



# 空间数据的关注问题 :从质量到可用性\*

李志林

(香港理工大学 土地测量与地理信息资讯学系,香港 九龙)

摘要 据统计,在一个地理信息系统(GIS)项目中,空间数据的花费将占总投入的80%左右。因此,如何保证所得到的数据能够满足项目需要是研究者、生产者及用户共同关心的一个重要问题。20世纪80年代以来,空间数据质量一直是GIS学术界研究的一个热门课题,并受到地理信息业界的重视。因此,数据生产者总是尽最大努力来生产高质量的数据。而实际上,数据用户经常对得到的空间数据不尽满意,尽管数据质量完全满足规范要求。笔者认为:可用性是空间数据最应优先考虑的问题,而质量仅仅是可用性的一个指标。其他的可用性指标包括可享用性、可得性、受益情况、完整性、熟悉性、业界最佳作业方法、可靠性等50多个。

关键词:空间数据;质量;可用性

中图分类号:P208 文献标识码:A

## Concerns with Spatial Data From Quality to Usability

LI Zhi-lin

(Dept. of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China)

Abstract: It has been widely accepted that about 80% of the budget of a GIS project is the cost of spatial data. Therefore, spatial data is a big concern to both researchers and practitioners. In recent years, the quality of spatial data has become a hot research topic. Data quality has attracted much attention from geoinformation community. Data producers try their best to produce data with highest quality. However, in practice, many users are not very satisfied with the spatial data available although the quality of the data is adequate according to specifications. In this paper, the author tries to explain that "quality" is only one of the many elements in a broader concept - "usability". Other fifty measures have been identified, such as accessibility, availability, benefits, completeness, familiarity, industry best practice, reliability, and so on.

Key words: spatial data; quality; usability



李志林(1960-),男,1982年毕业于西南交通大学,学士,1990年英国格拉斯哥大学获博士学位。香港理工大学教授,GIS领域国际权威杂志《International Journal of Geographical Information Science》的编委。  
E-mail:lszllie@polyu.edu.hk  
收稿日期:2006-05-11  
基金项目:香港理工大学基金(G-U045)

### 0 引言

据统计,在一个地理信息系统(GIS)项目中,空间数据的花费将占总投入的80%左右。因此,如何保证所得到的数据能够满足项目需要是研究者、生产者及用户共同关心的一个重要问题。

随着地理信息产业的迅速发展,地球空间数据已经成为一种基础设施。国家空间数据设施(NSDI)这一术语也应运而生。

20世纪80年代以来,空间数据质量一

直是GIS学术界研究的一个热门课题,并受到地理信息业界的重视。近年间,出现了专门讨论空间数据质量的学术会议,一些有关空间数据质量的专著(或文集)也相继出版,如Guptill & Morrison(1995),Heuvelink(1998),Mowrer & Congalton(2000),Shi et al.(2002)和Zhang & Goodchild(2002)。

对空间数据质量重视的一个直接结果是:数据生产者总是尽最大努力来生产高质量(或精度)的数据。而实际上,尽管数据质

\* 本文是由作者在2005年地理信息系统理论与方法研讨会上作的主题报告修订而成



量完全满足规范要求,数据用户经常对得到的空间数据感到不尽满意。因此,人们不禁要问究竟为什么?也就是说,人们想知道:“质量是空间数据最应优先考虑的问题吗?”其答案是:“不一定。”笔者认为:可用性是空间数据最应优先考虑的问题,而质量仅仅是可用性的一个指标。

## 1 质量与可用性的关系:用鞋子来作类比

实际应用中,由于各种原因,高质量的空間数据不一定适合用户的需要。这就像市场上的鞋子一样,五花八门、高质量的很多,但不一定可用。下面我们以前鞋子为例,对空间数据的质量和可用性进行较为全面的类比分析。

从鞋的尺寸来看,市场上可能有各种尺寸的高质量鞋(相当于多尺度的高质量空间数据),但对一个购买者(相当于空间数据用户)来说,可能只有某一尺寸的鞋才合适。也就是说,不管其他尺寸的鞋制作质量多高,对他来说都不适用,因此不可用。可以肯定,对于一个儿童(空间数据用户)来说,鞋商(空间数据生产者)提供的各种尺寸的成年人鞋,不管质量多高,都是不可用。这意味着,数据生产者提供的空间数据尺度或分辨率与用户的要

