




GDMD: A Disaster Data Hub Fabric the Open Data Network for Disaster Risk Reduction

RESEARCH PAPER

JUANLE WANG 
KUN BU
XIAODONG MIN

QIUYUAN WANG
YUELEI YUAN
FEIRAN SUN



*Author affiliations can be found in the back matter of this article

ABSTRACT

Disaster risk reduction is a priority in the Open Science recommendations proposed by UNESCO. Diverse types of scattered disaster data are the basis for disaster prevention and mitigation. The realization of these interconnections and links remains challenging. This study proposes an open directory mechanism that facilitates the interoperability of global disaster data resources, named Global Disaster Data Master Directory (GDMD). Based on open standards and open-source technology, the prototype GDMD system was developed and deployed in the Disaster Risk Reduction Knowledge Service (DRRKS) system at the UNESCO Category 2 centre. The GDMD realizes three types of disaster data resource aggregation: local resources built by DRRKS, manual cataloging from open disaster platforms worldwide, and protocol harvesting based on interoperable technology. Two types of retrieval services are implemented for users through local and federal searches. The master directory has interconnected over 1,400 resources. The open model of GDMD can realize Findable, Accessible, Interoperable, and Reusable data governance in disaster risk reduction, further pushing the Open Science mechanism and innovative applications of resources.

CORRESPONDING AUTHOR:

Juanle Wang

State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; China-Pakistan Earth Science Research Center, Islamabad 45320, Pakistan; Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

wangjl@igsnr.ac.cn

KEYWORDS:

Open Science; data sharing; data catalog; metadata; FAIR principle

TO CITE THIS ARTICLE:

Wang, J., Bu, K., Min, X., Wang, Q., Yuan, Y. and Sun, F. 2025 GDMD: A Disaster Data Hub Fabric the Open Data Network for Disaster Risk Reduction. *Data Science Journal*, 24: 8, pp. 1–12. DOI: <https://doi.org/10.5334/dsj-2025-008>

Disaster data are ubiquitous worldwide. In the face of emergency crises, a significant amount of the available data fails to play a role in disaster relief because they cannot be accessed promptly. Open disaster data form the foundation of Disaster Risk Reduction (DRR) activities. Although the nature of disaster data is massive and diverse, helping users find and access these open disaster data on time remains a huge challenge.

Disaster data are usually stored in a professional disaster database or related institutional platform. The USGS earthquake database (<https://earthquake.usgs.gov>) records the global real-time earthquake disaster events and provides an online access interface. The Fire Information Resource Management System (<https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms>) provides fire data in NASA Earth observation data. EM-DAT (<https://www.emdat.be>) is a historical disaster database recording the occurrence and impact of more than 22,000 mass disaster events globally since 1900 (Wang et al., 2019). The China Geological Survey (<https://www.cgs.gov.cn/>) has a geological cloud platform covering geological disaster resources, such as geo-hazards-related geological data, publications, standards, patent software, and popular science materials. The Global Assessment Report (GAR) data risk platform (<https://risk.preventionweb.net/>) shares spatial data concerning natural disasters, displaying multiple types of global disaster data visually. DesInventar (<https://www.desinventar.net/>) is a disaster information management system supporting the Sendai Framework implementation. Sigma (<https://www.swissre.com/institute/research/sigma-research.html>), part of the Swiss Re-insurance database, is open to the public and contains data records from 1970 to the present, originating from newspapers, direct insurance and reinsurance journals, publications, and reports. The Munich Re Disaster Database Natcat (<https://www.munichre.com/en/solutions/for-industry-clients/natcatservice.html>) records approximately 28,000 global disaster losses from 1980 to the present, mainly from the insurance industry, research institutions, governments, and non-governmental organizations. These DRR-related facilities are independent, scattered globally, and cannot be linked despite having similar disaster data products.

With the emphasis on fundamental data facilities for disaster prevention and mitigation, a more comprehensive database is needed. Several countries and agencies have established disaster databases. Among them, the Disaster Risk Reduction Knowledge System in the category 2 centre of UNESCO has built multiple kinds of disaster knowledge bases, including thematic databases for datasets, maps, videos, reports, experts, institutions, etc. (Wang et al., 2020a). These disaster resources are not only used for natural hazards reduction but also used for pandemic disaster control supported by social media big data mining (Wang et al., 2020b). As different databases are built by different institutions, these data cannot be openly accessed with a universal policy and are not easily connected by users or application systems. Users may be familiar with one or more of these databases but do not know other related datasets. Thus, it is inconvenient for them to retrieve and obtain the data required for an emergency response quickly. The facilitation of publicly accessible disaster data remains a common requirement.

Various disaster data-sharing mechanisms have been promoted by global and regional organizations through international cooperation. The United Nations Disaster Management and Emergency Response Space Information Platform has a 'Links and Resources' column that showcases disaster data and related software tool links. The International Mountain Comprehensive Development Center (ICIMOD, <https://regional.icimod.org/>) is a cross-regional intergovernmental knowledge-sharing center that provides DRR resource services to eight member states of the Hindu Kush Himalayas. The Global Facility for Disaster Reduction and Recovery (<https://www.gfdrr.org/en>) is an international organization that helps developing countries better understand and reduce their vulnerability to natural disasters and climate change. These cooperation networks are usually fixed for a specific scope and cannot be scaled to larger areas or more partners. A dynamic disaster data catalog discovery mechanism is missing from these distributed databases.

