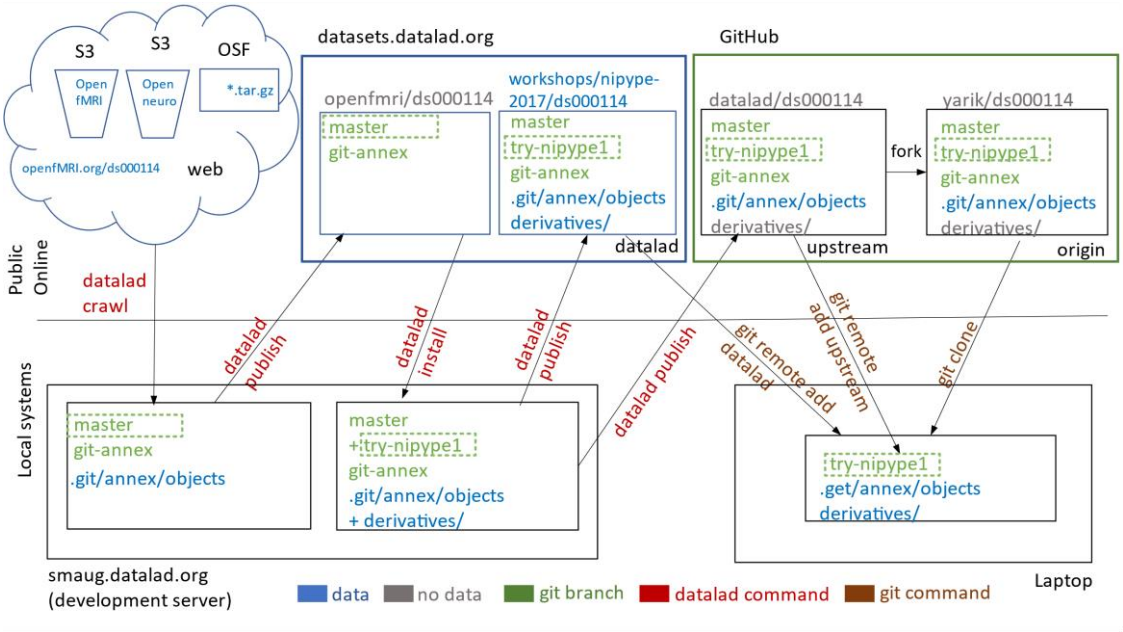

Figure 4 DataLad and git-anex [26]

2. Comparison

This section compares the previously introduced categories of approaches the requirements presented in section **Error! Reference source not found.**. Depending on how each requirement is fulfilled, each category will be rated on a scale of 3 different levels: "Not Fulfilled", "Partially Fulfilled", and "Completely Fulfilled", which we mentioned in section **Error! Reference source not found.**. In all evaluation tables, each scale will be represented by symbol minus (-), circle (O) and plus (+) respectively. At the end of this section, Table 6 will summarize the outcomes of the evaluation process.

Ownership of data storage

Regarding the ownership of data storage, the P2P based approaches have the disadvantage that research data is not only stayed in our node but also replicated in multiple nodes of the network. Besides, the data "can be accessed by everyone (by potentially untrusted peers) and used for everything (e.g., for marketing, profiling, fraudulence, or

for activities against the owner's preferences or ethics)." [27]. In Dat Project, there is an option that allows transferring research data directly from on researcher's server to another researcher's laptop, without replication data. However, this option is just a temporary solution and only applies to one specific use cases such as "data sharing between institutions" [28].

In the following, the Blockchain-based approaches have the same disadvantage. For instance, in the Datum Blockchain project, research data is stored off-chain, in a separate Storage Network, which based on the P2P network, shown in Figure 5. Because direct data storing on the blockchain (on-chain) is expensive and slow to access, and therefore only a hash of the data is stored on-chain for proof of storage. As a result, the user's data is also replicated in many storage nodes, e.g., three nodes by default in Datum.org.

Based on the disadvantage above, both P2P-based and the Blockchain-based approaches provide only partial ownership of data storage and therefore receive a "Partially Fulfilled" rating.

