

Challenge and Solution for Research Data Citation

Liu, C.^{1*} Ma, J. H.¹ Uhlir, P. F.² Shi, R. X.¹

1. Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;

2. Consultant on Information Policy and Management, New York 12723, USA

Abstract: The research data as an independent academic output raise the awareness of the need for proper data citation. Data citation in an open and standard form is also in accordance with the “full and open” data policy endorsed by many international organizations. Nowadays some data citation principals have been proposed, however, the implementations of these data citation principals are still rare. Challenges for data citation include using data without citation or citing data without standards. One of the solutions for data citation is making the data citation standardization. For practice in Global Change Research Data Publishing & Repository (GCdataPR), the solution is proposed by two ways for data citation with identifying datasets into 5 classes. The 5 classes include datasets published by certified publishers, datasets published by Word Data System (WDS) regular members, datasets published by national data centers or platforms, datasets published by voluntary individuals and teams and datasets are for internal use only. There are two ways for data citation: reference and footnote. The first three classes of datasets should be cited in form of reference, and the last two classes of datasets should be cited in form of footnote. In addition, the guide is explained in this paper for citing dataset, data record, and data paper.

Keywords: research data; data citation; reference; footnote; standardization

1 Challenge of Data Citation

Research data are both the basic input (resources)^[1] and one of the major outputs in a scientific research life cycle. Scientific data can also be the indispensable evidence for the scientific discoveries. With the advent of coming scientific data publishing, more data will be publicly available^[2]. Once these data are reused by other persons, should they be cited? If so, how to cite these data properly?

1.1 Status and Problems of Data Citation

At present, three approaches are applied in academic papers or monographs to cite the published data: (1) allocate an independent section for data source introduction; (2) explain data sources in “Acknowledgements”; and (3) introduce data sources in “Author Contributions”.

1.1.1 Data Citation Status 1: Allocate an Independent Section for Data Source Introduction

The most popular ways for most article authors to explain the data source is using a short

Received: 10-05-2018; **Accepted:** 01-06-2018; **Published:** 25-06-2018

***Corresponding Author:** Liu, C. L-3684-2016, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, lchuang@igsnr.ac.cn

Author ID: Liu, C. L-3684-2016; Ma, J. H. S-9830-2017; Uhlir, P. F. G-6246-2016; Shi, R. X. L-4389-2016

Article Citation: Liu, C., Ma, J. H., Uhlir, P. F., *et al.* Challenge and solution for research data citation [J]. *Journal of Global Change Data & Discovery*, 2018, 2(2): 123–134. DOI: 10.3974/geodp.2018.02.01.

paragraph or section.

Example 1^[3]:

Data source:

- ① The spatial extent of the Economic Park (EP) is based on the built-up land of 16 cities, interpreted from Landsat TM images of 1990, 1995, 2000, 2005, and 2010;
- ② The social-economic data of EP are from “Economic Park Year Book (1989–2015)” ;
- ③ The city social-economic data are from “China City Year Book (1985–2015)” .

Example 2^[4]:

Data source:

- ① Land use and cover of Xiong’an in 2015 (30 m) , from <http://www.resdc.cn>, used for evaluating environmental quality;
- ② ASTER GDEM (30 m), from <http://www.gscloud.cn/>, used for soil conservation analysis;
- ③ Meteorological Data (2015), from <http://data.cma.cn/>, used for water and soil conservation analysis;
- ④ The social-economic data of Xiong’an, from Year Books of relevant three counties, used for food yield estimation;
- ⑤ Open Street Map, from <https://www.openstreetmap.org/>; VIIRS/DNB (August, 2015) from <https://ncc.nesdis.noaa.gov/VIIRS/>, used for evaluating environmental quality.

1.1.2 Data Citation Status 2: Data Sources are Mentioned in “Acknowledgements”

Example 3^[5]:

“Acknowledgements

We would like to thank Drs. FAN Qiang, LI Wei-le and HUANG Yan for providing various datasets and advices throughout this research.”

Example 4^[6]:

“Acknowledgements

The authors would like to express appreciation to the Pakistan Meteorological Department (PMD) and the Capital Development Authority (CDA), Islamabad, for providing valuable data to conduct this research.”

