Review

High Quality Geographic Services and Bandwidth Limitations

Arzu Coltekin * and Tumasch Reichenbacher

GIScience-GIVA, Department of Geography, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland; E-Mail: tumach.reichenbacher@geo.uzh.ch

* Author to whom correspondence should be addressed; E-Mail: arzu.coltekin@geo.uzh.ch; Tel.: +41-44-6344550; Fax: +41-44-6355848.

Abstract: In this paper we provide a critical overview of the state of the art in human-centric intelligent data management approaches for geographic visualizations when we are faced with bandwidth limitations. These limitations often force us to rethink how we design displays for geographic visualizations. We need ways to reduce the amount of data to be visualized and transmitted. This is partly because modern instruments effortlessly produce large volumes of data and Web 2.0 further allows bottom-up creation of rich and diverse content. Therefore, the amount of information we have today for creating useful and usable cartographic products is higher than ever before. However, how much of it can we really use online? To answer this question, we first calculate the bandwidth needs for geographic data sets in terms of waiting times. The calculations are based on various data volumes estimated by scholars for different scenarios. Documenting the waiting times clearly demonstrates the magnitude of the problem. Following this, we summarize the current hardware and software solutions, then the current human-centric design approaches trying to address the constraints such as various screen sizes and information overload. We also discuss a limited set of social issues touching upon the digital divide and its implications. We hope that our systematic documentation and critical review will help researchers and practitioners in the field to better understand the current state of the art.

Keywords: online maps; virtual globes; high resolution data; bandwidth limitations; level of detail; cartographic design

1. Introduction and Background

1.1. Problem Overview

In recent decades, various forms of computer networks, along with other developments in technology, enabled almost any kind of information imaginable to be produced and distributed in unforeseen amounts. This *almost ubiquitous* availability of vast amounts of information at our fingertips enriched our lives, and continues to do so. However, despite very impressive developments, we have not yet perfected the art of transmitting this much information: we have to deal with bandwidth limitations. Moreover, by 2013, growth of global mobile data traffic is projected to rise by sixty-six times (projection relative to 2008, as estimated by International Telecommunication Union in April 2010), creating discussions on a possible "bandwidth crunch" [1,2]. Having limited bandwidth leads to long download times for large amounts of data and can have an impact on decision making, task fulfillment, performance and various usability aspects, such as effectiveness, efficiency, and satisfaction (e.g. [3–9]). Thus, we need to filter information. The word filtering in this context means having to selectively use information both for designing visual displays and for deciding what to transmit when. Therefore, filtering task is both a technical challenge and a cognitive one—*i.e.*, we work on algorithms that will manage the data intelligently, but at the same time we need to understand what our minds can process to customize the technology and the design accordingly. If we can *match* the cognitive limitations with bandwidth limitations; we may find a "sweet spot" where we can handle the data just right and possibly improve both human and machine performance considerably.

1.2. Geographic Data

Geographic data typically include very large chunks of graphic data, e.g., popular digital 2D cartographic products are often enhanced with aerial/satellite imagery and 3D objects, as well as annotations and query capabilities to non-graphic information that may be coupled with location input. Furthermore, the vision of *digital earth* [10] led to true 3D representations that are enriched with multi-media and multi-sensory data [11] exceeding these already large volumes of data. Longley *et al.* (2005) [12] list potential database volumes of some typical geographic applications as follows (p. 12):

- *1 megabyte*: A single data set in a small project database
- *1 gigabyte:* Entire street network of a large city or small country
- *1 terabyte:* Elevation of entire Earth surface recorded at 30 m intervals
- *1 petabyte:* Satellite image of entire Earth surface at 1 m resolution
- *1 exabyte:* A future 3-D representation of entire Earth at 10 m resolution?

For example, today a simple PNG (Portable Network Graphics) compressed shaded relief map of Switzerland is easily 20 MB [13] or Open Street Map dataset for Switzerland is 139 MB when compressed (bz2) and 1.7 GB uncompressed [14]. Considering these numbers; current geographic services on the Internet and mobile/wireless systems are impressive in speed. What such services can deliver today was hard to conceptualize a mere fifteen years ago (e.g., [15–17]). However, also today, when loading e.g., Google Earth [18], we have to wait even on "this side" of the *digital divide* (a term that refers to the gap between communities in terms of their access to computers and the Internet, high

quality digital content and their ability to use information communication technologies). There is a clear lag from the request time to the data viewing time; buildings slowly load, photo-textures come even later, terrain level of detail (LOD) switches are not seamless. Besides disturbing the user experience, such delays can also be financially costly. For example, when we are in a new location, mobile geo-location services for phones and other hand-held instruments are very convenient to have, but the data is so large that some of the geographic services quickly become prohibitively expensive [19–21].

1.3. Bandwidth Availability and Download Times

Bandwidth availability can be considered a social issue as well as a technological one. In this section we will provide a brief documentation of current state of the art from a technology perspective. We will touch upon the social and political aspects in the discussion section later. In the most basic sense, bandwidth is limited by speed of light, that is, latency cannot be reduced below t = (distance)/(speed of light). The maximum amount of (error free) data that can be transferred is often determined based on available bandwidth and signal-to-noise ratio. In information science this limit is formally defined by the *Shannon-Hartley theorem* [22,23].

For common *Internet access technologies*, the current bandwidth (net bit rate) varies from 56 kbit/s for modem and dialup connections at the lower (although not the lowest) end and 100 Gbit/s for Ethernet connections at the maximum. Between these two bit rates various technologies offer differing speeds, such as ADSL (Asymmetric Digital Subscriber Line) at 1.5 Mbit/s, T1/DS1 at 1.544 Mbit/s at the lower ends, and OC48 (Optical Carrier 48) at 2.5 Gbit/s, and OC192 at 9.6 Gbit/s.

Current data transfer rates for *mobile communication* vary from GSM (Global System for Mobile communication) at 9.6 kbit/s, GPRS (General packet radio service) up to 40 kbit/s, UMTS (Universal Mobile Telecommunications System) up to 384 kbit/s, HSDPA (High Speed Downlink Packet Access) and HSUPA (High Speed Uplink Packet Access) at 14.4 Mbit/s with High Speed Downlink Packet Access (HSDPA, 3.5G, 3G+ or UMTS-Broadband) and 600 Mbit/s for Wireless LAN (802.11n). The next generation of mobile network standards, such as the Long Term Evolution (LTE) will offer 300 Mbit/s for downlink and 75 Mbit/s for uplink. Another standard, similar to Wireless LAN is WiMAX (Worldwide Interoperability for Microwave Access) that currently offers data transfer rates up to 40 Mbit/s, but will reach with 802.16m expected rates up to 1 Gbit/s. To appreciate what these bit rates mean for geographic data volumes as suggested by Longley *et al.* (2005) [12] and reported in Section 1.2., we can calculate download times for a set of common bandwidths (Table 1).

