



Podcast Code for Thought*:

Research Software and Research Data in Open Science

Teresa Gomez-Diaz¹ and Tomas Recio²

¹University Gustave Eiffel-LIGM-CNRS, Est of Paris

²University Antonio de Nebrija, Madrid

Contact: Teresa.Gomez-Diaz@univ-eiffel.fr, trecio@nebrija.es

July 18th 2023

This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License

(CC-BY-SA 4.0), see <http://creativecommons.org/licenses/by-sa/4.0/>.

It contains two documents licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0), see <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Abstract

The goal of this text is to write down and extend the questions and answers that have been prepared for the Podcast Code for Thought registered in May 2023 at the Laboratoire d'informatique Gaspard-Monge (LIGM), located at the Gustave Eiffel University, at the Est of Paris. This podcast goal was to have an informal conversation about the recent published work on Research Software and Research Data in the Open Science context by T. Gomez-Diaz and T. Recio. The interview was conducted by Peter Schmidt, who provided the questions we do answer here, and that were answered by T. Gomez-Diaz in the recorded podcast. The podcast, entitled *Open Data, Open Software - with Teresa Gomez-Diaz*, is now published¹.

Keywords. Research Software, Research Data, Open Science, Research Evaluation, FAIR

1 Foreword

Code for Thought² is a podcast series launched by Peter Schmidt³ in January 2021⁴. The goals of Code for Thought podcasts are to support, explore and promote the growing community of Research Software Engineers (RSE) in the UK and abroad, and it is realized in close collaboration with the Software Sustainability Institute⁵ and the Society for Research Software Engineering⁶ as well as with other RSE communities outside the UK.

*Permission is granted by Peter Schmidt for the inclusion of the Code for Thought logo in this document.

1. <https://codeforthought.buzzsprout.com/1326658/13216530-en-open-data-open-software-with-teresa-gomez-diaz>

2. <https://codeforthought.buzzsprout.com>

3. <https://www.software.ac.uk/about/fellows/peter-schmidt>

4. The first Code for Thought podcast *And so it begins...* is available at <https://codeforthought.buzzsprout.com/1326658/7121884-and-so-it-begins>.

5. <https://www.software.ac.uk/>

6. <https://society-rse.org/>

Research Software and Research Data: dissemination, evaluation and reusability in the Open Science context

Teresa Gomez-Diaz, teresa.gomez-diaz@univ-eiffel.fr
 Université Gustave Eiffel-LIGM-CNRS, Est of Paris
<https://siteigen.univ-mv.fr/>

Tomas Recio, treccio@nebrija.es
 Universidad Antonio de Nebrija, Madrid
<https://www.nebrija.com/>



17th edition of the
 International Digital Curation Conference (IDCC22)
 13-16 June 2022

Our contribution: how to improve REUSABILITY conditions

The extended use of CDUR evaluation protocols for Research Data and Research Software will enhance the quality of the disseminated research outputs and will improve their reusability conditions. CDUR is one of the enablers to achieve Open Science objectives.

The Open Science Context

Open Science is the political and legal framework where research outputs are shared and disseminated in order to be rendered visible, accessible and reusable (Gomez-Diaz & Recio, 2020-21). Other definitions for Open Science available at (Méndez, 2021, UNESCO, 2021, Vicente-Saez et al., 2018).

Three steps: RS and RD definitions, dissemination procedures, CDUR evaluation protocols

I - Definitions

Research software (RS) is a well identified set of code that has been written by a (again, well identified) research team. It is software that has been built and used to produce a result published or disseminated in some article or scientific contribution. Can include: documentation, specifications, use examples... (Gomez-Diaz & Recio, 2019).

Research Data (RD) is a well identified set of data that has been produced (collected, processed, analyzed, shared and disseminated) by a (again, well identified) research team. The data has been collected, processed and analyzed to produce a result published or disseminated in some article or scientific contribution. Can include: documentation, use examples, provenance information, instrument information... and references to the software needed for the RD manipulation. (Gomez-Diaz & Recio, 2022).

