A methodological approach for data standardization and management of Open Data portals for scientific research groups: a case study on mobile and ubiquitous ecosystems

Vladimir Villarreal [1]*, Lilia Muñoz [1]*, Joseph González [1], Jesús Fontecha [2], Cosmin C. Dobrescu [2], Mel Nielsen [1], Dimas Concepción [1] and Marco Rodriguez [1]

[1]Grupo de Investigación en Tecnologías Computacionales Emergentes (GITCE), Universidad Tecnológica de Panamá, David, Chiriqui, Panama, [2]Modelling Ambient Intelligence Research Lab (MAmI Research Lab), Universidad de Castilla - La Mancha, Ciudad Real, Castilla - La Mancha, Spain

Introduction: In the scientific research domain, the Open Science movement stands as a fundamental pillar for advancing knowledge and innovation globally. This article presents the design and implementation of the GITCE Open Data Ecosystem (GITCE-ODE) Research Data Management System (RDMS), developed by the Research Group on Emerging Computational Technologies (GITCE) at the Technological University of Panama, as a platform for the long-term storage, publication, and dissemination of research products.

Methods: The architecture of the GITCE-ODE RDMS encompasses the entire data engineering lifecycle, facilitating information processing stages such as extraction, transformation, loading (ETL), as well as the management and analysis of diverse datasets and metadata.

Results: Compliance with the FAIR principles ensures that published data and products are Findable, Accessible, Interoperable, and Reusable, promoting automation in the discovery and reuse of digital resources. Key considerations of the web portal include file format standardization, data categorization, treatment of semantic context, and organization of resources to ensure efficient management and administration of open research data.

Discussion: Through this platform, GITCE aims to foster collaboration, transparency, and accessibility in scientific research, contributing to the ongoing advancement of knowledge transfer and innovation.

1 Introduction

In today's information age, the Open Science movement stands as a fundamental pillar for the progress of knowledge and innovation worldwide (Prieto, 2022). Committed to advocating transparency and open access to scientific knowledge and research data, Open Science main objective is to ensure that the results of research projects, scientific products, and associated data are available to all stakeholders. This is achieved by promoting their free access, reuse, and redistribution, provided that proper attribution to the origin

and authorship is maintained, and shared in a similar manner (Tzitzikas et al., 2021). Open Science not only proposes a transparent work philosophy but also seeks to foster an environment where collaboration and accessibility to scientific information are key elements for the continuous advancement of scientific research.

In this regard, academic research groups serve as catalysts for promoting Open Science initiatives, functioning as organizations dedicated to executing a diverse range of research projects across various domains of knowledge, fostering collaboration among researchers spanning the natural and social sciences, technology, engineering, and the humanities (Christensen et al., 2021). By functioning as organizational structures that facilitate collaboration among academia, government institutions, and private companies, these groups enhance research endeavors through the diversity of perspectives, resources implementation, and knowledge exchange. This collaborative environment drives the generation of robust and innovative research and development processes, thereby making significant impacts on society (Castrillón-Muñoz et al., 2020).

Academic research groups serve as the primary drivers of productivity in scientific endeavors, enhancing efficiency in the training of new researchers and formalizing organizational structures within academic institutions (Vabø et al., 2016). These groups cultivate a professional, social, and cultural working environment that accommodates a diverse array of members and experts, ranging from full-time professors and researchers to undergraduate and postgraduate students, as well as doctoral and postdoctoral candidates. They represent a strategic cornerstone in the contemporary research landscape, providing legitimacy to research activities and facilitating the acquisition of funding and access to essential resources necessary for conducting high-caliber research (Kyvik and Reymert, 2017).

However, to fully align the efforts of academic research groups with the philosophy and objectives of the Open Science movement, it is imperative to establish clear criteria for the management and supervision of research products and data (Manco, 2023). This includes addressing aspects such as data collection, annotation, publication, and long-term data preservation. Such measures not only facilitate knowledge sharing but also enhance efficiency in data management by minimizing dissemination and duplication of resources. Moreover, they promote the monitoring of research initiatives and foster scientific progress and innovation across various sectors, benefiting both the public and private domains (Tzitzikas et al., 2021). This approach is particularly crucial in the context of collaborative research, which is characterized by the integration of diverse perspectives and the participation of researchers from different disciplines. In this regard, effective data management emerges as a critical factor for the success and synergy of research projects within academic research groups (Finkel et al., 2020).

The increasing emphasis on Open Science, coupled with the imperative to preserve research knowledge over time, and the inherent complexities of collaborative scientific research, have led to the development of information platforms known as Research Data Management Systems (RDMS). These systems are specifically designed to streamline the collection, organization, preservation, and publication of scientific products (Nie et al., 2021). The

following is a summary of the reports in the scientific literature on RDMS implementations.

First, Piedra and Suárez (2018) introduces SmartLand-LD, a Semantic Web-based research data management framework. This framework encompasses the extraction, transformation, integration, and exploration of large and heterogeneous data, applied to decision-making for smart and sustainable development in territories of high biodiversity. SmartLand-LD integrates various data sources, such as sensor transmissions, geographic data, and scientific databases, using collaborative techniques for Open Science, knowledge graph construction, and semantic interoperability to integrate sustainable development indicators.

Similarly, Finkel et al. (2020) proposes a flexible framework for research data management, focusing on organizational measures, data management concepts, and technical solutions. The study highlights the importance of ensuring accessibility and fluidity in the exchange of data and metadata among researchers, as well as the effective publication of datasets and research products. The framework is based on data-type-specific metadata and a hierarchy using a common taxonomy, covering the workflow from data and metadata generation and preparation to long-term storage.

In the same vein, Mozgova et al. (2020, 2022) describe a framework for the development of RDMS platforms composed of a Research Data Management System and a Knowledge Management System, linked by a shared domain-specific vocabulary. This approach facilitates the contextualized storage of research data, documentation, and results in both systems, following the principles of FAIR (Findable, Accessible, Interoperable, and Reusable) data standardization across projects and simplifying workflows for data exchange throughout the research.

On the other hand, Nie et al. (2021) reports on the implementation process of the Peking University Open Research Data Repository (PKU-ORDR) project, applying the development of an Open Data RDMS platform at the university level. In this study, the various stages of the system implementation process were analyzed, from the project kick-off, requirements definition, software configuration, data cleaning, and training of the personnel involved. The authors also addressed key issues for the implementation of RDMS, such as the response to institutional policies, funding requirements and licensing, collaboration between administrative units and libraries, and concerns of researchers and data users. The above studies establish a robust foundation for examining how RDMS can be applied for the publication of open research data, promote scientific collaboration, and strengthen methods of preserving knowledge from university research groups.

This article outlines the design of a Research Data Management System (RDMS) developed following a framework aimed at standardizing and opening research data generated from scientific projects conducted by the Research Group on Emerging Computational Technologies (GITCE) at the regional center of the province of Chiriqui, within the Technological University of Panama (UTP) in Panama (GITCE, 2024). Since its inception, GITCE has dedicated its efforts to developing solutions in applied research areas related to significant societal challenges, including education, health, industry, and digital transformation. This has been made possible through active collaboration among

students, faculty members, researchers, professionals from private companies, and government institutions. However, the evolving composition of the group and the diversification of projects over time present challenges in effectively tracking the outcomes of previous research endeavors. This underscores the importance of implementing robust information and documentation management strategies to preserve and leverage the accumulated knowledge over the years.

This study offers a comprehensive overview of the design of the GITCE Open Data Ecosystem (GITCE-ODE) web portal. It adopts a rigorous systems design methodology, which involves the definition of research data sources, establishment of data integration strategies, and implementing mechanisms to ensure data and metadata quality. Additionally, it establishes a structured framework for acquiring contextual information related to research projects and their semantic context, along with defining metadata standardization protocols in accordance with the FAIR principles for Open Data initiatives. Furthermore, the study showcases a practical application of the GITCE-ODE web portal through the publication of research products derived from the development of a remote monitoring system for precision livestock farming within pastured poultry, as part of a broader research initiative focused on mobile and ubiquitous solutions ecosystems. The primary aim of this project is to create a platform that empowers the academic research group members to harness the information, knowledge, and resources accumulated over time. Moreover, the platform seeks to advance Open Science initiatives by facilitating scientific collaboration among researchers and stakeholders on both national and international contexts.

