

Data quality control specification of WDC-RRE

**Institute of Geographic Sciences and Natural Resources Research,
Chinese Academy of Sciences**

March 2016

Chapter one General rules

Clause 1 This specification is to ensure an accurate, reliable and authority data in WDC-Renewable Resources and Environment platform and the data normalization during the collecting, organizing and publishing process.

Clause 2 This standardization is for all data resources in the platform of WDC-Renewable Resources and Environment, which include initial data, research data and utility software and some other data.

1.Initial data refers to observation data, monitoring data, exploratory data, test data, experimental data, survey data and examining data, which come from the project.

2.Research data refers to processed data originated from initial data.

3.Utility software refers to software supported to be developed by the project, which can manipulate process and analysis and instructions of software.

The dataset mentioned in this standardization indicates discernible dataset, and those data sets which hold the same product specification come as dataset series.

Clause 3 This specification applies to the WDC-RRE platform and the data quality control during the process of data collection, reduction, dissemination and evaluation on WDC-RRE platform.

Chapter two Data reduction

Clause 4 In accordance to the general requirements of WDS and its self-image, the WDS platform will collect, reduce and process data resources.

Clause 5 There are textual data, numerical data, graphic data, image data and multimedia data in data resources. In order to ensure its facticity and integrity, different types of data must be collected and reduced in a highly strict way and complied with related quality standard and requirement.

Clause 6 All data resources must provide data documents. Data description (data record and quantity) ,database description(data type and data range, database structure, data dictionary, time limit for data protection, scientific value and field of use of data and so on), data instruction(data format and data access method and so on), related software(access, update, download and so on) and some other software are included in the data file. And its specific format adheres to the data file's requirement.

Clause 7 Data quality description is very important in data file, which embodies data sources, approaches of data production and process, and data accuracy. And its specific format adheres to the quality requirement in metadata standard.

Clause 8 Check data quality in line with data quality description file. Concretely, a check description of whether those data has been altered or

damaged need to be created, a more detailed data change record should be established, and the sources, approaches of data production and process, and data accuracy ought to be assessed.

Chapter three Metadata compilations

Clause 9 All datasets must to be traced to corresponding metadata description information.

Clause 10 For the sake of ensuring its facticity and integrity, metadata must obey the metadata standard of the WDS. With such condition in place, metadata items can be expanded to meet the data sets' needs.

Clause 11 Metadata can exist independently. All metadata must connect with relevant data service, such as file data service, database service, geographic information service, HTTP service, FTP service, offline services and so on.

Clause 12 Data must be reused no matter how the technology will develop. Basic metadata and data format of designated field must be built, and some metadata need to be retained to understand and use some data.

Chapter four Data publishment

Clause 13 There are three steps in data dissemination, they are, metadata audit, data audit and disseminating online. And regulations of data quality must be grimly observed.

Clause 14 Metadata must to be reviewed twice; one is its integrity,

which means that all required fields must have description. The other is its accuracy, that is, all metadata, no matter quantitative or non-quantitative need to be reviewed.

Clause 15 During the audit stage, not only the quality of metadata needs to be reviewed, but also the binding dataset entity.

Clause 16 Disseminating online is the last segment of data dissemination, which is executed by data management professionals in WDC-Renewable Resources and Environment. Data's integrity and uniformity will be reexamined. As for those obscure data sets, the technical general group and data expert group will put forward guidance. After being reexamined, data can be published ultimately. At the same time, data files and metadata will be provided to users who will anticipate in data value evaluation to set up a perfect data quality assessment system.

Chapter five Data sharing and application

Clause 17 Data will be assessed by experts and users during its sharing and application process. Assessing data indicates the judgment and recognition process of the delivered data, which helps to discover mistakes in datasets as well as enhancing its quality.

Clause 18 Data evaluation refers to assess data content, data quality, and its applicability.

Evaluate data content: Data content can be preliminarily understood

through its title, key words, abstract, sources and data provider. In some cases, the content can be further understood by downloading some data samples.

Evaluate data quality: Review and appraise data's quality elements, such as integrity, uniformity and accuracy. And the details please refer to data description requirements in Metadata's Content Standard in WDC-RRE Platform.

Evaluate applicability: According to the evaluate results in data content and quality and combined with aimed-needs, data can be judged whether it conforms to requirements.

Clause 19 User's data quality assessment is the feedback for dataset in the use of it, which can be accomplished by on-line or off-line.

Chapter six Appendix

Clause 20 This specification will be tentatively implemented since it has been issued.

Clause 21 The right to interpret of all those clauses is reserved by WDC-RRE platform.

If there are any suggestions or comments about this document, please contact us:

Tel:+86-10-64889048-8006

E-mail: wdc-rre@lreis.ac.cn

Address : 11A, Datun Road, Chaoyang District, Beijing, 100101, China,
Institute of Geographic Sciences and Natural Resources Research, CAS.

WDC-RRE