From spreadsheets to interfaces: redesigning clinical variable definition through interactive workflows

Andrea Vázquez-Ingelmo[1,*], Islem Román Nieto-Campo[1], Alicia García-Holgado[1], Francisco José García-Peñalvo[1], Antonio Sánchez-Puente[2] and Pedro L. Sánchez[2,‡]

[1]GRIAL Research Group, University of Salamanca, Paseo de Canalejas 169, 37008, Salamanca, Spain
[2]Cardiology Department, Hospital Universitario de Salamanca, SACyL. IBSAL, Facultad de Medicina, Universidad de Salamanca, and CIBERCV (ISCiii), Salamanca, Spain

Abstract

Spreadsheets remain a common but fragile foundation for clinical data management, often leading to errors and inefficiencies in defining and collecting structured variables. This paper presents a user-centered redesign of the variable definition workflow in a platform for managing structured clinical data and medical images. The proposed solution replaces manual spreadsheet-based schema creation with an interactive web interface that enables users to define, categorize, and reuse variables more effectively. It also introduces automated generation of validated spreadsheet templates based on the platform's internal schema, reducing the likelihood of formatting and semantic errors during data entry. A revised workflow illustrates the improved process, and the system addresses key usability issues previously identified through heuristic evaluations. Remaining limitations, such as continued reliance on offline data entry, are discussed, along with future work directions that include usability validation and AI-assisted variable generation.

Keywords

data management, variable definition, healthcare, interaction workflow, spreadsheets

1. Introduction

In medical research and clinical workflows, spreadsheets remain the predominant tool for defining, managing, and exchanging structured data [1]. Their accessibility and flexibility make them appealing for quick data entry and local management. However, these same characteristics introduce substantial challenges in terms of data quality, traceability, and interoperability.

As shown in [2], spreadsheets used in real healthcare environments frequently contain critical errors—more than 90% of audited files had at least one bottom-line error, with average cell error rates exceeding 13%. Such issues arise from poor structural design, inconsistent usage practices, and the absence of built-in validation mechanisms.

Insights gained from our previous development projects in the health domain—particularly those focused on managing structured clinical data alongside medical images—have revealed the limitations of spreadsheet-driven workflows [3, 4]. Defining variables through spreadsheet templates often required users to navigate a complex, multi-step process: manually encoding metadata across multiple sheets, conforming to strict naming and typing conventions, and aligning data definitions across patient, study, and file levels. These tasks were both time-consuming and highly error-prone, especially for users without technical backgrounds. Common issues such as unreserved or misformatted identifiers, inconsistent variable names, or incorrect date and value formats frequently resulted in failed uploads, requiring tedious revisions and interrupting the research workflow.

Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain
*Corresponding author.
andreavazquez@usal.es (A. Vázquez-Ingelmo); islemr@usal.es (I. R. Nieto-Campo); aliciagh@usal.es (A. García-Holgado); fgarcia@usal.es (F. J. García-Peñalvo); sanchezpu@saludcastillayleon.es (A. Sánchez-Puente); plsanchez@saludcastillayleon.es (P. L. Sánchez)
0000-0002-7284-5593 (A. Vázquez-Ingelmo); 0000-0001-9663-1103 (A. García-Holgado); 0000-0001-9987-5584 (F. J. García-Peñalvo); 0000-0002-6167-2041 (A. Sánchez-Puente); 0000-0002-1402-0526 (P. L. Sánchez)

Despite the use of templates and documentation, spreadsheets lacked interactive guidance, contextual support, and real-time feedback. The burden of ensuring data integrity was effectively placed on end-users—clinicians and researchers whose expertise often lies far from data modeling and validation.

This paper introduces a new system that redefines the variable definition workflow as part of an enhanced data and image management platform. Unlike earlier systems, the new solution places a strong emphasis on user-centric interaction. It provides an interface that guides users step-by-step through the process of creating, validating, and applying structured variables, embedding domain knowledge and validation rules directly into the workflow.

By reducing reliance on manual spreadsheet manipulation, the system aims at improving data quality, enhancing onboarding, and supporting collaborative research more effectively. Grounded in real-world requirements and prior experience, this approach answers to the need for more usable, transparent, and robust tools in clinical data environments.

The remainder of this paper is structured as follows. Section 2 reviews existing platforms for structuring health data and highlights their current limitations. Section 3 presents an analysis of the key challenges associated with current health data management systems, as well as insights drawn from previous development experiences. Section 4 introduces our proposed solution and describes the redesigned workflow. Section 5 discusses the implications of this user-centric approach, and Section 6 concludes the paper with a summary of contributions and directions for future work.

