Geo-Processing Networks
In a
European Territorial Interoperability Study

IST-1999-14146

Information Requirements and Data Strategy

Document Control Page

Document Title:	Information Requirements and Data Strategy (Formerly: Data Domain Model Report)
File Name:	D3.4.3_r10.doc
Compiled by:	Lance McKee and Guenther Pichler
Partner's Name:	OGCE
Position:	Principal Contractor
Contribution of:	All Partners
Deliverable number:	D 3.4.3
Action requested:	☐ to be revised ☒ for approval ☐ for information ☐ other
Status:	☐ internal revision ☐ rejected ☐ approved ☐ public ☒ restricted
History:	01/Nov/2002: First Submission For Review 01/Dec/2002: Revision after Review 13/Mar/2003: Re-Submission For Approval

Table of Contents

Table of Contents (cont.)

1 Summary

1.1 The GETIS Objective

The objective of GETIS is to show how it is possible in Europe today to build Web-based networks that enable people to quickly find and use geographic information for disaster management and many other purposes. Like the Web, the telephone system, and all other networks, geographic information networks depend on standards. This summary of the GETIS "Information Requirements and Data Strategy" document provides an overview of the basic methods and essential standards involved in building a Web-based geographic information network.

1.2 The GETIS Approach

GETIS was founded to better understand end-users' spatial information requirements and to inform and inspire data owners to embark on geoprocessing network pilot projects large and small across Europe, transforming isolated data stores into nodes on an interoperable information network. The GETIS partners recommend that organizations, singly or as consortia, undertake pilot projects to develop local experience with new methods, products and vendors prior to committing to major technology deployments. It especially makes sense for co-operating 'data-sharing communities' to work together in pilots, because pilots provide a focused way of addressing the specific local interoperability issues of greatest concern to those communities.

This above mentioned models and the related standards have been adopted in a GETIS Proof of Concept implementation, which highlights the benefits of adopting Standards Compliant Software Package and Components to link the various players in the information and data network for a chosen disaster management scenario. Of course, as in any system integration or software development project, it makes good sense to:

- Involve users in early stage planning as much as possible.
- Record requirements, goals and concepts in detail. It is good to get everything on paper and be sure that stakeholders thoroughly review and discuss the project.
- Experiment early and often. In ways that require little investment of time or money, learn what works and doesn't work and why. Disseminate these lessons and use them in early planning and during the pilot. Involve competing vendors. Get the commitment of the vendors who provided the legacy systems, which they must upgrade with interoperability interfaces if the legacy systems' capabilities are to become network-accessible resources for users of software developed by other vendors.

1.3 Model Architecture and Standards

The GETIS partners recommend that pilot project planners briefly review the ISO/IEC "Reference Model for Open Distributed Processing" (RM-ODP) and OGC's "OpenGIS Reference Model", released by OGC in February 2003, to develop their own architectures. The Reference Model for Open Distributed Processing is used by information technologists attempting to incrementally connect large numbers of heterogeneous component systems for the purpose of addressing a complex set of real world problems. Based on this, the OpenGIS Reference Model describes in detail the architectural model and interoperability standards, which are adopted by the industry and should be applied to achieve an interoperable information network, serving spatial information requirements.

1.4 GETIS Proof of Concept

In planning a geographic information network, one first looks at the network users' information requirements and the context of the users' workflow in which the requested information feeds into. The GETIS Proof of Concept is based on a simplified inland flooding incident in the South West of England. The scenario describes disaster management activities typical of how flooding incidents are dealt with at the local government level anywhere in the world. The schematic below shows the information flow for a simplified flooding scenario:

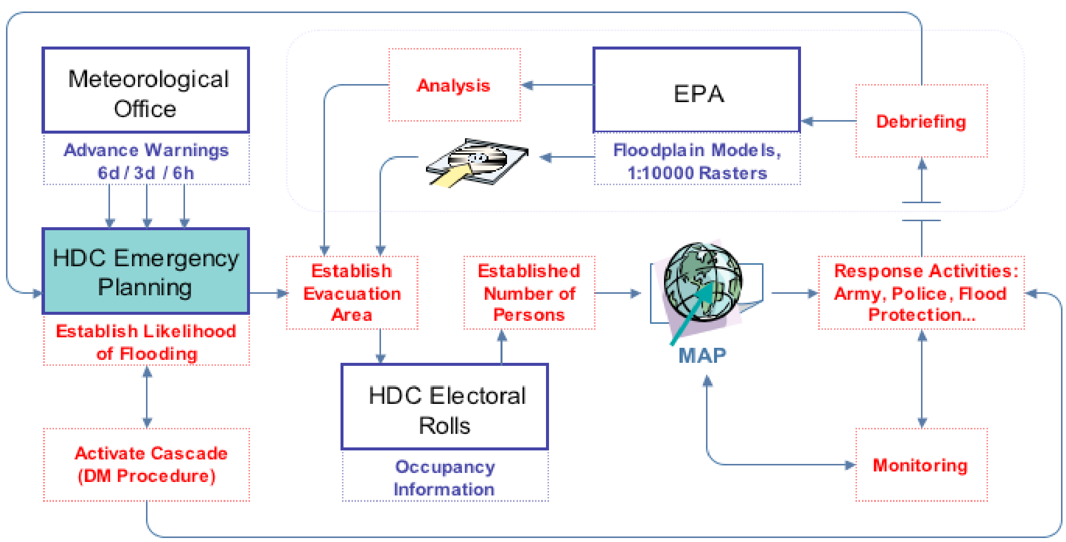


Figure 1: Information Flow in Horsham District (HDC) Flood Scenario

When advance warnings based on actual precipitation and groundwater levels reach the Horsham District Council (HDC) and likelihood of a flood is established, modelling of floodplains and anticipated water levels indicates the region to be evacuated. That area information is compared to occupancy data to produce a list of people and businesses who will need to be notified and helped. Geospatial and other forms of data come from and are used by a variety of organizations, including the Ordnance Survey, The District Council, and the Environment Protection Agency.

Below (Figure 2) is a general architecture diagram for the Proof of Concept, in which disaster managers use Web-based client applications to request information from (or add information to) a heterogeneous collection of data servers and processing services. These requests are mediated by a web services infrastructure that includes catalogues populated with data and service meta-information that describes the available data and information services. The infrastructure also provides the means for software "binding", i.e. communication between the service and the remote client application, facilitating the provision of information as opposed to the traditional exchanges of mere raw data sets:

