

Introduction to Environmental and Ecological Science Data Center for West China (Westdc)

Zhuotong Nan, Xin Li, Yongjian
Ding, Lizong Wu, Youhua Ran,
Liangxu Wang



5

Outline

- Background
 - Why NSFC China funded this data center?
- Components
 - Four platforms: Data Sharing platform, Knowledge Repository platform, Experience Exchange platform, and Data Science platform
- Implementation
 - Technical
 - Metadata profile: based on ISO 19115
 - User services: full & open data policy
 - Long archiving: save expertise with Knowledge Repository platform
- Summary
 - From data to knowledge

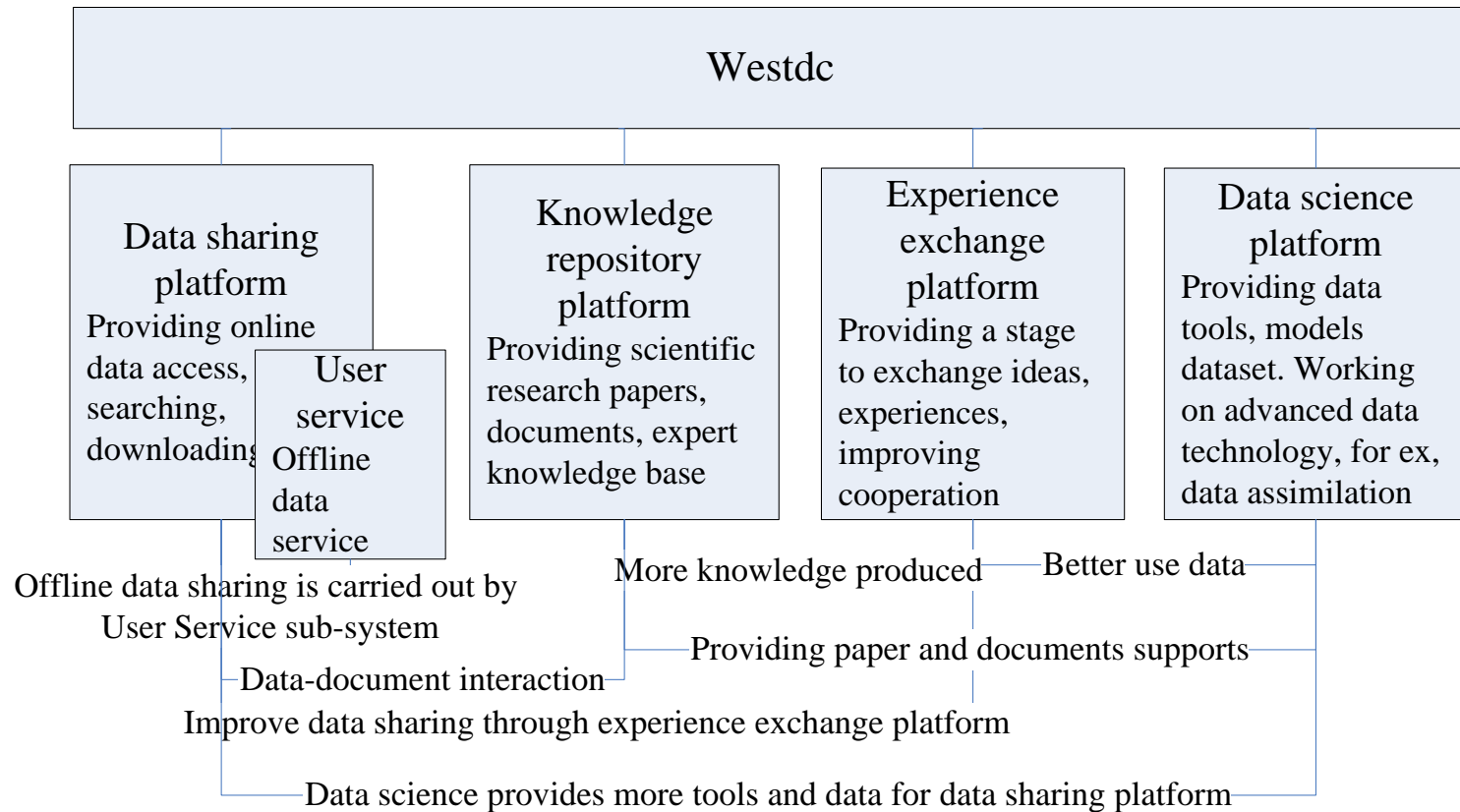


1. Background

- West Plan initiated by NSFC of China has collected and produced numerous precious scientific data about ecology and environment science in Western China
- Rich data center experience in CAREERI
 - Running 4 state-level data centers: WDC-D snow and ice subcenter, Data Sharing Network northwest center, Digital Heihe, Environmental resource data center
- CAREERI has over 40 years experience on environmental and ecological research
 - Experience on how to effectively use data



2. System components



2.1 Data sharing platform

- Functionality:
 - Online data access
 - Advanced data searching
 - Data collecting
- Including a specialized data service component, providing offline data service.
 - Some data cannot be online due to China's data policy restriction
- Each data could be associated with documents describing them or research papers using them. Such relationship is dual-way.
 - Documents are managed by Knowledge Repository platform.
- Data catalogue
 - Providing a general data classification, as well as a custom data cataloguing mechanism allowing users define their own classification
- A series of specifications & regulations to regulate data preparation and use have been created
 - Data sharing specification, use terms, specification for dataset creation



