This study constitutes the basis for multiple future research thrusts to validate the results in a larger scale and implement the proposed VDC planning and control framework. First, more cases of VDC work can be collected to form an acceptable sample for detailed statistical analysis that can test different quantitative metrics of the performance of the VDC tasks considering the project different attributes. Second, the proposed framework can be implemented in real projects and observe its impact on the performance of VDC tasks.

REFERENCES

Bosche F., Guillemet Al., Turkan Y., Haas C., and Haas R. (2014). "Tracking the Built Status of MEP Works: Assessing the Value of a Scan-vs.-BIM System." Journal of Computing in Civil Engineering, ASCE, 28(4), DOI: 10.1061/(ASCE)CP.1943-5487.0000343.

Chong H., Fan S., Sutrisna M., Hsieh S, and Tsai C. (2017). "Preliminary Contractual Framework for BIM-Enabled Projects." Journal of Construction Engineering and Management, ASCE, DOI: 10.1061/(ASCE)CO.1943-7862.0001278.

Chua D. K. and Yeoh J. K. (2015). "Understanding the Science of Virtual Design and Construction: What It Takes to Go beyond Building Information Modeling." Proceedings of Computing in Civil Engineering, Austin, TX, June 21-23, DOI: 10.1061/9780784479247.086.

Clevenger C. and Khan R. (2014). "Impact of BIM-Enabled Design-to-Fabrication on Building Delivery." Practice Periodical of Structural Design and Construction, ASCE, 19(1), 122 – 128, DOI: 10.1061/(ASCE)SC.1943-5576.0000176.

Christensen C. M. (1997). The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail. Boston, MA: Harvard Business School Press, 1997.

Du J., Liu R., and Issa R. (2014). "BIM Cloud Score: Benchmarking BIM Performance." Journal of Construction Engineering and Management, ASCE, 140(11), DOI: 10.1061/(ASCE)CO.1943-7862.0000891.

Hedrick T, Bickman L. and Rog D. J. (1993). Applied Research Design: A Practical Guide. Newbury Park, CA, Sage Publications Inc.

Herriott R.E. and Firestone W.A. (1983). "Multisite Qualitative Policy Research: Optimizing Description and Generalizability." Educational Researcher, 12, 14 – 19.

Hersen M. and Barlow D.H. (1976). Single-case Experimental Designs: Strategies for Studying Behavior. New York, Bergamon.

Holley P. and Mancill S. (2014). "Employing LiDAR Scanning for Drywall Prefabrication: A Qualitative Pilot Proof of Concept Study." Construction Research Congress, Atlanta, GA, DOI: 10.1061/9780784413517.025, 239 – 248.

Jung W. and Lee G. (2015). "Slim BIM Charts for Rapidly Visualizing and Quantifying Levels of BIM Adoption and Implementation." Journal of Computing in Civil Engineering, ASCE, 30(4), DOI 10.1061/(ASCE)CP.1943-5487.0000554.

Kam C., Song M.H., and Senaratna D. (2016). "VDC Scorecard: Formulation, Application and Validation." Journal of Construction Engineering and Management, ASCE, DOI: 10.1061/(ASCE)CO.1943-7862.0001233.

Kassem M, Iqbal N., and Dawood N. (2013). "A Practice oriented BIM Framework and Workflows." Proceedings of ASCE International Workshop on Computing in Civil Engineering, Los Angeles, CA, June 23-25, DOI: 10.1061/9780784413029.066.

Kim S., Chin S., Han J., and Choi C. (2017). "Measurement of Construction BIM Value based on a Case Study of a Large-Scale Building Project." Journal of Management in Engineering, ASCE, DOI: 10.1061/(ASCE)ME.1943-5479.0000551.

Liang C., Lu W., Rowlinson S., and Zhang X. (2016). "Development of a Multifunctional BIM Maturity Model." Journal of Construction Engineering and Management, ASCE, 142(11), DOI: 10.1061/(ASCE)CO.1943-7862.0001186.

Luth G. (2011). "VDC and the Engineering Continuum." Journal of Construction Engineering and Management, ASCE, 137(10), 906 – 915, DOI: 10.1061/(ASCE)CO.1943-7862.0000359.