2 Methodology

2.1 GITCE-ODE research data management system

The GITCE-ODE RDMS project focuses on the development of an Open Data platform for the long term storage and publication of products from research initiatives on the application of emerging technologies in key development areas. It includes the integration of projects focused on society priority areas, sustainable development goals, development plans, and public agendas, to create an environment that contributes to science and innovation through technological solutions and research. As shown in Figure 1, GITCE-ODE collects datasets, models, documents, code, and research products from sources such as:

- Catalog of research projects developed: constitutes the systematic compilation of the research products and results obtained from research projects carried out by GITCE from the year 2011 to the year 2024. Among some of these research initiatives, we can mention: AmIHEALTH, a web platform for the monitoring and control of patients with arterial hypertension problems in Panama (Villarreal et al., 2018), BEEBOT Project, an initiative based on the development of an interactive environment for the teaching of mathematics through the use of small robots (Muñoz et al., 2020), and Epidempredict for COVID19, an intelligent platform based on

Artificial Intelligence (AI) tools for the management of large volumes of data epidemiological studies (Munoz et al., 2021).
- Ecosystem of mobile and ubiquitous solutions: includes the integration of heterogeneous databases from information systems for data management in priority environments of Panama, focusing on the design of digital solutions based on mobile computing technologies for industry, health, education, and healthcare environments (Villarreal et al., 2023). This group of solutions includes ConnectedCoops, a distributed system for remote monitoring of poultry production environments (Gonzalez et al., 2023), MomApp, a platform for continuous monitoring of hypertensive disorders in pregnancy (Nielsen et al., 2023), AutismAR Discovery, an augmented reality application for teaching children with autism spectrum disorder (Patiño et al., 2023), and DOPDIVI, an obstacle-sensing device designed to enhance the mobility and independence of individuals with visual impairments (Rodríguez et al., 2023).
- Open Data and external collaboration portal: encompasses the technological infrastructure, tools, protocols, and standards, as well as the design of the user interface of the Open Data web portal to facilitate the publication of results on an ongoing basis by collaborators and researchers affiliated with GITCE.

2.2 Open data standardization considerations

In the context of the digital era and the growing importance of transparency and accessibility of information, Open Data portals have emerged as key tools to foster collaboration, innovation and accountability in various knowledge domains (Stagars, 2016). The Open Data movement advocates for the unrestricted access, use, and redistribution of data, aiming to remove barriers to information and promote transparency and collaboration. It emphasizes making data freely available to all stakeholders, and is crucial to Open Science as it aligns with the principles of transparency and openness, enabling researchers to validate findings, reproduce experiments, and build upon existing knowledge more efficiently (Burgelman et al., 2019). The increasing importance of Open Data has led to the development of various Open Data platforms such as CKAN, Socrata, OpenDataSoft, Kaggle, and others. These platforms offer a diverse range of features, technologies, and functionalities designed to streamline the management, publication, and sharing of digital resources (Ali et al., 2022).

As part of the design considerations for developing the GITCE-ODE RDMS, the implementation of the Comprehensive Knowledge Archive Network (CKAN) was chosen as the foundational technological platform for the Open Data web portal. CKAN is an open-source platform renowned for facilitating the management, publishing, access, and interoperability of digital resources and datasets. It stands out for providing and intuitive web interface, robust dataset metadata definition capabilities, support for multiple data formats, and seamless integration with external systems. Furthermore, CKAN provides features such as access control mechanisms, license management tools, built-in

FIGURE 1
Information sources for the research data management system.
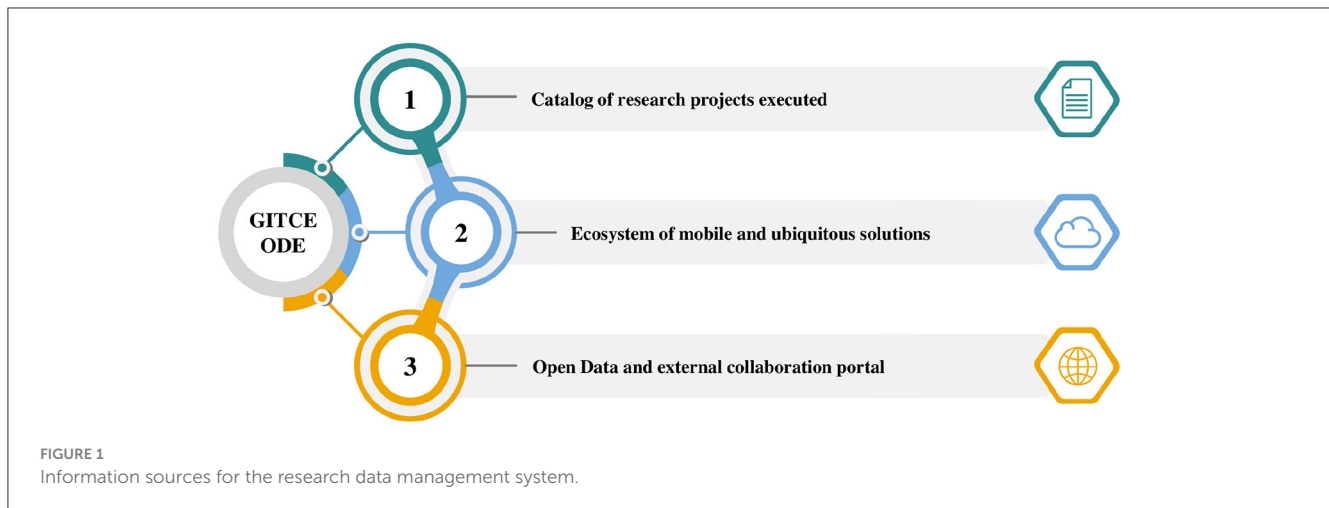
data visualization capabilities, versioning functionalities, and the flexibility to install custom extensions, making it a comprehensive solution for managing research data in an effective manner (Open Knowledge Foundation, 2024a).

The configuration of the GITCE-ODE RDMS is based on the FAIR principles for the effective management and administration of open research data. These principles promote that scientific data and products are Findable, Accessible, Interoperable and Reusable (FAIR), to facilitate automation in the discovery of digital resources, as well as the reuse and exploitation of scientific information (Wilkinson et al., 2016). Each of the FAIR principles integrates a set of implementation guidelines and recommendations for their use without strict dependence on specific technologies or standards. The FAIR Implementation Profile (FIP) provides a questionnaire to facilitate the definition of pathways to compliance for each of these guidelines, promoting flexibility in the design of systems and technological infrastructure according to the requirements of each project (Schultes et al., 2020). In Table 1 the FIP for the GITCE-ODE RDMS is presented in response to the considerations needed to satisfy the guidelines of each FAIR principle. The concept of each FAIR principle is described below, as well as the tools, standards and strategies chosen:

- (F)indable: scientific data and research products should be easy to find for both humans and machines. This involves assigning persistent unique identifiers, as well as rich, clear metadata that makes it easy to index and discover.

  - Archival resource key (ARK): system used to assign unique persistent identifiers to digital assets, such as datasets and metadata records. ARK provides resolvable links that allow digital resources to be reliably accessed and cited over time (Alliance, 2024).
  - Dublin Core (DC): metadata standard for the description of digital assets. It provides a set of terms, each with a simple, generic meaning, to describe resources such as documents, web pages, images, videos, and datasets. Each term include basic descriptors such as title, creator, subject, description, publisher, date, type, format, identifier, source, language, and licenses (DCMI, 2024).
  - CKAN Built-in search: CKAN has a built-in resource search module based on Apache Solr for locating and retrieving information by automatically indexing metadata such as dataset names, descriptions, and tags.

- (A)ccessible: scientific data and research products must be available for access and retrieval under specific conditions of use. This includes establishing appropriate access protocols, as well as authentication methods.

  - CKAN role-based authorization: as a web-based platform, CKAN leverages the Hypertext Transfer Protocol (HTTP) to facilitate access to metadata and datasets through standard web browsers and permissions granted to users through its built-in role authorization module.
  - CKAN REST API: CKAN also enables the generation of authentication keys for integration with external web systems, as well as code-based interaction methods through its Application Programming Interface (API). This API follows the principles of the Representational State Transfer (REST) web architecture and leverages standard HTTP methods (GET, POST, PUT, DELETE) for accessing, creating, updating, and deleting datasets, resources, and metadata.

- (I)nteroperable: scientific data and research products can be used and combined with other scientific data and products. This is achieved using common data standards, knowledge representation languages, and common vocabularies.

  - RDF (Resource Description Framework): is a standard for representing information about digital resources on the World Wide Web. RDF uses a triplet structure, consisting of subject-predicate-object statements, to describe resources and the connection between them. It provides a framework for expressing relationships between resources, allowing

TABLE 1   FAIR Implementation Profile (FIP) for the GITCE-ODE RDMS.

FAIR guide	Question	Resource type	Tool
F1	What globally resolvable, persistent, and unique identifiers do you use for metadata records and datasets?	Identifier Type	ARK
F2	What metadata schemas are used for searchability?	Metadata schema	DublinCore
F3	What is the technology that links the persistent identifiers of your data to the description of the metadata?	Metadata-data linking mechanism	CKAN Metadata Schema (Persistent Identifier Field)
F4	In which search engines are metadata records and datasets indexed?	Search engines	Solr (CKAN Integrated Search)
A1.1	What standardized communication protocol do you use for metadata records and datasets?	Communication protocol	HTTP (CKAN REST API)
A1.2	What authentication and authorization technique do you use for metadata records and datasets?	Authentication and Authorization Technique	CKAN's Built-in Role-Based Authorization Module
A2	What metadata longevity plan do you use?	Metadata Longevity	CKAN Metadata Schema (Data Unavailability Notice field)
I1	What knowledge representation languages (which enable machine interoperation) do you use for metadata records and datasets?	Knowledge representation language	RDF
I2	What structured vocabularies do you use to annotate your metadata records and datasets?	Structured vocabularies	DCAT
I3	What models and schemas do you use for metadata records and datasets?	Metadata Schema	CKAN Metadata Schema (Dataset Cross-reference field)
R1.1	What usage license do you use for metadata records?	Data usage license	Open Data Commons licenses
R1.2	What metadata schemas do you use to describe the provenance of metadata records and datasets?	Provenance model	PROV-O
R1.3	What metadata schemas do you use to describe minimum information in datasets?	Minimum Information Standard	DublinCore

for better semantic understanding of datasets and their metadata (W3C, 2024b).

