Exploring worldwide the impact of society on the success of national spatial data clearinghouses

Joep Crompvoets*[a], Arnold Bregt[a], Monica Wachowicz[a], Gert Jan Hofstede[b], Pieter Vereijken[c], Erik De Man[d]
[a]Centre for Geo-Information, P.O.Box 47, Wageningen UR, The Netherlands (joep.crompvoets@wur.nl)
[b]Information Technology Group, Wageningen UR, The Netherlands
[c]Biometris, Wageningten UR, The Netherlands
[d]ITC, Urban and Regional Planning and Geo-Information Management, Enschede, The Netherlands

Abstract
This paper provides the main results of a worldwide exploration of societal impact on the success of national spatial data clearinghouses for the situation of December 2002. The aim is to explore why several countries have successful clearinghouses, and to identify the societal factors for success in order to enhance the understanding of the societal factors contributing to future clearinghouse advancement. A clearinghouse suitability index was developed in order to indicate the success of clearinghouses. The society of each country was characterised by data of attributes relating to nine societal aspects (economy, education, technology, environment, culture, demography, institution, health care, and jurisdiction). Several explorative analyses were performed based on statistics (e.g. analysis of variance, partial least squares regression). Moreover, the success of potential clearinghouses was predicted for countries with no previously existing facility. These prediction results were compared with the results of the web survey of April 2005. The findings of the analyses highlight the importance of certain societal conditions for the success of national clearinghouses. However, society alone does not fully determine success, since clearinghouse-internal factors such as the introduction of web services, stability of funding and creation of user-friendly interfaces are also important. From a societal perspective, success is likely to be dependent on a combination of attributes from different societal aspects, in particular economy, education and technology. Critical factors for success could be those that raise the level of wealth in the country (e.g. taxes on income, gross domestic product, education expenditure and Internet accessibility). In the near future newly established national clearinghouses, in countries where they have not previously been established, are likely to face serious obstacles or may even fail as result of unfavourable societal conditions. This could, however, be reversed if critical clearinghouse-internal factors are implemented.

Keywords: national spatial data clearinghouse, societal impact, critical success factors, partial least squares regression

1. INTRODUCTION

In April 2005, 83 countries had established a national spatial data clearinghouse on the Internet. It is expected that more countries will establish a national clearinghouse in the near future. In order to facilitate the access, use and dissemination of spatial data and related services, these clearinghouses are regarded as key features of a National Spatial Data Infrastructure (Crompvoets and Bregt, 2003; Crompvoets et al, 2004). However, at present only a few clearinghouses are highly functional, in the sense that they provide efficient and effective facilities for spatial data/service accessibility, use, and dissemination (Crompvoets et al, 2005; Crompvoets and Bregt, 2007).

It appears that socio-technical issues including those which are economic, educational, cultural, institutional, legal, political and organisational, are currently the major impediments to the success of national spatial data clearinghouses, rather than technical issues alone (Nevodic-Budic, 1999; Bregt, 2000; Groot and McLaughlin, 2000; Williamson et al, 2003; Tait, 2005; Bernard et al, 2005b, Bouckaert et al, 2006). Implementing clearinghouses appears to be a complex task, fraught with difficulties in sustaining a shared language, a shared sense of purpose, and reliable financing. Because of these impediments, spatial data accessibility, use and dissemination in many countries is not optimal.

In order to improve this situation, it is essential that clearinghouse coordinators, practitioners and policy makers are made more aware of the factors, which could determine the success of a national clearinghouse. These factors could be societal or clearinghouse-internal. Examples of such clearinghouse-internal factors could be: the introduction of web services, stability of funding, and creation of user-friendly interfaces (Crompvoets et al, 2004). This paper focuses on the societal factors. Being aware of certain societal conditions for the success of clearinghouses as part of the National Spatial Data Infrastructure might help the understanding of coordinators, and practitioners to overcome difficulties associated with setting up efficient and effective clearinghouses and so National Spatial Data Infrastructures.

A growing body of literature reflects an increasing interest in societal issues as factors that are critical to successful implementation of national spatial data infrastructure and/or spatial data clearinghouses (Tosta, 1997; Masser, 1999; 2005; Groot and McLaughlin, 2000; De Man, 2000; De Man and van den Toorn, 2002; Crompvoets and Kossen, 2002; Rajabifard et. al, 2002; Ravi, 2003; Williamson et al, 2003; Spatial Application Division, Catholic University of Leuven, 2003; Craglia et al, 2003; Kok and Van Loenen, 2005; Wehn de Montalvo, 2004; Reece, 2004; Delgado et al, 2005; Bernard et al, 2005a,b; Tait, 2005; Bouckaert et al, 2006). However, comprehensive and systematic knowledge in this domain is currently limited.

A web survey focussing on the worldwide status of national clearinghouses was conducted in December 2002. It appeared that 67 countries had established a national clearinghouse on the Internet, 13 countries had projects to establish such an electronic facility, and 113 countries had taken no action to establish one (Crompvoets et al, 2004).

When considering the strategies for the establishment and maintenance of national clearinghouses, it is very likely that no single best solution or recipe exists since each country has its own unique society. However, exploring the societal impact on clearinghouse success and identifying societal factors for success might support clearinghouse coordinators and policy makers in the development of

successful strategies for establishing or maintaining national clearinghouses. To the best of the authors' knowledge, no systematic and comprehensive evaluation has taken place with regard to the societal impact and factors on the success of national clearinghouses. The purpose of the present paper is to fill this gap.

The main objectives are to explore societal impact worldwide on the success of national clearinghouses based on the situation of December 2002, and to identify the critical societal factors for success, and to predict clearinghouse success in countries where it has not previously been established.

2. METHODOLOGY

This study can be best described as empirical and explorative. The main methodological steps are: 1) Indexing suitability, 2) Collecting societal data, 3) Exploring societal factors for success, and 4) Predicting clearinghouse success in countries with no clearinghouse established. Each of these steps is described in more detail.

