

Global Developments of National Spatial Data Clearinghouses

J.W.H.C. Cromptvoets
Wageningen University, Centre for Geo-Information
P.O. Box 47, 6700 AA, Wageningen, The Netherlands
Joep.Cromptvoets@Staff.GIRS.WAU.nl

Abstract

In March 2000, 2001, 2002 and November 2000, 2001 a survey was carried out to assess systematically the state-of-the-art of national spatial data clearinghouses around the world. The aim for this survey was to evaluate the progress and spatial distribution of clearinghouse implementations and to find out similarities and differences between them. Each clearinghouse was characterized by the following criteria: year of first implementation; number of data suppliers; type of data accessibility; metadata-standard; number of datasets; most recently produced dataset; number of web references; monthly number of visitors; frequency of web updates; language used; use of maps for searching and registration-only access. For 59 countries national clearinghouses have already been established. These clearinghouses show a large variety in history, use, content and management. The implementation of national clearinghouses can be considered globally as successful. However, awareness to improve management is needed to keep them a success

INTRODUCTION

At present, out of the 192 countries in the world, 120 are working on their National Spatial Data Infrastructure (NSDI) in order to access spatial data. One of the main components of a NSDI is a national clearinghouse (McLaughlin 1991, Clinton 1994, FGDC 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 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 a fast development in implementation started around the world. Not many studies exist about the worldwide development of national clearinghouses. To the best of my knowledge, no systematic periodical research has taken place with regard to the development of national clearinghouses. The results of this survey are an attempt to fill this gap. The survey's main objectives are to assess the worldwide progress, the spatial distribution and to evaluate both similarities and differences of national clearinghouses. This paper presents the development results between the period of March 2000 and March 2002. Large parts of the text used is derived from the Cromptvoets and Bregt (2002)

SURVEY METHODOLOGY

In order to assess national clearinghouse developments around the world, a longitudinal survey was undertaken. This began in March 2000 and was conducted in a systematic and periodical manner (March 2000, November 2000, March 2001, November 2001 and March 2002). The main methodology 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 of the clearinghouses.

The inventory (step 1) is compiled by extensive browsing of the Internet, reading related literature and contacting experts and several webmasters. Clearinghouses are characterized (step 2) based on the following criteria: ease of measurement, objective character and clear presentation of history, content, use and management of the national clearinghouse. The following 12 characteristics are measured: 1) year of first implementation; 2) number of data suppliers; 3) type of data accessibility; 4) metadata-standard used; 5) number of datasets, 6) most recently produced dataset; 7) number of web references (Alta Vista and Google); 8) monthly number of visitors; 9) frequency of web updates; 10) languages used; 11) use of maps for searching and 12) registration-only access. Almost all the above information was sourced from clearinghouse web pages. Additionally, in cases of uncertainty or missing data the webmaster was contacted.

In next chapter, each of the used characteristics will be explained and its developments will be discussed in greater depth.

RESULTS AND DISCUSSION

The main results of this longitudinal survey based on the 12 mentioned characteristics are presented and discussed. Some of the results are presented by region. This division in regions is derived from the Dorling Kindersley's world atlas (1997)

1. Year of first implementation. This characteristic describes the history of the clearinghouse. From 1994, the number of national clearinghouses has been steadily increasing with time (Figure 1). Now 59 countries have an implemented version on the web (for 52 clearinghouses the year of first implementation is known). Additionally, it is known that 9 countries have projects for implementation. However, the variety in number between the different regions is considerably (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 is less than 5 %. Another important remark is that 124 countries did not conduct any initiative to build such a national facility.

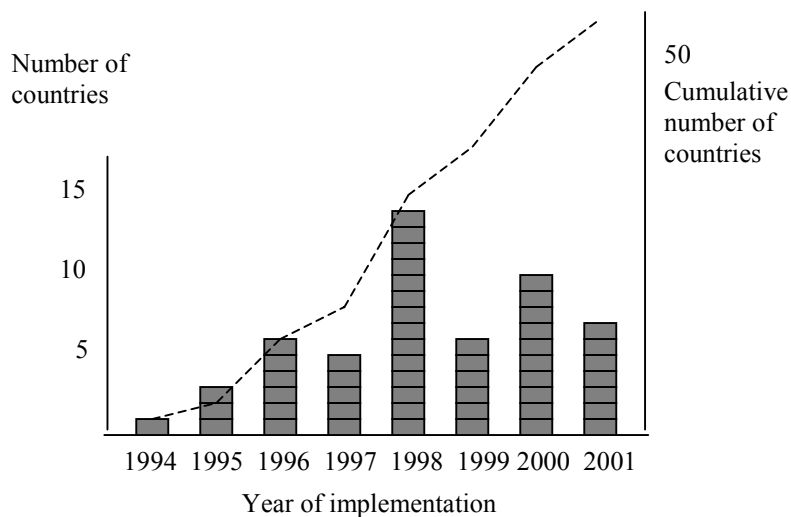


Figure 1 First Year of national clearinghouse implementation (per year (columns) and cumulative (dashed line)).

