

World Status of National Spatial Data Clearinghouses

Joep Crompvoets and Arnold Bregt

Abstract: In December 2001, a survey of the World Wide Web was carried out to assess systematically the state-of-the-art of national spatial data clearinghouses around the world. The aim of this Web survey was to evaluate the status and the spatial distribution of the clearinghouse implementations and to find out the similarities and differences between them. Each clearinghouse was characterized by the following criteria: the year of first implementation; the number of data suppliers; the type of data accessibility; the metadata-standard; the number of datasets; the most recently produced dataset; the number of Web references; the number of monthly visitors; the frequency of Web updates; the language used; and the use of maps for searching and registration-only access. For 59 countries, a national clearinghouse has already been established. The results of this Web survey show a large variety in history, use, content, and management.

Introduction

At present, of the 192 countries in the world, 120 are working on their national spatial data infrastructure (NSDI) in order to create an efficient environment for the access of spatial data. One of the main components of an NSDI is a national clearinghouse (McLaughlin 1991, Clinton 1994, Federal Geographic Data Committee 1997, Onsrud 1998). A spatial data clearinghouse can be defined as an electronic facility for searching, viewing, transferring, ordering, advertising, and disseminating spatial data from numerous sources via the Internet. Such a clearinghouse usually consists of a number of servers that contain information (metadata) about available digital data. A national clearinghouse aims to become a kind of shopping mall for all national available spatial data as acquired by governmental agencies and/or industrial bodies.

The first national clearinghouse was established in 1994 in the United States. From that moment, there has been fast development of other clearinghouses throughout the world. Few studies exist with information about the worldwide status of national clearinghouses. Onsrud (1998) and Lance and Hyman (2001) presented a list of the existing national clearinghouses; however, no detailed description was included. To the best of our knowledge, no systematic periodical research has taken place with regard to the status of national clearinghouses. In order to fill this gap, in 1999 we started a survey of the World Wide Web, which was repeated every 6 months. The Web survey's main objectives are to assess the worldwide progress, the spatial distribution, and the similarities and differences between national clearinghouses. Additionally, this Web survey can be considered as a starting point to gather information necessary for the analysis of the legal, economical, cultural, technical, and institutional impacts on the development of clearinghouses. Moreover, since clearinghouses are a key component of spatial data infrastructures, the evaluation of the findings of this Web survey might improve the planning and investing of spatial data infrastructures in a more strategic

way. This article presents only the results of data collected the month of December 2001.

Web Survey Methodology

In order to assess national clearinghouse developments around the world, a Web survey was undertaken. The main intention of this survey was to examine the condition (status) of national clearinghouses. Added to this survey is the word "Web" to emphasize that the needed information was mainly collected on the Web. Because of the easy access to the clearinghouse sites, the Web is an excellent means by which to gather the needed information quickly and objectively on a regular temporal basis. In this case, a traditional research survey, which would collect information by asking a set of pre-formulated questions in a predetermined sequence in a structured questionnaire to individuals (Hutton 1990), would not be a suitable approach to collect the needed information quickly, objectively, and easily.

The Web survey began in December 1999 and was conducted in a systematic and periodical manner. The procedure consisted of the following two steps: 1) making an inventory of all existing national clearinghouses on the Internet; and 2) measuring several characteristics to describe each clearinghouse.

The inventory (step 1) was compiled by extensive browsing of the Internet, reading related literature, and contacting experts and several Webmasters. Clearinghouses were characterized (step 2) based on the following criteria: ease of measurement, objective character, and clear presentation of history, content, and use and management of the clearinghouse. The following 12 characteristics were measured: 1) the year of first implementation; 2) the number of data suppliers; 3) the type of data accessibility; 4) the metadata-standard used; 5) the number of datasets; 6) the most recently produced dataset; 7) the number of Web references (AltaVista and Google); 8) the number of monthly visitors; 9) the frequency of Web updates; 10) the languages used; 11) the use of

maps for searching; and 12) registration-only access. Almost all of the above information was sourced from clearinghouse Web pages. Additionally, in cases of uncertainty or missing data, the Webmaster was contacted. The history of the clearinghouse is described by characteristic 1. The content is described by characteristics 2, 3, 4, 5, and 6. The use is described by characteristics 2, 7, 8, 10, 11, and 12. Finally, the management is described by characteristics 4, 6, 9, and 12. In the next section, each characteristic is explained and discussed in greater depth. Additionally, where available, information about funding and clearinghouse strategy has been incorporated.