While there are many additional factors that may affect download times (e.g., how many users are sharing the bandwidth, how far is the receiving device from the access point, channel inferences, etc.), it quickly becomes clear that for someone who would like to use high quality geographic data ubiquitously, bandwidth poses a serious problem. In other words; a reasonable assumption is this: using an average mobile device with a bit rate of 256 kbit/s, you will have to wait for more than eight hours (to be precise 08:40:50 and this number is without overhead) to download an "entire street network of a large city or small country", which "weighs" 1 gigabyte.

Table 1. Values are calculated without overhead [24]. Overhead depends on the application and can be as much as 50% (e.g., adds approximately 20 hours more waiting time for a 1 G street network on a 56 kbit/s connection or another 2h with a 622 Mbit/s connection). The data sizes are in measured in mega (MB), giga (GB), tera (TB), peta (PB) and exabytes (EB). We include a conversion of the hh:mm:ss format to a more "readable" format to help the reader to interpret the values more easily.

Bit rate → Data size	56 kbit/s	1.5 Mbit/s	622 Mbit/s	1 Gbit/s	10 Gbit/s
					Download times in hh:mm:ss
1MB Single dataset small project	00:02:22	00:00:05	00:00:00	00:00:00	00:00:00
1GB Street network large city	39:40:57	01:26:21	00:00:12	00:00:08	00:00:00
1TB Elevation entire Earth	39682:32:22 (>1653 days)	1439:15:47 (>29 days)	03:34:21	02:13:20	00:13:20
1PB Sat. image entire Earth	39682539:40:57 (> 4526 years)	1439263:05:50 (>164 years)	3572:42:16 (>148 days)	2222:13:20 (>92 days)	222:13:20 (>9 days)
1EB Future 3D Earth	39682539682:32:22 (>452696 decades)	1439263097:17:39 (> 16419 decades)	3572704:32:14 (>407 years)	2222222:13:20 (>250 years)	222222:13:20 (>25 years)

## 1.4. Network Performance and Response Time Limits ("Acceptable" Waiting Time)

Until this point we made a case that geographic data is large and despite the impressive technological developments, serving high quality data means some amount of waiting for the user. But how long do the users find it acceptable to wait, and how do we measure the quality of our service? User experience studies are helpful for the former question, and "network performance" measures are helpful for the latter. Network performance can be measured computationally, or based on users' *perceived performance* [25]. For both computational and perceptual measures, latency and throughput are important indicators, and often a *grade of service* and *quality of service (QoS)* is obtained within a project. Perceived performance is very important for the system usability; because it has implications as to how much a user is willing to wait for a download. In an earlier study Millar (1968) found that users consider a response time "instant" if the waiting was under 0.1 seconds [26]. In 2001, Zona Research reported that if the loading time exceeds 8 seconds (plus minus two), a user would seek for faster alternatives [27]. Similarly, Nielsen (1997) reported a 10-second rule [6], and in a 2010 study confirmed this with another usability study coupled with eye tracking [8]. In this study, Nielsen (2010) [8] also confirms the three important response time limits that he established earlier (Nielsen 1993) [28]: *0.1 seconds* feels "instant", under *1 second* enables "seamless work" and up to *10 seconds* most users will keep their attention on the task [28].

1.5. Information Overload

Human attention is a complex mechanism with known (but *not* too well understood) limitations, and when coupled with memory limitations, we observe a phenomenon called "information overload" in information science. The concept basically refers to the maximum amount of information we can handle, after which, our decision making performance declines (Figure 1) [17,29,30]. Human working

memory, for instance, is known to be able to handle maximum seven, in some more conservative measures, only three pieces of information for a very short period of time (*i.e.*, less than 30 s) [31–33].

Figure 1. A conceptual representation of the well-known inverted u-curve [29]. Figure is adapted (modified and re-drawn) from Eppler and Mengis (2003) [30].
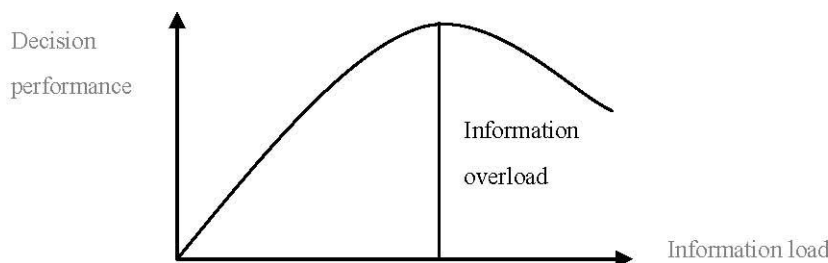
The linkage between cognitive limitations (perception, working memory) and the bandwidth usage/allocation is obvious, *i.e.*, bandwidth should not be "wasted" for information that cannot be processed by humans adequately.

1.6. Solutions?

What do researchers and practitioners do when they face low-bandwidth large-data scenarios? Despite the "bandwidth crunch" worries, it is likely that in the short term we can expect further development in the technology. That is, the bandwidth itself may get "larger" and cheaper. However, the rate of data production always competes with the availability of computational resources. Sensors get better, faster and cheaper, leading to high quality data in large quantities. Additionally, with Web 2.0 approaches more people create and publish new kinds of information as more data becomes available. There is a strong interdisciplinary effort in dealing with the problem of "large data sets" from various different aspects and there will always be an interest in intelligent data management in scientific discourse; particularly in fields such as landscape visualization and other domains where geographic data is used. As demonstrated, geographic data tend to be large, and since the amount of transferred geographic data available to transfer constantly increases, solutions need to be provided *beyond* the hope for increasing the bandwidth. In the remaining sections of this paper, we systematically review the current hardware, software, and user-centric approaches to tackle the bandwidth and resource issues when working with large data sets.

2. Technology and Design

2.1. General Purpose Graphics Processing Unit (GPGPU)

A hardware approach to processing very large graphic datasets has been the successful utilization of the graphics processing unit (GPU). A general purpose GPU is embedded in the majority of all modern computers including some mobile devices and handles all graphics processing, freeing considerable amount of computational resources in the central processing unit (CPU). The processes that are usually very "expensive" on the CPU can be transferred to the GPU, leading to much faster processing

times [34,35]. However the GPU is only relevant for local processes, that is, when the data is handled by the computer, e.g., a standalone device, a server or a client. If the term "bandwidth" is used for network transfer rates as opposed to multimedia bit rate, *i.e.*, local playback time, GPUs are only relevant for processes that are committed on the server and not for real time data streaming.

