IQRA': Jurnal Ilmu Perpustakaan Dan Informasi
Volume 20 Nomor 1 Mei 2026
ISSN: 1979-7737 E-ISSN: 2442-8175

Digital Preservation Strategy For Research Data At The Center For International Forestry Research (CIFOR) With The Dataverse Repository

Rafa Aqilah
Universitas Padjadjaran

Nurmaya Prahatmaja,
Universitas Padjadjaran

Kusnandar,
Universitas Padjadjaran

Email: rafa22004@mail.unpad.ac.id

ABSTRAK

Data penelitian kehutanan yang dihasilkan Center for International Forestry Research (CIFOR) merupakan data hasil observasi jangka panjang yang sulit untuk dilakukan kembali, sehingga apabila tidak dikelola dengan strategi preservasi digital yang tepat, data tersebut berisiko hilang atau tidak dapat diakses. Penelitian ini bertujuan untuk mengetahui strategi preservasi digital data penelitian melalui repositori Dataverse yang diterapkan oleh CIFOR. Metode penelitian yang digunakan adalah metode kualitatif dengan pendekatan studi kasus. Teknik pengumpulan data dilakukan melalui observasi, wawancara, dokumentasi, dan studi pustaka. Hasil penelitian ini menunjukkan bahwa CIFOR menerapkan tiga dari enam strategi preservasi digital menurut Deegan dan Tanner (2006), yaitu pemeliharaan perangkat lunak dan perangkat keras, penyegaran dan pencadangan (backup), serta migrasi. Adapun tiga strategi lainnya, yaitu emulasi, arkeologi digital, dan alih media ke bentuk analog, tidak diterapkan karena kebijakan preventif yang kuat sejak awal data didepositkan. Kendala yang dihadapi CIFOR mencakup keterbatasan kapasitas penyimpanan akibat penumpukan log files, persepsi peneliti terhadap prosedur Dataverse yang dianggap terlalu kaku, serta keterbatasan sumber daya manusia di bidang kurasi data penelitian.
Kata Kunci: preservasi digital, data penelitian, koleksi digital, repositori Dataverse.

ABSTRACT

The forestry research data produced by the Center for International Forestry Research (CIFOR) consists of long-term observational data that is difficult to replicate. Therefore, if not managed with an appropriate digital preservation strategy, this data is at risk of being lost or becoming inaccessible. This study aims to examine the digital preservation strategy for research data through the Dataverse repository implemented by CIFOR. The research method used is a qualitative approach with a case study design. Data collection techniques include observation, interviews, documentation, and literature review. The

results of this study indicate that CIFOR implements three of the six digital preservation strategies outlined by Deegan and Tanner (2006): software and hardware maintenance, refresh and backup, and migration. As for the other three strategies emulation, digital archaeology, and conversion to analog formats, they were not implemented due to strict preventive policies in place from the very beginning of data deposit. The challenges faced by CIFOR include limited storage capacity due to the accumulation of log files, researchers' perception that Dataverse procedures are too rigid, and limited human resources in the field of research data curation.
Keywords: *digital preservation, research data, digital collection, Dataverse repository.*

PENDAHULUAN

Indonesia has vast tropical forests that play a significant role in the global ecosystem. According to data from the Central Statistics Agency, forests cover more than 50 percent of Indonesia's total land area. This natural wealth has driven a high level of research activity in the forestry sector, where the number of publications and data collection has shown a much faster growth trend compared to many other scientific fields (Aznar-Sánchez et al., 2018). The Center for International Forestry Research (CIFOR) serves as a global research institution that consistently generates scientific data to manage tropical forests sustainably.

In every research activity, data is the primary asset, and the quality of its management will significantly determine the validity and sustainability of the research itself (Yudhanto & Mayesti, 2021). The collection of field observation data, spatial mapping, and socioeconomic surveys gathered by CIFOR constitutes valuable information that is difficult, or even impossible, to replicate in the future. Unfortunately, digital data management faces a serious threat from the rapid obsolescence of technology. Hardware and software that constantly undergo version updates make digital files vulnerable to corruption or becoming inaccessible (Marlina & Purwandari, 2019).

To address these threats, research institutions require a well-defined digital preservation strategy. The American Library Association (2008) defines digital preservation as a combination of policies, strategies, and technical actions to ensure digital content remains accessible despite technological changes or media failures. The theory by Deegan & Tanner (2006) offers six practical steps in this preservation process, including technological preservation, refreshing, migration, emulation, digital archaeology, and media conversion to analog formats.

