

PROCEEDINGS PAPER

A Study of the Application of Big Data in a Rural Comprehensive Information Service

Leifeng Guo¹, Wensheng Wang¹ and Nengfu Xie¹

¹ Key Laboratory of Digital Agricultural Early-Warning Technology, Agricultural Information Institute of Chinese Academy of Agricultural Sciences, Beijing 100081, China
guoleifeng@caas.cn, wangwensheng@caas.cn

Big data has attracted extensive interest due to its potential tremendous social and scientific value. Researchers are also trying to extract potential value from agriculture big data. This paper presents a study of information services based on big data from the perspective of a rural comprehensive information service. First, we introduce the background of the rural comprehensive information service, and then we present in detail the National Rural Comprehensive Information Service Platform (NRCISP), which is supported by the national science and technology support program. Next, we discuss big data in the NRCISP according to data characteristics, data sources, and data processing. Finally, we discuss a service model and services based on big data in the NRCISP.

Keywords: NRCISP; Rural area; Information service; Service model; Big data

1 Introduction

During the Eleventh-Five-Year period, China made remarkable achievements in agricultural and rural information. A rural information infrastructure developed rapidly, rural information service stations covered almost all administrative villages, models of information service focusing on agriculture and rural areas continued to be innovated, the information levels of agricultural production and operation grew increasingly, and e-government speeded up its extension to the township level. Local governments thoroughly pushed forward the construction of rural information according to their practical situations and the obvious effect it has had on various aspects, such as transferring surplus rural labor, enriching grass-roots cultural activity, improving new rural cooperative medical services, and promoting rural social security. However, the problems of information isolation and information separation have not been completely resolved. How to guarantee the information's authenticity, accuracy, integrity, and timeliness, how to protect farmers from being misled and harmed by false information, and how to ensure farmers can access information at the lowest cost are always problems for agriculture and rural information.

Agricultural information services, focused on providing value-added information to farmers, have been studied for many years. With the rapid development of information technology, the models and institutions of agricultural information services have changed, and the scope of agricultural information services has also been extended. Agro-climatic information services, vocational training, agro-technology information services, and agricultural advice are all included in these information services. In 2008, a rural comprehensive information service was introduced by researchers (Wang & Zhang, 2008; Zhang & Zhou, 2008). By 2010, Jiang, Luo, and Qi (2010) defined a rural comprehensive information service as a service provider that provides timely, accurate, and safe information products or services for service demanders in agricultural production, rural life, and farmer development with the help of information resources and proper information technology. This service aims to integrate various data resources from different top-level departments and provide services for farmers at the bottom-level. Data are the key to a rural comprehensive information service, especially when the data volume is huge.

Currently, data are regarded as new strategic resources, and people from all walks of life are keeping an eye on big data. Big data (Li & Cheng, 2012) is an all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process using hands on data management tools or traditional

data processing applications. High volume, high velocity, and high variety are considered essential features of big data (Laney, 2012). Additionally, a new V “Veracity” has been added by some organizations to describe it. In 2012, the Obama administration announced the Big Data Research and Development Initiative, which explored how big data could be used to address important problems faced by the government (Kalil, 2012). Research on big data is increasing. Some experts have even said that big data is the next “subversive” technological change in the IT industry after cloud computing and IOT, and it will produce a huge impact on social management, future predictions, and business decisions.

There are many aspects to rural comprehensive information services. Data sources vary, data volume is huge, and the data processing is complex. In this paper, we hope to improve rural comprehensive information services with big data and focus on rural information service data organization, data storage, data processing, and data visualization.

2 Background

In 2006, the Ministry of Agriculture (MOA) took “12316” as the unified, special phone number for public services in the agricultural system. Three years later, more than 20 provinces (city, district) opened “12316” hotlines serving “agriculture, village, and farmers”. In 2009, the MOA decided to use “12316” as the connection for making an overall plan for modern media, such as voice telephones, text messages, and the internet, to build a comprehensive information service platform for “agriculture, village, and farmers”.

