HEAL-Link and HELIX open collaboration to facilitate and promote scholarly communication through Open Access and engage Research Infrastructures in Open Science

Leonidas Pispiringas
HEAL-Link (Hellenic Academic Libraries Link), Aristotle University, Thessaloniki, Greece.
lpisp@heal-link.gr

Spiros Athanasiou
Information Management Systems Institute, Athena Research Center, Greece.
spathan@imis.athena-innovation.gr

Zisis Simaioforidis
HEAL-Link (Hellenic Academic Libraries Link), Aristotle University, Thessaloniki, Greece.
zsimaiof@heal-link.gr

Dimitrios Skoutas
Information Management Systems Institute, Athena Research Center, Greece.
dskoutas@imis.athena-innovation.gr

Abstract:

The quinquennium of 2018-2022 will bring dramatic changes in the field of scientific research, providing added value to global and national academic and research communities. Open Science, the leading policy priority for EU science, comprises inextricably linked horizontal e-Infrastructures and vertical Research Infrastructures to accommodate both the generic-purpose and domain-specific needs of the community. In this landscape, we observe an emerging convergence of scholarly communication with scientific data management, aiming to provide researchers with a comprehensive suite of services for the management, discovery, sharing, use, archiving of scientific data.

The Hellenic Data Service (HELIX) is a new national eInfrastructure supporting this paradigm, with a mandate to accommodate the entire lifecycle of data-intensive research. HELIX will deliver services for storing, managing, discovering, processing, analyzing, visualizing, and archiving diverse open and proprietary scientific data that will be shared and reused in a cross-domain and cross-discipline manner. HELIX will support scalable and interactive computing, lowering the entry barrier for researchers, and offering a single point to discover, share and reuse high-value scientific data assets linked with publications and software.

The HEAL-Link Consortium (Hellenic Academic Libraries Link) is actively involved in projects, initiatives, developments to the scientific publication landscape and its goal is to act in a way that is beneficial to its members and to the scientific community in Greece. HEAL-Link has a mandate to promote Open Access policies and power sustainable knowledge in the digital age. Its strategic goals are to facilitate and promote scholarly communication through Open Access, cultivate the digital skills of scientists, and engage Research Infrastructures in Open Science. Within this role, HEAL-Link has initiated an open collaboration with HELIX to engage its members, comprising all Academic institutions in Greece, and operate a scientific data catalogue/repository integrated within the broader HELIX ecosystem. HEAL-Link will also organize, coordinate and train the personnel of the Academic Institutions on: (a) generic and discipline-specific Data Management Plans (DMPs), (b) drafting and promoting policies to adopt and enforce Open Access/Open Science.

We argue that the joint activity of HEAL-Link and HELIX represents a new evolutionary step for scholarly communication, facilitating and accelerating the scientific process, while also streamlining the re(use) of scientific output.

Keywords: open science, research data repositories, open access, persistent identifiers, metadata

Introduction

During recent years there has been a big movement of the academic community towards Open Science. Open Access, Open Data, Open Research, Policies, Evaluation are the different aspects of Open Science, that when are effectively put together, can reach to a desired outcome. There have been many definitions regarding Open Science. Most of them rely on the same basis and differ on how the stakeholders evaluate each aspect as most important or realistic.

The Open Knowledge International (OKFN, https://okfn.org/) in 2005 issued the Open Definition as: "*Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)*". Wikipedia's article defines Open Science as "*the movement to make scientific research (including publications, data, physical samples, and software) and its dissemination accessible to all levels of an inquiring society, amateur or professional*" [Woelfle et. al. (2011)]. Open science is transparent and accessible knowledge that is shared and developed through collaborative networks [Vicente-Saez et. al. (2018)]. It encompasses practices such as publishing open research, campaigning for open access, encouraging scientists to practice open notebook science, and generally making it easier to publish and communicate scientific knowledge". The European Union funded project FOSTER (https://www.fosteropenscience.eu) defines Open Science as "*the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society*".

