

2023



Istituto di Scienza e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



ISTI Annual Reports

InfraScience Research Activity Report 2023

InfraScience Lab., CNR-ISTI, Pisa, Italy

ISTI-AR-2023/002



InfraScience Research Activity Report 2023

InfraScience Lab.

ISTI-AR-2023/002

InfraScience is a research group of the National Research Council of Italy - Institute of Information Science and Technologies (CNR - ISTI) based in Pisa, Italy. This report documents the research activity performed by this group in 2023 to highlight the major results. In particular, the InfraScience group engaged in research challenges characterising Data Infrastructures, e-Science, and Intelligent Systems. The group activity is pursued by closely connecting research and development and by promoting and supporting open science. In fact, the group is leading the development of two large scale infrastructures for Open Science, i.e., D4Science and OpenAIRE. During 2023 InfraScience members contributed to the publishing of several papers, to the research and development activities of several research projects (primarily funded by EU), to the organization of conferences and training events, to several working groups and task forces.

Keywords: Infrastructure, Open Science, Intelligent system, Research report.

Citation

InfraScience Lab. *InfraScience Research Activity Report 2023*, ISTI Annual Reports 2023/002.

DOI: 10.32079/ISTI-AR-2023/002

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" Area della Ricerca CNR di Pisa

Via G. Moruzzi 1 56124, Pisa Italy

<http://www.isti.cnr.it>

InfraScience Research Activity Report 2023

Michele Artini¹, Massimiliano Assante¹, Claudio Atzori¹, Miriam Baglioni¹, Alessia Bardi¹, Catherine Bosio, Pasquale Bove¹, Antonio Calanducci¹, Leonardo Candela¹, Giovanni Casini¹, Donatella Castelli*¹, Roberto Cirillo¹, Gianpaolo Coro¹, Michele De Bonis¹, Franca Debole¹, Andrea Dell'Amico¹, Luca Frosini¹, Ahmed Salah Tawfik Ibrahim¹, Sandro La Bruzzo¹, Lucio Lelii¹, Paolo Manghi¹, Francesco Mangiacrapa¹, Dario Mangione¹, Andrea Mannocci¹, Elisa Molinaro¹, Pasquale Pagano¹, Giancarlo Panichi¹, Maria Teresa Paratore¹, Gina Pavone¹, Tommaso Piccioli¹, Fabio Sinibaldi¹, Umberto Straccia¹, Gian Luca Vannini¹

Abstract




InfraScience is a research group of the National Research Council of Italy - Institute of Information Science and Technologies (CNR - ISTI) based in Pisa, Italy. This report documents the research activity performed by this group in 2023 to highlight the major results. In particular, the InfraScience group engaged in research challenges characterising Data Infrastructures, eScience, and Intelligent Systems. The group activity is pursued by closely connecting research and development and by promoting and supporting open science. In fact, the group is leading the development of two large scale infrastructures for Open Science, i.e., D4Science and OpenAIRE. During 2023 InfraScience members contributed to the publishing of several papers, to the research and development activities of several research projects (primarily funded by EU), to the organization of conferences and training events, to several working groups and task forces.

Keywords

Infrastructure — Open Science — Intelligent Systems — Research Activity

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", Consiglio Nazionale delle Ricerche, Via G. Moruzzi 1, 56124, Pisa, Italy

*Corresponding author: donatella.castelli@isti.cnr.it

This work is under   

Contents

1	Introduction	1
2	Research topics	2
2.1	Data infrastructures	2
2.2	eScience	2
2.3	Intelligent Systems	3
3	Papers	3
3.1	Contributions to Journals	3
3.2	Contributions to Conferences	14
3.3	Technical Reports	17
3.4	Other contributions	18
4	Projects	18
5	Infrastructures and Services	23
6	Software	25
7	Datasets	26
8	Organised Events	26
9	Training Activities	27
10	Working Groups, Task Forces, & Interest Groups	27
11	Conclusion	29

1. Introduction

Science today is characterized by its data and compute-intensive nature, AI assistance, participatory methods, and multi-disciplinary approaches. The practices of sharing and publishing scientific results are undergoing significant reconsideration to enhance openness, transparency, and reproducibility, while also providing incentives for scientists to disseminate their findings beyond traditional scientific articles. These evolving practices reflect a profound transformation in scientific methodologies, driven by and necessitating continuous innovation in IT tools and approaches.

InfraScience¹ is a research group working to contribute to this evolution by investigating, experimenting, and closely connecting research and development of innovative digital infrastructures, information systems, and smart solutions for fostering and empowering data-centered research. InfraScience is a research group of the National Research Council of Italy - Institute of Information Science and Technologies (CNR - ISTI)² based in Pisa, Italy. It consists of 27 members: 21 research staff and 6 technical staff. Moreover, it counts on

¹InfraScience website infrascience.isti.cnr.it

²Institute of Information Science and Technologies website www.isti.cnr.it

16 collaborators including postdocs, doctoral students, and research associates.

This report documents the research activity performed by the group in 2023, the resulting publications, the active research projects, and the services and infrastructures operated. In particular, Sec. 2 describes the topics characterising InfraScience research. Sec. 3 reports on the publications produced by the group. Sec. 4 documents the research projects InfraScience contributed to. Sec. 5 describes the major developments of the two infrastructures the team is responsible for. Sec. 6 reports on the software artefacts released by InfraScience. Sec. 7 describes the datasets released by InfraScience. Sec. 8 reports on the organised events. Sec. 9 details the training activity performed by InfraScience. Sec. 10 documents the working groups and task forces InfraScience members participate in. Finally, Sec. 11 concludes the report and gives prospects on future research activities.

2. Research topics

The research activities conducted by infraScience members revolve around three major topics: Data Infrastructures, eScience, and Intelligent Systems.

2.1 Data infrastructures

This research area encompasses a broad spectrum of models, approaches, and solutions that underpin the development and operation of data infrastructures suitable for both thematic and interdisciplinary scientific contexts. These contexts are often characterized by high variability, heterogeneity, reusability, and the presence of “big data,” requiring sophisticated methods for data integration, curation, and analysis. Addressing these complexities involves designing scalable architectures, adopting robust metadata standards, and implementing interoperable frameworks that facilitate seamless data exchange and collaboration across diverse research domains.

To address these challenges, the group is actively engaged in bridging the gap between research and development, ensuring that cutting-edge scientific methodologies are effectively translated into operational digital infrastructures. This involves fostering synergies between domain experts, data scientists, and software engineers to create solutions that are both scientifically rigorous and technologically sustainable. The group also prioritizes the development of governance models, policies, and best practices that enhance the usability, accessibility, and long-term sustainability of these infrastructures.

In fact, InfraScience is responsible for developing two significant large-scale infrastructures that support open science: D4Science and OpenAIRE (cf. Sec. 5).

By leading these initiatives, InfraScience plays a crucial role in shaping the future of open science, providing essential technological foundations for scientific collaboration, knowledge dissemination, and responsible data stewardship in the digital era.

The major themes and investigations include developments of the D4Science Infrastructure [24], development of a service for publishing compound objects with spatial and temporal extent [2], development of solutions for the ARIADNE infrastructure serving the archaeological research community [15], development of solutions for the IPERION HS research infrastructure supporting researchers in the field of heritage science [14], experiences in developing scholarly graphs [43], and development of author name disambiguation approaches in scholarly knowledge graphs [34].

2.2 eScience

This research domain encompasses a wide range of models, approaches, and solutions for conducting collaborative, data-driven, and reproducible analytical workflows. By integrating computational methods, statistical techniques, and digital infrastructure, it enables researchers to process, analyze, and interpret complex datasets efficiently. These workflows promote automation, interoperability, and scalability, ensuring that scientific investigations can be conducted transparently and with minimal barriers to collaboration.

Furthermore, this domain supports the sharing, publication, validation, and monitoring (usage and impact) of resulting scientific outcomes, including publications, datasets, software, and other related materials. It facilitates the adoption of FAIR (Findable, Accessible, Interoperable, and Reusable) principles, ensuring that scientific outputs are structured, well-documented, and easily discoverable by the global research community. Through the use of persistent identifiers, metadata standards, and open-access repositories, researchers can securely store and disseminate their findings, enhancing reproducibility and fostering new collaborations.

Additionally, mechanisms for impact assessment, citation tracking, and version control enable continuous monitoring of scientific contributions, allowing researchers and institutions to evaluate the reach and influence of their work. This also supports the broader Open Science movement, promoting inclusivity, transparency, and equitable access to knowledge across disciplines and geographical boundaries. Ultimately, by advancing digital tools and methodological frameworks, this research domain enhances scientific innovation, accelerates discovery, and contributes to addressing pressing global challenges.

The group studies and proposed approaches for several challenges belonging to the domain including development of solutions for FAIR data management contributing to the development of an open science graph [51], development of a collection of environmental, geophysical, and other marine-related data for marine ecological models and ecological-niche models [32], workflows to produce story maps from textual documents containing territory data [19], studies to examine data discovery requirements [49], and workflows for 2D marine environmental parameters [30].

2.3 Intelligent Systems

This research area explores advanced artificial intelligence (AI)-driven methodologies and computational approaches that facilitate the seamless discovery, access, processing, and interpretation of diverse forms of information. It encompasses techniques that enable both humans and automated systems to effectively interact with structured data, such as databases and spreadsheets, as well as unstructured data, including text, images, videos, and sensor-generated content. By leveraging machine learning, natural language processing, deep learning, and knowledge representation, these AI-powered solutions enhance decision-making, pattern recognition, and predictive capabilities across various domains. Furthermore, this field addresses challenges related to data integration, retrieval efficiency, semantic understanding, and adaptive learning, ultimately fostering more intelligent and autonomous information ecosystems.

InfraScience studied and proposed approaches for challenges including development of a self-training automatic infant-cry detector approach [31], studies on situated conditional reasoning [27], approaches for integrating Rational Closure within the triple language RDFS [26], workflows estimating potential unreported fishing activity hotspots in a marine area [33].

3. Papers

The following papers have been published by InfraScience members in collaboration with researchers from several Institutions and scientific disciplines. In particular, InfraScience contributed 25 articles in journals, 10 papers to conferences and workshops, and 5 publications including technical reports and other papers.

3.1 Contributions to Journals

InfraScience members contributed to the following papers published in journals.

A self-training automatic infant-cry detector [31] by Coro et al. for *Neural computing & applications*.

Summary: Infant cry is one of the first distinctive and informative life signals observed after birth. Neonatologists and automatic assistive systems can analyse infant cry to early-detect pathologies. These analyses extensively use reference expert-curated databases containing annotated infant-cry audio samples. However, these databases are not publicly accessible because of their sensitive data. Moreover, the recorded data can under-represent specific phenomena or the operational conditions required by other medical teams. Additionally, building these databases requires significant investments that few hospitals can afford. This paper describes an open-source workflow for infant-cry detection, which identifies audio segments containing high-quality infant-cry samples with no other overlapping audio events (e.g., machine noise or adult speech). It requires minimal training because it trains an LSTM-with-self-attention model on infant-cry samples automatically detected from the recorded audio through cluster

analysis and HMM classification. The audio signal processing uses energy and intonation acoustic features from 100-ms segments to improve spectral robustness to noise. The workflow annotates the input audio with intervals containing infant-cry samples suited for populating a database for neonatological and early diagnosis studies. On 16 min of hospital phone-audio recordings, it reached sufficient infant-cry detection accuracy in 3 neonatal care environments (nursery–69%, sub-intensive–82%, intensive–77%) involving 20 infants subject to heterogeneous cry stimuli, and had substantial agreement with an expert’s annotation. Our workflow is a cost-effective solution, particularly suited for a sub-intensive care environment, scalable to monitor from one to many infants. It allows a hospital to build and populate an extensive high-quality infant-cry database with a minimal investment.

Fig. 1 depicts the proposed workflow constituted by a sequence of five computational modules: (i) the “signal segmentation” module, which divides the signal into smaller units with acoustic characteristics that indicate the potential presence of infant cry; (ii) the “energy and pitch extraction” module, which estimates energy and intonation features to robustly represent infant-cry-related audio characteristics in noisy operational conditions; (iii) the “cluster analysis” module, which optimally clusters features to potentially distinguish between infant cry and other audio types; (iv) the “infant-cry cluster identification” module, which identifies the clusters that probably contain infant cry; (v) the “infant-cry detector” modules, which first train a complex machine learning model on the cry and non-cry cluster samples and then annotate the original audio file. A final module (“consecutive segment merging”) merges consecutive audio segments containing infant cry.

Global-scale parameters for ecological models [32] by Coro et al. for *Scientific data*.

Summary: This paper presents a collection of environmental, geophysical, and other marine-related data for marine ecological models and ecological-niche models. It consists of 2132 raster data for 58 distinct parameters at regional and global scales in the ESRI-GRID ASCII format. Most data originally belonged to open data owned by the authors of this article but residing on heterogeneous repositories with different formats and resolutions. Other data were specifically created for the present publication. The collection includes 565 data with global scale range; 154 at 0.5° resolution and 411 at 0.1° resolution; 196 data with annual temporal aggregation over 10 key years between 1950 and 2100; 369 data with monthly aggregation at 0.1° resolution from January 2017 to May 2021 continuously. Data were also cut out on 8 European marine regions. The collection also includes forecasts for different future scenarios such as the Representative Concentration Pathways 2.6 (63 data), 4.5 (162 data), and 8.5 (162 data), and the A2 scenario of the Intergovernmental Panel on Climate Change (180 data).

Fig. 1 summarises the implemented workflow. The authors harmonised and standardised geospatial data from their

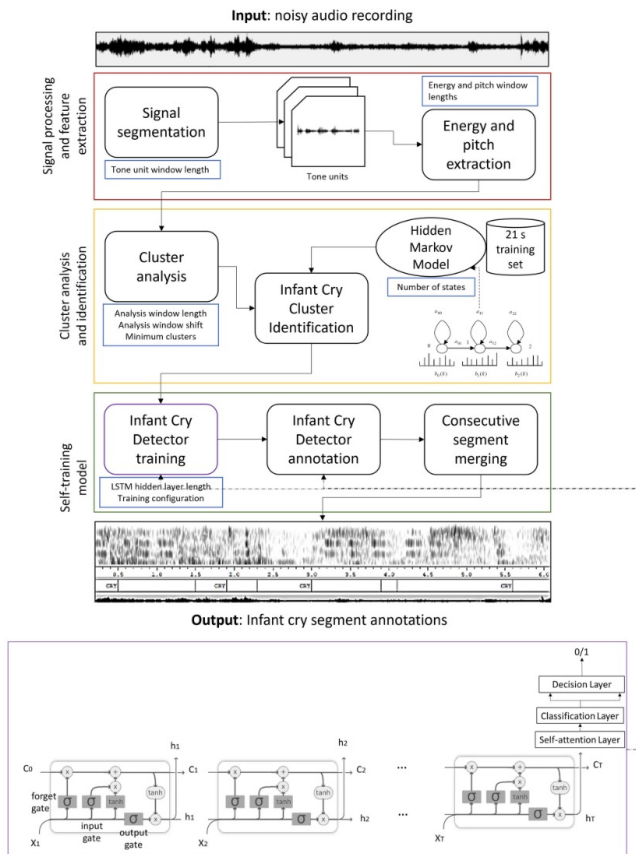


Figure 1. Infant-cry workflow proposed by [31]

own heterogeneous resources and publications that had newly produced or re-processed these data. Some data were previously available in custom formats (e.g., CSV or text files), which meant they were not as accessible as they could be. Additionally, they specifically produced other data to complement the collection. The primary sources involved were (i) environmental data produced for the AquaMaps ecological niche models, (ii) data from the Italian National Research Council (CNR) studies on marine science, Earth science, and epidemics that re-processed or newly produced open-access data based on other sources, and (iii) data produced by the Quantitative Aquatics (Q-quatics) non-governmental organisation for ecosystem and ecological models.

From unstructured texts to semantic story maps [19] by Bartalesi et al. for the International journal of digital earth.

Summary: Digital maps greatly support storytelling about territories, especially when enriched with data describing cultural, societal, and ecological aspects, conveying emotional messages that describe the territory as a whole. Story maps are interactive online digital narratives that can describe a territory beyond its map by enriching the map with text, pictures, videos, and other multimedia information. This paper presents a semi-automatic workflow to produce story maps from textual documents containing territory data. An expert first assembles one territory-contextual document containing

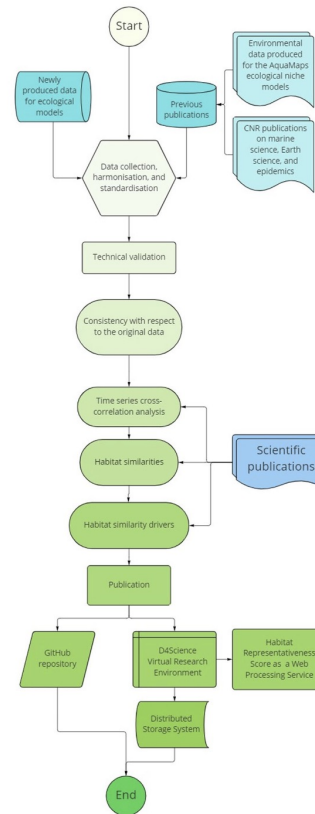


Figure 2. Conceptual flowchart of the data harmonisation, validation, and publication workflow implemented by [32]

text and images. Then, automatic processes use natural language processing and Wikidata services to (i) extract key concepts (entities) and geospatial coordinates associated with the territory, (ii) assemble a logically-ordered sequence of enriched story-map events, and (iii) openly publish online story maps and an interoperable Linked Open Data semantic knowledge base for event exploration and inter-story correlation analyses. Our workflow uses an Open Science-oriented methodology to publish all processes and data. Through our workflow, we produced story maps for the value chains and territories of 23 rural European areas of 16 countries. Through numerical evaluation, we demonstrated that territory experts considered the story maps effective in describing their territories, and appropriate for communicating with citizens and stakeholders.

Fig. 3 depicts the architectural schema of the proposed workflow for semi-automatically creating story maps.

A simple framework for the exploration of functional biodiversity [37] by Froese et al. for Cybium.

Summary: Key traits of functional biodiversity are examined for 31,134 species of fishes. These traits are maximum body weight, productivity, and trophic level. A new, simple framework is presented that shows the combined usage of these traits, in ordinal categories, for close to 90% of

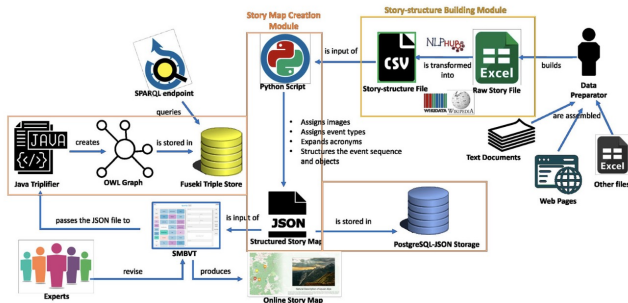


Figure 3. The architectural schema of the workflow for semi-automatically creating story maps by [19]

extant species of fishes. Most species are clustered tightly along an evolutionary axis in size-productivity-trophic space (SPT-space) from few large, evolutionary old species with very low productivity to many medium-sized newly evolved species with high productivity, superseding Cope’s rule of a within-lineages trend towards larger size and lower productivity. The across-lineages evolutionary axis is also found in the subsets of marine, freshwater, and Arctic species. Another notable prediction is the five-fold increase in top predators in Arctic waters in 2100, which could cause the extinction of endemic species. The main purpose of this study is to demonstrate the usefulness of the SPT-framework for comparing functional biodiversity patterns in ecosystems by salinity, geography or time. Also, the SPT-framework was used to explore correlations with other traits such as body shape, and to display the position of individual species, represented by pictograms of body shape and habitat, within SPT-space.

Fig. 4 depicts the SPT-plot for all of the 31,134 species of fishes with available data suggesting that the most used strategies follow a diagonal axis from large fishes with very low productivity to small and medium-sized fishes with high productivity.

What are researchers’ needs in data discovery? Analysis and ranking of a large-scale collection of crowdsourced use cases [49] by Mathiak et al. for Data science journal.

Summary: Data discovery is important to facilitate data re-use. In order to help frame the development and improvement of data discovery tools, we collected a list of requirements and users’ wishes. This paper presents the analysis of these 101 use cases to examine data discovery requirements; these cases were collected between 2019 and 2020. We categorized the information across 12 ‘topics’ and eight types of users. While the availability of metadata was an expected topic of importance, users were also keen on receiving more information on data citation and a better overview of their field. We conducted and analysed a survey among data infrastructure specialists in a first attempt at ranking the requirements. Between these data professionals, these rankings were very different, excepting the availability of metadata and data quality assessment.

Fig. 5 depicts the strategy for ranking of documented use

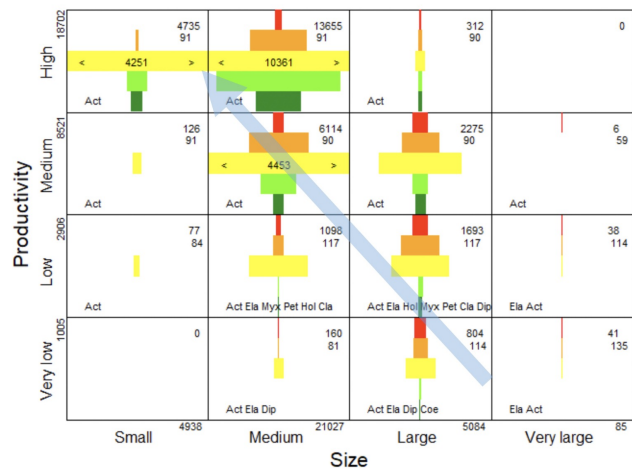


Figure 4. Functional biodiversity plot for 31,134 species of fishes with body weight, productivity and trophic level information in FishBase. The colours indicate the trophic groups from herbivore (dark green) to top predator (red). The width of the bars reflects the number of species in a trophic group, with cell width representing 2,000 species. Larger numbers per group are indicated by a number within pointy brackets. The numbers in the upper-right corners indicate the number of species assigned to a cell and the median evolutionary age (in millions of years) of the Orders the species belong to. The numbers left of the rows and below the columns indicate the sum of the respective species. The blue arrow indicates an evolutionary axis from few large and old species with low or very low productivity to many small and medium-sized recently evolved species with high productivity. [37]

cases. The authors pooled the two high and low ranking answers into one category each, and plotted the number of times each cluster was placed into either the high or low categories. Nearly all use case categories were designated at least once in the low and high ranking groups respectively (apart from the overview category). There was a relatively high consensus for placing Discoverability and Metadata for quality assessment high in the ranking list, while searching for data in a dataset or looking for specific researchers were both given a low ranking with the majority of the audience.