Faced with increasingly open and interconnected data user requirements, weaving a dynamic and updated open and interconnected data network is crucial for ubiquitous disaster data. Under the Open Science concept recommended by UNESCO, these dark data can be rekindled, improving all disaster prevention and mitigation efforts (UNESCO, 2011; Zheng, 2022). This study proposes an open sharing network of widely available disaster risk reduction data resources through the construction of the Global Disaster Data Master Directory (GDMD), playing an important role as a clearinghouse, hub, and master catalog for open disaster data sharing.

2.1 CONCEPT DEFINITION AND DEVELOPMENT TOOL

The GDMD refers to the open directory of interconnected global disaster data. It is an exchange system for global disaster data and an online tool for the presentation, search, release, mining, and analysis of a disaster data directory (Wang et al., 2014). Its function is mainly reflected in subject-related and rich metadata directory content through text or spatial location retrieval in the interconnected data portal. These data directories include locally published data, aggregated open data manually collected using the universal metadata standard, and they automatically harvest acquired metadata using an application programming interface (API). Among them, data from the federal database is the advantage of the open data master directory. In other words, the data hub can dynamically and automatically harvest data directories that keep them continuously updated.

The GDMD system design and development follows the open standard, using the open-source software system, specifically using the CSW (Catalog Service for Web) specification of the Open Geospatial Consortium (OGC) (Sun et al., 2012). The OGC catalog service is used to define access interfaces for geospatial data, services, or other related resources. Several OGC Web Service implementation specifications have been proposed that enable the integration and interoperability of datasets and are widely used in geospatial data sharing (Open Geospatial Consortium Inc, 2024). The CSW is the implementation specification for the network catalog service launched by the OGC, which describes the principles and standards of geospatial data and service discovery, access, and management (Open Geospatial Consortium Inc, 2024). It also specifies the interface, binding protocols, and framework of the directory service. As a digital catalog service, CSW can be deployed independently as servers or embedded in other applications (Sibolla et al., 2014).

GeoNetwork and pycsw are supporting platforms that provide OGC CSW open directory services. GeoNetwork is an early proposed, standard-based, free, and open geographic information directory service network application platform (Open Geospatial Consortium Inc, 2024). The pycsw is an OGC network catalog service standard implemented in Python. Originally released in 2010, it is now the most recommended tool for metadata management by OGC CSW. Its main functions include the retrieval, query, and browsing of spatial data. By supporting a variety of APIs (e.g., CSW2, CSW3, OAI-PMH, etc.), pycsw enables users to easily publish and discover geospatial metadata. Compared to GeoNetwork, pycsw has more advantages. Firstly, pycsw is a lightweight application platform, which is easily deployed and configured. Secondly, pycsw supports more types of metadata schemas than that of GeoNetwork, especially for those new metadata standards. Lastly, pycsw has a special tool for metadata generation automatically, while GeoNetwork can't. This study selected pycsw to provide the GDMD disaster information directory service.

2.2 DISASTER METADATA STANDARD

Metadata are the core content for data detail description in a master data catalog and are the textual information describing the data, facilitating quick understanding and the automatic discovery of the data by human or machine users. Metadata have different granularities. Fine-scale metadata cannot only describe the data content but also involve data production, processing methods, preservation, usage, and long-term management guidance. Typically, metadata have the functions of describing, retrieving, selecting, locating, integrating, cataloging, managing, and evaluating resources.

Owing to the wide range of disciplines involved in disaster prevention and mitigation, there are often various metadata standards for disaster data (Zhao and Wang, 2015). Open and universal metadata standards, such as Dublin Core Metadata and the ISO geographic information metadata standard, are recommended by most data management systems or platforms (Chen and Wang, 2007; Zhang and Xing, 2015). Following this principle, this study referred to the core metadata used in the Dublin Core Metadata and the ISO 19115 Metadata Standard (Wang et al., 2018). This international standard has wide compatibility for those data with geospatial information, like disaster data. The proposed disaster metadata elements for the master directory are listed in Table 1.

NUMBER	ELEMENT	DESCRIPTION	MANDATORY
1	Dataset name	Unique identity name of the dataset	Yes
2	Subject classification	Data classification within specific fields or subjects	Yes
3	Data sources	Original source or production agency information of the dataset	Yes
4	Spatial coverage	Spatial scope covered by the data, including longitude and latitude coordinates and administrative divisions	Yes
5	Temporal coverage	Time span of the data, including the start and end time	Yes
6	Resolution	Spatial resolution and time granularity of the data, such as hours, days, or months	Yes
7	Data type	The nature of the data, such as remote sensing images, vector data, and text data	Yes
8	Data format	Storage formats of the data, such as GeoTIFF, Shapefile, and CSV	Yes
9	Variable description	Description of each variable in the dataset, including name, type, and unit	Yes
10	Abstract	Basic information of the data, usually comprehensive and specific	Yes
11	Keyword	Words or phrases used for retrieving	Yes
12	Data service type	Description of the data access approaches, such as data file, database, and map access	Yes
13	Data sharing policy	Description of the rules and procedures for data sharing	Yes
14	Metadata created date	Record of the date of the creation of the metadata	Yes
15	Metadata update time	Last update time record of the metadata	Yes
16	Data responsible unit	Personnel or agency designation responsible for the update and maintenance of metadata	Yes

Table 1 Metadata elements for the Global Disaster Data Master Directory.

