

A DISCUSSION ABOUT EFFECTIVE WAYS OF BASIC RESIDENT REGISTER ON GIS

Naoya Oku^{a,*}, Yasuaki Nonaka^a, Yutaka Ito^a

^a Kokusai kogyo Co., Ltd., 2-24-1, Harumi-cho, Fuchu-shi, Tokyo, 183-0057, JAPAN – (naoya_oku, yasuaki_nonaka, yutaka_ito)@kk-grp.jp

Commission VI, WG VI/4

KEY WORDS: GIS, Basic Resident Register, Address Dictionary, Geocoding, Address Points

ABSTRACT:

In Japan, each municipality keeps a database of every resident's name, address, gender and date of birth called the Basic Resident Register.

If the address information in the register is converted into coordinates by geocoding, it can be plotted as point data on a map.

This would enable prompt evacuation from disaster, analysis of distribution of residents, integrating statistics and so on. Further, it can be used for not only analysis of the current situation but also future planning.

However, the geographic information system (GIS) incorporating the Basic Resident Register is not widely used in Japan because of the following problems:

▪ Geocoding

In order to plot address point data, it is necessary to match the Basic Resident Register and the address dictionary by using the address as a key.

The information in the Basic Resident Register does not always match the actual addresses. As the register is based on applications made by residents, the information is prone to errors, such as incorrect Kanji characters.

▪ Security policy on personal information

In the register, the address of a resident is linked with his/her name and date of birth. If the information in the Basic Resident Register were to be leaked, it could be used for malicious purposes.

This paper proposes solutions to the above problems. The suitable solutions for the problems depend on the purpose of use, thus it is important that the purpose should be defined and a suitable way of the application for each purpose should be chosen. In this paper, we mainly focus on the specific purpose of use: to analyse the distribution of the residents.

We provide two solutions to improve the matching rate in geocoding. First, regarding errors in Kanji characters, a correction list of possible errors should be compiled in advance.

Second, some sort of analyses such as distribution of residents may not require exactly correct position for the address point.

Therefore we set the matching level in order: prefecture, city, town, city-block, house-code, house, and decided to accept up to city-block level for the matching.

Moreover, in terms of security policy on personal information, some part of information may not be needed for the distribution analysis. For example, the personal information like resident's name should be excluded from the attribute of address point in order to secure the safety operation of the system.

1. INTRODUCTION

1.1 Background

In Japan, each municipality manages its administrative information as per the Basic Resident Register, which includes each resident's name, address, gender, date of birth, and so on, as part of its efforts to support the lives of residents. The Basic Resident Register is managed as a database of digital information in text format and is used for tax accounting, welfare services, and other operations in administrative systems.

The Register contains each resident's address information, which can be converted into coordinate position information and plotted as an address point on a map (Figure 1).

Distributing the information as points in a space would be useful for analysing concentration, dispersion, and aggregation.

* Corresponding author.

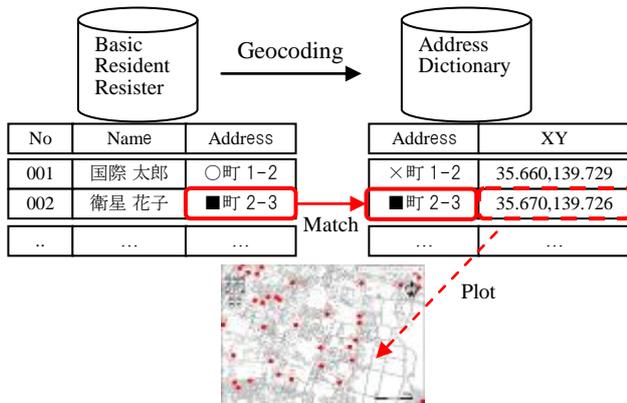


Figure 1. Method to be placed on the map the Basic Resident Register as Address Points

Furthermore, it would be useful not only for the analysis of the current regional situation, but for future planning.

Examples of such use of the Basic Resident Register are described below.

1.1.1 Grasping the number of people in need of care in disaster affected areas: Japan is subject to frequent destructive earthquakes as represented by the Great East Japan Earthquake. Recently, there have also been an increasing number of inundation events due to heavy rains as well as damage caused by sediment disasters; therefore, disaster control measures are urgently needed. For example, distributing positional information registered in the Basic Resident Register on a map would enable the following disaster control measure (Figure 2).

