An empirical study on the evaluation of the open sharing capability of meteorological data based on the data of 5 provincial meteorological departments

Danna Shen1∗, Jieying He2, Yan Li3 and Yicen Lu4
1Development and Planning Center, China Meteorological Administration, Beijing, China.
2Shanxi Institute of Meteorological Science, Shanxi, China.
3School of Statistics and Mathematics, Zhongnan University of Economics and Law, Hubei, China.
4Zhejiang Meteorological Information Internet Center, Zhejiang, China.
*Corresponding author: Shendanna@cma.gov.cn

Abstract. Data sharing benefit evaluation is an important guarantee for the continued stable construction and application of data sharing projects and an important basis for making data application policies. Although meteorological data is an essential part of the basic strategic resources of national big data, through domestic and foreign research, there is no comprehensive system research on the comprehensive capabilities of data sharing services, especially on the meteorological industry. Therefore, writers of this paper establish an evaluation model for open sharing of meteorological data, select the open sharing capability of meteorological data from five representative provincial meteorological departments in China for pilot evaluation and empirical analysis, and then propose the applicability of the meteorological data sharing policy, unified meteorological data sharing service platform, overall planning and allocation of meteorological data resources and other policy recommendations.

1. Introduction
Data sharing benefit evaluation is an important guarantee for the continued stable construction and application of data sharing projects and an important basis for making data application policies. As an essential part of the basic strategic resources of national big data, meteorological data play a vital role in national economic and social construction, national disaster prevention and mitigation, ecological civilization construction, and the construction of “The Belt and Road”. The China Meteorological Administration, with a long history of open sharing meteorological data, is the first State Council department to open professional data to the whole society. At present, the openly shared data mainly include basic meteorological data products of such as ground, high altitude, meteorological satellites, weather radar, numerical weather forecast, etc., which gradually realize the integration and application of meteorological departments with all kinds of meteorological data resources, promote the integration and application of big data in government affairs, support major national strategies such as disaster
prevention and mitigation, ecological civilization and “The Belt and Road” and “Smart City” construction, finally bringing huge economic and social values.

From the research status at home and abroad, some scholars have studied data economy. GOLDKINDL, MCNUTTIG [1] made research on sharing economy, employment, and economic justice and fairness from the perspective of data. Jonathan P. Tennant et al [2] studied the academic, economic and social impact of the OA systems. What’s more, the World Bank studied the economic growth of open data through case studies, and the European data department made a detailed analysis of the economic benefits of open data [3]. While in China, Zhao Wei et al. [4] proposed an evaluation index system for the sharing of scientific and technological resources under the sharing control framework based on the balanced scorecard, which provided a theoretical basis for the practical operation of the construction of publics & t resources and the evaluation of shared services. Based on the balanced scorecard principle and combined with the characteristics of the website, Zhu Li et al [5] established the performance evaluation model of the website for atmospheric environment monitoring data sharing. However, through domestic and foreign research, there have never been any comprehensive systematic studies on the comprehensive capabilities of data sharing services, especially empirical research on the meteorological industry. Hence, to clarify the main existed problems in the process of meteorological data governance so as to comprehensively measure and truly reflect the ability and level of meteorological data sharing in a certain period, writers of this paper establish an evaluation model of meteorological data open sharing capability, and select the meteorological data open sharing ability of five representative provincial meteorological departments in China as a pilot for evaluation, with a view to scientifically evaluating meteorological data open sharing ability, releasing meteorological scientific data dividends, optimizing the internal and external services of the meteorological industry, and lastly providing a reference for public services.

2. Construction of a valuation model for open sharing of meteorological data

The meteorological data involved in this valuation mainly refer to various meteorological observation (exploration) records collected and archived by meteorological authorities at all levels, as well as all kinds of meteorological data sets, climate statistics, numerical analysis data and products, etc. processed from those above records. As for data type, there are mainly 14 related categories (First-level classification) of meteorological data (materials), namely surface, high-altitude, marine meteorological data, meteorological radiation data, agricultural and ecological meteorological data, numerical forecast products, atmospheric composition data, historical climate substitute data meteorological disaster data, radar meteorological data, satellite meteorological data, scientific experiment and investigation data, meteorological service products and other data. For data sharing, it is mainly to provide different types of meteorological data and products sharing services at home and abroad according to the requirements of different users. Besides, sharing capability, the general ability to openly share meteorological data, includes the total benefits to industries, economy, society, scientific research, etc. brought by various platforms on account of data sharing policies and resources

2.1. Indicator selection

We should make the transition from “quantity” to “quality” supply into account when designing indicators for the sake of reflecting the idea of high-quality development of meteorological data. In consequence, the total amount of meteorological data resources, types and the possession and the delivery and distribution capabilities are considered to embody “quantity” in the design of the evaluation of the open sharing capability, and meanwhile, the “quality” is evaluated through the coverage of data sharing, the convenience of data use, and the ability to apply data. In addition, the capability to openly share meteorological data is not only reflected in the “external” benefit, but also in “internal”. Therefore, we strive to achieve both “external” and “internal” evaluations for designing indicators.

