Open Geospatial Consortium Inc.

Date: 2010-07-15

Reference number of this document: 10-124r1

OGC Name of this document: http://www.opengis.net/doc/WhitePaper/Identifiers/1.0

Category: OpenGIS® White Paper

Editor: Simon Cox

OGC Identifiers – the case for http URIs

Warning

This document is an OGC Whitepaper. It is subject to change without notice. This document is not an official position of the OGC membership on this particular topic.

Document type:	OpenGIS® Whitepaper
Document subtype:	n/a
Document stage:	Approved for Public Release
Document language:	English

Contents

Page

i. Preface

This document discusses implications of recent changes to the OGC Identifier policy, specifically as related to the use of http URIs for persistent OGC resources.

ii. Document terms and definitions

No special terms.

iii. Contributors

All questions regarding this whitepaper should be directed to the editor or contributors:

CONTACT	COMPANY
Simon Cox	European Commission – Joint Research Centre
Sven Schade	European Commission – Joint Research Centre
Luis Bermudez	Southeastern Universities Research Association
Ingo Simonis	IGSI

iv. Intellectual Property statement

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. The Open Geospatial Consortium Inc. shall not be held responsible for identifying any or all such patent rights.

Recipients of this document are requested to submit, with their comments, notification of any relevant patent claims or other intellectual property rights of which they may be aware that might be infringed by any implementation of the standard set forth in this document, and to provide supporting documentation.

OGC Identifiers – the case for http URIs

1 Introduction and context

The OGC provides a large number of resources to support the construction of spatial data infrastructures, including documents, specifications, schemas and concept definitions. When deployed, the infrastructures require persistent reference to these resources, enabled by persistent identifiers. This may be at various level of granularity.

Uniform Resource Names (URNs) have been used by OGC for the identification of resources governed by OGC, and also for some resources governed externally for which Uniform Resource Identifiers (URIs) were not available (especially EPSG definitions). This was formalized by RFC 5165 which establishes the OGC NID 'urn:ogc:' [URN-OGC], along with several OGC Best Practice and Policy documents that define the 'ResourceSpecificString' that composes the remainder of a name [see http://www.opengeospatial.org/ogc/policies/directives] .

Revisions to the OGC naming policy allow for OGC names structured as http URIs, as an alternative to URNs. The internal structure of the two forms is identical, with the full name differing only in the prefix and field separators, so the information content and governance is unchanged. According to the current policy either URNs or http URIs may be used in specifications [OGC-NA].

However, the use of http URIs (a) resolves some deployment challenges and (b) provides an opportunity for easier engagement with broader communities. So OGC should now consider taking the next step, and mandate the use of http URIs for persistent identifiers in OGC standards.

This white paper discusses a number of issues related to this proposal.

2 URIs, URNs and URLs

The Uniform Resource Identifier [URI] uses short strings to refer to resources in the web. Example of these resources can be documents, electronic mailboxes, images, downloadable files, concepts (e.g. bay, buoy, specification) and real objects (e.g. Monterey Bay, Mooring 1, OGC document). Various URI schemes have been devised, including Uniform Resource Locators (URLs) and Uniform Resource Names (URNs) [URN].

URLs use domain names to locate resources in the web. There is a well-defined governance structure. The Domain Name System (DNS) runs under the auspices of Internet Corporation for Assigned Names and Numbers (ICANN) to provide a service for mapping from names to machine addresses.

URNs are also URIs but do not necessarily imply web availability of the resource being identified. The Internet Assigned Numbers Authorities (IANA) maintains a registry of URN schemes, with each scheme described in an RFC document. The latter may provide details of a resolver service.

At the time that OGC started using URNs, it appeared to be a good option for persistent identification, avoiding some undesirable effects and expectations around URLs. In particular, URN NID registrations (governed by IETF through IANA) are forever, while http domain registration is periodic, and http server maintenance is a skilled job. There was also a principle at stake: identification and location are different functions. However, deployment and maintenance of a resolver service for URNs has been a challenge for OGC, particularly as there is no standard for a URN resolution service, and no best practice or precedent to fall back on.