 – DCAT (Data Catalog Vocabulary): provides a common vocabulary for describing datasets and resources in data catalogs, facilitating the discovery and sharing of datasets between different systems and platforms. DCAT implements RDF to represent metadata about datasets, including their distribution, structure, and relationships (W3C, 2024a).

• (R)eusable: scientific data and research products should be reusable for different purposes and by different people. This involves providing clear and detailed metadata about the semantic context, collection methods, and terms of use.

 – Open Data Commons (ODC): set of legal tools designed to provide a framework for sharing, publishing, and using Open Data. These licenses are designed to address the unique considerations associated with Open Data publication, ensuring that datasets can be used while respecting authors' rights (Open Knowledge Foundation, 2024b).
 – PROV-O (Ontology of provenance): is a specification that defines a set of terms and relationships to model the provenance of resources, including entities, activities, and

agents involved in the production or influence of a piece of data. It can be used for both metadata records and datasets (W3C, 2013).

Likewise, the definition of strategies for compliance with the FAIR principles as part of the development of the GITCE-ODE RDMS also involved the consideration of key factors for the management of the flow of information within the Open Data web platform, like:

• File formats: refers to the identification of the different types of research products to be shared, as well as their adaptation to formats or data structures that respond to Open Data standards, seeking to facilitate the interoperability, management and decoupling from proprietary formats subject to licensing. In this way, the data becomes actionable regardless of the data processing medium (Rudmark, 2020). Examples of Open Data formats include, for structured data, the use of CSV (Comma-Separated Values) to store tabular data and PDF (Portable Document Format) to represent structured information. For unstructured data, JSON (JavaScript Object Notation) is common due to its flexibility for transmitting data between systems, while the XML (eXtensible Markup Language) format is used with hierarchical data (Herrera-Cubides et al., 2023).

- Data categorization: this aspect focuses on the application of mechanisms that allow the classification of research products and datasets into categories and indicators that support the planning of long-term technological development strategies,

aligned with high-level initiatives from the governmental, industrial, societal and academic fields (Piedra and Suárez, 2018). In this project, a resource labeling strategy is employed through the implementation of a vocabulary based on the 17
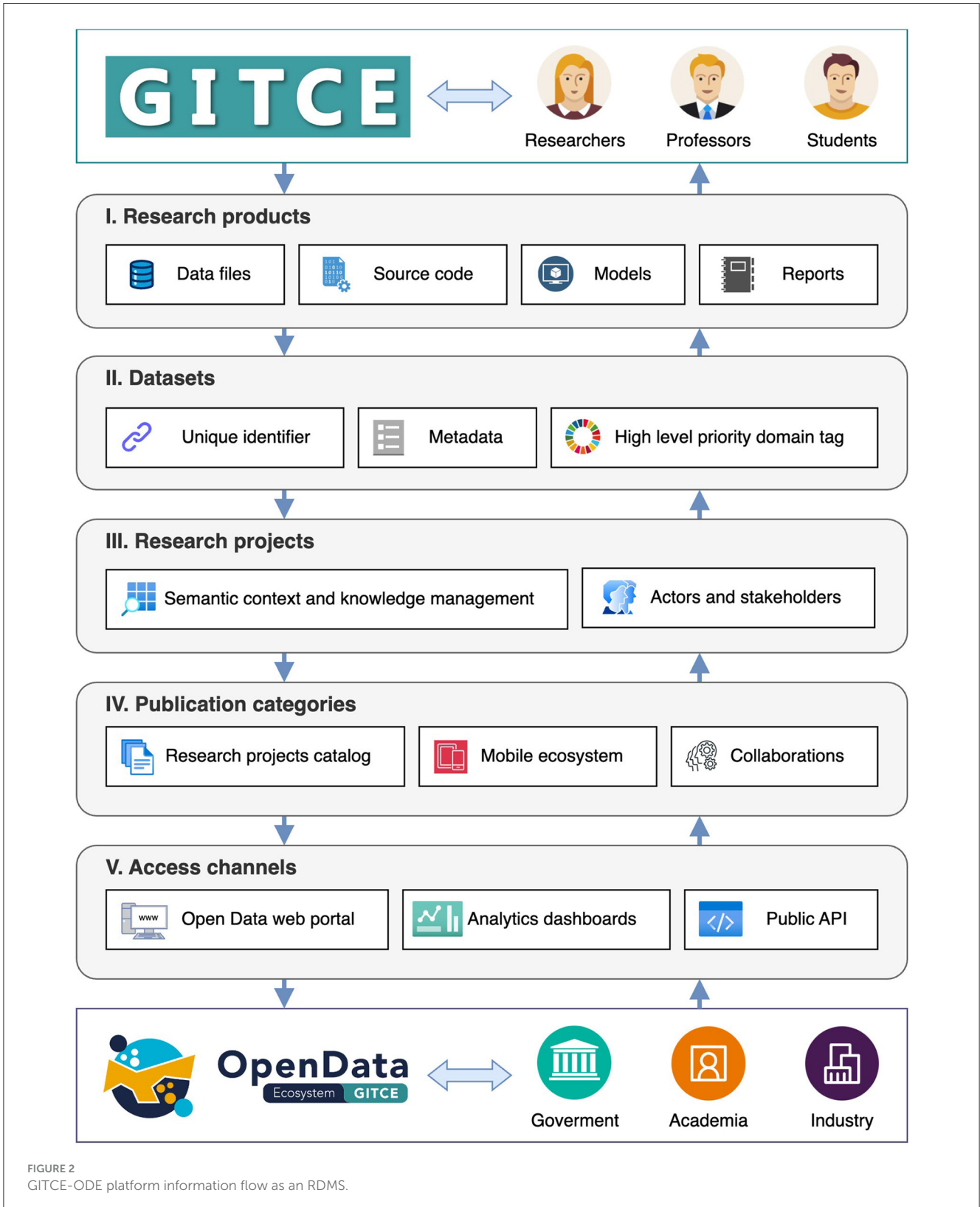

FIGURE 2
GITCE-ODE platform information flow as an RDMS.

Sustainable Development Goals (SDGs) of the United Nations (UN) 2030 agenda, adopted by the National Government of Panama in 2015, as a guiding framework for the planning and execution of social development actions in priority areas of the country. This approach seeks to facilitate the interoperability of resources, the reuse of research products and the long-term monitoring of the implementation of technological development initiatives, allowing the observation of priority areas that require attention and the evaluation of sustainable development strategies over time (Naciones Unidas Panamá, 2024).

• Treatment of the semantic context: this refers to the knowledge management process and involves exhaustively documenting the various factors related to the project in order to fully understand the magnitude and impact of the research carried out. Elements such as the context of development of

the project must be included, detailing the time frame and location in which it was carried out. The start and end date, the specific place of execution and the clear objectives of the research are crucial aspects. In addition, the initiative that motivated the project must be addressed, describing the need or problem that was sought to be addressed. Also highlighting the actors involved, such as institutions, collaborators, and sponsors, to provide a complete overview of collaboration dynamics. Details about the researchers in charge should be recorded, documenting their roles and contributions to the project (Mozgova et al., 2020).