2.1 Indexing suitability

A suitability index evaluating the suitability of clearinghouses to facilitate the spatial data/service accessibility, use, and dissemination was developed in order to indicate a measure for the success of a national clearinghouse implementation, using the 15 clearinghouse characteristics as described by Crompvoets et al (2004). The selection of these characteristics was based on the following criteria: ease of measurement, objective character and clear presentation of the people (suppliers, end-users), spatial data, technology, policy, and standards of national clearinghouses. The following characteristics were selected: Number of suppliers, Monthly number of visitors, Number of web references, Languages used, Frequency of web updates, Level of (meta)data accessibility, Number of datasets, Most recently produced dataset, Decentralised network architecture, Availability of view (web mapping) services, Number of mechanisms (alternatives) for searching, Use of maps for searching, Registration-only access, Funding continuity, and Metadata-standard applied. It is assumed that these characteristics represent the key criteria for determining the suitability of the national clearinghouse. However, with some imagination other criteria could be added. Nevertheless, it is an appropriate approach to index the suitability of all existing national clearinghouses of the world on an easy and objective way.

A survey was undertaken to determine weights that indicate the importance of each characteristic for evaluating the clearinghouse suitability. A questionnaire was distributed to ± 500 European representatives of the GI-community (e.g. ministries, municipalities, mapping agencies, cadastres, universities, public/private institutions, utilities, etc). This survey was strongly supported by the INSPIRE expert group (a group composed of representatives of the European Commission, and member states' Environmental and GI-communities) and EUROGI (the umbrella organization that represents the European GI-community). The representatives were asked to determine a weight between 0.00 and 1.00 of each clearinghouse characteristic. The summation of the 15 characteristic weights together should have been 1.00. The resulted weight values were averaged by the number of responders. The representatives were also asked to classify the domains of each characteristic into 3 classes taking into considering that the lowest suitability class should have a class weight of 0.00 and that the middle class should have a class weight half of the determined characteristic weight (see Table 1). For the quantitative characteristics

the median values of the classified borders were used to distinguish the classes since the median is less sensitive to extremes than the mean; for the nominal characteristics the most frequently mentioned answers were used. A description of each characteristic as presented in Crompvoets et al (2004) was enclosed to the questionnaire.

In order to index the suitability of the clearinghouses in December 2002, a web survey was undertaken consisting of an inventory of existing national clearinghouses on the Internet and measuring the above-mentioned characteristics. All the data collected during this web survey were classified into a characteristic class, and assigned a class weight. The result of the summation of the 15 class weights together formed the Clearinghouse Suitability Index (CSI) ranging between 0.00 and 1.00; 0.00 meaning that the national clearinghouse is not suitable; 1.00 means very suitable. In total, the suitabilities of all the 67 national clearinghouses existing in December 2002 were indexed.

The main limitation for using these clearinghouse suitability indices is that the scores only indicate the (general) suitability of national clearinghouses and do not directly support clearinghouse coordinators, practitioners, and policy makers in the development of successful strategies for establishing and maintaining national clearinghouses by specifying some recommendations for improvement.

2.2 Collecting societal data

Data were collected for 193 countries as existing in 2002 according to United Nations. Data as much linked to the year 2002 were collected from the following sources: The World Bank (2003; 2004; 2005), International Telecommunications Union (2003; 2004), Central Intelligence Agency (2003; 2004), United Nations Development Program (UNDP) (2003; 2004; 2005) and the Dorling Kindersley's Great World Atlas (2002). These sources are the key documents that describe the status of countries at a worldwide level using basic indicators to be measured. The data on national culture were taken from Hofstede (2001). All attributes together describe the societal conditions of a country. These attributes collected were classified into nine societal aspects using primarily the different sections in The World Bank's World Development Indicators books and UNDP's Human Development reports: 1) Economy, 2) Education, 3) Technology, 4) Environment, 5) Culture, 6) Demography, 7) Institution, 8) Health care and 9) Jurisdiction (including legal and judicial aspects). Each societal aspect describes a specific part of society and consists of a set of underlying societal attributes. The selection of attributes was based on the following criteria:

- Having a potential link to spatial data infrastructures, and in particular to clearinghouses.
- Not being outdated (most data before 2000 were disregarded).
- Not missing too many data values. In this sense, missing values mean that no data value is available describing a country. Attributes with too many data values not available were considered to be incomplete.
- Representing one of the nine societal aspects.

Prior to analyses, pre-processing of data took place. This pre-processing included:

- Classifying countries into two establishment classes for a national clearinghouse using the results of the web survey conducted in December 2002: 1) countries with a national clearinghouse established, and 2) countries with no national clearinghouses established.

- Handling missing values. Countries as well as attributes with more than 25% of missing values were removed. The missing data values of the remaining attributes were replaced in order to perform the analyses. They were replaced by the median or the data value that occurs to be the most frequent of each establishment class being aware that the resulted dataset would become (slightly) biased; median for the quantitative attributes, the data value that occurs to be the most frequent for the nominal attributes. Due to the highly skewed distribution of most attributes, the median was chosen since this number is less sensitive to extremes than the mean

2.3 Exploring societal factors for success

The Clearinghouse Suitability Index (CSI) was used as the main indicator for success. In order to analyse comprehensively the societal factors for success, correlation coefficients, analysis of variance (ANOVA) and Partial Least Squares regression (PLS) were used. The analyses were restricted to 60 countries since seven countries (Barbados, Brunei, Dominica, Guyana, Iceland, Luxembourg and Qatar) had too many missing values.

Correlation coefficients between the CSI and the quantitative societal attributes were calculated. These coefficients (r) measure the degree of linear relationship between two variables.

For the nominal attributes referring to jurisdiction, the ANOVA-method (Snedecor and Cochran, 1980) was used to test differences between CSI-means of several legal systems, e.g. civil versus common law (p-value = 0.01).

In this research, numerous societal attributes were collected; many more than the number of countries. It was expected that many societal attributes were strongly related to each other giving multi-collinear data, and that no single attribute alone would make a clearinghouse successful (Nebert, 2004; Reece, 2004). Therefore, the PLS regression method of analysis was chosen (Martens and Næs, 1989; Geladi and Kowalski, 1996; Massart et al, 1997; Tobias, 1997; Kooistra, 2004) being aware that the datasets are large enough and numerically sufficiently conditioned to provide reliable results. This method made it possible to explore relationships between the CSI and attributes of each societal aspect, as well as all attributes of society. PLS is one of the most widely used methods for multivariate analysis applied to a broad range of fields (econometrics, chemistry, education, marketing, and the social sciences). It is the method for constructing predictive models when data are multi-collinear. The strength of PLS is on predicting responses rather than analysing the underlying relationship between variables (Tobias, 1997). PLS is based on latent attribute decomposition using two blocks of variables, blocks X and Y, which contained the societal attribute data and CSI indices respectively. The objective of the method was to find a small number of latent PLS-factors that are predictive for the CSI indices (block Y) using the societal attribute data (block X) efficiently. All PLS-analyses were carried out in Matlab™. Some standard techniques were taken from the PLS_Toolbox of Matlab™.

