Mediating Data Conflicts on Integrated Land and Housing Information in China

Liu Lingling
March, 2007
Mediating Data Conflicts on Integrated Land and Housing Information in China

by

Liu Lingling

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-Information Management.

Thesis Assessment Board

Chairman: Prof. Ir. P. van der Molen
External Examiner: Dr. Mr. Ir. J.A. Zevenbergen
First supervisor: Dr. A.M. Tuladhar
Second supervisor: Dr. Ir. R.L.G. Lemmens
Disclaimer

This document describes work undertaken as part of a programme of study at the International Institute for Geo-information Science and Earth Observation. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.
Abstract

The rapid increase in the use of Information Technology has improved the methods of acquiring, storing and disseminating information for many applications, of which geographic data is part. This has led to the setting up of different databases and different formats for storing and exchanging data within and across many organisations. The major problem that had arisen from this global development is the sharing and exchange of data across these organisations. One of the major problems in data sharing today is that data is scattered separately in distributed heterogeneous environment, and therefore it is difficult to obtain data from them. Even the all necessary data are available at the users’ desk; there are still problems of integration such as duplication and mismatches of same data of different sources. This research focuses on resolving data conflicts using mediation technique within the concepts of federated data model where the user access integrated data without affecting any changes on the original data sources.

Cadastral and Housing data in the People’s Republic of China are managed by two different bureaus, where the data are separately stored and supplied to users. Such a system brings duplication of datasets between land administrative bureau and housing administrative bureau resulting in waste of money and inefficiencies. The research addresses this inefficiency through the development of federated data model to supply integrated data to the various users. Fieldwork is conducted for the detailed analysis of the As_Is situation of the Chinese cadastral and housing management system through structured and unstructured interviews and needs of integrated data is assessed from the viewpoint of the various stakeholders.

A Federated Data Model is then developed, and it includes design of cadastral data model, housing data model, integrated data model and mediation as a means to resolve data conflicts. The integrated schema defined according to the integrated data model and user requirements was created by XMLSpy. The federated data model was tested and implemented using XMLSpy, MapForce, Microsoft Access and Dreamweaver. This modest experiment shows that five differences including structure conflicts, domain conflicts, intentional conflicts, naming conflicts and missing attributes existed in these two types of datasets. Finally this research finds that the concepts of federation and federated data model bring a step forward for resolving data conflicts using mediation technique.

Key words: federated data model (FDM), heterogeneity, conflicts, mediation, and integrated data
Acknowledgement

First of all, I would like to express my heartfelt gratitude to ITC for giving me the opportunity to pursue my study in the Netherlands. Especially, Ir. M. C. Bronsveld and Prof. Ma Zhiming are making great efforts on this opportunity.

My deepest and foremost gratefulness is to my first supervisor Dr. A. M. Tuladhar. If without his encouragement, invaluable suggestions and constructive comments, it is impossible for me to finish this research. Besides, many thanks to him for concerns about study life in Netherlands, which make me feel like parents stay around. I am also thankful to my second supervisor Dr. Ir. R. L. G. Lemmens, whose support and positivism always encourage me. Special thanks to Prof. Ir. P. van der Molen who gave me encouragements and professional questions in proposal and mid-term defences, this makes me more focus on key points of research.

I am grateful to thank my dear friends who support me a lot in fieldwork in China; they are Prof. Ma Zhiming in Chang’an University; Mr. Guo Junsheng the Deputy Director-General in Xiamen Municipal State Resources and Housing Administrative Bureau (XMSRHAB); Ms. Zhao Daihong a staff of State Resources Information Centre; Ms. Zhang Ning a staff in Cadastre Department of Land Survey and Planning Institute; Ms. Shang Yaoling, General Director of National Geomatics Center of China (NGCC); Mr. Zhang Beifei, High Engineer of NGCC; Mr. Jiang Chunfa, Senior engineer in Survey and Geo-information Centre in XMSRHAB; Mr. Li Guiming, Senior engineer in housing transaction management in XMSRHAB; Ms. Liu Qing and Mr. Qiao Liang who are my schoolmates. I would also like to give my great thanks to my dear friends Ms. Zhu Qian and Ms. Lv Yanhong who gave me great support whenever I met with difficulties in research and life. Many thanks are also belonging to Mr. Mansour Ahmadi Foroushani, Mr. Anthony Arko Adjei and Mr. Pu Shi, for their help and encouragements during this research.

I wish to express my sincere appreciation to my dear teacher Ms. Li Xia who teaches me both scientific and life knowledge and skills that quite useful in my life. I also want to thank Mr. Zhang Wei for helping me a lot. My countrymen at ITC are so nice that they enrich my life, dismiss my homesick and give me precious advice in life. They are Ms. Li Xia, Ms. Yang Na, Mr. Song Yu, Ms. Bai Lei, Mr. Wang Tiejun, Ms. Zeng Yu, Mr. Wang Langyue, Mr. Yang Zhenshan, and Mr. Shang Xiao.

I still want to appreciate my thanks to Walter de Vries, Javier Morales and Sjef van der Steen who gave me lots of help during the modules. Thanks are non-doubtful also belonging to my dear classmates in GIM 2005, such as Mr. Abdul Hanan Iddirisu Abu, Mr. Augusto Cesar Darce Hernandez, Mrs. Julia Maria Galindo Coronado. My thanks also go to the ITC staff members who provide me a comfortable study and live environment for their efficient and careful work. Saskia Groenendijk, is very friendly and Vinke Bennie who impressed me a lot that he is always quite happy. Ms. A. W. S. M. Bettine Geerdink and Ms. J. M. Jacqueline provided great help to me.

Finally, my heart felt gratitude goes to my parents, my sisters and brothers, my uncles and aunts, my boyfriend who gave me the courage and support during every moment of my life. I want to dedicate my thesis to my parents for their ocean-deep love.
# Table of contents

1. Introduction.......................................................................................................................... 1
   1.1. Background.................................................................................................................. 1
   1.2. Brief description of current situation in China ....................................................... 2
   1.3. Research problem .................................................................................................... 3
   1.4. Prior Works............................................................................................................. 4
   1.5. Research scope........................................................................................................ 5
   1.6. Research Objective ............................................................................................... 6
   1.7. Research questions description............................................................................. 6
   1.8. Research Methodologies......................................................................................... 7
      1.8.1. Literature review.............................................................................................. 7
      1.8.2. Data collection in fieldwork ....................................................................... 7
      1.8.3. Design a model ............................................................................................ 7
      1.8.4. Demonstrating practical case as an prototype.......................................... 7
   1.9. Main structure of thesis ......................................................................................... 7

2. Data integration concepts..................................................................................................... 9
   2.1. Introduction............................................................................................................. 9
   2.2. Integrated data services......................................................................................... 9
   2.3. Real estate............................................................................................................. 10
      2.3.1. Land ........................................................................................................... 10
      2.3.2. Housing.................................................................................................... 10
      2.3.3. Means of services on land and housing ................................................... 10
   2.4. Spatial Data Infrastructure as a supporting infrastructure .................................... 11
      2.4.1. SDI definitions........................................................................................... 11
      2.4.2. SDI components....................................................................................... 11
   2.5. Database systems.................................................................................................. 12
      2.5.1. Distributed database system (DDBS) ....................................................... 13
      2.5.2. Federated database system (FDBS) ........................................................... 13
   2.6. Federated Data Model (FDM) ............................................................................. 14
   2.7. Conflicts and their solving approaches................................................................ 15
      2.7.1. Semantic conflicts.................................................................................. 15
      2.7.2. Semantic conflicts managements........................................................... 16
      2.7.3. Comparison of conflicts managements.................................................. 18
   2.8. Relevant Tools in research.................................................................................... 18
      2.8.1. UML ...................................................................................................... 18
      2.8.2. XMLSpy ................................................................................................ 18
      2.8.3. MapForce............................................................................................... 19
      2.8.4. Dreamweaver ......................................................................................... 19
   2.9. Concluding remarks............................................................................................... 19

3. Analysis of As_Is Situations in China............................................................................... 21
   3.1. Introduction.......................................................................................................... 21
   3.2. Fieldwork results ................................................................................................. 21
      3.2.1. Laws governing on land and housing ....................................................... 21
      3.2.2. Institutional framework ........................................................................... 22
      3.2.3. Technology used ..................................................................................... 25
      3.2.4. Data capture ............................................................................................ 27
      3.2.5. Human resources ...................................................................................... 28
3.3. Justification for FDM be used in China

3.3.1. Stability of political institutional framework

3.3.2. Possible infrastructure and governmental work emphases

3.3.3. Stimulation of economy

3.3.4. Effective and flexible service to present and potential users

3.3.5. Foster citizen’s trust on government

3.3.6. Helpful to real estate brokers

3.3.7. Accelerate the NSDI

3.3.8. Better services related to cadastral and housing data

3.4. Concluding remarks

4. Data Analysis on Cadastral Data and Housing Data

4.1. Introduction

4.2. Brief introduction on land data in Land Information System

4.2.1. Basic geographic data related to land

4.2.2. Thematic geometry data

4.2.3. Thematic attribute data

4.2.4. Other data related to land

4.3. Criteria for data analysis

4.4. Cadastral data

4.4.1. Brief introduction of CIS

4.4.2. Role and formats of cadastral data

4.4.3. Analysis of cadastral data

4.4.4. Datasets in cadastral managements

4.5. Housing Data

4.5.1. Basic geographic data

4.5.2. Housing thematic data

4.5.3. Housing property data and main types of data

4.6. Data covering both cadastral and housing

4.7. Main types of data conflicts in cadastral and housing databases

4.8. Concluding remarks

5. Development of Federated Data Model

5.1. Introduction

5.2. FDM as approach for solving conflicts

5.3. Framework for mediation

5.3.1. Pre-integration

5.3.2. Identifications

5.3.3. Schema integration

5.3.4. Mapping definition

5.4. UML data model diagrams for FDM

5.4.1. Data model for cadastral property data

5.4.2. Data model for housing data

5.4.3. Integrated data model

5.4.4. Mediation functions

5.5. Practical case for demonstrating FDM as prototype

5.5.1. Setting up cadastral property schema and data by XMLSpy

5.5.2. Setting up housing property schema and data by Microsoft Access

5.5.3. Setting up Export-schema based on a scenario requirements

5.5.4. Integration the two schemas by MapForce

5.5.5. Creation of an XSLT file and by Dreamweaver
List of Figures

Figure 1-1 As-is situation in China ......................................................................................................... 3
Figure 1-2 To-be situation in China ........................................................................................................ 4
Figure 2-1: Taxonomy of distributed data and multi-database systems (adapted from (Amit and James 1990)) .................................................................................................................................................... 12
Figure 2-2 FDM (Columbia 2005) ....................................................................................................... 15
Figure 3-1 Institutional structure in Chinese land resources management and relationships with local levels (left) ............................................................................................................................................. 22
Figure 3-2 Institutional structure in Chinese housing management and relationships with local levels (right) ..................................................................................................................................................... 22
Figure 3-3: Institutional frameworks of Survey and Mapping Bureaus in national and provincial level ............................................................................................................................................................... 24
Figure 3-4: Data produced by SMB ...................................................................................................... 25
Figure 3-5: Educational background of employees in land and housing administration ....................... 29
Figure 4-1: Classifications of cadastral data ......................................................................................... 38
Figure 4-2: Three core classes in cadastral management ...................................................................... 40
Figure 4-3: Illustrations of relationships among four types of data (Top view) ........................................ 43
Figure 4-4: Illustrations of relationships among four types of data (Front view) ......................................... 43
Figure 4-5: Housing properties data model ........................................................................................... 48
Figure 5-1: Integrated service supply (adapted from Tuladhar’s) .......................................................... 54
Figure 5-2: Framework for data integration (adopted from (Christine Parent 2000)) .......................... 54
Figure 5-3: Packages in FDM................................................................................................................ 56
Figure 5-4: Cadastral data model in Chinese city .................................................................................. 56
Figure 5-5: Housing data model in Chinese city ................................................................................... 57
Figure 5-6: Integrated data model in Chinese city ................................................................................ 58
Figure 5-7: Mediation package.............................................................................................................. 59
Figure 5-8: Cadastral schemas (some parts) ....................................................................................... 60
Figure 5-9: Housing property schema ............................................................................................... 61
Figure 5-10: Export schema ............................................................................................................... 62
Figure 5-11: The upper part of integration ............................................................................................ 63
Figure 5-12: The lower part of integration ........................................................................................... 63
List of Tables

Table 3-1: Types of services related to land and housing ................................................................. 32
Table 4-1: Dataset information on person.......................................................................................... 40
Table 4-2: Dataset information on user ship right ............................................................................ 41
Table 4-3: Dataset information on parcel ....................................................................................... 41
Table 4-4: Dataset information on land housing block ..................................................................... 44
Table 4-5: Dataset information on building .................................................................................... 45
Table 4-6: Dataset information on floor ............................................................................................ 46
Table 4-7: Dataset information on apartments .................................................................................. 47
Table 4-8: Basic datasets on housing ownership .............................................................................. 49
Table 4-9: Attributes associations for “person” in cadastral databases and housing databases ....... 50
Table 4-10: Attributes associations for “right” in cadastral databases and housing databases ........ 50
Table 4-11: Other relevant in cadastral databases and housing databases ........................................ 50
Table 5-1: Integrated data as output.................................................................................................. 65
## List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>P.R.C</td>
<td>People’s Republic of China</td>
</tr>
<tr>
<td>MLR</td>
<td>Ministry of Land and Resources</td>
</tr>
<tr>
<td>MC</td>
<td>Ministry of Construction</td>
</tr>
<tr>
<td>SDI</td>
<td>Spatial data infrastructure</td>
</tr>
<tr>
<td>FDM</td>
<td>Federated data model</td>
</tr>
<tr>
<td>LAB</td>
<td>Land Administrative Bureau</td>
</tr>
<tr>
<td>HAB</td>
<td>Housing Administrative Bureau</td>
</tr>
<tr>
<td>DDBS</td>
<td>Distributed database system</td>
</tr>
<tr>
<td>XMSRHAB</td>
<td>Xia’men Municipal State Resources and Housing Administrative Bureau</td>
</tr>
<tr>
<td>FDBS</td>
<td>Federated database system</td>
</tr>
<tr>
<td>UML</td>
<td>Unified Modeling Language</td>
</tr>
<tr>
<td>OWL</td>
<td>Web Ontology Language</td>
</tr>
<tr>
<td>XSLT</td>
<td>the Extensible Stylesheet Language Transformations</td>
</tr>
<tr>
<td>SOAP</td>
<td>Simple Object Access Protocol</td>
</tr>
<tr>
<td>WSDL</td>
<td>Web Services Description Language</td>
</tr>
<tr>
<td>LAL</td>
<td>Land Administration Law of the P.R.C</td>
</tr>
<tr>
<td>UREAL</td>
<td>Urban Real Estate Administration Law of the P.R.C</td>
</tr>
<tr>
<td>SMB</td>
<td>Survey and Mapping Bureaus</td>
</tr>
<tr>
<td>BMSRAB</td>
<td>Beijing Municipal State Resources Administrative Bureau</td>
</tr>
<tr>
<td>BHAC</td>
<td>Beijing Housing administrative committee</td>
</tr>
<tr>
<td>NSDI</td>
<td>National spatial data infrastructure</td>
</tr>
<tr>
<td>LIS</td>
<td>Land information system</td>
</tr>
<tr>
<td>CIS</td>
<td>Cadastral information system</td>
</tr>
<tr>
<td>HIS</td>
<td>Housing information systems</td>
</tr>
<tr>
<td>GCS</td>
<td>Geodetic Coordinate System</td>
</tr>
</tbody>
</table>
1. Introduction

1.1. Background

Given the increasingly demanding on land, the optimum use and sustainability of it becomes essential. In recent years, many nations around the world have intensified their activities in this field of Land Administration including the cadastre, the real estate and related information systems. The best way to achieve the optimum use and sustainability is through partnerships between the data managers, especially that land related data (Dale 2000).

Cadastre mainly concerns an inventory of a nation-wide basic information of land parcels about the property rights, use and tax situation. It has various important roles in the national economic development and governance. When governments raise tax revenue according to the cadastre, equity must be ensured through a clear and accurate cadastral system (Dale 2000). Property right is the core of cadastral registration that protects legal rights of owners or users of certain parcel. Cadastre is the first-hand tool for making land-use planning, land policy reform and so on. Besides, cadastre also provides quality, quantity and legal transaction data guaranteed by state and society at large for the sake of a prosperous real estate market.

Moreover, real estate, also well known as immovable property is a term that encompasses land along with anything permanently affixed to the land, such as buildings and housings (Wikipedia 2004). Thus, besides land, building is another important component in the real estate management. Governments through appropriate institutions govern the operation of the real estate market and provide more reasonable housing price, management of housing buying or selling and other related aspects. It is also important for environmental protection and spatial planning policies, as well as other activities within the framework of a rapid and efficient transition to a market economy.

