



The banner features six icons in a row: a globe (red), a book (orange), two spheres (yellow), a money bag with a Euro symbol (green), a scale of justice (blue), and a bicycle (purple). Below the icons, the text 'AIUCD 2021' is prominently displayed. Underneath, it reads 'DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale' and '10° congresso annuale PISA 19-22 gennaio'. On the right side, a list of topics is shown in various colors: 'DIGITAL PUBLIC HUMANITIES' (red), 'OPEN CULTURE' (orange), 'RETI SOCIALI' (yellow), 'TECH ECONOMY' (green), 'E-PARTICIPATION' (blue), and 'TECNOLOGIE ASSISTIVE' (purple). The background includes binary code and a classical building facade.

AIUCD 2021

DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale

10° congresso annuale **PISA** 19-22 gennaio

DIGITAL PUBLIC HUMANITIES
OPEN CULTURE
RETI SOCIALI
TECH ECONOMY
E-PARTICIPATION
TECNOLOGIE ASSISTIVE

Versione PROVVISORIA del contributo presentato al Convegno Annuale

DISCLAIMER

Questa versione dell'abstract non è da considerarsi definitiva e viene pubblicata esclusivamente per facilitare la partecipazione del pubblico al convegno AIUCD 2021

Il Book of Abstract contenente le versioni definitive e dotato di ISBN sarà disponibile liberamente a partire dal 19 gennaio sul sito del convegno sotto licenza creative commons.

The discovery platform GOTRIPLE: An EOSC service for social sciences and humanities research

Francesca Di Donato¹, Suzanne Dumouchel², Monica Monachini³, Stefanie Pohle⁴

¹ ILC-CNR, Italy - francesca.didonato(«»)ilc.cnr.it

² Huma-num-CNRS, France - suzanne.dumouchel(«»)huma-num.fr

³ ILC-CNR, Italy - monica.monachini(«»)ilc.cnr.it

⁴ Max Weber Stiftung, Germany - pohle(«»)maxweberstiftung.de

ABSTRACT

In this paper we present TRIPLE - Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration - an on-going project funded as part of the European Horizon 2020 programme INFRAEOSC-02-2019 “Prototyping new innovative services” (2019-2023). The project’s main objective is to develop a multilingual and multicultural discovery solution for the social sciences and humanities (SSH), which will provide a single access point that allows users to explore, find, access and reuse materials such as literature, data, projects and researcher profiles at European scale. The paper first provides an overview of TRIPLE’s main goals and impacts. It then describes the methodology adopted for the design and development of the project platform, GOTRIPLE. Finally, it contextualises the project within the European research landscape, and more specifically in the European Open Science Cloud (EOSC) ecosystem. In the conclusion, some current challenges and open issues are presented.

KEYWORDS

EOSC, Open Science, scholarly communication, discoverability, OPERAS, TRIPLE

1. INTRODUCTION

TRIPLE - Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration - is a project funded under the European Commission program INFRAEOSC-02-2019 “Prototyping new innovative services”.¹ It involves a consortium of 19 partners from 13 European countries². The TRIPLE project’s main objective is to design and develop a multilingual and multicultural discovery platform dedicated to social sciences and humanities (SSH) resources, which enables researchers to discover and reuse SSH macro-typologies data, i.e. research data and publications, profiles and projects. The GOTRIPLE platform is conceived to be composed of a core component, built upon the Isidore search engine³ developed by Huma-num [1], and integrated with a variety of innovative tools, which include a web annotation service (Pundit⁴ [2]), a Trust Building System, and a recommender system. The front-end visualizations are based on open technologies developed by Open Knowledge Maps⁵. Moreover, in order to access data about projects, a crowdfunding platform will be connected to GOTRIPLE. A Forum complements the innovative services by offering an agora for channeling internal and external discussions among users.

GOTRIPLE platform is not a stand-alone application, but is designed to be one of the innovative services provided by OPERAS - the research infrastructure supporting open scholarly communication in the social sciences and humanities in the European Research Area.⁶ OPERAS services adopt a common Authentication and Authorization Infrastructure (AAI), thanks to the implementation of Virtual Organization common identity which uses federated authentication mechanisms. Apart from GOTRIPLE, OPERAS services include:

- OPERAS Metrics Service⁷: The service collects usage metrics and altmetrics from many different sources (such as Google Books, Matomo analytics, World reader, etc.) about the usage of monographs.
- OPERAS Publishing Service Portal (Pathfinder)⁸: The service will be composed of two parts: a presentation of the services as a catalog, and a wizard that walks researchers through a series of questions towards the service offering that best fits their needs.

