

Case description from the project “FAIRify Humanities” of the National Data management Forum 2019.

Title: Consumer behaviour data on food shopping. Donate Your Food Data Lab

Name and affiliation: Bent Egberg Mikkelsen, AAU. Karsten Kryger Hansen, AAU.

Research area: Consumer behaviour

Data set(s): Proof of concept for consumer behaviour data.

What efforts was carried out to make your data more FAIR?

In general: The primary task has been to describe a proof of concept for a research infrastructure that could handle a case where consumers on a voluntary basis could provide data about their shopping, based on quantitative streamlined data. Data are sourced from credit cards or loyalty cards. There are two options in such a case for providing access to data; (i) made available as a web resource through a API like interface, (ii) made available for download. For the project, a dataset with 200.000 consumers’ data from a period of 3 months was tried for the proof of concept made available from a major Danish retailer’s loyalty card program.

More “Findable”:

If data like this is made available a resource, rather than being available as download, there is a still a need to able to issue persistent identifiers to identify datasets. This can be quite cumbersome, and the ability to do this, could be achieved by following e.g. the RDA standard for structuring data and system allowing for [the citation of evolving datasets](#). For instance a reference on the date for access could be used or a “n value” (referring to the number of consumer transactions) Furthermore there needs to be published value, e.g. in re3data.org about the existence of a database with consumer behavior data, and announced in other not so FAIR ways.

More “Accessible”:

A suggested way to provide access to this kind of data, would be by building a resource that can be queried by researchers that are approved for access. Access could be granted for instance by completing a tutorial. It could be considered to grant access to other than researchers for instance market bureaus or consultants. Or even consumers (providers of data) in an easy and portable way. But that needs to be negotiated with the donaters/providers of data. In addition, access rights could be at different levels. The research data can be governed on an individual end user basis, e.g. providing governing mechanisms for ensuring proper anonymization that can even rely on a script running e.g. differential privacy scripts on data. This approach would also allow each consumer to have some kind of feedback on consumption patterns, if there is a need for that.

More “Interoperable”:

The data in the project is very streamlined, almost like triples, and the interoperability between resources. Based on 200.000 consumer data as proof of concept, data could easily be linked via the EAN number, which is an international standard, to the database TradeSync. This database follows international rules defined by the EU regulation number 1169. This allow for standardized information on nutrition information about the specific product. In this case, the EAN is the shared key, and is considered the standard. The nutrition information can then be further queried in other resources to provide data like carbon footprint etc. In the test so far the interoperability between the two data sets – the consumer shopping data and the food database was proven.

More “Re-usable”:

There has not been selected a given license for the data, as this is described on a proof of concept level. However, it should be possible to assign a license to these kind of data, as the high interoperability and accessibility perspectives allows for the availability of many kinds of data, where e.g. an end user can use their own subscription to other services or own data, to query for further details, e.g. based on EAN numbers.

What was the biggest challenges to make your data more FAIR?

Technical challenges in interfacing the two data sets. Also given access rights for the pilots to the students, signing NDA’s check for GDPR compliance issues and contacting DPO/AAU’s proved to be some of the challenges. Also missing values in the TradeSync database was a challenge

Other comments or conclusions:

There seems to be great potentials in establishing such kind of an RI and bright opportunities for funding as well. It is important however to establish the needed training resources in order to get young researchers and students into the system and routine of using these kind of data. In addition, there could be great opportunities in looking into developing models for fair exchange of data since in our case we have more stakeholders involved – all with a kind of IP right. Consumers provide data.

Retailers process them and researchers analyse them. A good question is therefore that since everybody provides something, how can this “something” be exchanged into fair user and access rights.