Exploiting the haptic and audio channels to improve orientation and mobility apps for the visually impaired [52] by Paratore and Leporini for Universal Access in the Information Society.

Summary: Orientation and mobility apps for visually impaired people are well known to be effective in improving the quality of life for this target group. A mobile application that guides a visually impaired person step-by-step through a physical space is a valuable aid, but it does not provide an overview of a complex environment “at a glance,” as a traditional hard-copy tactile map does. The aim of this study is to investigate whether a smartphone GPS map, enriched with

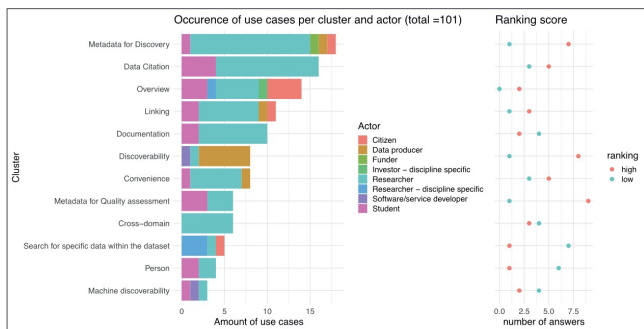


Figure 5. Summary of use case distribution (left) and ranking (right). Distribution is plotted depending on use case categories (cluster) and actor. Six use cases were reported for both researchers and students, and some use cases proposed for students may be relevant for researchers. The ranking of each category was assessed in a survey with 25 answers. [49]

haptic and audio hints, can facilitate cognitive mapping for visually impaired users. Encouraged by a preliminary study conducted in co-operation with two visually impaired volunteers, we designed and developed an Android prototype for exploration of an urban area. Our goal was to provide an affordable, portable and versatile solution to help users increase awareness of an environment through the positions of its landmarks and points of interest. Vibro-tactile and audio hints were linked to the coordinates on the map via the GeoJSON data format and were issued exploiting the text-to-speech and vibration features of the mobile device, as they were displayed through the operating system’s APIs. Test sessions and interviews with visually impaired users produced encouraging results. Results, to be verified by more extensive testing, overall confirm the validity of our approach and are in line with results found in the literature.

Fig. 6 shows a screenshot of the prototype as it was at the end of the three sessions, in which different POI categories are highlighted.

An exploratory approach to data driven knowledge creation [56] by Thanos et al. for the Journal of Big Data.

Summary: This paper describes a new approach to knowledge creation that is instrumental for the emerging paradigm of data-intensive science. The proposed approach enables the acquisition of new insights from the data by exploiting existing relationships between diverse types of datasets acquired through various modalities. The value of data consistently improves when it can be linked to other data because linking multiple types of datasets allows creating novel data patterns within a scientific data space. These patterns enable the exploratory data analysis, an analysis strategy that emphasizes incremental and adaptive access to the datasets constituting a scientific data space while maintaining an open mind to alternative possibilities of data interconnectivity. A technology, the Linked Open data (LOD), was developed to

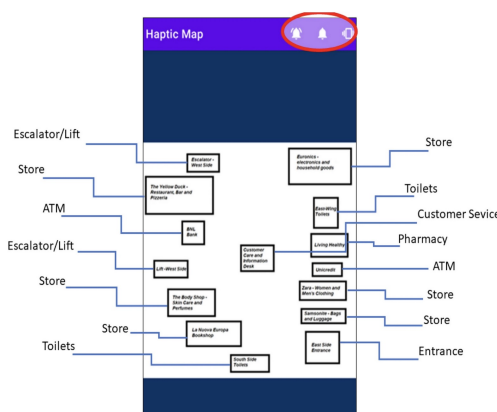


Figure 6. The prototype used for testing; different categories of POIs are highlighted, as well as buttons to toggle interaction modality [52]

enable the linking of datasets. We argue that the LOD technology presents several limitations that prevent the full exploitation of this technology to acquire new insights. In this paper, we outline a new approach that enables researchers to dynamically create data patterns in a research data space by exploiting explicit and implicit/hidden relationships between distributed research datasets. This dynamic creation of data patterns enables the exploratory data analysis strategy.

D4SCIENCE: a unique infrastructure delivering virtual research environments as a service [23] by Candela et al. for ERCIM News, Special theme: Data Infrastructures and Management.

Summary: Nowadays, research challenges – often based on the collaborative analysis of a large amount of data – require suitable infrastructures and user-facing solutions promoting multidisciplinary collaboration and appropriate communication and sharing of data, processes, and outcomes. The D4Science infrastructure and its virtual research environments proved to be a viable and effective solution for many communities of practice and use cases.

Fig. 7 shows the constantly growing trend of the D4Science user base in the period Jan 2019 - Feb 2023.

Estimating hidden fishing activity hotspots from vessel transmitted data [33] by Coro et al. for Frontiers in sustainable food systems.

Summary: Monitoring fishery activity is essential for resource planning and guaranteeing fisheries sustainability. Large fishing vessels constantly and continuously communicate their positions via Automatic Identification System (AIS) or Vessel Monitoring Systems (VMSs). These systems can use radio or Global Positioning System (GPS) devices to transmit data. Processing and integrating these big data with other fisheries data allows for exploring the relations between socioeconomic and ecosystem assets in marine areas, which is fundamental in fishery monitoring. In this context, estimating actual fishing activity from time series of AIS and VMS

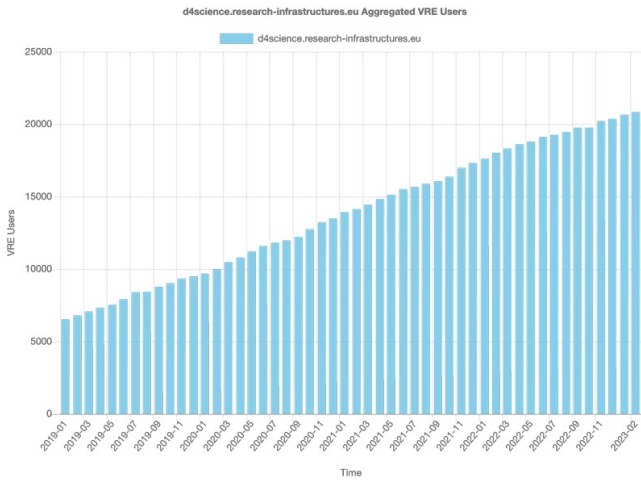


Figure 7. D4Science Users Jan 2019 - Feb 2023 [23]

data would enhance the correct identification of fishing activity patterns and help assess regulations’ effectiveness. However, these data might contain gaps because of technical issues such as limited coverage of the terrestrial receivers or saturated transmission bands. Other sources of data gaps are adverse meteorological conditions and voluntary switch-offs. Gaps may also include hidden (unreported) fishing activity whose quantification would improve actual fishing activity estimation. This paper presents a workflow for AIS/VMS big-data analysis that estimates potential unreported fishing activity hotspots in a marine area. The workflow uses a statistical spatial analysis over vessel speeds and coordinates and a multi-source data integration approach that can work on multiple areas and multiple analysis scales. Specifically, it (i) estimates fishing activity locations and rebuilds data gaps, (ii) estimates the potential unreported fishing hour distribution and the unreported-over-total ratio of fishing hours at a 0.01° spatial resolution, (iii) identifies potential unreported fishing activity hotspots, (iv) extracts the stocks involved in these hotspots (using global-scale repositories of stock and species observation data) and raises an alert about their possible endangered, threatened, and protected (ETP) status. The workflow is also a free-to-use Web Service running on an open science-compliant cloud computing platform with a Web Processing Service (WPS) standard interface, allowing efficient big data processing. As a study case, we focussed on the Adriatic Sea. We reconstructed the monthly reported and potential unreported trawling activity in 2019, using terrestrial AIS data with a 5-min sampling period, containing 50 million records transmitted by 1,600 vessels. The results highlight that the unreported fishing activity hotspots especially impacted Italian coasts and some forbidden and protected areas. The potential unreported activity involved 33 stocks, four of which were ETP species in the basin. The extracted information agreed with expert studies, and the estimated trawling patterns agreed with those produced by the Global Fishing Watch.

Fig. 8 depicts the proposed workflow that was entirely developed in R and released with the paper.

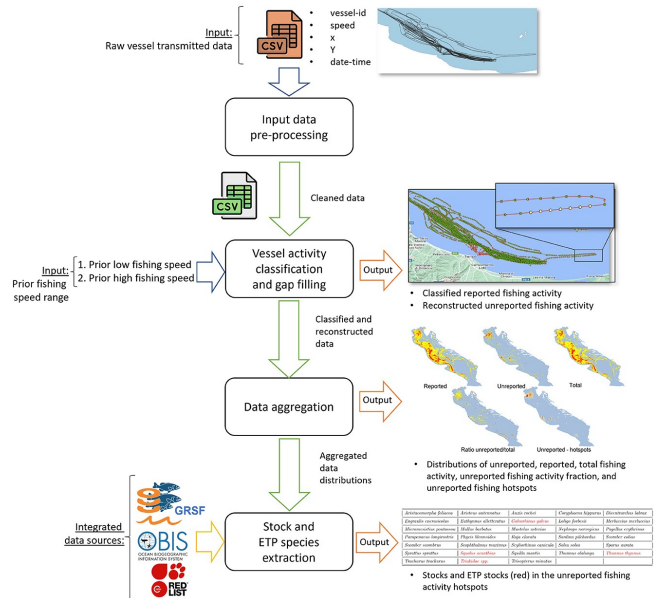


Figure 8. Conceptual schema of the workflow proposed in [33]

Situated conditional reasoning [27] by Casini et al. for Artificial intelligence.

Summary: Conditionals are useful for modelling many forms of everyday human reasoning but are not always sufficiently expressive to represent the information we want to reason about. In this paper, we make a case for a form of situated conditional. By ‘situated’, we mean that there is a context, based on an agent’s beliefs and expectations, that works as background information in evaluating a conditional, and we allow such a context to vary. These conditionals are able to distinguish, for example, between expectations and counterfactuals. Formally, they are shown to generalise the conditional setting in the style of Kraus, Lehmann, and Magidor. We show that situated conditionals can be described in terms of a set of rationality postulates. We then propose an intuitive semantics for these conditionals and present a representation result which shows that our semantic construction corresponds exactly to the description in terms of postulates. With the semantics in place, we define a form of entailment for situated conditional knowledge bases, which we refer to as minimal closure. Finally, we proceed to show that it is possible to reduce the computation of minimal closure to a series of propositional entailment and satisfiability checks. While this is also the case for rational closure, it is somewhat surprising that the result carries over to minimal closure.

Defeasible RDFS via rational closure [26] by Casini and Straccia for Information Science.

Summary: In the field of non-monotonic logics, the notion of Rational Closure (RC) is acknowledged as a notable

approach. In recent years, RC has gained popularity in the context of Description Logics (DLs), the logic underpinning the standard semantic Web Ontology Language OWL 2, whose main ingredients are classes, the relationship among classes and roles, which are used to describe the properties of classes. In this work, we show instead how to integrate RC within the triple language RDFS (Resource Description Framework Schema), which together with OWL 2 is a major standard semantic web ontology language. To do so, we start from pdf , a minimal, but significant RDFS fragment that covers the essential features of RDFS, and then extend it to pdf_{\perp} , allowing to state that two entities are incompatible/disjoint with each other. Eventually, we propose defeasible pdf_{\perp} via a typical RC construction allowing to state default class/property inclusions. Furthermore, to overcome the main weaknesses of RC in our context, i.e., the “drowning problem” (viz. the “inheritance blocking problem”), we further extend our construction by leveraging Defeasible Inheritance Networks (DIN) defining a new non-monotonic inference relation that combines the advantages of both (RC and DIN). To the best of our knowledge this is the first time of such an attempt. In summary, the main features of our approach are: (i) the defeasible pdf_{\perp} we propose here remains syntactically a triple language by extending it with new predicate symbols with specific semantics; (ii) the logic is defined in such a way that any RDFS reasoner/store may handle the new predicates as ordinary terms if it does not want to take account of the extra non-monotonic capabilities; (iii) the defeasible entailment decision procedure is built on top of the pdf_{\perp} entailment decision procedure, which in turn is an extension of the one for pdf via some additional inference rules favouring a potential implementation; (iv) the computational complexity of deciding entailment in pdf and pdf_{\perp} are the same; and (v) defeasible entailment can be decided via a polynomial number of calls to an oracle deciding ground triple entailment in pdf_{\perp} and, in particular, deciding defeasible entailment can be done in polynomial time.

Missing Plant Detection in Vineyards Using UAV Angled RGB Imagery Acquired in Dormant Period [36] by Di Gennaro et al. for Drones.

Summary: Since 2010, more and more farmers have been using remote sensing data from unmanned aerial vehicles, which have a high spatial-temporal resolution, to determine the status of their crops and how their fields change. Imaging sensors, such as multispectral and RGB cameras, are the most widely used tool in vineyards to characterize the vegetative development of the canopy and detect the presence of missing vines along the rows. In this study, the authors propose different approaches to identify and locate each vine within a commercial vineyard using angled RGB images acquired during winter in the dormant period (without canopy leaves), thus minimizing any disturbance to the agronomic practices commonly conducted in the vegetative period. Using a combination of photogrammetric techniques and spatial analysis tools, a workflow was developed to extract each post and vine

trunk from a dense point cloud and then assess the number and position of missing vines with high precision. In order to correctly identify the vines and missing vines, the performance of four methods was evaluated, and the best performing one achieved 95.10% precision and 92.72% overall accuracy. The results confirm that the methodology developed represents an effective support in the decision-making processes for the correct management of missing vines, which is essential for preserving a vineyard’s productive capacity and, more importantly, to ensure the farmer’s economic return.

Fig. 9 illustrates the workflow used in this study to identify the vines and missing vines from the UAV RGB imagery acquisition.

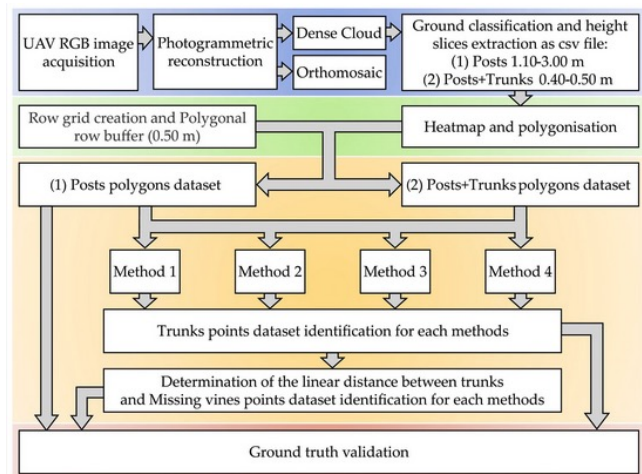


Figure 9. Scheme of the processing workflow used to identify vines and missing vines within the vineyard. The boxes show, in sequence, the functional steps of image processing for the extraction of dense cloud regions classified in height, the generation of polygon shapefiles from the point clouds, the application of the 4 methods and, finally, the validation of the results. [36]

The D4Science experience on virtual research environment development [24] by Candela et al. for Computing in science & engineering.

Summary: Today, complex research challenges, often based on the analysis of a large amount of data, require multidisciplinary collaboration and appropriate communication and sharing of data, processes and outcomes. Technologies and large-scale infrastructures provide stakeholders with computing capacity and data services to perform unprecedented levels of data-driven scientific activities. This opens the way to science gateways and virtual research environments supporting researchers in scientific and educational activities. This article describes our extensive experience with the Virtual Research Environments (VRE) operated by the D4Science infrastructure. It presents how this infrastructure supports their development, their basic functionalities and how they are easily customised to serve the needs of specific user communi-

ties. It also describes how they are used in real contexts. The article concludes by reporting how VREs are now progressively used as valuable instruments to support open science and how this role might become more relevant in the future.

Fig. 10 depicts the service-oriented view of the D4Science infrastructure architecture.

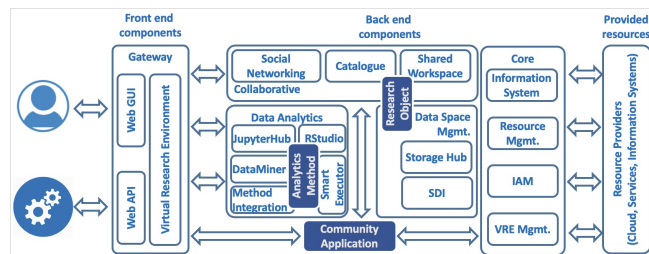


Figure 10. The D4Science Service-oriented Architecture. [24]

A Novel Curated Scholarly Graph Connecting Textual and Data Publications [43] by Irrera et al. for ACM journal of data and information quality.

Summary: In the last decade, scholarly graphs became fundamental to storing and managing scholarly knowledge in a structured and machine-readable way. Methods and tools for discovery and impact assessment of science rely on such graphs and their quality to serve scientists, policymakers, and publishers. Since research data became very important in scholarly communication, scholarly graphs started including dataset metadata and their relationships to publications. Such graphs are the foundations for Open Science investigations, data-article publishing workflows, discovery, and assessment indicators. However, due to the heterogeneity of practices (FAIRness is indeed in the making), they often lack the complete and reliable metadata necessary to perform accurate data analysis; e.g., dataset metadata is inaccurate, author names are not uniform, and the semantics of the relationships is unknown, ambiguous or incomplete. This work describes an open and curated scholarly graph we built and published as a training and test set for data discovery, data connection, author disambiguation, and link prediction tasks. Overall the graph contains 4,047 publications, 5,488 datasets, 22 software, 21,561 authors; 9,692 edges interconnect publications to datasets and software and are labeled with semantics that outline whether a publication is citing, referencing, documenting, supplementing another product. To ensure high-quality metadata and semantics, we relied on the information extracted from PDFs of the publications and the datasets and software webpages to curate and enrich nodes metadata and edges semantics. To the best of our knowledge, this is the first ever published resource, including publications and datasets with manually validated and curated metadata.

Fig. 11 reports the five-phases curation pipeline adopted for the work

ARIADNE Plus e il D4GNA-Dataset per il Geoportale

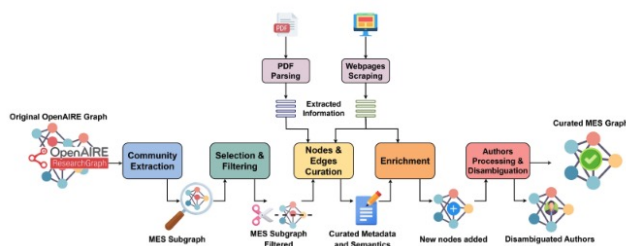


Figure 11. Curation pipeline. The output of each phase is the input of the subsequent phase. There are five main phases: (i) Community Extraction, (ii) Selection and Filtering, (iii) Nodes and Edges Curation, (iv) Enrichment, (v) Authors Processing and Disambiguation. The input of the entire pipeline is the OAG, while the final output is the curated MES graph. [43]

Nazionale per l’Archeologia [2] by Acconcia et al. for Digitalia.

Summary: The article describes the experience of the D4GNA - Dataset for the National Geoportal for Archaeology born within the framework of the ARIADNEplus (Advanced Research Infrastructure for Archaeological Dataset Networking in Europe, plus) project, which ended last 31 December. The contribution starts from the broader context of the European project to delve into the world of Italian archaeological data; the technological solution, standardization, dematerialisation and network sharing of data are the topics touched upon in this journey that illustrates the progress towards a virtuous objective: the Geoportale Nazionale per l’Archeologia, GNA. The GNA, created by the Istituto Centrale per l’Archeologia (ICA), released online July 10, 2023, is the national access point for receiving and consulting both archaeological interventions carried out under the scientific direction of the Ministry of Culture (MiC) as well as archaeological investigations conducted by universities and other research institutions.

Fig. 12 reports a screenshot of the Data-Viewer component.



Figure 12. GeoPortal Data-Viewer GUI [2]

ARIADNE: A data infrastructure for the archaeological research community [15] by Bardi et al. for ERCIM News,

Special theme: Data Infrastructures and Management.

Summary: The ARIADNE infrastructure provides tools and services for researchers to address archaeological grand challenges that require discovery and analysis of information scattered across different thematic and geographically distributed sources.

Fig. 13 reports a screenshot of the Data-Viewer component allowing any user to freely access and consult the documentation related to archaeological excavations in Italy, and authorised users to collaboratively manage these contents.

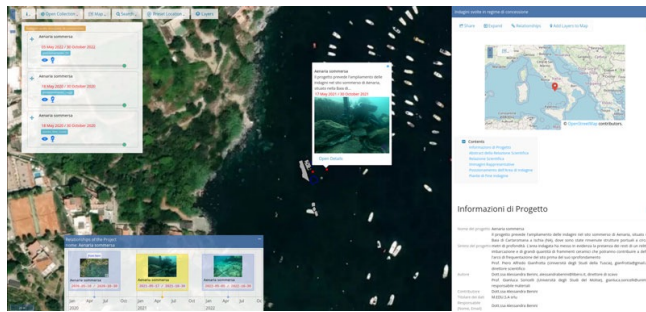


Figure 13. The geoport viewer of DGNA VRE [15]

Boosting Open Science in the IPERION HS research infrastructure with OpenAIRE [14] by Bardi and Benassi for ERCIM News, Special theme: Data Infrastructures and Management.

Summary: IPERION HS is a research infrastructure that supports researchers in the field of heritage science, an interdisciplinary domain studying cultural and natural tangible heritage. This article describes how IPERION HS makes the research outputs open and accessible to the community and monitors its impact thanks to the services offered by the OpenAIRE infrastructure.

Fig. 14 reports a screenshot of Open Research Gateway on Heritage Science where all research outputs of the domain are made discoverable, enabling searching and browsing across different sources.

Data management plans as linked open data: exploiting ARGOS FAIR and machine actionable outputs in the OpenAIRE research graph [51] by Papadopoulou et al. for Journal of biomedical semantics.