求不一致,便会导致空间数据的不适用性,因而产生不可用性。以中国西北黄土高原地区土壤侵蚀研究<sup>1)</sup>为例(如图1所示),在这个应用中理想的比例尺是1:50 000。但是根据生产规范可知,这个尺度的DEM分辨率是25 m。显然,对高原地区的地貌分析太过粗糙。因此,这种尺度的DEM数据虽然满足精度要求,但不能满足黄土高原地貌分析的应用要求。再如,1:10 000的DEM具有很高的精度,但不适用于全球范围的水文模型的建立。

从鞋的样式来看,可以分为女式和男式。并且男式的鞋子不适合女式穿,反之亦然。这就类比于空间数据的格式问题,一个矢量GIS不能有效地处理栅格数据。而对于这种GIS用户来说,尽管栅格数据的质量可能很高,但不可用。在文件格式和数据模型方面,也存在着类似的问题。这些问题也是实现不同系统之间数据共享和互操作必须解决的核心问题。

从鞋的使用功能来看,鞋子有休闲鞋和专业运动鞋之分。显然,休闲鞋不适用于专业运动。这就类比于一组从地形图获得的道路数据不一定适合用来交通建模,因为需要的信息内容不尽相同。

从鞋的时效性来看,一种鞋可能会因过时、用旧或穿破而被淘汰。类似

地,空间数据也具有现势性。现势性差的空间数据可能用处不大而需更新。这也是现势性被用作空间数据质量指标的一个重要原因。但是,笔者认为,现势性是数据可用性的一种度量指标,而不是数据质量的度量指标。图2表示了3张不同年代的香港海岸线地图。虽然这些地图的现势性都很差,但它们却都是高质量的,能很好地表达和反映海岸线在相应年代的实际情况。

也有这样的情况,两组或多组空间数据都具有高质量(精度),但他们在不同的坐标系统。如果不进行坐标系统转换预处理,它们就不能叠置或集成。

其他一些方面,如空间数据的属性信息、可得到性等,也都影响空间数据的应用范围和可用性。在属性方面,测绘部门生产的交通图对交通管理部门不适用,因为这些交通图缺乏其所需要的交通信息(如流量)。在可得到性方面,空间数据可能由于其安全性、使用权限或者价格等问题,使得很多用户难以得到。因而,这些空间数据对这类用户无法使用。

综上所述可知,高质量(或精度)空间数据不一定可用。并且高质量的空间数据不一定意味着具有高可用性。因此,空间数据的质量对数据用户来说不一定是最重要的一个方面,但是其可用性却一定是所有用户最应关心的一个方面。

## 2 空间数据可用性的度量指标

从上述分析可知,可用性是一个相对的概念。只有当空间数据的使用对象(即用户)和使用目的(即特定应用)确定的情况下,评价空间数据的可用性才更具有实际意义。对于一个用户来说,评价空间数据的可用性可以

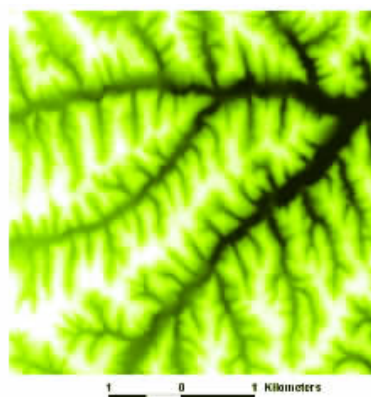


图1 中国西北黄土高原地区

Fig.1 Loess Plateau area in northwest China



根据这些空间数据在其应用中的使用效率、有效性以及满意度来衡量。

2001年,在捷克举办的第四届AGILE大会上,正式成立了一个空间数据可用性研究工作组 (<http://www.geo-informatie.nl/projects/agile/left.htm>)。工作组的目标为([http://www.geo-informatie.nl/projects/agile/about\\_us\\_mission\\_statement.htm](http://www.geo-informatie.nl/projects/agile/about_us_mission_statement.htm)):