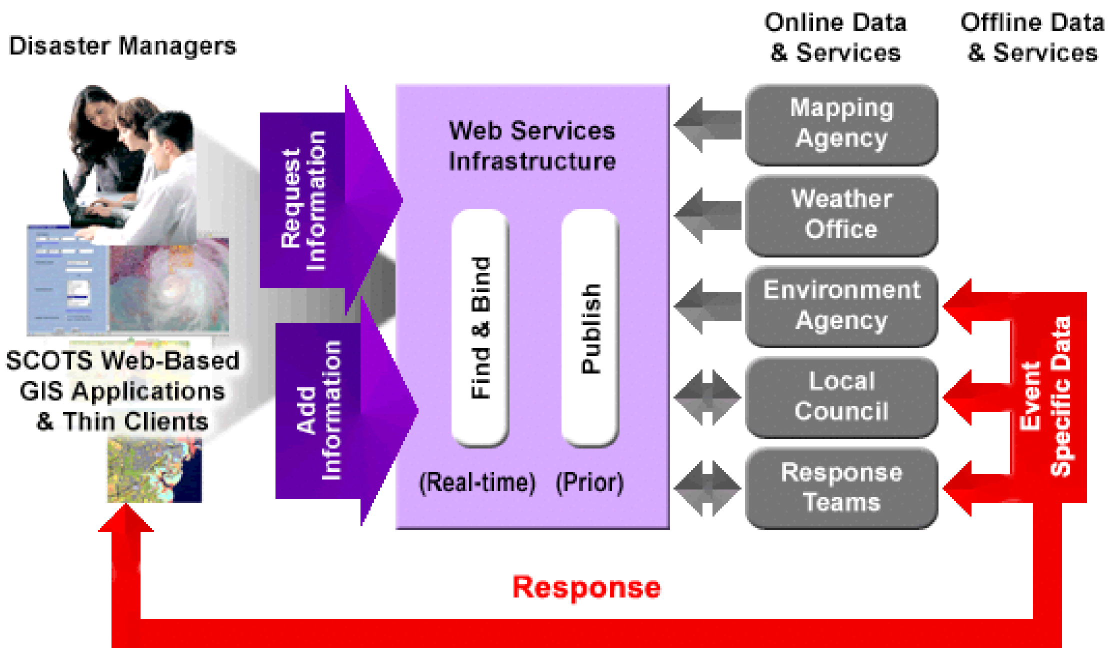


Figure 2: GETIS Architecture

Based on OpenGIS Consortium and ISO standards, the GETIS conceptual architecture provides a framework for unifying spatial information flow involving any number of Web-connected participants.

Adding information (lower purple arrow) typically means updating and correcting databases managed by the Local Council or Response Teams, verifying requests, making notes of secondary events and posting new observations online so they are available to anyone who needs them. The online data held by the various organizations typically includes the following:

Mapping Agency: Boundaries, Digital Elevation Models and Ortho Imagery.
Weather Office: Precipitation rates, wind speeds and weather warnings
Environment Agency: Flood Plain Models, Risk Area Data, Sensor Network Data
Local Council: Property lists, tax rolls, and contact information
Response Teams: Field reports and resource locations, local expertise

The Online Services available from these organizations' servers may be as simple as services enabling updates of databases and automatic coordinate transformation and overlay of data from multiple sources (demonstrated in the Proof of Concept), or they might be complex decision support systems (not part of the Proof of Concept). The two important things to note are:

1. Catalogues containing data and service meta-information describe the distributed online data and services in machine-readable XML text (see below). This enables users to find data quickly and enables client-side applications to find server-side processing capabilities to do the computing needed by the user.
2. Such data and services must be accessible across the Web to people using a variety of different vendors' GIS software products, or custom applications, or commercial database, facilities management or other geoprocessing software products. This is enabled by the Web Services infrastructure and open software interoperability interfaces as defined in OpenGIS Specifications.

This exemplifies the transition from data-centric systems that require expert manipulation of data to information-centric services that require much less data expertise.

1.5 Role of Geography Markup Language (GML)

XML (eXtensible Markup Language) is a system for encoding data in text. It is an essential part of the Web Services infrastructure and a key to next-generation Web features. The first lines in an XML file provide syntax, mechanisms and conventions that establish how subsequent lines are to be interpreted. Many industries are developing and reaching consensus on industry-specific XML namespaces, which are carefully defined sets of terms based on the data discovery, data integration and data presentation requirements of the industry. As plain text, XML can be read and understood by data managers and it can also be parsed by software programs (typically programs that are part of Web browsers).

OGC's Geography Markup Language (GML) is an XML encoding, or namespace, for geographic information, including both the spatial and non-spatial properties of geographic features. The GML specification defines the XML Schema syntax, mechanisms, and conventions that provide an open, vendor-neutral framework for the definition of geospatial application schemas and software objects. GML allows profiles or subsets of GML data schemas that can be geospatial application schemas for specialized domains and information communities. Commonality among data schemas enables the creation and maintenance of Internet-linked geographic schemas and datasets. Thus, more easily than before, organizations can maintain a practical level of commonality between their geographic application schemas, and this enables them to share their geographic data much more effectively and efficiently.

The main benefits of using GML are described below:

- One of the keys to GML's success is the OpenGIS Web Feature Service (WFS) Specification, a companion specification to the GML specification. To get GML data, a user issues a query to a web server with an OpenGIS Web Service Interface (a "Web Feature Server"). The standard interface enables access to the feature store and enables users to add, update or retrieve data, locally or across the Internet. The data can be retrieved as GML data, even if the data store uses a different format internally. Users no longer need to care whether the underlying store is from ESRI, SICAD, Oracle or IBM. This will be demonstrated in the GETIS Proof of Concept.

- GML maintains separation between feature data and portrayal (presentation style). This means, for example, that a certain road type defined in a data model does NOT carry with it a "hard-wired" color or pixel width. Road data from a server might be portrayed by one application client as a three-pixel wide red line, and it might be portrayed very differently by another client, perhaps a client that converts text to speech to say, for example, "two-lane road." This will be demonstrated in the GETIS Proof of Concept.
- GML is rapidly becoming an international de facto standard, and through OGC's procedural arrangement with ISO TC/211, GML is on track to become an ISO standard.
- In the future, GML will enable a new degree of semantic interoperability. GML makes it practical to allow different "application profiles" of the base data. That is, software can be written that translates between slightly different feature definitions, enabling one information community to use another information community's data with relative ease and confidence. (This is still in the research stage and is beyond the scope of the GETIS Proof of Concept.)

1.6 Conclusions

GETIS demonstrates how to apply accepted principles for building distributed geoprocessing systems in an incremental and interoperable manner. It shows how previously isolated legacy spatial systems and data repositories can be "wrapped" with OpenGIS interfaces to allow plug and play with other spatial systems via Web Services. Government departments at all levels can use the principles and standards described in the GETIS Project to express their complex requirements for open architectures, including architectures for decision support tools that need to share data and online services scattered across jurisdictions. It is hoped that Government departments at all levels in Europe will team to undertake pilot projects in the next year or two, aimed at prototyping elements of a European "Spatial Web" that will be useful for governments at all levels, and for individuals and businesses as well.

1.7 Report Structure

In its first part, this report describes a specific disaster management scenario in detail, i.e. the generalized inland flooding incident in West Sussex, UK. It summarises actors, information requirements and sources, and outlines an approach to provide relevant information utilising existing standards-based commercial off-the-shelf ("SCOTS") technology in conjunction with the existing data-sources in their current proprietary formats. A proof of concept for the use case scenario is being built on real world data provided by the involved agencies. It demonstrates how end-users benefit from the paradigm shift from data-centric systems towards information-centric services.

The second and major part outlines a GML-based approach to sharing some of the involved data-sets, directly providing interoperable information as opposed to standard data-sets with their inherent need for a certain level of pre- or post processing. GML encoded data used to explain this example has been obtained from the Ordnance Survey.

The final part reviews potential gaps between commercially provided GML data of a national mapping agency and the information requirements of the end user, trying to show a way forward to translate common base line data-sets into base line information sets.

1.8 Evolution of Objectives

The GETIS project as originally conceived included a "domain modelling" effort which would have resulted in a prescribed family of data schemas appropriate to the use cases identified as key information in Europe that can be shared by information communities.

