

Sustainable Information Technology for Global Sustainability

Keywords: Digital Earth, World map, Web mapping, Geographic Markup Language (GML), Web Feature Service (WFS), Distributed geodata management, Digital longevity

Gunnar Misund

Associate Professor

Østfold University College

Halden

Norway

gunnar.misund@hiof.no

<http://www.ia.hiof.no/~gunnarmi/>

Jan Høiberg

Associate Professor

Østfold University College

Halden

Norway

jan.hoiberg@hiof.no

Abstract

Information technology is used in many application areas as a means to support environmental sustainability, among them the various Digital Earth projects. Many of these initiatives are quite technology intensive, requiring sophisticated special purpose software and hardware. Other projects rely heavily on complex coordination of contributions from a large number of organizations. The authors are of the opinion that technology for supporting global sustainability in it self should be sustainable, or in other words, capable of being kept in existence over a sufficiently long span of time. We present some cases of both sustainable and non-sustainable systems and propose a set of criteria for analyzing and classifying the sustainability of a given technological framework and it's components. On this backdrop we present Project OneMap, which has been designed with longevity, accessibility, simplicity and other aspects of sustainable technology in mind. The main objective of this project is to provide free online access to a large, global map. The map will be incrementally build by seamlessly integrating a large number of independent and uncoordinated contributions of geodata. A sufficient level of quality and consistency is assured by applying a peer review process to all submissions. In many ways the project will be a grass root effort, thus orthogonal to many other more centralized mapping projects. We highlight some sustainable aspects of Project OneMap.

Table of Contents

[1. Introduction](#)

[2. Case Stories](#)

[2.1 The Domesday Project](#)

[2.2 Lost and Found](#)

[2.3 Internet and the Web](#)

[2.3.1 Internet](#)

[2.3.2 The Web](#)

[2.4 The Band Web Site](#)

[2.5 The Global Mapping Project](#)

[3. Sustainable Information Systems](#)

[4. Project OneMap](#)

[4.1 Repository](#)

[4.2 Gateway](#)

[4.3 Clearinghouse](#)

[4.4 Geodata and Global Sustainability](#)

[5. OneMap Sustainability](#)

[5.1 Longevity](#)

[5.2 Demand](#)

[5.3 Simplicity](#)

[5.4 Quality](#)

[5.5 Accessibility](#)

[5.6 Responsiveness](#)

[5.7 Adaptivity](#)

[5.8 Scalability](#)

[5.9 Robustness](#)

[5.10 Stability](#)

[6. Final Remarks](#)

[Footnotes](#)

[Acknowledgements](#)

[Bibliography](#)

1. Introduction

Designing and implementing the Digital Earth vision [\[GOR\]](#), or even parts of it, is a daunting task. Several large scale projects are aiming in this direction, and interesting and impressive results have started to emerge [\[DIG1\]](#), [\[DIG2\]](#). Many of the initiatives are quite technology intensive, like NASA Goddard's Digital Earth Workbench [\[NAS\]](#), requiring sophisticated special purpose software and hardware, often expensive and definitely not in the out-of-the-shelf category. Other projects rely heavily on detailed coordination of contributions from a large number of organizations, e.g. the Global Mapping Project (see section [\[GLO\]](#)). The authors are of the opinion that complex technology and/or complex management may be an obstacle in realizing the Digital Earth objectives.

The purpose of this paper [\[1\]](#) is two-fold: 1) To present an information system, Project OneMap [\[MIS1\]](#), which supports global sustainability, and 2) To advocate the notion that technology supporting sustainability should itself be sustainable.

The terms *sustainable* and *sustainability* have been frequently (mis)used since the Rio conference in 1992, where "Sustainable Development" was firmly placed on the international agenda. One of the most widely used definitions of sustainability in the environmental sense appeared in the so-called Brundtland Report in 1987 [\[BRU\]](#):

Meeting the needs of the present generation without compromising the ability of future generations to meet their needs.

In the more general setting, the meaning of sustainability is a lot broader than in the Brundtland Report. In this paper we use the term *sustainable information technology* when referring to technology that is capable of being maintained over a long span of time independent of shifts in both hardware and software.

The rest of the paper is outlined as follows. We first give some examples of both sustainable and non-sustainable information technologies, and outline key characteristics of sustainable

information systems. We then present Project OneMap, focusing on two its components called the Gateway and the Clearinghouse and how these may support global sustainability. In the last sections we describe how the OneMap infrastructure is implemented to satisfy sustainability criteria. We close the paper with some final remarks.

2. Case Stories

In this section we look at a few cases which may shed light on the sustainability aspect of digital information systems. We encourage the readers to reflect upon both the sustainable and non-sustainable elements in the stories told.

2.1 The Domesday Project

The Domesday Book and the The Domesday Project is an interesting case for illustrating sustainability, and lack of such. The original book, which was an inventory of more than 1300 settlements in England, was completed in 1086. It was initiated by William the Conqueror, and the purpose was probably of fiscal nature [\[DOM1\]](#). The Domesday Project was finished 900 years later and consists of interactive video, text, images and specialized computer programs. The purpose was to document a snapshot in time of English everyday life, and the inspiration was Williams book [\[DOM2\]](#).

The paradox is that the original book from 1086 is still readable, while the material collected in 1986 is only accessible through a specialized and now outdated technology. Analog videodiscs did not turn out to be a long-lived technology. This paradox raises a few interesting questions. One of which is what could have been done in 1986 to prepare for a flexible and changing technology? The answer to this is not obvious and it illustrates clearly the dilemma raised by the urge for accessibility. The original book(s) in handwritten Latin is not easily accessible now, and was probably not available to many readers at the time of writing. The accessibility has however not changed for nearly 1000 years.

We do clearly have problems to deal with on different levels. Basically it is clear that preserving video is another story than preserving text. We could however expect that the transformation of analog video footage to digital form would be a minor problem. It is also an interesting issue how we design solutions, typically how we design the relationship between logic and data.

The change of purpose or ambition from 1086 to 1986 is of course the major explanation of the paradox. Interactive video leads to other kinds of sophistication in the solution. The consequence was involving what must have been considered at the time modern technology. Interactive video was more or less tailored for the project, or probably vice versa.