2.2. Compression, Progressive Transmission and Level of Detail Management

High quality geographic services simply would not be possible on the Internet or on mobile devices without compression. Data compression is a continuously evolving field where algorithms can be *lossy* or *lossless*; and/or *adaptive* or *composite* [36]. While in principle all compression methods seek for data similarities and remove redundancies to minimize the data size [37], different types of data (e.g., spatial data, image or videos) may require specialized approaches [38–40]. For example, among the spatial data types, vector and raster data structures are inherently different. Thus, the approaches to compress them are also significantly different. In many cases compression is enhanced and/or supported by intelligent bandwidth allocation and data streaming approaches such as the progressive transmission where data is sent in smaller chunks that are prioritized based on deliberate criteria [40].
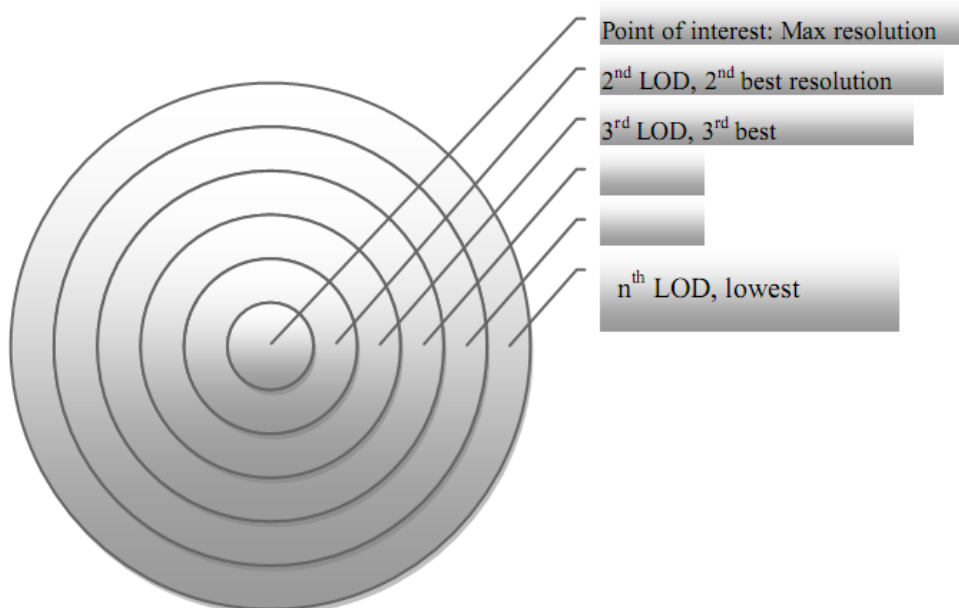
A proper review of all compression and transmission methods that are used in geographic services is out of scope for this paper because of the abundance of literature on the subject; however, we will provide a review of level of detail management, which is essentially an established computational approach where the data is manipulated in ways where technical and cognitive issues are both considered for data storage, transmission, and visualization.

Level of Detail Management

A software approach for creating "lighter" datasets to render or to stream over a network is called *level of detail management* (often referred to as LOD) in computer graphics literature. First attributed to Clark (1976) [40], LOD is a relatively old concept in computer years and is very commonly employed today. Most LOD approaches are concerned with partially simplifying the geometry of *three-dimensional* models (e.g., city or terrain models) where appropriate [41–43]. However, conceptually it is possible to draw parallels to other domains dealing with *two-dimensional* graph simplifications such as *progressive loading* (transmitting data incrementally) in streaming media, *mipmapping* in texture management or some of the cartographic *generalization* processes in map making. Figure 2 (below) shows a simplified model of a two-dimensional space partitioning for managing level of detail using a specific approach called *foveation* based on perceptual factors [44].

Foveation removes perceptually irrelevant information relying on the knowledge from human vision, that is, we see dramatically more detail in the center of our vision than in the periphery [45,46]. Many LOD approaches use a similar (though not identical) hierarchical organization with different constraints in mind. The inspiration of LOD management may come from perceptual considerations (e.g., distance to the viewer, size, eccentricity, velocity, depth of field) as well as practical constraints (e.g., priority, different culling techniques; visibility culling, occlusion culling, view frustum culling) [42,47,48].
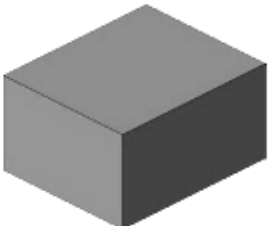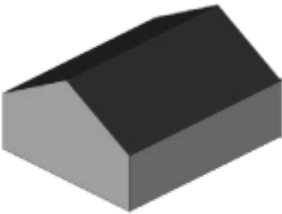
Figure 2. An equal step level of detail (LOD) model for 2D graphics.

Point of interest: Max resolution

2nd LOD, 2nd best resolution

3rd LOD, 3rd best

nth LOD, lowest

LOD and Geographic Information

The LOD concept as handled in computer graphics literature has also been used in geographic information visualization when working with terrain (e.g., mesh simplification) and city models [50,51]. In fact, for the city models, it has become a standard, namely, CityGML [49,52]. One can plan, model or deliver/order data in five levels of detail according to the CityGML standard, referred to as LOD 0, LOD 1, LOD 2, LOD 3, and LOD 4 (Table 2, Figure 3). LOD 0 typically is a regional model and expresses e.g., 2.5D digital terrain model or a "box" for a building without roof. LOD 1 is a city/site model that would have blocks of objects, i.e. buildings with roof structures. LOD 2 is also a city model, but *with textures*. LOD 3 adds further details to architecture of individual objects, and, LOD 4 is used for "walkable" architectural models including the building interiors [53].

Table 2. CityGML LOD categories. Note that LOD 4 is 3290 times larger than LOD 1. The figures are modified from [53]. LOD 0 is not represented here because it would not have a building object.

