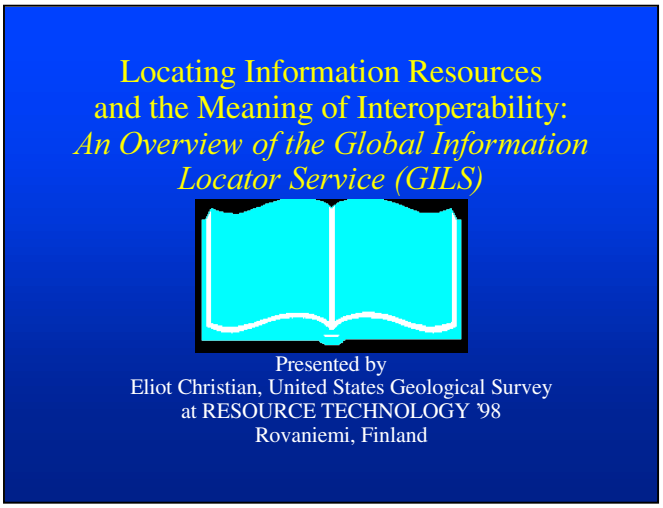
Locating Information Resources and the Meaning of Interoperability: An Overview of the Global Information Locator Service (GILS) – *Presentation*

Christian, E.

US Geological Survey, Reston, Virginia, USA, echristi@usgs.gov

Abstract

Interoperability can be understood in many distinct ways. This presentation focuses on the global and long-term problems faced by people trying to discover data and information resources relevant to the environment. Even with widespread access to public peer-computer networks, there remain fundamental problems in common semantics for characterizing information resources (sometimes called "metadata"). A surprising degree of consensus has now been reached in this area by adopting principles articulated in the Global Information Locator Service. These principles operate at levels of policy, standards, and technology, and provide immediate solutions as well as long-term strategies.



The U.S. Geological Survey is one of many participants in the international "Global Change Research Program". Global Change encompasses very large science, social, and political issues such as climate change and loss of biological diversity. In a sense, the Global Change Research Program is an attempt to understand how Earth systems "really work".

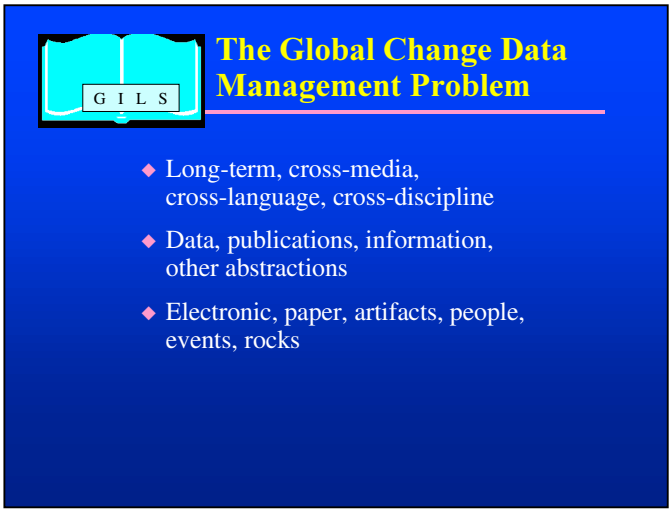
In 1991, a program plan on data management for Global Change was designed. This plan focused attention on how people find relevant data and information. The aim is to help make Earth science data and information more accessible to researchers and the public.

The Global Change Research program is not only science and technology. The program is also concerned with information policy issues. There is a parallel in public policy issues such as ensuring a free flow of information and public access to government information. These public policy issues in turn affect the social and political realities that help determine the long-term sustainability of Earth systems.

In dealing with whole Earth systems, understanding and action are intrinsically linked. Access to information is a critical part of that linkage.

In 1995, United States law established a Government Information Locator Service. This GILS adopted the international standards developed in the Global Change work.

Canada and others also adopted GILS. They added document-level search and multiple language support. This second version of GILS was a model for a Global Information Locator Service proposed through the G7 Global Information Society initiative. Aside from political scope, the Government Information Locator Service is identical to the Global Information Locator Service as endorsed in 1997.