MacLeamy, P. (2004). "Collaboration, integrated information and the project lifecycle in building design, construction and operation." ⟨http://code bim.com/wp-content/uploads/2013/06/CurtCollaboration.pdf⟩ (Aug. 28, 2014)

McGraw Hill Construction (MHC) (2012). SmartMaket Report: The Business Value of BIM in North America - Multi-Year Trend Analysis and User Ratings (2007 - 2012). McGraw Hill Construction, New York, NY.

Moon H., Kim H., Kamat V.R., and Kang L. (2014). "BIM-Based Construction Scheduling Method Using Optimization Theory for Reducing Activity Overlaps." Journal of Computing in Civil Engineering, ASCE, DOI: 10.1061/(ASCE)CP.1943-5487.0000342.

National Academy of Sciences (NAS) (2009). Advancing the Competitiveness and Efficiency of the U.S. Construction Industry. Committee on Advancing the Competitiveness and Productivity of the U.S. Construction Industry, National Research Council, http://www.nap.edu/catalog/12717.html, last accessed 8/19/2017, ISBN: 0-309-14192-3.

Nawari N. (2012). "Automating Codes Conformance" Journal of Architectural Engineering, ASCE, DOI: 10.1061/(ASCE)AE.1943-5568.0000049.

Park J., Kim K. and Cho Y.K. (2016). "Framework of Automated Construction-Safety Monitoring Using Cloud-Enabled BIM and BLE Mobile Tracking Sensors." Journal of Construction Engineering and Management, ASCE, DOI: 10.1061/(ASCE)CO.1943-7862.0001223.

Shourangiz E., Mohamad M.I., Hassanabadi M.S., Banihashemi S.S., Bakhtiari M., Torabi M. (2011). "Flexibility of BIM towards Design Change." 2nd International Conference on Construction and Project Management, IACSIT Press, Singapore.

Wu W. and Issa R. (2014). "BIM Execution Planning in Green Building Projects: LEED as a Use Case." Journal of Management in Engineering, ASCE, 31(1), DOI: 10.1061/(ASCE)ME.1943-5479.0000314.

Web-Based Data Federation, Archiving, and Curating of Construction Activity and Operation Sounds

Tianyi Zhang[1]; Yong-Cheol Lee[2]; and Seungwon Yang[3]

[1]Ph.D. Student, Dept. of Construction Management, Louisiana State Univ., Baton Rouge, LA 70803. E-mail: tzhan18@lsu.edu
[2]Assistant Professor, Dept. of Construction Management, Louisiana State Univ., Baton Rouge, LA 70803. E-mail: yclee@lsu.edu
[3]Assistant Professor, School of Library and Information Science, The Center for Computation and Technology, Louisiana State Univ., Baton Rouge, LA 70803. E-mail: seungwonyang@lsu.edu

Abstract

As the architecture and construction industries have increasing project requirements and complex processes, a consistent data storing, federating, and sharing method has become critical. However, the establishment of a robust platform for sharing design and construction data has been overlooked. To improve the current practice, this research study adopts a new data archive technology to federate and share the sound data of construction work activities and equipment operations. As the technology of sound recognition and identification has increasingly developed, the construction industry has recognized the significant implications of a sound data analysis of construction work activities and equipment operations. To establish a sustainable research foundation regarding a construction sound analysis and adopt a state-of-the-art sound analytics technology to the construction industry, this paper involves the investigation of a method to federate the distributed web data of construction activity sounds in a data archive and categorize them according to types of sounds and events. As an archive, the DSpace is used to collect the cleaned construction sound data from web and share them with the public. Establishing a sustainable and durable data archive containing sound data is expected to provide valuable and sustainable data assets for researchers and industry professionals to conduct construction sound-related research such as audio-based construction field monitoring.

INTRODUCTION

A data archive that aims to facilitate data storage, management, processing, and sharing, have been widely utilized as a data repository in diverse disciplines such as industrial, biomedical research, or education material produced by organizations or individuals (Smith et al. 2003; Alpert et al. 2016). Consistent data management and sharing has been a pivotal topic because it generally encourages interactive connection and collaboration between researchers and industry professionals and

