Social Networking, 2013, 2, 57-62
http://dx.doi.org/10.4236/sn.2013.22007 Published Online April 2013 (http://www.scirp.org/journal/sn)

Hakka Genealogical Metadata Principles and Practicalities

Chen-Chi Chang*, Ruey-Ming Chao, Ting-Yu Wu
Institute of Information and Society, National United University, Miaoli, Taiwan
Email: *kiwi@nuu.edu.tw

Received February 12, 2013; revised March 14, 2013; accepted April 9, 2013

ABSTRACT

Genealogy (Zupu in Chinese) is the history of families. Genealogy has been widely used for the tracing of their lineages and history. Hakka is a special Chinese ethnic group. To explore the origin and history of Hakka people and their culture, this study starts with the genealogy. After several Hakka genealogies collected, the content analysis is performed in this study. According to the qualitative data analyses, this study designs the pedigree metadata standards for Hakka to implement the union catalog. The Hakka Pedigree Metadata is comprised of many individual family and ancestors, each of which has a process in place for archiving and distribution of genealogy. Hakka genealogical metadata contains the following types of constructs. 1) Pedigree: use given name and lines expressed the genetically related. 2) Individual: describe the personal information. 3) Family History: the story of the ancestors is narrated. 4) Bibliography: description of the genealogical book. 5) Miscellany: included the unique documents that could not categorized in the above four classification. Results of this study can be applied to genealogies integration and implement the union catalog.

Keywords: Hakka Genealogy; Genealogy; Metadata; Pedigree; Pedigree Chart

1. Introduction

The language from the ancestors is the symbol and spirit of Hakka. Hakka language as the fundamental of Hakka culture, genealogical documents is important archives that recorded the ancestors' words, family instructions and surname is particularly important. Hakka is a special Chinese ethnic group. The Hakka people are one branch of the Han Chinese who lives predominantly in the provinces of Taiwan, Guangdong, Jiangxi, and Fujian in China. There is a proverbial saying in Hakka that, "It's better to sell one's ancestor's field than to forget one's ancestor's language; it's better to sell one's ancestor's house than to forget one's ancestor's voice". The concept of ancestor is really important for Hakka family. It makes the main barrier as the scholars doing the related research. The style of writing have different genealogy wording owing to the space and time difference, so there is no standard format. Several studies indicate that genealogy plays an important role in academic and social science [1,2]. Genealogical records related to the precious historical data of family. There are importance of historical value include economic, demographic, educational, ethnic and religious, etc. Genealogy (Zupu in Chinese) is the history of families. The genealogy is the historical trajectory of the family. The genealogy represents one of ethical rules for human survival and development. Recent years have seen increased attention being given to family tree writing in the Hakka genealogy literature. To explore the origin and history of Hakka people and their culture, this study will start from genealogy. Hakka genealogy doesn't have a writing format standard, so the preservation of Hakka Genealogy loss critical information. Scholars encounter certain degree of difficulty, who likes to collate these genealogies. A complete view of Hakka cultural characteristics is necessary. Therefore, establishment of the Hakka Pedigree Metadata is considerable. Well-designed Hakka Pedigree Metadata makes it is easier to merge genealogy and provide offspring easy to find their ancestors. Relatively little discussion in considerable amount of literature on Hakka Genealogy, the status of collected Hakka genealogies is a fundamental issue. Recent years have seen increased collected being given to Hakka genealogies in the literature. As the lack of union writing standard, family trees cannot integrity smoothly. Because the family tree recorded in each genealogies are separate. First, this study explores the genealogy metadata standards of countries in the World. To understand the differences between various genealogy,

*Corresponding author.

and the pedigree metadata standards for Hakka to implement the union catalog as the reference standard. Second, characteristics of the Hakka Genealogy will introduce in this study. Design the pedigree metadata standards for Hakka is claim of Hakka culture. Hakka genealogies of writing can be more standardized.

