International Journal of Digital Earth
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tjde20

Toward an improved data stewardship and service for environmental and ecological science data in West China

Xin Li [a], Zhuotong Nan [a], Guodong Cheng [a], Yongjian Ding [a], Lizong Wu [a], Liangxu Wang [a], Jian Wang [a], Youhua Ran [a], Hongxing Li [a], Xiaoduo Pan [a] & Zhongming Zhu [b]

[a] Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, 730000, China
[b] Scientific Information Center for Resources and Environment, Chinese Academy of Sciences, Lanzhou, 730000, China

Available online: 01 Jun 2011

To cite this article: Xin Li, Zhuotong Nan, Guodong Cheng, Yongjian Ding, Lizong Wu, Liangxu Wang, Jian Wang, Youhua Ran, Hongxing Li, Xiaoduo Pan & Zhongming Zhu (2011): Toward an improved data stewardship and service for environmental and ecological science data in West China, International Journal of Digital Earth, 4:4, 347-359

To link to this article: http://dx.doi.org/10.1080/17538947.2011.558123

International Journal of Digital Earth,
Vol. 4, No. 4, July 2011, 347–359

Taylor & Francis
Taylor & Francis Group

Toward an improved data stewardship and service for environmental and ecological science data in West China

Xin Li[a]*, Zhuotong Nan[a], Guodong Cheng[a], Yongjian Ding[a], Lizong Wu[a], Liangxu Wang[a], Jian Wang[a], Youhua Ran[a], Hongxing Li[a], Xiaoduo Pan[a] and Zhongming Zhu[b]

[a]*Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China;* [b]*Scientific Information Center for Resources and Environment, Chinese Academy of Sciences, Lanzhou 730000, China*

(*Received 24 May 2010; final version received 10 January 2011*)

Sharing of scientific data can help scientific research to flourish and facilitate more widespread use of scientific data for the benefit of society. The Environmental and Ecological Science Data Center for West China (WestDC), sponsored by the National Natural Science Foundation of China (NSFC), aims to collect, manage, integrate, and disseminate environmental and ecological data from western China. It also aims to provide a long-term data service for multidisciplinary research within NSFC's "Environment and Ecology of West China Research Plan" (NSFC West Plan). An integrated platform has been developed by the WestDC, and this has the function of data sharing, acting as a knowledge repository. Major data sets developed by the WestDC include basic geographic data, the regionalization of global data set for China, scientific data for cold and arid regions in China, scientific data for the cryosphere in countries that neighbor China, data relating to the inland river basins in northwestern China, and data submitted by the NSFC West Plan projects. In compliance with the "full and open" data sharing policy, most data in the WestDC can be accessed online. Highlights include detailed data documentation, the integration of data with bibliographic knowledge, data publishing, and data reference.

Keywords: data center; West China; ecology; environmental science; data sharing; digital infrastructure

1. Background

West China has one of the World's most diverse but fragile environments, with the Qinghai-Tibetan Plateau elevated in the south and high mountains, plateaus, deserts, and oases interspersed in the north. Its environmental status and evolution is always of great research interest. In 2001, the National Natural Science Foundation of China (NSFC) initialized a major research plan, the Environment and Ecology of West China Research Plan (hereafter referred to as the NSFC West Plan), which aims to improve the understanding of the evolution and future trends of the environment in western China and to detect the natural and human factors that have shaped the current environment in western China.

*Corresponding author. Email: lixin@lzb.ac.cn

ISSN 1753-8947 print/ISSN 1753-8955 online

DOI: 10.1080/17538947.2011.558123
http://www.informaworld.com

Establishing a data center was considered as a means of integrated study and has been assigned the highest priorities by the NSFC West Plan. Therefore, in 2006, the NSFC decided to establish the Environmental and Ecological Science Data Center for West China (hereafter referred to as WestDC). This was the first fully functioning data center for earth science developed by the NSFC. The objective of the data center is to systematically archive the data collected by individual projects of the NSFC West Plan and to provide a good data service for the integrated study of environmental and earth science in western China using these and other data sets (Li *et al.* 2008b).

Through efforts made over several years, the WestDC has achieved its objective as an operational data center for environmental and ecological data, and more broadly, for geoscientific data in western China. It has also accumulated experience in data stewardship and has been providing services that are believed to be useful for data sharing in China. This paper will introduce the objectives, institutional and technical constructions, main data sets, data stewardship, and services provided by the WestDC.

