Publishing as a First-Class Service in Virtual Research Environments: The D4Science Catalogue

Massimiliano Assante, Leonardo Candela, Andrea Dell'Amico, Luca Frosini, Francesco Mangiacrapa, Alfredo Oliviero, Pasquale Pagano, Giancarlo Panichi, Biagio Peccerillo, Tommaso Piccioli, Marco Procaccini
Istituto di Scienza e Tecnologie dell'Informazione (ISTI), Consiglio Nazionale delle Ricerche (CNR)
Via G. Moruzzi 1, 56124 Pisa, Italy
Email: name.surname@isti.cnr.it

***Abstract*—Virtual Research Environments (VREs) and Science Gateways support collaborative and data-intensive research by integrating data access, computing, and analytical services. However, publishing research outputs often remains external to these environments, requiring researchers to transfer datasets, workflows, software, and other artifacts to external repositories. This separation introduces friction and limits timely sharing and reuse.**

This paper presents the D4Science Catalogue, a configurable service designed to make publishing a native capability of Virtual Research Environments. The Catalogue enables communities to publish diverse research objects together with community-defined metadata and to support discovery through both human and machine access mechanisms. Experiences from multiple deployments illustrate how integrated catalogue-based services facilitate early sharing, improve discoverability, and support FAIR research practices within science gateways.

***Keywords*—Research Object Publishing, Metadata Catalogue, Virtual Research Environment, Open Science Infrastructure**

I. Introduction

Digital research infrastructures increasingly rely on Virtual Research Environments (VREs) and science gateways to support collaborative and data-intensive investigations [1], [2]. By integrating data access, computational resources, analytical tools, and collaboration services within unified platforms, these environments enable research communities to conduct complex investigations more efficiently and reproducibly.

Despite these advances, the publication of research outputs often remains external to the environments where research activities take place [3]. Researchers typically move datasets, workflows, software artefacts, and other research products from their working environments to external repositories once a study reaches a mature stage [4]. Moreover, intermediate research artefacts and workflows are often not systematically captured or shared, limiting reuse and reproducibility [5]. This separation introduces friction in the research lifecycle and can delay the dissemination and reuse of valuable research outputs. In practice, many research products—including intermediate datasets, experimental workflows, analytical scripts, and derived results—can already support collaboration, validation, and reuse well before the final publication stage.

The increasing adoption of Open Science practices and FAIR principles further highlights the importance of enabling timely sharing, discoverability, and reuse of research artefacts. However, supporting these practices requires infrastructures that allow research communities to publish diverse types of research products together with adequate contextual metadata, while maintaining flexibility to accommodate domain-specific needs.

This paper argues that publishing should be considered a first-class capability of Virtual Research Environments, rather than a separate step occurring outside the research workflow. Integrating publishing capabilities directly within research environments allows scientists to share and curate research products as early as convenient, preserving the context in which they were produced and facilitating collaboration within and across communities.

To support this vision, the D4Science infrastructure [6] provides a configurable catalogue service enabling research communities to publish, describe, discover, and retrieve research objects directly within their Virtual Research Environments. The D4Science Catalogue supports the publication of heterogeneous research products—including datasets, software, workflows, documents, and web resources—together with community-defined metadata schemas and flexible publishing workflows.

The remainder of the paper presents the catalogue-based publishing approach adopted in D4Science. Section II discusses the main publishing requirements emerging in Virtual Research Environments. Section III describes the design and main features of the D4Science Catalogue service. Section IV illustrates how the catalogue supports publishing workflows in operational Virtual Research Environments and discusses lessons learned from its adoption. Finally, Section V summarises the main conclusions and outlines future developments.

II. Publishing Requirements in Virtual Research Environments

Virtual Research Environments support collaborative and data-intensive research by integrating data access, analytical tools, and computing capabilities within unified platforms. To fully support Open Science practices, these environments must also enable effective publishing of research outputs. However, publishing within VREs introduces specific requirements that differ from traditional repository-based approaches.

Research activities within VREs produce a wide range of artefacts, including datasets, software components, workflows,

documents, visualisations, and links to external resources. These artefacts often evolve throughout the research lifecycle, from preliminary results to final outputs.

A publishing service embedded in a VRE must therefore support heterogeneous research objects that are often dynamic and evolving, allowing users to share not only final product but also intermediate, composite, and living artefacts that are progressively enriched during the research process. This capability is essential to enable collaboration, validation, and reuse at different stages of research.

