2022

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

ISTI Annual Reports

InfraScience Research Activity Report 2021

InfraScience lab., CNR-ISTI, Pisa, Italy

ISTI-AR-2022/001

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

InfraScience Research Activity Report 2021
InfraScience lab.
ISTI-AR-2022/001

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 2021 to highlight the major results. In particular, the InfraScience group confronted with 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 2021 InfraScience members contributed to the publishing of 25 papers, to the research and development activities of 18 research projects (15 funded by EU), to the organization of conferences and training events, to several working groups and task forces.

Infrastructure, Open Science, Intelligent Systems

Citation
InfraScience lab. *InfraScience Research Activity Report 2021*, ISTI Annual Reports 2022/001. DOI: 10.32079/ISTI-AR-2022/001

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

ISTI-AR-2022/001

Technical Report, October 2022
Consiglio Nazionale delle Ricerche - Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"

InfraScience Research Activity Report 2021

Michele Artini, Massimiliano Assante, Claudio Atzori, Miriam Baglioni, Alessia Bardi, Pasquale Bove, Leonardo Candela, Giovanni Casini, Donatella Castelli*, Roberto Cirillo, Gianpaolo Coro, Michele De Bonis, Franca Debole, Andrea Dell'Amico, Luca Frosini, Sandro La Bruzzo, Emma Lazzeri, Lucio Lelii, Paolo Manghi, Francesco Mangiacrapa, Dario Mangione, Andrea Mannocci, Enrico Ottonello, Pasquale Pagano, Giancarlo Panichi, Gina Pavone, Tommaso Piccioli, Fabio Sinibaldi, Umberto Straccia

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 2021 to highlight the major results. In particular, the InfraScience group confronted with 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 2021 InfraScience members contributed to the publishing of 25 papers, to the research and development activities of 18 research projects (15 funded by EU), to the organization of conferences and training events, to several working groups and task forces.

Keywords
Infrastructure — Open Science — Intelligent Systems

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

Contents

1. Introduction

Science is heavily data and compute-intensive, AI-assisted, participatory, and multidisciplinary. Sharing and publishing scientific results are activities subject to profound reconsideration to support openness, transparency and reproducibility, and to enable rewards for scientists who publish results of their work beyond the scientific articles. These approaches are expressions of a profound evolution of science practices that on the one hand is enacted by, and on the other demand for, a continuous innovation in IT instruments and approaches.

InfraScience[1] 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 16 collaborators including postdocs, doctoral students, and research associates.

This report documents the research activity performed by the group in 2021, the resulting publications, the active re-

[1] InfraScience website `infrascience.isti.cnr.it`

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

search 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 give prospects on future research activities.

2. Research topics

The research activities conducted by infraScience members revolves around three major topics: Data Infrastructures, e-Science, and Intelligent Systems.

2.1 Data infrastructures

This is a very broad research area including models, approaches and solutions underlying the development and operation of data infrastructures suitable for thematic and interdisciplinary scientific contexts characterized by variability, heterogeneity, reusability and presence of “big data”. The group is confronting with these challenges by closely connecting research and development. In fact, InfraScience is responsible for the development of two large scale infrastructures supporting open science, namely D4Science and OpenAIRE cf. Sec. 5. The major themes and investigations include approaches and solutions for Virtual Research Environment and Science Gateways, e.g., [7, 14], and approaches for distributed services management including workflows and transactions, e.g., [27, 16].

2.2 eScience

This is a wide research domain including models, approaches and solutions to carry out collaborative data-driven and reproducible analytical workflows while supporting, at the same time, sharing, publishing, validation, and monitoring (usage and impact) of the related scientific outcomes (publications, datasets, software, etc.). The group is confronting with several challenges belonging to the domain including data driven approaches for scientific challenges [2, 28, 23], and quantitative studies on scholarly communication practices, e.g., [8, 9, 36].

2.3 Intelligent Systems

This research area concerns AI-assisted methods and approaches to enable humans and systems to discover, access, process, and learn structured and unstructured information. InfraScience is confronting with challenges including the learning fuzzy concept inclusion axioms from ontologies, e.g., [17], languages for contextual conditional reasoning, e.g., [20], intelligent vision systems, e.g., [22], and speech recognition, e.g., [24].

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 14 articles in journals, 5 papers to conferences, 1 chapter in books, and 5 publications including technical reports and other papers.

3.1 Contributions to Journals

InfraScience members contributed to the following papers published in journals.

ReLock: a resilient two-phase locking RESTful transaction model [27] by Frosini et al. for Service Oriented Computing and Applications.

Summary: Service composition and supporting transactions across composed services are among the major challenges characterizing service-oriented computing. REpresentational State Transfer (REST) is one of the approaches used for implementing Web services that is gaining momentum thanks to its features making it suitable for cloud computing and microservices-based contexts. This paper introduces ReLock, a resilient RESTful transaction model introducing general purpose transactions on RESTful services by a layered approach and a two-phase locking mechanism not requesting any change to the RESTful services involved in a transaction.

Fig. 1 depicts the ReLock transaction management architecture highlighting the layering of the approach.

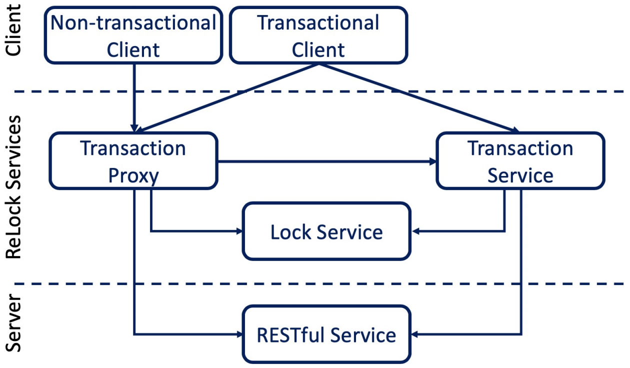
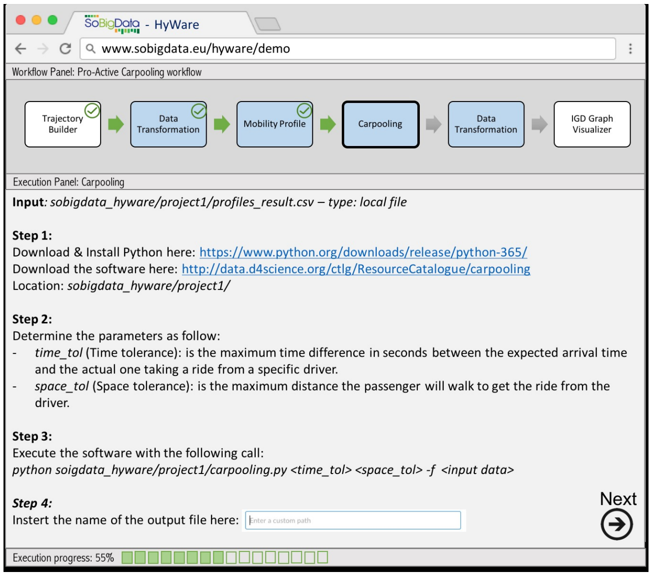


Figure 1. ReLock transaction model architecture [27]

A workflow language for research e-infrastructures [16] by Candela et al. for International Journal of Data Science and Analytics.

Summary: Research e-infrastructures are “systems of systems,” patchworks of resources such as tools and services, which change over time to address the evolving needs of the scientific process. In such environments, researchers carry out their scientific process in terms of sequences of actions that mainly include invocation of web services, user interaction with web applications, user download and use of shared software libraries/tools. The resulting workflows are intended to generate new research products (articles, datasets, methods, etc.) out of existing ones. Sharing a digital and executable representation of such workflows with other scientists would enforce Open Science publishing principles of "reproducibility of science" and "transparent assessment of science." This work presents HyWare, a language and execution platform capable of representing scientific processes in highly heterogeneous research e-infrastructures in terms of so-called hybrid workflows. Hybrid workflows can express sequences of "manually executable actions," i.e., formal descriptions guiding users to repeat a reasoning, protocol or manual procedure, and "machine-executable actions," i.e., encoding of the automated execution of one (or more) web services. An HyWare execution platform enables scientists to (*i*) create and share workflows out of a given action set (as defined by the users to match e-infrastructure needs) and (*ii*) execute hybrid workflows making sure input/output of the actions flow properly across manual and automated actions. The HyWare language and platform can be implemented as an extension of well-known workflow languages and platforms.

Fig. 2 shows a mock of the user interface proposed in the paper.



Input: *sobigdata_hyware/project1/profiles_result.csv – type: local file*

Step 1:
Download & Install Python here: https://www.python.org/downloads/release/python-365/
Download the software here: http://data.d4science.org/ctlg/ResourceCatalogue/carpooling
Location: *sobigdata_hyware/project1/*

Step 2:
Determine the parameters as follow:
- *time_tol* (Time tolerance): is the maximum time difference in seconds between the expected arrival time and the actual one taking a ride from a specific driver.
- *space_tol* (Space tolerance): is the maximum distance the passenger will walk to get the ride from the driver.

Step 3:
Execute the software with the following call:
python soigdata_hyware/project1/carpooling.py <time_tol> <space_tol> -f <input data>

Step 4:
Instert the name of the output file here: Enter a custom path

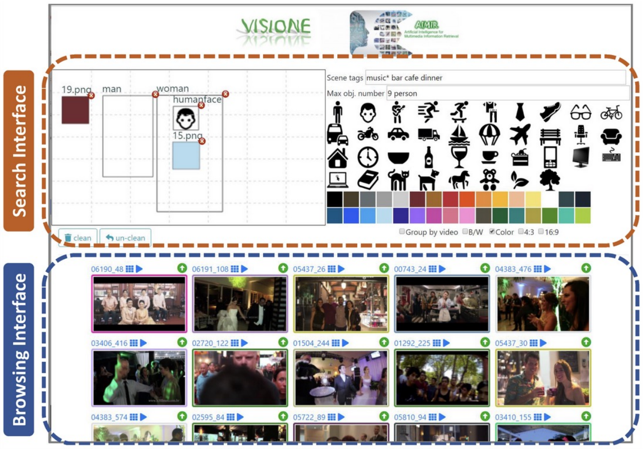
Execution progress: 55% 🟩🟩🟩🟩🟩🟩🟩🟩□□□□□□□□

Figure 2. An example of HyWare web GUI considering the execution of a human action [16]

The VISIONE video search system: exploiting off-the-shelf text search engines for large-scale video retrieval [1] by Amato et al. for Journal of Imaging.

Summary: This paper describes in detail VISIONE, a video search system that allows users to search for videos using textual keywords, the occurrence of objects and their spatial relationships, the occurrence of colors and their spatial relationships, and image similarity. These modalities can be combined together to express complex queries and meet users' needs. The peculiarity of our approach is that we encode all information extracted from the keyframes, such as visual deep features, tags, color and object locations, using a convenient textual encoding that is indexed in a single text retrieval engine. This offers great flexibility when results corresponding to various parts of the query (visual, text and locations) need to be merged. In addition, we report an extensive analysis of the retrieval performance of the system, using the query logs generated during the Video Browser Showdown (VBS) 2019 competition. This allowed us to fine-tune the system by choosing the optimal parameters and strategies from those we tested.

Fig. 3 depicts the VISIONE content-based retrieval system user interface allowing a user to search for a video describing the content of a scene by formulating textual or visual queries.



Scene tags music* bar cafe dinner
Max obj. number 9 person
□Group by video □B/W ☑Color □4:3 □16:9

Figure 3. A screenshot of the VISIONE User Interface composed of two parts: the search and the browsing [1]

Data science: a game changer for science and innovation [28] by Grossi et al. for International Journal of Data Science and Analytics.