2.3 DEVELOPMENT OF THE GDMD

GDMD provides metadata interoperability based on network interconnection to realize one-stop access and retrieval of resources and provides external directory services. Federated retrieval and metadata directory local services form the fundamental basis of the protocols and techniques for GDMD resource integration services. Federated retrieval is an ideal method for building a global data network. It is a type of information retrieval technology that uses a unified interface to help users search for and simultaneously access and retrieve information distributed in multiple data sources (such as databases, knowledge bases, and websites). However, the implementation of this method depends on the infrastructure in which each node deploys a directory service, which determines whether interconnection among different nodes can be achieved. Second, the metadata directory is deployed locally, and the service uses the OGC standard metadata protocol. The OGC standard provides a reasonable mechanism for geospatial data interconnection and interoperability. Although there is still a lag in these types of applications, they are sufficient to support metadata exchange and harvesting between data centers.

The core development of GDMD is based on the pycsw metadata server. Pycsw realizes all types of data services and protocols based on a variety of international general standards and uniquely identifies metadata records. In the design of the GDMD, pycsw is responsible for achieving federated retrieval, gathering, and harvesting metadata information from other data centers in its metadata catalog service framework (Figure 1). The GDMD uses PostgreSQL for database persistence and adds a PostGIS extension to support spatial features, enabling complex query optimization, complete data type support, JSON and extensible markup language (XML) data processing, full-text searching, and geospatial data processing. Server-side applications were developed using the Tornado Web framework, providing users with diversified operations, such as browsing and data download (Liu, 2016). Pycsw provides a full directory service but also queries the related metadata information and returns the query results for users based on their retrieval requirements.

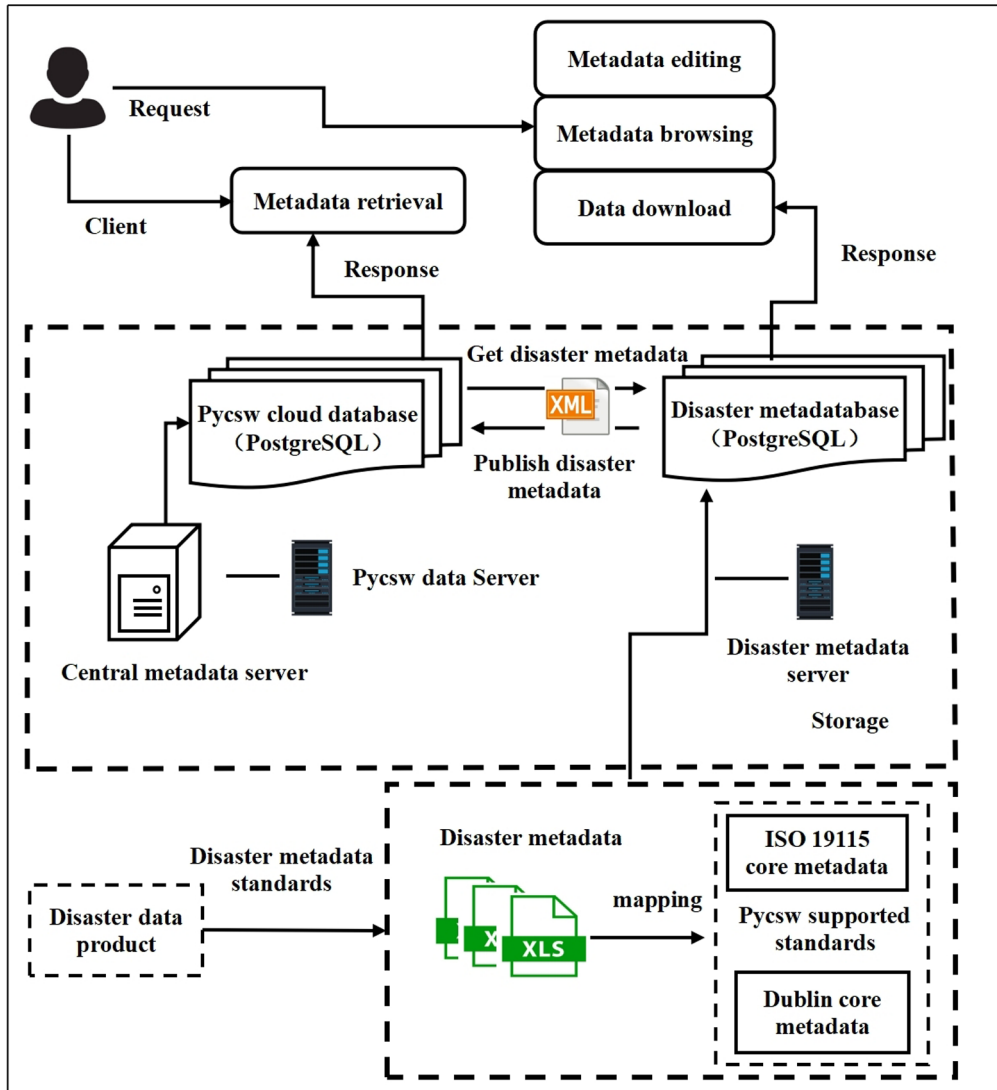


Figure 1 System architecture diagram of the GMDM.

2.4 CONTENT CONSTRUCTION OF THE GMDM

The GMDM prototype establishes a basic global disaster metadata module for metadata database management and sharing using classification catalog design and visualization technology. All metadata resources were encoded in the XML format, and the main coding tool was pycsw. The metadata compilation includes the required attribute information such as the dataset name, abstract, format, subject, keywords, and source (Table 1). All datasets were classified according to the universal classification system of disaster data in the GMDM (Table 2). There were eight first-level and 52 second-level classes. After metadata coding and classification, they can be archived in the GMDM and accessed after being retrieved by the users.

