Science & Society

TRANSPARENT PROCESS OPEN ACCESS

EMBO reports

Acknowledging and citing core facilities

Key contributions to data lifecycle should be recognised in the scientific literature

Katja Kivinen[1,2], Henri GAM van Luenen[1,3], Myriam Alcalay[1,4], Christoph Bock[1,5,6], Joanna Dodzian[1,7], Kateřina Hošková[1,8], Danielle Hoyle[1,9], Ondřej Hradil[1,8,10], Sofie Kjellerup Christensen[1,11], Bernhard Korn[1,12], Theodoros Kosteas[1,13], Mònica Morales[1,14], Krzysztof Skowronek[1,7], Vasiliki Theodorou[1,13], Geert Van Minnebruggen[1,15], Jean Salamero[16,17] & Lavanya Premvardhan[1,17,*]

Core facilities are an integral part of the life science research landscape as providers of centralised access to technological resources and expertise – such as equipment, databases, materials and organism collections, and other research-sustaining resources. The centralised operational model, spanning both physical and virtual sites, allows core facilities to pool finances and invest in expensive technologies and skilled staff with relevant expertise. In contrast to large Research Infrastructures (RIs) supported at the regional, national or international level, core facilities are primarily supported by their host institution whom they serve. By providing technological services and expertise, and through their support of research and the training of scientists, core facilities and their staff have a strong impact on the scientific performance and output of their host organisation. Their contribution to scientific research and innovation must thus be accordingly recognised.

The most straightforward way to recognise a core facility's contribution to research is an acknowledgement in all publications and other forms of dissemination that use data originating from the core facility. Acknowledgement for the use of this data is akin to the rights of co-authorship and arguably an ethical obligation. While co-authorship in life science journals is well defined, e.g. by ICMJE (https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html), there are no set rules that are systematically applied for acknowledgements. Although one may refer to published guidelines, e.g. by the Association of Biomolecular Resource Facilities (ABRF), on when to acknowledge core facilities versus when to include core facility staff as co-authors, the lack of formal rules still leaves core facilities being inadequately recognised. Journals are in a key position to help resolve this quandary and instate rules—similar to co-authorship rules—to acknowledge, and rightfully recognise, a core facility's contribution.

This practice will lead to a reliable acknowledgement 'count' that will have two key positive consequences: on the sustainability of core facilities and on their staff careers. In the absence of a high number of publications, particularly as lead or corresponding authors, acknowledgements are used as a measure of a core facility and its staff's output and impact. Funders and host organisations who use the acknowledgement 'count' as a Key Performance Indicator (KPI) will then be ensured of an accurate basis for attributing funds, which in turn impacts a core facility's ability to invest in equipment and salaries. Second, the acknowledgement of a core facility further motivates and incentivizes core facility staff to actively contribute to scientific research, notably when individual staff are acknowledged for their assistance. It is thus all the more important to make the acknowledgement count a reliable KPI, in particular because core facility staff use this indicator for their career advancement.

1 EU-LIFE Core Facilities Working Group, EU-LIFE Alliance, Barcelona, Spain
2 Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
3 The Netherlands Cancer Institute, Amsterdam, The Netherlands
4 IEO, European Institute of Oncology IRCCS, Milan, Italy
5 CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria
6 Center for Medical Data Science, Institute of Artificial Intelligence, Medical University of Vienna, Vienna, Austria
7 IIMCB, International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland
8 CEITEC MU, Brno, Czech Republic
9 Babraham Institute, Cambridge, UK
10 Masaryk University, Brno, Czech Republic
11 EU-LIFE Alliance, Barcelona, Spain
12 Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland
13 Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology Hellas (IMBB-FORTH), Heraklion, Greece
14 CRG, Center for Genomic Regulation, Barcelona, Spain
15 VIB Core Facility Program, VIB, Ghent, Belgium
16 CNRS, INRIA, Rennes Bretagne Atlantique, Rennes, France
17 Institut Curie, PSL Research University, Paris, France
*Corresponding author. E-mail: premvard@gmail.com; lavanya.premvardhan@curie.fr

