BRAINBANK METADATA SPECIFICATION FOR THE HUMAN BRAIN PROJECT AND NEUROINFORMATICS

Hu Lianglin[1]*, Hou Yufang[1], Li Jianhui[1], Yin Ling[2], Shi Wenwen[3]

*[1]Computer Network Information Center, Chinese Academy of Sciences, Beijing, China, 100080
  *Emails:*hll@sdb.cnic.cn, bnuxiaofang@sdb.cnic.cn, lijh@sdb.cnic.cn

[2] Neurology Department, General Hospital of PLA, Beijing, China, 100853
  *Email:*yinling301@126.com

[3]State Information Center, Beijing, China, 100045
Email: twiwnwaw@hotmail.com

ABSTRACT

Many databases and platforms for human brain data have been established in China over the years, and metadata plays an important role in understanding and using them. The BrainBank Metadata Specification for the Human Brain Project and Neuroinformatics provides a structure for describing the context and content information of BrainBank databases and services. It includes six parts: identification, method, data schema, distribution of the database, metadata extension, and metadata reference The application of the BrainBank Metadata Specification will promote conservation and management of BrainBank databases and platforms. it will also greatly facilitate the retrieval, evaluation, acquisition, and application of the data.

Keywords: BrainBank, Neuroinformatics, Metadata specification, Human brain project

1   BACKGROUND

Many databases and platforms for human brain and neuroinformatics had been established in China over the years, especially after the start of Chinese Human Brain Project, initiated in February 2000. As the number, complexity, and diversity of databases and platforms grows, a metadata specification for providing an understanding of them becomes more and more important. Metadata, which is defined as "data about data," plays an important role in describing the content, quality, condition, and other characteristics of the object described, and it is recognized as the infrastructure required for exchange and use of information. Metadata on brainbank databases is the method for helping the locating, accessing, understanding, and usage of those brain and neuroinformatics databases and platforms.

2   REQUIREMENTS ANALYSIS

Many general and domain-specific metadata standards have been developed or proposed by various user communities in the past few years. Even within the same subject domain or for the same type of resource, there are often multiple options of metadata standards. Until now, however, there has been no formal metadata specification for BrainBank datasets released.

Based on the needs of BrainBank datasets management and application, as well as the necessary and desirable characteristics for BrainBank datasets metadata schemes, five requirements for BrainBank metadata specification or application profile have been identified:

- Modularity
- Transparency

- Extensibility
- Interoperability

With the above requirements, a variety of general or dataset-oriented metadata schemes were reviewed for reference to develop metadata specification for BrainBank datasets. Examples of referenced databases are: Dublin Core (ISO 15836:2003), Ecological Metadata Language (EML 2004), Geographic information Metadata (ISO 19115:2003), Content Standard for Digital Geospatial Metadata (FGDC 1998), Scientific Databases Core Metadata (CNIC, 2005), Scientific Data Sharing Projects Metadata Specification (SDSP 2005) and its application profile for Chinese Medicine and Health. The last three metadata schemes were locally developed by the Chinese Academy of Sciences or the national Scientific Data Sharing Projects.
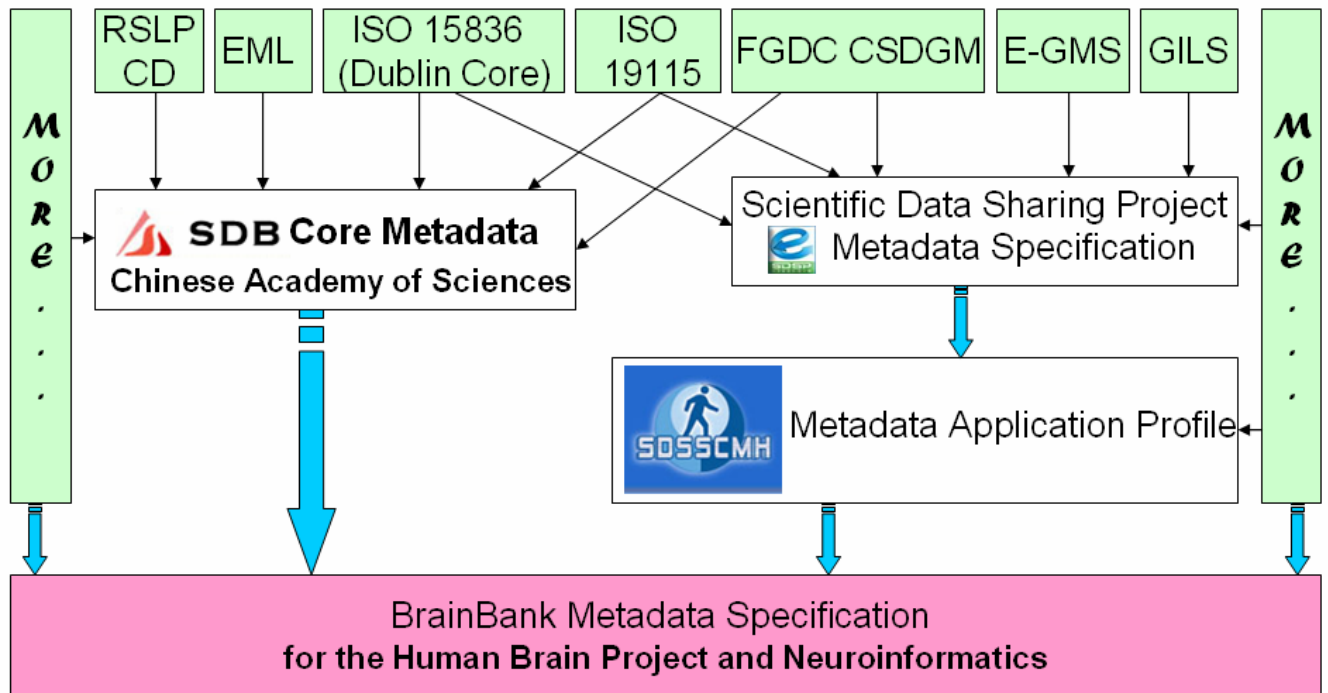