LOD 1 = 6 KB	LOD 2 = 15 KB	LOD 3 = 323 KB	LOD 4 = 19 737 KB

Figure 3. A conceptual representation of *distance LOD*: when the object is distant "enough" from the viewer, we can use a very low LOD and this should be perceptually indistinguishable for the viewer.
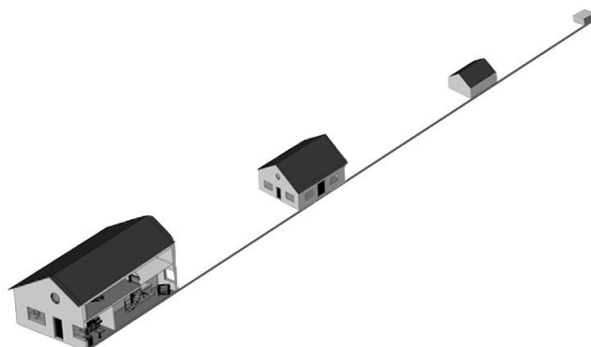
LOD management usually involves determining a parameter as to where and when to switch the LOD in the visualization process. When implemented well, LOD management essentially provides a much needed *compression* that works like an *adaptive* visualization. Ideally, a successful LOD management approach will also adapt to human perception, therefore the compression may be *perceptually lossless*. In many of today's applications, LOD switches help a great deal with the lag times and are relatively successfully utilized in progressive loading models, however the adaptation of the visualization is often not seamless.

2.3. Filtering and Relevance Approaches

Another, conceptually different but very useful approach to handle level of detail and reduce bandwidth consumption is to filter the data based on the location or the usage context, and thus its relevance to the user. Essentially we can discard the data that is not relevant for the user *perceptually* (e.g., [34,44,54]), and/or *contextually* (e.g., [55–58]). This kind of approach often needs to be "personalized"; e.g., for certain scenarios we need to know where the user is, where he or she is looking, or for what task the geographic visualization is needed. This way we can create an adaptive visualization for the person and his/her contextual and perceptual state, respecting the cognitive limitations as well as possibly avoiding "bandwidth overuse" when providing an online service.

Filtering can be applied on a database level or on a service level. A request to a geographic database can include for instance standard query language (SQL) clauses that filter the amount of data according to attribute or spatial criteria. Most mapping services follow standards introduced by the Open Geospatial Consortium (OGC) [59] within a framework for geoservices, named Open Web Services (OWS). The most relevant specifications are Web Map Service (WMS) and Web Feature Service (WFS). WMS is a portrayal service and specifically aimed at serving maps based on user requirements encoded in the parameters of a *GetMap* request. Maps from a WMS are served as raster images. WFS is a data service aimed at delivering the actual geospatial features as vector data instead of serving symbolized maps. Features are accessed by a *GetFeature* request that can incorporate filter expressions based on the *Filter Encoding Implementation Specification* [59]. Although WFS coupled with filter expressions are in general more flexible and powerful to filter geographic information, WMS also

allows to select layers that will be incorporated in the served map and therefore also offers a somewhat limited means to filter map content.

Filtering based on the context of use has been widely investigated in the last decade (e.g., [60–62]). Mainly the mobile usage profits from filtering. In the mobile case context is predominantly treated as spatial context, *i.e.*, the location of the mobile user. As a consequence so-called location-based services (LBS) have been developed. LBS filter content based on the location of a mobile user. Only information that is in a certain spatial proximity to the user are transmitted and presented to the user.

Although LBS are useful tools and are capable of filtering geographic information, they are restricted to the spatial dimension. More recent approaches look into a more general approach to the problem of information overload in mobile mapping services. Mountain and MacFarlane (2007) [63] propose additional filters for mobile geographic information beyond a binary, spatial filter used in LBS: spatial and temporal proximity, prediction of future locations based on speed and direction of movement, as well as the visibility of objects. A more recent approach that extends the idea of LBS and the filters proposed by Mountain and MacFarlane (2007) [63] is the concept of geographic relevance (GR). It has been introduced by Reichenbacher (2005, 2007) [55,64] and Raper (2007) [65]. GR is an expression of the relationship between geographic information objects and the context of usage. This context defines a set of criteria for the geographic relevance of objects, such as spatio-temporal, proximity, co-location, clusters, etc.

Filtering information based on the usage context and the resulting relevance of the information also has the advantage that the selected information is likely to fit the cognitive abilities of the users and can more easily be processed and connected to existing knowledge. Too much information, as well as irrelevant and hence useless information can actually bind cognitive resources for making sense of this information overload and limit higher-level cognitive processes, such as decision-making or planning [57]. Filtering, as discussed in the previous section, is one of the methods applied in the mapping process, specifically in the generalization process. Generalization is an abstraction process necessary to make maps of reduced scale still legible and understandable. In generalization filtering, often called *selection* or *omission*, is applied to reduce the amount of features that will be represented on the final map. The number of map features is one factor influencing the map's complexity. Other factors are the complexity of the phenomena to be represented and their relations [66,67]. Generalization also aims at reducing the complexity of the map in other ways, such as simplifying linear and areal features, aggregating areas, as well as reducing semantic complexity by aggregating categories to higher-level semantic units. In that way generalization also reduces the size of the data, e.g., less points, less attributes, less categories, less different colors, etc.

As discussed previously, a major problem of geographic services is information overload. Filtering is a powerful instrument to reduce the amount of data and potentially prevent information overload. However, for geographic information, and mainly for map representations of geographic information, the question remains how much information can be filtered and which information should not be represented. If we filter too much geographic information or if we represent too few spatial reference information the geographic context might get lost or it becomes impossible for the user to construct a consistent information structure without gaps from the representation. As in the case of information overload the result might be in the worst case a useless and unusable map representation.

Too little contextual information in a map can cause no or wrong references, errors in distance and direction estimates, invalid or wrong inferences, misinterpretations due the missing corrective function of context, problems in relating the geographic information on the map to the cognitive map, and incongruities between the perceived environment and the internal, mental representation of it.

Whereas the problem of too much filtering is evident, it is very difficult to know and even to measure a good map design and an appropriate degree of filtering. As theoretical guidelines in deciding which geographic information to represent on a map may serve research on spatial cognition, e.g., the elements that make up a city, such as paths, edges, districts, nodes, and landmarks [68] or the fundamental geographic concepts, such as identity, location, distance, direction, connectivity, borders, form, network, and hierarchies [69]. Nevertheless, there are some guidelines in designing geographic services that can handle the bandwidth limitations. Since different visual representation forms of geographic information—commonly maps, city models, digital shaded reliefs *etc.*—require different bandwidth capacities, services targeted at clients that have limited bandwidths could be restricted to light visual representation forms. Conversely, the different bandwidth capacities available for the transmission of visual representations of geographic information have a strong influence on the way such representations should be designed. Fixed line connections common for desktop applications usually offer more bandwidth than mobile network data connections. As a consequence different types of services have to be designed to meet the different capabilities.