2.2 Lost and Found

In one of my colleague's garage there are some boxes of punched cards which is a Simula67 program from 1970. It was written to simulate switching strategies in telephone networks. The boxes have been moved around for changing reasons. In the beginning because it had some general value as a simulation program. Later on it has been saved from the paper recycling process for nostalgic reasons.

Two things have changed that makes this program less accessible than it once was. Firstly it is difficult, if not impossible, to find a computer with a card reader. Secondly it is equally difficult to find a computer with a suitable Simula compiler. However, it is possible to reconstruct the program due to the fact that the top of the cards shows the characters punched. So if we assume that no one has dropped the cards and silently put them back in

wrong order and that the print on top of the cards are not too blurred, we can reconstruct the program fairly easy, without interpreting the holes in the cards. Once reconstructed as text, it will be straight forward to change it into Java or C++, if necessary. So the printed text survives.

2.3 Internet and the Web

The World Wide Web is perhaps the single most important application of information technology, and has completely penetrated the daily life of a huge percentage of the worlds population. The Web is a combination of an infrastructure for carrying digital information, the Internet, and a set of protocols and standards facilitating a higher level of interaction between user and machines connected to the Internet.

2.3.1 Internet

Out of the fear of the thermonuclear holocaust, interest and need for robust control and communication systems arouse. On this backdrop, Paul Baran at the Rand Corporation defined on paper what we now know as the Internet. In a series of eleven papers (1960 - 1964 !) he introduces and investigates the fundamental Internet concepts, such as distributed networks, packet switching and routing [BAR]. His main goal was to design a communication infrastructure that would survive even if a large number of the participation nodes where taken out. In addition, the system should be fully scalable, supporting an "infinitely" large number of nodes and links.

The first realization of the ideas from the Rand papers appeared in 1969 as the US ARPAnet. In 1983 the TCP/IP protocol where introduced as the standard transfer mechanism. Simultaneously, the US Defense Communications Agency split ARPAnet into a strictly military network, MILNET, and the civilian counterpart, officially termed Internet. Since then, no conceptually significant development have occurred to Internet other than an incredible growth, which it has indeed gracefully survived.

2.3.2 The Web

Tim Berners-Lee "invented" the Web in 1990 as a tool for managing a chaotic set of High Energy Physics related documents residing on a jungle of different machines on different nets [BER]. The main idea was the *hyperlink*, i.e. the possible to link parts of documents to arbitrary resources living anywhere on the net. He built on the existing Internet and carefully applied the legacy from a few visionaries such as Vanevar Bush and Doug Englebart. He defined a set of very simple protocols and formats, most important the URI (Universal Resource Identifier) that makes it possible to assign a unique identifier on any resource on net, the HTTP (Hyper Text Transfer Protocol), allowing efficient and flexible transport of different media types over the net, and finally HTML (Hyper Text Markup Language) which made it possible (and easy) to produce rich documents with hyperlinks. Both HTTP and HTML were merely extensions/simplifications of existing technology. When software to serve and browse HTML became widely accessible, the World Wide Web rapidly grew into what it is today, in a span of a decade.

Berners-Lee early recognized the the anarchistic nature of the Web called for some kind of government, and he founded the World Wide Web Consortium in 1994. This coordinating body has since then been able to keep the Web together.

2.4 The Band Web Site

A good example of a sustainable system is The Band web site that has been on the net since 1994 [\[BAN\]](#). This is a musical site with a wealth of information about the seminal rock group The Band, who had their heyday in the late '60s and early '70s, on their own and with Bob Dylan. It's a non-commercial project initiated by a fan (Jan Høiberg) and hosted by Østfold University College, Halden, Norway.

From its humble start the site quickly grew to become the definitive information resource on its selected topic. It currently consists of approximately 10.000 files (HTML, photos, audio and video) located on a server handling 170.000 file requests per day (February 2003), a significant portion of net traffic for any server in Norway. The average daily amount of data transferred is currently 4.5 Gbytes. Close to 100.000 distinct host machines from 150 different top-level internet domains access the site every month.

It is quite remarkable for a web site dedicated to a rather narrow topic to survive for almost 10 years (the average life-span of your non-commercial web site is assumed to be significantly shorter) and continue to experience growth in traffic, interest and contributions. According to the site's webmaster, the keystones that make this possible are:

Dedicated users and contributors

An incredible "root system" consisting of hundreds and hundreds of regular users of The Band site has developed over the years. Most of the material on the site, ranging from single photos and lyrics transcriptions to rare concert footage and long, detailed analysis of albums and films, have been submitted by these volunteers on the net. None of them have ever been paid a single kopek for their work.

Stable project leadership

Someone has to take on coordination responsibilities if these kinds of projects are to survive.

Proper recognition

The Band web site is listed as the most important resource on its topic in most internet catalog services. It has been given awards and positive reviews, has been the subject of a CNN news spot and is acknowledged by the artists themselves and their management.

Content and quality

Far too many sites (in the musical segment of the web, in particular) present a flashy facade with practically nothing behind it. Well-organized quality information is crucial in the long run, if you want to survive on the net.

The KISS Principle and platform independence

(KISS: Keep It Simple, Stupid) People, including product and service users, generally want things that are simple, meaning easy to learn and use. The Band site acknowledges this by avoiding meaningless animations and other "gadgets" requiring users to have special software or a particular operating system. Everything is written using standard HTML and a set of scripts to automate maintenance, no traces of (quoting the author of The Cathedral and the Bazaar [\[RAY\]](#), Eric Raymond) "creeping featurism."

Staying non-commercial

Of course commercial web sites could and should be sustainable. But a non-commercial site like theband.hiof.no probably would have disappeared eventually if the webmaster had accepted "offers" from the music industry to take over control of the content (Capitol Records, a major US label, wanted to move the site to their own servers and use it to promote a CD from former The Band member Robbie Robertson in 1998).

2.5 The Global Mapping Project

The Global Mapping Project [\[GLO\]](#) is an international effort started in 2000 to provide access to harmonized geodata on a global level. All national mapping agencies are encouraged to contribute data from their respective nations. Four types of raster data are offered (elevation, land cover, land use and vegetation) along with four vector data featurtypes (drainage systems, transportation, boundaries, populated places). Currently, completed data sets are available from 16 countries, see [Figure 1](#).

