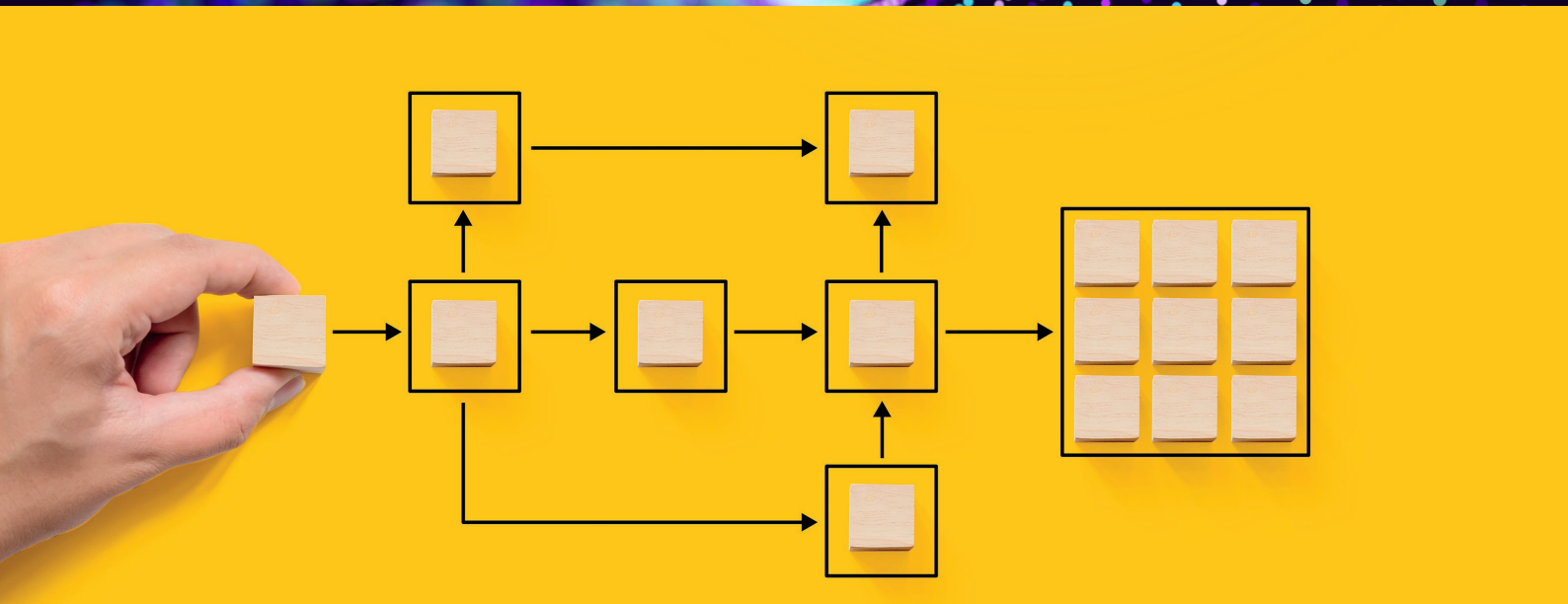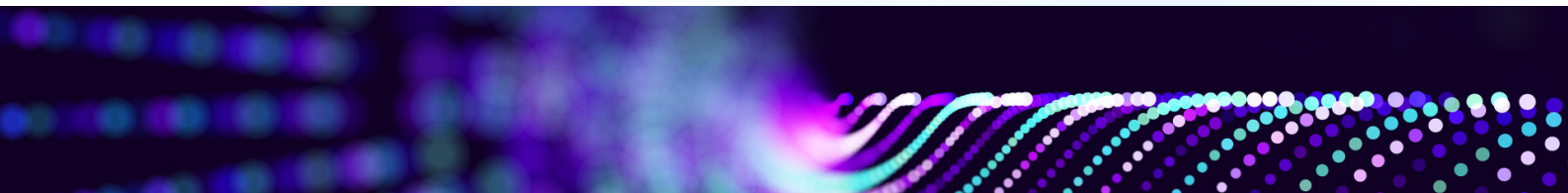# FAIRness assessment of Generations and Gender Programme longitudinal survey dataset

Authors names, affiliations and ORCIDs:

☐ Arianna Caporali, French Institute for Demographic Studies (INED) and Generations and Gender Programme (GGP), https://orcid.org/0000-0001-5770-1658

☐ Olga Grünwald, Netherlands Interdisciplinary Demographic Institute and Generations and Gender Programme (GGP), https://orcid.org/0000-0003-0591-4752

☐ Clara Linés, Digital Curation Centre (DCC), https://orcid.org/0000-0003-3437-5145

Support action:

AIRness Assessment Challenge. During the 3 month challenge participants took part in three virtual workshops to self-assess and incrementally improve the FAIRness of their selected outputs. During the support action, participants benefited from interacting with a group of mentors representing the various FAIRness assessment tools and methods.

Keywords:

F-UJI, FDMM, SHARC, social sciences

Summary:

A team involved in the Generations and Gender Programme tested F-UJI, FDMM and SHARC to assess the FAIRness of their longitudinal panel survey dataset.

Introduction

Our background is on social sciences and research data support, and we are both involved on the Generations and Gender Programme (GGP), which is a cross-national longitudinal panel survey on life-course and family dynamics, which comprises two rounds of data collection: Generations and Gender Survey-I (GGS-I), launched in 2000 and covering 19 countries, and GGS-II, initiated in 2017 and still ongoing. GGP is on the European Strategy Forum on Research Infrastructures (ESFRI) Roadmap and currently receives European funding to become a European Research Infrastructure Consortium (ERIC). As part of that process, GGP needs to deposit the Generation and Gender Survey (GGS) datasets and metadata files in a data repository. We believe that before depositing these datasets it is paramount to evaluate and improve their level of FAIRness.