- 1)了解和掌握空间数据的可用性;
- 2)加强空间可用性的意识,普及空间可用性的知识和技能;
- 3)解决空间数据存在的可用性问题,提升地球空间信息的使用价值;
- 4)发展一种有效方法来辨别、评价和可视化现有产品、数据和交互界面的可用性度量指标;
- 5)建立并普及空间数据可用性的评价标准;
- 6)加强这一领域国际重要问题的理解和研究;
- 7)建立这一领域研究的国际学术组织。

并且,工作组在大会上提出了空间数据可用性的52个度量指标 ([http://www.geo-informatie.nl/projects/agile/understanding\\_spatial\\_data\\_usability\\_elements.htm](http://www.geo-informatie.nl/projects/agile/understanding_spatial_data_usability_elements.htm)),包括:

- 1)可享用性(Accessibility);
- 2)精度(Accuracy);

- 3)增值(Adding-Value);
- 4)聚合程度(Aggregation Levels);
- 5)属性精度(Attribute Accuracy);
- 6)权威性(Authoritative);
- 7)可得性(Availability);
- 8)受益(Benefits);
- 9)数据内容的说明(Caveats on data content);
- 10)数据集的合格证(Certification of data sets);
- 11)便利性(Convenience);
- 12)完整性(Completeness);
- 13)成本(Costs);
- 14)现势性(Currency);
- 15)数据的集成性(Data Integration);
- 16)数据生产者的声誉(Data Producer's Reputation);
- 17)数据质量(Data Quality);
- 18)独占性(Exclusiveness);
- 19)熟悉性(Familiarity);
- 20)无错程度(Freedom from Error);
- 21)无错保证(Guarantee against error);
- 22)人机交互(Human Computer Interaction);
- 23)业界最佳作业方法(Industry Best Practice);
- 24)可集成性(Integration);
- 25)完美性(Integrity);
- 26)吸引力(Interestingness);
- 27)国际标准(International Standards);

- 28) 解读(Interpretation);
- 29) 合法性(Legality);
- 30) 逻辑一致性(Logical Consistency);
- 31) 元数据(Metadata);
- 32) 新服务(New Service);
- 33) 新颖性(Novelty);
- 34) 流行性(Popularity);
- 35) 位置精度(Positional Accuracy);
- 36) 生产目的(Product Purposes);
- 37) 可靠性(Reliability);
- 38) 满意度(Satisfaction);
- 39) 可查询度(Searchability);
- 40) 可提供的服务(Services provided);
- 41) 访问速度(Speed of Access);
- 42) 软件界面(Software Interfaces);
- 43) 时间精度(Temporal Accuracy);
- 44) 应用类型(Type of Application);
- 45) 决策类型(Type of Decision);
- 46) 可信赖度(Trust);
- 47) 意外性(Unexpectedness);
- 48) 模型和算法的可协作性(Use with models and algorithms);
- 49) 用户的技能水平(User Skill Levels);
- 50) 功用(Utility);
- 51) 有效性(Validity);
- 52) 可视化工具(Visualisation Tools);