Similarly to LOD, a widely used approach for map services is to design and hold maps of different scales. Lightweight, small-scale overview maps are presented to the user first. The user may then select an area of interest or zoom, which triggers the loading of a more detailed, large-scale map. This saves bandwidth, since only for a small area heavy data needs to be sent to the user.

As treated earlier in Section 2.2., for maps that are encoded as raster images *progressive loading* can be applied. A similar technique, *progressive rendering*, is also available for vector data. For example, the XML based Scalable Vector Graphics can render map content in the order it is coded in the document. A user will see parts of the map rendered right at the beginning of loading the document. Further map elements are successively rendered. This approach requires an intelligent map structure, *i.e.*, designing maps in a layered, prioritizing way. The map elements that are most important, have to be coded first in the document and have to be separated from less important objects that will be coded later. Sophisticated map applications make use of program logics that will load only small amounts of data for parts of the map that have changed, instead of sending a whole new map to the client. The first time the whole map is sent to a client, but for any further map updates usually only small data packages need to be transmitted and loaded. One technique supporting this kind of updating is *Asynchronous JavaScript and XML* (AJAX), which is, for instance, used in Google Maps [70].

In recent years, traditional GIS use has moved on to Spatial Data Infrastructures (SDI). SDI are a concerted effort on technological (e.g., OGC) and institutional standards on different spatial levels aiming at finding, distributing, serving, and using geographic information by diverse groups of users and for numerous applications, making geographic information more accessible [71]. Their objective is to share information rather than exchange data and therefore they are based on (web) services. The idea behind SDI is that data sources, data processing, and data provision is distributed in the Internet at different sites. Contrary to traditional GIS applications where complete, huge spatial data sets need to be stored locally first, the architecture of SDI generally requires only relatively small amount of data

(e.g., as a response of an OGC WFS request) to be transmitted from servers to clients when needed by the user and only the spatial extent and content required by the application or problem to be solved. If promises of SDI holds over time, then less bandwidth capacity will be needed, even though searching for data in catalogues or Geoportals, specifying requests and processes to be run on servers, or the conflation of different, distributed data sources may cause extra network traffic.

Another recent development that has to be considered in connection to bandwidth use is *cloud computing*. Cloud computing shares some characteristics with the client/server model, but essentially it is a marketing term that refers to the delivery of computational requirements (e.g., processing power, data storage, software and data access) as a utility over the Internet [72,73]. Cloud computing is currently very popular and it is likely that it will stay that way because of its many advantages, such as efficient distributed computing and ubiquitous services. However, it also introduces a great demand on the bandwidth by turning the traditionally local processes and services into network services.

3. Social Issues

As mentioned in the introduction section, besides technical aspects, discourse on bandwidth has a strong social dimension. Although technical tools and infrastructure for capturing and accessing geographic information have become cheaper in recent years, we still can observe an inhomogeneous and disparate availability, reliability, and capacity of network bandwidth on a global, regional, and local scale. On a global scale different measures of accessibility reveal a clear divide between the North and the South, the developed and developing counties respectively (Figure 3).

> **Figure 3.** A cartogram showing the number of Internet users in 2002 as published at worldmapper.org [74]. The size of the country is modified to match the number of Internet users. © Copyright SASI Group (University of Sheffield) and Mark Newman (University of Michigan).
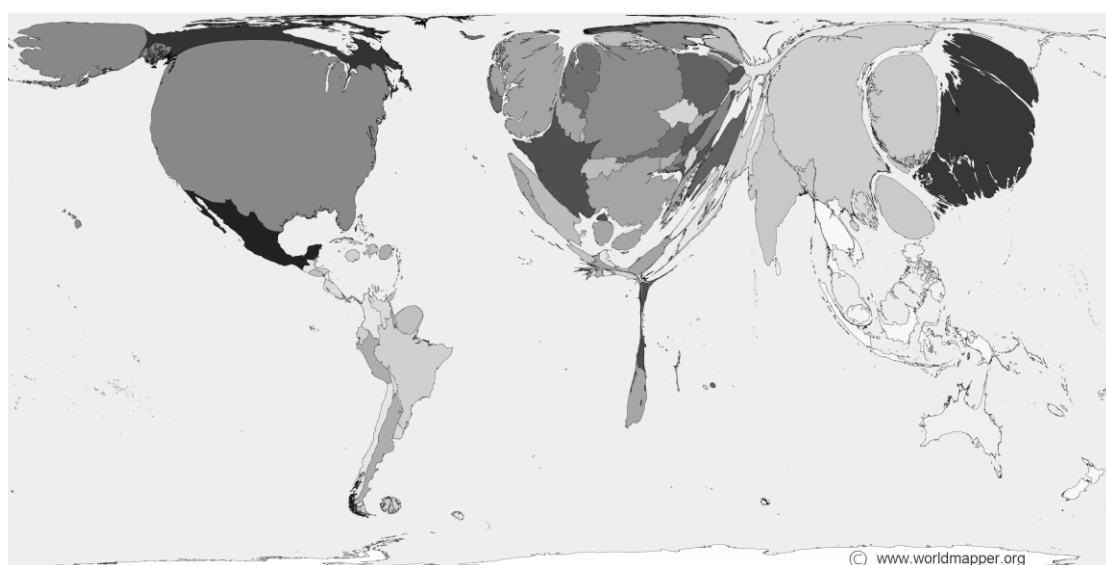
On a local scale, parts of the territory of a country may have no network access or only low bandwidth access. The reasons are mainly economical constraints faced by the network operators, such as topography, unpopulated areas, etc. An example of heterogeneous bandwidth supply takes place in

Switzerland [75], that is, mountainous areas in Switzerland are badly covered with high bandwidth offered by technologies such as HSPA.

There are many implications of the inequality of access to the Internet. In this particular example, we can see that that not all citizens have the same bandwidth at their homes, bandwidth supply varies over space, and ultimately that some types of services (rich data, e.g., maps, landscape visualizations) are not usable in all parts of the country. This example from Switzerland is interesting also from a global perspective as Switzerland is one of the highest "*bits per capita*" countries in the world. Bits per capita (BPC) is a measure that expresses the Internet use, taking the international Internet bandwidth as an indicator of Internet activity instead of number of Internet users considering many people, public organizations or commercial services share accounts [76]. A list of countries and their bit per capita measures can be found online in a variety of web pages [77].