In 2009, the Ministry of Industry and Information Technology (MIIT) issued basic regulations about the construction and service of rural comprehensive information service stations. The regulations specify the content of rural information services, including public services, information consulting services, farmer training services, culture and entertainment services, and proxy services.

In 2009, the Ministry of Science and Technology (MOST) started the rural information demonstration province program. Shandong and Hunan were chosen as the pilot provinces in 2010 and 2011. By 2012, another five provinces were included in the pilot program. By 2013, the number of pilot provinces had reached twelve. This program integrated more than 20 related departments of information resources in the pilot provinces, constructed a comprehensive information service platform in each province, and built a smooth resource sharing network and information transmission channel straight into the villages.

To integrate information resources and systems in the pilot provinces, “Construction and Application of National Rural Comprehensive Information Service Platform”, a national science and technology support program, was approved. The Agriculture Information Institute of the Chinese Academy of Agricultural Science (CAAS), together with another brother company, undertook this project. This program will build the national rural and agricultural information service platform using cloud technology, and complete the nationwide sharing of information resources. It will also provide agriculture product markets and agriculture technology information services directly to the farmers and solve last minute problems in rural and agriculture information.

3 NRCISP

NRCISP used cloud computing as the infrastructure with which to integrate data, application, and service resources from provincial level information service platforms. Services are further combined and optimized so as to serve farmers in rural areas in the “one portal site and N platforms” model. In NRCISP, rural data play a very important role. As data size is big and data type is diverse, big data technology is required for data processing and analysis.

3.1 Province-level rural comprehensive information service platform

Currently, there are twelve Chinese provinces who have promised to carry out the pilot program. They will provide data and resources to and share or interchange applications and services with the NRCISP. Construction of province-level rural comprehensive information service platforms follows the principle of “Platform Upward, Service Downward” and serves the local farmers. Most of the platforms include grass-roots information service stations, data transfer channels, special information service systems, and portal sites for rural comprehensive information services (**Figure 1**).

3.2 Infrastructure based on cloud computing

Cloud computing protects the heterogeneity of the hardware platform and operating system and provides uniformed “Infrastructure as a Service” for NRCISP. With the help of virtualization technology, distributed models, and parallel architecture, hardware resources are virtualized, and the computing, storage, and net-