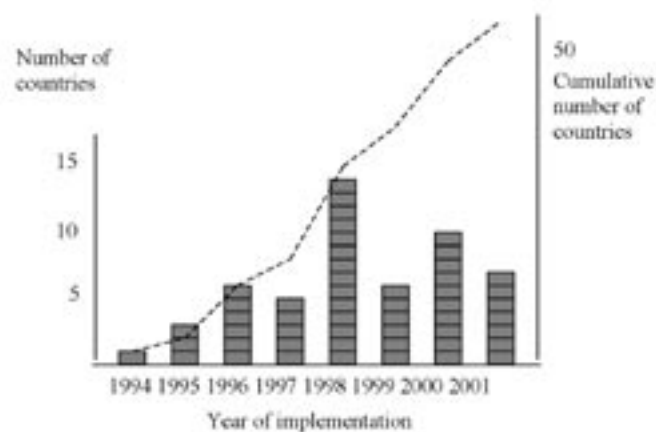


Figure 1 The first year of national clearinghouse implementation: per year (columns) and cumulative (dashed line).

Results and Discussion

The main results of this “December 2001” Web survey for each of the 12 mentioned characteristics are presented and discussed. Some of the results are presented by region. This division in regions is derived from Dorling Kindersley’s World Atlas (1997).

1. The Year of First Implementation

This characteristic describes the history of the clearinghouse. Beginning in 1994, the number of national clearinghouses has been steadily increasing (Figure 1). Currently, 59 countries have an implemented version on the Web (the year of first implementation is known for 52 of the clearinghouses). Additionally, it is known that nine countries have projects for implementation. However, the variety in number between the different regions is considerable (Table 1). For example, in Europe, North America, and South America, more than 50% of the countries have established a national clearinghouse, whereas in Africa this number is less than 5%. It is important to note that 124 countries have not conducted any initiative to build such a national facility.

In Figure 2, the global distribution of implementation status of national clearinghouses is presented. Here we can see that the main “hotspots” of implementation are the American continent, Europe (except Eastern Europe), Southeast Asia, and Australia. On the other hand, the main “holes” are the African continent and Middle East.

2. The Number of Data Suppliers

This characteristic describes the number and diversity of data suppliers. The power of a clearinghouse is that several data suppliers can disseminate their products via this facility. The average number of data suppliers participating in a