Due to the limited number of existing national clearinghouses, the dataset was not split into a training set and a separate test set. Instead 10-fold cross-validation was used to calibrate and assess the prediction capability of PLS-models.

The parameters used to assess the quality of fitting PLS-models (Kooistra et al, 2001) to CSIs were:

- The root-mean-square error of calibration (RMSEC), used to estimate the 'average' deviation of the model from the data.
- The percentage variance captured by the regression model (R^2).
- The root-mean-square error of cross-validation (RMSECV) used to measure 'average' predictive error. The optimal number of latent PLS-factors was determined by the lowest value of RMSECV.

2.4 Predicting clearinghouse success for countries with no clearinghouse established in December 2002

The strength of applying PLS-models is that they are predictive. The best PLS-model was applied to predict the CSI in countries where no clearinghouse had been established in December 2002. The results are an indication of society's capability of establishing successful national clearinghouses. The PLS-model was applied to 110 countries.

In order to evaluate the validity of the predictive results, a web survey was conducted in April 2005 (Crompvoets and Bregt, 2007). It appeared that then 83 countries had established a national clearinghouse on the Internet. CSI indices were calculated for those national clearinghouses that were newly established during the period December 2002 and April 2005. These CSI indices were compared with the prediction results.

3. RESULTS AND DISCUSSION

The presentation of the results is organised according to the methodology steps as presented before. The main results are discussed in turn.

3.1 Indexing suitability

Through indexing the clearinghouse suitability, an indication of clearinghouse success was achieved in an easy and transparent way. 126 European practitioners representing mainly mapping agencies, universities, (non)governmental organisations and public/private organisations responded to the survey, which is around 25% of the population. This percentage is in line with the responses to similar types of surveys (Hamilton 2003). The main limitation of this indexing is that only European representatives were approached. Approaching clearinghouse practitioners from other continents could have improved the quality of the index. The main results of this step are the characteristic weights, the domain classifications, and the CSI. Table 1 presents the description of the 3 classes, and their corresponding class weights of each clearinghouse characteristic. Be aware that the class weight of class 1 is the same as the characteristic weight. It appears that the Number of mechanisms (alternatives) for searching, Availability of view services, Level of (meta)data accessibility and Frequency of web updates have the highest weights, while Monthly number of visitors, Most recently produced dataset, and Funding continuity have the lowest. One reason for these differences could be that the meanings of services for searching, viewing or downloading are much more transparent and visible. This could have made the practitioners more aware of the need for these characteristics. Alternatively, the suitability of services may simply have been easier to evaluate.

Table 1: Description of the 3 classes and their corresponding class weights of each clearinghouse characteristic.

Clearinghouse Characteristic	Class 1	Class 1 weight*	Class 2	Class 2 weight	Class 3	Class 3 weight
Number of suppliers	> 16	0.08	2 - 16	0.04	1	0.00
Monthly number of visitors	> 4000	0.02	150 - 4000	0.01	< 150	0.00
Number of web references	> 250	0.04	20 - 250	0.02	< 20	0.00
Languages used	Multilingual including the national language	0.06	Monolingual using the national language	0.03	Monolingual using no national language	0.00
Frequency of web updates (in days)	< 4	0.10	4 - 365	0.05	> 365	0.00
Level of (meta) data accessibility	Data + standardised metadata	0.10	Standardised metadata	0.05	Non-standardised metadata	0.00
Number of datasets	> 1500	0.08	50 - 1500	0.04	< 50	0.00
Most recently produced dataset (in months)	< 2	0.02	2 - 60	0.01	> 60	0.00
Decentralised network architect.	Yes	0.08	Hybrid	0.04	No	0.00
Availability of view services	Yes	0.10	Prototype	0.05	No	0.00
Number of mecha-nisms (alternatives) for searching	> 5	0.18	2 - 5	0.09	1	0.00
Use of maps for searching	Yes, by locating an area of interest	0.04	Yes, by clicking on an area with predefined boundaries	0.02	No	0.00
Registration-only access	No	0.02	Partly	0.01	Yes	0.00
Funding continuity	Continuously funded	0.01	Piecemeal funded	0.01	Never funded	0.00
Metadata-standard applied	ISO/FGDC/CEN	0.07	National	0.03	No standard	0.00

* The same as the Characteristic weight

Table 2 presents the suitability indices of the 67 national clearinghouses in December 2002 by increasing CSI-values. From this evaluation, it appears that the national clearinghouses of Australia, Canada, Finland, UK and USA were evaluated to be the most suitable facilities (Table 2).

The low values of the mean (40) and median (24) of the CSI indicate that the suitability of the clearinghouses in December 2002 was still not high. These low values are in line with the findings of Crompvoets et al (2004) that the functional capabilities of clearinghouses in December 2002 did not fit the expectations of the practitioners. They expected more web services to have been provided and user-friendly interfaces.

The high diversity in values of the CSI is significant (standard deviation 22). It is important to be aware of the differences between countries like Denmark, Iceland,

Ireland, The Netherlands and Uruguay. From a technical perspective, these clearinghouses are similar (e.g. have similar network architectures and searching mechanisms). However, their suitability was evaluated differently due to differences in the other non-technical characteristics. Since these non-technical characteristics appear to be influenced by prevailing societal conditions, it is argued that the main reason for such CSI diversity is probably due to the fact that these clearinghouses were embedded within different societies, each with unique conditions.

Table 2: Clearinghouse Suitability Indices (CSI) of December 2002 by increasing CSI-values.