Figure 1. Roadmap of Establishment of BrainBank Metadata Specification

After examination of the above metadata standards and relevant projects, a new BrainBank Metadata Specification was proposed to be developed on the basis of SDB Core Metadata, the national Scientific Data Sharing Projects Metadata Specification, and its application profile for Chinese Medicine and Health. Figure 1 is the roadmap to establish the BrainBank Metadata Specification.
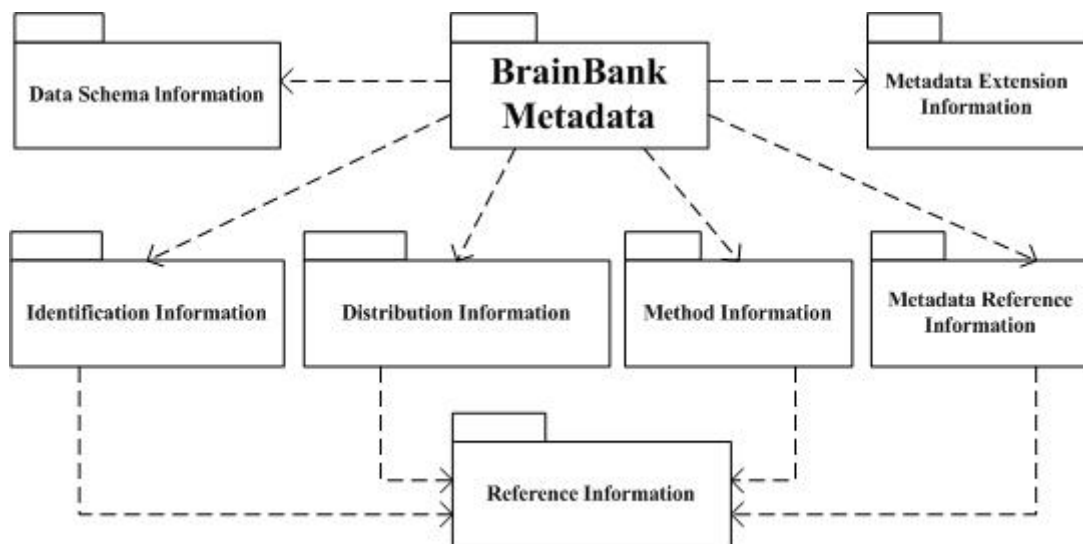
3   BRAINBANK METADATA SPECIFICATION

Based on the analysis and the roadmap to establish the metadata specification (Figure 1), the BrainBank Metadata Specification for the Human Brain Project and Neuroinformatics was developed. To provide a structure for describing the brainbank databases, the BrainBank Metadata Specification defines the schema required for describing BrainBank databases and services, and it includes six parts (as shown in Figure 2):

- Identification Information
- Data Schema Information
- Method Information
- Distribution Information
- Metadata Extension Information
- Metadata Reference Information

Identification Information contains information to uniquely identify a BrainBank dataset with 13 simple elements and 7 complex elements, including the title, an abstract, the purpose, language and character set, and the constraints to access the datasets. **Data Schema Information** defines the information on the database physical model, which is the essential foundation to access the databases by the users or to interoperate with others, and includes 7 simple elements describing the schema of the dataset, such as schema description language and restriction language. **Method Information** describes the methods used during data production, including process step information in detail and the method to assure data quality and provides the base information to assess datasets before usage. **Distribution Information** contains information about the distributor of and options for obtaining datasets. **Metadata Extension Information** is for specified extensions when the specification is insufficient, and **Metadata Reference Information** is the information on this metadata. **Reference Information** includes data types cited by the above six parts, and the data types specified here are responsible party, address, date, classification, process, and product. All those data types are cited more than once by the above six parts for corresponding purpose.

