

# Overcoming Geoinformatic Knowledge Fence: An Exploratory of Intelligent Geospatial Data Preparation within Spatial Analysis

Jian Wang<sup>1</sup>, Chun-jiang Zhao<sup>1,\*</sup>, Fang-qu Niu<sup>2</sup>, and Zhi-qiang Wang<sup>2</sup>

<sup>1</sup> National Engineering Research Center for Information Technology in Agriculture (NERCITA), Beijing 100097, China

Zhaocj@nercita.org.cn

<sup>2</sup> Institute of Geography Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100060 China

Wangj@nercita.org.cn

**Abstract.** The booming of earth observation provides decision-makers with more available geospatial data as well as more puzzles about how to understand, evaluate, search, process, and utilize those overwhelming resources. The paper distinguishes a concept termed geoinformatic knowledge fence (GeoKF) to discuss the knowledge-aspect of such puzzles and the approach to overcoming them. Basing on analysis of the gap between common geography sense and geoinformatic professional knowledge, the approach composes analysis space modeling and spatial reasoning to match decision models to the online geospatial data sources they need. Such approach enables automatically and intelligently searching of suitable geospatial data resources and calculating their suitability to given spatial decision and analysis. An experiment with geoservices, geo-ontology and rule-based reasoning (Jess) is developed to illustrate the feasibility of the approach in scenario of data preparation within decisions of bird flu control.

**Keywords:** spatial decision support, geo-ontology, spatial reasoning, geoservices, geography information metadata.

## 1 Introduction

With booming of earth observation and other earth-related activities, there is a dramatically increase of types and volume of available geospatial data in form of online data set and services. Compacting with such overwhelming data, users, esp. those without enough geoinformatic knowledge (termed here as n-geo users), find it more difficult to understand and utilize those dazzling resources [1-3]. Most difficulties can finally find the root to insufficiency of geoinformatic knowledge (such as those relating to temporal-spatio scale or precision) and so are defined here as geoinformatic knowledge fence (GeoKF). With continually increasing of available geospatial data, GeoKF will

---

\* Corresponding author.

become an important obstacle to earth observation application, geospatial information services and other geocomputing applications.

The experiment is an exploratory to overcome GeoKF by combining geoinformatic knowledge with computing reasoning in scenario of geospatial data preparation within spatial decision making. Five sections are laid out in the paper. Section 1 introduces and defines GeoKF, then section 2 explains the philosophy and conceptual model of the approach to overcoming GeoKF. In section 3 the design of experiment system is detailed following the conceptual model, the experiment result is analyzed in the same section. Finally there are a discussion (section 4) and a conclusion (section 5) at the end of the paper.

## 2 Philosophy and Conceptual Model of GeoKF Overcoming

### 2.1 Philosophy of GeoKF Overcoming

A questionnaire on 30 n-geo experts (mainly in agriculture engineering, information system and management) of NERCITA shows two sorts of geographic knowledge during their spatial analysis. One can be termed as common geographic sense (CGS) mainly by which people model their analysis space through assigning i.e. spatio-temporal domain, scale and precision, etc.; the other can be termed as geoinformatic professional knowledge (GPK) by which geoinformatic professionals can understand various geospatial data and utilize them in a more effective and suitable way, for example processing some ostensibly unsuitable data by generalization or scale transformation to meet spatial analysis. Most n-geo experts lack the later and so are prevented from efficiently geospatial data preparation within their spatial analysis.

The research philosophy comes from the knowledge-oriented analysis of GeoKF and spatial decision process (Fig.1). GeoKF roots from knowledge heterogeneity between n-geo users and geoinformatic professionals, and can be represented as a fence between CGS and GPK. CGS, together with domain knowledge, is generally employed by domain professionals (most are n-geo users) to construct space model of analysis (SMoA) within which the target objects or processes are modeled and analyzed [4]. Such models also determine the suitable geospatial data for analysis, namely, record the data requirements (both data sources and their properties) in form of CGS. On the other side, various geospatial data is traditionally described and represented by their GPK-based metadata in terms of i.e. FGDC. So GeoKF-overcoming practically means the mapping between SMoA and GPK-based metadata, which generally perform by cooperation of geoinformatic professionals and n-geo SMoA makers.

The paper aims to find an intelligent and automatic approach to replacing the cooperation by spatial reasoning. Specifically, the paper aims to verify two hypotheses below.