In WP2, a working group of experts was established to validate this approach. When defining application scenarios and identifying the relevant information needed in decision making processes, it was determined that most requirements can be satisfied by existing data sources, but that there's a distinct gap in deriving user-relevant Information from the existing data-sets.

Since the dawn of GIS, many data formats have been developed for the digital encoding of geographic information and almost all of them are proprietary to a vendor specific GIS. Though a few ad hoc standards were created to enable transfer of spatial data between different vendors' or organisations' GISs, no scheme for interoperability based on a common format or even interfaces has been agreed upon on a larger scale.

To overcome the problems associated with proprietary data formats, the Open GIS Consortium (OGC) has specified common interfaces that enable systems to interoperate despite their proprietary data formats and the Geography Markup Language (GML), an extension of the World Wide Web Consortium's XML (eXtensible Markup Language) standard, as a common exchange format.

GML is based on the abstract model of geography that is a central element of OGC's interface architecture. It is not an interface but an encoding designed for the transport and storage of geographic information, including both the geometry and properties of these features. Users develop *data schemas* consisting of a set of geographic feature descriptions and feature attributes appropriate to their application requirements. (In OGC parlance, a data schema designed for a particular application is called an *application schema*, to distinguish it from more general data schemas, such as those used in framework data sets.) Users who produce or use the same data schemas comprise "*information communities.*" With GML a new, more practical approach to creating similar yet more distinct "families" of data schemas is at hand, which allows standardised and efficient access to data sources, enabling and supporting the process of deriving and disseminating information.

Because XML provides a powerful platform for parsing and operating on structured text, the GML data modelling approach can greatly reduce the need for re-capturing and updating basic data-sets to comply with information community specific data models. Thus, the traditional approach to domain modelling is no longer a part of the GETIS project. To facilitate European adoption of this newer technical approach, the GETIS project was reoriented to demonstrate practical applications of GML using commercially available products.

2 Scenario Description

2.1 Scenario

The selected flooding use case scenario is based on process stages during an inland flooding incident in the area of Horsham District. It is not based upon one single specific event but on a description of typical process stages, which are more or less generic to how flooding incidents are dealt with at the Local Government level.

2.2 Information Flow

The schematic overview of the information flow for a highly simplified flooding scenario shows the number of involved parties and the different methods of communicating information that's based on a variety of data sources:

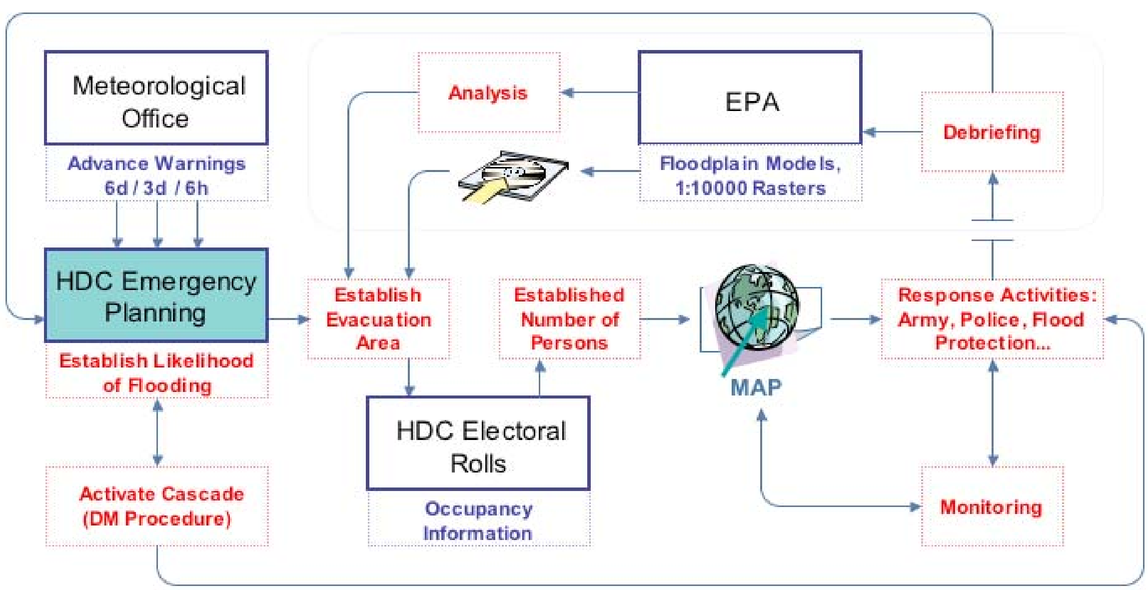


Figure 1: Information Flow in Horsham District Flood Scenario

When advance warnings based on actual precipitation and groundwater levels reach the Commission, and likelihood of a flood is established, modeling of floodplains and anticipated water levels yields information about the region that will need to be evacuated. That areal information is compared to occupancy data to produce a list of people and businesses who will need to be notified and helped.

Geospatial and other forms of data come from and are used by a variety of organizations:

The Ordnance Survey as National Mapping Agency provides raster and vector datasets with terrain information at several scale levels and with varying level of detail and accuracy. The traditional data distribution method is still to use data carriers such CD-ROM for bulk distribution as well as updates, though considerable efforts are currently spent to facilitate object oriented online provision of data for the large-scale vector data sets. The forthcoming new products, which will be suitable for commercial Internet based data dissemination models, would at the same time allow us to consider specific services, which provide only selected subsets of required features for an application domain. This might be a route to overcome licensing and funding issues, which at this point in time seem to impede the widespread application of high quality mapping products.

Horsham District Council represents a fairly typical UK Local Authority with widespread administrational responsibilities for its 200 square miles are just south of Gatwick Airport and the Greater London Area. It certainly differs from most other Authorities in the fact that corporate GIS has played a central role in its IT strategy and therefore a wealth of geospatial information is now available in digital formats, which can support a number of business processes well outside the initial application area of the Planning & Building Control Office, where most of these sets originated. Extensive data sets, such as contaminated land registers, potential hazard sites, demographics etc are held within the councils various IT databases and it can be assumed that this represents a typical cross-section of data found on a local authority level. However, most data has restricted availability for various reasons of administrational, legal or technical nature.

At the moment e.g. high quality OS data sets are used in the planning department under the local authority framework agreement with the OS, but they are not significantly shared with other department at the moment. Since in principal all geospatial data is hold in SICAD GIS systems, which are OGC WMS1 compliant, at least internal dissemination of this data through the intranet to simple browser based applications could be relative easily achieved.

In theory all data sets on this level could be shared at least on a proprietary GIS / DB exchange file level on CD-ROM as well.

Environment Protection Agency datasets on potential floodplains are published annually on CD. The datasets include areas modeled on the OS DTM and come with a 10k raster background supply, which however seems to be outdated by at least 5 years. In its current version the dataset is a suitable guidance or decision support set of data, but isn’t included in an automated process.

