HEALTH GIS AS AN APPLICATION OF AREA STUDIES

Shoichiro HARA

Center for Integrated Area Studies, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan
Email: shara@cias.kyoto-u.ac.jp

ABSTRACT

Area Study is one of the interdisciplinary sciences consisting of many disciplines to understand distinctive areas comprehensively. "Area Informatics" is the new paradigm to integrate individual disciplines of area studies and to create new knowledge on areas/countries using information science/engineering technologies. Individual health condition is the reflection of multi-factors of natural and socio-economical environments, and Area Informatics will be the potential paradigm to store, integrate and analyze variety kinds of health information. This paper will describe overviews on application of Area Informatics to public health.

1. INTRODUCTION

Area study is conceptually one of the interdisciplinary sciences consisting of many disciplines (e.g., ecology, environmentology, forestry, agriculture, medicine, health, geography, economics, literature, folklore, ethnology, history and sustainability etc.) to understand a distinct area/country comprehensively. However, due to differences of languages, terminologies and conceptions among disciplines, each area study researcher has been actually engaged in a rather narrow theme such as Thai agriculture, Vietnam Buddhism and so on. "Integrated Area Research" is a new approach to area study to pursue comparative studies on themes that are relevant across distinct areas using innovative integrated research procedures. The emphasis is on interregional relationships and variables and how they link with area-specific characteristics (CIAS 2008).

Correspondingly, "Area Informatics" is the new paradigm of information science for area studies. It will integrate various disciplines of area studies and create new knowledge on areas/countries by making good use of information science/engineering technologies such as metadata including spatiotemporal attributes, databases, ontology, geographical information systems/science (GIS), global positioning system (GPS), remote sensing (RS), geostatistics, data mining and so on. Individual health condition is the reflection of multi-factors (e.g., not only personal constitution but also accumulations of long term effects from natural and social-economical environments), and then many variables from various disciples should be collected, integrated and processed. Therefore, Area Informatics will be the potential paradigm for public health.

This paper will describe the application of Area Informatics to public health. First, "Area Informatics Model" will be defined and some spatiotemporal tools based on the model will be introduced in chapter 2. In chapter 3, an overview of health GIS in the context of area studies and importance of comparative approaches will be explained. Finally, some examples and problems for health GIS will be discussed in chapter 4.

2. OVERVIEW OF AREA INFORMATICS

Area Informatics is versatile model and it is applicable to public health. This chapter will define the notion of Area Informatics and introduce some spatiotemporal tools.

2.1 Area Informatics Model

Area Informatics can be modeled from the view of data, information and knowledge flows as shown in Figure 1 (Hara 2007). Every area study begins with "Field Research" to collect data and materials (resources) by interviewing persons, visiting places, observing events and searching for private/official collection holders in the area. Collected resources are organized according to "Model of Area Studies," that is, resources are digitized and encoded depending on their subjects (e.g., primary disciplines, contents or indexes), data types (e.g., texts, numeric, images, sounds, qualitative or quantitative), storage media, and collection types (relations with other resources). Organized resources are complied into "Databases" that will be opened on the Web. However, as each database is built on its own platform and has specific data structure and retrieval method, it is not effective for researchers to search many databases on the Web. "Resource Sharing System" is the new information system to integrate various databases on the Web and to provide researchers with a uniform interface to retrieve databases seamlessly by one operation (Hara 2002).