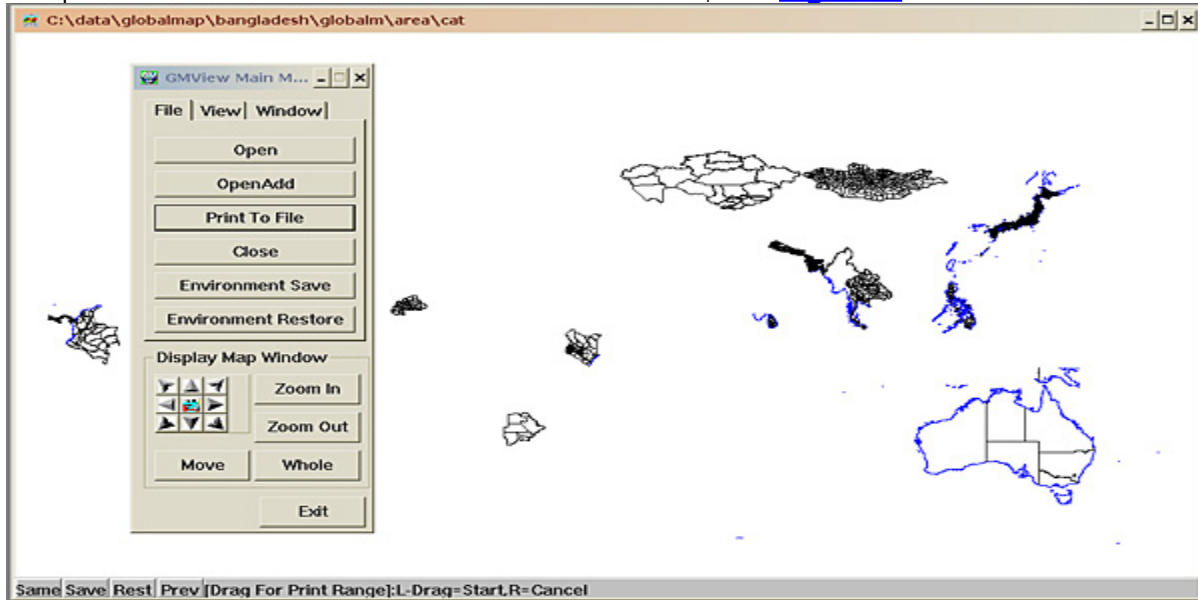


Figure 1: Progress of The Global Mapping Project

The Global Map data are available as a set of files for each nation on the Vector Product Format, one of many geodata standards defined and used by US (military) organizations, in this case The National Imagery and Mapping Agency (NIMA). The data is *not* considered to be in the public domain, but may be used for non-commercial and research purposes. It is possible to browse and view the data with a simple application available from the project homepages. However, further use of the data requires use of geographic systems capable of digesting VPF data.

3. Sustainable Information Systems

Based on the lessons learned in the previous section, we suggest the following characteristics of a sustainable information system. These criteria should indeed be applicable to a *digital* system, and in general for any infrastructure based on information technology. In the next section we will see how the OneMap infrastructure is designed and implemented to meet most of these criteria.

Longevity

A prime characteristic of a sustainable system is that it is to be able to survive over time. Digital longevity is an area of growing concern, see e.g. [\[BES\]](#) and [\[ROS\]](#). The case stories of the Domesday project and the lost and found punching cards indicate the graveness of this issue.

Demand

If there is no demand for the services offered from the system, the system will not fill any role in any context, and thus there would likely be little or no support of any kind. High demand is also a prerequisite for building a stable and significant user base, which definitely would help to sustain a system over time.

Simplicity

This is a common denominator for long-time survivors in the world of information technology systems. A sustainable system should be well planned with easy-to-use-and-understand design and clear, no-frills functionality.

Quality

Garbage in, garbage out...a system filled with low grade content will most likely end in the garbage bin, sooner or later.

Accessibility

One of the most prominent features of The Web is the easy and ubiquitous access. Services channeled through standard browsers reach the most distant corners of our globe.

Responsiveness

People needs feedback and the feeling of being heard and participating. Most systems will be strengthened by mechanisms allowing users to contribute to the development. The open-source "bazaar"-model used by the Linux project is one way of doing this [\[RAY\]](#).

Adaptivity

A system should have an underlying structure that allows for easy adaption of technological innovations and improvements. The information technology society experiences major paradigm shifts spaced only by a handful of years.

Scalability

Accelerated interest in and rapidly increasing number of request against a web server may bring a system to halt, and in several cases start a domino effect which eventually will crash the system fundamentally (the slashdot effect). A system has to gracefully handle an increasing work load, and also be prepared for a rapid growth of the content.

Robustness

Many of our most sustainable creatures of the earth may loose a limb or two, or even more significant body parts, and still function satisfactory. Robust systems are fault tolerant in the sense that sub systems may fail without bringing the system down. This might be implemented by exploiting redundancy or by providing backup systems/ methods. Robustness also implies that it should be easy to replace modules of the system without redesigning the whole system.

Stability

A sustainable system should be run by an organization (like the WWW Consortium or the Global Mapping Project), or one or more key persons (like Linus Torvalds or Richard Stallman and his GNU cohorts) responsible for "keeping the flame burning" and "passing the torch" if needed, with regards to maintenance and development. Commercial companies may administrate sustainable systems, but would most likely drop a project if becomes financially uninteresting.

4. Project OneMap

The work presented in this paper is part of an ongoing long term effort named Project OneMap [\[MIS1\]](#). The main objective is to incrementally build a large, distributed global online map. The intention is to provide free access to a repository of geographic information that may be used as base data for a wide variety of applications, such as GIS (Geographic Information Systems) analysis, environmental planning, tourist guides and transportation management. The main interface to the system, the Gateway, is shown in [Figure 2](#).