As stated above, land and housing data are geo-related and therefore have the attributes of geo-spatial data. In theoretical sense, both datasets must be delivered in an integrated fashion without any ambiguities or confusions to the users such as decision makers or buyers/sellers in the real estate market. In People’s Republic of China (P.R.C), both datasets are separately held by two ministries; the land data is handled by Ministry of Land and Resources (MLR) while building data is processed and disseminated by the Ministry of Construction (MC). Such environment poses interesting data integration challenges.

Data integration largely relates to the problems of interoperability between various data producing agencies. Interoperability can be defined as the ability of a system or components of a system to provide information sharing, and inter-application cooperative process control. This poses an urgent question to be answered because data from different agencies may be under different standards and in various formats. The heterogeneity problem occurs when communities willing to integrate the
different datasets have to contend with different views on the real world features, different modelling schemes, and different tools to represent, store, process, and manage geospatial data (Groot 2000).

Heterogeneities in geospatial data could be resolved under spatial data infrastructure (SDI) (Bishr 1997). SDI could be seen as a supporting mechanism for heterogeneities, thus in the context of SDI, integrating relevant data of lands and housings is meaningful. It would be ideal if integrated service on lands and housings is realized, and then the provision of the knowledge and insight required today would keep the users or the customers greatly aware of the dynamic nature of the real estate market.

Federated data model (FDM) which is featured for it can act as integrating layers on the top of the existing databases systems distributed at the different locations. FDM plays a crucial role to satisfy the global users, while individual databases at the different agencies keep satisfying their own local users. If such FDM implemented in a single synchronous and consistent way, users could get a consistent dataset for their uses at lowest possible cost (Tuladhar 2005).

This research is mainly concerned with the situation analysis of current Chinese land and housing management, and development of federated data model that is suitable for P.R.C to resolve the current integration problems of cadastral and housing data.

1.2. Brief description of current situation in China

As indicated in previous section, the Ministry of Land and Resources (MLR) is responsible for managing the land (i.e. land rights, owners/users), while the Ministry of Construction (MC) manages the real estate concerning houses in P. R. C. The conceptual difference is that the legal entities such as citizens or registered companies can have a right of use on land and at the same time they could retain ownership rights on the constructions (i.e. houses or buildings). Therefore, land and housing are two separate objects in Chinese mind, and real estate means only housing in many situations. Since the land and buildings are inseparable, both datasets are usually necessary for many applications in the land and buildings management.

The real estate industry has been a major source of strength for the improving economy in P. R. C especially after “the distributing apartment” mechanism was abolished in 1998. The concept of open market was then established for purchasing housing in real estate market. That is, everybody in China is required to obtain housing with charge from market instead of getting free of charge from the government. The importance of the real estate industry, as an engine of the nation’s growth, can be seen from the fact that it accounts for 7% in GDP in 2002 (Yin Xue 2004). Furthermore, the investments on real estate increase rapidly, $707.25 billion in the first 5 months in 2006, increased 21.8% to last year. The main determinants of the demand for housing in China are demographic. Demographers forecast in the coming years, the population will keep on increasing to 1331.1 million in 2010 (Eastern 2006). For the public, housing will be the most costly commodity they would ever buy. Therefore, today and even in the future, there is a high demanding for housing and its relevant data which are helpful to make decisions on investing, purchasing and management.
Since the land and housing markets are now growing issue due to open economic development policy in China, both of them have a large spectrum of stakeholders ranging from the government offices to ordinary people, such as: MLR, MC, State Bureau of Statistics, Tax Office, Municipality, Bank, Court and Citizen. Most of the stakeholders want complete data on land and housing including legal, economical and environmental data for their decision makings.

1.3. Research problem

Computer-based information systems connected to networks that increase access to a wide variety of data resources. These systems encounter several problems with cadastral and housing datasets separately processed and stored in MLR and MC respectively, the users or stakeholder have tremendous difficulties not only in retrieving them but also in integrating these datasets which are under different standards and formats. Consequently they are facing challenges in analyzing legal and economic consequences on certain real estate. Figure 1.1 shows As-Is situation where the stakeholders who have little ability in analysing are involved in the data integration. In fact, it is even inefficient to collect, process, store and disseminate data by two Ministries owned by government, duplications often occur as both of the two ministries collect the same datasets.

In order to provide one-stop service on land and housing market, the Chinese government decided to combine the two bureaus in most main cities around 1999. For example, Land Administrative Bureau (LAB) and Housing Administrative Bureau (HAB) were merged into State Resources and Housing Administrative Bureau in Beijing, Xi’an and other capital cities of many provinces. Because of the bureaucratic reasons, almost all of these bureaus have separated to their original situations around the year 2005. Currently only three cities (namely Shanghai, Tianjin and Chongqing) and Special Economic Zone (such as Xiamen and Zhuhai) have the combined bureaus for both land and housing. The three cities are of the four municipalities of the P.R.C, which are equivalent to provinces in China's administrative structure while the Special Economic Zone was created to be an experimental ground of capitalism in socialism with Chinese characteristics. However, they are taking initiatives in supplying land and housing services together.

If the concept of federation and FDM is introduced, the situation will change into a different one as Figure 1-2 shown. Every bureau could remain autonomous with their legal mandates and original
tasks, and therefore be not need to put these organizations in one bureau. The FDM concept resolves data integration problem by establishing intermediary through the concept of mediation that help the stakeholders to get a consistent and integrated data on land and housing. Such intermediary may be a professional agency.

Figure 1-2 To-be situation in China

Therefore, this research is about the studying on the As-Is situation, and the developing of FDM to accelerate the To-be situation will become true in China.

1.4. Prior Works

The few key articles that extended the interest for this research are discussed as follows:

Pilot projects are being carried out to ensure and support the operation and management of the tasks associated with land and housings. Heimbigner explained the vital concepts such as federal dictionary, federated architecture and the three basic data modelling primitives in federated data model and federated database architecture is also explained with detail schema description (Dennis and Dennis 1985).

In the 1990s, a working group of the Survey Authorities of the State of Germany developed a model of an integrated Official Cadastral Information System called ALKIS. ALKIS was used in combination with a redesign of the Official Topographic and Cartographic Information System ATKIS. This platform allowed cadastral data to be easily used for topographical mapping and vice versa. ALKIS adopted the integrated data model concept. User requirements were studied before the exact model design. At present, the ATKIS could satisfy demands from planning and environmental agencies as well as the military for such data independently (Kuhn 1991).

Yaser Bishr (1997) gave the definition of interoperability and further introduced interoperability at six different levels. GIS interoperability users can transparently access and share remote spatial databases and other spatial services, regardless of their underlying GIS platform.

Benchikha presented two approaches for the federated databases which supports the concept of viewpoint, the contribution to resolve of the problems of schema integration. Several constraints and
conflicts that usually occur in data management are intended to be solved by the integration of the viewpoint mechanism in federated databases (Fouzia, Mahmoud et al. 2001).

In Denmark, KMS which is a Danish governmental organization for national survey and cadastre together with other governmental organizations such as Ministry of Food and Ministry of Environment took the initiative to form a Spatial Data Service Community supplying services in a comprehensive way and also a national exchange format for geographic information. By solving such interoperability problems, services on several specifics are supplied through internet gradually (Hall 2002).

Tuladhar (2005) proposes a federated data model to land administration. The concepts on federated database management system and its architecture were briefly reviewed. The writer proposed the “three level federation architecture” by considering land administrative situation in Egypt.

The District of Columbia employs the Federated Geospatial Data Model to clarify the roles of government agencies in building and maintaining geospatial datasets. The goal is to share in the creation, use and maintenance of GIS datasets at the least possible cost; while providing District government staff, citizens, the media, and other users easy access to this resource. The Columbia government receives a success that satisfies their local and global users by the federated data model (Columbia 2005).

The Kadaster (Dutch Land Registry Office) had a stand-alone software application that clients could use to look up real estate Information which indicates it has a good partnership with the real estate business. More and more of Kadaster's clients are turning to Internet technology. They want their applications to be integrated as much as possible with those of Kadaster. Kadaster developed Kadaster-on-line in order to provide professional clients with up-to-date information that provides legal certainty (Kadaster 2006).

In this research, the FDM will be employed to Chinese land and housing management situation.

1.5. Research scope
According to Land Administration Law of the P.R.C, land resources management including cadastral management, land use planning, agrarian land protection, construction land management and other legally supervision. Housing management composes housing property registration management, housing land management, housing transaction management and other legally supervision. As the scope of land or housing management is wide, so there are some specifications for this research:

- Although such FDM concept can be applied to integrate any other land related data from many agencies, but in this research, land data is only referred to cadastral data while housing data refer to housing property data.

---

1 Housing land means land for housing construction
Therefore, this research specifically focuses on developing an approach in resolving semantic conflicts on land and building using the mediation concept.

As the process of data integration is both technical and institutional, this research is only focus on the technical part. For example, who will be the professional agencies and how it works are remaining research works.

As both the cadastral and housing data are spatial related, but as a modest research, only the non-spatial data are studied.

1.6. Research Objective

Research objective is set up by analyzing on current situation in China. Firstly, research scope will be listed out for later research study.

The main objective of this research is to develop and validate an approach for resolving semantic conflicts of land and housing data by means of federated data model with emphasis on customer-oriented services.

In order to achieve the above main objective, two specific objectives as have been set as follow:

- To assess existing situation of Chinese cadastral and housing management and datasets.
- To develop a federated data model that resolve semantic conflicts within federation architecture.

1.7. Research questions description

Based on the above specific objectives of this research, the following research questions have been formulated.

For the first Specific objective: the user requirements and data management situation are pre-requisite. The following questions should be answered:

Q1. What are the user classifications and requirements of cadastral and housing data?
Q2. Which departments take responsibility of cadastral and housing data?

For the second Specific objective: detail aspects on federated data model are discussed and then prototype using available software in demonstration to resolve conflicts in data integration.

Q3. What are the relevant concepts of data integration?
Q4. What are the difference and the relationship between cadastral and housing data?
Q5. What are the conflicts in provision of integrated cadastral and housing data?
Q6. How to resolve these conflicts?
Q7. How to create an integrated schema?
Q8. How to demonstrate the federated data model?
1.8. **Research Methodologies**

The methodologies used for this research are organized and described below.

### 1.8.1. Literature review

Literature review on:

- Present situation on land and housing data in general and related white papers.---Q4
- Concepts on real estate, database systems, data integration conflicts, federated architecture, federated data model and their requirements.---Q3, Q6, Q8
- Solutions on semantic conflicts issues.---Q6, Q7, Q8
- The previous achievements of federated data modelling. ---Q7
- Demonstration and related software. ---Q3, Q8

### 1.8.2. Data collection in fieldwork

- Grasp the current land and housing data in China, such as data types, data relationships, and data dissemination.---Q1, Q5
- Get more information of what kinds of different data are urgently needed by various customers.---Q2

Data collection in fieldwork was carried out through observations, interviewing, documentations and questionnaires.

### 1.8.3. Design a model

Design a model of federated data model to fulfil the main objective. Design a model is the main part of this research, and the Federated Data Model is carried out by United Modeling Language (UML).---Q6, Q7

### 1.8.4. Demonstrating practical case as an prototype

Case is the key methodology for Q8 and it is an important part for this research. The case demonstrating was taken in the following steps:

- Scenario: create a scenario based on situation of acquiring both cadastral and housing data.
- Setting up schemas: setting a cadastral schema in XMLSpy and a housing data schema in Microsoft Access.
- Integrating the schemas by the help of MapForce and make a output with the help of Dreamweaver.

### 1.9. **Main structure of thesis**

**Chapter 1**: Introduction
Mediating Data Conflicts on Integrated Land and Housing Information in China

This chapter provides an overview of the research. It presents the background to the study, the current situation in China, research problem, key references to prior works done. It further outlines the objectives of the research, the research questions and the methodologies used to answer these questions. Finally, it gives an overview of the structure of the thesis.

Chapter 2: Data integration concepts

This chapter critically reviews concepts of integrated services, real estate, SDI, database systems and federated data model with emphasis on conflicts and conflicts managements. Besides, relevant tools and their functions in this research will also be introduced.

Chapter 3: Analysis of As_Is Situations in China

This chapter describes the present situation about institutional situation, laws, technologies and data acquisitions related to land and housing management in China. Besides, justifications for FDM is used in P.R.C from 9 aspects, that are institutional framework, possible infrastructure, economy, effective and flexible service, citizen’s trust, real estate brokers, NSDI, better services and potential users. User classification and data mostly needed are analyzed according to questionnaires in field work will also be presented in the better services part.

Chapter 4: Data Analysis on Cadastral Data and Housing Data

This chapter introduces the present brief condition of land management system, cadastral information system and housing management system in China. It also describes datasets in cadastral management and housing data in housing management. The detailed cadastral and housing datasets and their relationships are also presented. In this way, data conflicts between these two kinds of data are analyzed.

Chapter 5: Development of Federated Data Model

This chapter details the description of datasets that are needed in the FDM modelling and discusses steps for developing federated data model in China based on the analysis in chapter 3 and chapter 4 using UML. Two simple databases schemas cadastral and housing are set up by MS Access and XMLSpy. At last, the integration process is demonstrated by MapForce.

Chapter 6: Conclusions and Recommendations

In this final chapter conclusions are drawn from previous chapters and recommendations for further researches will be proposed.
2. Data integration concepts

2.1. Introduction

The fast development of high-technology, ways of communication and its content take a great change in the past a few years. New ideas about how to improve the quality of communication are welling up a lot. Real estate industry as a foundational one in national economy is playing a more and more vital role through high-technology hardware and software. Geospatial data is a kind of popular data at present and ways communicating it varies from raw paper ones to digital ones. The necessity of geospatial data especially cadastral and housing data are understood in many area, purchasing, investing, urban planning, strategy planning or for just any business promotion. Therefore, data services and infrastructure associated with them attract more attentions.

Therefore, integrated data services, SDI, Federated data model (FDM) and conflicts are basic concepts in this research. Database systems are of vital importance in building any FDM. This chapter describes these concepts and further discusses the federated data model and the conflicts when integrating two different datasets. The chapter also introduces the tools for demonstrate FDM and tool functions.

2.2. Integrated data services

In general, integrated data service is a new conception which has just come into being in recent years and it is popular when referring to computer networking. In that context, integrated data service is an architecture that specifies the elements to guarantee quality of service on networks. Integrated service can for example be used to allow video and sound to reach the receiver without interruption (Wikipedia 2006).

Some other idea about the integrated data services is that they are almost exclusively with the time-of-delivery and are necessary to meet the growing need for real-time service for a variety of new applications, including teleconferencing, remote seminars, tele-science, and distributed simulation (R. Rajan 1998).

However, in this research, integrated data service means spatial related-data which is supplied by different organizations originally but now supplied together. Thus, from the user perspective, they received a comprehensive data which could fully satisfy their needs and without taking time and effort to get from various organizations. As far as users of land and housing data concerned, if they could obtain the processed data, their particular requirements can be satisfied through different kinds of media mass and save money and time as well, then they enjoy the convenient of integrated data services.
2.3. Real estate

The real estate industry developed dramatically in the whole world in recent years as a result of economic prosperous and high demand for housing in some urbanized areas. Real estate management is therefore becoming more and more important for it is a decision making associated with land and housing issues at both state and local levels. A sound and effective real estate management could supply reliable and accessible information on land and housing tailored to their users’ needs. Generally, real estate is composes of land and housing.

2.3.1. Land

Land is manifested as a set of rights for its use with a value that can be traded even though the physical object cannot be moved. Land values and land use are also of great concerned. All these types of information are high stressed in land market. Land market is the most direct manner with common cities and investors. As a result, the data related to land market is high-demanded. Furthermore, land transaction, land price and land supplies are three main aspects that account for land data. However, among all of the land data, property data about the relationship between human beings and the land is of fundamental importance in every society and is evident in the form of property rights (Dale 2000).

2.3.2. Housing

Human beings need shelter as a basic requirement for living. Besides, they also need housing with the purpose of commercial, industrial and military. In many countries, real estate mainly refers to housing. For the reason that housing gives an obvious impression to the public and its conditions are diverse dramatically depending on works people do on it. Thus, when purchasing housings, besides its location which referring to land, conditions of housing are key factors one shall have to consider these days. However, housing property data is also the most fundamental one among all of the housing data as it defines who owns housing and owner’s right over housing.

2.3.3. Means of services on land and housing

Services on both land and housing come through the simple ones to the complex one in hand writing and paper documentation to printer and computer management in the last decade. In cyberspace world, much of the work is done by computer; users could get the data they want directly from the databases. With the help of internet, services in some part of the world could be transferred to certain places distant away. In some developed countries, mass media could be used to display these services in an interactive way in which the customer could interact with the data supplier directly.