The evaluation mainly includes four parts: policy norms, data resources, data sharing and service benefits. First, policy norms mainly evaluate whether the relevant policies for data sharing are issued;
whether the policies are in line with the national policies; whether the policies such as standardization and quality control are issued internally; and whether the data are subject to strict quality and process control. Second, data resources mainly include the total amount of data resources and the total categories of data products, which are assessed from the integrity and richness of the data resources. Third, data sharing rate and coverage, data application ability and usage efficiency, sharing method and convenience are evaluated to measure the sharing capability and availability convenience of meteorological data in the industry and the global. Fourth, service benefits cover the benefits for scientific and technological research and development, the industry and the social benefits, and the benefits of safeguarding major national strategies. Table 1 shows specific indicators.

| First-Level Indicator | Second-Level Indicator                              | Third-Level Indicator |
|-----------------------|---------------------------------------------------|-----------------------|
| A1 Policy Norms       | B1 Policy Completeness                           | C1 Sharing Policy Implementation |
|                       | B2 Policy Applicability                          | C2 Sharing Policy Applicability |
| A2 Data Resources     | B3 Data Resources Completeness                   | C3 Total Data Products (Tb/Pb) |
|                       | B4 Data Resources Richness                       | C4 Types of Data Products (Category) |
| A3 Data Sharing       | B5 Total Amount of Data Sharing                  | C5 Total Amount of Data Open Sharing (Tb) |
|                       | B6 Sharing Coverage among Data Industries        | C6 Total Amount of Cloud Data Sharing (TB) |
|                       | B7 Global Data Sharing Coverage                  | C7 Data Exchange Volume among Industries (TB) |
|                       |                                                   | C8 Data Sharing Volume of Data Sites Among Industries (Number) |
|                       |                                                   | C9 Total Data Submission Volume (TB) |
|                       |                                                   | C10 Number of Data Exchange Industries (Number) |
|                       |                                                   | C11 Total Data Distribution (TB) |
|                       |                                                   | C12 Number of Data Distribution Industries (Number) |
|                       |                                                   | C13 Global Data Exchange Volume (TB) |
|                       |                                                   | C14 Number of Remote Sensing Monitoring Reports (Period) |
|                       |                                                   | C15 Number of Countries Applying Fengyun Satellite Data (Number) |
|                       |                                                   | C16 Number of Meteorological Data Service Countries Along the Belt and Road (Number) |
|                       |                                                   | C17 Data Platform Visits (Ten Thousand Times) |
|                       |                                                   | C18 Number of Data Platform Users (Number) |
|                       |                                                   | C19 Number of Data Platform Downloads (TB) |
|                       |                                                   | C20 Data Platform Downloads (Ten Thousand Times) |
| A4 Service Benefits   | B8 Data Resources Availability                   | C21 National S & T Support Projects (Number) |
|                       | B9 Technology R&D Support                        | C22 Provincial and Ministerial S & T Support Projects (Number) |
|                       |                                                   | C23 Number of S & T Projects Supported by Meteorological Departments and Provincial Bureaus |
|                       |                                                   | C24 Number of Published Scientific Research |

Table 1. Indicator System of Evaluation on the Open Sharing Capability of Meteorological Data
2.2. Weight determination

The weighted method is adopted to summarize each indicator step by step so as to calculate the open sharing capability of meteorological data and the total score of each sub-indicator. This paper uses a scale of 1-9 to convert the qualitative results into quantitative data to finally obtain the hierarchical order of first-level and second-level indicators through the design of the questionnaire and the method of expert opinion grading, which is shown in Table 2.

In hierarchical sorting, the consistency test is performed on the results to judge the rationality of the construction matrix. The specific steps are: first to calculate the consistency index CI: $CI = \frac{\lambda_{max} - n}{n-1}$; second to calculate the random consistency ratio CR: $R = \frac{CI}{RI}$. When CR < 0.1, the consistency test passes, indicating that the feature vector of the judgment matrix can be regarded as the weight vector. In accordance with the above steps, it is found that both the first- and second-level indicators have passed the consistency test. Besides, it can be seen from the weight of the first-level indicators that service benefits (0.28) have exerted the greatest influence on the open sharing capability of meteorological data, followed by data sharing (0.26) and data resources (0.24). Among the second-level indicators, policy applicability (0.53), data resources completeness (0.55), data resources availability (0.29), and major strategic support (0.28), etc. All remain the indicators with higher weights.