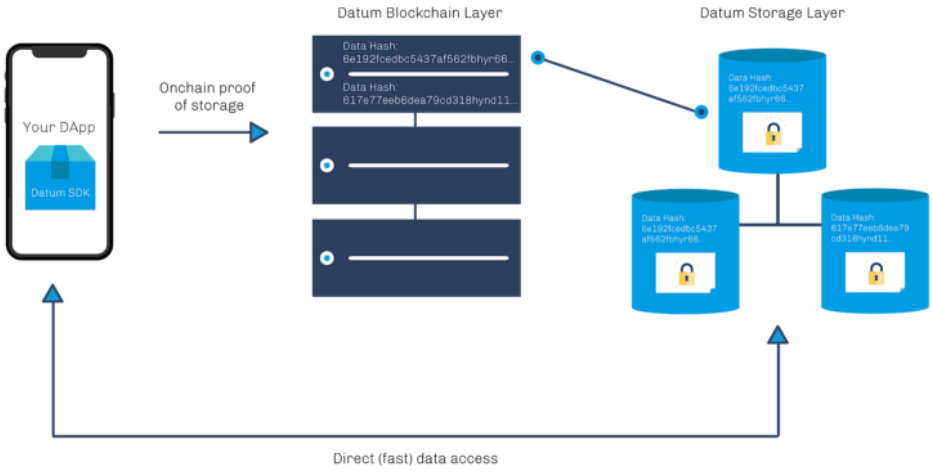

Figure 5 Data storage implementation by Datum.org

In contrast, the Git-based method is better in comparison. As its concept, Git itself is distributed. The user can store the research data on a local working repository on their computer. For a remote repository, the user has an option to set up their Git server, as well as using private or public data repositories. In other words, the choice of where to

store data belongs to the user. For that reason, this approach is rated "Completely Fulfilled", and Table 1 shows the final evaluation result.

	P2P based approach	Blockchain-based approach	Git-based approach
Ownership of data storage	O	O	+
+ Completely Fulfilled, O Partially Fulfilled, - Not Fulfilled			

Table 1 Evaluation of Data storage ownership

Data access control

In term of "Data access control", as shown in Table 0, no "Completely Fulfilled" rating is given to any of the categories because none of them can fully satisfy all criteria of this requirement. On the one hand, the Git-based approach and the Blockchain-based approach receive a "Partially Fulfilled" rating. The former supports data access control, which relies on filesystem permission[7]. A "*.gitignore"* file can be used to configure private mode by indicating which files or folder will not be pushed to the remote repository and only kept in the local machine. However, access control can only apply per-repository, not per-file or per-dataset. Whereas, in the latter, particularly in Datum Storage Network, users can "control over privacy settings and can fine-tune with whom to share data: share disabled (private mode), share with specific, identified and known data consumer, or share with everyone" [23]. However, this Blockchain implementation does not satisfy the criterion of cross-system because data consumer must have accounts on the same network to gain access.

On the other hand, the P2P based approach receives a "Not Fulfilled" rating because of the pure P2P system "offers limited guarantees concerning data privacy" [27]. Although there is some proposed mechanism to ensure data access control, e.g., OceanStore, Past, and Freenet, "these solutions remain insufficient" [27]. For example, in Dat Project, file shared is encrypted and only accessible by using a unique read key. Hence, if anyone has this read key, he or she will gain full access to the file, i.e., read,

[7] https://wincent.com/wiki/Git_repository_access_control

download, and re-share. The final evaluation result for this requirement is shown in Table 0.

	P2P based approach	Blockchain-based approach	Git-based approach
Full data access control	-	O	O
+ Completely Fulfilled, O Partially Fulfilled, - Not Fulfilled			

Table 0 Evaluation of Data access control

User support tools

In this requirement, three approaches can be rated as "Completely Fulfilled", shown in Table 3. In order to set up data storage, users can run the command line to install a local server instance (Dat Project, DataLad) or join their node to the Blockchain storage network (Datum). Moreover, all these approaches provide a useful documentation page with a clear and step-by-step installation guide.

In order to help the user to interact with the system, both of the Git-based and the P2P based approaches provide not only the command-line interface but also a web and desktop applications, e.g., Beaker browser[8], GitHub website, and GitHub desktop. While the Blockchain approach supports dApp (decentralized Application) that hides all the complexity of blockchain in the backend and allows the user interacts through GUI. For example, Datum provides dApp mobile application (iOS and Android apps) for end-user to manage their data.