- Hyp. #1. CGS can be regarded as dialect subsets of GPK, which means the data requirements of CGS-based SMoA can be ‘translated’ into GPK-based description of geospatial data. The translation gives the possibility of spatial reasoning to match SMoA and the data it need.
- Hyp. #2. Current GPK-based metadata can be modified or extended to support spatial reasoning relating to down or up-scaling, precision transform, etc.

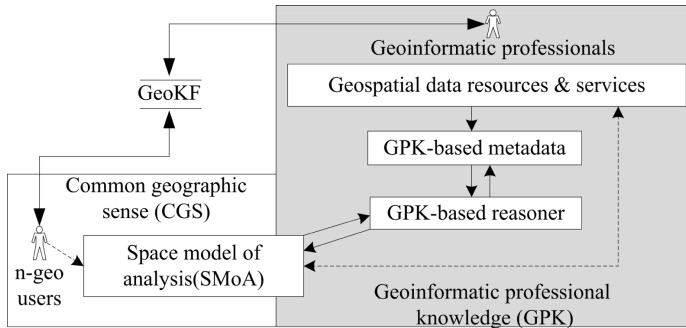


Fig. 1. GeoKF-overcoming philosophy

## 2.2 Conceptual Model

**Space Model of Analysis.** From the perspective of n-geo users, each spatial analysis task has its space models, which contain all analyzed geography features within certain spatial and temporal domain. The model can be expressed by a triple of (*features, space, time*). *Features* can be further expressed as (*focal features, background features*). *Focal features* are those analyzed such as roads in transport analysis; *background features* are those that function as context of *focal features* such as land use in transport analysis. Each dimension (*space* or *time*) is identified by domain, scale and precision. Although some geoinformatic professionals suggest that scale covers precision [5-7], most n-geo users are apt to confuse them and so the research employs both concepts contemporarily. Scale here means the size of analyzed or sampled unit. Precision here refers to minimum certain measure value of analysis. By this way, *space* dimension is the triple of (*spatial domain, spatial scale, spatial precision*), while *time* dimension is another triple of (*temporal domain, temporal scale, temporal precision*).

Table 1. Some main fundamental ontology

Fundamental ontology	Description
Geospatial theme-feature ontology	Extended ISO 19115[9], one example of extension are poultry feature of agriculture theme.
Geospatial scale and precision ontology	Describing the knowledge of geospatial scale and precision, including, metric scale (i.e. mapping scale, traditional airphoto scale, etc.), named scale (Global scale, continent scale, region scale etc.), and precision.
Geospatial domain ontology	Describing the knowledge of geospatial domain, including named domain (government precinct, geology zoning, geography zoning, etc.)
Time scale and precision ontology	Describing the knowledge of time scale and precision, including metric scale (second, day, year, etc.), named scale (Macro scale, geography scale, etc.), and precision (second, minute, ..., year, etc.)
Temporal domain ontology	Describing temporal domain, including named domain (geology, geography, etc.)

**Knowledge Organization and Reasoning.** All knowledge (CGS & GPK) should be organized in computable formats such as ontology, knowledge base with facts and rules[8], etc. The organization should consider both computing (i.e. spatial reasoning) and human operating. Some main fundamental knowledge is listed in Tab. 1 in form of ontology.

**Spatial Reasoning.** The spatial reasoning (Fig. 2) combines human intelligence and computing reasoning by an iteration cycle consisting of 1) rendering data requirements of SMoA to GPK-based reasoner, 2) reasoning and suitability calculating basing on data requirements, GPK, and metadata, and 3) returning result to n-geo users, who can re-render or accept result to end current cycle. Each cycle consists of three sub-reasoning tasks (theme-ontology evaluation, space evaluation, and time evaluation). Each task reasons on a section of SMoA requirement, corresponding knowledge base (represented as ontology in Fig.) and metadata of evaluated geospatial data source. The performance of each task will return the section suitability of current geospatial metadata. GPK-based resoner controls the running of sub-task and calculates the whole suitability of evaluated geospatial data sources by scanning metadata list of them. Formula (1) is used to calculate the suitability index (*SI*), while some GPK-based reasoning rules are listed in Tab.2.

$$SI = \sum ((suitability-item)_i \times W_i) (i=1, \dots, m) \tag{1}$$

*SI*: suitability index, the final suitability value of evaluated geospatial data.

*(Suitability-item)<sub>i</sub>*: the result of evaluating of certain (*i*) items of SMoA.

*W<sub>i</sub>*: the weight of each items of SMoA. The valued is appointed by decision makers to assign the importance of items. The total of all weight is 100%.