Country	CSI	Country	CSI	Country	CSI	Country	CSI
Greece	**8**	France	**29**	Croatia	**37**	Portugal	**48**
Panama	**9**	Barbados	**31**	Iceland	**37**	Dominica	**49**
Italy	**10**	Peru	**31**	Nicaragua	**37**	Czech Rep.	**51**
Luxemburg	**12**	Trinidad	**31**	El Salvador	**38**	Chile	**56**
Russia	**12**	Brunei	**33**	Estonia	**41**	Denmark	**57**
Ecuador	**13**	Costa Rica	**33**	Venezuela	**41**	Uruguay	**58**
Turkey	**14**	Dom. Rep.	**33**	Netherlands	**42**	Malaysia	**61**
South-Korea	**17**	Sweden	**33**	Argentina	**43**	Germany	**62**
Arab Emirates	**17**	Bolivia	**34**	Japan	**43**	Indonesia	**62**
Guyana	**18**	Spain	**34**	China	**44**	New Zealand	**63**
Uganda	**18**	Ethiopia	**35**	Singapore	**44**	South.-Africa	**64**
Slovak Rep.	**22**	Ghana	**35**	Norway	**45**	UK	**70**
Senegal	**24**	Honduras	**35**	Slovenia	**45**	Finland	**72**
Guatemala	**27**	Hungary	**35**	Austria	**46**	Australia	**83**
Qatar	**27**	Philippines	**36**	Colombia	**46**	Canada	**94**
Belgium	**28**	Switzerland	**36**	Ireland	**46**	USA	**97**
Iran	**28**	Brazil	**37**	Mexico	**46**		

3.2 Collecting societal data

Through data collected from 175 societal attributes, an indicative description considering the wide scope of society was achieved. Table 3 presents the number of attributes collected and some examples of attributes per societal aspect. The main sections of World Bank's World Development Indicators books (2003, 2004, 2005) and UNDP's Human Development reports (2003, 2004, 2005) were mainly used as the criteria for the attribute classification into the nine societal aspects. Crompvoets (2006) shows the full list of attributes collected (including units, attribute scales and sources). From the sources available, it appeared that more economic, technological, environmental and demographic attributes could be collected than cultural, institutional, and legal attributes.

Overall, collecting societal data was negatively affected by limited data availability and/or reliability. This lack of quality is primarily due to the weak statistical systems in many countries with the consequence that statistical methods, coverage, practices, and definitions differ widely (International Telecommunication, 2003, 2004; United Nations Development Programme, 2003, 2004, 2005; The World bank, 2003, 2004, 2005). Data coverage may not be complete for countries experiencing problems (such as stemming from internal or external conflicts) with the collection of data.

.

Table 3: Number of attributes collected and examples of attributes per societal aspect (between parentheses the units used).

Societal aspect	Number of attributes	Examples of attributes
Economy	48	Taxes on income, profits, capital gains (% of revenue), listed domestic companies, agricultural productivity, value added per worker ($), household final consumption expenditure ($ millions), gross national income (GNI, $ billions), gross domestic product (GDP, $ millions), armed forces personnel (% of labour force), military expenditure (% of GDP), net national savings (% of GNI)
Education	12	Education expenditure (% of GNI), net enrolment ratio (primary % of relevant age group), primary completion rate (% of relevant age group), average years of schooling, education index
Technology	30	Internet hosts per 10,000 inhabitants, Internet users per 10,000 inhabitants, Internet secure servers, personal computers per 1000 people, mobile phone subscribers, annual growth rate (%), air, aircraft departures (thousands), high technology exports ($ millions)
Environment	29	Energy use per capita (kg of oil equivalent), carbon dioxide emissions per capita (metric tons), arable land use (hectares per capita), surface area (thousand km_), irrigated land (% of cropland)
Culture	4	Individualism, uncertainty avoidance, power distance
Demography	26	Population density (people per km_), labour force gender parity index, average annual population growth rate (%, 2001-2015), total fertility rate (births per woman), life expectancy (years)
Institution	4	Institutional investor credit rating, type of government
Health care	11	Health expenditure per capita ($), prevalence of HIV (% of adults), hospital beds per 1000 people
Jurisdiction	11	Type of legal system

No data describing the geo-information (GI) community were collected. Judging by the available literature, it is expected that the GI-community in particular could have a high impact on clearinghouses (Steudler, 2003; Ravi, 2003; Reece, 2004; Delgado et al, 2005). It could also be argued that the behaviour of the GI-community and the success of national clearinghouses could depend on common causal factors. Unfortunately, collecting data referring specifically to this community on a worldwide scale appeared to be impossible.

23 countries were excluded from the analyses due to too many missing data values (Afghanistan, Andorra, Bahamas, Barbados, Brunei, Democratic Republic of Korea, Dominica, Guyana, Holy See, Iceland, Iraq, Liberia, Libya, Liechtenstein, Luxemburg, Monaco, Myanmar, Nauru, Qatar, San Marino, Somalia, and Timor-Leste, Tuvalu)

3.3 Exploring societal factors for success

Through the complementary use of correlation coefficients with PLS regression models, an exploration of how society could affect the success of national clearinghouses was undertaken.

Correlation coefficients between the CSI and 164 attributes were calculated. The attributes of societal aspect jurisdiction (11) were excluded from these analyses, because all legal attributes collected were nominal in nature.

Table 4 presents only those attributes with coefficient values above 0.40. A classical interpretation of coefficient values is that when these values are ranging between 0.40 and 0.60, they may be regarded as an indication of a moderate degree of correlation, and when higher than 0.60, as an indication for a marked degree of correlation (Franzblau, 1958).

Table 4: Correlation coefficients between CSI and societal attributes with values higher than 0.400.

Name attribute	Societal aspect	Correlation coefficient
Taxes on income, profits, capital gains (% of revenue)	Economy	0.665
Energy use per capita (kg of oil equival.)	Environment	0.632
Internet, hosts per 10,000 inhabitants	Technology	0.622
Carbon dioxide emissions per capita metric tons	Environment	0.589
Arable land (hectares per capita)	Environment	0.546
Personal computers per 1000 people	Technology	0.542
Education expenditure (% of GNI)	Education	0.540
Listed domestic companies	Economy	0.526
Surface area (thousand km_)	Environment	0.525
Agricultural productivity, value added per worker ($)	Environment	0.516
Individualism	Culture	0.512
Household final consumption expenditure ($ millions)	Economy	0.490
Gross national income ($ billions)	Economy	0.489
Gross domestic product ($ millions)	Economy	0.488
Internet, users per 10,000 inhabitants	Technology	0.478
Internet, secure servers	Technology	0.476
Air, Aircraft departures (thousands)	Technology	0.469
Internet, hosts total	Technology	0.453
Gross national income per capita ($)	Economy	0.443
Personal computers (thousands)	Technology	0.438
Internet, users thousands	Technology	0.430
High technology exports ($ millions)	Technology	0.429
Uncertainty avoidance	Culture	-0.428
Power distance	Culture	-0.427
Purchasing power parity, GNI ($ billions)	Economy	0.426
Purchasing power parity, per capita ($)	Economy	0.425
Armed forces personnel (% of labour force)	Economy	-0.417
Roads, total road network (km)	Technology	0.415
Health expenditure per capita ($)	Health care	0.414
Foreign direct investment ($ millions)	Economy	0.410
Average years of schooling, total	Education	0.405
Education index	Education	0.405
Institutional investor crediting rating	Institution	0.400