	P2P based approach	Blockchain-based approach	Git-based approach
Ease of setup and use	+	+	+
+ Completely Fulfilled, O Partially Fulfilled, - Not Fulfilled			

Table 3 Evaluation of User support tools

[8] https://beakerbrowser.com/

Support different file types and large file size

This requirement is fully satisfied by the P2P based approach. All nodes in a peer-to-peer network "share their resources, such as processing power, disk storage or network bandwidth"[29] so that they can handle easily with different file formats and large file size. Hence, it received a "Completely Fulfilled" rating.

Similarly, by using off-chain data storage, the Blockchain-based approach can effectively and efficiently handle the research data. The Storage layer, which builds on a distributed network of storage nodes, provides enough storage capacity and ability to deal with multiple file formats and large file sizes. Therefore, this approach is rated as "Completely Fulfilled".

In the Git-based approach, the original Git is less recommended for handling vast and frequently changing binary files because "every small change to a large binary file will add the complete file to the repository once more" [30]. Because of this reason, Git remote repositories need to set file size constraint, e.g., GitHub does not allow the user to push files larger than 100MB. In order to solve the issue of large file size, there are several third-party extensions, i.e., git-anex, git lfs, and git-bigfiles. One implementation is DataLad[9] which use git-annex and extends it to handle research data characteristics. Hence, this approach can also get a rating of "Completely Fulfilled".

	P2P based approach	Blockchain-based approach	Git-based approach
Support different file format and large file size	+	+	+
+ Completely Fulfilled, O Partially Fulfilled, - Not Fulfilled			

Table 2 Evaluation of Support different file formats and large file size

Data versioning control

Git is a distributed version control system, and it initially designed "for tracking changes in source code during software development" [31]. But Git also "facilitates

[9] https://www.datalad.org/about.html

scientific reproducibility across a wide range of disciplines, from archaeology to zoology. […], the tool allows researchers to store and share their code, analysis scripts, and data, and to ensure analyses are always executed using the appropriate versions of the files. Other researchers can then access those files to see how the work was done and to apply it to their own studies – features that advance research transparency" [32]. Therefore, the Git-based approach satisfies this requirement the best and therefore, is rated as "Completely Fulfilled".

In comparison, there is no data versioning control in the original concept of P2P based approach. However, there are several implementations of the P2P based approach which can integrate Git in the backend to achieve data versioning control, e.g., Dat Project and qri.io[10]. Therefore, it receives a rating of "Completely Fulfilled". In contrast, the Blockchain-based is rated as "Not Fulfilled" because it does not support for data versioning control.

	P2P based approach	**Blockchain-based approach**	**Git-based approach**
Data versioning control	+	-	+
+ Completely Fulfilled, O Partially Fulfilled, - Not Fulfilled			

Table 3 Evaluation of Data versioning control

Metadata integration

Nowadays, metadata becomes "an important part of the entire scientific research data management" [12]. As a result, all these implementations that we analyzed in this thesis, support metadata, but for different purposes and on a different level. For example, Dat-Project uses metadata to track file history and securely share the file. Datum.org supports metadata in JSON format and keep it public. Datalad does a step further by providing JSON-LD metadata, which provides better support. However, regarding the interdisciplinary aspect, there is no implementation of each approach which uses

[10] https://qri.io/

metadata or any artefacts to support the research data management in an interdisciplinary way. Therefore, all three categories only partially satisfy this requirement, which gets the rating "Partially Fulfilled". The evaluation result is shown in Table 4.

	P2P based approach	Blockchain-based approach	Git-based approach
Metadata integration	O	O	O
+ Completely Fulfilled, O Partially Fulfilled, - Not Fulfilled			

Table 4 Evaluation of Support metadata

Data exposure

In term of "Data exposure", there is no category that can adequately meet this requirement. In the Git-based approach, the "push" command is used to publish data from the local repository to remote repository, so that other users can access the data. The remote repository, for example, GitHub or Datalad dataset[11], will become a searchable source. However, data will be replicated and stored in a remote repository, which can lead to the issue of data privacy. Therefore, it is rated as "Partially Fulfilled".