Rich and Contextual Metadata: The value of published research objects depends on the availability of adequate contextual information. Metadata must capture not only basic descriptive elements (e.g., title, description, authorship) but also domain-specific attributes, provenance information, and relationships with other artefacts.

Given the diversity of research domains served by VREs, a one-size-fits-all metadata schema is insufficient. Instead, publishing services must support flexible and extensible metadata models, enabling communities to define their own schemas, controlled vocabularies, and validation rules. This approach ensures that published objects are both interpretable and reusable within their specific scientific contexts.

Community-Driven Publishing Workflows: Different research communities adopt different practices for sharing and validating research outputs. Some favour rapid, informal sharing to support collaboration, while others require structured review and curation processes.

Publishing services in VREs must therefore support configurable workflows, ranging from immediate publication to moderated processes involving roles such as editors and reviewers. These workflows should accommodate activities such as submission, validation, approval, and revision, allowing communities to balance openness with quality control.

Integration with the Research Workflow: In many current infrastructures, publishing is performed outside the environment where research is conducted, requiring manual transfer of artefacts to external repositories. This separation introduces friction, disrupts workflows, and often leads to the loss of contextual information.

To overcome these limitations, publishing capabilities should be tightly integrated within VREs, enabling users to publish research objects directly from their working environment. Such integration supports early and continuous sharing of artefacts, preserving their context and reducing the effort required for dissemination.

Discoverability and Access: Publishing within VREs must ensure that research objects are easily discoverable and accessible by both humans and machines. This requires advanced search and browsing capabilities, including support for filtering by metadata attributes, as well as mechanisms for navigating collections of related items.

At the same time, access control mechanisms must accommodate different visibility requirements, ranging from private sharing within a community to public dissemination. A flexible visibility model is therefore essential to support collaboration while respecting data sensitivity and ownership constraints.

Interoperability and External Dissemination: While VREs provide a controlled environment for research activities, published outputs must often be shared beyond the boundaries of the platform. Publishing services must therefore support interoperability with external systems, including repositories, catalogues, and aggregators.

This requires the adoption of standard data models and protocols, as well as mechanisms for exporting or synchronising content with external infrastructures. Supporting both human-readable interfaces and machine-oriented access is essential to maximise the visibility, reuse, and long-term impact of published research objects.

Support for FAIR and Open Science Practices: Finally, publishing services in VREs must align with the principles of Open Science and FAIR (Findable, Accessible, Interoperable, Reusable) data management [7]. This implies enabling early sharing of research artefacts, providing rich metadata, ensuring persistent access, and supporting reuse across communities.

Embedding publishing capabilities within VREs allows these principles to be operationalised throughout the research lifecycle, rather than being applied only at the final stage of dissemination.

III. The D4Science Catalogue

The D4Science Catalogue is conceived as a core service of Virtual Research Environments, enabling publishing to be seamlessly integrated into the research lifecycle. Rather than treating publication as a final, external step, the Catalogue supports the continuous sharing, curation, and discovery of research objects within the same environment where they are produced.

The design of the Catalogue is guided by three main principles:

- *Publishing as a native capability*: Researchers are enabled to publish research objects directly from their working environment, including intermediate artefacts, thus reducing friction and supporting timely dissemination.
- *Community-driven configurability*: Each research community can define its own types of research objects, metadata schemas, and publishing workflows, ensuring flexibility and domain appropriateness.
- *Interoperability and openness*: The Catalogue supports both human- and machine-oriented access, enabling integration with external systems and fostering reuse in line with FAIR principles.

The D4Science Catalogue is architected as an open and integrated service within the D4Science Infrastructure (cf. Fig. 1). Exposed through the VRE user interface by means of a dedicated Catalogue Portlet, it provides specialised components for publication, moderation, sharing, and interoperability with external repositories. Its core access layer is represented by *gCat* (cf. Sec. III-A), a RESTful service enabling programmatic interaction with the Catalogue and supporting both internal