Technological preservation involves maintaining the software and hardware used to store digital collections. Refreshing, commonly known as backup, is the process of copying digital collections from one medium to another. Migration is the transfer of digital collections to a medium that ensures a longer lifespan. Emulation involves recreating a computer system

or program so that the digital collection can be read again. Digital archaeology involves retrieving digital collections to read information from those that were not previously migrated or backed up. Finally, the strategy of converting to an analog format is employed when digital collections cannot be preserved through other methods. Of the six strategies outlined, not all need to be implemented by an institution; rather, institutions can adapt them to their specific practices.

Unlike most academic institutions that manage final research outputs such as theses or articles (Sari, 2017), CIFOR specifically stores raw and generative datasets using the open-source repository platform Dataverse. The Dataverse repository has an architecture designed for data preservation in accordance with the FAIR (Findable, Accessible, Interoperable, Reusable) principles (Boyd, 2021).

Several previous studies have addressed digital preservation, such as the research by Denanty et al. (2023), which focused on the preservation of ancient cultural manuscripts on a national portal, as well as Safri (2020) study on university repositories. However, studies that specifically examine digital preservation practices for international-level research data using the Dataverse platform remain very limited. Therefore, this study aims to address this gap by critically evaluating the digital preservation strategies implemented by CIFOR, along with the accompanying technical and managerial challenges.

METODE

This study employs a qualitative research design using a case study approach. The research focuses on digital preservation practices for research data managed in the Dataverse repository of the Center for International Forestry Research (CIFOR). Research subjects were selected through purposive sampling based on criteria of direct involvement and operational understanding. The research subjects included four internal CIFOR staff members: a research data manager, a repository publications librarian, a data systems developer, and an IT specialist.

The data used is descriptive qualitative data, which was gathered from two sources: primary data obtained directly from the field and informants, and secondary data obtained through literature and supporting documents. Data collection procedures were conducted through three main methods: passive participatory observation to directly observe the workflow of data managers within the CIFOR environment without intervening; in-depth and flexible semi-structured interviews using an interview guide; and a documentary study examining written evidence such as data management policies, file format guidelines, and repository operational procedures (Sugiyono, 2013).

To ensure the credibility of the findings, this study applied data validity checks using source triangulation and methodological triangulation techniques. Source triangulation cross-checked statements among informants across divisions, while methodological triangulation cross-referenced results obtained from interviews, observations, and policy documents. The validated information was then processed using the Miles and Huberman data analysis model. This analysis process included data reduction to select and simplify raw information, narrative data presentation to identify policy patterns and challenges, and the systematic formulation of conclusions and verification to address all research questions (Sugiyono, 2013).

HASIL DAN PEMBAHASAN

Digital preservation activities at the Center for International Forestry Research (CIFOR) begin with the management of research data, which follows a structured workflow from the very start. From the outset, data deposited into the Dataverse repository has been identified and verified by the Knowledge, Research, and Information Service unit. Through this identification process, CIFOR only accepts and manages research data that is purely derived from the research itself; if the research uses data from third parties, that data is not managed in the Dataverse repository. A preservation strategy is then implemented using the six strategies proposed by Deegan and Tanner (2006).

Technology Preservation

Technology preservation activities at CIFOR focus on maintaining the Dataverse software and its supporting infrastructure. These practices are carried out through cross-departmental collaboration. The Knowledge, Research and Information Service unit is fully responsible for managing the library's content and functionality. Meanwhile, the Information and Communication Technology (ICT) team acts as the administrator, managing servers, networks, and security systems.

In practice, Dataverse software updates are never performed automatically whenever a new release becomes available. The managing unit always conducts a thorough evaluation to compare the added value and advantages of the new version. If the differences are not significant and the old system is still functioning well, they choose to postpone the update. This approach is highly efficient and aligns with the perspective of Deegan and Tanner (2006), who state that continuous technology updates incur high operational costs, making them an unsuitable sole long-term solution.

Delaying this software update is not a problem since it is merely a refinement, as long as the repository's core functions, ensuring data accessibility, continue to operate properly. Additionally, the Dataverse platform has included built-in features since version 4 that strongly support long-term preservation, such as metadata export and data consistency checks

(fixity checks or checksums), which ensure the repository remains stable even if the software is not always running the latest version.

Backup (Refreshing)

CIFOR's data backup strategy is based on a very strict initial policy. Before data is uploaded to the Dataverse system, the managing unit conducts a thorough quality check. This check covers data completeness, document naming standards, and format compliance to ensure preservation-friendly formats. Researchers are required to convert files in proprietary formats to open formats. For example, tabular data in Excel format is converted to .csv format, while text documents are converted to .pdf.