Similarly, the "Data Exposure" of the Blockchain-based approach can be achieved after data is uploaded to the Storage node and a block which contains a hash of data, is added to the Blockchain network. Searching function, which is based on metadata, theoretically is possible. However, there is no implementation of this approach works on the direction to support the interdisciplinary research project, which yields a rating of "Partially Fulfilled".

In contrast, the P2P-based approach receives a "Not Fulfilled" rating. Each protocol that uses the P2P network has a different way to expose, index and search data, e.g., the BitTorrent indexes a .torrent file that contains information about how files should be accessed in a BitTorrent P2P network. Therefore, we rate this approach based on one implementation for research data management- Dat protocol. When the researcher wants to publish and share one folder on his laptop, "*dat share*" command is used to

[11] http://datasets.datalad.org/

index all the files and subfolders, generate and share a Dat link to other researchers to instantly access the data. In other words, the researcher exposes and shares data directly to his peers. Therefore, no data is indexed to any searchable source, which results in the "Not Fulfilled" rating.

	P2P based approach	Blockchain-based approach	Git-based approach
Data exposure	-	O	O

+ Completely Fulfilled, O Partially Fulfilled, - Not Fulfilled

Table 5 Evaluation of Data exposure

3. Conclusion

Based on the evaluation per approach, there are three decentralized approaches to manage research data, and each of them has it owns a share of merits and demerits. As a result, shown in Table 6, the P2P-based approach has the most disadvantaged with two "Not Fulfilled" ratings and two "Partially Fulfilled" ratings. Similarly, the Blockchain-based approach is received one "Not Fulfilled" rating, but it partially fulfilled four requirements of Ownership of data storage, Data access control, Metadata Integration, and Data Exposure. On the contrary, the Git-based method tops the list of approaches shown in the comparison table and thus indicates the highest potential in all the categories detailed. However, there are three requirements (Data Access Control, Metadata Integration, and Data Exposure) that the Git-based can only partially fulfilled.

Based on the evaluation per requirement, the requirements of "User support tools" and "Support different file formats, and large file size" are easier to achieve with three "Completely Fulfilled". In contrast, two requirements of "Data Access Control" and "Data Exposure" seems to be the hardest with two "Partially Fulfilled" and one "Not Fulfilled" ratings. Besides, there is no approach can fully meet the requirement of Metadata Integration (three "Partially Fulfilled" ratings). And the Ownership of data storage requirement is only achieved by one group (the Git-based approach).

To sum up, several decentralized approaches were already presented and evaluated. However, each of them has its limitations and cannot satisfy all requirements, mainly for two crucial requirements of Data access control and Data exposure. Consequently, there is room for improvement, and a research gap visible to check, especially in the context of an interdisciplinary research project.

	P2P based approach	Blockchain-based approach	Git-based approach
Ownership of data storage	O	O	+
Data access control	-	O	O
User support tools	+	+	+
Support different file types and large file size	+	+	+
Metadata integration	O	O	O
Data versioning control	+	-	+
Data exposure	-	O	O

+ Completely Fulfilled, O Partially Fulfilled, - Not Fulfilled

Table 6 Evaluation of the decentralized RDM approaches

Bibliography

[1] Caroline Ingram, "How and why you should manage your research data: a guide for researchers," *Jisc*, 07-Jan-2016. [Online]. Available: https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data. [Accessed: 16-May-2019]

[2] V. Steeves, "Research Guides: Data Management Planning- Research Lifecycle," *NYU DataServices* [Online]. Available: https://guides.nyu.edu/data_management. [Accessed: 20-Jun-2019]

[3] U. Utrecht, "Research Data Management Support: Publishing and sharing data," *Utrecht University*, 2016. [Online]. Available: https://www.uu.nl/en/research/research-data-management/guides/publishing-and-sharing-data. [Accessed: 29-May-2019]

[4] Wiley, "Wiley Open Science Researcher Insights," p. 2016, 2016 [Online]. Available: https://authorservices.wiley.com/asset/photos/licensing-and-open-access-photos/Wiley Global Data Sharing Infographic June 2017.pdf

[5] R. Rylance, "Grant giving: Global funders to focus on interdisciplinarity," *Nature*. 2015.

[6] W. Wang, T. Göpfert, and R. Stark, "Data Management in Collaborative Interdisciplinary Research Projects—Conclusions from the Digitalization of Research in Sustainable Manufacturing," *ISPRS Int. J. Geo-Information*, vol. 5, no. 4, p. 41, 2016.