LEVEL 1 CLASS	SECOND CLASS	LEVEL 1 CLASS	SECOND CLASS
Geohazards	Earthquake	Hydrometeorological hazards	Drought (hydrological, meteorological, agricultural, socio-economic)
	Volcanic		Flood
	Ground collapse		Typhoon
	Collapse		Rainstorm
	Landslide		Low temperature
	Ground subsidence		Wind
	Mudslide		Hail
	Ground fissuring		Lightning
	Others		Sand-dust storm

Table 2 Classification of disaster data in GMDM.

LEVEL 1 CLASS	SECOND CLASS	LEVEL 1 CLASS	SECOND CLASS
Maine hazards	Storm tide		Snow storm
	Ocean wave		High temperature
	Sea ice		Fog
	Tsunami		Others
	Others	Technological hazards	CBRNE (chemical, biological, radiation, nuclear, or explosion)
Biological hazards	Plant pests		Infrastructure failure
	Disease		Cyber hazard
	Rodent pest		Industrial failure/Non-compliance
	Weed		Waste
	Red tide		Transportation
	Wildfires		Others
	Others	Societal hazards	Conflict
Environmental hazards	Water and soil loss		Post-conflict
	Desertification		Behavioral
	Salinization		Economic
	Stony desertification		Others
	Others	Other	Others

3. IMPLEMENTATION AND APPLICATION OF GDMD

3.1 CONSTRUCTION OF DATA DIRECTORY RESOURCES

The resource collection in the GDMD can be divided into three sources, as shown in Figure 2. The first part comprises the self-built resources of the host institution, the Disaster Risk Reduction Knowledge Service System of the International Knowledge Centre for Engineering Science and Technology auspices of UNESCO (IKCEST-DRR). In 2016, the IKCEST-DRR disaster database was built; since then, the disaster datasets have been accumulated (Wang et al., 2018). These disaster data are distributed in different countries and regions but are cataloged in a universal metadata standard and can be accessed online with a visualization map. All datasets were converted into GDMD by field mapping in the database.

The second originates from global disaster platforms with open access to the network through a manual search. This data catalog is also called manually cataloging resources. There are a large number of disaster databases on the web, but they are diverse and cannot be linked. We collected disaster-related information from the internet by manually retrieving and sorting disaster-related websites open to the public. The data catalogs were extracted, edited, and sorted using the same disaster metadata standards. The metadata were saved, and the data were curated and published through the OGC CSW standard. In addition to collecting core metadata information, additional attribute information was also processed, such as data documents and thumbnails.

The third type is the resources harvested through open standard protocols that follow both the GDMD and the target harvested database. This approach integrates the main directory database into the data center, which publishes the directory service using an open protocol. By harvesting metadata information, relevant open disaster data catalogs can be retrieved from the GDMD portal. Currently, open access catalog resources related to disaster data from the World Data Center for Renewable Resources and Environment (WDC-RRE) and the World Food and Agriculture Organization (FAO) have been harvested.

To date, 1,463 data points have been released in the GDMD. IKCEST-DRR contains 329 items for self-construction data, 420 items for manually cataloged data, and 714 items for machine-harvested resources. Federated retrieval data can be presented directly in the static directory of the platform. After compilation, the main disaster data catalog system was classified into eight

categories: 273 hydrometeorological disaster datasets, 162 geological and earthquake disaster datasets, 11 marine disaster datasets, 95 biological disaster datasets, 41 ecological and environmental disaster datasets, five technical disaster datasets, 12 social disaster datasets, and 150 other disaster datasets.

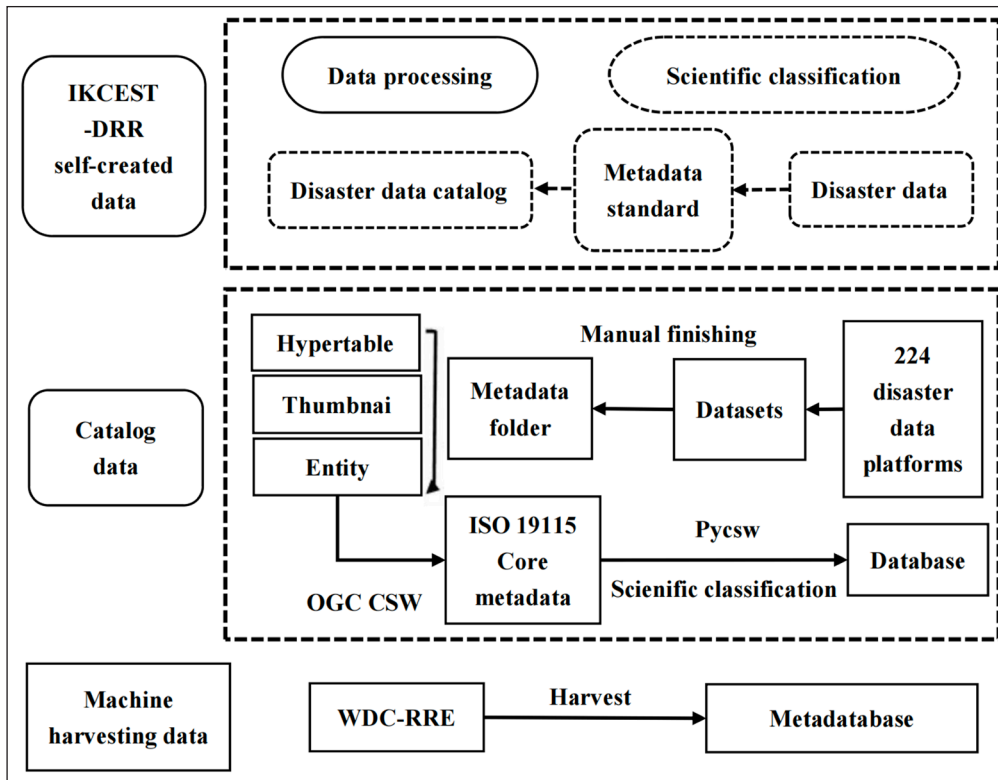


Figure 2 Data compilation diagram of GDMD.

