

THE FAIR PRINCIPLES AND THE EOSC CONCEPT IN THE RESEARCH COMMUNITY OF DIGITAL HUMANITIES, LANGUAGE STUDIES AND CULTURAL HERITAGE

AN EXPEDITIONARY SURVEY



PARTHENOS

Pooling Activities, Resources and Tools
for Heritage E-research Networking,
Optimization and Synergies

This report is distributed under a CC – BY 4.0 license.

The financial support of the European Commission under the H2020 grant 654119 is duly acknowledged.

Introduction: the survey

Performing a survey on the FAIR principles and the EOSC within the research community of the domain of digital humanities, language studies and cultural heritage (DH+L+CH) – the remit of the PARTHENOS project – was motivated by the need of taking the pulse of these researchers about the EOSC, its development and purpose, in the light of the statement made in the EU EOSC Declaration that EOSC “services and functionalities shall be user-driven”.

The concept was to rapidly get the researchers’ feedback on very basic issues, such as their knowledge and appreciation about the FAIR principles and the EOSC, and get indications on urgent actions as well as on more precise ways of collecting their perspective and expectations. To the best of our knowledge, it is the first time that researchers in this wide domain are asked to give their opinion.

Two actions have been undertaken.

The first one consists in a questionnaire sent to some mailing lists totalling several hundred researchers, most of which are involved in projects of Digital Humanities and Digital Heritage. Such lists included, but were not limited to, the internal PARTHENOS and ARIADNE mailing lists.

Addressees were asked to distribute the survey to any other list to maximize the dissemination effect: as it is well known, questionnaires are important not only for the feedback but also for raising awareness about the issue surveyed among the perspective respondents. Since the sample did not aim at statistical significance, this approach does not seem to bias the survey. On the contrary, since the target of the survey was mainly formed by researchers involved in projects with a strong digital component, i.e. a “learned” minority, it may be expected that negative aspects such as limited knowledge or pessimism about the EOSC impact are even more serious among the wider population of all researchers. In other words, if such aspects are evidenced in the survey, it is just the tip of the iceberg.

To date¹ some 100 answers (approximately 15% of the total) were received. Thus, as already mentioned, the sample is *not* representative of the whole research community: it indicatively represents, instead, those actively involved in digital research projects. The geographic distribution of respondents shows some imbalance, in part because the issues surveyed are perceived as more important in some countries, either because of the maturity of the research community or for the perceived absence of political actions on a subject that, instead, researchers feel as important. However, the questionnaire results already provide important information, and suggest to plan a more detailed data collection.

¹ The present report is based on the answers received by 16/11/2017. The survey started on 25/10/2017.

The second action was a workshop titled “Focus on the EOSC” held on 8 November 2017 by the PARTHENOS project, in the occasion of another project meeting. Participants were asked to provide their opinion about three groups of issues:

- Empowerment: the role of the DH+L+CH community in building the EOSC
- EOSC services: what should we do to obtain the implementation in EOSC of the really useful ones
- EOSC architecture: which requirements do we state for the EOSC architecture

The participants grouped in break-out sessions on the above themes. The reports are in preparation, and from the preliminary outcomes organizing a more comprehensive workshop already seems to be very useful.

Such activities were carried out in the framework of the PARTHENOS (www.parthenos-project.eu), an EU-funded H2020 cluster project, by a team led by the project coordinator Franco Niccolucci, with the active collaboration of WP2 “Community involvement and requirements”, WP6 “Services and tool” and WP7 “Skills, professional development and advancement”. Other colleagues substantially helped in distributing the questionnaire to different email lists.

1. The questionnaire

The questionnaire was created as an online Google form, including:

- a first optional question (#1) about the respondent name and email: a large majority provided their personal information, while others preferred to remain anonymous;
- a question (#2) about the country where they worked;
- 8 closed-answer mandatory questions (#3 to #10)
- 5 open-answer optional questions (#11 to #15) on the EOSC

Respondents had the opportunity to skip the last 5 questions if they felt they had no contribution to give. The mandatory part of the questionnaire required a few minutes to complete. The results are examined below.