2. WestDC – a geoscientific and environmental data center

2.1. Why the WestDC?

Data sharing is very important in helping science research to flourish (Nelson 2009). "Research cannot flourish if data are not preserved and made accessible. All concerned must act accordingly," as stated in the special issue of *Nature* on data sharing (Editorial, 2009). Data sharing is particularly important in China, as there is no strong culture of data sharing. Data have usually been retained by the scientists themselves rather than deposited in data centers. Fortunately, this situation has significantly improved in recent years. Data sharing has been emphasized by both policymakers and scientists, and several new data centers have been established in addition to the existing World Data Centers in China (Wang and Sun 2007). A nationwide scientific data-sharing program has been implemented by the Ministry of Science and Technology of China (Xu 2003), and a consensus has been reached that it is inappropriate to retain data as personal property (Huang and Guo 2002, Sun and Shi 2003, Ran *et al.* 2007).

It is one of the research sponsors' responsibilities to ensure the integrity, accessibility, and stewardship of research data (NRC 2009). As the most important agency that supports basic research in China, the NSFC has decided to play an institutional role in data sharing in China. The WestDC can be viewed not only as the data hub for the NSFC West Plan (i.e. a technical infrastructure) but also as a permanent data center featuring distinctive regional representativeness and characteristic data sets. In this way, it will help in creating a data-rich environment and even a data-sharing culture, thereby acting as an institutional infrastructure of the NSFC for providing data services.

The immediate goals of the WestDC that focus on western China are as follows: (1) collecting and standardizing the existing environmental and ecological data; (2) providing full and open data access, stewardship, and services in line with the NSFC West Plan and in support of other environmental researchers; (3) formulating effective guidelines for data sharing; and (4) facilitating data submissions from the

various projects covered by the NSFC West Plan and other organizations and individuals.

2.2. Data platforms in the WestDC

The technical infrastructure of the WestDC consists of four platforms: (1) data sharing; (2) a knowledge repository; (3) cooperation and exchange; and (4) data science (see Figure 1).

(1) Data sharing platform. This is a Web-based portal using the state-of-the-art computer technology. By reviewing existing geographic metadata standards, the ISO 19115 metadata profile has been adopted in the WestDC to serve the data-sharing requirement. Using the metadata, the portal enables user-friendly data display, navigation, and query, and offers both online and offline data services to users. To ensure the quality of metadata, a peer-review procedure is applied for metadata in the WestDC. Specific metadata are first edited by the data manager in the WestDC, and this information is then made available on a metadata interactive editing system created by GeoNetwork. The data authors and some experts are then invited to edit and correct the metadata online. It has been proven that metadata quality can be greatly improved with the input of expertise from two or more related experts. The metadata are then formally published. A feedback mechanism is also offered so that data users can comment on both the data and the associated metadata. The details of the technical architecture of the WestDC are introduced in Wang *et al.* (2010). The data-sharing platform also provides applicable online map services, catalog services, and other Web services.

(2) The knowledge repository platform is an attempt to integrate the data with bibliographic knowledge. Literature-based knowledge relating to data collection, generation, processing, and application is generally not fully represented by the current metadata architecture. Compared with the itemized metadata structure, literature knowledge is usually presented in a

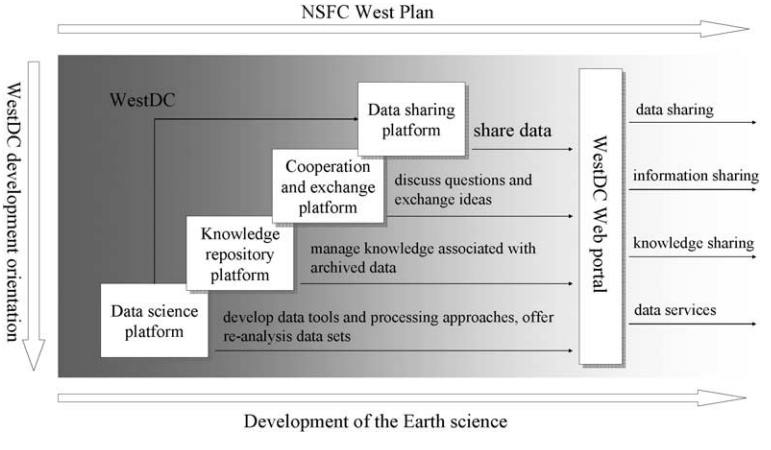

Figure 1. General scheme of the data platforms in the Environmental and Ecological Science Data Center for West China.

concrete scientific context but is essentially useful for understanding the science behind the associated data. These kinds of knowledge can be brought to users only by linking the data and the relevant literature. The knowledge repository platform was built upon the Open Archival Information System (OAIS) reference model. A digital knowledge resource library (the core of the knowledge repository platform) was established to cover the environmental and ecological research literature of western China (Zhu *et al.* 2007, Nan *et al.* 2010). To avoid copyright issues, this library was developed in collaboration with the Lanzhou Branch of the Chinese National Science Library. WestDC has implemented connections between the metadata, data documents, and published papers by keyword query based on the Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH) protocol. Such technically complicated connections are transparent to the end user.