Summary: This paper shows data science's potential for disruptive innovation in science, industry, policy, and people's lives. We present how data science impacts science and society at large in the coming years, including ethical problems in managing human behavior data and considering the quantitative expectations of data science economic impact. We introduce concepts such as open science and e-infrastructure as useful tools for supporting ethical data science and training new generations of data scientists. Finally, this work outlines SoBigData Research Infrastructure as an easy-to-access platform for executing complex data science processes. The services proposed by SoBigData are aimed at using data science to understand the complexity of our contemporary, globally interconnected society.

Fig. 4 depicts the data science pipeline discussed in the paper.

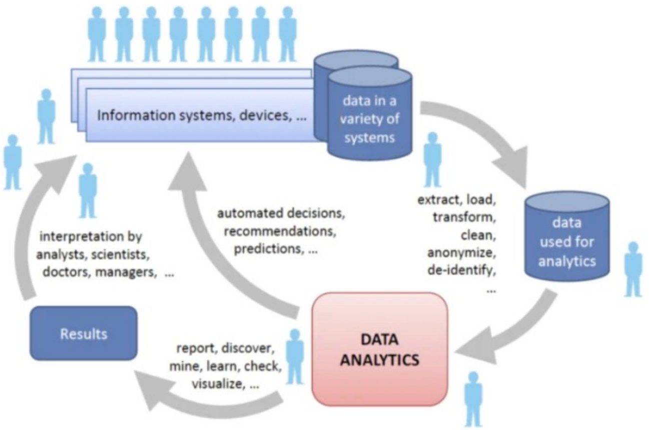

Information systems, devices, ...
data in a variety of systems
extract, load, transform, clean, anonymize, de-identify, ...
data used for analytics
automated decisions, recommendations, predictions, ...
interpretation by analysts, scientists, doctors, managers, ...
DATA ANALYTICS
report, discover, mine, learn, check, visualize, ...
Results

Figure 4. The data science pipeline starts with raw data and transforms them into data used for analytics. The next step is to transform these data into knowledge through analytical methods and then provide results and evaluation measures [28]

Data Poor Approach for the Assessment of the Main Target Species of Rapido Trawl Fishery in Adriatic Sea [2] by Armelloni et al. for Frontiers in Marine Science.

Summary: Information on stock status is available only for a few of the species forming the catch assemblage of rapido fishery of the North-central Adriatic Sea (Mediterranean Sea). Species that are caught almost exclusively by this gear, either as target (such as Pectinidae) or accessory catches (such as flatfishes apart from the common sole), remain unassessed mainly due to the lack of data and biological information. Based on cluster analysis, the catch assemblage of this fishery was identified and assessed using CMSY model. The results of this data-poor methodology showed that, among the species analyzed, no one is sustainably exploited. The single-species CMSY results were used as input to an extension of the same model, to test the effect of four different harvest control rule (HCR) scenarios on the entire catch assemblage, through 15-years forecasts. The analysis showed that the percentage of the stocks that will reach Bmsy at the end of the projections will depend on the HCR applied. Forecasts showed that a reduction of 20% of fishing effort may permit to most of the target and accessory species of the rapido trawl fishery in the Adriatic Sea to recover to B$_{m}sy$ levels within 15 years, also providing a slight increase in the expected catches.

Fig. 5 depicts the forecasts of alternative HCRs from the CMSY extended analysis, performed on the entire catch assemblage.

Fuzzy OWL-Boost: Learning fuzzy concept inclusions via real-valued boosting [17] by Cardillo and Straccia for Fuzzy Sets and Systems.

Summary: OWL ontologies are nowadays a quite popular way to describe structured knowledge in terms of classes, relations among classes and class instances. In this paper, given an OWL ontology and a target class T, we address the problem of learning fuzzy concept inclusion axioms that describe sufficient conditions for being an individual instance of T (and to which degree). To do so, we present Fuzzy OWL-Boost that relies on the $\mathbb{R}$eal AdaBoost boosting algorithm adapted to the (fuzzy) OWL case. We illustrate its effectiveness by means of an experimentation with several ontologies.

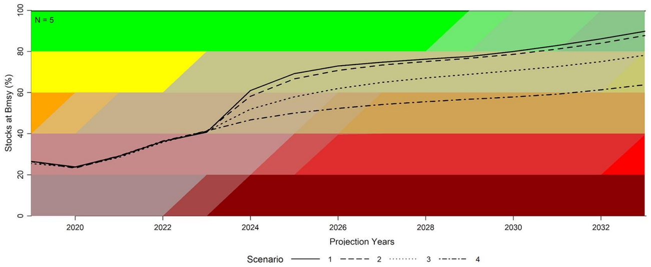


Figure 5. Forecast of alternative HCRs from the CMSY extended analysis on the catch assemblage: percentage of stocks at Bmsy. Stronger the effort reduction, shorter the range of time in which 80% of the stocks will reach the Bmsy. Scen. (1): 50% of effort reduction; Scen. (2): 40% of effort reduction; Scen (3): 20% of effort reduction; Scen. (4): 5% of effort reduction. [2]

Fig. 6 depicts the comparison between Fuzzy OWL-Boost and FOIL-$\mathscr{DL}$.

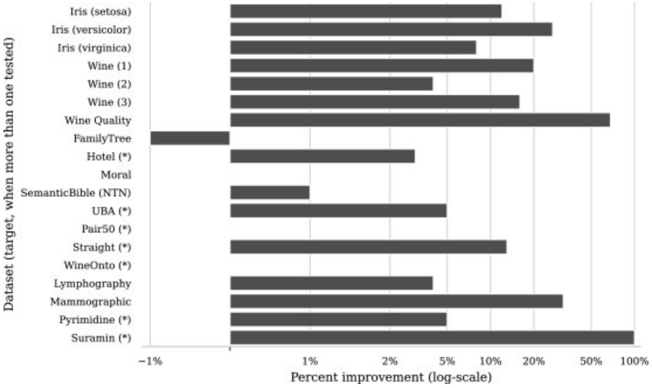


Figure 6. Improvement of Fuzzy OWL-Boost over FOIL-$\mathscr{DL}$ according to the fF1F1 measure. [17]

An intelligent and cost-effective remote underwater video device for fish size monitoring [22] by Coro and Bjerregaard Walsh for Ecological Informatics.

Summary: Monitoring the size of key indicator species of fish is important to understand ecosystem functions, anthropogenic stress, and population dynamics. Standard methodologies gather data using underwater cameras, but are biased due to the use of baits, limited deployment time, and short field of view. Furthermore, they require experts to analyse long videos to search for species of interest, which is time consuming and expensive. This paper describes the Underwater Detector of Moving Object Size (UDMOS), a cost-effective computer vision system that records events of large fishes passing in front of a camera, using minimalistic hardware and power consumption. UDMOS can be deployed underwater, as an unbaited system, and is also offered as a free-to-use Web Service for batch video-processing. It embeds three different alternative large-object detection algorithms based on deep learning, unsupervised modelling, and motion detection, and can work both in shallow and deep waters with infrared or visible light.

Fig. 7 depicts examples of events in the test case videos of large fishes passing in front of the camera and correctly captured by the proposed workflow.

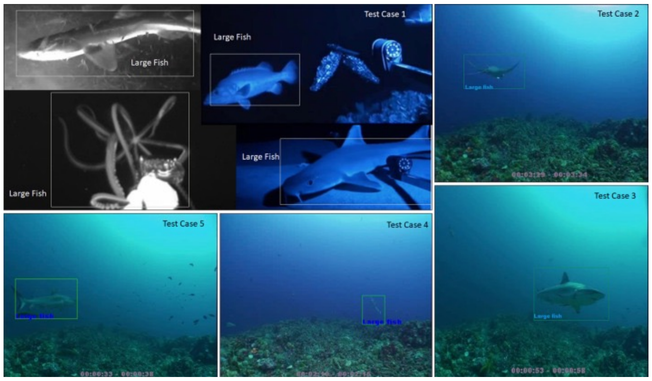


Figure 7. Examples of events in the test case videos of large fishes passing in front of the camera and correctly captured by the proposed workflow. [22]

Psycho-acoustics inspired automatic speech recognition [24] by Coro et al. for Computers & Electrical Engineering.

Summary: Understanding the human spoken language recognition process is still a far scientific goal. Nowadays, commercial automatic speech recognisers (ASRs) achieve high performance at recognising clean speech, but their approaches are poorly related to human speech recognition. They commonly process the phonetic structure of speech while neglecting supra-segmental and syllabic tracts integral to human speech recognition. As a result, these ASRs achieve low performance on spontaneous speech and require enormous costs to build up phonetic and pronunciation models and catch the large variability of human speech. This paper presents a novel ASR that addresses these issues and questions conventional ASR approaches. It uses alternative acoustic models and an exhaustive decoding algorithm to process speech at a syllabic temporal scale (100–250 ms) through a multi-temporal approach inspired by psycho-acoustic studies. Performance comparison on the recognition of spoken Italian numbers (from 0 to 1 million) demonstrates that our approach is cost-effective, outperforms standard phonetic models, and reaches state-of-the-art performance.

Fig. 8 depicts the proposed ASR that uses several possible alternative acoustic models.

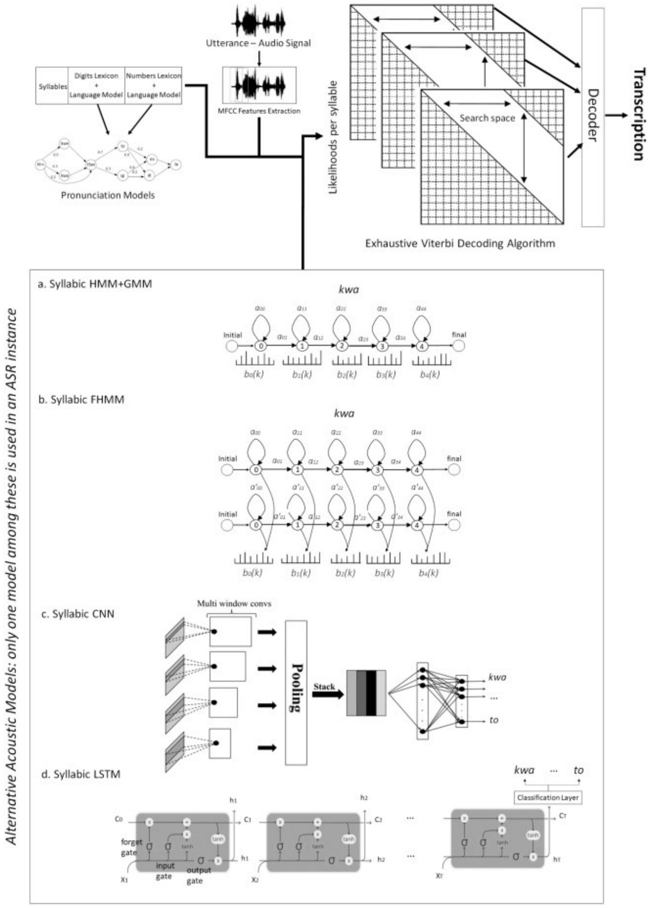


Figure 8. Diagram of the proposed ASR with acoustic models used alternatively (only one in an ASR instance): (*a*) syllabic HMMs using GMM emission probabilities, (*b*) syllabic Factorial HMMs, (*c*) Convolutional Neural Network using multi-temporal windows, with one output neuron for each syllable, and (*d*) Long Short Term Memory model, with one output for each syllable. [24]

Information and Research Science connecting to Digital and Library Science: Report on the 17th Italian Research Conference on Digital Libraries, IRCDL2021 [26] by Dosso et al. for ACM SIGMOD Record.

Summary: Since 2005 the Italian Research Conference on Digital Libraries is a yearly date for researchers on Digital Libraries and related topics, organized by the Italian Research Community. Over the years, IRCDL has become an essential national forum focused on digital libraries and associated technical, practical, and social issues. IRCDL encompasses the many meanings of the term digital libraries, including new forms of information institutions; operational information systems with all manner of digital content; new means of selecting, collecting, organizing, and distributing digital content; and theoretical models of information media, including document genres and electronic publishing.