2.3 Information Requirements and Related Datasets

The following table provides an overview of data sources available to one or more departments of Horsham District Council. The column 'used' refers to the availability of a specific data set to the emergency planning office, 'restricted use' indicates whether there are any potential issues with sharing this data set internally:

Dataset	supplied on-line	supplied on data carrier	supply formats or standards	used	update cycle	restricted use
Floodplains	no	CD-ROM	SHP, MIF	yes	annually	no
OS Raster 10k via EPA	no	CD-ROM	TIF	yes	annually[1]	no
OS Raster 250k, 50k, 10k and 5k via OS	no	CD-ROM	GEOTIF	no	?	yes
OS Land Line via OS	no	CD-ROM	DXF or NTF	no	monthly	yes
OS Raster 250k, 50k, 10k and 5k via HDC internal	yes	CD-ROM	OGC WMS1 or TIF	no	as OS	?
OS Land Line via HDC internal	yes	CD-ROM	OGC WMS1 or DXF	no	as OS	no
HDC Corporate Property Database	yes	no	OGC WMS1	no	weekly	no
HDC Electoral Register	no	no	?	yes	?	A
HDC Planning Constraints Database	yes	no	OGC WMS1	yes	weekly	no
HDC Contaminated Land Register	…					

Restriction types: T=technical, A=administrational, L=legal

[1] Though this data set is supplied annually, it usually seems to be at least 5 years out of date.

3 Proof of Concept

To demonstrate the different stages in moving from traditional data-centric set-ups to information networks serving user requirements, a Proof of Concept is being implemented as a tool to aid dialogues with user communities. The demonstration will:

- Provide an overview of a set-up with traditional data sources to address the requirements of the selected use case scenario specifically and other scenarios in general. That is, the demonstration will go through the customary steps of opening and converting data files with conventional desktop mapping tools.
- Simulate potential online information services based on real world information requirements and real world commercial data sets. This will involve taking the standard data and plugging it into web services using "SCOTS" (Standards-based Commercial Off-The-Shelf) products. This is the beginning of the conversion from data-centric systems that require expert manipulation of data to information-centric services that require much less data expertise. This web services approach will also:
    - Illustrate how geospatial information can feed into mainstream ISTs.
    - Help users visualize the way to move towards an interoperable information service network supporting end-users in production environments.
    - Demonstrate the benefit of European vendors' and other countries' vendors' SCOTS products interoperating because of international participation in OGC standards initiatives.
- Show how GML can provide further benefits, particularly in the area of data sharing, as described below.

3.1 Proposed Architecture

The Proof of Concept and the entire GETIS project are based on an architecture that outlines how a particular set of user needs are to be addressed with a particular set of technologies that work together in an open distributed computing network.

3.1.1 Open Distributed Processing Reference Model (ODP-RM) applied to GETIS

To best way to create such an architecture is to use the "Open Distributed Processing Reference Model" (ODP-RM) that has been defined by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). The ODP-RM is applied by information technologists attempting to connect large numbers of heterogeneous component systems for the purpose of addressing a complex set of real world problems. "It creates an architecture within which support of distribution, interworking and portability can be integrated." (The Reference Model of Open Distributed Processing, ITU-T Recommendations X.901 to X.904 | ISO/IEC 10746.) An introduction to the ODP framework can be found at http://www.iso.ch:8000/RM-ODP/part2/toc.html.

The ODP-RM defines a hierarchy of six “viewpoints” from which to view, think about, design, and plan a distributed processing system:

1. *Contextual viewpoint*: the pre-existent and external contexts within which the resultant ODP must be developed. In the case of GETIS, this includes the overall social, cultural, political, legal, bureaucratic and financial environment in which disaster management operates in Europe.

2. *Enterprise viewpoint*: A viewpoint on an ODP system and its environment that focuses on the purpose, scope and policies for that system. The GETIS proof of concept enterprise viewpoint is described in the roles and activities of the data users and providers. The *purpose* of the system developed for the proof of concept is to make it easier for disaster managers to quickly get the information they need at every step in the disaster management process. The *scope* is limited to a specific set of operations that are deemed to be "typical" during an inland flooding incident in the area of Horsham District. The *policies* in the proof of concept likewise reflect typical inter-organizational policies of participating agencies during a flooding incident.

3. *Information viewpoint*: A viewpoint on an ODP system and its environment that focuses on the semantics of information and information processing. The information viewpoint is described generally in the discussion of GML below and will be given specific detail when particular GML data samples are created according to certain schemas and provided with metadata.

4. *Computational viewpoint*: A viewpoint on an ODP system and its environment, which enables distribution through functional decomposition of the system into objects which interact at interfaces. The computational viewpoint for the GETIS proof of concept will show a set of software objects or functions (in web clients and servers) interoperating through interfaces that implement particular OpenGIS® Implementation Specifications, including OpenGIS Catalog Service, Coordinate Transformation Service, Web Map Service, Web Feature Service and Web Coverage Service in publisher client, viewer client, map server, catalog server and image server. Other web services and their associated standards infrastructure are also part of this viewpoint.

5. *Engineering viewpoint*: A viewpoint on an ODP system and its environment that focuses on the mechanisms and functions required to support distributed interaction between objects in the system. In our case, such mechanisms and functions would include the programming details and interface mechanics involved when the web is used as a distributed computing platform. It will include descriptions of thin servers and thick servers, thin clients and thick clients, dial up and broadband connections, and platforms like http (and probably not COM and CORBA). It will also show the underlying hardware systems – PCs, workstations, and handheld computers.

6. *Technology viewpoint*: A viewpoint on an ODP system and its environment that focuses on the choice of technologies in that system. This viewpoint describes details of particular geoprocessing software products selected to meet the requirements defined in the other viewpoints, and also details about databases, hardware, operating system, development tools, and programming languages.

3.1.2 Architecture Diagram

Below is an architecture diagram for the proof of concept, in which disaster managers use application servers to request information from (or add information to) a collection of data servers and processing servers. These requests are mediated by a web services infrastructure that relies on catalogues populated with data and service meta-information that describes the available data and services.

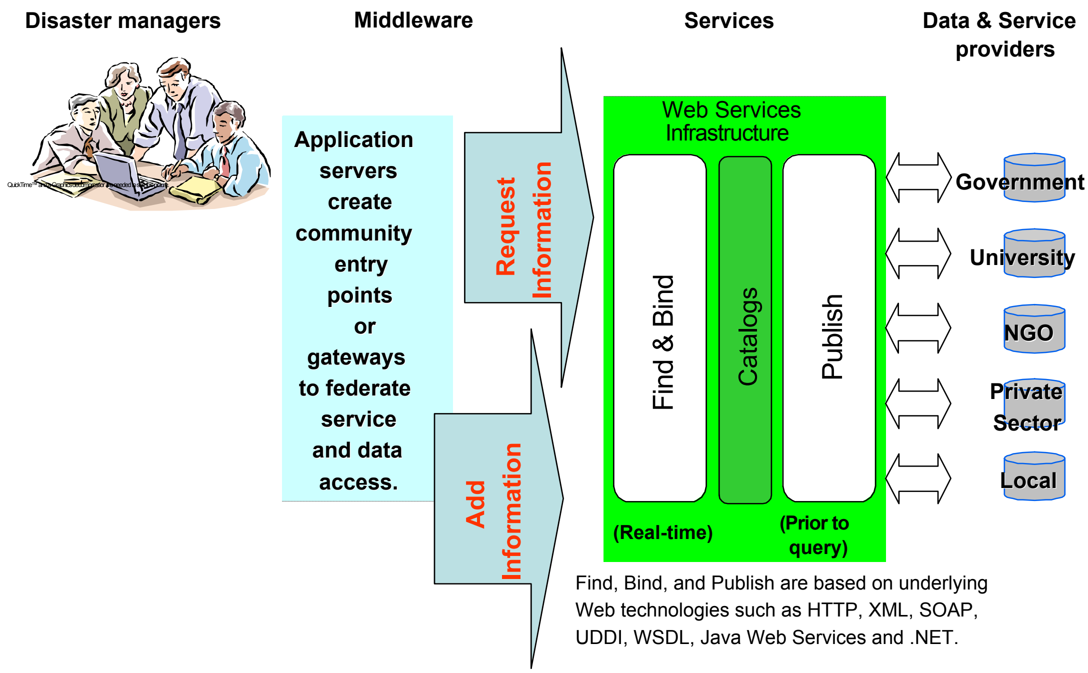


Figure 1: GETIS Architecture

As the Proof of Concept is described in more detail, this diagram can be elaborated to show particular requirements of particular people and roles and the particular application servers and data & service providers they employ to get the information they need. A more fully elaborated diagram will look more like the information flow diagram in Figure 1, but each information request and response will be based on this conceptual diagram.