1.1.3 Data Citation Status 3: Data Sources are Introduced in “Author Contributions”

Example 5^[7]:

“Contributions

W. W., Y. H., D. L., and X. G. designed the study. W. W. wrote the paper. D. L., Y. M., Y. H., P. X., H. D., and X. G. reviewed the paper and contributed to the data analysis. Y. Y., Z. H., and Y. M. helped examine and improving the dataset. Y. Y. and Z. H. collected part of the first version of the 2014 sub-dataset. P. X. and H. D. contributed to the collection of the 1960s and 2005 sub-datasets. X. G. helped to provide the 2014 GF-1 data.”

These examples demonstrate that the current data citation is far from consistent both in form and in content.

2 Open Data for Data Citation

With more scientific data being shared, the calls to recognize the data creators’ contribution to the scientific community and to protect the data intellectual property are getting louder.

Among many feasible steps, standardization of data citation is one of the most important. In 2010, Committee on Data for Science and Technology (CODATA) and International Council for Scientific and Technical Information (ICSTI) formed “CODATA-ICSTI Data Citation Standards and Practices” task group. Over 10 international workshops organized by this task group since then have ushered in the worldwide scientific data citation movement.

Three events occurred in 2014 may herald the coming era of scientific data publishing. They were the kicking-off of the Scientific Data by Springer Nature, the Geosciences Data by Royal Meteorological Society, and the Global Change Research Data Publishing & Repository, which is co-sponsored by Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, and the Geographical Society of China. Although the publishing styles vary for three publishers, one thing is common: data and data paper are published concomitantly. Data papers are published the same way as the other academic articles, thus they are easily citable. Data can be easily located and downloaded through network.

In order to promote the standardized data citation, the Future of Research Communications and *e*-scholarship (FORCE 11) proposed the “Data Citation Principles” in 2014. These Principles consist of eight parts, with its focus on the importance of the data citation. For instance, its Principle 3 states: “In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited”. FORCE 11 aside, other national or international organizations are active to advance the standardized data citation: UK Digital Curation Center, RDA, etc.

In China, two national standards were issued recently which are closely related to data citation. They are: Information and Documentation—Rules for Bibliographic Reference and Citations to Information Resources (GB/T 7714—2015)^[8]; Information Technology—Scientific Data Citation (GB/T 35294—2017)^[9].

Although there are principles about published data citation, more works need to be done to transform abstract principles and guidelines into operational actions.

The CODATA, WDS of International Science Council (ISC), Research Data Alliance (RDA), Group of Earth Observations (GEO), World Meteorology Organization (WMO), etc. release the open data principle. The Cape Town Declaration of UN World Data Forum (2017) indicated that open quality and timely data are vital for enabling governments, international organizations, civil society, private sector and the general public to make informed decisions and to ensure the accountability of representative bodies^[10]. So, open data is a common understanding for the data management.

The data FAIR (Findable, Accessible, Inter-operational, Re-useful) principles^[11] and Nairobi Data Sharing Principles (Data Sharing Principles in Developing Countries, Nairobi, Kenya, 2014)^[12] also emphasize the open data principle. It is obviously, without open data, the data-inter-operational is impossible.

There is a difference between the data citation and article citation. Mostly, the article citation is taking the scientific ideas or results as references, however, the data citation is the data reuse in all or parts directly. In most cases, different results could come from the same datasets. For example, the IPCC reports indicated the different results, because of different simulating models although using the same datasets^[13]. In this progress, more fundamental work is the quality data, which is directly reused in the research. However, without open data, it is impossible to process the data quality checking, dataset peer review, as well as the original intellectual property protection^[14].

No matter in which way, open data should be emphasized for data citation. In some cases, the datasets may not be available to open to all, when those datasets were used, should those dataset be cited? We believe these datasets are also valuable to be cited, but they should be separated from that of open datasets in ways of citation.

3 A Solution in Practice: Two Ways for Data Citation

Unlike the citation of scientific research articles, it is more complex to cite research data. We propose a two-steps approach: first, identify research datasets into five classes; second, choose the proper citation ways for each data class.

3.1 Five Data Classes

After exploration of the publicly available research data, five data classes can be identified to serve the data citation:

- (1) Data published, long-term archived and openly available by certified publishers
- (2) Data published, long-term archived and openly available by Regular Member of WDS
- (3) Data published, long-term archived and openly available by a National Data Center (NDC)
- (4) Data openly available through internet by individuals (or a group)
- (5) Data for internal use only

3.1.1 Data Published by a Certified Publisher

The data falling into this category are formal publication from certified publishers, either publishing house (for academic books) or publishing office (for academic journals). The forms of publication cover datasets, data papers, and data included in or attached to academic papers. Advantages of this class are: (1) all dataset are peer reviewed and data quality should be trustable; (2) all datasets can be long-term archived and managed and low data lost risk; (3) all data are openly available in any possible way.