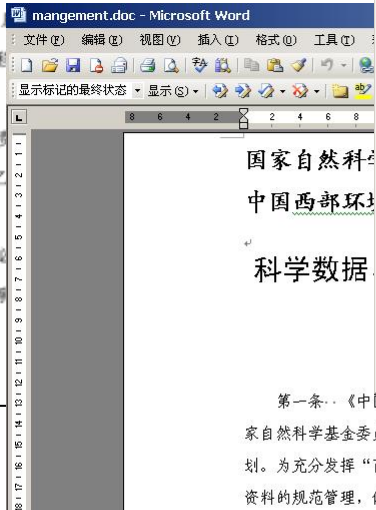
国家自然科学基金委“中国西部环境与生态 科学数据中心”项目

网箏 (NK-800) 使用协议

甲方：国家自然科学基金委“中国西部环境与生态科学数据中心”项目
乙方：

“中国西部环境与生态科学数据中心”（以下简称“数据中心”，<http://westdc.westgis.ac.cn>）在国家自然科学基金委（以下称“基金委”）经费支持下，收集、整理和归档基金委“西部计划”开展以来的产生的环境与生态领域的科学数据，并开展数据科学研究，作为“数据中心”文献服务的一个重要内容，“数据中心”从中国科学院资源环境信息中心（以下称“信

双方特签订如下协议：
一、甲方方向乙方免费
年3月26日，期满后乙
后续使用事宜。
二、乙方必须提供必
“信息中心”备案；网



中国国家自然科学基金委（以下简称“基金委”）“中国西部环境与生态科学数据中心”（以下简称“西部数据中心”）采用“完全与开放”（Full & Open）的数据共享政策，即所有的科学家或研究项目都有权无差别地获得数据中心的数据以及包括数据处理过程等的技术文档，以不高于复制和邮寄成本或免费的形式，向基金委“中国西部环境与生态科学研究计划”（以下简称“西部计划”）内项目和人员，以及对对中国西部环境生态科学研究感兴趣的研究团队或个人提供数据。

使用条款

本着尊重知识产权和维护网络安全的原则，我们向数据用户提出以下使用条款：

第一条 “西部数据中心”网站上的一切数据和资料版权归数据和资料原生产单位所有。

第二条 为保障数据开发者的著作权，用户在使用全部或部分“西部数据中心”所提供的数据的基础上产出的研究成果中（包括公开发表的论文、论著、数据产品和未公开发表的研究报告、数据产品、系统开发等），须在相关成果的显著位置上明确注明数据来源。除对数据来源署名有特殊要求以外，用户须依据以下规范注明数据来源：

中文成果：“数据来源于国家自然科学基金委员会“中国西部环境与生态科学数据中心”（<http://westdc.westgis.ac.cn>）”

英文成果：“This data set is provided by Environmental & Ecological Science Data Center for West China, National Natural Science Foundation of China”（<http://westdc.westgis.ac.cn>）”

第三条 数据仅供用户用于科研目的，不得用于商业目的。不得将数据转让给第三方，由此引起的一切后果由数据用户承担。

第四条 从数据中心获取的数据如属数据保密范围，用户须与“西部数据中心”签署保密协议，并有义务遵守国家相关法律法规，如出现间题，责任由用户完全承担。

第五条 用户在申请数据时，应填写真实资料，如用户资料发生变更，应及时更改相关资料。若您提供的资料不齐全或不正确，“西部数据中心”有权拒绝提供服务。

第六条 “西部数据中心”用户须遵守中华人民共和国计算机安全、互联网通讯的有关法律法规，用户不得利用西部数据中心网站传输、出版、下载、复制或传播任何非法、有害、恐吓、骚扰、侵权、诽谤、淫秽、中伤、违反他人隐私或宣传、违反他人知识产权 或者任何其他

完成

第一条 “中国西部环境与生态科学研究计划”（以下简称“西部计划”）是国家自然科学基金委员会组织实施的关于中国西部环境和生态建设的重大科学研究计划。为充分发挥“西部计划”所获取数据与资料的作用并保护其产权，实现数据与资料的规范管理，依据《中华人民共和国知识产权保护法》、《中华人民共和国保守

等、平等互利”的基本原则，制定本条例，以切实实现“西部计划”的数据与资料的集成与共享。

第二条 “科学数据中心”（以下简称“数据中心”）是服务于“西部计划”的

Some specifications and regulations created for Westdc



Data available at present stage

Earth observing data

- MSS, TM, ETM+ imageries (1970s, 1990s, 2000s);
- ASTER image (2002)
- AVHRR-NDVI data (1982 to 2004);
- Daily NOAA AVHRR data for northwest China, 2002, with resolution 1km
- SPOT Vegetation data (1998 to 2005)
- MODIS data products (2001-2004)
- SSMR、SSM/I数据(1978-2004)

