# DATA MANAGEMENT PLAN IN EUROPEAN PROJECTS – METHODOLOGY AND TEST CASES IN NANOTECHNOLOGY COLLABORATIVE PROJECTS

**A. Gondikas[1,*], S. Makri[2], I. Katsavou[2], A. Zoikis-Karathanasis[2], I. Deligkiozi[1]**

[1]AXIA Innovation UG, Germany

[2]Creative nano, Greece

(*ago@axia-innovation.com)

## ABSTRACT

Preserving and providing access to research data is essential for promoting quality research and advancing scientific innovation. However, this concept is often perceived with skepticism, because data are the building blocks of research and innovation and making it openly available may result in loss of competitive advantages from the data owner. This misconception is frequently met in collaborative European research projects and the way to resolve it is by gradually introducing the new concepts through the development of the data management plan, starting early in the project. Here we present the main objectives of data management, potential issues, and discuss approaches to address them.

## INTRODUCTION

Data management in research projects has attracted attention in the past decade. Although scientific results are widely disseminated, the data used to produce these results often remain in the dark. More importantly, negative results, i.e. data that disqualify the research hypothesis are almost always neglected and lost with time. However, the value of all data is undisputable and there is an urgent need to promote data management in order to preserve and reuse information that is already available, thus avoiding repetition and duplication of work. The importance of this issue is reflected by the recent push for open access research and the introduction of mandatory data management plans (DMP) in research projects funded by the European Commission (EC). This plan is developed early in the project and all partners, both academic and industrial, are required to comply. However, depending on the planned paths to exploitation of the research results, not all partners are eager to comply with open access and data management principles.

Traditionally, academic institutions focused on basic research to improve science and the exploitation of research results was through publishing in scientific journals. On the other hand, industries used to focus on applied research to improve the company value and the exploitation of their research results was through patents, licenses and corporate secrets. In current times, the boundaries between academic and industrial exploitation of research results are blurred. Patent filing has become popular in academia and industries have taken interest in basic research hoping to take advantage of emerging technologies.

Therefore, data management is emerging as a key aspect in European research projects and is linked to the exploitation of research results. In order to boost economic growth, the European Commission (EC) is seeking to empower collaborations between European industries and academic institutions in research projects; however, the results from these projects are lost with time due to inefficient data management. For example, experimental data kept in the private computer of a PhD student may not be available after the student has graduated. This situation results in research outcomes that are not exploited further, duplication of work, and unnecessary spending of research funds. The EC is therefore pushing for the creation of a uniform system for data storage that facilitates tracking of data through the use of peripheral information (metadata) and open access data repositories.

Here, the main principles and methodology for data management in European projects is discussed and examples are given from EU-funded collaborative projects that aim to develop novel nano-enabled products, or to utilize nanotechnological solutions to industrial and societal challenges. The projects consortiums consist of universities, research organizations, and industries from the manufacturing, environmental, automotive, aeronautic, and infrastructure sectors. The inter-sectoral and interdisciplinary nature of these projects creates a dynamic environment where data management principles can be tested and advanced.

The EC has published guidelines for promoting the sound management of data, while introducing the concept of FAIR data, i.e. Findable, Accessible, Interoperable, and Reusable[1]. This concept refers to the handling of data during and after a research project, the applications of standards and ontologies, methods for curation and preservation, and the establishment of access rights. The benefit of FAIR data principles is the support of new research and further generation of data (Figure 1), without wasting time and resources repeating experiments that have already been made.