Open Science, the leading policy priority for European Union science, comprises inextricably linked horizontal e-Infrastructures and vertical Research Infrastructures to accommodate both the generic-purpose and domain-specific needs of the global and national academic and research communities. The European Union actively participates through the European Open Science Cloud (EOSC, https://www.eosc-portal.eu), a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines by federating existing scientific data infrastructures, currently dispersed across disciplines and the EU Member States. EOSC is the

implementation of the European Union's concept of Open Science. There are projects, for infrastructures and services, that are financially supported by the EU Framework Programme for Research and Innovation (Horizon 2020) in order to implement the EOSC (EOSCpilot, EOSC-hub, OpenAIRE-Advance, FREYA, RDA Europe 4.0, et. al.).

Research Data

Research data is any information that has been collected, observed, generated or created to validate original research findings. Although usually digital, research data also includes non-digital formats such as laboratory notebooks and diaries. In this paper, we will focus on digital data, as this is the only form used by Research Data Repositories.

Research data can take many forms. It might be:

- documents, spreadsheets
- laboratory notebooks, field notebooks, diaries
- questionnaires, transcripts, codebooks
- audiotapes, videotapes
- photographs, films
- test responses
- slides, artefacts, specimens, samples
- collections of digital outputs
- data files
- database contents (video, audio, text, images)
- models, algorithms, scripts
- contents of an application (input, output, logfiles for analysis software, simulation software, schemas)
- methodologies and workflows
- standard operating procedures and protocols

Research data can be generated for different purposes and through different processes. The HEAL-Link - HELIX Research Data Repository will accommodate all aforementioned types of data produced by HEAL-Link's members (Greek academic institutions and research centres).

Some main groups of data are:

- Observational data; captured in real-time, and is usually irreplaceable, for example sensor data, survey data, sample data, and neuro-images.
- Experimental data; captured from lab equipment. It is often reproducible, but this can be expensive. Examples of experimental data are gene sequences, chromatograms, and toroid magnetic field data.
- Simulation data; generated from test models where model and metadata are more important than output data. For example, climate models and economic models.
- Derived or compiled data transformed from pre-existing data points. It is reproducible if lost, but this would be expensive. Examples are data mining, compiled databases, and 3D models.
- Reference or canonical data; a static or organic conglomeration or collection of smaller (peer-reviewed) datasets, most probably published and curated. For example, gene sequence databanks, chemical structures, or spatial data portals.

Research Data Lifecycle

HEAL-Link and HELIX proposes a very specific research data lifecycle in order to be used by all Academic Institutions that will be used as data sources. The purpose of this, is for them to be able to:

- plan effectively
- be aware of existing data and how data adds to it
- think carefully about how data will be gathered and processed