¹ <https://cordis.europa.eu/project/id/863420>.

² <https://www.gotriple.eu>.

³ <https://www.huma-num.fr/service/isidore>.

⁴ <http://thepund.it>.

⁵ <https://openknowledgemaps.org>.

⁶ <http://operas-eu.org>.

⁷ <https://www.operas.unito.it/services/metrics-service/>.

⁸ <https://www.operas.unito.it/services/publishing-service-portal-pp/>.

- OPERAS Certification Service⁹. It collects the variety of peer reviewing practices from hundreds of monograph publishing houses, categorizes them, and provides a single access point to the list of certified peer reviewed monographs available in Open Access in the world.
- OPERAS Research for Society¹⁰: By developing the practices of academic blogging for SSH researchers, offered by the hypotheses.org platform, it is designed to be an interactive platform between SSH researchers and the society at large.
- OPERAS XML Toolbox: The toolbox provides tools to achieve interoperability between two standards, XML JATS, adopted by the academic publishing industry, and certain flavours of XML TEI (Text Encoding Initiative) adopted by the humanities research community for books and digital editions.

Within the OPERAS-P project¹¹, active discussions and collective think-tanks on the development of sustainable models of governance for infrastructures, business models for open scholarly publishing, groundbreaking concepts to address the fairification of SSH data, multilingualism, the future of scholarly writing as well as quality assessment of novel research outputs are on-going. As a next step, OPERAS aims at becoming a mature community, based on a set of stable national nodes, grounded on innovative plans for future development, and providing a set of services compatible with the European Open Science Cloud (EOSC)¹².

The following section presents TRIPLE's approach, focusing on the methodology we are following in the design and development of GOTRIPLE, based on a user-centered design approach. In section 3 we then contextualize TRIPLE as an OPERAS service in the European landscape. In the conclusion, some critical aspects are listed, which are offered for discussion.

2. METHODOLOGY

TRIPLE's work plan is based on a strict and strong interaction between different activities: the core is made by a double-helix structure: one is dedicated to data acquisition, and the other one to the co-design processes [5]. Both of them are like the TRIPLE DNA - which produces the discovery platform and its innovative services. This double-helix structure nurtures as well other activities such as the definition of business models for the platform's long-term sustainability and the integration of GOTRIPLE into the EOSC. Other aspects of the work plan include project management and the communication and dissemination of the project outputs and results. These interactions are described in Figure 1, where activities are grouped under eight work packages.

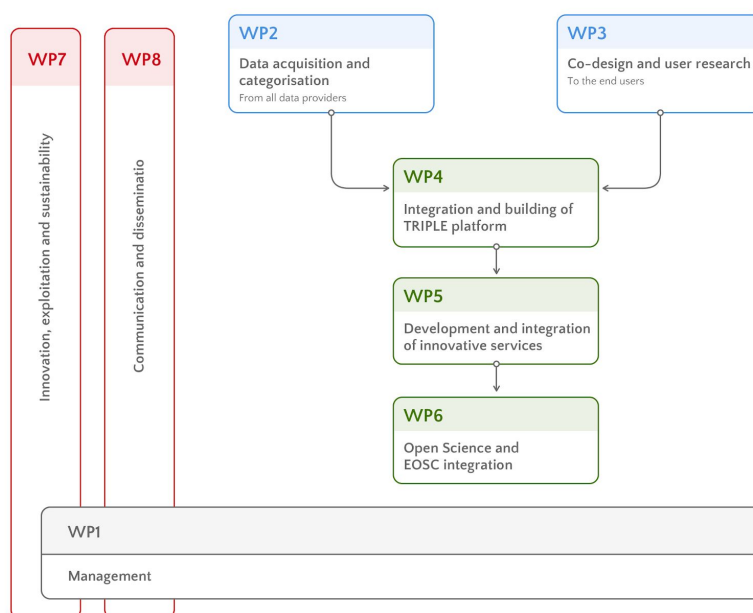


Fig.1 - TRIPLE pert chart

⁹ <https://www.operas.unito.it/services/certification-service-doab/>.

