Boosting geoscience data sharing in China

Enabling public sharing of scientific data in China not only needs top-down mandates but also incentive mechanisms that boost confidence and willingness to engage in data-sharing practices among Chinese researchers.

Xin Li, Guodong Cheng, Liangxu Wang, Juanle Wang, Youhua Ran, Tao Che, Guoqing Li, Honglin He, Qiang Zhang, Xiaoyi Jiang, Ziming Zou and Guofeng Zhao

Data flow and information sharing are important for developing a borderless scientific community. However, policy, culture and technical barriers have obstructed the free flow of scientific data. To ensure research integrity and maximize research impact, good data sharing and management should be prioritized¹.

Geoscience data sharing is particularly important for addressing today's most challenging and unprecedented global environmental problems, such as climate change, and for achieving the United Nations Sustainable Development Goals². For example, as early as the 1990s, twelve distributed active archive centres were launched by the National Aeronautics and Space Administration of the United States to store data and information in climate research. The World Data System currently has 86 data centres, 57% of which are in the Earth sciences.

As China plays an increasingly important role in tackling these challenges, the country has recently adopted a more proactive policy in data sharing and transparency. For example, in 2018, the Chinese Ministry of Science and Technology released a policy that sets data sharing as a principle for research funded by the government³. This policy has now brought online first-generation national data centres, ten of which are focussed on the Earth and environmental sciences (Supplementary Material 1).

But to what extent will these actions really boost the practice of data sharing in China? A recent survey of more than 2,000 Chinese researchers reveals both opportunities and challenges⁴. The survey showed that while researchers in China are willing to share research data, they are concerned about misuse of data and violation of copyright and licensing⁴. Instead of wider public sharing, private sharing of data with immediate colleagues and collaborators is more common in China⁴. This suggests that a lot of work is still needed to increase the visibility of new data centres and to build confidence in

Fig. 1 | Reinforcing feedback for data sharing by encouraging and benefiting both data providers and data users. Data centres are the mediators that link policy makers, data contributors, data, and data users in the ecosystem of data sharing. In this system, the purpose of data centres is to provide good management that can turn the loop into reinforced feedback that eventually benefits science and society.

data-sharing practices more broadly among Chinese researchers.

Two pioneering projects shed light on the key to boosting wider sharing of scientific data in China. The first involved two Earth science programmes funded by the National Natural Science Foundation of China, where data sharing was a mandate. The project required all data obtained from the programmes to be deposited in the foundation's geoscience data centre for public access and data reuse. The data submission and data quality were evaluated during annual, interim and final evaluations of the project. Most importantly, the key mechanism was giving credit to data contributors by clearly acknowledging their contribution through data citations via data DOIs and the associated paper publications. To date, more than 2,500 scientific papers have cited these datasets (Supplementary Material 2). The legacy of this programme is mandatory

data sharing, credit to data contributors, and respect for intellectual property.

Another project was launched by the Chinese Academy of Sciences (CAS), called CASEarth (<http://data.casearth.cn/>). The project aims to build a cyberinfrastructure for data on the Earth, environmental, ecological and biological sciences⁵. By collecting data from CAS institutions, the CASEarth repository has now stored more than 5 PB of data, and its data have been downloaded more than 500,000 times.

The success of these pioneering projects suggests that policies that support public data sharing from the top down, and bottom-up incentives that credit data contributors, are key to enabling wider data-sharing practices in China. We also argue that more specific actions are needed in policy, management and technological aspects to roll out data-sharing mandates more broadly in the big data era in China.

Policy

Clearer policies on data sharing and restrictions are needed. In particular, it's crucial to set clear rules for classified information⁶. Geoscience data can be sensitive in nature, especially those with regards to national security, business secrets and individual privacy. To maximize data-sharing practices, it is important to have a clear definition of sensitive data and specific rules for their sharing limitations and restrictions. For data outside the sharing restrictions, sharing practices should be fully based on findability, accessibility, interoperability and reusability (FAIR) principles^{7,8}. Additionally, new protocols of intellectual property for open science, such as creative commons, should be introduced as a commonplace.

Management

Data-sharing practices should be incentivized by fully crediting data contributors. First, the evaluation mechanism should be changed to credit the success of a researcher or grant not only on the basis of publications but also on data availability and the quality of the shared data. Second, data centres should incentivize data contributors by promoting data publication and citation, and track data use by quantifying the impact of each specific dataset with data-reuse metrics⁹. Only if data contributors are properly evaluated, credited and encouraged can data sharing be turned into a voluntary practice.