The schematics of the Research Data Lifecycle proposed is depicted below:
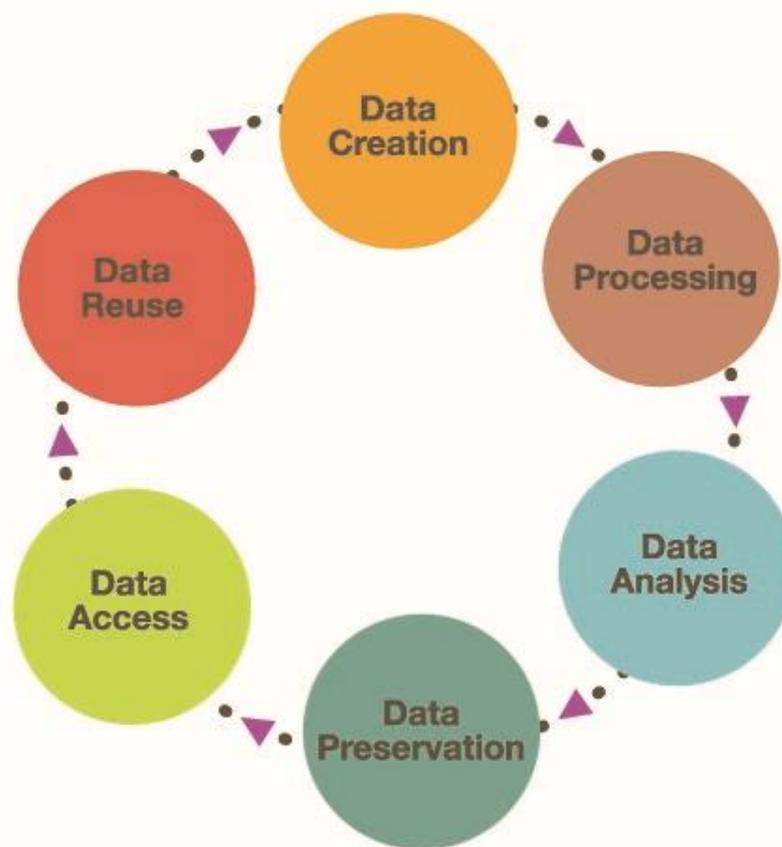

Figure 1: Research Data Lifecycle

The Research Data Lifecycle model describes and identifies the steps to be taken at the different stages of the research cycle to ensure successful data curation and preservation. There are several stages in the research data lifecycle, e.g., data creation, data processing, data analysis. There are a few models that one can make use to plan the data management activities, for example, DCC Curation Lifecycle Model [The International Journal of Digital Curation, Sarah Higgins, 2008]

It is also highlighted in Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information [Pennock (2007)] that the lifecycle approach is necessary because:

- Digital materials are fragile and susceptible to change from technological advances throughout their lifecycle, i.e. from creation onwards;
- Activities (or lack of) at each stage in the lifecycle directly influence our ability to manage and preserve digital materials in subsequent stages;
- Reliable re-use of digital materials is only possible if materials are curated in such a way that their authenticity and integrity are retained.

Additionally, each stage of a research data lifecycle may comprise many small activities. For example, data creation may involve new data collection, reuse of existing data, capturing and creating metadata.

For all the previous reasons, it is an extreme priority for us to facilitate all of these steps to all HEAL-Link members participating in the project.

FAIR Principles

Another major guideline and prerequisite that is proposed to all Academic institutions that will be used as data sources, is that all data must be FAIR, which means that they must be Findable, Accessible, Interoperable and Reusable.

As proposed by [Wilkinson, et al., (2016)] the FAIR Data Principles are the following:

1. Findable:

     F1. (meta)data are assigned a globally unique and persistent identifier

     F2. data are described with rich metadata (defined by R1 below)

     F3. metadata clearly and explicitly include the identifier of the data it describes

     F4. (meta)data are registered or indexed in a searchable resource

2. Accessible:

     A1. (meta)data are retrievable by their identifier using a standardized communications protocol

          A1.1 the protocol is open, free, and universally implementable

          A1.2 the protocol allows for an authentication and authorization procedure, where necessary

     A2. metadata are accessible, even when the data are no longer available

3. Interoperable:

     I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

     I2. (meta)data use vocabularies that follow FAIR principles

I3. (meta)data include qualified references to other (meta)data

4. Reusable:

R1. meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards

HEAL-Link - HARDMIN

The Hellenic Academic Libraries Link Consortium (HEAL-Link, https://www.heal-link.gr/) operates since 1998. Its members are all the Academic institutions in Greece, all the research institutions operating under the auspices of the Hellenic Ministry of Education, as cooperating members, the Academy of Athens, the National Library of Greece and the Institute of Educational Policy. HEAL-Link is actively involved in projects, initiatives, developments to the scientific publication landscape and its goal is to act in a way that is beneficial to its members and to the scientific community in Greece. HEAL-Link has a mandate to promote Open Access policies and power sustainable knowledge in the digital age. Its strategic goals are to facilitate and promote scholarly communication through Open Access, cultivate the digital skills of scientists, and engage Research Infrastructures in Open Science.