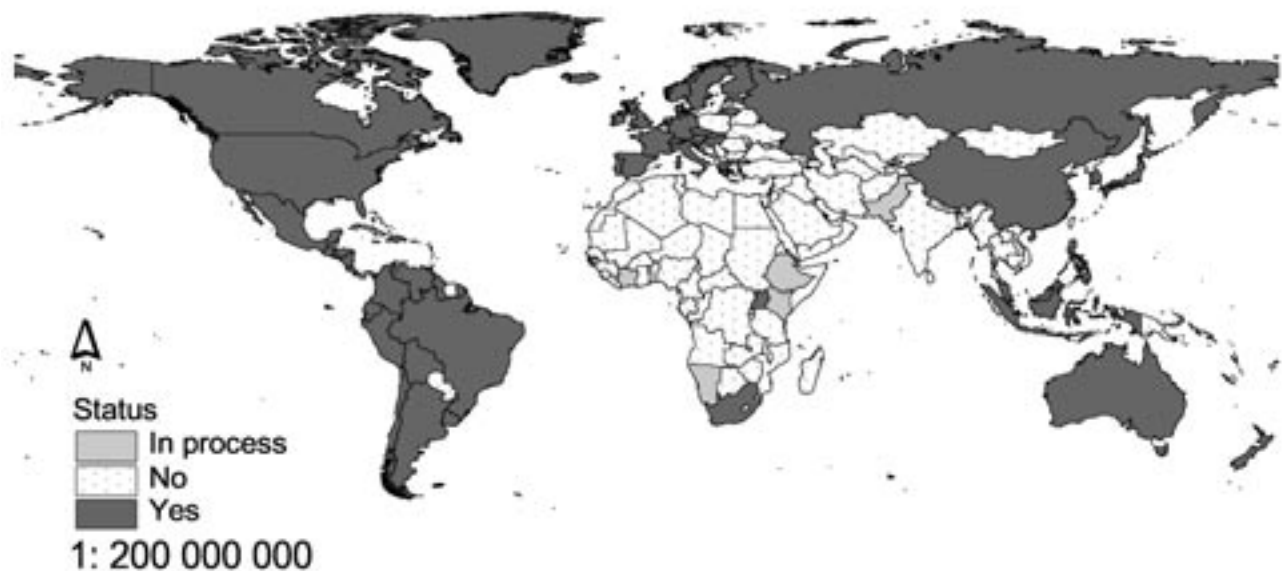


Figure 2 Global distribution of status of national clearinghouses

First Year of Implementation	Total No. of Countries	No. of African Countries	No. of Asian Countries	No. of Australian Countries	No. of European Countries	No. of North American Countries	No. of South American Countries
1994	1					1	
1995	3				3		
1996	6	1	3	1	1		
1997	5		2		2		1
1998	14	1	3	1	6	2	1
1999	6				3	1	2
2000	10		1		3	4	2
2001	7				3	1	3
Date unknown	7				3	3	1
1994 – 2001	59	2	9	2	24	12	10
Building phase	9	4	2	0	1	2	0
No implementation	124	45	38	12	18	9	2

Table 1 The first year of national clearinghouse implementation (distributed per region).

Region	Average	Standard Deviation	Median	Maximum number
Africa (2)	11	11	10	19
Asia (9)	7	6	5	16
Australasia & Oceania (2)	14	17	14	26
Europe (24)	33	49	12	133
North America (12)	204	551	6	1758
South America (10)	4	3	2	8
World (59)	54	239	6	1758

Table 2 The number of data suppliers of national clearinghouse per region. The number in parentheses is the number of clearinghouses analyzed per region.

Region	Abstract	Metadata	Data (+ metadata)
Africa (2)	1	1	0
Asia (9)	1	6	2
Australasia & Oceania (2)	0	1	1
Europe (24)	6	16	2
North America (12)	0	9	3
South America (10)	2	8	0
WORLD (59)	10	41	8

Table 3 The type of data accessibility of national clearinghouse per region. The number in parentheses is the number of clearinghouses analyzed per region.

clearinghouse is high; however, there is great variety between the clearinghouses (Table 2). For Austria, the Czech Republic, Slovenia, and the U.S., the number of data suppliers exceeds 100. In Canada, there are 1758 data suppliers. This contrasts with the 35 clearinghouses that have fewer than 10 suppliers (notably in South America and Asia with their powerful national mapping agencies).

3. The Type of Data Accessibility

This characteristic describes the presentation of the content. Not all existing clearinghouses give access to data or metadata. For example, in some cases the clearinghouse presents only a simple (not standardized) description of the datasets. For this reason, three classes of accessibility are distinguished: 1) abstract (simple/short description about the databases without using any formal meta-data description); 2) metadata; and 3) data (+metadata). In most clearinghouses, the user has access to metadata (Table 3). However, in eight countries (Australia, Canada, Dominica, Finland, Malaysia, Portugal, Singapore, and the U.S.), an option exists to access the data itself.

Region	CEN	FGDC	National	Others
Africa (1)		1		
Asia (8)		3	2	3
Australasia & Oceania (2)			1	1
Europe (18)	7		4	7
North America (12)		11	1	
South America (8)		8		
WORLD (49)	7	23	8	11

Table 4 Metadata-standards as used by national clearinghouses per region. The number in parentheses is the number of clearinghouses analyzed per region.

Region	Average number	Standard deviation	Total number of data sets	Median	Minimum	Maximum
Africa (2)	1561	2198	3122	1561	6	3115
Asia (6)	676	857	4056	260	9	1782
Australasia & Oceania (2)	15,031	21,249	30,062	15,031	5	30,056
Europe (22)	440	867	9768	118	8	3011
North America (10)	11,802	31,089	118,020	211	8	99,649
South America (8)	721	1646	5768	38	7	4735
WORLD (50)	3616	14,618	170,796	111	5	99,649

Table 5 The number of datasets described within national clearinghouses per region. The number in parentheses is the number of clearinghouses analyzed per region.