2. Literature Review

2.1. Genealogy

Family history can be a very interesting and as a research field has not yet been much explored. The Genealogy is one kind of family history record. In previous study, Liaw (2003) provides extensive discussions of genealogy, overview of genealogical documents reorganization, antiquarianism, as well as the genealogy of the contribution. Liaw (2003) also noted that there are important historical value of genealogy include economic, demographic, educational, ethnic and religious, etc [3]. Nowadays, several studies indicated that genealogy as the critical reference because of their potential to disentangle complex population histories [4-6]. However, the genealogy is a gray literature (non-conventional literature). It can't be obtained through purchasing channels. To collect these genealogies, it need people to provide or donated voluntarily. Field investigation is another way to collect the genealogical data. In the past, genealogy often consists of a family or a clan as a unit and writing in the paper. Various social and environmental factors, such as war, natural disasters and migration lead genealogies has been damaged or disappeared. In order to save and protect the genealogies, it is necessary to make these genealogies in the digital form. In the Western, the family tree is expanding pedigree in graphic form. The family tree chart presents a different history and cultural connotations that consisted of religious and secular. Family tree is one of documentation and digitization of genealogical to present. The genealogy employed the family tree as the structure for family members. The graphic family tree chart linked according to the genetically relationship of family member by information commutation technology. The family tree concept is similar to Chinese genealogy. Several sites, such as Ancestry.com [7], and Taiwan Genealogy Online [8] offer the services that genealogy writer and provide offspring to find their ancestors.

2.2. Genealogy Metadata

The regular definition of metadata is data about data. This explanation derived from the Metadata Workshop Report in March 1995. Renato Innella (1998) proposed a new concept of metadata is structural data about data. Metadata plays an important role not only in the digital library and museum environment, but also other areas [9]. For example, metadata is control of medicines management in the medical profession [10]. Metadata has different name in various domain fields. In library science, metadata is called Library Catalogue. In museum science, the metadata is called documentation and data standard [11]. Lorcan Dempsey and Rachel Heery introduced three types of metadata standards: simple formats, structured formats, and rich formats. Simple formats are that the data structure of a company exclusivity, emphasis on full-text indexing. Structured format is that standardized in formats. Characteristics based on the field as the main structure, such as, Dublin Core, IAFA the templates, RFC1807, SOIF, LDIF, etc. The formats design has a problem that lack of relational functions. Rich formats have been compliance with international standards. Its characteristics are detailed and complex flag to present the contents of the resource. For example, ICPSR, CIMI, EAD, TEI, MARC, etc. [12]. However, Dublin Core is the most suitable Document-Like Objects. Dublin Core originated during the 1995 invitational Online Computer Library Center (OCLC)/National Center for Supercomputing Applications (NCSA) Metadata seminar. Dublin Core design is based on a simple description applies to different metadata standards [13]. Therefore, Dublin Core can also apply to genealogy metadata. There are many genealogy metadata formats at present. **Table 1** shows different genealogy metadata standards.

2.3. Hakka Genealogy

Hakka is a special Chinese ethnic group. Instead of living in the same region, Hakka people were connected by their culture, language and heritage. The Hakka people are one branch of the Han Chinese who lives predominantly in the provinces of Taiwan, Guangdong, Jiangxi, and Fujian in China [18]. There will experience many difficulties in migration of Hakka people. Hakka people have been consciousness that not afraid of predicaments and courageous spirit. The reason for Hakka people special attention to the genealogy records is that they leave hometown frequently. Hakka genealogy contents include migration process of the family, genetically related, marital status, Family Instructions, and so on. The contents of Hakka genealogy are inseparable from Hakka culture. If you understand Hakka culture, you will comprehend Hakka genealogy what elements is necessary Hakka culture has the following characteristics [19]. Hakka culture has the following characteristics [20,21]. First, respect for ancestors, mourn for ancestors, the lessons from the ancestors is the symbol and spirit of Hakka. Consequently, it recorded Family Instructions in the genealogy. Hakka people pay careful attention to own parent's funerary rites. Second, believe that mysterious power is the common tradition in the Central Plains. Hakka people believe that Feng Shui has a mysterious power and will affect their fortune. Third, this is a glorious to become

Table 1. Genealogy metadata standards.