HEAL-Link's Research Data Infrastructure, co-developed with HELIX, is called HARDMIN, the research data repository for the Academic institutions in Greece (https://hardmin.heal-link.gr).

The streams of action of the HEAL-Link's Research Data Infrastructure are (a) Hardware, (b) Software, (c) Permanent Identification Services, (d) Support, Promotion, Publicity & Training Plans, (d) Research Data Management Plan and (f) Policies. The first three streams of actions have already been implemented, with the software part undergoing testing and evaluation and expected to be available to all HEAL-Link members by the end of 2019. The rest of the actions that are needed to facilitate and promote scholarly communication through Open Access and engage Research Infrastructures in Open Science for the for the HEAL-Link's members are going to be planned accordingly.

Hardware
A central storage infrastructure has been supplied and installed at the IT Center of Aristotle University with a capacity of ~1PiB (Peta Byte) in high availability, which is necessary to ensure optimal operational performance (uptime).

Software
HEAL-Link has initiated an open collaboration with HELIX to engage its members, comprising all Academic institutions in Greece, and operate the HARDMIN repository integrated within the broader HELIX ecosystem, which focuses on the National Roadmap for Research Infrastructures and is aligned with the implementation of the European Open Science Cloud.

Permanent Identification Services

The further enhancement, traceability and recognition of the content of the repository from the global research community depends on the need to exploit and incorporate good practices and technologies that have already been implemented by research centers, libraries and scientific publishers globally (DOIs: https://doi.org via Datacite: https://datacite.org and ORCID IDs via ORCID: https://orcid.org).

The necessity and obligation of assigning Digital Object Identifiers (DOIs) to the research data and the accompanying metadata arises, as appropriate. A Digital Object Identifier (DOI) is a unique alphanumeric string to identify research data and other research output and provide a persistent link to its location on the Internet. DataCite provides persistent identifiers (DOIs) for research data and other research outputs. HEAL-Link joined DataCite as a Consortium member in order to be able to assign DOIs to the research outputs of its members and to the research data populated to the HARDMIN repository.

ORCID provides a persistent digital identifier that distinguishes the researchers and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between the researchers and their professional activities ensuring that their work is recognized. HEAL-Link's participation to ORCID representing the Greek Academic Community, is indispensable both for the unambiguous identification of the researchers of HEAL-Link members and for the proper functioning of the HEAL-Link's research data infrastructure.

HELIX

HELIX (Hellenic Data Service) is the cloud-based scientific eInfrastructure for data-intensive research developed and operated by the Athena Research Center (Athena RC), handling the data management, analysis, sharing, and reuse needs of researchers and innovators in a cross-sector, scalable, and cost-effective manner. HELIX provides its services directly to its users (researchers, scientists, students) as an autonomous cloud infrastructure in support of data sharing, open access publishing, and data experimentation. Further, it is an integral building block for all national Research Infrastructures and industrial platforms, offering its services in a horizontal manner to thematic and domain-specific infrastructures and communities engaging research organizations (Research Centers, Universities), scientific infrastructures (e.g., ELIXIR, OpenAIRE), data catalogues and repositories (e.g., HEAL-link, geodata.gov.gr), as well as the private sector, with a mandate to promote Data Science and the Data Economy. As such, HELIX achieves economies of scale, added value, and maximum return of national human and capital investments. HELIX provides cloud services for sharing, discovering, managing, processing, analyzing, visualizing, and archiving Big Data. Further, it assumes the role of OpenAIRE for Greece, and being in close interaction with RDA, provides existing pathways for implementing EU's Open Access policies. In addition, HELIX supports interactive and scalable computing paradigms for researchers and innovators, exposing large-scale HPC and distributed computing infrastructures over standards-based service end-points and Jupyter notebooks.