3.2 Designing a Pilot

The goal of GETIS is to provide a model for organizations that will be developing or investing in the development of distributed geoprocessing systems in Europe. These will not, of course, be limited to disaster management, but will be created for a wide variety of purposes. In the beginning, it is likely, and it is recommended by the GETIS partners, that organizations will singly or as consortia undertake pilot projects to help develop local experience with new methods, products and vendors prior to committing to major technology deployments. It particularly makes sense for cooperating data sharing communities to work together in pilots, because pilots provide a very focused way of addressing the specific local interoperability issues that have highest priority.

The GETIS partners recommend that pilot project planners study the ODP-RM generally and the GETIS architecture more specifically to sketch and then fill in the details of their own architectures. As in any system integration or software development project, it makes good sense to:

- Involve users as much as possible in early stage planning.
- Spend a good amount of time recording requirements, goals and concepts in detail. It is important to get everything on paper and be sure that all concerned stakeholders have thoroughly reviewed and discussed the project.
- Experiment early and often. In ways that require little investment of time or money, try to learn what works and what doesn't and why. Disseminate these lessons and use them in early planning and in "mid-course corrections" during the pilot. Involve competing vendors. Be sure those vendors who have provided the legacy systems that are currently in use are committed to the project's interoperability objectives.

4 Using GML

4.1 XML and GML Explained

XML (eXtensible Markup Language) is a means of encoding data in text. Although it started out as a language for “marking up” or encoding a document for selection and presentation (HTML has been written in XML), it has quickly evolved into a mechanism for general data description. Today, XML is used in a variety of industries including finance, chemistry, e-business, document publishing, multimedia, telecommunications, graphics, and e-government. Each of these domains is developing and reaching consensus on an XML *namespace,* which is a carefully defined set of terms, a *vocabulary* or *ontology,* based on the needs of the domain. Like domains, organizations and information communities are developing their own XML namespaces. As plain text, XML can be read and understood by data managers and it can be parsed by software programs. XML is easily transformed by simple programs, and it is easy to integrate and combine XML-based data from many disparate sources. Virtually all web browsers include software for processing XML, which is a major reason for its success.

The Geography Markup Language (GML) is an XML encoding for the transport and storage of geographic information, including both the spatial and non-spatial properties of geographic features. The GML specification defines the XML Schema syntax, mechanisms, and conventions that provide an open, vendor-neutral framework for the definition of geospatial application schemas and software objects. GML allow profiles or subsets of GML data schemas. Such profiles can be geospatial application schemas for specialized domains and information communities. Commonality among data schemas, with standard ways of creating profiles, enables the creation and maintenance of Internet-linked geographic schemas and datasets. Thus, much more easily than before, organizations can maintain a practical level of commonality between their geographic application schemas, and this enables them to share their geographic data much more effectively and efficiently.

Implementers may decide to store geographic application schemas and information in GML, or they may decide to convert from some other storage format on demand and use GML only for schema and data transport.

4.2 GML vs. DXF Data

Data sets in the essential disaster management data domains exist in many formats. In the old (and still dominant) paradigm of data sharing, data sets like these are converted in batch mode from one format to another so the data can be used in different software systems. Frequently data are stored in "interchange formats," formats designed to be convertible into many other formats. The best of these are supported by most vector GIS software packages and are sufficiently "rich" enough in capability to exchange spatial information with relatively little information loss. One commonly used interchange format is DXF (Drawing eXchange Format), introduced many years ago by Autodesk. Below we provide a brief comparison of DXF and GML.

We begin with the similarities, though it is the differences that are most important. GML is, viewed in the old paradigm, a data format, and commercial tools are available to convert DXF data into GML data or vice versa, just as DXF can be converted to or from perhaps a hundred other data formats. (See a list of these on Safe Software's web site, http://www.safe.com.) Both DXF and GML are hierarchical data files that are in some ways like databases. In DXF there are Headers, Classes, Tables, Blocks, Entities, Objects, and also a thumbnail image to provide a quick view of the drawing. DXF is a drawing format, but because it provides a flexible, hierarchical system for defining these categories of elements, it is much more sophisticated than pure drawing formats like HPGL, which provide only a simple linear series of plotter commands. Because DXF provides a formal system for defining complex components of various kinds, software programs operating on DXF files can discriminate among different data components in the files for the purpose of performing operations on specific components, such as changing the line colour used to represent a certain class of road. DXF files can contain and transfer much of the complex information in a multi-layer GIS data file that may have been originally created and stored in a GIS's sophisticated spatial database.

(A DXF tutorial can be found at http://astronomy.swin.edu.au/~pbourke/geomformats/dxf2000/, and the UK Ordnance Survey's Land Line DXF data User Guide can be downloaded from http://www.ordsvy.gov.uk/downloads/vector/landline/Land-Line%20user%20guide.pdf)

Zeros begin and end definitions in DXF. Below is a declaration of an object "171_Laubbaum" (in English, "deciduous tree") in an AutoCad file:

```
0
LAYER
  5
27
100
AcDbSymbolTableRecord
100
AcDbLayerTableRecord
  2
171_LAUBBAUM
 70
     0
```

In GML (an XML "namespace"), elements begin with a name enclosed in angle brackets and end with the same name enclosed in angle brackets where the first angle bracket is followed by a "/". A corresponding declaration of a "171_Laubbaum" feature type in a GML file looks like this:

```
<Feature>
<featureType>171_LAUBBAUM_line</featureType>
<property name="gml2_type">gml2_line</property>
```

```
</Feature>
```

Like DXF, GML includes a hierarchical set of abstractions: global declarations, common aliases for geometric properties, feature descriptors, abstract supertypes, etc. So GML files, like DXF files, can be parsed and manipulated for a variety of purposes.