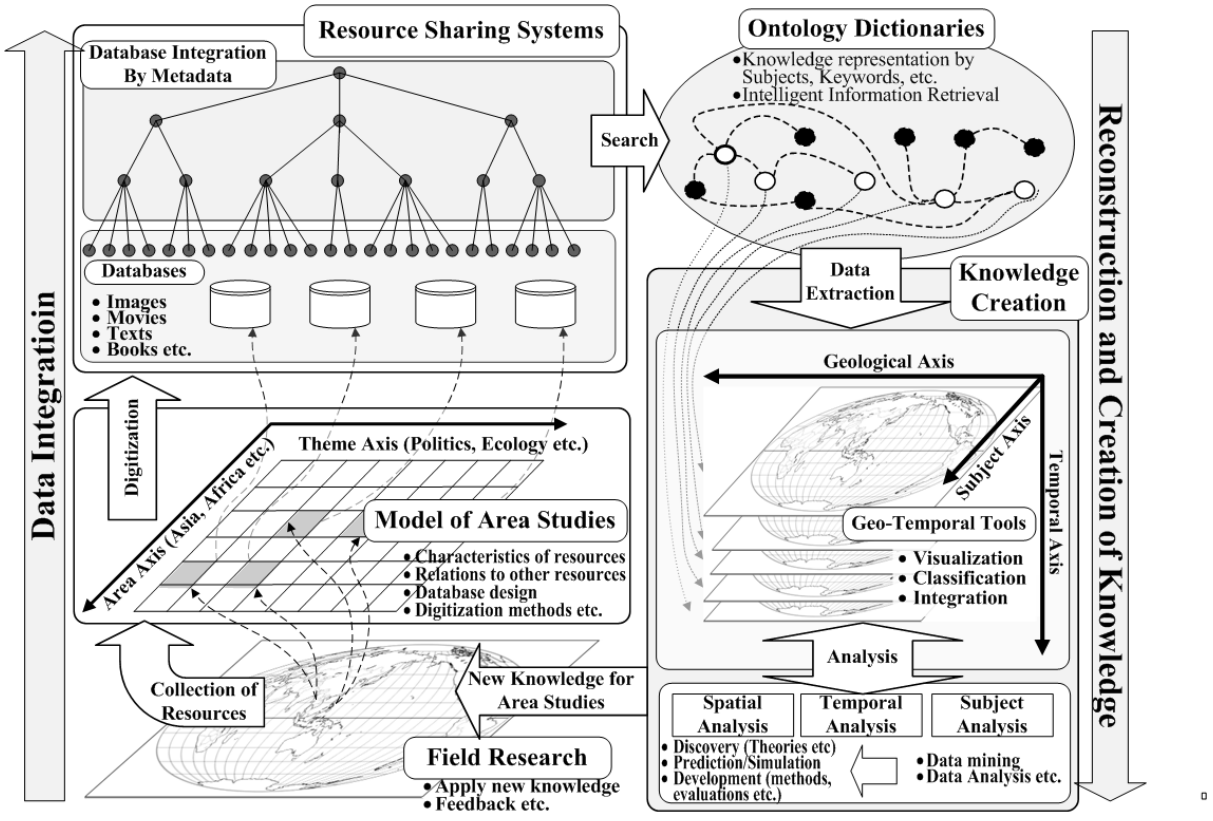

Figure 1. Area Informatics Model

By the way, as each database encodes data depending on specific language and terminology, Resource Sharing System is not enough to search data across disciplines. "Ontology Dictionaries" are introduced to solve this problem. Ontology Dictionary contains terms (vocabulary) and their attributes to relate terms depending on their meanings (synonym, antonym), hierarchical conceptions (taxonomy) and so on, that is, they can change data into

information. Resources Sharing System and Ontology Dictionaries provide advanced information accessing /filtering methods.

"Knowledge Creation" is the final phase in the Model. All information is integrated, classified, visualized, compared and analyzed to create new knowledge. Spatiotemporal attributes are crucial part of information, and GIS, geostatistics and data mining techniques will be applied to these attributes. Results of this phase will be fed back to field as new knowledge, and new research process starts.

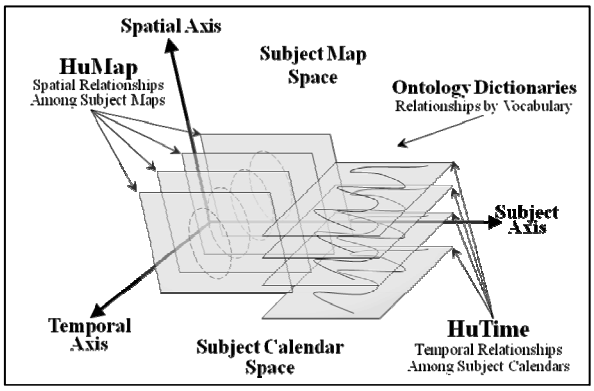

Figure 2. Spatiotemporal Model

2.2 Spatiotemporal Tools

In Area Informatics Model, all data objects are described as a distribution in the 3D space created by three axes as shown in Figure 2 (Sekino 2007). The subject axis arranges data objects by vocabulary (semantics), and it is suitable to catalogues. The spatial axis arranges data objects by locations, and it is suitable to maps. The temporal axis arranges data objects by time, and it is suitable to the chronological table. We are developing three spatiotemporal tools based on the 3D model.