¹⁰ <https://www.operas.unito.it/services/research-for-society/>.

¹¹ <https://cordis.europa.eu/project/id/871069>.

¹² <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>.

The methodology adopted by the TRIPLE project is based on user-centered design techniques. TRIPLE's approach in fact puts users' needs at the core of the design process. We are using a mix of social sciences and design research approaches in order to study the future GOTRIPLE users and to elicit their needs as well as to co-design main functionalities of the platform. In particular, TRIPLE has tried to identify the user needs and to better understand the working practices of SSH researchers and researchers in order to design the GOTRIPLE platform. Through an online questionnaire over 925 survey responses could be collected. Current and future user needs also emerged from approx. 40 qualitative interviews. From the interviews analysis, we started defining the personas which are associated with their scenarios. Additionally, to obtain a comprehensive picture of existing working practices of researchers (e.g. what tools they are already using, how they structure collaborations, etc), to gather further information on user needs and to better identify potential user communities, a benchmarking of TRIPLE based on the results of the questionnaire is on-going. The consortium will also conduct a number of co-design/co-creation workshops and design research activities on the user journey in order to increase the appropriation capacity of SSH researchers towards the GOTRIPLE platform. Here, the plan is to use a double approach, based on hands-on workshops with the goal to sketch prototypes of interfaces, in an agile and iterative approach.

The second helix in the construction of GOTRIPLE is data acquisition through efficient coordination with data providers to ensure high quality indexation of content in six languages in addition to those already managed by ISIDORE (French, English, Spanish), i.e. Portuguese, Italian, German, Polish, Croatian, Greek. This objective is divided in two parts: to achieve effective data acquisition during the timeframe of the project; and to set up a framework that will be used after the end of the project to continue the data acquisition process by adding new content from the current data sources, adding/removing data sources and to provide guidelines for further additions of languages through machine learning and thesauri alignments.

We aim at overseeing the effective gathering of metadata and enrichment in the platform. A data acquisition plan defines the general specifications. On the one hand, support is provided to beneficiaries, OPERAS members and potential providers to comply with the platform specifications and OpenAire¹³. On the other hand, data discoverability is enhanced through the use of controlled vocabularies. A minimal set of categories is used for machine learning for all the languages of the platform. A mapping of existing classifications in the field is going to be realised, and new ones created, especially to facilitate searching in new languages on the platform. Links with the most used controlled vocabularies are provided through automated semantic annotation and completed with disambiguation tools. This work will make it possible to provide discipline-based metrics able to document SSH production scope and visibility.

At the beginning of the project, the first envisaged scenario had the objective to ensure the global harvesting process in the GOTRIPLE platform, including: a presentation of three protocols used for harvesting (OAI-PMH protocol, OAI ResourceSync and Sitemaps protocol pointing to RDFa metadata); the metadata of the resource (Dublin Core Metadata Element Set, DCMI Metadata Set), embedded metadata in the files (pdf, jpg); a description of the global harvesting process: harvesting of metadata, classification and annotation, enrichment with vocabularies, then indexing and conversion of metadata in a structured way in RDF model. However, by using feedback experience from the ISIDORE platform, this scenario has changed for a new one, i.e a push/pull model collecting metadata through existing aggregators (ISIDORE, OpenAIRE, Narcis etc.) with a delivery platform and not anymore acquiring metadata by a harvesting process. Nevertheless, the enrichment process including normalization, classification and annotation, is still made by the TRIPLE pipeline which gathers data provided by aggregators.

All of the information above converged into the "Data acquisition plan"¹⁴ which develops a full methodology to allow data providers and aggregators to be interoperable with both the GOTRIPLE platform according to the technical specifications and especially by adopting the TRIPLE data model and with OpenAIRE at the EOSC level. Indeed, on this second level of interoperability, the OpenAire guidelines will serve as a basis for the management and interoperability of metadata. Accordingly, the European Publication Guidelines and Aggregation Scheme is used as a management framework for cultural content. Adapted support (tutorials, webinars, workshops, documentation) will be proposed to fulfill the needs of each category of stakeholder to prepare their metadata. The OPERAS Executive Assembly, that gathers National Nodes, will be instrumental for the achievement of this task, particularly the University of Coimbra, that

¹³ <https://explore.openaire.eu/>.