2. Background and related work

Defining structured variables is a critical component of clinical data management, underpinning research validity, interoperability, and downstream analytics. Traditionally, this process has relied heavily on spreadsheet-based templates or rigid electronic report form tools, often demanding significant technical expertise from users. Over the last decades, a wide range of platforms have emerged to facilitate structured data collection in medical settings. While these systems offer varying degrees of flexibility for custom variable definition, most still suffer from usability and integration issues—particularly when it comes to dynamic or user-driven workflows.

One of the most widely adopted tools in clinical research is REDCap (Research Electronic Data Capture) [5], which allows researchers to define custom data fields through a user-friendly web interface or by uploading a data dictionary in spreadsheet form. Its modular design and built-in validation make it a practical solution for many studies. However, REDCap imposes certain limitations once a project enters production mode, restricting changes to the schema without administrator approval. It also lacks advanced user interaction features, such as contextual feedback or real-time support for collaborative form design.

Other platforms such as OpenClinica (https://www.openclinica.com/) and Castor EDC (https://www.castoredc.com/) also offer flexible data modeling and support for variable definition. OpenClinica uses Excel-based templates to define report forms while Castor, a more modern commercial alternative, provides drag-and-drop interfaces and rapid form creation. However, both systems introduce challenges around scalability, integration, or cost, particularly for projects seeking seamless interoperability with imaging data or third-party systems.

In the domain of medical imaging, platforms like XNAT (Extensible Neuroimaging Archive Toolkit) play a central role [6]. XNAT supports the management of imaging data and allows limited extensions to metadata through schema customization. While powerful, it requires considerable IT expertise to deploy and maintain, and its support for general clinical variable definition is minimal compared to dedicated EDC systems.

In response to these limitations, we previously developed CARTIER-IA [3], an integrated platform that enables the management of both medical images and structured data for collaborative research across institutions. The system allowed users to define variables across different levels—patient, study, and file—via structured Excel templates. While the platform introduced mechanisms for automatic validation, pseudonymization, and multi-level data modeling, it still relied on manual workflows and demanded that users conform to strict spreadsheet conventions. This created usability barriers for clinicians and researchers unfamiliar with the technical details of data formatting. Errors such as invalid identifiers, inconsistent variable definitions, or incorrect formats often disrupted uploads and required tedious manual corrections, especially in large-scale, multi-center studies.

In short, while there is no shortage of platforms for clinical data capture, most continue to reflect design assumptions that privilege rigid structures or technical expertise over usability and flexibility.

Building on the limitations observed in prior platforms, this paper introduces a preliminary proposal for a redesigned approach to defining custom clinical variables, framed within a platform that enables their integration and association with imaging data. By embedding domain knowledge and real-time validation directly into the interface, the system seeks to reduce common errors and lower the technical barrier for clinical researchers and healthcare professionals involved in data modeling tasks.

3. Problem analysis

The current variable definition workflow in CARTIER-IA was designed to align with requirements gathered from clinical partners and domain experts involved in its initial deployment. Specifically, these stakeholders expressed a preference for managing data definitions via spreadsheet templates. In practice, platform administrators or technical staff first define a spreadsheet file that specifies the variables to be collected, which is then distributed to researchers and collaborators for data entry.

While this approach reflects familiar workflows and tools used in many research contexts, it introduces several usability and process-related challenges when integrated into a digital platform. These challenges become particularly apparent when analyzing the full upload workflow—outlined in Figure 1—and are reinforced by findings from heuristic evaluations [7].

3.1. Fragmented and error-prone variable definition workflow

As depicted in the BPMN workflow (Figure 1), the variable definition process in CARTIER-IA is heavily dependent on a sequential and spreadsheet-based workflow. Users must first manually define tables at different data levels (patient, study, or file) via the interface, and then prepare an external Excel file that encodes the metadata for each variable (including name, type, range, allowed values, etc.) using a strict multi-column schema. This process demands exact compliance with sheet naming and column conventions, which are not visually guided within the platform itself.

This disconnected, multi-step approach leads to a high probability of failure during the upload phase, especially when users inadvertently introduce inconsistencies—such as misformatted identifiers, missing required fields, or incorrect variable typing. These issues typically result in a rejection of the upload, forcing the user into a trial-and-error loop without in-context remediation or live preview of expected results.