References

- Gomez-Diaz T. (2014). Free software, Open source software, licenses. A short presentation including a procedure for research software and data dissemination. Zenodo
- Gomez-Diaz T. & Recio T. (2019). On the evaluation of research software: the CDUR procedure. F1000Research
- Gomez-Diaz T. & Recio T. (2020-21). Towards an Open Science definition as a political and legal framework on the sharing and dissemination of research outputs. Version 2, PLOS N. 19, pp. 3-25, 2020. Version 3, 28/02/2021. Zenodo.
- Gomez-Diaz T. & Recio T. (2022). Research Software vs. Research Data I: Towards a Research Data definition in the Open Science context. F1000Research
- Gomez-Diaz T. & Recio T. (2022). Research Software vs. Research Data II: Protocols for Research Data dissemination and evaluation in the Open Science context. F1000Research
- Méndez E. (2021). Open Science por defecto. La nueva normalidad para la investigación. ARBOR Ciencia, Pensamiento y Cultura.
- UNESCO (2021). UNESCO Recommendation on Open Science. SC/PC/SPP/2021/106.UROS.
- Vicente-Saez R., & Martínez-Fuentes C. (2018). Open Science now: A systematic literature review for an integrated definition. J. Business Research, V. 88.

This work is partially funded by the CNRS-IEA PREOSI project (2021-22).

II - Dissemination Procedures

- Choose a name, avoid trademarks and proprietary names; associate date, version...
- (*) Research team. Establish list of authors, contributors, participants to produce the RS or RD. Consider legal issues related to copyright, and other RD legal issues.
- (*) Establish the list of included software & data components, their licenses.
- Choose a license, have an agreement (signed) with rightholders and authors. Beware of licence compatibility and inheritance issues.
- Choose a website, large, or deposit for dissemination, indicate licences and how to cite the work. Use PIDs if possible.
- (*) Research work. Establish the list of the main functionalities to facilitate REUSE.
- Inform your laboratories and head institutions (if not done in the license step).
- Set and indicate clearly your contact address.
- Distribute the software or data component. Inform the target community.

(*) To review for each new RS version.
 (Gomez-Diaz, 2014, Gomez-Diaz & Recio II, 2022).

III - CDUR Evaluation Protocols

The CDUR protocol includes four steps:

(C) **Citation**, measure if the RS or RD is well identified as a research output: good citation form, but also metadata, best citation practices...
 Legal point: authors, contributors, affiliations, copyrights...
 Policy point: Open Science. Legal point: licenses

(D) **Dissemination**, best dissemination practices, in agreement with the scientific policy of the evaluation context
 Policy point: Open Science. Legal point: licenses

(U) **Use**, "software" or "data" aspects of RS or RD: correct results, facilitate reuse, best software or data practices, can include documentation, use cases, test, install...
 Reproducibility point: validation of scientific results, REUSABILITY

(R) **Research**, "research aspects": quality of the scientific work, proposed and coded algorithms & data structures, related publications, collaborations...
 Research point: measures research impact

Flexibility of application: each decision maker or evaluation committee sets its own CDUR protocol adapted to the evaluation context and goals.
 (Gomez-Diaz & Recio, 2019, Gomez-Diaz & Recio II, 2022).



FIGURE 1 – Research Software and Research Data: dissemination, evaluation and reusability in the Open Science context, IDCC22. Work licensed under the CC-BY-NC-ND Creative Commons License.

This particular podcast is dedicated to recent work by T. Gomez-Diaz and T. Recio on Research Software and Research Data in the Open Science context that was published in F1000Research⁷ last year [8, 9].

The poster [10] entitled *Research Software and Research Data: dissemination, evaluation and reusability in the Open Science context* summarizes these two articles and further enhances the importance of the work presented there in the context of reusability of these research outputs. It was presented at the 17th edition of the International Digital Curation Conference (IDCC22) that took place in June 2022⁸. This poster is included in Figure 1.

The questions to prepare the podcast were sent by Peter Schmidt one week before the recording was done and the answers were started to be arranged in a draft document which became a first version of the present document.

The above mentioned publications on Research Data [8, 9, 10] do follow our work on Research Software [5] and on Open Science [6, 7] that was also the object of another Code for Thought podcast: *Open Science and Research Software*⁹. It was also recorded in May 2022 at the Laboratoire d'informatique Gaspard-Monge and released in July 2022. Furthermore, the poster [7] entitled *The future of Open Science asks for a common understanding* was presented at the EGI Virtual Conference 2021¹⁰ and it is included in Figure 2.