4. The Metadata-Standard Used

This characteristic describes the metadata-standard used. With the diverse sources from which spatial databases are built, it is extremely important to maintain information about the content, quality, source, and lineage of the data. A number of standard organizations have developed (or are in the process of developing) standards for storing and maintaining metadata. The most mature of these have been developed by the Federal Geographic Data Committee (FGDC) (1995) and the European Committee for Standardization (CEN/287 1996). These metadata-standards form the backbone of national clearinghouses. The FGDC metadata-standard is the most applied and distributed standard around the world (Table 4). The CEN standard is only applied in Europe. Recently, the International Organization of Standardization has created the ISO19115 standard (ISO/TC-211 2001). Currently, 10 countries have started a project to apply this last-mentioned standard for their national clearinghouse

5. The Number of Spatial Datasets

A means to quantify the content of a clearinghouse is the number of datasets. However, it does not represent the importance of the accessible datasets to the economic and social development of the country. The variety in the number of datasets is enormous (Table 5). For example, the U.S. federal clearinghouse can give access to almost 100,000 datasets (December 6, 2001), while the average of the 24 European clearinghouses is 440. The difference in the total number of accessible datasets between the U.S. and Europe is easily noticed (100,000 vs. 10,000). In total, the clearinghouses describe 170,000 spatial datasets together. 10 clearinghouses have more than 1000 datasets described (Australia, Austria, Canada, the Czech Republic, Japan, Mexico, South Africa, Switzerland, Uruguay, and the U.S.).

6. The Most Recently Produced Dataset

This characteristic describes the up-to-date nature of content and the management of content in the clearinghouse. It is

Region	Average Duration (months)	Standard Deviation	Median (months)	Mini-mum	Maxi-mum
Africa (2)	31	41	31	2	60
Asia (6)	47	19	55	23	66
Australasia & Oceania (2)	2	1	2	1	2
Europe (22)	21	31	9	1	126
North America (10)	15	12	17	1	29
South America (8)	73	106	43	1	257
WORLD (50)	28	44	15	1	257

Table 6 The duration in months between date of the Web survey (December 2001) and the date of the most recently produced dataset. The number in parentheses is the number of clearinghouses analyzed per region.

Region	Number of Web References "AltaVista"				Number of Web References "Google"			
	Ave-rage	STD	Median	Maxi-mum	Ave-rage	STD	Median	Maxi-mum
Africa (2)	50	56	50	89	41	43	41	71
Asia (9)	151	148	105	477	59	112	52	175
Australasia and Oceania (2)	3084	2851	3084	5100	1315	913	1315	1960
Europe (24)	320	792	42	3642	123	129	67	502
North America (12)	96	168	40	480	146	351	55	1080
South America (10)	112	140	48	428	76	78	53	213
WORLD (59)	312	857	50	5100	145	309	54	1960

Table 7 The number of Web references by AltaVista and Google search engines per region (STD = STANDARD deviation). The number in parentheses is the number of clearinghouses analyzed per region.

the difference in months between the date of Web survey and the date of the most recently produced dataset described in the national clearinghouse. On average, the time of the production of the most recent dataset is more than 2 years (Table 6). However, 22 national clearinghouses describe spatial datasets produced within 1 year of the Web survey. However, for 12 national clearinghouses, this duration is longer than 3 years (mainly countries located in South America or Asia).

7. The Number of Web References

This number can be interpreted as a means to measure the popularity (use) of the clearinghouse site within the Internet network. The "Free Link Popularity Service" <http://www.linkpopularity.com> (The PC Edge, Inc.) is used, which measures the number of links to the home page of the national clearinghouse that can be checked by the following search engines: "AltaVista" and "Google." A well-linked popularity can dramatically increase traffic to the specific

Web site. The link popularity of national clearinghouse is high, which means that they are an excellent source of consistent and targeted Web traffic. However, the variety is enormous (Table 7). The "Number of Web references" does not differ that much between the regions and so the popularity of a national clearinghouse can be considered as universal. The following national clearinghouses have high link popularity: Australia, Canada, Colombia, Finland, New Zealand, Norway, the U.S., and Venezuela.