finally result in new intellectual knowledge or research findings in various domains. In addition, as numerous materials which produced by faculty and researchers are often saved on a personal hard drive or a department server, the sharing and preserving of the data is often not easy (Tansley et al. 2003). Particularly, data management and archiving provide long-term benefits to industries and academia and a great leap forward of society. Likewise, the architecture, engineering, construction, and facility management (AEC-FM) industries also require a reliable data management and sharing approach that allow project participants to seamlessly share necessary information and safely store project data for a long period of time. Even though several data archiving technology has been developed, the AEC-FM industries have not received the benefits of its application. Because of the increasing complexity of research materials such as construction work images or task sounds, the collection, preservation, indexing and distribution of those data are demanding (Smith et al. 2003). As an initial effort, this study involves the application of new data archive technology for establishing the data management and sharing framework. The data source used in this study is the sound data of construction work activities and equipment operations. The importance of sound identification techniques for enhancing the construction work productivity, project control, and safety surveillance has been stressed and studied in several previous studies (Cheng et al. 2016; Cho et al. 2017). Using the sound data, this paper investigates the implementation method and potential benefits of data sharing and archiving for leveraging data management practices in the AEC-FM industries.

LITERATURE REVIEW

For supporting consistent data sharing, management and long-term preservation, diverse data archive technology has been significantly developed for many years. The LONI Image and Data Archive (IDA) is a repository for neuroimaging and biomedical research data which holds data from more than a hundred institutions (Crawford et al. 2016). This data management system plays a pivotal role as a primary data repository to facilitate active research collaboration. NUNDA is a secure data storage system that aims to centralized data management, standardized data processing, and data sharing (Alpert et al. 2016). DSpace is also one of advanced data management and archiving platforms that can explore and develop standards, manage, archive, and preserve digital asset for ongoing research (Smith et al. 2003). Tansley et al. described the current status and highlighted the areas in which DSpace needs work to address an institution's long-term digital repository needs (Tansley et al. 2003). DSpace reflects the current trends in scholarly communication and education, and offers new methods of distributing research material (Barton and Walker 2006). In this paper, DSpace is used to archive the cleaned construction sound data from Web and share them with the public.

DATA ARCHIVING AND DATA CURATING

Data archiving is the process of hosting data in a computer system through curation processes for ensuring both the long-term maintenance and continued

accessibility of the data (Whitlock et al. 2010; Hammersley 1997). In this process, data curation techniques are applied following the data lifecycle model so that research data could be selected, cleaned, processed, deposited, and disseminated to interested communities. After going through the curation steps, the deposited data will have increased quality, reusability, and potentially be shared benefiting much larger scope of the community members. Data archiving through quality curation is becoming the essential part of research projects because there are so much information and data, which are valuable but not easily accessible. Inaccessible data mean that they are not reusable, and thus practically useless. Technologies are changing in an unprecedented speed. As a result, valuable data may not become accessible considering that new technologies replace older ones and new software tools and standards may not always be backward compatible. In addition, due to cybercholarship (or e-science), researchers from various academic fields such as Physics, Biology, or Computer Science are conducting research in a new way by generating massive scale of data and related documentations (Lord et al. 2004). Such massive scale of data can become valuable resources not only for the researchers of the particular project, but for many others in the community by going through the data curation processes and being hosted in a data archive.

A common practice of managing data in an archive is to follow the procedures of the Data Lifecycle Model. Higgins illustrates the detailed processes involved in successful curating of digital objects and data in her Digital Curation Center (DCC) Curation Lifecycle Model (Higgins 2008). Figure 1 shows Key elements of Digital Curation Center (DCC) Curation Lifecycle Model. This model provides a graphical overview of the steps involved in curating and preserving data in a hosting data archive. Major steps include: creating and receiving data; appraisal and selection based on policies and legal requirements; ingestion by transferring data into an archive following policies; preservation actions for data authenticity, reliability, usability, and integrity via data cleaning, validation, assignment of metadata; storing of data securely following standards; promoting access, use and reuse; and transforming data by creating new data from the original. Occasionally other actions such as disposal, reappraisal, and migration of the data occur to maintain the data quality in the archive.