(3) Cooperation and exchange platforms provide space to facilitate communication and the exchange of ideas using modern network tools, such as Web forums, mailing lists, and other current, advanced online conferencing approaches.

(4) The data science platform is designed as an advanced data processing and re-analysis component. It provides various data processing, provision, and assimilation tools and attempts to produce ready-to-use model data sets and high-resolution re-analysis data sets through data assimilation.

3. Main data sets

There are 331 data sets available in WestDC. The main data sets can be categorized as follows:

(1) Basic geographical data. These are defined as the fundamental data used in land surface hydrological, ecological, and other types of environmental analysis and modeling. These data were provided by research institutions and individual scientists or digitized from existing paper maps. Generally, they cover the entire land territory of China. The data include the following:

- Administrative boundaries, river basin boundaries, topography (digital topographic maps and DEM in raster format), transportation, rivers, lakes, reservoirs, and other water bodies. These data were archived using map scales from 1:100,000 to 1:4,000,000 for vector data and at resolutions from 100 m to 1 km for raster data.
- Land cover/land use data of China using a map scale of 1:100,000 (Liu *et al.* 2003) covering three periods: the late 1980s, 1999/2000, and 2005. A detailed description of the mapping procedure and classification system can be found in Liu *et al.* (2005).
- A soil map of China using a map scale of 1:1,000,000, based on attributes such as soil texture, bulk density, organic matter content, and other information (Shi *et al.* 2006).
- A vegetation map of China using a map scale of 1:1,000,000 (Editorial Board of Vegetation Map of China, Chinese Academy of Sciences 2001).

- A wetlands map of China using a map scale of 1:100,000 (Zhang 2002).
- A map of the lakes in China using a map scale of 1:100,000 (He *et al.* 2005).
- A new, 1-km land cover map using the IGBP (International Geosphere – Biosphere Programme) classification system based on integrating multi-source information including the land cover/land use, vegetation, and wetland maps introduced above, and the MODIS (Moderate Resolution Imaging Spectroradiometer) landcover map (Ran *et al.* 2009, 2010).
- A geomorphological map of western China using a map scale of 1:1,000,000 (e.g. Chai *et al.* 2009).
- Much high-resolution remote sensing data have been collected, such as the MSS/TM/ETM + and ASTER data in western China.

(2) Regionalization of global data sets for the land territory of China

- A subset of AVHRR Pathfinder NDVI (1981–2001), AVHRR GIMMS NDVI (1981–2006), and SPOT Vegetation NDVI (1998–2008) covering China's land territory.
- A subset of daily SMMR and SMM/I brightness temperature data covering China's land territory from 1978.
- A subset of NECP/NCAR, ECMWF ERA-40, and JMA JRA-25 re-analysis data.

(3) Scientific data for cold and arid regions in China. These refer to the Qinghai-Tibetan Plateau, the mountain cryosphere in western China, and arid regions (particularly the deserts). These data are usually derived from the results of scientific research in this region and are among the most innovative data in the WestDC. All the data were archived with detailed documentation.

- Daily snow depth data set for China derived from passive microwave remote sensing from 1978 to 2008 (Che *et al.* 2008) at a spatial resolution of 25 km.
- Daily surface soil freeze/thaw data set for China derived from passive microwave remote sensing from 1994 to 2008 (Jin *et al.* 2009) at a spatial resolution of 25 km.
- Chinese Glacier information System using a map scale of 1:100,000 (Wu and Li 2004). Both the glacier distribution maps and the attributes in the *Concise Chinese Glacier Inventory* (Shi 2005) were archived.
- Map of Chinese deserts using a map scale of 1:100,000 (Wang *et al.* submitted). This was compiled using more than 500 TM/ETM + images in 1999/2000.
- Historic permafrost maps of China as a whole and regional permafrost areas from the 1970s, using map scales ranging from 1:600,000 to 1:4,000,000 (e.g. Qiu *et al.* 2000, Wang *et al.* 2006).
- The Chinese cryosphere information system, which archives data relating to the cryosphere and related environmental factors for three test areas, that is the Qinghai-Tibetan Plateau, the Qinghai-Tibet Highway, and the Urumqi River Basin in the Tianshan Mountains (Li *et al.* 2003, 2008a).