This policy of standardizing to open formats has a crucial impact on CIFOR as an environmental research institution. Forestry data typically has a strong longitudinal nature, where research is based on specific time periods that are often long-term to achieve research objectives. The value of this field observation data actually increases over time because it can be used to obtain the most up-to-date research results. If data from the past cannot be accessed because the format or storage medium of the digital data has become obsolete, this will hinder researchers who wish to conduct studies on that topic. This strategy implemented by CIFOR is highly effective in ensuring that research data remains usable in the future for the latest research or forestry policies.

Once the standard format is implemented, the IT team runs a multi-tiered backup cycle to protect data against system failures. Daily backups are specifically implemented for frontend servers to ensure that restore points are always available based on the previous day's data. Meanwhile, the database is backed up on a weekly or biweekly cycle as an additional verification measure. CIFOR operates four primary servers that support one another: two for the interface and two database servers. CIFOR highly values the research conducted by its researchers, so even research data dating back decades is retained without any deletion process.

Migration

CIFOR is currently undergoing a massive infrastructure transition, migrating its systems from local physical servers (on-premise) to Azure cloud

computing services. This strategic decision was made to address challenges with local hardware and software, where minor failures often resulted in service outages or downtime that hindered researchers' activities.

During this migration process, the IT team is utilizing containerization technology using Docker. This technology allows the Dataverse application to run in an isolated environment, making the system highly flexible and resilient to changes in hardware or software. To maintain stable access during the migration, CIFOR has implemented a parallel-running mechanism. The old servers and the new cloud-based system are run simultaneously to ensure uninterrupted service while data transfer is in progress.

The biggest challenge in this migration is the issue of system updates. The old server runs on Dataverse version 4, while the new environment uses version 6. This version upgrade cannot be performed directly due to fundamental differences in the database structure. The system development team must gradually adjust the database column layout to comply with the latest version's guidelines. Following that, a verification process for five terabytes of data was conducted using random sampling to ensure the file counts matched, which was then revalidated by the management team to detect any files that were counted in the system but turned out to be empty or corrupted upon opening. This step was taken to prevent data from becoming inaccessible in the future.

Digital Preservation Strategies Not Implemented

Based on the results of the field analysis, CIFOR does not implement the three advanced preservation strategies from Deegan and Tanner's theory: emulation, digital archaeology, and analog media migration. This decision was made based on cost calculations and infrastructure suitability.

In the context of emulation, CIFOR does not attempt to replicate or rebuild obsolete software. Emulation practices are considered extremely costly and prone to increasing long-term system maintenance burdens (Rosenthal, 2015). As an alternative solution, the CIFOR development team chose to rebuild the Application Programming Interface (API) available in the old Dataverse and then replicate it into the new database. For some hidden

files that the API failed to retrieve automatically, the infrastructure team performed manual copying directly from the backend without having to build an emulator.

Emergency recovery strategies through digital archaeology were also unnecessary. Such measures are typically employed to extract information from media that has already become obsolete. However, the strong policy of open-format requirements from the outset and the daily backup routine proved highly effective. Research files at CIFOR were always recovered before reaching a critical point that would require these costly recovery methods. In the event of a system failure, data can be fully restored directly from the previous day's backup.

Furthermore, the practice of converting data to analog formats has been completely abandoned. Printing digital documents on paper or copying them to microfilm is irrelevant for CIFOR datasets that were created entirely in digital format (born digital). Converting data into a physical form would actually undermine the core value and functionality of the information. This demonstrates the research institution's flexibility in selecting preservation mechanisms that are truly aligned with the characteristics of its data.

Challenges

Although the infrastructure is well-established, the implementation of digital preservation in the CIFOR repository still faces three challenges. The first challenge is technical and relates to the storage capacity on the local server, which is only about three terabytes. This storage space shortage is apparently not caused by an overwhelming volume of research data uploaded by researchers. The root of the problem is the automatic accumulation of system activity log files. Every user interaction is recorded by the system, generating log files that can swell to as much as four gigabytes in a single day. This accumulation frequently causes the server to crash due to running out of space, and so far, the cleanup must still be performed manually by the IT team without a fixed schedule.

The second challenge are from users' cultural perceptions of the system. Some researchers feel that the data upload procedures in Dataverse are overly

rigid and administratively burdensome. The requirement to document audit trails in a sequential and detailed manner often draws complaints, particularly from researchers accustomed to instant-save methods. A lack of awareness regarding the importance of standardizing data preservation from the outset leads them to view this platform as more suitable for administrators than for scientists.

The most significant operational challenge is the limited number of experts specializing in data curation. The workload of verifying researchers' documents often falls on a single managerial staff member, resulting in response delays of several days. This accumulation of verification tasks directly creates long queues or bottlenecks in the system's workflow.