Meteorological and hydrological datasets

- Daily meteorological monitoring data from 1955 (all meteorological stations in West China)
- Hydrological stations observing data in Heihe River Basin since 1980s

Thematic data

- DEM (1:250 000 for most Western areas, partially 1:100 000);
- Land use data of West China, 1:100 000
- Soil data 1: 1000 000 and soil erosion data
- 1:1 000 000 vegetation classification data
- 1:100 000 glacial database
- 1: 1 000 000 desert, wet land, and lake data
- 1:250 000 geological and hydro-geological maps

Other

- Datasets for four counties of Gansu Province, ie., Dingxi, Tongwei, Weiyuan, and Longxi counties, including land use in 1:100 000, DEM and remote sensing data
- Re-analysis meteorological data in Heihe River Basin, West China
- Seismologic inventory data
- Stalagmite data provided by Prof. Tang, etc



Featured dataset

- **Featured disciplinary dataset**
 - **Cryospheric datasets**
 - National wide glacier inventory data, glacier and ice lake data from neighboring countries, including India, Nepal, Pakistan, etc
 - site observation data of West China, national wide snow distribution map, and snow water equivalent data
 - Frozen soil data, 1:4 000 000 permafrost distribution data of China, permafrost map for Qinghai-Tibet Plateau, along Qinghai-Tibet railway, etc
 - Snow depth data of China
 - **Arid region data resource**
 - 1: 100 000 China deserts distribution map
- **Regional thematic data**
 - Heihe river basin
 - Remote sensing, observation, thematic, and field experiment data
 - Shiyanghe River Basin
 - Remote sensing, thematic, and field experimental data
 - Soil profiles
 - Background data along the Qinghai-Tibet railway
 - Other areas
 - South Asia, Central Asia
- **Assimilation data (re-analysis datasets)**
 - Re-analysis meteorological datasets.





- ### 中心动态
- 北京2006CODATA会议 (2006-10-23)
 - 贵州地化所同志来西部数据中心学习 (2006-10-19)
 - 西部数据中心第四次例会 (2006-10-14)
 - 数据中心开发BLOG (2006-10-06)
 - 数据中心紧急会议 (2006-9-30)
- [>>更多...](#)

搜索

搜索选项:

关键词:

[高级搜索](#)

- ### 快速链接
- [如何申请数据](#)
 - [如何向数据中心提交数据](#)
 - [数据使用反馈表单](#)
 - [申请移动IP文献查询服务](#)

移动文献查询服务

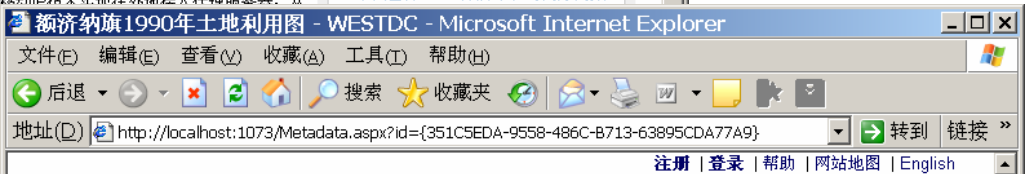
移动文献查询服务采用先进的移动IP技术实现在外地插入任何服务器，而不受IP地址的限制阅读和查作为数据中心“知识积累平台”项目成员使用。使用该网等，的国内外文献资源了。

[>>详细...](#)

[申请移动文献查询服务](#)

- ### 关键词
- 中国雪深25km分辨率长时间序列数据集
 - 石羊河土壤实验资料
 - 黑河流域土壤数据
 - 黑河流域土壤数据
 - 中国雪

- ### 数据类别
- 矢量数据**
专题矢量图件; 地理背景矢量数据如交通、行政区划; 等
 - 栅格数据**
遥感影像(如MODIS数据)、DEM及派生
 - 台站监测数据**
气象、北京、地震等台站观测数据
 - 实验参数资料**
土壤、植被等



额济纳旗1990年土地利用图

元数据标识: {351C5EDA-9558-486C-B713-63895CDA77A9}

元数据语言: 中文

元数据字符集: utf16 - 16位UCS传输格式

元数据描述的数据范围: 数据集

自定义范围名称: dataset

元数据联系单位(人):
联系人姓名: 冉有华
单位: 中国科学院寒区旱区环境与工程研究所遥感与地理信息科学研究室
联系人角色: 拥有者
联系信息:
地址:
城市: 兰州市
E-mail地址: ranyh@lzb.ac.cn
邮政编码: 730000
通信地址: 甘肃省兰州市东岗西路320号
电话:

Screenshot of the homepage of Westdc Website

Metadata document detailed page



2.2 Knowledge Repository platform

- **Functionality**
 - Repository of research papers related to environmental and ecological research in West China, especially those used dataset provided by Westdc.
 - Repository of documents describing a kind of data, for ex. MODIS data, and manuals of data tools
 - Advanced data mining research
- Link data from Data Sharing platform to literature
- Link tools from Data Science platform to documents
- OAIS model based
- **Right screenshot shows the mobile bibliography service**
 - With installation of Netkite, a proxy client for bibliographic query, scholar can access literature and documents FREELY, without the restriction of IP address.
 - This service is free of charge for West Plan projects