- Large-scale desert and desertification maps of major deserts in China. Some of these maps date back as early as 1958.
- Large-scale topographic maps of some important glaciers in China, using map scales ranging from 1:10,000 to 1:100,000

(4) Scientific data for the cryosphere in countries neighboring China. These data include the following:

- Glacier and glacier lake inventories for Nepal, Bhutan, Pakistan, and the Indian Himalayas.
- Permafrost maps of Russia using a map scale of 1:2,500,000, Kazakhstan (map scale 1:10,000,000), and Mongolia (map scale 1:12,000,000).

(5) Heihe River Basin data, a comprehensive database relating to the Heihe River Basin, which is the second largest inland river basin in the arid region of China. There is also a separate Web site known as the Digital Heihe River Basin (http://heihe.westgis.ac.cn) devoted specifically to sharing data relating to the Heihe River Basin. This shares the same database as the WestDC.

- Basic data relating to the Heihe River Basin, including topography (DEM), land cover/land use map, soil map, other thematic maps, meteorological and hydrological (surface water and groundwater) observation data, socioeconomic statistics, and remote sensing data obtained using various satellite sensors.
- The Watershed Allied Telemetry Experimental Research (WATER) database. This archives data collected during the WATER campaigns. WATER is a simultaneous airborne, satellite-borne, and ground-based remote sensing experiment taking place in the Heihe River Basin (Li *et al.* 2009).
- Atmosphere-Land Surface Processes Experiment at the Heihe River Basin (HEIFE) database. HEIFE, as a WCRP/IGBP project and the first arid region, land surface experiment in Asia, was conducted from 1988 to 1992. Numerous data relating to boundary layer meteorology, micrometeorology, energy flux, and hydrology were collected (Hu and Gao 1994, Tao *et al.* 1994).

(6) Data submitted by NSFC West Plan projects. In total, 66 projects were conducted within the framework of the NSFC West Plan. Among them, 26 projects have submitted their data to the WestDC. After detailed documentation based on interactive discussions with the project principal investigator (PI) and co-PIs, specific data sets are then produced using the data submitted by one project or group of projects that have similar objectives. Of these data sets, some become immediately and freely available; others become publicly available after a 2-year privileged period. Some examples of these data sets are introduced below:

- Cyclic rapid warming on the centennial scale revealed by a 2650-year stalagmite record of warm season temperatures. A 2650-year (BC 665–AD 1985) warm season (May, June, July, and August) temperature reconstruction is derived from correlating thickness variations in annual layers of a stalagmite from Shihua Cave, Beijing, China, and instrumental meteorological records (Tan *et al.* 2003).
- An assimilated land surface hydrological data set for western China. This archives the data generated by the Land Data Assimilation System for western China (Li *et al.* 2007). The variables recorded include near surface air temperature, humidity, wind velocity and direction, radiation, precipitation, soil moisture, snow water equivalent, soil temperature and its freeze–thaw status, and energy fluxes at a spatial resolution of 1/4 degree and temporal resolution of 1 hour.
- A paleoclimate data set. This archives data from seven NSFC West Plan projects. These projects investigated the environmental evolution of western China. Numerous data, particularly proxy records from loess, stalagmite, ice-core, pollen, tree-ring, lithological sediments, isotope, and historical references, were collected.
- An ecohydrological data set for desert vegetation in western China. This archives data from four NSFC West Plan projects. These projects investigated the biological, ecological, morphological, and physiological characteristics of typical desert vegetation types in west China and their relationship with hydrology and climatology. Numerous experimental and long-term monitoring data were collected.

Other data sets are very useful for environmental, ecological, and other types of earth science research in western China. All of these data can be found on the official Web site at http://westdc.westgis.ac.cn. Currently, the Web site is available only in Chinese. However, support for the use of bilingual metadata (Chinese and English) has been designed into the system at a fundamental database level. The translation of all the metadata and major data documents into English is ongoing and is expected to be completed by mid-2011.

4. Data service

4.1. Data distribution

The WestDC complies with the "full and open" data sharing policy, and its data are accessible using two types of data services.