3.2 IMPLEMENTATION OF GDMD

(1) GDMD prototype

The GDMD prototype was developed based on the IKCEST-DRR platform as a new module in the ‘Directory’ column. This module was developed, followed by a metadata standard designed by the GDMD, three types of data collection approaches, and open-source technologies. This process involves data collection, collation, standardization, and storage. Through the GDMD module, we can realize the function of online searching and accessing disaster metadata and support background platform management for these resources, such as metadata editing, updating, and deleting. The module interface of the GDMD is shown in Figure 3.

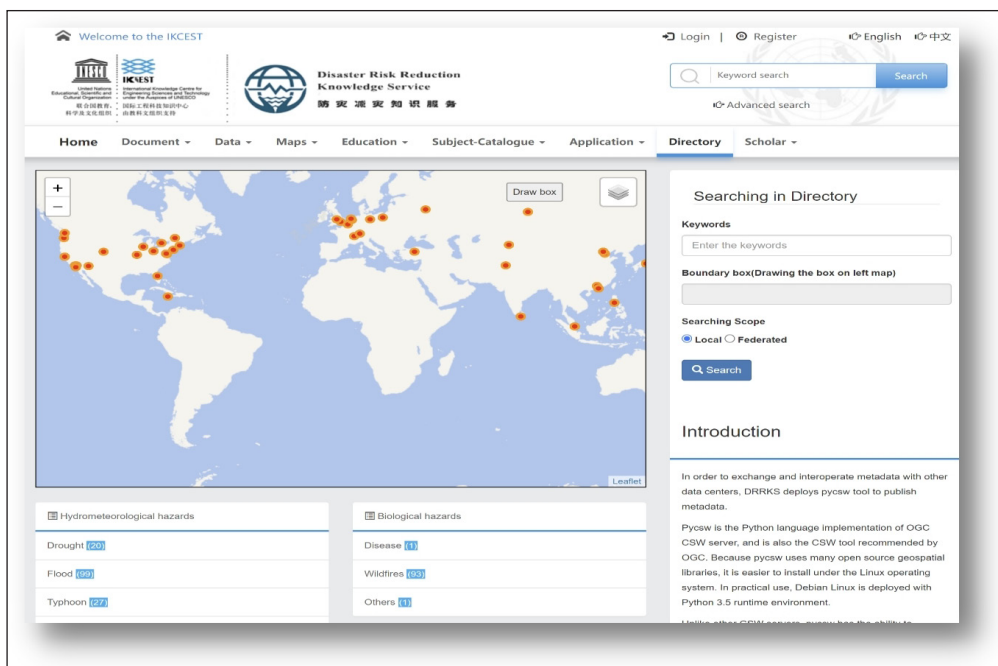


Figure 3 Main interface of the GDMD prototype.

The GDMD can be accessed using an open metadata interface for external services. Protocols that can be requested include OGC CSW, SOAP, OGC OpenSearch, and OAI-PMH. Pycsw is responsible for harvesting disaster metadata in the background and storing it in the database to ensure its centralized management and efficient usage. The metadata retrieval client of the GDMD prototype, as the main interface for interaction with the user, is responsible for receiving the query conditions and displaying the corresponding retrieval results. It provides various retrieval methods, including keyword retrieval and spatial queries (Figure 4).

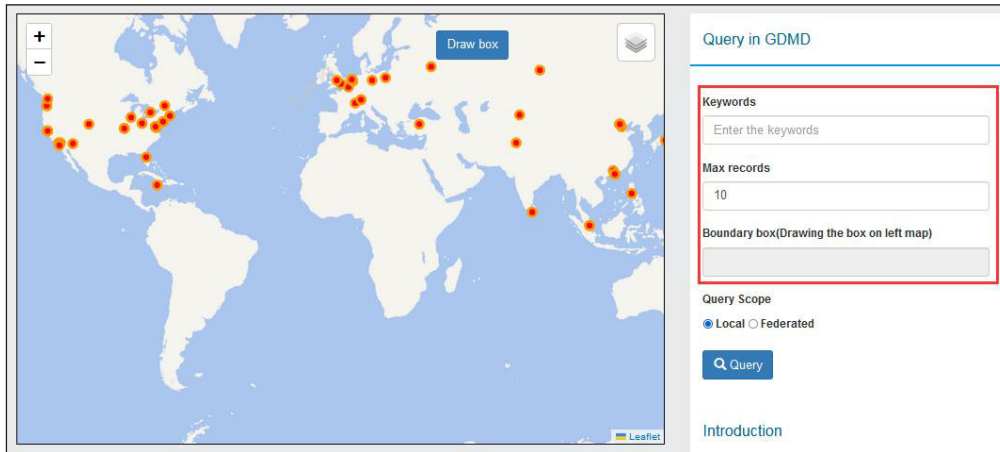


Figure 4 Joint data retrieval by geographic location or attributes.

In the GDMD prototype, the implementation of retrieval by the map uses the plug-in ‘Leaflet. Draw’, which has the complete spatial scope selection function. For the operation, the button ‘Drawbox’ is clicked to trigger the drawing rectangle event. The latitude and longitude ranges are then obtained, and the data query function is realized through the selected range.