3.1.2 Data Published by World Data System

The WDS is an interdisciplinary body of the International Council for Science (ICSU) created by its 29th General Assembly in Maputo, Mozambique in 2008, and the mission of WDS is to support ICSU's vision by promoting long-term stewardship of, and universal and equitable access to, quality-assured scientific data and data services, products, and information across a range of disciplines in the natural and social sciences, and the humanities (<http://www.icsu-wds.org/organization>). WDS Regular Members must fulfill the WDS membership requirements and undergo a transparent and objective certification procedure to ensure their trustworthiness, and their memberships will be reviewed on a biannual basis. Data published by WDS Regular Members are internationally recognized, reliable data. In addition, these data are permanently archived and can be accessed through network. Until March 2018, WDS has 72 Regular Members, and another 7 members are under evaluation (Table 1).

3.1.3 Data Published by a National Data Center (NDC)

National Data Centers are certified by specific administrations and tasked with permanent archive and data sharing of certain scientific datasets.

Normally, data archived at National Data Centers fall into one of three classes in terms of publicity (availability): (1) data publically available for all; (2) metadata publically available

for all, but datasets only available for a few qualified persons; and (3) both metadata and datasets are only internally open within certain groups.

Here, Data published by national scientific data centers mainly refer to class (1). Data from class (2) and (3) belong to internal use.

Table 1 List of Regular Member of World Data System

| Member name | Country | Date to be member | Host organization |
|---|--------------------|-------------------|--|
| Incorporated Research Institutions for USA Seismology (IRIS), Data Services | USA | 13-12-2011 | National Science Foundation (NSF), USA |
| WDC-Geoinformatics and Sustainable Development | Ukraine | 13-12-2011 | National Technical University of Ukraine |
| ISRIC-WDC Soils | Netherlands | 13-12-2011 | ISRIC-World Soil Information |
| WDC for Climate | Germany | 13-12-2011 | German Climate Computing Center |
| WDC-Meteorology, Asheville | USA | 13-12-2011 | NOAA National Centers for Environmental Information (NCEI) |
| Center de Donnees astronomiques de France Strasbourg (CDS) | France | 13-12-2011 | University of Strasbourg / CNRS |
| World Glacier Monitoring Service, Zurich | Switzerland | 13-12-2011 | University of Zurich |
| Australian Antarctic Data Center | Australia | 13-12-2011 | Australian Government, Antarctic Division |
| Chinese Astronomical Data Center | China | 13-12-2011 | Chinese National Astronomical Observatories, MOST/CAS |
| WDC-Renewable Resources and Environment | China | 13-12-2011 | Institute of Geographic Sciences and Natural Resources Research, CAS |
| Flanders Marine Institute, Data Center | Belgium | 13-12-2011 | Flanders Marine Institute |
| World Data Service for Oceanography | USA | 13-12-2011 | NOAA National Centers for Environmental Information (NCEI) |
| International Earth Rotation and Reference Systems | Germany | 13-12-2011 | Federal Agency for Cartography and Geodesy, Germany |
| Fish Database of Taiwan (Academia Sinica, Taiwan) | Taipei, China | 13-12-2011 | Academia Sinica Center for Digital Culture |
| WDC-Oceanography, Tianjin | China | 13-12-2011 | National Marine Data and Information Service |
| World Data Service for Geophysics | USA | 13-12-2011 | NOAA National Centers for Environmental Information (NCEI) |
| PANGAEA-Data Publisher for Earth & Environmental Science | Germany | 13-12-2011 | Center for Marine Environmental Sciences, University of Bremen; and Alfred Wegener Institute for Polar and Marine Research |
| WDC-Solar-Terrestrial Physics, Moscow | Russian Federation | 13-12-2011 | Geophysical Center of the Russian Academy of Sciences |
| WDC-Sunspot Index and Long-term Solar Observations (SILSO) | Belgium | 24-02-2012 | Solar Influences Data analysis Center-Royal Observatory of Belgium |
| WDC-Oceanography, Obninsk | Russian Federation | 24-02-2012 | All-Russian Research Institute of Hydro-meteorological Information |
| WDC-Remote Sensing of the Atmosphere | Germany | 24-02-2012 | German Aerospace Center (DLR) |
| WDC-Geomagnetism, Copenhagen | Denmark | 24-02-2012 | National Space Institute |
| International Service of Geomagnetic Indices | France | 24-02-2012 | EOST, Université de Strasbourg/CNRS, France |
| WDC-Geomagnetism, Edinburgh | United Kingdom | 24-02-2012 | British Geological Survey |