(1) Online data services. These are data sets that are accessible, together with their metadata and available data document(s), from the WestDC data portal. Since becoming operational in 2007, most of the data (725.4 GB, 90% of all data) in the WestDC have become downloadable, and online access keeps increasing: more than 7.0 terabytes of data have been downloaded as of September 2010. For instance, the subset of SPOT Vegetation NDVI for China has been downloaded more than 1540 times. The desert map of China

has been downloaded more than 730 times. The number of registered users has increased rapidly to over 3800 as of September 2010.

(2) Offline data services. In accordance with Chinese laws and regulations, some data, such as large-scale topographic maps and high-resolution DEMs, cannot be placed online. In addition, some scientists and institutions who have submitted data demand that access to these data remain offline to retain full awareness of the use of their data. Therefore, a small set of data is offered offline for which a contract needs to be signed between the user and the WestDC. The requested data will then be sent to the user via DVD or other media. The total amount of data provided offline by the WestDC accounted for 2.13 terabytes as of September 2010.

4.2. Data documentation, reference, and publishing

Traditionally, the data themselves have not been cited in scientific publications. Insufficient and informal citation of data hinders the willingness of researchers to submit data because it adds little credit to their academic career as compared with formal papers in peer-reviewed journals (Parsons *et al.* 2010). To encourage increased submission of data sets, the WestDC has instituted the following measures:

(1) Data identification. The digital object identifier (DOI) makes unique data identification possible. This makes protection of intellectual property of data much easier and data publication more likely. The WestDC is among the first Chinese data centers to attempt the use of DOI for data identification. In cooperation with the Institute of Scientific and Technical Information of China, DOIs are applied to uniquely identify each data set within the WestDC.

(2) Data documentation. Two types of data documents exist. The first is peer-reviewed papers that are published in connection with the scientific data sets, and these are usually requested to be cited when using the data. The second type refers to data documents that describe the data sets, usually prepared by the data center. The latter is considered as gray literature and is very important for ensuring data usability (Parsons and Duerr 2005). As such, it is imperative that the full scope of the metadata be provided, and the data documentation should (a) clearly describe the data collection and production process; (b) clearly describe the properties of the data and the related operations; and (c) include references, citations, and other related information, including any scientific papers related to the data. Such a data document must be prepared by the data managers at WestDC in consultation with the scientists who contributed the data via an iterative dialog. Data documentation is also encouraged to be cited.

(3) Data publishing. Whereas data identification and data documentation enable the citation of data, it is desirable that the data be published as peer-reviewed papers in scientific journals. This level of review will significantly increase interest in the data sets, and scientists will be much more willing to share their data (Editorial, 2009). Data can be published in two ways. The first way would be to establish a data journal to publish scientific data, as is done by some other

data centers. For example, EGU has published a new data journal, *Earth System Science Data*. The WestDC is also considering cooperating with some peer-reviewed scientific journals for data publication. The second way is to publish data on the data center's data portal; this is what most data centers, including the WestDC, are doing now, and by using DOIs, the data can be formally cited. Obviously, quality control (including proof-reading, correction, and peer review) is a prerequisite for publication. However, scientific journals are generally considered to be the best medium to foster data sharing among scientists.

(4) Data reference. The use of data should always be acknowledged. In the WestDC, the data reference information. In particular, the data contributors' and producers' names are emphasized in both metadata and data documents. A data user is required to make the necessary references to the data set(s) he or she uses, together with the document describing the data and related papers, in which the data contributors usually are major authors. In addition, users are encouraged to acknowledge the data center and this will add credibility to the WestDC.

4.3. Cooperation with other data centers

It is very important for the WestDC to cooperate with domestic and international data centers to provide a better data service for its users. It collaborates with other data centers in different ways.

WestDC has a close relationship with several world data centers (WDCs). The WDC for Glaciology and Geocryology in Lanzhou (http://wdcdgg.westgis.ac.cn) is affiliated with the same institute as the WestDC. The cryosphere data of western China are available at both data centers, with the WDC for Glaciology and Geocryology also providing information and data downloading services in English. The WestDC also collaborates closely with the WDC for Glaciology in Boulder, CO, which is maintained at the National Snow and Ice Data Center (NSIDC). An important data set of the WestDC, the subset of daily SMMR and SMM/I brightness temperature data for China from 1978, was derived from the DMSP Pathfinder Daily EASE-Grid brightness temperatures (Armstrong *et al.* 2010). Meanwhile, one of the flagship data sets in the WestDC, the *Concise Chinese Glacier Inventory*, contributes significantly to NSIDC's World Glacier Inventory. In NSIDC's Global Land Ice Measurements from Space (GLIMS) database, approximately 49% of the glacier records come from the *Concise Chinese Glacier Inventory*.