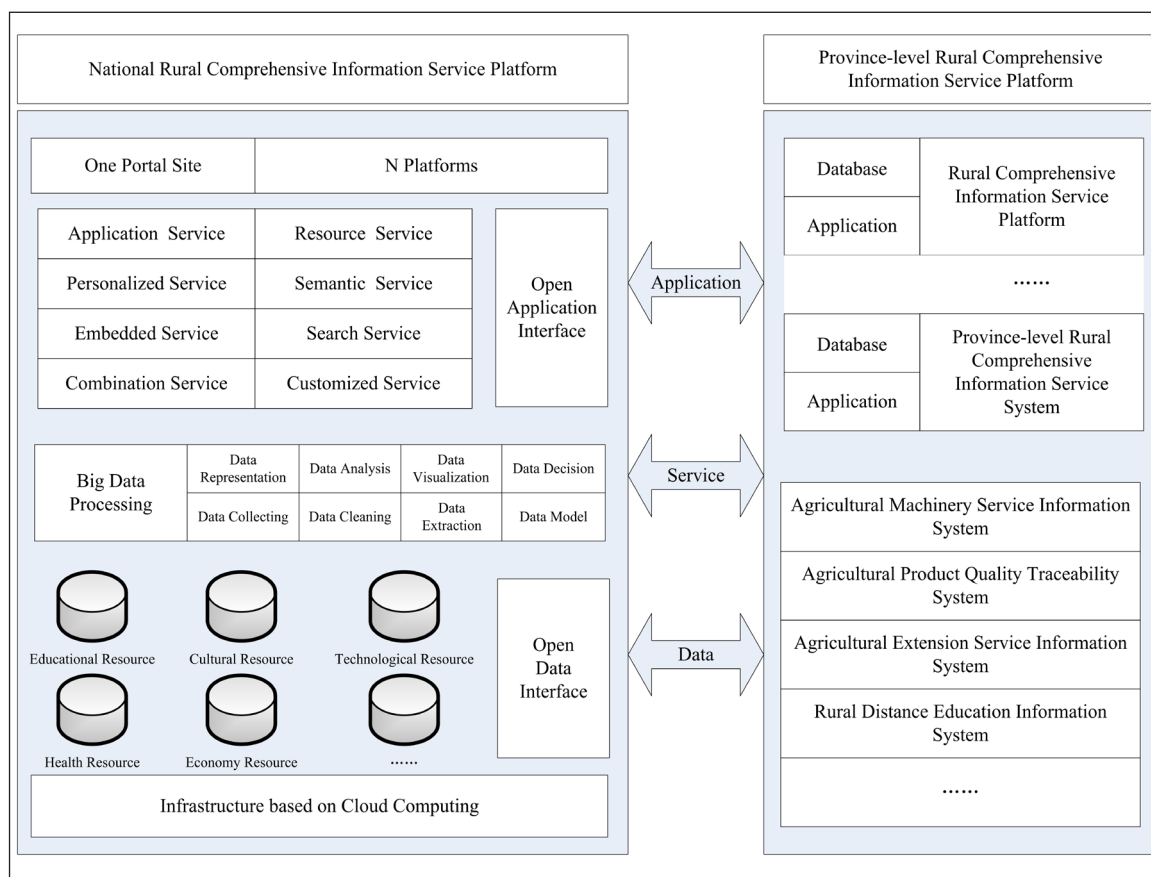


Figure 1: Architecture of NRCISP.

work resources are integrated into a resource pool. This consolidation causes unified scheduling, monitoring, allocation, and management of multiple IT resources (**Figure 2**).

3.3 One portal site and N platforms

One portal site and N platforms is the main carrier of information services for the majority of users related to agriculture and is an important component of NRCISP (**Figure 3**).

One portal site means a comprehensive information service platform with universal access functions. The core of the portal site relies on website group integration, such as a website group for rural public services, a website group for agriculture specialty industries, and a website group for villages and towns.

N platforms focuses on presenting information service platforms or channels. Smart call centers provide an important channel for information communication and advice. Currently there are many hot-lines oriented to serving “agriculture, rural area, farmers”, such as the number 12316 from MOA, Xinghuo Science and Technology’s 12396, agriculture technology’s 110, and nongxintong, meteorological information’s 12121. Smart call centers exploit the organic integration of video calls, network calls, and remote video and visual calls to share resources and allocate calls throughout the whole communications network. Rural distance education service platforms have greatly affected the training of party members in rural areas. Many information receiving stations have been built in rural areas, and the service system is relatively perfect. It is comparatively simple to carry out distance education in rural areas. The Agriculture Information Institute of CAAS has had an agricultural extension information service for many years. It provides information services with text, voice, and videos to grass-roots agriculture extension practitioners. There are three major categories of applications. The first category focuses on information exchange, knowledge sharing, self-learning, and self-improvement for grass-roots extension practitioners; the second category pays close attention to management and performance appraisals of agricultural technicians based on WEBGIS technologies; the third category follows information collection, such as the coverage of crops, crop pests and animal diseases, the market supply and demand of agriculture products and materials, etc.