(To be continued on the next page)

(Continued)

| Member name | Country | Date to be member | Host organization |
|---|--------------------|-------------------|---|
| WDC for Solid Earth Physics, Moscow | Russian Federation | 24-02-2012 | Geophysical Center of the Russian Academy of Sciences, |
| WDC-Meteorology, Obninsk | Russian Federation | 24-02-2012 | All-Russian Research Institute of Hydro-meteorological Information |
| WDC-Solar Activity / BASS2000 | France | 24-02-2012 | LESIA, Observatoire de Paris |
| WDC-Geomagnetism, Kyoto | Japan | 17-04-2012 | Data Analysis Center for Geomagnetism and Space Magnetism, Kyoto University |
| Interdisciplinary Earth Data Alliance | USA | 26-06-2012 | Lamont-Doherty Earth Observatory of Columbia University. |
| WDC-Space Weather, Australia | Australia | 13-07-2012 | The Bureau of Meteorology, Australian Government |
| NSIDC DAAC | USA | 13-07-2012 | University of Colorado |
| Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) | USA | 20-08-2012 | Oak Ridge National Laboratory (ORNL) |
| World Stress Map Project | German | 20-08-2012 | German Research Center for Geosciences |
| WDC-CU | USA | 04-10-2012 | National Snow and Ice Data Center, University of Colorado |
| WDC-Ionosphere and Space Weather | Japan | 22-11-2012 | National Institute of Information and Communications Technology (NICT) |
| Ukrainian Geospatial Data Center | Ukraine | 26-11-2012 | Space Research Institute of Ukraine |
| Data Center for Geography, Moscow | Russian Federation | 30-11-2012 | Lomonosov Moscow State University, Faculty of Geography |
| WDC-Earth Resources Observation and Science (EROS) Center | USA | 29-01-2013 | USGS Earth Resources Observation and Science (EROS) Center |
| The Language Archive | Netherlands | 29-01-2013 | Max Planck Institute for Psycholinguistics |
| World Data Service for Paleoclimatology | USA | 29-01-2013 | NOAA National Centers for Environmental Information (NCEI) |
| DataFirst | South Africa | 29-01-2013 | University of Cape Town |
| WFCC-MIRCEN World Data Center for Microorganisms | China | 30-01-2013 | World Federation for Culture Collections (WFCC) |
| Goddard Earth Sciences Data and Information Services Center (GES DISC) | USA | 23-03-2013 | NASA/GSFC |
| Crustal Dynamics Data Information System (CDDIS) | USA | 27-03-2013 | NASA GSFC |
| Chinese Space Science Data Center | China | 28-05-2013 | Center for Space Science and Applied Research, CAS |
| Cold and Arid Regions Science Data Center at Lanzhou (CARD) | China | 03-06-2013 | Cold and Arid Regions Environmental and Engineering Research Institute, CAS |
| Global Hydrology Resource Center (GHRC) | USA | 03-06-2013 | NASA |
| Italian Center for Astronomical Archive-IA2 | Italy | 03-06-2013 | INAF |
| Inter-university Consortium for Political and Social Research (ICPSR) | USA | 03-06-2013 | University of Michigan |
| Atmospheric Science Data Center (Distributed Active Archive Center) | USA | 26-06-2013 | NASA |
| WDC-Geomagnetism, Mumbai | India | 28-03-2014 | Indian Institute of Geomagnetism, New Panvel, India |
| Canadian Astronomy Data Center/ Canadian Virtual Observatory | Canada | 15-04-2014 | Canadian Astronomy Data Center/ Canadian Virtual Observatory |
| Alaska Satellite Facility | USA | 19-05-2014 | Geophysical Institute, Fairbanks |

(To be continued on the next page)