In order to complete the Open Science framework in which we place our work, we would like to mention here the UNESCO Recommendation on Open Science adopted in November 2021 [12].

7. <https://f1000research.com/>
 8. <https://www.dcc.ac.uk/events/idcc22>
 9. <https://codeforthought.buzzsprout.com/1326658/10822132-open-science-and-research-software>
 10. <https://indico.egi.eu/event/5464/>

The future of Open Science asks for a common understanding

Teresa Gomez-Diaz, CNRS/LIGM, Est of Paris
Tomas Recio, University Nebrija, Madrid

A definition is missing

Definition of Open Science remains vague, probably because objectives are vague... (Aspesi, COASP 2014)

...there is no single, accepted, unified definition or vision of 'open science...' (Fennell, 2018)

...there is a lack of awareness about what Open Science is, mainly due to the fact that there is no formal definition of Open Science... (Vicente-Saez & Martinez-Fuentes, 2018)

Recent, inclusive & complete visions

(Méndez 2021, Image: Méndez 2017, CC-BY)

For the purpose of this Recommendation, Open Science is defined as an inclusive construct [...] aiming to make scientific knowledge openly available, accessible and reusable for everyone. [...] It includes all scientific disciplines and aspects of scholarly practices, [...] it builds on the following key pillars: open access to scientific knowledge, open science infrastructures... (UNESCO, 2021)

Our contribution

Goal

To contribute to the adoption of a common, unified vision.

Definition proposal

Open Science is the political and legal framework where research outputs are shared and disseminated in order to be rendered visible, accessible and reusable (Gomez-Diaz & Recio, 2020-21).

Three steps supporting this proposal

- I - Three selected pillars for a common understanding
- II - Towards a political and legal framework
- III - Enablers: three cornerstones to get to a working framework

I - Three selected pillars for a common understanding

Free Software Foundation, 1985

The free software definition presents the criteria for whether a particular software program qualifies as free software: the freedom to run the program as you wish, the freedom to study how the program works, and change it... = importance of licenses

Budapest Open Access Initiative, 2002

By "open access" to [open-reviewed] research literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link...

Committee on Data for Sciences and Technology, 1966

CODATA est un Comité au niveau scientifique international le plus élevé [...] à cause de l'importance qui s'attache à l'évaluation des données...

II - Towards a political and legal framework

...sets out the action that intends to improve access to scientific information and to boost the benefits of public investment in research. [...] To improve access to scientific information, Member States, research funding bodies, researchers, scientific publishers, Universities and their libraries, innovative industries, and society at large need to work together [...] so that the "fifth freedom" of the EU – the free circulation of knowledge – can become a reality.

Policy

BOE N. 131 de 2 de junio de 2011, Artículo 37

- Los agentes públicos impulsarán el desarrollo de repositorios de acceso abierto a las publicaciones...
- Los investigadores cuya actividad haya sido financiada mayoritariamente con los Presupuestos Generales del Estado están obligados a publicar en acceso abierto una versión electrónica de los contenidos aceptados para publicación...
- La versión electrónica pública podrá ser empleada en los procesos de evaluación.

Law

LOI du 7 octobre 2016 pour une République numérique, Article 30

- Lorsqu'un écrit scientifique issu d'une activité de recherche financée au moins pour moitié par des dotations publiques est publié, son auteur dispose, même après avoir accordé des droits exclusifs à un éditeur, du droit de mettre à disposition gratuitement...
- Dès lors que les données issues d'une activité de recherche financée au moins pour moitié par des dotations publiques [...] sont des données publiques, leur diffusion est libre.

Plan national pour la Science Ouverte, Juillet 2018 & 2ième Plan, juillet 2021

- Rendre obligatoire la publication en accès ouvert des articles et livres issus de recherches financées par appels d'offres sur fonds publics.
- Rendre obligatoire la diffusion ouverte des données de recherche issues de programmes financés par appels à projets sur fonds publics.

Policy

Other Member States, see: Open Science overview in Europe, OpenAIRE, <https://www.openaire.eu/en-eu-countries>

III - Enablers: three cornerstones to get to a working framework

Institutional policies: the required evolution of policies of Universities and RPOs.