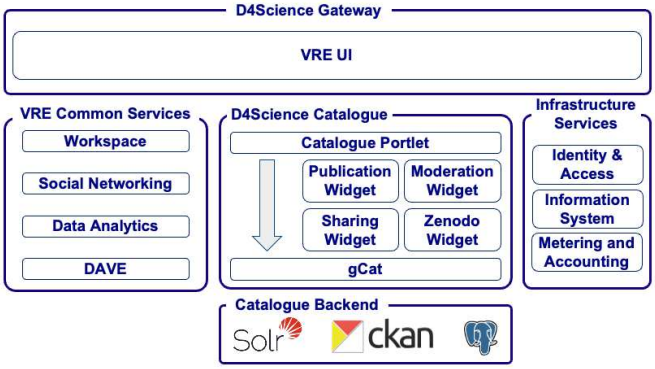

Fig. 1. D4Science Catalogue Architecture

integration and external machine-oriented access. The Catalogue is tightly integrated with D4Science common services, including the *Workspace*, where collaborators share datasets and research results; *Social Networking*, where community interactions take place through posts and replies; *Data Analytics*, where analytical methods are shared and executed; and DAVE [8], the D4Science conversational agent supporting the exploitation and uptake of VRE services through a wide range of user workflows. It also relies on infrastructure services, including Identity and Access, the Information System, and Metering and Accounting. At the backend level, the Catalogue builds on open technologies such as Solr [9], CKAN [10], and PostgreSQL [11].

At the core of the Catalogue is the concept of a *type-based catalogue item* (cf. Fig. 2), representing a published research object. The model is designed to accommodate heterogeneous artefacts, including datasets, software, workflows, documents, and web resources, and to support their organization into collections.

Catalogue items are organised into *Organizations* and *Groups*, which support both the structuring and the governance of catalogue contents. In particular, each catalogue item belongs to exactly one Organization and may belong to zero, one, or many Groups. Organizations provide a higher-level container representing the community, institution, or project responsible for publishing the item, and are associated with the corresponding Virtual Research Environment context. Groups provide a finer-grained subdivision mechanism used to organise catalogue items according to thematic, functional, or community-specific criteria.

Each catalogue item consists of the following components:

- *Basic metadata*: General descriptive information such as title, description, tags, and licensing information.
- *Type-specific metadata*: Additional metadata fields defined according to the type of research object and community-specific requirements.
- *Resources*: One or more associated resources, including files stored within the VRE or external links to web-accessible content.
- *Management information*: System-level information supporting item identification (PURL), ownership, visibility, and lifecycle management.

The model supports extensibility through community-defined profiles, allowing different research domains to tailor metadata structures and validation rules according to their needs. In fact, a key feature of the D4Science Catalogue is its support for *community-defined metadata profiles*, i.e., a specification-based mechanism enabling to characterise the types of objects that can be published by defining the type-specific metadata component of each envisaged catalogue item type. Rather than enforcing a fixed schema, the Catalogue allows each VRE to configure: (*i*) the types of research objects to be published (e.g., datasets, workflows, services); (*ii*) the type-specific metadata fields associated with each type; (*iii*) controlled vocabularies, tags, and classification schemes. This is achieved by creating an XML-based metadata profile for each type of item comprising a list of metadata field specifications each defining:

- *field identifier*: (optional) the identifier to used to refer to the field in the metadata record. If not provided, the field name is used;
- *field name*: the name of the metadata field;
- *mandatory flag*: a boolean value indicating whether the field is mandatory (`true`) or optional (`false`);
- *data type*: the type of value that the field can assume. Supported types include `String`, `Time`, `Time_Interval`, `ListOfTimes`, `Text`, `Boolean`, `Number`, and `GeoJSON`;
- *maximum cardinality constraint*: the maximum number of occurrences allowed for the field. Supported values are any positive integer or $\star$ to denote an unbounded number of occurrences;
- *default value*: (optional) the value assigned to the field when no explicit value is provided;
- *note*: a descriptive text clarifying the purpose of the field and providing guidance to support its correct completion;
- *vocabulary*: (optional) a predefined list of admissible values for the field. The `isMultiSelection` attribute specifies whether multiple values can be selected;
- *validator*: (optional) a validation rule constraining the field values through a regular expression;
- *tagging directive*: (optional) a rule specifying how tags are generated for the metadata record from the field. Supported values include `onFieldName` (the field name becomes a record Tag), `onValue` (the field value becomes a record Tag), `onFieldName_onValue` (a combination of the field name and the field value becomes a record Tag); `onValue_onFieldName` (a combination of the field value and the field name becomes a record Tag). The `separator` attribute specifies the symbol use when combining field name and field value to generate the Tag.
- *grouping directive*: (optional) a rule specifying how the metadata record is associated with a group on the basis of the field. It supports the same values as the tagging