When viewed through the lens of the library and information science profession, this challenge is a clear manifestation of the unique qualifications required of a Data Librarian. Institutions cannot simply hire pure IT staff for this position. Curation work requires experts who master technical skills (database management and metadata manipulation), while also being able to think critically to assess the methodological completeness and utility of research data for future studies. The gap between these high competency requirements and the availability of experts in the job market confirms the Digital Preservation Coalition's (DPC) warning regarding the threat of a "Skills Crisis" in global information governance. Despite these human resource constraints, CIFOR has proven resilient in managing intellectual property rights by classifying data ownership from the outset and recommending the use of the Creative Commons (CC-BY) open license to prevent the misuse of researchers' work.

PENUTUP

Simpulan

First, CIFOR has successfully implemented three of the six digital preservation strategies. Technology maintenance is not carried out by simply updating software, but rather involves prior evaluation to ensure the system remains efficient. For data refreshment and backup, CIFOR requires the use of easily accessible file formats from the moment data is first submitted. This approach is reinforced by a schedule of daily and weekly backups securely stored on overseas servers and in the cloud. Additionally, to manage large-scale data, CIFOR is gradually migrating its data from local physical servers to cloud services to ensure greater stability and resilience.

Second, CIFOR has decided not to use emulation, digital archaeology, or analog (physical) media migration strategies. Emulation is not used because it is expensive and risky; instead, they use API technology to transfer and reorganize data between servers. Digital archaeology or the recovery of damaged data has also never been performed because CIFOR's file formats and backup systems are already very robust. Furthermore, converting data into physical form is not done, as printing research data could actually eliminate the original functionality of the data.

Third, this data preservation process still faces several challenges. The local server became full not because of the volume of research data, but due to the accumulation of log files (system activity records) generated automatically. From the users' perspective, some researchers have raised objections, finding the data upload and documentation rules too rigid. However, the most pressing issue is the lack of specialized personnel such as Data Librarians or data curators. Adding staff with these specialized skills is crucial to ensure the verification process is not hindered and that the management of forestry research data runs more smoothly in the future.

Saran

For future research, there are two areas of study with significant potential for further exploration. First, exploring researchers' perspectives as repository users to analyze the factors driving and hindering their compliance

with existing data preservation policies. Second, specifically examining the issue of the human resources crisis among data curators in research institutions to map out ideal competency standards, the availability of professionals in the field, and projections for their career paths.

DAFTAR PUSTAKA

Aznar-Sánchez, J. A., Belmonte-Ureña, L. J., López-Serrano, M. J., & Velasco-Muñoz, J. F. (2018). Forest Ecosystem Services: An Analysis of Worldwide Research. Forests, 9(8), 453. https://doi.org/10.3390/f9080453

Boyd, C. (2021). Use of Optional Data Curation Features by Users of Harvard Dataverse Repository. Journal of eScience Librarianship, 10(2), e1191. https://doi.org/10.7191/jeslib.2021.1191

Deegan, M., & Tanner, S. (2013). Digital Futures: Strategies for the Information Age. Facet Publishing.

Denanty, S. A., Kusnandar, & Cms, S. (2023). STRATEGI PRESERVASI DIGITAL PADA KOLEKSI PUSTAKA NUSANTARA DI PORTAL KHASTARA. Jurnal Ilmiah Multidisiplin, 2(04), 35–42. https://doi.org/10.56127/jukim.v2i04.751

Marlina, E., & Purwandari, B. (2019). Strategy for Research Data Management Services in Indonesia. Procedia Computer Science, 161, 788–796. https://doi.org/10.1016/j.procs.2019.11.184

Rosenthal, D. S. H. (2015). Emulation & Virtualization as Preservation Strategies. UNT Digital Library. https://digital.library.unt.edu/ark:/67531/metadc799755/

Safri, T. M. (2020). Strategi Preservasi Digital di Perpustakaan STMIK AMIKOM Yogyakarta. Jurnal Adabiya, 21(2), 84. https://doi.org/10.22373/adabiya.v21i2.6612

Sari, D. P. H. (2017). ANALISIS BENTUK REPOSITORI INSTITUSI DI PERPUSTAKAAN PERGURUAN TINGGI (Studi Kasus di Perpustakaan Universitas Ma Chung Malang) [Fakultas Ilmu Administrasi Universitas Brawijaya]. http://repository.ub.ac.id/id/eprint/8461

Yudhanto, S., & Mayesti, N. (2021). Deskripsi Metadata dalam Manajemen Data Penelitian: Studi Kasus pada Sistem Repositori Ilmiah Nasional. Tik Ilmeu : Jurnal Ilmu Perpustakaan dan Informasi, 5(1), 35. https://doi.org/10.29240/tik.v5i1.2486

Sugiyono. (2013). Metode Penelitian Kuantitatif Kualitatif dan R&D. Alfabet Bandung.

American Library Association. (2008). Preservation Policy.