# Reflection on FAIR Principles and Digital Humanities through Scientific Pre-Initiation Projects in Vocational Education

*Azevedo, Láyla*
*Fundação de Apoio à Escola Técnica do Estado do Rio de Janeiro*
*layla-azevedo@hotmail.com*
*Bessa, Alessandra*
*Universidade Federal Rural do Rio de Janeiro*
*alefbessa@gmail.com*

## Abstract

Numerous projects have been developed in the field of Digital Humanities, including studies related to the FAIR Principles, (acronym for Findable, Accessible, Interoperable and Reusable) and Data Provenance. Bessa (2021) presented a related recent research on soils data security, in Digital Humanities tools based on FAIR Principles and Data Provenance, using OpenSoils plataform. The present article intends to reflect on an initiative to present to the students of a public education institution, which offers vocational high school, in the external and subsequent concomitant modality, the dynamics of Digital Humanities and the FAIR Principles, through a scientific initiation program.

Keywords: Educational Tools, FAIR Principles, Scientific Initiation, Vocational High School, Young Talents.

## 1 Introduction

Numerous projects have been developed within the scope of the digital humanities. Studies involving DH's are comprehensive: research on the state of the art; use of Natural Language Processing (NLP) for textual analysis, applied to the areas of History and Literature; wiki-like collaborative creations and; studies on repositories, demonstrating the concern with the storage of data that we produce daily in large quantity and speed (big data).

In the context of this article, it is worth emphasizing the importance of repositories, as digital informational environments for the management and control of the academic and scientific production of institutions and/or communities, offering advantages such as access and preservation of information in the long term (Camargo; Vidoti, 2008 apud Bessa, 2021). The FAIR principles and data provenance have also been the object of study of HDs. We can mention the studies developed by Bessa (2021), on the OpenSoils platform[1], a digital repository in the area of soil safety, with input to the FAIR Principles and Data Provenance (Cruz, 2018 apud Bessa, 2021).

However, a search in the literature revealed that there are still no studies or projects developed within the scope of Vocational High School that address digital humanities or the FAIR Principles and adequate data management.

Although the Principles are presented as good practices to be considered when working with data and not a guideline or determination to be followed, it would be interesting for students of technical courses to be aware of this issue at the beginning of their careers.

The present article intends to reflect on an initiative to present to the students of a public education institution, which offers vocational high school, in the external and subsequent concomitant modality, the dynamics of Digital Humanities and the FAIR Principles, through a scientific initiation program. And for that, it is structured as follows: in section 2 we describe the FAIR Principles and Data Provenance, section 3 presents the concepts related to scientific initiation projects carried out in the context of secondary education, then in section 4 we demonstrate the structure of the project being conducted in a Vocational Technical School and finally, in section 5 we present the conclusion of the study.

## 2 The FAIR Principles and Data Provenance

The FAIR Principles, acronym for Findable, Accessible, Interoperable and Reusable, have been disseminated since 2016 as a way of proposing new practices for researchers with regard to Data Management (Santos, 2021). In the same year, the FAIR Guiding Principles for scientific data management and stewardship was published in the journal Scientific Data with the prerogative of establishing guidelines to improve the application of the FAIR Principles – (Wilkinson et al., 2016). This article brings the approach to good data management in order to benefit new discoveries and knowledge innovation, as well as data integration and reuse.

The FAIR Principles are composed of fifteen high-level guiding principles, with divisions and subdivisions, and fifteen characteristics of their applicability in repositories (Bessa, 2021). Knowledge and data shared from the beginning of a survey contribute to the dissemination of the most recent results, in accordance with European Commission (website) policy. In this way, the project should build on this basis of good practices for data management.

Considering the digital environment, where research has been developed, it is essential that data have credibility. Otherwise, no result will be validated, without having an origin, a traceable history. Provenance determines the path that information has taken. How it was created, where and by whom. Thus, it can be said, from a computing point of view, that data provenance is "information about entities, activities and people involved in the production of a data or thing, which can be used to form assessments about its quality, reliability" (W3C Consortium).

In the context of this article, the project will greatly benefit from working with information generated by provenance data. According to Wilkinson (2016), the "Reusable" principle is directly linked to provenance metadata as a condition for reusing digital data.

## 3 Scientific Initiation Programs for High School

In general, scientific initiation programs seek to encourage students to practice research or to lead them to learn topics that complement the learning taught in the classroom. According to the CNPq[2] "it is necessary that, from the first years of formal education, students are put in contact with scientific culture, that is, with the scientific way of producing knowledge and with the main human activities that have shaped the environment and human life throughout history". In Brazil, at the national level, scientific initiation is encouraged in public high schools, through projects such as IC-Jr[3].

