Cryosphere monitoring and data management in Cold and Arid Regions

Yaonan Zhang
Cold and Arid Regions Environmental and Engineering Research Institute, CAS
yaonan@lzb.ac.cn
Outline

1. The data come from field observation site and survey
2. The data management of Data Center of Sciences In Cold and Arid Regions
3. The on-line data sharing and off-line application
4. Prospect
1. Where the data come from

- The data come from

  The data of Data Center of Science in Cold and Arid Regions mainly come from field observation stations, spatial observations, and internet other service institutions, the data sources and the data formats are multiplicity and complex.
1. Where the data come from

- We have 16 long-term field observation stations and these located in cold and arid regions.
- The data acquisition from these field observation sites, such as Tianshan Glaciological, Qinghai-Tibet plateau Permafrost and so on, is very hard for researcher.
- We have deployed a data transmission test bed in our data center to verifying these technologies stability.
1. Where the data come from

- **The Permafrost data come from**
  - Permafrost physics, cold regions engineering, freezing-thawing processes and its control.
  - Long-term field observation datasets, such as Golmud, Bailuhe, Tianshan Permafrost observation.
  - Permafrost Engineering
    - Tibet railway
    - Tibet road
  - Special Environment observation areas
1. Where the data come from

- The water data come from inland river basin, Cold and Arid Regions
- Such as surface water, and ground water.
- Heihe river basin, Shiyanghe river basin, Shulehe river basin and so on. Some water data come from other organizations or institutes.
Outline

1. The data come from field observation site and survey

2. The data management of Data Center of Sciences In Cold and Arid Regions

3. The on-line data sharing and off-line application

4. prospect
(1) The data using and managing problem

- A large amount of observation data have been accumulated by now, but the management of these data are difficult, and the use of these data are very complex.
- Because of
  - The data scattered owned distribution.
  - The field observation stations are geographically distributed.
  - The observation data are real-time, multiplex, long-term, and massive.
  - The data management systems of the stations are diversified.

- These data are usually maintained in different formats.
- These data are usually store in disparate systems.

So once the datasets are collected, scientists in practice then waste considerable time on repetitive, time-consuming operations to integrate such disparate datasets rather than focusing on real scientific analysis and decision-making.
Another challenging issue that scientists are facing is visualizations

- Scientists are likes to have useful and effective data visualizations.
- Users, in particular, prefer to visualize multi-dimensional datasets so that they are able to have a better understanding of the quality and characteristics of data before having them downloaded.
(3) Virtual Joint Observing System

A. **Device networking layer**

B. **Network transport layer**

C. **Data aggregation layer**

D. **System application layer**
XML-based metadata description schematic of the VJOS

Encapsulation the metadata according the data transmit form sensors to the data center
The Data Platform in the Data Center

The tasks of the data platform are principally:
- data collection
- data management
- data assimilation
- data share service
- data application

The modules in the data platform mainly include the:
- user-management system
- information-service system
- metadata-management system
- data-management system.
The Data Platform Architecture

- The server layer mainly provides data services, including archive file data, local data, online data, data’s auxiliary catalogue, and metadata-relevant services.

- The middleware layer offers the discovery service of metadata, data-replicating service, metadata-replicating service, requesting-management service, certification and authorization service, visual service, and data-analysis service.

- The interface layer mostly offers the services on data release, data discovery, data analysis, and data application.
Metadata service

- Data sharing and analysis service is based on metadata.
- The platform provides a variety of metadata-based queried approach allows users to more easily and simply access the data and models.

- Query type:
  - Simple
  - Complex
  - Map

- Data classification
- Data navigation
- Data search
- Data preview
- Data service
Data quality control

- Simple but not easy, you can through on line chart evaluated the data quality of the datasets as a whole.
Outline

1. The data come from field observation site and survey
2. The data management of Data Center of Sciences In Cold and Arid Regions
3. The on-line data sharing and off-line application
4. prospect
The data sharing management

- The data authentication (whose data?)
- On-line data query management
- Off-line data query management
- Datasets
- Metadata custom made (data citing methods, limited using, reference, figure……)
- Statistics
Data management and sharing frame

Diagram:

1. Data汇集
   - 元数据汇集
   - 数据实体汇集
   - 数据审核
   - 数据发布
   - 数据共享
   - 数据反馈
   - 数据出版
   - 数据挖掘
   - 数据可视化
   - 数据服务
   - 数据统计
Data aggregation process
Data examine and verify process
Data on-line sharing service process

- The data organization according on Discipline, Topical Subject, Location, Time
- User can discovery they needs data by the key words, location, field station, spatial extent, datasets, data catalogue, document, figure
Data query process

- We provide six methods for user to query data
  - key words, location, field station, spatial extent, datasets, data catalogue

Based on field observation station, you can use field stations to find data. Based on atmosphere station, you can use station’s information to search need’s data. Based on spatial extent, use bounding boxes to search need’s data.
Data-discovery service

Here, we offer a unanimous data view for an application program through the storage system abstract and the Application Programming Interface (API), so as to the application program can access these data by the unanimous data view in the data platform.
Data-replicating service