• Organization of resources: this involves establishing a clear and coherent structure that facilitates the organization and classification of data to ensure accessibility and efficiency in the search for information. This includes creating a hierarchy that reflects the thematic structure of projects, defining
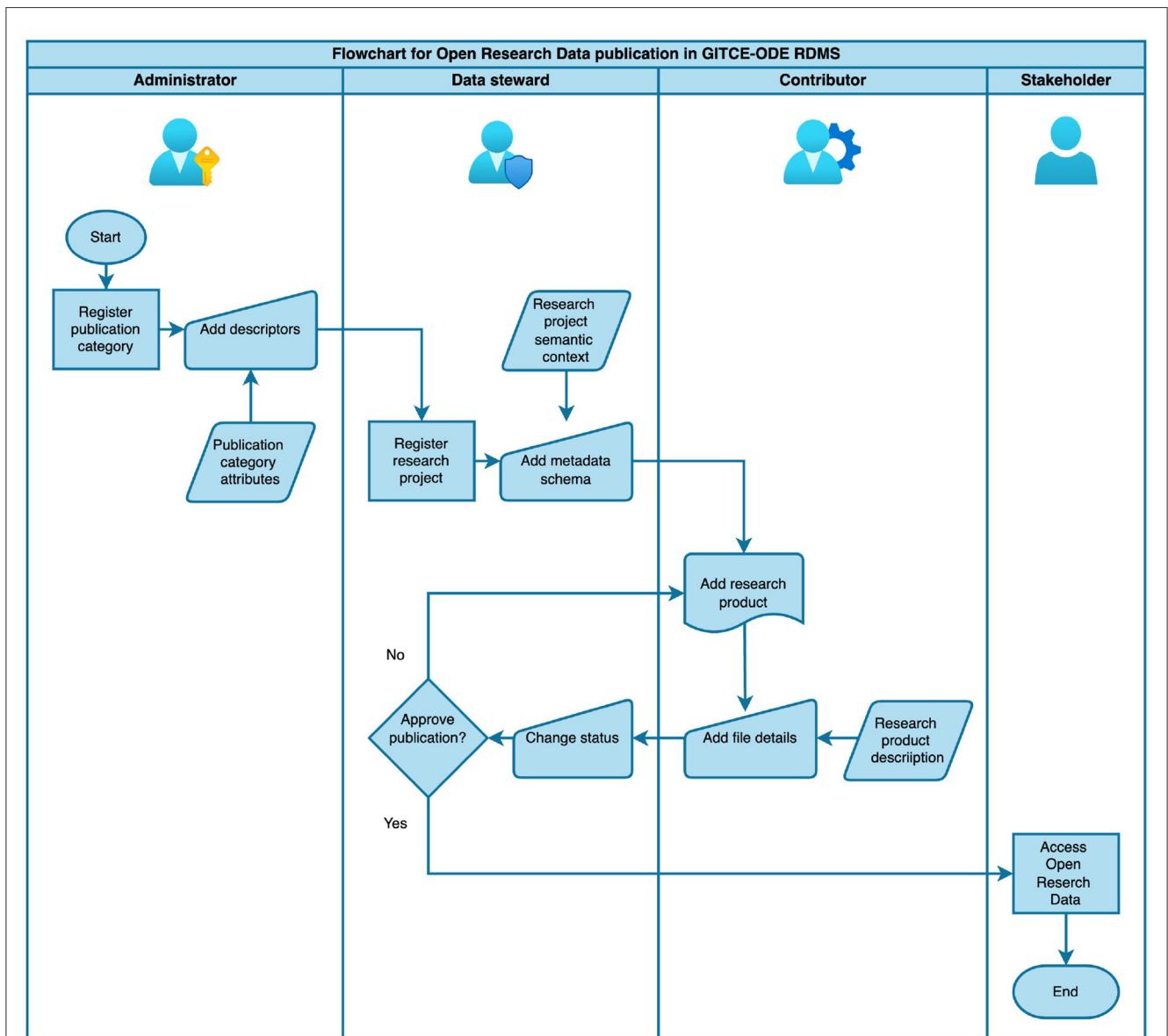

FIGURE 3
Roles and steps for Open Data publication in the GITCE-ODE RDMS.

categories, and logical groupings of resources. In Figure 2 the hierarchy designed for the classification of the elements participating in the information flow of the GITCE-ODE RDMS is presented.

2.3 Organizational planning

The operation of the GITCE-ODE Open Data web portal as an RDMS involves the collaboration of various roles specialized in information systems management (Finkel et al., 2020). These roles play vital functions in the operation and continuous maintenance of the system, fulfilling a set of predefined responsibilities which include:

- The Administrator oversees the overall functioning of the web portal, manages user accounts, configures system settings, and ensures security and performance standards. Its responsibilities also include organizing publication categories, tags, and data catalogs, providing technical support, and compliance with data governance policies.
- The Data Steward is responsible for managing the data lifecycle within the platform. This includes registering research projects, validating semantic context, and approving resource publication. They work with contributors to maintain metadata quality, comply with standards, protocols, and data management practices.
- The Contributor, as primary data producer, plays a fundamental role in populating the web portal with datasets and other scientific resources. They create, upload, and describe research products, ensuring the relevance, integrity, and quality of shared resources, as well as accuracy while defining their corresponding metadata.
- The Stakeholders represent end-users and beneficiaries of the GITCE-ODE Open Data web portal. They access and utilize published datasets and resources for various purposes such as research, analysis, and decision-making.

The flowchart presented in Figure 3 illustrates the process of publishing Open Data in GITCE-ODE. This diagram describes the series of steps necessary to register, prepare, and manage research projects and resources for publication, grouped by the role responsible. The process entails the registration of descriptive data for each element within the resource classification hierarchy, comprising publication categories (CKAN Organizations), research projects (CKAN Datasets), and research products (CKAN Resources). To achieve this, a specific set of attributes and descriptors outlined in Table 2 were defined for each element of the hierarchy. The establishment of these attributes and descriptors adhered to the recommended model presented in Karimova et al. (2019) for documenting research data repositories using the CKAN metadata scheme in conjunction with Dublin Core standard terms. The resulting GITCE-ODE RDMS metadata schema also includes the use of a custom catalog for SDG tagging and supporting custom fields for extended resource description when necessary.

TABLE 2 Metadata schema for the GITCE-ODE RDMS.

Attribute	Descriptor	Source
Publication category	Title, description, image	CKAN
Research project	Title, description, tags, license, visibility, author, author email, Mantainer, Mantainer email	CKAN
	Availability, identifier, language, provenance, temporal, spatial	DublinCore
Research product	Name, description, format	CKAN

2.4 System design

The diagram depicted in Figure 4 illustrates the technological architecture designed for the analysis and management of data within the GITCE-ODE RDMS. It encompasses the entire lifecycle of the data engineering process, structured into distinct stages including extraction, transformation, loading (ETL), and data analysis. This architecture features a comprehensive data processing pipeline aimed at facilitating the loading of diverse data types, their subsequent refinement and preparation, storage in databases and file systems, API interactions, metadata management, alignment with standards, and eventual analysis or export for external use beyond the GITCE-ODE Open Data web portal.

Detailed descriptions of the various stages of data processing and the responsibilities of the associated components are provided below.

- (E)xtract: corresponds to the recovery of data from various sources, in different formats and structures, with the aim of capturing data accurately and without loss.

  - Automated upload: involves data being automatically loaded from NoSQL and SQL databases, using scheduled processes or triggers to initiate data integration.
  - Manual upload: corresponds to the process of user's interaction with the system through a web interface for manually uploading files and defining metadata.

- (T)ransform: focuses on refining and improving the original raw data to ensure it is in the best possible format for data analysis. Data preparation at this stage includes:

  - Validation: refers to confirming the integrity and authenticity of data to ensure that it is accurate and usable.
  - Normalization: involves bringing disparate data into a common format or standard, making it easier to be compatible and consistent with data analysis tools.
  - Cleaning: corresponds to correcting or eliminating errors, inaccuracies, duplicates, or irrelevant data points to improve data quality.
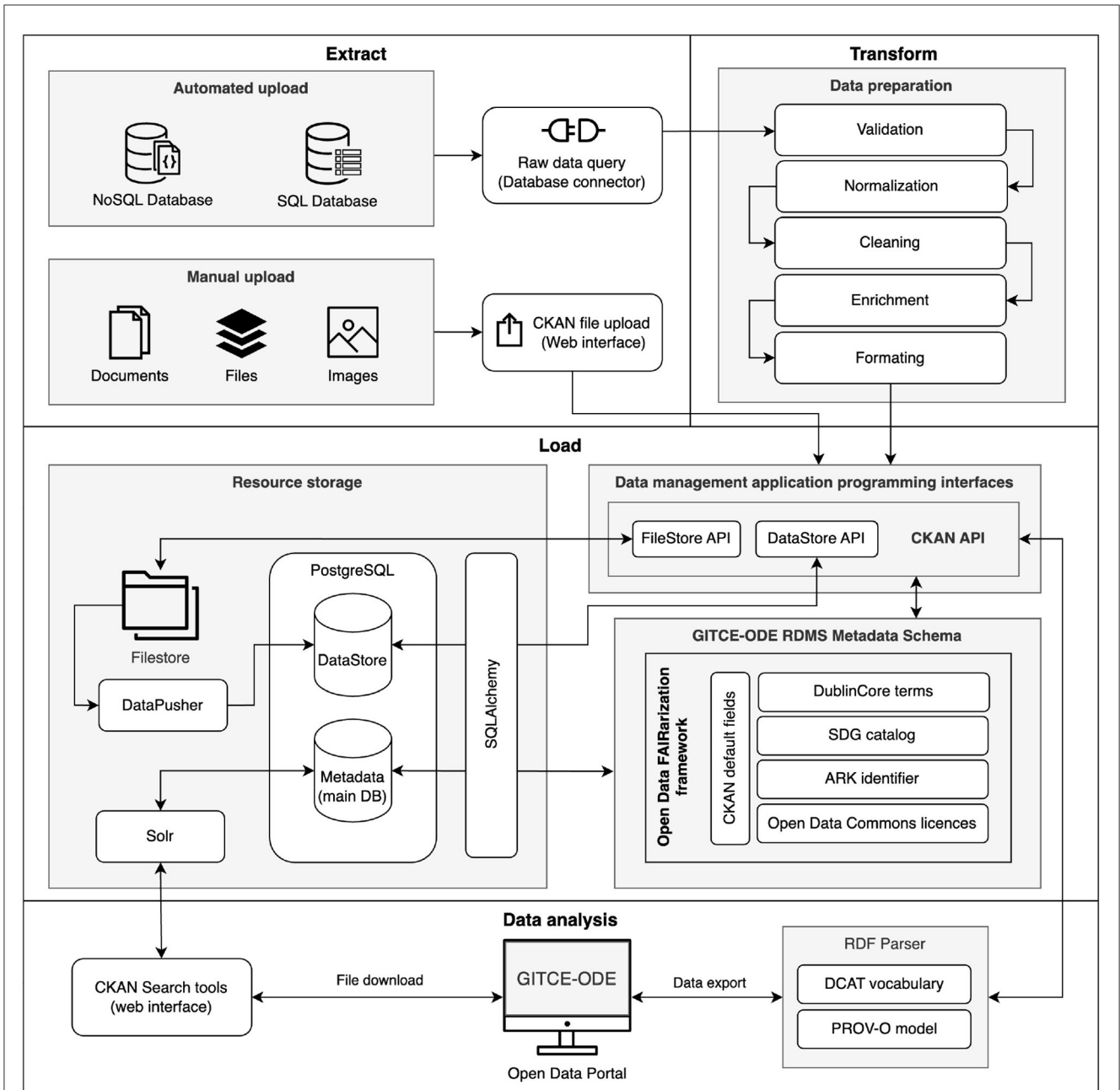
FIGURE 4
GITCE-ODE RDMS technological architecture.