Below is another example, data – mainly four vertices – describing the location of a particular school. In DXF:

```
  0
VERTEX
  5
3BCD7
100
AcDbEntity
  8
SCHOOL
100
AcDbVertex
100
AcDb2dVertex
 10
490776.989999
 20
5463274.007635
 30
0.0
  0
VERTEX
  5
3BCD8
100
AcDbEntity
  8
SCHOOL
100
AcDbVertex
100
AcDb2dVertex
 10
490783.767999
 20
5463284.043635
 30
0.0
  0
VERTEX
  5
3BCD9
100
AcDbEntity
  8
SCHOOL
100
AcDbVertex
```

```
100
AcDb2dVertex
 10
490793.683999
 20
5463277.347635
 30
0.0
  0
VERTEX
  5
3BCDA
100
AcDbEntity
  8
SCHOOL
100
AcDbVertex
100
AcDb2dVertex
 10
490786.905999
 20
5463267.310635
 30
0.0
  0
SEQEND
  5
3BCDB
100
AcDbEntity
  8
SCHOOL
  0
POLYLINE
  5
3BCDC
100
AcDbEntity
  8
```

and in GML:

```
<gml:featureMember>
        <exp:School fid="School5057">
            <gml:description>Buildings school</gml:description>

            <gml:name></gml:name>
            <gml:extentOf xmlns:gml="http://www.opengis.net/gml">
                <gml:Polygon >
                    <gml:outerBoundaryIs>
                        <gml:LinearRing>

<gml:coordinates>490776.989999387,5463274.007635
```

```
490783.767999392,5463284.04363501 490793.683999386,5463277.347635
490786.905999392,5463267.310635
490776.989999387,5463274.007635</gml:coordinates>
                        </gml:LinearRing>
                    </gml:outerBoundaryIs>
                </gml:Polygon>
            </gml:extentOf>
        </exp:School>
    </gml:featureMember>
```

The essential high-level difference between GML and DXF is that GML is not an artefact of the computer-aided drawing world, but an artefact of the Internet world. GML is an extension of (or more precisely, a namespace in) XML (eXtensible Markup Language). In the world of the Internet, applications must be able to talk to each automatically and in near real-time, irrespective of their existing formats and platforms. XML has emerged as the standard World Wide Web data format, recognized by all the Web platforms & applications across the world. Virtually all web browsers can parse XML. But "data format" is an imprecise term to use in describing XML. A "markup language" is more than a data format, and XML is in fact a set of rules for defining markup languages. That is, the first lines in an XML file establish how an application is to interpret subsequent lines in the file. This is a very powerful basic concept. "Namespace" is central to this concept.

An XML "namespace" is an application-specific set of global declarations, aliases, feature descriptors, and abstract supertypes used by people and organizations whose communication and collaboration can be supported by such a shared vocabulary. Or more precisely, this set of definitions enables people in such a community to set up their software and data to interoperate semantically, thanks to the shared vocabulary. GML is a namespace that establishes, by its technical comprehensiveness (it is based on the OpenGIS Abstract Specification) and its early and well promoted leap into the XML world, a global standard set of geospatial declarations, aliases, descriptors and supertypes. Version 2.0 of the specification bears the status of "adopted specification" within OGC, which means that GML is mature enough to be used in other implementation activities or incorporated into software products.

4.3 Extended GML Capabilities

Reviewing the DXF-GML comparison in closer detail, we see a number of significant features in GML that are not available in DXF:

In DXF, different feature types are "layers" in the data file. Features with areal extent are identified by a point or "seed" within the area. The software must infer the relationship between the seed and the area. To capture a feature, which is the only one of its type, a new layer is necessary. In GML, by contrast, the geometry associated with a feature instance in the real world is stored with the feature instance in the data. In GML, "portrayal" – the line style, polygon color, font, etc. used to portray a feature – is defined independently of geometry and feature attributes, so the same data can be programmed to "look" different in different applications. Similarly, feature labels and other displayable text can be translated from one language to another.

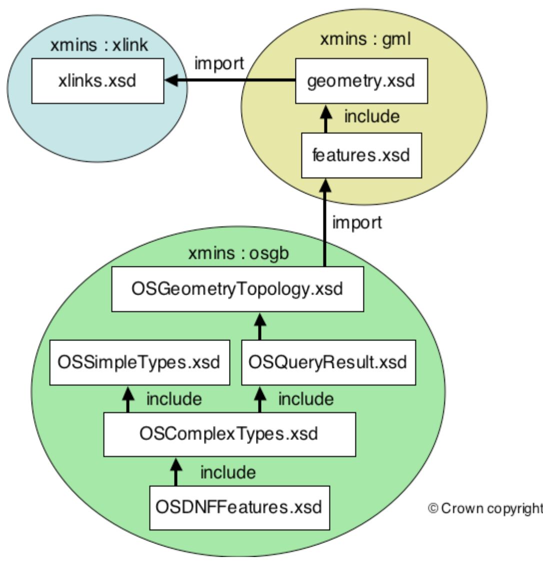
A far more dramatic difference is that in GML, a point can be provided with a link to a spatial reference system that is defined externally to the data file, somewhere out on the web. For example, in this specification of a point:

```
<gml:location>
<gml:Point srsName="http://www.opengis.org/srs/epsg.xml#4326">
<gml:coordinates>-0.125764,51.511289</gml:coordinates>
</gml:Point>
</gml:location>
```

the point's location is given in terms of a spatial reference system that an application can access in all its technical detail at the url provided. “epsg” is the European Petroleum Survey Group, the geodesy authority that has worked with OGC to make its database of coordinate reference systems available online as a resource to GML applications worldwide. This is a key enabler of automatic coordinate transformation, which is necessary for automatic overlaying of map layers from different web-resident data servers. And this is just one example of a network-resident resource linked to a GML file. Many other kinds of resources can be made available on web servers for use by GML applications via links embedded in GML files. Such resources can be data, metadata, data or metadata schemas, or processing resources. More than a fixed data format, XML is a language for writing self-describing objects, where an object can include both data and software that operates on the data. Online catalogs that will direct queries to appropriate resources will use XML as the language that describes all the various resource types.

“Domain modelling” in the GML context will involve such links to schemas, including both standard “framework” (base layer) schemas and application schemas. Schema repositories for core GML are available at http://www.opengis.net/gml and http://www.opengis.net/namespaces/gml/core/. Other GML schemas are implementations of GML, just as GML is an implementation of XML.

The Schema Structure diagram below (from p.116 in the OS MasterMap user guide v2.0.pdf) shows how schemas in GML use inheritance. GML is an instance of XML and the OS MasterMap schema is an instance of GML.



XML namespaces

xlink – http://www.w3.org/1999/xlink

gml – http://www.opengis.net/gml

osgb – http://www.ordnancesurvey.co.uk/xml/namespaces/osgb

DXF files and other GIS, CAD, and image formats do not take advantage of this object-oriented, web-based architecture.