Modern data supplying methods make the possibility of more complex data supplying. At present, the services on land and housings are required to be supplied in an integrated way. The communication between different data management systems should be compatible with each other. In this way, comprehensive and professional data could satisfy more users.
2.4. **Spatial Data Infrastructure as a supporting infrastructure**

It is clear that data the incoming trends require data to be made available and easily accessible to many users. SDI principles ensure that use of data is optimized and cost associated with collecting already existing data is eliminated. Synthesis of geospatial information with other information is available these years and its devices improved from paper to digital mobile, wireless and location-aware devices. Thus, the information technology infrastructure will foster the information services sector.

SDI is a basic infrastructure on which countries are making great effort and it is a factor in fostering land and housing data provisions for it provides an available infrastructure to make data provisions become true.

2.4.1. **SDI definitions**

SDI is understood and described differently by stakeholders from different disciplines and different administrative/political levels (Tai On Chan1 2000). In the SDI cookbook Nebert defines SDI as the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data, thereby providing a basis for spatial data discovery, evaluation and application for users and providers within all levels of government, the commercial sector, non-profit sector, academia and citizens in general (Nebert 2004).

2.4.2. **SDI components**

The components of SDI are policy framework, technology framework, data acquisition and display and human resources (Groot 2000). These aspects work together to enable data acquisition, store, evaluation, distribution, accessibility and use smoothly.

2.4.2.1. **Policy and Institutional framework**

As globalization is taking place everywhere in the world, technology goes hand in hand in most industrialized countries. Countries that lack technology find it difficult to catch up various hitech methods. Policy framework comprises legal perspective, funding perspective, standards, management and cultural factors are confronted with some degree of incongruence between various legislation, regulations and cultural conflicts. Policy and institutional framework make a promise for any reform or novel idea become true.

2.4.2.2. **Technology framework**

Spatial data infrastructures only function effectively when reliable and efficient computing and communication technologies are in place (Groot 2000). Various databases have to be linked together to transfer data and provide a basic software and hardware systems. Data capture technologies like GPS, scanners, digitizers etc ensure that data is acquired. Work benches for abstraction, modeling and validation of data, as well as logical and physical storage technologies are important for sustainable spatial data availability. Communication technologies ensure that users could get what they want. Computer technology, network technology and database technology are playing dominant role in SDI.
2.4.2.3. Data acquisition

Data is the most essential part in SDI for without spatial data, the concept of SDI loses its meaning. Data acquisition as the initial step plays a vital role in the whole process. Geo-database and metadata form two core elements of data development. Geo-database refers to a database which stores spatial data in addition to attribute data. Metadata describes spatial data to the users and facilitates data understanding in an SDI. Efficiency, consistency, and accuracy of positioning methods and systems affect the quality of data integration when doing exchange and sharing of data amongst users, thus, ultimately affect the effectiveness of SDI.

2.4.2.4. People

People are the users who apply spatial data and whose needs are met using the spatial data in an SDI environment. Demand for skilled persons familiar with their operation and management has greatly increased with high techniques applied into SDI domain. Skilled persons, to some extent, are the key factor driving SDI development. Processes of spatial data production and relevant services must be geared towards users needs.

2.5. Database systems

With the explosion of information, database systems play a vital role in data management. Database systems are efficient in maintaining records and presenting data to customers to support any purposes. Both federated database system and distributed database system are great systems in computer field, but have some relationships and differences. As Amit (1990) discussed in his paper, federated database system is a kind of distributed data system. With relationships clearly showed in Figure: 2-1 below and the detail description of Figure 2-1 are in 2.5.1 and 2.5.2.
2.5.1. Distributed database system (DDBS)

Generally speaking, a DDBS allows its components such as databases or users to share or manage data and enforce some operations in certain ways. In this research, Ozsu’s definition on DDBS was adopted as a collection of multiple, logically interrelated databases distributed over a computer network (Ozsu and Valduriez 1999).

As shown in Figure 2-1, homogeneous and heterogeneous are two unique characteristics of a DDBS. Homogeneous means each of the distributed databases has the same data models or software while heterogeneous refers to the distributed databases employing their own data model or software.

Heterogeneity could be resolved by integration via systems or via gateways. Full DBMS functionality and partial DBMS functionality compose the method via systems. Federated database system is a kind of partial DBMS.

The prevailing utility of DDBS gives its advantages as follows (H.A.K.N. Priyadarshana 2006):

- Reflects organizational structure: database fragments are located in the departments they relate to, with high autonomy.
- Local autonomy: a department can control the data, including updating, processing and storage.
- Improved availability: a fault in one database system will only affect one fragment, instead of the entire database.
- Improved performance: data is located near the site of greatest demand, and the database systems themselves are parallelized, allowing load on the databases to be balanced among servers. (A high load on one module of the database won’t affect other modules of the database in a distributed database.)
- Economics: it costs less to create a network of smaller computers with the power of a single large computer.
- Modularity: systems can be modified, added and removed from the distributed database without affecting other modules (systems).

The growth of the internet has increasingly affected the availability and use of databases on the internet. Besides the DDBS, federated database systems are becoming widely used these days.

2.5.2. Federated database system (FDBS)

As shown in the lower part of Figure 2-1, two forms of federated system architecture are in existence. Loosely coupled systems as component database users construct their own federated schema while tightly coupled systems take independent processes to conduct and publicize an integrated federated schema.

With the integration of multiple database systems becoming the hot topic in research area and commercial database companies, federated database systems attract more attention.
A federated database system is a kind of system which is an integration of multiple autonomous database systems on meta-database management; meanwhile, local applications are supported of accessing from global applications in it while fulfill its local accesses. However, the application fails sometimes due to heterogeneity (Ismail 2001).

Heterogeneity issues as one of the main character of FDBS described by Bishr are syntactic, schematic, and semantic due to different components at the database level, including data model, query language and so on (Bishr 1997). Syntactic heterogeneity arises for various system platform including hardware and software or for transaction management. Schematic refers to diversity in data model or schemas. A major research issue is semantic heterogeneity is mainly caused by problem definition and resolution.

Besides heterogeneity, distribution and autonomy are also features of the FDBS appropriately. Data distribution in FDBS is based on bottom-up design process owing to preexistence of multiple heterogeneous database systems to FDBS making data distribution efficient. DBMS are often as autonomous in federated systems, managing databases and making data sharing between different component databases.

In a FDBS, no centralized control on account of each component databases control access to their data and local operations are submitted to the local component databases. However, there is a global conceptual schema or federated schema generated from all schemas. Global operations involve data access using the FDBMS (Ismail 2001).

2.6. Federated Data Model (FDM)

Federated data model acting as integrating layers on top of the existing cadastral database systems distributed at the different locations. If such FDM is implemented in a single synchronous, consistent federated database system, the users would be relieved from the complex tasks of integration and yet get a consistent dataset for their uses at lowest possible cost (Tuladhar, Radwan et al. 2005). Each database could satisfy its specific users while servicing its global users who needed integrated data generated from each databases.

The basic elements of the federated architecture are components, which represent individual information systems that wish to share and exchange information in the lower level (Dennis and Dennis 1985). Mediation level usually as the middle level takes the role solving interoperability or heterogeneity while the upper level is the output level with integrated data.

Federated Geospatial Data Model which has been implemented by the District of Columbia government is about three levels in that data model. The outer rim as the first level as several components covering numerous source agencies which will share and exchange their data as shown in Figure 2-2. As participants in the federation, each of them takes responsibility for particular datasets. The center is the GIS steering committee and GIS staff who host common services to global users. Mediation is connector of the rim and the center. Heterogeneity problems are solved according to the standards or procedures in mediation part.
2.7. Conflicts and their solving approaches

Integrating datasets is becoming the many information system application. However, the main problem is the heterogeneity between data sources which were created and developed separately. The three main types of conflicts are syntactic conflicts, schematic conflicts and semantic conflicts; semantic conflict which is the main component of this research work will be described in detail.

2.7.1. Semantic conflicts

Schema integration is the process of generating one homogeneous database schema from several, heterogeneous source schemas (Ingolf Geist 2001). Detecting semantic conflicts and then solving these conflicts by designing new schemas and corresponding schema mappings. After having solved all conflicts, the schemas can be merged into an integrated schema. Extensional and intentional conflicts are two major types of semantic conflicts. Extensional conflicts refer to potential object redundancies among different cases whereas intentional conflicts are caused by different object types of semantically related classes.

Semantic conflict is the most popular one since many more types of conflicts have the relationship with it. Therefore, the common subtypes of semantic conflicts’ are (Ismail 2001):

**Structure conflict** which indicates an attribute in one schema is a class in the other schema.
**Domain conflict** happens when many different attribute value representations. One attribute can encompass more information than others. For example, mapping functions are required to bridge different attribute values semantically.
**Meta conflict** means in different schema that all objects have implicitly the same met value.
**Intentional conflicts**: equivalent classes of integrated schema and a local schema can have different sets of attributes.
**Naming conflicts**: database adopts different names to present the same concepts.
Missing attributes: some attributes missed, e.g. the gender is included in one system but not in another related one.

2.7.2. Semantic conflicts managements

With the cooperation between different organizations, integration of data is gaining more attention. However, conflicts are caused by heterogeneity between data sources which were set up and developed differently. The various methods for solving semantic conflicts are available but the most promising ones are introduced as follows:

2.7.2.1. Classifying heterogeneity with view definition

Accessing heterogeneous files and databases in navigational database systems or relational database systems, through a unified data definition created great difficulties during the past 20 years. Unless the structural or representational discrepancies named conflicts are solved, users could not access such a multi-database system smoothly. Won Kim (1991) developed a complete framework for enumerating and classifying the kinds of structural and representational conflicts.

Two basic classifications of conflicts have been made. As a database is defined by its schema and data, Kim (1991) classifies conflicts at these two levels.

A. Schema conflicts:
   1. Table vs. table conflicts
      a. One to one table conflicts
         1) Table name conflicts
         2) Table structure conflicts
         3) Table constraint conflicts
      b. Many-to-many table conflicts
   2. Attribute vs. attribute conflicts
      a. One to one attribute conflicts
         1) Attribute name conflicts
         2) Default value conflicts
         3) Attribute constraint conflicts
      b. Many-to-many attribute conflicts
   3. Table vs. attribute conflicts

B. Data conflicts:
   1. Wrong data
      a. Incorrect-entry data
      b. Obsolete data
   2. Different representations for the same data
      a. Different expressions
      b. Different units
      c. Different precisions

By addressing schematic heterogeneity, the writer defines views on the schemas of more than one component database and to formulate queries against the views. View definitions could specify how to homogenize the schematic heterogeneity (Won Kim 1991).
2.7.2.2. Model-based methods

One of the most popular ways of data integration is modeling the entities in real world and their relationships. Model-based methods intend to define procedures and heuristic algorithms for schema integration by integrating database views or database integration. The differences between these two kinds of integration is that database view integration means producing a global view in one database and database integration creates a global view from distributed databases (Lawrence 1999).

Model integration approaches are referred to relational/ functional models and semantic models. In relational models, integrators made the universal relational assumption that every attribute had a unique name in a database and this allows integrators to ignore naming conflicts. Moreover, some algorithms could then be developed using the relational model’s set properties. In schematic models, more conflicts will be solved and did not assume certain naming features. Taking Semantic Data Model (SDM) as an example, entities in real world are organized and modelled into classes, given a semantic meaning with the definitions of the entity. In this way, attributes are defined on both classes and entities (Lawrence March 22, 1999). Classes’ relationships are basically referred to super-class/sub-class relationship and grouping values which based on the other classes’ values or entity presence. Attributes in various classes can be defined attribute interrelationships or using mappings from different classes. Model-based methods are gaining popularity these days.

2.7.2.3. Schema mapping

For conflicts solving, information systems and program comprehension tools use mappings to translate data from one schema to another. Schema mapping is a kind of conflicts solving by integrated schemas into a canonical model or being mapped into the same model. Sets of inter-schema correspondences are translated into a set of mapping that capture the design features made in the two schemas. The design features include the hierarchical organization of the data and schema constraints (i.e., foreign key constraints). These mappings are then translated into queries (SQL, or XSLT) over the source schema that generates data to make up of the target schema. The schemas perform transformations until the schemas are identical or similar by the common concepts. Two or more data models produces a matching between elements of the two schemas that correspond semantically to each other (Giunchigilia 2003). A significant characteristic of the mapping algorithm is that it takes into consideration target schema constraints in order to guarantee that the generated data will not violate the integrity of the target schema (Mylopoulos 2005). Presently, many tools used for schema mapping such as, Clio, ToMAS and MapForce.

2.7.2.4. Ontology approaches

Ontology models the vocabulary and meaning of domains of interest: the objects (things) in domains, the relationships among those things; the properties, and constraints on and rules about those things (Lemmens, de Vries et al. 2003). Ontologies enable designers to specify the semantics of their domain or across communities in great and arbitrarily greater detail.

In World Wide Web, ontology can be used to publish and share data in Web Ontology Language (OWL). OWL facilitates machine interpretability of Web content for it is designed to process the content of information but not just presenting information. As OWL is written in XML, OWL
information can be changed between different kinds of computers using various application languages and operating systems. Moreover, OWL is able to model complex relationships between different classes of kinds of data models. In Dutch National Mapping Agencies, a project about schematic integration used ontology to integrate data at different scales gain successful (Stoter, Lemmens et al. 2006).

2.7.3. Comparison of conflicts managements

New view definitions should created to specify heterogeneities in the method of Classifying heterogeneity with view definitions. This requires much more insight understanding of views. A model-based method is a way of data integration by modelling the entities in real world and their relationships. Careful study should take on the real world study. Schema mapping is based on the content of attribute values and this avoids the construction of ontologies for a particular knowledge domain. FDM is a combination of model-based method and schema-mapping as entities and their relationships will be studied and modelled in FDM and the mediation part of FDM could be fulfilled by schema mapping.

2.8. Relevant Tools in research

In this research, the Unified Modeling Language (UML) would be used to develop a conceptual model. Schemas will be set up by the help from XMLSpy and then Mapforce as an integration tool which are discussed as follows.

2.8.1. UML

The UML is a standard language administered by Object Management Group (OMG). It is a graphical language for visualizing, specifying, constructing, and documenting the artifacts of a software-intensive system, as well as a process and tool independent language that presents model elements, diagrams and rules associated with them (OMG 2003). It is used in facilitating the ability to exchange data models between software development systems. UML offers a standard way to write a system's blueprints, including conceptual things such as business processes and system functions as well as concrete things such as programming language statements, database schemas, and reusable software components.

The Class diagram which will be used is by far the most used and best known of the object-oriented diagrams. The Class diagram represents classes, their component parts, and the way in which classes of objects are related to one another. It is the source for generating code and the target for reverse engineering code (Pender 2002).

2.8.2. XMLSpy

XMLSpy is a industry-standard XML editor and XML development environment for modeling, editing, debugging and transformation all XML technologies including XML Schema, XSLT, XQuery, SOAP, WSDL, Web service and so on.
In this research, XMLSpy will be used to:
- Create a simple schema from scratch
- Generalize the schema using simple and complex types
- Create schema documentation
- Create an XML documentation based on the schema file
- Validate the XML document against the XML schema
- Transform the XML document
- Import and export database data to and from XMLSpy
- Create a schema from an MS Access database

2.8.3. MapForce

MapForce 2006 Professional Edition is a visual data mapping tool for advanced data integration projects. Thus, MapForce can generate custom mapping code in XSLT 1.0 and 2.0, XQuery, Java, C#, and C++, and supports the listing area used in this research (GmbH 2006):

- Schema-to-Schema mapping
- Database-to-Schema/XML mapping
- XML-Schema-to-Database mapping
- Database-to-Database mapping
- Definition of user-defined functions/components, having complex in/outputs
- Advanced search and replace functions in transformation preview data

2.8.4. Dreamweaver

Dreamweaver is a kind of software that could build dynamic, database-driven web applications using server technologies such as CFML, ASP.NET, ASP, JSP, and PHP (Varese 2005). In this research, the tools provide in Dreamweaver are used to create XSLT pages, attach XML files, and display XML data.

2.9. Concluding remarks

This chapter discussed the relevant concepts in developing FDM. Real estate is mainly bout land and housing which supplies the living environment and investment market. A federated database system is a kind of system which is an integration of multiple autonomous database systems on meta-database management. The components of SDI are policy framework, technology framework, data acquisition and display human resources. The components work together to ensure that spatial data is discovered, evaluated and used. Land and housing data are two kinds of framework data which have spatial attributes. After introduce the SDI perspectives, the FDM and approaches how to solve data integration conflicts, the idea of making integrated data on land and housing data through FDM is possible.
3. Analysis of As_Is Situations in China

3.1. Introduction

This chapter gives the as-is situation of services on land and housing in China and the related data need. For effective integrated services put in place, the situation analysis on land and housing, especially the four main components as mentioned previously in chapter 2 should be carefully studied. During the fieldwork in China, these kinds of data were collected through structured and unstructured interviews, observations, document collection, questionnaires and gathering of documents.

The detail activities including location of fieldwork, interviews, document collection and questionnaires are shown in appendix 2. On the analysis of fieldwork findings, section 3.2 gives results on the institutional framework of Chinese land and housing management, technologies, data acquisition and display and on human resources.

