COMMENTS OF THE ACM EUROPE TECHNOLOGY POLICY COMMITTEE
ON DRAFT IMPLEMENTING LEGISLATION FOR DIRECTIVE (EU) 2019/1024
LAYING DOWN A LIST OF SPECIFIC HIGH-VALUE DATASETS

The Association for Computing Machinery (ACM) is the world's largest and longest estab-lished professional society of individuals involved in all aspects of computing. It annually bestows the ACM A.M. Turing Award, often popularly referred to as the "Nobel Prize of computing." ACM's Europe Technology Policy Committee ("Europe TPC")[1] is charged with and committed to providing objective technical information to policy makers and the general public in the service of sound public policymaking. ACM and Europe TPC are non-profit, non-political, and non-lobbying organizations. Europe TPC is pleased to submit the following Comments[2] in response to the Commission's above-captioned consultation, "Open data – availability of public datasets"[3]:

Capsule Conclusion and Recommendation

While established "FAIR" principles[4] are used as a structure to define open access in the Com-mission's policy Directive (EU) 2019/1024 on open data and the re-use of public sector informa-tion, key aspects of these principles are missing from the proposed implementing Act, particu-larly provisions addressing challenges to the findability and accessibility of data. Europe TPC respectfully recommends that this omission be addressed in the final proposed Act. Europe TPC also notes that the Commission recognizes that *adherence to FAIR principles alone will not by itself assure that critical data sets will be truly "open" in all manners intended and appropriate.* We therefore also urge the Commission to clarify to what extent FAIR principles must be applied in this context.[5]

---

[1] See https://europe.acm.org/europe-tpc.

[2] The principal author of this document for Europe TPC was Committee Member Julie Williamson, Senior Lecturer in Computer Science at The University of Glasgow. Also contributing were: Chair Chris Hankin, Fellow of the Institute for Security Science and Technology and Professor of Computing Science at Imperial College London and Committee Member Gerhard Schimpf of SMF Team. *(All affiliations listed for personal identification purposes only.)*

[3] See DRAFT "Commission Implementing Regulation . . . laying down a list of specific high-value datasets and the arrangements for their publication and re-use," pertaining to Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. [https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12111-Open-data-availability-of-public-datasets_en]

[4] See FAIR principles for more detail.

[5] For example, the 'A' in FAIR stands for 'Accessible under well-defined conditions. There may be legitimate reasons to shield data and services generated with public funding from public access. These include personal privacy, national security, and competitiveness. The FAIR principles, although inspired by Open Science, explicitly and deliberately do not address moral and ethical issues pertaining to the openness of data. In the envisioned Internet of FAIR Data and Services, the degree to which any piece of data is available, or even advertised as being available (via its metadata) is entirely at the discretion of the data owner.

"FAIR" Analysis and Application

Findability

The final Act should require that data and metadata be assigned a globally unique and consistent identifier, such as a Digital Object Identifier (DOI), Universal Resource Name (URN), or Archival Resource Key (ARK). The findability principle also requires that data and metadata be appropriately indexed, allowing potential users to find datasets using meaningful queries. The requirements for a globally unique and consistent identifier, along with existing recommendations for rich metadata within the Act and Annex, will improve the findability of datasets made available under this policy. Requiring that all datasets be indexed by an appropriate platform also should be codified.

Accessibility

The accessibility of open data depends not only on using recognised open and machine-readable formats, but also requires rigorous archiving practices to ensure its long-term availability. The Act makes clear requirements for formats and application processing interfaces but does not impose any clear requirements on long term archiving practices. Preservation services like CLOCKSS and Portico ensure digital archives are protected against loss in the event that the original publishing organisation fails to preserve digital content. At a minimum, FAIR principles encourage that metadata remain consistently available even if the original data can no longer be archived. The proposed Act should fully reflect this principle.

Interoperability

Interoperability is discussed in the preamble of the Act in the context of directives relating to specific areas, such as spatial data, but the Act and its appendices do not address this issue explicitly. They should do so as appropriate standards are critical to ensuring interoperability.

Reusability

The goal of the FAIR principles is to optimize the reuse of data. To achieve this: 1) metadata and data should be described precisely so that they can be replicated and/or combined in different environments; and 2) provenance of the data, a key issue in assuring data reusability, should be made a clear requirement for metadata in the Act's Annex.

Conclusion

ACM's Europe Technology Policy Committee stands ready to leverage the expertise of its thousands of European members to assist the European Commission in its further consideration of access to high value data sets in this proceeding, or otherwise with respect to technical matters implicating any aspect of computing and its societal impacts. To request such technical, apolitical input please contact ACM's Director of Global Policy & Public Affairs, Adam Eisgrau, at acmpo@acm.org or +1 202.580.6555.