Advanced Architecture of Spatial Data Infrastructure in Republic of Korea

GONIL RHO, KIYUN YU
Department of Civil & Environmental Engineering
Seoul National University
599, Gwanak-ro, Gwanak-gu, Seoul
Republic of KOREA
ginho03@snu.ac.kr, kiyun@snu.ac.kr

Abstract: - There are two major systems for national spatial data infrastructure in Republic of Korea; National Spatial Information Clearinghouse (NSIC) and National Spatial Data Infrastructure (NSDI). We founded some limitations like redundant data construction, limited data access and search ability, impossible to integrate two dataset for interoperability from those two systems. It is proposed that new architecture of spatial data infrastructure after analyzing advanced spatial infra structures in the world. The proposed architecture has some advantages compared to current one; dynamic evaluation mechanism, interoperability of spatial datasets, open interface for distributing channel.

Key-Words: - spatial data infrastructure, architecture design, dynamic evaluation mechanism, interoperability, open standard

1. Introduction
Spatial data represents information about natural and artificial space around us. Because most of the events in the real world are related with space, we need reliable spatial datasets to analyze and understand what happens around us. From 1996 to 2010, government of Republic of Korea has implemented a National Geographic Information System Project. As a result of the project, a large amount of high-quality spatial datasets were constructed. Now, it is registered about one hundred thousand dataset (about one million maps) from 165 themes in the national spatial data catalogue.

It requires a lot of times and costs to construct spatial dataset. But spatial dataset can be reused multiple times for various purposes after once constructed. The government of each country in the world already realized this characteristic of spatial dataset as public goods, and they have developed national spatial data infrastructure to share spatial dataset in the national level [1] [2]. Also, government of Republic of Korea has implemented its own systems for managing national spatial dataset. They aim to promote sharing and reusing of spatial dataset in the national level. But there are still some limitations like redundant data construction, limited data access and search ability and impossibility to integrate two dataset for interoperability.

In this study, we are going to investigate the situation of the spatial data infrastructure in Republic of Korea and point some limitations out. Then we want to propose an improved architecture of spatial data infrastructure based on analysis of advanced systems which are being developed in the world.

2. Spatial data infrastructure in Republic of Korea
There are two major systems for national spatial data infrastructure in Republic of Korea; National Spatial Information Clearinghouse (NSIC) and National Spatial Data Infrastructure (NSDI). NSIC is a system which collects spatial data and provides online searching and downloading functionalities to the users. The purpose of NSIC is to distribute spatial dataset to the public area and encourage usage of national spatial datasets [3]. NSDI is a system that integrates spatial dataset from multiple governmental agencies and provides spatial data service framework to support administrative duties of local governments [4]. The major functionalities of two systems are summed up in the table 1.

NSIC has collected most of spatial dataset that is constructed in the public domain, but it does not evaluate and organize datasets. The users may access to NSIC and try to know which dataset is fit for his/her purpose. But NSIC does not provide any criteria of selecting proper dataset to the users. Also, dataset like topographic map, cadastral map and forest map which is collected from various...
providers have different data scheme, format, spatial reference system and spatial management unit from each other. Because each of them represent unique point of view of the real world, datasets have to be integrated with each other to analyze real world phenomenon. But NSIC does not provide any functionality for integrating datasets. On the other hand, NSDI provides integrated fundamental dataset to the user. But the range of dataset is limited only to small part of national spatial datasets; topographic map, cadastral map, Arirang satellite image. Besides, purpose of implementing NSDI is limited to the supporting administrative duties of local governments. Users in the public and commercial area cannot access to the dataset in the NSDI directly.

3. Analysis on the recent development of spatial data infrastructures

Geospatial Platform has been developed from November 2010 to offer access to a suite of geospatial assets including data, service, applications, and infrastructure in the USA [5]. Its main purposes are promoting sharing of geospatial assets, enhancing interoperability and ensuring accountability and transparency of geospatial and related investments through portfolio management. In this stage, they implemented prototyped platform and define National Geospatial Data Asset under portfolio management system.

As a conceptual model, Location Information Infrastructure (UKLII) is an effective and efficient environment for the re-use of location information in the U.K [6]. It covers the discovery of what data exists, the creation of foundational, coordinated core geographic reference datasets, dataset interoperability, data publishing and data sharing.

Government of Japan has developed Digital Japan which is a concept of virtual geo-space where users can easily utilize geospatial information by selecting from datasets constructed by different organizations and integrating them for their purpose [7]. And they define Fundamental Geospatial Data (FGD) to implement Digital Japan. FGD is a reference spatial dataset composed of control point of surveying, shore line, public facility, administrative boundary, road, river, railway, DEM, building and etc. They developed a strategic plan named ‘Grand Design for Fundamental Geospatial Data’ to maintain and update FGD systematically.

There are lessons learned from the recent development of spatial data infrastructure in the world.

First, they focus on selecting and managing list of fundamental spatial dataset. Portfolio management system of USA is the result of adapting a business centric perspective to the management of spatial datasets. It maps the relation between consumer and provider of spatial dataset to that of business partners. Fundamental spatial datasets can be intensively managed and the requirements of the users can be easily applied in the portfolio management system, contrast to the voluntary and temporary management system from the past. Japan developed centralized management system of fundamental spatial dataset. They built and distributed reference spatial dataset (FGD) to prepare integrated usage of spatial information.