Metadata Standards	Introduction
GEDCOM	GEDCOM Stands for GEnealogical Data COMmunication. It is a Common file format that allows users to exchanges genealogical files between different Programs [14].
GedML	These pages describe GedML, a way of encoding genealogical data sets in XML. It combines the well-established GEDCOM data model with the XML standard for encoding complex information. The result is a representation that can easily be converted to and from GEDCOM, but can be manipulated much more easily using standard tools: notably, by using an XSLT processing such as Saxon [15].
GEDC	GEDC is an XML format for genealogical data. Although it's relatively new, GEDC has already stood the test of time. It has been in everyday use since early 2003, in parallel with its development, and its evolution has been guided by practical needs as they arose. GEDC is currently being used to maintain several genealogies as large as 25,000 persons. GEDC originated as an outgrowth of the GEDCOM 6.0 Beta spec from the Church of Jesus Christ of Latter Day Saints. It still bears a superficial resemblance to its predecessor, but many minor changes have been made as well as several major ones. Especially, GEDC has followed the lead of Gentech's Genealogical Data Model (GDM) and broken away from the mindset implicit in GEDCOM: that genealogical data is entirely consistent and correct [16].
GEDCOM X	GEDCOM X is partitioned into different data models that each supports a distinct aspect of the genealogical proof standard. Specific serialization formats (e.g. XML, JSON) are defined to describe the ways the data is written to a file or exchanged over the Internet. You may also be interested in learning more about how GEDCOM X is being extended to provide for specific use cases and requirements such as field-based image extraction and standard Web Service APIs [17].

officials in the early China. They think became officials can upgrade themselves status in society. If they don't have any ancestors of celebrities or aristocrats, it humiliates them. It is shame at they don't have any ancestors of celebrities or aristocrats. Forth, patriarchal clan culture is unique culture of the Hakka. The extended family consists of a lots nuclear family. It will write on paper that the name of the ancestral hall. These points will affect the contents of Hakka genealogy. Genealogy writer will be carefully recorded that ancestors of celebrities, family instructions, Feng Shui and The name of Ancestral Hall. The nature of Hakka people owe to ancestors. It will affect the emphasis on their genealogy.

3. Research Methodology

To establish the Hakka Pedigree Metadata and merge genealogy, this study starts with the content analysis of genealogy. Both meta-analysis and content analysis were performed to gain the genealogy metadata standards of Hakka. This study also identified which surname is Hakka surname, and chooses for the representative of Hakka genealogy. Data were collected primarily from the Taiwan Memory database in National Central Library. National Central Library has rich collection of historic documents in Taiwan. Recent years, the rich range of historic documents on Taiwan all collected through a variety of channels. National Central Library entered into a cooperative agreement with Genealogical Society of Utah (GSU) in 2004. Also acquired copies of materials in microfilm collected from civil society in Taiwan in 1970—there are contain a total of more than 794 microfilms and more than 9000 records of family genealogy [8]. In order to properly preserve these genealogies of Taiwan, National Central Library used digital technology. It makes the expedience as the scholars doing the related research. Scholars can study the genealogy by browsing features. When data is collected, Hakka surname as judgment standard by the expert opinions. Study on genealogies and the genealogies catalogs are classified. After the analysis and filter the unnecessary catalogs and Hakka culture catalogs has be reserved. This study designs the pedigree metadata standards for Hakka to implement the union catalog.

4. Research Results

4.1. Hakka Genealogical Metadata

This research study collected related literature about Hakka surnames, concluded of the 28 Hakka surnames. The reference genealogy surname standards in accordance with 28 Hakka surnames. It had difficulty to obtain because genealogy is gray literature. This study only refers to 25 genealogies. In this study, reference 25 genealogies compiled pedigree metadata. In addition, the catalog of Hakka culture will be preserved. In this study, the genealogy of information is program to divide into five constructs, these constructs contains many sub-constructs. The genealogy of information is consists of pedigree, individual, family history, bibliography, miscellany. Key concepts are presented in **Figure 1**.

Genealogy of category contains the following types of constructs. Pedigree: Use given name and lines expressed the genetically related. Individual: Describe the personal information. Family History: The story of the ancestors is

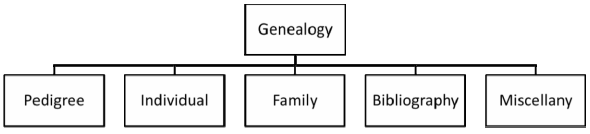

Figure 1. The sub-constructs of genealogy.

narrated. Bibliography: Description of the genealogical book. Miscellany: It's not categorized in the above four classification, including some unique documents (**Figure 1)**.