At the state level, the Carlos Chagas Foundation for Research Support of the State of Rio de Janeiro - FAPERJ also promotes such programs, through a partnership with the Science Center Foundation of the State of Rio de Janeiro - Cecierj. Created in 1999, the Young Talents Program has the "objective of contributing, through the offer of internship scholarships for scientific pre-initiation to students of the public high/technical education network in universities and teaching and research institutions based in the state of Rio de Janeiro". (Faperj, 2022)

By including scholarships for Young Technological Talents, targeting the work carried out in schools of the Technical School Support Foundation – FAETEC, JT arrived in 2014 at the State Technical School of Paracambi – ETE Paracambi[4]. Through these scientific pre-initiation grants, several projects were developed with the students of the Computer Technician course, among them: Development and application of a software: thinking and action game approaching computational content; Devices for alarm and detection of fuel gas leaks using the Open Source[5] Arduino UNO platform and Educational Robotics: an essay in the construction of algorithms (ETE Paracambi, 2022). The Open Source Definition was originally derived from the Debian Free Software Guidelines (DFSG)[6].
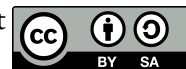
In 2021, the unit also started to offer the Environmental Technician course and, seeking integration and interdisciplinarity between the courses, there was a need to develop a scientific initiation project that would achieve this purpose. The project, whose name is FAIR Principles applied in research in Vocational Technical High School, is a proposal expected to start in the second half of 2022.

## 4 The Project

The FAIR Principles and Data Provenance are a relatively recent topic. The school community of ETE Paracambi is unaware of the subject, this fact was verified during the meetings to choose the theme of the projects, where the teachers informed that they did not know the theme that was being proposed and, during the presentation to the students who were reluctant to accept part of the project because they had no idea what they were dealing with. Ignorance of the subject occurred even among students and professors in the IT area. For this reason, the project proposes to disseminate this knowledge, in addition to allowing, influenced by the Digital Humanities, the integration between the courses taught at the unit. It is very relevant to mention that this is an embrionary project and it is in its very first beggining. Expectation is increasing and we are very hope on its development.

Since the unit started to receive scholarships from the Young Talents Program, all projects have focused on the Computer Technician course. With the implementation of the Environmental Technical course, the unit realized the need to propose a project that also included students of this course.

The project, with an interdisciplinary character, consists of an exploratory research using bibliographic research so that students have a first contact with the proposed theme.

Program fellows will be encouraged to investigate the concept of repository. At a conceptual level, how an institutional repository can have FAIR characteristics and how the indexing of Course Completion Works - TCC can be done following these principles, taking into account the description of data and metadata in the research.

Another point that should be explored concerns the format of the data so that it can be read on other computers at any time, regardless of the software version being used.

With the analysis and discussion of the possibilities of applying good practices in Data Management in research carried out by students of Technical courses, it is expected to expand the research possibilities of students inserted in the courses taught at ETE for the elaboration of their final papers, allowing collaboration between the areas of information technology and the environment.

## 5 Conclusion

It is understood that the more the Students' Completion Assignments are well developed, indexed and stored in institutional repositories, with FAIR characteristics and Data Provenance, these works can be found, accessible, reused, interoperable. Assisting in other researches, through the correct description of data, such as metadata. In this way, other students/students will be able to have access to these researches.

In this way, it will be possible to give greater visibility to the work of students and to prepare them better for the world of work, through the acquisition of new knowledge regarding the ability to integrate with other areas of knowledge.

It is necessary to follow this initiative to verify if it is possible, through scientific pre-initiation programs such as the Young Talents Program in Vocational High School courses, to disseminate the concepts related to the FAIR Principles and the Provenance of data and to promote the understanding of the importance of data for conducting scientific research and preparing Course Completion Works carried out at this level of education.

## Notes

[1] OpenSoils hosts, connects and shares vast amounts of data and soil knowledge curated at a

Brazilian and South American level. Available at: https://opensoils.org

[2] CNPq - National Council for Scientific and Technological Development.

[3] IC-Jr – Junior Scientific Initiation.

[4] In 2014 the unit was a Center for Vocational and Technological Education - CETEP, becoming an ETE in 2021 through Decree No. 47.478 of 02/05/21.

[5] "Open source is a way to plan and collaborate within the free software community. This philosophy is based on intellectual freedom and fundamental principles: transparency, collaboration, delivery, inclusion and community."
https://www.ibm.com/br-pt/topics/open-source

[6] Debian Free Software Guidelines (DFSG) originally allowed Open Source definition to be created.
https://opensource.org/docs/definition.php

### References

Bessa, A C F 2021. Princípios FAIR e Proveniência em repositórios de dados voltados para a Agricultura Digital: um estudo exploratório à luz das Humanidades Digitais. 2021. 88 p. Dissertação (Mestrado em Humanidades Digitais). Instituto Multidisciplinar, Universidade Federal Rural do Rio de Janeiro, Nova Iguaçu, 2021, https://tede.ufrrj.br/jspui/bitstream/jspui/5674/5/2021%20-%20Alessandra%20Castro%20Fiorini%20Bessa.pdf

CNPq. Iniciação Científica, viewed 25 abril 2022,<http://memoria.cnpq.br/web/guest/iniciacao-cientifica>

ETE Paracambi 2022. Relatório de Atividades.

Faperj 2022. Programa Jovens Talentos. viewed, 25 abril 2022, <https://siteantigo.faperj.br/?id=1305.2.7.

Santos et al. 2021. Princípios FAIR aplicados à gestão de dados de pesquisa. Rio de Janeiro: Ibict. p. 5 -6. DOI: 10.22477/9786589167242.pref

Wilkinson et al. 2016 The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018, <https://doi.org/10.1038/sdata.2016.18>

W3C. The PROV Data Model. <https://www.w3.org/TR/prov-dm/>