Galdos Systems, Inc. (Vancouver, Canada) offers on their web site (http://galdossystems.com) the following list of the "Top 10 Benefits of Using GML." These benefits are described in the context of explaining how the GML approach is a great improvement over simple web-delivered GIF/JPG image maps, but they provide a good description of GML's overall benefits:

1. *Better quality maps*. GML encodes information about geographic features or objects, and these can be displayed to as fine a resolution as required. Thus, screen-based maps generated from GML appear crisp and easy-to-read. Such maps can also be saved as local files, emailed, or printed.
2. *Works on a browser, without the need to purchase client-side software*. When a GML file is received at the client, it is converted to a set of drawing objects and rendered as a map on the browser. Typically, Scalable Vector Graphics (SVG) is used as the drawing object language. As long as the browser supports vector graphics, then the map can be displayed without any additional software. Currently, SVG is supported by means of a free downloadable plug-in from Adobe Corporation (just like their PDF reader). No other special software is needed to view GML files.
3. *Custom map styling*. GML contains map "content" only (e.g., where features are, their geometry, type and attributes), but it does not provide any information about how that map data should be displayed. This is actually a benefit because different "stylesheets" can be applied to the geographic data to make it appear however the user wishes. One user may choose thin black dashed lines for roads and a cross symbol to depict churches, whereas another might choose thick solid red lines for roads and a CH sign for churches. The important point is that the GML data is the same in both cases, each group just invokes a different stylesheet to render the map. Choosing the stylesheet may be done automatically, or users may be given a choice of stylesheets to use.
4. *Editable maps*. It is quite straightforward to annotate GML-based maps that have been downloaded and rendered on a browser. Once GML has been converted to SVG, the user can apply graphic editing tools on the client to add text (of any font, size and color), highlight features, and draw virtually any kind of shape on the map. The annotated map graphic can then be saved as a file, emailed, or printed.
5. *More sophisticated linking capabilities*. One of the benefits of GML is that you can embed links associated with features. These links can be simple URL addresses, or they can be more sophisticated. But, at its simplest, this means that you can associate any Web address with a feature. When a user clicks on a feature, the user is transferred to that address. This capability can be used to initiate a feature-based query or to take a user to a new Web page (e.g., click on the opera house and be taken to its home page).
6. *Better query capability*. Often users want to be able to click on a feature on a map and find out more about it (e.g., what is the name of this river?). For GIF/JPG-style maps either this cannot be done at all, or if it can be done, the query mechanism is quite rudimentary and involves measuring the pixel position of the cursor, translating this to geographic coordinates on the server, and then looking for the required feature in the server-side GIS. Limited pixel resolution and the vagaries of different browsers often limit the accuracy of this method, meaning that frequently the user does not get back information about the feature they clicked on. This problem does not arise in the world of feature-based GML. When you click on a feature, you will always unambiguously identify that feature. And, by means of turning on and off different feature themes, it's easy to identify features within features (e.g., a house within a lot).
7. *Control over content*. Because GML is feature-based, it is quite easy to provide a filtering function that allows users to download only the feature-types that they want to appear on their maps. For example, if you don't care about railway lines, then you don't have to download them from the server. This filtering can reduce data transfer time. Map content can also be controlled after the geographic information has been delivered to the client's Web browser. Using a click-

able legend, a user can display/hide information themes instantly, and without the need to call the server to generate and deliver a new map. This cannot be done with GIF/JPG maps.

8. *Animated features*. Objects and features that change over time can be accommodated in GML, and can be rendered as animated graphics using SVG. For instance, if you want to show the path of an aircraft, its position at different times can be recorded in GML as separate features. Simple code on the client can be used to create an icon of a plane and display it as a moving object against a background map. The user might use a time slide bar to move the plane. GML also lets you define an object's position algorithmically. In your GML file, you could have a feature representing an oil spill, and include a dispersion algorithm as an attribute of the feature. Then you would just need a simple client-side utility to render the oil spill and show how it spreads over time.
9. *You don't have to target just a Web browser*. GML is a non-proprietary geographic file format that can encode most types of geographic information. As such, you can use it as a general geo-spatial data interchange format. In fact, geographic data in GML can be sent to any device with an XML interface. So, for instance, you could use GML to send geographic data from one GIS to another. GML can also be displayed on XML-enabled devices like the new-generation PDAs and cell phones. The benefit for the provider is that one format suits all uses.
10. *Enabling service chaining*. An example of service chaining is where you take some geo-spatial data, send it to a site to convert from the NAD27 to the NAD83 reference system, send it to another site to convert from geographic coordinates to UTM, send it to another site to add administration boundaries and demographic data, and finally pass it along to another site for display or storage. Each site is providing a discrete service. This concept is supported well by GML because (a) GML is a general format, so sites don't need to support lots of proprietary data formats, and (b) GML is extensible and XML-based, which makes it easy to manipulate, change, and add to its contents.

These ten points help explain how GML is different from DXF and all other data formats. GML exploits Internet technologies that have become available only in the last few years. [2]

4.4 Commercial GML Adoption

The Ordnance Survey, Britain's national mapping agency, noted a year ago on their website: "Following extensive consultation we can confirm that it is our intention to produce all DNF (Digital National Framework) data in GML format." They are now the first commercial data provider to offer data in the GML format on a national scale.

The Ordnance Survey web site FAQ (http://www.ordnancesurvey.gov.uk) includes this question:

"Why is MasterMap only available in GML format?

Answer: The Open GIS Consortium, a global organization of GIS users and developers, has created Geography Mark-up Language (GML) with the intention that GI users use the same format. Software will be written around GML format, and it is derived from XML, which is a web-friendly format...."

[2] The author wishes to acknowledge that this section was written with help provided by Martin Huber of Geotask, AG, Basel, Switzerland.

Other data provider organizations, including those listed below, are evaluating GML or preparing to offer their products in GML:

- The US Census Bureau is considering using it to encode and distribute the Bureau's TIGER files.
- The Netherlands Society for Earth Observation and Geoinformatics (KvAG) organized a "GML Relay" in June to explore the movement of GML data through several vendors' software implementations.
- GeoConnections Canada, a division within Natural Resources Canada, is in the process of integrating GML into the Canadian Spatial Data infrastructure.

The Ordnance Survey provides the following list of GML tools providers:

Galdos Systems, Inc. (Canada)
www.galdosinc.com

Ionic Software SA (Belgium)
www.ionicsoft.com

Intergraph Corporation (US)
www.intergraph.com/gis

Laser-Scan (UK)
www.laser-scan.com

PCI Geomatics (Canada)
www.pcigeomatics.com

Safe Software (Canada)
www.safe.com

Snowflake Software (UK)
www.snowflakesoft.co.uk

4.5 GML in the Proof of Concept

The Proof of Concept that is part of the GETIS exploitation and dissemination plan will help to demonstrate how the roll-out of GML will take place through incremental migration to GML-based information integration and access. In almost all cases, agencies undertaking the shift to GML will be concerned with the migration of diverse information systems and substantial quantities of data in multiple formats.

It is not in the scope of this study to undertake large-scale conversion of legacy data into GML. In the interest of demonstrating as much capability as possible within the resource limitations of the study, some of the demonstration data will be encoded in XML but will not use the full GML core schema. The proof of concept will document and demonstrate a practical, realistic set of steps that will provide early success and a solid foundation for further integration of web-based distributed capabilities layered on working legacy systems.

4.6 Raster Data

The OpenGIS Grid Coverages Implementation Specification has been available for more than a year. It specifies the details of interfaces that enable interoperability between systems that produce and use satellite images of the Earth, digital aerial photos, digital elevation data, scanned (rasterized) maps, and other kinds of data represented in a grid cell or "raster" image coordinate system that is tied (or not tied) to an Earth coordinate system. But it does not address all kinds of coverages and has not been implemented in as many commercial products as the specifications that address vector data and simple raster maps on the web. A more comprehensive coverages specification is in process in OGC, as is a specification for interfaces for image exploitation functions. In many cases, users need only to see or visually overlay simple maps, and in such cases the OpenGIS Web Map Service usually provides the necessary interoperability.