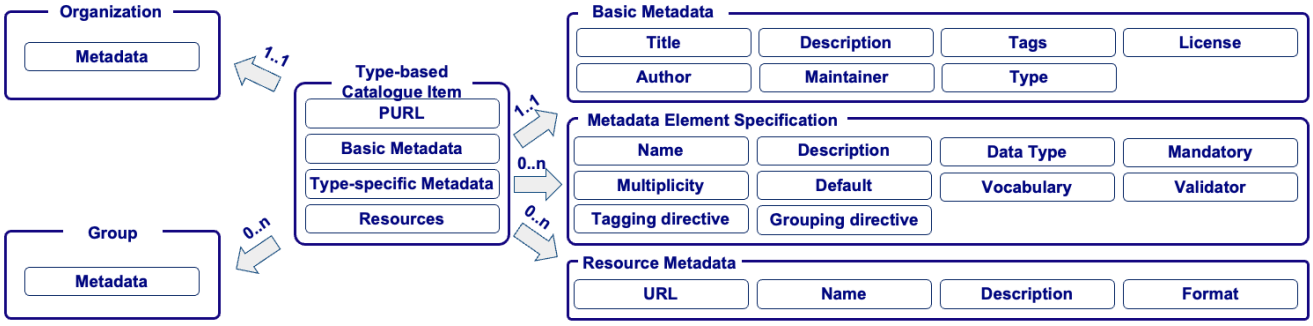

Fig. 2. D4Science Type-based Catalogue Item Information Model

directive. The `separator` attribute specify the symbol used when combining field name and field value to identify the target group.

Metadata elements can be organised into namespaces, allowing related elements to be grouped and managed coherently. Each namespace is characterised by: (*a*) a *unique identifier*, (*b*) a *name*, (*c*) a *title* introducing the related elements, and (*d*) a *description* providing an explanation of those related elements in the metadata profile.

This approach enables communities to capture the contextual information necessary for proper interpretation and reuse of research objects, while maintaining consistency within the community. Metadata profiles can be updated over time, supporting the evolution of community practices and requirements.

The Catalogue adopts a role-based access model in which each user is assigned one or more roles, with role assignment performed by the VRE Manager. The hierarchical roles include *Catalogue-Member*, mainly entitled to list and read items; *Catalogue-Editor*, entitled to manage the items he or she creates besides listing and reading items; *Catalogue-Admin*, responsible for administering several aspects of the Catalogue including the management of all the items and the accompanying resources, the management of groups and organizations; and *Catalogue-Manager*, entitled to fully manage a catalogue service but interferring to the content moderation. In addition, the Catalogue defines the non-hierarchical role of *Catalogue-Moderator*, specifically devoted to moderation actions such as approving or rejecting submitted items in the contexts where this feature is enabled. This role model enables the Catalogue to combine broad access to published content with controlled governance, curation, and moderation workflows.

Publishing is typically performed through a wizard-based interface (the Publication Widget), guiding users through a sequence of steps: (*i*) completion of basic metadata; (*ii*) completion of type-specific metadata; (*iii*) association of resources (files and/or links).

Publishing can be initiated directly from the Catalogue interface or from other VRE services, such as the Workspace, enabling users to publish artefacts as they are produced, or the Analytics Platform, enabling users to publish implementations of analytics methods as-a-Service.

The Catalogue supports flexible and configurable publishing workflows that reflect different levels of curation and quality control. In particular, it supports both immediate publishing and moderated workflows. In immediate publishing workflows, authorised actors (i.e., users having at least the Catalogue-Editor role) are entitles to directly publish items into the catalogue, the item becomes immediately visible and available for the catalogue users. In moderated scenarios, distinct phases are envisaged: (*a*) a *submission* phase, where authorised users submit items for publication; (*b*) a *moderation* phase, where authorised users (users with *Catalogue-Moderator* role) review submissions and decide whether to approve or reject them. Only approved items are published.

This flexibility allows communities to adopt workflows ranging from lightweight sharing to more formal review processes, including peer-review-like mechanisms.

The Catalogue can also be configured to promote newly published items through the VRE Social Networking component. In particular, whenever an object is published, the system may automatically generate a post announcing its availability, thereby increasing its visibility within the community and fostering timely awareness, discussion, and reuse.