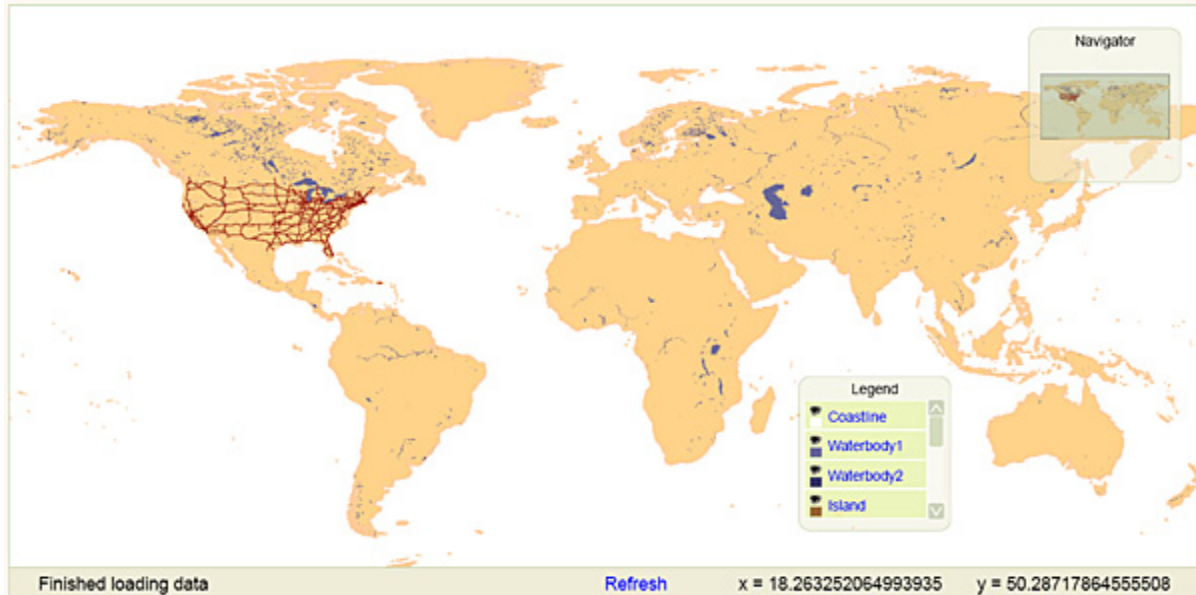


Figure 2: OneMap Gateway

Project OneMap is an open project in every respect, as it is based on Open Source [\[OSRI\]](#) and Open Content [\[OCN\]](#) principles, and also adopts a distributed management model. All major software components are completely platform independent. The project is currently hosted by and coordinated from Østfold University College, School of Computer Science. The OneMap infrastructure is functionally comprised of three main components: *Repository*, *Gateway* and *Clearinghouse*. The alpha version has been up and running since 2002, and the beta version was released in August 2003 [\[ONE\]](#). The overall architecture is depicted in [Figure 3](#).

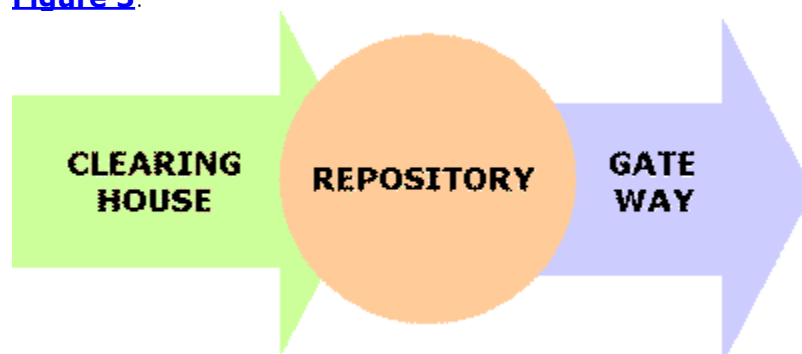


Figure 3: OneMap Architecture

4.1 Repository

The Repository is the heart of OneMap. This is the core legacy part of the infrastructure that should gracefully survive the rapid shifts in the information technology society. It is a distributed storage infrastructure, basically a (huge) set of plain text files structured to efficiently support retrieval and updating of the geodata comprising the world map. Each file is more or less self-explanatory, using the XML based metastandard GML (Geographic Markup Language) [\[GML\]](#). GML will become an international ISO standard later this year or early 2004 [\[ISO\]](#). The storage structure is illustrated in [Figure 4](#). Each feature type, e.g. roads, buildings, rivers, is represented in a logically consistent *layer* with global extent. Each layer is then preprocessed into a set of layers with increasing level of detail, corresponding to a range of scales. Each scale layer is then tiled by adaptive quadtree subdivision, such that the size of each tile is below a given threshold. A single tile is termed a *sheet*.

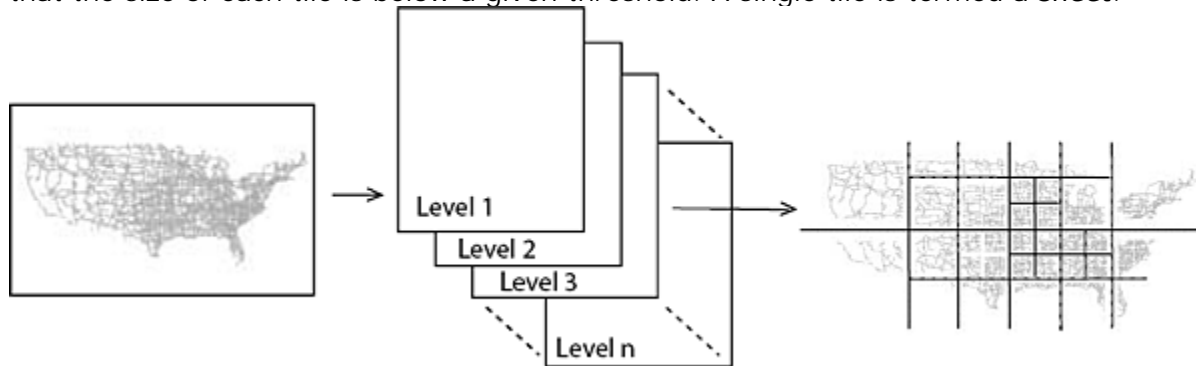


Figure 4: Repository storage scheme

The total amount of base data is currently in the magnitude of 100 M points. The data is structured in levels of detail to facilitate global overviews as well as street level inspections. The files are distributed redundantly on a set of servers. This facilitates both storage scalability and parallel processing of retrieval queries and updating procedures. Currently we use a straight forward directory based management of the distributed data. However, work is in progress to implement a peer-2-peer infrastructure to avoid bottlenecks and single-points-of-failure. The communication within the OneMap infrastructure is based on Web Services paradigms and technologies. The main interface to the repository is defined as an OGC (Open GIS Consortium) WFS (Web Feature Service) [\[WFS\]](#). The services are based on Java Servlet technology.