3.2. Lack of interaction, feedback, and task visibility

The heuristic evaluations conducted in [7] highlight several critical usability flaws. Notably, CARTIER-IA lacks sufficient system feedback during key operations. Under Nielsen's heuristic [8] HR1 (Visibility of System Status), experts identified the absence of progress indicators and status updates for long-running actions such as file uploads. HR9 (Help users recognize, diagnose, and recover from errors) received one of the highest severity scores, due to vague or missing error messages when uploads failed or when data inconsistencies were found.

Furthermore, the system provides little to no support for previewing how uploaded variable definitions will appear within the platform before committing. Users are expected to recall and manually reproduce structural details, violating HR6 (Recognition rather than recall). This undermines user confidence and increases reliance on external documentation, which itself was rated insufficient under HR10 (Help and documentation).

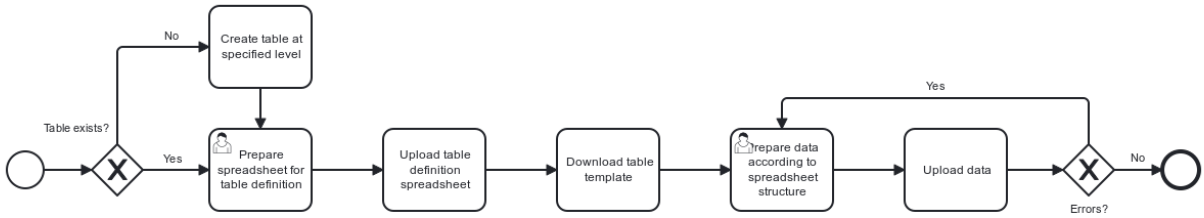

Figure 1: Workflow for variable definition and structured data upload in CARTIER-IA.

3.3. Accessibility and scalability concerns for non-technical users

The current variable definition workflow in CARTIER-IA poses substantial barriers for users without a background in data modeling, clinical informatics, or spreadsheet engineering. Specifically, the process of preparing a valid Excel definition file is not intuitive and requires familiarity with a number of internal conventions that are not self-explanatory.

For each table defined within a project (at the patient, study, or file level), users must create a corresponding spreadsheet sheet whose name matches the exact table name. Within each sheet, the variable metadata must be encoded using a specific column schema that includes headers such as `SV0` through `SV6`. These codes represent different properties of each variable—e.g., `SV0` is the variable name, `SV1` indicates the data type using numeric codes (1 for integer, 2 for real number, etc.), and other columns capture optional attributes like permissible values or measurement units.

According to [3], even experienced clinical data scientists struggled with template complexity and semantic ambiguities in variable representation.

This manual overhead becomes especially problematic in collaborative or longitudinal studies where variable sets are frequently updated. The platform does not provide any in-built support for versioning, collaborative editing, or reusing variable schemas across projects, making it difficult to maintain consistency or traceability over time.

4. Proposed solution

To address the limitations identified in the previous section, we introduce a redesigned workflow and interface for defining structured variables within CARTIER-IA. This solution adopts a user-centric approach aimed at reducing cognitive load, preventing errors, and promoting task visibility. It consists of two core innovations: an integrated visual editor for variable definition, and the automated generation of validated data entry templates.

4.1. A visual interface for variable definition

The new interface replaces the external spreadsheet encoding of metadata with an interactive web-based variable editor. As shown in Figure 2, users can add, modify, or delete variables directly within the platform. The interface supports three data types: binary, categorical, and quantitative, each associated with intuitive visual cues and custom input fields. This design reduces reliance on external documentation and eliminates the need for users to memorize column codes or typing conventions.

Each variable includes optional descriptive metadata such as units, allowed values, and range limits. These parameters are no longer entered manually in spreadsheet cells but are configured via dedicated input components. The variable list is persistently visible on the left panel, providing immediate feedback on the project schema and enabling quick revisions.