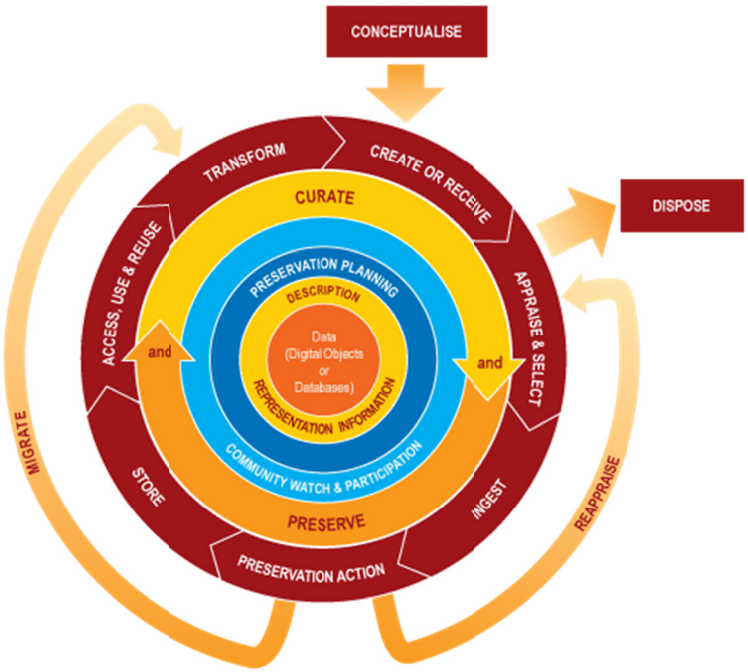

Figure 1 Key elements of Digital Curation Center (DCC) Curation Lifecycle Model (Higgins 2008).

There are hundreds of data archives in every academic field. Among them, two example data archives, from Social Science and Machine Learning communities, might be briefly described here to help understanding of the readers. The Inter-University Consortium for Political and Social Research (ICPSR) is a social sciences archive that was founded in 1962 in the University of Michigan (ICPSR 2005). It hosts approximately 500,000 files in its database, which makes it the largest archive in social science field. Their funding is received from federal grants as well as from the private sector. ICPSR requires users to register before being allowed to download and upload data. ICPSR is free to use for registered users, but some files are restricted and access is granted only with permissions. The archive requests that users submit data files in the format produced by statistical software tools (e.g., SAS, SPSS, Stata), or in ASCII format; however, datasets in other formats are also accepted. Variables and value labels in the dataset should be described as a document, and missing data codes should also be defined prior to deposit of the dataset. Documentation that is accepted includes codebooks, data collection instruments, summary statistics, project summaries, and bibliographies of publications pertaining to the data. Privacy policy indicates that user information is collected and stored securely.

Another popular data archive from Machine Learning field is the University of California, Irvine Machine Learning (UCI-ML) data archive (Asuncion 2007). UCI-ML began as a File Transfer Protocol (FTP)-based archive in 1987 at the University of California, Irvine. The data archive underwent a redesign in 2007 in collaboration with the University of Massachusetts Amherst. The archive stores datasets and related documentations for machine learning research. They host over 330 datasets, with rows of data ranging from 15 rows to 4 million. UCI-ML does not

require a login or registration in order to access the data. It is free to use, with the appropriate access privileges. The archive grants users access to restricted datasets only through direct contact with them. Anonymous uploads are allowed via FTP, email for smaller submissions, and by designating a Web location for retrieving the data. Data providers may request that the archive restrict access to the dataset that they deposited. A potential drawback of this archive is that there is no privacy or security policy.

DSPACE FOR SOUND DATA COLLECTION AND SHARING

DSpace is one of the most popular data archive application software, which was developed by MIT and HP Labs in 2002 (Smith et al. 2003; Tansley et al. 2003). It is an opensource tool for mainly scholarly publications and any other digital data. DSpace supports long-term preservation, access, and storage of the digital content hosted by it.

Figure 2 represents DSpace established on LSU network. To evaluate this system for the AEC-FM industries, this study collected the diverse types of construction sound data including breaking concrete, concrete mixing, concrete pouring, dozer, dumper, excavating ground, grinding concrete, hammering, pilling, and reinforcement welding. These sound data were collected from Web and cleaned to maintain the data quality. The collected sound data are uploaded on DSpace and categorized according to their characteristics.

