Journal of Information and Data Management, 2025, 16:1, doi: 10.5753/jidm.2025.4358

APProve: Facilitating Active Monitoring, Versioning and Provenance Tracking of Data Management Plans

Annatercia Gomes Pinheiro [**Universidade Federal do Rio de Janeiro** | ***annatercia@ufrj.br***]
Maria Luiza Machado Campos [**Universidade Federal do Rio de Janeiro** | ***mluiza@ufrj.br***]
Sérgio Manuel Serra da Cruz [**Universidade Federal do Rio de Janeiro** | ***serra@ppgi.ufrj.br***]

Programa de Pós-Graduação em Informática (PPGI/UFRJ)- Universidade Federal do Rio de Janeiro (UFRJ) – Rio de Janeiro, RJ – Brasil

Programa de Pós-Graduação em Humanidades Digitais (PPGIHD/UFRRJ)-Universidade Federal Rural do Rio de Janeiro (UFRRJ) – Nova Iguaçu, RJ – Brasil

Received: 8 April 2024 • **Published:** 23 August 2025

Abstract As modern societies become increasingly data-driven, reliance on research data generated by complex experimental setups and scientific systems has grown. However, applying this scientific knowledge to diverse contexts can be challenging. This paper introduces the Active Plans Provenance (APProve) framework, which offers solutions and mechanisms to monitor variations and trace the lineage of Data Management Plans (DMPs) in a simple and dynamic manner. APProve facilitates the synchronized evolution of DMPs with research projects, ensuring accessibility, shareability, and long-term maintainability for researchers. To assist researchers in tracking DMP versions, we developed a loosely coupled architecture that integrates seamlessly with traditional DMP generation tools. To evaluate the framework's effectiveness, we conducted two distinct experiments, both involving DMPs generated by the ARGOS system. The initial experiment assessed APProve functionalities using a static DMP from the VODAN BR project. The second experiment analyzed an a-DMP from the OpenSoils platform project, a Brazilian soil data governance system. APProve empowers users to monitor changes, visualize version histories, and trace retrospective provenance related to DMP modifications. Additionally, it simplifies the DMP import process and provides comprehensive project visualization, including automated comparison of DMP versions across projects. The experiments demonstrated the framework's efficacy in tracking and visualizing DMP version histories, thereby enhancing the management and evolution of these plans. In the VODAN BR case, APProve captured provenance details and enabled comparisons across revisions. For OpenSoils, the framework facilitated dynamic updates, ensuring alignment with ongoing project changes and highlighting discrepancies. Both cases confirmed APProve's ability to improve accessibility and usability for researchers.

Keywords: Data Management Plans, FAIR Data Principles, Versioning, Data Provenance, Data Governance.

1 Introduction

In modern societies, there is a growing dependence on research data obtained from intelligent experimental apparatuses and various scientific systems. However, repurposing portions of scientific knowledge can present challenges for both the general public and scientists alike. The necessity to access and manipulate research data is clear; nevertheless, researchers often face difficulties in processing data that lacks descriptors and well-defined repositories, thereby hindering data comprehension and reuse.

Amidst this landscape, there has been a notable surge in initiatives promoting the governance of scientific data and research, particularly fueled by recent advancements in artificial intelligence, data science, big data processing technologies, distributed data storage, and open science initiatives [Abraham *et al*., 2019].

According to [Schultes *et al*., 2018], [Papadopoulou *et al*., 2020], [de Souza *et al*., 2020], and [Pinheiro *et al*., 2023], the governance of scientific data has emerged as a cornerstone of Open Science-based endeavors. A diverse set of stakeholders, including academia, industry, funding agencies, and various civil organizations such as the Research Data Alliance, GO FAIR, Force11 Data Citation, and the International Science Council (ICSU), actively advocate for best practices in managing scientific data [Council, 2024]. Furthermore, today, several initiatives focused on research data repositories, such as Re3data, DataCite and OpenAIRE, have proliferated [Sales *et al*., 2020]. These initiatives not only provide repositories for scientific data but also offer curation, citation and mentoring services to scientists to manage the complexities of data storage, organization, and sharing [Abraham *et al*., 2019], [Pinheiro *et al*., 2023].

Hence, the adoption of Data Management Plans (DMPs) as obligatory components by national or international funding agencies is not arbitrary. For instance, since 2017, the São Paulo Research Foundation (FAPESP) has mandated the incorporation of a DMPs in all project submissions as an integral facet of research best practices advocated by the foundation. Similarly, the Brazilian National Council for Scientific and Technological Development (CNPq) has recently enforced a requirement for DMPs in new project research proposals.

A Data Management Plan is a formal document that out-

lines how research data will be handled during a research project and after its completion. It serves as a guideline for researchers, prompting them to plan and describe the phases of their research comprehensively [Simms *et al*., 2017]. This includes considerations such as data collection or generation methods, the methodologies, scientific protocols, and data standards utilized, the conditions for data sharing or openness to the research community, and strategies for data curation and preservation [Cardoso *et al*., 2018].

From the funders' perspective, DMPs aim to ensure transparency and facilitate the location and reuse of datasets and digital artifacts by other researchers and society at large [de Oliveira Veiga *et al*., 2019], [Sales *et al*., 2020]. However, two main problems persist regarding the management and monitoring of DMPs.