Pedigree consists of the following three constructs: Surname, First Name, Generation and Lang (Male). "Lang" is the digital name transformed from the legacy of the Central Plains culture, is unique characteristics of Hakka Genealogy (**Figure 2**).

Individual comprises the following constructs: Generation, Given Name, Lang (Male), Parents, Present Place, Date of Birth, Date of Death, Age, Mate, Children, Experience, Career. Tomb is divided into Tomb Place and Feng Shui. It is worth mentioning that Feng Shui is a unique culture of Hakka. Hakka people feel that Feng Shui has a mysterious power and will affect their fortune (**Figure 3**).

The history of Hakka family comprises the following constructs: Origins are divided into Origin of Surname, Native Place, Migration Process, The name of Ancestral Hall. Honor included files record that offspring to become the officer. Regulation contains the Family Instructions. Ancestral Hall contains Family Temple. Ancestors of Celebrities to be recorded Given Name, Birthplace, Portrait, and Experience. It's characteristics of the Central Plains culture that don't forget own origins. Therefore, Hakka people will carefully record own origins and ancestors. Hakka people comply with the teachings of own ancestors and build family temple. In addition, Hakka people pay attention to the glory. They hope family is proud of its own. Hakka people will record the glorious deeds of their ancestors, and regarded this as his glory (**Figure 4**).

Bibliography comprises the following constructs, Title, Genealogy Introduction and Creator (**Figure 5**).

Miscellany consists of various files. This study can't give a fixed form. In this work, we provide collection of document as a reference. Miscellany comprises the following constructs, Renovation tomb of the Log and Contract. The deceased was mourning article, Petition, Surname of Songs, Door couplet of the Ancestral Hall, and Ancestral Worship Rules. It will write on couple that the name of the Ancestral Hall implied meaning. The couple will be posted at the door (**Figure 6**).

The pedigree metadata standards for Hakka to implement the union catalog are presented in **Table 2**. The pedigree metadata have a total of five constructs and 31 sub-constructs.

4.2. Case Study: Tseng's Genealogy Design

This study selected one Hakka Family's genealogy as the case. After the complete content analysis, this study takes Tseng's genealogy to exhibit how this designed metadata has been applied in a real situation. This study used three

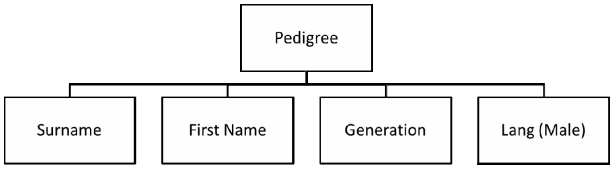

Figure 2. The sub-constructs of pedigree.

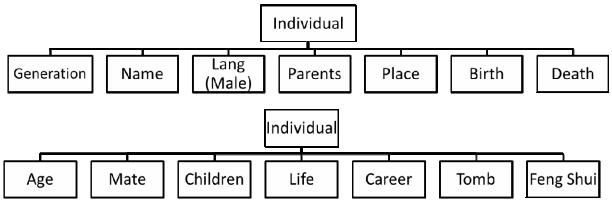

Figure 3. The sub-constructs of Individual.

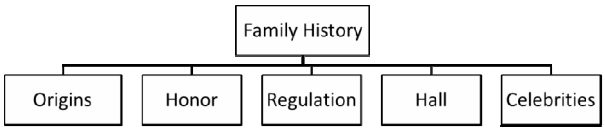

Figure 4. The sub-constructs of Family History.

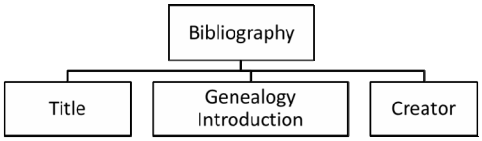

Figure 5. The sub-constructs of Bibliography.

