GILS:
What is it?
Where's it going?

presentation by

Eliot Christian,
United States Geological Survey

I am very pleased to be invited here to Australia and to help introduce GILS.

So, What is GILS?

The goal of GILS is to make it easier for people to locate information, in all its forms, in all languages, for all time.

This is perhaps a somewhat grand statement, but the free flow of information really <u>is</u> the touchstone of GILS. This leads into all sorts of interesting implications--ranging over technology policy, scientific research, socio-economics, geo-politics, and so on.

Whether you call it the "Government Information Locator Service", the "Generic Information Locator Service", or the "Global Information Locator Service", the overarching goal of GILS is help people locate information.

What is a Locator?

Locator: an information resource that identifies other information resources, describes the information available in those resources, and provides assistance in obtaining the information

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.

If you like, you can regard locator records as metadata, meta-information, directories, catalogs, or abstracts. All of these descriptive mechanisms are "bibliographic" in a broad sense, and GILS is carefully contrived to be interoperable with existing bibliographic systems.

We should make one distinction, though, between locator records and "metadata". Metadata is a very broad concept that includes all sorts of information that might be associated to data. A locator record, on the other hand, only deals with one kind of metadata.

To use a warehouse analogy, locator metadata is like a shipping list. It tells you what is in some container, and something about how it gets from the producer to the end user. If you were to open up such a container, you might find assembly instructions or a user's manual for the product inside. That, too, is a kind of metadata. But, it is NOT locator information in the sense that we use the term in GILS.

So, GILS focuses only on this notion of locator records that help people find information and does not have much to say about the deep metadata needed to help people actually make use of the referenced data and information.

Overview of GILS

- ◆ Focused on how resource characteristics are exposed to searching

- ◆ Uses existing open network and information standards (Z39.50 subset, TCP/IP, URI's, MIME, MARC), defined at servers only

- ◆ Server content not constrained, extra functions available for "locator records"--pointers to all kinds of information resources

- ◆ Search server content by full-text and/or well-known elements (including Dublin Core equivalents) if available

- ◆ About 150 elements are registered--any set of locator records may use any or none of these, and may use any locally-defined elements

- ◆ Servers support delivery in a structured syntax, unstructured text, and MARC if available

GILS focuses on how the characteristics of information resources are exposed to searching. GILS adopts existing open standards to achieve search interoperability.

In GILS, interoperability is based on common reference to existing semantics for the "elements" that characterize information resources. The semantics of these elements (or fields, or tags) have been developed over many decades in libraries worldwide.

GILS adopts the Z39.50 standard (ISO 10163) to specify how servers expect searches to be expressed, and how servers respond with search results. This standard is growing very rapidly. There are well over 400 information sources supporting public Z39.50 access--many of which have many millions of bibliographic or GILS records.

GILS is silent about clients, and so does does not have a user interface at all.

There is no constraint on what content is accessible from a GILS-compliant server, but GILS does provide functions for handling metadata--known in GILS as "locator records".

In essence, locator records are annotated pointers to any kind of information resources, whether online or not. You might think of GILS locator records as the "virtual card catalog" for a global library and archive, containing not only collections of document-like objects, but organizations of people and their activities and interactions.

GILS does not constrain how servers generate or manage content. Locator records may be hand crafted catalog records or the records may be generated dynamically.

GILS-compliant servers must handle full-text search and must recognize registered numbers for certain well-known elements (including equivalents to Dublin Core elements). Any specific set of locator records may use any or none of these. Any GILS-complaint server may also use any number of locally-defined elements.

GILS requires servers to deliver locator records in at least three forms: a structured syntax, unstructured text, and MARC if available. (There is a one-to-one mapping of available GILS elements to MARC tags.)

Information Resources

- **GILS locator records are for ALL kinds of information in any form possible, <u>not</u> just Internet information resources**

- **GILS can also handle networked information resources including, but not limited to, HTML and other objects, with or without Meta tags or other metadata encoding, including, but not limited to, Dublin Core metadata**

- **GILS allows any level of aggregation, but communities of interest may constrain the domain and cataloging rules by adopting more specific "Usage Guidelines"**

- **Using the Linkage elements, separate GILS locators comprise a network for distributed search**

I want to be clear that GILS should not be viewed as an alternative variety of "META" tagging for Internet information resources. GILS does not prescribe tagging.