First year of implementation	Total Number countries	Number African countries	Number Asian countries	Number Australian countries	Number European countries	Number North-america countries	Number 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
Unknown date	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 First year of national clearinghouse implementation (distributed per region).

2. 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 clearinghouse is high, however the variety between the clearinghouses is extremely high (Table 2). For example, Canada clearinghouse is supported by more than 1800 suppliers. This is in contrast with the more than 35 clearinghouses, which have less than 10 suppliers. Remarkable is the fact that the median number of data suppliers of all the clearinghouses (a possible means which describes the value for the most representative national clearinghouse) is decreasing during the time.

	2000 November	2001 November	2002 April
Number clearinghouses	35	59	60
Average	34	54	52
Standard deviation	89	239	238
Median	7	6	5
Maximum number	486	1758	1826

Table 2 Number of data suppliers of national clearinghouse.

3. 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 and only in a few cases he/she has direct access to data (Table 3). The development seems to be that more clearinghouses give the user access to metadata and data.

Accessibility type	2000 March	2000 Nov.	2001 Nov.	2002 March
Number of clearinghouses	35	46	59	59
Abstract	12	15	10	9
Metadata	17	24	41	42
Data (+ metadata)	6	7	8	8

Table 3 Type of data accessibility of national clearinghouse.

4. 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 are developed by the Federal Geographic Data Committee (FGDC 1995) and the European Committee for Standardization (CEN/287 1996). These meta-standards form the backbone of national clearinghouses. The FGDC metadata-standard is the most applied and distributed one around the world (Table 4). Additionally, the use of this standard is steadily increasing during time. Recently, the International Organization of Standardization TC/211 has created the ISO19115 standard (ISO/TC-211 2001). Now, 10 countries have already started projects to apply this last-mentioned standard for their national clearinghouse.

	2000 March	2000 November	2001 November
Total number of clearinghouses	25	30	39
CEN	7	8	7
FGDC	14	18	23
National	4	4	9
Others	3	4	11

Table 4 Metadata-standards as used by national clearinghouses per time.

5. 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, US federal clearinghouse gives the user access to more than 135 000 datasets, while the 24 European clearinghouses together give access to $\pm 10\,000$ datasets. Remarkable is that the average and median number per clearinghouse is decreasing during time. Nevertheless, the total summed number of datasets of all the clearinghouses together is steadily increasing.

	2000 March	2000 Nov.	2001 Nov.	2002 March
Number clearinghouses	32	37	50	56
Average	4 699	4 427	3 616	3 653
Standard dev.	16 102	16 987	14 618	18 418
Median	153	105	111	74
Minimum	5	5	5	5
Maximum	84 675	96 787	99 649	135 049
Summation of all datasets	150 359	163 786	170 796	204 597

Table 5 Number of datasets described per time.

6. Most recently produced dataset. This characteristic describes the up-to-date nature of content and the management of content in the clearinghouse. It is the difference in months between the date of survey and date of most recently produced dataset described in the national clearinghouse. The average time of the production of the most recent dataset is more than 2 years and the median is 15 months (Table 6). The development seems that no much is changing during time. This is only based on 2 times of survey.

	20001 Nov.	2002 March
Number clearinghouses	50	51
Average	28	29
Standard deviation	44	44
Median	15	15
Minimum	1	1
Maximum	257	261

Table 6 Duration in months between date of survey (November 2001 or March 2002) and date of most recently produced dataset per time.

7. 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. Used is the 'Free Link Popularity Service' <http://www.linkpopularity.com> (The PC Edge, Inc.) which measures the number of links to the homepage of the national clearinghouse that can be checked by the following search engines: 'Alta Vista', 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 development is that the number of web references by using AltaVista search engine is decreasing since March 2001, meanwhile the number by Google is steadily increasing during time.

AltaVista	2000 March	2000 Nov.	2001 March	2001 Nov.	2002 March
Number	34	42	43	56	58
Average	77	182	318	301	252
Standard dev	103	458	828	850	637
Median	26	29	57	46	54
Maximum	389	2500	4 926	5100	3 264
Google	2000 March	2000 Nov.	2001 March	2001 Nov.	2002 March
Number			43	56	58
Average			62	147	251
Standard dev			123	309	488
Median			35	63	94
Maximum			1	1960	2750

Table 7 Number of web references by AltaVista and Google search engines per time (STD: Standard deviation).