As for basic data, such as meteorological and hydrological data, WestDC archives these only for specific regions, for example typical inland river basins in northwestern China. Requests from users for these data for other regions are generally redirected to the data centers that host them. For example, meteorological data requests are redirected to the China Meteorological Data Sharing Service System.

In future, the WestDC should collaborate more broadly with other data centers. For example, several paleoclimate data sets (Tan *et al.* 2003) submitted by the NSFC West Plan project are also available at the WDC for Paleoclimatology.

Currently, all these cross-platform collaborations have been operated manually by the data managers at the different data centers. Potentially, these centers could

interoperate using open geospatial standards (Nativi 2010, Percivall 2010). Because a standard metadata format is used at the WestDC and other data centers, it should be easy for data centers to harvest metadata from each other and develop further interoperable functions. In future, the creation of a data center league is also possible.

5. Summary and prospects

Indeed, data sharing is important. However, this sharing requires the institutional and technical infrastructure to facilitate it. We have to develop new types of data centers and introduce new ideas regarding data reference and data publishing, both within China and globally, to deal with the challenges that new technologies bring. The WestDC, as NSFC's first fully functioning data center, should pioneer this transition in China.

This paper introduces the objectives, institutional and technical construction, main data sets, and data stewardship and services of the WestDC. When referring to data stewardship, we include data archival, preservation, and publishing but also involve efforts to make the data accessible, including the provision of data documents, the linking of data to a digital literature library, the preparation of ready-to-use model data sets and re-analysis data sets, and the provision of a number of Web services. After several years, the WestDC has achieved its objectives in regard to the following aspects: the provision of a technically sound data portal that aids in facilitating data sharing, the development of a knowledge repository and online scientific collaboration tools, the collection of large amounts of innovative scientific data that are useful in environmental, ecological, and (more broadly) geoscientific research in western China, the facilitation of stable and supportive data users, the generation of useful experiences in data stewardship and services, and the rewarding of efforts in data referencing, publishing, and sharing.

Some weaknesses remain in the WestDC. The data sets are not yet fully integrated. For example, the lack of high-resolution and vegetation and soil maps that are compatible with international standards are still a bottleneck in geoscientific research for west China. The WestDC should collaborate with the scientists who are working in this field to make these data available – this is also an important objective of the NSFC West Plan. Other issues that should be prioritized include forming consortia with other data centers, data publishing, developing a bilingual Web site (Chinese and English), and strengthening the visibility of the data center.

Data sharing in China needs to be improved further. To create a data-rich environment in which data sharing and associated referencing are more common, many difficult issues remain to be overcome, and considerable actions need to be taken. The WestDC, as a representative data center of NSFC and a practitioner of data science, should be and is able to pioneer this effort. However, to grow from a data center of regional importance to one of global importance, the WestDC needs to make further efforts in collaborating more closely with scientists and scientific policymakers. "Data management should be woven into every course in science" (Editorial 2009). It is incumbent upon the entire scientific community to ensure that we foster the scientific data infrastructure required for scientific research to flourish and ultimately be a benefit to society, to China and the world.

Acknowledgements

This work is financially supported by the NSFC (National Science Foundation of China) (grant number: 40925004) and the Chinese Academy of Sciences Action Plan for West Development Project "Watershed Allied Telemetry Experimental Research (WATER)" (KZCX2-XB2-09). We thank the editor and the anonymous reviewers for their helpful and constructive comments on the manuscript.

Notes on contributors

Xin Li received his PhD degree from the Chinese Academy of Sciences (CAS) in 1998. He is currently a professor at the Cold and Arid Regions Environmental and Engineering Research Institute (CAREERI), CAS. He is also the director of the World Data Center for Glaciology and Geocryology at Lanzhou. His primary research interests include land data assimilation, application of remote sensing and GIS in hydrology and cryosphere science, and integrated watershed modeling.

Zhuotong Nan received his PhD degree in physical geography from the CAS in 2003. He was a post-doc research associate in the University of Pittsburgh, USA, in 2007–2009. Currently he holds an associate professor position at CAREERI. His research interests are related to data sharing, integrated modeling environment, and decision support systems for water resources management.

Guodong Cheng is currently the professor of CAREERI and the director of Lanzhou branch of CAS. He became an academician of CAS in 1993. He has been working for permafrost and related engineering problems for over 40 years and was the chief scientist of the Qinghai-Tibet railway research project. His current research interests are in hydrology and water resources in arid regions, and environment and engineering in cold regions.