An Open Science approach to infer fishing activity pres-

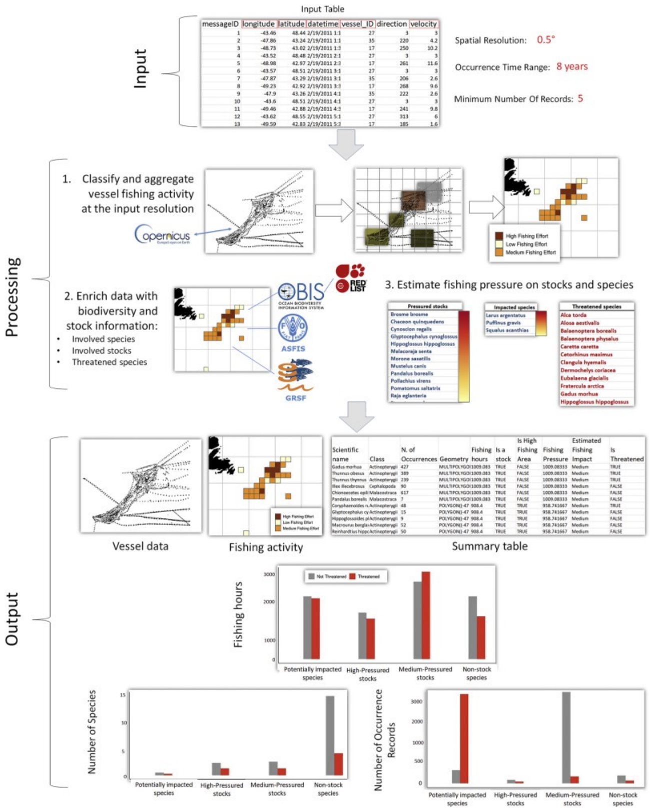
sure on stocks and biodiversity from vessel tracking data [23] by Coro et al. for Ecological Informatics.

Summary: Vessel tracking data help study the potential impact of fisheries on biodiversity and produce risk assessments. Existing workflows process vessel tracks to identify fishing activity and integrate information on species vulnerability. However, there are significant data integration challenges across the data sources needed for an integrated impact assessment due to heterogeneous nomenclatures, data accessibility issues, geographical and computational scalability of the processes, and confidentiality and transparency towards decision making authorities. This paper presents an Open Science data integration approach to use vessel tracking data in integrated impact assessments. Our approach combines heterogeneous knowledge sources from fisheries, biodiversity, and environmental observations to infer fishing activity and risks to potentially impacted species. An Open Science e-Infrastructure facilitates access to data sources and maximises the reproducibility of the results and the method's reusability across several application domains. Our method's quality is assessed through three case studies: The first demonstrates cross-dataset consistency by comparing the results obtained from two different vessel data sources. The second performs a temporal pattern analysis of fishing activity and potentially impacted species over time. The third assesses the potential impact of reduced fishing pressure on marine biodiversity and threatened species due to the 2020 COVID-19 lockdown in Italy. The method is meant to be integrated with other systems through its Open Science-oriented features and can rapidly use new sources of findable, accessible, interoperable, and reusable (FAIR) data. Other systems can use it to (*i*) classify vessel activity in data-limited scenarios, (*ii*) identify bycatch species (when catchability data are available), and (*iii*) study the effects of fisheries on habitats and populations' growth.

Fig. 9 depicts the schema of the methodological workflow proposed, with processing divided into three separate steps.

The role of technology and digital innovation in sustainability and decarbonization of the Blue Economy [15] by Campana et al. for Bulletin of Geophysics and Oceanography.

Summary: The development of a sustainable technology for the Blue Economy (a new Blue Technology) sets out three core research objectives, reflecting key challenges to be tackled by the sea industries and scientific and technological communities: The fast development of doable decarbonization processes through development and demonstration of deployable, competi ti ve, and sustainable technological solutions for energy transition (climate neutral blue economy), a sustainable exploitation and exploration of oceans, seas and coastal areas to provide new resources, from raw materials to products, including food (sustainable use and management of marine resources), and the development and exploitation of digital-based knowledge while accumulating data from new observation networks (persistent monitoring and digitalization of seas and oceans). To meet these operational objectives, different topics and related technologies need to be further developed.



messageID	longitude	latitude	datetime	vessel_ID	direction	velocity
1	-43.46	48.44	2/19/2011 1:1	27	3	3
2	-47.86	43.24	2/19/2011 1:1	35	220	4.2
3	-48.73	43.02	2/19/2011 1:3	17	250	10.2
4	-43.52	48.48	2/19/2011 2:1	27	3	3
5	-48.98	42.97	2/19/2011 2:3	17	261	11.6
6	-43.57	48.51	2/19/2011 3:1	27	3	3
7	-47.87	43.29	2/19/2011 3:1	35	206	2.6
8	-49.23	42.92	2/19/2011 3:3	17	268	9.6
9	-47.9	43.26	2/19/2011 4:1	35	222	2.6
10	-43.6	48.51	2/19/2011 4:1	27	3	3
11	-49.46	42.88	2/19/2011 4:3	17	241	9.8
12	-43.62	48.55	2/19/2011 5:1	27	313	6
13	-49.59	42.83	2/19/2011 5:3	17	185	1.6

Pressured stocks
Brosme brosme
Chaceon quinquedens
Cynoscion regalis
Glyptocephalus cynoglossus
Hippoglossus hippoglossus
Malacoraja senta
Morone saxatilis
Mustelus canis
Pandalus borealis
Pollachius virens
Pomatomus saltatrix
Raja eglanteria

Impacted species
Larus argentatus
Puffinus gravis
Squalus acanthias

Threatened species
Alca torda
Alosa aestivalis
Balaenoptera borealis
Balaenoptera physalus
Caretta caretta
Cetorhinus maximus
Clangula hyemalis
Dermochelys coriacea
Eubalaena glacialis
Fratercula arctica
Gadus morhua
Hippoglossus hippoglossus

Scientific name	Class	N. of Occurrences	Geometry	Fishing hours	Is a stock	Is High Fishing Area	Fishing Pressure	Estimated Fishing Impact	Is Threatened
Gadus morhua	Actinopterygii	427	MULTIPOLYGOI	1009.083	TRUE	FALSE	1009.08333	Medium	TRUE
Thunnus obesus	Actinopterygii	389	MULTIPOLYGOI	1009.083	TRUE	FALSE	1009.08333	Medium	TRUE
Thunnus thynnus	Actinopterygii	239	MULTIPOLYGOI	1009.083	TRUE	FALSE	1009.08333	Medium	TRUE
Illex illecebrosus	Cephalopoda	90	MULTIPOLYGOI	1009.083	TRUE	FALSE	1009.08333	Medium	FALSE
Chionoecetes opili	Malacostraca	617	MULTIPOLYGOI	1009.083	TRUE	FALSE	1009.08333	Medium	FALSE
Pandalus borealis	Malacostraca	7	MULTIPOLYGOI	1009.083	TRUE	FALSE	1009.08333	Medium	FALSE
Coryphaenoides ru	Actinopterygii	48	POLYGON((-47	908.4	TRUE	TRUE	958.741667	Medium	TRUE
Glyptocephalus cy	Actinopterygii	15	POLYGON((-47	908.4	TRUE	TRUE	958.741667	Medium	FALSE
Hippoglossoides pl	Actinopterygii	9	POLYGON((-47	908.4	TRUE	TRUE	958.741667	Medium	FALSE
Macrourus berglai	Actinopterygii	52	POLYGON((-47	908.4	TRUE	TRUE	958.741667	Medium	FALSE
Reinhardtius hippc	Actinopterygii	50	POLYGON((-47	908.4	TRUE	TRUE	958.741667	Medium	FALSE

Figure 9. Schema of the methodological workflow, with processing divided into three separate steps. [23]

Measuring success for a future vision: Defining impact in science gateways/virtual research environments [14] by Calyam et al. for Concurrency and Computation: Practice and Experience.

Summary: Scholars worldwide leverage science gateways / virtual research environments (VREs) for a wide variety of research and education endeavors spanning diverse scientific fields. Evaluating the value of a given science gateway/VRE to its constituent community is critical in obtaining the financial and human resources necessary to sustain operations and increase adoption in the user community. In this article, we feature a variety of exemplar science gateways/VREs and detail how they define impact in terms of, for example, their purpose, operation principles, and size of user base. Further, the exemplars recognize that their science gateways/VREs will continuously evolve with technological advancements and standards in cloud computing platforms, web service architectures, data management tools and cybersecurity. Correspondingly, we present a number of technology advances that could be incorporated in next-generation science gateways/VREs to enhance their scope and scale of their operations for greater success/impact. The exemplars are selected from owners of science gateways in the Science Gateways Community Institute (SGCI) clientele in the United States, and from the owners of VREs in the International Virtual Research Environment Interest Group (VRE-IG) of the Research Data Alliance. Thus, community-driven best practices and technology advances are compiled from diverse expert groups with an international perspective to envisage futuristic science gateway/VRE innovations.

Exploring the status of the Indonesian deep demersal fishery using length-based stock assessments [25] by Dimarchopoulou et al. for Fisheries Research.

Summary: The deep demersal snapper-grouper fishery in Indonesia is a data-poor fisheries resource that provides food security and a source of income to millions globally. Owing to an ongoing crew-operated data recording system implemented in Indonesia since 2015, the stocks of this fishery can now be assessed using length-frequency data and updated life-history parameters. Here, we use two length-based methods, one that is fishery-specific and another that is more generalized, to assess the status of Indonesian stocks. Specifically, we develop a literature-based assessment method based on a patchwork of conventional approaches but tailored to the studied stocks, and compare it with a newly established and broadly applicable length-based Bayesian biomass estimation method (LBB). The methods were applied to 16 stocks from 4 Indonesian Fisheries Management Areas and were compared based on simulations, as well as the convergence of the resulting stock status classification and uncertainty of the results. Analyzing the effect of using the literature-based species/family-specific life-history parameter values for asymptotic length (Linf) and relative natural mortality (M/K) in LBB showed that different values do affect the estimated biomass indicator. Nevertheless, in more than half the cases, the stock status classification did not differ between the two methods, while LBB results became more reliable with narrower confidence limits. Simulations, as well as similar status indicators between the two models support the value of the literature-based approach as an assessment methodology for the Indonesian deep demersal fisheries. Narrower confidence ranges highlight the importance of using fishery-specific information when applying generalized stock assessment methods. While most catches had few immature fish, half of the assessed stocks were consistently shown to have low biomass, indicating that important Indonesian stocks are at high risk of overfishing.

Fig. 10 depicts the catch length frequency distributions for the CODRS samples collected in 2020 and life-history parameters as estimated with the customized length-based approach for the analyzed Indonesian stocks.