[7] M. Bower, "LibGuides: Research Data Management @ Pitt: Sharing Data."

[8] D. C. Robinson, J. A. Hand, M. B. Madsen, and K. R. McKelvey, "The dat project, an open and decentralized research data tool," *Sci. Data*, vol. 5, pp. 1–4, 2018 [Online]. Available: http://dx.doi.org/10.1038/sdata.2018.221

[9] E. Graham-Harrison and C. Cadwalladr, "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach," *Guard.*, Mar. 2018 [Online]. Available: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election. [Accessed: 29-Jul-2019]

[10] J. A. Martin, "What is access control? A key component of data security," *IDG Communications, Inc.*, 2018. [Online]. Available:

https://www.csoonline.com/article/3251714/what-is-access-control-a-key-component-of-data-security.html. [Accessed: 01-Jun-2019]

[11] A. Rajabifard, M. Kalantari, and A. Binns, "SDI and Metadata Entry and Updating Tools," *SDI Converg. Res. Emerg. Trends, Crit. Assess.*, vol. 41, no. 2007, pp. 121–136, 2009.

[12] C. Curdt, D. Hoffmeister, G. Waldhoff, C. Jekel, and G. Bareth, "Development of a metadata management system for an Interdisciplinary Research Project," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 1, no. September, pp. 7–12, 2012.

[13] M. Assante, L. Candela, D. Castelli, and A. Tani, "Are Scientific Data Repositories Coping with Research Data Publishing?," *Data Sci. J.*, vol. 15, pp. 1–24, 2016.

[14] L. Stuart, "Open data repository – file size analysis," *Edinburgh Research Data Blog*, 2014. [Online]. Available: http://datablog.is.ed.ac.uk/2014/06/19/open-data-repository-file-size-analysis/. [Accessed: 06-Jun-2019]

[15] Harvard University, "ODAP- The Open Data Assistance Program at Harvard," *Harvard University Blog*, 2014. [Online]. Available: https://projects.iq.harvard.edu/odap/frequently-asked-questions. [Accessed: 11-Jun-2019]

[16] D. S. Falster, R. G. Fitzjohn, M. W. Pennell, and W. K. Cornwell, "Versioned data: why it is needed and how it can be achieved (easily and cheaply)," *PeerJ Prepr.*, p. 5:e3401v1, 2017.

[17] J. (NISO) Riley, *Understanding Metadata - What Is Metadata?* 2017.

[18] M. D. Wilkinson *et al.*, "Comment: The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, pp. 1–9, 2016.

[19] T. Wissik and M. Ďurčo, "Research Data Workflows : From Research Data Lifecycle Models to Institutional Solutions," *Clar. 2015 Sel. Pap. Linköping Electron. Conf. Proceedings, Annu. Conf. 2015, Oct. 14–16, 2015, Wroclaw, Pol.*, no. 123, pp. 94–107, 2015 [Online]. Available: http://www.ep.liu.se/ecp/123/008/ecp15123008.pdf

[20] M. Wu, F. Psomopoulos, S. J. Khalsa, and A. de Waard, "Data Discovery Paradigms: User Requirements and Recommendations for Data Repositories," *Data Sci. J.*, vol. 18, pp. 1–13, 2019.

[21] J. P. Cohen and H. Z. Lo, "Academic Torrents," in *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment - XSEDE '14*, 2014, pp. 1–2 [Online]. Available:

http://dl.acm.org/citation.cfm?doid=2616498.2616528. [Accessed: 16-Jun-2019]

[22] Tara Vancil, "Dweb: Serving the Web from the Browser with Beaker - Mozilla Hacks - the Web developer blog," *Mozilla*, 2018. [Online]. Available: https://hacks.mozilla.org/2018/08/dweb-serving-the-web-from-the-browser-with-beaker/. [Accessed: 16-Jun-2019]

[23] Datum, "Datum - White Paper," 2017 [Online]. Available: https://datum.org/

[24] K. Ram, "Git can facilitate greater reproducibility and increased transparency in science," pp. 1–8, 2013.