Summary: Open Science Graphs (OSGs) are scientific knowledge graphs representing different entities of the research lifecycle (e.g., projects, people, research outcomes, institutions) and the relationships among them. They present a contextualized view of current research that supports discovery, re-use, reproducibility, monitoring, transparency and omni-comprehensive assessment. A Data Management Plan (DMP) contains information concerning both the research processes and the data collected, generated and/or re-used during a project's lifetime. Automated solutions and workflows that connect DMPs with the actual data and other contextual information (e.g., publications, fundings) are missing

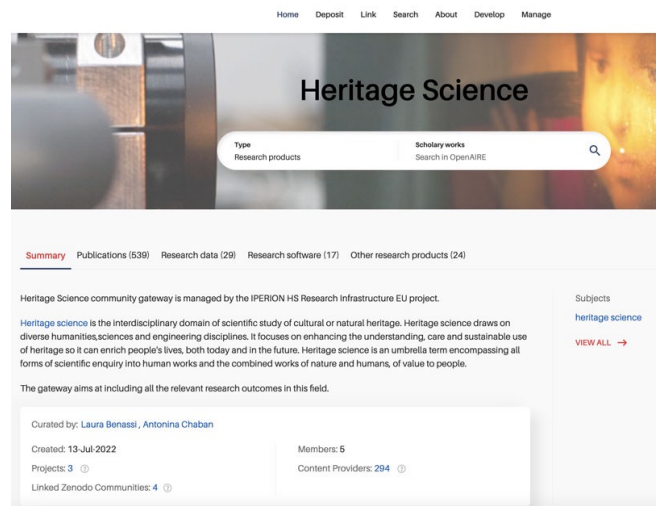


Figure 14. Home page of the Open Research Gateway on Heritage Science. [14]

from the landscape. DMPs being submitted as deliverables also limit their findability. In an open and FAIR-enabling research ecosystem information linking between research processes and research outputs is essential. ARGOS tool for FAIR data management contributes to the OpenAIRE Research Graph (RG) and utilises its underlying services and trusted sources to progressively automate validation and automations of Research Data Management (RDM) practices.

Fig. 15 reports the mappings between the ARGOS machine actionable DMP entities and the OpenAIRE Research Graph entities.

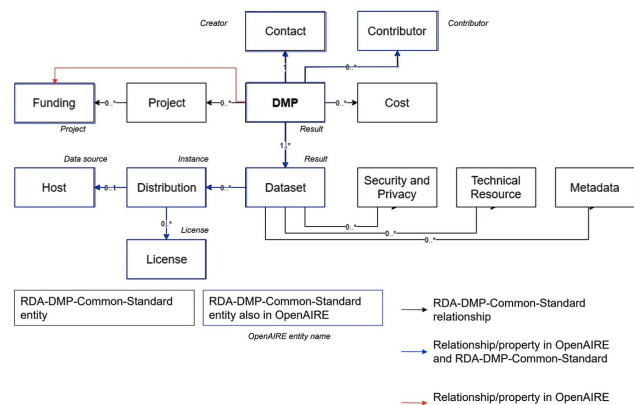


Figure 15. RDA standard and the OpenAIRE Research Graph Model. [51]

New developments in the analysis of catch time series as the basis for fish stock assessments: the CMSY++ method [38] by Froese et al. for Acta Ichthyologica et Piscatoria.

Summary: Following an introduction to the nature of fisheries catches and their information content, a new development of CMSY, a data-limited stock assessment method for

fishes and invertebrates, is presented. This new version, CMSY++, overcomes several of the deficiencies of CMSY, which itself improved upon the “Catch-MSY” method published by S. Martell and R. Froese in 2013. The catch-only application of CMSY++ uses a Bayesian implementation of a modified Schaefer model, which also allows the fitting of abundance indices should such information be available. In the absence of historical catch time series and abundance indices, CMSY++ depends strongly on the provision of appropriate and informative priors for plausible ranges of initial and final stock depletion. An Artificial Neural Network (ANN) now assists in selecting objective priors for relative stock size based on patterns in 400 catch time series used for training. Regarding the cross-validation of the ANN predictions, of the 400 real stocks used in the training of ANN, 94% of final relative biomass (B/k) Bayesian (BSM) estimates were within the approximate 95% confidence limits of the respective CMSY++ estimate. Also, the equilibrium catch-biomass relations of the modified Schaefer model are compared with those of alternative surplus-production and age-structured models, suggesting that the latter two can be strongly biased towards underestimating the biomass required to sustain catches at low abundance. Numerous independent applications demonstrate how CMSY++ can incorporate, in addition to the required catch time series, both abundance data and a wide variety of ancillary information. We stress, however, the caveats and pitfalls of naively using the built-in prior options, which should instead be evaluated case-by-case and ideally be replaced by independent prior knowledge.

Fig. 16 depicts examples of graphical output of CMSY++.

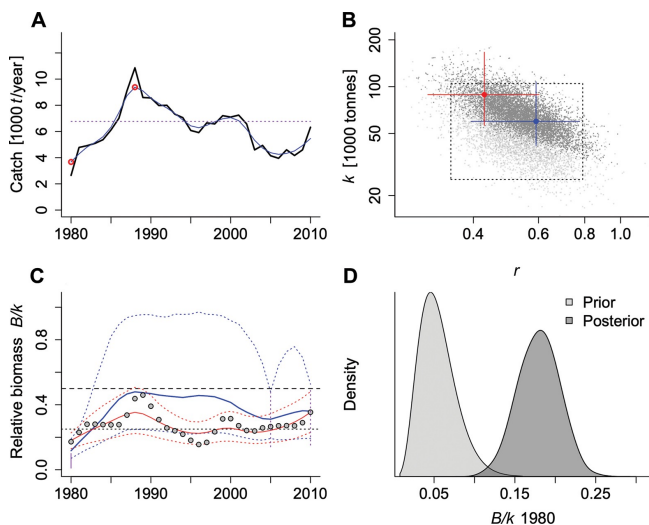


Figure 16. Examples of graphical output of CMSY++, here for European plaice (*Pleuronectes platessa*) in the eastern English Channel. [38]

Evaluation of operational ocean forecasting systems from the perspective of the users and the experts [29] by Ciliberti et al. for State of the Planet.

erti et al. for State of the Planet.

Summary: The Intergovernmental Oceanographic Commission (IOC) has an Ocean Decade Implementation Plan (UNESCO-IOC, 2021) that states seven outcomes required for the ocean we want, with the fourth outcome being “A predicted ocean where society understands and can respond to changing ocean conditions.” To facilitate the achievement of this goal, the IOC has endorsed Mercator Ocean International to implement the Decade Collaborative Center (DCC) for OceanPrediction³, which is a cross-cutting structure that will work to develop global-scale collaboration between Decade Actions related to ocean prediction. To have a predicted ocean, the OceanPrediction DCC understands that is critical to co-design ocean forecasting architecture that will permit different services to deliver as one and that could take advantage of the concept of digital twinning (European Union, 2022). This architecture will be designed to overcome the present-day limitations of our systems in terms of interoperability and tools sharing. This will translate into a new scenario for ocean forecasting, where more robust systems will be easier to implement thanks to a common set of agreed tools, standards, and best practices. This new architecture will serve as inspiration for the development targets of the different decadal actions related to ocean forecasting, such as ForeSea⁴, DITTO⁵, CoastPredict⁶, Global Environment Monitoring System for the Ocean and Coasts⁷, Ocean Best Practices⁸, and others. To develop this architecture, the OceanPrediction DCC has implemented the Ocean Forecasting Co-Design Team (OFCT), which is composed of 43 international experts on all of the different aspects of the ocean forecasting value chain. The first task of this group is to analyze the present status of ocean forecasting at a global level, in order to properly identify the existing gaps before moving into the design phase. One of the first steps in this process has been to explore the degree of satisfaction of both users and experts with respect to the existing ocean forecasting systems. This has been done by launching a series of surveys among the members of the OFCT and another one among the users of the forecasting services. This paper describes the findings derived from the analysis of these surveys. Section 2 introduces the surveys, while Sect. 3 presents the results. Section 4 establishes a discussion and identifies some conclusions as part of the outlook for future exploitation.

Fig. 17 depicts depicts the experts’ average rates and standard deviations per Operational Ocean Forecasting Systems property.

Scientific knowledge gaps on the biology of non-fish marine species across European Seas [1] by Abucay et al. for

³<https://www.mercator-ocean.eu/oceanprediction/>
⁴<https://oceanpredict.org/un-decade-of-ocean-science/foresea/>
⁵<https://ditto-oceandecade.org/>
⁶<https://www.coastpredict.org/>
⁷<https://www.unep.org/explore-topics/oceans-seas/what-we-do/ocean-and-coastal-observations>
⁸<https://www.oceanbestpractices.org/>

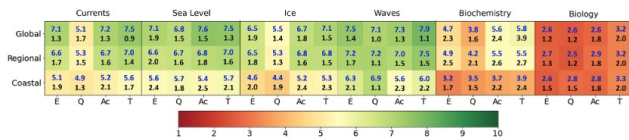


Figure 17. The experts’ average rates (in blue) and standard deviations (in black) per Oofs property (E is for existence, Q is for quality, Ac is for accessibility, and T is for timeliness), considering a proposed list of EOVs (e.g., currents, sea level, ice, waves, biochemistry, and biology) at a given scale (e.g., global, regional, and coastal). The color scale corresponds to mean values. [29]

Frontiers in Marine Science.

Summary: Available information and potential data gaps for non-fish marine organisms (cnidarians, crustaceans, echinoderms, molluscs, sponges, mammals, reptiles, and seabirds) covered by the global database SeaLifeBase were reviewed for eight marine ecosystems (Adriatic Sea, Aegean Sea, Baltic Sea, Bay of Biscay/Celtic Sea/Iberian Coast, Black Sea, North Sea, western Mediterranean Sea, Levantine Sea) across European Seas. The review of the SeaLifeBase dataset, which is based on published literature, analyzed information coverage for eight biological characteristics (diet, fecundity, maturity, length-weight relationships, spawning, growth, lifespan, and natural mortality). These characteristics are required for the development of ecosystem and ecological models to evaluate the status of marine resources and related fisheries. Our analyses revealed that information regarding these biological characteristics in the literature was far from complete across all studied areas. The level of available information was nonetheless reasonably good for sea turtles and moderate for marine mammals in some areas (Baltic Sea, Bay of Biscay/Celtic Sea/Iberian Coast, Black Sea, North Sea and western Mediterranean Sea). Further, seven of the areas have well-studied species in terms of information coverage for biological characteristics of some commercial species whereas threatened species are generally not well studied. Across areas, the most well-studied species are the cephalopod common cuttlefish (*Sepia officinalis*) and the crustacean Norway lobster (*Nephrops norvegicus*). Overall, the information gap is narrowest for length-weight relationships followed by growth and maturity, and widest for fecundity and natural mortality. Based on these insights, we provide recommendations to prioritize species with insufficient or missing biological data that are common across the studied marine ecosystems and to address data deficiencies.

Fig. 18 depicts the number of species by number of biological characteristics found for each.

Editorial: Ecocentric fisheries management in European seas: data gaps, base models and initial assessments, volume I [57] by Tsikliras et al. for Frontiers in Marine Science.

Summary: This is the editorial of the Research Topic “Ecocentric fisheries management in European seas: Data

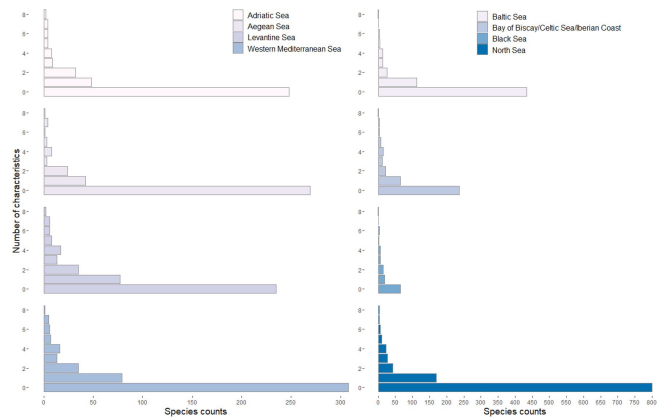


Figure 18. Species counts by number of biological characteristics (0 to 8) studied for non-fish marine organisms in the Adriatic Sea. [1]

gaps, base models and initial assessments”. This topic included six original articles, two reviews, and one policy and practice review. The six original articles focused on Mediterranean ecosystems, where data gaps and deficiencies are wider. Spatiotemporal and environmental modelling, that can be applied to areas with varying heterogeneity, was used to identify and fill gaps in trawl surveys in the Adriatic Sea. In the same area, a timely manuscript examined the (beneficial) effects of the COVID-19 lockdown related fishing restrictions and the involuntary fishing effort reduction on the status of target stocks. The impact of fisheries on ecosystem structure and functioning was examined in the Sea of Marmara over a period of thirty years, while data availability and the participatory approach for promoting fisheries sustainability were reported for the central and western Mediterranean Sea. Two trophic structure models were developed using the EwE modelling approach, one focusing on the oligotrophic ecosystem of the Balearic Islands in the western Mediterranean Sea and the other evaluating fisheries management policies in the alien-rich eastern Mediterranean Sea. An exhaustive review of all EwE models applied in European marine ecosystems stressed the importance of ecosystem modelling in studying the structure and dynamics of ecosystems and examining management and climate scenarios. Another article reviewed all available biological information on non-fish marine organisms and identified gaps in knowledge across European Seas. Finally, the policy and practice review included the global and European policies and implementing bodies which have a repercussion on the implementation of EBFM, highlighting specific stakeholder needs, which ecosystem models could help address.

Using semantic story maps to describe a territory beyond its map [21] by Bartalesi et al. for Semantic Web.

Summary: The paper presents the Story Map Building and Visualizing Tool (SMBVT) that allows users to create story maps within a collaborative environment and a usable Web interface. It is entirely open-source and published as

a free-to-use solution. It uses Semantic Web technologies in the back-end system to represent stories through a reference ontology for representing narratives. It builds up a user-shared semantic knowledge base that automatically interconnects all stories and seamlessly enables collaborative story building. Finally, it operates within an Open-Science oriented e-Infrastructure, which enables data and information sharing within communities of narrators, and adds multi-tenancy, multi-user, security, and access-control facilities. SMBVT represents narratives as a network of spatiotemporal events related by semantic relations and standardizes the event descriptions by assigning internationalized resource identifiers (IRIs) to the event components, i.e., the entities that take part in the event (e.g., persons, objects, places, concepts). The tool automatically saves the collected knowledge as a Web Ontology Language (OWL) graph and openly publishes it as Linked Open Data. This feature allows connecting the story events to other knowledge bases. To evaluate and demonstrate our tool, we used it to describe the Apuan Alps territory in Tuscany (Italy). Based on a user-test evaluation, we assessed the tool’s effectiveness at building story maps and the ability of the produced story to describe the territory beyond the map.

Fig. 19 depicts the story map building and visualising tool (SMBVT) interface for story map creation and gives a preview of a story map produced by the tool and a story timeline.

An Open Science oriented Bayesian interpolation model for marine parameter observations [30] by Coro for Environmental Modelling & Software.

Summary: Ecological and ecosystem modellers frequently need to interpolate spatiotemporal observations of geophysical and environmental parameters over an analysed area. However, particularly in marine science, modellers with low expertise in oceanography and hydrodynamics can hardly use interpolation methods optimally. This paper introduces an Open Science oriented, open-source, scalable and efficient workflow for 2D marine environmental parameters. It combines a fast, efficient interpolation method with a Bayesian hierarchical model embedding the stationary advection-diffusion equation as a constraint. Our workflow fills the usability gap between interpolation software providers and the users’ communities. It can run entirely automatically without requiring expert parametrisation. It is also available on a cloud computing platform, with a Web Processing Service compliant interface, supporting collaboration, repeatability, reproducibility, and provenance tracking. We demonstrate that our workflow produces comparable results to a state-of-the-art model (frequently used in oceanography) in interpolating four environmental parameters at the global scale.

Fig. 20 depicts the BIMAC workflow with data provisioning and pre-processing, prior parameter distribution calculation, and final interpolated distribution estimation.

Graph-based methods for Author Name Disambiguation: a survey [34] by De Bonis et al. for PeerJ Computer Science.

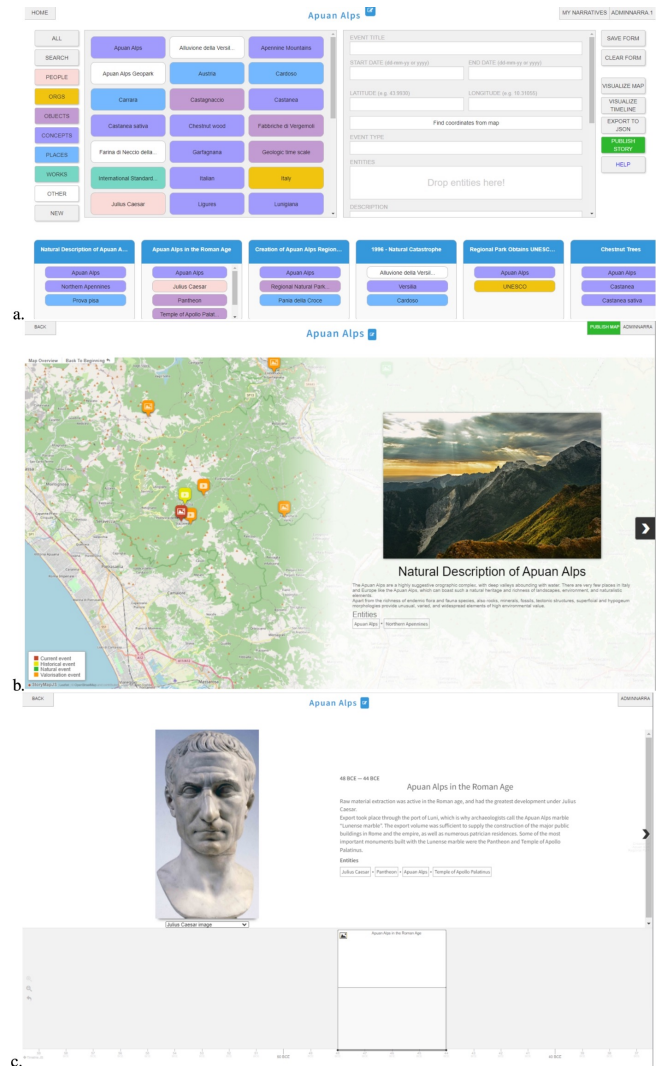


Figure 19. a. The story map building and visualising tool (SMBVT) interface for story map creation; b. A preview of a story map produced by our tool; c. A preview a story timeline produced by our tool. [21]

Summary: Scholarly knowledge graphs (SKG) are knowledge graphs representing research-related information, powering discovery and statistics about research impact and trends. Author name disambiguation (AND) is required to produce high-quality SKGs, as a disambiguated set of authors is fundamental to ensure a coherent view of researchers’ activity. Various issues, such as homonymy, scarcity of contextual information, and cardinality of the SKG, make simple name string matching insufficient or computationally complex. Many AND deep learning methods have been developed, and interesting surveys exist in the literature, comparing the approaches in terms of techniques, complexity, performance, etc. However, none of them specifically addresses AND methods in the context of SKGs, where the entity-relationship structure can be exploited. In this paper, we discuss recent graph-based methods for AND, define a framework through

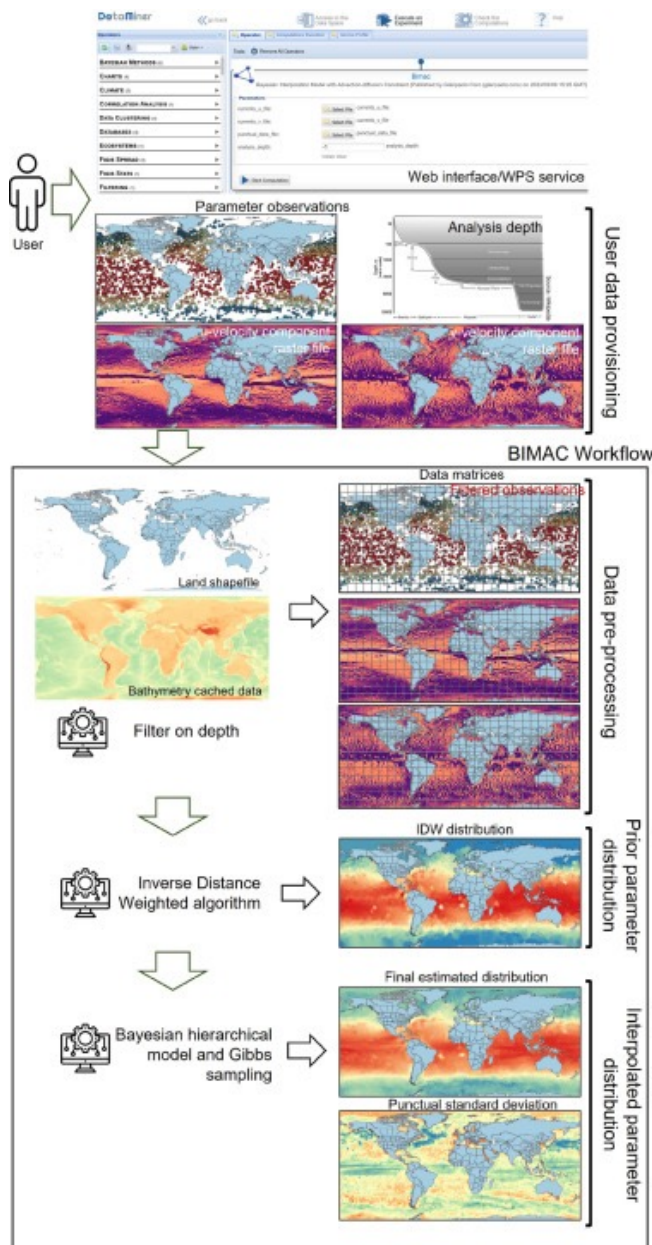


Figure 20. General schema of the proposed workflow. [30]

which such methods can be confronted, and catalog the most popular datasets and benchmarks used to test such methods. Finally, we outline possible directions for future work on this topic.

Fig. 21 depicts the taxonomy classifying AND approaches where each method can be characterized with respect to the learning strategy, the evidence explored, i.e., the type of information used to create the author representations, and the node representation strategy.

3.2 Contributions to Conferences

InfraScience members contributed to the following papers presented at international and national conferences.