*m*: the number of evaluated items of SMoA.

**Table 2.** Main rules for suitability reasoning and calculation

Items of SMoA	Suitability-item	Description
Feature	$M/N \times 100\%$	<i>N</i> : the number of expected features in SMoA <i>M</i> : the number of existence features of evaluated geospatial data
Space-Domain	$(SMoA-D/Data-D) \times 100\%$ (0-100%)	<i>SMoA-D</i> : the domain of spatial analysis <i>Data-D</i> : the domain of evaluated geospatial data
Space-Scale & Time-scale	0%, 100%	100%: the scale of evaluated geospatial data is equal directly or by upscaling. 0%: other occasions
Space-Precision & Time-Precision	0%, 100%	100%: the precision of evaluated geospatial data directly or more precise to SMoA. 0%: other occasions
Time-domain	0%, 100%	100%: the time period of evaluated data content can cover time of SMoA completely 0%: other occasions

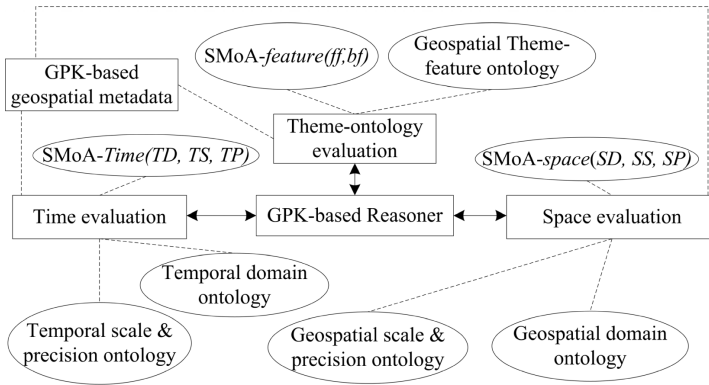


Fig. 2. Spatial reasoning

**Conceptual Model.** The model have 5 key components including geospatial data source agent, n-geo users and geoinformatic professionals, knowledge base (facts and rules), GPK-based reasoner, and human-computer interface (Fig.3). Geospatial data sources agent functions as a list of enrolled GPK-based metadata (Tab.3) of online

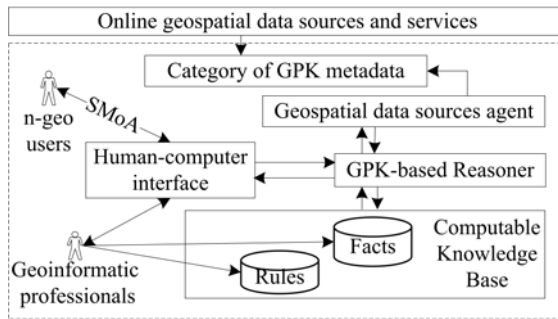


Fig. 3. Conceptual Model

Table 3. GPK-based FGDC metadata (limited to extended or modified items only)

FGDC metadata item	Extended or modified metadata Items	Description of extension or changes
Entity & Attributes and Keyword-Theme	Theme-features	Bounding Entities to Theme-feature ontology
Identification-spatial domain	Spatial domain	Bounding to geospatial domain ontology
Null	Spatial Scale	Bounding to geospatial scale & precision ontology
Null	Temporal scale	Bounding to temporal scale & precision ontology

geospatial data sources. CGS and GPK are organized in computable knowledge base for GPKreasoning. Human-computer interface is a GUI web page that lets decision makers (mainly n-geo users) render (and re-render) their SMOA as well as check the suitability of reasoned results.

### 3 Experiment Design and Analysis

An experiment, with a software system as its core, is designed and implemented to check the feasibility of philosophy. The comparison of reasoned results with evaluation of both n-geo users and geoinformatics professionals is employed as the basic method for the verification.