- Enrichment: involves augmenting the dataset with additional relevant information from complementary sources to add value.
- Formatting: corresponds to the transformation of data into a structure or format suitable for analysis or subsequent processing.

- (L)oad: involves transferring the processed data to a storage system where it can be effectively accessed, managed, and queried, encompassing the following stages:

- Resource storage: different backup options for data, metadata and resources.

  * FileStore: file system for storing and managing media files, as well as unstructured or binary data.
  * DataStore: PostgreSQL database for storing and managing structured data such as records and tables.
  * Metadata (main DB): PostgreSQL database used as a main central repository for dataset metadata and general system configurations.

* SQLAlchemy: toolkit for executing statements in Structured Query Language (SQL) and object relational mapping (ORM) library that facilitates interaction and programmatic management of databases through Python instructions.
* Solr: searchable index for efficient and scalable metadata indexing, query optimization, and data retrieval.

- Data Management Application Programming Interfaces (APIs): includes the different APIs, such as the FileStore API, the DataStore API, and the CKAN API interface,

required to interact with the corresponding storage resources.
- GITCE-ODE RDMS Metadata Schema: comprises the framework and set of standards that specify how dataset metadata is organized according to the implementations chosen to follow the FAIR principles and corresponding guidelines.

• Data analysis: includes leveraging the stored resources and data management and retrieval tools provided by the GITCE-ODE RDMS to obtain research outputs, gain insights, and make informed decisions, including:
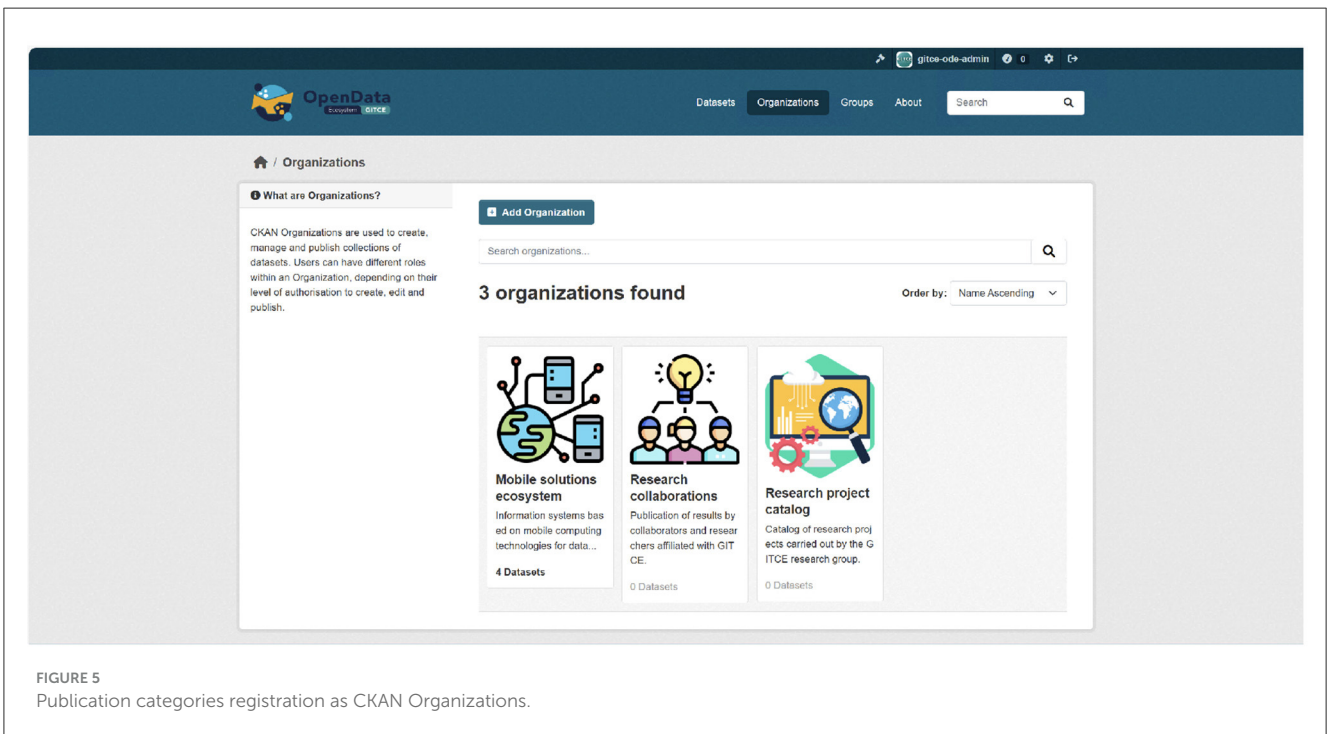

FIGURE 5
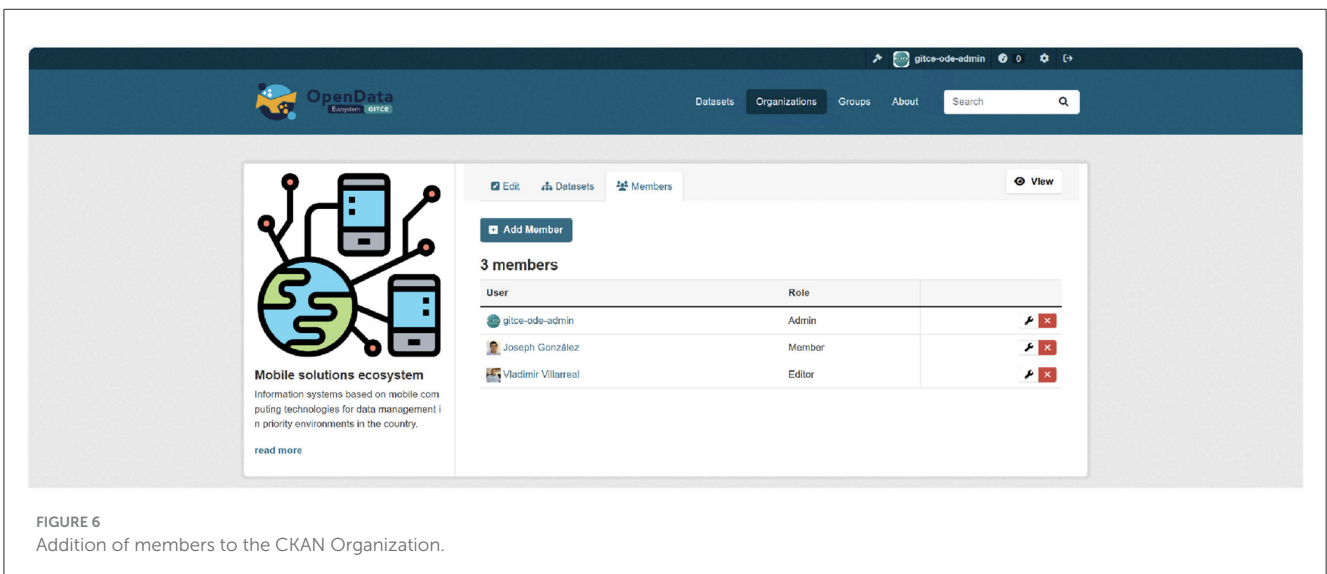Publication categories registration as CKAN Organizations.

FIGURE 6
Addition of members to the CKAN Organization.

– Open Data Portal: represents the web interface for the search and analysis of datasets and research products that have been processed and uploaded.

* File Download: feature that allows users to search for and retrieve files that can be accessed and used offline with other tools.
* Data Export: option to programmatically export data from the system using a parsing tool that transforms the stored data and metdata to RDF format using the DCAT vocabulary and PROV-O, to allow information to be shared or processed

in different systems oriented to the analysis of scientific products utilizing a linked data format for knowledge representation.