4.2 Gateway

The Gateway is a browser based user interface for retrieval of OneMap data. We are currently using an SVG/Javascript implementation, which provides simple but sufficient navigation and query possibilities. In addition to the SVG view, the users may download the GML formatted version of the response. The Gateway is based on the same principles as the GML Editor presented in [\[MIS3\]](#). The zooming functionality is illustrated in [Figure 5](#).

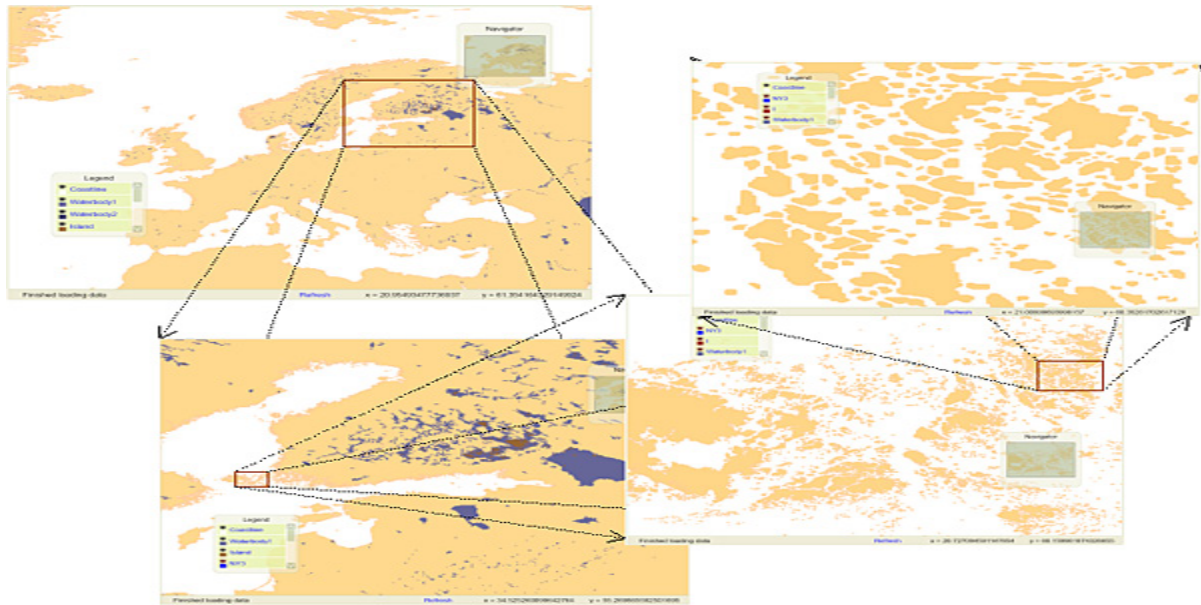


Figure 5: Multiple levels of detail

The following featuretypes are currently available through the Gateway:

Coastline

The initial population of OneMap was the (free) full resolution version of the DCW (Digital Chart of the World) from

[\[GSH\]](#), for short the GSHHS data. The original data set consists of around 10 million points and 200.000 polygon features.

Waterbody 1

Main rivers and lakes from the GSHHS source.

Island

Main GSHHS islands.

Waterbody 2

Rivers and lakes on islands from GSHHS.

A1

This is a complete set of US roads classified as "Primary Highway With Limited Access", converted from Tiger Line Files from US Census Bureau [\[UCS\]](#).

A5

US roads classified as "Vehicular Trail", also from the same source as the A1 roads.

Landuse

This data set covers most of Norway, and is land cover/land use data originally compiled for 1:250.000 scale uses, kindly provided by the Norwegian Institute of Land Inventory (NIJOS) [\[NIJ\]](#).

4.3 Clearinghouse

One of the main objectives of OneMap is to make it possible to build a huge map in an incremental and uncoordinated manner with contributions from a wide variety of parties while still guaranteeing a reasonable level of reliability and quality. The major obstacle is integration of submissions with existing geodata. We address the problem by using a framework based on the well known principles of *peer review*. The process of contributing new geodata or updating existing data is illustrated in [Figure 6](#).

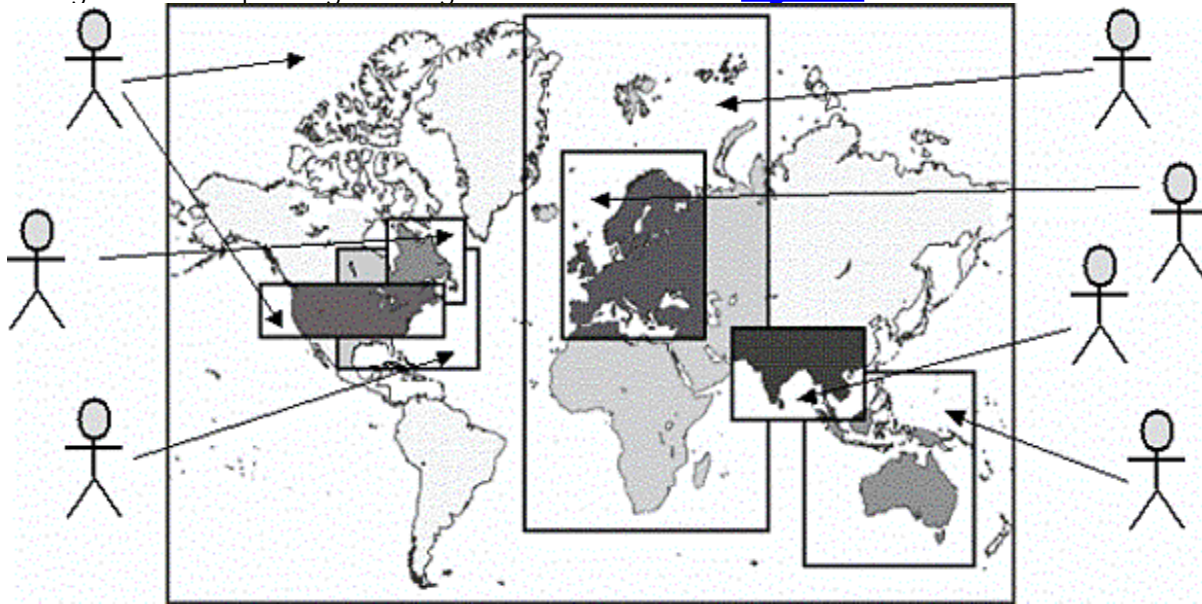


Figure 6: Incremental Map Construction

The peer review process is outlined in [Figure 7](#) and described in more detail in [\[MIS3\]](#).