3.2. Fieldwork results

Data in this section were all collected during the fieldwork in China. Cadastral data as a fundamental part of framework data in SDI attracts more attention from the government, as NSDI is being built in China. In this section the main components such as institutional framework, technology related to land and housing in NSDI are presented.

3.2.1. Laws governing on land and housing

Two fundamental laws concerning land and housing were passed by the State Bureau; the Land Administration Law of the P.R.C (LAL) and the Urban Real Estate Administration Law of the P.R.C (UREAL).

Under the China’s LAL, which was firstly drafted in 1986 and amended in 1998, the state owns all urban land, while farmers collective own all rural lands. Before 1986 when the land tenure system was introduced to China, all land use was treated as allocation. The user could extend land use right by paying land use tax annually or transfer the land use right into “granted” by paying the land grant premium (Roy Prosterman 1998). The land owner but the land-use rights may be separated, and the state remains the land ownership and local government may transfer the land use rights by law on behalf of the state (Council 1986).

Urban Real Estate Administration Law of P.R.C was later passed in July 1994. The law allows housing and land to be two separate entities and are managed by two different bureaus. Housings could be transferred in the real estate market and the owner could have its ownership for certain years according to the housing type. For example, owner could enjoy his owner rights of the housing for living in 70 years.
3.2.2. **Institutional framework**

In P.R.C., land and housing are mainly managed at four levels: national level, provincial level, municipal level and county level. Each level of management is discussed as follows:

3.2.2.1. **Institutional framework on land management**

As presented in Figure 3-1, land is under the management of Ministry of land resources (MLR) at the national level. MLR is in the highest level in land management. Its main roles which have relationship with land are four items derived from more than 30 specific roles in land administration (MLR 2000) and as described below.

- Compile and implementation of the national comprehensive planning for land and resources, overall plan for land use and other specific plans;
• Enactment of relevant laws and regulations and promulgation of rules governing the management of land;
• Formulation of cadastral regulations, organization of land and resources survey, cadastral survey, land statistics and dynamic monitoring;
• Administration of land titles, land grading and registration;

Provincial departments of land resources are taking charge of municipal and country level of land resources management, and are under the supervision of MLR. The lower three levels have almost the same section as MLR on land management. The municipal land administrative for example, has section for technology, land use planning and cadastral.

3.2.2.2. Institutional framework for housing management

Housing is under the management of Ministry of construction (MC) which is directly under the vice Minister of MLR. The MC has three main roles in housing management derived from specific role (MC 2006).

• Enactment of relevant laws and regulations governing the management of various industries related to housings;
• To make the whole construction market standardize and under guidance;
• To make the future plan, short-time plan, reformative plan and technique policy for construction enterprise.

As clearly shown in the Figure 3-2, housing management constitutes four levels, provincial level, municipal level and county level in housing management. At each level, the same sections like housing, construction are composed of the housing administration. The Figure 3-2 again shows the top-down pattern administrative levels with each level playing its role in housing management.

3.2.2.3. Roles and data production of Survey and Mapping Bureau

As shown in Figure 3-1, the State Bureau of Survey and Mapping is at the same level as the departments of ministries of land and resources which is also directly under MLR. However, at the provincial level, the Survey and Mapping Bureaus’ (SMB) structures are quite different. Detailed structural information is as follows (SMB 2005):

From Figure 3-3, it is obvious that the institutional framework of SMB is really complex in provincial level administration. China has 31 provinces and municipalities in main land with various local features and uneven development in technology and economy. SMB roles related to land and housing mainly covering the follows:

• To enact relevant regulations and promulgate the rules governing the management of surveying and mapping in its own administrative levels;
• To organize or cooperate in cadastral surveying and mapping;
• To organize or cooperate in housing surveying and mapping;
To examine the competency and authenticate surveying companies or organizations of surveying;

Figure 3-3: Institutional frameworks of Survey and Mapping Bureaus in national and provincial level

From the first two Figures, it is observed that the bureaucratic structures of MLR and MC are almost the same. Each of them has a complete and different perspective on management of their own operating target-land or housing. The national level managements cooperated with the country, municipal, and provincial levels managements. Each level has its corresponding management section with the upper level. For diversity in local places, different provinces acquire their own feature administrative on surveying and mapping. The LAB or HAB gets cooperation from SMB in surveying and mapping sometimes, however, they also like to get help from some private organizations which are authenticated by SMB.

4D productions which are employed in different applications are the main productions in SMB. 4D productions are Digital Elevation Model (DEM), Digital Ortho-photo Map (DOM), Digital Line
Graphic (DLM) and Digital Raster Graphic (DRM). DEM can be used to analyze elevation and accuracy, slope and exposure; measure coordinates, distance, areas, and volumes; make perspective drawings and section maps; perform physiognomy analysis, and overlay related vector data and image data to make various thematic maps. DOM is an image which integrates both the map geometric precision and image characteristics characterized with high precision, and rich information. DLM is a vector dataset of fundamental geographic features on existing topographic maps. DRM can be used in related fields as fundamental data to carry out measuring, analyzing and digital mapping activities. It can also be used for integrating with digital ortho-images, and digital elevation models to obtain new visual information (SXSM 2006). These productions have been applied in cadastral and housing agencies with the above mentioned functions.

The detailed datasets which has a close relationship with land and housing produced by SMB in both provincial and municipal level SMB are shown in Figure 3-4. Datasets related to cadastral and housing will be discussed in section 4.2 and 4.5.1. It is clear from Figure 3-4 that ortho-images, topographic data and spatial framework are all produced in SMB. Data about cadastral parcel like boundaries and parcel area are also produced by SMB.

![Figure 3-4: Data produced by SMB](image)

### 3.2.3. Technology used

With the rapid development of technology in recent 20 years, a great change took place in land and housing management. Especially in an area of Geo-ICT, data acquisition, data processing and data storage are much more efficient. This section examines the different technological aspects for achieving data in China.
3.2.3.1. Approaches in collecting data

Obtaining data is the initial step in data management. Some data should be transformed to be accepted by computers. Generally, data is divided into spatial data and attribute data. Approaches for spatial data are mainly through field survey, remote sensing, fieldwork and documentation while approaches for attribute data also encompass questionnaires and statistics.

Technology for data acquisition is much like in the whole country. GPS, RS, GIS cooperated to evaluate the resources and forecast hazards. Laser Range Finder and Whole Stop are two tools widely used in survey. Communication technology is developing fast together with the most prominent one-Internet. However, the land survey part is mostly taken by local SMB bureaus or local surveying organizations, so local regulations or software are not the same.

3.2.3.2. Means for supplying services

In supplying the data on land and housing, both conventional and modern technologies are used. Conventionally that data are mainly paper-stored or hand-writing while modern means for services are computer, internet and mass-media. This previously happens sometimes in country level while computer-based data provisions are quite normal in many places. Stakeholders could access internet to get some of the data they want or they can go to the bureau directly to get the exact type of data this bureau provided.

3.2.3.3. Brief introduction about systems concerning cadastral and housing management

The land tenure system was introduced into China after 1980s, since then pioneers in computer and land management field started the research of computer technologies implemented in land administration. In the 1990s, enterprises on software designing play an important role in management system designing. Cadastral database, land tenure system, land transaction systems are widely used in every province for land management (Wang 2006).

Cadastral information system (CIS) is a type of system which realizes the cadastral data input, store, searches, processing and analysis. CIS is a great supportive for land administration as it assists the LABs with huge cadastral data management for various usages. However, the design and usage of CIS in China started around 1990s. On account of a big country with various features in each local place, MLR allow local government to develop their own CIS according to local features and finance status. Thus, CIS in each city are almost differently designed.

CIS design is normally based on Arc/Info and Arc/View. The database platform is oracle and GIS platform is micro-station / geo-graphics. Shanghai, Shenzhen and Nanjing led the field in CIS design and setting up. The main problems of CIS built are uneven development, no unique technical standard and low designing level. Besides, duplication design is also a prevailed problem.

When referring to housing management system, there are not as many fine and complex computer management systems as land management systems. Microsoft Access or Microsoft Excel systems are often used in the housing administration. Besides, some of housing maps are stored on paper while in
big cities they are stored in AutoCAD. An example is Guangzhou HAB, located in Guangzhou, a fast developing and dominate city in northwest of P.R.C as an example. In the 1990s, each section of this bureau developed their own management systems in accordance with their business rules for example the housing transition management system, file of housing management system and housing pre-sale management system. Systems are fragmented in different places that are difficult for cooperation, even in one bureau there is duplication and different data standards. Most of the systems are based on Wintel PC and SQL Server database infrastructure and adopted C/S schema (Zhu 2006).

3.2.4. Data capture

A consistent definition of the position of different geospatial datasets is a necessary condition for their geospatial integration (Groot 2000). Different datasets are spatially referenced by different techniques, so they have different map projections. The quality of the decisions based on data integration depends on applications that bring these different datasets together. Geodetic coordinate system and projections are important in data acquisition.

3.2.4.1. Geodetic Coordinate System (GCS)

Beijing GCS 1954 (P-54) and Xi’an GCS 1980 (NGS-80) are two main coordinate systems in cadastral survey and mapping in the nation wide. P-54 is the first GCS referenced from Krassovsky spheroid in former Soviet Russia since the P.R.C set up and NGS-80 is set up according to the recommendation of IAG 75 from IAG. NGS-80 is built based on the average error in national wide and the geodetic origin located in Jingyang, Shaanxi province. The two GCS are playing their roles today with the NGS-80 more widely used. Meanwhile, transformation between the two systems is possible. Since the GPS satellite navigation is employed, observation value as the localization survey achievement is based on the WGS-84 world earth coordinate system. It is possible to get the transformation data from WGS-84 to NGS-80.

3.2.4.2. Projections

Map projection is used to represent round earth on a flat surface. Besides using projections of mapping and other special requirements, the mapmaker must select a map projection also best suited to the cartography scope, shape and location reduced distortion of the most important features. In P.R.C, if the basic scale of topographic maps is over 1:500,000, Gauss-Krueger projection also named Transverse Mercator projection will be adopted.

The Gauss-Krueger projection has the following properties:

- The projection is conformal. Due to a break in the series development, however, this property does not hold exactly.
- Only small stripes (1.5°) to the west and east of the central meridian (with ½° overlap) are used for a single projection.
- The central meridian is mapped in true length and serves as the ordinate axis of the strip system (Ressl 2002).

Lambert Conformal Conic projection is used for cadastral map and topographic maps with scales under 1:500,000. A Lambert Conformal Conic projection superimposes a cone over the sphere of the
earth with two reference parallels secant to the globe and intersecting it. This minimizes distortion from projecting a three dimensional surface to a two-dimensional surface. Distortion is least along the standard parallels, and increases further from the chosen parallels. As the name indicates, maps using this projection are conformal (Wikipedia 2006).

In cities and towns, cadastral map scales are basically 1:500, 1:1000 and sometimes are 1:200 (MLR 2006). Remotely-sensed maps and word documentations which explanations the maps are stored in paper and land management systems. The results of housing survey are mapped into housing floor map and housing apartment map. These two kinds of maps are key data in housing administration.

### 3.2.4.3. Data capture methods

A variety of data capture methodologies have been used to build cadastral and housing databases. Based on the analysis of question 3.9 in questionnaires, five main methods are generalized. The choice of the four methods to capture data depends on influential factors various from the target data accuracy requirements, funding and technical resources.

- **Measurement:** measurement is a survey method which uses professional equipment, such as whole-stop. This method is often used when some new spatial data is acquired through accurate surveying. Such maps are normally used for base.
- **Construction using coordinate geometry:** Data collected by this method could be very close fit with digital ortho-imagery.
- **Construction and best fit to base map:** The data produced from this survey is less accurate than the two methods mentioned above. However, the accuracy could be maintained and documented. It is suitable for most assessment uses as well.
- **Digitizing and best fit to base map:** This method is popular for it requires lower level of technical knowledge and least costly.
- **Documentation:** Many attribute data are gained from documentation used in former management.

### 3.2.5. Human resources

The demand for skilled personnel in land and housing management are familiar with their operation. China has a large population and with great development on professional and high education, professionals in land and housing field have increased not only in quantity but also in quality. High requirements of the staffs in this field, including well-master professional knowledge, inter-subject knowledge, and computer science. Every year, there is a national-wide official examination for the selection of government officers. 407 persons were taking part in an exam last year for one job place. Competition also happens in MC and SMB. The requirements for human resources are not only in the professional aspect but also personality are very strictly, as showed in Figure 3-5. This information was obtained from the questionnaires in fieldwork, with 50 leases out and 32 returned.
3.3. **Justification for FDM be used in China**

In the era of new public management, Geo-ICT is a crucial technology in improving management. Moreover, federation concept is a kind of new concept for accessing distributed databases. FDM as an efficient model in facilitating e-government could be used in P.R.C. In this section, justification for FDM in China is expatiated from eight aspects various from administrative point of view to reality.

3.3.1. **Stability of political institutional framework**

In order to get a sound management on land and housing, the Chinese government made a great effort in institutional framework. In the beginning years of P.R.C, there was a National Land Management Bureau before housing management became an important part.

Before 1998 Land and housing were managed in two different bureaus named LAB and HAB. Since then, as the great economic development, more services began to cover both land and housing data. In order to make efficient governance, the Chinese government decided to merge LAB and HAB into State Resources and Housing Administrative Bureau (SRHAB) in most of the fast developing cities, like Beijing, Shanghai.

For some reasons, most of these bureaus are departed around 2002 with only four SRHAB remained. Among these, SRHABs are located in Economic Zone besides the three municipalities. Even the bureaus are combined, the problems of getting data on land and housing is still difficult. The reason is that most of the data in SRHAB are still stored in former format. It means that although they combined their services together, the data are still as before. Therefore, whether the bureaus combined or not, they need FDM to be an integrating layer on the top of the existing cadastral database system and housing database system no matter where they are distributed. Thus, the government could sustain the present political institutional framework.

3.3.2. **Possible infrastructure and governmental work emphases**

Intensifying information systems on state resources management and the realization of one-stop information service to society is one of the five main goals made at the beginning when MLR was founded and also as law for land administration (Council 1986). The “digital state resources” project started since 1999. 1000 counties had built their own local land management systems tailored to local features (Dai 2006). However, these systems as isolated systems and cannot work consistently. Mainly, there are about two reasons for this awkward situation. One is local regulations and administrative method are different and another one is the design of databases are various as for
different designers. Furthermore, housing management systems in local places are also having these problems.

As the Chinese government’s emphases on building national basic database system for land and housing, FDM becomes very important to Chinese situation. Each local system could be autonomous system at the local level which satisfies their own users and could also satisfy their global users through FDM. Furthermore, according to the investigation about status of current database systems in national wide, there are only two kinds of management systems; they distributed data system and federated data system (Dai 2006). Thus, the current situation of systems makes fulfilment of FDM possible.

3.3.3. Stimulation of economy

The investment on land and housing is increasing dramatically these years, which could be seen clearly from the following numbers in the real estate blue book. In 2003, more than 1 trillion RMB was invested on land and housing, that is 29.7% more investment than in 2002. The investment accounts for 9% of GDP in 2002 which indicates that land and housing is an important support in the national economic (Niu 2004). Compared to 2003, in 2004, the investment on land and housing was 1315.825 billion, 2.2% less than 2003. But the increasing ratio is still higher than GDP increasing (Niu 2005). When coming to the year of 2005, 3405 billion RMB were invested in land and housing during the first 4 months, 25.9% higher than the same period of last year. It is forecasted that the number will be increasing in the next 2 years (Niu 2006). The real estate developer and the common customer need more data on land and housing to help their decision making. At present, as the data are supplied by two different bureaus in different format, the common customers have some difficulty in getting both datasets in time and also combing these two datasets becomes a tedious work for them. FDM will help the customers to obtain the combined land and housing data they want. If there is no difficulty in getting data, more people would like to have transactions on land and housing and therefore improve economy.

3.3.4. Effective and flexible service to present and potential users

As a large country with 31 provinces and 1.3 billion citizens, Chinese government’s intend to supply one-stop service through e-government is quite rough. There are about 3 classes of land classification. In the first class, 8 main kinds of land are defined and 46 kinds are outlined in the second class(Bi 2001). Different parts featured with different kinds of land. Housing as the engine in economic development is attracting more attention to all citizens. Facing with such a giant group of users and more potential users who are willing to participate in land and housing business, the government needs to provide effective and flexible services to the users with different land and housing data requirements. With FDM, users request can easily been met. Data duplications between various organizations could be avoided. FDM also saves money and time for both of users and organizations.

3.3.5. Foster citizen’s trust on government

As real estate market is quite hot these years, some players make use of their ability to get some data to earn money while others cannot get comprehensive data. These kinds of things have a strong
influence on the emotional feelings of common citizens and some communities about the role the government plays. To them even they could have some data, it is still hard for them to do some analysis and then make decision. What the common citizens want is getting some data meet their requirements and being both efficient and effective. With the establishment, and NSDI and better data supplement with FDM, citizens could trust on government that they are making efforts to give a perfect public service.