Since the Basic Resident Register contains the birth date information of the residents, their ages can easily be calculated. Figure 2 below indicates the distribution of elderly people aged 65 or over with red dots. The red frame enclosing the dots is drawn by using a function of the GIS to indicate the area in which a disaster has occurred. The GIS can extract the attributes of each point in the area, and therefore can confirm the safety of the residents represented by the points by using their names registered in the Basic Resident Register as a key.

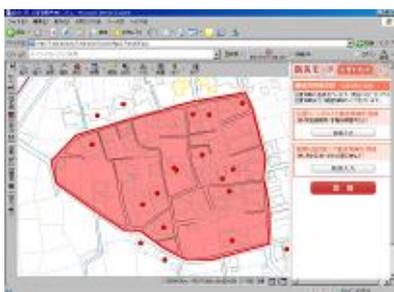


Figure 2. Grasping the number of people in need of care in disaster affected area

1.1.2 Grasping the number of preschool children by town: In Japan, there are children on waiting lists for child care-center with that is not available due to the population concentration in urban areas. This is a difficult problem requiring various measures, but described below is one example of a solution which maps and utilizes the information registered in the Basic Resident Register (Figure 3).

As with the case of elderly people mentioned above, the age of each resident can be calculated from their date of birth registered in the Basic Resident Register. Then, those aged three (3) or under, i.e., preschool children in Japan, can be identified and plotted on a map, showing where preschool children are concentrated. In reality, to solve the problem, it is also necessary to consider the wishes of the relevant residents and the distribution and capacity of child care-center, and therefore the problem cannot be solved quickly, but this solution is meaningful because it provides basic information for consideration.



Figure 3. Grasping the number of preschool children by town

1.2 Previous studies and focus of this study

1.2.1 Previous studies: There are previous studies on the use of the Basic Resident Register for GIS and discussions on the methods of address matching.

For example, APPLIC introduced the processing method, editing method, and application examples of an address dictionary as a guideline.

Shinichi, H acknowledged the effect of using the Basic Resident Register for GIS, but pointed out that its successful use depended on solving the problem of the geocoding matching rate and the difference between the actual residence and the Basic Resident Register, which is based on the applications made by the residents themselves.

The International Society for Photogrammetry and Remote Sensing (ISPRS) introduced the Greenland addressing system and described the method of geocoding.

1.2.2 Focus of this study: As described above, there are reports introducing the effect of using the Basic Resident Register for GIS, but few are based on actual examples of use. As mentioned in previous studies, this is because of the poor accuracy of the address information, which is the result of the address system of Japan, the time and cost required for editing, the security policy on personal information, and so on.

Among these problems, this study concretely discusses the relationship between geocoding and privacy, and proposes a solution.

We selected one city in Japan as a target area for geocoding. Since Basic Resident Register includes the privacy information, the city name should be kept as anonymous. However, the selected city is an average-scale core city with a population of approx. 200,000. Therefore the case described below can be regarded as an average case in Japan.

2. PROBLEMS

2.1 Geocoding

To convert the Basic Resident Register into positional information, it is necessary to match the address information registered in the Basic Resident Register with that in the Address Dictionary table, which associates address information with coordinate position information. However, since the Basic Resident Register is based on the applications made by the residents themselves, the information is prone to errors, such as incorrect or various styles of Kanji characters and inaccurate addresses; therefore, the matching rate is generally low.

2.2 Security policy on personal information

The Basic Resident Register contains not only the address, but also the name and birth date of each resident. Such information can be used to identify the age and place of residence of the relevant resident. Therefore, care must be taken to ensure not to be leaked. Considering such risks, many municipalities hesitate to introduce the Basic Resident Register with the GIS. Therefore, the method is not yet used nationwide.

3. METHOD OF RESOLVE

Our proposed solution to the problem is as follows.

3.1 Geocoding

3.1.1 Correspondence table: The address information registered in the Basic Resident Register is based on the applications made by the residents themselves. In Japanese notation, however, the same thing can sometimes be notated in different styles, for instance, in Kanji, Hiragana, and Katakana characters. When matching character strings, even a slight difference in notation is judged as a mismatch. Nevertheless, it is difficult to modify the application procedures immediately. Therefore, we decided to devise a way of geocoding, i.e., to make a correspondence table for absorbing the gap of the notation of addresses (Figure 4).