Meanwhile, however, the use and interpretation of URIs on the web has evolved significantly. The contemporary version of the URL story is [W3C/IETF, URI-URN]

- a 'http URI' (often, though incorrectly, abbreviated to just URI) may serve as a persistent identifier
    - used in this sense, the URI is interpreted trivially as a text-string: the presence of '/' and ':' characters is unimportant in that context
- a http URI has the highly desirable characteristic that it is also a URL and thus provides the opportunity for automatic resolution, following standard internet conventions
    - use of the DNS system and http protocol mean that no special resolver service is required

The latter overcomes a legitimate criticism of all non-http URI schemes.

For OGC this creates a significant opportunity. Generic tools (browsers, reasoners) have built-in support for binding to http URIs, without any customization, so the adoption of http URIs would allow documents referring to resources with OGC identifiers to be used more effectively in the generic web context. There is no requirement for the consumer to use a specific resolver service, so no requirement that everyone who encounters an OGC document to know where the OGC resolver is, what the invocation method is, or to cache definitions corresponding to OGC names in advance of encountering them.

3 Resolution of OGC URIs

Http URIs provide the *opportunity* for resolution using standard tools, but also create an *expectation* of immediate resolvability. To meet user expectations, a service should be deployed at the http address that handles requests for resources and responds with useful representations. In order to preserve coherence both conceptually and practically this requires attention to a number of issues.

3.1 Domain management and long term persistence

For an identifier to be persistent it requires the governing body to arrange for the identifier to be available for the long term. This will give confidence to application developers to use those strings as identifiers, and will provide the necessary persistence to knowledge bases (e.g. semantic web graphs) that use those concepts.

Use of http URIs as persistent identifiers requires a commitment on the part of the organization issuing the identifier to maintain the http domain registration, and a strategy for managing the domain and the web servers. Organizational changes may occur on a shorter timescale than the than that over which the identifiers are intended to be useful, which can create instability in domain names change (e.g. opengis.org changed to opengeospatial.org). Persistent identifiers may use domain names that are not tied directly to the organization name to protect against this (e.g. geosciml.org vs. auscope.org). This must be planned for when the scheme is established [URIlife].

3.2 OGC identifiers for resources served by other providers

With the consent of some external authorities, the OGC 'def' name scheme provides identifiers for certain resources where that authority is not affiliated directly with OGC, in particular EPSG (for coordinate reference systems) and UCUM (for units-of-measure) [OGC-def]. Requests to OGC for representations of these resources may be handled by redirection to a service provided by the governing authority. For example, the OGC name

> http://www.opengis.net/def/crs/EPSG/0/4326

redirects to

> http://www.epsg-registry.org/indicio/query?request=GetRepositoryItem&id=urn:ogc:def:crs:EPSG::4326

The http response code should be 302 (Found) or 307 (Temporary Redirect) [HTTP]. The redirect is 'temporary' since while the OGC Name is persistent, the URL that the request is redirected to may change over time.

This illustrates an important principle: the OGC identifier should be persistent, and useful in the long term. The redirect leads to only the current service that provides representations of the resource. This requires that OGC maintains the current state of mappings from the http URIs to the URLs.

3.3 Representation of resources

Under REST principles there may be alternative representations of the same resource [Webarch; LD]. The http protocol supports content negotiation between the client and server to allow delivery of a representation that the client is capable of processing (in particular using the *Accept* and *Accept-Language* headers) [HTTP].

Within the 'def' branch of the OGC scheme, OGC provides alternative representations for the key resources for which there may be alternative client applications. For example, a http GET request for

http://www.opengis.net/def/cs/OGC/1.0/GridCS2d

results in different responses depending on whether the http Accept value is *text/html*, *application/gml+xml* or *application/rdf+xml*. The variant representations are also resources in their own right, addressed directly as

http://www.opengis.net/def/cs/OGC/1.0/GridCS2d.html

http://www.opengis.net/def/cs/OGC/1.0/GridCS2d.gml

http://www.opengis.net/def/cs/OGC/1.0/GridCS2d.rdf

If the client expresses no preference, the web-page (HTML) representation is delivered, since this representation is supported by the most common web-client – the browser. Client applications that have more specific requirements (e.g. GML, RDF) are expected to also know how to negotiate for these.