Yongjian Ding received his MS degree in physical geography from CAS in 1987. He is currently the professor, doctoral advisor, and vice director of CAREERI. His research interests include glacier and periglacial environment, hydrology and water resources, and hydrogeology and engineering geology.

Lizong Wu received his PhD degree in remote sensing and GIS from CAS in 2010. Currently he holds an associate professor at CAREERI. His research interests are related to data sharing, glacier inventory, and glacial lake outburst flood.

Liangxu Wang received his MS degree in remote sensing and GIS from the CAS in 2005. Currently he holds an assistant professor at CAREERI. His research interests are related to scientific data sharing, land data assimilation, and high performance computing in scientific research.

Jian Wang received his PhD degree in remote sensing and GIS from the CAS in 2004. He is currently a professor at CAREERI. His research interests are related to remote sensing of snow and ice, hydrology of cold region, and integrated research of cryosphere and climate change.

Youhua Ran received the BS degree from the CAS in 2008. His main research interests are the application of the remote sensing and GIS in ecosystems analysis and modeling; land cover, plant function types and permafrost mapping; spatiotemporal analysis of remotely sensed data.

Hongxing Li is currently a PhD student of the CAS. Her research interest is related to snow pollution research by remote sensing.

Xiaoduo Pan is currently a PhD student of the CAS. Her current research interests are development of atmospheric forcing data for land surface model and downscaling methods.

Zhongming Zhu is a research scientist, the head of information technology department of the Scientific Information Center for Resources and Environment, CAS. His research interests include digital information environment and metadata applications, digital objects and repositories, knowledge management in academic and research institutions, and application of semantic web technologies to information integration.

References

Armstrong, R.L., *et al.*, 2010. *DMSP SSM/I Pathfinder Daily EASE-Grid brightness temperatures*. Boulder, CO: National Snow and Ice Data Center. Digital media.

Chai, H.X., *et al.*, 2009. Digital regionalization of geomorphology in Xinjiang. *Journal of Geographical Sciences*, 19 (5), 600–614. doi: 10.1007/s11442-009-0600-4.

Che, T., *et al.*, 2008. Snow depth derived from passive microwave remote sensing data in China. *Annals of Glaciology*, 49, 145–154.

Editorial Board of Vegetation Map of China, Chinese Academy of Sciences, 2001. *Vegetation atlas of China* (1:1,000,000). Beijing: Science Press.

Editorial, 2009. Data's shameful neglect. *Nature*, 461 (7261), 145–145. doi: 10.1038/461145a.

He, L.H., Jiang, N., and Shi, J.J., 2005. On the design of China lake database. *Modern Surveying and Mapping*, 28 (5), 3–5.

Hu, Y.Q. and Gao, Y.X., 1994. Some new understandings of processes at the land surface in arid area from the HEIFE. *ACTA Meteorologica Sinica*, 52 (3), 285–296.

Huang, D.C. and Guo, Z.Y., 2002. *Study on the management of scientific data sharing*. Beijing: China Science and Technology Press. 389 pp.

Jin, R., Li, X., and Che, T., 2009. A decision tree algorithm for surface soil freeze/thaw classification over China using SSM/I brightness temperature. *Remote Sensing of Environment*, 113 (12), 2651–2660. doi: 10.1016/j.rse.2009.08.003.

Li, X., *et al.*, 2003. Modeling Chinese cryospheric change by using GIS technology. *Cold Regions Science and Technology*, 36 (1–3), 1–9.

Li, X., *et al.*, 2007. Development of a Chinese land data assimilation system: Its progress and prospects. *Progress in Natural Science*, 17 (8), 881–892.

Li, X., *et al.*, 2008a. Cryospheric change in China. *Global and Planetary Change*, 62 (3–4), 210–218. doi: 10.1016/j.gloplacha.2008.02.001.

Li, X., *et al.*, 2008b. Environmental and ecological science data center for west China: Integration and sharing of environmental and ecological data. *Advances in Earth Science*, 23 (6), 628–637.

Li, X., *et al.*, 2009. Watershed allied telemetry experimental research. *Journal of Geophysical Research*, 114, D22103. doi: 10.1029/2008JD011590.

Liu, J.Y., *et al.*, 2003. Study on spatial pattern of land-use change in China during 1995–2000. *Science in China Series D-earth Sciences*, 46 (4), 373–384 + Plate III.