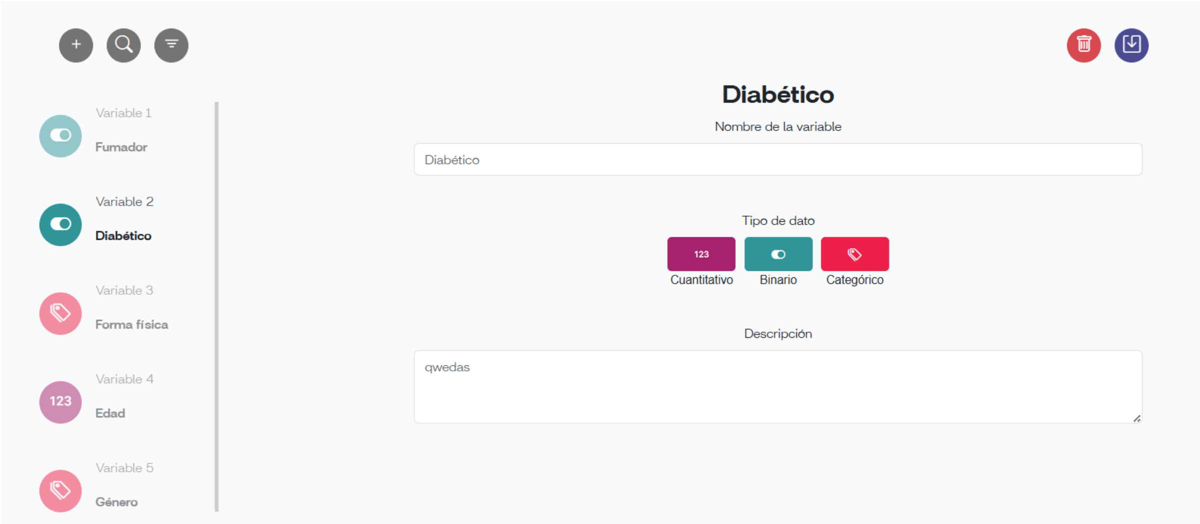

Figure 2: Initial version of the interface for variable creation and editing in the new workflow. Contents in Spanish.

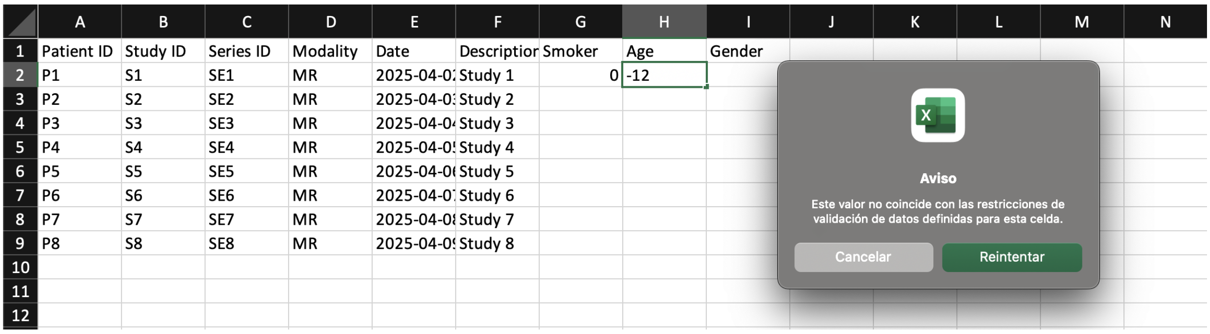

Figure 3: Error message (in Spanish) provoked by introducing a negative age in the spreadsheet template.

4.2. Automated generation of validated templates

Once variables have been defined via the web interface, the system can automatically generate a data entry template with embedded validation rules. This functionality is implemented in Python using the `openpyxl` library, and ensures that each column reflects the semantic constraints defined by the user, as shown in Figure 3 (e.g., binary fields are restricted to `0` or `1`, quantitative fields to user-defined numeric ranges, and categorical fields to predefined values).

This template generation step bridges the gap between schema design and data collection, reducing the likelihood of errors during upload. Unlike previous templates that had to be manually constructed by users, the new version is programmatically derived from the platform's internal schema, guaranteeing consistency and compliance.

4.3. Updated workflow overview

The restructured process is illustrated in Figure 4. Compared to the earlier version (Figure 1), the new workflow eliminates the need for users to manually encode metadata or align multiple sheets. Instead, table creation and variable definition are seamlessly integrated into the platform, followed by a guided download of the appropriate data template.

The preparation of data is still performed locally using spreadsheet software, but the burden of correct structure and validation is now offloaded to the template logic. Upon uploading the completed file, the system validates the data against the pre-defined schema and provides immediate, actionable feedback in case of errors.