- Sometimes the data we need are distributed in the node of different field observation stations. So we need to gather the data from nodes for each use.
- We collect the data that accord with certain conditions into a dataset.
Data off-line sharing service process

1. **开始**
2. **收到用户的申请**
3. **邮件通知用户**
4. **委托数据中心管理**
   - **有数据作者**
     - **专题数据集**
     - **数据管理委员会**
     - **从元数据提取联系人**
5. **发送管理邮件**
6. **两个工作日后或所有申请都有反馈**
7. **接受或拒绝申请**
8. **联系人 PoC**
9. **管理者 custodian**
   - **资源提供者**
   - **数据所有者 owner**
10. **数据发布者**
11. **数据分发者**

结束
(2) The functions will be include in management

- **Data publication**: DOI
- **Data information mining**: data related to data, data related to document, data related to search engine
- **Preparation specialist data for driver model modeling**
Data Citation Method AS DOI

邀请信

合作协议

DATA PROVIDER AGREEMENT

This Agreement is dated as of May 6, 2014 between Thomson Reuters (Scientific) LLC ("Thomson Reuters"), with a principal place of business at 1500 Spring Garden Street, Fourth Floor, Philadelphia, Pennsylvania 19103, and Cold and Arid Regions Science Data Center at Lanzhou (CARD), with a principal place of business at 330 Dongguan Road, Lanzhou, Gansu, China.

WHEREAS, Thomson Reuters has developed and owns various electronic database products, which allow users to search and access information gathered from data repositories containing data from research studies in the areas of science, social sciences, and arts and humanities;

WHEREAS, Data Provider wishes to provide Thomson Reuters with access to data from its repository for the purpose of allowing Thomson Reuters to index such data for its databases;

NOW, THEREFORE, in consideration of the agreements and obligations set forth herein and for other good and valuable consideration, the receipt and sufficiency of which are hereby acknowledged, Thomson Reuters and Data Provider agree as follows:

1. Definitions:

   In this Agreement, the following terms shall have the following meanings:

   "Data" means the metadata, abstracts and citation data (if available) contained in the Repository.

----

WATER: Dataset of automatic meteorological observations at the A'rou freeze/thaw observation station

作者：Zeyong, Hu; Mingguo, Ma; Rui, Jin, 等。

Cold and Arid Regions Science Data Center at Lanzhou

来源 URL: http://card.westgis.ac.cn/data/50a6fe84-deeb-49b0-b8d8-8f784bc27a89

文献类型: Data set
Cold and Arid Regions Data Science Center has been confirmed as a recommendation data storing and publishing organization by Nature publishing group.
The data related to the publication
The data related to the publication

- 黑河综合遥感联合试验: 蓝科灌区绿洲站自动气象站数据集
- 黑河综合遥感联合试验: 蓝科灌区绿洲站涡动相关通量数据集

数据文献库

1. 庄金鑫, 王维真, 王介民. 涡动相关通量计算及三种主要软件的比较分析 [J]. 水电学报, 2013(12): 3774-3782. 查看 相关数据
2. 朱乔, 梁睿, 晋华, 等. 基于SWAT模型的堰河流域径流模拟[J]. 水电学报, 2013(12): 3774-3782. 查看 相关数据
4. 周伟, 王倩, 章超斌, 等. 黑河中上游草甸区NDV点数据时空变化规律及其对
   数据
7. 赵岩, 王思远, 毕海芸, 王辉, 殷慧. 基于ASTER与TM数据的黑河流域中游区典型区域植被覆盖度研究, 测绘通报, 2013(04): 4-7. 查看 相关数据
10. 张虎, 焦子飞, 李小文, 等. 光学信息估算HJ-1 CCD数据地表反照率[J]. 遥感学报. 2013(02): 295-305. 查看 相关数据
1. The data come from field observation site and survey

2. The data management of Data Center of Sciences In Cold and Arid Regions

3. The on-line data sharing and off-line application

4. prospect
4. Prospect

- **Further study is focused on**
  - Data replication, data copy, metadata management, key data application techniques based on the preliminarily established data platform
  - Completing tasks of data collection, management and integrating for field observation stations.
  - Data analysis in graphical methods, on-demand data preparation by means of integrating and processing, and multi-source data assimilation and data fusion in river basin scale.
4. Prospect

- **Establishing a model integration framework to supporting Model application**
  - Should be consistent with models import, the hierarchical levels management, spatial data analysis and Multi-dimensional explicit representing simulation of the process;
  - Should provide appropriate services to facilitate users to use the data, can integrate various kinds of models and methods.
  - Establish model metadata to manage models in the platform for users to view and search relevant model rapidly.
4. Prospect

- **Building a data assimilation system**
  - to address some problems related to uncertainties in model and data, helping us to assess the worth of various data sources and to use this data in an optimal way.
  - The data assimilation system should be composed of
    - the model operator
    - the observation operator
    - data assimilation algorithm
    - data sets
Thank you!