(2) Directory Resource Compilation System Interface

The IKCEST-DRR self-built datasets were collated in a standard metadata format. This portion of the data is directly mapped to the main directory prototype system. The 1981–2012 set is taken as an example (Figure 5) to present the key metadata elements in this dataset. Manually cataloged data are compiled, classified, and stored based on a certain metadata standard format.

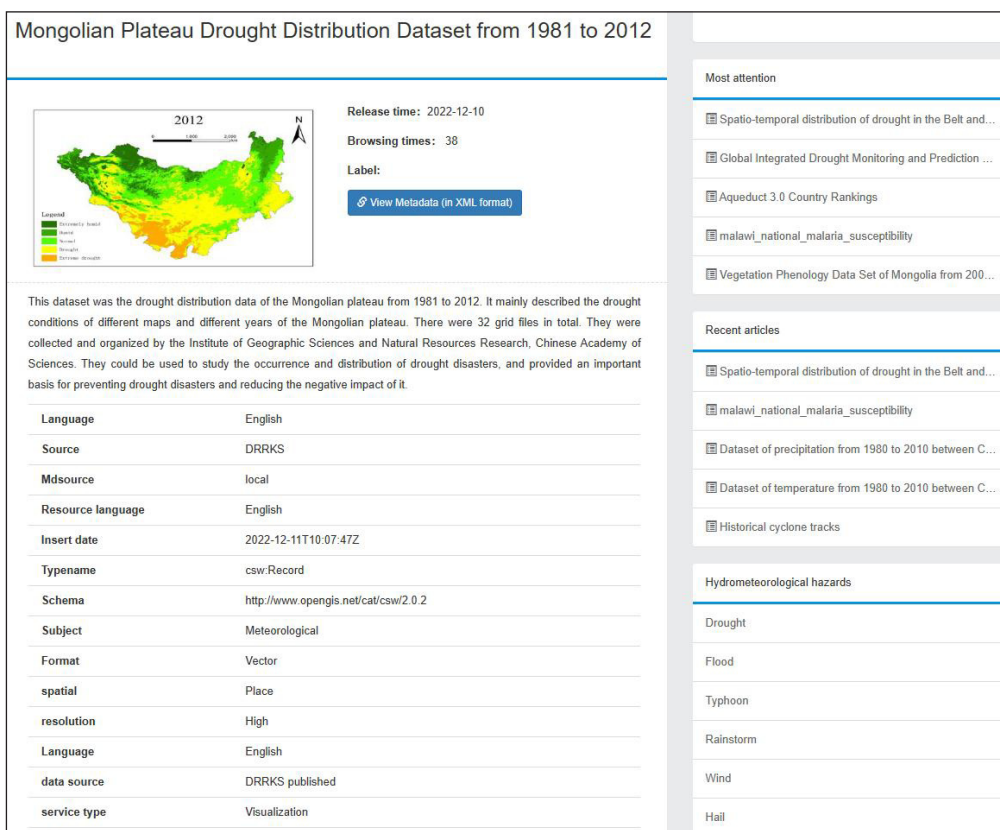


Figure 5 IKCEST-DRR self-built data-map of the drought distribution data set in the Mongolian Plateau from 1981–2012.

(3) Data search

Two types of data search approaches are provided by the GDMD. One was to query the local database, and the other was an online joint federal data search. The local search is primarily for data resources already centralized in the GDMD module in IKCEST-DRR. Federated retrieval can simultaneously search and integrate multiple data sources to obtain the target data catalog. These existing data sources remain independent in federated retrieval and have their own data storage, retrieval engines, and query interfaces. The federated retrieval system is responsible for forwarding query requests to various data sources and integrating the results for the user. Taking the rime disaster in Hangzhou from 1997 to 2013 as an example, the joint search data and core metadata fields are listed in Table 3.

FIELD_NAME	METADATA DESCRIPTION
Title	Fog and Rime disaster in Hangzhou 1997–2013
Identifier	drks-9cd36
Abstract	Fog and Rime disaster in Hangzhou 1997–2013 covers the major fog and haze disasters since 1997, mainly including the time, date, and visibility of fog and haze disasters in Hangzhou.
Xml	<csw:SummaryRecord xmlns:csw = "http://www.opengis.net/cat/csw/2.0.2" xmlns:dc = "http://purl.org/dc/elements/1.1/" xmlns:dct = "http://purl.org/dc/terms/" xmlns:gmd = "http://www.isotc211.org/2005/gmd" xmlns:gml = "http://www.opengis.net/gml" xmlns:gml32 = "http://www.opengis.net/gml/3.2" xmlns:ows = "http://www.opengis.net/ows" xmlns:xs = "http://www.w3.org/2001/XMLSchema" xmlns:xsi = "http://www.w3.org/2001/XMLSchema-instance"><dc:identifier>drks-9cd36</dc:identifier><dc:title>Fog and Rime disaster in Hangzhou 1997–2013</dc:title><dc:type/><dc:abstract>Fog and Rime disaster in Hangzhou 1997–2013 mainly covers the major fog and haze disasters since 1997, mainly including the time, date and visibility of fog and haze disasters in Hangzhou.</dc:abstract></csw:SummaryRecord>

Table 3 Metadata table of Hangzhou rime disaster from 1997–2013.

4 DISCUSSION

4.1 GDMD PROMOTES THE FAIR DATA GOVERNANCE

The FAIR (findable, accessible, interoperable, and reusable) guiding principles for Scientific Data management and stewardship were published in *Scientific Data* in 2016 (Wilkinson et al., 2019). The FAIR principle defines a set of rules that enables both machines and humans to find, access, interoperate, and reuse data and metadata. It involves four parts and 15 clauses and is an important theoretical achievement in scientific data governance. After the FAIR principle was proposed, many international organizations and disciplinary repositories followed it, pushing their own data governance.