The Catalogue provides rich mechanisms for accessing and discovering published items, supporting both end users and applications.

For human users, the Catalogue offers a web-based interface featuring: (*a*) Keyword-based search; (*b*) Faceted browsing based on metadata attributes such as type, tags, format, and license; (*c*) Domain-specific features such as geospatial search when applicable.

For machine access, the Catalogue exposes its content through multiple interfaces: (*a*) A RESTful API (gCat) supporting programmatic access to items, resources, and metadata profiles; (*b*) Standard-based representations such as DCAT, facilitating interoperability with data catalogues; (*c*) Plug-in mechanisms supporting additional protocols (e.g., OAI-PMH, CSW) and domain-specific integrations.

The Catalogue is designed to both harvest and be harvested, enabling integration with external infrastructures and repositories. Furthermore, items can be shared beyond the VRE through persistent links or deposited in external repositories

(e.g., Zenodo), supporting broader dissemination and long-term preservation.

The Catalogue also provides per-item and per-resource usage indicators, including a *recent count* (views in the last 14 days) and a *total count* (all tracked views). To avoid inflated statistics, at most one view per user per item or resource per day is counted.

A. The gCat Service

gCat is the REST service providing the machine-oriented access layer of the D4Science Catalogue. Its API[1] exposes the main catalogue entities and functions through a uniform service interface. In particular, gCat supports operations on catalogue *items*, their attached *resources*, item-type *profiles*, metadata *namespaces*, *licenses*, *groups*, *organizations*. Moreover, it supports management oriented operations on *users*, *trash*, *configuration*, and *extensions*. This makes the Catalogue accessible not only through the web interface, but also programmatically to internal components and external applications.

The item APIs support listing, filtered search, retrieval, creation, full and partial update, moderation, and deletion, including both soft deletion through the trash and irreversible purge operations. Item listing supports pagination and filtering, including full-text queries and the inclusion of private items for authorised users. Nested resource APIs allow resources associated with an item to be created, retrieved, updated, patched, and deleted. Beyond content management, gCat also supports the management of XML-based metadata profiles and their validation schemas, namespace discovery and schema retrieval, and context-dependent catalogue configuration persisted in the infrastructure Information System.

Overall, gCat plays a key role in the openness and interoperability of the D4Science Catalogue. It enables the implementation of publication, moderation, discovery, and integration workflows in a service-oriented way, thereby turning the Catalogue into a reusable infrastructure component rather than a purely user-facing application. Moreover, thanks to a plug-in based mechanisms it is possible to extend the protocols through which the catalogue content is exposed by complementing and extending what's natively supported by CKAN.

IV. Catalogue-Based Publishing in Practice

The practical value of the D4Science Catalogue emerges most clearly when it is considered as part of operational Virtual Research Environments. Across D4Science deployments, the Catalogue has been used to support community-specific publishing scenarios in which heterogeneous research objects, rich metadata, and configurable workflows are needed to make research products discoverable, reusable, and interoperable. At the time of writing, a total of 19 catalogue instances are operational overall hosting more that 510k items, 88 organizations, 238 groups and more than 100 types.

In the majority of cases, the catalogue was exploited as a service for disseminating research artefacts according to the FAIR principles where the artefacts were developed and implemented by specific VREs. The *SoBigData Catalogue*[2] developed by the SoBigData Research Infrastructure [12] features a broad collection of research resources including datasets, methods, experiments, and applications supporting advanced data science experimentation and reproducible research workflows. The *Blue-Cloud Catalogue*[3] developed by the Blue-Cloud initiative [13] promotes the publication of research results including datasets, methods, and notebooks as well as domain specific services and Virtual Laboratories. The catalogue developed in the context of the EOSC-Pillar supported the development of a rich array of training materials [14]. The catalogue developed by the AGINFRA [15] supported the publication of models, datasets, methods, services, and other research objects, and also supported more advanced dissemination patterns, such as the publication of *executable papers* through the integration of catalogue items with scholarly communication platforms [16].

The D4GNA Catalogue[4] [17] supports the publication and consultation of archaeological datasets for the Geoportale Nazionale per l'Archeologia. It is aimed at datasets produced by archaeological investigations in concession and by Italian archaeological missions abroad, and more generally serves as a digital access point for standardised, shareable archaeological documentation.