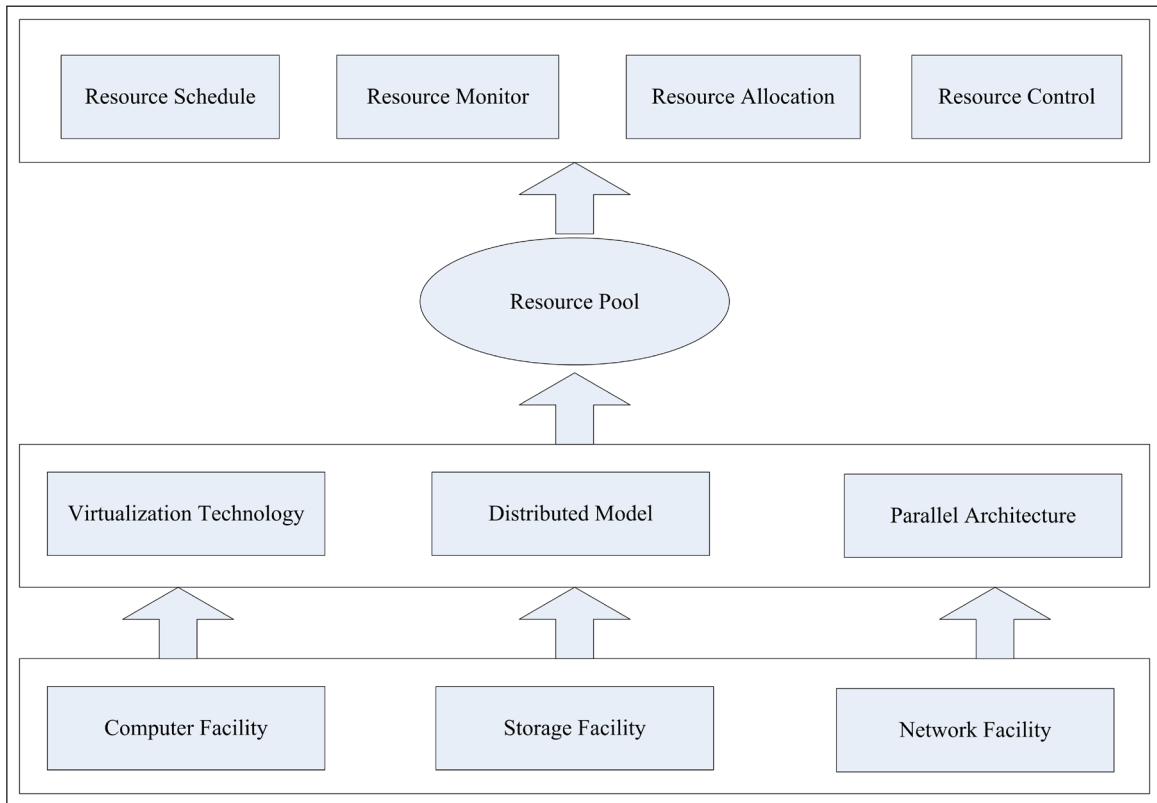


Figure 2: Cloud computing infrastructure.