2.2.1 *HuMap (Humanities Map)*

HuMap is a newly developed GIS system that handles data objects on the planes created by the subjects and spatial axes (i.e., some subject maps). It overlays many maps and images (layers), and we can easily grasp distributions and relations of events from different layers by locations. It can also retrieve features on layers, find and import necessary layers from the resource sharing database, and carry out some arithmetic and logical operations on features among layers (Figure 3). For example, we will find the geographical relationships between patients' distribution and land uses using HuMap.

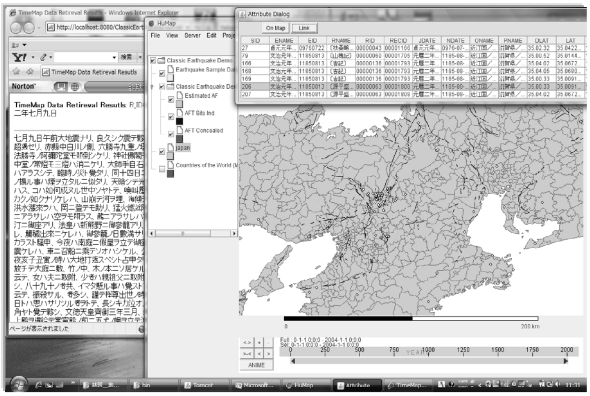

Figure 3. HuMap

2.2.2 *HuTime (Humanities Time)*

HuTime is a evolutionally temporal oriented system that handles data objects on the planes created by the subjects and temporal axes (i.e., some subject calendars). It arranges many calendars according to a timeline, and we can glance over many events of different calendars simultaneously. It can also retrieve events on layers, find and import necessary layers from the resource sharing database, and carry out some arithmetic and logical operations on events among layers (Figure 4). For example, we can find a time order of disease prevalence among areas using HuTime.

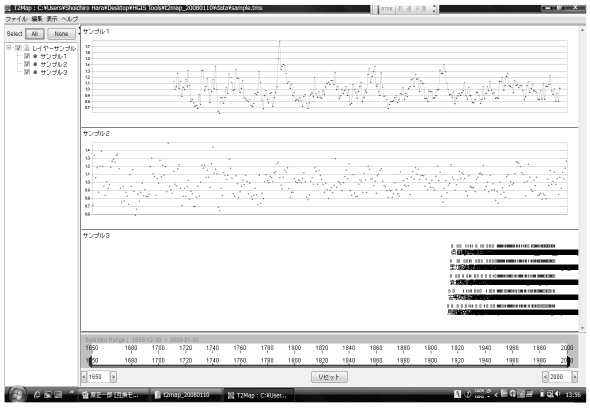

Figure 4. HuTime

2.2.3 *Ontology Databases*

Two ontology databases are created to support spatiotemporal data organization. The one is "Japanese Historical Gazetteer Database." This database includes place names of present and past with their attributes (pronunciation of a place name, its location types, and the broader place name, etc.) and locations (longitudes and latitudes). It is used to convert a place name into the pair of longitude and latitude, and to select a correct place from others that have the same name and so on. The second is "Calendar Table." This table organizes all dates (Japanese dates, Gregorian dates etc.) according to Julian dates. It is used to convert a date of a calendar to the date of another calendar.

3. COMPARATIVE APPROACH TO HEALTH GIS

Development and prevalence of diseases are multifactorial. Even infectious diseases whose etiologies are relatively simple are subject to not only the relationship between personal constitution and vectors but also relationship between physical, biological and social-economical environments. If there is a pathogen in an area, the disease might propagate rapidly. But if the area is well developed and/or its sanitary condition is good, the propagation might be suppressed. Actually, many epidemiological histories suggest the importance of environments (i.e., SNOW 1855.) In developed countries, infectious diseases are not crucial because water supply and sewerage systems are fully equipped, some vectors in residential, forest and agricultural areas are almost controlled by pesticides, and people's nutritional status is well. These countries achieved their well statuses by consuming immeasurable amount of resources to improve/reform environments. However, we gradually recognize that this traditional development model is no longer possible because it is unsustainable development process, wastes much energy and resources and puts heavy burdens to future generations. Then, if we will carry on heath promotion, especially in under developing countries, we have to take into account environments and sustainability.