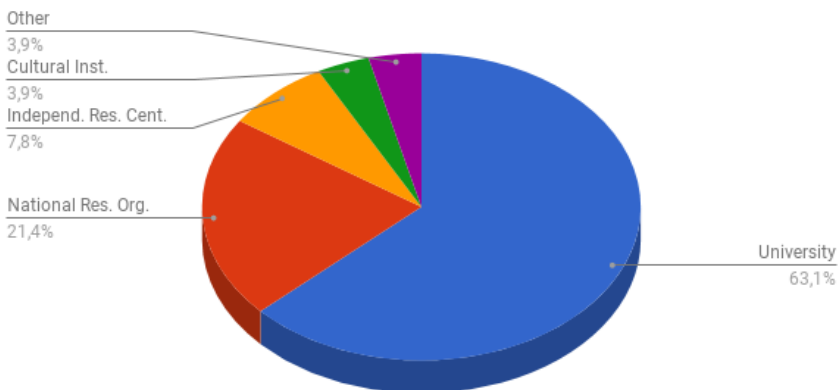
2. General questions on respondents (Questions 3-4)

Question 3. Where do you work?

- University*
- Department of a National Research Centre (e.g. CNR, CNRS, National Academy, etc.)*
- Independent Research Centre (public or private)*
- Cultural Institution (Library, Museum, Archive, etc.)*
- Other: _____*

About two thirds of the respondents work in a University; one-fifth works in a national research centre.

Affiliation



Question 4. In which domain is your main research interest?

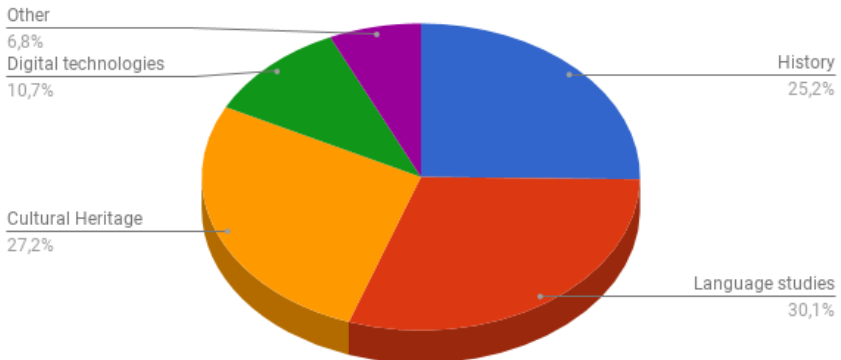
The domain is defined according to the PARTHENOS definition in D2.1. They include also the application of digital technologies to the domain. Please indicate the most relevant subject.

- History (in a broad sense: including Medieval Studies, Recent History, Art History, Epigraphy, etc.)
- Language Studies (including Literature, Linguistics, Philology, Language Technology, etc.)
- Archaeology, Heritage & Applied Disciplines (including Cultural Heritage, Museums, Preservation / Conservation, etc.)
- Digital technologies
- Other: _____

The domain definition is taken from the PARTHENOS report “Report on users requirements”, which defines the groupings listed above. It was assumed that the explanation summarily given in the form was clear enough also for respondents not knowing the PARTHENOS report.

The distribution² is rather balanced among Language studies (32%), History (22%) and Cultural Heritage (27%). The others mainly work on digital technologies (11%), with a few not belonging to (or not recognising themselves in) any of the PARTHENOS categories.

Research domain



² Note that the percentages in the diagrams are automatically calculated by the software by rounding the figures, and thus they sometime may total to 99.9% or 100.1%. Such a rounding error is irrelevant for our purposes.

3. Importance of computer-based work (Questions 5-6)

Questions #5 and #6 concern the respondents' perception of the importance of digital technologies in their work.