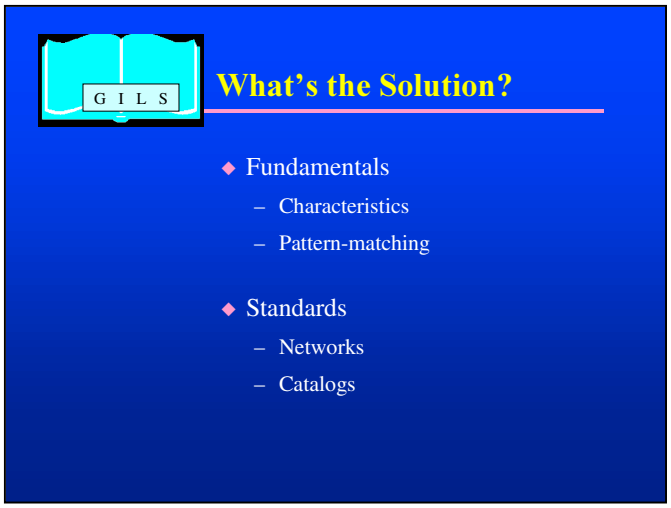
The Global Information Locator Service is shaped by public policy issues that are especially evident in the Global Change Data Management arena.

For instance, How should we define the data and information user community? A user is virtually anyone, anywhere who makes decisions affecting, or affected by, the environment. Users range from children to politicians to specialized researchers. They communicate in any language and in any discipline.

We know that Global Change research will be in a formative stage for decades. It may be 40 years before scientists even know the right questions to ask in some areas. This means most of our users have not even been born. Obviously, we cannot interview them to determine user requirements.

What data and information resources do we need? Relevant resources range from small data tables to massive, global observations. Resources also include directories of organizations and people, chronicles of events, and bibliographic references for publications, books, and maps. Other relevant resources are in natural history museums and archives of all kinds, from seed banks to butterfly collections to genetic libraries to the USGS rock library.

At this point, it may seem that finding data and information is already a huge problem and that this broadly scoped research program only makes it worse. This broad perspective is not only necessary, but also essential because it keeps us focused on fundamentals. In such an important global program, mere technology must not drive public policy.



In GILS, networks are acting as a vast, interconnected catalog that people use to find information. For example, a person wants to pick a face out of a crowd, find a library book, or search the Internet. Fundamentally, the person is searching for a pattern. The searcher looks for such patterns among the observable characteristics of objects.

There are two kinds of opportunities here to make it easier for people to find information. We can improve the pattern-matching process, and we can make it easier to observe object characteristics.

In the context of electronic networks, we are seeing rapid evolution in pattern-matching and other sorely needed components such as natural language processing. Yet, there is an immediate need for standards to exploit the huge and diverse collections of information already available. Essential to such standards is a common way to perform searches.

Bibliographic searches are very common – publications with a certain title, by a certain author, on a certain subject. GILS adopted the international standard protocol for information search. ISO 23950 (ANSI Z39.50). Major players worldwide already support this standard, including most information services and libraries.

The ISO 23950 standard grew out of network search in library catalogs, but it supports much more than bibliographic patterns. It is already used for searching Web addresses, telephone directories, gene sequence libraries, maps by latitude/longitude, imagery by content, and chemicals by structure. Such a growth path is critical for GILS.

Even the best processes for pattern matching need a way to pick out specific object characteristics. Bibliographic objects have well-known characteristics – author, title, subject, date published, and so on. Such characteristics have been developed over many decades and are widely used throughout publishing, libraries, and archives, and other settings. The GILS standard also adopted this widespread set of bibliographic semantics, extended to apply broadly to networked information discovery.

The center of the GILS strategy is to adopt relevant parts of existing international standards dealing with information search. These standards allow us to use networks as catalogs for all manner of data and information resources.

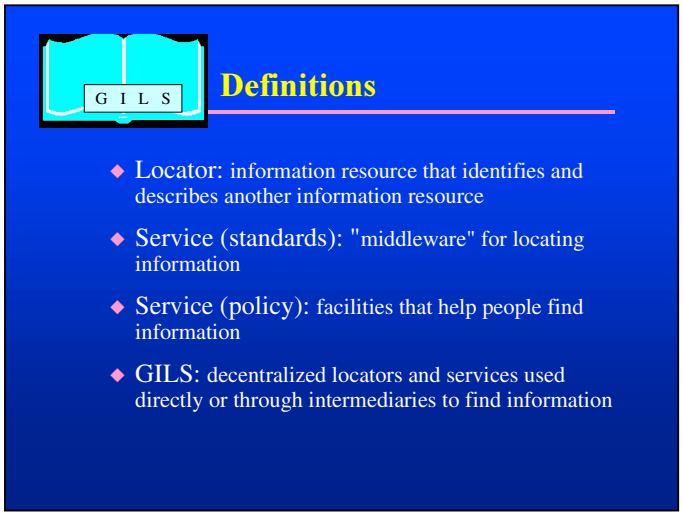
What is the strategy for deploying GILS broadly?