Firstly, many researchers are unaware of DMPs or their significance in new grant applications. Some researchers lack the knowledge to produce DMPs or manage them throughout the lifecycle of their research projects [Wittenburg *et al*., 2019].

Another challenge arises from the functional limitations of current DMP generation tools. Most tools produce static DMPs, typically created before the start of a research project, which fail to correlate with the evolving nature of plans throughout scientific investigation [Henning *et al*., 2021].

These issues underscore the need for DMPs to reflect researchers' efforts, methods, and datasets accurately. In response, we have conceived the Active Plans and Provenance (APProve) framework to address these challenges.

APProve is a modular architecture designed to support principal investigators, researchers, data analysts or funding institutions by simplifying the display of versioning in static or active DMPs. Leveraging traditional data provenance methods [Moreau, 2010], [da Cruz *et al*., 2018] and FAIR Data Principles [Wilkinson, 2016], [Wittenburg *et al*., 2019].

APProve enhances scientific data governance by ensuring digital objects, particularly data, are findable, accessible, interoperable, and reusable. By integrating FAIR Principles into scientific data governance, APProve fosters increased data reuse potential for new research endeavors while upholding ethical, legal, and contractual constraints.

This article extends the work presented at the Brazilian e-Science and Data Science Symposium (BRESCI)/Brazilian Symposium on Database Systems (SBBD) 2023 [Pinheiro *et al*., 2023], showcasing how the APProve architecture enhances existing DMP generation tools. Furthermore, we present a proof of concept through two experiments in the Health and Digital Agriculture domains . The experiments were developed under the GO FAIR Brazil umbrella [Sales *et al*., 2020]. We investigated versions and changes in the static DMP of the Virus Outbreak Data Network Brazil (VODAN-Br) project [Campos, 2020; Borges and Campos, 2023; de Oliveira *et al*., 2022] and the a-DMP (active DMP) of the OpenSoils project [da Cruz *et al*., 2018] and [Cruz *et al*., 2018].

This article is structured as follows: Section 2 compares the main DMPs generation tools in terms of data curation, FAIR Data Principles, and data provenance. Section 3 outlines the materials and methods of the investigation. Section 4 introduces the APProve architecture. Section 5 presents the experimental results and discussions for the aforementioned projects. Finally, Section 6 concludes with remarks, limitations, and future work considerations.

2 Related Works

In contemporary scientific research, there is a growing trend towards adopting a data-centric approach. Data-centric scientific experiments prioritize the collection, analysis, and interpretation of data as their central focus. In these experiments, the generation and manipulation of data take precedence over other aspects of the research process.

Data-centric scientific experiments are challenging scientists, forcing them to acquire novel skills, such as digital curation and tracking data provenance [Mattoso *et al*., 2010; Moreau, 2010; Cruz *et al*., 2009]. These elements are pivotal in the research lifecycle, ensuring the long-term preservation of data and other digital assets, with the overarching goal of ensuring reusability, reproducibility, transparency, and other essential aspects [Poole, 2015].

However, despite the traditional nature of data curation in Open Science and the extensive literature available, there is currently a lack of identified strategies for producing Data Management Plans (DMPs) or Data Management Plan Annotations (a-DMPs) that explicitly support versioning and lineage. This absence impedes researchers from effectively tracking their plans [Simms *et al*., 2017], [Cardoso *et al*., 2018], [Miksa, 2021].

As defined by [Cardoso *et al*., 2018], a DMPs is a formal document of a static nature, typically up to two pages in length. It comprises descriptions of project-produced data and metadata, legal and ethical constraints regarding data sharing, privacy policies, security measures, and intellectual property considerations. Furthermore, it may encompass policies for data preservation and sharing, along with mechanisms, formats, and standards for storing data to ensure accessibility to third parties.

According to [Miksa, 2021], an a-DMPs is a dynamic and machine-actionable variation of the DMPs, aligned with FAIR Principles, facilitating data exchange among various entities throughout the entire research lifecycle.

In this section, we present a comparative study examining the functionalities of prominent DMPs generation tools, following the method outlined by [Jones *et al*., 2020]. Notably, tools such as DMPOnline (https://dmponline.dcc.ac.uk/), DMPTool (https://dmptool.org/), Data Stewardship Wizard (https://ds-wizard.org/), and ARGOS (https://dmptool.org/) are among the widely used platforms in the scientific community.

In line with the findings of [Sharma, 2023], our analysis indicates that the available DMPs tools predominantly generate static plans in closed PDF formats, lacking indications of versioning. Additionally, current tools fail to establish correlations between the provenance of plan changes and do not offer visualization features for changes in DMPs to researchers.

Our preliminary investigations [Pinheiro *et al*., 2023] indicated that the tool that best suited the needs for adding metadata annotations on Data Management Plan was AR-

GOS [Papadopoulou *et al*., 2020]. It is an open-source and extensible tool that offers services for creating, validating, monitoring, and publishing exportable textual DMPs in JSON format (standard, interoperable across different machines/systems) but, unlike the aforementioned tools, allows editing of datasets and plans.

Our comparative analyses revealed that within the scope of this work, ARGOS possesses features that classify it as one of the best options for supporting the entire research lifecycle in Open Science projects, such as VODAN-Br and OpenSoils projects (Table 1).

ARGOS allows for the creation of DMPs correlated with the datasets associated with the research. By default, its plans have five sections (*project, grant, authors, description, and license*) that are created privately and can later, if necessary, be shared with project teams. The DMPs and their datasets are initially created as drafts files in JSON format and treated as updatable documents before being published.