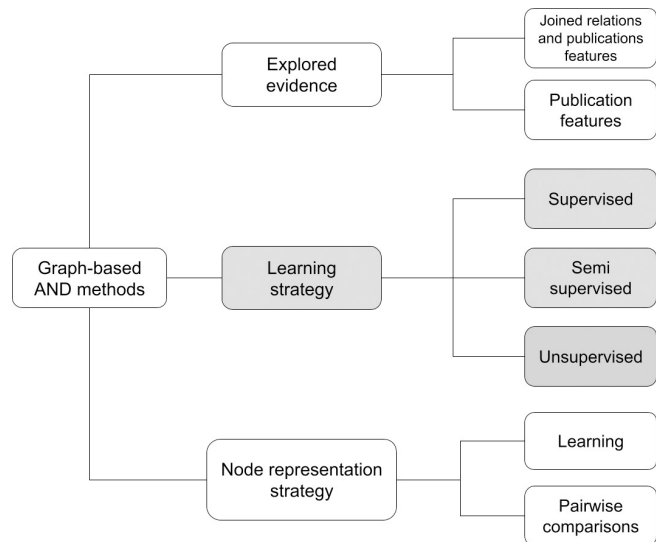


Figure 21. Proposed taxonomy for graph-based AND methods. [34]

Self-assess momentary mood in mobile devices: a case study with mature female participants [55] by Senette et al. for 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT-2023).

Summary: Starting from the assumption that mood has a central role in domain-specific persuasion systems for well-being, the main goal of this study was to investigate the feasibility and acceptability of single-input methods to assess momentary mood as a medium for further interventions in health-related mobile apps destined for mature women. To this aim, we designed a very simple android App providing four user interfaces, each one showing one interactive widget to self-assess mood. Two widgets report a hint about the momentary mood they represent; the last two do not have the hints but were previously refined through questionnaires administered to 63 women (age 45-65) in order to reduce their expressive ambiguity. Next, fifteen women (age 45-65 years) were recruited to use the app for 15 days. Participants were polled about their mood four times a day and data were saved in a remote database. Moreover, users were asked to fill out a preliminary questionnaire, at the first access to the app, and a feedback questionnaire at the end of the testing period. Results appear to prove the feasibility and acceptability of this approach to self-assess momentary mood in the target population and provides some potential input methods to be used in this context.

Fig. 22 depicts the proposed widgets.

Haptic-based cognitive mapping to support shopping malls exploration [53] by Paratore and Leporini for 8th EAI International Conference, GOODTECHS 2022.

Summary: This paper describes a study, which is currently underway, whose aim is to investigate how the haptic channel can be effectively exploited by visually impaired users in a mobile app for the preliminary exploration of an

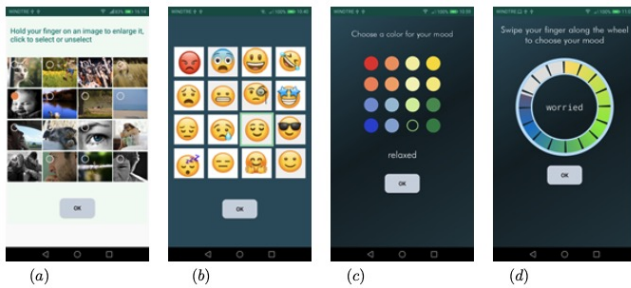


Figure 22. Designed tools: (a) PAM widget; (b) Emoji widget; (c) Color picker widget; (d) Color wheel widget. [55]

indoor environment, namely a shopping mall. Our goal was to use haptics to convey knowledge of how the points of interest (POIs) are distributed within the physical space, and at the same time provide information about the function of each POI, so that users can get a perception of how functional areas are distributed in the environment “at a glance”. Shopping malls are typical indoor environments in which orientation aids are highly appreciated by customers, and many different functional areas persist. We identified seven typical categories of POIs which can be encountered in a mall, and then associated a different vibration pattern each. In order to validate our approach, we designed and developed a prototype for preliminary testing, based on the Android platform. The prototype was periodically debugged with the aid of two visually impaired experienced users, who gave us precious advice throughout the development process. We will describe how this app was conceived, the issues emerged during its development and the positive outcomes produced by a very early testing stage. Finally, we will show that the proposed approach is promising and is worthy of further investigation.

A primer on open science-driven repository platforms [18] by Bardi et al. for 16th Research Conference on Metadata and Semantic Research, MTSR 2022.

Summary: Following Open Science mandates, institutions and communities increasingly demand repositories with native support for publishing scientific literature together with research data, software, and other research products. Such repositories may be thematic or general-purpose and are deeply integrated with the scholarly communication ecosystem to ensure versioning, persistent identifiers, data curation, usage stats, and so on. Identifying the most suitable off-the-shelf repository platform is often a non-trivial task as the choice depends on functional requirements, programming and technical skills, and infrastructure resources. This work analyses four state-of-the-art Open Source repository platforms, namely Dryad, Dataverse, DSpace, and InvenioRDM, from both a functional and a software perspective. This work intends to provide an overview serving as a primer for choosing repository platform solutions in different application sce-

narios. Moreover, this paper highlights how these platforms reacted to some key Open Science demands, moving away from the original and old-fashioned concept of a repository serving as a static container of files and metadata.

(Semi)automated disambiguation of scholarly repositories [13] by Baglioni et al. for the 19th Conference on Information and Research science Connecting to Digital and Library science, IRCDL 2023.

Summary: The full exploitation of scholarly repositories is pivotal in modern Open Science, and scholarly repository registries are kingpins in enabling researchers and research infrastructures to list and search for suitable repositories. However, since multiple registries exist, repository managers are keen on registering multiple times the repositories they manage to maximise their traction and visibility across different research communities, disciplines, and applications. These multiple registrations ultimately lead to information fragmentation and redundancy on the one hand and, on the other, force registries’ users to juggle multiple registries, profiles and identifiers describing the same repository. Such problems are known to registries, which claim equivalence between repository profiles whenever possible by cross-referencing their identifiers across different registries. However, as we will see, this “claim set” is far from complete and, therefore, many replicas slip under the radar, possibly creating problems downstream. In this work, we combine such claims to create duplicate sets and extend them with the results of an automated clustering algorithm run over repository metadata descriptions. Then we manually validate our results to produce an “as accurate as possible” de-duplicated dataset of scholarly repositories.

Fig. 23 depicts the proposed methodology.

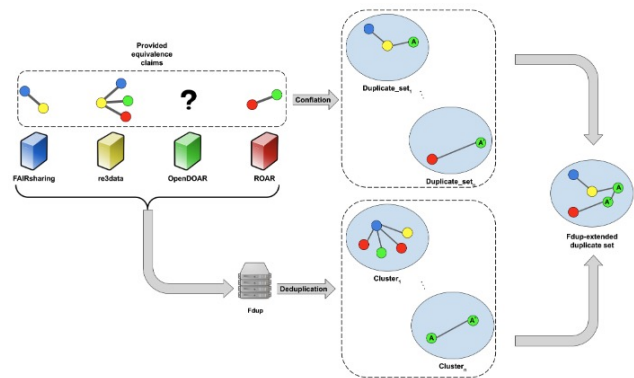


Figure 23. Overview of the methodology. Firstly, the claims provided by the registries are conflated in order to derive duplicate sets; then, the identified duplicate sets are extended with the clusters provided by the FDup automatic de-duplication algorithm running against the content of the registries. In this example, FDup discovers an equivalence between repository profiles A and A*, and this triggers a merger between the two duplicate sets represented above. [13]

Revising Typical Beliefs: One Revision to Rule Them All [39] by Heyninck et al. for the 20th International Conference on Principles of Knowledge Representation and Reasoning, KR2023.

Summary: Propositional Typicality Logic (PTL) extends propositional logic with a connective \bullet expressing the most typical (alias normal or conventional) situations in which a given sentence holds. As such, it generalises e.g. preferential logics that formalise reasoning with conditionals such as “birds typically fly”. In this paper, we study revision of sets of PTL-sentences. We first show why it is necessary to extend the PTL-language with a possibility operator, and then define the revision of PTL-sentences syntactically and characterise it semantically. We show that this allows us to represent a wide variety of existing revision methods, such as propositional revision and revision of epistemic states. Furthermore, we provide several examples showing why our approach is innovative. In more detail, we study revision of a set of conditionals under preferential closure, and the addition and contraction of possible worlds from an epistemic state.

A Discovery Hub for Diamond Open Access publishing [16] by Bardi et al. for the 19th Conference on Information and Research Science Connecting to Digital and Library Science, IRCDL 2023.

Summary: Open Access (OA) publishing is the set of practices thanks to which research publications are accessible freely without barriers. With Diamond Open Access, authors can publish free of charge as the institutional sector with universities, research institutions or libraries provide the necessary technological infrastructure. However, the Diamond OA landscape continues to be fragmented, is often underfunded, and is not always technically proficient enough to develop its full potential for science and society. The CRAFT-OA project, started in January 2023, aims to consolidate the Diamond OA publishing landscape both from the technical and organisational point of views. In this paper we describe the context and architecture of the Diamond Discovery Hub that will be released by the project to increase visibility, discoverability and recognition of Diamond OA institutional publishers and their content. The Diamond Discovery Hub will facilitate the integration with the wider scholarly communication ecosystem and the European Open Science Cloud to enlarge visibility, discoverability and reach of open access publications as part of the emerging Open Science paradigm.

Fig. 24 depicts the architecture of the Diamond Discovery Hub..

Towards digital twins of territories through semantic story maps [20] by Bartalesi et al. for the BUILD-IT 2023.

Summary: Digital maps greatly support storytelling about territories, especially when enriched with data describing cultural, societal, and ecological aspects, conveying emotional messages that describe the territory as a whole. Story maps are interactive online digital narratives that can describe a territory beyond its map by enriching the map with text, pictures, videos, and other multimedia information. This paper

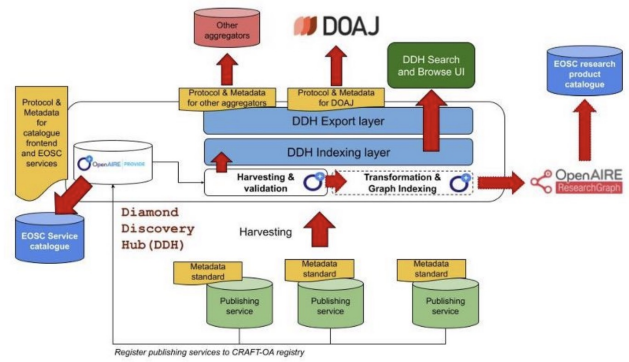


Figure 24. Architecture of the Diamond Discovery Hub. [16]

outlines how online story maps can fill the gap between a map and a territory in narratives to create a digital twin of different territories as inter-connected semantic stories.

A Graph Neural Network Approach for Evaluating Correctness of Groups of Duplicates [35] by De Bonis et al. for the International Conference on Theory and Practice of Digital Libraries, TPD L 2023.

Summary: Unlabeled entity deduplication is a relevant task already studied in the recent literature. Most methods can be traced back to the following workflow: entity blocking phase, in-block pairwise comparisons between entities to draw similarity relations, closure of the resulting meshes to create groups of duplicate entities, and merging group entities to remove disambiguation. Such methods are effective but still not good enough whenever a very low false positive rate is required. In this paper, we present an approach for evaluating the correctness of “groups of duplicates”, which can be used to measure the group’s accuracy hence its likelihood of false-positiveness. Our novel approach is based on a Graph Neural Network that exploits and combines the concept of Graph Attention and Long Short Term Memory (LSTM). The accuracy of the proposed approach is verified in the context of Author Name Disambiguation applied to a curated dataset obtained as a subset of the OpenAIRE Graph that includes PubMed publications with at least one ORCID identifier.

Fig. 25 depicts the final architecture for the classification of a group of duplicates proposed in the paper.

Blue-Cloud-2026, a Federated European Ecosystem to deliver FAIR & Open data and analytical services, instrumental for the Digital Twins of the Oceans [54] by Schaap et al. for the 10th EuroGOOS International Conference.

Summary: The pilot Blue-Cloud H2020 project combined interests of developing a thematic marine EOSC cloud and serving the Blue Economy, Marine Environment and Marine Knowledge agendas. It deployed a versatile cyber platform with smart federation of multidisciplinary data repositories, analytical tools, and computing facilities in support of explor-

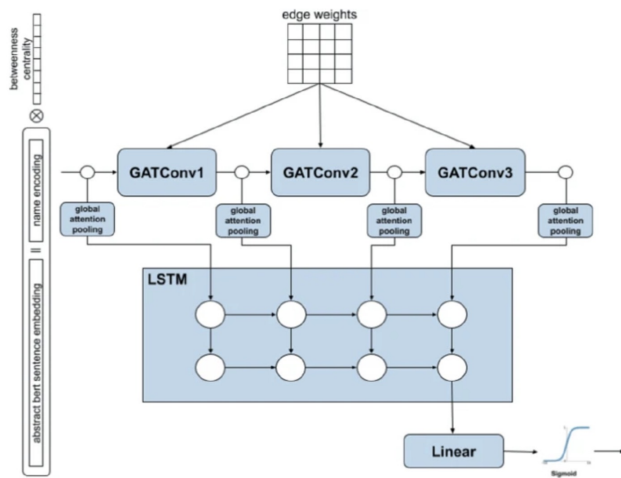


Figure 25. Final architecture for the classification of a group of duplicates. [35]

ing and demonstrating the potential of cloud based open science for ocean sustainability, UN Decade of the Oceans, and G7 Future of the Oceans. The pilot Blue-Cloud delivered: 1) Blue-Cloud Data Discovery & Access service (DD&AS), 2) Blue-Cloud Virtual Research Environment infrastructure (VRE) and 3) Five multi-disciplinary Blue-Cloud Virtual Labs (VLabs). Since early 2023, Blue-Cloud 2026 aims at a further evolution into a Federated European Ecosystem to deliver FAIR & Open data and analytical services, instrumental for deepening research of oceans, EU seas, coastal & inland waters, and building a major data ground segment for the Digital Twins of the Oceans (DTO's). The EMODnet Data Ingestion portal plays a role in the pathways towards the EMODnet data portal. Specifically, the services it provides to data holders include: (a) data submission, with integrated services such as the online submission form, user management service, tracking service, (b) discovery and access, operating on the ingested and completed data submissions, and (c) operational data integration.

Tracing Data Footprints: Formal and Informal Data Citations in the Scientific Literature [42] by Irrera et al. for the International Conference on Theory and Practice of Digital Libraries, TPDL 2023.

Summary: Data citation has become a prevalent practice within the scientific community, serving the purpose of facilitating data discovery, reproducibility, and credit attribution. Consequently, data has gained significant importance in the scholarly process. Despite its growing prominence, data citation is still at an early stage, with considerable variations in practices observed across scientific domains. Such diversity hampers the ability to consistently analyze, detect, and quantify data citations. We focus on the European Marine Science (MES) community to examine how data is cited in this specific context. We identify four types of data citations: formal, informal, complete, and incomplete. By analyzing the usage

of these diverse data citation modalities, we investigate their impact on the widespread adoption of data citation practices.

Fig. 26 depicts the final architecture for the classification of a group of duplicates proposed in the paper.

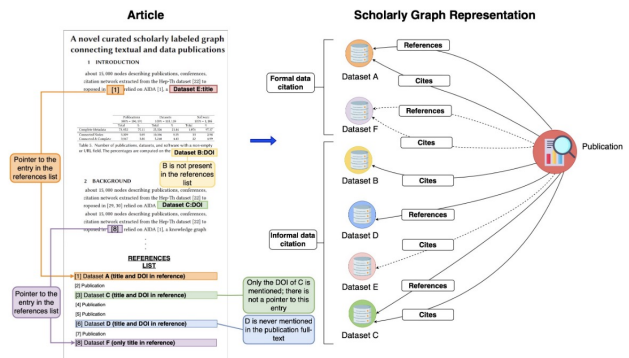


Figure 26. Final architecture for the classification of a group of duplicates. [42]

3.3 Technical Reports

InfraScience members contributed to the following Technical Reports.

Implementation of a drug discovery pipeline on the D4-Science platform [50] by Orro et al., ISTI Technical Report ISTI-TR-2023/001.

Summary: This report documents the implementation of drug discovery pipeline in the D4Science platform realised in the context of the EOOSC-Pillar project. In particular, it documents the pipeline and its constituents. Moreover, it describes how this pipeline has been integrated into the D4Science platform and exploited to create a dedicated Virtual Research Environment facilitating its exploitation and promoting a collaborative oriented approach for screening activities.

Janet: a prototype of a conversational agent for a virtual research environment [40] by Ibrahim and Candela, ISTI Technical Report ISTI-TR-2023/006.

Summary: Conversational agents have been undergoing a lot of development where they have reached their best nowadays thanks to the large language models that have been developed over the past years. Thanks to this, the capabilities of such agents have become more diverse and they are now able to perform a wide range of tasks which can be harnessed to reduce or even replace the human effort in performing those tasks. Therefore, in an effort to equip scientific communities with such a powerful tool, we developed a prototype conversational agent, Janet, and deployed it into one virtual research environment (VRE) with the aim of supporting the scientists in exploiting the knowledge contained within these environments. Janet is still in the preliminary phase where there is a need for extensive testing and improvement; however, this report shows the main aspects of Janet including the architecture, implementation and deployment.

Conversational agents for virtual research environments: a survey of the literature [41] by Ibrahim and Candela, ISTI Technical Report ISTI-TR-2023/007.

Summary: Conversational artificial intelligence is becoming a rather hot topic in the field of artificial intelligence especially after the release of ChatGPT by OpenAI. That is why it is useful to conduct a survey of the literature regarding conversational agents in order to have an idea about the techniques used to develop such systems. Therefore, the aim of this report is to aggregate and summarize the previous research efforts in order to provide insight that would aid future researchers. In addition, the survey was conducted within the context of developing Janet, a conversational agent for virtual research environments; thus, it can be used as a starting point for similar projects.

The gCube geoportal platform [25] by Candela et al., ISTI Technical Report ISTI-TR-2023/012.

Summary: The gCube Geoportal platform is a component of the gCube open source software system conceived to support the creation and publication of georeferenced research objects, i.e., multi-part and multimedia research objects characterised by geospatial and temporal features. The Geoportal platform enables user communities to fully customise the data model characterising their instance by defining the structure, the content and the workflow of the potential research objects to be managed. This report carefully describes the technology and documents how it was exploited to serve the development of a national catalogue for archaeological artifacts.

3.4 Other contributions

InfraScience members contributed to the following publications.

Roadmap Scienza Aperta (Open Science Roadmap) [28] by Castelli et al., CNR Report.

Summary: Open science is a paradigm that influences the practices of knowledge production and sharing. The objective of this roadmap is to outline a path for the implementation and dissemination of open science practices and policies within the National Research Council of Italy.

4. Projects

InfraScience was an active member of the consortiums proposing and implementing 21 research projects (18 were European Union’s supported projects) all focusing on the development of data infrastructures and solutions for various communities of practice.

Blue-Cloud⁹ is a European Union’s Horizon 2020 project (grant agreement No. 862409) started in October 2019 and ending in March 2023. It was funded to implement a practical approach to address the potential of cloud based open science to achieve a set of services identifying also longer term

challenges to build and demonstrate the Pilot Blue Cloud as a thematic EOSC cloud to support research to better understand and manage the many aspects of ocean sustainability, through a set of five pilot Blue-Cloud demonstrators. It seeks to capitalise on what exists already and to develop and deploy, through a pragmatic workplan, the pilot Blue Cloud as a cyber platform bringing together and providing access to (i) multidisciplinary data from observations and models, (ii) analytical tools, and (iii) computing facilities essential for key blue science use cases. The *InfraScience* team is leading the work package “Developing and operating the Blue Cloud VRE, its services and Virtual Labs” called to (a) develop and operate the Blue Cloud Virtual Research Environment, (b) develop and integrate in the Blue Cloud VRE a data taming service, (c) develop and integrate in the Blue Cloud VRE a data analytics service, (d) develop and integrate in the Blue Cloud VRE a research object publishing service, (e) develop facilities interfacing the Blue Cloud services catalogue with EOSC.

Blue-Cloud 2026¹⁰ is a European Union’s Horizon Europe project (grant agreement No. 101094227) started in January 2023 and ending in June 2026. The project aims at further evolving the pilot Blue-Cloud project into a federated European ecosystem to deliver FAIR and open data as well as analytical services for research of oceans, EU seas, and coastal and inland waters. It will develop a thematic marine extension to the EOSC in support of the Green Deal, Destination Earth, the EU Mission “Restore our Oceans and Water by 2030” and UN SDGs, ultimately providing a core data service for the Digital Twin Ocean. It develops additional analytical services, new Virtual Labs and new data sets from a multitude of data originators (SeaDataNet, EurOBIS, Euro-Argo, ELIXIR-ENA, SOCAT, EcoTaxa and ICOS-Ocean) and major e-infrastructures (EUDAT, D4Science, WEkEO). The *InfraScience* team was leading the development of the project by playing the role of scientific and administrative coordinator. Moreover, it is leading the work package “Blue-Cloud VRE platform evolution and integration with EOSC resources and services” called to extend and strengthen the services offered by the Blue-Cloud VRE and the facilities for integrating them with other infrastructures and service providers.

CODECS¹¹ is a European Union’s Horizon Europe project (grant agreement No. 101060179) started in October 2022 and ending in September 2026. CODECS works with farmers to develop user-friendly approaches, methods and tools able to document the co-benefits and costs of technologies applied to real contexts. Specifically, it develops a vision of ‘sustainable digitalisation’. By assessing a full range of social, economic and environmental costs and benefits, the CODECS platform will host search, demonstration and assessment tools. The project applies an action research methodology and tests digital technologies via a demonstration farm network. The *InfraScience* team was responsible for the co-

⁹Blue Cloud website <https://blue-cloud.org>

¹⁰Blue Cloud website <https://blue-cloud.org/>

¹¹CODECS website <https://www.horizoncodecs.eu/>

ordination of task 8.3 “Ethics, Open Science, data management and gender perspective”, and for gathering information needed for data management and to establish a proper Data Management Plan. Support was also offered at this stage for management choices of research products in line with Open Science and FAIR principles.