Liu, J.Y. *et al.*, 2005. Spatial and temporal patterns of China's cropland during 1990–2000: An analysis based on Landsat TM data. *Remote Sensing of Environment*, 98 (4), 442–456.

Nan, Z.T., *et al.*, 2010. Experiences of knowledge integration from the environmental and ecological science data center for west China. *China Science & Techonology Resources Review*, 42(5), 15–21. (In Chinese)

Nativi, S., 2010. The implementation of international geospatial standards for earth and space sciences. *International Journal of Digital Earth*, 3 (Suppl. 1), 2–13. doi: 10.1080/17538941003764412.

Nelson, B., 2009. Data sharing: Empty archives. *Nature*, 461 (7261), 160–163. doi: 10.1038/461160a.

NRC, 2009. *Ensuring the integrity, accessibility, and stewardship of research data in the digital age*. Washington, D.C.: National Academies Press. 152 pp.

Parsons, M.A. and Duerr, R., 2005. Designating user communities for scientific data: Challenges and solutions. *Data Science Journal*, 4, 31–38.

Parsons, M.A., Duerr, R., and Minster, J.-B., 2010. Data citation and peer review. *EOS*, 91 (34), 297–298. doi: 10.1029/2010EO340001.

Percivall, G., 2010. The application of open standards to enhance the interoperability of geoscience information. *International Journal of Digital Earth*, 3 (Suppl. 1), 14–30. doi: 10.1080/17538941003792751.

Qiu, G.Q., *et al.*, 2000. *The map of geocryological regionalization and classification in China*. Beijing: Science Press. (In Chinese)

Ran, Y.H., Li, X., and Lu, L., 2009. China land cover classification at 1 km spatial resolution based on a multi-source data fusion approach. *Advances in Earth Science*, 24 (2), 192–203.

Ran, Y.H., Li, X., and Lu, L., 2010. Evaluation of four remote sensing based land cover products over China. *International Journal of Remote Sensing*, 31(2), 391–401. doi: 10.1080/01431160902893451.

Ran, Y.H., Li, X., and Wang, J., 2007. The current key problems and potential solutions for geosciences data sharing in china. *Data Science Journal*, 6 (Supp l), S250–S254.

Shi, X.Z., *et al.*, 2006. Cross-reference system for translating between genetic soil classification of china and soil taxonomy. *Soil Science Society of America Journal*, 70 (1), 78–83. doi: 10.2136/sssaj2004.0318.

Shi, Y.F., 2005. *Concise Chinese glacier inventory*. Shanghai: Shanghai Popular Science Press. 194 pp. (In Chinese)

Sun, J.L. and Shi, H.Z., 2003. Building up a scientific data sharing network for the Earth system science in China. *China Basic Science*, (1), 75–81.

Tan, M., *et al.*, 2003. Cyclic rapid warming on centennial-scale revealed by a 2650-year stalagmite record of warm season temperature. *Geophysical Research Letters*, 30 (12), 1617. doi: 10.1029/2003GL017352.

Tao, Z.H., Zuo, H.C., and Hu, Y.Q., 1994. HEIFE data base (HDB). *Plateau Meteorology*, 13 (3), 369–376.

Wang, J.H., *et al.*, (submitted). A large-scale (1:100,000) desert map of China. *Journal of Arid Environments*.

Wang, J.L. and Sun, J.L., 2007. Development of China WDC systems for data sharing. *China Basic Science*, (2), 36–40. (In Chinese)

Wang, T., Wang, N.L., and Li, S.X., 2006. *Map of the glaciers, frozen ground and desert in China, 1:4,000,000*. Beijing: Chinese Map Press. (In Chinese)

Wang, L.X., *et al.*, 2010. Application of open source technologies in geoscientific data centers. *China Science & Technology Resources Review*, 42 (3), 17–23. (In Chinese)

Wu, L.Z. and Li, X., 2004. *China glacier information system*. Beijing: Ocean Press of China. 135 pp. (In Chinese)

Xu, G.H., 2003. Implement the scientific data sharing program and enhance the national science and technology competitiveness. *China Basic Science*, (1), 5–9. (In Chinese)

Zhang, S.Q., 2002. Brief introduction of the wetland science database of China. *Scientia Geographica Sinica*, 22 (2), 189–189. (In Chinese)

Zhu, Z.M., *et al.*, 2007. A Dspace-based sharing enviromental and ecological knowledge space. *Library and Information Service*, 51 (4), 71–74, 108. (In Chinese)