Additionally, DMPs and a-DMPs generated by ARGOS can be identified through ORCID and DOI unique indentifiers. Despite these advantages, it does not offer native features to track the evolution of plan versions. In other words, it is not possible to verify through the tool itself what has been changed or even who and when a plan was modified. In summary, ARGOS overwrites the files of a given DMP, does not correlate its versions, and does not add provenance metadata to the plans.

Conceptually, the stages of elaborating a DMPs in ARGOS are: *draft, validation, finalization*, and *publication*; finalized with preservation in the Zenodo repository, allowing the plans to be directly shared in an open data curation environment.

Finally, an additional significant observation pertains to the DMP templates provided by ARGOS. While they are extensive and may be somewhat cumbersome to complete, they offer comprehensive customization options. In order to assist researchers, we suggest an approach aimed at streamlining the management of evolving DMPs and tracking changes in research projects.

3 Materials and Methods

This research is characterized by being of a theoretical-applied nature. Following a systematic literature review, we adopted the method proposed by [Jones *et al*., 2020] to compare the functionalities of DMP-producing tools (Table 1).

The materials used in the construction of APProve include PHP, HTML5, Javascript languages, Python libraries for generating and visualizing provenance graphs, and manipulating JSON files. The code extensively utilizes frameworks like Bootstrap to ensure responsiveness and page styling, as well as jQuery, a fast, lightweight, and feature-rich JavaScript library for interactivity.

The data schema of APProve is persisted locally in an instance of the PostgreSQL DBMS v. 13. It was modeled based on the relational schema of ARGOS itself to store DMPs and a-DMPs and their versions, as well as project provenance data and metadata. The source code of the tool and its datasets are freely available at https://github.com/annatercia/Approve.

Table 1. Comparative table of DMP tools (up to march 2024).

Attribute	DMPOnline	DMPTool	DSW	ARGOS
Template	Partial	No	Partial	Yes
Export	Partial	No	Partial	Yes
Ease of use	Partial	Yes	Yes	Yes
Support	Yes	Yes	Yes	Yes
Actionable	No	No	Yes	Yes
FAIR	Yes	Yes	Yes	Yes
Open-source	Yes	Yes	Yes	Yes
Edit Data	No	No	No	Yes

As a proof of concept, we evaluated the static DMPs of the VODAN-Br project [Campos, 2020] and the a-DMPs of the OpenSoils project [da Cruz *et al*., 2018]. In summary, the former project establishes a federated data infrastructure aligned with FAIR principles to support the collection of data from medical records of patients infected by the SARS-CoV-2 virus, while the latter is a cloud platform that allows for the secure tracking of scientific experiments and soil mapping.

4 APPROVE Architecture

APProve is a modular and loosely coupled architecture designed to seamlessly integrate with the data layer of ARGOS. It provides funding agencies, research institutions, researchers, or administrators with a comprehensive overview of the projects and associated Data Management Plans stored within. Through streamlined interfaces, stakeholders can efficiently monitor the evolution of plan versions throughout the project lifecycle.

Figure 1 delineates the core components of the architecture and elucidates the data flows (denoted from 1 to 5) signifying the trajectory of a DMPs within ARGOS and its interfacing with the proposed tool. Unnumbered dashed lines denote the data exchanges between the elements of APProve.

The first flow depicts the creation of a static DMPs in JSON format, while the subsequent flow entails the loading of this file for validation and publication by ARGOS (flows 3 and 4). Flow 5 denotes the loading of static DMPs to be manipulated within APProve.

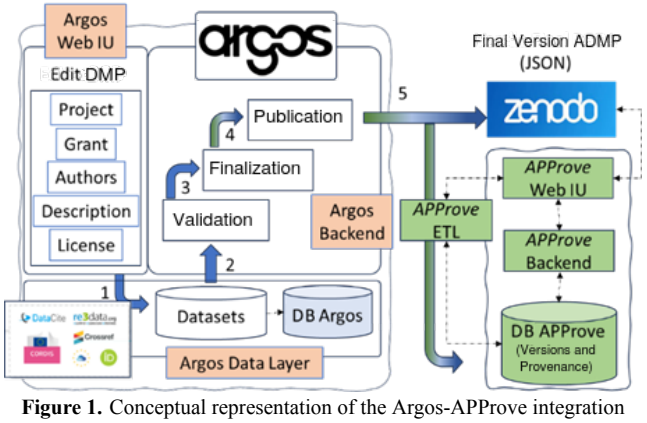

Figure 1. Conceptual representation of the Argos-APProve integration

Among the essential functionalities of APProve, we emphasize the capabilities of its loading component (ETL pipeline), which can import and process JSON files of DMPs from ARGOS, persisting them in APProve's data layer.

APProve provides a comprehensive top-down visualization of previously registered research projects and details about the researchers in the working group. Additionally, it offers editing functionalities in the Web UI frontend and includes features for user management, permissions, and version administration in the backend.

Furthermore, APProve automatically gathers retrospective provenance metadata for each operation performed on the DMPs versions. These provenance metadata are locally stored in the database, persisting in the tool's data layer, and linked with the changes made to the DMPs sections and file versions.