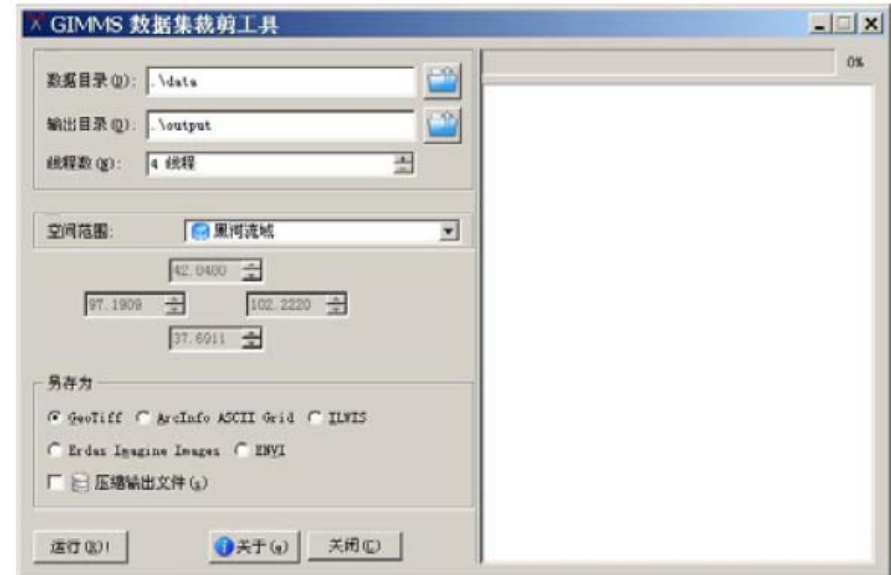
2.3 Experience Exchange platform

- **Functionality:**
 - Share ideas, exchange data between individuals, improve cooperation between individuals
- **Forum (BBS)**
- **mailList**
 - subscribe online
- **Forum <->mailList**
 - Enable senior scientists, who are generally not interested to use forum, to join discussion
- **Blog for developers**
 - <http://westdc.westgis.ac.cn/blog/>



2.4 Data Science platform

- Developing data tools
- Sharing model datasets
 - SWAT datasets for Heihe River basin, etc
 - Datasets ready to run without any change.
- Providing live map and webservices
 - WMS, Web-services, etc
- Advanced data technology
 - Data assimilation
 - Datasets produced by data assimilation
 - Etc
- Right screenshot shows a tool to clip GIMMS dataset with specifying spatial extent

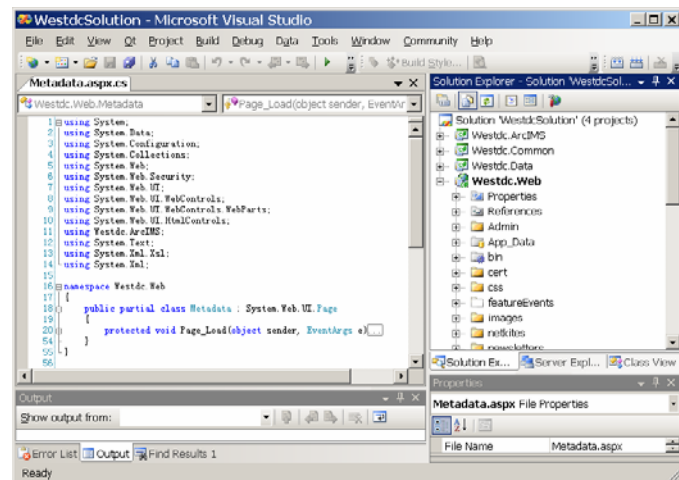
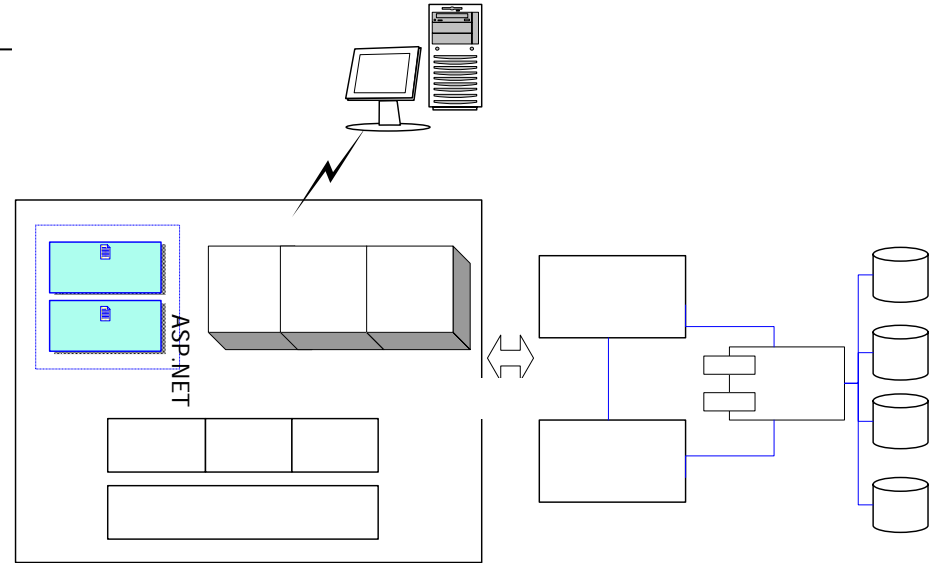