(Continued)

| Member name | Country | Date to be member | Host organization |
|---|--------------------|-------------------|---|
| Ocean Networks Canada | Canada | 14-07-2014 | University of Victoria |
| Environment Climate Data Sweden | Sweden | 13-08-2014 | Swedish National Data Service |
| Socioeconomic Data and Applications Center (SEDAC) | USA | 20-04-2015 | Center for International Earth Science Information Network, Columbia University |
| UNAVCO, Inc. | USA | 20-04-2015 | UNAVCO, Inc. |
| Land Processes Distributed Active Archive Center | USA | 26-06-2015 | USGS |
| Permanent Service for Mean Sea Level (PSMSL) | United Kingdom | 02-10-2015 | National Oceanography Center |
| DANS | Netherlands | 06-11-2015 | KNAW |
| Research Institute for Sustainable Humanosphere, Kyoto University | Japan | 17-02-2016 | Research Institute for Sustainable Humanosphere, Kyoto University |
| The Cambridge Crystallographic Data Center | United Kingdom | 29-07-2016 | The Cambridge Crystallographic Data Center |
| Global Biodiversity Information Facility | Denmark | 29-08-2016 | Global Biodiversity Information Facility |
| WDC for Geophysics, Beijing | China | 09-09-2016 | Institute of Geology and Geophysics, Chinese Academy of Sciences |
| Global Change Research Data Publishing and Repository | China | 23-09-2016 | Institute of Geographical Sciences and Natural Resources Research, CAS, Geographical Society of China |
| Canadian Cryospheric Information Network/Polar Data Catalogue | Canada | 11-10-2016 | Canadian Space Agency, Meteorological Service of Canada, Natural Resources Canada |
| Swedish National Data Service | Sweden | 14-12-2016 | Swedish Research Council |
| Center for Astronomical Data of the Institute of Astronomy of the Russian Academy of Sciences | Russian Federation | 09-02-2017 | Institute of Astronomy of the Russian Academy of Sciences |
| Ocean Biology Data Active Archive Center (OB.DAAC) | USA | 26-05-2017 | National Aeronautics and Space Administration |
| Neotoma Paleoecological Database | USA | 01-09-2017 | The Neotoma Community |
| ImmPort Repository | USA | 07-02-2018 | DAIT, NIAID, NIH |
| Worldwide Protein Data Bank (wwPDB) | USA | 19-02-2018 | Worldwide Protein Data Bank (wwPDB) |

Note: The table contents are from: <https://www.icsu-wds.org/community/membership/regular-members>.

Since there are various strict regulations for all NSDCs, including the management of data quality, copyright, data security, etc., it is regarded that data in class (1) have two traits: data are permanently archived and can be accessed through internet; data are reliable.

3.1.4 Data Made Openly Available through Internet by Individuals or Groups

The data within this class are usually up-to-date and professional, but the problem with this kind of data is that there is little or no guarantee for permanent data archive, and the data quality might be questionable due to the absent of peer-review.

3.1.5 Data for Internal Use Only

Out of certain considerations, some data may only be used by selected people or within specific period. The citation way for this kind of data should be different.

3.2 Two Ways for Data Citation

Based on three criteria (the data in question are openly available, permanently archived, and

reliable/peer-reviewed), the aforementioned five types of data can be put into two citation forms: cited in form of reference; and cited in form of footnote.

3.2.1 Data Citation in Creating References

In general, the data cited in reference are the major data sources both for the authors and data reviews. Either of the data falling one of the following group should be cited in reference.

(1) Data formally published by certified publishers

The data cited in this class may come from three forms:

- Digital monograph and datasets published by publishing house;
- Datasets published by digital academic journals specialized in data publication;
- Datasets attached to academic papers published by academic journals.

In China, “certified publishers” means they have ISSN serial number (for international version)¹ or CN serial number (for Chinese version) for their publications.

(2) Data published by World Data System (WDS) Regular Members

The data published by WDS Regular Members should be cited in form of reference. Table 1 provides the detailed information about WDS Regular Members until July 31, 2018.

(3) Data published by National Data Centers

As a follow-up to the UN big data White Paper (*UN Global Pulse [May 2012] Big Data for Development: Challenges and Opportunities*), many countries have been improving their NDCs: NASA-DAACs in the USA, National Science and Technology Infrastructure in China, National Open Science Sharing Network in South Africa. Part of the data published from NDCs can be cited in form of reference.

3.2.2 Data Cited in the Form of Footnote

Two classes of data should be cited in form of footnote: data published by voluntary individuals or groups, and data used internally.

The major shortcomings of the former are the potential problems resulted from the lack of peer-review, and the lack of assurance for long-term data archive. Citation in footnote forms intends to remind the users of the potential problems with data quality and availability.

As for the latter, citation in footnote form informs users there are special data policy for some data, although these data have been used by some authors.

4 Elements and Places of Data Citation in References

If the used or re-used datasets from a publisher or WDS or NDC, they may be cited in the references. Four situations should be dealt with: (1) elements and place of dataset citation; (2) elements and place of data records citation; (3) elements and place of data paper citation; and (4) elements and place of data file citation, while the data file as supplementary information.