CRAFT-OA¹² is a European Union’s Horizon Europe project (grant agreement No. 101094397) started in January 2023 and ending in December 2025. The CRAFT-OA project aims to consolidate the Diamond OA publishing landscape. The project focuses on four threads of activities to improve the technical and organisational infrastructure of Diamond OA: (i) Provide technical improvements for journal platforms and journal software; (ii) Build communities of practice to foster overall infrastructure improvement; (iii) Increase visibility, discoverability and recognition for Diamond OA publishing; and (iv) Integrate Diamond OA publishing with EOSC and other large-scale data aggregators. CRAFT-OA’s 23 consortium partners from 17 European countries are all engaged in institutional publishing and its infrastructures, and committed to sustaining and developing capacities in the field. Many CRAFT-OA partners are leading organisations at national and European levels in terms of open publishing, and represent internationally visible centres of expertise in Open Science and FAIR implementation with strong connections to EOSC. CRAFT-OA plans to deliver technical and community tools, training events, training materials, information, and services for the Diamond OA institutional publishing environment. It will foster communities of practice with the capacity to sustain the project improvements over time. The InfraScience team contributed as an affiliated entity of OpenAIRE to WP5 on Journal Visibility and Discoverability Improvement and WP6 on Journals’ data re-use and uptake in EOSC. The team brought its expertise on aggregative infrastructures, metadata interoperability, and Scientific Knowledge Graphs for the design of (i) the Diamond Discovery Hub [16], a catalogue of Diamond OA publishers and journals that ease the integration with indexes and aggregators; and (ii) the OpenAIRE Publisher Dashboard, a monitor dashboard where publishers and journals can find statistics about different Open Science aspects (e.g., FAIRness, openness, links to datasets and software, projects, funders, citations, etc.).

DESIRA¹³ is a European Union’s Horizon 2020 project (grant agreement No. 818194) started in June 2019 and ending in May 2023. It was funded to develop a methodology - and a related online tool - to assess the impact of past, current and future digitalization trends of agriculture and rural areas, using the concept of socio-cyber-physical systems – which connect and change data, things, people, plants and animals. Impact analysis will be linked directly to the United Nation’s Sustainable Development Goals. It also contributes to the promotion of the principles of Responsible Research and Innovation. The *InfraScience* team is leading the activity

“Knowledge Infrastructure: the DESIRA Virtual Research Environment” to design, deliver, and operate the Virtual Research Environment envisaged to serve the needs of the Living Labs. This VRE, a ready-to-use infrastructure for communication exploiting the resources and services operated by D4Science, offers (i) a private cloud storage area, equipped with an easy-to-use workspace application designed for use by a wide set of different actors, and the capability to store either private or shared data; (ii) social networking applications, where each project member can share posts (text, images, and files annotated with hashtags) with VRE members and to collect them in a dedicated News Feed (as in Twitter and Facebook); (iii) a private messaging application integrated with cloud storage to exchange large amounts of data securely; (iv) an activity tracker and collaborative wiki.

EcoScope¹⁴ is a European Union’s Horizon 2020 project (grant agreement No. 101000302) started in September 2021 and ending in August 2025. It aims to develop an interoperable platform and a robust decision-making toolbox, available through a single public portal, to promote an efficient, ecosystem-based approach to the management of fisheries. It will be guided by policy makers and scientific advisory bodies, and address ecosystem degradation and the anthropogenic impact that are causing fisheries to be unsustainably exploited across European Seas. In compliance with the Open Science practices, the EcoScope Platform will organise and homogenise climatic, oceanographic, biogeochemical, biological and fisheries datasets for European Seas to a common standard type and format that will be available to the users through interactive mapping layers. The EcoScope Toolbox, a scoring system linked to the platform, will host ecosystem models, socio-economic indicators, fisheries and ecosystem assessment tools that can be used to examine and develop fisheries management and marine policy scenarios as well as maritime spatial planning simulations. Multi-disciplinary groups of end-users and stakeholders will be involved in the design, development and operation of both the platform and the toolbox. Novel assessment methods for data-poor fisheries, including non-commercial species, as well as for biodiversity and the conservation status of protected megafauna, will be used to assess the status of all ecosystem components across European Seas and test new technologies for evaluating the environmental, anthropogenic and climatic impact on ecosystems and fisheries. A series of sophisticated capacity building tools, such as online courses, documentary films, webinars and games, will be available to stakeholders through the EcoScope Academy. By filling these knowledge gaps and developing new methods and tools, the EcoScope project will provide an effective toolbox to decision makers and end-users that will be adaptive to their capacity, needs and data availability. The toolbox will incorporate methods for dealing with uncertainty and deep uncertainty; thus, it will promote efficient, holistic, sustainable, ecosystem-based fisheries management that will aid towards restoring fisheries

¹²CRAFT-OA website <https://www.craft-oa.eu/>

¹³DESIRA website desira2020.eu

¹⁴EcoScope website <https://ecoscopium.eu/>

sustainability and ensuring a balance between food security and healthy seas. The *InfraScience* team manages WP4 by contributing to (i) environmental data production and harmonisation, (ii) data mining of vessel transmitted information, (iii) ecological niche modelling via AI models, (iv) biodiversity monitoring indexes, and (v) ecosystem risk assessment.

EOSC Future¹⁵ is a European Union’s Horizon 2020 project (grant agreement No. 101017536) started in April 2021 and ending in September 2023. It aims to integrate, consolidate, and connect e-infrastructures, research communities, and initiatives in Open Science to develop further the EOSC Portal, EOSC-Core and EOSCExchange of the European Open Science Cloud (EOSC). EOSC Future will unlock the potential of European research via a vision of Open Science for Society by (i) bringing all major stakeholders in the EOSC ecosystem together under one project umbrella to break the disciplinary and community silos and consolidate key EOSC project outputs, (ii) developing scientific use cases in collaboration with the thematic communities showcasing the benefits and societal value of EOSC for doing excellent and interdisciplinary research, (iii) engaging the wider EOSC community and increasing the visibility of EOSC through communications campaigns, marketing strategies, and physical and online engagement events, and (iv) including the EOSC community in developing the EOSC Portal (including the long tail of science, public and private sectors, and international partners) via co-creation open calls. The *InfraScience* team is contributing in the WP3 to define the architecture and interoperability guidelines and frameworks and in WP4 to the design and development of the Portal Supply Layer (back-office) adapting the existing OpenAIRE services to the EOSC-Core requirements.

FAIRCORE4EOSC¹⁶ is a European Union’s Horizon Europe project (grant agreement No. 101057264) started in June 2022 and ending in May 2025. It was funded to develop and introduce new components seamlessly integrated with the existing EOSC-Core services, bridging gaps identified in the EOSC Strategic Research and Innovation Agenda (SRIA). It uses existing technologies and services to develop nine new EOSC-Core components to enable EOSC persistent identifiers, an EOSC research software infrastructure and support for advances in EOSC repositories – all of which are important for the FAIR research life cycle. *InfraScience* is responsible for the Research Discovery Graph component (RDGraph). The RDGraph utilises the content of the EOSC catalogue and expands upon it by incorporating additional entities such as Research Activity Identifiers (RAiDs). It provides a range of advanced functionalities that leverage multiple intelligent community-oriented discovery tools developed by the partners of the FAIRCORE4EOSC project.

FOSSR¹⁷ is a project funded by the EU Next Genera-

tion Funding in the context of the National Recovery and Resilience Plan. It was funded to become an Italian Open Science Cloud, along the lines of the European Open Science Cloud project, in which to integrate innovative services developed by the project for data collection, data curation and Fairness, and data analysis on economic and societal change. FOSSR fosters the building of an integrated knowledge sharing platform, a single point of access to all the tools and services made available by the Italian nodes of social science infrastructures: ERIC CESSDA, ERIC SHARE and RISIS adopting the common theme of the development of Open Science in the Italian context with the goal of creating a framework of tools and services for the social science scholar community. *InfraScience* contributes to developing and operating the Virtual Research Environment (VRE) service, which nicely complements the FOSSR e-infrastructure offering by enabling the creation of Virtual Laboratories where Open Science practices are transparently promoted. The VRE service is operated by D4Science and comprises a set of services conceived as a collaborative research platform to boost the capacity to connect to all targeted audiences and increase substantially the interaction among them.

GraspOS¹⁸ is a European Union’s Horizon Europe project (grant agreement No. 101095129) started in January 2023 and ending in December 2025. It aims to further extend the reach and influence of Open Science. Specifically, it will develop a data infrastructure that will help policy reforms and bring about the adoption of a responsible research assessment system that embeds OS practices. Additionally, the project will create the necessary tools to expand the reach of the European Open Science Cloud ecosystem and improve monitoring services. These services will also allow for policy-level rewards that encourage OS practices and mindsets. The *InfraScience* team coordinates a pilot aimed at assessing the evaluation processes for the career progression of CNR researchers and technologists. After signing the Agreement on Reforming Research Assessment (ARRA) and joining the Coalition for Advancing Research Assessment (CoARA), CNR launched an internal assessment campaign in 2023, currently ongoing, to evaluate the promotion of over 3,000 permanent research staff members. The pilot’s objective is to study how the current competitive call aligns with the principles of ARRA and CoARA, moving beyond the mostly bibliometric practices of previous selections. *InfraScience* is also contributing to the development of the RA-SKG¹⁹ extension within the Scientific Knowledge Graphs Interoperability Framework²⁰. This extension enhances core model entities to incorporate relevant details for research assessment, facilitating a more structured and semi-automated approach to composing narrative CVs and product portfolios in assessment exercises.

¹⁸GraspOS website <https://graspos.eu/>

¹⁹RA-SKG repository <https://github.com/skg-if/ext-ra-skg>

²⁰Scientific Knowledge Graphs Interoperability Framework <https://skg-if.github.io/>

¹⁵EOSC Future website <https://eoscfuture.eu/>

¹⁶FAIRCORE4EOSC website <https://faircore4eosc.eu/>

¹⁷FOSSR website <http://www.fossr.eu/>

I-GENE²¹ is a European Union’s Horizon 2020 project (grant agreement No. 862714) started in November 2019 and ending in October 2023. It proposed a new concept of genome editing based on nanotransducers (NTs), aiming to make previously impracticable applications of genome editing and transcriptional regulation by Cas9 safe. This methodology relies on the laser activation of a NT, which triggers consequently a thermo-switchable DNA double strand break or cleavage. The proposed technology implements a concept of multi-input AND gates, where the output (gene editing) is true only if multiple inputs are true at the same time (e.g., NT activation and recognition of 2 different loci). *InfraScience* provides the I-GENE community with a dedicated gateway and a series of Virtual Research Environments fostering large-scale collaborations where many potentially geographically distributed co-workers can access and process large amounts of data, also by promoting the public debate to support the design of a new strategy/technology for genome editing, ethically acceptable, sustainable and society desirable.

MOVING²² is a European Union’s Horizon 2020 project (grant agreement No. 862739) started in September 2020 and ending in August 2024. It builds capacities and co-develop policy frameworks across Europe to assess how European mountain areas – playing a central role in the well-being of many highly populated European regions –are being impacted by climate change. It establishes new or upscaled value chains to boost resilience and sustainability of mountain areas. The first step will be to screen traditional and emerging value chains in all European mountain areas. The next step will involve in-depth assessment of vulnerability and resilience of land use, production systems and value chains in 23 mountain regions. The project will use a virtual research environment to promote online interactions amongst actors and new tools to ensure information is accessible by different audiences. *InfraScience* supports the development and operation of the virtual research environment.

NAVIGATOR²³ is a project funded by Regione Toscana, started in October 2020 and ending in October 2023. It is called to set the first, regional Virtual Research Environment (VRE) to advance a personalized vision of the clinical management of malignant, solid tumors. The core component will be a Tuscan Biobank of cancer images and imaging biomarkers, set as a shared infrastructure that will support the discovery, test and proof of new models, biomarkers and predictive methods for a better understanding of cancer biology, risk and care. *InfraScience* is involved as service provider, to deliver the NAVIGATOR VRE via the D4Science infrastructure and enable integration of the biobank with the data analysis tools of the platform. The work is performed in collaboration with the SILab team of ISTI.

OpenAIRE Nexus²⁴ is a European Union’s Horizon 2020

project (grant agreement No. 101017452) started in January 2021 and ending in June 2023. The objective of the project is to assemble a suite of services to support researchers, research communities, research performing organisations, policy makers and SME at the adoption, implementation, and monitoring of Open Science practices. The suite is composed of fourteen services, onboarded to the European Open Science Cloud (EOSC), organised in three portfolios: PUBLISH (Zenodo.org, episciences.org, AMNESIA, Argos), DISCOVER (PROVIDE, EXPLORE, CONNECT), MONITOR (OpenAIRE Research Graph, Research Impact Monitoring, UsageCounts, OpenCitations, ScholeXplorer, OpenAPC, Open Science Observatory, OpenAIRE AAI). The project also establishes synergies with INFRAEOSC-07 and INFRAEOSC-03 projects to contribute to the interoperability framework for the EOSC. *InfraScience* leads the technical coordination of the project, is responsible for the integration of the services with the EOSC to provide Virtual Access (WP3) and for the contribution to the EOSC Interoperability framework in collaboration with the INFRAEOSC-07 and INFRAEOSC-03 projects (WP7). The group is responsible for the provision of the following services: OpenAIRE Research Graph, ScholeXplorer, Broker service (integrated in PROVIDE) and contributes to the delivery of the Research Impact Monitoring, the Open Science Observatory, PROVIDE, EXPLORE, and CONNECT.

SciLake²⁵ is a European Union’s Horizon Europe project (grant agreement No. 101058573) started in January 2023 and ending in December 2025. SciLake aims to extend technical work in the field of Science Knowledge Graphs (SKGs). The project leverages SKGs as a basis to establish the concept of the scientific lake: a research ecosystem that facilitates the creating, combining, and querying of cross-domain and domain-specific SKGs. This ecosystem includes tools capable of extracting knowledge from unstructured (e.g. textual) information and tools for creating, interlinking, and managing SKGs. It empowers interoperability among SKGs, supporting knowledge transformation, unifying and simplifying how SKGs can be queried, and accelerating SKG graph processing and analysis. The project builds a prototype and delivers an additional tier of smart services to assist the discovery of knowledge and improve research reproducibility. The work is done in close consultation with four research communities (Neuroscience, Cancer, Transportation, and Energy). All services are tailored to consider domain-specific requirements, ensuring their relevance to community needs and expectations. Each of these research communities also demonstrates and evaluates selected SciLake services in the context of a domain-specific pilot. Finally, the project leverages EOSC functionalities (e.g., comply with the European Open Science Cloud Interoperability Framework for monitoring, accounting, and Authentication and Authorization Infrastructures). It integrates its open-source services into the OpenAIRE and EOSC Core services portfolio. The *Infra-*

²¹I-GENE website i-geneproject.eu

²²MOVING website www.moving-h2020.eu

²³NAVIGATOR website <http://navigator.med.unipi.it/>

²⁴OpenAIRE Nexus website <https://www.openaire.eu/>

openaire-nexus-project

²⁵SciLake website <https://scilake.eu/>

Science team is involved in several work packages, with a specific focus on requirement acquisition and architectural design (WPI), led by WP leader. The team collaborates in defining technical specifications and designing innovative solutions for managing and interoperating scientific knowledge graphs. They also collaborate on creating tools that facilitate efficient retrieval, analysis, and integration of research data, making it easier for scientists and researchers to access, share, and collaborate. Additionally, InfraScience members contribute to the project by providing the SciLake resource catalog, leveraging D4Science technologies.

SerGenCOVID-19²⁶ is an Italian project funded by CNR, started in January 2021 and ending in December 2024. Its goal was to systematically collect clinical data on the Italian population affected by Covid-19 to learn more about the individual response to the virus and develop protocols for the management of future patients. *InfraScience* is responsible for the design and development of an IT platform, with the following main components: (a) store and retrieve surveys and results of serological tests; (b) a web interface for the participants to fill the anamnestic questionnaire elaborated by the experts, and to access the results of their serological test; (c) a working environment for data collection by a Virtual Research Environment.

Skills4EOSC²⁷ is a European Union's Horizon Europe project (grant agreement No. 101058527) started in September 2022 and ending in August 2025. It was funded to unify the existing training landscape into a common and trusted pan-European ecosystem of Competence Centers on open science and data, to accelerate the upskilling of European researchers and data professionals in the field of FAIR and Open Data, intensive-data science, and scientific data management. *InfraScience* is coordinating the contribution of the Italian National Research Council research unit comprising three Institutes: Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (ISTI), Istituto di Linguistica Computazionale "A. Zampolli" (ILC), and Istituto di Fisica Applicata "N. Carrara" (IFAC). In addition to that, *InfraScience* is leading the task "Collaboration, Coordination and Sharing Platform" called to select a set of agreed and interoperable IT services (e.g., shared workspace, Virtual Labs, Catalogues, Registries, Repositories, Videoconference services facilitating the collaboration, services for supporting common training activities on specific topics) hosted by the Competence Centers to facilitate the continuous alignment, sharing and reuse of material, methodologies and best practices across Competence Centers.

SoBigData RI PPP²⁸ is a European Union's Horizon Europe project (grant agreement No. 101079043) started in October 2022 and ending in September 2025. It was funded to move the SoBigData RI forward from the simple awareness of ethical and legal challenges in social mining to the

development of concrete tools that operationalize ethics with value-sensitive design, incorporating values and norms for privacy protection, fairness, transparency, and pluralism. *InfraScience* is responsible for the SoBigData central hub, the component that oversees the technological, administrative, and governance aspects of the RI. The central hub, located in Italy, coordinates all the national nodes, train the staff who will work in them, and establish the methods to provide cutting-edge dynamic digital assets to remote locations without requiring costly on-site expertise. The SoBigData RI follows the System of Systems model, and the national nodes will join and participate based on these principles: autonomy of constituents (independence and evolution); openness (join and leave; dynamic reconfiguration), and distribution (interdependence and interoperability).

SoBigData-PlusPlus²⁹ is a European Union's Horizon 2020 project (grant agreement No. 871042) started in January 2020 and ending in December 2023. It was funded to develop a distributed, Europe-wide, multidisciplinary research infrastructure. This is coupled with the consolidation of a cross-disciplinary European research community. The project builds upon the EU-funded SoBigData project set out to create a research infrastructure delivering an integrated ecosystem for advanced applications of social data mining and Big Data analytics. SoBigData-PlusPlus strengthens infrastructure tools and services by establishing an open platform for the design and performance of large-scale social mining experiments. It delivers specific tools approaching ethics with value-sensitive design integrating values for privacy protection, transparency, and pluralism. *InfraScience* contributes with its infrastructures supporting Open Science, namely D4Science and OpenAIRE. Specifically, D4Science not only operates the SoBigData e-infrastructure, it enables virtual access to the integrated resources, including existing and newly collected datasets, tools and methods for mining social data. *InfraScience* VRE technology supports scientists in benefitting from the integration of the integrated resources and from the access to the computational resources, such as the social mining computational engine and the online coding and workflow design frameworks, needed to process these resources. Within this context, OpenAIRE provides an online science monitoring dashboard, which monitors and quantifies the outputs of the SoBigData research infrastructure in the scholarly communication ecosystem. It identifies every research product (publications, datasets, software, and other types) produced thanks to the OpenAIRE Research Graph and acts as a single entry point for users to discover, search, browse, and get access to research products related to the infrastructure hosted in several scholarly communication sources (e.g., repositories, journals, archives).

SoBigData.it³⁰ is a project funded by the EU Next Generation Funding in the context of the National Recovery and Resilience Plan started in November 2022 and ending in Oc-

²⁶SerGenCovid19 website <https://sergencovid.iit.cnr.it/>

²⁷Skills4EOSC website <https://www.skills4eosc.eu/>

²⁸SoBigData website sbigdata.eu

²⁹SoBigData website sbigdata.eu

³⁰SoBigData.it website <https://pnrr.sbigdata.it/>

tober 2025. The project aims to strengthen the Italian node of the SoBigData research infrastructure, with the goal of enhancing interdisciplinary and innovative research on the multiple aspects of social complexity by combining data and model-driven approach. SoBigData emphasizes the concept of “responsible data science”, considering the ethical values as one of the pillars of reliable use of big data analytics and artificial intelligence technologies. Consequently, SoBigData RI develops methodologies and approaches to put into practice the FAIR (Findable, Accessible, Interoperable, and Reusable) and FACT (Fair, Accurate, Confidential, and Transparent) principles. The strengthening will concern four main aspects of the infrastructure: (i) Technological aspect: strengthening and creating new computational nodes distributed on the Italian territory with centralized and networking technologies; (ii) Scientific aspect: strengthening and creating new virtual laboratories (VL) to explore specific aspects of society through data science; (iii) Ethical Aspect: Promote the new generation of “responsible data scientists” collaborating with the academia and industry (i.e., providing ethical and legal counseling and facilitating the technology transfer); finally (iv) the creation of DIGIT services for multidisciplinary communities by exploiting the research produced in the VL and the new computational and technological resources. The RI will tackle complex social phenomena such as (but not limited to): Social Impact of AI and explainable machine learning, Societal Debates, and Misinformation Analysis, Disaster response and recovery, Societal and industrial impact of Next-Generation Internet beyond 5G networks. The *InfraScience* team is leading the work package “Empowering the Digital Backbone” that aims to empower and create the computational nodes of SoBigData at the Italian level through data centers connected to the RI network. The objective is to increase the capability of the RI in terms of storage and computing power needed to support short and long-term scalability, robustness, availability, and reliability of services. Moreover, it is supporting the development of virtual laboratories and the provisioning of the RI services.

TAILOR³¹ is a European Union’s Horizon 2020 project (grant agreement No. 952215) started in September 2020 and ending in August 2023. Its purpose was to building the capacity of providing the scientific foundations for Trustworthy AI in Europe by developing a network of research excellence centres leveraging and combining learning, optimization and reasoning. *InfraScience* is leading the Trustworthy AI work package aiming at establishing a continuous interdisciplinary dialogue for investigating the methods and methodologies to design, develop, assess, enhance systems that fully implement Trustworthy AI with the ultimate goal to create AI systems that incorporate trustworthiness by design. This activity is organized along the six dimensions of Trustworthy AI: explainability, safety and robustness, fairness, accountability, privacy, and sustainability. Each task aims at advancing knowledge on a specific dimension and puts it in

relationships with foundation themes. The overall mission for Trustworthy AI is to combine the various dimensions in the TAILOR research and innovation roadmap. Moreover, to maximize this overall goal and take advantage of any effort in Europe, TAILOR will also interact and collaborate with the activities related to “AI Ethics and Responsible AI” of the proposal Humane-AI-net and will lead the organization of joint scientific actions.