During the support action, we wanted to evaluate the FAIRness of the datasets and metadata files belonging to the second round of the GGS, which are stored in the GGP Colectica Portal (https://ggp.colectica.org/) according to the Data Documentation Initiative (DDI) standard version 3.3. GGP Colectica portal allows users to browse the GGS data and metadata online.

Approach taken:

We tried a tool, F-UJI1, and two methods: FDMM2 and SHARC3. We first used the F-UJI tool, and we obtained a very low level of FAIRness (4%). We had three hypotheses as to the possible reasons for the low score: F-UJI tool being sensitive to the software we use to develop our catalogue (Colectica), the lack of persistent identifiers, such as DOI, for our datasets and metadata files, and the type of DDI version that we use (DDI version 3.3).

To examine these hypotheses, we tested two of our datasets from the first round of the GGS (https://doi.org/10.17026/dans-z5z-xn8g and https://doi.org/10.17026/dans-xm6-a262) that are stored in the Dutch data archive, DANS, which uses Dataverse. There our datasets have DOIs and are in DDI version 2. For those datasets we obtained a much higher score (60%).

We also tested in F-UJI the datasets from another project, the European Social Survey (ESS), that we know are documented in DDI version 3, just like GGS datasets. We tried entering the weblink in F-UJI and got a low score, while with the DOI we got a moderate score.

Based on these tests, we deduced that the tool is sensitive to both the software and the DDI version, and that the use of persistent identifiers is necessary.

We also used the FDMM method, after clarifications from the mentor and reading the guidelines published by the RDA [1]. From that we learned that the level of our scores was affected by the lack of persistent identifiers, controlled vocabularies, and external references to other data and metadata. We are already working on these enhancements that are either under consideration or in the implementation phase. Compared to FDMM, in SHARC we performed better in reusability and interoperability. We found SHARC easier to apply compared to FDMM.

---

1   https://www.f-uji.net/
2   FAIR Data Maturity Model Working Group. (2020). FAIR Data Maturity Model. Specification and Guidelines (1.0). Zenodo. https://doi.org/10.15497/rda00050
3   David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., Jonquet, C., Dollé, L., Jacob, D., Bailo, D., Bravo, H., Gachet, S., Gunderman, H., Hollebecq, J.-E., Ioannidis, V., Le Bras, Y., Lerigoleur, E., Cambon-Thomsen, A., & SHARC Community. (2020). Templates for FAIRness evaluation criteria - RDA-SHARC ig (1.1). [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3922069

Challenges encountered and addressed:

From discussions with Colectica and the mentors we learned that the problem we were having with F-UJI was that the tool was not detecting machine actionable metadata that is included on the GGP portal (i.e., standardised DDI XML or JSON files). The software (i.e., Colectica), file formats and metadata standard we use are well established in our community and fully recognised by CESSDA (Consortium of European Social Science Data Archives), so we do not see a need to change that.

With FDMM and SHARC we were uncertain about how to apply the final scores and how to interpret the criteria and some of the options. The methods are both quite complete and useful, but also rather subjective.

Some aspects of FAIRness are very case-specific, and each method and tool has its own peculiarities. We were new to the field and were surprised to learn that there is not a single official standard or unique method to assess FAIRness of datasets, because of that it is important to use as many tools or methods as possible and compare the scores obtained with all of them in order to figure out take-home messages and key actions required to improve FAIRness.

Impact:

Overall, the assessments helped us identify that we need to apply persistent identifiers, use controlled vocabularies and improve external references (i.e., linking the datasets to, for example, bibliography and references, e.g., scientific articles that are based on that data), and gave us a stronger argument to focus on these improvements. Something that we learned from this is the distinction between global identifiers and persistent identifiers. The global ones are unique identifiers applied by the tool used to put the data and metadata online, in our case Colectica. Then there are the persistent identifiers, for example DOIs, which we don't have yet.

The assessment also gave us a better understanding of the different domains of FAIRness as well as the criteria that we need to meet to be FAIR. However, we missed concrete and detailed examples of application of the methods and tools to a case study. This could have been of help to better understand how to apply the methods and tools to our case and how to interpret the results. This would be particularly helpful to those trying to carry out the assessment independently, especially if they are not familiar with the terms used in the field, as it was our case.

We presented this work at a conference for users of DDI [2], which is mainly for social science data providers, and talked about the tools and methods we tested. Some people may already know about these methods but we hope we helped others learn about them.

The exercise has helped us plan our future activities and we will be implementing persistent identifiers and controlled vocabularies. In the long term, ideally, we would have a new GGP data platform with all the elements necessary to be FAIR set up in place. We also would like to have the datasets deposited to a trusted repository.

Key messages:

Contacting the developers of the tools and methods you want to use is useful to clear any doubt and correctly apply the method.

It is important to use as many tools or methods as possible and compare the scores obtained with all of them in order to figure out take-home messages and key actions required to improve FAIRness.

Links:

[1] FAIR Data Maturity Model Working Group. (2020). FAIR Data Maturity Model. Specification and Guidelines (1.0). Zenodo. https://doi.org/10.15497/rda00050

[2] Caporali, A., & Grünwald, O. (2023, November 28). FAIRness assessment and the role of DDI in the Generations and Gender Programme. EDDI2023: The 15thAnnual European DDI User Conference (EDDI2023), Ljubljana, Slovenia. Zenodo. https://doi.org/10.5281/zenodo.10263500