A data subsetting utility for GIMMS data, available through the Data Science platform



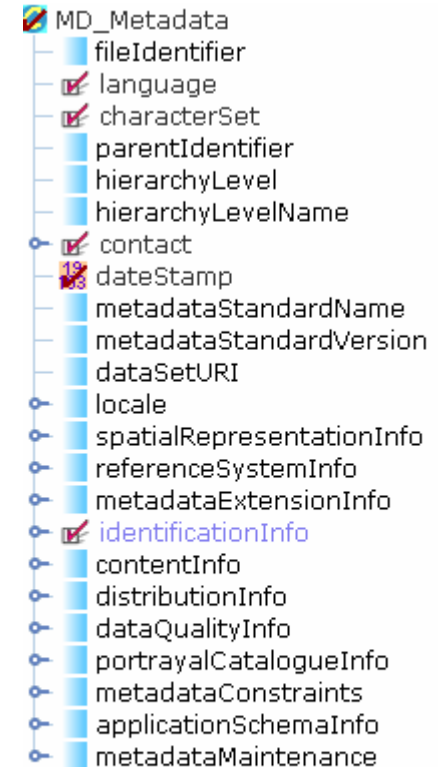
3. Implementation

- Metadata standard
 - Iso 19115 with extension
- Metadata server
 - ArcIMS 9
- Web programming language
 - Asp.net 2.0 (c#)
- Web Server
 - IIS 6 under windows 2003 server
- Database
 - Sql Server 2k sp4



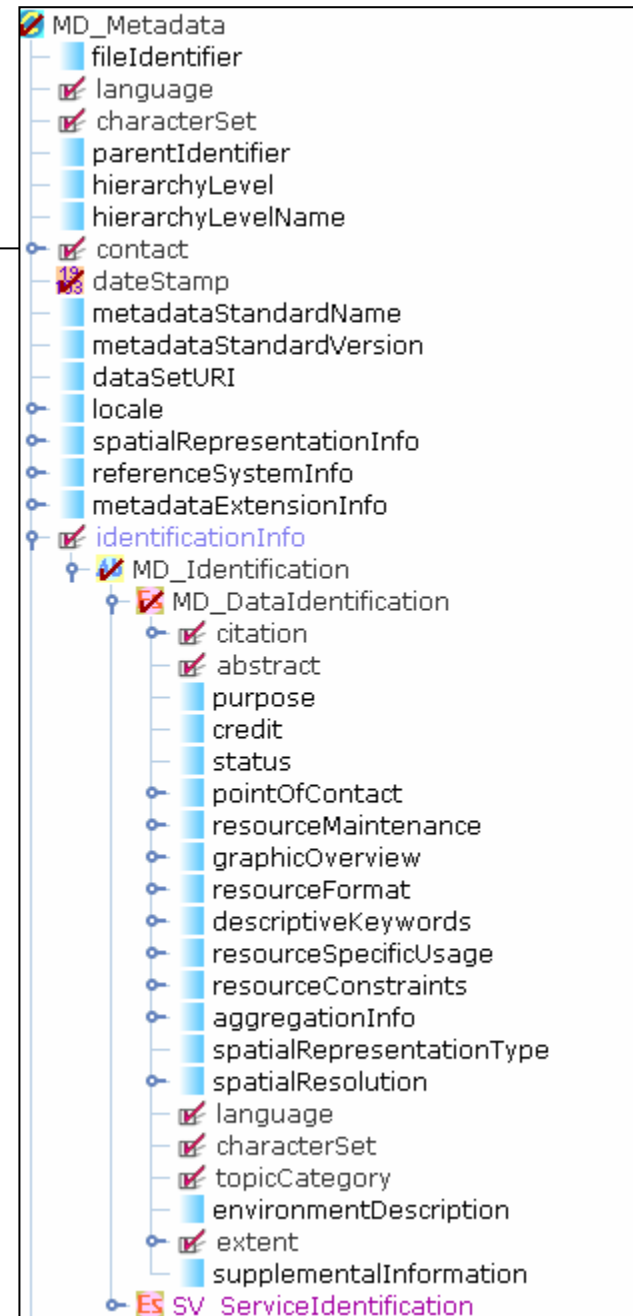
3.1 Metadata profile

- ISO 19115 based
- ESRI ISO profile with extension
 - Extension in terms of field observation data, such as field description, learning from the GB metadata standard for Ecological data
- Why choose ESRI ISO profile with extension method?
 - ISO 19115 is an international metadata standard, effective especially for spatial data. Most of data in Westdc is spatial
 - We are familiar with ESRI products
 - ESRI ISO profile can be easily implemented by ArcIMS md server
 - ArcCatalog has an excellent ISO metadata editor, which can be used as metadata input method. (sure, has to be customized, for ex. to add more pages)
 - ESRI ISO different from ISO 19139 in xml implementation, but easy to build a metadata crosswalk.