5 Conclusions

5.1 Recommendations

5.1.1 Transforming Data into Information

It would be a good thing if every community had some cheap and simple tools for publishing and viewing relevant information to defined user communities, using GML as a common open exchange format. With that in place, a next logical step would be to "GML-ize" sample data elements and begin a dialog among information sharing partners regarding the encoding of their data models in GML. It would be important to include user communities when defining attributes in models, so that base data models could be built that would effectively support variations, but with the goal of making these translatable. It is important that everyone share the goal of going beyond the mere migration of data-sets to the maintenance and evolution of data models based on applications, but codified in machine-readable GML data.

5.1.2 Feedback Mechanisms / Continuous User Validation

Opportunities exist for governments at all levels, and other kinds of organizations, to participate in the shaping of GML and other OGC standards. This opportunity is no longer focused on technical committee work, but on short term, fast paced "interoperability initiatives" such as testbeds and pilot projects. These initiatives provide hands-on experience, help develop and test specifications, help develop networks of people with related expertise, and provide a platform or mechanism for European requirements to be fed into standardization processes. GETIS follow up activities involving an exploitation team will explore these opportunities.

5.2 Benefits of GML

The main benefits to using GML are described below:

1. GML plays a role in **technical interoperability** between systems that have different internal formats. Sharing and exchange of geographic information historically depends on the source and destination datasets having a common data format, but now this is not an issue. One of the keys to GML's success is the OpenGIS Web Feature Service (WFS) Specification, a companion specification to the GML specification. To get GML data, a user issues a query to a web server with an OpenGIS Web Service Interface (a "Web Feature Server"). The standard interface enables access to the feature store and enables users to add, update or retrieve data, locally or across the Internet. The data can be retrieved as GML data, even if the data store uses a different format internally. Users no longer need to care whether the underlying store is from ESRI, SICAD, Oracle or IBM.
2. GML enables a new degree of **semantic interoperability**. GML gives information communities 1) new tools for data coordination, and 2) a way to avoid adopting a "one size fits all" data model that defines features and feature relationships in ways that don't suit everyone's application needs. It is still necessary to agree on the elements of the basic *framework* or *base* data. (All GML-encoded data models use the same OpenGIS geometry model, which eliminates the need to resolve a great many highly technical geometry issues.) XML makes it easy to write tools that support a web-based process of presenting, reviewing and maintaining feature definitions and overall data models. But the most important thing is that GML makes it practical to allow different "application profiles" of the base data. That is, software can be written that translates between slightly different feature definitions, enabling one information community to use another information community's data with relative ease and confidence.
3. GML is instrumental in integrating spatial services into the World Wide Web's portfolio of web services. Most major organizations are beginning to develop, or have already developed, web-based capabilities that are integrated into their legacy information systems. In disaster man-

agement, public networks and virtual private networks will provide people with information services that will utilize web services and XML. OGC Web Services and GML provide the standards platform for integration of commercial spatial data and processing capabilities into these information services. In one scenario, a thick client on a desktop system will take advantage of packets of XML and GML data delivered to be processed by powerful desktop software. In another scenario, a handheld device will receive an applet along with the XML and GML data, and that applet will handle portrayal and provide the user with an interface to powerful remote server software.

4. GML maintains separation between feature data and portrayal (presentation style). This means, for example, that a certain road type defined in a data model does NOT carry with it a "hard-wired" color or pixel width. Road data from a server might be portrayed by one application client as a three-pixel wide red line, and it might be portrayed very differently by another client, perhaps a client that converts text to speech to say, for example, "two-lane road."
5. GML became "adopted technology" of the OGC in March 2001. It is rapidly becoming the international industry standard for encoding geo-spatial information, and through OGC's procedural arrangement with ISO TC/211, it is on track to become an ISO standard. Standards provide market stability, and thus more user choice (and lower prices) through the proliferation of standards-based products. Proliferation occurs because vendors' risks and costs are reduced by the stable technology platform. The "commodity" platform enables them to focus on adding more application-specific value. Traditionally, a few major vendors have dominated the GI software industry. These vendors largely controlled the manner in which all geographic data was encoded. With GML, this is changing.

5.3 Identified Gaps

The GETIS study reveals that the realities of data sharing and geomatics software use are in a period of rapid transition. The "benefits of GML" listed above provide a good summary of the changes that are occurring. An important task for GETIS is to assess at this point in time what is possible and feasible for users now, and to assess the migration path toward capabilities that are likely to become available in the near term or later. Below are some of the areas that users should watch as they endeavour to pace their technology investments.

5.3.1 GML Uptake

As a first step, communities need to be able to do basic visualization of GML with standard browser functionality. This is available now in various online demonstrations, some implemented as part of operational programs. But more data needs to go online, with publicity where appropriate to help encourage active use. As commercial GML data providers come online, they will, of course provide significant momentum.

5.3.2 Adopted GML Definitions vs. Information Requirements

The following table lists common feature groups that the GETIS Working Group considers to be important in disaster management. A "yes" in the "Required" column indicates that the feature group is in the "Minimum Required" list:

Feature Group	Required
Transport Infrastructure	
Airports	**YES**
Roads	**YES**
Railway Lines	**YES**
Rivers / Canals	**YES**
Supply Infrastructure	
Gas	**YES**
Electricity	**YES**
Water	**YES**
Heating	NO
Telecom	**YES**
Buildings	
Public Shelters	**YES**
Residential	**YES**
Response Infrastructure	
Disaster Area	**YES**
Real-Time Location of Response Resources	**YES**
Public Shelters	**YES**
Hazardous Objects	**YES**
Flood Models	**YES**
Coverages	
Topography	**YES**
Hydrology	**YES**
Geology	NO
Land Use	NO
Aerial Photographs	**YES**
Satellite Images >5m Resolution	NO
Satellite Images <5m Resolution	NO
Soil Types	NO
DEM	**YES**

GML can be thought of as a flexible language for writing application schemas. All of these types of data can be represented in GML. It is advantageous in terms of data sharing, of course, if the schemas are standard schemas in the larger community of organizations and agencies that might use or produce such data.

5.3.3 A Need for Service Chaining

The proof of concept shows how a single web service chain is built. Clearly, the inputs to it and outputs from it can be improved, but the concept is very powerful. It’s important that the Proof of Concept (or any web-service for that matter) is not seen only as a singular element, but as a flexible link in a chain that can also be attached to other chains.

5.3.4 A Need for SAR Data and a Service to Derive "Flooded Area" features

Synthetic Aperture Radar (SAR) data can be valuable both in analysing terrain prior to flood events and in mapping flooded areas during flood events. The lack of interoperability arising from the availability of SAR data only in proprietary formats is a serious constraint. It would be useful to have a service available that extracts the “flooded area” feature from SAR raw data. If such a service is not available, this report recommends to earth observation and remote sensing service vendors that they develop such as service.

Such a capability would employ, when provided as an online service, the OpenGIS Web Coverage Service Specification in providing access to the raster data, and it would employ the OpenGIS Web Feature Service in providing access to the derived flooded area features.

5.3.5 Stimulating the Use of EO Data and Information

The Earth Watching Programme of the European Space Agency (see http://earth.esa.int/ew/) makes satellite products available to civil agencies in case of disasters. However, the addition of online access to the relevant data sets through interactive and interoperable web mapping and data services might help to promote a more widespread use of EO data and derived information in disaster incidents.

A good example for making EO data and derived products available on a national scale to promote education in its application and more widespread use can be found at the web site of Natural Resources Canada: http://geogratis.cgdi.gc.ca/frames.html