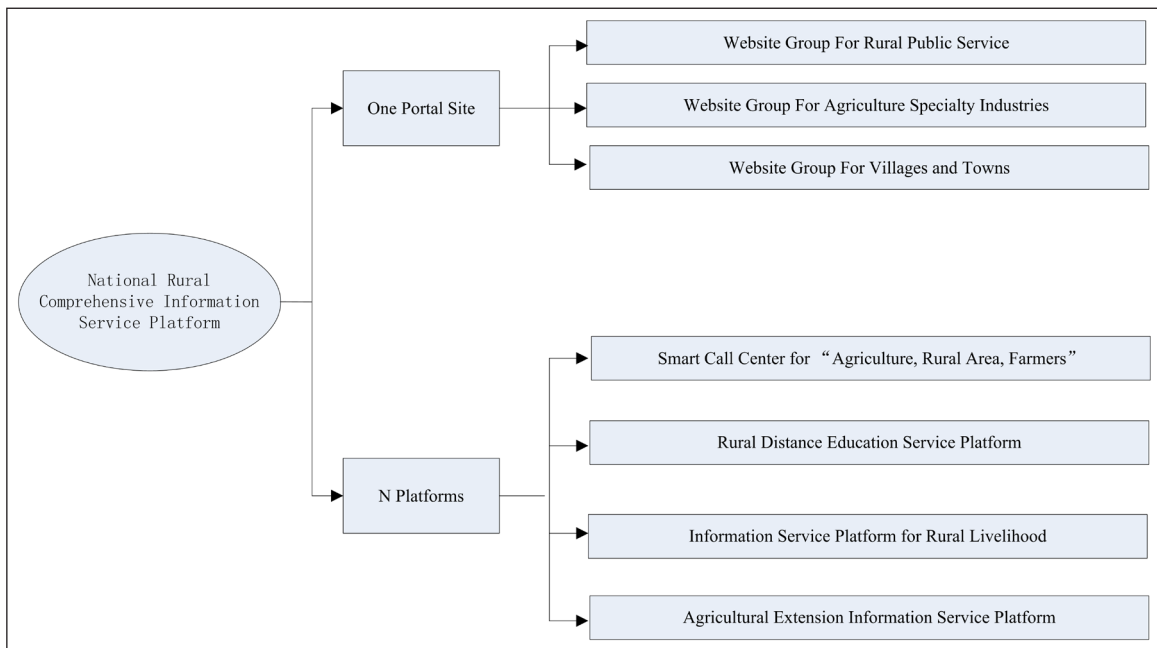


Figure 3: One portal site and N platforms.

4 Big Data in NRCISP

Agriculture data in NRCISP share the attributes of huge data volume, complicated data structure, and various data types. They possess the essential characteristics of big data. Comprehensive rural information services have strong needs for information personalization for their wide service content and objects. The application of big data in comprehensive rural information services will not only bring revolutionary advances for information service technology but also will improve the overall progress of agriculture information.

4.1 Data characteristics

As data sets in NRCISP get bigger and bigger, data are not as simple as in the past. First, data volume is increasing exponentially. Ten years ago, a gigabyte was enough for agriculture data. When cloud technology and smart devices began to become popular in agriculture, the data were stored by terabytes. Today, there is petabyte or even exabyte data in NRCISP. Second, data are becoming increasingly complex. They have various data formats, types, and structures. With the further development of agricultural information, the amount of unstructured and semi-structured data such as text, images, audio, video, sequences is growing rapidly. Finally, the data are exhibiting non-linear characteristics such as clustering and bursting. The uncertainty due to data inconsistencies, ambiguities, and latencies is becoming more acute. Streaming data in information services intersect with each other, and data are being generated fast and need to be processed fast.

4.2 Multi-sourced data

The huge amounts of diverse, heterogeneous, and inconsistent data bring great challenges to NRCISP; traditional data storage and processing based on relational databases can't meet the needs of data analysis. NRCISP obtains data from multiple sources, such as web crawlers, data integration, user interactions, and special databases. Data collecting is very important for an information service platform. A web crawler is deployed to gather specific information, such as news, prices of agricultural products, and agricultural product quality data, from web pages. NRCISP integrates the data from the province-level comprehensive rural information service platform with the existing information system. Specific knowledge concerning field crops, livestock, disease prevention, aquaculture, fruit growing, vegetable growing, and fish processing in the form of video or courseware is upload to NRCISP. NRCISP users also generate a lot of communication data, accessing data, log data, and so on. These data are very important for NRCISP's analysis of users' behavior so as to provide better services. NRCISP also provides specific data services from professional databases, for example, pest databases, meteorological databases, and agricultural machinery databases.

4.3 Data processing

Data processing is the key to extracting value from the big data in NRCISP. There are two steps to transforming data into services (Figure 4). The first step is data preprocessing, such as data de-duplication, data de-noising, data cleaning, and data representation. Unlike traditional data processing, NRCISP introduces a big data processing framework, such as MPP, Hadoop, and stream, based on the type of data being processed. Thus, the data processing capacity is strengthened. The second step is data analysis based on big data techniques or tools. Depending on the type of data, either data fusion, data clustering, data statistics, data visualization, or data mining is chosen to extract value from prior stage results. Analysis results, including charts, curves, and reports, will be provided to the farmers who need the services.