The results presented suggest that several economic, environmental, educational and technological attributes could be important factors affecting success of national clearinghouses success. On the other hand, demographic, institutional and health care attributes are likely to be less important. The attributes with the highest correlation coefficients are related to taxes, energy use, Internet and agriculture. Taxes refer to the main source of revenue for many governments with the potential of influencing incentives and thus economy's competitiveness (The World Bank, 2003). The use of energy refers indirectly to people's wealth even though energy consumption has environmental consequences (e.g. carbon dioxide emissions). The Internet attributes refer to the information and communication technologies that offer vast opportunities for economic growth and better service delivery. The agricultural attributes refer to the agricultural productivity, and indirectly to the standard of living of farmers who produce the needed food and other agricultural products of a country. The relatively high number of cultural attributes indicates that culture could also have an impact on the success of a national clearinghouse, in particular individualism.

The main result of the ANOVA-applications in order to test differences in CSI between several legal systems is that the mean of countries with a common law (52) is significantly higher than the mean of countries with a civil law (35).

PLS-regression models were applied to predict the CSI from the societal attributes of each societal aspect. Since they were nominal in nature, attributes of societal aspect jurisdiction were excluded from these PLS-analyses. The percentages of variance captured by the best eight PLS-regression models (R_) appear to be low for each societal aspect (Table 5). This indicates that none of the societal aspects alone could explain the success of a clearinghouse. This is in line with literature on spatial data infrastructures (Groot and McLaughlin, 2000; Williamson et al, 2003; Spatial Application Division, Catholic University of Leuven, 2003; Craglia, 2003; Masser, 2005; Delgado et al, 2005; Bouckaert et al, 2006). Societal aspects with relatively high percentages are: economy, education and technology. On the other hand, societal aspects with relatively low values are: demography, institution and health care.

Table 5: Number of attributes used, percentages of variance captured by the PLS- models (R_), optimal number of latent PLS-factors, and attributes with the highest PLS-regression coefficients of each societal aspect.

Societal aspect	Number of attri-butes	R_	Number of PLS-factors	Attributes with highest PLS-regression coefficients
Economy	48	0.43	2	- Taxes on income, profits, capital gains (% of revenue) - Armed forces personnel (% of labour force) - Listed domestic companies
Education	12	0.42	3	- Net enrolment ratio (primary % of relevant age group) - Primary completion rate (%)
Technology	30	0.39	2	- Internet, secure servers - Mobile phone subscribers, annual growth rate (%)
Environment	29	0.31	1	- Arable land, hectares per capita
Culture	4	0.30	1	- Individualism - Uncertainty avoidance
Demography	26	0.26	2	- Population density (people/km_) - Labour force gender parity index
Institution	4	0.14	1	- Institutional investor crediting rating
Health care	11	0.11	1	- Health expenditure per capita ($)

PLS-regression was also applied to predict the CSI from societal conditions of each country considering all 164 attributes, covering the eight societal aspects. The percentage variance captured by the PLS-model (R_) with the optimal number of 3 latent PLS-factors (PLS3-model) is 0.77, which is considerably higher than those from PLS-models of each societal aspect. This suggests that a combination of several attributes from different societal aspects form a better explanation for the success. The attributes with the highest PLS-regression coefficients are mainly economic, educational and technological (Table 6). In addition, this R_-value (0.77) also suggests that the CSI cannot be predicted from societal attributes alone (assuming that the quality of determining CSI is high). The value of RMSEC is 10.5, and that of RMSECV 15.0 (section 2.3). These values confirm that this index cannot be fully predicted from societal attributes alone.

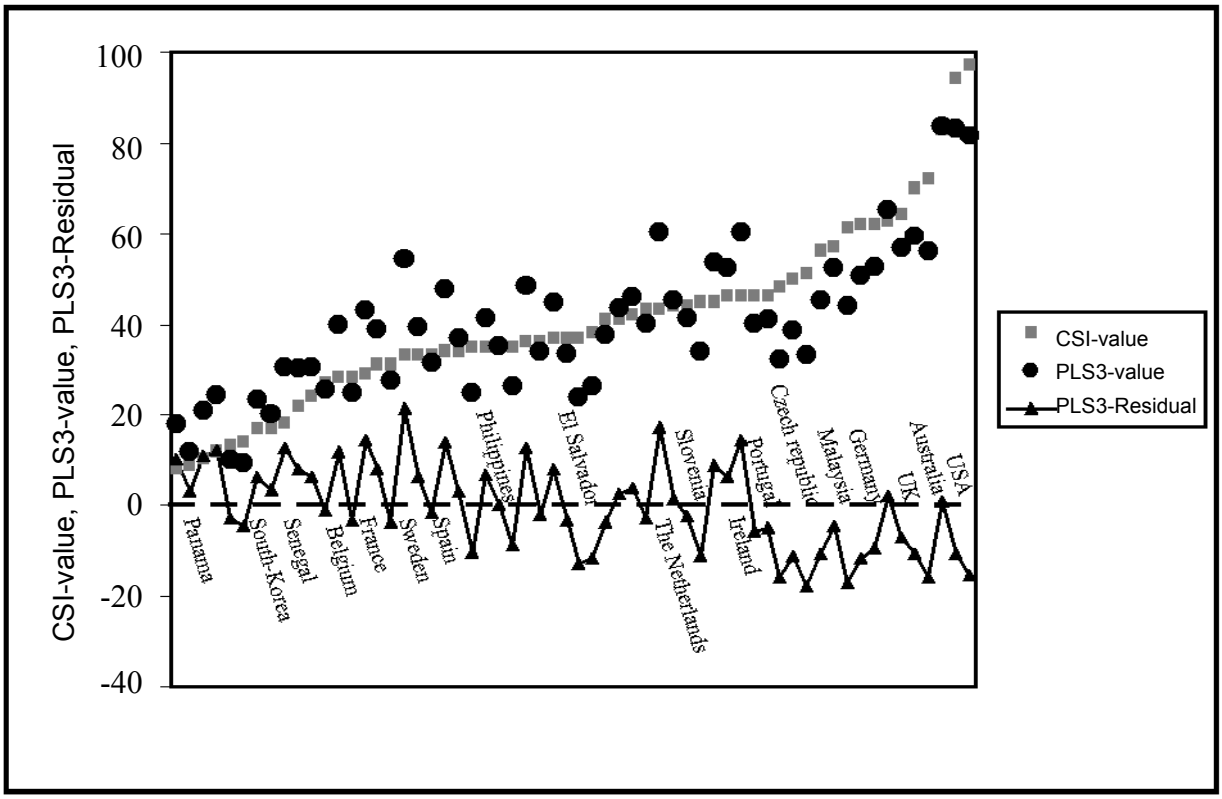
Table 6: Attributes with the highest PLS-regression coefficients (in order of importance) using the PLS3-model of society considering 164 societal attributes.