[25] DataLad Team, "DataLad," *datalad.org*, 2016. [Online]. Available: https://www.datalad.org/. [Accessed: 21-Jun-2019]

[26] Yaroslav O. Halchenko, "Using version control for code and data," *nipy*, 2017. [Online]. Available: http://nipy.org/workshops/2017-03-boston/lectures/version-control/#1. [Accessed: 16-Jun-2019]

[27] M. Jawad *et al.*, "Supporting Data Privacy in P2P Systems," pp. 0–51, 2013.

[28] Dat Project, "Data sharing between institutions," *Dat Project Blog*, 2018. [Online]. Available: https://blog.datproject.org/2018/04/24/data-sharing-at-institutions-and-beyond-with-dat/. [Accessed: 20-Jun-2019]

[29] R. Schollmeier, "A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications," *Proc. - 1st Int. Conf. Peer-to-Peer Comput. P2P, 2001*, no. September, pp. 101–102, 2001.

[30] git-tower, "How can I handle large binary files (efficiently) with Git LFS?," *fournova Software GmbH*. [Online]. Available: https://www.git-tower.com/learn/git/faq/handling-large-files-with-lfs. [Accessed: 25-Jun-2019]

[31] S. Chacon and B. Straub, "Pro Git," 2014.

[32] J. Perkel, "Git: The reproducibility tool scientists love to hate," *Naturejobs blog*, 2018. [Online]. Available: http://blogs.nature.com/naturejobs/2018/06/11/git-the-reproducibility-tool-scientists-love-to-hate/. [Accessed: 15-Jun-2019]

[33] C. Bizer and F. U. Berlin, "Linked Data - The Story So Far," *Int. J. Semant. Web Inf. Syst.*, 2009.

[34] C. Wiljes *et al.*, "Towards Linked Research Data : An Institutional Approach," *SePublica*, pp. 1–12, 2013.

[35] A. V. Sambra *et al.*, "SoLiD: A Platform for Decentralized Social Applications Based on Linked Data," 2018 [Online]. Available: https://diasporafoundation.org

[36] E. Mansour *et al.*, "A Demonstration of the Solid Platform for Social Web Applications," pp. 223–226, 2017.

[37] S. Andrei, T. Berners-Lee, and D. Zagidulin, "GitHub - solid/solid-spec: The Solid spec and architecture," *GitHub*, 2015. [Online]. Available: https://github.com/solid/solid-spec. [Accessed: 30-Jun-2019]

[38] Henry Story, "WebAccessControl - W3C Wiki," *W3C Wiki*, 2012. [Online]. Available: https://www.w3.org/wiki/WebAccessControl. [Accessed: 18-Jun-2019]

[39] Stanford University, "Data versioning | Stanford Libraries," *Stanford University*. [Online]. Available: https://library.stanford.edu/research/data-management-services/data-best-practices/data-versioning. [Accessed: 10-Jul-2019]

[40] P. Ciccarese, S. Soiland-Reyes, K. Belhajjame, A. J. Gray, C. Goble, and T. Clark, "PAV ontology: provenance, authoring, and versioning," *J. Biomed. Semantics*, vol. 4, no. 1, p. 37, Nov. 2013 [Online]. Available: http://jbiomedsem.biomedcentral.com/articles/10.1186/2041-1480-4-37. [Accessed: 28-Jun-2019]

[41] S. L. Weibel and T. Koch, "The Dublin Core Metadata Initiative," *D-Lib Mag.*, vol. 6, no. 12, Dec. 2000 [Online]. Available: http://www.dlib.org/dlib/december00/weibel/12weibel.html. [Accessed: 08-Jul-2019]

[42] A. Chatzigeorgiou, T. Chaikalis, G. Paschalidou, N. Vesyropoulos, C. K. Georgiadis, and E. Stiakakis, "A Taxonomy of Evaluation Approaches in Software Engineering," pp. 1–8, 2015.

[43] J. Nielsen and T. K. Landauer, "A mathematical model of the finding of usability problems," *Proceeding INTERCHI '93 Proc. INTERCHI '93 Conf. Hum. factors Comput. Syst.*, pp. 206–213, 2009 [Online]. Available: https://dl.acm.org/citation.cfm?id=169166

[44] O. Hartig, "Linked Data Query Processing Based on Link Traversal," no. May 2014, pp. 263–283, 2015.