5 Service Model

5.1 Service Model

The basic model of an information service refers to descriptions of the elements and relationships among information services (Chen, 2003), and its main purpose is to improve the quality and level of service. Most information services focus only on sending information or data to users. It differs from big data in that service strategies, service channels, and service objects vary. Big data can extract the potential value by deep data mining, which is used to add and discover value during the information service process. Finally, big data offers data resources to users in the form of services. In NRCISP, multi-source data sets become professional data sets after big data processing and the value of the data resource is added. Then, through big data

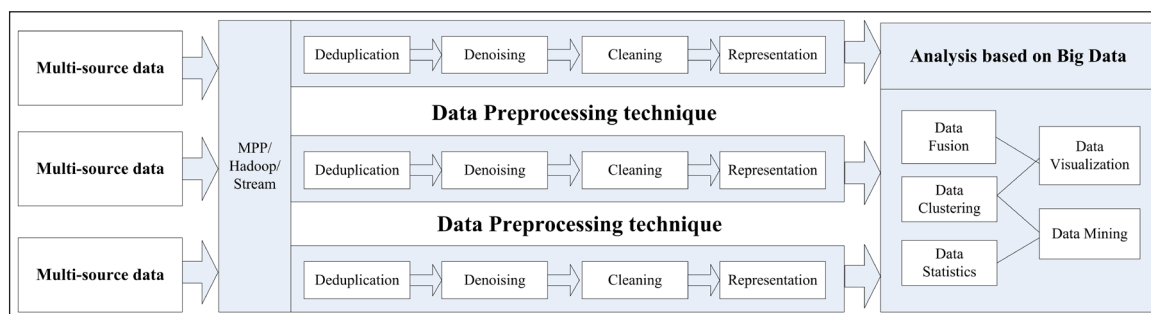


Figure 4: Big data intelligent processing and analysis.

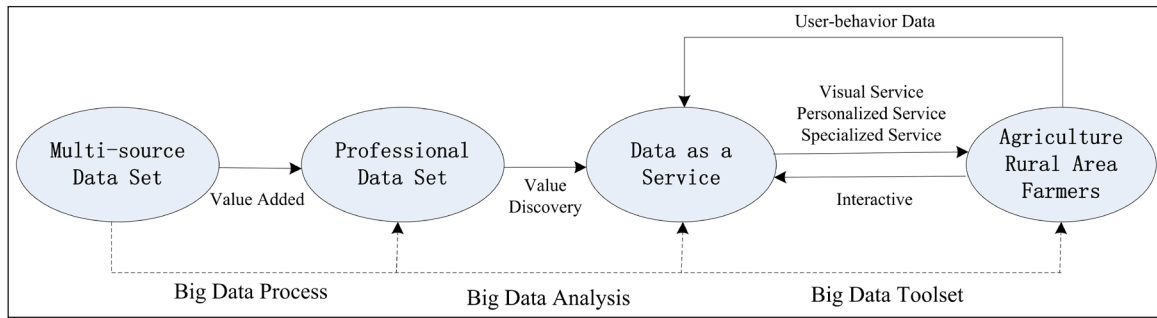


Figure 5: Service model based on big data.

analysis, the value is discovered, and data resources are changed into services. By providing various big data tools, NRCISP can provide visual, personalized, and specialized service to agriculture, rural areas, and farmers (Figure 5).

5.2 Services

The kernel of agricultural information is information service, which satisfies the needs of specialization, individualization, low cost, and sustainability (Wang, 2013). Combined with big data technology, NRCISP provides a better information service. According to the service content, information services are divided into four types.