¹⁴ Cf. TRIPLE deliverable D4.1 - Data Acquisition Plan. The deliverable, and all the project public deliverables released, will be published on Zenodo.

leads the Multilingualism Special Interest Group of OPERAS¹⁵, and IBL-PAN, that reaches out to Central European countries and languages.

3. TRIPLE IN THE CONTEXT OF OPEN SCIENCE RESEARCH

As mentioned before, GOTRIPLE is not conceived to be a standalone platform, but a fundamental brick of the OPERAS Research Infrastructure (RI), complementing its tools and services; moreover, it is an SSH service of EOSC, and linked with the SSHOC project. Since the beginning of the project, existing research infrastructures for SSH (ERICs), namely DARIAH, CLARIN and CESSDA¹⁶, have been members of the project consortium. Moreover, the partnership with EGI, actively involved in the definition of the European Open Science Cloud (EOSC), ensures a full compliance of the platform with EOSC and its marketplace. More specifically, while planning GOTRIPLE's integration into the EOSC, we are working on two levels of integration.

The first level of integration concerns the integration of services in the EOSC Marketplace. This integration will be performed via the protocol for the integration of OPERAS RI services into the EOSC [3]. In the context of the EOSC, onboarding resources means that providers connect them with and through the EOSC, with a wide range of communities interested in knowledge production, and by following a protocol for integration. This protocol consists of a standardised administrative process aiming at collecting information about the resource in question and a validation process assessing the compliance of the provided information from service providers, thanks to a set of inclusivity criteria. Onboarding services to the EOSC will also be possible through regional or thematic portals and their marketplaces as well, but current and established work around on-boarding is centered around the central route, namely, the integration through the EOSC Portal Marketplace, which is currently under development in the EOSC Enhance project¹⁷. At a later stage, GOTRIPLE is also expected to be integrated within the SSHOC Marketplace, which is planned to be released in Autumn 2021. Within the framework of the EOSC Enhance project, this protocol will be extended to datasets and other scholarly resource types (such as training resources), while integration through the regional and thematic clouds will enable giving access to specialised, context-dependent knowledge structures, e.g. implementing language services in regional portals or detailed, domain-specific ontologies and vocabularies in the thematic portals. In the EOSC, data repository and data discovery services, like GOTRIPLE, are onboarded as services within the EOSC Marketplace for the benefit of SSH Researchers.

A second level integration concerns the integration of EOSC Core services, and can be achieved by implementing some of the EOSC Federation Services into our Services, such as the GOTRIPLE platform and innovative services. Such EOSC Federation Services include, but are not limited to, the Authorization and Authentication Infrastructure (AAI), the Helpdesk, the Accounting Service, the Monitoring Service, etc. The AAI and the Helpdesk are important additions to the GOTRIPLE platform.

Further activities of the TRIPLE project are focused on Open Science training. Planned to start at the time of writing this article, this activity will produce, adapt and reuse general and specific guidelines, as well as provide training mainly in the form of webinars - also linking to existing active projects which are already providing them such as OpenAIRE, FOSTER, DARIAH, CLARIN, and the SSHOC Training Community¹⁸ - to the project members, in order to ensure they all have competencies on Open Science practices and workflows. Webinars will tackle the different technical aspects of GOTRIPLE, covering the following topics: multilingualism and vocabularies, web annotation, FAIR data, certification, visualisation, API data, and data reuse.

4. CONCLUSION

In the process of development of GOTRIPLE we are facing four specific difficulties. The first one concerns GOTRIPLE's integration into the EOSC, while the latter is still under definition through a complex iterative process, which involves many different stakeholders. As an example, we can mention the on-boarding process. At this point of maturity of the EOSC Portal, the onboarding workflow is under continuous development which means that although

¹⁵ <https://www.operas.unito.it/special-interest-groups/multilingualism/>.

¹⁶ <https://www.dariah.eu/>; www.clarin.eu; www.cessda.eu.

¹⁷ <https://www.eosc-portal.eu/enhance>.