The discrepancy between the developed and developing world (*i.e.*, digital divide) has many short, mid, and long term political and possibly ethical implications in terms of "information poverty" that is out of scope for this paper (entire books have been written on the subject, e.g., see [78]). However, it is important to be aware of the presence of the problem and further acknowledge the need for designing light-weight, but equally informative geographic services (e.g., level of detail management and filtering) to increase their accessibility.

4. Discussion and Conclusions

In this study we surveyed the current state of the art in the approaches to deal with bandwidth limitations when high quality geographic information is being provided. The topic is interdisciplinary; and this survey provides only a small window to the vast literature. However, we tried to identify critical topics and approach them systematically.

A most important concept related to bandwidth availability is the system responsiveness. The system responsiveness is directly linked to "response time" (efficiency); which is a basic metric in usability, and is measured in almost all user studies (e.g., [28,79]). In any service that is provided online, the system response time heavily depends on bandwidth, and user's response time depends on the system response time. Increasing the bandwidth and speeding the system up in other ways such as the level of detail management or filtering will positively correlate with user performance, up to some level as illustrated in the "inverted u-curve". That is, we need to keep in mind that there is a point when the system can be *too fast* for the user to process the provided information; *i.e.* when designing interaction, we have to pay attention that the changes occur in the right speed. For complex tasks, moderate waiting times can facilitate thinking [7,8,80]. Similarly, too little information just is not helpful for good decisions, but we can also provide *too much* information. In both cases, the main message is that we need to take care of the cognitive limitations of humans and test our systems properly for the target audience.

As we covered in this paper, there are many hardware, software, and design approaches to providing high quality geographic information such as the level of detail management and filtering techniques. Another aspect for the information providers as well as researchers to consider on this point is that the pre- or post-processing of information ideally should not be more "expensive" than the bandwidth related waiting times. *I.e.*, if we compress a package to stream it faster, we may be pleased

with the financial aspect as we would pay less for the data transfer, but if the mobile device that receives the package has a very small processor, we may wait just as long to decompress the data on the client side.

Mobile devices have limited processing power as well as smaller screen sizes, less storage and battery problems [20]. In some user studies, perhaps not too surprisingly, people preferred desktop networking to mobile browsers [9]. However, the advantages of mobility are self-explanatory in many scenarios. Besides, in some of the poorest parts of the world where people do not have access to electricity, they can only use mobile phones to access online services.

We demonstrated that even with moderately fast bit rates, geographic data can create very long waiting times. This gets amplified with the roaming charges in mobile networks—while maps are clearly much needed in unknown territories, the current price plans for international roaming makes the services prohibitively expensive and this problem should be addressed not only in collaboration with policy makers, but also with interdisciplinary science teams to find ways to provide the right amount of relevant data at the right time, preferably customized for the user.

Among these ways, we provided a more in depth review of level of detail management and filtering as well as relevance approaches. Both of these topics are studied by geographic, computational as well as cognitive science communities and are fairly complex, multi-faceted topics. Despite their complexity, however, both domains are already well established and carry further promise in the future to ease the bandwidth limitations (at least) to a degree.

While the bandwidth availability is directly related to user and system performance, it also has technical and socio-financial constraints. From a political perspective, we need to remember that not everyone has equal opportunities to access the rich, very high quality raw information, and try to improve our designs to be accessible and informative at the same time. This is also a usability concern, as Nielsen put it, "a snappy user experience beats a glamorous one" [81]—*i.e*, users engage more if the information overload can be avoided successfully, thus decision making can be better facilitated [82].

To conclude; we contend that being aware of technical, perceptual, and social topics related to bandwidth availability and limitations should help designers and researchers, as well as practitioners to create, design, and serve better geographic products as well as understand their use and usefulness in a human-centric manner.

Acknowledgments

References

1. Zhao, H. Globalizing trend of China's mobile internet. Available online: http://www.itu.int 2011 (accessed on 5 December 2011).
2. Gahran, A. FCC warns of looming mobile spectrum crunch. 2010. Available online: http://articles.cnn.com, *CNN Tech.* (accessed on 5 December 2011).
3. Ramsay, J.; Barbesi, A.; Preece, I. A Psychological investigation of long retrieval times on the world wide web. *Interact. Comput.* **1998**, *10*, 77–86.

4.   Lazar, J.; Norcio, A. To err or not to err. that is the question: Novice user perception of errors while surfing the web. In *Proceedings of the Information Resource Management Association 1999, International Conference*, Hershey, PA, USA, May 16–19, 1999; pp. 321–325.

5.   Jacko, J.; Sears, A.; Borella, M. The effect of network delay and media on user perceptions of web resources. *Behav. Inf. Technol.* **2000**, *19*, 427–439.

6.   Nielsen, J. The need for speed. 1997. Available online: http://www.useit.com (accessed on 5 December 2011).

7.   Selvidge, P.R.; Chaparro, B.S.; Bender, G.T. The world wide wait: Effects of delays on user performance. *Int. J. Ind. Ergon.* **2002**, *29*, 15–20.

8.   Nielsen, J. 2010 Website response times. 2010. Available online: http://www.useit.com (accessed on 5 December 2011).

9.   Öztürk, Ö.; Rızvanoğlu, K. How to improve user experience in mobile social networking: A user-centered study with turkish mobile social network site users. In *Design, User Experience, and Usability*; Marcus P.I., Ed.; Springer-Verlag: Berlin/Heidelberg, Germany, 2011; LNCS 6769, pp. 521–530.

10. Gore, A. The digital earth: Understanding our planet in the 21st century. California Science Center, Los Angeles, CA, 1998. Available online: http://www.isde5.org/al_gore_speech.htm (accessed on 5 December 2011).

11. Coltekin, A.; Clarke, K.C. A representation of everything. *Geospat. Today* **2011**, *01*, 26–28.

12. Longley, P.A.; Goodchild, M.F.; Maguire, D.J.; Rhind D.W. *Geographic Information Systems and Science*; John Wiley & Sons: New York, NY, USA, 2005.

13. Wikimedia. Reliefkarte Schweiz. 2011. Available online: http://de.wikipedia.org/wiki/Datei:Reliefkarte_Schweiz.png (accessed on 5 December 2011).

14. OSM. 2011. Available online: http://www.openstreetmap.org (accessed on 5 December 2011).

15. Imielinski, T.; Badrinath, B.R. Mobile wireless computing: Challenges in data management. *Commun. ACM* **1994**, *37*, 18–28.

