Data Sharing Principles in Developing Countries
(The Nairobi Data Sharing Principles)

CODATA Task Group of Preservation of and Open Access to S&T Data in/for/with Developing Countries (CODTA-PASTD)

All public funded data and information have public good characteristics; they cannot be depleted with use and can be inefficient to exclude others from access. The value of data and information generally increases with their reuse by others. Public data and information generated by governments and by activities outside government that are publicly funded, including research, are especially inefficient to exclude from access. This is because both the data or information, and the public-sector activities that generated them, are undertaken with public money and in the public interest.

Data, or factual information that has been created or collected in a structured database or compilation of information, are particularly valuable to share (i.e., make openly available) on digital networks. Since the advent of the internet, many studies have shown that open data online have economic, social, educational, and research values that can greatly benefit society and the progress of science. The data and information policies and management have not kept pace with the technological changes, however.

Less developed countries and those with economies in transition have much to gain from the formation of open data policies in the public sector. In addition to the general values that are inherent from open access and reuse, schools, universities, research organizations, governments and all of societies in the developing world can improve governance and decision making, empower and educate citizens, promote capacity building, and generally increase the opportunities for innovation and the return on public investments. It is not possible for nations in the developing world to reach the post-2015 U.N. Sustainable Development Goals, to play an equal role in international cooperation programs, and to close the digital divide without such data and information policies. Moreover, governments, foundations, and in some cases even the private sector should open their data as much as possible for the public welfare of users in developing countries. All these justifications and issues are elaborated in greater detail in the accompanying Guidelines.

It is for these reasons that we, the participants in the International Workshop on Open Data for Science and Sustainability in Developing Countries, agree on the following ten Data Sharing Principles in Developing Countries which we also refer to as "The Nairobi

Data Sharing Principles":

(1) Data should be open and unrestricted

Data generated with public support, including private foundations, should be openly accessible and subject to unrestricted (re)use, absent specific, justified reasons to the contrary (see Principle 10). Openness is especially beneficial for development purposes and research uses, but can benefit all society equally and have a multiplier effect on the economy.

(2) Data should be free to the end users

In most cases, any cost for access is an insurmountable barrier to users in the developing world. Therefore, data should be free online to the user. In some special cases, access to data may be no more than the marginal cost of filling a user request. At the same time, it is recognized that adequate preparation and open availability of data require sufficient financial support (see Principle 7).

(3) Data should be informative and assessed for quality

Data should be of known quality and integrity, and should be organized and described (with metadata) in datasets sufficient to allow them to be understood and effectively (re)used by others. Baseline technical and management standards need to be established, especially in the developing world where state-of-the art practices are not yet as prevalent. Adequate preparation and the use of non-proprietary software are especially important for any datasets expected to have long-term value.

(4) Data sharing should be timely

Once datasets are sufficiently informative and quality controlled, they should be released as quickly as possible. This can be done in steps, starting with the metadata to avoid duplication. In some cases, such as public emergencies and disasters, open release of relevant data should be an immediate priority. In other cases, such as research, data should be openly available no later than upon the publication or patenting of results. Users in developing countries have the most to gain from such policies.

(5) Data should be easy to find and access

Upon the public release of any dataset, the provider should promote ease of access by the broadest user base. Diverse means of publication should be considered in recognition of potential connectivity and other technological challenges.

(6) Data should be interoperable, when necessary

If data from a dataset are likely to be combined with data from one or more other datasets (e.g., in geospatially referenced research), special attention should be given to make such data technically, semantically, and legally interoperable.

(7) Data should be sustainable

The life-cycles of any datasets that are expected to be reused by others should be planned at the outset with support sufficient to successfully implement the first six Principles. The lower availability of funding in developing countries, especially for long-term preservation, makes this a key priority so that valuable datasets remain intelligible and are not lost or in need of rescue. Cost recovery for data archiving and availability should not be borne by the users, consistent with Principle 2, but by other entities in the data lifecycle.

(8) Data contributors should be given credit

A significant incentive for the open disclosure and "publication" of a dataset is the ability

to properly cite and attribute the contributor(s), whether internal or external to an organization. Any subsequent user of the data has at least an ethical obligation—and possibly a legal one—to cite and attribute the source of the data whenever they are reused, and not to misuse the data in any way. Such practices can also improve the integrity of the data sets made available by the contributors, in support of Principle 3. Data contributors in the developing world require greater recognition and rewards for such disclosure, and this should become common practice. A persistent digital identifier, attached to the dataset online, is the best way to promote this goal.