HELIX currently comprises three broad service offerings.

- HELIX Pubs. Provides OA publications of Greek research and scientific organizations harvested via OpenAIRE's facilities, enabling the cross-sector nation-

wide discovery and linking of publications with their output, as well as the provision of analytics and KPIs for national research funders.

- HELIX Data. Provides catalogue and repository services FAIR scientific and industrial data assets, enabling users to discover, view, download, (re)use, and subsequently analyze them within HELIX Lab.
- HELIX Lab. Offers a collection of both generic-purpose and domain-specific services and APIs for data analysis, processing, and experimentation, for interactive and scalable computing.

The novelty of HELIX as a Research eInfrastructure, lies within its comprehensive coverage of the scientific process, enabling researchers to discover, use, experiment, and contribute their output through a single integrated service. For example, a user discovering a publication from HELIX Pubs can find and download the related data assets and software from HELIX Data, and experiment with the same data assets within HELIX Lab. A high-level architecture of HELIX is presented in the following figure. The reader is invited to visit https://www.hellenicdataservice.gr for more detailed information regarding HELIX and its services.

Implementation Methodology

HEAL-Link has initiated an open collaboration with HELIX to engage its members, comprising all Academic institutions in Greece, and operate a scientific data catalogue and repository integrated within the broader HELIX ecosystem, the HARDMIN research data repository. Towards this, the HELIX team has designed and deployed a federated scientific data catalogue and repository aiming to (a) harvest and provide scientific data assets from HEAL-Link's members, and (b) make these data assets available for integration and ingestion within HELIX's service offerings. This approach perfectly serves the requirements and mandate of HEAL-Link and its members, while ensuring economies of scale and interoperability by design. In addition, it enables HEAL-Link and its users to tap into the output of investments in infrastructures that are orders of magnitude greater, ensure uniform de facto FAIR principles, and serve the emerging landscape of the European Open Science Cloud (EOSC).

The data catalogue and repository available to HEAL-Link members is a customized installation of the HELIX Data catalogue, suitably customized (e.g., theme, vocabularies) to provide a suite of comprehensive services for scientific data management. The software is based on CKAN, the most popular data catalogue platform, which powers among others, data.gov (the first open data portal) and data.europa.eu (EUs own open data portal). Due to its popularity and application worldwide, CKAN is considered as a de facto technical solution for open data publishing with a broad community and excellent support for complex open publishing workflows. CKAN is implemented in Python and supports an extensible plug-in architecture to introduce new functionality (e.g., schemas, hierarchies, visualization, analytics). A series of CKAN plugins and extensions have established scientific data in CKAN as first-class citizens, streamlining the publishing and discovery/reuse lifecycle for publishers and users. This output is further extended in HELIX, introducing support for Big Data processing workflows, integrated interactive and scalable computing for geospatial assets, and demand-aware resource management over IaaS/HPC infrastructures.

In the following, we briefly present the core functionalities supported by the HARDMIN research data repository, currently deployed and validated at its own cloud infrastructure.