The Global Record of Stocks and Fisheries (GRSF) Catalogue[5] [18] provides a unified view over heterogeneous fisheries and stock information harvested from multiple external sources. This catalogue is part of a broader workflow including fetch, transform, normalise, clean, dissect, merge, publish, curate, validate, and exploit phases. After integration and normalization, the resulting records are published into a data catalogue offered through a dedicated VRE, where domain experts inspect, curate, approve, or reject them before final publication by another catalogue. The process preserves provenance, supports the assignment of globally unique and human-readable identifiers, and turns the catalogue into both a dissemination mechanism and a collaborative validation environment. This case is particularly relevant because it highlights how the Catalogue can support publication workflows in which the main value lies not only in exposing information, but also in enabling expert-mediated curation and quality control over integrated records.

Taken together, these experiences indicate that the D4Science Catalogue can support rather different publication settings while preserving a common architectural approach. In some cases, the emphasis is on data entry and visibility control; in others, it is on data integration, normalization, curation, and validation; in others, it is on FAIR dissemination of datasets, models, methods, and even executable research

[1] gCat Catalogue Service Documentation https://api.d4science.org/catalogue/docs/index.html

[2] SoBigData Gateway https://sobigdata.d4science.org

[3] Blue-Cloud Gateway https://blue-cloud.d4science.org/

[4] D4GNA Gateway https://gna.d4science.org

[5] GRSF VRE https://i-marine.d4science.org/web/grsf

outputs. This variety confirms that publishing as a first-class VRE capability is not limited to a single domain or object type, but can be configured to serve the concrete practices of diverse research communities.

V. Conclusion

This paper presented the D4Science Catalogue as a configurable service designed to make publishing a first-class capability of Virtual Research Environments. The proposed approach addresses key requirements emerging in science gateways, including support for heterogeneous and evolving research objects, community-defined metadata profiles, flexible publishing workflows, and both human- and machine-oriented access mechanisms.

By embedding publishing capabilities directly within the environments where research is carried out, the Catalogue reduces the separation between producing and disseminating research artefacts. This makes it possible to publish datasets, methods, workflows, software, documents, and other resources as early as convenient, while preserving the contextual information needed for their interpretation, discovery, and reuse. The experiences discussed in the paper show that the same architectural approach can support rather different publication settings, ranging from FAIR dissemination of community assets to curation-oriented workflows and expert-mediated validation scenarios.

Overall, the D4Science Catalogue demonstrates that publishing can be effectively integrated into the daily operation of Virtual Research Environments, thereby strengthening collaboration, improving discoverability, and supporting Open Science and FAIR practices across diverse research communities.

Future works include promoting catalogue item authors to first-class entities within the Catalogue, so as to better support attribution, discovery, and the interlinking of authors with the research objects they contribute to. Further developments will also address interoperability with external repositories and open scholarly infrastructures, and the refinement of community-driven curation workflows.

Acknowledgment

The work presented in this paper was supported by BlueCloud 2026 (grant agreement No. 101094227) funded by the European Union's Horizon Europe.

References

[1] K. A. Lawrence, M. Zentner, N. Wilkins-Diehr, J. A. Wernert, M. Pierce, S. Marru, and S. Michael, "Science gateways today and tomorrow: positive perspectives of nearly 5000 members of the research community," *Concurrency and Computation: Practice and Experience*, vol. 27, no. 16, pp. 4252–4268, 2015.

[2] L. Candela, D. Castelli, and P. Pagano, "Virtual Research Environments: An Overview and a Research Agenda," *Data Science Journal*, vol. 12, no. 0, pp. GRDI75–GRDI81, 2013.

[3] M. Assante, L. Candela, D. Castelli, P. Manghi, and P. Pagano, "Science 2.0 repositories: Time for a change in scholarly communication," *D Lib Mag.*, vol. 21, no. 1/2, 2015.

[4] C. C. Austin, T. Bloom, S. Dallmeier-Tiessen, V. K. Khodiyar, F. Murphy, A. Nurnberger, L. Raymond, M. Stockhause, J. Tedds, M. Vardigan, and A. Whyte, "Key components of data publishing: using current best practices to develop a reference model for data publishing," *Int. J. Digit. Libr.*, vol. 18, no. 2, pp. 77–92, Jun. 2017.