16. Smith, T.R. A digital library for geographically referenced materials. *Computer* **1996**, *29*, 54–60.

17. Barrie, J.M.; Presti D.E. The WWW as an instructional tool. *Science* **1996**, *274*, 371–372.

18. Google Earth. 2011. Available online: http://earth.google.com (accessed on 5 December 2011).

19. Bertolotto, M.; Egenhofer, M.J. Progressive transmission of vector map data over the world wide web. *GeoInformatica* **2001**, *5*, 345–373.

20. Burigat, S.; Chittaro, L. Visualizing the results of interactive queries for geographic data on mobile devices. In *Proceedings of the 13th annual ACM international workshop on Geographic information systems (GIS '05)*, New York, NY, USA, November 2005; pp. 277–284.

21. Burigat, S.; Chittaro, L. Interactive visual analysis of geographic data on mobile devices based on dynamic queries. *J. Vis. Lang. Comput.* **2008**, *19*, 99–122.

22. Shannon, C.E. Communication in the Presence of Noise. *Proc. of Inst. Radio Eng.* **1949**, *37*, 10–21.

23. Hartley, V.L. Transmission of information. *Bell Syst. Tech. J.* **1928**, *7*, 535–563.

24. Kessels, J.C. Download time calculator. Available online: http://www.numion.com/calculators/time.html (accessed on 5 December 2011).

25. TM Technologies. Perceived performance—tuning a system for what really matters. Available online: http://www.tmurgent.com/WhitePapers/PerceivedPerformance.pdf (accessed on 5 December 2011).

26. Millar, R.B. Response in man-machine conversational transactions. In *Proceedings of AFIPS Fall Joint Computer Conference*, San Francisco, CA, USA, December 1968; pp. 267–277.

27. Zona Research. *The Economic Impacts of Unacceptable Web-Site Download Speeds*; Zona Research: Redwood City, CA, USA, 1999.

28. Nielsen, J. *Response times: The 3 Important Limits*; Morgan Kaufmann: San Francisco, CA, USA, 1993.

29. Schroder, H.M.; Driver, M.J.; Streufert, S. *Human Information Processing—Individuals and Groups Functioning in Complex Social Situations*; Holt, Rinehart, & Winston: New York, NY, USA, 1967.

30. Eppler, M.J.; Mengis, J. Side-effects of the e-society: The causes of information overload and possible countermeasures. In *Proceedings of IADIS International Conference e-Society*, Avila, Spain, 16–19 July 2004; pp. 1119–1124.

31. Miller, A.G.; The magical number seven, plus or minus two: Some limits on our capacity for processing information, *Psychol. Rev.* **1956**, *63*, 81–97.

32. Baddeley, A.D.; Hitch, G. Working Memory. In *The Psychology of Learning and Motivation: Advances in Research and Theory*; Bower, G.H., Ed.; Academic Press: New York, NY, USA, 1974; Volume 8, pp. 47–89.

33. Ware, C. *Information Visualization—Perception for Design*, 2nd ed.; Elsevier Inc., Morgan Kaufmann Publishers: San Francisco, CA, USA, 2004.

34. Duchowski, A.T.; Coltekin, A. Foveated gaze-contingent displays for peripheral LOD management, 3D visualization and stereo imaging. *ACM Trans. Multimedia Comput. Commun. Appl. (TOMCCAP)* **2007**, *3*, 1–18.

35. Gobron, S.; Coltekin, A.; Bonafos, H.; Thalmann, D. GPGPU computation and visualization of three-dimensional cellular automata. *Vis. Comput.* **2010**, *27*, 67–81.

36. Ali, A.; Khelil, A.; Szczytowski, P.; Neeraj, S. An adaptive and compsite spatio-temporal data compression approach for wireless sensor networks. In *Proceedings of ACM MSWiM'11,* Miami, FL, USA, 31 October–4 November 2011.

37. Kattan, A. Universal intelligent data compression systems: A review. In *Proceedings of 2010 2nd Computer Science and Electronic Engineering Conference*, Essex, UK, 8–9 September 2010; pp. 1–10.

38. Pereira, F. Video compression: Discussing the next steps. In *Proceedings of the 2009 IEEE international conference on Multimedia and Expo* (ICME'09), New York, NY, USA, June 28–July 3, 2009; IEEE Press: Piscataway, NJ, USA, 2009; pp. 1582–1583.

39. Peyre, G. A review of adaptive image representations. *IEEE J. Selected Top. Signal Process.* **2011**, *99*, 896–911.

40. Hou, P.; Hojjatoleslami, A.; Petrou, M.; Underwood, C.; Kittler, J. Improved JPEG coding for remote sensing. In *Proceedings of Fourteenth International Conference on Pattern Recognition*, Brisbane, QLD, Australia, 16–20 August 1998; Volume 2, pp. 1637–1639.

41. Clark, J.H. Hierarchical geometric models for visible surface algorithms. *Commun. ACM* **1976**, *19*, 547–554.

42. Heok, T.K.; Daman, D. A review on level of detail. In *Proceedings of the International Conference on Computer Graphics, Imaging and Visualization, (CGIV 2004)*, Penang, Malaysia, 26–29 July 2004; pp. 70–75.

43. Constantinescu. Z. Levels of detail: An overview. *J. LANA* **2000**, *5*, 39–52.

44. Coltekin, A. Foveation for 3D visualization and stereo imaging. Ph.D. Thesis, Helsinki University of Technology, Helsinki, Finland, February 2006.

45. Perry, J.; Geisler W.S. Gaze-contingent real-time simulation of arbitrary visual fields. *Proc. SPIE* **2002**, *4662*, 57–69.

46. Sanghoon, L.; Bovik, A.C.; Kim, Y.Y. High quality low delay foveated visual communication over mobile channels. *J. Vis. Commun. Image Represent.* **2005**, *16*, 180–211.

47. Reddy, M. Perceptually modulated level of detail for virtual environments. Ph.D. Thesis, University of Edinburgh, Edinburgh, UK, July 1997.

48. Luebke, D.; Reddy, M.; Cohen, J.D.; Varshney A.; Watson, B.; Huebner, R. *Level of Detail for 3D Graphics*. Elsevier Science Inc.: New York, NY, USA, 2002.

49. CityGML. 2011. Available online: http://www.citygml.org (accessed on 5 December 2011).

50. Koller, D.; Lindstrom, P.; Ribarsky, W.; Hodges, L.F.; Faust, N.; Turner, G. Virtual GIS: A real-time 3D geographic information system. In *Proceedings of the 6th Conference on Visualization '95 (VIS '95)*. Atlanta, GA, USA, 29 October–3 November 1995; pp. 94–100.