The GDMD data hub significantly facilitated the implementation of FAIR principles on the IKCEST-DRR platform. More than 1400 global disaster data catalogs were collected beyond independent resources in the IKCEST-DRR. More than 120 global disaster platforms are involved in the GDMD system, which provides more than 700 compiled data entries. UNESCO recommended the Disaster Risk Reduction Knowledge Service System as a global demonstration platform for open knowledge in 2023, supported by the GDMD. It can now be accessed through the UNESCO open knowledge portal (<https://www.unesco.org/en/open-science/knowledge-sharing>). We determined that GDMD is a typical function in IKCEST-DRR, comparing it with other open knowledge platforms in the disaster risk reduction field recommended by UNESCO.

Currently, more than 4.6 million disaster data items, 23 thematic online knowledge service applications, and 370,800 disaster literature databases have been released in IKCEST-DRR. By the end of 2023, users had covered 235 countries and regions worldwide. Users accessed the platform more than 2,248,000 times, and data were downloaded more than 1,013,000 times. The top ten countries accessing the platform are the United States, India, the Philippines, Britain, Canada, Turkey, Germany, Australia, Pakistan, and France.

GDMD has an open interface for data catalog interoperability. This means that not only can human users access data using search engines, but machines (algorithms or application systems) can also access data using a unified protocol. Currently, GDMD is mainly used manually, but it has the potential for model and artificial intelligence (AI) applications in the near future. The diverse resources collected in GDMD have not only traditional databases or datasets but also digital resources ready for artificial intelligence usage, such as remote sensing images annotations (Arnaudo et al., 2024). With the continuous increase of high-quality AI-

ready resources in GDMD in the future, it will enable human or machine users to push large-scale data or model-driven applications for disaster risk reduction. In addition, it will help more partners share the cooperation and exchange of disaster data resources and jointly respond to global disaster challenges.

GDMD will try to find a sustainable development model for its long-term updating. The pycsw supports automatic data harvesting and metadata interoperability, which is the main method to ensure metadata growth. As the IKCEST-DRR is a stable international platform in UNESCO, it has sustainable funding and human resource support. Based on this premise, manually cataloging data is still an effective maintenance method that can ensure data catalog quality in the current stage.

4.2 ANALYSIS OF THE APPLICATION SCENARIOS OF THE GDMD

The GDMD is a worldwide directory, hub, and clearinghouse for disaster data. It is also a tool for presenting, searching, publishing, mining, and analyzing disaster data. Presentation entails the catalogs displaying in the same classification system with rich metadata. Searching is the ability to retrieve catalogs in local and federal databases using keywords or spatial map locations. Mining can be performed for text tag production and knowledge graph mapping, supported by Natural Language Processing technologies for large-scale disaster knowledge discovery.

In terms of disaster data discovery, the GDMD offers multiple paths for open resources to converge and exchange. The standards used were the OGC standard and XML format, which are openly used for spatial and text information exchange. The GDMD module is developed and deployed on an open online platform programmed by pycsw, a popular open-source tool. Users can access the integrated global disaster catalog in the GDMD and use these resources in different scenarios. For example, users can find available open data in various ways, including local or federated databases by text or spatial searches, access the data using classification systems or rich metadata, learn the data gap among these existing data catalogs, build a catalog exchange mechanism with other partners who host those datasets, or obtain the global landscape of disaster data distribution and characterization using all the information above.

With the increase of global climate change and extreme weather events, the uncertainty of major disasters such as extreme heat and massive floods has increased as well. Incorporating more data catalogs that provide future climate predictions into GDMD can enrich its time series data catalog, including not only historical data but also current and future data. Long time sequences, including future predictions, will be an important time dimension reference for GDMD data networks. These time tags or stamps can be registered in the GDMD data management system.

The GDMD also supports emergency responses for disaster relief and post-disaster rebuilding. A global heat wave and extreme precipitation in 2022 led to massive flooding in Pakistan, killing approximately 1,700 people. IKCEST-DRR conducted flood relief support work in Pakistan and organized the production and release of disaster prevention and reduction data in combination with GDMD technology. A total of 70 GB of disaster prevention and mitigation data were compiled and processed into three major categories and 21 subcategories. These datasets included the water system, traffic network, land cover, forest distribution, flooded farmland, flooded roads, flooded buildings, soil distribution, and dam breakage risk map datasets. The online emergency sharing service data directory system provides the data distribution. By the end of 2023, the directory system served more than 3,000 personal users.

Another case of emergency response was that of the earthquake in Turkey and Syria. The earthquake in Turkey and Syria in 2023 killed approximately 60,000 people. It injured more than 120,000 people and left approximately 1.5 million people homeless, damaging an area of approximately 350,000 km² and causing economic losses of approximately 120 billion dollars. The IKCEST-DRR Platform was rapidly activated for disaster data collection and compilation in response to and in support of earthquake responses in Turkey and Syria. Forty-five datasets were completed and shown in an accessible catalog online.

Based on the prototype system construction of the GDMD, it is necessary to further improve the disaster data catalog resource aggregation, data content mining, and disaster risk reduction knowledge discovery abilities. For example, more in-depth applications can be developed to promote the quick construction of knowledge graphs, forming a more comprehensive and efficient data catalog organization and search system and improving the application ability and effect of GDMD.