5. Infrastructures and Services

InfraScience leads the development of two large scale and well known infrastructures supporting Open Science, namely *D4Science* and *OpenAIRE*. Moreover, the team actively contributed to the development of the European Open Science Cloud by participating in key projects, initiatives and task forces (cf. Sec. 10).

*D4Science*³² [5] is an IT infrastructure specifically conceived to support the development and operation of Virtual Research Environments by the as-a-Service provisioning mode. The underlying distributed computing infrastructure is spread across four main sites, geographically distributed, and managed across different administrative domains. The Pisa site is conceived to be the core element of the D4Science computing infrastructure. It realizes a cloud infrastructure completely based on open source technologies aiming at guaranteeing the dynamic allocation of the hardware resources and high availability of the services. Three sites are operated on GARR premises, i.e., the Italian National Research and Education Network. D4Science-based VREs are web-based, community-oriented, collaborative, user-friendly, open-science-enabler working environments for scientists and practitioners willing to work together to perform a certain (research) task. From the end-user perspective, each VRE manifests in a unifying web application (and a set of Application Programming Interfaces (APIs)) (a) comprising several components made available by portlets organized in custom pages and menu items and (b) running in a plain web browser. Every component is aiming at providing VRE users with facilities implemented by relying on one or more services possibly provisioned by diverse providers. In fact, every VRE is conceived to play the role of a gateway giving seamless access to the datasets and services of interest for the designated community while hiding the diversities originating from the multiplicity of resource providers. Among the components each VRE offers there are some basic ones enacting VRE users to perform their tasks collaboratively, namely: (a) a *workspace* component to organise and share any digital artefact of interest; (b) a *social networking* component to communicate with coworkers by posts and replies; (c) a *data analytics* platform to share and execute analytics methods; (d) a *catalogue* component to document and publish any worth sharing digital artifact. In 2023 its user base reached 22,626 active users (+2,893 users wrt December 2022) (see

³¹TAILOR website tailor-network.eu

³²D4Science website www.d4science.org

Fig. 27). These users executed a total of 73,664 working sessions (circa 6,138 working sessions per month) (see Fig. 28), a total of 103,670,677 analytics tasks (circa 8.6 millions tasks per month), a total of 3696 JupyterLab working sessions, and a total of 3789 RStudio working sessions.

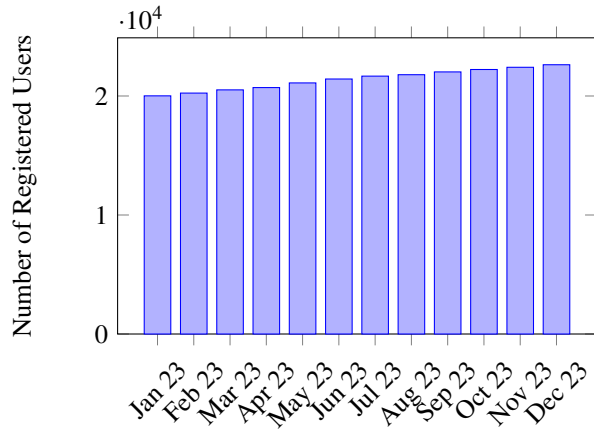


Figure 27. D4Science registered users in 2023.

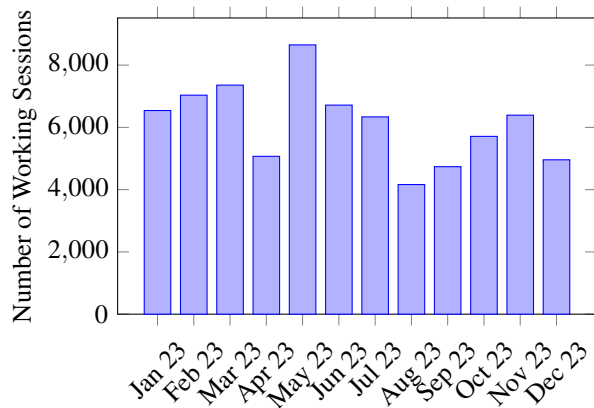


Figure 28. D4Science working sessions in 2023.

OpenAIRE³³ is the technical infrastructure developed and operated by OpenAIRE AMKE, a legal entity composed of 50 institutions working to promote and support a sustainable implementation of Open Access and Open Science policies for reproducible science, transparent assessment and omni-comprehensive evaluation. OpenAIRE AMKE supports the implementation and alignment of Open Science policies at the international level by developing and promoting the adoption of global open standards and interoperability guidelines to realize a sustainable, participatory, trusted, scholarly communication ecosystem, open to all relevant stakeholders (e.g., research communities, funders, project coordinators) and capable of engaging society and foster innovation. Thanks to the network of National Open Access Desks (NOADs), OpenAIRE supports the implementation of Open Science at the

³³OpenAIRE website www.openaire.eu

local and national level, supporting researchers, project coordinators, funders and policy makers with training and support activities. Furthermore, the technical infrastructure materializes the OpenAIRE Graph³⁴: an open, de-duplicated, participatory scientific knowledge graph of interlinked scientific products (including research literature, datasets, software, and other types of research products like workflows, protocols and methods), with access rights information, linked to funding information, research communities and infrastructures. The graph is materialised by collecting more than 240 millions of metadata records from 129k scholarly data sources worldwide. In addition to the information collected from trusted scholarly data sources, the graph includes metadata and links that are (i) asserted by users of the OpenAIRE portals, and (ii) inferred by full-text and metadata mining algorithms. Added-value services are built on top of the graph to offer Open Science services to different stakeholders. During 2023, 249 new data sources implemented the OpenAIRE guidelines for metadata exchange and registered to use the PROVIDE dashboard, 14 research communities started using the CONNECT service to offer a thematic discovery portal to their researchers, and 7 research initiatives used the CONNECT & MONITOR service bundle to track their impact; the OpenAIRE EXPLORE portal, which offers search & discovery functionalities over the OpenAIRE Graph, has exceeded one million visitors from 78 countries. ScholXplorer continued its operation, expanding its citation data with more data sources, reaching 487 millions relationships and 542 millions of API hits. At the end of 2023 the OpenAIRE Graph contained bibliographic records for more than 170Mi publications, 59Mi datasets, and 390K software.

InfraScience was also responsible for the development and operation of services for the ISTI community, namely, the *ISTI IT Infrastructure & services* and the *ISTI Open Portal*.

InfraScience was responsible for the management and operation of the *ISTI IT Infrastructure* and its *services* via the S2I2S Working Group (Servizio Infrastruttura Informatica ISTI e Supporto ai Servizi). InfraScience guaranteed the operation of basic services including e-mail, mailing lists, DNS and centralized authentication. In addition to these basic services, the group designed, implemented and made available other research support services (e.g., content collaboration platform, software development service, flexible project management) and provided extensive and timely support concerning their exploitation. The group started the development of the new institute’s OpenStack-based IaaS platform, i.e., a modern platform conceived to host the Institute services and facilitate their management in the near future. During the last months of 2023 the OpenStack-based infrastructure was released to selected users (see Fig. 29 for VM numbers).

*ISTI Open Portal*³⁵ is a gateway to the scientific production of the Institute of Information Science and Technologies.

³⁴The OpenAIRE Graph <https://graph.openaire.eu/>

³⁵ISTI Open Portal <https://openportal.isti.cnr.it>

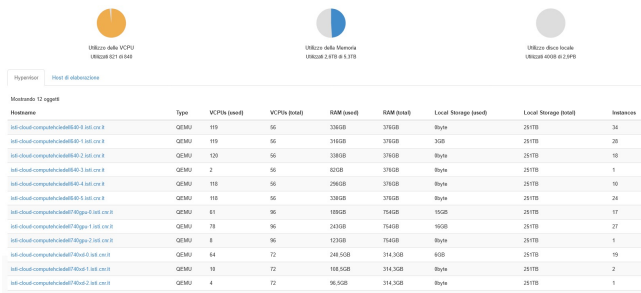


Figure 29. ISTI IT Infrastructure OpenStack Dashboard.

The gateway is an instance of the RepOSGate technology [3]. It (a) systematically collects the ISTI scientific production from the CNR Institutional Repository, (b) enriches the ISTI products metadata by using information from OpenAIRE, Scioexplorer [22], and Altmetric³⁶, and (c) make available the open access (self archived) version(s) of ISTI products. In 2023, the gateway had a total of 16,702 item page views and 7971 downloads.³⁷ Fig. 30 displays the item page views by month. Fig. 31 reports the downloads by month.

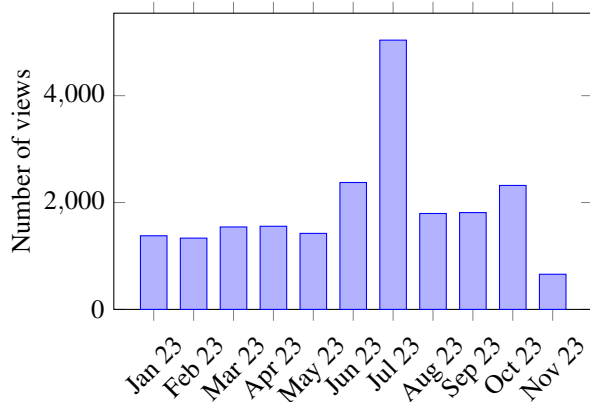


Figure 30. ISTI Open Portal 2023 monthly views.

During 2023 ISTI supported the operation of the *ISPC Open Portal*³⁸, a gateway replicating the ISTI OpenPortal initiative and technology for the needs of the Institute of Heritage Science (Istituto di Scienze del Patrimonio Culturale) of the National Research Council of Italy. Moreover, it supported the creation of the *INO Open Portal*³⁹, a gateway replicating the ISTI OpenPortal initiative and technology for the needs of the National Institute of Optics (Istituto Nazionale di Ottica) of the National Research Council of Italy.

InfraScience was responsible for *open-science.it*⁴⁰, the Italian portal dedicated to Open Science and Open Access.

³⁶Altmetric website <https://www.altmetric.com>
³⁷These figures were collected via OpenAIRE and do not capture the December activities.

³⁸ISPC Open Portal <https://openportal.ispc.cnr.it/>

³⁹INO Open Portal <https://openportal.ino.cnr.it/>

⁴⁰Open-science.it website open-science.it

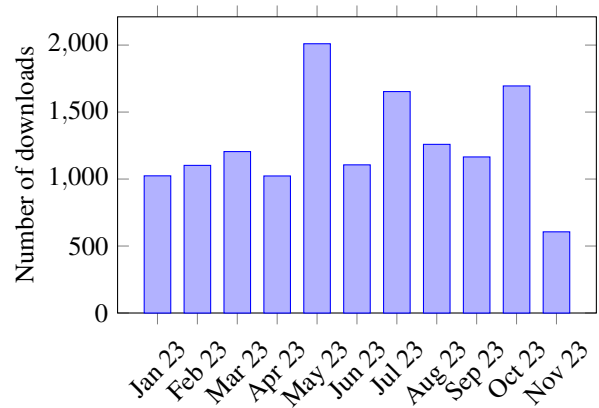


Figure 31. ISTI Open Portal 2023 monthly downloads.

The portal is the result of an initiative developed by the Institute of Information Science and Technologies of the National Research Council of Italy to promote Open Science topics. It originated from the activities of OpenAIRE, the European infrastructure for Open Access, and is supported by the Italian Computing and Data Infrastructure (ICDI) community⁴¹ comprising stakeholders and experts from 28 Italian Universities and Research Performing Organizations. The portal aims to be a point of reference for the Italian scientific community on issues related to Open Science, Open Access and in general to innovations in academic and scientific communication. The portal was officially launched in December 2021. Since then, the portal has published in-depth articles and news, a special section of FAQs on Open Science and Open Access with a specific legal perspective⁴², hosts a resource catalogue⁴³ containing more than 50 open science policies of Italian universities and several other documents of interest, and keeps an up-to-date calendar of Open Science events that it makes available for embedding at interested external sites.

6. Software

InfraScience leads the development of two large scale software systems going hand in hand with the two infrastructures described above.

*gCube*⁴⁴ [4] is an open source software toolkit used for building and operating Hybrid Data Infrastructures (namely D4Science) enabling the dynamic deployment of Virtual Research Environments. It consists of hundreds of web services and software libraries overall offering functions including infrastructure development and operation, science gateways development, VRE creation and management, users management, data management, analytics, and open science support.

⁴¹ICDI website <https://www.icdi.it>

⁴²<https://open-science.it/faq>

⁴³<https://open-science.it/catalogue>

⁴⁴gCube website www.gcube-system.org

According to OpenHub⁴⁵ (statistics collected in December 2023) this software (i) has had 28,393 commits made by 54 contributors representing 1,468,978 lines of code; (ii) is mostly written in Java with an average number of source code comments; (iii) has a well established, mature codebase maintained by a large development team with decreasing Y-O-Y commits; (iv) took an estimated 411 years of effort (CO-COMO model) starting with its first commit in October, 2008 ending with its most recent commit in December, 2023. During 2023, 16 releases of this technology have been released (from gCube 5.14.1 in January 2023 up to gCube 5.16.0 in October 2023). All these releases have been exploited to enhance the service offered by the D4Science Infrastructure.

D-Net⁴⁶ [44] is a framework toolkit designed to empower developers in constructing customized aggregative infrastructures in a cost-effective manner. It offers comprehensive data management services that enable access to diverse external data sources, storage and processing of information objects across various data models, conversion into standardized formats, and seamless exposure of information objects to third-party applications through a suite of standardized access APIs. D-Net's infrastructure enabling services streamline the development of domain-specific aggregative infrastructures by facilitating the selection and configuration of required services and enabling their effortless integration into autonomic data processing workflows. This combination of out-of-the-box data management services and workflow assembly tools makes D-Net an attractive starting point for developers tasked with creating aggregative infrastructures. In 2023, D-Net powered eight installations running aggregation systems for: (a) Recolecta, the Spanish National aggregator⁴⁷; (b) Research networks, associations, and infrastructures – EAGLE (Europeana network of Ancient Greek and Latin Epigraphy)⁴⁸, EFG (European Film Gateway)⁴⁹, OpenAIRE (Open Access Infrastructure for Research in Europe)⁵⁰; (c) EU projects – ARIADNEplus (Advanced Research Infrastructure for Archaeological Data Networking in Europe - plus - EC H2020 project GA 823914)⁵¹; (d) Institutions – ISTI Open Portal⁵², ISPC Open Portal⁵³, and INO Open Portal⁵⁴.

7. Datasets

InfraScience released the following datasets.

OpenAIRE Research Graph Dataset The dataset represents

⁴⁵gCube on Open Hub <https://www.openhub.net/p/gCube>

⁴⁶D-Net website d-net.research-infrastructures.eu

⁴⁷RECOLECTA website <https://recolecta.fecyt.es>

⁴⁸EAGLE website <https://www.eagle-network.eu>

⁴⁹European Film Gateway website <https://www.europeanfilmgateway.eu>

⁵⁰OpenAIRE Explore website <https://explore.openaire.eu/>

⁵¹ARIADNEplus website <https://ariadne-infrastructure.eu/>

⁵²ISTI Open Portal <https://openportal.isti.cnr.it>

⁵³ISPC Open Portal <https://openportal.ispc.cnr.it/>

⁵⁴INO Open Portal <https://openportal.ino.cnr.it/>

the OpenAIRE Graph⁵⁵, a free and open resource that brings together and interlinks hundreds of millions of metadata records from over 100k data sources trusted by researchers. Two new releases of the OpenAIRE Research Graph were published in August [47] and October [45].

OpenAIRE Research Graph: Dumps for research communities and initiatives This dataset contains metadata records of the OpenAIRE Graph relevant for the research communities and initiatives collaborating with OpenAIRE and with a public Open Research Gateway. In particular, one version was released in August [48].

OpenAIRE Covid-19 publications, datasets, software and projects metadata This dataset provides access to the metadata records of publications, research data, software and projects that may be relevant to the Corona Virus Disease (COVID-19) fight. The dataset contains the OpenAIRE COVID-19 Gateway records, identified via full-text mining and inference techniques applied to the OpenAIRE Graph. The OpenAIRE Graph is one of the largest Open Access collections of metadata records and links between publications, datasets, software, projects, funders, and organizations, aggregating 12,000+ scientific data sources world-wide, among which the Covid-19 data sources Zenodo COVID-19 Community, WHO (World Health Organization), BIP! FINDER for COVID-19, Protein Data Bank, Dimensions, scienceOpen, and RSNA. One new version was released in August [17].

OpenAIRE Research Graph Dump: new collected projects

The dataset is updated at every update of the OpenAIRE Graph and contains the new project grants available in the OpenAIRE Graph. One of the main users of the dataset is Zenodo, which uses it to feed the list of grants that users can choose when filling in the deposition form. Seven versions were released [8, 9, 10, 11, 12, 6, 7].

OpenAIRE Graph: Dataset of funded products This dataset contains the metadata records about research products (research literature, data, software, other types of research products) with funding information available in the OpenAIRE Graph. Records are grouped by funder in a dedicated archive file. One new version was released in August [46].

8. Organised Events

A. Bardi was Program Chair of the 19th Conference on Information and Research Science Connecting to Digital and Library Science. M. Assante, M. Baglioni, L. Candela, P. Manghi, and A. Mannocci were members of the Program Committee.

G. Casini was Chair of the first Workshop on AI-driven heterogeneous data management: Completing, merging, handling inconsistencies and query-answering (ENIGMA-2023) co-located with the 20th International Conference on Principles of Knowledge Representation and Reasoning (KR2023). U. Straccia was member of the Program Committee.

⁵⁵<https://graph.openaire.eu/>

L. Candela was a member of the Program Committee of the 15th International Workshop on Science Gateways, 13-15th June 2023, Tübingen, Germany.

L. Candela was a member of the Program Committee of the 2nd Workshop on Artificial Intelligence for Cultural Heritage, held in conjunction with the 22nd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2023).

U. Straccia was a member of the Program Committee of the 13th Conference of the European Society for Fuzzy Logic and Technology jointly with the AGOP and FQAS conferences.

U. Straccia was a member of the Program Committee of the Survey Track of the 32nd International Joint Conference on Artificial Intelligence.

U. Straccia was a member of the Program Committee of the 39th Conference on Uncertainty in Artificial Intelligence.

U. Straccia was a member of the Program Committee of 21st International Workshop on Nonmonotonic Reasoning, September 2-4, 2023, Rhodes, Greece.

U. Straccia was a member of the Program Committee of Declarative AI 2023, a conference bringing together the 7th International Joint Conference on Rules and Reasoning (RuleML+RR 2023), DecisionCAMP 2023, and the 19th Reasoning Web Summer School (RW 2023) taking place at Oslo, Norway during 18 - 24 September, 2023.

U. Straccia was a member of the Program Committee of 36th International Workshop on Description Logics, September 2-4, 2023, Rhodes, Greece.

U. Straccia was a member of the Program Committee of the 20th International Conference, ESWC 2023, Hersonissos, Crete, Greece, May 28-June 1, 2023.

9. Training Activities

In 2023, the InfraScience Lab delivered a series of targeted training sessions aimed at fostering Open Science practices and improving Research Data Management (RDM) across diverse academic and research communities. These initiatives, spanning multiple institutions and international collaborations, provided early career researchers, PhD students, and research professionals with essential knowledge and practical tools to navigate the evolving landscape of Open Science.

In particular, the following training activities were organised:

- “Open Science and Research Data Management”. 10-hour course for PhD students at Scuola Normale Superiore. Period: January 2023. Modules covered: (i) Open Science, Open Access, and research assessment, (ii) Research Data Management and FAIR principles, (iii) Data Management Plan, (iv) Ask Me Anything!
- “Introduction to Open Science and RDM for early career researchers”. 3-hour training as part of the winter school of the MSCA-ITN network SMART-X project (Grant Agreement 860553), Bormio, January 12, 2023.

- “On the importance of Open Science and good Research Data Management. Rationale and tools”. Presentation within the seminar “Research data: European perspectives”, University of Lille, March 23, 2023.
- “OpenOrgs curation sprint”. Coordination of a training session and group disambiguation of organization data. Webinar for OpenAIRE NOAD. Date: April 19, 2023.
- Modules “FAIR Principles and application” and “Data Management Plan”, totaling 6 hours, as part of the Open Science course in the transversal teaching for PhD students 2023 at the University of Pisa. Dates: May 11 and 12, 2023.
- “Open Science and Data Management Plan”. 7-hour course for Area Science Park within the scientific and innovation system of Friuli Venezia Giulia. Dates: May 15 and 18, 2023.
- Organization of the 12-hour course “News and Insights from Open Science, and How to Exploit Them in Your Scientific Work” for MSCA ITN SMART-X, grant agreement no. 860553. Period: June-July 2023.
- “Fulfilling Open Science. Compliance with European Commission mandates and Data Management Plan in CODECS project”. Webinar for the Codecs project, November 30, 2023.
- “Etica nella pubblicazione scientifica. Workshop su Open Access e prevenzione dell’editoria predatoria”. Workshop organised by CID Ethics-CNR, December 7, 2023, Più Libri Più Liberi.

10. Working Groups, Task Forces, & Interest Groups

InfraScience members chaired the following Working Groups, Task Forces, and Interest Groups:

- *EOSC Future - Research Product Publishing Framework Working Group* (A. Bardi, P. Manghi) – a WG to define a Research Publishing framework to simplify the adoption of that practice, by enabling the services of research infrastructures to seamlessly integrate repository deposition workflows in the context of the EOSC.
- *GOFAIR - Discovery Implementation Network* (A. Bardi) – a GO FAIR consortium called to provide interfaces and other user-facing services for data discovery across disciplines;
- *Gruppo di Lavoro “Roadmap per la scienza aperta del CNR”* (D. Castelli) – A WG called to develop the roadmap leading to the implementation of open science practices by the National Research Council of Italy.

