# Institutions Matter: The Impact of Institutional Choices Relative to Access Policy and Data Quality on the Development of Geographic Information Infrastructures

BASTIAAN VAN LOENEN AND JITSKE DE JONG

*DELFT UNIVERSITY OF TECHNOLOGY, DELFT, THE NETHERLANDS*

ABSTRACT

Access to geographic information is critical for the development of geographic information infrastructures (GIIs). Two access policy options are dominant: open access and cost recovery. Cost recovery policies are generally thought to be associated with high-quality datasets, while open-access policies are thought to be associated with poor-quality data. However, how data collection is organized is more critical for the quality and the use of the dataset than access policies are. We compared parcel and large-scale topographic datasets in five jurisdictions of comparable sizes, population densities, and socioeconomic levels and argue that institutional choices affect GII development.

## INTRODUCTION

Access policies have been deemed critical for geographic information infrastructures (Borgman 2000 p. x; Masser 1999, p. 81; Tosta 1999, p. 23). Two access policy options have been dominant: open access and cost recovery. Cost recovery policies have been thought to be associated with high-quality datasets (GITA 2005; Lopez 1998, p. 79), while open-access policies have been thought to be associated with poor-quality data (GITA 2005; Aslesen 2002).

These findings are based on examples from the United States and the European Union. According to the Geospatial Information and Technology Association (GITA) survey, the cost recovery policies of the UK Ordnance Survey are justified by the quality of its products, which "far exceed the quality, in terms of accuracy and timeliness, of most products given away in the United States" (GITA 2005; see also Lopez 1998, p. 79). In addition, the experiences of academics in the United States suggest that access policies become more restrictive as the level of information detail increases. Figure 1 shows the assumed relationship between access policy and data quality.

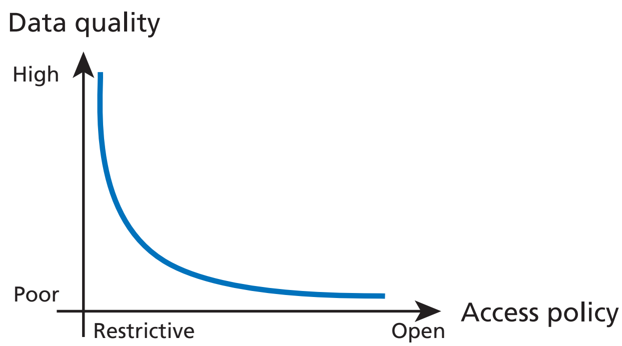

Figure 1. Assumed relationship between access policy and data quality.

We sought to test the hypothesis that the quality and the access policy of a dataset are related: datasets with restrictive access policies have excellent quality, while datasets with open access policies have poor quality. The hypothesis assumes that access policy is decisive for the quality of the dataset and that the different access policies of European and U.S. jurisdictions may account for the differences in data quality. The hypothesis conforms to the economic theory predicting that quality of information is related to the price and use restrictions.

We report the findings of a study (Van Loenen 2006) comparing parcel and large-scale topographic datasets in five jurisdictions of comparable sizes, population densities, and socioeconomic levels (the Netherlands, Denmark, the German state of North Rhine–Westphalia; and in the United States, Massachusetts and the Metropolitan region of Minneapolis and St. Paul).[1]

## OPEN ACCESS VERSUS COST RECOVERY

In the open-access model, information held by a government can be accessed by those outside the government for a price not exceeding the cost of reproduction and distribution (marginal cost of dissemination), with the imposition of as few restrictions as possible. The information is available to all (nonexclusively) on a nondiscriminatory basis (see also NRC 1997, p. 15). Acceptable restrictions include national security, trade secrets, and privacy.

Although the open-access model may initially have been enacted to control government, "[it] fosters a process for adding value to raw government information resources" (Lopez 1998, p. 58). This spin-off effect promotes the use of the information, which results in higher (income, company, or value-added) tax revenues going to the government. The increased use of geographic information resulting in innovative solutions to societal needs encourages GII development and the information economy (Pira et al. 2000; Weiss and Pluijmers 2002).

However, open policies make government entities responsible for the collection of geographic data fully dependent on general budgets. This is a precarious position, especially in an economic climate of recession. For example, the U.S. national mapping agency, U.S. Geological Survey (USGS), has suffered from significant real budget reductions, causing it to scale back updates of the 1:24,000 map series (NRC 2003, p. 22). The diminished quality of the data may result in diminished use.

Cost recovery approaches seek profits from the sale of information to support the development and maintenance of the datasets (Lopez 1998, p. 43; Onsrud 1992). Information collection, maintenance, and dissemination are not fully supported by public funds, and the costs must be covered through other means. The agency is forced to generate income from the sales of information or products or through the provision of services. The cost recovery model assumes that sufficient income will be generated for the creation and maintenance of the dataset.

