

# IC\_9 Quantitative accounting of Open Data use

**Author:** Markus Fiebig

## **Description:**

Research infrastructures have often used data policies emphasizing the correct attribution of scientific merit to the data originator responsible for generating a high quality dataset. These policies usually allow use of the data free of charge, but often rule out e.g. commercial use of the data, and thus are not open in the sense as defined e.g. by the creative commons licenses. Frameworks and funding agencies now push towards truly open data policies, and may have some success in forcing their implementation. On the other hand, providing an incentive for the data providers in exchange for an open data policy may be more successful. Here, correct quantification of the use of a dataset across applications may provide such an incentive in order to yield a quotable metric to quantify the scientific merit associated with providing a highly used scientific dataset. Efforts in this direction make use of DOIs associated with datasets, and also services counting the use of DOIs, also those for data, are emerging (e.g. <http://stats.datacite.org/>). DOIs can either be issued to the data in a flat fashion and chosen granularity on import into the archive, or by providing the option of coining DOIs for a user defined selection of data on export. The first option allows for the exact quantification of data use, while the second option is more convenient when quoting the data, but doesn't give an exact quantification, i.e. puts large data providers in a data selection at a disadvantage. In order to have the advantages of both options, DOIs need to be associated with the data in a flat fashion, and DOIs coined for data selections need to refer to all DOIs of the original datasets the selection consists of. This approach would allow for the exact quantification of use for each original dataset while allowing for creating DOIs for arbitrary selections of data, even across repositories. Provisions for such use of DOIs are already made in the DOI metadata specification. The implementation case would thus need to:

- Specify in the ENVIplus reference model that RI data repositories issue DOIs to archived data in a flat fashion and chosen granularity on import into the archive (which requires an archive capable of version tracking).
- Specify in the reference model that, when providing the service to coin DOIs for data selections, to include references to the original dataset DOIs in the DOI of the dataset selection. This specification would also need to apply to portals connecting several repositories if they provide this service.
- Work with initiatives providing DOI use tracking, e.g. DataCite, to resolve the references to original datasets in DOIs of data selections when counting data use.
- Work with standardisation bodies, e.g. RDA, to include respective specifications in their standards.

**Issue:** Markus: "My problem is that I clearly see the need to such a service, but I'm lacking virtually all connections into the relevant bodies, i.e. DataCite, RDA, ... , so I may not even be the ideal candidate for leading this implementation case."

## External Link

1. IC\_9 Notebook: <https://envriplus.manageprojects.com/projects/wp9-service-validation-and-deployment-1/notebooks/633+>