51. Dadi, U.; Cheng L.; Vatsavai, R.R. Query and visualization of extremely large network datasets over the Web using quadtree based KML regional network links. In *Proceedings of 17th International Conference on Geoinformatics*, Fairfax, VA, 12–14 August 2009; pp. 1–4.

52. Kolbe, T.H.; Gröger, G.; Plümer L. CityGML – Interoperable access to 3D city models. In *Proceedings of the Int. Symposium on Geo-information for Disaster Management*, Oosterom, P., Zlatanova, S., Fendel, E.M. Eds.; 21–23 March 2005; Springer-Verlag: Berlin/Heidelberg, Germany, 2005; pp. 883–900.

53. Benner, J.; Geiger, A.; Häfele, K.-H. Concept for building licensing based on standardized 3D geo information. In *Proceedings of the 5th International 3D GeoInfo Conference*, Berlin, Germany, 3–4 November 2010.

54. Bektas, K.; Coltekin, A. An approach to modeling spatial perception for geovisualization. In *Proceedings of STGIS 2011*, Tokyo, Japan, 14–16 September 2011.

55. Reichenbacher, T. Adaption in mobile and ubiquitous cartography. In *Multimedia Cartography*, 2nd ed.; Cartwight, W., Peterson, M., Gartner G., Eds.; Springer-Verlag: Berlin/Heidelberg, Germany, 2007; pp. 383–397.

56. Reichenbacher, T. The concept of relevance in mobile maps. In *Lecture Notes in Geoinformation and Cartography*; Gartner, G., Cartwright, W., Peterson, M., Eds.; Location Based Services and TeleCartography, Springer-Verlag: Berlin/Heidelberg, Germany, 2007; pp. 231–246.

57. Swienty, O.; Reichenbacher, T.; Reppermund S.; Zihl, J. The role of relevance and cognition in attention-guiding geovisualisation. *Cartogr. J.* **2008**, *45*, 227–238.

58. Reichenbacher, T. Die Bedeutung der Relevanz für die räumliche Wissensgewinnung. In *Geokommunikatiion im Umfeld der Geographie, Wiener Schriften zur Geographie und Kartographie*; Kriz, K., Kainz, W., Riedl, A., Eds.; Institut für Geographie und Regionalforschung, Universität Wien: Vienna, Austria, 2009; Volume 19, pp. 85–90.

59. Open Geospatial Consortium. Available online: http://ogc.org (accessed on 5 December 2011).

60. Nivala, A.; Sarjakoski, T. An approach to intelligent maps: Context awareness. In *Proceedings of the 2nd Workshop on HCI in Mobile Guides*, Udine, Italy, 8 September 2003.

61. *Map-Based Mobile Services. Theories, Methods and Implementations*, Meng, L., Zipf, A., Reichenbacher, T., Eds.; Springer-Verlag: Berlin/Heidelberg, Germany, 2005.

62. Keßler, K.; Raubal, M.; Wosniok, C. Semantic rules for context-aware geographical information retrieval. In *Proceedings of 4th European Conference on Smart Sensing and Context, EuroSSC*, Guilford, UK, 16–18 September 2009; Volume 5741, pp. 77–92.

63. Mountain, D.; MacFarlane, A. Geographic information retrieval in a mobile environment: Evaluating the needs of mobile individuals. *J. Inf. Sci.* **2007**, *33*, 515–530.

64. Reichenbacher, T. The importance of being relevant. In *Proceedings of XXII International Cartographic Conference*, Coruna, Spain, 11–16 July 2005.

65. Raper, J. Geographic Relevance. *J. Doc.* **2007**, *63*, 836–852.

66. Castner, H.W.; Eastman, R.J. Eye-movement parameters and perceived map complexity. *Am. Cartogr.* **1985**, *12*, 29–40.

67. MacEachren, A. Map complexity: Comparison and measurement. *Am. Cartogr.* **1982**, *9*, 31–46.

68. Lynch, K. *The Image of the City.* The MIT Press: Cambridge, Cambridge, MA, USA, 1960.

69. Golledge, R. *Geographical Perspectives on Spatial Cognition. Behavior and Environment: Psychological and Geographical Approaches*; Golledge, R., Gärling, T., Eds.; North-Holland: Amsterdam, The Netherlands, 1993; pp. 16–45.

70. Google maps. Available online: http://maps.google.com (accessed on 5 December 2011).

71. Craglia, M. Introduction to the international journal of spatial data infrastructures research. *Int. J. Spat. Data Infrastruct. Res.* **2006**, *1*, 1–13.

72. Mirashe, S.P.; Kalyankar, N.V. Cloud computing. *Commun. ACM* **2010**, *51*, 9–22.

73. Wu, J.; Ping, L.; Ge, X.; Wang, Y.; Fu, J. Cloud storage as the infrastructure of cloud computing. In *Proceedings of the International Conference on Intelligent Computing and Cognitive Informatics*, Kuala Lumpur, Malaysia, 22–23 June, 2010; pp.380–383.

74. Worldmapper. Available online: http://www.worldmapper.org (accessed on 5 December 2011).

75. Unique level of coverage in Switzerland. Available online: http://www.swisscom.ch (accessed on 5 December 2011).

76. International development research center. Available online: http://www.idrc.ca/cp/ev-6568-201-1-DO_TOPIC.html (accessed on 5 December 2011).

77. International Internet bandwidth > Mbps (per capita) (most recent) by country. Available online: http://www.nationmaster.com/graph/int_int_int_ban_mbp_percap-international-bandwidth-mbps-per-capita (accessed on 5 December 2011).

78. Norris, P. *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*; Cambridge University Press: Cambridge, UK, 2001.

79. Coltekin, A.; Heil, B.; Garlandini, S.; Fabrikant, S.I. Evaluating the effectiveness of interactive map interface designs: A case study integrating usability metrics with eye-movement analysis. *Cartogr. Geogr. Inf. Sci.* **2009**, *36*, 5–17.

80. Stokes, M. Time in human–computer interaction: Performance as a function of delay type, delay duration, and task difficulty. PhD Dissertation, Texas Tech University, Lubbock, TX, USA, May 1990.

81. Nielsen, J. Website response times, 2010. Available online: http://www.useit.com (accessed on 5 December 2011).

82. Chewning, E.C., Jr.; Harrell, A.M. The effect of information load on decision makers' cue utilization levels and decision quality in a financial distress decision task. *Acc. Org. Soc.* **1990**, *15*, 527–542.