In fact, GILS is silent about server record formats. Servers may hold SGML or HTML documents, with or without Dublin Core tags, or Microsoft Word documents, or MARC records, or closed caption text streams--literally anything that can be made searchable might serve as a locator record. GILS only defines the behavior of the server at its interface to a searching client.

Like bibliographic records generally, GILS locator records are understood to refer not to themselves but to information elsewhere.

GILS locator records must have a representation on electronic networks; but the referenced information resources need not. I have a GILS locator record for the USGS Rock Library--a place with drawers full of completely non-electronic rocks. For this resource, the pointer is a physical address. I also have a GILS locator record for the USGS home page. For that resource, the pointer is a URL.

GILS has no prescription for what is an appropriate level of aggregation. However, GILS implementations for communities of interest usually constrain their specific collections by adopting a "Usage Guideline".

GILS locator records already describe information resources ranging from individual pamphlets up to to multi-national programs. North Carolina is using GILS locator records to describe individual fields within databases throughout the state. The U.S. Government Printing Office has created GILS locator records at the level of entire agencies. The Usage Guideline for Canada's GILS is different than the one for the US Federal GILS, and both will be different than the ones for Japan GILS, Singapore GILS, Australia GILS, and so on.

Servers acting as GILS locators are themselves information resources and so can be described by a GILS locator record in another GILS locator. Using the Linkage element, the separate GILS locators form a network for distributed search.

This recursive feature makes GILS useful in addressing the "query routing" problem of Internet searching. This distributed search is being exploited at the U.S. Government Printing Office, and is also a central feature of the U.S. Advanced Search Facility procurement and a similar initiative underway here in Australia.

GILS Law and Policy Initiatives

- **United States**
  - Federal law (44 USC 3511), National Spatial Data Infrastructure, National Biological Information Infrastructure, Geospatial Information Infrastructure, Global Change Data and Information System, National Environmental Data Index
  - North Carolina, Florida, Massachusetts, Missouri, New York, South Carolina, Texas, Washington, Southern Growth Policy Board
- **Australia, Canada, Denmark, Japan, Singapore**
- **International: G7 Environment and Natural Resources Management, Clearinghouse Mechanism for the Convention on Biological Diversity, European Environment Agency "Catalogue of Data Sources"**

GILS is required by law at the U.S. Federal level, which gives us perhaps10 or 20 years to get it right. There are many U.S. initiatives using or building on GILS, including the the National Biological Information Infrastructure and the National Spatial Data Infrastructure that focuses on helping people to find and use data referenced to places on the Earth. The U.S. Library of Congress has committed to make its products and search engines GILS-compliant. Also, the 1,400 Federal Depository libraries in the U.S. are required to provide GILS-aware client software for public access this year.

Several states and regional organizations are pursuing GILS for their specific purposes. There are even more state and city participants in the National Spatial Data Infrastructure. The Southern Growth Policy Board has a regional GILS initiative underway to make the region more competitive in global markets.

On the international front, there is a recommendation to adopt GILS and compatible profiles for locating environmental information globally, as part of the G7 Global Information Society project on Environment and Natural Resources Management. This infrastructure also links other efforts such as the "Clearinghouse Mechanism for the Convention on Biological Diversity" and the European Environment Agency "Catalogue of Data Sources" and some UN initiaitves.

Diversity of applications is a GILS design feature. Different social, political, economic, and/or thematic organizations are free to independently organize relevant parts of information space for their own particular needs. They do not need to subscribe to a single view of information, nor do they need to subordinate their information locators to some "mother of all GILS".

The various GILS collections organized for different purposes have different policies, and these are typically reflected in their specific Usage Guidelines. Although distinct, each shares in common with all other GILS that their locators are interoperable for searching--because they all reference the same set of well-known semantics.