3 Results

To test the GITCE-ODE RDMS configuration, the developed web platform integrated with one of the systems from the GITCE ecosystem of mobile and ubiquitous solutions. This aimed to validate the Open Data web portal's functionality in sharing data
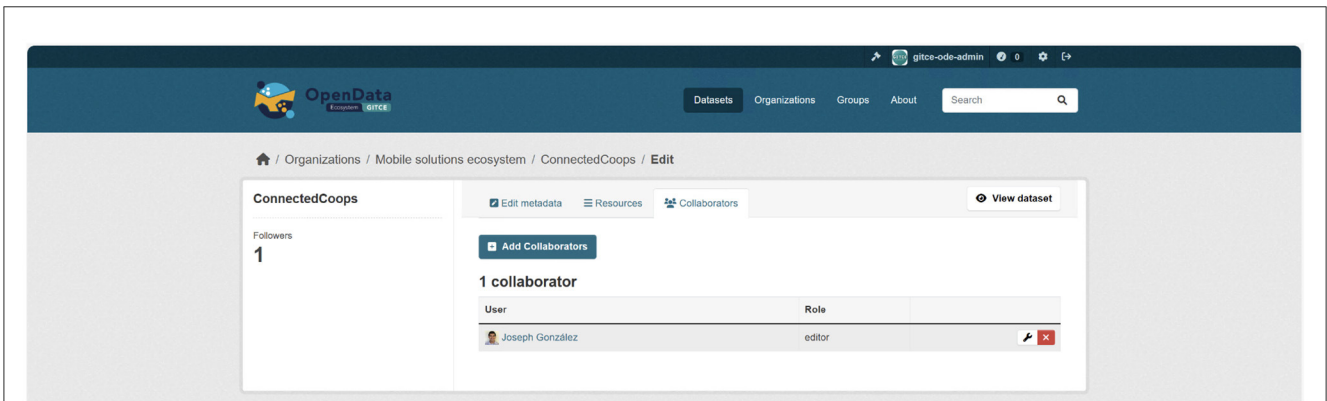

FIGURE 7
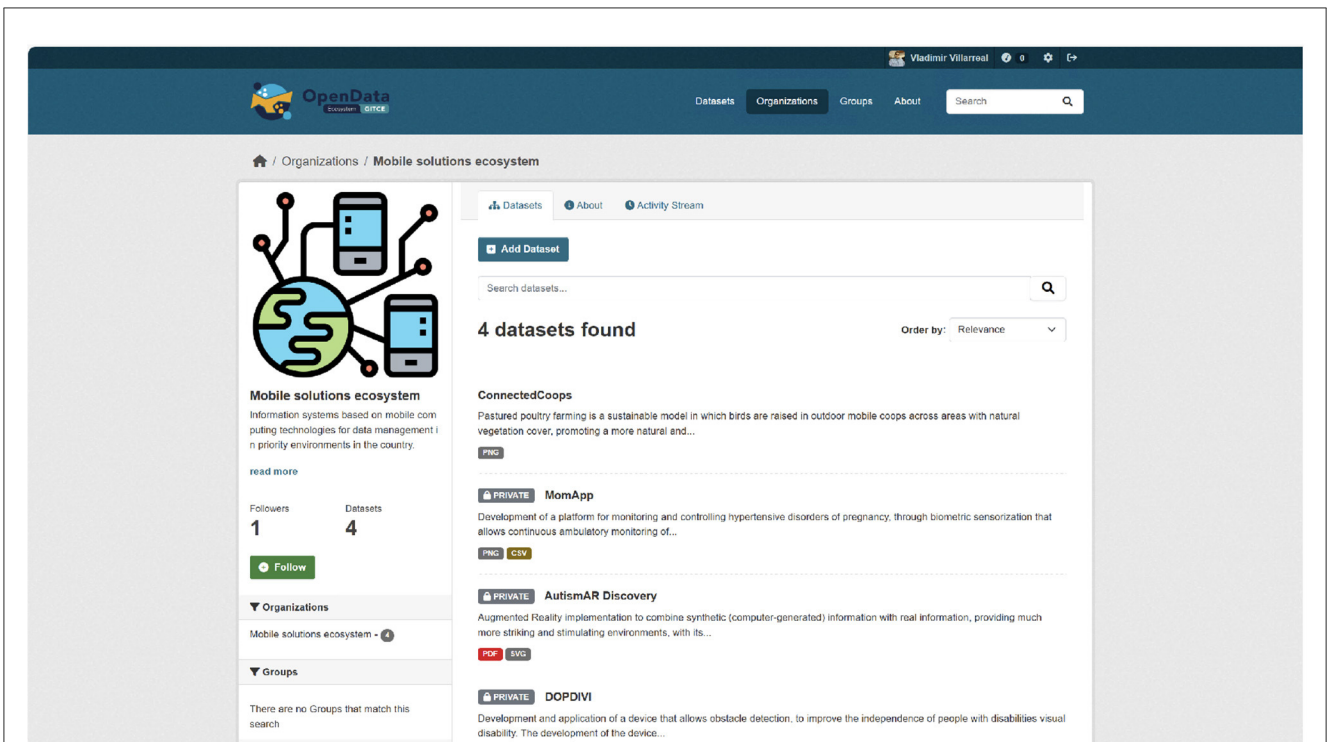Addition of a research project contributor as CKAN Dataset collaborator.

FIGURE 8
Research projects registration as CKAN Datasets.

and research information, as well as its capability to dynamically receive data from external databases and information systems.

3.1 ConnectedCoops: precision livestock farming for remote monitoring of pastured poultry

The ConnectedCoops research project is part of GITCE's Ecosystem of Mobile and Ubiquitous Solutions for the integration of information systems based on mobile computing technologies for data management in priority environments of society. It encompasses the development of a distributed information system to facilitate remote monitoring of environmental and animal welfare conditions present in production spaces dedicated to pastured poultry farming. It involves the deployment of a network of wireless sensors to monitor temperature, humidity, and lighting levels in infrastructures known as mobile chicken coops designed for the rearing of birds in open fields. The wireless network extends the deployment of four nodes made up of three sensors in each, adding up to a total of 12 sensors installed in the production space. Environmental measurements are captured throughout the day in 30-min periods and transmitted via Bluetooth Low Energy (BLE) technology to the network's hub node and gateway to the Internet. This node is responsible for replicating the information remotely to a Cloud Firestore database in the Firebase cloud through the 4G/LTE cellular data network. The resulting stored information can then be consulted by users using a mobile application installed in their Android smartphones.

3.1.1 Semantic context treatment

The testing process commenced with the execution of the tasks outlined in the research data opening flow, leveraging the tools offered by the CKAN platform to organize resources.
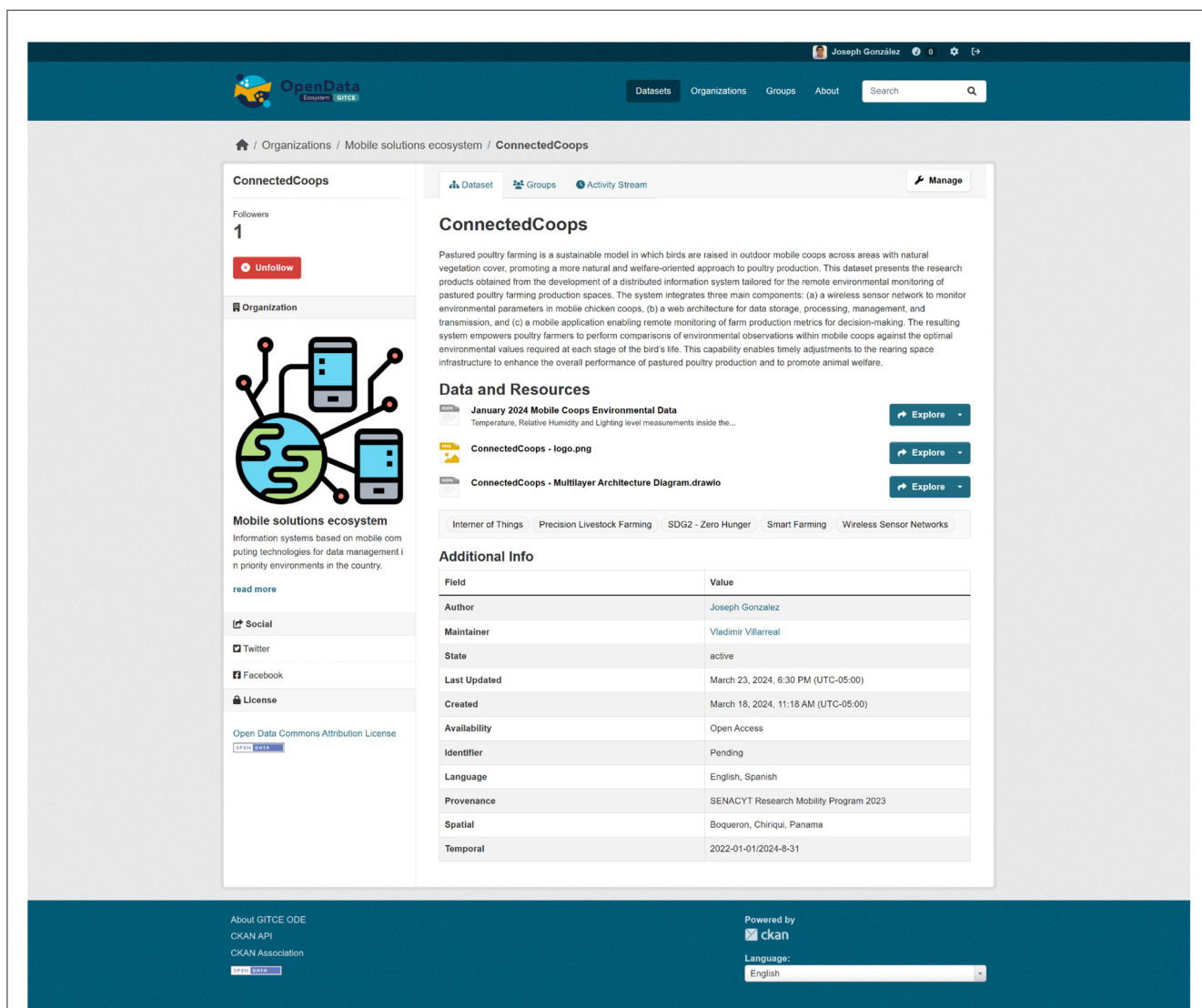

FIGURE 9
Addition of the research project metadata schema and research products.