To assure that the free flow of ideas is sustainable over the long term, it is not sufficient just to have government mandates. Instead, we must find and promote solutions that respond not only to public sector requirements but to commercial and entertainment interests.

There is now good consensus that information standards intended for global use should presume not central authorities. Instead, everyone is trying to build decentralized and interoperable systems. It now seems that diversity is a defining characteristic of the emerging Global Information Society.

Current bibliographic cataloging practice is a good example of an effective decentralized approach. Cataloging standards provide interoperability across independently maintained libraries, while allowing wide latitude in how collections are developed and organized. Over the last decade, open standards in the library and information services communities have evolved to take advantage of public networks such as the Internet. In particular, the GILS Z39.50 standard is a good example of an open, international standard.

Much progress has been made, but we all recognize the bare fact of the currently primitive facilities for handling complex information, whether at the personal, corporate, community, national, or global scale. The focus of GILS on fundamentals is a conscious strategy to deal with the certainty of continuous technology evolution, and occasional drastic revolution.



A "locator" is defined as an information resource that identifies other information resources, describes the information available in those resources, and provides assistance in obtaining the information.

A locator can be modeled as a database of locator records, each of which is a set of related data elements. Locator records are also known as metadata, meta-information, directory entries or catalog entries. All of these descriptive mechanisms are "bibliographic" in a broad sense. Again, I would point out that GILS is designed to work with existing bibliographic systems.

In the standards context, GILS defines an application profile as part of a service definition. This "middleware" service performs specific functions useful for locating information. It is used by higher level applications such as library catalogs, and it uses lower level components such as networks.

Of course, public policy has a less technological understanding of the term "service" and this carries over into GILS at the policy level.

Bringing these together, we can define GILS as: "A decentralized collection of locators and associated information services used by the public either directly or through intermediaries to find information."

What information resources are typically represented through GILS?

GILS is a very useful way to simplify the locating of networked information resources. For example, Web documents can act as GILS locator records. Such documents can be searched via GILS using embedded metadata.

But, GILS differs from most Internet metadata schemes. GILS locator records are designed to act as pointers not just to Internet pages, but to ALL kinds of information – including people, organizations, events, books, artifacts, paper documents, and so on. Static HTM documents on the Web is just one part of the information addressed by GILS.

GILS itself is open to any level of aggregation. A GILS locator record may describe an information resource ranging from one pamphlet up to an entire multi-national program. North Carolina is using GILS locator records to describe individual fields within tables within databases throughout the state. The U.S. Government Printing Office created GILS locator records at the level of entire agencies. Each community can specify its own preferred level of aggregation in their own GILS guidelines.

Also, a collection of GILS locator records can be described by a record in another GILS locator. Using the Linkage element, the separate GILS locators then form a network for distributed search. This feature addresses the “query routing” problem of Internet searching. GILS is used this way in the public domain “Advanced Search Facility” now under development.

There are also searchable databases giving locator records for hundreds of information sources that are searchable with Z39.50 – many of which have millions of bibliographic or GILS records. Such locators can encompass all library catalogs, Web crawlers, and public directories on the Internet, yet also cover non-electronic resources like the USGS Rock Library.

Just how does GILS make information searching interoperable?