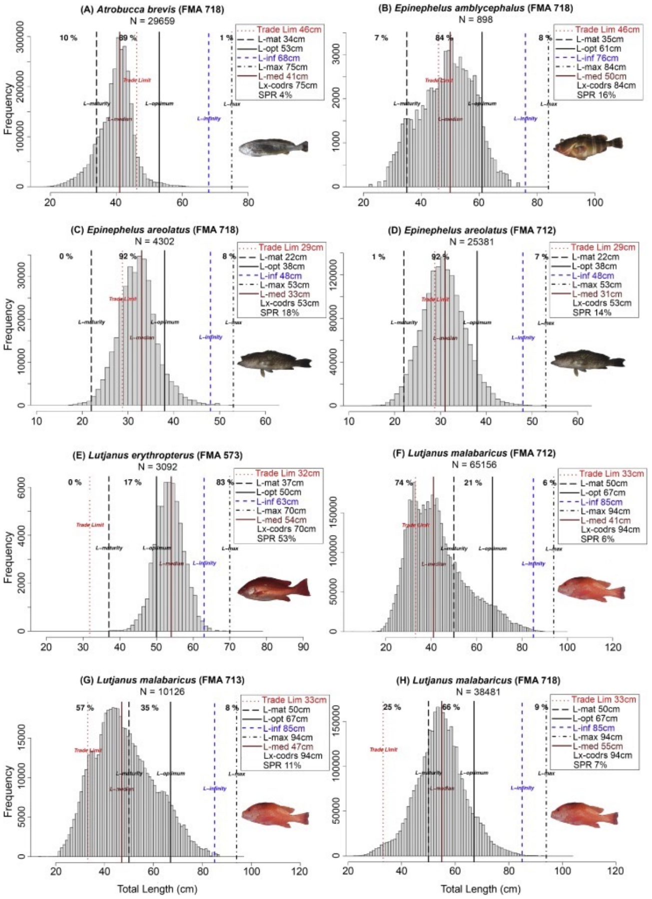


Figure 10. Catch length frequency distributions, life-history parameters and reference points as estimated with the highly customized length-based approach presented in the paper for selected Indonesian stocks. [25]

We Can Make a Better Use of ORCID: Five Observed Misapplications [8] by Baglioni et al. for Data Science Journal.

Summary: Since 2012, the "Open Researcher and Contributor ID" organisation (ORCID) has been successfully running a worldwide registry, with the aim of "providing a unique, persistent identifier for individuals to use as they engage in research, scholarship, and innovation activities". Any service in the scholarly communication ecosystem (e.g., publishers, repositories, CRIS systems, etc.) can contribute to a non-ambiguous scholarly record by including, during metadata deposition, referrals to iDs in the ORCID registry. The OpenAIRE Research Graph is a scholarly knowledge graph that aggregates both records from the ORCID registry and publication records with ORCID referrals from publishers and repositories worldwide to yield research impact monitoring and Open Science statistics. Graph data analytics revealed "anomalies" due to ORCID registry "misapplications", caused by wrong ORCID referrals and misexploitation of the ORCID registry. Albeit these affect just a minority of ORCID records, they inevitably affect the quality of the ORCID infrastructure and may fuel the rise of detractors and scepticism about the service. In this paper, we classify and qualitatively document such misapplications, identifying five ORCID registrant-related and ORCID referral-related anomalies to raise awareness among ORCID users. We describe the current countermeasures taken by ORCID and, where applicable, provide recommendations. Finally, we elaborate on the importance of a community-steered Open Science infrastructure and the benefits this approach has brought and may bring to ORCID.

Fig. 11 reports the number of records registered every year, as derived from the ORCID public data.

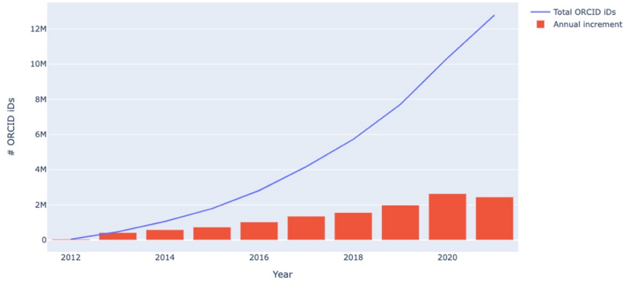


Figure 11. Number of ORCID records per year. The bar plot indicates the annual increment, while the line reports the total number of ORCID iDs through the years. (source: ORCID's public data file, October 2021). [8]

3.2 Contributions to Conferences

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

Reflections on the misuses of ORCID iDs [9] by Baglioni et al. for the 17th Italian Research Conference on Digital Libraries (IRCDL 2021).

Summary: Since 2012, the "Open Researcher and Contributor Identification Initiative" (ORCID) has been successfully running a worldwide registry, with the aim of unequivocally pinpoint researchers and the body of knowledge they contributed to. In practice, ORCID clients, e.g., publishers, repositories, and CRIS systems, make sure their metadata can refer to iDs in the ORCID registry to associate authors and their work unambiguously. However, the ORCID infrastructure still suffers from several "service misuses", which put at risk its very mission and should be therefore identified and tackled. In this paper, we classify and qualitatively document such misuses, occurring from both users (researchers and organisations) of the ORCID registry and the ORCID clients. We conclude providing an outlook and a few recommendations aiming at improving the exploitation of the ORCID infrastructure.

BIP! DB: A Dataset of Impact Measures for Scientific Publications [37] by Vergoulis et al. for WWW '21: Companion Proceedings of the Web Conference 2021.

Summary: The growth rate of the number of scientific publications is constantly increasing, creating important challenges in the identification of valuable research and in various scholarly data management applications, in general. In this context, measures which can effectively quantify the scientific impact could be invaluable. In this work, we present BIP! DB, an open dataset that contains a variety of impact measures calculated for a large collection of more than 100 million scientific publications from various disciplines.

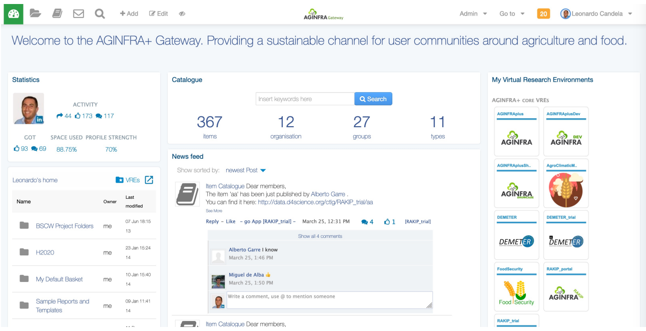
Contextual Conditional Reasoning [20] by Casini et al. for the 35th AAAI Conference on Artificial Intelligence, February 2–9, 2021.

Summary: We extend the expressivity of classical conditional reasoning by introducing context as a new parameter. The enriched conditional logic generalises the defeasible setting in the style of Kraus, Lehmann and Magidor, and allows for a more refined representation of an agent's epistemic state, distinguishing, for example, between expectations and counterfactuals. In this paper we introduce the language for the enriched logic, and define an appropriate semantic framework for it. We analyse which properties generally associated with conditional reasoning are still satisfied by the new semantic framework, provide an appropriate representation result, and define an entailment relation based on Lehmann and Magidor's notion of Rational Closure.

Realising a science gateway for the Agri-food: the AGINFRAplus experience [7] by Assante et al. for 11th International Workshop on Science Gateways.

Summary: The enhancements in IT solutions and the open science movement are injecting changes in the practices dealing with data collection, collation, processing and analytics, and publishing in all the domains, including agri-food. However, in implementing these changes one of the major issues faced by the agri-food researchers is the fragmentation of the "assets" to be exploited when performing research tasks, e.g., data of interest are heterogeneous and scattered across several repositories, the tools modellers rely on are diverse and often make use of limited computing capacity, the publishing practices are various and rarely aim at making available the "whole story" with datasets, processes, workflows. This paper presents the AGINFRA PLUS endeavour to overcome these limitations by providing researchers in three designated communities with Virtual Research Environments facilitating the use of the "assets" of interest and promote collaboration.

Fig. 12 depicts the AGINFRA PLUS *gateway* realising the single access point to the rest of the platform discussed in the paper.



Welcome to the AGINFRA+ Gateway. Providing a sustainable channel for user communities around agriculture and food.

Name	Owner	Last modified
BSCW Project Folders	me	07 Jun 18:15 13
H2020	me	23 Jan 15:24 14
My Default Basket	me	10 Jan 15:40 14
Sample Reports and Templates	me	09 Jan 11:41 14

Item Catalogue Dear members,
The item 'aa' has been just published by Alberto Garre .
You can find it here: http://data.d4science.org/ctlg/RAKIP_trial/aa
See More
Reply - Like - go App [RAKIP_trial] - March 25, 12:31 PM 4 1 [RAKIP_trial]
Show all 4 comments
Alberto Garre I know
March 25, 1:46 PM
Miguel de Alba 👍
March 25, 1:50 PM
Write a comment, use @ to mention someone
Item Catalogue Dear members,

Figure 12. AGINFRA PLUS Gateway: the Dashboard. [7]

KLM-style defeasibility for restricted first-order logic [19] by Casini et al. for NMR 2021 - 19th International Workshop on Non-Monotonic Reasoning.

Summary: We extend the KLM approach to defeasible reasoning to be applicable to a restricted version of first-order logic. We describe defeasibility for this logic using a set of rationality postulates, provide an appropriate semantics for it, and present a representation result that characterises the semantic description of defeasibility in terms of the rationality postulates. Based on this theoretical core, we then propose a version of defeasible entailment that is inspired by Rational Closure as it is defined for defeasible propositional logic and defeasible description logics. We show that this form of defeasible entailment is rational in the sense that it adheres to our rationality postulates. The work in this paper is the first step towards our ultimate goal of introducing KLM-style defeasible reasoning into the family of Datalog+/- ontology languages.

3.3 Contributions to Books

InfraScience members contributed to the following chapters in books.

Detection, Analysis, and Prediction of Research Topics with Scientific Knowledge Graphs [36] by Salatino et al. for Predicting the Dynamics of Research Impact.

Summary: Analysing research trends and predicting their impact on academia and industry is crucial to gain a deeper understanding of the advances in a research field and to inform critical decisions about research funding and technology adoption. In the last years, we saw the emergence of several publicly-available and large-scale Scientific Knowledge Graphs fostering the development of many data-driven approaches for performing quantitative analyses of research trends. This chapter presents an innovative framework for detecting, analysing, and forecasting research topics based on a large-scale knowledge graph characterising research articles according to the research topics from the Computer Science Ontology. We discuss the advantages of a solution based on a formal representation of topics and describe how it was applied to produce bibliometric studies and innovative tools for analysing and predicting research dynamics.

3.4 Technical Reports

InfraScience members contributed to the following Technical Reports.

Progettare un evento divulgativo online. L'esperienza di ”Fai una domanda su Covid-19, gli esperti rispondono” [35] by Pavone et al.

Summary: Il documento raccoglie le domande ricevute e le risposte date in occasione del webinar “Fai una domanda su Covid-19, gli esperti rispondono”, realizzato il 27 novembre 2020. Inoltre il report descrive gli aspetti principali dell'organizzazione di un webinar “Ask Me Anything”, che voleva essere un evento informativo, online e rivolto ai giovani sul tema pandemia di Covid-19. A distanza di diversi mesi dall'inizio della pandemia, erano ancora tanti i dubbi e le incertezze diffuse tra le persone, anche riguardo ad aspetti noti e chiariti in sede di ricerca scientifica. In occasione dell'edizione 2020 della Notte Europea dei Ricercatori e delle Ricercatrici, è stato fatto uno sforzo divulgativo di apertura e confronto con alcuni esperti impegnati nella ricerca su Sars-Cov-2 e Covid- 19.

Situated conditional reasoning [21] by Casini et al.

Summary: Conditionals are useful for modelling, but aren't always sufficiently expressive for capturing information accurately. In this paper we make the case for a form of conditional that is situation-based. These conditionals are more expressive than classical conditionals, are general enough to be used in several application domains, and 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 situation-based 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 proceed to define a form of entailment for situated conditional knowledge bases, which we refer to as minimal closure. It is reminiscent of and, indeed, inspired by, the version of entailment for propositional conditional knowledge bases known as rational 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.

A rational entailment for expressive description logics via description logic programs [18] by Casini and Straccia.

Summary: Lehmann and Magidor's rational closure is acknowledged as a landmark in the field of non-monotonic logics and it has also been re-formulated in the context of Description Logics (DLs). We show here how to model a rational form of entailment for expressive DLs, such as SROIQ, providing a novel reasoning procedure that compiles a non-monotone DL knowledge base into a description logic program (dl-program).

Competence Centre ICDI per Open Science, FAIR, ed EOSC - Mission, strategia e piano d'azione [29] by Lazzeri et al.