(9) Data access should be equitable

Open access and use of data in developing countries, especially for public purposes, should be supported by the governments and institutions in the more economically developed nations. Capacity building of essential experts and infrastructure in developing countries should be a priority of international organizations. Similarly, experts in developing countries should join and actively participate in the relevant regional and international organizations.

(10) Data may be restricted for a limited time, if adequately justified

Restrictions may be placed on access to and uses of publicly funded data and datasets for specified periods of time. Justified restrictions may include specific protections of national security, personal privacy, intellectual property, confidentiality, and other values, such as indigenous peoples' rights or location of endangered species. Nevertheless, the default rule should be one of openness, consistent with Principle 1, and any restrictions should be minimized to the extent possible.

August 8, 2014, Nairobi, Kenya

Background:

CODATA Task Group on Preservation of and Open Access to S&T Data in/for/with Developing Countries (CODATA-PASTD for short) was approved to be established by the CODATA General Assembly in 2002. This Task Group has organized an impressive series of research and training workshops in a host of developing countries (South African, China, Brazil, Mongolia, Cuba, Colombia, Kennya and Indian) over the last fifteen years. In August 2014, CODATA-PASTD cooperated with and supported by UNESCO, Kenyan Ministry of Information and Communication and number of international organizations to held the International Workshop of Open Data for Science and Sustainability in Developing Countries (Nairobi SSDC Workshop for short), which was hosted at the United Nations Offices in Nairobi. More than 150 worldwide experts, especially from African, jointed the workshop. They were representatives of the following organizations: CODATA Task Group on Preservation of and Access to S&T Data in/for/with Developing Countries (CODATA PASTD), Communication and Information Committee, World Federation of Engineering Organization (WFEO-CIC), Ministry of Communication and Information of Kenya (MCIK), United Nations Educational, Scientific and Cultural Organization (UNESCO), Jomo Kenyatta University of Agriculture and Technology, Kenya (JKUAT), International Council for Sciences Regional Office for Africa (ICSU), World Data System of the International Council for Sci-

ences (ICSU-WDS), The Group on Earth Observations (GEO), International Cartographic Association (ICA) Commission on Geoinformation (GI), International Society for Photogrammetry and Remote Sensing (ISPRS), International Society of Digital Earth (ISDE), Research Data Alliance (RDA), CODATA-China (Chinese Academy of Sciences, Beijing) and USA National Committee for CODATA.

The Data Sharing Principles in Developing Countries (or The Nairobi Data Sharing Principles) is one of the outcomes of the Nairobi SSDC workshop. The final version was completed by Uhlir, Paul, Liu, Chuang and Hodson, Simon based on the workshop discussions and comments from broad communities of the world.

The Nairobi SSDC Workshop participates include: Dr. Matiang'I, Fred (Kenya Cabinet Secretary for the Ministry of Education, Science and Technology in Kenya); Gu, Xingfa (CODATA-PASTD); Gong, Ke (WFEO); and Odido, Mika (UNESCO) the co-chairs of the workshop scientific committee; and Liu, Chuang (CODATA PASTD and CODATA China); Imbuga, Mable (CODATA PASTD and JKUAT); and Uhlir, Paul F. (CODATA PASTD and US CODATA), the co-chairs of the Program Committee; Waweru Mwangi from JKUAT, chair of the Local Committee. More representatives of co-organizers are: Djelid, A Mohamed (UNESCO); Minster, Jean-Bernard (WDS/ICSU); Madela-Mntla, Edith (ICSU African); Wittenburg, Peter (RDA); Hodson, Simon (CODATA International); Hasan, Nordin (ICSU Asia); Chen, Xuehong (ISPRS); Zhou, Xiang (CODATA-PASTD); Kwast, Johannes van der (UNESCO); Tikunov, Vladimir (ICA); Kurban, Alishir (ICA); Wang, Changlin (ISDE); Zhu, Yunqiang (ICSU/WDS); Tomoko Doko (CODATA-PASTD); Chen, Wenbo (CODATA-PASTD); Laughton, Paul (ICSU/WDS); Noasila-laonomenjahary, Ambinintsoa (GEO); and more from China, Germany, Japan, Kenya, Malaysia, Madagascar, Netherland, Nigeria, Russia, Rwanda, South African, Tanzania, Togo, Uganda and USA.