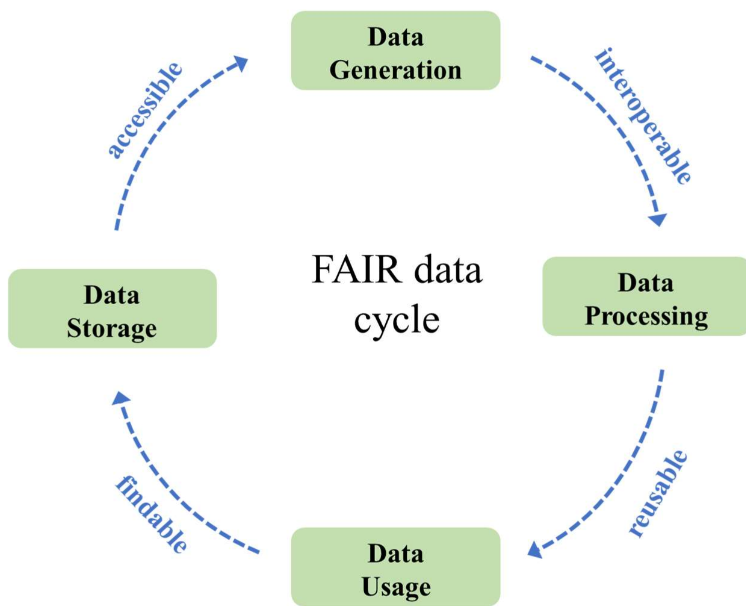


***Figure 1.*** *Schematic representation of the FAIR data cycle*

A key parameter in FAIR data is metadata, i.e. a list of relevant information accompanying the dataset. This information may refer to the process of data collection, e.g. the equipment used, the type of experiment, the person conducting the experiment, the institution, department, and facility where it takes place. It may also refer to the project, e.g. the funding source, the objectives, and the coordination team or the location, e.g. the country, postal address, and language. Finally, metadata may refer to the outcome of the experiment, e.g. a publication, patent, or product, and whether the data is openly available, confidential, or with an embargo time. A simplified example of metadata links to data is shown in Figure 2.

The use of metadata and the FAIR data principles may be readily compared to the example of publication indexing in libraries. Not too long ago, in order for someone to search for publications relevant to their interests, they had to go through several card catalogues to find what was available in a specific library. In order to find more, they had to travel to other locations and visit more libraries. This process has been radically changed and researchers can now search for publications from all over the world, simply by searching keywords in their computer, at a location of their choice. This ease of access has dramatically improved research quality, minimized duplication of work, and advanced scientific output rate. Therefore, it is easy to comprehend the immense benefits to research and innovation when the same concept is applied to research data. Although challenges still remain[2], the concepts of FAIR data and open access, in combination with data repositories and relevant infrastructures are the tools that will allow us to achieve this goal.