Summary: This document (in Italian) presents the mission and strategy of the Italian Competence Centre on Open Science, FAIR, and EOSC. The Competence Centre is an initiative born within the Italian Computing and Data Infrastructure (ICDI), a forum created by representatives of major Italian Research Infrastructures and e-Infrastructures, with the aim of promoting sinergies at the national level, and optimising the Italian participation to European and global challenges in this field, including the European Open Science

Cloud (EOSC), the European Data Infrastructure (EDI) and HPC. This working paper depicts the mission and objectives of the ICDI Competence Centre, a network of experts with various skills and competences that are supporting the national stakeholders on topics related to Open Science, FAIR principles application and participation to the EOSC. The different actors and roles are described in the document as well as the activities and services offered, and the added value each stakeholder can find the in Competence Centre. The tools and services provided, in particular the concept for the portal, though which the Centre will connect to the national landscape and users, are also presented. An english translation of this document was produced [30].

ICDI Competence Centre for Open Science, FAIR and EOSC - Mission, strategy and action plan [30] by Lazzeri et al.

Summary: This document presents the mission and strategy of the Italian Competence Centre on Open Science, FAIR, and EOSC. The Competence Centre is an initiative born within the Italian Computing and Data Infrastructure (ICDI), a forum created by representatives of major Italian Research Infrastructures and e-Infrastructures, with the aim of promoting synergies at the national level, and optimising the Italian participation to European and global challenges in this field, including the European Open Science Cloud (EOSC), the European Data Infrastructure (EDI) and HPC. This working paper depicts the mission and objectives of the ICDI Competence Centre, a network of experts with various skills and competencies that are supporting the national stakeholders on topics related to Open Science, FAIR principles application and participation to the EOSC. The different actors and roles are described in the document as well as the activities and services offered, and the added value each stakeholder can find the in Competence Centre. The tools and services provided, in particular the concept for the portal, through which the Centre will connect to the national landscape and users, are also presented. This record is the English translation of the original Italian document [29].

4. Projects

InfraScience was an active member of the consortiums proposing and implementing 18 research projects (15 were European Union's supported projects) all focusing on the development of data infrastructures and solutions for various communities of practice.

ARIADNEplus[3] is a European Union's Horizon 2020 project (grant agreement No. 823914) started in January 2019 and ending in December 2022. It extends the previous ARIADNE Integrating Activity, which successfully integrated archaeological data infrastructures in Europe, indexing in its registry about 2.000.000 datasets. It extends and supports the research community that the previous project created and further develops the relationships with key stakeholders such as the most important European archaeological associations, researchers, heritage professionals, national heritage agencies and so on. The ARIADNEplus data infrastructure is conceived to offer the availability of Virtual Research Environments where data-based archaeological research may be carried out. The project will furthermore develop a Linked Data approach to data discovery. Innovative services will be made available to users, such as visualization, annotation, text mining and geo-temporal data management. Innovative pilots will be developed to test and demonstrate the innovation potential of the ARIADNEplus approach. Fostering innovation will be a key aspect of the project, with dedicated activities led by the project Innovation Manager. The *InfraScience* team is leading two work packages: "Data Integration and Interoperability" to develop, deliver and maintain the ARIADNEplus data and knowledge Cloud and the ARIADNEplus Data Infrastructure; "ARIADNEplus Infrastructure Operation and Management" to (*i*) manage the set of technologies required to operate the ARIADNEplus e-infrastructure, by exploiting the set of services and computational resources provided by the D4Science infrastructure and by supporting the integration of tools, facilities, and services provided by the present project; (*ii*) provide access to the stack of such facilities via Virtual Research Environments, by exploiting the procedures and policies tested and already used by D4-Science; (*iii*) manage the software release process covering all stages from integration, through documentation and validation, up to provisioning.

Blue Cloud[4] 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.

DESIRA[5] is a European Union's Horizon 2020 project

[3] ARIADNEplus website `ariadne-infrastructure.eu`

[4] Blue Cloud website `blue-cloud.org`

[5] DESIRA website `desira2020.eu`

(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 has the possibility to 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 the cloud storage to exchange large amount of data securely; (*iv*) an activity tracker and collaborative wiki.

EcoScope[6] 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 sustainability and ensuring 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 Future[7] 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, connect e-infrastructures, research communities, and initiatives in Open Science to further develop 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.

EOSC-Pillar[8] is a European Union's Horizon 2020 project (grant agreement No. 857650) started in July 2019 and ending in December 2022. It was funded to establish an agile and efficient federation model for open science services covering the full spectrum of European research communities by building on representatives of the fast-growing national initiatives for coordinating data infrastructures and services in Italy, France, Germany, Austria and Belgium. The project aims to contribute to the development of EOSC within a science-driven approach which is efficient, scalable and sustain-

[6] EcoScope website `https://ecoscopium.eu/`

[7] EOSC Future website `https://eoscfuture.eu/`

[8] EOSC-Pillar website `www.eosc-pillar.eu`

able and that can be rolled out in other countries. The *InfraScience* team is coordinating the contribution of the Italian National Research Council research unit comprising four Institutes: Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (ISTI), Istituto Officina dei Materiali (IOM), Istituto di Biomembrane, Bioenergetica e Biotecnologie Molecolari (IBIOM), and Istituto di Tecnologie Biomediche (ITB). Moreover, InfraScience is leading the research and development tasks leading to the development of a model and a prototype of a nation service catalog interoperable with EOSC, the development of a catalog driven solution to discover and access the items of a FAIR data space across scattered and heterogeneous data providers, the provisioning of Virtual Research Environments supporting the implementation of case studies in diverse domains.

EOSCsecretariat.eu[9] is a European Union's Horizon 2020 project (grant agreement No. 831644) started in January 2019 and ended in June 2021. It was funded to provide support for the European Open Science Cloud (EOSC), addressing all the specific needs of this digital platform. The project works alongside the community to deliver many of its activities, and it has reserved a substantial portion of its budget for organisations not in the consortium. The EOSCsecretariat.eu's responsibilities include the organisation of EOSC-related events, press and media services, coordination with other related projects, liaison with non-EU countries, efforts to increase pan-European awareness and the provision of a sound legal framework. The project is carried out with the support of highly experienced partners from academia and industry. The *InfraScience* team is responsible of the stakeholders engagement. In this role it coordinated activities across projects involved in the implementation of EOSC and established an exchange dialogue with EOSC potential users, like researchers and industry, to provide requirements and feedback during the EOSC development process. It also supports the project by developing and operating a dedicated gateway making available a set of virtual research environments facilitating the collaboration and communication among the members of the task forces the project set up.

I-GENE[10] 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). The *InfraScience* team 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.

MOVING[11] 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. The *InfraScience* team supports the development and operation of the virtual research environment.

NAVIGATOR[12] 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. The *InfraScience* team 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 Nexus[13] 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 syn-

[9] EOSCsecretariat.eu website `www.eoscsecretariat.eu`

[10] I-GENE website `i-geneproject.eu`

[11] MOVING website `www.moving-h2020.eu`

[12] NAVIGATOR website `http://navigator.med.unipi.it/`

[13] OpenAIRE Nexus website `https://www.openaire.eu/openaire-nexus-project`

ergies 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.

OpenAIRE-Advance[14] is a European Union's Horizon 2020 project (grant agreement No. 777541) started in January 2018 and ended in February 2021. It was funded to continue the mission of OpenAIRE to support the Open Access and Open Data mandates in Europe. By sustaining the current successful infrastructure, comprised of a human network and robust technical services, it consolidates its achievements while working to shift the momentum among its communities to Open Science, aiming to be a trusted e-Infrastructure within the realms of the European Open Science Cloud. In this next phase, OpenAIRE-Advance strives to empower its National Open Access Desks (NOADs) so they become a pivotal part within their own national data infrastructures, positioning Open Access and Open Science onto national agendas. On the technical level OpenAIRE-Advance focuses on the operation and maintenance of the OpenAIRE services, and radically improves the OpenAIRE services on offer by: (*a*) optimizing their performance and scalability, (*b*) refining their functionality based on end-user feedback, (*c*) repackaging them into products, taking a professional marketing approach with well-defined KPIs, (*d*) consolidating the range of services/products into a common e-Infra catalogue to enable a wider uptake. *InfraScience* was responsible for (*i*) the technical coordination; (*ii*) the operation of the Italian National Open Access Desk (NOAD)[15], which participated to the Research Data Management Task Force; (*iii*) the product management of OpenAIRE-CONNECT[16], (*iv*) work package "Participatory Scholarly Communication", with pilots and technical collaborations with EOSC-Hub, EGI and research infrastructures (EPOS-IT, DARIAH-IT, ELIXIR-GR), (*v*) work package "Optimization & Upgrade of OpenAIRE Technical Services", devoted to technical improvements, scalability optimisation and development of new features of OpenAIRE products (e.g., extension of the Broker service for aggregators in a pilot with LaReferencia, enabling the realization of country portals in a pilot with the Canadian repository network). InfraScience was also responsible for the operation of workflows for the generation of the OpenAIRE Research Graph[17] and the maintenance of the infrastructure for metadata aggregation and full-text collection. InfraScience contributed to the curation of the OpenAIRE Research Graph, integrated project metadata from funders' databases, and developed services to detect duplicates (de-duplication framework), delete wrong links (blacklisting), curate organisation entities (OpenOrgs)[18].

PerformFISH[19] is a European Union's Horizon 2020 project (grant agreement No. 727610) started in May 2017 and ending in April 2022. It was funded to increase the competitiveness of Mediterranean aquaculture by overcoming biological, technical and operational issues with innovative, cost-effective, integrated solutions, while addressing social and environmental responsibility and contributing to "Blue Growth". It adopts a holistic approach constructed with active industry involvement to ensure that Mediterranean marine fish farming matures into a modern dynamic sector, highly appreciated by consumers and society for providing safe and healthy food with a low ecological footprint, and employment and trade in rural, peripheral regions. The project brings together a representative multi-stakeholder, multi-disciplinary consortium to generate, validate and apply new knowledge in real farming conditions to substantially improve the management and performance of the focal fish species, measured through Key Performance Indicators. At the core of PerformFISH design are, (*a*) a link between consumer demand and product design, complemented with product certification and marketing strategies to drive consumer confidence, and (*b*) the establishment and use of a numerical benchmarking system to cover all aspects of Mediterranean marine fish farming performance. *InfraScience* is leading the activity "Building a Virtual Research Environment (VRE) to Host and Manage Project Data" to deliver (*i*) a set of VREs offering workspace capabilities for supporting the collection, management and controlled sharing of datasets produced by experiments carried out in WPs 1,2,3,4,6. Data sharing will be enabled either between the members of a VRE or between selected users (e.g. colleagues and companies); (*ii*) a VRE supporting KPI data analysis and benchmarking based on production data collected by private companies and securely managed using advanced cryptography and pseudo-anonymisation techniques; (*iii*) a VRE providing access to aggregated and anonymised data to authorised members only.

RISIS 2[20] is a European Union's Horizon 2020 project (grant agreement No. 824091) started in January 2019 and ending in December 2022. It was funded to develop an e-infrastructure that supports full virtual transnational access by researchers in the field of science, technology and innovation to (*a*) an enlarged set of services aimed at meeting field-specific needs (for exploring open data and supporting researchers' analytical capabilities) and (*b*) a set of datasets. *InfraScience* contributes to the development of the RISIS 2 infrastructure with its infrastructures supporting Open Sci-

[14] OpenAIRE-Advance website `www.openaire.eu/advance`

[15] OpenAIRE NOADs website `openaire.eu/noad-activities`

[16] OpenAIRE-CONNECT website `connect.openaire.eu`

[17] OpenAIRE Research Graph website `graph.openaire.eu`

[18] OpenOrgs website `https://orgs.openaire.eu/`

[19] PerformFISH website `performfish.eu/`

[20] RISIS 2 website `www.risis2.eu`

ence, namely D4Science and OpenAIRE. Specifically, the Open Data Virtual Research Environment, empowered by the D4Science, has been equipped with the capability of bridging the RISIS Core Facility Framework and OpenAIRE. This VRE allows delivering tailored and specific datasets collected by OpenAIRE, and selected to satisfy the needs of the RISIS community, to the RISIS project members and community. The Open Data Virtual Research Environment is part of a wider setting involving and all three infrastructures, namely OpenAIRE, D4Science, and RISIS. Its goal is to enrich the RISIS e-Infrastructure in terms of datasets and tools available for the RISIS Community.