DOI 10.15252/embr.202255734 | EMBO Reports (2022) 23: e55734 | Published online 23 August 2022

Thus, all research-resource providers that follow the core facility model, with shared technological resources, need to be recognised for their contribution to scientific research. We propose that this recognition of core facilities be deployed via two actions and implemented in two phases: first, with the systematic acknowledgement of core facilities in all scientific publications, and second, by including core facilities and their staff in data citations (Cousijn *et al*, 2018).

> *"By providing technological services and expertise [. . .] core facilities and their staff have a strong impact on the scientific performance and output of their host organisation."*

The first step can be accomplished at the manuscript-submission stage by asking the corresponding author to confirm if any data (and associated metadata) used in the manuscript originated from a core facility, and if yes, to identify the associated core facility. EMBO Press has recently included a question in the author checklist to confirm whether the work in the publication "benefited from core facilities" and that the core facility be acknowledged accordingly. To our knowledge, this is the first life science journal to do so; an action that we fully endorse.

The next step would be to make it compulsory for authors to respond to such a query and explicitly identify the core facility and relevant data (and associated metadata). The MDAR (Materials, Design, Analysis, Reporting) form (Macleod *et al*, 2021), wherein one needs to provide information about data availability in the Analysis section, could likewise include a question to explicitly identify core facilities involved. Eventually, the information in the author checklist could be *automatically* fed into the acknowledgement section.

The acknowledgement of a core facility goes beyond professional courtesy: identifying the origin of data (and associated metadata) is essential for data traceability and reproducibility particularly since core facilities are major generators of data in life science research; indeed, data is their *leitmotif*. Our working group has estimated that between 40 and 80% of imaging, proteomics and genomics data at our institutes are generated at core facilities. This lack of precision, from an individual institution or in the literature, is itself due to the fact that the role of core facilities in data generation is not sufficiently recognised, documented and acknowledged.

The exponential growth of data during the past two decades has created an increasing need for transparency and traceability of research data (https://www.cbd.int/abs/) to ensure reproducibility and reliability. In response, many journals incorporated the use of the MDAR form (Macleod *et al*, 2021), specifically conceived for experimental studies in the life sciences. Whereas research journals act as de facto 'guardians' of data findability and accessibility at the manuscript-submission stage, authors and institutions implement Research Data Management (RDM) plans to ensure that data in their manuscript is FAIR (findable, accessible, interoperable and reusable). However, the extent of core facility involvement in data generation is not always clear: not only do core facilities generate the data used in numerous research articles, but they are also responsible for the quality control (QC) of raw data, and, depending on the technology and available expertise, staff may also be involved in data analysis and interpretation.

The contribution and responsibility of core facilities in the data cycle can thus be progressively identified at three key stages: raw data acquisition at, or by, a core facility; data processing, data analysis and interpretation from initial treatment up to the creation of figures; and data archiving and sharing of raw and processed data.

> *"In the absence of a high number of publications, particularly as lead or corresponding authors, acknowledgements are used as a measure of a core facility and its staff's output and impact."*

Knowing the precise contribution of core facilities at these different stages would ensure greater data traceability and transparency. This process may be formalised in the second phase, for which scientific journals can provide the necessary impetus. For example, the cited core facility could complete a core facility-specific checklist (see Box 1). A more robust solution would be to trace and link a data set, or the metadata, to an individual core facility. To do so, one could use a formal data citation, a dataset DOI (https://datacite.org/cite-your-data.html), a data accession number, or a link to a data set (the latter is linked directly from the figure and not within the text as for data citations). In fact, several scientific publishers have previously endorsed a data-citation roadmap (Cousijn *et al*, 2018), which is currently used by EMBO Press and a few others. We recognise the added value of data citations in identifying staff involved in data acquisition, and not just the core

Box 1. Core facility checklist at article submission stage

- Core facility ID: Name in English and PID if available
- Certification if any (at time of data acquisition)
- Inter/national RI: Yes/No (If yes, which one)

QC of raw data

- Was instrumentation/equipment used to acquire referenced data optimised: Yes/No (If yes, identify instrument RFID, PID . . .)
- Data acquired by core facility staff: Yes/No (If yes, can you confirm QC of samples. Explanation optional).