Infrastructures: the development of Open Science-oriented infrastructures and services.

Research evaluation: the transformation of evaluation policies and practices.

Policy

ESI Virtual Conference 2021
19-21 October 2021

Some references: Aspesi C. (2014). A financial analyst's perspective on open access, COASP 2014. European Commission (2012). Towards better access to scientific information: Boosting the benefits of public investments in research, COM(2012) 401 final, European Commission, Brussels. European Commission (2014). The state of the art in peer review, E405 Microbiology Letters, Volume 350, Issue 19, May 2014. Margam E, Caser R, Ducato R, Guarà P & Moscon V. (2018). Open Access, Open Science, Open Society, University of Trento. Méndez J. P. (2021). Open Science per definition. La nuova normalità per la investigación. IADIR Ciencia, Humanidades y Cultura. UNESCO (2021). Final report on the draft text of the UNESCO Recommendation on Open Science. CU-0498. Vicente-Saez R., & Martínez-Fuentes C. (2018). Open Science reap: A systematic literature review for an integrated definition. J. Business Research, V. 88. More information and references at Gomez-Diaz & Recio (2020-21). This work is partially funded by the CNRS-REA PNEOSI project (2021-22). 888 989 Paris is licensed under the CC-BY-NC-ND Creative Commons License.

FIGURE 2 – The future of Open Science asks for a common understanding, EGI Virtual Conference 2021. Work licensed under the CC-BY-NC-ND Creative Commons License.

2 Questions and Answers

Q1. The introduction to your paper mentions Borgman's Conundrum. What is the Conundrum, and how does it relate to your paper?

A1. Christine Borgman explained in her article entitled *The conundrum of sharing research data* [1] that to share research data is a difficult problem and asked some fundamental questions: *which Research Data might be shared, by whom, with whom, under what conditions, why, and to what effects.*

In the recently published article [8] we have proposed a *Research Data (RD)* definition and a framework in which we can provide answers to the Borgman's Conundrum questions. The formulation of this RD definition finds a similar formulation that the one also proposed for *Research Software (RS)* in [5], in 2019.

Q2. You are linking Research Software and Research Data. Could you explain the link - i.e. how they are connected.

A2. One link is that both productions, RS and RD, find similar problems in the scientific context. For example there is not an accepted way to publish them, that is, a publication system that is so generally and widely accepted as the one that exists for articles and scientific publications since longtime.

Another important point is that researchers are evaluated only and mainly regarding their publications but still without taking into consideration the RD or RS production. This is changing a lot nowadays, but nevertheless it will take time to arrive to a system so widely adopted and accepted as the one existing for articles.

What we show in these two articles [8, 9] and in the poster [7] (see Figure 1) is that the framework in which we define RS and propose the CDUR evaluation protocols to give value to this production [5], this framework is also valid for RD. Which means that we can propose a similar formulation for the RD and the RS definitions, and similar dissemination [4, 9] and CDUR evaluation protocols [5, 9].

Q3. What are the difficulties to define Research Data

A3. Well, the problem is not to define RD, but to define what data is. Data is a complex object from many points of view: what is data, what is information, what is evidence... It is also complex from the scientific point of view, as data manipulated by a chemist can have little to do with data manipulated by a sociologist.

But the difficulties are also from the legal point of view. One of the points that have helped us is that we do not need to define software. In our RS work [5], our RS definition stands over a software definition that is provided by the legal framework of author rights and copyright (as defined in France, Spain and by the EC). But it has taken us a lot of time to understand that we have the same schema for RD although for different reasons. We do not try to define data, but its legal context is not the same as in software. Software legal context is more or less easy and regards mainly copyright and licenses, and data is far too much complex. In fact legal context of data to study temperature in a region for a year usually does not raise many legal issues. But data can have associated author rights, or maybe not, and there are also other legal issues like personal data, medical data, *sui generis* database rights...

Yet, we can provide a RD definition with similar formulation than the RS definition [8] if we forget about how to define data. Sometimes you need mathematicians to understand and deal correctly with what means a definition and which is the right way to propose one.