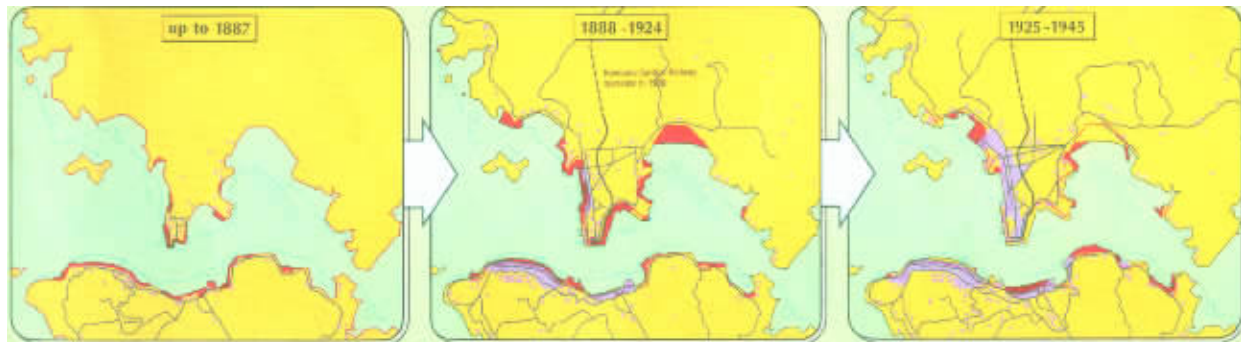


图2 香港海岸线随时间的变化情况  
Fig.2 Changes in the coastal lines of Hong Kong over time





### 3 空间数据可用性的研究日程表

尽管对空间数据可用性的研究只是近几年的事,但已得到一些相关研究机构的重视,如美国国家地理信息中心(NCGIA)和欧洲 GISDATA 资助了有关这方面的研究专题<sup>[2]</sup>,如“GIS 传播和应用的障碍 (Impediments to GIS diffusion and adoption)”、“信息的成本 (Cost of information)”、“信息价值评价 (Assessing the value of information)”,“信息有偿使用的自愿性 (Willingness to pay for information)”,以及“经提炼的信息对决策支持的影响 (Effects of improved information on decision making)”。

特别是 2001 年 11 月在荷兰召开了一个有关空间数据可用性的工作组会议 (<http://www.geo-informatie.nl/cgi/news/workshop/workshop-page.htm>)。工作组会议包括以下几个专题<sup>[3]</sup>:

- 1) 什么是“可用性”?
- 2) 为什么“可用性”是重要的?
- 3) 空间数据可用性的特征是什么?
- 4) 空间数据可用性有待解决的问题有哪些?
- 5) 关于空间数据可用性,需要优先研究的问题有哪些?

2003 年 2 月,工作组就有关空间数据可用性的问题在《Data Science Journal》杂志上组织出版了一本专集<sup>[2]</sup>。

AGILE 工作组列出的一些主要研究问题为 ([http://www.geo-informatie.nl/projects/agile/understanding\\_spatial\\_data\\_usability\\_research\\_questions\\_and\\_priorities.htm](http://www.geo-informatie.nl/projects/agile/understanding_spatial_data_usability_research_questions_and_priorities.htm)):

1) 所列的 52 个可用性指标是否完备,是否为理解空间数据可用性的最基本起点?

2) 如果这些是可用性的指标,那么它们的度量是什么?如何应用这些度量?

3) 这些指标是否可以分组,以得到较为一般的分类?

4) 这些指标之间存在什么关系?

5) 这些指标是否存在等级性?是否可以利用自上而下的方法评价可用性,例如,首先可以从应用的角度评价,然后再从数据质量的角度评价等。

6) 从工程学的角度,可用性指的是什么?

7) 我们能否利用形式化规则来描述和确保空间数据的可用性?

8) 可用性指标随时间如何变化?

9) 在可用性度量和用户任务之间是否存在一个可供参考的连接方案?

10) 如何定义空间和非空间数据的可用性?

### 4 有关空间数据可用性研究的一些刍议

在空间可用性的评价方面,笔者认为 AGILE 工作组提出的 52 个指标是非常复杂的。其中一个关键问题是,这些评价指标是否具有相同的重要性,即同等级的,事实上未必如此。最近 Bregt (2001) 的研究表明,专业人员和非专业人员对空间数据可用性的评价角度通常不一样<sup>[4]</sup>。对专业人员来说,“可靠性 (Reliability)”和“精度 (Accuracy)”是两个最主要的指标,其次是“可应用性 (Useful for goal)”、“元数据 (Meta data)”、“数据集成的可能性 (Possibility for data integration)”以及“与决策的相关度 (Relevant to decision making)”等指标。而对非专业人员,“简单明了又不复杂 (Simple, clear not complex)”和“容易得到 (Easy accessible)”是两个主要指标。

到目前为止,中国政府已经投入了大量资金建设国家空间数据基础设施,并且各种类型、专题的地理空间数据框架已经基本建立起来了。下一步的一个主要工作将是评价这些空间数据的可用性。笔者对这方面的研究也非常感兴趣,并且正在主持一个大学基金项目,主要是针对国家 1:50 000 比例尺 DEM 的可用性进行系统地调查研究。希望在这方面能够与有兴趣的同仁合作。此外,其他类型的空间数据可用性也值得调查研究。

### 5 结 论

空间数据是 GIS 项目成果的关键。实践中用户可能经常遇到这样的情形,即现有的数据不是自己想要的,得到的数据又可能不是自己所需要的,需要的数据自己可能又得不到,能得到的数据自己又可能负担不起。