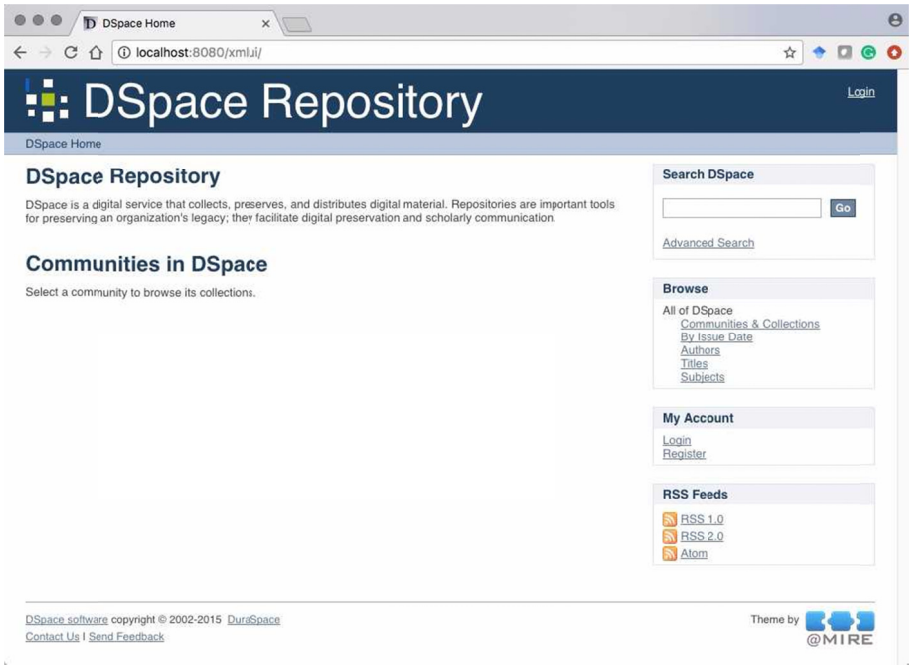

Figure 2 DSpace archive established on LSU

Figure 3 and Figure 4 shows the browsing user interface of the uploaded construction process sound data. To implement the DSpace archive for hosting construction sound files, a Ubuntu Linux operating system was installed to a high-performance rack server, along with prerequisite software components such as Java JDK 8, a relational

database (i.e., Postgre SQL), Apache Maven for building Java programs, and a servlet engine (i.e., Apache Tomcat 7). DSpace is composed of multiple modules including Java API source module, Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) source module, REST API source module, Resource Description Framework (RDF) source module, common services module, and two user interface modules (JSP-UI and XML-UI) to name a few.

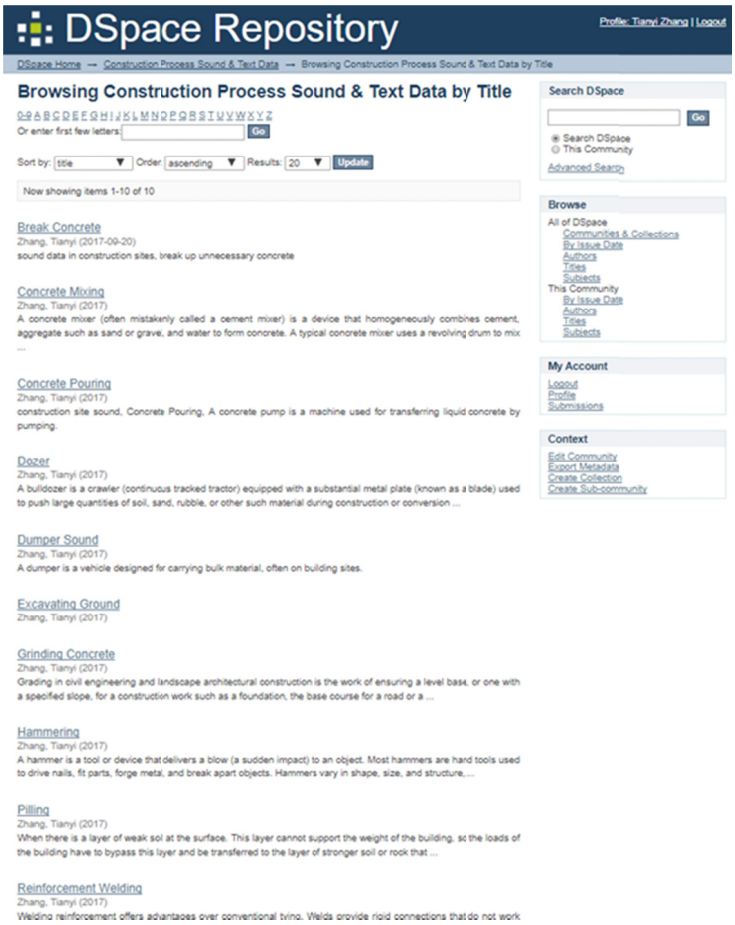

Figure 3 DSpace browsing user interface of construction process sound

Figure 4 Grinding concrete sounds at construction sites

The DSpace archive can host multiple communities and their sub-communities so that different groups of users could share their data. For each community, data collections can be created along with their metadata information detailing the description of the collection, creators, data publication date, and so on. Administrators can assign different permission levels to the users in a community. DSpace has a wide spectrum of user base and there exists thousands of data archives developed using DSpace application software worldwide. Therefore, users and administrators can be supported with tutorials and training programs, as well as from Q/A communities. Due to its opensource nature, the quality of DSpace software is continuously controlled by the developer and user communities.