- Initially, the resource publication categories of the GITCE-ODE RDMS were created through the web portal administrator role. Each publication category was configured

as a CKAN Organization, serving as logical entities to group and manage datasets and resources associated with specific initiatives, groups, or projects. Figure 5 illustrates the interface
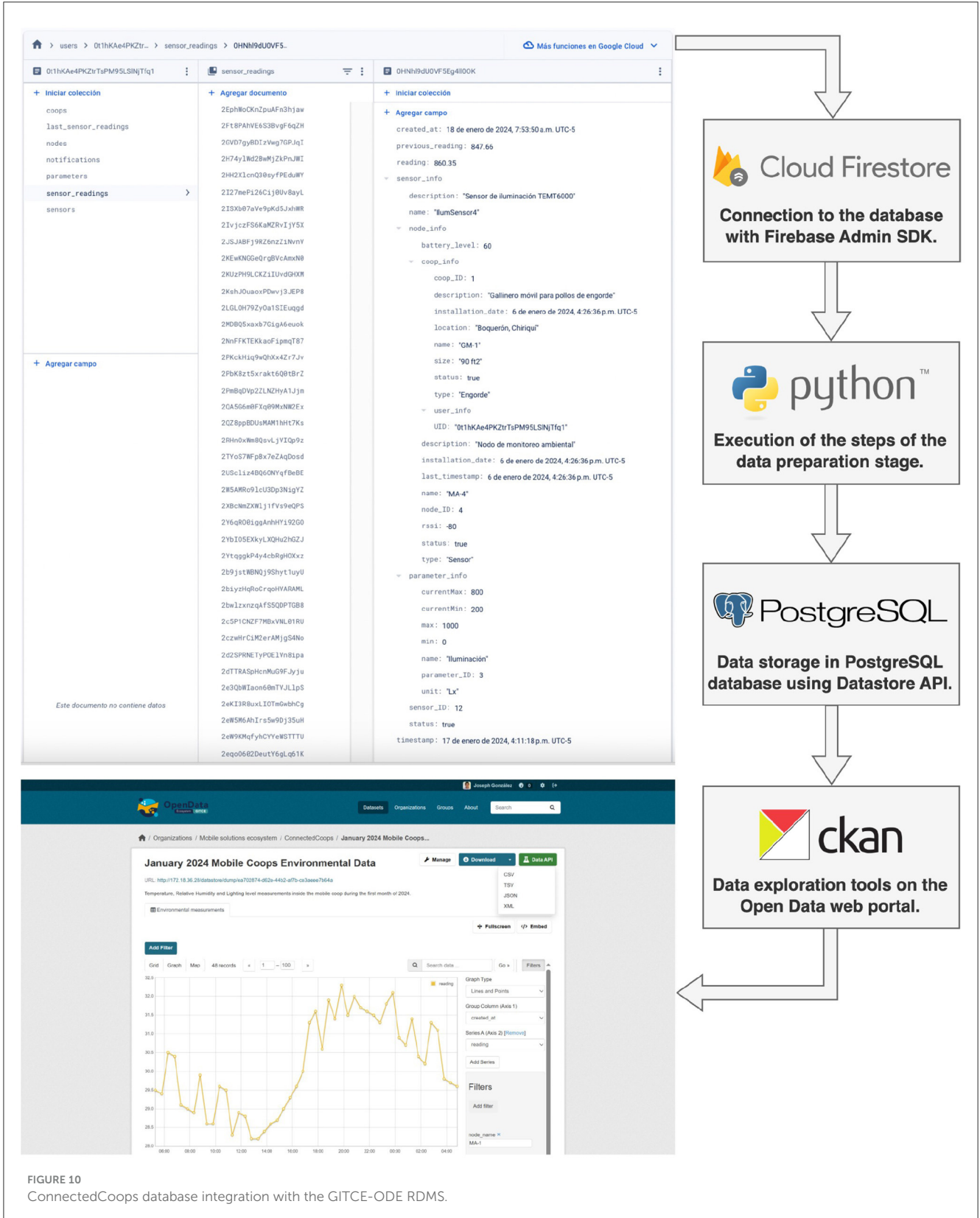

FIGURE 10
ConnectedCoops database integration with the GITCE-ODE RDMS.

of the web portal displaying the three initially identified publication categories, each with their corresponding title, description, and image metadata.

- After creating the publication category, members were added with authorized access for collaboration within the corresponding organization. Figure 6 displays the list of organization members for the Mobile Solutions Ecosystem project category. This list comprises the Administrator of the web portal, a user assigned with the role of organization editor responsible for acting as a Data Steward for the management of the publication category, and a user with the role of organization member responsible for acting as Contributor of research data and resources to the publication category.

- To facilitate the addition of resources and research products by a Contributor within a research project, the necessary authorization must be granted. This entails the web portal Administrator first adding each user as a member of the organization corresponding to the publication category of the research project dataset. Subsequently, each user is designated as a collaborator of the specific research project dataset, as illustrated in Figure 7. This authorization empowers users to upload data and documents both manually and programmatically, and to provide the required metadata.

- Subsequently, the data steward of the publication category proceeded to register the research projects belonging to the mobile solutions ecosystem as datasets within the organization, fulfilling the role of its editor. Figure 8 showcases the listed datasets, which encompass the four research projects constituting the GITCE mobile solutions ecosystem to date. This process involved registering data according to the predefined descriptors outlined as part of the metadata scheme for the standardization of data in the GITCE-ODE RDMS datasets, adhering to the FAIR principles. This resulted in the research project documentation and resources structure presented in Figure 9.

3.1.2 ETL process and data analysis

For the integration of the data from the ConnectedCoops system for its publication on the Open Data web portal, a script in Python language was programmed for the periodic and automated execution of the stages of the ETL process described in the technological architecture of the GITCE-ODE RDMS. In the Figure 10 the original semi-structured organization of the data stored as JSON documents in the Firestore database is presented, as well as its transformation to a tabular structure for data exploration in the Open Data web portal. The execution of this process entailed the use of the Firebase Admin SDK in its version for Python, through which the connection with the project's database was established in Firestore, allowing the execution of query requests to the records stored in the corresponding document collection within a specific period of time. The records obtained were subjected to the different stages of the system's data preparation process to ensure that the information had the necessary quality for publication. Subsequently, the DataStore API was used for the manipulation, grouping and structuring of the data in tabular format suitable for storage in the DataStore PostgreSQL database. Finally, the data can

be accessed from the Open Data web portal using the CKAN data exploration tools.

4 Conclusion

The development of the GITCE-ODE RDMS stands as a pivotal achievement in advancing Open Science initiatives within the Research Group on Emerging Computational Technologies (GITCE) at the Technological University of Panama. This comprehensive platform serves as a cornerstone for enhancing transparency, collaboration, and accessibility in research endeavors, in line with the large scale vision of Open Science toward open and reproducible scientific practices and Open Data initiatives in benefit of society.

GITCE-ODE's integration of diverse research projects, spanning societal priority areas and sustainable development goals, underscores its commitment toward facilitating the long-term storage and dissemination of research products originating from applied research initiatives with the use of emerging technologies. Anchored in the FAIR principles, the platform ensures that scientific data and products are easily discoverable, accessible, interoperable, and reusable, thereby streamlining the process of data discovery, reuse, and exploitation. This commitment to standards enhances the efficiency of data sharing and encourages the broader scientific community's engagement with GITCE's research outputs.

The technological architecture of GITCE-ODE RDMS enables seamless data processing, from extraction to analysis, providing researchers with tools to find, manage, explore and extract diverse datasets and digital resources. Integration with external platforms, exemplified by the connection with a system from a mobile and ubiquitous solution ecosystem, underscores GITCE-ODE's adaptability and interoperability in handling diverse data sources. Moreover, the collaborative efforts of key stakeholders, including administrators, data stewards, contributors, and stakeholders, play a crucial role in ensuring the effective management of research projects semantic context and metadata, as well as the utilization of research data within the platform. This collective endeavor not only empowers academic research group members to harness accumulated knowledge but also opens the door toward fostering scientific collaboration on both national and international scales. Future works include enhancing metadata registration and project tracking through custom modifications to the web portal interface, and storing data from previous projects.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Writing – original draft. JG: Investigation, Methodology, Visualization, Writing – original draft. JF: Conceptualization, Resources, Validation, Writing – review & editing. CD: Data curation, Formal analysis, Methodology, Validation, Writing – review & editing. MN: Data curation, Formal analysis, Software, Writing – review & editing. DC: Writing – review & editing. MR: Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ali, M., Alexopoulos, C., and Charalabidis, Y. (2022). "A comprehensive review of open data platforms, prevalent technologies, and functionalities," in *ICEGOV '22: Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance* (New York, NY: ACM), 203–214. doi: 10.1145/3560107.3560142

Alliance, A. (2024). *Archival resource key* (ARK). Available online at: https://arks.org/ (accessed February 20, 2024).