Q4. What are the problems you see with current definitions of Research Data. For instance the definition by UKRI

A4. One of the problems is, as far as we understand, that these definitions try to say what data is. But data is a too much complex object, that *should be studied in a case-by-case basis* [3], and this is not only from the legal point of view, but also from the scientific point of view, taking into consideration a specific scientific community, a concrete problem. We have provided examples in [8]: temperature related data, the STRENDA Guidelines for enzyme activities, the linguistic research work at LIGM, what is environment information as defined by the French *Code de l'environnement*.

The Research Data definition coming from the Concordat on Open Research Data that is signed by the research councils of the UK Research and Innovation (UKRI) organisation [2] is the following one:

Research data are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical). These might be quantitative information or qualitative statements collected by researchers in the course of their work by experimentation, observation, modelling, interview or other methods, or information derived from existing evidence. Data may be raw or primary (e.g. direct from measurement or collection) or derived from primary data for subsequent analysis or interpretation (e.g. cleaned up or as an extract from a larger data set), or derived from existing sources where the rights may be held by others.

As you can see, there is a lot of information in the above text that is related to what data is. A good point of this definition is that the goal is research. But how to provide answers to the Borgman's Conundrum questions? Who shares the data? Which data?...

Q5. What is your definition of Research Data?

A5. The RD definition proposed in [8] is the following:

Research Data is a well identified set of data that has been produced (collected, processed, analyzed, shared & disseminated) by a (again, well identified) research team. The data has been collected, processed and analyzed to produce a result published or disseminated in some article or scientific contribution. Each RD encloses a set (of files) that contains the dataset maybe organized as a database, and it can also include other elements as the documentation, specifications, use cases, and any other useful material as provenance information, instrument information, etc. It can include the research software that has been developed to manipulate the dataset (from short scripts to research software of larger size) or give the references to the software that is necessary to manipulate the data (developed or not in an academic context).

We can summarize the above definition in the following three main characteristics:

- the goal of the collection and analysis is to do research, that is, to answer a scientific question (which includes the validation of research findings),
- it has been produced by a research team,
- the RD is involved in the obtention of the results presented in scientific articles (as the most important means for scientific exchange are still articles published in scientific journals) or by any other kind of recognized scientific means.

This definition tackles then with three points: which is the research object (RD), who is the research team and which is the RD related research work to answer a research question. In this context it is easier to provide answers to the Borgman's Conundrum questions, as presented in the Conclusions section of [8].

Q6. Your second paper discusses protocols that can be used to make sure research software and data adhere to FAIR principles. It mentions the CDUR principles. What are they?

A6. The question is how to give value to the RD and/or the RS production.

The CDUR evaluation protocol has been initially formulated for RS in [5] and has been extended to RD in [9]. It has been designed to help evaluators, members of evaluation committees, and evaluated researchers, members of the research teams responsible of the RS or RD.

There are four steps in the proposed evaluation protocol, which is flexible enough to be applied in different evaluation contexts:

(C) Citation. This step measures if the RS or RD are well identified as a research output, i.e. if there is a good citation form, or good metadata. We look here to best citation practices.

This is a legal related point where we ask for authors (if any) are well identified, which are their affiliations, and for example the % of their participation in software writing.

(D) Dissemination. In this point we look to best dissemination practices, in agreement with the scientific policy of the evaluation context. The dissemination of RS and RD needs a license to set the sharing conditions. For RD there are maybe further legal issues to look at (personal data, *sui generis* database rights...).

This is a policy point in which we look at Open Science requirements.

(U) Use. This point examines “software” or “data” aspects, in particular the correct results that have been obtained, and we can also look if their reuse has been facilitated, the output quality, best software/data practices such as documentation, testing, installation or reuse protocols, up to read the code, launch the RS, use examples...

This is the reproducibility point that looks at the validation of scientific the results obtained with the RS and/or the RD.

(R) Research. This point examines the research aspects associated to the RS and/or RD production: the quality of the scientific work, the proposed and coded algorithms and data structures, which are the related publications, the collaborations, the funded projects...

This point measures the impact of the RD and/or RS related research.

The CDUR protocol is flexible enough to be adapted to different evaluation situations. Each evaluation committee sets its own protocol adapted to a specific evaluation context with its own goals (recruitment, career evolution, publication...).

It is our understanding that, if followed correctly, CDUR may clearly contribute towards making RS and RD more visible, accessible and reusable, but also towards providing tools to build more solid implementations of FAIR principles for RS and RD [9].