- Metadata. The catalogue extends the basic core CKAN schema to inherently support metadata for published data assets according to DataCite Metadata Schema 4.2. Support for recommended and optional DataCite properties is simple via a lightweight programmatic extension ensuring that all catalogue services and UI/UI elements are updated in a uniform manner. In addition, a custom plugin allows the support for domain-specific metadata schemas (e.g., INSPIRE for geospatial data), allowing the user to select the type of metadata she wants to create (or upload), with the corresponding metadata mapped to the core DataCite schema. This functionality allows the reuse of (any) existing metadata already available, thus further lowering the entry barrier and repurposing existing investments. Moreover, it allows visitors to request the on-demand transformation of metadata in other schemas, bridging the gap to domain-specific data infrastructures.
- Publishing workflows. A user can publish a data asset via a lightweight web-based GUI, enabling her to either create the metadata from scratch or import them (if available) at a metadata schema of her choosing. Any metadata attributes not supported from the source schema that are required for DataCite, are explicitly requested. At a next step, the user uploads or links to the corresponding data assets and can define, among others, their license, embargo period, etc. The new resource is available for curation from the administrator of the user's organization (see user hierarchies next), may undergo a series of revisions, and becomes publicly available after it has been accepted.
- Identifiers. Resources receive DOI identifiers automatically by the system and immediately before they become public. In addition, the catalogue maintains separate unique identifiers with a local context (surrogates) to assist maintenance, potential migration, and support for multiple identifier schemes to ensure interoperability.
- Organizations. An organization (e.g., HEAL-Link member) is a first-class entity in the catalogue, with published data assets being linked to them (based on user affiliation), and with organization-level user management being entirely handled by organization administrators. These can import and invite users, assign and revoke publishing roles, customize the publishing workflow, and overall, implement any provisions required from their open science policies and data management plans (horizontal or at a project level).
- Harvesting. The catalogue can automatically and periodically harvest data assets from any existing data catalogues, repositories and infrastructures, importing the corresponding metadata and data assets. The harvesting policy can be configured from the organization and up to the data asset level, affording a highly grained control over the harvesting frequency and scope. A harvester job can, for example, retrieve only the metadata, or a thematic subset of the data assets, or the most recent version of data assets, as needed. Currently, the catalogue supports harvesting both from de facto (e.g., OAI-PM, OGC) and de jure standards (e.g., CKAN catalogue), as well as any web-accessible application via an extendable and customizable harvester plugin.
- Authentication. User authentication and authorization is based on the federated AAI of GRNET, allowing every user to sign up and login using her institutional credentials. In addition, a user may link third-party credentials to her account (e.g., ORCID), which will form the basis for a future series of integrated services regarding user and asset management.
- Asset storage. Published data assets are automatically assigned and automatically transferred between one of two different storage end-points, with different speed and cost characteristics (fast, slow). The transfer of data assets follows configurable

policies from the organization, down at the asset level, depending on the data asset's popularity, size, or significance.

- HELIX integration. Given the out-of-the-box interoperability between the HEAL-Link and HELIX data catalogues, their integration has been exceptionally simple and a best-practice example for similar federated catalogues at a national or domain level. HELIX automatically harvests HEAL-Link's data catalogue, imports all available metadata and data assets, stores them at its own cloud-based infrastructure and selectively makes them available (i.e., ingest into appropriate data engines, transfer to HPC) to allow their use across the services of HELIX Lab. For example, a user may integrate a data asset for analysis and experimentation in the HELIX Lab Jupyter notebook facility without the need to download or import the data asset, which is extremely important especially for Big Data collections. In another setting, the user may submit an HPC job using the same data asset directly within HELIX Lab's DASK framework, without again needing to deal with complex and resource-intensive data handling and wrangling operations.

Future work

The HARDMIN repository is currently undergoing user testing and validation, with the complete suite of its services expected to become available to HEAL-Link's members by the end of 2019. This activity is expected to serve as a best-practice example for national research and industrial data catalogues, paving the way for the implementation of the Greek EOSC offerings. Our current efforts are therefore focused on expanding the support for de facto and de jure metadata and data standards in a cross-sector manner, further simplifying and automating publishing workflows and operations/maintenance to lower the operating expenses for organizations, as well as expanding the automated ingestion of data assets into HELIX Lab services for specialized domains and user communities (e.g., neuroscience, nanotechnology).

In the following part we describe the aforementioned streams of actions that are very important in order for the HEAL-Link's research data infrastructure and the joint activity of HEAL-Link and HELIX to take the next evolutionary step for the scholarly communication in close collaboration with Athena Research & Innovation Center for the purposes of "OpenAIRE-Advance" the project regarding Open Access and Open Data in Europe.