Name variable	Societal aspect
Taxes on income, profits, capital gains (% of revenue)	Economy
Net enrolment ratio (primary % of relevant age group)	Education
Armed forces personnel (% of labour force)	Economy
Gross capital formation, average annual growth (%)	Economy
Household final consumption expenditure ($ millions)	Economy
Internet, secure servers	Technology
Education expenditure (% of GNI)	Education
Military expenditure	Economy
Individualism	Culture
Labour force gender parity	Demography
Listed domestic companies	Economy
Gross domestic product, average annual growth (%)	Economy
Arable land (hectares per capita)	Environment
Irrigated land (% of cropland)	Environment
Population density (people/km_)	Demography
Surface area (thousand km_)	Environment
Average annual change in Consumer price index (%)	Economy
Domestic credit to private sector (% of GDP)	Economy
Net national savings (% of GNI)	Economy
Uncertainty avoidance	Culture

Figure 1 presents the CSI-values, the predicted values of PLS3-model considering 164 societal attributes, and the PLS3-residual values of each national clearinghouse. From this figure, it appears that the PLS3-model overestimates slightly the values of national clearinghouses with a low CSI and underestimates the values of clearinghouses with a high CSI. This supports the suggestion that the CSI cannot be predicted from societal attributes alone.

A reason that societal conditions might not have a full impact on clearinghouse success could be due to the fact that critical clearinghouse-internal factors also need consideration. Regarding Figure 1, the underestimation of the PLS3-model for national clearinghouses with a high CSI is likely to be caused by to the implementation of clearinghouse-internal success factors, such as: the introduction of web services, the stability of funding, the provision of good communication channels, the formulation of a clear vision of the national clearinghouse function, the creation of user-friendly interfaces with less discipline-specific terminology, as well as trust in the clearinghouse. On the other hand, the slight overestimation of the clearinghouses with a low CSI could be due to the absence or partial implementation of these success factors.

Figure 1: Clearinghouse Suitability Index values (CSI-values), PLS3-values and PLS3-residual values of each national clearinghouse. Samples ordered by increasing CSI-values.



Another PLS-regression model was applied to predict the CSI from societal conditions taking only into account the attributes of aspects economy, education and technology (90 attributes). The percentage variance captured by this PLS-model (R_) is 0.70. The RMSEC and RMSECV values of this PLS-model are slightly worse than the one considering all 164 societal attributes, respectively 11.5 and 16.0. These results suggest that a combination of economic, educational and technological attributes only, predicts the CSI reasonably well.

Finally, a PLS-regression model was applied to predict the CSI using the latent PLS-factors of the eight societal aspects. In total, the thirteen latent PLS-factors presented in Table 5 were used. The percentage variance captured by this PLS-model (R_) is 0.67. The RMSEC and RMSECV values of this PLS-model are similar to the PLS-model considering all 164 societal attributes, respectively 11.0 and 15.0. The latent PLS-factors with the highest PLS-regression coefficient values are (in order of importance): economic, technological and educational. This result seems to confirm the findings that the success of national clearinghouses is likely to be influenced by a combination of several factors from different societal aspects, in particular economy, technology and education. On the other hand, the latent PLS-factors with very low values are: institutional and health care.

Examples of economic, educational and technological attributes that could be critical for clearinghouse success are: taxes on income, profits and capital gains (% of revenue), listed domestic companies, gross domestic product ($ billions), education expenditure (% of GNI) and Internet secure servers (Table 4 and Table 6). In addition, arable land per capita (environment) as well as individualism (culture) could also be critical.

The suggestion of these explorative analyses is that wealthier countries, where communities are comprised of calculating citizens focussed on self-interest, are more likely to have a successful clearinghouse. These countries have a sound investment climate with good macro-economic management, trade and investment policies that promote openness, and high quality (technological) infrastructure and services. In addition, they have good education that enables people to take advantage of new opportunities. Moreover, they have supportive legal and regulatory systems (e.g. common law) that support the day-to-day operations of governments and firms by protecting property rights, promoting access to credit, and ensuring tax, customs, and judicial services. Finally, these countries still have the opportunity to increase their agricultural productivity by expanding arable land or by using their land more intensively. Investing in successful national clearinghouses could improve the spatial data/service accessibility, use, and dissemination, in order to raise the level of wealth or respond to the demand of citizens. But whether the successful implementation of a national clearinghouse indeed contributes to increasing wealth – in other words, the causal direction of the association – is not ascertained by the present exploration

3.4 Predicting clearinghouse success in countries with no clearinghouse established

Through the application of the PLS3-model, which takes all 164 societal attributes into consideration, the establishment of successful national clearinghouse could be predicted in countries where it has not previously been established. As a result of the web survey conducted in April 2005, it appeared that 83 countries had a national clearinghouse. This means that the following 16 countries established a national clearinghouse between December 2002 and April 2005: Afghanistan, Belarus, Botswana, Burkina Faso, Cambodia, Cuba, India, Israel, Madagascar, Malawi, Namibia, Nepal, Oman, Thailand, Togo, and Yemen. Table 7 presents the predicted PLS3-values, CSIs, and PLS3-residuals of these newly established national clearinghouses ordered by increasing CSI-values. Due to too many missing values, it was not possible to predict the suitability of the Afghanistan clearinghouse. When comparing the PLS3-values with the calculated CSI-values then a strong relationship between these 2 variables can be observed with a correlation coefficient of 0.72. This could indicate that the PLS3-model predicts reasonable well the success of national clearinghouses in countries where it has not previously been established. It also confirms that the societal conditions might have an impact on clearinghouse success.

A reason for the high CSI-values of Afghanistan and Madagascar clearinghouses could be that these initiatives were partly funded by foreign agencies such as the US Agency for International Development.