Burgelman, J-., Pascu, C., Szkuta, K., Schomberg, R. V., Karalopoulos, A., Repanas, K., et al. (2019). Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. *Front. Big Data* 2:469872. doi: 10.3389/fdata.2019.00043

Castrillón-Muñoz, A. J., Infante-Moro, A., Zúñiga-Collazos, A., and Martínez-López, F. J. (2020). Generación de empresas derivadas de base tecnológica (spin offs), a partir de los resultados de i+d+i de los grupos de investigación de la universidad del cauca, colombia. *Inf. Tecnol.* 31, 67–78. doi: 10.4067/S0718-07642020000100067

Christensen, J., Ekelund, N., Melin, M., and Widén, P. (2021). The beautiful risk of collaborative and interdisciplinary research. a challenging collaborative and critical approach toward sustainable learning processes in academic profession. *Sustainability* 13:4723. doi: 10.3390/su13094723

DCMI (2024). *Dublincore.* Available online at: https://www.dublincore.org/ (accessed February 20, 2024).

Finkel, M., Baur, A., Weber, T., Osenbrück, K., Rügner, H., Leven, C., et al. (2020). Managing collaborative research data for integrated, interdisciplinary environmental research. *Earth Sci. Inform.* 13, 641–654. doi: 10.1007/s12145-020-00441-0

GITCE (2024). *Grupo de investigación en tecnologías computacionales emergentes.* Available online at: https://gitce.utp.ac.pa/ (accessed February 14, 2024).

Gonzalez, J., Villarreal, V., and Muñoz, L. (2023). Microservice architecture for a remote management platform for pastured poultry farming using amazon web services and wireless mesh sensor networks. *Ing. Solidar.* 19, 1–22. doi: 10.16925/2357-6014.2023.01.02

Herrera-Cubides, J. F., Gaona-García, P. A., Montenegro-Marin, C. E., and Sánchez-Alonso, S. (2023). The relevance of open data principles for the web of data. *J. Electr. Comput. Eng.* 2023, 1–17. doi: 10.1155/2023/4854965

Karimova, Y., Castro, J. A., and Ribeiro, C. (2019). "Data deposit in a CKAN repository: a Dublin core-based simplified workflow," in *Digital Libraries: Supporting Open Science, Vol. 988*, eds. P. Manghi, L. Candela, and G. Silvello (Cham: Springer Verlag), 222–235. doi: 10.1007/978-3-030-11226-4_18

Kyvik, S., and Reymert, I. (2017). Research collaboration in groups and networks: differences across academic fields. *Scientometrics* 113, 951–967. doi: 10.1007/s11192-017-2497-5

Manco, A. (2023). Prácticas de ciencia abierta vistas desde la perspectiva de las comunidades de investigadores de las ciencias básicas de perú. *Rev. Cient.* 48, 40–55. doi: 10.14483/23448350.20905

Mozgova, I., Koepler, O., Kraft, A., Lachmayer, R., and Auer, S. (2020). "Research data management system for a large collaborative project," in *The Design Society* (Glasgow). doi: 10.35199/NORDDESIGN2020.48

Mozgova, I., Altun, O., Sheveleva, T., Castro, A., Oladazimi, P., Koepler, O., et al. (2022). Knowledge annotation within research data management system for oxygen-free production technologies. *Proc. Des. Soc.* 2, 525–532. doi: 10.1017/pds.2022.54

Muñoz, L., Villarreal, V., Morales, I., Gonzalez, J., and Nielsen, M. (2020). Developing an interactive environment through the teaching of mathematics with small robots. *Sensors* 20:1935. doi: 10.3390/s20071935

Munoz, L., Villarreal, V., Nielsen, M., Caballero, Y., and Sitton, I. (2021). "Knowledge management applying disruptive technologies as a response to the COVID-19 crisis in public administration," in *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)* (Chaves: IEEE), 1–6. doi: 10.23919/CISTI52073.2021.9476616

Naciones Unidas Panamá (2024). *Acerca de nuestro trabajo para los objetivos de desarrollo sostenible en panamá.* Available online at: https://panama.un.org/es/sdgs (accessed February 29, 2024).

Nie, H., Luo, P., and Fu, P. (2021). Research data management implementation at Peking University library: foster and promote open science and open data. *Data Intell.* 3, 189–204. doi: 10.1162/dint_a_00088

Nielsen, M., Villarreal, V., Muñoz, L., and González, J. (2023). "Design of a platform for the continuous monitoring of hypertensive disorders in pregnancy," in *2023 VI Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC)* (Cali: IEEE), 1–6. doi: 10.1109/AmITIC60194.2023.10366363

Open Knowledge Foundation (2024a). *Ckan.* Available online at: https://ckan.org/ (accessed February 20, 2024).

Open Knowledge Foundation (2024b). *Open data commons: legal tools for open data.* Available online at: https://opendatacommons.org/ (accessed February 22, 2024).

Patiño, D. H. C., Muñoz, L., Villarreal, V., and Pardo, C. (2023). "Proposal for the evaluation of the teaching/learning process in children with autism spectrum disorder through a mobile application with augmented reality," in *Proceedings of the International Conference on Ubiquitous Computing & Ambient Intelligence (UCAmI 2022), Vol. 594* (Cham: Springer Science and Business Media Deutschland GmbH), 113–118. doi: 10.1007/978-3-031-21333-5_11

Piedra, N., and Suárez, J. P. (2018). Hacia la interoperabilidad semántica para el manejo inteligente y sostenible de territorios de alta biodiversidad usando smartland-ld. *Rev. Iber. Sist. Tecnol. Inf.* 26, 104–121. doi: 10.17013/risti.26.104-121

Prieto, D. (2022). Ciencia abierta: desafíos y oportunidades para uruguay y el sur global. *Information* 27, 254–283. doi: 10.35643/Info.27.1.5

Rodríguez, M., Muñoz, L., Villarreal, V., and Concepción, D. H. (2023). "Proposal of a device for obstacle detection applied to visually impaired people," in *Proceedings of the 15th International Conference on Ubiquitous Computing & Ambient Intelligence, Vol. 835* (Cham: Springer Science and Business Media Deutschland GmbH), 215–220. doi: 10.1007/978-3-031-48306-6_22

Rudmark, D. (2020). "Open data standards: vertical industry standards to unlock digital ecosystems," in *Proceedings of 53rd Hawaii International Conference on System Sciences*, volume 2020 (Hawaii, HI: IEEE Computer Society), 2063–2072. doi: 10.24251/HICSS.2020.252

Schultes, E., Magagna, B., Hettne, K. M., Pergl, R., Suchánek, M., Kuhn, T., et al. (2020). "Reusable FAIR implementation profiles as accelerators of FAIR convergence," in *Advances in Conceptual Modeling, Vol. 12584* (Cham: Springer Science and Business Media Deutschland GmbH), 138–147. doi: 10.1007/978-3-030-65847-2_13

Stagars, M. (2016). "Promises, barriers, and success stories of open data," in *Open Data in Southeast Asia* (Cham: Springer International Publishing), 13–28. doi: 10.1007/978-3-319-32170-7_2

Tzitzikas, Y., Pitikakis, M., Giakoumis, G., Varouha, K., and Karkanaki, E. (2021). What process can a university follow for open data the University of Crete Case. *Int. J. Metadata Semant. Ontol.* 15:254. doi: 10.1504/IJMSO.2021.125886

Vabø, A., Alvsvåg, A., Kyvik, S., and Reymert, I. (2016). The establishment of formal research groups in higher education institutions. *Nord. J. Stud. Educ. Policy* 2016:33896. doi: 10.3402/nstep.v2.33896

Villarreal, V., Muñoz, L., Nielsen, M., Gonzalez, J., Concepcion, D., Rodriguez, M., et al. (2023). "Towards a digital and ubiquitous ecosystem of mobile technology-based solutions to facilitate data management based on sustainable development goals," in *Proceedings of the 15th International Conference on Ubiquitous Computing & Ambient Intelligence, Vol. 835* (Cham: Springer Science and Business Media Deutschland GmbH), 112–117. doi: 10.1007/978-3-031-48306-6_11

Villarreal, V., Nielsen, M., and Samudio, M. (2018). Sensing and storing the blood pressure measure by patients through a platform and mobile devices. *Sensors* 18:1805. doi: 10.3390/s18061805

W3C (2013). *Prov model primer*. Available online at: https://www.w3.org/TR/prov-primer/ (accessed February 22, 2024).

W3C (2024a). *Data catalog vocabulary (dcat) - version 3*. Available online at: https://www.w3.org/TR/vocab-dcat-3/ (accessed February 22, 2024).

W3C (2024b). *Rdf 1.2 concepts and abstract syntax*. Available online at: https://www.w3.org/TR/rdf12-concepts/ (accessed February 22, 2024).

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18