4.1 APProve Core Components

This section presents the essential components of the APProve architecture. They perform functions to manage, visualize, and track provenance throughout the lifecycle of a project. The main components are:

1. **Versioning**: APProve offers the capability to track the versioning of DMPs, addressing a gap in traditional DMP and PGDA tools. This feature allows researchers to monitor changes made to DMPs throughout the project's lifecycle,filling a gap in traditional DMPs tools.
2. **ETL - Extraction, Transformation, and Loading module**: When integrated with ARGOS, this component facilitates the Extraction, Transformation, and Loading of DMP into the APProve database. The component simplifies the import, transformation, and storage of JSON files in the database, ensuring semantic integration at the data layer level.
3. **Provenance metadata collection and storage agents**: APProve can collect and locally store retrospective provenance metadata in the form of DAG (Directed Acyclic Graph) graphs. These graphs record each change made in the versions of the DMPs. The plan metadata is compatible with the W3C PROV-DM standard [Moreau, 2013], representing entities, agents, activities, and relationships between entities.
4. **User interface**: APProve offers a user-friendly and responsive web interface accessible across various types of devices. This interface facilitates the visualization, tracking, and comparison of different DMPs versions through interactive boards and dashboards. It provides indicators of variation points and changes over time, incorporating the use of third-party libraries for generating graphs of retrospective provenance.

In short, APProve enables the visualization of retrospective provenance graphs associated with changes in DMPs or a-DMPs versions. The incorporation of advanced provenance practices in tracking each operation is a distinctive feature of our proposal, contributing to maintaining the integrity, authenticity, and reliability of DMPs over time.

4.2 DMP Versioning in APProve

We have adopted the versioning approach proposed by Preston-Werner [2015]. Versioning enables the classification of changes to plans, ranging from minor corrections that have no significant impact on research outcomes to major changes that have notable effects on results or the overall conduct of the project.

APProve's versioning follows the following pattern for versions:

1. **Major changes**: This indicates whether the original DMPs version has undergone substantial alterations, making it incompatible with previous versions. These changes are structural updates that have the potential to "break" a DMPs. A change is considered substantial if it corresponds to a change rate exceeding 30% of the DMPs files.
2. **Minor changes**: This indicates that a new version has been created or that a significant improvement has been added to the DMPs, without presenting compatibility issues with previous versions. A change is considered an improvement if it corresponds to a change rate below 10% of the DMPs files.
3. **Patch**: This represents a specific change in part of the DMP, such as variations in scientific software or methodology, bug fixes, or similar changes that do not substantially alter a DMP or introduce new steps in the research. A change is considered a specific alteration if it corresponds to a change rate below 1% of the DMPs files.

Upon any type of modification, an indication of the versioning type is stored alongside the DMPs file, and a new version of the DMPs is created by APProve. Additionally, these changes are represented in the graphs of alterations of a DMP version.

5 Experimental Results and Discussion

This section presents the computational experiments conducted to evaluate the functionalities of APProve. We considered static and active DMPs of two research projects being conducted by the authors research groups. For each project, we used two new versions of the plan to assess APProve's capabilities. We established, as an experimental condition, that a new version would be generated if any changes occurred after the completion of the previous version of the DMPs.

The results of the computational experiments aimed to evaluate APProve's functionalities using the DMPs of the VODAN-Br project (https://vodanbr.github.io/) and the DMPs of the OpenSoils project (https://www.opensoils.org/), both created from the ARGOS tool using the Horizon 2020 and RDA-madmp-only Model templates, respectively. Both experiments considered the five standard sections of a DMPs as the starting point for the computational experiments.

5.1 Static DMP experiments

APProve provides a top-down view of projects, diplay information about researchers, workgroups, and registered DMPs, offering an interactive interface to list and visualize detailed information about specific projects. For the first experiment, we considered the static DMPs of the VODAN-Br project.

Figure 2 identifies the main descriptive attributes of the VODAN-Br project with the two query options of ”Plans” and ”Researchers”.

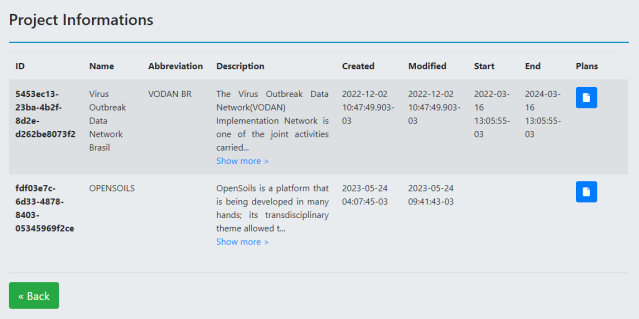

Figure 2. APProve screen fragment with visualization of the DMP descriptors for VODAN-Br project

When navigating through the “*Plans*” interface, a user will be able to compare the three versions of a VODAN-Br DMPs and identify the variability and percentages of variation in the sections of a plan (Figure 3). The “*Researchers*” interface displays all researchers registered in DMP. APProve automatically compares plan versions and displays variability sections and percentages in dynamic, online DMPs variability tracking reports.

The similarity percentage is calculated by comparing Strings. Therefore, this comparison is calculated automatically using the functions: *strcasecmp* and *similar_text*. This way, the similarity between the strings is calculated and printed using the round function, passing the percentage calculated with the *similar_text* function.