8. The Number of Monthly Visitors

This characteristic describes the use of national clearinghouses for accessing spatial datasets. This amount is related to the number of visitors who have visited the homepage of the clearinghouse. The average number of visits of this page exceeds the 5000 visitors. It is worth noting that the variety between the implementations is high due to some particularly popular clearinghouses (Table 8). The following national clearinghouses are visited the most: Canada, Finland,

Region	Average	Standard Deviation	Median	Minimum	Maximum
Africa (1)	423		423		
Asia (5)	1 055	382	1 120	618	1576
Australasia & Oceania (1)	4 378		4 378		
Europe (10)	10 521	18 571	1 743	410	60 000
North America (6)	5 384	7 492	1 973	328	18 700
South America (3)	1 684	944	1 517	835	2 700
WORLD (26)	5 871	12 337	1334	328	60 000

Table 8 The monthly number of visitors per region. The number in parentheses is the number of clearinghouses analyzed per region. (note: the number of clearinghouses is much lower as illustrated in the other tables due to the fact that not all clearinghouse sites are able to count the number of visitors).

Region	Average Duration (days)	Standard Deviation	Median (days)	Clearinghouses Updated Within 1 Day	Clearinghouses Updated more than 100 Days Ago
Africa (2)	902	1270	902	0	1
Asia (9)	482	723	7	4	3
Australasia & Oceania (2)	12	13	12	0	0
Europe (22)	195	312	27	6	7
North America (9)	3	3	2	4	0
South America (10)	37	42	26	3	1
WORLD (54)	201	440	15	17	12

Table 9 The frequency of Web updates per region (days). The number in parentheses is the number of clearinghouses analyzed per region.

Portugal, Slovenia, and the U.S.; Portugal's clearinghouse has approximately 60,000 visits per month.

9. The Frequency of Web Updates

This characteristic describes the management of the content in the clearinghouse. One possible indication of a well-managed clearinghouse can be seen by the frequency of updated information. The average number of days of last update is high for the whole population of clearinghouses due to instances of poor management (with some updates exceeding 100 days) in Europe and Asia (Table 9). The variety between clearinghouses is high as, alongside the poorer managed clearinghouse, numerous excellently managed facilities operate (update within 1 day).

10. The Languages Used

This characteristic describes the number and diversity of users able to access data because of their familiarity and knowledge of the given language. 30 clearinghouses do not have a search mechanism written in English (in addition, five of these are written in Arabic, Chinese, Greek, Japanese,

and Korean script). 29 clearinghouses use only their home language. These language problems reduce the accessibility to data (for English-speaking people).

11. The Use of Maps for Searching

The use of this facility can improve the accessibility to data. In 18 clearinghouses, maps can be used as an option to search for (meta)data. This relatively advanced alternative for searching is popular in Europe and Asia.

12. Registration-Only Access

This characteristic describes the management and possible limitations of use. Before accessing the data, users must register themselves by entering personal details. This characteristic could have a negative impact on accessibility. For eight national clearinghouses, the user is required to register to access metadata or data (Canada, El Salvador, Finland, Hungary, Malaysia, Singapore, Spain, Canada, and Uruguay).

Conclusions

Since 1994, the number of national clearinghouses has steadily increased to a total of 59. Looking at the trend of implementation, countries can expect to see additional national clearinghouses established. In fact, building clearinghouses is a global activity (with the exception of Africa and the Middle East (as well as Australasia and Oceania)). Most existing clearinghouses are established in Europe, Southeast Asia, and North and South America. The main initiatives for establishment come from Anglo-Saxon countries, such as the U.S., South Africa, and Australia. The U.S., in particular, which is supported by the FGDC, has stimulated many (American continent) countries to build a clearinghouse. However, 124 countries have still not shown any initiative to build one. There are several reasons for this. For example, a country may not have appropriate network architecture or there may be institutional bottlenecks for implementation. The differences in content, use, and management between the clearinghouses are broad. An example of such broad difference in content is the total number of accessible datasets described in a clearinghouse. In the U.S. clearinghouse, this number is 10 times as high as the total number of all 24 European clearinghouses. The reason for such difference is due to each country's unique historical, institutional, economic, legal, technical, and cultural setting. Especially in Europe, there are great contrasts in the number of datasets, suppliers, visitors, Web references, and frequency of Web updates, probably as a result of the high institutional, economic, legal, technological, and cultural diversity within this region. However, similarities between clearinghouses do exist (for example, the type of data accessibility and the metadata-standard used).

The most applied metadata-standard is the FGDC. However, looking to the numerous projects to apply the ISO standard, it is likely that ISO19115 will be the most applied standard in the future. This international consensus standard reflects FGDC, CEN, and other inputs. It provides detail that goes beyond FGDC and CEN metadata, including special coverage of raster and imagery information. Currently, there are several initiatives to create implementable subsets and extensions of ISO19115 so that conversion of FGDC-support tools and implementations to meet ISO conformance requirements is facilitated (Federal Geographic Data Committee Metadata Staff Coordinator 2001). Looking to the average number of data suppliers, Web references, and visitors, we can conclude that national clearinghouses are a popular facility to distribute and access spatial data.