In case of number	
0, 0, 〇, 零	=> 0
1, 1, 一, 壹	=> 1
In case of house-code	
2 - 2, 2-2, 2 番地の 2	=> 2-2
In case of hyphen	
-, ~, -, -	=> -

Figure 4. Correspondence table

3.1.2 Matching level: Even after absorbing the gap of the notations, inaccurate addresses are still judged as a mismatch. Also, wrong numbers cannot be completely absorbed by a correspondence table.

Japanese address is composed by prefecture, city, town, city block, house-code level, as well as the other countries. However the address notation in Japan is written in order from larger to smaller-scale level: prefecture, city, city block, house-code, and the resident name (Table 1).

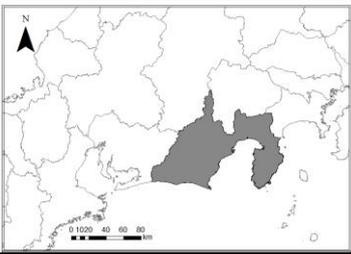
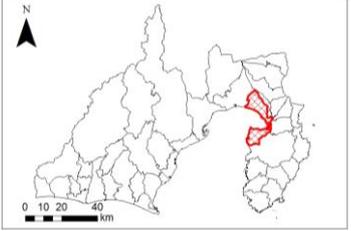
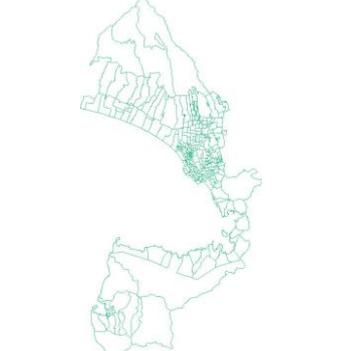
Level	Capture	Address notation
prefecture		●● 県
city		▲▲ 市
town		○○ 町
city-block		16
house-code		16-3
house		The name of building
●● 県 ▲▲ 市 ○○ 町 16-3 ●● 県 ▲▲ 市 ○○ 町 16 -3 prefecture city town city-block house-code		

Table 1. Address structure in Japan

A solution to the problem is to match the address information by “begins-with” matching, i.e., provide matching levels of city, town, city-block, and house-code, and record the level at which the address information matches the coordinate position information.

If the address information matches at the house-code or house level, it means that the address information completely matches the coordinate position information, and therefore the address point matches the center of the house on the map. The house is almost that of the relevant resident.

On the other hand, if the address information matches the coordinate position information only to the city-block level, the address point is plotted at the center of the city-block. In this case, the address point does not always represent the place of the resident. This also applies to the town levels (Table 2).

Adopting a complete match, the matching rate becomes low resulting in a decreased number of cases registrable with the GIS. Therefore, we decided to adopt a partial match in order to maximize the cases registrable with the GIS. Of course there are some cases which it is necessary to match address points at the house level, but in other cases, it is desirable to increase the number of dots on the map to maximize the accuracy of aggregation. This study adopts such a procedure.

Matching Level	Address notation
house-code	●●県▲▲市○○町2-1/
city-block	●●県▲▲市○○町2/- 1
town	●●県▲▲市○○町/2-1
City	●●県▲▲市/○○町2-1
/ separate Address into each a matching level	

Table 2. Matching level

3.1.3 Security policy on personal information: Security is in a trade-off relationship with necessity. The more detailed the information required, the higher the security risk. Therefore, it is desirable to clarify the purpose and extract only the minimum necessary information to avoid unnecessary security risks. The information to be extracted from the Basic Resident Register is classified into the following three levels in accordance with the purpose of geocoding.

3.1.4 Identification of individuals: If the address point of each resident is located on the center of a house on a map and the resident's name and date of birth are accessible on the Basic Resident Register, it is possible to identify who lives where at the individual level. Such information is very useful for confirming safety in the event of a disaster (Figure 2).

However, if such information is leaked and used for malicious purposes, the resident could be exposed to criminal activity.