Support, Promotion, Publicity & Training

HEAL-Link will organize, coordinate and train the personnel of the Academic Institutions on: (a) generic and discipline-specific Data Management Plans (DMPs), (b) drafting and promoting policies to adopt and enforce Open Access/Open Science.

Academic Libraries will undertake actions to support the Academic community and to promote the project on site. They will be in direct contact with local and regional promoters of the research of the Academic community, e.g. with the Research Committees of the Academic Institutions, to ensure that all partners are informed in a timely and up-to-date manner and that the data flow to the repository is constantly enhanced.

The promotion and publicity part will aim to meet those needs that are necessary to maximize the dissemination of the results, to inform the Academic community, to promote action abroad, and in general any action may offer added value.

The training will include the education of the Academic Libraries staff to the two main axes, (a) Research Data Management Plan and (b) Data Literacy. There will be held seminars by people with expertise in research data management. The initial training will be help with on-site seminars that will be organized in geographic locations in Greece so that staff from all Academic Libraries will be able to attend. The staff that will be trained will handle, at institutional level, the responsibility of the enrichment of the HARDMIN repository, the general promoting of the service and the actions to encourage the research staff to use the HARDMIN repository.

Research Data Management Plan

Through this action a Research Data Management Plan (DMP), that will be solicitous to all the phases of the research data life cycle, will be developed. The aim will be to initiate a DMP which will take into account the academic community's needs per field of science for the preparation, production, harvesting, maintenance, preservation and withdrawal of research data. This plan will be developed following a study of the specific needs expressed by the Academic community to provide solutions to all issues that may hinder the FAIR principles of research data. Academic Libraries have the ability and the opportunity, with a well-trained staff, to be an important link of the chain between the Academic community and the National Roadmap for Research Infrastructures.

Policies

The purpose of this action is to compose all the necessary policies that need to be enforced in order to produce meaningful results. These policies will cover all possible levels, e.g. institutional level, national level, and will be composed after a thorough study has been carried out in cooperation with all the research stakeholders.

Conclusion

HEAL-Link' Research Data infrastructure and HELIX is the beginning of a new era in scholarly communication. Various stakeholders are going to participate in the implementation of the European Open Science Cloud project in Greece and under the auspices and guidance of the Athena Research & Innovation Center and "OpenAIRE-Advance" project regarding Open Access and Open Data in Europe (https://www.openaire.eu/advance/), there will be successful and constructive collaborations for facilitating and accelerating the scientific process in Greece.

Acknowledgements

References

- Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016)
- Woelfle, Michael & Olliaro, Piero & H Todd, Matthew. (2011). Open science is a research accelerator. Nature chemistry. 3. 745-8. doi: 10.1038/nchem.1149.
- Vicente-Saez, Ruben & Martinez-Fuentes, Clara, 2018. "Open Science now: A systematic literature review for an integrated definition," Journal of Business Research, Elsevier, vol. 88(C), pages 428-436. doi: 10.1016/j.jbusres.2017.12.043
- Higgins, S. The DCC curation lifecycle model. International Journal of Digital Curation 3 (1), 134-140, 2008. doi: 10.2218/ijdc.v3i1.48
- Pennock, Maureen. (2007). Digital curation: A life-cycle approach to managing and preserving usable digital information. Library & Archives. 1.
- Pispiringas L, Simaioforidis Z, Ollandezou L, Florou P, Araka I. HEAL-Link: Research data Infrastructure. (2018) 24th Panhellenic Conference of Academic Libraries, TEI of Thessaly, Larissa, Greece
- ORCID, Connecting Research and Researchers https://orcid.org/organizations
- DataCite, Locate, identify, and cite research data with the leading global provider of DOIs for research data. https://datacite.org
- Open Knowledge International (OKFN, https://okfn.org/)
- The European Union funded project FOSTER (https://www.fosteropenscience.eu)
- The OECD Library (https://www.oecd.org)