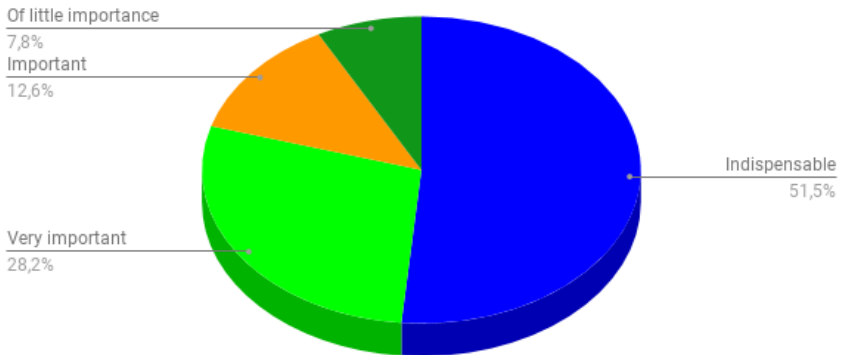
Question #5 aims at identifying the characteristics of the "digital learnedness" of the sample: only a small fraction does not acknowledges to be a "digital researcher".

Question 5. How important is computer-based work for your research?

- Indispensable*
- Very important*
- Important*
- Of little importance*
- Of no importance*

About 80% state it is absolutely or very important; only for less than 10% it is of little importance. This confirms the hypothesis that the researchers interviewed belong to the category of "digital researchers".

Importance of computer-based activity for own research



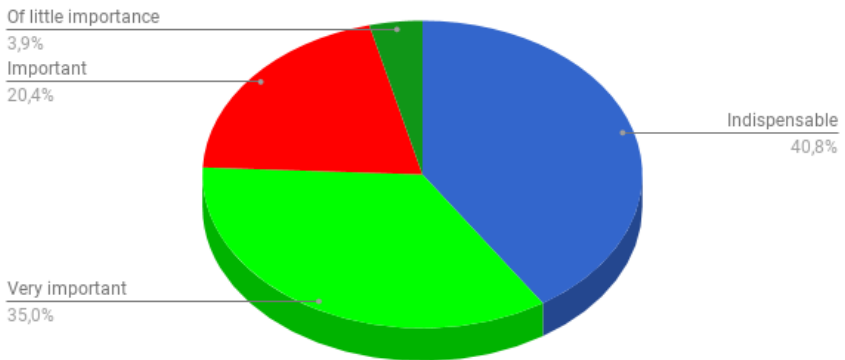
Question #6 concerns instead the researchers' appreciation of the importance of digital technologies in their research domain, regardless of their individual activity.

Question 6. How important is computer-based work for research in your domain?

- Indispensable*
- Very important*
- Important*
- Of little importance*
- Of no importance*

A large majority (77%) believes that computers are indispensable or very important for their domain, with almost all the others qualifying it as important. Since the scale of importance may be a little fuzzy, we may state that the importance of computer-based research is, not surprisingly, acknowledged by about 90% of the answers.

Importance of computer-based activity for domain research



It is interesting to note that by inspecting individual answers, it results that the few ones who indicate digital technologies as being of little importance in their domain also declare themselves as not being computer-skilled; on the contrary, some researchers who do not consider themselves as computer users deem, nonetheless, such technologies as very important for their research domain

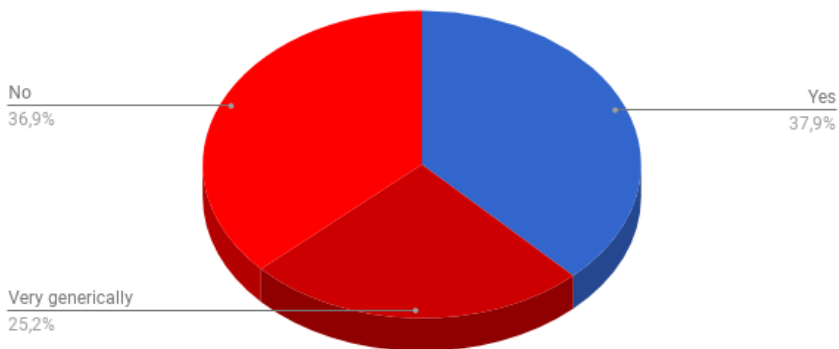
4. The FAIR principles (Question 7)

Question 7. Do you know what are the FAIR principles?

- Yes
- Yes, but in a very generic way
- No

The FAIR principles are relatively little known. Only about one third states they know them, while another third declares to know them only in a generic way, and the remaining third has no information about them.

