



Characteristics of Data Relevant to Access Conditions

By

Paul F. Uhlir

Director, Board on Research Data and Information

National Academy of Sciences, US

puhlir@nas.edu

March 2009



Data Characteristics

The amount and nature of data availability depends on a number of factors particular to the relevant scientific community and its practices. These characteristics may be juxtaposed in a way that helps analyze whether they are appropriate or beneficial for broad open access, or not. However, these pairings are not mutually exclusive and in many cases form a continuum. They also typically should not be considered by themselves, but in conjunction with all the other relevant characteristics in such an analysis.

- **1) interdisciplinary** vs. single discipline research;
- **2) large research projects** vs. small research projects;
- **3) the level of public investment** vs. private investment;
- **4) whether the data** have other non-IP statutory restrictions or **are otherwise open**;
- **5) whether the research is highly specialized or the data are easy to interpret by others**;
- **6) the data have quality controls** or are have poor or unknown quality or accuracy;
- **7) the data are generated by communities with longstanding practices of sharing** or do not practice sharing norms;
- **8) whether the data are considered competitive or non-competitive**;
- **9) whether there is an institutional mandate for data preservation**;
- **10) whether there is funding for long-term data preservation** vs. no preservation funding;



Data Characteristics (cont.)

- 11) whether the data are appropriately documented for broad use or not;
- 12) whether there are foreseeable future uses for the data or not;
- 13) whether institutions (e.g., tenure committees) value data work or not;
- 14) whether the data are collected for non-scientific purposes (e.g., marketing) but are useful for scientific research, or not;
- 15) whether data are intended for narrow/local or international research;
- 16) the openness of the format or data structure used for storing information;
- 17) organized opposition to openness or favorable to openness;
- 18) centralized or decentralized data sources;
- 19) whether the data compromise confidentiality requirements or not;
- 20) real-time vs. retrospective data (or longitudinal vs. single event);
- 21) experimental vs. observational data;
- 22) automated sensor vs. PI-generated;
- 23) human vs. non-human subjects (lack of privacy concerns);
- 24) spatially referenced vs. not;
- 25) data product is copyrightable (or has copyrightable elements) or not
- 25) whether the data have national security implications or not
- 26) other...