To support metadata interchange among different systems, the specification also is described in XML schema.

Figure 2. Architecture of BrainBank Metadata Specification

4   APPLICATION

The application of the BrainBank Metadata Specification promotes conservation, management of BrainBank databases and also greatly facilitates the retrieval, evaluation, acquisition, and application of brainbank databases. Figure 3 is a snapshot of the International Neuroinformatics Network – the China BrainBank website, which adopts this metadata specification to provide service on dataset retrieval and access and offers the services based on metadata. The redlined "Metadata Search" is one example of the included capability.
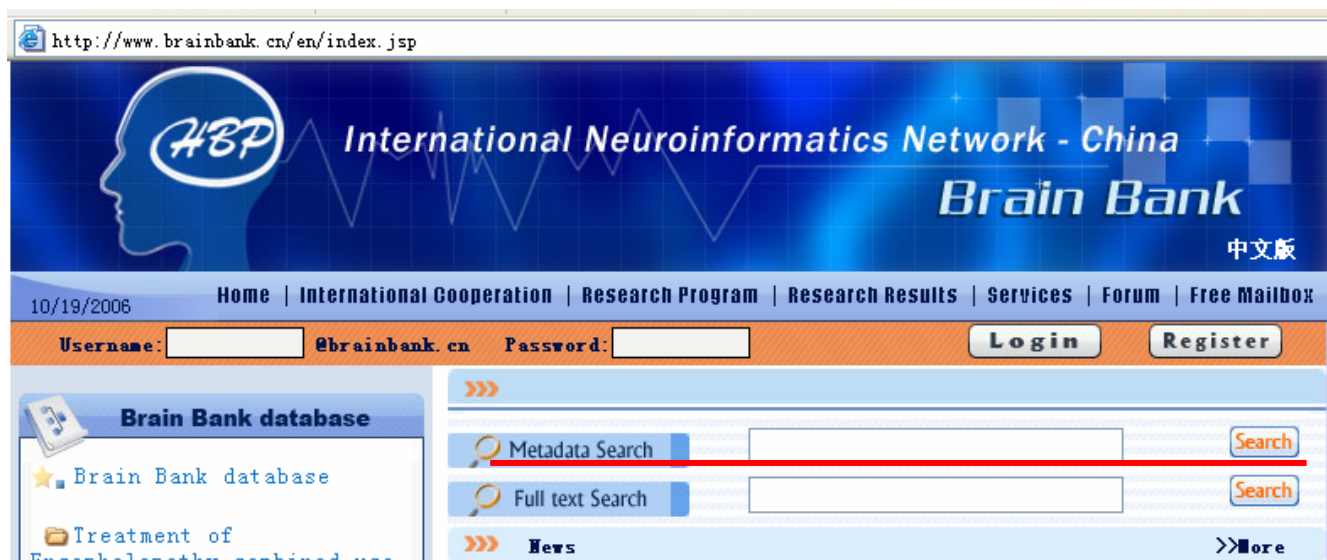
Figure 3. Snapshot of BrainBank Metadata Application

5   CONCLUSION

As a domain-specific specification, the BrainBank Metadata Specification is intended to be used widely. Feedback and comments are important to its improvement. The BrainBank Metadata Specification will be revised according to the feedback, and then it will hopefully be accepted and used by more groups. With its application, this specification will promote intercommunication among the platforms and interoperatibility among the databases about brain and neuroinformatics.

6   REFERENCES

Computer Network Information Center, Chinese Academy of Sciences. (2005) *Scientific Databases Core Metadata* (Version 2.0)

EML Project. (2004) *Ecological Metadata Language (EML) Specification* (Version 2.0.1). Retrieved November 27, 2006 from the World Wide Web: http://knb.ecoinformatics.org/software/eml/eml-2.0.1/index.html

Federal Geographic Data Committee. (1998) *Content Standard for Digital Geospatial Metadata*. Retrieved November 27, 2006 from the World Wide Web:
http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf

ISO 15836:2003, Information and documentation - *The Dublin Core metadata element set*

ISO 19115:2003, Geographic information – *Metadata*

Scientific Data Sharing Program, China. (2005) *Core Metadata Standard*