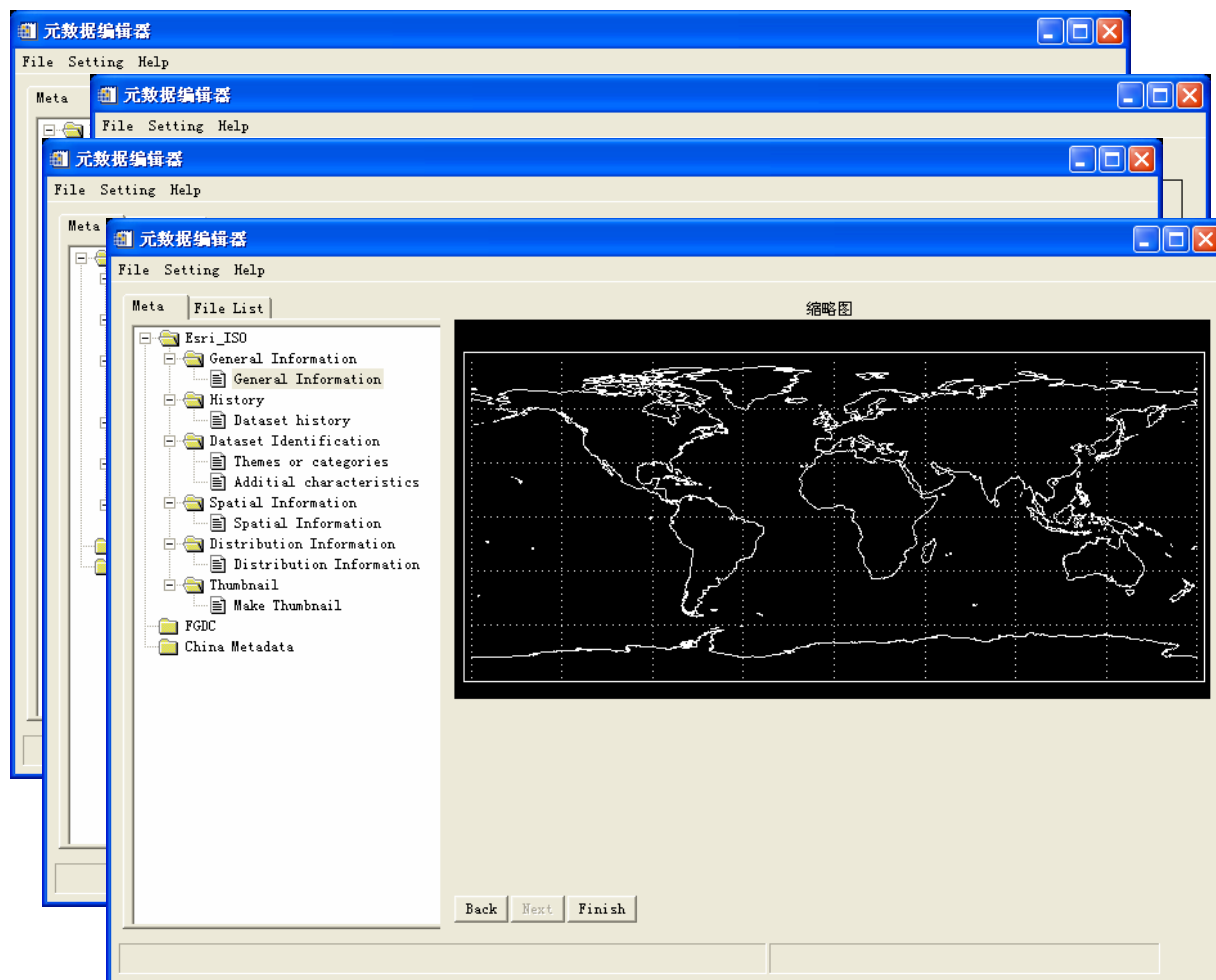
Minimum metadata for Westdc

- Dataset title
- Dataset reference date (citation.date)
- Dataset language/character set
- Topic category
- Abstract
- Metadata point of contact (metadata author)
- Metadata date stamp
- Metadata language/character set
- Geo-location of the dataset



Customized ISO editor

- IDL required to run
- Features
 - Batch processing documents
 - Get information from files automatically
 - For ex., proj, thumbnail, etc
 - Template support



3.2 User service

- What user cares?

!!DATA!! How many and how easily can they obtain data from this data center?

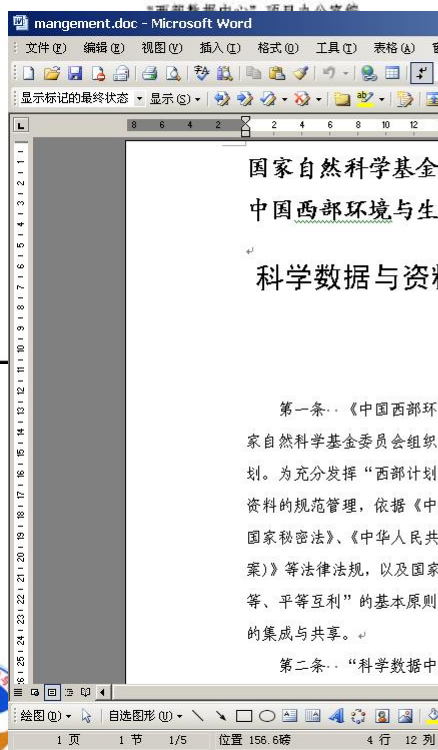
- Insist in full & open data policy
 - Data freely available for every body
 - Anybody can access the data center equally (try to do, but due to data policy in China, actually we cannot really reach this purpose)
- A dedicated user service group
 - Dedicated persons responsible for this group
- Quarterly newsletter (paper and electronic)