Knowledge about FAIR principles



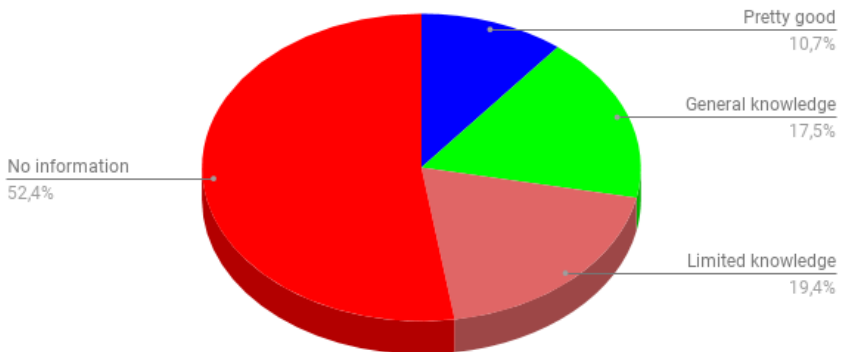
5. The EOOSC concept (Questions 8-9)

Question 8. How well do you know the EOOSC concept?

- Pretty well*
- I have some general knowledge of it*
- I have only a limited knowledge of it*
- I have no information at all*

Only little more than 10% knows about the concept. An additional 20% has some information, while the remaining 70% has very little or no information at all (50%). Comparing such results with those of the previous question, it emerges the existence of a group of about one third of the community of the total which is reasonably informed, while the remaining two thirds are unaware. The FAIR principles are more popular than the EOOSC.

Knowledge about EOOSC

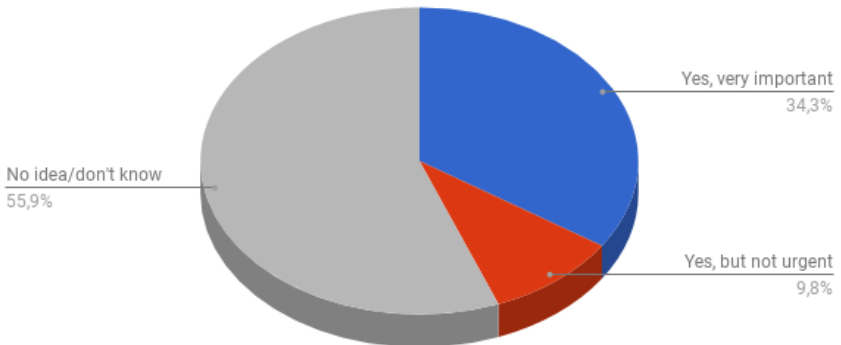


Question 9. Do you think that the EOSC concept should be better known by researchers?

- Yes, it is of primary importance that researchers know better the EOSC and how it will impact on their work
- Yes, but it is not an urgent matter, a general information on the EOSC is probably sufficient
- I have no precise idea about this issue
- I think it is not so important that researchers know about EOSC as their focus is on research
- I think that EOSC is irrelevant to improve the quality of research in my domain
- I do not know

It must be noted that nobody chose the very negative answers “It is not so important to know about EOSC” or “EOSC is irrelevant”; only 9% think that improving knowledge about the EOSC is not an urgent matter. The majority simply stated “no idea”.

Importance of knowing more about EOSC



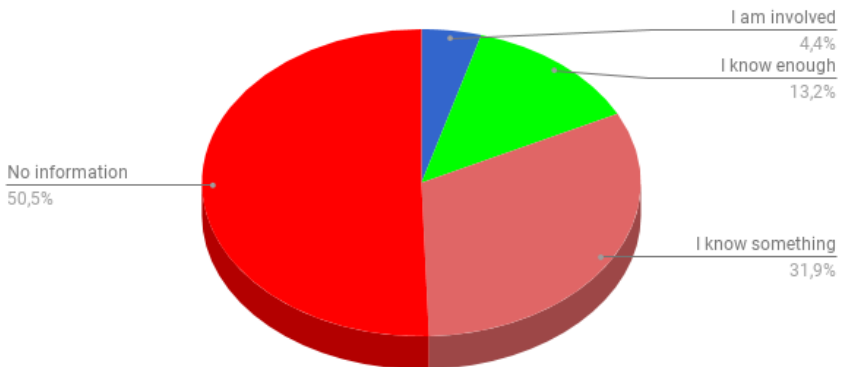
6. The EOSeopilot project (Question 10)

Question 10. Do you know about the EOSeopilot EU project?