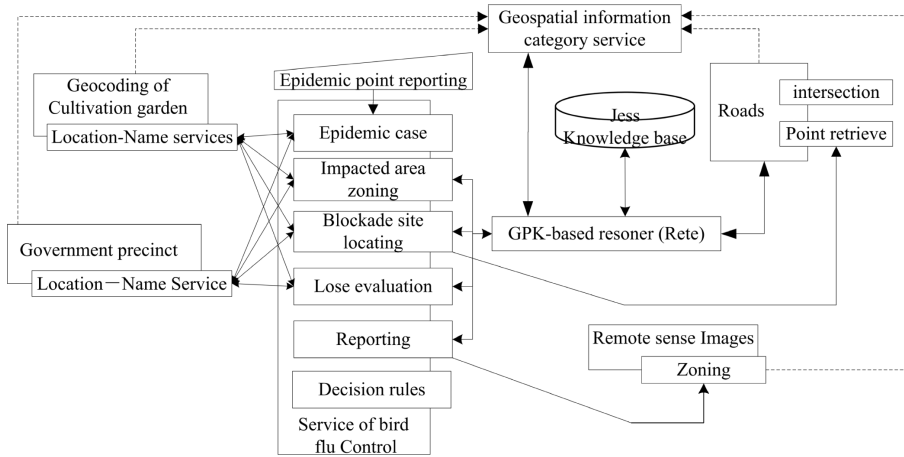
#### 3.1 Experiment Design

The experiment is a simplified spatial decision support of bird flu control in an agri-intensive county of Beijing area. Three independent and geographically distributed government agencies, which hold different geospatial and agricultural data sources, will collaborate with their data sources. The spatial analysis aims to zone epidemic impacted area and set block stations to control epidemic diffusion. Some main items of SMOA of the decision are listed in Tab.4. The available online geospatial data can be seen in Tab.5.

**Table 4.** SMOA items of bird flu control decision

SMoA items	Value	Description
Feature-focal	Poultry – livestock agriculture	From geospatial theme-feature ontology
Feature-background	Roads-Transportation	From geospatial theme-feature ontology
Feature-background	Resident area - land use - plan	From geospatial theme-feature ontology
Feature-background	Government precinct – named domain	From Geospatial domain ontology
Space-domain	Daxing District-Beijing city-China-Government precinct-Named domain	From geospatial domain ontology
Space-Scale	1:10000, Mapping scale-metric scale	From geospatial scale & precision ontology
Space-Precision	100M – precision	From geospatial scale & precision ontology
Time-domain	after. 2004	Null
Time-scale	Year-metric scale	From Temporal scale & precision ontology
Time-precision	Null	Null

**Software System Design.** As an experiment prototype, the system adopts Jess and Rete algorithm [8] because that Jess is a classic and fundamental technology for knowledge organization and reasoning. Ontology, and other knowledge technologies, is only used to organize knowledge to apart the proposed philosophy from implementation technologies as possible as it can. The logic structure of software system is listed in Fig 4.



**Fig. 4.** The system is an implementation (with 7 web services and 13 endpoints) of conceptual model (Fig.3). The service *geocoding of cultivation garden*, *government precinct*, *roads*, and *remote sense images* are all online geospatial data services and enrolled into *geospatial information category service*. The service of *bird flu control* is an application service (with a GUI page) that performs the spatial analysis triggered by a geo-referenced *epidemic point reporting*. The endpoint of *epidemic case* will initialize the spatial analysis of bird flu control as well as construct general SMOA basing on epidemic report. The endpoint of *impacted area zoning*, *blockade site locating*, *lose evaluation* and *reporting* will undertake the spatial analysis within which their detailed SMOA will be constructed basing on general SMOA and then its requirements will be transported o *GPK-based resoner* for obtaining the suitable geospatial data sources. The resoner will return reasoned result to corresponding endpoint for refining if needed.

### 3.2 Result and Analysis

Tab. 5 shows the comparison of reasoned results with human evaluation. Such comparison displays a generally consistence between reasoned results and expert evaluations, which can preliminarily give feasibility for further research. While as a philosophy-proof experiment, there still exist many defects such as small sample of online geospatial data, relatively simple reasoning rules (esp. on scale down- and up-scaling and precision, etc.) and facts, etc. Additionally, a strong suggestion from experts for further research is to add quality of data services (QoS) into SMOA and GPK-reasoning to get dynamic optimal data set.

**Table 5.** Comparison of human evaluation with reasoned result

Available geospatial data	Human evaluation of suitability	Reasoned suitability index(%)								
		total	Feature%		Space			Time		
			FF 30%	BF 5%	SD 30%	SS 10%	SP 5%	TD 10%	TS 5%	TP 5%
Cultivation garden (2004)	High	95	100	0	100	100	100	100	100	100
Fundamental geography information (2000: 1:10M)	Mid	85	50	100	100	100	100	100	100	100
Beijing fundamental geography information(1998 1:10000)	High	75	50	100	100	100	100	0	100	100
Beijing Quickbird image (2008-12: R=0.61M)	High	100	100	100	100	100	100	100	100	100
Beijing DEM (1:10000, 2001)	Low	0	0	0	100	Null	Null	Null	Null	Null
Beijing MODIS (2008-12, 30M)	Mid	70	50	100	100	0	0	100	100	100

## 4 Discussion

According to the exploratory research, GeoKF-overcoming is a synthetic issue relating to some traditional geocomputing problems including intelligent geospatial data generalization [10-11], scale and precision transformation [12], etc. In other words, GeoKF research will depend on and contribute to those relating studies. Such relationships among them also discover some features of further researches.