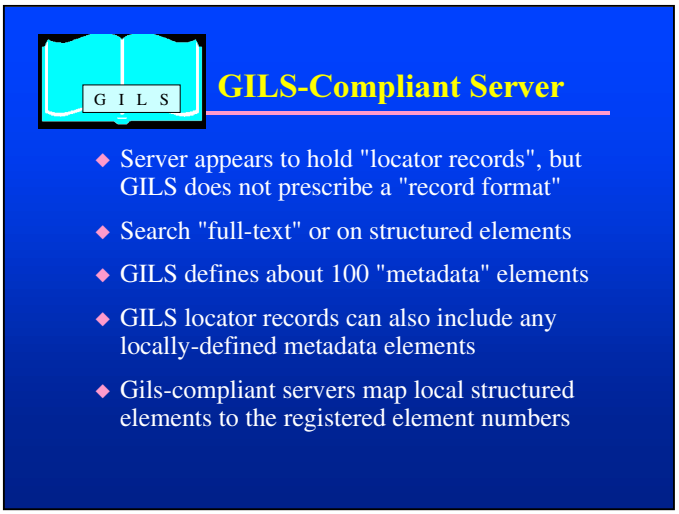
A key technology for GILS is a network-based architecture called "client-server". This is the same architecture used for the World Wide Web,and many other network applications as well.

Interoperability among the many different GILS-compliant servers is defined through the GILS Profile. The GILS Profile is part of the ISO 23950 standard and is coordinated through a GILS Special Interest Group within the Open Systems Environment Implementors Workshop. (The GILS discussion list is open to anyone to join – the address is on my last slide.)

The GILS Profile is defined on networks, including the Internet. But, instead of working at the presentation level like Web pages, GILS defines a machine-level network interface.

The GILS Profile only specifies the behaviors of server software at the client/server interface – it does not constrain clients at all. This means user interfaces can be anything you want, including self-operating client software like information gatherers.

The operation of a GILS-compliant server is only defined as an "abstraction layer". This means GILS doesn't care how servers actually manage content. You can have relational databases like Oracle, Informix, or Microsoft Access. You can have your own favorite search engine, or an SGML or GIS database, or a Web crawler. You could even operate a search broker and have content only developed as needed to respond to a specific search. Again, the Advanced Search Facility provides a good freeware base for developing new information discovery tools in the context of GILS.



GILS-compliant server appears to hold a set of locator records. Each locator record characterizes some information resource. The locator records may be handcrafted catalog records, or they may be created automatically using machine-aided indexing. GILS locator records describing Web sites use a listing of words drawn from the site contents.

A GILS search can be full-text, as is done with Internet-wide search services. Yet a GILS search can also exploit bibliographic elements such as Title, Author, Subject, Publication Date, Language – over 100 elements are registered in GILS with one-to-one mappings to well-known MARC tags.

GILS locator records can also have locally-defined elements – whatever and as many as you need. GILS-compliant servers must support the required search attributes and all registered elements, and must not degrade locally defined elements.

GILS is based on a common search interface rather than a common record format. This means that you can use GILS to search virtually any metadata structures – HTML, XML, SGML, X.500, SQL databases, Internet mail, IAFA, and so on. GILS-compliant servers simply map local semantics to the registered elements.

Also, multi-lingual searching is supported in GILS because the elements are referenced by number – to display a tag, the user interface simply looks up the number for the particular language in use.

Where are we now in terms of GILS implementations?

There is no central registry of GILS implementations, so the following listings are just a sampler of the GILS implementations.

"National GILS" are showing up in Australia, Canada, the European Union, Germany, Japan, and some other countries. Also Information System Russia is adopting GILS.

GILS is especially well-known for access to environmental information. GILS has been adopted by international organizations such as the Committee on Earth Observing Satellites, the Consortium for International Earth Science Information Network, the Great Lakes Information Network, and the Inter-American Biodiversity Information Network.

In the Global Environmental Information Locator Service, G7 countries and others have agreed to use GILS for environmental information worldwide. This includes the European Environment Agency "Catalogue of Data Sources", and United Nations initiatives such as the Framework Convention on Climate Change and the Convention on Biological Diversity.

Cuba has translated GILS materials into Spanish and it is being implemented throughout Latin America for access to health care information. This includes a Spanish gateway to "Medline" using Z39.50.

In addition, there are many other GILS implementations focused primarily on spatial, remote sensing, biological, and other environmental and Earth science information. These include national initiatives for geospatial (map) data in countries such as Australia, Canada, Japan, and Malaysia, as well as a genetic information initiative in Singapore.

There are quite a few GILS implementations already underway in the United States at the state and regional levels. Others at the local level are starting.

The Southern Growth Policy Board is using GILS in the context of economic development. Just as libraries act as a trusted source of diverse information, this regional GILS will provide a trusted source of diverse economic, demographic, and other commercially significant public information – enhancing the region's competitiveness in attracting business investment on a global basis.