Access to information may therefore be restricted to cope with the financial conditions established by government rules. In practice this entails a charge for the information at more than the marginal cost of dissemination, and restrictions are imposed on the use of government information through copyright and database rights. Further, use restrictions are often imposed through contractual or licensing provisions. Government expertise may be used to respond to private requests for specific geographic products.

In several cost recovery models, individual government agencies are in control of their budgets, making them independent of the fluctuating budgets of the national governments (Onsrud 1998, p. 146). As a consequence, government agencies are able to offer (access to) accurate, consistent, standardized databases with sufficient coverage (Aslesen 2002). Therefore, the cost recovery model may provide sustainable funding to individual government agencies, allowing them to maintain their information collection activities over time (Onsrud 1992). This is especially true in instances where "income is generated through actions required by statute" (Pira et al. 2000, p. 44; see also Coopers Lybrand 1996).

The cost recovery model may be summarized as follows: "[it] benefits end-users who are interested and able to acquire high-quality geographic information, directly from government" (Lopez 1998, p. 58). The continuous availability of high-quality datasets supports GII development.

**GII DEVELOPMENT**

A GII may be defined as a framework continuously facilitating the efficient and effective generation, dissemination, and use of needed geographic information within a community or between communities (after Kelley 1993). The framework consists of six interdependent components: (framework) datasets, institutional framework, technology, standards, financial resources, and human resources. Interaction between the components is a condition for the further development of the infrastructure.

In the early stages, GII development is data-centric: data creation and integration, reduction of duplication, effective use of resources, and creation of a base from which to expand the productivity of the geographic information sector and the geographic information market (Rajabifard et al. 2003, p. 101, 107; Rajabifard et al. 2002, p. 14). After a dataset has achieved sustainable quality in meeting the needs of primary users, value-added use of the information is the driver (Rajabifard et al. 2003; cf. Masser 2000; Van Loenen 2006).

**Users of datasets.** Users of the GII "will probably be the most mentioned group and yet actually the least considered" (McLaughlin and Nichols 1994, p. 72). This, however, does not imply that all potential user groups or applications need to be identified. It does mean that the user community has to be considered as part of the total infrastructure, and that real, rather than purely academic requirements have to be met (McLaughlin and Nichols 1994, p. 72).

We distinguish four user groups:

1. Primary users: the data collector and major users
2. Secondary users: incidental users with similar purposes as the primary user
3. Tertiary users: users that use the dataset for other than primary purposes
4. End users: users that use the end product, such as a map

Primary users are those that use the dataset in line with the initial purpose of information collection on a continuous basis. They are typically members of a government organization that has collected and processed the information. Secondary users may be found in government, the private sector, or academia. They use the information for similar purposes but only incidentally. Tertiary users are typically found in the private sector. They add value to the dataset. A tertiary use may be the integration of several topographic datasets into one layer for a jurisdiction, the linkage of a dataset with several thematic layers, provision of user-friendly access to the dataset (e.g., adding search facilities, explanation, help desk), or intermediary services for distributing the dataset. End users are citizens, decision-makers, and others that use the end product of geographic information, for example, a map, an animation, or an answer to a question.

**Technical and nontechnical characteristics of framework datasets.** Framework datasets are a key component of GIIs. The value of a framework dataset rests on its "coverage, the strengths of its representation of diversity, its truth within a constrained definition of that word, and on its availability" (Longley 2001, p. vii). A user decides to use the dataset or not to use it on the basis of its technical and nontechnical characteristics (Van Loenen 2006). Although each user category and within each category each user may have specific needs for framework data characteristics, the technical and nontechnical data characteristics described below are assessed to satisfy most users.

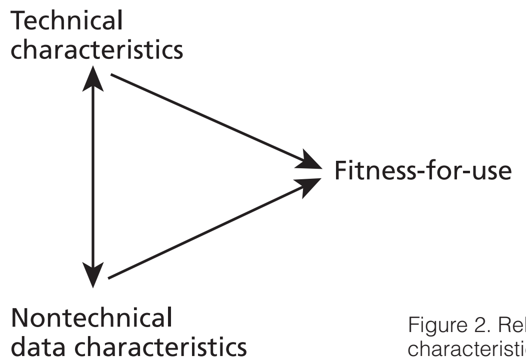


Figure 2. Relationship between data characteristics and use.

Technical data characteristics are commonly referred to as "data quality." From a technical perspective, an ideal framework dataset includes full jurisdictional coverage; a jurisdiction-wide uniform data model; adherence to open data formats; current, accurate, clearly demarcated coverage area; comprehensive metadata; and a high level of interoperability with other datasets. In the ideal dataset, each of these characteristics is sustainable over time (Van Loenen 2006).