- I am involved in it
- I am not involved but know enough about this project I have heard something about it
- I have no information about this project

Among the respondents, there are 6 people involved in EOSeopilot in various ways. There are also a few outliers who don't know about the EOSe, but claim to know about EOSeopilot. Apart from that, the popularity of EOSeopilot follows the one of EOSe, with similar, not encouraging, percentages; but almost all those who have an idea of the EOSe concept also declare to know the EOSeopilot project.

Knowledge about EOSeopilot

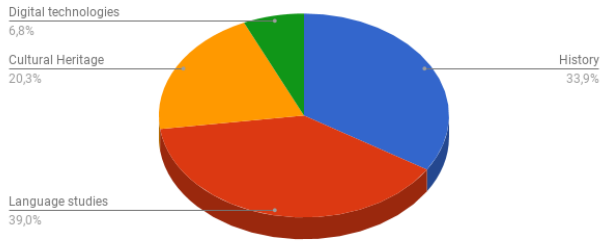


7. FAIR and EOSe knowledge by domain

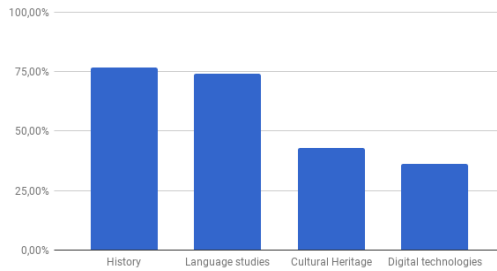
The cases of limited knowledge of these two concepts was analysed by domain and the results are shown below. The pie charts refer to the domain percentage on the total of respondents with limited or no knowledge; the bar chart shows instead the percentage of researchers with little or no knowledge on the total of respondents, by domain: for example, of all the respondents having limited knowledge of the FAIR principles, 19% belong to the domain of CH (pie chart); of all the respondents from the CH domain, 26% states to have limited or no knowledge of the FAIR principles (bar chart).

The proportion of the uninformed is similar on both topics, with a surprising presence of researchers in the field of digital technologies. There is, in general, a slightly better awareness for both topics among CH researchers.

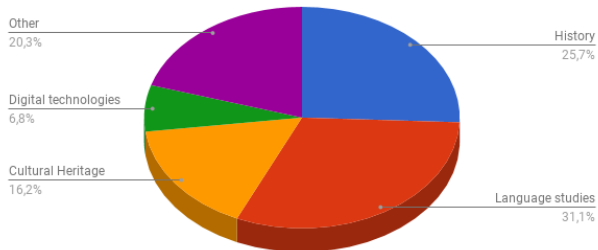
Domain percentage of researchers with limited knowledge of FAIR principles



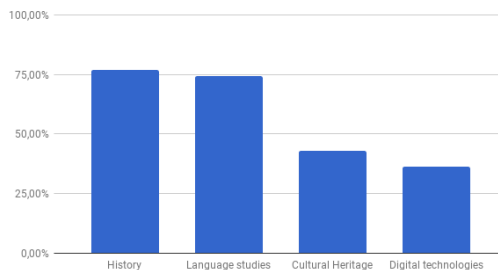
Limited knowledge of FAIR principles in each domain



Domain percentage of researchers with limited knowledge about EOSC



Limited knowledge of EOSC within each domain



8. Answers to open questions

As already mentioned, answering to these questions was optional, so about 30% of the respondents gave a significant answer here (i.e. excluding “no idea” and the like): in practice, who responded to one question also responded to the others. The questions are the following.

Question 11. What are the services that your research collaboration would benefit from as consumer of EOSC?

About one third of the answers indicated storage as the main service, with some adding data access and sharing. Searching is also a popular answer. Some mention Open Publication. Only few mention other services, in most cases mentioned generically as “collaboration services”, in others summarily described. On the negative side, three answer state that EOSC would be of no use, because local or national infrastructures already do the job.

Question 12. Do you envisage the need of shared services that EOSC could provide with central coordination, to take advantage of economies of scale and ensure harmonization across different providers?