Table 2. Benefit Evaluation Index and Weight Score of Meteorological Sharing Service

| First-Level Indicator | Policy Norms | 0.22 |
|-----------------------|-------------|------|
| Open and Sharing Capabilities of Meteorological Data | Data Resources | 0.24 |
|                       | Data Sharing | 0.26 |
|                       | Service Benefits | 0.28 |

| Second-Level Indicator | Policy Norms | Policy Completeness | 0.47 |
|------------------------|--------------|--------------------|------|
|                        | Policy Applicability | 0.53 |

| Data Resources | Data Resources Completeness | 0.55 |
|----------------|-----------------------------|------|
|                | Data Resources Richness    | 0.45 |
2.3. Construction of evaluation model

After the evaluation indicators and weights are determined, the writer adopted the weighting method to summarize each indicator to calculate the open sharing capability of meteorological data and the total score of each sub-indicator.

(1) To standardize the original data: Considering that the indicators involved in this research are all positive, the writers use the following formula to standardize the data: 

$$Z_{ij} = \frac{X_{ij} - \text{min} X_{ij}}{\text{max} X_{ij} - \text{min} X_{ij}},$$

where $X_{ij}$ means the original data; $\text{max} X_{ij}$ and $\text{min} X_{ij}$ respectively refer to the maximum and minimum values of the original data.

(2) To calculate the scores of policy norms: 

$$S = \sum_{i=1}^{n} W_i Z_{ij},$$

among which, $W_i$ represents the weight of the factors affecting the policy norms; $Z_{ij}$ is the standardized data of the policy norms, and $S$ stands for the score of the local policy norms.

(3) To calculate the scores of data resources, data sharing and service benefits by the same way, and finally get the scores of open sharing capability of meteorological data in different places according to the weight of policy norms, data resources, data sharing and service benefits in the sharing service ability of meteorological data.

3. Empirical analysis of the open sharing capability of meteorological data

3.1. Data acquisition

Due to the wide range of industries involved in the open sharing of meteorological data, its potential social and economic benefits are so difficult to quantify statistically. Prior to this, there has never been any specific statistical study on the benefits of open sharing of meteorological data in China. Hence, this paper collects statistical meteorological data of 2019 on the ground of determined indicators via the field survey and the organization of the pilot units so as to fill in the statistics. In addition, expert scoring methods are adopted to turn them into quantitative statistics for the qualitatively described indicators such as the major strategic benefits of the meteorological support country.

3.2. Analysis of evaluation results

This paper calculates the scores of all-level indicators and total scores of the open sharing capabilities of meteorological data of the five pilot provinces by standardizing the data obtained from the open sharing capability of meteorological data, and in line with the evaluation model mentioned in the previous section. And the total scores of the five pilot provinces are sorted from large to small, as shown in Table 3 that presents the average scores of the four first-level indicators in the five pilot provinces.
Table 3. Open Sharing Capability Scores of Meteorological Data

| Pilot Unit | Policy Norms | Data Resource | Data Sharing | Service Benefit | Total Score |
|------------|--------------|---------------|--------------|----------------|-------------|
| AMA(AMA)   | 28.1         | 41.2          | 24.7         | 23.1           | 28.9        |
| BMA(BMA)   | 28.1         | 20.6          | 27.5         | 26.4           | 25.7        |
| XMA(XMA)   | 17.6         | 14.0          | 20.9         | 14.7           | 16.8        |
| YMA(YMA)   | 11.8         | 15.8          | 15.2         | 14.4           | 14.4        |
| ZMA(ZMA)   | 12.4         | 8.4           | 11.6         | 21.4           | 13.8        |
| A VERAGE   | 19.6         | 20.0          | 20.0         | 20.0           | 19.9        |

From the above table, we can know that the average score of the open sharing ability of meteorological data is 19.9 and the scores of different units vary greatly with the highest score 28.9 and the lowest score 13.8. In general, open sharing capabilities of meteorological data vary greatly among provinces.

From the perspective of policy norms, the five provincial meteorological departments scored an average of 19.6, two of which are higher than the average score. Both of them have developed big data implementation plans and meteorological data usage specifications. The sharing policy and standards are basically applicable to the local meteorological departments internally and externally applicable to certain industries and fields. What's more, from the view of data resources, although the provincial meteorological departments all have about 10 categories of meteorological products, the total amount of data resources owned by them varies greatly. AMA holds the most resources, close to Level PB; BMA and XMA with medium resources, at Level 100TB; YMA and ZMA with least resources, at Level 10TB. Moreover, from the data sharing standpoint, BMA has the largest amount of total open resources of 97.5TB, accounting for 57.5% of the total provincial unit open resources, while XMA holds the least amount of only 0.3TB, taking up 0.2% of that (See the above table). From the viewpoint of meteorological data industrial coverage, the total distribution of meteorological data to various industries far exceeds the total amount of industry submission, 14 data distribution industries and 8 data archiving industries. The number of inter-industry sites shared by AMA provincial meteorological department is the highest, with 5,351 sites, and AMA and XMA also reach thousands of sites. Besides, seeing from the data exchange industry number and the total exchange amount, AMA holds the largest exchange amount of 10.6TB, involving 6 different industries. The data resource availability scores are XMA, BMA, AMA, ZMA and YMA in descending order. More importantly, AMA has continuously optimized its provincial data sharing platform to improve data sharing capabilities. In 2019, the CIMISS platform was accessed 12 billion times and downloaded 4 million times, making it become the most complete and authoritative big data cloud platform of the AMA provincial meteorological department.