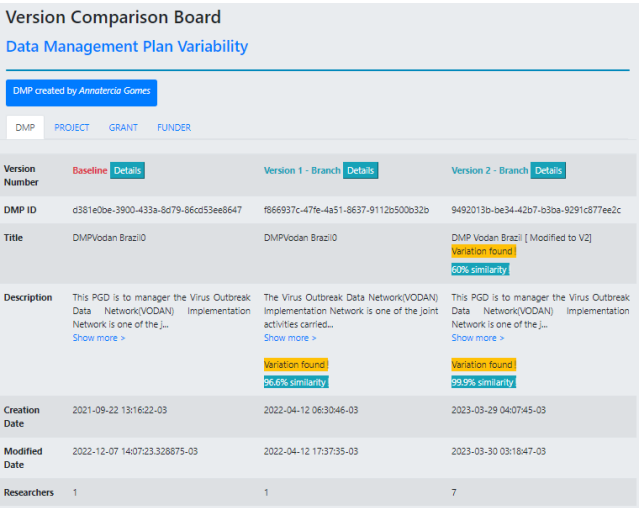

Figure 3. Fragment of the APProve screen with a comparison of three versions of the same DMP for VODAN-Br project

Through the ’*Show more* >’ links (embedded in the dynamic reports), it is possible to individually view the altered sections of a DMPs. For example, in the ’Version 0’ column of Figure 3, the user can see who was responsible for creating the first version of the DMPs and the subsequent changes to that version, along with when and which actions occurred, as illustrated in Figure 4. Additionally, the user can review the textual structure of the DMPs version by selecting the ’JSON Structure’ interface. Moreover, it is possible to view the graph of the provenance of the changes occurred in the DMP by selecting ’Provenance Graph’.

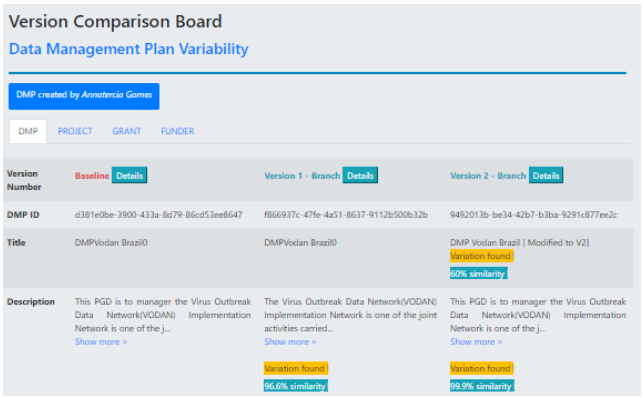

Figure 4. APProve screen fragment showing as example of variability tracking in the Description section of the DMP for VODAN-Br project

When comparing different versions of the plan (Figure 3), it not only simplifies the illustration of altered sections but also provides the percentage of variation in each section compared to the previous version. By clicking on the ’*Show more* >’ links, users can delve into the changes made in the updated DMPs version. For instance, in Figure 5, users can identify the individual responsible for the modifications, the timestamp of the changes, and their specific locations. Moreover, users have the option to download intermediate or final versions of the DMPs.

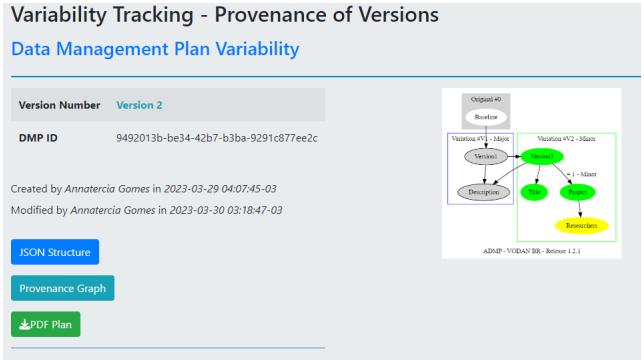

Figure 5. Detailed view of the versions of a DMP

When accessing the ’*Provenance Graph*’ interface, users can download the retrospective provenance graph generated by APProve which adheres to the W3C PROV standard and tracks the alterations made to the DMPs. Figure 6 illustrates a segment of a provenance graph depicting the action of the agent ’Researcher: AnnaterciaGomes,’ who modified the description of the entity ’Vodan-Br:PGD-v0.json’, resulting in the creation of a new version ’Vodan-Br:PGD-v1.json’ that incorporates a new dataset named ’Sirio_Libanes001.zip’.

The primary outcome reveals that APProve facilitates the swift automatic comparison of DMPs versions. Additionally, the tool effectively captures discrepancies between plan ver-

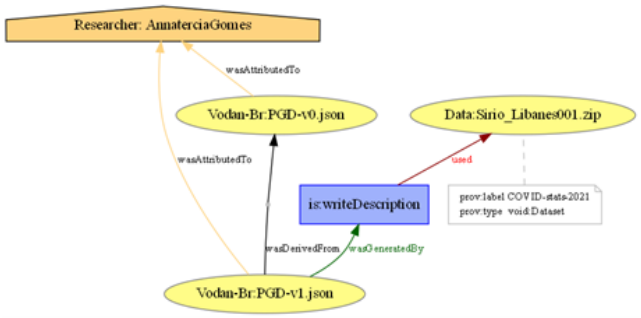

Figure 6. Fragment of provenance graph of an ADMP change

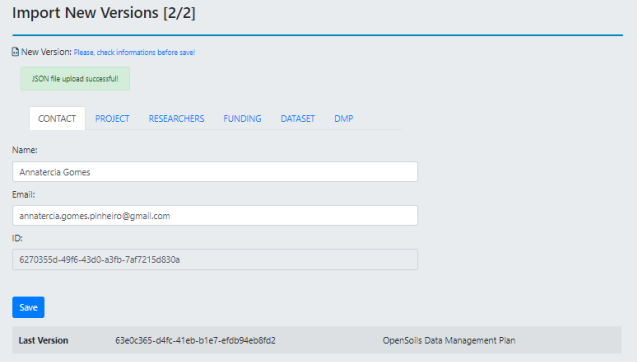

Figure 7. Screenshot of APProve for Importing Version of OpenSoils DMPA - Contact Tab with loaded contact information