Finally, in the future, it is highly probable that many national clearinghouses will give access to spatial data itself and provide complementary services such as online mapping. However, a concern could be the low frequency of Web updates of several clearinghouses due to poor management. Therefore, special attention has to be given to keep clearinghouse managers motivated for having a well-managed clearinghouse. Based on the 12 characteristics used, we can conclude that Australia, Canada, Portugal, and the U.S. have the best existing national clearinghouses. Additionally, this Web survey shows that not only the richest countries have good clearinghouses. Examples of relatively poorer countries with

suitable national clearinghouse are El Salvador, Nicaragua, and Uruguay. Based on the above research, for all countries, it seems that one of the keys for successful clearinghouse implementation is high political support and interest by means of funding and long-term strategy.

About the Authors

Joep Crompvoets is Assistant Professor of Geo-Information Science at the Wageningen University in The Netherlands. From 1999, he has been monitoring the development of national clearinghouses and spatial data infrastructures around the world.

Corresponding Address:
Wageningen University
Arnold Bregt
Centre for Geo-Information
P.O. Box 47, 6700 AA, Wageningen
The Netherlands
joep.crompvoets@wur.nl

Arnold Bregt is Professor of Geo-Information Science at Wageningen University. Following more than 15 years of experience in the field of geographic information systems research and applications, his current areas of interest are spatial data quality, dynamic modeling of land use change, and spatial data infrastructures. From 1996 to 1998, he was one of the project leaders to develop a national clearinghouse for spatial data in The Netherlands.

Corresponding Address:
Wageningen University
Arnold Bregt
Centre for Geo-Information
P.O. Box 47, 6700 AA, Wageningen
The Netherlands
arnold.bregt@wur.nl

Acknowledgements

The authors would like to thank Dr. Gert Jan Hofstede and Dr. Monica Wachowicz of Wageningen UR for their critical comments and suggestions on earlier versions of this article. Additionally, the authors appreciate very much the numerous reactions of the Webmasters to all their questions.

References

- CEN/TC 287, 1996, Geographic Information-Data Description-Metadata, Technical Report, prEN12657 (Brussels: CEN).
- Clinton, W., 1994, Coordinating Geographic Data Acquisition and Access to the National Spatial Data Infrastructure. Executive Order 12096, Federal Register 59, 17671-4, (Washington, D.C.).
- Dorling Kindersley, 1997, World Atlas (London: Dorling Kindersley Limited), 402 pp.
- Federal Geographic Data Committee, 1995, Content Standards for Digital Geospatial Metadata Workbook (Washington, D.C.).
- Federal Geographic Data Committee, 1997, Metadata to Clearinghouse Hands-On Tutorial. <http://www.fgdc.gov/clearinghouse>
- Federal Geographic Data Committee (Metadata Staff Coordinator), 2001, Proposal for a National Spatial Data Infrastructure Standards Project, June 2001. http://www.fgdc.gov/standards/status/iso_metadata.doc
- Hutton, 1990, Survey Research for Managers: How to Use Surveys in Management Decision-Making, 2nd edition (Basingstoke: Macmillan)
- ISO/TC-211, 2001, Geographic Information: Metadata, International Standard 19115.
- Lance, K. and G. Hyman, 2001, Adoption and Implementation of National Spatial Data Infrastructure in Latin America and the Caribbean. 5th Global Spatial Data Infrastructure Conference, Cartagena de Indias, Colombia, May 2001. http://codazzi4.igac.gov.co/gsdi5/m_sesion7.htm
- McLaughlin, J., 1991, Towards National Spatial Data Infrastructure. Proceedings of the 1991 Canadian Conference on GIS, Ottawa, Canada, Canadian Institute of Geomatics, Ottawa, Canada, March 1991, 1-5.
- Onsrud, H.J., 1998, Compiled Responses by Question for Selected Questions. Survey of National and Regional Spatial Data Infrastructure Activities Around the Globe. Global Spatial Data Infrastructure survey. <http://www.spatial.maine.edu/~onsrud/gsdi/Selected.html>