SerGenCOVID-19[21] 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 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.

Snapshot[22] is an Italian project funded by CNR, started in September 2020 and ending in December 2022. Its goal was to provide a quantitative assessment of the effects of the reduced anthropogenic pressure on marine systems during the lockdowns that responded to the COVID-19 pandemic. The 2020 restrictions generated unprecedented, and partially unexpected, human and marine ecosystem dynamics at various levels besides those related to fisheries. By analysing these dynamics in the Italian marine ecosystems, specific cause-effect relationships can be identified and extended to other world ecosystems. The aim of the project is to measure these relationships and the multiple factors involved – including pollution, the economy, fisheries and ecosystem services – to design novel strategies for a more sustainable future. *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.

SoBigData-PlusPlus[23] 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 strengthen 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 the 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).

TAILOR[24] 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.

[21] SerGenCovid19 website `https://sergencovid.iit.cnr.it/`

[22] Snapshot website `http://snapshot.cnr.it/`

[23] SoBigData++ website `sobigdata.eu`

[24] TAILOR website `tailor-network.eu`

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[25] [6] 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 2021 its user base reached 17,492 active users (+3,891 users wrt December 2020) (see Fig. 13). These users executed a total of 98,038 working sessions (circa 8,169 working sessions per month) (see Fig. 14) and a total of 373,135,457 analytics tasks (circa 31 millions tasks per month).

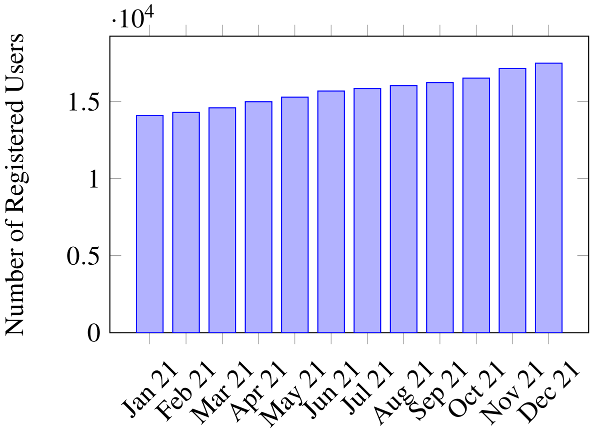


Figure 13. D4Science registered users in 2021.

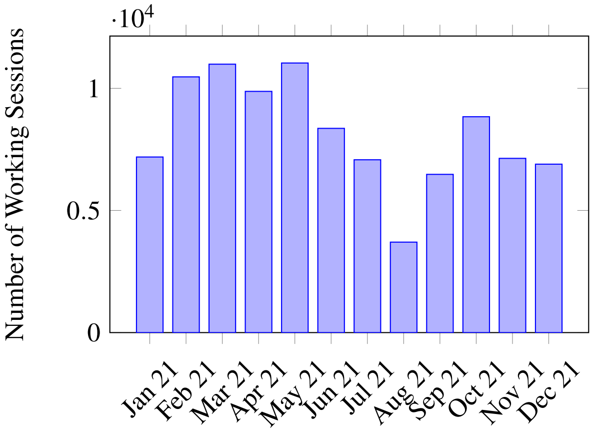


Figure 14. D4Science working sessions in 2021.

OpenAIRE[26] is a legal entity composed of 49 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. It 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 local and national level, supporting researchers, project coordinators, funders and policy makers with training and support activities. Furthermore, the technical infrastructure materializes the OpenAIRE Research Graph: an open, de-duplicated, participatory metadata research 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 materialized by collecting more than 210 millions of metadata records from more than 9,000 scholarly data sources worldwide. In addition to the information collected from trusted

[25] D4Science website `www.d4science.org`

[26] OpenAIRE website `www.openaire.eu`

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. In 2021, more than 1K repositories implemented the OpenAIRE guidelines for metadata exchange and registered to use the PROVIDE dashboard, 19 research communities used the CONNECT service to offer a thematic discovery portal to their researchers, and 10 research initiatives used the CONNECT & MONITOR services to track their impact; an average of 40,000 monthly users visited the OpenAIRE EXPLORE portal, offering search & discovery functionalities over the Research Graph to 18,300 registered users. At the end of 2021 the Research Graph contained bibliographic records for more than 128Mi publications, 15Mi datasets, 9Mi other research products and 260K software. As part of the activities of the National Open Access Desk of OpenAIRE, 13 training events were organized in 2021, targeting different audiences. These included a course for EC-financed project coordinators, different seminars for research support staff, two events on the topic of tackling the Covid-19 pandemic, and a couple of events aimed at early career researchers and university students on scholarly communication. In addition, several training sessions were organized on OpenOrgs, an OpenAIRE service for disambiguating organizations.

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. Moreover, in 2021 it completed the design and started the implementation of a technological infrastructure supporting virtualization platform, distributed storage, and integration of services. The group begins to install OpenStack, the free and open source IaaS cloud platform that handles cloud compute, storage and network resources exploiting TripleO (OpenStack On OpenStack) project on a machine park composed of fifteen servers for a total amount of 250TB of disk, 5.3TB of memory and 840 VCPU.

ISTI Open Portal[27] is a gateway to the scientific production of the Institute of Information Science and Technologies. 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, Scholexplorer [13], and Altmetric[28], and (*c*) make available the open access (self archived) version(s) of ISTI products. In 2021, the gateway had 8,340 users, 12,617 sessions and 31,210 page views. Fig. 15 displays the 2021 users distribution across countries. Fig. 16 reports the downloads by month.

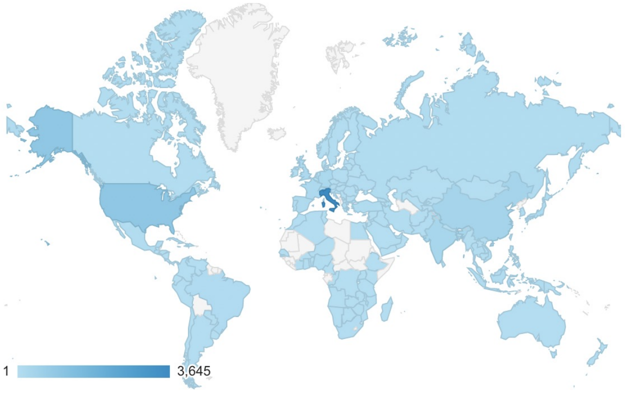


Figure 15. ISTI Open Portal 2021 Users geographic distribution

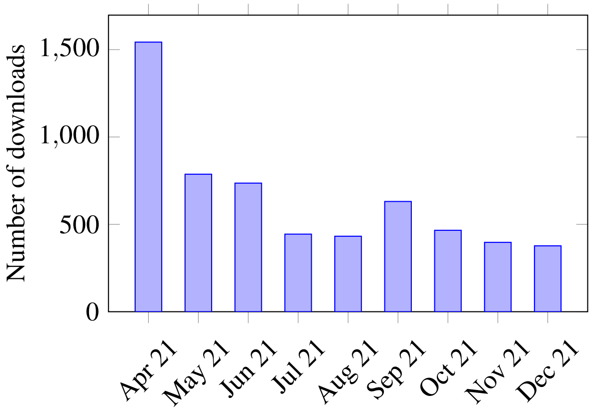


Figure 16. ISTI Open Portal 2021 monthly downloads.

InfraScience was responsible for ***open-science.it***[29], the Italian portal dedicated to Open Science and Open Access. 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) community[30] 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.

[27] ISTI Open Portal `https://openportal.isti.cnr.it`

[28] Altmetric website `https://www.altmetric.com`

[29] Open-science.it website `open-science.it`

[30] ICDI website `https://www.icdi.it`

6. Software

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

gCube[31] [5] 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. According to OpenHub[32] (statistics collected in October 2022) this software (*i*) has had 22,328 commits made by 49 contributors representing 1,160,934 lines of code (*ii*) is mostly written in Java with a low number of source code comments (*iii*) has a well established, mature codebase maintained by a large development team with stable Y-O-Y commits (*iv*) took an estimated 322 years of effort (COCOMO model) starting with its first commit in October, 2008 ending with its most recent commit. During 2021, 13 releases of this technology have been released (from gCube 4.28 in February 2021 up to gCube 5.6.1 in December 2021). All these releases have been exploited to enhance the service offered by the D4Science Infrastructure.

D-Net[33] [31] is a framework toolkit designed to support developers at constructing custom aggregative infrastructures in a cost-effective way. D-Net offers data management services capable of providing access to different kinds of external data sources, storing and processing information objects of any data models, converting them into common formats, and exposing information objects to third-party applications through a number of standard access APIs. Its infrastructure enabling services facilitate the construction of domain-specific aggregative infrastructures by selecting and configuring the needed services and easily combining them to form autonomic data processing workflows. The combination of out-of-the box data management services and tools for assembling them into workflows makes the toolkit an appealing starting platform for developers having to face the realization of aggregative infrastructures. In 2021, D-Net featured 8 installations running aggregation systems for (*a*) National aggregators: CeON in Poland, Recolecta in Spain; (*b*) research networks, associations, and infrastructures: EAGLE (Europeana network of Ancient Greek and Latin Epigraphy), EFG (European Film Gateway), OpenAIRE (Open Access Infrastructure for Research in Europe); (*c*) EC projects: PARTHENOS (Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies - EC H2020 project GA 654119), ARIADNEplus (Advanced Research Infrastructure for Archaeological Data Networking in Europe - plus - EC H2020 project GA 823914); (*d*) institutions: ISTI Open Portal.

7. Datasets

InfraScience released the following datasets.

OpenAIRE Research Graph: Dumps for research communities and initiatives. This dataset contains dumps of the OpenAIRE Research Graph containing metadata records relevant for the research communities and initiatives collaborating with OpenAIRE. In particular, three versions were released in January [32], April [33], and July [34].

OpenAIRE Covid-19 publications, datasets, software and projects metadata. This dump 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 dump contains records of the OpenAIRE COVID-19 Gateway[34], identified via full-text mining and inference techniques applied to the OpenAIRE Research Graph. The 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. In particular, three versions were released in January [10], May [11], and June [12].

8. Organised Events

P. Manghi was Program Chair for the *17th Italian Research Conference on Digital Libraries (IRCDL 2021)*, Padova, Italy - 18-19 February 2021. A. Bardi, L. Candela, and A. Mannocci were members of the Program Committee.

P. Manghi and A. Mannocci were in the Organising Committee of the *1st International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment (Sci-K 2021)*, co-located with The Web Conf 2021, April 13, 2021. A. Bardi was member of the Program Committee.

A. Bardi, M. Baglioni, and P. Manghi organised the "*OpenAIRE-CONNECT: co-design workshop with research communities*", in the context of the 3rd Open Science Fair Conference, 20 - 23 September 2021.

9. Training Activities

InfraScience members organised several training activities and courses mainly related to Open Science:

- Open Science e finanziamenti europei – Come ottemperare agli obblighi nei progetti H2020 e in Horizon Europe, February 2021, 4 webinars;
- OpenAIRE Connect Service: empowering Research Communities, 7th July 2021, Webinar;

[31] gCube website www.gcube-system.org

[32] gCube on Open Hub https://www.openhub.net/p/gCube

[33] D-Net website d-net.research-infrastructures.eu

[34] OpenAIRE COVID-19 Gateway website https://covid-19.openaire.eu/

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) – 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.