4.1 Elements and Places of Dataset Citation

4.1.1 Places of Dataset Citation

If the dataset was used or re-used, the dataset should be cited in the references using [] with

^① ISSN Center of China, National Library, No. 33 South Street, Zhongguancun, Beijing 100080.

the order number.

Example 6^[15]:

“The summaries of the monthly datasets of global marine NPP at 9 km based on MODIS and ABPM model (2003–2012) (GlobalMarineABMP_NPP)^[8] are presented in Table 1, which includes the dataset full name, short name, authors, geographical region, year of the dataset, temporal resolution, spatial resolution, data format, data size, data files, data publisher, and data sharing policy.”

4.1.2 Elements of Dataset Citation

Elements of dataset citation in references should include author(s), title of dataset, publisher or name of serial publication, year of publishing, DOI.

Example 7^[16]:

Sui, C. J., Zhang, Z. H., Wu, H. D., *et al.* Arctic seasonal sea ice statistics dataset (1979–2012) [DB/OL]. Global Change Research Data Publishing & Repository, 2016. DOI: 10.3974/geodb.2016.03.05.V1.

4.2 Elements and Places of Data Records Citation

4.2.1 Both Ways for Citing Data Records

In the data sharing policy of Global Change Research Data Publishing & Repository, if data are used to compile new datasets, the “ten percent principal” should be followed such that data records utilized should not surpass 10% of the new dataset contents, while sources should be clearly noted in suitable places in the new dataset. Two types of citations are usually required, namely, not only to mark the citation position and list in the reference, but also to note the source of the records in the dataset.

4.2.2 Place of Data Records Citation

Data record citation should be marked “*” in the data record and noted in the suitable position of the dataset. If the data is in a table, the note should be placed at the bottom of the table.

Example 8^[17]:

In the statistics of the relief degree of land surface (Rdls) in provinces of China in “Relief degree of land surface dataset of China (1 km)” (DOI: 10.3974/geodb.2018.03.16.V1), one column is the code, which is cited and need mark the “*” and add the note at the bottom of the table (Figure 1).

4.3 Elements and Places of Data Paper Citation

4.3.1 Labeling Place of Data Paper Citation

If a published data paper is cited in an article or other documents (contribution from serial publications of academic journals), the data paper needs to be inserted in the text as numerical or chronological designations (Example 9).

Example 9^[18]:

Labeling place of data paper citation:

Scientific research on global change has thus become one of the most active fields of research in Earth Sciences. International studies on global change started in the 1980s, led by a number of international scientific research programs. The current one is the “Future Earth Plan (Future Earth) (2014–2023)”^[2].

| Province Name | 省名 | CODE* | Rdls |
|---|---------|--------|--------|
| | | 行政编码* | 地形起伏度 |
| Beijing | 北京市 | 110000 | 0.6553 |
| Tianjin | 天津市 | 120000 | 0.0247 |
| Hebei Province | 河北省 | 130000 | 0.6519 |
| | | | |
| Hong Kong Special Administrative Region | 香港特别行政区 | 810000 | 0.8006 |
| Macao Special Administrative Region | 澳门特别行政区 | 820000 | 0.1532 |

*备注: 行政编码引自文献: 中华人民共和国民政部. 中华人民共和国行政区划简册2014. 中国地图出版社. 2014.

*Note: The code was cited from the reference: Ministry of Civil Affairs of the People's Republic of China. People's Republic of China administrative division 2014. SinoMaps Press. 2014.

Figure 1 Example 8 for the place of data records citation

4.3.2 Elements of Data Paper Citation

The elements of data paper citation includes the main creator (authors), title of item and contribution information of documents [media designation], title of serial publication, other title information, year, volume (issue), pages, and digital object unique identifier (DOI).

Example 10^[19]:

Lin, L. N., Wang, Y. J., Zhang, J., *et al.* Hydrologic CTD dataset obtained through Chinese Antarctic expeditions [J]. *Journal of Global Change Data & Discovery*, 2017, 2(3): 157–164. DOI: 10.3974/geodp.2017.02.04.

4.4 Elements and Place of Data File Citation, while the Data File as Supplementary Information

A growing number of academic journals require authors to publish their papers together with associated data files, which are published as supplementary information. The labeling position of data file citation is the same as that of this paper, but the documentation format is different. The documentation format of citation data paper includes the main creators (authors), title of item and contribution information of documents [media designation], title of serial publication, other title information, year, volume (issue), pages, and digital object unique identifier (DOI), inserting position of the text, data format, file name of data, file extension, data size, data acquisition and access path (Example 11).