- *ISTI IT infrastructure (S2I2S)* (F. Debole) – a WG called to drive the development of the Institute IT and services;
- *ISTI Open Access* (L. Candela) – a WG called to drive the development of the Institute open access and open science policies and practices;
- *Research Data Alliance - Scientific Knowledge Graphs – Interoperability Framework (SKG-IF) WG* (A. Mannocci) – a WG targeting the definition of a framework to enable a seamless exchange of information among diverse initiatives regarding Scientific Knowledge Graphs, intended as knowledge bases of scholarly knowledge content (e.g., repositories, databases, catalogues, knowledge graphs, LOD collections).
- *Research Data Alliance - Open Science Graphs for FAIR Data IG* (P. Manghi, A. Mannocci) – an IG investigating the open issues and identifying solutions towards achieving interoperability between services and information models of Open Science Graph initiatives.

InfraScience members contributed to the following Working Groups, Task Forces, and Interest Groups:

- *CoARA WG - Towards Open Infrastructures for Responsible Research Assessment (OI4RRA)* (A. Mannocci) – This working group’s mission is to enable institutions to move from proprietary infrastructure and research information, to open (interoperable) alternatives – in support of the transition to responsible research assessment practices. This effort will take into consideration the wide range of research outputs and open science practices, and address the diversity of the global research community.
- *Commission expert group on National Points of Reference on Scientific Information* (D. Castelli) – Commission lead by EU CNECT - DG Communications Networks, Content and Technology and EU RTD - DG Research and Innovation of Member States’ National Points of Reference (NPRs) whose tasks would be to (i) co-ordinate the measures listed in the Recommendation C(2012) 4890 final (relating to open access to publications, open research data, preservation of scientific information, and e-infrastructures); (ii) to act as interlocutor with the Commission; and (iii) to report on the follow-up of the Recommendation.
- *European Innovation Forum Working Group on Data* (D. Castelli) – The EIC Forum serves as a platform where Member States, Associated Countries, and other stakeholders can exchange experiences and ideas regarding these standards and common data policies. Participants discuss their needs, pinpoint gaps, and explore strategies to address these challenges. During discussions, members identified key obstacles, such as discrepancies in data definitions and data accessibility. The Data working group aims to contribute to overcoming the gap characterising the management of innovation-related data by aligning with the objectives outlined in the New European Innovation Agenda, which advocates for the creation and utilization of comprehensive, comparable datasets and the establishment of a common data repository.
- *EOSC Task Force on Technical Interoperability of Data and Services* (P. Manghi) – a TF taking the EOSC Interoperability Framework (EIF) recommendations around technical architecture as their starting point to help develop the EOSC Core and Exchange as described in the SRIA.
- *EOSC Task Force on Infrastructures for Quality Research Software* (L. Candela) – a TF fostering the development and deployment of tools and services that allow researchers to properly archive, reference, describe with proper metadata, share and reuse research software, as well as to improve their quality, both from the technical and organizational point of view.
- *Gruppo di Lavoro “Roadmap per la scienza aperta del CNR”* (L. Candela) – A WG called to develop the roadmap leading to the implementation of open science practices by the National Research Council of Italy.
- *Helmholtz Metadata Collaboration (HMC)* (D. Castelli) – Crossing the lines between the research fields, the Helmholtz Association decided to strengthen activities in the area of metadata significantly. To do this, it set up the Helmholtz Metadata Collaboration (HMC) with an annual budget of 4.9 million Euros. HMC provides funding for innovative metadata approaches in a competitive selection process. The evaluation is conducted by an international panel of experts.
- *International Scientific Committee of the CCSD* (D. Castelli) – The Center for Direct Scientific Communication (CCSD) is a French organization providing the higher education and research community with the tools needed to archive, disseminate and capitalise on scientific publications and data. The international scientific committee is made of 11 qualified personalities, French and international experts in the fields of open science, publication, open archives and research data that provide advices and recommendations, contribute to the scientific and technological watch and to the international visibility of the programs of the CCSD, and advise on partnership prospects with third parties.
- *Research Data Alliance - Scientific Knowledge Graphs – Interoperability Framework (SKG-IF) WG* (A. Bardi, M. Baglioni, P. Manghi) – a WG targeting the definition of a framework to enable a seamless exchange of

information among diverse initiatives regarding Scientific Knowledge Graphs, intended as knowledge bases of scholarly knowledge content (e.g., repositories, databases, catalogues, knowledge graphs, LOD collections).

- *OpenAIRE AMKE Services and Technologies Standing Committee* (C. Atzori, A. Bardi, M. Baglioni) – a committee providing the strategic framework to define, assess, expand, maintain and improve the OpenAIRE services and enhance their interoperability with international, national, regional, and sub-regional services.

11. Conclusion

This report documented the research activities carried out by the InfraScience research group at the National Research Council of Italy - Institute of Information Science and Technologies (CNR - ISTI) during the year 2023. Throughout the year, the group has made significant strides in addressing the complexities associated with Data Infrastructures, eScience, and Intelligent Systems.

InfraScience has actively contributed to the advancement of open science practices by leading the development of key infrastructures such as D4Science and OpenAIRE, which facilitate collaboration and enhance the accessibility of scientific resources. The research conducted during this year has resulted in a substantial number of publications, participation in various research projects, and organization of conferences and training events, highlighting our commitment to open and transparent scientific methodologies.

Additionally, the group's involvement in several working groups and task forces underscores our dedication to fostering a collaborative environment conducive to innovation and knowledge sharing within the scientific community.

Looking ahead, InfraScience aims to build on these accomplishments, further enhancing the technological foundations that support data-centered research and promoting sustainable practices in open science. We remain committed to exploring new avenues for collaboration and innovation, ensuring that our contributions continue to evolve in response to the dynamic landscape of scientific research.

As we move into the future, we anticipate greater advancements in the integration of AI, data sharing, and infrastructure development, positioning InfraScience at the forefront of these transformative changes in the scientific ecosystem.

Acknowledgments

InfraScience received funding from the European Union's Horizon 2020 research and innovation programme under: Blue Cloud project (grant agreement No. 862409), DESIRA project (grant agreement No. 818194), EcoScope project (grant agreement No. 101000302), EOSC Future project (grant agreement No. 101017536), I-GENE project (grant agreement No. 862714), MOVING project (grant agreement No. 862739),

OpenAIRE Nexus project (grant agreement No. 101017452), PerformFISH project (grant agreement No. 727610), RISIS 2 project (grant agreement No. 824091), SoBigData-PlusPlus project (grant agreement No. 871042), TAILOR (grant agreement No. 952215).

InfraScience received funding from the European Union's Horizon Europe research and innovation programme under: Blue-Cloud 2026 (grant agreement No. 101094227), CODECS project (grant agreement No. 101060179), CRAFT-OA project (grant agreement No. 101094397), EOSC Future (grant agreement No. 101017536), FAIRCORE4EOSC project (grant agreement No. 101057264), GraspOS (grant agreement No. 101095129), SciLake (grant agreement No. 101058573), Skills-4EOSC project (grant agreement No. 101058527), SoBig-Data RI PPP project (grant agreement No. 101079043).

References

- [1] L. R. Abucay, P. Sorongon-Yap, K. Kesner-Reyes, E. C. Capuli, R. B. Reyes, E. Daskalaki, C. Ferrá, G. Scarcella, G. Coro, F. Ordines, P. Sánchez-Zulueta, G. Dakalov, S. Klayn, L. Celie, M. Scotti, D. Grémillet, C. Lambert, G. Gal, M. L. D. Palomares, D. Dimarchopoulou, and A. C. Tsikliras. Scientific knowledge gaps on the biology of non-fish marine species across European Seas. *Frontiers in Marine Science*, 10:1198137, Oct. 2023. ISSN 2296-7745. doi: 10.3389/fmars.2023.1198137.
- [2] 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*, 18(1):129–140, June 2023. ISSN 1972-621X. doi: 10.36181/digitalia-00064.
- [3] M. Artini, L. Candela, P. Manghi, and S. Giannini. Re-posgate: Open science gateways for institutional repositories. In M. Ceci, S. Ferilli, and A. Poggi, editors, *Digital Libraries: The Era of Big Data and Data Science*, pages 151–162, Cham, 2020. Springer International Publishing. ISBN 978-3-030-39905-4. doi: 10.1007/978-3-030-39905-4_15.
- [4] M. Assante, L. Candela, D. Castelli, R. Cirillo, G. Coro, L. Frosini, L. Lelii, F. Mangiacrapa, V. Marioli, P. Pagano, G. Panichi, C. Perciante, and F. Sinibaldi. The gcube system: Delivering virtual research environments as-a-service. *Future Generation Computer Systems*, 95(n.a.):445–453, 2019. doi: 10.1016/j.future.2018.10.035.
- [5] 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*, 101: 555–563, Dec. 2019. ISSN 0167739X. doi: 10.1016/j.future.2019.05.063.


- [6] M. Baglioni, A. Bardi, H. Dimitropoulos, C. Atzori, and P. Manghi. OpenAIRE Research Graph Dump: new collected projects, Oct. 2023.
- [7] M. Baglioni, A. Bardi, H. Dimitropoulos, C. Atzori, and P. Manghi. OpenAIRE Research Graph Dump: new collected projects, Nov. 2023.
- [8] M. Baglioni, A. Bardi, H. Dimitropoulos, C. Atzori, and P. Manghi. OpenAIRE Research Graph Dump: new collected projects, Jan. 2023.
- [9] M. Baglioni, A. Bardi, H. Dimitropoulos, C. Atzori, and P. Manghi. OpenAIRE Research Graph Dump: new collected projects, Feb. 2023.
- [10] M. Baglioni, A. Bardi, H. Dimitropoulos, C. Atzori, and P. Manghi. OpenAIRE Research Graph Dump: new collected projects, Apr. 2023.
- [11] M. Baglioni, A. Bardi, H. Dimitropoulos, C. Atzori, and P. Manghi. OpenAIRE Research Graph Dump: new collected projects, Apr. 2023.
- [12] M. Baglioni, A. Bardi, H. Dimitropoulos, C. Atzori, and P. Manghi. OpenAIRE Research Graph Dump: new collected projects, Aug. 2023.
- [13] M. Baglioni, A. Mannocci, G. Pavone, M. De Bonis, and P. Manghi. (semi)automated disambiguation of scholarly repositories. In A. Bardi, A. Falcon, S. Ferilli, S. Marchesin, and D. Redavid, editors, *Proceedings of the 19th Conference on Information and Research Science Connecting to Digital and Library Science*. CEUR-WS.org, 2023.
- [14] A. Bardi and L. Benassi. Boosting open science in the IPERION HS research infrastructure with OpenAIRE. *ERCIM News*, 133, April 2023.
- [15] A. Bardi, M. Assante, and F. Mangiacrapa. Ariadne: A data infrastructure for the archaeological research community. *ERCIM News*, 133, April 2023.
- [16] A. Bardi, M. Bargheer, and P. Manghi. A discovery hub for diamond open access publishing. In A. Bardi, A. Falcon, S. Ferilli, S. Marchesin, and D. Redavid, editors, *Proceedings of the 19th Conference on Information and Research Science Connecting to Digital and Library Science, Bari, Italy, February 23-24, 2023*, volume 3365, pages 162–166, 2023.
- [17] A. Bardi, I. Kuchma, G. Pavone, M. Artini, C. Atzori, A. Bäcker, M. Baglioni, A. Czerniak, M. De Bonis, H. Dimitropoulos, I. Fofoulas, M. Horst, K. Iatropoulou, A. Kokogiannaki, S. La Bruzzo, E. Lazzeri, P. Manghi, A. Mannocci, N. Manola, J. Schirrwagen, T. Vergoulis, S. Chatzopoulos, and D. Pierrakos. OpenAIRE Covid-19 publications, datasets, software and projects metadata., Aug. 2023.
- [18] A. Bardi, P. Manghi, A. Mannocci, E. Ottonello, and G. Pavone. A Primer on Open Science-Driven Repository Platforms. In E. Garoufallou and A. Vlachidis, editors, *Metadata and Semantic Research*, volume 1789, pages 222–234. Springer Nature Switzerland, Cham, 2023. ISBN 978-3-031-39140-8 978-3-031-39141-5. doi: 10.1007/978-3-031-39141-5_19. Series Title: Communications in Computer and Information Science.
- [19] V. Bartalesi, G. Coro, E. Lenzi, P. Pagano, and N. Pratelli. From unstructured texts to semantic story maps. *International Journal of Digital Earth*, 16(1):234–250, Oct. 2023. ISSN 1753-8947, 1753-8955. doi: 10.1080/17538947.2023.2168774.
- [20] V. Bartalesi, G. Coro, E. Lenzi, N. Pratelli, and P. Pagano. Towards digital twins of territories through semantic story maps. In *Proceedings of the BUILDing a Digital Twin: requirements, methods, and applications workshop, BUILD-IT 2023*, 2023.
- [21] V. Bartalesi, G. Coro, E. Lenzi, N. Pratelli, P. Pagano, F. Felici, M. Moretti, and G. Brunori. Using semantic story maps to describe a territory beyond its map. *Semantic Web*, 14(6):1255–1272, Dec. 2023. ISSN 22104968, 15700844. doi: 10.3233/SW-233485.
- [22] A. Burton, A. Aryani, H. Koers, P. Manghi, S. La Bruzzo, M. Stocker, M. Diepenbroek, U. Schindler, and M. Fenner. The Scholix Framework for Interoperability in Data-Literature Information Exchange. *D-Lib Magazine*, 23 (1/2), Jan. 2017. ISSN 1082-9873. doi: 10.1045/january2017-burton.
- [23] L. Candela, D. Castelli, and P. Pagano. D4SCIENCE: a unique infrastructure delivering virtual research environments as a service. *ERCIM News*, 133, 2023.
- [24] L. Candela, D. Castelli, and P. Pagano. The D4Science Experience on Virtual Research Environment Development. *Computing in Science & Engineering*, 25(2):12–19, Mar. 2023. ISSN 1521-9615, 1558-366X. doi: 10.1109/MCSE.2023.3290433.
- [25] L. Candela, R. Cirillo, F. Mangiacrapa, P. Pagano, F. Sinibaldi, and G. L. Vannini. The gcube geoportal platform. Technical Report 012, Istituto di Scienza e Tecnologie dell’Informazione “A. Faedo”, 2023.
- [26] G. Casini and U. Straccia. Defeasible RDFS via rational closure. *Information Sciences*, 643:118409, Sept. 2023. ISSN 00200255. doi: 10.1016/j.ins.2022.11.165.
- [27] G. Casini, T. Meyer, and I. Varzinczak. Situated conditional reasoning. *Artificial Intelligence*, 319:103917, June 2023. ISSN 00043702. doi: 10.1016/j.artint.2023.103917.


- [28] D. Castelli, G. De Simone, F. Cancedda, L. Candela, V. Colcelli, R. Conte, F. Di Donato, S. Giannini, E. Lazzeri, S. Mangiaracina, R. Puccinelli, and M. A. Ranchino. Roadmap scienza aperta. Technical report, Consiglio Nazionale delle Ricerche, 2023.
- [29] S. A. Ciliberti, E. Alvarez Fanjul, J. Pearlman, K. Wilmer-Becker, P. Bahurel, F. Ardhuin, A. Arnaud, M. Bell, S. Berthou, L. Bertino, A. Capet, E. Chassignet, S. Ciavatta, M. Cirano, E. Clementi, G. Cosarini, G. Coro, S. Corney, F. Davidson, M. Drevillon, Y. Drillet, R. Dussurget, G. El Serafy, K. Fennel, M. Garcia Sotillo, P. Heimbach, F. Hernandez, P. Hogan, I. Hoteit, S. Joseph, S. Josey, P.-Y. Le Traon, S. Libralato, M. Mancini, P. Matte, A. Melet, Y. Miyazawa, A. M. Moore, A. Novellino, A. Porter, H. Regan, L. Romero, A. Schiller, J. Siddorn, J. Staneva, C. Thomas-Courcoux, M. Tonani, J. M. Garcia-Valdecasas, J. Veitch, K. von Schuckmann, L. Wan, J. Wilkin, and R. Zufic. Evaluation of operational ocean forecasting systems from the perspective of the users and the experts. *State of the Planet*, 1-osr7:2, 2023. doi: 10.5194/sp-1-osr7-2-2023.
- [30] G. Coro. An Open Science oriented Bayesian interpolation model for marine parameter observations. *Environmental Modelling & Software*, 172:105901, Jan. 2024. ISSN 13648152. doi: 10.1016/j.envsoft.2023.105901. URL <https://linkinghub.elsevier.com/retrieve/pii/S1364815223002876>.
- [31] G. Coro, S. Bardelli, A. Cuttano, R. T. Scaramuzzo, and M. Ciantelli. A self-training automatic infant-cry detector. *Neural Computing and Applications*, 35(11):8543–8559, Apr. 2023. ISSN 0941-0643, 1433-3058. doi: 10.1007/s00521-022-08129-w.
- [32] G. Coro, P. Bove, and K. Kesner-Reyes. Global-scale parameters for ecological models. *Scientific Data*, 10(1):7, Jan. 2023. ISSN 2052-4463. doi: 10.1038/s41597-022-01904-3.
- [33] G. Coro, L. Sana, C. Ferrà, P. Bove, and G. Scarcella. Estimating hidden fishing activity hotspots from vessel transmitted data. *Frontiers in Sustainable Food Systems*, 7:1152226, May 2023. ISSN 2571-581X. doi: 10.3389/fsufs.2023.1152226.
- [34] M. De Bonis, F. Falchi, and P. Manghi. Graph-based methods for Author Name Disambiguation: a survey. *PeerJ Computer Science*, 9:e1536, Sept. 2023. ISSN 2376-5992. doi: 10.7717/peerj-cs.1536.
- [35] M. De Bonis, F. Minutella, F. Falchi, and P. Manghi. A graph neural network approach for evaluating correctness of groups of duplicates. In O. Alonso, H. Cousijn, G. Silvello, M. Marrero, C. Teixeira Lopes, and S. Marchesin, editors, *Linking Theory and Practice of Digital Libraries*, pages 207–219, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-43849-3.
- [36] S. F. Di Gennaro, G. L. Vannini, A. Berton, R. Dainelli, P. Toscano, and A. Matese. Missing Plant Detection in Vineyards Using UAV Angled RGB Imagery Acquired in Dormant Period. *Drones*, 7(6):349, May 2023. ISSN 2504-446X. doi: 10.3390/drones7060349.
- [37] R. Froese, G. Coro, M. L. D. Palomares, N. Bailly, M. Scotti, T. Froese, C. Garilao, and D. Pauly. A simple framework for the exploration of functional biodiversity. *Cybium*, 47(3):271–286, 2023. doi: 10.26028/CYBIUM/2023-003. Publisher: Cybium.
- [38] R. Froese, H. Winker, G. Coro, M.-L. D. Palomares, A. C. Tsikliras, D. Dimarchopoulou, K. Touloumis, N. Demirel, G. M. S. Vianna, G. Scarcella, R. Schijns, C. Liang, and D. Pauly. New developments in the analysis of catch time series as the basis for fish stock assessments: The CMSY++ method. *Acta Ichthyologica et Piscatoria*, 53:173–189, Oct. 2023. ISSN 1734-1515, 0137-1592. doi: 10.3897/aiep.53.105910.
- [39] J. Heyninck, G. Casini, T. Meyer, and U. Straccia. Revising Typical Beliefs: One Revision to Rule Them All. In *Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning*, pages 355–364, 8 2023. doi: 10.24963/kr.2023/35. URL <https://doi.org/10.24963/kr.2023/35>.
- [40] A. S. T. Ibrahim and L. Candela. Janet: a prototype of a conversational agent for a virtual research environment. Technical Report 006, Istituto di Scienza e Tecnologie dell’Informazione “A. Faedo”, 2023.
- [41] A. S. T. Ibrahim and L. Candela. Conversational agents for virtual research environments: a survey of the literature. Technical Report 007, Istituto di Scienza e Tecnologie dell’Informazione “A. Faedo”, 2023.
- [42] O. Irrera, A. Mannocci, P. Manghi, and G. Silvello. Tracing data footprints: Formal and informal data citations in the scientific literature. In O. Alonso, H. Cousijn, G. Silvello, M. Marrero, C. Teixeira Lopes, and S. Marchesin, editors, *Linking Theory and Practice of Digital Libraries*, pages 79–92, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-43849-3.
- [43] O. Irrera, A. Mannocci, P. Manghi, and G. Silvello. A Novel Curated Scholarly Graph Connecting Textual and Data Publications. *Journal of Data and Information Quality*, 15(3):1–24, Sept. 2023. ISSN 1936-1955, 1936-1963. doi: 10.1145/3597310.
- [44] P. Manghi, M. Artini, C. Atzori, A. Bardi, A. Mannocci, S. La Bruzzo, L. Candela, D. Castelli, and P. Pagano.

The D-NET software toolkit: A framework for the realization, maintenance, and operation of aggregative infrastructures. *Program*, 48(4):322–354, 2014. doi: 10.1108/PROG-08-2013-0045.