- *EOSC Glossary Interest Group* (D. Castelli) – an IG called to collaboratively provide contribution and feedback towards the creation and development of the EOSC Glossary;

- *Skills and Training Task Force* (E. Lazzeri) – a TF set up within the collaboration agreement among 7 H2020 projects funded by the INFRAEOSC-05 call to share and coordinate the common efforts in the topics of training and skills.

- *EOSC Service and Research Product Catalogues Interest Group* (A. Mannocci) – an IG called to provide a set of "enabling catalogues" that will facilitate the use and re-use of services in support of data-driven research (e.g., computing, storage, scholarly communication, thematic, etc.), data available in a multitude of sources (e.g., repositories, data archives, software archives, libraries, publishers, etc), and scientific products (e.g., publications, research data, research software, other products).

- *GOFAIR Discovery Implementation Network* (A. Bardi) – a GO FAIR consortium called to provide interfaces and other user-facing services for data discovery across disciplines;

- *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;

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

- *G6 Network Expert group on Open Science* (D. Castelli) – G6 is a network of collaborating research performing organisations (CNR, CNRS, Helmholtz, Leibnitz, MaxPlanck, CSIC). Open Science has been recognised as one of the topics of interest for all the G6 organizations. An appropriate expert group has been created to address this need composed of one representative per organization.

- *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 CCDS, and advise on partnership prospects with third parties.

- *Gruppo di Lavoro Piano Nazionale per la Scienza Aperta* (D. Castelli) – a WG called to develop the italian national plan for Open Science.

- *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 Commission Open Science Monitor Advisory Board* (E. Lazzeri) – a committee called to support the development of the EU Open Science Monitoring platform by giving advices and suggestions;

- *EOSC Architecture* (P. Manghi) – a WG called to define the technical framework required to enable and sustain an evolving EOSC federation of systems;

- *EOSC Rules of Participation* (P. Pagano) – a WG called to design the Rules of Participation that shall define the rights, obligations governing EOSC transactions between EOSC users, providers and operators;

- *EOSC Skills & Training* (E. Lazzeri) – a WG called to provide a framework for a sustainable training infrastructure to support EOSC in all its phases and ensure its uptake;

- *EOSC Task Force on Scholarly Infrastructures of Research Software* (L. Candela, P. Manghi) – a TF called to established a set of recommendations to allow EOSC to include software, next to other research outputs like publications and data, in the realm of its research artifacts;
- *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 activity performed by the InfraScience research group of the National Research Council of Italy - Institute of Information Science and Technologies (CNR - ISTI) in 2021.

During 2021 InfraScience members contributed to the publishing of 25 papers, to the research and development activities of 18 research projects (15 funded by EU), to the organization of conferences and training events, to several working groups and task forces.

Moreover, the group led the development of two large scale infrastructures for Open Science, i.e., D4Science and OpenAIRE.

Acknowledgments

InfraScience received funding from the European Union's Horizon 2020 research and innovation programme under: ARIADNEplus project (grant agreement No. 823914), 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), EOSC-Pillar project (grant agreement No. 857650), EOSCsecretariat.eu project (grant agreement No. 831644), I-GENE project (grant agreement No. 862714), MOVING project (grant agreement No. 862739), OpenAIRE Nexus project (grant agreement No. 101017452), OpenAIRE-Advance project (grant agreement No. 777541), PerformFISH project (grant agreement No. 727610), RISIS 2 project (grant agreement No. 824091), SoBigData-PlusPlus (grant agreement No. 871042), TAILOR (grant agreement No. 952215).

References

[1] G. Amato, P. Bolettieri, F. Carrara, F. Debole, F. Falchi, C. Gennaro, L. Vadicamo, and C. Vairo. The VISIONE Video Search System: Exploiting Off-the-Shelf Text Search Engines for Large-Scale Video Retrieval. *Journal of Imaging*, 7(5):76, Apr. 2021. ISSN 2313-433X. doi: 10.3390/jimaging7050076. URL `https://www.mdpi.com/2313-433X/7/5/76`.

[2] E. N. Armelloni, M. Scanu, F. Masnadi, G. Coro, S. Angelini, and G. Scarcella. Data poor approach for the assessment of the main target species of rapido trawl fishery in adriatic sea. *Frontiers in Marine Science*, 8, 2021. ISSN 2296-7745. doi: 10.3389/fmars.2021.552076. URL `https://www.frontiersin.org/article/10.3389/fmars.2021.552076`.

[3] M. Artini, L. Candela, P. Manghi, and S. Giannini. Reposgate: 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. Artini, M. Assante, C. Atzori, M. Baglioni, A. Bardi, L. Candela, G. Casini, D. Castelli, R. Cirillo, G. Coro, F. Debole, A. Dell'Amico, L. Frosini, S. La Bruzzo, E. Lazzeri, L. Lelii, P. Manghi, F. Mangiacrapa, A. Mannocci, P. Pagano, G. Panichi, T. Piccioli, F. Sinibaldi, and U. Straccia. InfraScience Research Activity Report 2020. Annual Reports 002, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", 2021. URL `https://doi.org/10.32079/ISTI-AR-2021/002`.

[5] M. Assante, L. Candela, D. Castelli, R. Cirilllo, 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. URL `http://www.sciencedirect.com/science/article/pii/S0167739X17328364`.

[6] M. Assante, L. Candela, D. Castelli, R. Cirillo, G. Coro, L. Frosini, L. Lelii, F. Mangiacrapa, P. Pagano, G. Panichi, and F. Sinibaldi. Enacting open science by d4science. *Future Generation Computer Systems*, 101:555–563, 2019. ISSN 0167-739X. doi: https://doi.org/10.1016/j.future.2019.05.063. URL `https://www.sciencedirect.com/science/article/pii/S0167739X1831464X`.

[7] M. Assante, A. Boizet, L. Candela, D. Castelli, R. Cirillo, G. Coro, E. Fernandez, M. Filter, L. Frosini, G. Kakaletris, P. Katsivelis, R. Knapen, L. Lelii, R. M. Lokers, F. Mangiacrapa, P. Pagano, G. Panichi, L. Penev, F. Sinibaldi, and P. Zervas. Realising a science gateway for the agri-food: the aginfraplus experience. In V. Stankovski and S. Gesing, editors, *11th International Workshop on Science Gateways, Ljubljana, Slovenia, 12-14/06/2019*. CEUR Workshop Proceedings, 2021.

[8] M. Baglioni, P. Manghi, A. Mannocci, and A. Bardi. We Can Make a Better Use of ORCID: Five Observed Misapplications. *Data Science Journal*, 20:38, Dec. 2021. ISSN 1683-1470. doi: 10.5334/dsj-2021-038.

URL http://datascience.codata.org/articles/10.5334/dsj-2021-038/.

[9] M. Baglioni, A. Mannocci, P. Manghi, C. Atzori, A. Bardi, and S. La Bruzzo. Reflections on the misuses of ORCID iDs. In D. Dosso, S. Ferilli, P. Manghi, A. Poggi, G. Serra, and G. Silvello, editors, *Proceedings of the 17th Italian Research Conference on Digital Libraries, Padua, Italy (virtual event due to the Covid-19 pandemic), February 18-19, 2021*, volume 2816, pages 117–125, 2021.

[10] A. Bardi, I. Kuchma, G. Pavone, M. Artini, C. Atzori, A. Bäcker, M. Baglioni, A. Czerniak, M. De Bonis, H. Dimitropoulos, I. Foufoulas, M. Horst, K. Iatropoulou, P. Jacewicz, A. Kokogiannaki, S. La Bruzzo, E. Lazzeri, A. Löhden, P. Manghi, A. Mannocci, N. Manola, E. Ottonello, and J. Schirrwagen. OpenAIRE Covid-19 publications, datasets, software and projects metadata., Jan. 2021. URL https://zenodo.org/record/4439663. Version Number: 1.0 Type: dataset.

[11] A. Bardi, I. Kuchma, G. Pavone, M. Artini, C. Atzori, A. Bäcker, M. Baglioni, A. Czerniak, M. De Bonis, H. Dimitropoulos, I. Foufoulas, M. Horst, K. Iatropoulou, P. Jacewicz, A. Kokogiannaki, S. La Bruzzo, E. Lazzeri, A. Löhden, P. Manghi, A. Mannocci, N. Manola, E. Ottonello, and J. Schirrwagen. OpenAIRE Covid-19 publications, datasets, software and projects metadata., May 2021. URL https://zenodo.org/record/4736827. Version Number: 2.0 Type: dataset.

[12] A. Bardi, I. Kuchma, G. Pavone, M. Artini, C. Atzori, A. Bäcker, M. Baglioni, A. Czerniak, M. De Bonis, H. Dimitropoulos, I. Foufoulas, M. Horst, K. Iatropoulou, P. Jacewicz, A. Kokogiannaki, S. La Bruzzo, E. Lazzeri, A. Löhden, P. Manghi, A. Mannocci, N. Manola, E. Ottonello, and J. Schirrwagen. OpenAIRE Covid-19 publications, datasets, software and projects metadata., June 2021. URL https://zenodo.org/record/6638745. Version Number: 2.0 Type: dataset.

[13] 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. URL http://www.dlib.org/dlib/january17/burton/01burton.html.

[14] P. Calyam, N. Wilkins-Diehr, M. Miller, E. H. Brookes, R. Arora, A. Chourasia, D. M. Jennewein, V. Nandigam, M. Drew LaMar, S. B. Cleveland, G. Newman, S. Wang, I. Zaslavsky, M. A. Cianfrocco, K. Ellett, D. Tarboton, K. G. Jeffery, Z. Zhao, J. González-Aranda, M. J. Perri, G. Tucker, L. Candela, T. Kiss, and S. Gesing. Measuring success for a future vision: Defining impact in science gateways/virtual research environments. *Concurrency and Computation: Practice and Experience*, 33(19), 2021. doi: https://doi.org/10.1002/cpe.6099. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.6099.

[15] E. Campana, E. Ciappi, and G. Coro. The role of technology and digital innovation in sustainability and decarbonization of the blue economy. *Bulletin of Geophysics and Oceanography*, 62(3):123–130, September 2021.

[16] L. Candela, V. Grossi, P. Manghi, and R. Trasarti. A workflow language for research e-infrastructures. *International Journal of Data Science and Analytics*, 11(4):361–376, 2021. doi: 10.1007/s41060-020-00237-x. URL https://doi.org/10.1007/s41060-020-00237-x.

[17] F. A. Cardillo and U. Straccia. Fuzzy OWL-Boost: Learning fuzzy concept inclusions via real-valued boosting. *Fuzzy Sets and Systems*, 438:164–186, June 2022. ISSN 01650114. doi: 10.1016/j.fss.2021.07.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0165011421002426.

[18] G. Casini and U. Straccia. A rational entailment for expressive description logics via description logic programs. Technical Report 019, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", 2021. URL https://doi.org/10.32079/ISTI-TR-2021/019.

[19] G. Casini, T. Meyer, and G. Paterson-Jones. KLM-style defeasibility for restricted first-order logic. In L. Amgoud and R. Booth, editors, *Proceedings of 19th International Workshop on Non-Monotonic Reasoning*, pages 184–193, 2021.

[20] G. Casini, T. Meyer, and I. Varzinczak. Contextual Conditional Reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6254–6261, May 2021. doi: 10.1609/aaai.v35i7.16777. URL https://ojs.aaai.org/index.php/AAAI/article/view/16777.

[21] G. Casini, T. Meyer, and I. Varzinczak. Situated conditional reasoning. Technical Report 009, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", 2021. URL https://doi.org/10.32079/ISTI-TR-2021/009.

