**List of publications from aquatic mesocosm experiments investigating where, when and how FAIR data get published**

Heidrun Feuchtmayr1, Kelly Le Quesne1 & A. Katharina Makower2

1 UK Centre for Ecology & Hydrology, Lake Ecosystems Group, Lancaster, UK

2 FVB-IGB, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Stechlin, Germany

192 publications from two EU projects focused on aquatic mesocosm facilities were analysed for a) the number of data publications in a repository on its own and the number of data publications associated with a scientific paper, b) the time lag between mesocosm experiments and data/paper publication, and c) adherence of scientific papers to FAIR principles of data publication. The table published here refer to the three research questions and are linked to the peer reviewed paper:

Feuchtmayr H, Le Quesne K, Makower A.K. (2025), Where, when and how FAIR do data from aquatic mesocosm experiments get published? BioScience, DOI: 10.1093/biosci/biaf081

**Method**:

The AQUACOSM and AQUACOSM-plus projects involved a network of aquatic mesocosm facilities across Europe from 2017 to 2020 and 2020 to 2024, comprising 21 and 30 partners, respectively, and over 60 mesocosm facilities.

In order to obtain the data to download here, we searched for datasets (or links to datasets) within journal articles using Web of Science and Google Scholar and within data repositories (PANGAEA, Dryad/Zenodo, EmodNet, SEANOE, Mendeley) as well as data holdings for genetic, enzyme and taxa data (NCBI, European Nucleotide Archive, EMBL-EBI Metabolights, Ecotaxa Obs, MassIVE). The search terms we used were “AQUACOSM” and “AQUACOSM-plus”. A cross-check was performed to identify if any datasets were linked to journal articles missed in our search and *vice versa.*. All paper publications were manually checked for data availability by looking at the supplementary materials and their data availability statements.

Journal articles that contained data in the supplementary material and in a data repository were considered a single item and categorised as data in repository. 2024 data only comprises January to mid-March.

The table presented here lists all DOI links of open access publications from the AQUACOSM or AQUACOSM-plus project for all years (total = 192) together with the year of publication. For each analysis, (a) Figure 1 and 2: location of dataset publication and occurrences of data published to online repositories versus data published in association with a scientific paper, b) Figure 3: time lag between experiment and publication, and c) Figure 4 and 5: FAIR data publication) the table contains a column stating if the paper or dataset was included in the analysis (column C, G, L; Y – Yes, N – No). It further includes a column to give the reason why a paper or dataset was excluded from analysis a), b) or c) (column D, H, M).

For analysis a), column E provides the information if the data was published with or linked to a scientific paper (data repository or supplementary material), or published in a data repository without link to a journal article (see figure 2 of the paper publication). Note that journal articles stating ‘data available on request’ were not considered as accessible data and thus were excluded from this analysis. Other reasons for exclusion were that the experiment finished prior to AQUACOSM (i.e. 2017; n=9), no data was generated or available (n=8; e.g. model studies), the DOI link was not working (n=1), or to avoid duplication, i.e. that the paper was linked to other datasets which had already been included in the analysis (n=29). Journal articles that did provide some data in the paper or their supplementary material were included. Journal articles that contained data in the supplementary material and in a data repository were considered a single item and categorised as data in repository. The location of the publication is given in column F (see figure 1).