[5] A. Shaon, S. Callaghan, B. Lawrence, B. Matthews, A. Woolf, T. Osborn, and C. Harpham, "A linked data approach to publishing complex scientific workflows," in *IEEE 7th International Conference on E-Science, e-Science 2011, Stockholm, Sweden, December 5-8, 2011*. IEEE Computer Society, 2011, pp. 303–310.

[6] M. Assante, L. Candela, D. Castelli, R. Cirillo, G. Coro, L. Frosini, L. Lelii, F. Mangiacrapa, P. Pagano, G. Panichi, and F. Sinibaldi, "Enacting open science by D4Science," *Future Generation Computer Systems*, vol. 101, pp. 555–563, Dec. 2019.

[7] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. ́t. Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data*, vol. 3, no. 1, p. 160018, Dec. 2016.

[8] M. Assante, L. Candela, A. Dell'Amico, L. Frosini, F. Mangiacrapa, A. Oliviero, P. Pagano, G. Panichi, B. Peccerillo, T. Piccioli, and M. Procaccini, "Deploying conversational agents in virtual research environments: Approaches and lessons learned," *SN Computer Science*, vol. 7, no. 4, p. 339, 2026.

[9] Apache Solr Project, "Apache Solr: Open Source Search Platform," https://solr.apache.org/, accessed: 2026-04-10.

[10] CKAN Project, "CKAN: The Open Source Data Management System," https://ckan.org/, accessed: 2026-04-10.

[11] PostgreSQL Global Development Group, "PostgreSQL: The World's Most Advanced Open Source Database," https://www.postgresql.org/, accessed: 2026-04-10.

[12] V. Grossi, B. Rapisarda, F. Giannotti, and D. Pedreschi, "Data science at SoBigData: the European research infrastructure for social mining and big data analytics," *International Journal of Data Science and Analytics*, vol. 6, no. 3, pp. 205–216, Nov. 2018.

[13] D. Schaap, M. Assante, P. Pagano, and L. Candela, "Blue-Cloud: Exploring and demonstrating the potential of Open Science for ocean sustainability," in *2022 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea)*. Milazzo, Italy: IEEE, Oct. 2022, pp. 198–202.

[14] P. O. Garcia, L. Berberi, L. Candela, I. Van Nieuwerburgh, E. Lazzeri, and M. Czuray, "Developing the EOSC-Pillar RDM Training and Support Catalogue," in *Linking Theory and Practice of Digital Libraries*, G. Silvello, O. Corcho, P. Manghi, G. M. Di Nunzio, K. Golub, N. Ferro, and A. Poggi, Eds. Cham: Springer International Publishing, 2022, vol. 13541, pp. 274–281, series Title: Lecture Notes in Computer Science.

[15] M. Assante, A. Boizet, L. Candela, D. Castelli, R. Cirillo, G. Coro, E. Fernández, M. Filter, L. Frosini, T. Georgiev, G. Kakaletris, P. Katsivelis, R. Knapen, L. Lelii, R. M. Lokers, F. Mangiacrapa, N. Manouselis, P. Pagano, G. Panichi, L. Penev, and F. Sinibaldi, "Realizing virtual research environments for the agri-food community: The AGINFRA PLUS experience," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 19, Oct. 2021.

[16] M. Filter, L. Candela, L. Guillier, M. Nauta, T. Georgiev, P. Stoev, and L. Penev, "Open Science meets Food Modelling: Introducing the Food Modelling Journal (FMJ)," *Food Modelling Journal*, vol. 1, p. e46561, Sep. 2019.

[17] V. Acconcia, V. Boi, A. Falcone, F. Massara, L. Candela, F. Mangiacrapa, P. Pagano, and F. Sinibaldi, "ARIADNE Plus e il D4GNA - Dataset per il Geoportale Nazionale per l'Archeologia," *DigItalia*, vol. 18, no. 1, pp. 129–140, Jun. 2023.

[18] Y. Tzitzikas, Y. Marketakis, N. Minadakis, M. Mountantonakis, L. Candela, F. Mangiacrapa, P. Pagano, C. Perciante, D. Castelli, M. Taconet, A. Gentile, and G. Gorelli, "Methods and Tools for Supporting the Integration of Stocks and Fisheries," in *Information and Communication Technologies in Modern Agricultural Development*, M. Salampasis and T. Bournaris, Eds. Cham: Springer International Publishing, 2019, vol. 953, pp. 20–34, series Title: Communications in Computer and Information Science.