Minnesota, North Carolina, Texas, and the state of Washington have law and executive orders mandating GILS. Several other states and regional organizations also have GILS initiatives, including California, Florida, Massachusetts, Missouri, New York, Ohio, and South Carolina.

We are now seeing the beginning of city-level GILS implementations in cities such as Fort Lauderdale, Greensboro, Madison, Philadelphia, and Tallahassee.

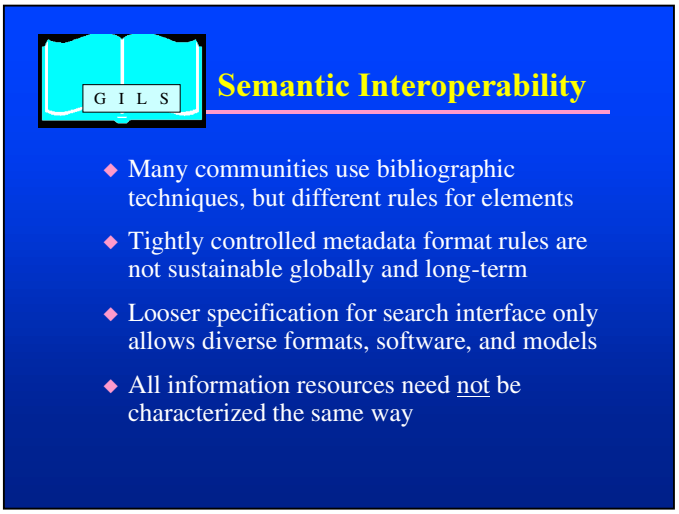
There are many additional states and cities involved in the National Spatial Data Infrastructure and National Biological Information Infrastructure as well, and all of these are also linked by common support of GILS.

As mentioned before, the U.S. Federal GILS is established in law. The rest of these United States initiatives are following GILS without a legal mandate. For example, both the Government Printing Office and the Library of Congress are making their search services GILS-compliant.

GILS is a basic component of the National Spatial Data Infrastructure, the National Biological Information Infrastructure, the National Environmental Data Index, and several other national initiatives.

Although there is not a complete list available, it is clear that there are already many and diverse GILS implementations.

In adopting GILS, these various organizations do not submit to a single view of information, nor do they subordinate themselves to some "mother of all GILS". Although their different Usage Guidelines reflect their disparate goals, they are all interoperable with other GILS as well as libraries, information services, and other major resources worldwide.



Why do we need mechanisms like abstraction layers and search interfaces?

Simply put – people over the centuries and throughout the world have built an incredibly rich and complex treasure house of information. To preserve that richness and diversity, we must step back and see the commonalities.

Many separate communities worldwide over many decades have used bibliographic techniques to characterize data and information resources. Unfortunately, when each community uses different tags for the bibliographic or metadata elements, this commonality becomes obscured.

The usual result of this independent development is that there is no functional interoperability between the catalog services, unless some organization is able to force the communities to accept imposition of a common format. Such strong-arm tactics may be useful at times, but they are clearly inappropriate on a long-term and global scale.

GILS takes a gentler approach – GILS leverages existing systems instead of trying to superceded them. It provides interoperability at the semantic level but only in this specific interface for a machine-level search service.

With its semantic interoperability approach, GILS allows each community to develop as much additional interoperability as they want. Often, this is merely coming to agreement on a Usage Guideline. Sometimes, a more specific profile is constructed on top of GILS – as in the Geospatial Profile used for maps and spatial data, or the Catalog Interoperability Profile for imagery.

Recently, the GILS community is looking at application of ISO 11179 as a way to enhance the realization of interoperability using semantic registries. The GILS elements have been mapped into the ISO Basic Semantic Registry and the EPA Environmental Data Registry, for example.

The vision of a Global Information Locator Service operates at several levels, including public policy, standards, and technology.

In all these aspects, we expect GILS to evolve over the decades, within some basic principles:

- Enhancing the free flow of information globally.
- Assuring access to all media and legacy sources
- Using open, international standards
- Balancing the needs of searchers, intermediaries, and content owners
- Avoiding preferred views or central control mechanisms

On a global, long-term scale, these are not simple platitudes. Just look back a few decades and consider how fragile is this notion of a free flow of information – and how fragile it is for most people yet today.