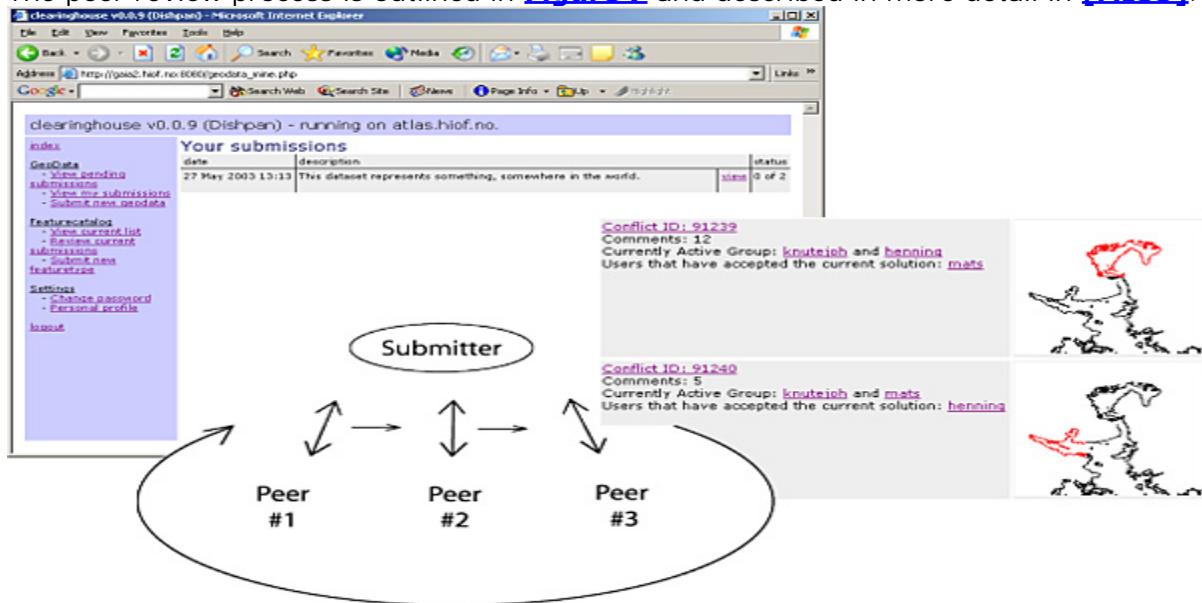


Figure 7: Clearinghouse submission process

4.4 Geodata and Global Sustainability

In the process of trying to achieve global sustainability, the need for information on the state of our environment plays a key role. Virtually all environmental data is somehow linked to a location, and will in most cases benefit from being presented in a map context. The environmentally oriented UN activities, in particular United Nations Environmental Program (UNEP) and the Rio conference in 1992 on sustainable development, have spawned several important geodata projects. Eight chapters of Agenda 21 deal with the importance of geodata. In particular, chapter 40 aims at decreasing the gap in availability, quality and standardization of geodata between nations [\[UN1\]](#). The need for global mapping with public access is further emphasized in [\[UN2\]](#).

However, the supply of general map data on a global scale is sparse. The main bulk of geodata is indeed not in the public domain, and is in many cases too expensive for environmental use. In addition, a virtual tower of Babel of different formats and standards, many of them proprietary, provides a major obstacle to widespread use of geodata. Some formats does in fact require special purpose software in the 10K USD range and upwards (and operation of this software often requires equally expensive training), effectively barring usage in smaller companies and public departments on tight budgets.

The OneMap Project is meeting this challenge with an orthogonal approach. First of all, we insist on keeping an open and free access to all our data. Secondly, we do not rely on coordinating all the mapping authorities around the globe, but rather encourage parties to submit the data that already is available and which they think might be of general interest. The third principle is to serve the data on an open format, currently being GML, the upcoming standard for representation of geodata [\[GML\]](#).

5. OneMap Sustainability

We now outline how different parts of the OneMap infrastructure meet the sustainability criteria outlined in the previous section.

5.1 Longevity

The OneMap Repository is basically a huge set of sheets, i.e. XML files following the GML specification (the soon-to-come international geodata standard). If deployed with care, this yields a set of documents that are fairly self-explanatory. XML is a way of structuring information in a hierarchical (recursive) manner by marking parts of the content with tags indicating the semantics of the content. The following example is a snippet from the XML document that this paper is generated from:

```
...
<title>Sustainable Information Technology for Global Sustainability</title>
<keyword>Digital Earth</keyword>
<author>
  <fname>Gunnar</fname>
  <surname>Misund</surname>
  <jobtitle>Associate Professor</jobtitle>
  <address>
```

```

        <affil>#216;stfold University College</affil>
        <city>Halden</city>
        <country>Norway</country>
        <email>gunnar.misund@hiof.no</email>
        <web>http://www.ia.hiof.no/~gunnarmi/</web>
    </address>
</author>
...

```

GML files requires a little more knowledge of the geodata domain, but the following excerpt from a Repository file should, hopefully, illustrate the readability of the sheet files:

```

...
<one:FeatureInfo>
    <one:Metadata>
        <dc:source>GSHHS - Global Self-consistent, Hierarchical,
        High-resolution Shoreline Database
        http://www.ngdc.noaa.gov/mgg/shorelines/gshhs.html</dc:source>
        <dc:publisher>Gunnar Misund gunnar.misund@hiof.no</dc:publisher>
        <dc:rights>notify</dc:rights>
        <dc:date>200303</dc:date>
    </one:Metadata>
</one:FeatureInfo>
...
<one:FeatID>Coastline_12</one:FeatID>
<one:OmLinearRing>
    <one:GeomID>Coastline_12#0.1.1.1.2.1.2.2.1</one:GeomID>
    <gml:coordinates>
        7762,0
        7762,34
        7769,41
        7769,55
        ...
        7762,0
    </gml:coordinates>
</one:OmLinearRing>
...