Example 11^[20]:

Zhang, Q. Q., Ying, G. G., Pan, C. G., *et al.* Comprehensive evaluation of antibiotics emission and fate in the River Basins of China: Source analysis, multimedia modeling, and linkage to bacterial resistance [J]. *Environ. Sci. Technol.*, 2015, 49 (11): 6772–6782. DOI: 10.1021/acs.est.5b00729. Supporting Information, MS Excel: es5b00729_si_001.xlsx (2.43 MB), <https://pubs.acs.org/doi/10.1021/acs.est.5b00729>.

5 Data Citation in the Form of Footnote

References to data published by individuals or teams in the computer network and data for internal use (not publicly published) are cited as footnotes. The footnotes format is automatically generated by the serial number computer in the form of a coil code superscript. In text in PDF data format, footnotes appear at the bottom of the page, keeping them away from the text with a split line (Example 12).

Example 12 (PDF format)^[21]: Please see Figure 2.

Table 3 Proposed strategy for the re-zoning and re-classification of Chinese ecological-populated residential-industrial land

| Land use type | Definition | Data sources |
|----------------------------------|--|--|
| Water conservation land | Sources of major rivers and important areas of water | Ecological function zoning ^① , China's Natural Reserve Areas ^[17] , Land use map |
| Soil conservation land | Key protection and administration of water and soil conservation | Ecological function zoning ^① , National Water and Soil Conservation Planning ^[18] , Land use map |
| ... | ... | ... |
| Urban built-up area | Urban built-up area in the Land use map | Land use map |
| Rural populated residential land | Rural populated residential land in the Land use map | Land use map |

- ① Ministry of Environmental Protection of P. R. China, Chinese Academy of Sciences. National Ecological Function Zone. 2008.
- ② State Forestry Administration of P. R. China. National Sandification Control Program (2011-2020). 2013.
- ③ Vector map of river system at scale of one: four million in China provided by Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences.
- ④ State Forestry Administration of P. R. China, Ministry of Finance of P. R. China. National Public Forest Division delineation. 2009.
- ⑤ WCMC Natural Reserve Areas (<http://protectedplanet.net/>).

Figure 2 Example 12 for data citation in the form of footnote in PDF format

In the HTML format, all footnotes are grouped into the end part of the text in sequence numbers, distancing from the text by a split-line (Example 13).

Example 13 (HTML format)^[21]: Please see Figure 3.

-
- ④ Ministry of Environmental Protection of P. R. China, Chinese Academy of Sciences. National Ecological Function Zone. 2008.
 - ② State Forestry Administration of P. R. China. National Sandification Control Program (2011-2020). 2013.
 - ③ Vector map of river system at scale of one: four million in China provided by Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences.
 - ④ State Forestry Administration of P. R. China, Ministry of Finance of P. R. China. National Public Forest Division delineation. 2009.
 - ⑤ WCMC Natural Reserve Areas (<http://protectedplanet.net/>).

Figure 3 Example 13 for data citation in the form of footnote in HTML format

6 Conclusion

With the coming of big data era, more and more scientific data are published. Standardization of data citation will promote the sustained scientific data sharing by giving credit to data producers (authors, creators) and by protection of intellectual property for the original creators, and improve the data quality by peer review and re-uses. It is of great significance to avoid the phenomenon of data using without citation, of data citation without documentation, of data documentation without standardization.

References

- [1] National Council of P. R. China. Scientific data management law of China [OL]. 2018. http://www.gov.cn/gongbao/content/2018/content_5283177.htm.
- [2] Liu, C. Global change research data publishing [J]. *Acta Geographica Sinica*, 2014, 69(Sup): 3-11. DOI: 11821/DLXB2014s000.