Nontechnical data characteristics include access policy (price and use restrictions) and ease of access (e.g., access through a clearinghouse). From a nontechnical point of view, an ideal framework dataset is available, findable, easily assessable, and readily accessible (i.e., the dataset can be accessed immediately from a one-stop shop with few legal restrictions at an affordable price) (Van Loenen 2006). The access policy should be transparent and generic throughout the government.

Together, the nontechnical and technical characteristics determine the dataset's fitness for use (figure 2).

## RESEARCH METHODOLOGY

To assess the impact of access policies on the technical characteristics and use of a (framework) dataset, a multiple-case-study design was used. The case studies were selected on the basis of their similarities to the Netherlands in socioeconomic level, jurisdiction size, population density, and government type. In maximizing variance in access policies, both literal and theoretical analogies were sought. We also assumed that jurisdictions of comparable geographic sizes and population densities have similar needs for geographic datasets. Therefore, jurisdictions comparable to the Netherlands in geography were chosen: the Netherlands, Denmark, the German state of North Rhine–Westphalia, the U.S. state of Massachusetts, and the U.S. metropolitan region of Minneapolis–St. Paul (Metro region).

We focused on parcel and large-scale topographic datasets for four reasons:

1. These datasets are important for local levels of GIIs (Rajabifard et al. 2000).
2. The high level of detail at the local level can be used as the basis for higher GII levels.

3. These datasets are relatively expensive to collect, process, and maintain.
4. They have barely been addressed in access policy research.

The case study was conducted in 2004.

## FINDINGS FOR PARCEL INFORMATION

The findings for parcel information are organized in three categories: technical characteristics, nontechnical characteristics, and use.

**Technical characteristics.** The technical data characteristics were judged to be advanced for the Danish and Dutch parcel datasets. The Netherlands and Denmark have full digital parcel information coverage. In each country, one national public organization has the responsibility for parcel information. The datasets adhere to one national data standard; have a legal basis, comprehensive content, full topology, and generally no gaps or overlaps; and are fully harmonized and current for their purposes. In addition, the Danish dataset has comprehensive metadata documentation.

The Metro dataset also meets the technical data requirements of a GII. However, it's less comprehensive than the Danish and Dutch datasets. Because its parcel information comes from seven counties, the Metro dataset is not as consistent with respect to positional accuracy and content. However, Metro has full digital parcel information coverage, and all county datasets adhere to the U.S. parcel information standard, although individual counties have modified it to meet their specific needs. The Metro dataset has comprehensive metadata documentation.

North Rhine–Westphalia has the most comprehensive parcel dataset, including both parcel information and full topographic detail. Its 54 counties have created their own parcel datasets, with different technical characteristics. The state mapping agency has, together with the counties, initiated the integration of these 54 datasets into one statewide dataset. However, as of 2006, this dataset did not cover the entire jurisdiction of North Rhine–Westphalia, and some parts were still in analogue format. This inconsistency led to poor scores on technical data characteristics. With full digital coverage, the North Rhine–Westphalian dataset could be comparable to those of Denmark and the Netherlands. As with the Metro dataset, differences between some of the North Rhine–Westphalian component datasets may lead to gaps or overlaps in the integrated dataset.

The Netherlands, Denmark, Metro, and North Rhine–Westphalia datasets have legal foundations.

The Massachusetts' dataset has been judged inadequate in all technical categories. A significant percentage of the 351 component datasets are not in digital format and/or do not adhere to a standard data model. The datasets vary so much that harmonized, statewide parcel information coverage is not expected soon.

**Nontechnical characteristics.** The open-access policies of Massachusetts state government are conducive to GII development. However, obtaining parcel information covering the entire state is difficult because of the need to contact the 351 dataset sources.

The other four jurisdictions have restrictive access policies. The Danish and Dutch parcel datasets, for example, cost €2,400,000 and €3,500,000, respectively, and cannot be resold without prior consent of the data provider. However, the datasets can be acquired more easily than the datasets in Massachusetts. To obtain the dataset for all of Denmark or the Netherlands, only one organization needs to be contacted. The Metro integrated parcel dataset is available to MetroGIS participants; others need to contact each of the seven counties. The North Rhine–Westphalia datasets can potentially be made available from one contact point, but currently the 54 data providers need to be contacted separately. In none of the cases is access immediate.