The presented PLS3-residual values confirm also that the PLS3-model overestimates the values of national clearinghouses with a low CSI and underestimates the values of clearinghouses with a high CSI. Another indication that the CSI cannot be predicted from societal attributes alone.

In addition, the PLS3-model was also applied to predict clearinghouse success in 95 countries that did not establish a national clearinghouse before April 2005. The mean and median of PLS3-values were very low for this group, respectively 24 and 10. These values could indicate that new national clearinghouses established in the near future are unlikely to succeed and may eventually fail, unless the clearinghouse-internal factors for success are implemented. Poland, Saudi Arabia, and Paraguay were the countries with the highest PLS3-values, respectively 42.9, 40.7 and 40.2. On the other hand, more than half of the countries had a PLS-value lower than 10.

Table 7: PLS3-values, CSI-values, and PLS3 Residual values of countries that established a national clearinghouse between December 2002 and April 2005.

Country	PLS3-value	CSI	PLS3-Residual
Cambodia	29.4	8	21.4
Burkina Faso	23.6	10	13.6
Togo	29.9	12	17.9
Belarus	19.5	17	2.5
Oman	21.5	17	4.5
Yemen	22.1	18	4.1
Israel	37.1	18	19.1
Nepal	23.8	19	4.8
Botswana	40.3	26	14.3
Malawi	29.5	26	3.5
Thailand	40.7	39	1.7
Namibia	54.5	42	12.5
Afghanistan	No value	44	No value
Cuba	39.5	48	-8.5
India	49.9	48	1.9
Madagascar	34.8	53	-18.2

The strength of the applied PLS3-model appears to be in predicting potential clearinghouse success in countries that have no clearinghouse established. It also strengthens where, in addition to societal conditions, clearinghouse-internal factors need to be implemented in order to compensate for the lack of societal support.

4. CONCLUSIONS

The findings of these explorative analyses highlight the importance of societal conditions on the success of national spatial data clearinghouses. However, society alone does not fully explain the success of national clearinghouses, since clearinghouse-internal factors are also an important consideration. It is likely that the success of national clearinghouses depends on a combination of critical attributes from different societal aspects, in particular economy, education and technology. Critical factors for clearinghouse success are likely to be those that raise the level of wealth in the respective country (e.g. gross domestic product, education expenditure, Internet servers, and taxes on income, profits, and capital gains). In this exploration, it appeared to be difficult to identify the critical factors due to the complex interaction of societal (and clearinghouse-internal) factors. Subsequent research may help us understand the causal hierarchies between these societal factors. To conclude, the results of the present exploration suggest that there is no simple solution or uniform approach for setting up a successful clearinghouse. However, being aware of the societal impact on clearinghouse success and the identified societal factors might help practitioners to overcome difficulties associated with setting up efficient and effective clearinghouses, and support clearinghouse coordinators and policy makers in the development of successful strategies for establishing or maintaining national clearinghouses.

The strength of applying the PLS-model appears to be the prediction of potential clearinghouse success in countries where it has not previously been established. In the near future, newly established national clearinghouses are likely to face serious obstacles, if not failure, due to unfavourable societal conditions. This could however be reversed by additional efforts to implement the critical clearinghouse-internal factors.

This study was explorative in nature. In order to improve the societal impact assessment and identification of societal factors for clearinghouse success, it is

strongly recommended to involve alternative clearinghouse characteristics in the CSI, non-European practitioners for determining the class weights and domains, more intelligence in the analyses using data mining and/or factor analysis approaches, the nominal (legal) attributes in the overall analysis, and maximum likelihood estimation or multiple imputation as advanced approaches for imputing the missing values (Acock, 2005; Tempelman, 2007)

Moreover, it is strongly recommended to explore the societal factors periodically. Critical (societal) factors for success may change over time. Exploring them should take their dynamic character into account. It is likely that new critical (societal) factors for clearinghouse success differ than the ones determined for the situation in December 2002. This is in line with the diffusion of innovation model developed by Rogers (1995). For example, the pioneering countries of national clearinghouses appeared to be in a position to handle the high uncertainties associated with these innovations (Masser 2005). This is unlikely to have been the case in later years, since the later adopting countries are less likely to have managed such high uncertainty.

Acknowledgement

The authors would like to thank dr. Wies Akkermans, dr. Cajo ter Braak, dr. Lammert Kooistra of Wageningen UR, ir. Harm Kossen and Judith Uwimana MSc. for comments and suggestions on earlier versions of this paper. The authors also gratefully wish to acknowledge the support of the European representatives of the European GI-community in the preparation of the Clearinghouse Suitability Index. Their contributions were crucial for the results in this paper. The authors also acknowledge the financial support by the Bsik innovation programme 'Space for Geo-Information'.

REFERENCES

Acock, A,C.(2005). Working with missing values, *Journal of Marriage and Family*, 67: 1012-1028

Bernard, L., Crompvoets, J. and J. Fitzke (2005a). "Geodateninfrastrukturen – ein Überblick", in Bernard, L., J. Fitzke and R.M. Wagner (Eds). *Geodaten-infrastruktur,Grundlagen und Anwendungen*. Herbert Wichmann, Heidelberg, Germany, pp. 4-8 (in German).

Bernard, L., Kanellopoulos, I., Annoni, A. and P. Smits (2005b). The European geoportal –one step towards the establishment of a European Spatial Data Infrastructure, *Computers, Environment and Urban Systems*, 29:15-31.

Bregt, A. (2000). "The Dutch clearinghouse for geospatial information: cornerstone of the national geospatial data infrastructure", In: Groot, R. and J. McLaughlin (Eds). *Geospatial Data Infrastructure: Concepts, Cases and Good Practice*. Oxford University Press, Oxford, UK, pp. 262-267.

Bouckaert, G., Van Orshoven, J., Dumortier, J., Van Hootegem, G., Macharis, C. (2006). *SBO Project Proposal, Spatial Data Infrastructure and Public Sector Innovation*. Katholieke Universiteit Leuven and Vrije Universiteit Brussel.

Central Intelligence Agency (2003, 2004). *The World Factbook.*

Craglia, M., Annoni, A., Klopfer, M., Corbin, C., Hecht, L., Pichler, G., and P. Smits, (Eds.) (2003). *Geographic Information in the Wider Europe*. Geographic Information Network In Europe (GINIE).

Crompvoets, J. (2006). *National spatial data clearinghouses, worldwide developments and impact*. Ph. D. Thesis, Wageningen University, The Netherlands.