3.3.6. Helpful to real estate brokers

The first real estate agency was set up in P.R.C in Dec. 1988. With the growing development of real estate industry these years, being a broker is a hot job with millions of people involved. But there are also many problems in this field. Fees are charged without a standard regulation but according to some regulations made by their own agencies. Many of the brokers are not professionals. And therefore require professionals with rich knowledge on real estate services and laws are needed urgently. Hardware and software for brokers need to be improved. Many agencies do not have modern communication technology and some data are processed by hand. The management of brokers should be intensified (Dong 2002). If FDM is employed, it will be a great help to the brokers to get the exact data they want that do not need them to process. Moreover, this could also help the brokers more professional on data they work with.

3.3.7. Accelerate the NSDI

In the implemention of digital State Land project these years, each state resources bureaus and HABs in the local level enjoy its own internet web site to make e-government services. In 2000, authored by the State council, 11 ministries including MLR and MC are coordinator of the NSDI to foster inter-communication and cooperation between different ministries (MLR 2002). Service scheme of Geospatial basic data was made as illustrated below. Cadastral data as a key framework data plays an important role in the build of NSDI. In the process of NSDI establishment and NSDI application, heterogeneity issues must be a problem that must be solved. FDM is an acceleration of NSDI built for it is a solution for heterogeneity problems solving.

3.3.8. Better services related to cadastral and housing data

As land and housing are inseparabe, services concerned with land and housing take a large proportion in services supplied by LAB or HAB. After introducing the analysis of services provided by Xiamen Municipal State Resources &Housing Administrative Bureau (XMSRHAB), Beijing Municipal State Resources Administrative Bureau （BMSRAB）and Beijing Housing administrative committee（BHAC）, the main types and specific services related to cadastral and housing data are discussed as follows.

3.3.8.1. Present services

Services supplied at present related to land and housing are grouped into 12 main types and 111 specific types as shown in Table 3-1.
From the kinds of services listed, it is clear that they almost cover all service areas related to land and housings by LAB and HAB. The question is if these data could satisfy the users and their requirements on data provision.

Table 3-1: Types of services related to land and housing

<table>
<thead>
<tr>
<th>No</th>
<th>Main Types of Services</th>
<th>Amount of specific services</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Housing reform and fund management</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>Tax</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Property management</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Land assignment and land price management</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>State-owned housing administration</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>Public fund for housing</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>Mortgage</td>
<td>13</td>
</tr>
<tr>
<td>8</td>
<td>Migrate for requisition</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>Property right</td>
<td>24</td>
</tr>
<tr>
<td>10</td>
<td>Survey (related to land and housing)</td>
<td>26</td>
</tr>
<tr>
<td>11</td>
<td>Transfer</td>
<td>13</td>
</tr>
<tr>
<td>12</td>
<td>Unsafe housing reconstruction</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Total: 111 kinds of services</td>
<td></td>
</tr>
</tbody>
</table>

3.3.8.2. Stakeholder types and shortcomings of present data provision

From the questionnaires and interviews accomplished in the fieldwork, three main types of land and housing data user classifications and services they need could be summarized as follows:

**Government agency:** For the housing and implementation of NSDI and achieving the goal of e-governance, government agencies face the phenomenal challenge of providing accurate, up-to-date and comprehensive geo-information not only to the common users but also government agencies themselves. The government takes macroscopical management all over the country, so they need comprehensive analysis on land conditions and situations as the context for decision-making all over the country.

In Planning and Reforming Bureau for an example, cadastral data, and housing data are needed every year to make a statistic of how many houses are transferred or mortgaged and what are the situations of their attachment-land. After they get the data from the two bureaus, they need to make comparisons by themselves to find what kinds of data are the same and then just keep one copy; they find the different datasets as well to analyse the relationship between land and housing. In such kind of scenario, duplications happen usually and therefore it is also a waste of human resources and time.
During the fieldwork, one real case shows that a borrower mortgages his housing by giving a lender his housing ownership certificate and land use certificate. When the borrower cannot pay back money to the lender, the lender find that the certificates are illegal because the borrower lied to HAB and LAB he lost that certificates and then make a new one. The problem here is that the lender cannot get data about whether the mortgages are legal from both bureaus. Besides, both bureaus issue new certificates without check if the other certificate is in normal situation just for communication delay. Though, this is a special case and hardly happens, it indicates that bureaus should share some data or easy accessing for some data.

Organizations: Organizations are mainly referred to as enterprises set up by citizens. The most proper representation of these kinds of organizations is real estate companies who buy use rights of land from LAB and build house on it under the permission of Urban Planning Bureau. The real estate companies then sell the houses after transaction registration. At last, the real estate company obtain the housing certificate in HAB and land certificate in LAB. In this processing, the real estate companies also need to penetrate the land market to study which land they should buy and where housings are in high demand. The detailed data about the land and housing are stored separately in LAB and HAB. It costs about 1-2 months to get the data from the bureaus normally.

From the interviews in the fieldwork, the real estate company said that they have to collect or buy data from two bureaus by their employees whose jobs are mainly focused on getting data. However, some of the data they bought were the same data type but the value cannot match. Another problem for them is duration for applying for use rights, approval of permits and getting certificate. The reason for these is that there is seldom any operation communication among the bureaus.

Citizens: Citizens who have no background knowledge of land and housing data are the largest client group for both LAB and HAB. They need data for purchases or investment but most of them cannot afford the fee paid to consulting agencies. From the questionnaires, they need easy data which do not need professional background and the data can be accessible through internet without difficulties.

Based on the kinds of user groups and their needs and the shortcomings mentioned above, if FDM is applied, two or more datasets could satisfy their common users though various export schemas to reach the goal of saving time, money and efforts and improve the data delivery.

Potential users

As showed in Figure 1-1, there are various data on land and housing that influence decision making, such as the legal consultative office, environmental office and transportation office. If FDM can be employed in other area, it will welcome more users in many fields from normal citizens to governmental officers.
3.4. Concluding remarks

Through the discussion of present situations about land and housing data above, it is clear that these two kinds of data are separate managed by MLR and MC. Data collection are taken by three main bureaus, LAB and HAB produce data for themselves and Survey and Mapping Bureaus also provides the data for LAB and HAB. The main requirements of three groups of users, integrated data are in great need and if they are implemented, it will enhance e-governance. On the analysis of the As-Is situation of the land and housing in P.R.C, eight aspects on justifications for FDM be used in P.R.C are discussed. However, the most challenging issues are the conflicts. FDM will be an efficient data model tool for solving them. An analysis of the detailed land and housing data would be introduced in next chapter.
4. Data Analysis on Cadastral Data and Housing Data

4.1. Introduction

Land information system (LIS), Cadastral information system (CIS) and Housing information systems (HIS) are based on spatial database systems with the help from software and hardware to collect, manage, operate, analysis, simulate and display land or housing related data. The systems provide various data to broad stakeholders. Therefore, the kind of data in these systems and the relationships are important for the study of these systems.

This chapter is made up of four main components. Data in LIS will be introduced in general in section 4.2 as CIS is an application of LIS. Then criteria for data analysis will be discussed in section 4.3 before the detail analysis on cadastral data and housing data are introduced in section 4.4 and 4.5. Similar data types between these two kinds of data are analyzed in section 4.6 for the preparation to generalize some conflicts between cadastral and housing database in section 4.7. Based on the analysis on cadastral data and housing data, the conflicts will be solved by mediation in federated data model in Chapter 5.

4.2. Brief introduction on land data in Land Information System

LIS is based on land spatial database and it could manage, operate, analyse, simulate and display land data with support from software and hardware. In order to make an efficient land data management, MLR in P.R.C makes various land data classifications for its different usages and make several systems to fulfil one particular objective. The systems vary from Land Price System (LPS), Land Use Information System (LUIS), Land Tenure System, to Land Registration System and CIS. In information management systems, generally, land data are divided into four types for systems’ development and maintenance. Basic geographic data, graph data, thematic attribute data and other attributed data related to land are four basic types of land data in LIS (Liu 2003). All of the land management systems for example, land price management system or land planning management systems are all based on LIS. Therefore, a clear brief understanding of data in LIS is necessary to comprehend the overall situation of land administrative in China. Four types of basic data are listed below with explanation of what they are made of.

4.2.1. Basic geographic data related to land

Basic geographic data provide uniform background of geo-spatial reference frame and thematic data for land data management and applications. Normally, basic geographic data include:

- Geodetic Coordinate System: A coordinate system in which the location of a point in space is defined by distances from the point to the reference plane.
- Topography: Topographic data is about land surface features.
- Physiognomy: Physiognomy is the interpretation of outward appearance of the earth.
- Water systems: Kind of water systems exists on land.
Transportation data: This kind of data encompasses a wide range of spatial data entities which are related to land.

4.2.2. Thematic geometry data

When building the land management system, land related data are presented in different layers. It is easy to inquire and update according to these kinds of layers. There are about 7 different layers in land management systems.

- Cadastral data and land use plan;
- Floor plan of the building (data on each floor of a building);
- Land use planning and thematic planning;
- City planning and thematic planning;
- Land use types;
- Urban land grading and benchmarking land price;
- Administrative region: data about administrative boundaries in national and provincial wide have close relationship with land as the region data is a kind of spatial data.

4.2.3. Thematic attribute data

Thematic attribute data is about the data which could basically show the status of a certain piece of land. In most cases, it covers the following:

- Land right and restrictions: Land right indicates what person could do with the land while restrictions shows what cannot be done on the land, the regulations are made by government. Land right registration is of great importance as it not only a record of rights concerned ground material but also underground material like mine.
- Land value and tax: Land value and tax records play a vital role in national economic and development.
- Land use in urban and rural: These are records that give the clear history of land use situation.
- Housing and building: This records the height of the building and floors.

4.2.4. Other data related to land

Opinion these years is that cooperation between different organizations is becoming more and more important which indicates that data related to each other are needed. In land information management system, there are some easy and simple data covering different application fields.

- Census: census map is designed to present census data.
- Climate data: general data about climate.
- Hydrographical data: hydrographical data covers a large range of spatial data entities that related to land.
- Plant: plant refers to the plant on the surface on land.
4.3. Criteria for data analysis

In order to explain the aspects of cadastral data and housing data, criteria for data analysis is worked out. Some of these criteria will be picked and employed to analyze land and housing data. The criteria are mainly made from 3 aspects which indicates detail situation of data.

1. Data source:
   1) Who makes these data: indicate who surveys the data
   2) How to collect these data: the method of data collection.

2. Data characteristic:
   1) What kind of formats: how the data are stored.
   2) What is the Geodetic Coordinate System?
   3) What is the scale of these data: this mainly refers to map as well.
   4) What is the resolution of these data: this mainly refers to map.
   5) What is the error scope?

3. Present situation:
   1) How often does update time circle.
   2) To whom does these data: explain the user of data
   3) What condition to use these data: in what kind of situation does these data used.
   4) Are these data free or not: when the data are used by customers, whether it charged or not.

4.4. Cadastral data

In this part, cadastral data will be analyzed in detail after a brief introduction about the CIS in China.

4.4.1. Brief introduction of CIS

CIS is a crucial system in China for it manages the cadastral datasets which play an important role in land administration. The CIS is also design by different designers according to local situations. However, the basic hardware requirements for the system are host computer and equipment for display, import, export, and data storage. Besides, memory, had disk, tape machine, scanner, graph plotter, CD and monitor are also needed.

4.4.2. Role and formats of cadastral data

Cadastral data is a collection of information about parcel and its location, quality and quantity. It has a core of land property which indicates the owner’s rights, restrictions and responsibilities. Cadastre is established for different purposes and have various records(Dale 2000):

- Fiscal records: This is mainly for public sector as the basic record for the full and accurate taxation of land.
- Legal records: Cadastre plays a dominant role in land transfer and it confirmations ownership and other rights of land.
- Land use records: Cadastre is the primary materials which record the story of land use. It is a great help in land use management.
- Multi-purposes: Cadastral data also enables sustainable development for e.g. environmental protection.
Cadastral data normally can be divided into spatial data and attribute data which are integrated together to accomplish its roles mentioned above. The cadastral data can be classified as shown in Figure 4-1 as follows:

![Figure 4-1: Classifications of cadastral data](image)

Cadastral data mainly contain three data formats: cadastral data files, cadastral maps and cadastral datasheets (Wu 2002).

- **Cadastral data files**: Cadastral data files contain records of where the land and its attachments located and land conditions in the form of descriptions.

- **Cadastral maps**: The maps clearly indicate the location relationship between land and its attachments. These maps contain points, lines, polygons and annotations. Thematic map, parcel map, land use status map and land quality appraise map are all included in it.

- **Cadastral datasheets**: They constitute tables showing the relationships about land and its attachments. Legal aspects, land value and other situations are clearly recorded.

### 4.4.3. Analysis of cadastral data

The four types of cadastral data will be presented under the criteria mentioned in section 4.3 to a clear understanding of the cadastral data in P.R.C.

#### 4.4.3.1. Parcel data

Parcel data mainly covers the basic information about the location, ID and area, property data and processing data. Property data make clear records of land property status and property history. Processing data is about the land data which are changing with time passing by. These kinds of data are in textual or attribute files which can be accessed by the unique parcel codes shown on the cadastral map. These data are mainly from the LAB which stores the long history of land. The parcel data are recorded in the registration book and therefore most of these data are in forms of tables.
Mediating Data Conflicts on Integrated Land and Housing Information in China

recorded in books. Nowadays, it is stored in CIS as metadata to record the parcel situation. The update of parcel data is depended on how frequent the parcel changes.

4.4.3.2. Non-parcel data
Non-parcel data refers to laws, regulations and statistics related to parcel. These kinds of data are normally records in forms of paper or digital document which are introduced at the background when the parcel records were set up. The government offices especially the MLR makes over-all laws and regulations while provincial administrative departments make local regulations according to MLR’s laws and local situations. This kind of data as red-head document is quite important and the local bureaus must obey which are issued by local provincial and municipal level administrative bureaus. All of the bureaus must keep them even when it is expired.

4.4.3.3. Cadastral map
Cadastral maps use graphical indices of the parcels which show the relative location of all parcels in a given region. Cadastral maps as key proofs of present parcel condition and they are prepared by professionals. SMBs are authorized by MLR to take the responsibility of cadastral survey and cadastral mapping while they can also authorize other private companies or institutions to do same. These organizations have to pass strict examinations and receive their ranking certificate which indicates their ability level to survey and mapping cadastral data. For example, the provincial SMB is professional in surveying and mapping, so its ranking certificate is the first level certificate. Cadastral data acquisition is usually taken by modern hi-tech equipments, such as theodolite, water level machine or global station machine. Beijing GCS 1954 and Xi’an GCS 1980 are two main coordinate systems in the whole nation as introduced in 3.2.3.1. Cadastral maps commonly range from scales of 1:10,000 to 1:500. Cadastral maps are managed in and belonged to LABs while SMBs only take the responsibility of surveying and mapping. If the SMB intend to use the cadastral map, it also has to pay for but with a lower price for government use. Cadastral map has a large range of users from government to common citizens become the records are used for fiscal, legal, land use or other purposes.

4.4.3.4. Thematic map
Thematic map displays the spatial pattern of a theme and attributes of this particular theme. The thematic maps portray spatial variations and interrelationships of geographical distributions. It provides specific information and its spatial patterns about particular locations. Data source referring to map maker and method of making thematic map is almost the same with cadastral map. GCS adopts Beijing GCS 1954 and Xi’an GCS 1980 as well. As the thematic map is used frequently, so LAB and SMB both have the ownership and they could supply the map to the whole society. However, the scale of thematic map should be between 1:250,000 to 1:1,000,000 (State 2003). Thematic maps satisfy the largest group of users who use them in daily life for transportation, travel, invest, learning, research and so on. These kinds of thematic map should be paid but not expensive. Nowadays, most persons could access to internet which enables some particular websites about thematic maps.
4.4.4. **Datasets in cadastral managements**

Three basic datasets covering parcel, right and person are the main datasets in the Chinese CIS. The main relationship is depicted in Figure 4-2 which indicates that person could only have user ship right over parcel according to LAL.

![Diagram showing the relationship between Parcel, Usership right, and Person](image)

**Figure 4-2: Three core classes in cadastral management**

Some sample datasets of cadastral management as shown in Table 4-1. The datasets are taken from the Urban Cadastral Database Standards. Local cadastral databases are built under this standard; therefore, the datasets type could be representations of cadastral datasets in national wide.