- [3] Sun, Z. X., Luan, W. X., Ma, Y., *et al.* Study on the spatial expansion characteristics of development zones in different coastal areas of China [J]. *Journal of Natural Resources*, 2018, 33(2): 262–274. DOI: 10.11849/zrzyxb.20170052.
- [4] Peng, J., Li, H. L., Liu, Y. X., *et al.* Identification and optimization of ecological security pattern in Xiong'an New Area [J]. *Acta Geographica Sinica*, 2018, 73(4): 701–710. DOI: 10.11821/dlxb201804009.
- [5] Meng, Q. K., Miao, F., Zhen, J., *et al.* Impact of earthquake-induced landslide on the habitat suitability of giant panda in Wolong, China [J]. *Journal of Mountain Science*, 2016, 13(10): 1789–1805. DOI: 10.1007/s11629-015-3734-0.
- [6] Abbas, T., Nabi, G., Boota, W. M., *et al.* Uncertainty analysis of runoff and sedimentation in a forested watershed using sequential uncertainty fitting method [J]. *Science in Cold and Arid Regions*, 2016, 8(4): 297–310. DOI: 10.3724/SP.J.1226.2016.00297.
- [7] Wan, W., Long, D., Hong, Y., *et al.* A lake data set for the Tibetan Plateau from the 1960s, 2005, and 2014 [J]. *Scientific Data*, 2016, 3: 160039. DOI: 10.1038/sdata.2016.39.
- [8] General Administration of Quality Supervision, Inspection and Quarantine of the P. R. China, Standardization Administration of the P. R. China. Information and documentation: rules of references citation: GB/T 7714—2015 [S]. Beijing: Standards Press of China, 2015: 5.
- [9] General Administration of Quality Supervision, Inspection and Quarantine of the P. R. China, Standardization Administration of the P. R. China. Information technology: scientific data citation: GB/T 35294—2017 [S]. Beijing: Standards Press of China, 2017: 12.
- [10] High-level Group for Partnership, Coordination and Capacity-Building for statistics for the 2030 Agenda for Sustainable Development (HLG-PCCB). Cape Town Global Action Plan for Sustainable Development Data [OL]. 2017. <https://undataform.org/WorldData-Form/wp-content/uploads/2017/01/Cape-Town-Action-Plan-For-Data-Jan2017.pdf>.
- [11] FORCE 11. The FAIR Data Principles [OL]. <https://www.force11.org/group/fairgroup/fairprinciples>.
- [12] CODATA PASTD Task Group. Data Sharing Principles in Developing Countries [J]. *Journal of Global Change Data & Discovery*, 2017, 1(1): 12–15. DOI: 10.3974/geodp.2017.01.03.
- [13] IPCC. Climate Change 2014: Synthesis Report [M]. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R. K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 2014: 151.
- [14] GCdataPR Editorial Office. GCdataPR Data Sharing Policy [OL]. DOI: 10.3974/dp.policy.2014.05 (Updated 2017).
- [15] Tao, Z., Zhou, X., Ma, S. Monthly datasets of global marine NPP at 9 km based on MODIS and ABPM model (2003–2012). [J]. *Journal of Global Change Data & Discovery*, 2017, 1(2): 149–156. DOI: 10.3974/geodp.2017.02.03.
- [16] Sui, C. J., Zhang, Z. H., W. H. D., *et al.* Arctic seasonal sea ice statistics dataset (1979–2012) [DB/OL]. Global Change Research Data Publishing & Repository, 2016. DOI: 10.3974/geodb.2016.03.05.V1.
- [17] You, Z., Feng, Z. M., Yang, Y. Z. Relief degree of land surface dataset of China (1 km) [DB/OL]. Global Change Research Data Publishing & Repository, 2018. DOI: 10.3974/geodb.2018.03.16.V1.
- [18] Xu, G. H., Ge, Q. S., Fang, X. Q., *et al.* Global Change Research Program of P. R. China (2011–2020) [J]. *Journal of Global Change Data & Discovery*, 2017, 1(1): 16–26. DOI: 10.3974/geodp.2017.01.04.
- [19] Lin, L. N., Wang, Y. J., Zhang, J., *et al.* Hydrologic CTD dataset obtained through Chinese Antarctic Expeditions [J]. *Journal of Global Change Data & Discovery*, 2017, 2(3): 157–164. DOI: 10.3974/geodp.2017.02.04.
- [20] Zhang, Q. Q., Ying, G. G., Pan, C. G., *et al.* Comprehensive evaluation of antibiotics emission and fate in the river basins of China: source analysis, multimedia modeling, and linkage to bacterial resistance [J]. *Environmental Science & Technology*, 2015, 49 (11): 6772–6782. DOI: 10.1021/acs.est.5b00729.
- [21] Zhang, H. Q., Xu, E. Q., Zhu, H. Y. A new land use classification system dataset based on the ecological-populated residential-industrial uses of China (2008) [J]. *Journal of Global Change Data & Discovery*, 2017, 1(4): 419–425. DOI: 10.3974/geodb.2017.04.06.