sions by visualizing the provenance of plan alterations. In the comparison process, apart from promptly illustrating the areas and extent of divergence between plan versions, users can identify the individual responsible for the changes, the timestamps of their occurrence, and their specific locations through the visualization of the provenance graph.

5.2 Active DMP Experiments

In the second experiment, a different type of model, specifically the RDA-madmp-only model, was utilized for the active Data Management Plan (a-DMPs) of the OpenSoils project. Unlike the static Horizon 2020 model employed in the VODAN-BR project, the machine-actionable RDA model yields a more comprehensive output file. This output encompasses various elements, including the contact metadata of the researcher who completed the plan, which can offer additional insights into the described data in the a-DMP, as well as metadata pertaining to the research dataset.

Following the export of the JSON file from ARGOS, it is transferred to APProve. Subsequently, the system undertakes the importation and transformation of the a-DMPs through the Extract, Transform, Load (ETL) components, adapting the data to conform to the structure of APProve's local database.

After this stage, researchers gain the capability to access and contrast various versions of the a-DMPs via dynamic reports, which showcase the variability of the plans while highlighting the points of variation along with the corresponding percentage of variation or similarity.

Figures 8 and 9 depict two segments of the APProve interface, providing a visual juxtaposition between two versions of the a-DMP directly within the tool's interfaces. Figure 8 presents details of the section, offering crucial information

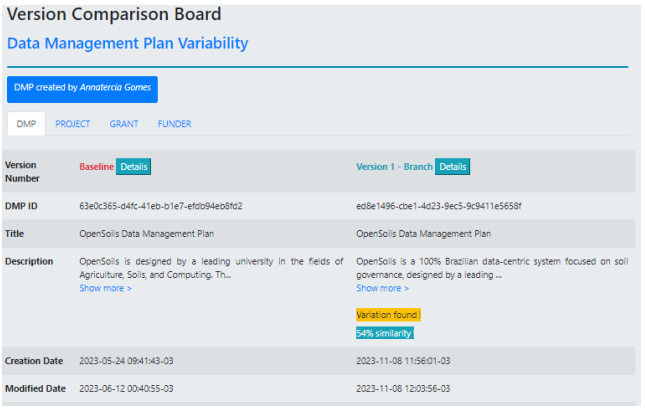

Figure 8. Fragment of the APProve screen showing a comparison of two versions of the a-DMP from the OpenSoils project.

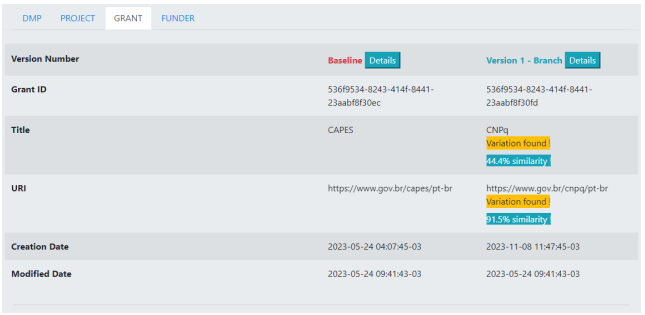

Figure 9. Screenshot of APProve displaying a comparison of the GRANT section in two versions of the a-DMP for the OpenSoils project

about the a-DMPs, including its title, description, and creation and modification dates.

Figure 9 shows the variations found in the Grant section, for example. The percentage of variability is calculated by comparing strings. In the context of the OpenSoils a-DMP, the GRANT that was originally inserted in the Baseline had a mistake, which was corrected in Version 1.

Despite the change, the associated URI did not undergo significant changes. Intuitively, the user can identify, in a matter of seconds, the exact point where the modification occurred and evaluate its respective impact.

Provenance data is depicted through DAG graphs, which outline the relationships between plans, activities, agents, and data. Each alteration in the plan is annotated with temporal details, providing a chronological overview of the plan's progression over time. Consequently, each activity is documented and visualized within the graph, establishing accountabilities for all parties involved in the process.

These accountabilities extend beyond the initial proposers of the plan to encompass reviewers, approvers, and implementers. This allocation of responsibilities contributes to effective governance and accountability throughout the plan's lifecycle.

Throughout the project's evolution, nuances in the initial plan filling can be identified, indicating potential inaccuracies (patches) or, in certain scenarios, substantial changes in its conception (major changes). Such modifications can range from adjustments in team composition, with the inclusion or exclusion of members, to the adoption of emerging technologies (minor changes).