QC of data analysed (and steps)

By core facility staff or non-core facility staff

- Initial data anaylsis: Yes/No
- Analysis for data used in figure(s): Yes/No (If yes, indicate analysis and figures involved)

Data storage and archiving

- At the facility/organisation level: raw/meta data, local/cloud server, archived for how long, access, etc.
- At an inter/national dB archiving facility: Yes/No (If yes, which one, with link to site).

Box 2. Persistent Identifiers (PIDs)

Various RIs have endorsed the use of PIDs to ensure data traceability and accurate citation of the resource provider, and one may freely obtain a PID from, e.g. RRID, ROR or WOLF. Note however that there is a lack of interoperability between these PIDs. We nevertheless support the use of PIDs and, as recommended by project FREYA for "mature" entities, suggest that it becomes a requirement for core facilities that are part of inter/national RIs.

facility, as is the case for a data DOI. We thus recommend implementing full data citations, which not only ensures FAIR data, but importantly also identifies, and recognises, the scientific contribution of core facility staff by crediting them with a full literature citation for the data in the publication.

"The acknowledgement of a core facility goes beyond professional courtesy..."

To earn a citation, core facilities would be responsible for uploading the relevant metadata to curated data repositories, such as those listed by EMBO Press journals. However, the roles and responsibilities of different parties in the data lifecycle, and the incumbent financial burden, should be covered in an RDM plan set up at the institutional level. A full data citation would also bypass the need to find a unique PID, which we nevertheless strongly recommend for 'mature' core facilities (see Box 2). With the impetus provided by key stakeholders, including major publishers and funders, the uptake of data citations in publications will be more widespread.

Acknowledging and recognising the contribution and responsibility of core facilities thus goes well beyond professional courtesy, to improve research quality (Kos-Braun *et al*, 2020) and to contribute to implementing the Horizon Europe objectives, specifically its data roadmap and EOSC initiative.

Acknowledgements

The authors from the EU-LIFE Core Facilities Working Group (https://eu-life.eu/research-excellence/working-groups-task-forces/core-facilities) acknowledge the support of EU-LIFE (www.eu-life.eu). EU-LIFE is an alliance of research institutes whose mission is to support and strengthen European research excellence by acting as a voice for researchers and research institutions in European science policy and by promoting and sharing good institutional practices in scientific research and the research ecosystem across Europe. The commentary has received the endorsement of the Core for Life Network (https://coreforlife.sites.vib.be/en) and by the France BioImaging Research Infrastructure (https://france-bioimaging.org/).

Disclosure statement and competing interests

Members of the EU-LIFE Core Facilities Working Group—in their professional capacity as managers, coordinators or directors of core facilities—, declare their vested interest in this commentary.

References

Cousijn H, Kenall A, Ganley E, Harrison M, Kernohan D, Lemberger T, Murphy F, Polischuk P, Taylor S, Martone M *et al* (2018) A data citation roadmap for scientific publishers. *Sci Data* 5: 1–11

Kos-Braun IC, Gerlach B, Pitzer C (2020) A survey of research quality in core facilities. *eLife* 9: e6221

Macleod M, Collings AM, Graf C, Kiermer V, Mellor D, Swaminathan S, Sweet D, Vinson V (2021) The MDAR (Materials Design Analysis Reporting) Framework for transparent reporting in the life sciences. *Proc Natl Acad Sci USA* 118: e2103238118