Toward Semantic Interoperability

- ◆ Many communities characterize information using bibliographic techniques

- ◆ Metadata elements may have the same meaning, but we can't relate them because different communities use different nomenclatures

- ◆ Interoperability is required for cross-community applications

- ◆ Tightly controlled approaches: require single software implementations, formats, or "common denominator" elements

- ◆ Loose federation approach: semantic interoperability allows diverse formats, software, data models, and fully distributed data management

What does it mean to say that GILS interoperability is based on adopting common semantics?

Many separate communities worldwide over many decades have used what are essentially bibliographic techniques to characterize data and information resources. (For example, the current Dublin Core initiative is focused on HTML documents on the World Wide Web.) Unfortunately, as each community independently chooses different tags for their bibliographic or metadata elements, any commonality that may have existed can become completely obscured.

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

GILS takes a gentler approach. GILS encourages interoperability at the semantic level but stops short of constraining how information is represented, managed, or presented. Each community can develop as much additional interoperability as they want, up to and beyond the full scope of MARC itself. For most communities, the available registered elements together with their own locally-defined elements have been more than sufficient. Such communities specify GILS-compliance for their servers and then may work out their own Usage Guidelines. For other communities, a more specific profile can be constructed on top of GILS, as has been done with the Geospatial Profile.

Let me give a typical example of "semantic interoperability: in practice. The European Environment Agency's "Catalogue of Data Sources" is organized by a different community than the library community's "OPAC (Online Public Access Catalog) Network- Europe". But, for most combinations of resource, cataloger, and searcher, the semantics of "system name" in the Catalogue of Data Sources is equivalent to "title" in the OPAC, "principal investigator" is equivalent to "author", and "keyword" is equivalent to "subject term". So, there is a ripe opportunity for interoperability because both systems can simply map to the well-known attribute number for those elements. They don't have to change the way they represent or manage their data--they just add-on a separate function that supports interoperability for searching. And, they are not only interoperable with each other--they are interoperable with all other GILS-complaint servers as well as all of the libraries, museums, archives, and other institutions worldwide that have used MARC for decades.

GILS Usage Guideline
selecting metadata elements

(1) GILS Core Elements that are always searchable:
Title, Subject Terms Controlled, Subject Terms Uncontrolled,
Originator (Creator), Distributor Name (Publisher),
Record Source, Date of Last Modification

(2) Other registered GILS (and Dublin Core) elements:
Contributors, Abstract (Description), Source, Spatial Domain,
Time Period, Date of Publication, Medium (Type),
Language of Resource, Linkage (Identifier), Cross Reference,
Access Constraints, Use Constraints

(3) Other registered Z39.50 elements:
GILS (includes Bib-1), Geospatial, etc.

(4) Other MARC or AACR2 defined elements

(5) Related community-defined elements

(6) Locally-defined elements

Server and Client Software

- ◆ **Freeware Servers: CNIDR, IndexData, Blue Angel Technologies**

- ◆ **Commercial Servers: Fulcrum Technologies, OCLC, Sovereign-Hill, SIRSI, Verity**

- ◆ **Gateways available for World Wide Web and X.500**

- ◆ **Clients: library systems with Z39.50 access, Znavigator, BookWhere 2000, SIRSI Vizion**

- ◆ **Freeware Java client being enhanced for spatial search**

By applying an existing standard in wide use for many years, GILS takes advantage of existing networks and software to access a vast array of valuable resources, including hundreds of libraries, as well as museums and archives worldwide. Together with spatial data catalogs, these professionally maintained resources provide free access to information resources collectively valued in the tens of billions of dollars, with even more information available on a fee basis from commercial information services.

GILS-compliant server freeware is available from the Clearinghouse for Networked Information and Discovery in North Carolina and from IndexData in Denmark. Blue Angel Technologies is offering a new freeware GILS-compliant server for Windows NT and a companion Meta Data Manager for handling metadata in ODBC databases. (Other facilities for GILS metadata collection are also available.)

The Online Computer Library Center SiteSearch server is being made GILS-compliant in support of the Solinet Public Information Project. GILS-compliance is also being developed by Sovereign-Hill, SIRSI, and Verity. The old WAISserver commercial software was acquired by Fulcrum Technologies, which will be dsitributing the GILS-compliant release (2.2). Fulcum is also making Surfboard GILS-compliant.