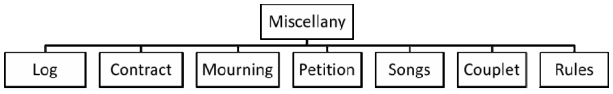

Figure 6. The sub-constructs of Miscellany.

dimensions to describe Tseng's genealogy metadata which contains the following data:

1) Pedigree consists of the following constructs: First Name (Yue-Yun), Last Name (Tseng), Father (Ming-Yu), Children (Yue-Fun), Generation (22). Please see **Table 3**.

2) Individual consists of the following constructs: Generation (22), First Name (Yue-Yun), Lang (Hang Shao Three Lang), Parents Father (Jin-Sheng Tseng), Mother (Mi -Mei Yu), Present Place (Wugu Rd., Miaoli City, Miaoli County, Taiwan), Date of Birth (1934/7/25), Date of Death (1986/9/12), Age (53), Tomb Place (Jhonghua Rd., Miaoli City, Miaoli County, Taiwan), Feng Shui (West toward East), Career (Policeman), Mate (Si-Mei Liu), Children (Chang-Ren Tseng), Life (Born in Miaoli. He graduated from Miaoli Elementary School.). Please see **Table 4**.

3) Family History consists of the following constructs: Origins (Noble offspring), Native Place (Miaoli), Migration Process (Family migrated in 1785 from Taipei to Miaoli because the war), The name of Ancestral (Hall

Table 2. The pedigree metadata standards for Hakka to implement the union catalog.

Constructs	Items	Subitems
Pedigree	First Name	
	Generation	
Individual	Generation	
	First Name	
	Lang	
	Parents	Father
		Mother
	Present Place	
	Date of Birth	
	Date of Death	
	Age	
	Tomb	Tomb Place
		Feng Shui
	Career	
	Mate	
	Children	
	Life	
Family History	Origins	Origin of Surname
		Native Place
		Migration Process
	Honor	The name of Ancestral Hall
		Files record that offspring to become the officer
	Regulation	Family Instructions
	Ancestors of Celebrities	First Name
		Birthplace
		Portrait
		One's Life

Table 3. The Pedigree metadata for Tseng's genealogy.

Constructs	Items	Subitems	Value
Pedigree	First Name		Yue-Yun
	Last Name		Tseng
	Father		Jin-Sheng
	Children		Chang-Ren
	Generation		22

Pengcheng Hall), Honor (Files record that offspring to become the officer Chin Tseng, General), Regulation Family Instructions (Failure is the mother of success). Please see **Table 5**.

5. Discussion and Implications

In this paper, we present designs the pedigree metadata standards for Hakka to implement the union catalog. Well-designed Hakka Pedigree Metadata makes it is easier to merge genealogy and provide offspring easy to find their ancestors. The Hakka Pedigree Metadata is comprised of many individual family and ancestors, each of which has a process in place for archiving and distribution of genealogy. Hakka genealogical metadata contains the following types of constructs. 1) Pedigree: use given

Table 4. The individual metadata for Tseng's genealogy.

Constructs	Items	Subitems	Value
Individual	Generation		22
	First Name		Yue-Yun
	Lang		Hang Shao Three Lang
	Parents	Father	Jin-Sheng Tseng
		Mother	Mi-Mei Yu
	Present Place		1, Lienda, Miaoli 36003, Taiwan
	Date of Birth		1934/7/25
	Date of Death		1986/9/12
	Age		53
	Tomb	Tomb Place	The Miaoli Shihtan Hiroshi Court
		Feng Shui	West toward East
	Career		Policeman
	Mate		Si-Mei Liu
	Children		Chang-Ren Tseng
	Life		Born in Miaoli. He graduated from Miaoli Elementary School.

Table 5. The Family History metadata for Tseng's genealogy.

Constructs	Items	Subitems	Value
Family History	Origins	Origin of Surname	Noble offspring
		Native Place	Miaoli
		Migration Process	Family migrated in 1785 from Taipei to Miaoli because the war.
	Honor	The name of Ancestral Hall	Pengcheng Hall
		Files record that offspring to become the officer	Chin Taeng General
	Regulation	Family Instructions	Failure is the mother of success
	Ancestors of Celebrities	First Name	Zi Tseng
		Birthplace	1956/1/1
		Portrait	
		One's Life	

name and lines expressed the genetically related. 2) Individual: describe the personal information. 3) Family History: the story of the ancestors is narrated. 4) Bibliography: description of the genealogical book. 5) Miscellany: included the unique documents that could not categorized in the above four classification. Results of this study can be applied to genealogies integration and implement the union catalog. In the future, information technology can be applied to genealogy integration and implement. The Hakka Pedigree Union Catalog is an attempt to make these individual family pedigree charts appear as one seamless genealogy to family and researchers seeking out the connection and family history.