Table 4-1: Dataset information on person

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence number</th>
<th>Field Title (Chinese code)</th>
<th>Field Title (Translated into English)</th>
<th>Field Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>名称</td>
<td>Name</td>
<td>String</td>
<td>user name</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>身份证号码</td>
<td>ID</td>
<td>String</td>
<td>A unique identifier for person as defined by the jurisdiction</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>电话</td>
<td>telephone</td>
<td>String</td>
<td>latest phone number</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>地址</td>
<td>Address</td>
<td>String</td>
<td>latest address</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>出生日期</td>
<td>Birth</td>
<td>String</td>
<td>birth date</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>性别</td>
<td>gender</td>
<td>String</td>
<td>female or male</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>邮件</td>
<td>E-mail</td>
<td>String</td>
<td>a address in cyber space</td>
</tr>
</tbody>
</table>
### Table 4-2: Dataset information on user ship right

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence number</th>
<th>Field Title (Chinese code)</th>
<th>Field Title (Translated into English)</th>
<th>Field type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>User ship right</td>
<td>1</td>
<td>土地证编号</td>
<td>Certificate no.</td>
<td>Int</td>
<td>A unique user ship right identifier for a user as defined by the jurisdiction</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>设定日期</td>
<td>Begin date</td>
<td>Date</td>
<td>since when use right is valid</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>注销日期</td>
<td>Expire date</td>
<td>Date</td>
<td>since when use right is valid</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>使用权人</td>
<td>user name</td>
<td>String</td>
<td>who is user</td>
</tr>
</tbody>
</table>

### Table 4-3: Dataset information on parcel

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence number</th>
<th>Field Title (Chinese code)</th>
<th>Field Title (Translated into English)</th>
<th>Field type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parcel</td>
<td>1</td>
<td>宗地号</td>
<td>parcel ID</td>
<td>String</td>
<td>A unique parcel identifier for the parcel as defined by the jurisdiction</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>要素代码</td>
<td>feature code</td>
<td>Int</td>
<td>code number about the description of the source reference for the parcel</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>使用权代码</td>
<td>area code</td>
<td>Int</td>
<td>code of administrative area where the parcel located</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>土地用途</td>
<td>land use type</td>
<td>String</td>
<td>usage of land</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>土地坐落</td>
<td>land location</td>
<td>String</td>
<td>precise description of where the land is</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>建筑物密度</td>
<td>building density</td>
<td>String</td>
<td>building area divide parcel area</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>建筑物容积率</td>
<td>building occupied area</td>
<td>String</td>
<td>building area</td>
</tr>
</tbody>
</table>
4.5. **Housing Data**

Housing data is about the housing location, housing condition, and housing property data. Housing data is not only for governance but also is important for economic as it could reflect on past and current situation and predict future trend. Therefore, housing data are widely used in city planning, land management and economic management.

Housing data is also a kind of spatial data and it could be divided into two basic types, such as basic geographic data and housing thematic data. Basic geographic data is simply introduced while housing thematic data is described in detail in this research. All of these data were collected during fieldwork in P.R.C.

4.5.1. **Basic geographic data**

Basic geographic data are about the related data which have close relationship with housing data. These data are mainly collected by SMB through the methods mentioned in section 3.2.3.3 and could be bought in under certain regulations in SMB. Basic geographic data could be applied in broad fields such as city planning and environment analysis. Basic geographic data of housing are mainly 9 kinds, they are as follows:

- Survey reference point: points used to be referenced in surveying.
- Residential parcel: zone of residential
- Buildings for factory and mine including their facilities;
- Transportation infrastructure;
- Pipeline infrastructure;
- Water systems;
- Boundary; boundaries among objects.
- Physiognomy and terrene;
- Land use type and vegetation;

4.5.2. **Housing thematic data**

Land housing block, building, floor and apartment constitute the housing thematic data and they are the four main parts in housing management system in China. Relationships between four types of data are shown as flowing Figure 4-3 from top view and front view and these data will be modeled in chapter 5, section 5.3.2. Datasets in this section are derived from Housing Attribute Data Structure Standard in Beijing.
4.5.2.1. Land housing block

Land housing block which is shown as “A” in Figure 4-3 and 4-4 refers to a piece of land with limitary boundary as the maximal unit in housing survey. Land housing block is widely used in housing survey and it is collected and update by SMB authorized by HAB. There is no overlap between two pieces of land in spatial parameters while incorporation is allowed. It is clear that a land housing block could be a composition of many pieces of parcels which is presented in Figure 4-3 as “B”. Land housing block attributes are produced and maintained by the survey department and attributes query is allowed in every application but no revisions.

In order to make a clear record of land housing block, the changed situation should be recorded in its attribute list. The land housing block code should be a new one after several pieces of land housing blocks are incorporated. For example, if two land housing blocks are incorporated, the new piece should have a new land code. Attributes of land housing block changes is listed in Table 4-4 below. Land housing block changing records which records the changing is in appendix 4-2.
Table 4-4: Dataset information on land housing block

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence number</th>
<th>Attribute (Chinese code)</th>
<th>Attribute (Translated into English)</th>
<th>Field type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land housing block</td>
<td>1</td>
<td>丘编码</td>
<td>Code</td>
<td>Int</td>
<td>A unique identifier for land housing block as defined by the jurisdiction</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>区编码</td>
<td>area code</td>
<td>Int</td>
<td>code of administrative area where the parcel located</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>图幅号</td>
<td>map no.</td>
<td>Int</td>
<td>no. for the land housing block map</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>房屋占地面积</td>
<td>occupation area of building</td>
<td>string</td>
<td>building area</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>测绘日期</td>
<td>survey date</td>
<td>Date</td>
<td>when the surveying is happens</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>丘状态</td>
<td>land status</td>
<td>string</td>
<td>Use valid or incorporate describes its’ status.</td>
</tr>
</tbody>
</table>

4.5.2.2. Building

Building is shown as “C” in Figure 4-3 and 4-4 referring to physical constructions which could be identified by its boundary. It could be a flat, a story building or a house. So on a piece of land housing block may build many buildings. Building is the basic unit in housing survey. Building attribute data are produced and maintained by HAB and attributes query and citation is allowed in every application but no revisions. The spatial data is produced and updated by SMB authorized by HAB. Building data is the most needed one as it required by common citizens who need some information to help purchase decision making and government agencies to help management.

Building data rely on land housing block record and is normally stored in database systems. Whenever land housing block record is invalidated, the building record on it will also be invalidated. Building code remains as the building exists no matter whether the land is incorporated or not. The attributes of building data is shown as follows in Table 4-5.
### Table 4-5: Dataset information on building

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence number</th>
<th>Attribute (Chinese code)</th>
<th>Attribute (Translated into English)</th>
<th>Field type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>64</td>
<td>No.</td>
<td>Int</td>
<td>a unique number for building generated by the system</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>area code</td>
<td>Int</td>
<td>code of administrative area where the parcel located</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>丘编码</td>
<td>land housing block code</td>
<td>Int</td>
<td>block number in local usage</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>门牌号</td>
<td>doorplate No.</td>
<td>Int</td>
<td>Number generated by police</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>地址</td>
<td>location</td>
<td>string</td>
<td>description about where the building is</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>使用类型</td>
<td>use type</td>
<td>string</td>
<td>what the building used for, e.g. inhabitation, commercial housing</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>结构</td>
<td>structure</td>
<td>string</td>
<td>steel structure or brick structure</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>密度</td>
<td>density</td>
<td>string</td>
<td>building area divide parcel area</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>面积</td>
<td>Area</td>
<td>string</td>
<td>building area</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>建筑年代</td>
<td>built year</td>
<td>Date</td>
<td>when it is built</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>测绘员</td>
<td>surveyor</td>
<td>string</td>
<td>who did surveying of building</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>测绘日期</td>
<td>survey date</td>
<td>Date</td>
<td>when the surveying is happens</td>
<td></td>
</tr>
</tbody>
</table>

As mentioned above, when land housing blocks are incorporated, the buildings which lie on these lands without code changed if their physical status remains the same as before. However, building should be indicated attaching to new land. Building change is about the records of a building changing situation. Building changing attribute list is in appendixes 4.3.

### 4.5.2.3. Floor

Floor is referred to “D” in Figure 4.3 and 4.4. Floor is used to describe a certain layer in a building. However, geometrical attribute of floor is a layer layout without geospatial reference. Relative apartment position, exits and stairs are the main presents in a floor map. Common area and balcony area are also main data recorded in floor datasets. Table 4-6 shows the main attribute sets of floor.
### Table 4-6: Dataset information on floor

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence number</th>
<th>Attribute (Chinese code)</th>
<th>Attribute (Translated into English)</th>
<th>Field type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>1</td>
<td>ID</td>
<td>Int</td>
<td>ID is the key number generated by system and it can be used for enquiry floor data.</td>
<td></td>
</tr>
<tr>
<td>floor code</td>
<td>2</td>
<td>floor code</td>
<td>Int</td>
<td>A unique number to identify floor</td>
<td></td>
</tr>
<tr>
<td>Building code</td>
<td>3</td>
<td>Building code</td>
<td>Int</td>
<td>The building which is the floor belonged to.</td>
<td></td>
</tr>
<tr>
<td>Amount of apartment</td>
<td>4</td>
<td>Amount of apartment</td>
<td>string</td>
<td>How many apartments in this building.</td>
<td></td>
</tr>
<tr>
<td>Common area</td>
<td>5</td>
<td>Common area</td>
<td>string</td>
<td>E.g.: communal aisle, stairs, elevator space, water tank space.</td>
<td></td>
</tr>
<tr>
<td>Balcony area</td>
<td>6</td>
<td>Balcony area</td>
<td>string</td>
<td>Consistent with building which the floor in.</td>
<td></td>
</tr>
<tr>
<td>Floor map no.</td>
<td>7</td>
<td>Floor map no.</td>
<td>Int</td>
<td>Consistent with building which the floor in.</td>
<td></td>
</tr>
</tbody>
</table>

### 4.5.2.4. Apartment

Apartment is referred to “E” in Figure 4-3 and 4-4. Apartment is the smallest survey unit and it should be in accordance with a particular building. If the apartment has the same attributes with its building, then the common attributes would incarnate in the attribute list of building, such as the location of apartment. Generally, apartment is the main data in housing management as it more familiar to person and therefore attracts more attention. The main attributes of apartment is shown in table 4-7. When some changing happens, there are also records on it. The apartment changing record is in Appendix 4.1.
## Table 4-7: Dataset information on apartments

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence number</th>
<th>Attribute (Chinese code)</th>
<th>Attribute (Translated into English)</th>
<th>Field type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>ID</td>
<td>ID</td>
<td>Int</td>
<td>Apartment ID is generated by system and is as the main key.</td>
</tr>
<tr>
<td>1</td>
<td>户编码</td>
<td>Apartment code</td>
<td>Int</td>
<td></td>
<td>A unique number to identify apartments</td>
</tr>
<tr>
<td>2</td>
<td>丘编码</td>
<td>land housing block code</td>
<td>Int</td>
<td></td>
<td>A unique number to identify land housing block</td>
</tr>
<tr>
<td>3</td>
<td>建筑编码</td>
<td>Building code</td>
<td>Int</td>
<td></td>
<td>A unique number to identify building</td>
</tr>
<tr>
<td>4</td>
<td>使用权编号</td>
<td>Property ID</td>
<td>string</td>
<td></td>
<td>A unique identifier for owner ship as defined by the jurisdiction</td>
</tr>
<tr>
<td>5</td>
<td>分层编码</td>
<td>Floor code</td>
<td>Int</td>
<td></td>
<td>A unique number to identify floor</td>
</tr>
<tr>
<td>6</td>
<td>房屋总数</td>
<td>Room amounts</td>
<td>string</td>
<td></td>
<td>Total room apartments in one floor</td>
</tr>
<tr>
<td>7</td>
<td>测量日期</td>
<td>survey date</td>
<td>Date</td>
<td></td>
<td>When the floor survey took place</td>
</tr>
</tbody>
</table>

### 4.5.3. Housing property data and main types of data

Different services require different types of housing data. But almost every type of services has some relationship with housing property data as housing property is totally different from land property, obligee could enjoy the ownership about housing. Property relationship is modeled in Figure 4-5.
In this model three main classes are obligee, housing and housing property. Obligee could be a person or an organization like army or school. Generally, obligee attributes are some basic ones such as name, ID, birthday and address. More detail information is also available in housing management.

Housing here could be buildings, floors and apartments. Its main attributes are shown in the class diagram. Housing status mainly choose between valid, logout and part restricted. Valid means property is recognized by the government, in most cases, it has a default. Logout indicates property has been cancelled some time. Part restricted demonstrates some parts of the housing are under mortgage or sequestrated. Full restricted means all of the housing are mortgaged or sequestrated. In order to make detailed relationship between property and each apartment, there is a housing property detailed record. It makes clear about each kind of rights and its related information.

Housing has three main types of property: housing property, property right of common and other rights. Right of common and other rights are subclass of property right. Obligee and housing are the two main objects of property right. Property links obligee and housing together, a housing could belong to several obligees and one obligee could has several housings, so the relation between housing and obligee is one to many. So, in housing property management, a property belonged to one or many obligees and an obligee could cover several properties, in this situation the property is called common property. One right of common uniquely matches one property while one property could match several rights of common. Attribute list of right of common is shown in Appendix 4.4.

Other rights include leasehold, mortgage and heirdom rights on the housing. Other rights correspond to part or all housing under one property or several properties. However, a piece of housing could only with one other right in a certain period. Attribute list of other rights are shown in Appendixes 4.5.
Table 4.8 presents main datasets in housing managements covering obligee and ownership right.

Table 4-8: Basic datasets on housing ownership

<table>
<thead>
<tr>
<th>Association</th>
<th>Sequence number</th>
<th>Attribute (Chinese code)</th>
<th>Attribute (Translated into English)</th>
<th>Field type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oblige</td>
<td>1</td>
<td>姓</td>
<td>family name</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>名</td>
<td>given name</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>生于</td>
<td>Year of birth</td>
<td>Date</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>使用权编号</td>
<td>Identification No.</td>
<td>Int</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>电话号码</td>
<td>tel.</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>性别</td>
<td>Gender</td>
<td>string</td>
</tr>
<tr>
<td>Housing property</td>
<td>1</td>
<td>使用权编号</td>
<td>ID no.</td>
<td>Int</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>房屋所有权证</td>
<td>Property certificate No.</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>土地所有权证</td>
<td>Land certificate No.</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>建号</td>
<td>Building no.</td>
<td>Int</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>宗地号</td>
<td>Parcel No.</td>
<td>Int</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>开始日期</td>
<td>start date</td>
<td>Date</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>结束日期</td>
<td>finishing date</td>
<td>Date</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>交易日期</td>
<td>transfer date</td>
<td>Date</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>发证日期</td>
<td>Issue organization</td>
<td>string</td>
</tr>
</tbody>
</table>

4.6. Data covering both cadastral and housing

As stated previously, it is easy to find out that some data exist in both cadastral and housing management. Same datasets can be derived though the previous introductions and documentation, some of them have different names or different types. The common datasets related to a certain association are generated as follows in Table 4-3, 4-4 and 4-5.
Table 4-9: Attributes associations for “person” in cadastral databases and housing databases

<table>
<thead>
<tr>
<th>Association name in cadastral database</th>
<th>Attributes in cadastral database</th>
<th>Association name in housing databases</th>
<th>Attributes in housing database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person</td>
<td>Name</td>
<td>obligee</td>
<td>family name</td>
</tr>
<tr>
<td></td>
<td>ID</td>
<td></td>
<td>given name</td>
</tr>
<tr>
<td></td>
<td>telephone</td>
<td></td>
<td>Year of birth</td>
</tr>
<tr>
<td></td>
<td>Address</td>
<td></td>
<td>Identification No.</td>
</tr>
<tr>
<td></td>
<td>Birth</td>
<td></td>
<td>tel.</td>
</tr>
<tr>
<td></td>
<td>gender</td>
<td></td>
<td>Gender</td>
</tr>
<tr>
<td></td>
<td>E-mail</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4-10: Attributes associations for “right” in cadastral databases and housing databases

<table>
<thead>
<tr>
<th>Association name in cadastral database</th>
<th>Attributes in cadastral database</th>
<th>Association name in housing database</th>
<th>Attributes in housing database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right</td>
<td>Certificate no.</td>
<td>owner ship right</td>
<td>ID</td>
</tr>
<tr>
<td></td>
<td>Commencement date</td>
<td></td>
<td>building no.</td>
</tr>
<tr>
<td></td>
<td>Expire date</td>
<td></td>
<td>Parcel No.</td>
</tr>
<tr>
<td></td>
<td>user</td>
<td></td>
<td>start date</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>finishing date</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>transfer date</td>
</tr>
</tbody>
</table>

Table 4-11: Other relevant in cadastral databases and housing databases

<table>
<thead>
<tr>
<th>name</th>
<th>difference</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cadastral database</td>
<td>housing database</td>
<td></td>
</tr>
<tr>
<td>building</td>
<td>attribute</td>
<td>class</td>
<td></td>
</tr>
<tr>
<td>floor</td>
<td>attribute</td>
<td>class</td>
<td></td>
</tr>
<tr>
<td>apartment</td>
<td>attribute</td>
<td>class</td>
<td></td>
</tr>
<tr>
<td>map no.</td>
<td>string</td>
<td>Int</td>
<td></td>
</tr>
</tbody>
</table>
4.7. **Main types of data conflicts in cadastral and housing databases**

Based on the three tables above and the two databases, five main conflicts between cadastral databases and housing databases are generalized. The conflicts are grouped according to the conflicts types which were introduced in section 2.7.1. These conflicts are solved in demonstration in section 5.5.4.