Gateway freeware and toolkits fort accessing GILS-compliant servers are available for World Wide Web, either through CGI scripts or integrated into the Apache Web server. There is a gateway between GILS and X.500 developed by DSTC here in Australia, and we're anticipating a gateway to WhoIs++.

Any client software capable of access to a server compliant with Z39.50 version 2 or 3 can access GILS-compliant servers. Extra capability is provided by GILS-aware clients such as the freeware Znavigator, the commercial BookWhere 2000 product from Sea Change Corporation, and products under development such as SIRSI Vizion and the new freeware Java client.

The GILS application profile specifies, though it does not mandate, spatial searching by a "bounding box" of latitude, longitude pairs. This extension of search beyond text is an important feature of Z39.50. The protocol has already been applied to imagery, to searching for chemicals by bond angles, and to searching gene sequences--other pattern-matching is certainly feasible.

Design Principles

- Policy and technology choices must support the diversity of sources, and points of view, in our Global Information Society
- Give no preference to any particular way of organizing information
- Support the role of intermediaries as well as content owners
- Use open, international, voluntary standards
- Be sensitive to the world's many languages and technical standards, international and national legal and financial issues
- Employ public, peer computer networks, but point to information in all media and forms
- Extend and evolve with technical advances in understanding how people handle information
- Design for the future, but preserve access to libraries, museums, and archives worldwide
- Expect evolutionary and sometimes revolutionary changes, but do not compromise basic principles

As a long-term infrastructure initiative, it is clear that GILS will be evolving in many ways. Yet, there are some basic principles from which GILS should not diverge over the decades:

Policy and technology choices must support the diversity of sources, and points of view, in our Global Information Society. There should remain no preference toward any particular hierarchy or other way of organizing information, but many organizing structures should co-exist.

Intermediaries will have a crucial role--it is not feasible nor desirable to establish master repositories to serve all the world's information needs. Content owners and intermediaries should be able to draw from other locators and also make their value-added products known through the same mechanism.

GILS must continue to use open standards, and be fully coordinated through the international voluntary standards processes. GILS must continue to be sensitive to the world's many languages and technical standards, and it must accommodate international and national issues such as copyright, security, and payments.

GILS should employ networks, but point to information in all media and forms. Because information changes meaning according to context, GILS must be extensible into the many ways people extract information from data.

GILS must look to the future, but preserve access to accumulated knowledge represented in libraries, museums, and archives worldwide. Adopted standards must accommodate the variable pace of technology worldwide. We cannot today predict a single technical basis for GILS over the decades. Evolutionary and sometimes revolutionary changes must be expected and accommodated, but not at the expense of these basic principles.

References, Contacts, and Further Information

- Australia's Information Management Steering Committee (IMSC) report <**http://www.nla.au/imsc/**>
- December, 1996 article in D-Lib magazine <**http://www.dlib.org/december96-christian**>
- Government Information Locator Service (GILS) <**http://www.usgs.gov/gils/**>
- GILS Contacts <**http://www.usgs.gov/gils/contacts.html**>
- GILS Core Elements <**http://www.usgs.gov/gils/elements.html**>
- G7 Environment and Natural Resources Management (ENRM) <**http://www.g7.fed.us/enrm**>
- G7 ENRM Experts Group <**http://www.g7.fed.us/enrm/experts.html**>
- Press Release "World's Experts Plan Global Locator Service for Environmental Information" <**http://www.g7.fed.us/enrm/press.txt**>

Australia's Information Management Steering Committee (IMSC) report
<**http://www.nla.au/imsc/**>

December, 1996 article in D-Lib magazine
<**http://www.dlib.org/december96-christian**>

Government Information Locator Service (GILS) <**http://www.usgs.gov/gils/**>

GILS Contacts <**http://www.usgs.gov/gils/contacts.html**>

GILS Core Elements <**http://www.usgs.gov/gils/elements.html**>

G7 Environment and Natural Resources Management (ENRM) <**http://www.g7.fed.us/enrm**>

G7 ENRM Experts Group <**http://www.g7.fed.us/enrm/experts.html**>

Press Release "World's Experts Plan Global Locator Service for Environmental Information"
<**http://www.g7.fed.us/enrm/press.txt**>