Table 1 lists the findings for parcel information. Table 2 lists the overall scores.

| Characteristic | Score[a] | | | | |
|---|---|---|---|---|---|
| | DK | NL | NRW | MA | Metro |
| **Technical** | | | | | |
| Internal | | | | | |
| Content | + | + | ++ | - | 0 |
| Horizontal positional accuracy | + | ++ | ++ | - | - |
| Currency | ++ | ++ | ++ | - | ++ |
| Structure | 0 | + | -- | -- | + |
| Quality consistency throughout the (integrated) dataset | + | + | -- | -- | - |
| Average | + | + | 0 | - | 0 |
| External | | | | | |
| Digital coverage (vector format) | ++ | ++ | + | - | ++ |
| Number of datasets for jurisdiction coverage | ++ | ++ | - | -- | + |
| Standard adherence | + | + | - | - | + |
| Data model | + | + | - | - | + |
| Metadata documentation | ++ | - | -- | -- | ++ |
| Quality assurance | ++ | ++ | ++ | - | + |
| Average | + | + | - | - | + |
| **Nontechnical** | | | | | |
| Access policy | | | | | |
| Legal access | - | - | - | ++ | - |
| Financial access[b] | - (++) | - | - (++) | ++ | - (++) |
| Average | - (+) | - (+) | - (+) | ++ | - (+) |
| Accessibility | | | | | |
| Publication | ++ | - | - - | - - | ++ |
| Points to contact for maximum coverage of jurisdiction | ++ | ++ | - | - - | + |
| Acquisition procedure | + | + | + | + | + |
| Time between request and access | + | + | + | + | + |
| Average | + | + | -/ 0 | - | + |

a DK, Denmark; NL, Netherlands; NRW, North Rhine–Westphalia; MA, Massachusetts. ++, ideal; --, far from ideal.
b Free access for designated user groups.

Table 1. Findings for parcel information.

| Characteristic | Score[a] | | | | |
|---|---|---|---|---|---|
| | DK | NL | NRW | MA | Metro |
| Technical characteristics | + | + | 0/ - | - | 0/ + |
| Nontechnical characteristics | 0 | 0 | 0/ - | + | 0 |

a Users were mainly primary and secondary. DK, Denmark; NL, Netherlands; NRW, North Rhine–Westphalia; MA, Massachusetts.

Table 2. Overall scores for parcel information.

**Use.** Parcel information in all five cases was found to be used primarily by state and local governments and secondary users such as real estate managers, notaries public, utilities, architects, and engineering companies. We found few tertiary users creating value-added products. Some potential users indicated that use restrictions do not allow this; others indicated that the prices are too high to create value-added products. In Massachusetts, where prices and use restrictions are intended to promote reuse, the investment needed to integrate the 351 datasets into one is too high for the creation of value-added services.

## FINDINGS FOR TOPOGRAPHIC INFORMATION

As with parcel information, the findings for topographic information are organized in three categories: technical characteristics, nontechnical characteristics, and use.

No one central organization is responsible for the collection and processing of large-scale topographic information in any of the five jurisdictions. Instead, data is collected cooperatively, and costs are shared. The Netherlands has strong cooperation between the public and private sectors, North Rhine–Westphalia has strong cooperation within the public sector, Denmark has cooperation within the public sector and between the public and private sectors, while in the Metro region and in Massachusetts cooperation is limited.

**Technical characteristics.** The technical characteristics of the European topographic datasets were found to be reasonably good. The consistency of quality needs to be improved, but each dataset is potentially sufficient for use as a framework layer in a GII.

Full coverage and the integration of local topographic datasets distinguish the Dutch dataset. The metadata documentation, however, does not meet the standards of a GII. The Danish datasets in combination also cover Denmark completely. However, users interested in full coverage need to integrate the individual datasets themselves and face the likelihood of geometric and/or semantic incompatibility. Several Danish datasets have good metadata documentation, while others have none. As with parcel information, North Rhine–Westphalia has the most comprehensive dataset for topography. However, as of 2004, this dataset did not cover the entire jurisdiction of North Rhine–Westphalia, and some parts were still in analogue format. This inconsistency led to poor scores on technical data characteristics. With full digital coverage, the North Rhine–Westphalian dataset could be comparable to those of Denmark and the Netherlands.

The Massachusetts and Metro region datasets have poor technical quality (e.g., data models are not harmonized, and data is collected in an ad hoc fashion) and large-scale topographic information for parts of the jurisdiction.

**Nontechnical characteristics.** Because of the involvement of utilities, all five jurisdictions have generally restrictive-access policies. The large-scale basemap of the Netherlands costs between €1,000,000 and €2,000,000. In North Rhine–Westphalia the large-scale topographic dataset costs €3,400,000. The Danish equivalent sells for approximately €5,000,000. Redistribution is generally prohibited.