Q7. How do the practical CDUR steps differ for Research Software and Research Data

A7. As you can see in the previous answer, the formulation is pretty similar. For example, in the **(D) Dissemination** step, the licenses that should be used for RS or RD are not the same, but there is the need of a license in both cases to set the conditions in which the outputs are shared and disseminated [4].

The **(U) Use** step is the one in which we concentrate the software or the data specific issues, but it is the same step for both outputs, to deal basically with quality and reuse issues but also with the reproducibility issues of the related scientific work. This step is intimately related to the Reusable FAIR principle [9].

Q8. From your perspective: how big is the problem, that is, how much research software and research data do NOT follow Open Science/FAIR principles

A8. This is still a big problem!! There are still a lot of RS and RD that do not follow Open Science/FAIR principles!! As said simply in a recent OpenAire webinar¹¹ by Kostas Glinos (April 2023): *you get what you reward*. Or as we have explained in the poster included in Figure 1 [10], there is an *evaluation/dissemination loop*. The dissemination of the research output will follow evaluation rules, and the dissemination provides outputs to be evaluated. If the evaluation rules evolve, the dissemination of the research outputs will adapt to the new rules.

There is now, as far as we understand, a largely accepted position to say that research evaluation should be changed, in particular to include other outputs than publications, namely RS and RD. But how? A lot of work is ongoing now to study how to propose these changes in the Open Science context, for example with the Coalition for Advancing Research Assessment (COARA)¹² work, a recent coalition that gathers many Universities and other institutions.

There is need for a new, large consensus on these changes in order to be adopted extensively. Our work [5, 9] proposes precise RS and RD evaluation protocols which can help to build a new research assessment system and, therefore, they can help to foster Open Science/FAIR policies and best practices.

Q9. Do you still encounter scientists and research software engineers who are NOT familiar with these principles?

A9. Truly, the construction, development and implementation of these principles are coming in the context of very recent, ongoing work, and changes take time in the scientific community. Some research communities, institutions, countries can go faster than others to adopt and install these changes. But it is clear that there is need for a large, general consensus, and the COARA coalition can provide the necessary context for this.

Q10. Are there people opposed to the principles?

A10. As far as we understand, most of the people are not opposed to Open Science or FAIR principles, many do not know them very well, or if you know them, how to comply? And how much time should you spend for this? And which are the consequences for the career (evaluation...)

11. <https://www.openaire.eu/blogs/openaire-in-greece-and-cyprus-webinar-on-the-european-reform-on-research-assessment>

12. <https://coara.eu/>

Q11. France has legislation to make research software (and data?) for publicly funded projects Open Source. Do you think this legislation has helped? Is it enough?

A11. Legislation for data is in the Digital Republic Law¹³ (October 2016), that gives a new duty for Universities and other research performing organizations. Its Article 6 says that *open data should be the default for all publicly funded data, including research*. So, for example, if your research project is publicly funded for at least half of the total funding, the data produced within the project should be open.

For software it was a bit more unclear, but the Second National Plan for Open Science of the Minister of Higher Education, Research and Innovation includes RS (July 2021) [11].

This new legal and political context helps to change the system, but still the system needs deep changes like, for example, in the evaluation, to consider these productions and how well they are disseminated to facilitate their reuse and the validation of the associated research findings.

Q12. What do you and your colleagues do to promote CDUR/FAIR further

A12. At the moment we try to make this work [5, 6, 7, 8, 9, 10] better known, presenting it to several conferences etc. This is why we thank you so much, Peter, as well as the Code for Thought team, for helping us to promote our work and to give us this opportunity to talk about it.

There is also a very recent news in the F1000Research journal, as there is a new collection of publications called *Innovations in Research Assessment Collection*¹⁴ and our three articles [5, 8, 9] have been included there last week. This will help a lot to give more visibility to this recent work, and, again, we thank the Collection Advisors, Simon Hettrick and David Moher, for the inclusion of our articles in this collection.