Eventually, the highest score is 23.1 for AMA and the lowest is 14.4 for ZMA with an average score of 20, among which, BMA and AMA performed significantly in serving enterprise scope and quantity, more than 200 companies in 17 industries; Besides, the benefits of ZMA's public livelihood are more powerful, and YMA, as well as AMA, are equipped with outstanding national strategic capabilities. Unfortunately, ZMA is slightly deficient in supporting scientific research papers and serving enterprises.

4. Conclusions and policy suggestions

Writers find that the sharing capabilities vary greatly among regions, especially in terms of policy norms, data resources ownership and data sharing, through the evaluation of the open sharing abilities of meteorological data of the five provincial meteorological departments. It is mainly presented in the following aspects:

4.1. Strengthening the research on the applicability of the meteorological data sharing policy

The implementation of meteorological data sharing policies in various places is relatively high, but
adaptability is not strong. For example, units and the public, the existing policy documents of local meteorological departments at all levels can no longer meet the needs of data governance in face of the ever-increasing diversified and personalized weather data service requirements of the government. And there is indeed an urgent requirement to reorganize the development of weather data sharing services policy. In addition, in terms of data exchange, the meteorological detection activities carried out by local governments, relevant departments, social organization and individuals (except meteorological departments) have been increasing, and some indoor and outdoor micro-meteorology detection equipment, such as smart phone, smart wearable device and air fruit, etc., can also make meteorological detection at any time and in anywhere. These detection activities generate a large amount of meteorological data. However, there is no unified data exchange specification currently for effective management of these meteorological data. That is, the inapplicability of existing policy has hindered the intensive integration and the sharing of data resources to a certain extent, which affects the discovery and release of its potential value. Therefore, we should further clarify the applicability of the meteorological data sharing policy, and issue more standardized meteorological data sharing policies.

4.2. United planning for the construction of the meteorological data sharing platform
It is difficult to form a unified data export and the statistical work becomes more difficult due to the existing of so many platforms for sharing meteorological data. According to the statistics reported by the five provincial departments, all their meteorological data platforms alone have involved 20 kinds of platforms. Whereas, these platforms with rich variety pose the following questions for us: first, whether these platforms have played their due role. Although the provincial platforms are with rich contents, some of them are of low usage value; second, despite their variety, unified data export has not been formed. There is an urgent need to establish a unified public meteorological data service platform across the whole country, which can not only provide industry users with convenient, efficient and standardized official interface services of meteorological data, but also facilitate unified management and service.

4.3. Balancing the allocation of data resources among regions
The open sharing capabilities of meteorological data vary significantly from region to region. Regional differences are obvious in data resource ownership, sharing coverage and resource availability rather than in service benefits. The highest score of data resource ownership is three times of that of the lowest score. And the scores of data sharing indicators are also different greatly too, especially the data platform visits and downloads, with 20 times differences. Hence, it is essential for meteorological departments, on the ground of providing better services, to adjust the allocation of meteorological data resources in a reasonable way, like increasing the amount of resource ownership and building data sharing capacity for regions with low resource allocation.

References
[1] Goldkindl, McNuttjg. (2019) Vampires in the technological mist: the sharing economy, employment and the quest for economic justice and fairness in a digital future [J]. Ethics and Social Welfare, 13(1): 51-63
[2] Jonathan P. Tennant, Francois Waldner, Damien C. Jacques, Paola Masuzzo, Lauren B. Collister, Chris. H. J. Hartgerink. (2016) The academic, economic and societal impacts of open access: an evidence-based review [J]. F1000 Research, 5: 632
[3] Andrew Scott. (2015) Open data for economic growth [R]. The World Bank, 2: 1-20
[4] Zhao Wei, Peng Jie, Wang Yunhong. (2009) Research on science and technology resource sharing evaluation based on balanced scorecard [J]. Science and Technology Management Research, 29(07): 104-106+91
[5] Zhu Li, Yi Min, Huang Yanmin, Yao Shenjun, Wu Jianping. (2018) Performance evaluation research on atmospheric environmental monitoring data sharing website based on the balanced scorecard [J]. The Administration and Technique of Environmental Monitoring, 30(04): 8-11