In response to identified inaccuracies or substantial scope changes, it becomes imperative to review and refine the plan

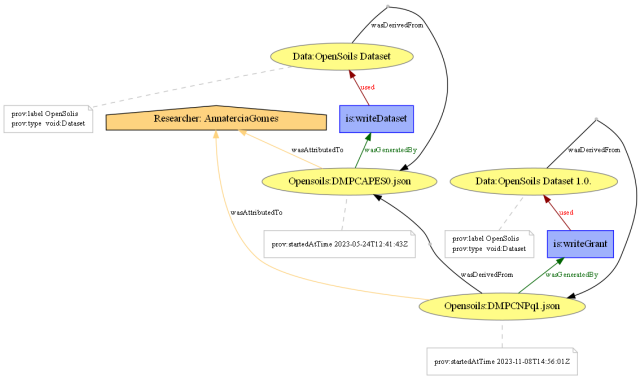

Figure 10. Provenance graph of Version 1.0 of the OpenSoils DMP

to uphold its fidelity and relevance. Consequently, as the project progresses, the plan undergoes modifications, culminating in the generation of a new version to accurately reflect the project's reality.

Within the specific context of the OpenSoils project, a need arose to rectify information initially entered inaccurately by the individual responsible for completing the plan. This correction was facilitated by generating a new version of the document.

One of the elements adjusted in version 1.0 of the plan pertains to the "Grant" associated with the project. In the original version, a granting agency for a scholarship was erroneously indicated, which did not align with the actual circumstances. However, in version 1.0, this discrepancy was duly rectified. This modification is readily discernible in the provenance graph, as illustrated in Figure 10.

6 Final Remarks

The convergence of big data, data science, artificial intelligence, and the exponential growth of scientific data production underscores the critical need for DMPs to organize, collect, preserve, and share research information effectively.

Various stakeholders and funding agencies recognize the importance of research data governance as an essential part of good research practices [Wilkinson, 2016]. Therefore, they encourage the sharing of research project data, aiming for the greatest possible benefit to the scientific, technological, socio-economic, and cultural advancement of a nation.

APProve emerges as one of the pioneering integrated tools designed to support managers, researchers, and institutions in effectively managing DMPs and a-DMPs versioning, facilitating visual comparisons between different versions, and recording change provenance. Our experiments demonstrate the seamless integration of APProve with ARGOS, streamlining typical data management tasks and alleviating the workload on stakeholders.

In our view, a-DMPs represent an innovative approach that is essential for dynamic research environments. By tracking versions, a-DMPs evolve beyond traditional static and active DMPs, becoming dynamic resources that adapt to changes in data and datasets throughout the research lifecycle. As scientific research grows in volume and complexity, a-DMPs can be considered a crucial initial step in helping researchers address ethical concerns, such as bias, privacy, data ownership, intellectual property, and transparency in research development and deployment.

As future work, we aim to introduce new functionalities: including generating reports of a-DMPs in PDF format, incorporating new graphical representations to visualize changes, and expanding testing to encompass a broader spectrum of a-DMPs from diverse domains, involving more users across distributed environments in the AWS cloud.

Lastly, we advocate for further research to explore the impact of successful a-DMP implementation, facilitated by APProve, on fostering collaboration among different institutions, researchers, and working groups.

Acknowledgements

The authors thank CNPq (processes 306115/2021-2 and 400044/2023-4), the PET-SI/UFRRJ program maintained by the National Education Development Fund, the AI4AGROIB project of the Red CYTED and PPGI/UFRJ Program.

Funding

This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Financing Code 001.

Authors' Contributions

AGP contributed to the conception of this study. AGP, MLMC and SMSC performed the experiments. AGP, MLMC and SMSC are the main contributors and writers of this manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they do not have competing interests.

Availability of data and materials

The datasets and source-codes generated and/or analysed during the current study are available in https://github.com/annatercia/Approve.

References

Abraham, R., Schneider, J., and vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49:424–438. DOI: https://doi.org/10.1016/j.ijinfomgt.2019.07.008.

Borges, V. and Campos, M. L. M. (2023). Vodan br – uma plataforma de apoio para dados covid-19 seguindo os princípios fair.

Campos, M. L. M., e. a. (2020). Vodan brazil - the brazilian experience at the virus outbreak data network.

Cardoso, J., Miksa, T., and Borbinha, J. (2018). Debunking active data management plans. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 5308–5310. DOI: 10.1109/BigData.2018.8621860.

Council, I. S. (2024). We bring the world together through the power of science.

Cruz, S. M. S., Ceddia, M. B., Tàvora Miranda, R. C., Rizzo, G., Klinger, F., Cerceau, R., Mesquita, R., Cerceau, R., Marinho, E. C., Schmitz, E. A., Sigette, E., and Cruz, P. V. (2018). Data provenance in agriculture. In Belhajjame, K., Gehani, A., and Alper, P., editors, *Provenance and Annotation of Data and Processes*, pages 257–261, Cham. Springer International Publishing.