```

Simple data structures and simple algorithms have good survivability odds. The structure of the Repository file storage system is straight forward and intuitive, and thus fairly easy to realize, at least as a naive variant. The algorithms used for preprocessing, search and retrieval are correspondingly simplistic and easily implementable.

The main programmatic interface to the Repository is defined according to the Web Feature Service Specification [\[WFS\]](#) from the OpenGIS® Consortium [\[OGC\]](#). This is a standard that enables platform independent exchange of geodata. Following relevant standards is an important way to increase longevity.

The principle of uncoordinated, incremental map construction ensures that the content base grows in an "organic" way. We think the system have a better chance to survive when not imposing external constraints.

The distributed coordination model should hopefully prevent a "single-person-of-failure" situation. However, there is a serious challenge in how to maintain a stable project leadership under this model.

5.2 Demand

Geodata with global extent should be accessible in a set of scale levels. The Global Mapping Project offers their data in one single resolution, and this is barring usage of the data in many applications. The OneMap Gateway provides access from the bird's view to street level detail (providing that the original data were sufficiently detailed), thus facilitating "infinite" zooming.

All feature types in OneMap are considered to have global coverage. As an example, the coastline is one single "entity", which is topologically coherent, with no gaps or slivers between national borders. Consistent data on a global level is a prerequisite when performing regional or other cross-boundary analysis.

The Gateway will provide the user with the option of several output formats (currently SVG (Scalable Vector Graphics, the WWW Consortiums recommendation for efficient exchange of 2D graphics on the Web [\[SVG\]](#) and GML). Different delivery formats will inherently attract more users.

5.3 Simplicity

The KISS principle (Keep It Simple and Stupid) is governing much of the work in OneMap. The Repository uses a very intuitive file system structure, and algorithms for preprocessing, search and retrieval are easy to understand and implement. We also design the Gateway and Clearinghouse services with user friendliness in mind.

5.4 Quality

All the geodata in the Repository is available in a long range of preprocessed scale levels. The coarser levels are constructed from the most detailed version available. This procedure, most often called *map generalization*, ensures that the data is available at any scale with "optimal" quality.

The integrated modeling of features, i.e. that all data is logically consistent on a global level, obviously yields a higher quality than serving data that may overlap or be disconnected. Submission of new data, or revision of existing data is governed by peer review processes, which is a well known and well used technique for quality assurance.

The distributed coordination model leaves the management transparent and open to critique and corrections, and may thus enhance the overall quality of the project.

OneMap strictly follows the open content principles, most prominently by providing free and indiscriminating access to any data in the Repository for any user. We hope that more usage will produce more feedback, thus facilitating a continuous enhancement process.

It is a well known fact that widely used open source software often is of high quality, due to the fact that more eyes gives less bugs. All software produced in OneMap will be released under free licenses, and all core software used in the project should also be open source, if possible.

5.5 Accessibility

The core system of simple XML sheet files makes it possible to access and retrieve the content with a wide range of software tools, either directly or through the Web Feature Service (WFS) interface.

We plan to store the data on a distributed set of independent servers. In doing so, we are able to perform search and retrieval by spawning several parallel processes, and merge the outcome of these in the final response. This may reduce response times significantly. The Gateway is implemented as a thin client running in any standard Web browser. This makes it very easy for even an inexperienced user to browse and download geodata. The submission and peer review processes are also supported by correspondingly simple browser based client software.

5.6 Responsiveness

OneMap is in its nature extremely aware of its users, and is in fact based on an ongoing dialogue between the users and the system and its coordinators. The principle of incremental map construction supported by online submission, review and conflict resolution will make the content grow according to the supply and demand situation in the user group.

5.7 Adaptivity

The core file store is fairly easy to convert to other formats, thus making transitions to other base formats easier. The simple structure also facilitates migration to other algorithms and software used for search and retrieval in the Gateway.

5.8 Scalability

We plan to implement the Repository as a distributed and redundant structure, with a number of replicas of each file. Many copies of each information entity makes it possible to implement parallel processing of search and retrieval routines. By increasing the number of nodes it is possible to store more data and process "infinitely" huge datasets, and serve a high hit rate.

All our preprocessing, search and retrieval methods are fairly simplistic in their nature, on the verge of being naive. The algorithms are working algorithms with "optimal" complexity and makes it possible handle huge datasets efficiently.

Distributed coordination makes it possible to extend the management network, so that more work may be undertaken and thus spreading an increased load on more hands.

5.9 Robustness

The principle of redundant storage reduces the risk of permanent loss of data according to the LOCKSS (Lots of Copies Keep Stuff Safe) principle. The search and retrieval processes are distributed on many servers, and the system will continue to function even if a significant number of the nodes goes down. Work is currently in progress to exploit contemporary peer-2-peer techniques in order to remove any single-point-of-failure situation.

The algorithmic simplicity allows for easy replacement of buggy components, thus minimizing the risk for serious system malfunction.

By not having a traditional, centralized management model, we will hopefully be able to smoothly change and replace function in the coordination network, and avoid becoming totally dependent of key personnel.

5.10 Stability

Project OneMap is currently coordinated by a single dedicated person, but we are investigating models for distributed management, e.g. the one used in the WWW consortium.

6. Final Remarks

Project OneMap is still in its infancy. Considerable efforts have been made to design and implement with sustainability criteria in mind, but potential evidence of success will only start to emerge after some years. However, the content base will soon reach a level, regarding quality and quantity, where it should be of interest for many users. Hence, it will no longer be a toy/academic case, but a real-life sized information system, and may form a good foundation for further investigation and research into sustainable systems.

One of the core sustainability features is the decision to use XML as the ubiquitous "language" in the Repository, or more precisely, the not-yet-official geodata standard GML. The information technology community have adopted and adapted to XML with an astonishing force, but it is still a rather fresh technology, with many open questions attached. Still, if XML should leave the scene in some near future, it would be a fairly straight forward (however overwhelming) task of converting the core content to the new favorite of the day.

Another open challenge, is how to design and implement the distributed management model. The project is currently managed in a traditional, hierarchical manner with a single-person-of-failure root. We will be grateful for all suggestions and advice on this topic. Finally we would like to invite any persons or parties to participate in Project OneMap. You may contribute with data sets, pieces of software, storage nodes, processing servers, research projects, student projects/thesis, or just comments or ideas. Information is available at the OneMap site [\[ONE\]](#).