有关空间数据的诸多问题也一直受到 GIS 学术界和应用部门的高度重视。当前的大量研究主要集中在空间数据的质量方面。而用户关心是否满足其需要,即空间数据的可用性,不一定是空间数据的质量。并且,高质量的数据不一定能够满足用户需要。这里必须强调质量仅仅是衡量空间数据可用性的一个指标。希望在今后的研究工作中空间数据的可用性问题能得到更多的重视。

### 参考文献:

- [1] Tang, G.A., Shi, J., Zhao, M. and Zhang, Y.. Modeling slope uncertainty derived from DEM: A case study in China Loess Plateau area [C]/Proceedings of ISPRS Commission II Symposium, 21-23 August 2003, Xi'an, China. 455-461.

(下转第 27 页)



技术的发展历程,从数据格式转换到直接访问再到网络服务技术的引入,使得地理信息共享从独立过程转向协作完成、从局部范围的数据共享转向大范围的乃至全球的数据共享。针对基于网络服务技术的标准化需求,本文介绍了与地理信息共享和互操作相关的标准和规范,包括元数据标准、GML、简单要素访问和网络数据服务规范等,并在此基础上,阐述了基于网络服务的地理信息共享和互操作的实现技术,并从物理实现的角度给出实现方案。

目前,网络服务技术已经比较完善,相关地理信息标准的制定工作也逐渐完成,地理空间数据共享与互操作的技术与标准已基本成熟,需要解决的问题是各GIS厂商和数据提供者尽快采用相关的标准来实现空间数据服务。尽管W3C关于网络服务的定义并没有指定需要SOAP和HTTP协议,也不一定需要UDDI来注册和发布服务信息,但是为了便于实现地理信息共享和互操作,不同网络服务实现程序应该遵循OGC和ISO定义的、被普遍公认的地理信息网络服务标准,方便客户端以同样的方式访问和调用地理信息服务,这也利于网络服务和地理信息系统开发商推广产品。