For analysis b), the duration of dataset publication was determined as the time span (number of months) between the last month of the last experiment and the month of publication of the scientific paper or dataset was used (see figure 3 of the paper publication). Mesocosm facilities within AQUACOSM and AQUACOSM-plus were all large-scale, and experiments were clearly defined with a start and end date, either provided in the metadata of datasets or the method section of journal articles for most publications. For most experiments, the end date of an experiment coincides with the last day of sampling. For repositories, the dataset publication date was taken from the date the data were uploaded into the repository and the end of the experiment was determined by manually checking though all dataset metadata and sometimes the data itself. For journal articles, the publication date was obtained from the first page of the journal article or the details given on the journals’ webpage and the methods section of the journal articles. Column I contains the publication month and year and column J provides the last month and year of the experiment, while column K provides the difference (in months) between column I and J, given as Months Lag. The time lag analysis included journal articles where data were made available via open access as well as via direct request. Experiments which ran before AQUACOSM started (i.e. 2017) were removed from the time lag analysis, even though the data and/or journal article was published during the run-time of AQUACOSM or AQUACOSM-plus. Journal articles not reporting on experimental results (e.g. meta-analyses) were excluded from the time lag analysis, along with articles where no experimental dates were given. If a single dataset was used for multiple scientific papers, the time lag for each paper was determined individually. A single scientific paper was considered a single item (i.e. time lag) even though it might make use of multiple datasets in repositories. Where there was one scientific paper that used multiple datasets, this was counted as a single time lag and the datasets which were not used are labelled “1 paper multiple datasets” in the table.

For analysis c), we determined if data from open access scientific paper publications were published and were published FAIR. We adopted the DATA (Discoverable, Accessible, Transparent and Actionable) scoring criteria from Van Tuyl & Whitmire (2016) (see their Table 1) for FAIR principles (see Table 1 below) and manually scored each paper accordingly (column N to Q). Each article was scanned for information about shared data in the acknowledgements and supplementary material, and any data links were followed, and data were downloaded and checked for completeness (compared to figures in the results section). To the best of their ability, the same person scored all scientific papers. The use of shared vocabulary was determined as usage of common vocabulary by the authors, however uncommon or unclear parameters or abbreviations were checked with EnvThes (controlled vocabulary developed in the context of LTER Europe for ecological research, monitoring and experiments; <https://vocabs.lter-europe.net/envthes/en/>). All scores were summed to determine how many scientific paper publications reached the highest scores of 7 and 8 (column R). Papers which were not part of AQUACOSM and papers which did not include any data were excluded along with dataset publications.

Table 1: Scoring criteria for the effectiveness of data sharing from Van Tuyl & Whitmire (2016) (see their Table 1), adopted for Journal Article Scoring of shared data under the FAIR principles.

|  |  |
| --- | --- |
| Findable |  |
| 0 | No link or indication of data source in the paper OR non-actionable mention of data location (e.g. broken link, mention of source without link) OR no raw data sharing |
| 1 | A reference to the location or source of the data but no specific indication of the data used (e.g. link to researchers’ home page or external database) OR data shared are not all of the data used in the paper OR summarised data instead of raw data is presented. |
| 2 | A direct link to a dataset with a persistent identifier (e.g. DOI) OR all data provided in supplementary material OR link to OA genetic database |
| Accessible |  |
| 0 | Data shared through closed or subscription access platform or accessed by request OR no raw data sharing |
| 1 | Data shared through a platform that requires some barrier to access, e.g. obtain permission to use the data or closed-access source (e.g. journal, repository) |
| 2 | Data shared in an open repository or platform or source (e.g. supplementary material of OA scientific publication) |
| Interoperable |  |
| 0 | No documentation/metadata provided for the data OR no raw data sharing |
| 1 | Some documentation/metadata provided for the data but lacks detail (e.g. how data was collected, analysed/processed, description of units etc) OR metadata provided in journal article OR terminology cannot be verified |
| 2 | Detailed metadata or read me file; using shared vocabulary (i.e. is the vocabulary accessible for humans and machines) |
| Reusable |  |
| 0 | Data are not in a format that is usable in an analysis application (e.g. PDF or Figure) OR no raw data sharing |
| 1 | Data are in a usable format in an analysis application but are formatted in a way that makes use difficult OR data are shared in a non-open format (e.g. .xls, .doc, .sas, .mat, .shx) OR data in open format but large proportion of data missing |
| 2 | Data are in an open or non proprietary format (e.g. .csv, .xlsx, .txt, etc) with usable formatting OR genetic data deposited in OA genetic database |

Reference:

Van Tuyl S, Whitmire AL. 2016. Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia. PLoS one E 11(2): e0147942.