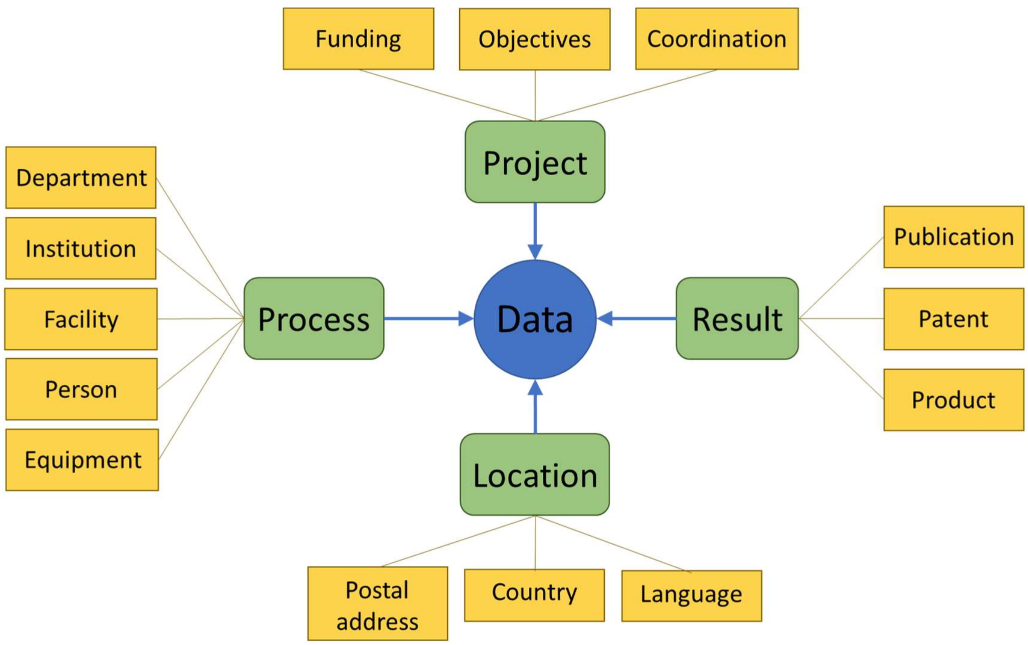


***Figure 2.*** *A simplified example of metadata links to data generation*

In parallel to FAIR data, the EC is pushing for the concept of Open research, which means that all publications must be made open access and the data should be "as open as possible and as closed as necessary". Open access in publications is achieved through publication in open access journals (often a fee has to be paid), or by depositing the manuscript that has passed peer-review in an open access repository[3,4]. This concept is readily adopted by researchers, because it increases the visibility and citations of their work, which improves their productivity metrics. On the other hand, open access in data is achieved by depositing the data in an open access repository. This concept is often met with skepticism, especially when the research is likely to result in a marketable product or valuable know-how, which is often the case in EC funded research projects. For example, industrial partners that collaborate in a research project expect to gain innovative knowledge that will provide them a competitive advantage in the market. Similarly, a researcher that has made a scientific discovery, which can lead to a new product in the market and the creation of a spin-off company will certainly want to keep this knowledge away from established companies. Finally, there is a general trend with scientists to keep their data private. Against these arguably strong cases, the EC has proposed a very reasonable solution. First, all data that is required to validate the results presented in a publication must be made openly available. Since the information is included in a publication, there should be no objection to making the raw data available for everyone to use. And second, data not used in publications should be made as open as possible and as closed as necessary. This recommendation allows the owners of the data to decide whether it will be made open access, when there is no risk of losing the intellectual property rights (IPR) on a certain invention or to keep it confidential otherwise.

## CONCLUSIONS

In our experience, from several EC-funded research projects, there is still a trend of researchers to keep data private. This trend persists partly due to insufficient understanding of the FAIR data and open access principles. The DMP is developed very early in the project, introducing concepts of open access, when the research results are unknown. Researchers are thus hesitant to commit to open access, not knowing if their research will produce marketable results. Furthermore, we have observed that researchers are hesitant to incorporate new concepts, such as metadata. The key to resolving this issue is the process of developing the DMP, where three stages are involved. An initial DMP is drafted within the first six months of the project, where the basic principles of FAIR data and open access are introduced. At this stage, it is important to stress the difference between data management and open data, in order to address concerns of researchers that are not yet familiar with these concepts. The concept and importance of metadata along with examples is also necessary to explain. Later on, a questionnaire may be distributed to collect information about how partners are considering to treat their data. Building on this information, a detailed DMP is developed close to the middle of the project, where guidelines are provided in order for partners to follow the EC recommendations and be compliant with the grant agreement. The final DMP is delivered towards the end of the project, providing details on what has been achieved, so that future projects may utilize and build on these procedures. Through the gradual development and implementation of the DMP, industries and the scientific community will eventually get on board the effort to develop and feed the necessary infrastructure for creating a virtual research environment with a multitude of information that is findable, accessible, interoperable, and reusable either openly or under terms.

## REFERENCES

[1] European Commission, Directorate-General for Research & Innovation (2016) Guidelines on FAIR Data Management in Horizon 2020

[2] S. Karcher, E. Willighagen, J. Rumble, F. Ehrhart, C. Evelo, M. Fritts, S. Gaheen, S. Harper, M. Hoover, N. Jeliazkova, N. Lewinski, R. Marchese Robinson, K. C. Mills, A. P. Mustad, D. G. Thomas, G. Tsiliki, C. Ogilvie Hendren. NanoImpact 9 (2018) 85-101

[3] P. Manghi, N. Manola, W. Horstmann, D. Peters. The Grey Journal (TGJ) : An International Journal on Grey Literature 6 (2010) 31-40

[4] P. Principe, N. Rettberg, E. Rodrigues, M. Elbaek, J. Achirrwagen, N. Houssos, L. Holm Nielsen, B. Jorg. Procedia Computer Science 33 (2014) 92-94

## ACKNOWLEDGEMENTS

This work has received funding from the European Union Horizon 2020 Programme projects MASTRO (Grant Agreement No 760940), OActive (Grant Agreement No 777159), and PROCETS (Grant Agreement No 686135)