5 CONCLUSION

Open Science is a common understanding that facilitates data sharing in all science and technology fields, including disaster risk reduction. However, improving the ability to share disaster data online is a significant challenge because they are scattered and ubiquitous worldwide. This study proposes the concept of the Global Disaster Data Master Directory, which serves as a global hub for disaster data catalog connection networks. Based on the unified metadata and disaster data classification standards, pycsw was used to publish the data online and build a prototype system of the GDMD. Combined with the three methods for cataloging resource collection in the IKCEST-DRR platform, namely, self-built resources, manual curation catalogs, and harvesting with a unified protocol, the interconnection of the 1,400+ global disaster data directory has been realized. The open model of GDMD helps to realize FAIR data governance in the field of disaster risk reduction and further pushes the Open Science mechanism and innovative applications of resources in more scientific communities.

CODE AVAILABILITY

Code can be available at GitHub: <https://github.com/bukun/GDMD>.

ACKNOWLEDGEMENTS

We thank all members of the IKCEST-DRR team who participated in the development and validation of the GDMD.

FUNDING INFORMATION

This research was funded by the National Key R&D Program of China (2022YFF0711600) and the Construction Project of the China Knowledge Centre for Engineering Sciences and Technology (grant number CKCEST-2023-1-5).

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

Conceptualization: Juanle Wang. Methodology: Juanle Wang, Kun Bu, Formal analysis: Xiaodong Min. Software: Kun Bu, Qiuyuan Wang. Validation: Yuelel Yuan, Feiran Sun. Writing–review and editing: Juanle Wang, Xiaodong Min, Feiran Sun. Supervision: Juanle Wang. All authors contributed to the final paper.

AUTHOR AFFILIATIONS

Juanle Wang  orcid.org/0000-0002-5641-0813

State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; China-Pakistan Earth Science Research Center, Islamabad 45320, Pakistan; Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

Kun Bu

Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China

Xiaodong Min

State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; School of Civil Engineering and Geomatics, Shandong University of Technology, Zibo 255049, China

Qiuyuan Wang

Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China

REFERENCES

- Arnaudo, E., Vaschetti, J.L., Innocenti, L., Barco, L., Lisi, D., Fissore, V. and Rossi, C.** (2024) *FMARS: Annotating Remote Sensing Images for Disaster Management using Foundation Models*. ArXiv 2405.20109. Available at: <https://doi.org/10.1109/IGARSS53475.2024.10641130>
- Chen, Z. and Wang, X.** (2007) 'Preliminary study of ocean core metadata standards', *Modern Computer*, 06, pp. 120–122.
- Liu, Y.** (2016) *Design and Implementation of Operation and Maintenance IT Equipment Monitoring System of Jinan Bank of Tianjin*, Doctoral dissertation, University of Electronic Science and Technology of China.
- Open Geospatial Consortium Inc.** (2024) *OpenGIS catalog Service Implementation Specification*. Version 3.0.
- Sibolla, B., Van Zyl, T., Mcferren, G. and Hohls, D.** (2014) 'Adding temporal data enhancements to the advanced spatial data infrastructure platform', *Tenth International Conference of the African Association of Remote Sensing of the Environment (AARSE 2014)*. University of Johannesburg, South Africa, 27–31 October 2014.
- Sun, L., Zhang, C. and Fang, C.** (2012) 'Design and implementation of the resource registry based on service metadata', *Science of Surveying and Mapping*, 37, pp. 155–157.
- UNESCO** (2011) *UNESCO recommendation on open science*. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>.
- Wang, J., Bu, K., Yang, F., Yuan, Y., Wang, Y., Han, X. and Wei, H.** (2020a) 'Disaster risk reduction knowledge service: A paradigm shift from disaster data towards knowledge services', *Pure and Applied Geophysics*, 177(2), pp. 135–148. Available at: <https://doi.org/10.1007/s00024-019-02229-w>
- Wang, J., Han, X., Bu, K., Zhang, M., Wang, X. and Yuan, Y.** (2020b) 'Knowledge service system on disaster risk reduction and its application in social media analysis', *Journal of Global Change Data & Discovery*, 4(1), pp. 25–32.
- Wang, J., Lin, H., Ran, Y., Zhou, Y., Song, J. and Du, J.** (2014) 'A Study of Earth System Science Data Classification for Data Sharing', *Advances in Earth Science*, 29(2), pp. 265–274.
- Wang, Y., Bu, K. and Wang, J.** (2018) 'Design and prototype implementation of disaster metadata management system based on open source pycsw', *E-science Technology & Application*, 9, pp. 60–70.
- Wang, Y., Wang, J. and Bu, K.** (2019) 'Research on disaster data management technology and platform progress and the demand it faces', *Journal of Catastrophology*, 34, pp. 205–210.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B.** (2019) 'Addendum: The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data*, 6, p. 6. Available at: <https://doi.org/10.1038/s41597-019-0009-6>
- Zhang, A. and Xing, L.** (2005) 'The application and standardization of metadata', *Standard Science*, 10, pp. 52–53.
- Zhao, H. and Wang, J.** (2015) 'Domestic and foreign scientific data metadata standard and content analysis', *Information Research*, 2, pp. 21–24 + 30.
- Zheng, S.** (2022) 'UNESCO has adopted the Open Scientific Proposal', *Science & Technology in China*, 296, pp. 102–104.

TO CITE THIS ARTICLE:

Wang, J., Bu, K., Min, X., Wang, Q., Yuan, Y. and Sun, F. 2025 GDMD: A Disaster Data Hub Fabric the Open Data Network for Disaster Risk Reduction. *Data Science Journal*, 24: 8, pp. 1–12. DOI: <https://doi.org/10.5334/dsj-2025-008>

Submitted: 31 October 2024

Accepted: 24 January 2025

Published: 05 February 2025

COPYRIGHT:

© 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.