There are many research papers on health GIS that are concerned with spatiotemporal relationships between environments and infectious diseases (e.g., Tripathi 2006, Jeefoo 2008 and Chaikaew 2008.) These papers choose variables of physical environments such as climate, land use and elevation, and are targeted at relatively narrow areas. Then these papers build numerical models to estimate risk zones and to predict number of patients, and attain some progresses. However, as these models use only physical environment variables in narrow areas, it is difficult to apply them across areas and times, that is, these models are not suitable to comparative studies. Some papers mention the importance of socio-economical variables, but their models do not include these variables (e.g., Razeghi 2008.) Other papers indicate that non-climate factors are necessary for auto-regression (time-series) models to predicate the scale of epidemics (number of patients) precisely (e.g., Matsuda 2008).

These imply that if we want to pursue sustainable health development, we have to learn many matters from development histories, that is, we have to compare many facts/events across times and areas, and to derive new knowledge. This is the same paradigm of area studies that pursue comparative studies on themes across areas emphasizing on interregional relationships and variables and how they link with area-specific characteristics. From Area Informatics points of view, this will correspond to collecting various resources, organizing them into databases, comparing/analyzing data to select variables that can explain common and different development/propagation of diseases across times and areas. Then, Area Informatics Model can be applicable to the research on health and sustainable development.

4. EXAMPLES OF AREA INFROTMATICS APPROACH TO HEALTH GIS

For sustainable health development, we have to know not only present status but also our development histories. This is important to predict the future directions of socio-economical environment changes such as population compositions, community behaviors, sanitary conditions, nutrition conditions, environmental pollutions and so on. Following two sections show the examples of reconstructing past conditions from historical data.

4.1 Change of Infant Mortality by Smallpox in Suburban Tokyo in 19th Century

Japan and East Asian countries accumulate huge numbers of historical documents and materials, and they are valuable resources to reconstruct past status/environments to some extent. Kawaguchi estimated the change of infant prevalence and mortality by smallpox at a suburban village in western Tokyo area in 19th century using diaries written by a Shinto priest (Kawaguchi 2008). Figure 5 shows infant prevalence and mortality by smallpox from 1835 to 1870. Bovine vaccine was introduced into the village in 1852, and prevalence and mortality were subsequently reduced. Kawaguchi investigated diaries carefully, and concluded that this rapid reduction is the result of vaccine.

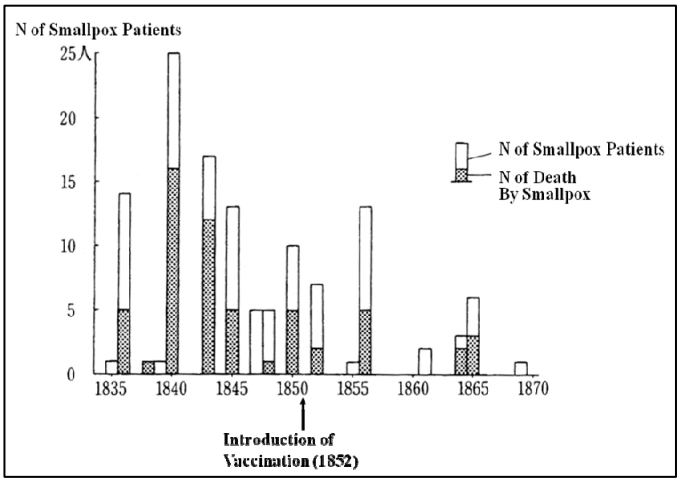

Figure 5. Estimation of Infant Prevalence and Mortality by Smallpox (Kawaguchi 2008)

4.2 Change of Water Quality in 20th Century Tokyo

Many literary works describe natural and artificial landscapes and living creatures around areas. These literal works are valuable resources to estimate environment changes quantitatively. Taniguchi examined many Japanese novels written by great writers in early 20th century and extracted words related to hydrological landscapes and water creatures. He also developed a water quality rating table that was used to convert qualitative expressions of water qualities (words in literary works) to semi-quantitative values. Then he estimated the changes of water qualities of main rivers in Tokyo (Taniguchi 2008). Figure 6 shows water quality change at the several points along Sumida River from 1900 to 2000. According to this figure, Sumida River constantly increased in pollution from 1900 to 1970. But there are two periods when the pollution was improved, i.e., around 1925 when Tokyo was hit by the Great Kanto Earthquake, and around 1945 when Tokyo was destroyed by air arracks, and economic activities declined during these periods. Thus this figure reconstructs the change of water quality very well.