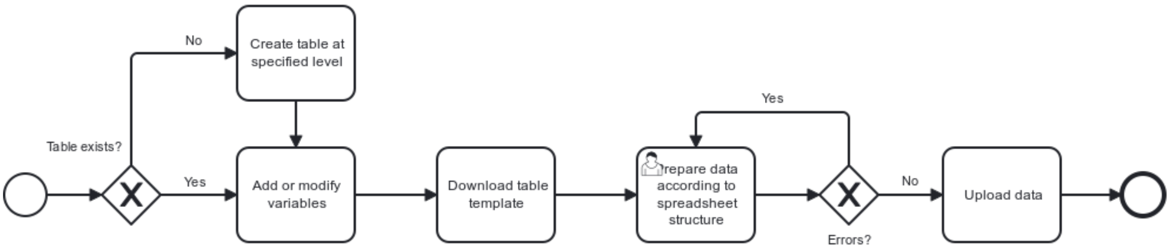

Figure 4: Updated BPMN workflow for structured data upload using the new variable editor and validated templates.

5. Discussion

The proposed redesign directly tackles several critical usability and workflow issues identified through the heuristic evaluation of the original CARTIER-IA platform. By embedding variable definition into a guided, web-based interface and automating the generation of validated templates, the system shifts complexity away from the end-user and embeds it into the platform's logic—where it can be controlled, audited, and evolved.

This approach enhances compliance with multiple Nielsen heuristics [8]:

- **HR1 (Visibility of system status):** Users receive immediate visual feedback during variable creation and can download templates with confidence that they match the internal schema.
- **HR5 (Error prevention):** Validations are embedded both in the interface and the generated Excel files, reducing the opportunity for input mistakes or misinterpretation of requirements.
- **HR6 (Recognition rather than recall):** The interface minimizes the need to remember codes or structural rules by surfacing configuration options contextually.
- **HR9 (Help users recognize, diagnose, and recover from errors):** Template errors are prevented by construction, and future versions could embed in-line feedback during data upload.
- **HR10 (Help and documentation):** Visual affordances and the structure of the interface itself reduce the need for external documentation, simplifying onboarding.

The benefits extend beyond individual usability principles. By aligning variable modeling with intuitive user workflows, the redesign reduces onboarding time, minimizes support burden, and facilitates collaboration between technical and clinical stakeholders. Moreover, the clear delineation of responsibilities—where domain experts define variables and the system enforces structure—represents a more sustainable and scalable model for managing structured data in medical research.

In terms of extensibility, this architecture opens new opportunities. For example, variables could be annotated with semantic tags (e.g., SNOMED [9], LOINC [10], etc.) to support interoperability. The system could incorporate versioning to track changes in variable definitions over time, or offer a central repository of reusable variable sets to support multi-project harmonization.

Importantly, the shift to platform-guided modeling also lays the foundation for more advanced features, such as AI-assisted variable creation. By analyzing previously defined schemas or dataset patterns, the system could recommend candidate variables, detect redundancy, or suggest optimal encoding strategies for downstream analytics.

Another of the major benefits of the new system is the support for reusability of variable definitions across projects. Variables defined in one context can be cloned or imported into another with minimal effort, significantly reducing redundancy and accelerating the setup of new studies. This is particularly advantageous in multi-center or longitudinal research environments, where consistency of definitions is crucial.

Variables can also be grouped into categories or semantic domains (e.g., demographics, comorbidities, imaging features), allowing users to organize large schemas more effectively.

Nonetheless, limitations remain. While the redesign simplifies definition and validation, it still relies on spreadsheet-based data entry for the actual records. This dependency continues to impose certain risks and constraints: users must still manipulate spreadsheet files manually, which can lead to accidental modifications of headers, incorrect encodings, or loss of validation rules if files are edited with incompatible tools.

Moreover, the spreadsheet remains disconnected from the platform's internal logic during the data entry phase, offering no real-time guidance, progressive data saving, or collaborative editing features. As a result, although the template ensures structural correctness at the point of generation, the responsibility for maintaining data integrity during entry still lies with the user.

Overall, this redesign reflects a transition from tool-centric to user-centric thinking, reimagining structured data modeling as a cooperative, intelligible, and iterative process. It sets a precedent for the evolution of CARTIER-IA and similar platforms toward more inclusive, robust, and intelligent research infrastructures.

6. Conclusions and future work

This paper has presented a user-centered redesign of the variable definition workflow in CARTIER-IA, a platform that integrates structured data and medical imaging management for clinical research. The proposed solution replaces the error-prone, spreadsheet-driven schema definition with an interactive, web-based interface that enables users to define, categorize, and reuse variables with greater ease and reliability. In parallel, a backend mechanism automatically generates validated data entry templates, minimizing formatting errors and enforcing semantic constraints.