- **Structure conflict:** Apartment is an attribute in cadastral databases while it is a class in housing databases. (see Table 4-5)
- **Domain conflict:** Land in cadastral database and land housing block in housing database have the same pronunciation and characters in Chinese. However, land in cadastral database means a piece of parcel while in housing databases land housing block means a block for housing built and as the maximal unit in housing survey. In this situation one attribute can encompass different information than another one.
- **Intentional conflicts:** Apartment is recorded in both of the cadastral and housing databases while the attributes of apartment are different. (see Table 4-5)
- **Naming conflicts:** Person in the cadastral database and oblige in housing database have the same concept that means who enjoy the user right or ownership. (see Table 4-3) Commencement date and start date also have the same meaning while using different expressions. (see Table 4-4)
- **Missing attributes:** Some attributes missed, e.g. the e-mail is included in cadastral data but not in housing one. (see Table 4-3)

4.8. **Concluding remarks**

In this chapter, situation on LIS, CIS and housing systems were introduced briefly and datasets on cadastral and housing presented in details. Cadastral data is about parcel entity data, parcel non-entity data, cadastral map and thematic map. Housing thematic data are referring to land housing block, building, floor and apartment. Based on the cadastral and housing data analysis, five main conflicts from structure, domain, intentional naming and missing attributes are generalized. The conflicts are solved by FDM which is developed in chapter 5.
5. Development of Federated Data Model

5.1. Introduction

Having discussed the concepts of FDM and the As-Is situation in the previous chapters, this chapter describes the development of FDM to resolve heterogeneity problem using the Chinese cadastral and housing data. Mediation as a method for schema integration and a federated schema will be the output schema which is the result of schema integration of cadastral and housing databases.

This chapter seeks to develop an approach for resolving semantic conflicts of land and housing data through federated data model. Framework for developing a FDM is introduced in section 5.3. Besides the development of FDM, mediation part is also proposed in 5.3. Practical cases with cadastral and housing data are used to demonstrating the FDM in 5.4.

5.2. FDM as approach for solving conflicts

FDM which generates federated schema is seen as basis for mapping the schemas coming from cadastral databases and housing databases and it is used to resolve the conflicts between cadastral data and housing data derived in section 4.7. Schema mapping is a kind of conflicts solving by creating integrated schemas introduced in 2.6.3.3. Schema mapping is based on the content of attributes values instead of rich domain knowledge and requires data managers to familiar with semantics between the various elements and concepts. So, this avoids the construction of ontology for a particular domain.

A proposed FDM architecture is presented in Figure 5-1. In this architecture, cadastral DB and housing DB are the components viewed as an autonomous database. Each component has schema which is a collection of datasets types. The main goal here is the efficient and effective application of resources to deliver the required cadastral data and housing data though federated schema to different users with the help from mediation.

Federation architecture is a bottom-up structure with three levels:

- The first one is database level, including cadastral DB and housing DB, each of them collects and maintains their data to meet its unique missions but coordinates to satisfy the various requirements from global users.
- The second level is mediation layer, solving schematic heterogeneity issues.
- The third level is the conceptual federated schema layer which meets with the users requirements is derived with the help of mediation.

Users are in top to receive the federated schema. The users could be government agencies, real estate companies or institutes who need both cadastral data and housing data.
5.3. Framework for mediation

Mediation refers to the dynamic interface function which composes processes needed to make interfaces work, knowledge structures drive transformations and intermediate storage that is demanded. Among many functions, five main types of mediation functions have been fostered: transformation and sub setting of databases using view definitions and object terms; methods to access and merge data from multiple databases; computations which support abstraction and generalization over underlying data; intelligent directories to information bases and methods to deal with uncertainty and missing data (Wiederhold 1992). As mediation is a main part in this research, this section will provide the framework for it to show the schema integration steps.

As presented in Figure 5-2, there are four main steps composes of data integration as follows.
5.3.1. Pre-integration

This is the first step in which input schemas are re-arranged in various ways to make them more homogenous. This step refers to schema transformation in Figure 5-2. Local heterogeneous schemas are transformed to local homogeneized schemas under transformation rules.

5.3.2. Identifications

Identification as the correspondence investigation in Figure 5-2 is a step focus on precise description on related items in the input schemas and identifying their inter-schema relationships. This is done by matching rules about analyzing schematic properties of attributes, types and relationships in the to-be integrated schema. Besides, communicating with designers and exploiting the understanding of the application domain is also highly requested. After this step, the local homogeneized schemas have changed into inter-schema correspondence assertions.

5.3.3. Schema integration

In this step, the inter schema relationships identified previously are used to generate an integrated representation of the output schema under integration rules based on requirements of users. Schema integration devotes to resolving various forms of heterogeneity (or conflicts) that exist in two related databases.

5.3.4. Mapping definition

Mapping definition as the final step unifies corresponding items into an integrated schema and produces the associated mappings.

5.4. UML data model diagrams for FDM

Microsoft Office Visio is used to make UML data model diagrams in this research. Three packages are created for drawing the models of cadastral, housing and integrated data before the modelling as shown in Figure 5-3. Entities, attributes, relationships are modelled under these packages. And mediation package is used to present the functions of integration.
5.4.1. Data model for cadastral property data

Cadastral property data is mainly about person, use right and parcel. Person is referring to a legal individual or an organization that can enjoy use right on land (Bureau 1986). Their relationship is presented as follows together with their attributes. The attribute names in the model are what they are in real CIS systems.

In this cadastral data model as shown in Figure 5-4, nine classes have been created under the land DB package: parcel, user-ship right, person, individual, organization, ownership right, housing, apartment and floor. This model makes clear that person could only have user-ship on land and have owner ship on housing. Person could be legal individual or organizations which may be government agencies or enterprises. Housings in cadastral record are the generalization about housing, apartment and floor while in housing data records the four are independent ones which will be showed in the housing data model.

Figure 5-4: Cadastral data model in Chinese city
5.4.2. Data model for housing data

As has discussed before, land housing block, building, layer and apartment are four kinds of entities that describe housing nature attribute. Among these four, building is the core of housing survey. Apartment and layer are the details of building while land housing block is the organizational part. Their relationships are shown in Figure 5-5 below. Beside the four classes, ownership right class, classes for building, land housing block and apartment change and oblige class together with its sub classes named individual and organizations are set up.

In housing data model, it is clear that the main classes are about housings while the land data are just attributes in land housing block, building and floor classes. Moreover, buildings, apartments and floors are the sole classes which are different from the cadastral data model. The meanings of these class names have already been introduced in section 4.5.2.

![Figure 5-5: Housing data model in Chinese city](image_url)
5.4.3. Integrated data model

Based on the study of present cadastral and housing management and the idea, an integrated data model is proposed as Figure 5-6 shown. In this data model, person is the key class connected both the ownership right which is mainly about housing property part and user right which about the cadastral part. Through this data model, it is easy to search both the housing and cadastral data by person ID, parcel ID, building ID or any ID. However, the housing and cadastral data are separately stored and maintained in different systems in HAB and LAB. Professional or non-professional stakeholders could get the data they want an easy and understandable format required for their appreciations. The demonstration will show how it works in section 5-6.

Figure 5-6: Integrated data model in Chinese city

5.4.4. Mediation functions

The idea of mediation is developed as a way of resolve the data conflicts mentioned in 5.3. Mediation is a package with many functions presented in Figure 5-7.
The functions of mediation are mainly about these:

**Test**: test means to check whether two or more source schemas are available to integrating so it is the first step in mediation.

**Mapping**: the mediation has a function of mapping same items or schemas together to get a united export schema.

**Logical**: logical could check whether the various source schemas are equal, greater or less to each other, and then make further operations.

**Conversion**: if the field types in source schemas and target schemas are different, mediation could converse the types to a unit one through some conversion.

**Math**: math referring to functions about add, minus, divide, multiply. Math could support the Boolean functions strongly.

**Translate**: this function means that mediation could understand some domain field glossaries.

### 5.5. Practical case for demonstrating FDM as prototype

In demonstrating the FDM, two databases schemas concerning on cadastral and housing property are set up according to the Urban Cadastral Database Standard and Housing Attribute Data Structure Standard in Beijing.

#### 5.5.1. Setting up cadastral property schema and data by XMLSpy

Cadastral property schema was set up by XMLSpy using cadastral datasets for demonstration. The cadastral property database includes usual datasets which are absolutely necessarily in daily cadastral management. Cadastral-Data is set up as a root element and person, right, and parcel are elements under the root element. Two complex type about address type and basic information type are set up as well. Some parts of the elements are show as follows in Figure 5-8:
Figure 5-8: Cadastral schemas (some parts)
Data on cadastral property schema is made according to some rules about land and housing. Such as rules for building coding and parcel number coding. Other data such as user names, ID and address are made as usual information of a Chinese person.

5.5.2. Setting up housing property schema and data by Microsoft Access

In many local bureaus, HAB employs Microsoft Access databases as their housing management systems as in Xiamen where the fieldwork was taken. According to the housing property data in reality, a simple Microsoft Access schema is set up to do the demonstrating. HousingDataFirst.mdm as presented in Figure 5-9 is built up with data about building, floor, land housing block, obligee, and ownership etc.

![Figure 5-9: Housing property schema](image)

Data on housing property schema is also made according to some rules about land and housing. Such as rules for land housing block coding, building coding, floor coding, apartment coding, parcel number coding. Owner names, e-mail and address are made as usual information of a Chinese person.

5.5.3. Setting up Export-schema based on a scenario requirements

The Export-schema was set up according to a scenario that a Court as a type of government stakeholder needs to investigate the person’s property condition urgently to check whether the person breaks the law or not. The court has no time to check this information from LAB and HAB. However, the FDM could help the Court through the designed format presented by the integrated data. The Export-schema is designed as follows in Figure 5-10:
5.5.4. Integration the two schemas by MapForce

The XML cadastral schema, database housing schema and Export-schema are imported. After the operations, conflicts in two schemas are solved. The output XML file is in appendixes 5. In this step, the conflicts mentioned in section 4.7 are solved.

As the integration is between two different schemas, priority context enclosed in a circle is designated to make a context node that is searching data firstly in the cadastral schema. This avoids repeated search in the two schemas.
5.5.5. Creation of an XSLT file and by Dreamweaver

In order to give the stakeholders a clear and easy interpreted view of the integrated dataset, an XSLT file will be used to present the XML file output. First a new html is created by design format which will be accessible to the users. The design output format should be done with careful consideration.
the saved designing could be seen by “preview” function key in IE-explorer. Second, the saved file is converted into an XSLT 1.0 and then attach XML source document which is created by MapForce in the integrating step which is shown in section 5.5.4.

5.5.6. The integrated schema output

The output of the integrated schema is shown as in Table 5-1. The Court could clearly see the person information including its basic information, property information and mortgage information. If more detail information about mortgage is needed, just enquire with Mortgage Certificate Number, Apartment ID or any data in this format again. Then, integrated detail data will satisfy the user’s requirement.

<table>
<thead>
<tr>
<th>Person Information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Basic Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
</tr>
<tr>
<td>Name</td>
</tr>
<tr>
<td>Birth Date</td>
</tr>
<tr>
<td>Email</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Tel</td>
</tr>
<tr>
<td>Address</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Property Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land User ship Certificate Number</td>
</tr>
<tr>
<td>Commencement Date</td>
</tr>
<tr>
<td>Expiry Date</td>
</tr>
<tr>
<td>Housing Ownership Certificate Number</td>
</tr>
<tr>
<td>Commencement Date</td>
</tr>
<tr>
<td>Expiry Date</td>
</tr>
<tr>
<td>Apartment ID</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mortgage Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortgage Certificate Number</td>
</tr>
<tr>
<td>Borrower Name</td>
</tr>
<tr>
<td>Mortgage Beginning Date</td>
</tr>
<tr>
<td>Mortgage Finishing Date</td>
</tr>
<tr>
<td>Apartment ID</td>
</tr>
<tr>
<td>Floor Map No</td>
</tr>
<tr>
<td>Building ID</td>
</tr>
<tr>
<td>Building Area</td>
</tr>
</tbody>
</table>
Table 5-1: Integrated data as output

<table>
<thead>
<tr>
<th>Building Density</th>
<th>Doorplate No</th>
</tr>
</thead>
<tbody>
<tr>
<td>89</td>
<td>126</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parcel Area</th>
<th>Improvement Type</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>YanTa</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parcel Value</th>
<th>Assessor Name</th>
<th>Assessor Tel</th>
</tr>
</thead>
<tbody>
<tr>
<td>3564444</td>
<td>Xu Na</td>
<td>13044542354</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assessor ID</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>345959198109233348</td>
<td></td>
</tr>
</tbody>
</table>

5.6. Discussion about the practical case

The conflicts discussed in section 4.7 are solved in this practical case. Test function of mediation first check whether the two resources schemas are available to integrating. If there are some problems with the schema, they can not be imported into MapForce. The structure conflict is solved when design the data model of integrated data that apartment is defined as an attribute of housing. Domain conflict is solved is also by integrated data model in FDM which is presented as the export-schema. Naming conflict and missing attribute conflict are solved by both mapping and logical function of mediation. The items are checked whether they are equal to each other and if not some logical functions will make them match each other and then mapping to the export part.

Conversion function which could converse the different field types into a unite one and math function which could do some mathematics are not demonstrate in this practical case as the sample data is an easy one.

5.7. Concluding remarks

The chapter has discussed the FDM developed by UML under the Chinese cadastral and housing situation. In this model, four packages are put forward and three data model covering cadastral, housing and integrated are developed by UML and mediation package functions are proposed through different packages. A scenario with the conflicts which are derived in chapter 4 is created and conflicts are solved by demonstrating FDM with help of XMLSpy, MapForce and Dreamweaver.
6. Conclusions and Recommendations

The rapid increase of information technology and different means of capturing and storing spatial information has led to the spring of different stems for storing data in organisation and government agencies. These often result in overlapping sources of data making and data exchange a problem to users when acquiring data. It is important to have systems capable of querying multiple databases rapidly. Moreover, the possibility to integrate and exchange data in a distributed environment is in high demand therefore, making research on data integration very vital in recent times.

Motivated by the necessity of developing data models to solve the duplication, low-efficacy and fragmented data supplying problems in cadastral and housing data provision in P.R.C, this research has reviewed and discussed in detail the As-Is situation in the institutional, technological and data aspects in the context of spatial data infrastructure. The Chinese government is making a great effort to put SDI in place. This infrastructure will greatly influence and support integrated data supply. Based on the derived conflicts from data analysis on the data collected in field in P.R.C, a Federated Data Model was developed and demonstrated with sample datasets presented in the context of mediation.

6.1. Conclusions

In order to fulfil the objectives of this research, seven research questions are answered during the whole work. Some conclusions are made according to these and then make some recommendations for future research study.

6.1.1. To assess existing situation of Chinese cadastral and housing management and datasets

Data management situation is pre-requisite for answering the objective which seeks to develop federation architecture for integrated data by assessing user requirements. Data on user kinds, most needed data, user requirements and the organizations responsible for the data were collected through observations, interviewing, documentations and questionnaires.

Three main kinds of users of cadastral and housing data grouped as government agencies, organizations and citizens were identified. Government agencies are willing to provide and receive comprehensive, up-to-date and accurate data to satisfy users need and foster their governance. Organizations referred to as enterprises such as Real Estate Company are willing to obtain data as quick as possible without going through any difficulty in data mismatching. Citizens most of whom are non professionals in the analyses of cadastral and housing data, after obtaining the data have to require the comments from the professionals to help decision making.

Cadastral and housing data are managed in different bureaus named LAB and HAB. The data in these bureaus are collected by themselves, SMB or the organizations which have been issued the survey and mapping certificate by SMB. They collect the data through many different approaches. Spatial data are obtained from the survey through measurement, construction using coordinate geometry and best
fit to base map and digitizing to base map. Attribute data are not only obtained from surveying but also from some regulations that are issued by upper government or documentations covering the history of the data.

6.1.2. To develop a federated data model that resolve semantic conflicts within federation architecture

In order to solve the problems in As-Is situation, a FDM was developed. This requires a carefully study of the data types and data relationship.

Object, person and right are main data in both cadastral and housing data and these datasets are almost the same. Data differences are the structure, name and attribute. For example, parcel, person and user ship right are three main types of data in cadastral management while land housing block, building, apartment, floor, obligee and ownership right are main ones in housing management.

Five differences existed in these two types of data; structure conflict occurs when an attribute in one schema is a class in another schema; domain conflict occurs when there is different attribute value representations; intentional conflicts occurs when the integrated schema and resource schema have different sets of attributes; naming conflicts occurs when different names presents the same concept; missing attributes occurs when some attributes are missing.