Crompvoets, J. and H. Kossen (2002). The impact of culture on national spatial data clearinghouses. *GISDECO-conference, Governance and the use of GIS in Developing Countries*. Enschede, The Netherlands, pp. 9.1–9.3.

Crompvoets, J. and A. Bregt (2003). World status of National Spatial Data Clearinghouses, *URISA Journal*, Special Issue on Public Participation GIS, 15, APA I: 43-50.

Crompvoets, J., Bregt, A., Rajabifard, A. and I. Williamson (2004). Assessing the worldwide developments of national spatial data clearinghouses*, International Journal of Geographical Information Science*, 18(7): 665-689.

Crompvoets, J., Bregt, A., and M. van Adrichem (2005). Disappointing NSDI National Clearinghouse Survey, International Developments, Status, Suitability and Spatial Distribution, *GIM International*, 19 (9): 35-37.

Crompvoets, J. and A. Bregt (2007). "National spatial data clearinghouses, 2000 – 2005" in: Onsrud H.J. (Ed). *Research and Theory in Advancing Spatial Data Infrastructure Concepts*. ESRI Press, Redlands, pp. 141-154.

Delgado Fernández, T., Lance, K., Buck, M. and H.J. Onsrud (2005). Assessing an SDI Readiness Index. *From Pharaohs to Geoinformatics, FIG Working Week 2005 and 8th International Conference on Global Spatial Data Infrastructure. Cairo, Egypt*.

De Man, W.H.E. (2000). Institutionalization of geographic information technologies: unifying concept? *Cartogr. Geogr. Inf. Sci*., 27(2): 139-151.

De Man, W.H.E. and W.H. van den Toorn (2002). Culture and the adoption and use of GIS within organisations, *International Journal of Applied Earth Observation and Geoinformation*, 4: 51-63.

Dorling Kindersley (2002). *Great World atlas*. Dorling Kindersley Limited.

Franzblau, A. (1958). *A Primer of Statistics for Non-Statisticians*. Harcourt, Brace & World, New York.

Geladi, P. and B.R. Kowalski (1986). Partial least-squares regression: a tutorial *Analytica Chimica Acta,* 185: 1-17.

Groot, R. and J. McLaughlin (Eds) (2000). *Geospatial Data Infrastructure: Concepts, Cases and Good Practice*. Oxford University Press, Oxford, UK.

Hamilton, M., 2003, Online Survey Response Rates, at http://www.supersurvey.com/papers/supersurvey_white_paper_response_rates.htm, [accessed 18 December 2007].

Hofstede, G. (2001). *Culture's consequences, Comparing Values, Behaviors, Institutions, and Organisations Across Nations*. 2nd edition. Sage Publications, Thousand Oaks.

International Telecommunications Union (2003, 2004). *World Telecommunications Indicators*. International Telecommunication Union, Geneva.

Kok, B. and B. Van Loenen (2005). How to assess the success of National Spatial Data Infrastructures? *Computers, Environment and Urban Systems*, 29: 699-717.

Kooistra, L. (2004). *Incorporating spatial variability in ecological risk assessment of contaminated river floodplains*. Ph.D. Thesis. Catholic University Nijmegen, The Netherlands.

Kooistra, L., Wehrens, R., Leuven, R.S.E.W. and L.M.C. Buydens (2001). Possibilities of VNIR spectroscopy for the assessment of soil contamination in river floodplains*, Analytica Chimica Acta*, 446: 97-105.

Martens, H. and T. Næs (1989). *Multivariate calibration*. John Wiley & Sons, Chichester, UK.

Massart, D.L., Vandeginste, B.G.M., Buydensm L.M.C., De Jong, S., Lewi, P.J., and J. Smeyers-Verbecke (1997). *Handbook of Chemometrics and Qualimetrics: Part A*. Elsevier Science, Amsterdam.

Masser, I. (1999). All shapes and sizes: The first generation of National Spatial Data Infrastructures, *International Journal Geographical Information Science*, 13: 67-84.

Masser, I. (2005). *GIS Worlds, Creating Spatial Data Infrastructures*. ESRI Press, Redlands, California.

Nebert, D.D. (Ed), 2004*. Developing Spatial Data Infrastructures: The SDI Cookbook*, Version 2.0.

Nedovic-Budic, Z. (1999). Evaluating the Effects of GIS Technology: Review of Methods. *Journal of Planning Literature, 13*(3): 284-295.

Rajabifard, A., Feeney, M.E. and I.P. Williamson (2002). Future directions for SDI Development. International *Journal of Applied Earth Observation and Geoinformation*, 4(1): 11-22.

Ravi (2003). *Space for Geo-information*, Bsik knowledge project proposal. Amersfoort, Ravi.

Reece, K. (2004). *Defining and evaluating the criteria for effective implementation of national geospatial data clearinghouse in Jamaica*. MSc. Tthesis, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands.

Rogers, E. M. (1995). *Diffusion of Innovations*. New York, The Free Press

Spatial Applications Division of Katholieke Universiteit Leuven (2003). *Spatial data infrastructures in Europe: state of play during 2003*, Summary report.

Steudler, D. (2003). "Developing evaluation and performance indicators for SDIs" in: IWilliamson, I., Rajabifard, A. and M.E. Feeney (Eds). *Developing Spatial Data Infrastructures: From Concept to Reality*. Taylor & Francis, London.

Tempelman, C. (2007). Imputation of restricted data; applications to business surveys. Ph.D. Thesis. Rijksuniversiteit Groningen, The Netherlands.

Tait, M.G. (2005). Implementing geoportals: applications of distributed GIS *Computers, Environment and Urban Systems*, 29: 33-47.

The World bank (2003, 2004, 2005). *World Development Indicators*. The World bank, Washington.

Tobias, R.D. (1997). *An introduction to partial least squares regression*. SAS Institute, Cary, NC:.

Tosta, N. (1997). Data relevations in Qatar: why the same standards won't work in the United States, *Geo Info System*, 7:5.

United Nations Development Programme (2003, 2004, 2005). *Human Development Report*. New York, Oxford University Press, Oxford.

Wehn de Montalvo, U. (2004). “Chapter Eight, Outreach and Capacity Building” in: Nebert D.D. (Ed), *Developing Spatial Data Infrastructures: The SDI Cookbook*. Version 2.0. pp. 96-116.

Williamson, I., Rajabifard, A. and Feeney, M.E. (Eds) (2003). *Developing Spatia Data Infrastructures: From Concept to Reality*. Taylor & Francis, London.