Technology

In the big data era, the role of data centres needs to change from data warehouses to smart information providers. For example, Google Earth Engine, a platform for analysing geospatial information on the basis of Google Earth, sets a good example for such a transition. By incorporating the technologies that emerge from big data and machine learning, data centres can turn big data into useful information and knowledge that serves users more efficiently. Data centres can also cooperate and interoperate with publicly available tools, for example, Google Dataset Search¹⁰, to make data more findable, widely accessible and friendly to both humans and machines. Additionally,

data provision should be strengthened, and automatic data quality control and smart information services should be enabled using artificial intelligence.

Internationalization

Currently, less than 10% of the geoscience metadata created in China are available in English (Supplementary Material 3). This language issue has prevented data centres in China from absorbing datasets more widely due to low visibility. To maximize data visibility and reuse, all major geoscience data centres in China should be encouraged to publish metadata and data bilingually in both Chinese and English. Developing data search tools in English for data centres is also essential.

As a hub in the system of data sharing (Fig. 1), data centres play a key role in realizing many of the above proposed actions. Numerous data centres in China have already been working with international communities to advance data sharing. For example, the National Tibetan Plateau/Third Pole Environment Data Center (<https://data.tpdc.ac.cn/en/>) has signed up to the Enabling FAIR Data Project and DataCite, and the National Earth System Science Data Center and National Space Science Data Center have been certified by CoreTrustSeal.

The data explosion in the big data era has posed both challenges and opportunities to the global geoscience community. Making research data FAIR is essential to tackling challenges and leveraging data resources. While progress has been made in public data sharing in China, vigorous efforts from government, researchers and data centres are still needed to achieve a paradigm shift. The more we honour the data and the data creators, the more we benefit science and society.

Xin Li¹✉, Guodong Cheng^{1,2,3}, Liangxu Wang^{1,2}, Juanle Wang⁴, Youhua Ran^{1,3}, Tao Che^{1,3}, Guoqing Li⁵, Honglin He⁶, Qiang Zhang⁷, Xiaoyi Jiang⁸, Ziming Zou⁹ and Guofeng Zhao¹⁰

¹National Tibetan Plateau Data Center, State Key Laboratory of Tibetan Plateau Earth System, Resources and Environment, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China. ²Institute of Urban Study, Shanghai Normal University, Shanghai, China. ³Cold and Arid

Region Data Center, World Data System, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, China.

⁴National Earth System Science Data Center, State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. ⁵National Earth Observation Data Center, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. ⁶National Ecological Science Data Center, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. ⁷National Meteorological Information Center, China Meteorological Data Service Center, China Meteorological Administration, Beijing, China. ⁸National Marine Data Center, China Oceanic Information Center, Ministry of Natural Resources of China, Tianjin, China. ⁹National Space Science Data Center, National Space Science Center, Chinese Academy of Sciences, Beijing, China. ¹⁰National Earthquake Data Center, China Earthquake Networks Center, China Earthquake Administration, Beijing, China.

✉e-mail: xinli@itpcas.ac.cn

Published online: 5 August 2021

<https://doi.org/10.1038/s41561-021-00808-y>

References

- Hanson, B., Sugden, A. & Alberts, B. *Science* **331**, 649 (2011).
- Alexander, S. M. et al. *Nat. Sustain.* **3**, 81–88 (2020).
- Scientific Data Administrative Measures* (State Council of China, 2018); http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm
- Lucraft, M., Allin, K., Baynes, G. & Sakellariopoulou, R. *Challenges and Opportunities for Data Sharing in China* (Springer Nature, 2020).
- Guo, H. *Bull. Chin. Acad. Sci.* **33**, 818–824 (2018).
- Chen, R., De Sherbinin, A. & Ye, C. *Science* **348**, 981 (2015).
- Stall, S. et al. *Nature* **570**, 27–29 (2019).
- Nat. Geosci.* **12**, 961 (2019).
- Pierce, H. H., Dev, A., Statham, E. & Bierer, B. E. *Nature* **570**, 30–32 (2019).
- Castelvecchi, D. *Nature* **561**, 161–162 (2018).

Acknowledgements

We are grateful to generous help from V. Khodiyar and X. Jiang.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41561-021-00808-y>.

Peer review information *Nature Geoscience* thanks Ruishan Chen and Shelley Stall for their contribution to the peer review of this work. Primary Handling Editor: Xujia Jiang.