A FDM was developed by designing cadastral data model, housing data model, integrated data model and mediation as a means to solve these conflicts. Several classes are set up with relationships between them. The integrated schema defined according to the integrated data model and user requirements was created by XMLSpy.

The model was tested and implemented using sample cadastral and housing datasets in the context of a scenario using MapForce, XMLSpy, Microsoft Access and Dreamweaver. Cadastral property schema was defined by XMLSpy while the housing property schema is defined by Microsoft Access. Integrated data was achieved with the conflict solved by mediation.

6.2. Recommendations

The process of data integration is both technical and institutional. FDM is only the technical aspect of data integration; therefore how the condition of institutional aspect of FDM should studied in future research. Policy, stands or legislation and who will be the professional agency to supply integrated data.

The development of FDM in this research is a modest one for some initial study. Thus, only attribute data are taken into consideration in this research. Since land and housing are typical spatial-referenced, cadastral data provide a spatial representation for a piece of land with defined boundaries. The spatial data related to parcel and housing are also displayed by polygons, lines and points. Other annotations are necessary as well. Therefore, how to integrate the spatial data is an issue for future research.
The mediation part only presents its functions by several packages in this research. As an initial study on FDM, the detail part on mediation has to study in the future research. Mediation could be developed based on the model of OCL/UML. Object Constraint Language (OCL) which is a part of UML can aid in the unambiguous modelling of database constraints.
Bibliography and Reference:


MLR (2002). Researches on digital China NSDI National Geomatics Center of China(NGCC): 36.


Ressl, C. (2002). "THE IMPACT OF CONFORMAL MAP PROJECTIONS ON DIRECT GEOREFERENCING."


## Appendix

### Appendix 1: Activities in fieldwork in P.R.C

<table>
<thead>
<tr>
<th>Time</th>
<th>City</th>
<th>Activity</th>
<th>Contact Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>July30~ Aug.1</td>
<td>Xi’an</td>
<td>Interview</td>
<td>Prof. Ma Zhiming (Professor in Chang’an University)</td>
</tr>
<tr>
<td>Aug.4</td>
<td>Beijing</td>
<td>Interview Questionnaires</td>
<td>Zhang Ning (staff in Cadastre Department of Land Survey and Planning Institute. MLR)</td>
</tr>
<tr>
<td>Aug.7~10</td>
<td>Xiamen</td>
<td>Interview Questionnaires</td>
<td>Guo Junsheng (Deputy Director-General in Xia’men Municipal State Resources and Housing Administrative Bureau)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Jiang Chunfa (Senior engineer in Survey and Geo-information Centre, XMSRHAB)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Li Guiming (Senior engineer in housing transaction management in XMSRHAB)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lin Shunv (staff in cadastral section of XMSRHAB)</td>
</tr>
<tr>
<td>Aug.21~25</td>
<td>Beijing</td>
<td>Interview Questionnaires</td>
<td>Zhao Daihong (staff in Information Centre in MLR) &amp; Zeng Yuandi (staff in Beijing Municipal State Resources Administrative Bureau)</td>
</tr>
</tbody>
</table>
Appendix 2: Questionnaires for users

2.1 Basic information of users.

2.1.1 What is your occupation?
- real estate practitioner
- architecture practitioner
- IT practitioner
- artistic practitioner
- financial practitioner
- governor
- professional
- businessman
- teacher
- peasant
- student
- army man
- jobless man
- other type

2.1.2 How much is your income one month? (RMB)
- \(\leq 1000\)
- 1001～2000
- 2001～3000
- 3001～4000
- 4001～5000
- 5001～6000
- 6001～8000
- 8001～10000
- \(\geq 10000\)

2.1.3 How old are you?
- \(\leq 20\)
- 21～30
- 31～40
- 41～50
- 51～60
- \(\geq 61\)

2.1.4 Where are you graduated from?
- middle school
- technical school
- university

2.2. Users’ requirements

2.2.1 What kinds of land data do you need?
- land use type
- land location
- land price
- use time
- land area
- four reaches
- surroundings
- transportation
- others

2.2.2 What kinds of housing data do you need?
- housing location
- four reaches of housing
- building number
- building layer
- user area
- architecture
Mediating Data Conflicts on Integrated Land and Housing Information in China

☐ face ward  ☐ architectural area
☐ housing price  ☐ housing service
☐ others

2.2.3 Why you need for these kinds of data?
☐ living  ☐ investing
☐ other

2.2.4 Do you know where to get these kinds of data?
☐ yes  ☐ no

2.2.5 Where do you get these kinds of data?
☐ TV  ☐ book  ☐ internet  ☐ administrative bureaus
☐ libraries  ☐ agency  ☐ other

2.2.6 Are you satisfied with the present services?
☐ yes  ☐ no

2.2.7 What other kinds of services do you want?
☐ comprehensive data about land and housing
☐ consult data in land and housing bureaus
☐ obtain data in paper, CD forms though post
☐ searches, query and download data in internet
☐ according to your own requirements, supply individual services
☐ others

3 Questionnaires for bureau staffs

3.1 What are the main types of data users?
☐ citizens  ☐ real estate industry
☐ court  ☐ government
☐ investment trust  ☐ others

3.2 What are the supplied data used for?
☐ resources planning and administration
☐ economic developing
☐ environment protection
☐ land use planning
☐ agriculture, herd, forest
☐ economic trade
☐ scientific study
☐ scientific training
☐ city construction planning
☐ planning for important project
☐ prepare study for project
☐ others

3.3 Which part of your organization in charge of the land data collection, storage, analyze, issuance and updating?

Collect : ------
Store : ------
Analysis : ------
Issue : ------
Update : ------

3.4 Which part of your organization in charge of the housing data collection, storage, analyze, issuance and updating?

Collect : ------
Store : ------
Analysis : ------
Issue : ------
Update : ------

3.5. How to issue the data:

---------------------------------

3.6. What kinds of land data are users’ most favorite?

☐ usage ☐ land classification
☐ property right ☐ bench mark price
☐ applied price ☐ location
☐ four reaches ☐ expiration date
☐ neighbor parcel ☐ parcel title
☐ other

---------------------------------

3.7. What kinds of housing data are users’ most favorite?

☐ location ☐ four reaches
☐ building number ☐ building type
☐ obligee ☐ architecture area
☐ use area ☐ architecture
☐ face ward ☐ surroundings
☐ housing price ☐ information of the land where the housing is
☐ other data

---------------------------------

3.8 What else do you supplied?
3.9 What are the methods of collecting data?

3.10 What kinds of management systems related to land?

3.11 What kinds of management systems related to housing?

3.12 What are the advantages and disadvantages of these systems?

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.13 Are there any similarities of data type in these two types of systems?

- [ ] yes
- [ ] no

3.14 If integrate the land and housing management systems, what problems will happen?

3.15 Are there any conflicts on these aspects and could you give an example?

- [ ] structure conflict example:
- [ ] attribute conflict example:
- [ ] meta conflict example:
- [ ] intensional conflict example:
- [ ] other

3.16 How do you think of present services supplied?

- [ ] satisfied
- [ ] not satisfied

3.17 How do you think of the info-service of land and housing?

- [ ] good
- [ ] ok
- [ ] bad

Could you give some advices?

---

79
Appendix 3: Specific services covering both cadastral and housing data

- **Services on survey**: services on survey are popular these years especially concerning to housing.
  Topographic map supply: map is only in a certain distance scope
  Geo-information data (scale, number, documentation);
  Thematic map supply: map is only in a certain distance scope
  Transformative survey: mainly concerning housings and buildings
  Municipal road survey: for municipality
  Private housing renewes;
  Cadastre survey: for organizations and private ones;
  Examination of private cadastre survey;
  Leveling: capable of indicating the level of a surface between horizontal and vertical to the nearest degree.

- **Services on land assignment**: means land allowance.
  Land, housing transfer, exchange and donatives’ information;
  Land, housing of company’s trade information;
  Land, housing cession: Cession is action of ceding land.
  Economic housing\(^2\) and private state-owned housing\(^3\) trade;
  Collective housing trade;
  Pre-sell to sell;
  Changing information of selling housing;
  Contract for housing sell and buy.

- **Services on property right**:
  Owner right registration of new house;
  Modify owner right registration;
  Multi-owner registration of house;
  Modify multi-owner right registration;
  Registration for houses in cases of without some certificates, land transferred from farming land, in legal problems and successive houses.

- **Services on acquisition**:
  Qualified enterprise for implement acquisition;
  Intermediation data;

- **Services on mortgage**:
  Cancel (part of) and modify mortgage;

---

\(^2\) Economic housing: a kind of housing of equal quality with the others in transaction but quite cheap, especially for local couples who can’t afford the high amount of housing.

\(^3\) Private -owned state housing: a kind of former state-owned housing belong to private after 1998 when “the distributing apartment” mechanism was abolished and then purchasing housing in real estate market. That is, obtaining housing with charge from market instead of from government free of charge.
Private house and enterprise in village for mortgage;  
Land mortgage;  
Second-hand house mortgage;  
State-owned land mortgage;  
Use right of badlands mortgage;  
Mortgage by preselled housing.

- Services on public fund for house:  
  Loan information for house;  
  Public fund in an organization for house;  
  Apply for public fund;  
  Refund loan;  
  Account for public fund.

- Services on unsafe housing:  
  Confirmation of unsafe housing;  
  Substitute house for unsafe housing.

- Services on state-owned housing:  
  Purchasing state-owned housing;  
  Apply for use right ship of state-owned housing.

- Services on land assignment and land price:  
  Land assignment procedure;  
  Postpone a deadline and de-rate for land money.

- Services on land use right management:  
  Exceed use of land;  
  Certificate for land use right;  
  Division the land use right.

- Services on tax:  
  Tax drawback, payment and de-rate.

- Services on housing policy reform and fund management:  
  Organizations apply for fund use;  
  Subsidy for staffs in organization;  
  Organization-owned house sell;
Appendix 4: Attribute list

<table>
<thead>
<tr>
<th>Association</th>
<th>Sequence number</th>
<th>Attribute (Chinese code)</th>
<th>Attribute (Translated into English)</th>
<th>Field type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apartment changing</td>
<td>1</td>
<td>BGBH</td>
<td>ID</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>YSHH</td>
<td>Previous apartment ID</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>XHH</td>
<td>New apartment ID</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>BGSJ</td>
<td>Changing time</td>
<td>date</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>BGYY</td>
<td>Changing reasons</td>
<td>string</td>
</tr>
</tbody>
</table>

4.1 Apartment changing attribute list:

<table>
<thead>
<tr>
<th>Association</th>
<th>Sequence number</th>
<th>Attribute (Chinese code)</th>
<th>Attribute (Translated into English)</th>
<th>Field type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land housing block changing</td>
<td>1</td>
<td>YSQH</td>
<td>Previous land housing block code</td>
<td>integer</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>XQH</td>
<td>New land housing block code</td>
<td>integer</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>HBRQ</td>
<td>Incorporating time</td>
<td>date</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>HBYY</td>
<td>Incorporating reasons</td>
<td>string</td>
</tr>
</tbody>
</table>

4.2 Land Housing Block changing attribute list is as follows:

<table>
<thead>
<tr>
<th>Association</th>
<th>Sequence number</th>
<th>Attribute (Chinese code)</th>
<th>Attribute (Translated into English)</th>
<th>Field type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building change</td>
<td>1</td>
<td>ZBM</td>
<td>Building ID</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>YSQH</td>
<td>Previous land housing block ID</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>BGSJ</td>
<td>Changing time</td>
<td>date</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>BGYY</td>
<td>Changing reason</td>
<td>string</td>
</tr>
</tbody>
</table>

4.3 Building changing attribute list is as follows:
### 4.4 Attribute list of right of common:

<table>
<thead>
<tr>
<th>Association</th>
<th>Sequence number</th>
<th>Attribute (Chinese code)</th>
<th>Attribute (Translated into English)</th>
<th>Field type</th>
</tr>
</thead>
<tbody>
<tr>
<td>right of common</td>
<td>1</td>
<td>GYQBH</td>
<td>Right of common ID</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>GYQZSBH</td>
<td>Certificate No. of right of common</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>SYQBH</td>
<td>Property ID</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>SYRSFZ</td>
<td>Obligee ID</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>GYQL</td>
<td>Percentage for common right</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>QYQZT</td>
<td>Status of common right</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>......</td>
<td>......</td>
<td>string</td>
</tr>
</tbody>
</table>

### 4.5 Attribute list of other rights:

<table>
<thead>
<tr>
<th>Association</th>
<th>Sequence number</th>
<th>Attribute (Chinese code)</th>
<th>Attribute (Translated into English)</th>
<th>Field type</th>
</tr>
</thead>
<tbody>
<tr>
<td>other rights</td>
<td>1</td>
<td>TXQBH</td>
<td>Other right ID</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>TXQZSBH</td>
<td>No. of other rather certificate</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>TXQDJBH</td>
<td>Registration No. for other rights</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>TXQL</td>
<td>Types of other rights</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>JZ</td>
<td>Value</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>SYQ</td>
<td>Obligee</td>
<td>string</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>KSRQ</td>
<td>Start date</td>
<td>date</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>JESRQ</td>
<td>Expire date</td>
<td>date</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>TXQZT</td>
<td>Status of other right</td>
<td>string</td>
</tr>
</tbody>
</table>
Appendix 5: Final export schema in xml

```xml
<?xml version="1.0" encoding="UTF-8" ?>
<IntegratedDataOfCadastralAndHousing xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="C:/DOCUME~1/Ling/桌面/HOUSIN~1/Export-schema.xsd">
  <Person>
    <ID>612724196508060076</ID>
    <Name>Liu Xue</Name>
    <TEL>13536278849</TEL>
    <Email>xue@163.com</Email>
  </Person>
  <BirthDate>1965-08</BirthDate>
  <Gender>female</Gender>
  <BuildingID>012345600789</BuildingID>
  <Address>
    <Street>YanTa</Street>
    <City>XIAN</City>
    <Province>Shaanxi</Province>
  </Address>
  <ParcelID>40000090028</ParcelID>
  <Parcel>
    <ParcelID>40000090028</ParcelID>
    <ParcelCode>456983</ParcelCode>
  </Parcel>
  <ParcelLocation>
    <Street>YanTa</Street>
    <City>XIAN</City>
    <Province>Shaanxi</Province>
  </ParcelLocation>
  <Assessments>
    <Assessor>
      <ID>345959198109233348</ID>
      <Name>Xu Na</Name>
      <TEL>13044542354</TEL>
    </Assessor>
    <value>3564444</value>
  </Assessments>
  <Improvements>
    <Name>telecom company</Name>
    <Type>website</Type>
  </Improvements>
</IntegratedDataOfCadastralAndHousing>
```
<Street>HanGuang</Street>
<City>XIAN</City>
<Province>Shaanxi</Province>
</Address>
</Improvements>
<Area>YanTa</Area>
</Parcel>
- <Useright>
  <ParcelID>40000090028</ParcelID>
  <CertificateNumber>7693423</CertificateNumber>
  <CommencementDate>1968-06</CommencementDate>
  <ExpireDate>2038-06</ExpireDate>
  <TaxAmount>7883</TaxAmount>
- <Mortgage>
  <DocumentNumber>5497774</DocumentNumber>
- <Borrower>
  <ID>384928195903248819</ID>
  <Name>Song Jian</Name>
  <TEL>13988773934</TEL>
</Borrower>
- <Apartment>
  <Street>YanTa</Street>
  <City>Xi'an</City>
  <Province>Shann'Xi</Province>
</Apartment>
</Mortgage>
- <ProfitTaker>
  <ID>612324197809238898</ID>
  <NAME>Liu Li</NAME>
</ProfitTaker>
</Useright>
- <OwnerRight>
  <ID>768846532</ID>
  <BuildingNo>012345600789</BuildingNo>
  <ParcelNo>40000090028</ParcelNo>
- <Mortgage>
- <borrower>
  <Name>Ma Qiang</Name>
</borrower>
- <lender>
  <ID>612724196508060076</ID>
  <NAME>Liu Xue</NAME>
</lender>
</Mortgage>
<CommencementDate>2002-03</CommencementDate>
<FinishingDate>2006-09</FinishingDate>

<Floor>
  <ID>0089</ID>
  <ApartmentsAmount>28</ApartmentsAmount>
  <FloorMapNo>3498-09</FloorMapNo>
</Floor>

<Apartment>
  <ApartmentID>77834</ApartmentID>
  <LandHousingBlockCode>123456</LandHousingBlockCode>
  <BuildingCode>012345600789</BuildingCode>
</Apartment>

<Building>
  <BuildingNumber>012345600789</BuildingNumber>
  <DoorplateCode>126</DoorplateCode>
  <Density>89</Density>
  <AREA>789</AREA>
</Building>
</IntegratedDataOfCadastralAndHousing>
Appendix 6: Certificates

Certificates of land use ship and house ownership in Xiamen

Certificates of co-land use ship and house ownership in Xiamen