3 Epilog

This document reviews the questions and answers that were the basis of the podcast Code for Thought on Research Software and Research Data in Open Science and that was recorded in May 2023 at the Gustave Eiffel University. The podcast goal was to have an informal conversation about our recent published work and the interview was conducted by Peter Schmidt, who provided the questions. Our intention here is not to realize an accurate transcription of this podcast, but rather to complete and accompany the recorded audio, and to extend and clarify few points that could be difficult to grasp or that we forgot to mention there.

The work presented in the podcast and in this document has been partially funded, among others, by the LIGM, the University Gustave Eiffel, and the CNRS-International Emerging Action (IEA) PREOSI (2021-22). It has also enjoyed from the hospitality of the Departamento de Matemáticas, Estadística y Computación de la Universidad de Cantabria (Spain).

This work was also presented at the *Journée de clôture du printemps de la donnée*, June 30th 2023¹⁵.

We hope that you have enjoyed listening to the podcast and reading this text, many thanks for listening and/or reading about our work!

References

- [1] Borgman CL: The conundrum of sharing research data. J. Am. Soc. Inf. Sci. Technol. 2012; 63: 1059-1078. <https://doi.org/10.1002/asi.22634>
- [2] Concordat on Open Research Data, that was signed by the research councils of the UK Research and Innovation (UKRI) organisation, 2020, <https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf>
- [3] de Cock BM, van Dinther B, Jeppersende Boer CG, et al. The Legal Status of Research Data in the Knowledge Exchange Partner countries, Knowledge Exchange report. 2011. <https://repository.jisc.ac.uk/6280/>

13. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746/>

14. <https://f1000research.com/collections/innovations-in-research-assessment/about-this-collection>

15. <https://printempsdeladonnee.fr/events/journee-de-cloture-du-printemps-de-la-donnee/>

- [4] Teresa Gomez-Diaz. Free software, Open source software, licenses. A short presentation including a procedure for research software and data dissemination. 2014. Presented at the Workshop on open licenses: Data licencing and policies, EGI Conference 2015, Lisbon, May 2015, <https://zenodo.org/record/11709>
Spanish version: Software libre, software de código abierto, licencias. Donde se propone un procedimiento de distribución de software y datos de investigación, Septiembre 2015, <https://zenodo.org/record/31547>
- [5] Teresa Gomez-Diaz, Tomas Recio. On the evaluation of research software: the CDUR procedure, *F1000Research* 2019, 8:1353, <https://doi.org/10.12688/f1000research.19994.2>
- [6] Teresa Gomez-Diaz, Tomas Recio. Towards an Open Science definition as a political and legal framework: on the sharing and dissemination of research outputs.
V3 dated 28th February 2021 is available at <https://zenodo.org/record/4577066>
V2 dated December 2020 is published in POLIS N. 19, <https://uet.edu.al/polis/wp-content/uploads/2022/01/polis-19.pdf>
V1 dated September 2020 and entitled *A policy and legal Open Science framework: a proposal* is available at <https://zenodo.org/record/4075106>
- [7] Teresa Gomez-Diaz, Tomas Recio. Poster. The future of Open Science asks for a common understanding, EGI Virtual Conference 2021, <https://zenodo.org/record/6433533>
- [8] Teresa Gomez-Diaz, Tomas Recio. Research Software vs. Research Data I: Towards a Research Data definition in the Open Science context, *F1000Research* 2022, 11:118, <https://doi.org/10.12688/f1000research.78195.2>
- [9] Teresa Gomez-Diaz, Tomas Recio. Research Software vs. Research Data II: Protocols for Research Data dissemination and evaluation in the Open Science context, *F1000Research* 2022, 11:117, <https://doi.org/10.12688/f1000research.78459.2>
- [10] Teresa Gomez-Diaz, Tomas Recio. Poster. Research Software and Research Data: dissemination, evaluation and reusability in the Open Science context, 17th International Digital Curation Conference 2022 (IDCC22), <https://zenodo.org/record/6778872>
- [11] Minister of Higher Education, Research and Innovation, Second National Plan for Open Science of the Research Ministry (July 2021),
<https://www.enseignementsup-recherche.gouv.fr/fr/le-plan-national-pour-la-science-ouverte-2021-2024-vers-une-generalisation-de-la-science-ouverte-en-48525>
- [12] UNESCO Recommendation on Open Science, November 2021, <https://unesdoc.unesco.org/ark:/48223/pf0000379949>