3.1.5 Aggregation of distribution of residents: If the address point of each resident is located on the center of a house on a map but the resident's name is not accessible on the Basic Resident Register, it cannot be used for purposes that require individual identification, but the aggregation of residents by age structure is possible by using the spatial aggregation function of GIS. In this case, it is not necessary to extract the names from the Basic Resident Register, which is beneficial for security.

3.1.6 Displaying aggregation results by town and city-block: Almost all the cases registered in the Basic Resident Register match the coordinate position information at the town or city-block level (see section 4 for details).

In this case, all that needs to be done is to put the address point at the center of the town or city-block and to enter the aggregate by age structure as the attribute of the polygon data of the town or city-block. This is highly secure since address points are not plotted for individuals, but it is of limited use because it only reflects the aggregation results and therefore cannot be used for detailed analysis using the spatial aggregation function of GIS or visualization of distribution.

4. RESULTS AND DISCUSSION

The solution proposed above concerning the balance between geocoding and security can be summarized as follows:

- Geocoding
Carry out matching at each level and accept the level suitable for the purpose of geocoding.
- Security
Extract and display only the information necessary for the purpose of geocoding.

The results of matching are as follows (Table 3).

At the house-code level, the number of matched cases is 153,705 and the matching rate is 76.45%, which means that only three-fourths of all the cases completely match. This is not sufficient for the purpose of geocoding. At the city-block level, the number of matched cases increases by 17,032 and the total matching rate is 84.92%. Furthermore, at the town level, the number of matched cases increases by 30,278 and the total matching rate is 99.98%. Thus, almost all the cases match the coordinate position information.

Matching Level	The number of matched cases	The matching rate	The total matching rate
house-code	153,705	76.45%	76.45%
city-block	17,032	8.47%	84.92%
town	30,278	15.06%	99.98%
city	42	0.02%	100%

Table 3. Matching rate

On the other hand, there still remains a problem with reliability of matching at the city-block and town levels. To discuss this problem, Table4: shows an example (Table 4).

At the city-block level, address information is plotted on the center of a city-block, which is assumed to match a house in the relevant city-block. If the size of the city-block is several meters and spatial aggregation is not carried out across the city-block, it is judged acceptable.

On the other hand, it is difficult to judge the reliability of matching at the town level. As shown in (Table 3), the matching rate differs by approx. 15% depending on whether the town level is included or not. The size of a town varies from town to town. For small towns, spatial aggregation can be performed covering the center of the town, but for large towns, there may be a case which the center of the town is not covered by spatial aggregation resulting in inaccurate analysis. Therefore, considering the size of the town at the time of matching may help improve the reliability of matching. This is an issue for this study.

Level	Legend	Capture
house -code		
city- block		
town		
city		

Table 4. Mapping after matching (at each level)

5. CONCLUSION AND FUTURE WORKS

In the previous section, we proposed solutions to two problems. Each solution is only effective within an acceptable level depending on the purpose of geocoding. This study does not deal with solutions to the root of the problems, but clarifies the problems and proposes alternative solutions.

There are many methods of geocoding other than those introduced in this paper. Further studies are expected.

Purpose of use	Accuracy level of matching	Necessary information (for Basic Resident Resister)	Security risk
identify individual (for prompt)	house code	name, address,	High

evacuation from (disaster)	level	gender, date of birth	
analysis of distribution (of residents)	city block level	address, gender, date of birth	Middle
integrating statistics	no level (not plot)	gender, date of birth	Low

Table 5. Suitable ways chosen for each purpose of use

In Japan, the "My Number" system was implemented in November 2015. At present, however, there is much concern over the possible leakage of personal information. There will also be increasing concerns about GIS incorporating the Basic Resident Register. Therefore, it is necessary to continuously consider the balance between security and convenience.

REFERENCES

The Association for Promotion of Public Local Information and Communication., 2011. Guide and Application Examples of Address Dictionary maintenance, Japan, <http://www.applic.or.jp/2013/tech/APPLIC-0002-2011-07.pdf>

Shinichi, H., 2014, Benefit to convert Basic Resident Register as geographic information, Report on the 66th Conference of Population Association of Japan, Tokyo, Japan, <http://paj66th.web.fc2.com/abs/sp4.pdf>

Janis, S., 2011, Towards Introducing A Geocoding Information System For Greenland, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXVIII-4/W25, 2011 ISPRS Guilin 2011 Workshop, 20-21 October 2011, Guilin, China*