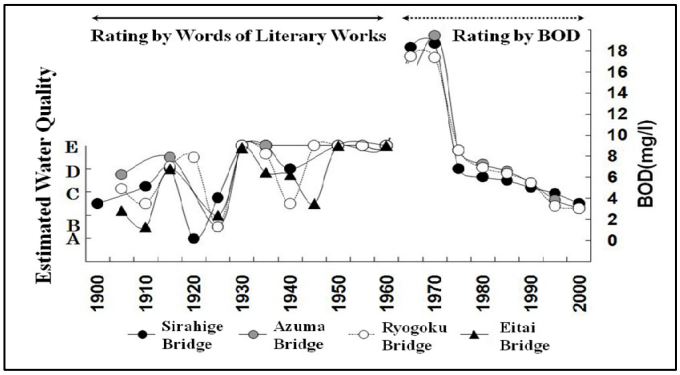

Figure 6. Estimation of Water Quality (Taniguchi 2008)

5. DISCUSSIONS AND CONCLUSIONS

Notions and availabilities of Area Informatics to health GIS were described. Researches according to this new paradigm have just begun, i.e., three basic spatiotemporal tools have been built (HuMap, HuTime and Digital Gazetteer), and several studies to reconstruct past environments (water pollution) and disease images (malaria, schistosomiasis japonica and so on) using historical materials are going on. As for an application of Area Informatics to current public health, we are trying to set some research fields in Southeast Asian countries and Japan. We will collect/compile data sets that include socio-economical attributes and can be comparative across the fields, then push forward comparative studies by applying many models to these data sets.

6. ACKNOWLEDGMENT

This research has been operated by H-GIS (Humanities GIS) research group, and special thanks to Dr. Kawaguchi and Dr. Taniguchi for providing Figures 5 and 6.

7. REFERENCES

Chaikaew, N., Tripathi, N. K., Souris, M., 2008. Point Pattern analysis of Malaria Incidence in Chiang Mai, Thailand. *Proceedings of Second International Conference on Health GIS.* 137-140.

CIAS, 2008. The Center for Integrated Area Studies. http://www.cias.kyoto-u.ac.jp/index.php/english.

Hara, S., Yasunaga H., 2002. Resource Sharing System for Humanity Researches. *Proceedings of the Third International Conference on Language Resources and Evaluation,* 51-58.

Hara, S., Shibayama, M., 2007. Development of Area Informatics and Spatiotemporal Analysis Tools (in Japanese). *Proceeding of the 2007 Symposium on Computers and the Humanities (JINMONKON 2007)*, 71-78.

Jeefoo, P., Tripathi, N. K., Souris, M., 2008. Exploring Geospatial Factors Contribution to Malaria Prevalence in Kanchanaburi, Thailand. *Proceedings of Second International Conference on Health GIS.* 130-133.

Kawaguchi, H., 2008. Necessity of Spatiotemporal Information System to Understand Mortality Structure during Transitional Period to Modern Age (in Japanese). *Intriguing ASIA*, No.113, 36-43.

Matsuda. F., Ishimura, S., Wagatsuma, Y., Higashi, T., Hayashi, T., Faruque, A. S. G., Sack, D. A., Nishibuchi, M. 2008. Prediction of epidemic cholera due to Vibrio cholerae O1 in children younger than 10 years using climate data in Bangladesh. *Epidemiology and Infection,* Volume 136, Issue 01, January 2008, 73-79.

Razeghi S. M., Sadrabadi, E. S., Dizabadi, F. B., Zadeh, A. SM, 2008. Prediction of Malaria Incidence using GIS. *Proceedings of Second International Conference on Health GIS.*17-21.

Sekino, T., Kubo, M., 2007. T2Map – A tool for temporal information system (in Japanese). *Proceeding of the 2007 Symposium on Computers and the Humanities (JINMONKON 2007),* 183-188.

Snow, J., 1855. On the Mode of Communication of Cholera. *London, England, John Churchill.*

Taniguchi, T., 2008. River Environment in Tokyo described in Modern Literal Works (in Japanese). *Intriguing ASIA*, No.113, 60-67.

Tripathi, N. K., Nakhapakorn K., 2005. Finding Spatial Relationship of Environmental and Climate factors with Dengue Epidemic using GIS. Program of *First International Symposium on "Area Informatics 2005: The Potential of GIS/RS in Area Studies."*