8. Monthly number of 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 (however the median value is just above 1 000). It is worth noting that the variety between the implementations is high due to some particularly popular clearinghouses (US, Portugal and Slovenia) (Table 8). The recent development is that the number is stabilizing and not increasing.

	2000 March	2000 Nov.	2001 March	2001 Nov.	2001 Apr,
Number	18	27	24	26	20
clearinghouses					
Average	5 012	5 552	5 700	5 871	5 876
Standard dev.	12 842	11 833	12 443	12 337	13 700
Median	555	835	1 088	1 334	1 280
Minimum	89	101	307	328	115
Maximum	53 557	57 222	59 121	60 262	60 679

Table 8 Monthly number of visitors per time.

(* note: the number of clearinghouses is much lower as illustrated in the other tables).

9. 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). However, the median value is low (Table 9). The variety between clearinghouses is high as, alongside the poorer managed clearinghouse, numerous excellently managed facilities operate (update within 1 day). The development is that the clearinghouses are less updated, especially the median value shows almost a continuous increase.

	2000 March	2000 Nov.	2001 March	2001 Nov.	2002 March
Number of clearinghouses	24	31	54	54	54
Average	122	169	122	168	190
Standard dev.	244	332	338	398	447
Median	9	6	15	18	27
Clearinghouses updated < 1 day	5	4	8	17	11
Clearinghouses updated > 100 days	7	11	7	12	17

Table 9 Frequency of web updates per time (days).

10. 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. Many clearinghouses do not have a search mechanism written in English. Most of these use only the language(s) used in the country. Happily, the number of multilingual clearinghouses is increasing.

	2000 March	2000 Nov.	2001 March	2001 Nov.	2002 March
Number analyzed clearinghouses	31	42	43	56	58
English	17	22	22	33	34
Spanish	6	12	13	15	17
Multilingual	8	11	11	19	21
Only home language	20	26	27	30	31

Table 10 Languages used per time

11. Use of maps for searching. The use of this facility can improve the accessibility to (meta)data. This implementation encourages visible success. The number of maps used as search mechanism is slightly increasing during time.

	2000 Nov.	2001 March	2001 Nov.	2002 March
Number analyzed Clearinghouses	42	43	56	58
Maps used	16	16	20	22

Table 11 Maps used for searching per time

12. Registration-only access. This characteristic describes the management and possible limitations of use. Before accessing the data, users have to register themselves by entering personal details. This characteristic could have a negative impact on accessibility. Only for a few clearinghouses, users have to register themselves.

	2000 Nov.	2001 March	2001 Nov.	2002 March
Number analyzed Clearinghouses	42	43	56	58
Registration needed	6	7	8	7

Table 12 Registration-only access per time.

CONCLUSIONS

From presented developments of national spatial data clearinghouse implementations, we can expect that more and more national clearinghouses will be established. In the future, many national clearinghouses will give access to spatial data itself and provide complementary services such as online mapping.

Building clearinghouses is a global activity (with the exception of Africa (and Australasia & Oceania)). Most existing clearinghouses are established in Europe, Southeast Asia, North and South America. However, numerous 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 have institutional bottlenecks for implementation.

The differences in content, use and management between the clearinghouses are broad. The main reason, for these differences, is due to each country's unique historical, institutional, economic, legal, technical and cultural setting. Especially in Europe, the contrasts in number of datasets, suppliers, visitors, web references and frequency of web updates are high, probably resulting from the high institutional, economic, legal, technological and cultural diversity within this region.

Looking to all presented numbers, we can conclude that the introduction of national clearinghouses for the distribution and access of spatial data was successfully implemented by different countries. However, we were able to observe a recent decrease of interest in managing national clearinghouses, which is a warning development. This observation is based on the decreasing number of data suppliers and frequency of web updates, the stabilizing number of datasets, number of visitors and most recently produced datasets. These trends are almost all related to a more poor management of data providers and (especially) clearinghouse managers. Therefore, their awareness to improve this situation is needed. Especially now, the challenge is to keep spatial data providers, clearinghouse managers and users motivated for and informed about this mechanism. This is really needed when we consider all the global activities related to the implementation of national spatial data infrastructures wherein clearinghouses are the key component.

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, DC, 2pp.

Crompvoets, J. and A. Bregt, 2002. World status of National Spatial Data Clearinghouses. Urban and Regional Information Systems Association. Submitted.

Dorling Kindersley, 1997. *World atlas*. Dorling Kindersley Limited, London. 402 pp.

Federal Geographic data committee, 1997, *Metadata to Clearinghouse Hands-on Tutorial* (<http://www.fgdc.gov/clearinghouse>).

Federal Geographic data committee, 1995, *Content Standards for Digital Geospatial Metadata Workbook*, Washington DC.

ISO/TC-211, 2001, *Geographic Information: Metadata, International Standard 19115*.

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, pp. 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>.