

Including open data as an additional indicator for the performance-based allocation of funds

Criteria for datasets to qualify as open data for performance-oriented funding at the Charité

Miriam Kip*, Nico Riedel*, Sabine König*, Evgeny Bobrov*

*Berlin Institute of Health at Charité – Universitätsmedizin Berlin, QUEST Center for Responsible Research, Berlin, Germany

For the allocation of the open data incentive as an additional indicator of the LOM (*Leistungsorientierte Mittelvergabe*, performance-oriented funding at the Charité – Universitätsmedizin Berlin), below criteria regarding data and publications need to be fulfilled. Additional Charité-specific criteria for the respective researcher's eligibility to receive LOM funding are not considered here.

The criteria for the open data incentive are as follows:

- Research data have been made freely accessible by researchers of the Charité
- OR the data have been shared with restricted access and meet the following requirements:
 1. Data is stored in an external repository (or archive, database, registry)
 2. A standardized access route is named, i.e. the access requirements, the procedure for a request and the responsible persons or offices are described
 3. The reason for the restricted access is stated or is directly evident from the data being subject to data protection
 4. Access is possible for all academic researchers – at least from the European Economic Area
 5. Co-authorship of articles is not a condition for the provision of the data
 6. The access to the data is free of charge or maximally requiring compensation of expenses

- The data can be raw, primary, or secondary data (e.g. from analyses of freely available datasets, meta-analyses, or health technology assessments); the data would thus allow the analytical replication (retracing of analysis steps) for at least a part of the study's results; reporting of statistical values (means, standard deviations, p-values etc.) is not sufficient
- Data have been shared in the context of an article publication; thus, stand-alone datasets without reference to an article are not considered

- The data can also be found independently of the publication; thus, Supplementary Materials are only permissible if they are stored in a repository (archive) and can also be found via this repository.
- The publication contained an explicit reference to the dataset(s); a reference to e.g. supplementary materials without further explanation is not sufficient, nor is a reference to a database without naming the data set, accession code or exact search settings.
- The data are indeed available and can be accessed at the time of checking (for data under embargo, this must expire no later than July 31st)
- Data have been shared in a machine-readable format; for tables e.g. CSV, Excel or Word files, but not PDFs or image formats

The open data definition applied does not include:

1. Analysis scripts, computer programs, and other methods, materials, and protocols, even if their development was the goal of the research project and/or their presentation was the focus of the publication; if data has been collected and shared for development or validation, these can, however, fall under the open data definition.
2. Data contained within the article text itself, as long as these are not embedded tables, which can be accessed as digital objects for themselves as well.
3. Image, audiovisual, and other data which primarily serve illustrative purposes
4. Data supporting case reports, unless these were shared in repositories (archives) of the respective discipline.
5. For systematic reviews and meta-analyses: lists of sources or other general information on the studies, such as survey method or number of participants (eligible for LOM, however, are datasets newly compiled from the original literature that ensure traceability of the analysis, such as extracted text passages or statistical values).
6. Data which is only accessible on request or when fulfilling certain requirements, unless it is personal or otherwise sensitive data that is available via a standardized access route (also see above; this does not include, for example, availability "upon request").
7. Data from data collections of consortia ("data pools"), if it is unclear whether the authors themselves have contributed to the pool.
8. Data for which only a "private link" is shared, so that it cannot be found in the repository, but is only accessible via the publication.
9. Data shared before the LOM period under consideration (this is to ensure that only a limited number of articles can be rewarded for sharing of a particular dataset).