- The combination of artificial and human intelligence. Because of the immature of scale transform, automatically geospatial data generalization and other relative issues, the overcoming of GeoKF is still a combination of human and computing intelligence, namely, increasing the level of automation by adding more knowledge and reasoning ability is a key research direction.
- The dependence of GeoKF-overcoming practice on fundamental GPK research and practices. Although more and more discussions of geo-ontology or other geo-knowledge relative researches appear recently, the practice works, such as construction



of geo-ontology (including geospatial theme-feature ontology, geospatial scale & precision ontology, etc.), have not yet functioned. Furthermore, the workings relating to geo-knowledge components, such as adding more semantic to metadata to support computing reasoning, are still in their way to practice. It is clear that the power of proposed approach of the paper will grow with increase of those workings.

## 5 Conclusion

With booming of available online geospatial data resources, the bottleneck of net-centric geocomputing has moved gradually from network and computing capability to efficiently utilization of data resources by more users, esp. n-geo users. As a consequence, geoinformatic knowledge and intelligent geo-computing (including geo-ontology, spatial reasoning, intelligent generalization of geospatial data, etc.) has been attracting increasing attentions when facing overwhelming online resources.

The research suggests that GeoKF is an emerging issue within intelligent geo-computing. It shares one aim of geo-cognition research, namely, how to provide n-geo users with correct understanding and utilization of dazzling online geospatial data sources, as well as to help geoinformatic professionals improve their efficiency and capability. Featuring with knowledge-oriented analysis of geospatial data process and spatial analysis, the conceptual model with SMoA, GPK-based reasoning and extended GPK metadata, and the combination of human and computing intelligence, the proposed approach shows its feasibility for further research of GeoKF.

**Acknowledgments.** The paper is supported by Beijing sci-tech Nova plan (2006B27) and national key technology R&D program (2006BAJ05A09, 2006BAD10A05).

## References

1. Rahimi, S., Cobb, M., Ali, D., Paprzycki, M., Petry, F.E.: A knowledge-based multi-agent system for geospatial data conflation. *Journal of Geographic Information and Decision Analysis*, 1480–8943 (2002)
2. Wei, Y., Yue, P., Dadi, U., Min, M., Hu, C., Di, L.: Effective Acquisition of Geospatial Data Products in a Collaborative Grid Environment. In: *IEEE International Conference on Services Computing (SCC 2006)*, pp. 455–462. IEEE Press, New York (2006)
3. McMaster, R.B., Lynn, E.: *A Research Agenda for Geographic Information Science*. CRC Press, Boca Raton (2004)
4. Longley, P., Batty, M.: *Advanced Spatial Analysis*. ESRI Press, Redlands (2003)
5. Zhilin, L.: A theoretic discussion on the scale issues in geospatial data handling. *J. Geomatics world*, 1–5
6. Sun, q.-x., Li, m.-t., Lu, j.-x., Guo, d., Fang, t.: Scale issue and its research progress of geospatial data. *J. Geography and geo-information science* 23, 5–56 (2007)
7. Meng, b., Wang, j.-f.: A review on the methodology of scaling with Geo-data. *J. ACTA geographic SINICA* 60(2), 277–288 (2005)
8. Friedman-Hill, E.: *Jess in Action: Java Rule-Based Systems*. Manning Publications Co. (2003)
9. ISO/TC211 19115: *Geographic information – Metadata* (2003)

10. Lehto, L., Sarjakoski, L.T.: Real-time generalization of XML-encoded spatial data for the Web and mobile devices. *J. Geographical information science* 19(8,9), 957–973 (2005)
11. Jones, C.B., Abdelmoty, A.I., Lonergan, M.E., van der Poorten, P., Zhou, S.: Multi-scale spatial database design for online generalisation. In: *Proceedings of the Spatial Data Handling Symposium, Beijing*, pp. 7b.34–7b.44 (2000)
12. Li, S.-l., Cai, y.-l.: Some scaling issues of geography. *J. Geographical research* 1, 11–18 (2005)