国家自然科学基金委员会
中国西部环境与生态科学研究计划
中国西部环境与生态科学数据中心

数据通讯

2005年6月 总第1期

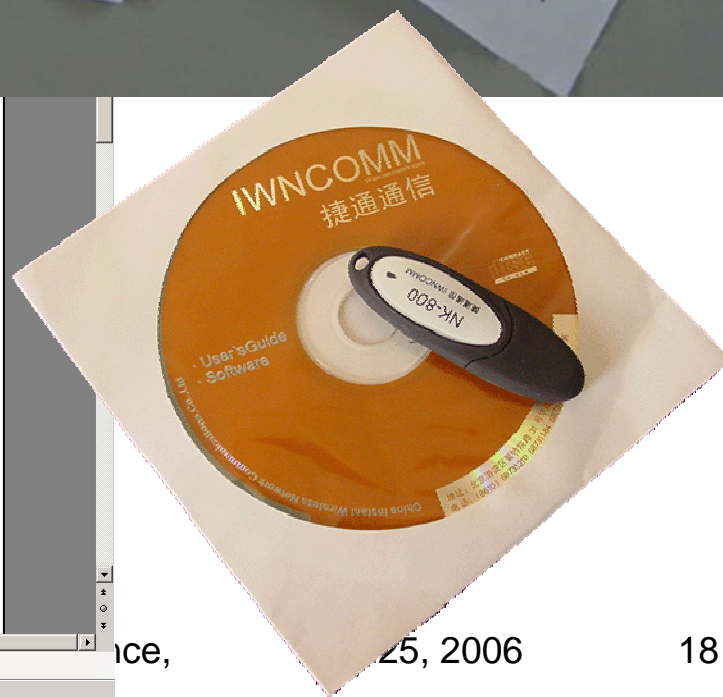
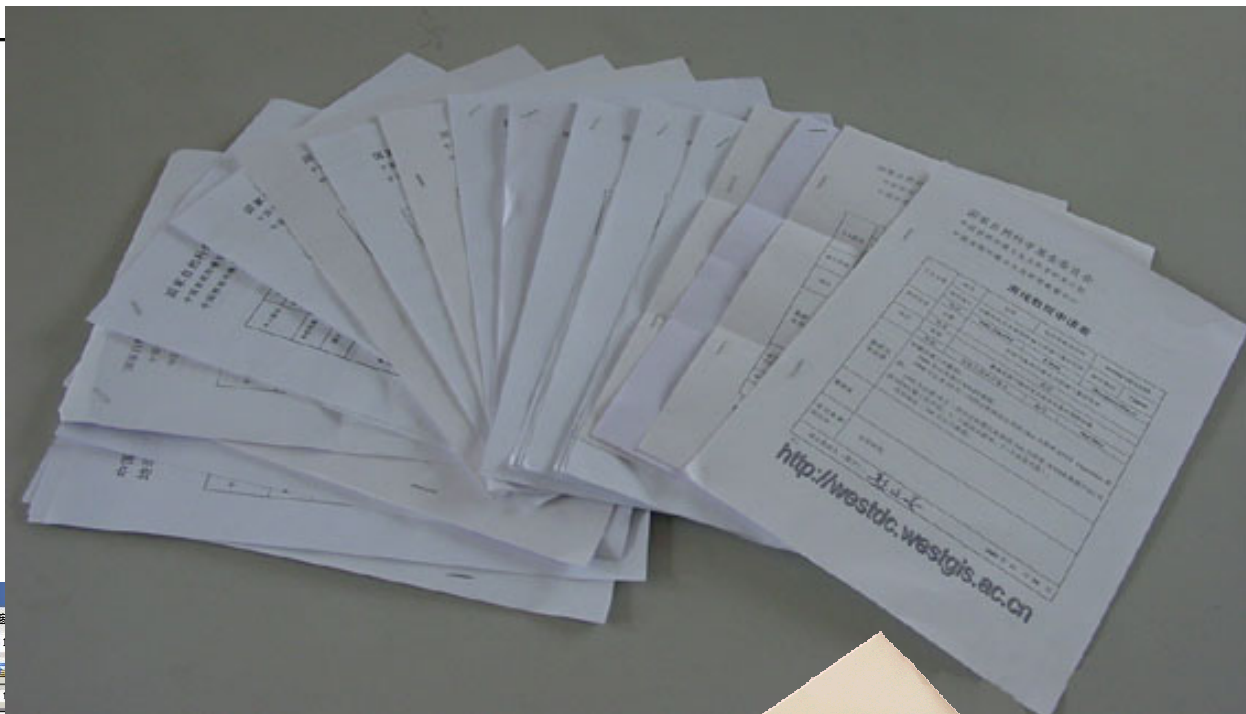