The Netherlands topographic dataset is published only on the provider's Web site and is sufficiently accessible: only one point needs to be contacted. As with parcel information, the North Rhine–Westphalia topographic datasets can potentially be made available from a central location, but currently the 54 information providers need to be contacted separately. Only a few of the 54 datasets are included in the state's clearinghouse. In Denmark, the 68 data sources have to be contacted separately, but some of them may be found through the national clearinghouse. In the Metro region, county datasets are sufficiently accessible, and comprehensive metadata is available online in the region's clearinghouse. In Massachusetts, large-scale topographic information is difficult to find and is not provided by the state's clearinghouse. Consequently, the 351 local authorities need to be contacted separately. However, in many instances local governments rely on topographic data provided by the utilities.

Because of incomplete coverage, the quality of the Massachusetts and Metro region datasets is judged to be poor.

Table 3 lists the findings for topographic information. Table 4 lists the overall scores.

| Characteristic | Score[a] | | | | |
|---|---|---|---|---|---|
| | DK | NL | NRW | MA | Metro |
| **Technical** | | | | | |
| Internal | | | | | |
| Content[b] | + | + | ++ | -- (+) | -- (+) |
| Horizontal positional accuracy | + | ++ | ++ | - - | - - |
| Currency | + | ++ | ++ | - - | - - |
| Structure[b] | - | - | - - | - - (+) | - - (0) |
| Quality consistency throughout the (integrated) dataset | - | -/ 0 | - - | - - | - - |
| Average | 0 | 0/ + | 0 | - - | - - |
| External | | | | | |
| Digital coverage (vector format) | ++ | ++ | 0 | - - | - - |
| Number of datasets for jurisdiction coverage | - | ++ | - | - - | - - |
| Standard adherence | 0 | + | - | - - | - - |
| Data model | 0 | + | - | - - | - - |
| Metadata documentation | + | - - | - - | - - | - - |
| Quality assurance | - | - | ++ | - - | - - |
| Average | 0/ + | + | - | - - | - - |
| **Nontechnical** | | | | | |
| Access policy | | | | | |
| Legal access | -/ - - | -/ - - | - | - - (++) | - |
| Financial access[c] | - | - | - (++) | - (++) | - |
| Average | - | - | - (+) | - (++) | - |
| Accessibility | | | | | |
| Publication[c] | - | - - | - - | - - | N/A (++) |
| Points to contact for maximum coverage of jurisdiction | - | ++ | - | N/A | N/A |
| Acquisition procedure | + | ++ | + | + | + |
| Time between request and access | + | + | + | + | + |
| Average | 0 | + | -/ 0 | N/A | N/A |

a DK, Denmark; NL, Netherlands; NRW, North Rhine–Westphalia; MA, Massachusetts. ++, ideal; --, far from ideal; N/A, not applicable.
b Individual datasets.
c Free access for designated user groups.

Table 3. Findings for large-scale topography.

| Characteristic | Score[a] | | | | |
|---|---|---|---|---|---|
| | DK | NL | NRW | MA | Metro |
| Technical characteristics | 0 | 0/+ | 0/ - | - - | - - |
| Nontechnical characteristics | - /0 | 0 | - /0 | - - | 0 |
| Users | Primary and secondary | Primary and secondary | Primary and secondary | Primary and secondary | Primary and secondary |
| a DK: Denmark; NL, Netherlands; NRW, North Rhine–Westphalia; MA, Massachusetts. | | | | | |

Table 4. Overall scores for large-scale topography.

**Use.** Uses in the five jurisdictions were similar. The primary users of large-scale topographic information are local governments and utilities. Secondary users (those that use the data incidentally) are engineers and planning agencies. Tertiary users are few or nonexistent. Restrictive policies and pricing, incomplete datasets, and lack of transparency may be the causes (the lack of transparency may also explain the findings of duplicate information collection in four cases). On the other hand, the products may be so large-scale and application-specific that they are of limited value to outside users.

## SUMMARY OF FINDINGS

Both technical and nontechnical characteristics may affect the use of parcel information. All five jurisdictions had significant levels of primary and secondary uses but few value-adding activities, possibly because of restrictive-access and cost-recovery policies.

We had hypothesized that the technical quality of a dataset and its access policy are related: that technically excellent datasets have restrictive-access policies and technically poor datasets have open-access policies.

The case studies yielded conflicting findings. We identified several technically advanced datasets with less advanced nontechnical characteristics: the topographic dataset in the Netherlands and the parcel datasets of Denmark, the Netherlands, and the Metro region. We also identified technically insufficient datasets with restrictive-access policies: the topographic datasets in Denmark, North Rhine–Westphalia, Massachusetts, and Metro region. Thus, cost recovery does not necessarily signify excellent quality. We did not obtain sufficient information to draw any conclusions about open-access policies.