A most important part of DSpace is data ingest. The DSpace ingest process is shown in Figure 5. A web-based submission user interface is adopted to construct the Submission Information Package (SIP). The approval status of the submission will be determined by reviewers, approvers and edited by editors in the process of Workflow. Once the workflow is approved, the DSpace SIP is converted into DSpace AIP by the Item Installer.

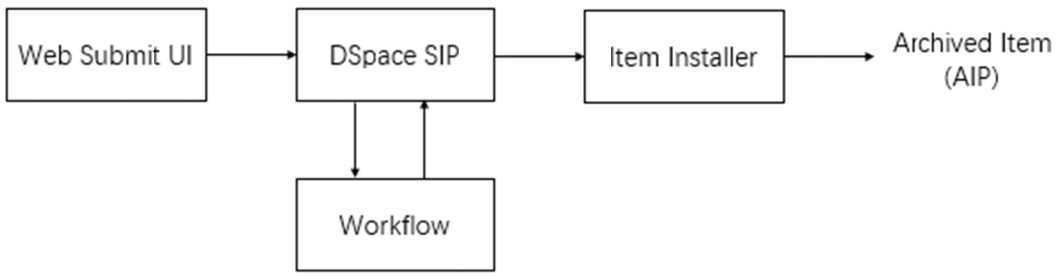

Figure 5 The DSpace ingest process (Tansley et al. 2003 Aug.)

CONCLUSION

Each construction project entails distinct sets of project information, requirements, and facility data. However, since the information contains unique knowledge and capabilities of project participants, the construction industry has a reluctance to share the construction project data with the researchers, industry professionals, students, and the public. However, construction project data sharing and archiving in a robust platform will be increasingly encouraged for improving the construction industry. Specifically, as highly digitalized documents and research data are generated in recent years, the data management and sharing of the AEC-FM industries is vital for project managers, educators, and researchers to rapidly retrieve previous research materials and encourage research collaboration. With the help of the data archive technology, these data can be continuously collected and seamlessly shared by the uses of the defined category description, access permission, and sharing options. The data can be saved in a highly organized format and well preserved by a server, preventing the loss of data caused by individual factors. In addition, the sharing of sound data can avoid repeated labor-intensive work such as sound data collection.

To validate the data sharing approach and platform that can be applied for the construction discipline, this paper collected and archived sound data of construction work activities and equipment operations in the data archive platform, DSpace. The platform provides diverse tools and resources for identifying, integrating, searching, sharing, and visualizing a diverse range of construction sound data, assisting facilitate collaborations between researchers in various disciplines. This web-based data sharing and archiving approach for construction sound data will be a valuable asset for researchers and industry experts who need relevant sound signal data of construction activities and operations. In addition, this initial effort to sharing construction project and field data will be a great foundation that can encourage people in the construction industry to share their knowledge and take an advantage of others' information. The authors recognized the required next steps for settling down this platform for sharing construction project data. First, for sharing sensitive data, data encryption supported in the platform is required. And, an organized management plan for encouraging people to share data such as categorizing users' data access levels according to their data sharing amount is also considered for controlling user accesses to data. Last, but not least, an automated data collection and retrieval feature that can find and acquire the accurate data from Web should be developed. Particularly, it should involve the feature to address a copyright issue of each data source.

REFERENCE

Alpert, K., Kogan, A., Parrish, T., Marcus, D., & Wang, L. (2016). "The northwestern university neuroimaging data archive (NUNDA)." *NeuroImage*, *124*, 1131-1136.

Asuncion, Arthur, and David Newman. "UCI machine learning repository." (2007).

Borgman, C. L., Ashley, E., & Sands, M. S. G. (2017). "Users and uses of a digital data archive: A case study of DANS."