Dataverse@unimi.it: a path of gradual and steady growth

The Research Data Management @unimi project (https://dataverse.unimi.it/) was born in 2018 to respond to the demands of funders for transparency and those of reproducibility or replicability of research.

At that time, there was no centralised service to support the management of FAIR data, nor tools to manage it. This meant that each researcher or research group tried to provide for data management, archiving and storage independently and according to common sense.

From the interviews conducted in the departments, however, it soon became clear that there was a strong need for archiving tools and training in FAIR data management and rights management (especially in collaborations with third parties).

The first step was therefore to identify a tool that would be suitable for all subject areas. The tool had to cover those areas that had no reference repository and ensure data management in FAIR mode.

After various experiments, the choice fell on Dataverse, an open-source software platform developed by Harvard University that is widely used by national research systems (e.g. in Denmark and the Netherlands).

The software was implemented with the support of 4Science.com, which manages the hosting and maintenance and updating processes.

At the same time as the software was analysed, the policy on RDM was approved, which demanded FAIR data management from the researchers receiving funding.

The publishing environment was also placing increasingly precise demands on researchers with respect to the data underlying the results presented in the articles.

Any tool that one intends to implement, although excellent in terms of the functionality it offers, is of little use if it is not brought to the attention of researchers and if it is not demonstrated how the tool can be useful to them.

On this point, the University and the Directorate that manages the Open Science processes have put a great deal of effort into a series of actions:

- Preparation of instructions and guidelines for data deposit.
- Creation of a site dedicated to RDM which, in addition to providing instructions for using the repository, also offers a series of information on the subject of data management, tools and services available.
- Monthly information meetings on RDM, FAIR Data and the use of Dataverse.
- Specific skills courses (compulsory) for PhD students, repeated every cycle.
- Presentations on request to Departments and specific areas.
- A pilot project for PhD students dedicated to the drafting of the DMP (Data Management Plan) for their research project.
- *Ad hoc* training projects dedicated to international UniMI-led research groups.
- Support for researchers who need to archive their data in a FAIR manner.
- Monitoring report on the use of Dataverse (who - what - how) including the many downloads.

- Recruitment of two (soon to be three) data stewards to support the RDM project.

Particular attention was paid to the implementation of the software: first of all, attention was paid to the connection with persistent identifiers (ORCID, DOI), then an attempt was made to facilitate the access of our users by allowing them to register to the system both through the connection with the university authentication system and through ORCID.

The workflow for loading data into Dataverse, described in one long and one short guide, is periodically reviewed, based on user suggestions and also trying to make the process as robust as possible.

The two data stewards worked on the data and metadata quality of the repository. The arrival of two data stewards who were trained in research data and its management made it possible to implement a validation workflow for the datasets both already published and being published, and this facilitated the certification process (Core Trust Seal) that is underway.

The service was by no means quick nor easy to implement, partly because the tendency of those producing datasets is to omit all activities that are not mandatory due to the intensity of research processes. However, as requests from funding bodies and journals have grown, the use of Dataverse has also begun to spread.

The dissemination activities described above, however, continue with the same intensity. To work effectively, the service must have staff dedicated to supporting data management and promoting the service among the university's researchers.

Dataverse is a tool that, if carefully managed, allows widespread dissemination of data. This is demonstrated by the statistics: in the middle of last year there were a few thousand downloads of the archive, while one year on we have almost reached 300 thousand downloads.