- [45] P. Manghi, C. Atzori, A. Bardi, M. Baglioni, H. Dimitropoulos, S. La Bruzzo, I. Foufloulas, A. Mannocci, M. Horst, K. Iatropoulou, A. Kokogiannaki, M. De Bonis, M. Artini, A. Lempesis, A. Ioannidis, N. Manola, P. Principe, T. Vergoulis, S. Chatzopoulos, and D. Pierrakos. OpenAIRE Research Graph Dataset, Oct. 2023.
- [46] P. Manghi, C. Atzori, A. Bardi, M. Baglioni, J. Schirrwagen, H. Dimitropoulos, S. La Bruzzo, I. Foufloulas, A. Czerniak, M. Horst, K. Kiatropoulou, A. Kokogiannaki, M. De Bonis, M. Artini, A. Lempesis, A. Mannocci, A. Ioannidis, T. Vergoulis, S. Chatzopoulos, and D. Pierrakos. OpenAIRE Graph: Dataset of funded products, Aug. 2023.
- [47] P. Manghi, C. Atzori, A. Bardi, M. Baglioni, J. Schirrwagen, H. Dimitropoulos, S. La Bruzzo, I. Foufloulas, A. Mannocci, M. Horst, A. Czerniak, K. Iatropoulou, A. Kokogiannaki, M. De Bonis, M. Artini, A. Lempesis, A. Ioannidis, N. Manola, P. Principe, T. Vergoulis, S. Chatzopoulos, and D. Pierrakos. OpenAIRE Research Graph Dataset, Aug. 2023.
- [48] P. Manghi, C. Atzori, A. Bardi, M. Baglioni, J. Schirrwagen, H. Dimitropoulos, S. La Bruzzo, I. Foufloulas, A. Mannocci, M. Horst, A. Czerniak, K. Kiatropoulou, A. Kokogiannaki, M. De Bonis, M. Artini, A. Lempesis, A. Ioannidis, T. Vergoulis, S. Chatzopoulos, and D. Pierrakos. OpenAIRE Graph: Dumps for research communities and initiatives., Aug. 2023.
- [49] B. Mathiak, N. Juty, A. Bardi, J. Colomb, and P. Kraker. What are Researchers’ Needs in Data Discovery? Analysis and Ranking of a Large-Scale Collection of Crowdsourced Use Cases. *Data Science Journal*, 22:3, Feb. 2023. ISSN 1683-1470. doi: 10.5334/dsj-2023-003.
- [50] A. Orro, P. D’Ursi, P. Fossa, L. Candela, and G. Panichi. Implementation of a drug discovery pipeline on the d4science platform. Technical Report 001, Istituto di Scienza e Tecnologie dell’Informazione “A. Faedo”, 2023.
- [51] E. Papadopoulou, A. Bardi, G. Kakalettris, D. Tziotzios, P. Manghi, and N. Manola. Data management plans as linked open data: exploiting ARGOS FAIR and machine actionable outputs in the OpenAIRE research graph. *Journal of Biomedical Semantics*, 14(1):17, Nov. 2023. ISSN 2041-1480. doi: 10.1186/s13326-023-00297-5.
- [52] M. T. Paratore and B. Leporini. Exploiting the haptic and audio channels to improve orientation and mobility apps for the visually impaired. *Universal Access in the Information Society*, Feb. 2023. ISSN 1615-5289, 1615-5297. doi: 10.1007/s10209-023-00973-4.
- [53] M. T. Paratore and B. Leporini. Haptic-Based Cognitive Mapping to Support Shopping Malls Exploration. In I. M. Pires, E. Zdravevski, and N. C. Garcia, editors, *Smart Objects and Technologies for Social Goods*, volume 476, pages 54–62. Springer Nature Switzerland, Cham, 2023. ISBN 978-3-031-28812-8 978-3-031-28813-5. doi: 10.1007/978-3-031-28813-5_4. Series Title: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering.
- [54] D. Schaap, S. Pittonet, and P. Pagano. Blue-cloud-2026, a federated european ecosystem to deliver FAIR & open data and analytical services, instrumental for the digital twins of the oceans. In D. Eparkhina and J. Nolan, editors, *Proceedings of the 10th EuroGOOS International Conference*, pages 304–311, 2023.
- [55] C. Senette, M. C. Buzzi, and M. T. Paratore. Self-assess momentary mood in mobile devices: a case study with mature female participants. In *3rd International Conference on Intelligent Communication and Computational Techniques, January 19-20, 20203 Jaipur*. 2023.
- [56] C. Thanos, C. Meghini, V. Bartalesi, and G. Coro. An exploratory approach to data driven knowledge creation. *Journal of Big Data*, 10(1):29, Mar. 2023. ISSN 2196-1115. doi: 10.1186/s40537-023-00702-x.
- [57] A. C. Tsikliras, G. Coro, G. Daskalov, D. Grémillet, M. Scotti, and G. Sylaios. Editorial: Ecocentric fisheries management in European seas: Data gaps, base models and initial assessments, volume I. *Frontiers in Marine Science*, 10:1295733, Oct. 2023. ISSN 2296-7745. doi: 10.3389/fmars.2023.1295733.


InfraScience Members


Michele Artini  is a member of the Technical Staff at the Istituto di Scienza e Tecnologie dell’Informazione “A. Faedo” (ISTI), an institute of the Italian National Research Council (CNR). His skills concern Digital Libraries, e-Infrastructures, Data Management, Web Services, Web Applications and Mobile Applications. Michele joined ISTI in 2005, he worked for several EU Projects such as DELOS, DRIVER, EFG and OpenAIRE, currently, he is working in OpenAIRE-Nexus (EU H2020).


Massimiliano Assante  is Senior Technology Researcher (Primo Tecnologo) at the Istituto di Scienza e Tecnologie della Informazione “A. Faedo” (ISTI), an institute of the Italian National Research Council (CNR). He has strong theoretical foundation in Computer Science and Technology, encompassing mathematics, logic, and cross-platform coding, and

over 15 years of experience working on distributed systems, e-infrastructures and Virtual Research Environments. Massimiliano's academic journey includes a Ph.D. in Information Engineering, a Master's degree (M.Sc.) in Information Technologies, and a Bachelor's degree (B.Sc.) in Computer Science, all earned from the University of Pisa. Throughout his career, Massimiliano has been actively involved in several European Union (EU) projects, including ARIADNE-Plus, Blue-Cloud, AGINFRAPlus, BlueBRIDGE, DESIRA, PARTHENOS, SoBigData, EOSCPilot, PerformFISH, iMarine, EU-BrazilOpenBio, D4Science II, D4Science, and DILIGENT. He has held various roles in these projects, showcasing his versatility and adaptability. Currently, Massimiliano is engaged in multiple EU projects with responsibility roles, including Blue-Cloud 2026, FOSSR PNRR, SoBigData-Plus-Plus, MOVING, and RISIS2.

Claudio Atzori  is a computer science researcher at the National Research Council of Italy, Istituto di Scienza e Tecnologia dell'Informazione "A. Faedo". His research activity focuses on digital library management systems, data curation in digital libraries, autonomic service-oriented data infrastructures, and the disambiguation of digital objects in big data graphs. Moreover, he has participated in several EC funded R&D projects: DRIVER-II, EFG, EFG1914, HOPE, EAGLE, OpenAIRE, OpenAIRE-Plus, OpenAIRE2020, OpenAIRE-Advance, OpenAIRE-Connect, OpenAIRE-Nexus, Data4Impact, EOSC-Future as developer, software architect, data analyst, task and work package leader, where his work contributed to the realisation of aggregative data infrastructures for e-science and scholarly communication.


Miriam Baglioni  is a (PhD) researcher at InfraScience Laboratory of the Italian National Research Council - Institute of Information Science and Technologies (CNR-ISTI) since 2016. She is currently participating in the EU funded projects OpenAIRE-Nexus, Ariadne Plus and RISIS2. She has worked on Data Mining, Knowledge Discovery, ontologies, social networks and bioinformatics. Her current research interests include data e-infrastructure for science, and science reproducibility.


Alessia Bardi  is a PhD researcher in computer science at the Institute of Information Science and Technologies of the Italian National Research Council. She has been involved in EC funded projects for the realisation and operation of aggregative data infrastructures for research communities in the Humanities and Studies of the past (e.g., HOPE - Heritage of the People's Europe, PARTHENOS, Ariadne+) and for the realization of Open Science services like OpenUP, EOSC Future and OpenAIRE projects. In particular, for OpenAIRE she also has the product manager role for the OpenAIRE CONNECT service. Her research interests include service-oriented architectures, data and metadata interoperability and data infrastructures for e-science and scholarly communication.

Pasquale Bove  is a PhD researcher in computer science


at the Institute of Information Science and Technologies of the Italian National Research Council. His research focuses on Data Mining and Ecological Niche Modeling. His work is currently focused on the experimentation of models and methodologies to process biological and environmental data, especially in the marine field, with an Open Science and science reproducibility-oriented approach.

Catherine Bosio is a member of the Technical Staff at the "A. Faedo" Institute of Information Science and Technologies (ISTI). She is French mother tongue and graduated in Applied foreign languages at the University of Angers (France). She joined ISTI in 2004. Her skills concern administrative management, she worked for several EU projects (FP6, FP7, Horizon 2020 and Horizon Europe) such as Diligent, D4Science and D4Science II, Belief and Belief II, Driver and Driver II, Envri and Envriplus, Venus-C, RDA and RDA2, EU Brazil OpenBio, BlueBridge, Ecoscope, Blue-Cloud and Blue-Cloud 2026, SoBigData, EoscSecretariat, EoscPillar. Over the years she gave also logistics support to the organization of more than 50 events.


Antonio Calanducci  is a senior research fellow at the Institute of Science and Information Technologies "A. Faedo" of the National Research Council of Italy. He holds a PhD in Computer Science and has been working in research institutions (INFN, INAF, CNR) since 2005 supporting the activities of scientific research communities belonging to very different fields, such as Cultural Heritage, Astrophysics, Earth Observation, Agriculture, Health and Life science, High Energy Physics in the context of several EU funded projects. He recently joined the InfraScience team as a Cloud Infrastructure Engineer, supporting the Technical Staff in the management and provisioning of the D4Science Infrastructure with Infrastructure-as-Code approaches. He has 20+ years of experience in the field of grid/cloud computing and distributed data infrastructure, covering different roles, such as community consultant, software engineer, dissemination officer, developer relation and technical instructor. Since 2009 he is a contract professor of Computer Science at the University of Catania, teaching courses on mobile app development, Python programming, Database Management Systems. He was a mentor at the Apple Developer Academy and owned a company for several years delivering cross-platform mobile app development training and consultancy to public and private companies and enterprises.


Leonardo Candela  is computer science senior researcher at the National Research Council of Italy, Institute of Information Science and Technologies. His research interests are driven by the development of systems and services supporting research infrastructures for science. In particular, he is intertwining virtual research environments, data infrastructures, collaborative working environments, reference models for complex systems, information retrieval, data analytics, data publishing and innovative scholarly communication practices. His research activity is developed by closely con-


necting research and development. In fact, he has been involved in several EU-funded projects called to develop Digital Libraries & Data Infrastructures and is the Strategy and Portfolio Manager of the D4Science.org infrastructure.

Giovanni Casini  is a researcher at the National Research Council of Italy, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo". His main research topic is Knowledge Representation and Reasoning, with a particular focus on logical formalisms for uncertain reasoning, belief change, and the Semantic Web. Previously he has worked as a researcher at Scuola Normale Superiore, CSIR (South Africa), University of Pretoria (South Africa), and University of Luxembourg (Luxembourg).


Donatella Castelli  is Research Director at Istituto di Scienza e Tecnologie dell'Informazione, "A. Faedo" of the National Research Council of Italy where she leads the InfraScience research team. Under her supervision, the InfraScience team coordinated and participated in several EU and nationally funded projects on Digital Libraries and Research Data Infrastructures. She has participated as an expert to the shaping of the Italian National Plan for Open Science and she is currently the Italian member of the EU Group of National contact points for scientific Information. Her research interests include open science data infrastructures and open science scientific approaches. She authored several research papers in these fields.


Roberto Cirillo  is researcher at the Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy. His scientific and professional activity involves the research and development on Data Infrastructures. His research interests include e-Infrastructures, Cloud-based technologies, Virtual Research Environments and NoSQL Data Stores. He is currently member of the BlueCloud EU Project. He was involved in various EU-funded projects including BlueBridge, iMarine, EUBrazil-OpenBio, ENVRI, EGI-Engage. In the past, he has been working on Language Technologies.


Giampaolo Coro  is a Physicist with a Ph.D. in Computer Science. His research focuses on Artificial Intelligence, Data Mining and e-Infrastructures. Since 2002, he works on machine learning and signal processing with applications to computational biology, brain-computer interfaces, language technologies and cognitive sciences. The aim of his research is the study and experimentation of models and methodologies to process biological data with an Open Science oriented approach. His approach relies on distributed e-Infrastructures and uses parallel and distributed computing via Cloud-based technologies.


Michele De Bonis  is a research fellow at the Institute of Science and Information Technologies 'A Faedo' (ISTI) of the CNR of Pisa, and a PhD student of Information Engineering at the University of Pisa. He graduated in Computer Science at the University of Pisa and his research focuses on


entity deduplication on big scholarly communication graphs. In particular, the aim of his studies is to find solutions for author name disambiguation and entity linking based on Artificial Intelligence and Deep Learning techniques. Michele joined ISTI in 2017 and he worked for the projects in the OpenAIRE Infrastructure.

Franca Debole  is a researcher at the Institute of Science and Information Technologies "A. Faedo" of the CNR of Pisa. Graduated in Computer Science at the University of Pisa, she received a PhD in Information Engineering. He has participated in international and national research projects in the field of information retrieval, in the creation of content management systems for multimedia digital libraries and in the field of multilingual search engines. Over the years, she has been a technical director and involved in several European and National projects. Her current research activities range from digital image processing to techniques for image retrieval and automatic annotation tools. Her technical knowledge ranges from design tools stand-alone to web programming techniques. She is also head of a group for IT infrastructure at ISTI-CNR.


Andrea Dell'Amico  is a member of the Technical Staff at the Istituto di Scienza e Tecnologie dell'Informazione A. Faedo (ISTI), an institute of the Italian National Research Council (CNR). His skills concern systems administration and integration, automation of systems and services provisioning, configuration and maintenance of large compute and storage infrastructures. He manages the computing and storage facilities of the D4Science.org project. Andrea joined ISTI in 2013 and worked on several EU projects such as BlueBRIDGE, OpenAIRE, Parthenos.


Luca Frosini  is researcher at the Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy. He has relevant expertise in the area of Virtual Research Environments development. He was involved in various EU-funded projects, including DILIGENT, D4Science, EAGLE, PARTHENOS, SoBigData and BlueBRIDGE. His research interests include Data Infrastructures, Virtual Research Environments, Information Systems, Accounting Systems, and Grid and Cloud Computing.


Ahmed Salah Tawfik Ibrahim  is a computer engineer and a researcher at the National Research Council of Italy, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo". With a master's degree in artificial intelligence, his research work focuses on equipping virtual research environments with artificial intelligence capabilities as part of the SoBigData RI PPP project. His interests include information retrieval, large language models, conversational agents and multimodal models.


Sandro La Bruzzo  is a member of the Technical Staff at the Institute of Information Science and Technologies "Alessandro Faedo" (ISTI). His skills concern Big Data, Data Analytics & Data infrastructure, Data curation, and aggregation.


He is the technical manager of Scholexplere Service. Sandro joined ISTI in 2010; he worked for several EU Projects such as EFG, EAGLE, and OpenAIRE. Currently, he is working in OpenAIRE-Nexus (EU H2020).

Lucio Lelli  is Researcher at the Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy. His scientific and professional activity involves the Research and Development on Data Infrastructures.


Paolo Manghi  is a (PhD) Researcher in computer science at Istituto di Scienza e Tecnologie dell'Informazione (ISTI) of Consiglio Nazionale delle Ricerche (CNR), in Pisa, Italy. He is the CTO of OpenAIRE AMKE, involved in coordination and/or activities in the H2020 projects FAIRCORE4EOSC, EOSC-Future, EOSC-Enhance, OpenAIRE-Nexus, OpenAIRE-Connect, OpenAIRE-Advance, OpenAIRE2020. His research areas of interest are today data e-infrastructures for science and scholarly communication infrastructures, with a focus on technologies supporting open science publishing within and across different disciplines, i.e., computational reproducibility and transparent evaluation of science.


Francesco Mangiacrapa  is a computer scientist and researcher at the Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy. He has a background in geospatial data, technologies, models and standard OGC (like WMS, WFS and so on) for spatial data representation and exchange. His scientific and professional activity includes study and research on Virtual Research Environments and Data Infrastructure, Data Publication, GeoSpatial Data and Open Science. Moreover, his work involves the design and development of (Web-)GUI based on several frameworks (like GWT, Material, Bootstrap and so on) to support his research activity and able to improve community collaboration and exchange of scientific data. Currently, he is working in several EU projects (BlueCloud, SoBigData, PARTHENOS) and is responsible for: Data Access and Exchange (Workspace Area), Data Catalogue and Publishing (Catalogue Area).


Dario Mangione  is a library and information scientist and a graduate fellow at the National Research Council of Italy, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo". His research activity is focused on the study and development of models, solutions, and systems enabling and fostering open and Findable, Accessible, Interoperable, and Reusable (FAIR) practices and ultimately an open science approach. His research interests include semantic Web oriented controlled vocabularies and metadata standards. He has been involved as a terminology expert in the EC-funded EOSC-Secretariat.eu project for supporting the development of standardisation solutions within the scope of the creation of the European Open Science Cloud (EOSC). He is currently working on FAIR digital objects evaluation practices.


Andrea Mannocci  is a research fellow at ISTI-CNR in Italy. He currently works as a data scientist within the frame-

work of the EU project OpenAIRE Nexus. His research interests span from the analysis of enabling services for Open Science, to Science of Science, complex networks and the analysis of research as a global-scale phenomenon inserted in a delicate socioeconomic and geopolitical context. He obtained his Ph.D. degree in Information Engineering from the University of Pisa (Italy) researching on systems for data flow quality monitoring in data infrastructures. He co-organised the international workshop series on Reframing Research (Refresh-2018-2020) held at the European Computational Social Science symposium, and at SocInfo 2020 respectively.


Elisa Molinaro  is a project manager at the National Research Council of Italy, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo". Elisa has an extensive experience in the ICT sector, gained through consultancy roles and within end-user companies. Specialising in the management of IT projects related to Business systems for Finance, HR and Legal areas. Her expertise covers the entire project lifecycle, from conception to implementation and deployment, with a particular focus on project management officer, ERP SAP and Oracle JDEdwards solutions, ensuring the execution and adherence to timelines and budgets. Currently, she supports the research team with project management activities, focusing on the financial aspects, resource allocation to align with project budgets, meeting administrative deadlines, and ensuring quality expectations, in addition to handling the technical aspects of contracts with external parties.


Pasquale Pagano  is Senior Researcher at CNR-ISTI. He has a strong background and experience on models, methodologies and techniques for the design and development of distributed virtual research environments (VREs) which require the handling of heterogeneous computational and storage resources, provided by Grid and Cloud based e-Infrastructures, and management of heterogeneous data sources. He participated in the design of the most relevant distributed systems and e-Infrastructure enabling middleware developed by ISTI - CNR. He is currently the Technical Director of the D4Science Data Infrastructure. In the past, he has been involved in the iMarine, EUBrazilOpenBio, ENVRI, Venus-C, GRDI2020, D4Science-II, D4Science, Diligent, DRIVER, DRIVER II, BELIEF, BELIEF II, Scholnet, Cyclades, and ARCA European projects.


Giancarlo Panichi  is a member of the Technical Staff at the Istituto di Scienza e Tecnologie dell'Informazione A. Faedo (ISTI), an institute of the Italian National Research Council (CNR). His skills concern e-Infrastructures, Web Processing Service, Virtual Research Environments, Data Management, Data Analytics, Web Services, Web Applications and Mobile Applications. Giancarlo joined ISTI in 2013. He worked for several EU Projects including iMarine, BlueBRIDGE, EUBrazilOpenBio and ENVRI. He is currently mainly involved in BlueCloud and EOSC-Pillar projects.


Maria Teresa Paratore  is a technology researcher at the Institute of Information Science and Technologies "Alessan-


dro Faedo” (ISTI), an institute of the Italian Research Council (CNR). She graduated in Telecommunications Engineering and since then she has been involved in several research projects. Maria Teresa’s interests range from neural networks to eHealth. In recent years, she has been involved in research projects focused on accessibility and inclusivity. Her activities are documented in various scientific publications.

Gina Pavone  is a research fellow focusing on Open Science, Open Access and Research Data Management. She is in charge as National Open Access Desk of OpenAIRE and she coordinates the editorial board of open-science.it, the Italian portal dedicated to the many components of Open Science. She is also member OpenAIRE Community of Practice of Training Coordinators and she is involved in the structuring of a national Competence Centre for Open Science, FAIR data and EOSC within the ICDI (Italian Computing and Data Infrastructure). Her activities range from the definition of strategies and tools for the support and training of researchers to the dissemination of Open Science activities and initiatives. She has worked in several international projects such as OpenAIRE, EOSC Pillar, EOSC Secretariat and RDA Europe. She is a journalist with expertise in data analysis, she holds a master’s degree in publishing and journalism at the Sapienza University of Rome and a second-level master’s degree in big data analytics and social mining at the University of Pisa. She has been involved in campaigns for open data and transparency in public institutes and administrations and she has worked as a data analyst and data journalist for the European Data Journalism Network (EDJNet).

Tommaso Piccioli  is a member of the Technical Staff at the A. Faedo Institute of Information Science and Technologies (ISTI). He graduated in Computer Science, with knowledge and responsibility in hardware and software infrastructures design and management, from server farm maintenance to networking, data backup, virtualization environments and systems integration. He was involved since 2005 in the technological support to many projects of the research group including DELOS, Diligent, D4Science and D4Science II, iMarine, EUBrazilOpenBio, various OpenAIRE projects, EFG, PerformFISH, PARTHENOS, BlueBRIDGE, RISIS 2, SoBig-DataPlus, AriadnePlus.

Fabio Sinibaldi  is a Researcher at CNR-ISTI. He holds a degree in computer science engineering with specialization in business management technologies received from the University of Pisa. In his research studies, he worked on designing and developing distributed environments’ services aimed at managing scientific data, with special attention to Ecological Niche Modelling approaches. These studies involved the exploitation of federated Grid and Cloud e-Infrastructures along with Digital Libraries oriented workflow analysis and design, leading to the development of D4Science’s Spatial Data Infrastructure. He currently works as Spatial Data Infrastructure designer for D4Science Data Infrastructure. In the past he has been involved in the iMarine, EAGLE, EU-BrazilOpenBio, ENVRI, Venus-C, D4Science-II, D4Science projects.

Umberto Straccia  is Research Director at ISTI - CNR (the Istituto di Scienza e di Tecnologie dell’Informazione - ISTI, an Institute of the National Research Council of Italy - CNR). He received a Ph.D. in computer science from the University of Dortmund, Germany. His research interests include logics for Knowledge Representation and Reasoning (Description Logics, Logic Programming, Answer Set Programming), Semantic Web Languages (OWL, RDFS, RuleML), Fuzzy Logic, Machine Learning (Statistical Relational Learning, Ontology-based Machine Learning), their combination and application.

Gian Luca Vannini  is a research fellow at the Institute of Information Science and Technologies “A. Faedo” (ISTI) of the National Research Council (CNR), Pisa, specializing in agronomy, urban greenery, landscape, and geospatial data. Engaged in the “SoBigData-PlusPlus” and “InfraScience” programs, his current work focuses on designing and developing computer applications for the management, analysis, and publication of georeferenced data, with a particular emphasis on agronomic, environmental, and archaeological data. His expertise ranges from resolving geolocational issues to standardizing spatial data formats and analyzing geospatial data from remote sensing. His research contributes to Open Science-oriented research infrastructures, strengthening the connection between computer science and environmental disciplines.