Cruz, S. M. S. d., Campos, M. L. M., and Mattoso, M. (2009). Towards a taxonomy of provenance in scientific workflow management systems. In *2009 Congress on Services - I*, pages 259–266. DOI: 10.1109/SERVICES-I.2009.18.

da Cruz, S. M. S., Ceddia, M. B., Schmitz, E. A., Rizzo, G. S., Miranda, R. C. T., Cruz, S. O., Correa, A. C., Klinger, F., Marinho, E., and Cruz, P. V. (2018). Towards an e-infrastructure for open science in soils security. DOI: 10.5753/bresci.2018.3273.

de Oliveira, N. Q., Borges, V., Rodrigues, H. F., Campos, M. L. M., and Lopes, G. R. (2022). A practical approach of actions for fairification workflows. In Garoufallou, E., Ovalle-Perandones, M.-A., and Vlachidis, A., editors, *Metadata and Semantic Research*, pages 94–105, Cham. Springer International Publishing.

de Oliveira Veiga, V. S., Henning, P., Dib, S., Dib, S., da Costa Lima, J., da Silva, L. O. B., and Pires, L. F. (2019). *Liinc em Revista*, 15(2). DOI: 10.18617/liinc.v15i2.5030.

de Souza, D. L., Zambalde, A. L., Mesquita, D. L., de Souza, T. A., and da Silva, N. L. C. (2020). A perspectiva dos pesquisadores sobre os desafios da pesquisa no brasil. *Educação e Pesquisa*, 16.

Henning, P., da Silva, L. O. B., Pires, L. F., van Sinderen, M., and Moreira, J. L. R. (2021). The fairness of data management plans: an assessment of some european dmps. *Revista Eletrônica de Comunicação, Informação amp; Inovação em Saúde*, 15(3). DOI: 10.29397/reciis.v15i3.2270.

Jones, S., Pergl, R., Hooft, R., Miksa, T., Samors, R., Ungvari, J., Davis, R. I., and Lee, T. (2020). Data management planning: How requirements and solutions are beginning to converge. *Data Intelligence*, 2(1-2):208–219. DOI: 10.1162/dint$_{a0}$0043.

Mattoso, M., Werner, C., Travassos, G. H., Braganholo, V., Ogasawara, E., Oliveira, D., da Cruz, S. M. S., Martinho, W., and Murta, L. (2010). Towards supporting the life cycle of large scale scientific experiments. 5(1):79–92. DOI: 10.1504/IJBPIM.2010.033176.

Miksa, T., e. a. (2021). Application profile for machine-actionable data management plans. *Data Science Journal*, 20.

Moreau, L. (2010). The foundations for provenance on the web. *Foundations and Trends® in Web Science*, 2(2–3):99–241. DOI: 10.1561/1800000010.

Moreau, L., M. P. e. a. (2013). Prov-dm: The prov data model.

Papadopoulou, E., Kakaletris, G., Tziotzios, D., Moa, H., and Hasan, A. (2020). Argos: a collaborative tool to plan and follow your data.

Pinheiro, A. G., Cerceau, R., Campos, M. L. M., and da Cruz, S. M. S. (2023). Uma abordagem de acompanhamento da evolução de planos de gestão de dados ativos. pages 17–24. DOI: 10.5753/bresci.2023.233895.

Poole, A. H. (2015). How has your science data grown? digital curation and the human factor: a critical literature review. *Archival Science*, 15:101–139.

Preston-Werner, T. (2015). Semantic versioning 2.0.0.

Sales, L., Henning, P., Veiga, V., Costa, M. M., Sayão, L. F., da Silva, S., Bonino, L. O., and Pires, L. F. (2020). GO FAIR Brazil: A Challenge for Brazilian Data Science. *Data Intelligence*, 2(1-2):238–245. DOI: 10.1162/dint$_{a0}$0046.

Schultes, E. A., Strawn, G. O., and Mons, B. (2018). Ready, set, go fair: Accelerating convergence to an internet of fair data and services. In *International Conference on Data Analytics and Management in Data Intensive Domains*.

Sharma, S. e. a. (2023). Evaluating tools for data management plans: A comparative study of the dart rubric and the belmont scorecard. pages 26–46.

Simms, S., Jones, S., Mietchen, D., and Miksa, T. (2017). Machine-actionable data management plans (madmps). *Research Ideas and Outcomes*, 3:e13086. DOI: 10.3897/rio.3.e13086.

Wilkinson, M., e. a. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3(1).

Wittenburg, P., Sustkova, H. P., Montesanti, A., Bloemers, S. M., de Waard, S. H., Musen, M. A., Graybeal, J. B., Hettne, K. M., Jacobsen, A., Pergl, R., Hooft, R. W. W., Staiger, C., van Gelder, C. W. G., Knijnenburg, S. L., van Arkel, A. C., Meerman, B., Wilkinson, M. D., Sansone, S.-A., Rocca-Serra, P., McQuilton, P., Gonzalez-Beltran, A. N., Aben, G. J. C., Henning, P., Alencar, S., Ribeiro, C., Silva, C. R. L., Sayao, L., Sales, L., Veiga, V., Lima, J., Dib, S., Xavier, P., Murtinho, R., Tendel, J., Schaap, B. F., Brouwer, P. M., Gavai, A. K., Bouzembrak, Y., Marvin, H. J. P., Mons, A., Kuhn, T., Gambardella, A. A., de Miranda Azevedo, R., Muhonen, V., van der Naald, M., Smit, N. W., Buys, M. J., de Bruin, T. F., Schoots, F., Goodson, H. J. E., Rzepa, H. S., Jeffery, K. G., Shanahan, H. P., Axton, M., Tkachenko, V., Maya, A. D., Meyers, N. K., Conlon, M., Haak, L. L., and Schultes, E. A. (2019). The fair funder pilot programme to make it easy for funders to require and for grantees to produce fair data.