Second, they emphasize the importance of transformation and integration between spatial datasets. They try to implement functionality more than simply collect and accumulate spatial data to the database. One of major goal of Geospatial Platform is developing some tools to integrating and visualizing spatial datasets from multiple agencies. A tool which can transform coordinate and data
scheme from one model into another is defined as a basic functionality of UKLII for the integration of spatial datasets.

Third, they follow open standards to improve interoperability of spatial data and service. Geospatial Platform and UKLII actively adopted open standard from Open Geospatial Consortium (OGC) like Web Map Service (WMS) and Web Feature Service (WFS) for the open interface of spatial data infrastructure.

4. Architecture for advanced spatial data infrastructure in Republic of Korea

The advanced architecture of spatial data infrastructure in Republic of Korea can be suggested from lessons learned (figure 1). It is composed of two channels (spatial data collecting and distributing channel) and four modules (evaluation, catalogue management, transformation, and matching). And it stores three types of data: metadata, standardized spatial data, linked spatial data.

The new architecture has three major improvements compared to the current one.

The first one is a dynamic evaluation mechanism composed of four activities: collecting, evaluating, managing and feedback. There are two opposing approaches to collect and manage spatial data in the current architecture. NSIC collects as many as possible datasets, but there is no management system to evaluate the priority of each dataset. It can be difficult to choose the right dataset for their application to the users. On the other hand, NSDI has built a fundamental dataset, but its range is limited to specific area. The fundamental dataset of NSDI can’t satisfy the diverse needs of users today. The dynamic evaluation mechanism is an alternative of current approaches. The mechanism is composed of cyclic four-step activities. 1) At first, as many as spatial datasets are collected from data sources by Spatial Data Collecting Channel. 2) Then collected datasets are evaluated in the Evaluating Module. The Evaluation Module categorizes datasets by grade: fundamental dataset, major dataset, and additional dataset. 3) After evaluation, the Catalogue Management Module updates the metadata of each dataset and stores it into Metadata DB. 4) A feedback which is the most important activity in the mechanism is started from Spatial Data Distributing Channel. The feedback information created based on distribution statistics of spatial datasets. The information goes to the Evaluation Module and has an effect on the grading criteria of datasets dynamically.

An improvement for interoperability of spatial datasets is another advantage of new architecture. Spatial data is an abstract model of information about real world space. There may be some differences between models of datasets, and it can be divided into syntactic difference and semantic difference. These differences are the biggest obstacle in achieving high-level interoperability between spatial datasets. The syntactic difference comes from different spatial reference system and data format. The Transformation Module provides the functionality for transforming spatial reference system and data format into another one. In addition, standard for spatial reference system and data format has to be defined. The survey-law of Korea defines ITRF2000 as a standard spatial reference system for national official surveying. Geography Markup Language (GML) is a standard data format for spatial domain from OGC. ITRF2000 and GML can be a good starting point of standard syntax of spatial dataset in advanced SDI. The semantic difference comes from different perspective of view to the same spatial entity in the real world. The semantic difference can be solved by feature-level
matching process. The Matching Module can be implemented based on matching algorithms that have been suggested by many researchers. After matching process is done, connection information between features can be stored in Resource Description Framework (RDF) format. RDF is graph-based data model for Linked Data from the semantic web technology. Relations between objects can be stored with object itself at the same place using RDF. Recently, web applications using spatial data has increased. RDF-based Linked Spatial Data which is composed of spatial features and relations between themselves can be easily consumed by web-based applications because RDF is a standard format for the web technology.

The last improvement is an open interface for distributing spatial datasets. Spatial Data Distributing Channel is a component that open standards have to be actively adopted. The open interface using open standards is the best way to provide the functionalities and data to all possible users. OGC defines some open standards for spatial web service [8]. Catalogue Service for the Web (CSW) is interface to discover and query metadata about data and service. Web Map Service (WMS) and Web Feature Service (WFS) are interface to retrieve maps and features respectively. Also, there is SPARQL Endpoint which is standard query interface for Linked Data defined by World Wide Web Consortium (W3C). It is a semantic enablement approaches for spatial data infrastructure that implementing Linked Data and SPARQL Endpoint [9].

5. Conclusion
This study investigated the limitations of current spatial data infrastructure in Korea, and analyzed three lessons from advanced spatial data infrastructures in the world. Then new architecture of spatial data infrastructure was designed based on the result of analysis. It has some advantages compared to current one; dynamic evaluation mechanism, interoperability of spatial datasets, open interface for distributing channel.

Recently, location based big data is collected from smart sensors like GPS embedded mobile device (e.g. smart phone), traffic sensor, CCTV and etc. This kind of data can be analyzed based on reliable and qualified spatial datasets to create high value-added information about human activity and environment surrounding us. This is the one of reasons that advanced spatial data infrastructure become more important.

Discussions in this study are limited to technical point of view for advanced spatial data infrastructure. In the future study, it is necessary to discuss about the way to improve spatial data infrastructure in political and organizational perspective.

References:
[6] Rod Kedge, Integrating the UK Location Information Infrastructure and data.gov.uk, UK Location Programme, 2011