

Quality Control in Microbial Data Integration: Global Catalogue of Microorganisms as for an Example

Linhuan Wu,

Institute of Microbiology, Chinese Academy of Sciences, China

Global Catalogue of Microorganisms (GCM) is a project proposed by World Data Center of Microorganisms (WDCM) to help organize, unveil and explore the data resources of its member collections. The GCM currently contains strain information from 27 collections in 15 countries. GCM contains (1) catalogue information about strains provided by culture collections, some of the data items are manually classified by GCM staff which allows easier access of catalogue information (2) Data related to strains extracted from public data sources such as Pubmed and Patents, (3) Links to external database such as NCBI (4) Tools for bioinformatic analysis and also tools to better explore resources in GCM.

GCM sets the WDCM Minimum Data Sets (MDS) and Recommended Data Sets (RDS) based on widely applied standards such as OECD Best Practice Guidelines for Biological Resource Centres and Microbial Information Network Europe (MINE). Each participant collection transferred its catalogue information according to WDCM MDS either by Excel template, XML template or database files directly. WDCM worked on the strain information and published on global catalogue web page.

Before integration of catalogue information from different data sources in different data format, the important measurements on data quality control should be taken. These measurements include the checking data of organism type with its species information, the sequence information and the nomenclature information for microorganisms. In this presentation, the author will take GCM as an example to introduce the data quality control issues in microorganism data integration.