Footnotes

1. This paper is part of the OneMap Documentation Series (OMD 00503). The OneMap publications serve two purposes: 1) To disseminate knowledge and experiences that might be useful for other parties and applications, and 2) To encourage feedback and discussion in a broad setting. The presented results are *not* to be regarded as "final solutions", rather as a starting points for further discussions and developments.

Acknowledgements

The authors are grateful for the financial support offered by Østfold University College, which made attendance on the Digital Earth 2003 conference possible. They would also like to thank their colleague Børre Stenseth for interesting conversations on the topic and for contributing to [Chapter 2](#), and Grete Rindahl for proof reading and valuable comments.

Bibliography

[BAN]

The Band. Jan Høiberg, Østfold University College, Halden, Norway
<http://theband.hiof.no>.

[BAR]

On Distributed Communications: I. Introduction to Distributed Communications Network. Paul Baran, The RAND Corporation, Memorandum RM-3420-PR August 1964
<http://www.rand.org/publications/RM/RM3420/>.

[BER]

The World Wide Web: Past, Present and Future. Tim Berners-Lee, Director of the World Wide Web Consortium, August 1996 <http://www.w3.org/People/Berners-Lee/1996/ppf.html>.

[BES]

Digital longevity. Howard Besser, in Handbook for Digital Projects: A Management Tool for Preservation and Access, Maxine Sitts (ed.), Andover MA: Northeast Document Conservation Center, 2000, pages 155-166 <http://www.gseis.ucla.edu/~howard/Papers/sfs-longevity.html>.

[BRU]

Our common future. World Commission On Environment And Development, Oxford University Press, 1987.

[DIG1]

Digital Earth HomePage. Digital Earth <http://www.digitalearth.gov/>.

[DIG2]

Proceedings Digital Earth 2001. Digital Earth 2001, Fredericton, New Brunswick, Canada http://www.digitalearth.ca/ses1_10.html.

[DOM1]

The Domesday Book. www.domesdaybook.co.uk <http://www.domesdaybook.co.uk/index.html>.

[DOM2]

The Domesday Project. Atchison Topeka and Santa Fe Limited <http://www.atsf.co.uk/dottext/domesday.html>.

[GLO]

International Steering Committee for Global Mapping. International Steering Committee for Global Mapping <http://www.iscgm.org/html4/index.html>.

[GML]

OpenGIS®® Geography Markup Language (GML) Implementation Specification, version 3.0. Open GIS Consortium, Inc., 2003-01-29 (OGC 02-023r4) <http://www.opengis.org/techno/documents/02-023r4.pdf>.

[GOR]

The Digital Earth: Understanding our planet in the 21st Century. Al Gore. California Science Center, Los Angeles, Ca, January 31, 1998 <http://www.digitalearth.gov/VP19980131.html>.

[GSH]

GSHHS, A Global Self-consistent, Hierarchical, High-resolution Shoreline Database. Dr. Paul Wessel, SOEST, University of Hawai'i, Honolulu, HI <http://www.ngdc.noaa.gov/mgg/shorelines/gshhs.html>.

[LOC]

Permanent Publishing on the Web. Vicky Reich, Director LOCKSS Program, Stanford University Libraries and Academic Resources <http://lockss.stanford.edu/>.

[ISO]

Geographic information - Geography Markup Language. International Organization for Standardization <http://www.isotc211.org/scope.htm#19136>.

[MIS1]

The One Map Project. Gunnar Misund and Knut-Erik Johnsen. In online proceedings from GML Dev Days, Vancouver, July 2002 http://www.gmldev.org/GMLDev2002/presentations/MakingMapsfromGML/johnsen/one_map_misund_johnsen.html.

[MIS2]

Hierarchical GML Modeling of Transportation Networks. Gunnar Misund, Knut-Erik Johnsen and Mats Lindh. In online proceedings from GML Dev Days, Vancouver, July 2003 <http://not-yet-published>.

[MIS3]

Distributed GML Management with SVG Tools. Gunnar Misund, Henning Kristiansen, Mats Lindh, Presentation, SVG Open 2003 Conference and Exhibition, Vancouver, Canada, July 13-18, 2003 <http://www.svgopen.org/2003/papers/DistributedGmlManagementWithSVG/>.

[NAS]

Digital Earth Workbench. NASA Goddard <http://webserv.gsfc.nasa.gov/DE-Workbench>.

[NIJ]

Arealressurskart AR 250. NIJOS (Norwegian Institute of Land Inventory) <http://dmk.nijos.no/arfakta.htm>.

[OCN]

OpenContent License (OPL). OpenContent, Version 1.0, July 14, 1998. <http://opencontent.org/opl.shtml>.

[OGC]

OGC Homepage. OpenGIS® Consortium <http://www.opengis.org/>.

[ONE]

Project OneMap <http://www.onemap.org>.

[OSR]

The Open Source Definition. Open Source Initiative (OSI), Version 1.9 <http://www.opensource.org/docs/definition.php>.

[RAY]

The Cathedral and the Bazaar. Eric S. Raymond <http://www.catb.org/~esr/writings/cathedral-bazaar/>.

[ROS]

The Rosetta Project. The Rosetta Project/The Long Now Foundation
<http://www.rosettaproject.org/live>.

[SVG]

Scalable Vector Graphics (SVG) 1.1 Specification. W3C Recommendation 14 January 2003 <http://www.w3.org/TR/SVG11/>.

[UCS]

2002 TIGER/Line Files. U.S. Census Bureau <http://www.census.gov/geo/www/tiger/tiger2002/tgr2002.html>.

[UN1]

Information For Decision-making. United Nations, Agenda 21, chapter 40, United Nations Conference on Environment and Development, Rio de Janeiro, 1992
<http://www.igc.org/habitat/agenda21/ch-40.html>.

[UN2]

Programme for the Further Implementation of Agenda 21. United Nations General Assembly <http://www.un.org/documents/ga/res/spec/aress19-2.htm>.

[WFS]

Web Feature Service Implementation Specification. Open GIS Consortium, Inc., 19-September-2002 (OGC 02-058) <http://www.opengis.org/techno/specs/02-058.pdf>.