国家自然科学基金委员会。
中国西部环境与生态科学数据中心。
科学数据与资料集成和共享管理条例(暂行)。

第一章·总则。

第一条·《中国西部环境与生态科学研究计划》(以下简称“西部计划”)是国家自然科学基金委员会组织实施的关于中国西部环境和生态建设的重大科学研究计划。为充分发挥“西部计划”所获取数据与资料的作用并保护其产权,实现数据与资料的规范管理,依据《中华人民共和国知识产权保护法》、《中华人民共和国保守国家秘密法》、《中华人民共和国著作权法》、《中华人民共和国科学技术普及法(草案)》等法律法规,以及国家自然科学基金项目管理相关规定,本着“权利与义务对等、平等互利”的基本原则,制定本条例,以切实实现“西部计划”的数据与资料的集成与共享。

第二条·“科学数据中心”(以下简称“数据中心”)是服务于“西部计划”的



Data center quarterly newsletters

NSFC-WESTDC



国家自然科学基金委员会
中国西部环境与生态科学数据中心

项目简报

总第1期 2005年06月26日

中心简介

国家自然科学基金委员会“中国西部环境与生态科学数据中心”，简称“西部数据中心”，是国家自然科学基金委员会地学部“中国西部环境与生态科学计划”（以下简称“西部计划”）资助下建立的数据共享平台，其任务是收集、管理“西部计划”项目数据产出，并面向国内外特别是基金项目有关人员提供科学数据服务。“西部数据中心”受基金委地学部领导，由中国科学院寒区旱区环境与工程研究所和中国科学院地理科学与资源研究所承建和运行。

本期导读

发布大量基础数据	(第1页)
开通了星版的数据中心网站	(第2页)
“西部数据中心”数据目标	(第2页)
“西部数据中心”建设纲要	(第3页)
如何获取数据	(第4页)
数据使用声明	(第4页)
联系方式	(第4页)

最新数据

- 基础地理背景数据
 - 西部地区 1:25 万 DTM (DEM、坡度、坡向)；
 - 甘肃省 1:25 万基础地理背景数据集；
 - 黑河流域基础地理背景数据集,包括公路、铁路、等高线、居民点等；
 - 甘肃省定西四县(定西、通渭、渭源、陇西)1:10 万基础地理背景数据集；
 - 西部地区 1:100 万基础地理背景数据集；
 - 西部地区 1:40
- 陆地表面数据
 - 西部地区 1:10
 - 西部地区 1:10
 - 西部地区 1:10
 - 西部地区 1:10
 - 西藏、青海、
 - 黑河流域土壤
 - 甘肃省定西四
- 遥感数据

Data service group visiting major West Plan projects

We have spent about **2 month** to visit users across the country

NSFC-WESTDC



国家自然科学基金委员会
中国西部环境与生态科学数据中心

项目简报

总第2期 2006年07月16日

中心简介

国家自然科学基金委员会“中国西部环境与生态科学数据中心”，简称“西部数据中心”，是国家自然科学基金委员会地学部“中国西部环境与生态科学计划”（以下简称“西部计划”）资助下建立的数据共享平台，其任务是收集、管理“西部计划”项目数据产出，并面向国内外特别是基金项目有关人员提供科学数据服务。“西部数据中心”受基金委地学部领导，由中国科学院寒区旱区环境与工程研究所和中国科学院地理科学与资源研究所承建和运

本期导读

我国平均积雪深度	(第1页)
2650 年北京石笋重建地月温度	(第2页)
互生河土壤剖面	(第2页)
中国地震目录	(第2页)
平台开发进展顺利	(第3页)
GIMMS CLIP Tools V1.0 发布	(第3页)
上门服务第一阶段工作顺利完成	(第4页)
如何获取数据	(第5页)
数据使用声明	(第5页)



3.3 Long-term data archiving

- Technically
 - Make use of latest technology
 - Hardware assurance: disk array,
 - Using unique data identifier
 - Both this identifier and associated metadata will be not changed even if the data becomes obsolete
 - Make detailed metadata, associate data and documents to save necessary information to use this data many years later. SAVE EXPERTISE/KNOWLEDGE on the Knowledge Repository platform
 - **Data science platform** will carry out data science research activity, for ex., to develop data tools to convert different formats.
- Institutionally
 - Establish regulations for metadata creation and improvement. Metadata should be created and complemented as detailed as possible.
 - Consideration about the long term data archiving problems, and try to find a way out.



5. Summary

- Westdc consists of 4 components. Each component functions separately and interacts with others.
- Data is the kernel of a data center; data center should really serve users. Westdc has carried out many measures to realize this principle.
- Full and open is Westdc's pursuing, although there are many obstacles. To achieve this purpose, a dedicated user service group has been established. Data specifications and regulations are created.
- Technology will be updated very soon, system and service should use those technology in time.
- As for the trend of from data to knowledge, we built Knowledge Repository platform and data science platform, try to link data to literature and improve understanding. Long term data archiving problems have been considered. Westdc has done some tentative works.



[Http://westdc.westgis.ac.cn](http://westdc.westgis.ac.cn)

Thanks!