[22] G. Coro and M. Bjerregaard Walsh. An intelligent and cost-effective remote underwater video device for fish size monitoring. *Ecological Informatics*, 63:101311, 2021. ISSN

1574-9541. doi: 10.1016/j.ecoinf.2021.101311. URL https://www.sciencedirect.com/science/article/pii/S1574954121001023.

[23] G. Coro, A. Ellenbroek, and P. Pagano. An open science approach to infer fishing activity pressure on stocks and biodiversity from vessel tracking data. *Ecological Informatics*, 64:101384, 2021. ISSN 1574-9541. doi: https://doi.org/10.1016/j.ecoinf.2021.101384. URL https://www.sciencedirect.com/science/article/pii/S1574954121001758.

[24] G. Coro, F. V. Massoli, A. Origlia, and F. Cutugno. Psycho-acoustics inspired automatic speech recognition. *Computers & Electrical Engineering*, 93:107238, 2021. ISSN 0045-7906. doi: https://doi.org/10.1016/j.compeleceng.2021.107238. URL https://www.sciencedirect.com/science/article/pii/S0045790621002251.

[25] D. Dimarchopoulou, P. J. Mous, E. Firmana, E. Wibisono, G. Coro, and A. T. Humphries. Exploring the status of the Indonesian deep demersal fishery using length-based stock assessments. *Fisheries Research*, 243:106089, Nov. 2021. ISSN 01657836. doi: 10.1016/j.fishres.2021.106089. URL https://linkinghub.elsevier.com/retrieve/pii/S0165783621002174.

[26] D. Dosso, S. Ferilli, P. Manghi, A. Poggi, G. Serra, and G. Silvello. Information and Research Science connecting to Digital and Library Science: Report on the 17th Italian Research Conference on Digital Libraries, IRCDL2021. *ACM SIGMOD Record*, 50(2):44–47, Aug. 2021. ISSN 0163-5808. doi: 10.1145/3484622.3484635. URL https://dl.acm.org/doi/10.1145/3484622.3484635.

[27] L. Frosini, P. Pagano, L. Candela, M. Simi, and C. Bernardeschi. Relock: a resilient two-phase locking restful transaction model. *Service Oriented Computing and Applications*, 15(1):75–92, 2021. doi: 10.1007/s11761-020-00311-z. URL https://doi.org/10.1007/s11761-020-00311-z.

[28] V. Grossi, F. Giannotti, D. Pedreschi, P. Manghi, P. Pagano, and M. Assante. Data science: a game changer for science and innovation. *International Journal of Data Science and Analytics*, 11 (4):263–278, May 2021. ISSN 2364-415X, 2364-4168. doi: 10.1007/s41060-020-00240-2. URL https://link.springer.com/10.1007/s41060-020-00240-2.

[29] E. Lazzeri, F. Tanlongo, G. Pavone, F. Alpi, A. Ansuini, E. Bertazzon, D. Bonaccorsi, F. Cappelluti, S. Casati, D. Castelli, R. Cippitani, V. Colcelli, A. Costantini, S. Cozzini, E. Degl'Innocenti, F. Di Donato, S. Di Giorgio, I. Fava, S. Fiore, M. Forni, G. Galimberti, E. Giglia, A. Giorgetti, S. Kurapati, M. Landoni, M. Lavitrano, C. Marras, F. Niccolucci, M. Occioni, E. Osmenaj, G. Paolini, V. Pasquale, C. Petrillo, R. Pugliese, E. Ripepi, G. Rivoira, G. Rossi, S. Salon, A. Sarretta, A. Sartori, D. Spiga, D. Tamagno, A. M. Tammaro, M. Vellico, M. Vignocchi, and D. Zane. Competence Centre ICDI per Open Science, FAIR, ed EOSC - mission, strategia e piano d'azione. Technical Report ISTI-TR-2021/023, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", 2021. URL https://doi.org/10.32079/ISTI-TR-2021/022.

[30] E. Lazzeri, F. Tanlongo, G. Pavone, F. Alpi, A. Ansuini, E. Bertazzon, D. Bonaccorsi, F. Cappelluti, S. Casati, D. Castelli, R. Cippitani, V. Colcelli, A. Costantini, S. Cozzini, E. Degl'Innocenti, F. Di Donato, S. Di Giorgio, I. Fava, S. Fiore, M. Forni, G. Galimberti, E. Giglia, A. Giorgetti, S. Kurapati, M. Landoni, M. Lavitrano, C. Marras, F. Niccolucci, M. Occioni, E. Osmenaj, G. Paolini, V. Pasquale, C. Petrillo, R. Pugliese, E. Ripepi, G. Rivoira, G. Rossi, S. Salon, A. Sarretta, A. Sartori, D. Spiga, D. Tamagno, A. M. Tammaro, M. Vellico, M. Vignocchi, and D. Zane. ICDI Competence Centre for Open Science, FAIR and EOSC - Mission, strategy and action plan. Technical Report ISTI-TR-2021/023, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", 2021. URL https://doi.org/10.32079/ISTI-TR-2021/023.

[31] 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.

[32] P. Manghi, C. Atzori, A. Bardi, M. Baglioni, J. Schirrwagen, H. Dimitropoulos, S. La Bruzzo, I. Foufoulas, A. Löhden, A. Bäcker, A. Mannocci, M. Horst, A. Czerniak, K. Kiatropoulou, A. Kokogiannaki, M. De Bonis, M. Artini, E. Ottonello, A. Lempesis, A. Ioannidis, and F. Summan. OpenAIRE Research Graph: Dumps for research communities and initiatives., Jan. 2021. URL https://zenodo.org/record/4439644. Version Number: 1.0 Type: dataset.

[33] P. Manghi, C. Atzori, A. Bardi, M. Baglioni, J. Schirrwagen, H. Dimitropoulos, S. La Bruzzo, I. Foufoulas, A. Löhden, A. Bäcker, A. Mannocci, M. Horst, A. Czerniak, K. Kiatropoulou, A. Kokogiannaki, M. De Bonis, M. Artini, E. Ottonello, A. Lempesis, A. Ioannidis, and F. Summan. OpenAIRE Research Graph: Dumps for research communities and initiatives, Apr. 2021. URL https://zenodo.org/record/4722961. Version Number: 2.0 Type: dataset.

[34] P. Manghi, C. Atzori, A. Bardi, M. Baglioni, J. Schirrwagen, H. Dimitropoulos, S. La Bruzzo, I. Foufoulas, A. Löhden, A. Bäcker, A. Mannocci, M. Horst, A. Czerniak, K. Kiatropoulou, A. Kokogiannaki, M. De Bonis, M. Artini, E. Ottonello, A. Lempesis, A. Ioannidis, and F. Summan. OpenAIRE Research Graph: Dumps for research communities and initiatives, July 2021. URL https://zenodo.org/record/5095707. Version Number: 2.1 Type: dataset.

[35] G. Pavone, E. Lazzeri, M. Circella, and F. De Leo. Progettare un evento divulgativo online. l'esperienza di "fai una domanda su covid-19, gli esperti rispondono". Technical report, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", 2021.

[36] A. A. Salatino, A. Mannocci, and F. Osborne. Detection, Analysis, and Prediction of Research Topics with Scientific Knowledge Graphs. In Y. Manolopoulos and T. Vergoulis, editors, *Predicting the Dynamics of Research Impact*, pages 225–252. Springer International Publishing, Cham, 2021. ISBN 978-3-030-86667-9 978-3-030-86668-6. doi: 10.1007/978-3-030-86668-6_11. URL https://link.springer.com/10.1007/978-3-030-86668-6_11.

[37] T. Vergoulis, I. Kanellos, C. Atzori, A. Mannocci, S. Chatzopoulos, S. La Bruzzo, N. Manola, and P. Manghi. BIP! DB: A Dataset of Impact Measures for Scientific Publications. In *Companion Proceedings of the Web Conference 2021*, pages 456–460, Ljubljana Slovenia, Apr. 2021. ACM. ISBN 978-1-4503-8313-4. doi: 10.1145/3442442.3451369. URL https://dl.acm.org/doi/10.1145/3442442.3451369.

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 a Research Technologist of the "Istituto di Scienza e Tecnologie della Informazione A. Faedo" (ISTI), an institute of the Italian National Research Council (CNR). He is skilled in leadership, with a strong foundation in math, logic, and cross-platform coding. He holds a Ph.D. on Information Engineering, a Master degree (M.Sc.) on Information Technologies, and a degree (B.Sc.) on Computer Science from University of Pisa. He has more than 15 years of experience working on distributed systems, e-infrastructures and Virtual Research Environments. Massimiliano is also the Operations and VRE Manager of the D4Science Infrastructure where he develops policies, procedures, and staff awareness as means to maintain performance and meet end users demands. Massimiliano joined ISTI in 2007, he worked for several EU Projects such as AGINFRAPlus, BlueBRIDGE, PARTHENOS, SoBigData, EOSCPilot, iMarine, EU-BrazilOpenBio, D4Science II, D4Science and DILIGENT. Within these projects, he progressively covered different positions. Currently, he is involved with various responsibility roles in the EU H2020 projects ARIADNEPlus, Blue-Cloud, DESIRA, MOVING, SoBigData-PlusPlus and RISIS2.

Claudio Atzori is a computer science researcher at the National Research Council of Italy, Istituto di Scienza e Tecnologie 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 role of product manager 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.

Leonardo Candela is computer science senior researcher at the National Research Council of Italy, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo". 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 connecting research and development. In fact, he has been involved in several EU-funded projects called to develop Digital Libraries & Data Infrastructures and he 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 group. Under her supervision, the InfraScience team coordinated and participated in several EU and nationally funded projects on Digital Libraries and Research Data Infrastructures. In particular, she has been the co-ordinator of the EU projects that have developed the D4Science infrastructure and technical coordinator of those that have developed the OpenAIRE one. She has participated in experts groups dedicated to the shaping of the European Open Science Cloud. 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 is author of 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 is graduated in Computer Science at the University of Pisa and his research focuses on the 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 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 technical director and involved on several European and National project. Her current research activities range from the digital image processing to techniques for image retrieval and automatic annotation tool. 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.

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 Scholexpler 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).

Emma Lazzeri is Open Science Manager at Consortium GARR defining strategies, tools and providing training and information on issues related to open science and research data management. She was affiliated researcher at the Institute of Information Science and Technologies of the Italian National Research Council in Pisa Italy up to May 2022. Emma coordinates the ICDI (Italian Computing and Data Infrastructure) Task Force for the National Competence Center for Open Science, FAIR and EOSC. She is part of numerous international boards linked to EOSC and Open Science, she collaborates in both national and European projects and initiatives. She is the Coordinator of Skills4EOSC (2022-2025), the project funded by the Horizon Europe programme that will build a network of Competence Centers in Europe for the training and updating of European researchers and professionals in the field of FAIR data and open science.

Lucio Lelii 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. He is currently member of the BlueCloud EU Project.

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 background on 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 involve design and development of (Web-)GUI based on several framework (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 framework 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 (Refresh2018-2020) held at the European Computational Social Science symposium, and at SocInfo 2020 respectively.

Enrico Ottonello is a research fellow at Istituto di Scienza e Tecnologie dell'Informazione A. Faedo (ISTI), an institute of the Italian National Research Council (CNR), since December 2018. He graduated in computer science at the University of Pisa in 2002. He played the role of software engineer for several companies since 2003 up to 2018. At ISTI-CNR he worked for several EU Projects including OpenAIRE-Advance and OpenAIRE-Connect. His research interests and activities are focusing on data cleaning, anomaly detection, and enrichment in scholarly communication graphs.

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.

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, EOSCSecretariat 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, SoBigDataPlus, 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 the design and development of distributed environments' services aimed to manage scientific data, with special attention to Ecological Niche Modelling approaches. These studies involved 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, EUBrazilOpenBio, ENVRI, Venus-C, D4Science-II, D4Science projects.

Umberto Straccia is a 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.