REFERENCES

[1] R. C. Griffiths and P. Marjoram, "Ancestral Inference from Samples of DNA Sequences with Recombination," *Journal of Computational Biology*, Vol. 3, No. 4, 1996, pp. 479-502. doi:10.1089/cmb.1996.3.479

[2] A. J. Drummond, *et al*., "Estimating Mutation Parameters, Population History and Genealogy Simultaneously from Temporally Spaced Sequence Data," *Genetics*, Vol. 161, No. 3, 2002, pp. 1307-1320.

[3] C. L. Liaw, "The Chinese genealogy," SMC Co., Ltd., Taipei, 2003.

[4] J. M. Meskill, "The Chinese Genealogy as a Research Source," Family and Kinship in Chinese Society, Calif, 1970, pp. 139-161.

[5] K. C. Zatloukal and N. J. A. Harvey, "Family Trees: An Ordered Dictionary with Optimal Congestion, Locality, Degree, and Search Time," *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, New Orleans, 11-14 January 2004, pp. 308-317.

[6] M. K. Kuhner, "Coalescent Genealogy Samplers: Windows into Population History," *Trends in Ecology & Evolution*, Vol. 24, No. 2, 2009, pp. 86-93. doi:10.1089/cmb.1996.3.479

[7] "Ancestry," 2012. http://www.ancestry.com/

[8] National Central Library, "Taiwan Memory," 2012. http://memory.ncl.edu.tw/tm_cgi_test/hypage.cgi?HYPAGE=about_tm_e.hpg

[9] C. Chen, *et al*., "The Design of Metadata for the Digital Museum Initiative in Taiwan," *Online Information Review*, Vol. 26, No. 5, 2002, pp. 295-306. doi:10.1108/14684520210447868

[10] C. C. Yang and W. W. M. Chan, "Metadata Design for Chinese Medicine Digital Library Using XML," *Proceedings of the* 33*rd Annual Hawaii International Conference on System Sciences*, Maui, 4-7 January 2000, p. 10

[11] Y.-N. Chen and S.-J. Chen, "An Analysis of Metadata Research and Application Development in LIS Based on a Hybrid Approach of Quantity and Quality," *Bulletin of Library and Information Science*, No. 55, 2005, pp. 51-77.

[12] L. Dempsey and R. Heery, "A Review of Metadata: A Survey of Current Resource Description Formats," The UK Office for Library and Information Networking, 1997.

[13] DCMI, "Dublin Core Metadata Initiative," 2012. http://dublincore.org/documents/usageguide/

[14] GEDCOM, 2012. http://www.gedcoms.com/

[15] GedML, "GedML: Genealogical Data in XML," 2012. http://homepage.ntlworld.com/michael.h.kay/gedml/index.html

[16] GEDC, "GEDC—An XML-Based Standard for Genealogy," 2012. http://www.sunflower.com/~billk/GEDC/

[17] GEDCOMX, "GEDCOM X," 2012. http://www.gedcomx.org/Data-Models.html

[18] J.-X. Xie, "The Hakkas Whose Roots Are in Central Plains," *Journal of Huanghe S&T University*, Vol. 9, No. 3, 2007, pp. 15-19.

[19] X.-J. Li and X.-Y. Li, "On Hakka's Values from Family Rules and Instructions of Their Pedigree Records," *Journal of Longyan University*, Vol. 23, No. 5, 2005, pp. 71-73.

[20] S. Xie, "Hakka formation Development Outline of History," South China University of Technology Press, Guangzhou, 2001

[21] X.-Y. Li, "Values Reflected in the Family Rules and Regulations of Hakka Genealogy," *Journal of Guangxi University for Nationalities* (*Philosophy and Social Science Edition*), Vol. 28, No. 3, 2006, pp. 77-80.