Although the links between access policy and use and between quality and use are apparent, we did not find convincing evidence for a direct relation between the access policy and the quality of a dataset (figure 3).

## INSTITUTIONAL SETTING IS CRITICAL

The institutional setting of a jurisdiction affects the way data collection is organized (e.g., centralized versus decentralized control), the extent to which data collection and processing are incorporated in legislation, and the extent to which legislation requires use within government.

We have focused on the organization of data collection. The differences in data characteristics may be explained by the different way each jurisdiction has

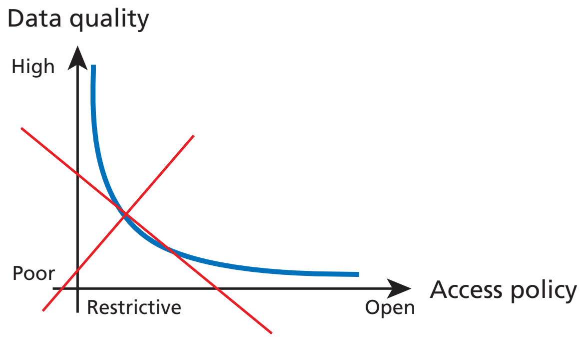


Figure 3. No convincing evidence was found supporting a causal relationship between access policy and data quality.

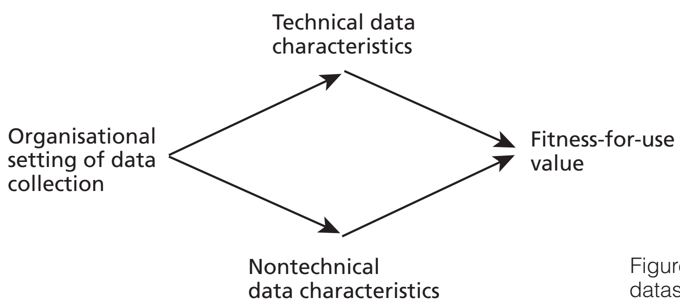


Figure 4. Factors determining dataset use.

organized its data collection. We found a direct link between the institutional setting and the characteristics of the datasets.

In jurisdictions where information collection was centralized in a single public organization, datasets (and access policies) were more homogenous than datasets that were not controlled centrally (such as those of local governments). Ensuring that data is prepared to a single consistent specification is more easily done by one organization than by many.

The parcel datasets of the Netherlands and Denmark are good examples of central control and high technical quality. They cover the entire jurisdictions in digital format and have harmonized content, accurate information, one standard data model, and no gaps or overlaps. Each has a consistent restrictive access policy. If information collection is not centralized or at least under the control of a central body (as in the case of surveys and mapping projects subcontracted to private companies by a central government agency), the likelihood of heterogeneous data characteristics increases and the fitness of the data for (tertiary) use diminishes (figure 4). For example, if information is collected by 300 local entities and each entity adheres to a different information model and policy, cross-jurisdictional users (such as tertiary users) need to find 300 datasets, contact 300 organizations, navigate 300 access policies, and ultimately integrate 300 datasets differing in content, currency, accuracy, and format. This would require significant investments, which may be difficult to recover. The 351 parcel datasets in Massachusetts, the 68 *Tekniske Korte* in Denmark, and the 54 *Automatisierten Liegschaftskarten* in North Rhine–Westphalia are examples of decentralized heterogeneous datasets.

**OVERCOMING INSTITUTIONAL BARRIERS**

The disadvantages of decentralized information collection can be redressed by institutional reform that would centralize responsibility for a specific dataset and formalize data collection arrangements (e.g., legislation), including the maintenance of the data over time.

Centralizing responsibility for a dataset has the advantage of establishing one point of contact with far-reaching authority that would enforce standards, execute agreements, and improve data quality. Access policies of central government agencies can be improved relatively easily. Moreover, a central organization would allow for investment in research and development activities specifically aimed at improving the dataset and developing the GII. In addition to contributing to a product-based GII strategy, a central organization is better able to appropriate resources for GII development. It can also coordinate activities efficiently and effectively. Increased trust may also accelerate the pace of GII development, especially with support from the highest levels of the central organization.

Based on the above factors, centrally organized datasets seem preferable to those that are not managed centrally. However, even centrally managed datasets can hinder GII development since it is not directly depending on developments in society. For example, the central agency may not readily adopt open standards or may be unwilling to provide metadata documentation. In a decentralized setting such practices may be corrected by best practices of some individual datasets. In a centrally managed situation, which in practice is a one-of-a-kind dataset, such best practices may not be stimulated. Central management may prevent individual datasets from having flexible policies and procedures and may stifle change. On the other hand, information societies increasingly communicate across borders, and organizations that operate centrally may learn from the best practices of other organizations.