## 参考文献:

- [1] Gong Jianya, Shi Lite, Du Daosheng, Rolf A. de By. Technologies and standards on spatial data sharing [C]/Proceedings of ISPRS 7/2004, B4, pp118-128.
- [2] Allan Levinsohn. Geospatial Interoperability: The Holy Grail of GIS [EB/OL]. GeoEurope, available online, <http://www.geoplance.com/gw/2000/1000/1000data.asp>
- [3] Y. Bishr. Overcoming the semantic and other barriers to GIS interoperability [J]. International Journal of Geographic Information Science, 1998, 12(4):299-314.
- [4] ESRI White Paper. Interoperability in Enterprise GIS [EB/OL]. 2005. <http://www.esri.com/library/whitepapers>
- [5] 杨涛,刘锦德. Web Services 技术综述一种面向服务的分布式计算模式 [J]. 计算机应用 2004, 24(8).
- [6] IBM Redbook. Patterns: Service-Oriented Architecture and Web Services [EB/OL]. 2004. <http://www.redbooks.ibm.com/red-books.nsf/redbooks/>
- [7] Roger A Longhorn. Geospatial Standards, Interoperability, Metadata Semantics and Spatial Data Infrastructure [C]. background paper for NIEES Workshop on Activating Metadata 6-7 July 2005, Cambridge, UK.
- [8] ESRI White Paper. Spatial Data Standards and GIS Interoperability [EB/OL]. 2003. <http://www.esri.com/library/whitepapers>
- [9] ESRI White Paper. Spatial Data Standards and GIS Interoperability [EB/OL]. 2003. <http://www.esri.com/library/whitepapers/pdfs/spatial-data-standards.pdf>
- [10] Open GIS Consortium (OGC). OpenGIS Abstract Specification—Topic 12: OpenGIS Service Architecture, Version 4.3 [S]. 2002.
- [11] Open GIS Consortium (OGC). Web Coverage Service, Version 1.0. [S]. 2003.
- [12] Open GIS Consortium (OGC). Web Map Service, Version 1.1. [S]. 2004b.
- [13] Open GIS Consortium (OGC). Web Feature Service Implementation Specification, Version 1.1. [S]. 2005d.
- [14] Open GIS Consortium (OGC). OpenGIS Geography Markup Language (GML) Implementation Specification, Version 3.1.0 [S]. 2004c.
- [15] Open GIS Consortium (OGC). OpenGIS Implementation Specification for Geographic Information—Simple feature access—Part 1: Common architecture, Version 1.1. [S]. 2005b.
- [16] Open GIS Consortium (OGC). OpenGIS Implementation Specification for Geographic Information—Simple feature access—Part 2: SQL option, Version 1.1.0 [S]. 2005c.
- [17] Open GIS Consortium (OGC). OpenGIS Metadata, Version 1.1. [S]. 2001.
- [18] FGDC. Content Standard for Digital Geospatial Metadata (CSDGM) [EB/OL]. Federal Geographic Data Committee, 2000. <http://www.fgdc.gov/metadata/contstan.html>
- [19] Open GIS Consortium (OGC). OWS 2 Common Architecture: WSDL SOAP UDDI, Version 1.0. [S]. 2005e.
- [20] R. Laurini. Spatial multi-database topological continuity and indexing: a step towards seamless GIS data interoperability [J]. International Journal of Geographic Information Science, 1998, 12 (4) 373-402.
- [21] Open GIS Consortium (OGC). OpenGIS Abstract Specification—Topic 0: Abstract Specification Overview, Version 1.0. [S]. 2004a.
- [22] Open GIS Consortium (OGC). OGC Web Services Common Specification [S]. 2005a.
- [2] Hunter, G., Wachowicz, M. and Bregt, A.. Understanding spatial data usability [J]. Data Science Journal, 2003, 2:79-89.
- [3] Wachowicz, M. AI-GEOSTATS: Call for participation - Workshop on Spatial Data Usability [EB/OL]. <http://www.mail-archive.com/ai-geostats@unil.ch/msg00374.html>.
- [4] Bregt, A.. Users perception of spatial data usability [Z]. Presentation at Workshop on Spatial Data Usability, November 19-20, 2001, The Netherlands.
- [5] Guptill, S., and Morrison, J.L.. Elements of Spatial Data Quality [M]. Pergamon Press Inc, 1995.
- [6] Heuvelink, G.. Error Propagation in Environmental Modelling with GIS [M]. CRC Press, 1998.
- [7] Mowrer, H. T. and Congalton, R. G. (eds.). Quantifying Spatial Uncertainty in Natural Resources: Theory and Applications for GIS and Remote Sensing [M]. Chelsea, MI, USA: Ann Arbor Press, 2000.
- [8] Shi, W.Z., Goodchild, M. F. and Fisher, P. (eds.). Spatial data quality [M]. London: Taylor & Francis, 2002.
- [9] Zhang, J.X. and Goodchild, M.F.. Uncertainty in Geographical Information [M]. CRC Press, 2002.

(上接第 17 页)

- [2] Hunter, G., Wachowicz, M. and Bregt, A.. Understanding spatial data usability [J]. Data Science Journal, 2003, 2:79-89.
- [3] Wachowicz, M. AI-GEOSTATS: Call for participation - Workshop on Spatial Data Usability [EB/OL]. <http://www.mail-archive.com/ai-geostats@unil.ch/msg00374.html>.
- [4] Bregt, A.. Users perception of spatial data usability [Z]. Presentation at Workshop on Spatial Data Usability, November 19-20, 2001, The Netherlands.
- [5] Guptill, S., and Morrison, J.L.. Elements of Spatial Data Quality [M]. Pergamon Press Inc, 1995.
- [6] Heuvelink, G.. Error Propagation in Environmental Modelling with GIS [M]. CRC Press, 1998.
- [7] Mowrer, H. T. and Congalton, R. G. (eds.). Quantifying Spatial Uncertainty in Natural Resources: Theory and Applications for GIS and Remote Sensing [M]. Chelsea, MI, USA: Ann Arbor Press, 2000.
- [8] Shi, W.Z., Goodchild, M. F. and Fisher, P. (eds.). Spatial data quality [M]. London: Taylor & Francis, 2002.
- [9] Zhang, J.X. and Goodchild, M.F.. Uncertainty in Geographical Information [M]. CRC Press, 2002.