The updated workflow and interface directly address key usability issues identified in previous evaluations, particularly those related to error prevention, recognition, feedback, and documentation. By embedding domain knowledge and constraints into both the UI and the template generation process, the system reduces the cognitive load on users and improves accessibility for clinicians and non-technical collaborators.

Future work will follow two main avenues. On the one hand, we plan to carry out a new round of usability studies—combining expert heuristics and user-centered evaluations—to quantitatively measure improvements in efficiency, accuracy, and user satisfaction over the original workflow. The results will guide further enhancements to the interface and feedback mechanisms.

On the other hand, we will explore the integration of AI-assisted tools for variable definition, capable of recommending names, types, and constraints based on existing schemas or textual study descriptions. These features aim to streamline early project setup and promote reuse of established patterns across studies.

Acknowledgments

This research was partially funded by the Spanish Ministry of Science and Innovation through the AVisSA project grant number (PID2020-118345RB-I00). This work was also supported by competitive community grants (GRS 2033/A/19, GRS 2030/A/19, GRS 2031/A/19, GRS 2032/A/19) from the SACYL, Junta Castilla y León; by competitive national grants (PI14/00695, PIE14/00066, PI17/00145, DTS19/00098, PI19/00658, PI19/00656, PI21/00369) from the Institute of Health Carlos III, Span-ish Ministry of Science and Innovation and co-funded by ERDF/ESF, "Investing in your future" and; by the CIBERCV (CB16/11/00374) from the Institute of Health Carlos III, Spanish Ministry of Science and Innovation.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT, Grammarly in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] S. P. Iyengar, H. Acharya, M. Kadam, Big data analytics in healthcare using spreadsheets, in: Big Data Analytics in Healthcare, Springer, 2019, pp. 155–187.

[2] E. Dobell, S. Herold, J. Buckley, Spreadsheet error types and their prevalence in a healthcare context, Journal of Organizational and End User Computing 30 (2018) 20–42.

[3] F. J. García-Peñalvo, A. Vázquez-Ingelmo, A. García-Holgado, J. Sampedro-Gómez, A. Sánchez-Puente, V. Vicente-Palacios, P. I. Dorado-Díaz, P. L. Sánchez-Fernández, et al., Application of artificial intelligence algorithms within the medical context for non-specialized users: The cartier-ia platform, International Journal of Interactive Multimedia and Artificial Intelligence 6 (2021) 46–53.

[4] F. J. García-Peñalvo, A. Vázquez-Ingelmo, A. García-Holgado, J. Sampedro-Gómez, A. Sánchez-Puente, V. Vicente-Palacios, P. I. Dorado-Díaz, P. L. Sánchez, Koopaml: A graphical platform for building machine learning pipelines adapted to health professionals, International Journal of Interactive Multimedia and Artificial Intelligence (2024). In press.

[5] P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, J. G. Conde, Research electronic data capture (redcap)—a metadata-driven methodology and workflow process for providing translational research informatics support, Journal of Biomedical Informatics 42 (2009) 377–381.

[6] R. Herrick, W. Horton, T. Olsen, M. McKay, K. A. Archie, D. S. Marcus, Xnat central: Open sourcing imaging research data, NeuroImage 124 (2016) 1093–1096.

[7] A. Vázquez-Ingelmo, J. Alonso, A. García-Holgado, F. J. García-Peñalvo, J. Sampedro-Gómez, A. Sánchez-Puente, V. Vicente-Palacios, P. I. Dorado-Díaz, P. L. Sánchez, Usability study of cartier-ia: A platform for medical data and imaging management, in: P. Zaphiris, A. Ioannou (Eds.), Learning and Collaboration Technologies: New Challenges and Learning Experiences, Springer International Publishing, Cham, 2021, pp. 374–384.

[8] J. Nielsen, Finding usability problems through heuristic evaluation, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1992, pp. 373–380.

[9] K. A. Spackman, K. E. Campbell, R. A. Côté, Snomed rt: A reference terminology for health care, in: Proceedings of the AMIA Annual Fall Symposium, 1997, p. 640.

[10] C. J. McDonald, S. M. Huff, J. G. Suico, G. Hill, D. Leavelle, R. Aller, A. Forrey, K. Mercer, G. DeMoor, J. Hook, et al., Loinc, a universal standard for identifying laboratory observations: A 5-year update, Clinical Chemistry 49 (2003) 624–633.