Central governments may be unable or unwilling to counteract strong feelings of local independence in order to enforce institutional reform for the sake of GII development. Nevertheless, Dutch and Danish municipalities have been forced to merge (Tweede Kamer 1998–1999; Christoffersen 2005) in order to increase efficiency through the economies of scale, to organize knowledge better within the public sector, and to better address complex societal issues. GIIs will likely benefit from institutional reforms aimed at centralizing government functions.

The transformation of voluntary efforts into formal arrangements (e.g., legislation) can also be considered institutional reform. Institutional arrangements for collecting geographic data, maintaining its quality, and ensuring its accessibility are a way to promote the development of a GII. "In order to function as a foundation framework datasets should have guaranteed qualities, and central control over these qualities should exist" (Philips et al. 1999; see also NRC 1995, pp. 25, 27). To ensure data quality, standards and requirements can be legislated. For example, the Dutch government has assumed responsibility for the collection, processing, and dissemination of framework information for a wide range of framework datasets, the so-called national base registers (Stroomlijning Basisgegevens 2004). Government agencies are required to use this information and to provide feedback. The Dutch large-scale basemap (GBKN) is an example of a voluntary effort of public and private organizations, which is about to become part of the national base registries. This will result in a legislation-based dataset with guaranteed qualities, central control, and mandatory use. However, inclusion of data requirements in legislation may be against the interests of users, since they will be tied to the standard of a specific time frame and may not meet future needs. Legislation should therefore be flexible to allow for necessary changes to meet future needs.

For parcel information collection that's not centrally organized, strong cooperation between public information providers may facilitate GII development. We found that centrally organized datasets had better technical quality than decentralized ones. Information sharing and other forms of cooperation need to be initiated where centralization is lacking. A common understanding of the needs of all primary and secondary users makes successful cooperation more likely. Champions are key to the success of the effort (Craig 2001; Rietdijk 2000, p. 222).

The topographic dataset of the Netherlands and the integrated parcel dataset of Denmark offer examples of cooperative efforts in creating one jurisdiction-wide homogeneous dataset. Although they are not centrally organized, the datasets of North Rhine–Westphalia and Metro are integrated and harmonized (to some extent) by the central organizations *Landesvermessungsamt* and MetroGIS, respectively. In Massachusetts, the centrally operating organisation, MassGIS lacks the resources to accomplish comparable results. Technological developments allowing organizations to exchange information without losing their autonomy (Onsrud 1990) may aid cooperation.

## CONCLUSIONS

Technical and nontechnical data characteristics are important factors on the basis of which users decide to use or not to use a dataset. We hypothesized that the technical and nontechnical components were independent of the specific institutional setting, but we found that the way information is organized has major implications for a dataset. The institutional setting can affect access policy, accessibility, technical quality, and consequently, the type and number of users.

The direct link between the institutional setting and the technical and nontechnical characteristics of datasets suggests that choices in institutional settings many years (or even centuries) ago may affect GII development.

## ENDNOTE

1. Research reported here was conducted as part of Bastiaan van Loenen's PhD dissertation, Delft University of Technology.

## REFERENCES

Aslesen, L. 2002. The U.S. experience. Data policy and legal issues position paper for Infrastructure for Spatial Information in Europe (INSPIRE). http://inspire.jrc.it/reports/position_papers/inspire_dpli_pp_v12_2_en.pdf.

Borgman, C. L. 2000. *From Gutenberg to the global information infrastructure; Access to information in the networked world.* Cambridge, MA: The MIT Press.

Christoffersen, H. 2005. The structural reform: A reform of the organisation and management of local authorities in Denmark. http://www.akf.dk/dk2005/summary/strukturreform.htm.

Coopers Lybrand. 1996. Economic aspects of the collection, dissemination and integration of government's geospatial information. A report arising from work carried out for Ordnance Survey by Coopers and Lybrand. http://www.ordnancesurvey.co.uk/oswebsite/aboutus/reports/coopers/index.html.

Craig, W. J. 2001. Spatial data infrastructure in Minnesota: Institutional mission and individual motivation. International Symposium on Spatial Data Infrastructure. University of Melbourne. http://www.metrogis.org/documents/presentations/craig_2001.pdf.

GITA. 2005. Free or fee: The governmental data ownership debate. http://www.gita.org/resources/white%20papers/Free_or_fee.pdf.

Kelley, P. C. 1993. A national spatial information infrastructure. Proceedings of the 1993 Conference of the Australasian Urban and Regional Information Systems Association. Adelaide, Australia.