- 1) **Public services based on big data visualization**
Public services in rural areas include sharing of cultural information resources, openness in village affairs, family planning, employment, bringing science and technology to the countryside, etc. Most public services are promoted by a specific government department to improve the service level in a certain respect. Using big data visualization tools, NRCISP can display the basic data of public services in the form of charts, tables, or videos, which will make it easy for farmers in rural areas to understand and accept these public services.
- 2) **Information consultation services based on big data prediction**
Information consultation services focus on providing policies, regulations, education, medical services, and market information to farmers and solving the practical problems that arise during agricultural production and management. Big data can predict the future development of agriculture, including market changes and technological development, with huge samples of basic agriculture data. Through receiving accurate agricultural predictions, farmers can deal effectively with big markets.
- 3) **Distant training services based on big data organization**
Distant training services provide remote training on modern breeding techniques, production and management, cultural knowledge, and information techniques for farmers to improve their cultural quality and information skills. A large amount of data needs to be efficiently organized in case of unlimited expansion for quick retrieval and precise positioning. Big data tools are very suitable for unstructured data storage such as videos, course wares, text, and pictures.
- 4) **Culture and entertainment services based on big data personalization**
Culture and entertainment services provide interactive cultural and entertainment programs using videos and network technology. Local tourism, natural resources, industrial advantages, and historical culture are also reported. The key to culture and entertainment services lies in providing different services to different people. Personalized services based on big data can deliver related programs to farmers based on their consumer behaviors.

6 Conclusion

Rural comprehensive information service plays a very important role in the economic and social development of rural areas. With the fast development of new information technology, we should improve our ideas and methods to improve service levels and provide more specialized, individualized, and low-cost services.

Now, big data has shown its potential in data analysis and has been employed in many fields. The process of a rural comprehensive information service involves big data collecting, storage, processing, and analysis, and it is very necessary to apply big data to the rural comprehensive information service.

7 Acknowledgments

This paper is supported by the National Science and Technology Support Program “Research and application of cloud storage and cloud computing technology for rural information service”(2013BAD15B02), the “public sector (Agriculture) scientific research funding program” integration and demonstration of agricultural extension service based on ICTs”(201303107) and public research institutes for basic research funds program” research on the key technology of agricultural information service based on email”(2014-J-002).


8 References

- Chen, J. L. (2003) A Study of Information Service Models. *Journal of Peking University (Philosophy and Social Sciences)* 40(3), pp 124–132.
- Jiang, Y., Luo, L. P., & Qi, C. J. (2010) Model and Institutions of Modern Integrated Information Service in Rural Areas: An Analysis of Hubei Province. *Journal of Intelligence* 29(3), pp 173–178, 138.
- Kalil, Tom. (2012) “Big Data is a Big Deal”. White House. Retrieved 26 September 2012 from the World Wide Web: <http://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal>
- Li, G. J., & Cheng, X. Q. (2012) Research Status and Scientific Thinking of Big Data. *Bulletin of Chinese Academy of Sciences* 27(7), pp 647–657.
- Wang, R. J. (2013) Bottleneck of Agricultural Informatization Development in China and the Thinking of Coping Strategies. *Bulletin of Chinese Academy of Sciences* (5), pp 337–343.
- Wang, W. Z., & Zhang, C. L. (2008) Design of Villages Integrated Information Service Platform. *Journal of Agricultural Mechanization Research* (7), pp 137–139.
- Zhang, J. L., & Zhou, W. (2008) The design and implementation of the integrated platform of rural information service application. *Journal of Zhejiang University of Technology* 36(4), pp 370–373, 381.

How to cite this article: Guo, L, Wang, W and Xie, N 2015 A Study of the Application of Big Data in a Rural Comprehensive Information Service. *Data Science Journal*, 14: 12, pp.1-8, DOI: <http://dx.doi.org/10.5334/dsj-2015-012>

Published: 22 May 2015

Copyright: © 2015 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License (CC-BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/3.0/>.

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

OPEN ACCESS 