¹⁸ See: <https://www.openaire.eu/frontpage/webinars>; <https://www.openaire.eu/workshops>; <https://www.fosteropenscience.eu/>; <https://www.dariah.eu/activities/training-and-education/>; <https://www.clarin.eu/content/funding-call-participation-clarin-trainer-network-programme>; <https://sshopencloud.eu/training>.

drastic changes are not expected, certain aspects of the workflow and the currently used Service Description Template - which is work-in-progress - are expected to change.

The release of EOSC-related outputs by the six EOSC Working Groups is the result of a participatory process to which the TRIPLE's consortium - are asked to provide comments and feedback very frequently. In order to be able to monitor the production of the main documentation, a reporting methodology has been established, by assigning a responsible partner to each working group [4]. This process allows the team to have a constant update on relevant outcomes and the progress of different EOSC Working Groups. Another difficulty derives from the richness and variety of the EOSC ecosystem, which is composed of several projects, research infrastructures and initiatives and actions, where a hierarchy of authoritative sources is not always clear. We are addressing this problem thanks to the involvement in our consortium of partners such as EGI, CLARIN, CESSDA and DARIAH, whose different areas of expertise and perspectives, as well as their participation in relevant related projects, helps to understand the complexity of the scenario which composes the EOSC ecosystem. In this perspective, a strong collaboration with the SSHOC consortium is expected to develop common knowledge of EOSC requirements, to ensure the cohesion between different SSH services and even to consider deeper links between the services such as mixed governance models and business models.

The second main difficulty concerns how to deal with the pandemic situation. In order to deal with this problem, we are implementing some strategies, such as creating informal sessions where we have open discussions on various topics (TRIPLE Café) and publishing an internal monthly newsletter to inform the different partners about the progress of the project.

The third main difficulty is related to the ambition of the TRIPLE project: having a multilingual and interdisciplinary platform to highlight multiculturalism practices. To achieve this goal, important work has to be done on the alignment of vocabularies (for each discipline of the SSH - currently 27 following the MORESS categories and in nine languages). In itself, this task is a challenge which will contribute to the whole SSH community.

Last but not least, TRIPLE specificity is to harvest many small SSH data repositories in the different European countries in order to make them more visible and reusable by the scientific community. However, the practices of the repositories are very different, and there is no harmonisation yet between them on their data model. This means for the TRIPLE consortium to support the data providers and aggregators in their transition to a more common data model, the future TRIPLE data model.

5. ACKNOWLEDGEMENTS

The authors want to thank Yoann Moranville (DARIAH), Paula Forbes (Abertay University) and Mélanie Bunuel (Huma-num - CNRS) for their contributions to this paper. The project is funded by the European Commission, under Grant Agreement No. 863420.

REFERENCES

- [1] Dumouchel, Suzanne. «The EOSC as a knowledge marketplace: the example of ISIDORE: A virtuous data circle for users and providers». *EUDAT conference: Putting the EOSC vision into practice, Porto, Portugal (2018)*. <https://hal-amu.archives-ouvertes.fr/hal-01708784/> {hal-01708784}
- [2] Grassi, Marco, Morbidoni, Christian, Nucci, Michele, Fonda, Simone, Di Donato, Francesca. «Pundit: Creating, Exploring and Consuming Semantic Annotations». *Proceedings of the 3rd International Workshop on Semantic Digital Archives, SDA (2013)*
- [3] Deliverable 4.5, Protocol for the integration of OPERAS RI services into EOSC (OPERAS-P) (2020). 10.5281/zenodo.4005677
- [4] Deliverable 6.1, Report on the General Interoperability Requirements (TRIPLE) (30-09-2020).
- [5] Dumouchel, Suzanne, Blotière, Emilie, Breiffuss, Gert, Chen, Yin, Di Donato, Francesca, Eskevich, Maria, Forbes, Paula, Georgiadis, Harris, Gingold, Arnaud, Gorgaini, Elisa, Moranville, Yoann, Pohle, Stefanie, de Paoli, Stefano, Petitfils, Clara, Toth-Czifra, Erzsébet, «GOTRIPLE: A User-Centric Process to Develop a Discovery Platform». *Information 2020*, 11, 563. doi:10.3390/info11120563