Longley, P. A., M. F. Goodchild, D. J. Maguire, and D. W. Rhind, eds. 2001. *Geographic information systems and science*. Chichester, England: John Wiley & Sons Ltd.

Lopez, X. R. 1998. *The dissemination of spatial data: A North American-European comparative study on the impact of government information policy*. London: Ablex Publishing Corporation.

Masser, I. 1999. All shapes and sizes: The first generation of national spatial data infrastructures. *International Journal of Geographical Science* 3:67–84.

Masser, I. 2000. What is a spatial data infrastructure? Proceedings of the 4th Global Spatial Data Infrastructure conference. http://gsdidocs.org/docs2000/capetown/masser/masser.ppt.

McLaughlin, J., and S. Nichols. 1994. Developing a national spatial data infrastructure. *Journal of Surveying Engineering* 120:62–76.

NRC (National Research Council). 1995. *A Data Foundation for the National Spatial Data Infrastructure*. Washington, DC: National Academy Press.

NRC (National Research Council). 1997. *Bits of power: Issues in global access to scientific data*. Washington, DC: National Academy Press.

NRC (National Research Council). 2003. Weaving a national map: Review of the U.S. *Geological Survey concept of the national map*. Washington, DC: The National Academy Press.

Onsrud, H. J. 1990. The cadastral mapping challenge for surveyors in the U.S.A. Proceedings of the FIG XIX conference. http://www.oicrf.org.

Onsrud, H. J. 1992. In support of cost recovery for publicly held geographic information. *GIS Law* 1:1–7. http://www.spatial.maine.edu/~onsrud/pubs/Cost_Recovery_for_GIS.html.

Onsrud, H. J. 1998. The tragedy of the information commons. In *Policy issues in modern cartography,* ed. D. R. F. Taylor, 141-158. Oxford: Elsevier Science. http://www .spatial.maine.edu/%7eonsrud/pubs/tragedy42.pdf.

Phillips, A., I. Williamson, and C. Ezigbalike. 1999. Spatial Data Infrastructure Concepts. *The Australian Survey*or 44 (1): 20–28. http://www.geom.unimelb.edu.au/research/publications/IPW/SDIDefi nitionsAusSurv.html.

Pira International Ltd., University of East Anglia, and KnowledgeView Ltd. 2000. Commercial exploitation of Europe's public sector information. Final report for the European Commission Directorate General for the Information Society. ftp://ftp.cordis.lu/pub/econtent/docs/commercial_final_report.pdf.

Rajabifard, A., I. P. Williamson, P. Holland, and G. Johnstone. 2000. From local to global SDI initiatives: A pyramid of building blocks. Proceedings of the 4th Global Spatial Data Infrastructure Conference. http://www.gsdi.org/docs2000/capetown/abbas.rtf.

Rajabifard, A., M.-E. F. Feeney, and I. P. Williamson. 2002. Directions for the future of SDI development. International Journal of Applied Earth Observation and Geoinformation 4:11-22. http://www.geom.unimelb.edu.au/research/publications/IPW/ITC%20Journal%202002.doc.

Rajabifard, A., M.-E. F. Feeney, and I. P. Williamson, and I. Masser. 2003. National SDI initiatives. In *Developing spatial data infrastructures: From concept to reality,* eds. Ian Williamson, A. Rajabifard, and M.-E. F. Feeney, 95–110. London: Taylor & Francis.

Rietdijk, M. 2000. Gemeentelijke vastgoedinformatievoorziening en regelgeving. Dissertation. Delft: Delft University Press.

Stroomlijning basisgegevens. 2004. De stelselnota: Op weg naar een werkendstelsel van basisregistraties. http://www.stroomlijningbasisgegevens.nl/pdfs2/Stelselnota_versie_14_4.doc.

Tosta, N. 1999. NSDI was supposed to be a verb: A personal perspective on progress in the evolution of the U.S. national spatial data infrastructure. In *Integrating information infrastructures with GI technology,* ed. B. Gittings, 13–24. Philadelphia: Taylor & Francis.

Tweede Kamer. 1998–1999. *Gemeentelijk herindelingsbeleid* (26331, nr. 1). Den Haag: Sdu Uitgeverij.

van Loenen, B. 2006. Developing geographic information infrastructures: The role of information policies. Dissertation. Delft: Delft University Press. http://www.library.tudelft.nl/dissertations/dd_list_paged/dd_metadata/index.htm?docname=088301.

Weiss, P., and Y. Pluijmers. 2002. Borders in cyberspace: Conflicting public sector information policies and their economic impacts. http://www.spatial.maine.edu/GovtRecords/cache/Final%20Papers%20and%20Presentations.