Most of the answers reply with a simple “Yes”. A few give some hints on why such services are needed, such as “cross-repository data access”, “standardization”, “cost reduction” and so on. One is skeptical because is afraid of the risk that rules bureaucratically imposed from top limit the freedom of choosing the most appropriate processing tools. A couple of answers indicate a preference for centralized, but domain-related services (and repositories). On the pessimist side, there is concern about privacy, fit to domain needs, or generic fear that the drawback is bigger than the benefits.

Question 13. What criteria should the EOSC services meet in order to be appealing to your research collaboration?

Here there is more diversity in answers. In general fit to user needs is a common feature of many of them, with one advocating “user-push” vs “technology-push”. Usability and ease of use is also a common point. Others mention data openness, standardization, interoperability, data quality and trustworthiness, scientific recognition to data creators, and easy access. One respondent describes here (probably because it was the most appropriate place to do so) the difficulty of operating without national political support, what for example makes a general adoption of the FAIR principles a difficult objective. He requires EOSC to advocate this good cause at the political level, perhaps meaning the mainstreaming of its principles at all levels, although it is unclear who should be in charge of this.

Question 14. What services and resources (e.g. data, publications, software and other digital artefacts) could you provide for sharing, third party access and use through the EOSC?

Perhaps unexpectedly, most of the respondents offer tools and other resources of various kind, but all potentially significant, implicitly suggesting that a super-service to support the migration of such resources to the EOSC framework could be relevant.

Question 15. How do you see your role in fostering the uptake of the FAIR principles in your environment?

Answers to this question show a great willingness to advocate the FAIR principles. People state that they avail of their role or position to foster their adoption. A suggested, interesting way of doing so concerns education and training. Some answers state that the adoption of the principles by the respondent is an exemplary practice that others may be induced to follow. On the contrary, a couple of respondents highlight the issues deriving from the lack of political support, what prevents them to be able to foster the uptake of the FAIR principles.

Overall, the fact that about 30% of the respondents tried to give concrete suggestions, although sometimes generic and imprecise, shows that there is an active expectation for the EOSC and people is convinced about the FAIR principles and ready to advocate them.

9. Conclusions and suggested actions

Even from qualitative results such as those resulting from this survey, it is possible to get some insights about the current attitude of the DH+L+CH research community towards the EOSC and the actions required to improve it.

The survey results may be summarized as follows:

- The respondents are committed to computer-based research; about 80% of them state that digital technology is very important for their work and their community;
- In the answers to the open questions, a majority sees in the EOSC a mean to foster access and data sharing, also acknowledging that this requires interoperability and standardization;
- Indications on services are more generic, with a clear preference for search-and-find, and data post-processing and re-processing;
- The FAIR principles and the EOSC are concepts still unclear to many.

Thus, a first set of actions required, and viable in the short term, should aim at raising the awareness of the research community about the formalization of what is already perceived as a generic need. This could be accomplished by setting up a series of thematic workshops introducing the FAIR principles and the EOSC through domain-related use-cases. Their preparation should build on the knowledgeable ones who should set up demos of how existing services may be improved in a cloud environment. They should also include demonstrations of the currently available cloud services proposed by different technological actors.

A second set of actions, also possible with the available resources and in a reasonably short time frame, should examine and propose training and educational solutions to improve the knowledge of the community, for example about topics such as data stewardship from the domain perspective, in preparation of educational programs on such topics.

Such actions should ultimately lead to the definition of user requirements in a bottom-up way, which is still difficult due to the gaps evidenced by the survey, and therefore is still delegated to a few experts trying to interpret, in earnest, the needs of their reference community.

On the other hand, such actions should prepare the evaluation of currently available tools in the light of EOSC and the definition of new ones to be implemented in the cloud framework. The main action, on this regard, should include the definition of a pipeline to bring already available services in a cloud environment, as for example it is being done with the EOSCpilot Science Demonstrators. Creating a standard procedure for such tasks could greatly improve the effectiveness of the job.

Finally, complete and exemplary use cases should be developed in the cloud environment. This would serve as convincing demonstrators towards researchers, and stimulate the implementation of other workflows, typical of the research activities of the DH+L+CH community.

Contacts

The PARTHENOS project: www.parthenos-project.eu

Project coordinator: franco.niccolucci@pin.unifi.it