3.4 Information vs. non-information resources

In the 'semantic web' a distinction is made between 'things on the web' and 'things in the world', also referred to as information resources and non-information resources. Representations of information resources can be transmitted electronically, as web pages, GML encoded feature representations, RDF documents, PDF documents, etc. Non-information resources such as human beings, real world objects, or abstract concepts can be named by an URI, but a representation on the wire can only be of a *description* of the resource. Browsers can display representations of information resources directly, or *descriptions* of non-information resources.

Http response codes should be used to indicate if a representation of an information or non-information resource was returned when dereferencing a URI. If a client dereferences an information resource, http code 200 ('OK') gets returned. In case of a non-information resources, http code 303 ('See other') gets returned [RDFpub; SemURIs]

A key topic area for which OGC provides standard identifiers is 'definitions' of key concepts used in geographic information services. While there is some contention in the semantics community about whether concepts are information resources or non-information resources, OGC identifies *definitions* rather than concepts. Each of these an information resource, corresponding to 'the definition of the concept according to OGC', so there is no ambiguity in this context[1].

Nevertheless, when deploying feature services, the GI community may want to consider whether they need to make this distinction[2].

---

[1] A corollary of this is that OGC does not provide URIs for the (abstract) concepts themselves.

[2] OGC AS Topic 5 Features (also known as ISO 19101) is ambiguous on the issue, defining a feature as "A feature is an abstraction of a real world phenomenon". Of course a particular feature-service provides a representation of a real-world feature *as it is modeled for a particular application-domain*. The same real-world feature may be modeled differently for different applications, and none of these can claim to be comprehensive. It could be argued that the representation delivered by a specific service is strictly an (incomplete) feature description, and thus is by definition an *information resource*.

4 Governance of OGC URIs

Every web site or service has a URL policy, typically embodied in their 'sitemap', though it is rarely documented explicitly. However, when the purpose of a 'http URI' is to serve as a persistent identifier, such a policy is important. There have been many discussions of URI design [e.g. CoolURIs; CHIPS; SemURIs; IVOAid; WebName; URI-URN; UKURI].

Three audiences for the URI policy or scheme must be considered:

1. the party that wishes to mint an identifier as part of an orderly publication process;
    - the URI policy will usually embody aspects of a publication process which is a core part of the resource governance process
2. the party that wishes to use identifiers minted by the first group, within expressions of their data
    - for this group the stability of the URI scheme is most important. If the scheme uses a predictable pattern for new identifiers, then this is bonus.
3. the party that consumes identifiers, usually embedded in a dataset or other document as links;
    - it is common for URIs to be parsed by the user, who will construct variants with a reasonable prospect of finding something useful.

The OGC-NA [http://www.opengeospatial.org/ogcna] oversees the assignment of OGC Names to resources, primarily OGC resources including documents, specifications, and definitions. OGC Names are composed of a sequence of fields, with specific governance arrangements for the values of each field. This means that a complete OGC Name can only be assigned when certain pre-conditions are met. Functionally this institutes and enforces a delegation policy. Since the name production rules have been designed for use in either http URIs or URNs, OGC is in a strong position to transition to http URIs without compromising governance.

Note that persistent identifiers in the OGC URI scheme are based on 'slash namespaces'. The register of OGC identifiers does not include 'hash namespaces', which lead to URIs with a fragment after a # [RDFpub]. The latter allow you to address a secondary resource in the context of the resource identified by the URI before the #, but since the fragment is not normally communicated by an http client as part of a request [URI clause 3.5; Webarch clause 3.2.1] the http server is not required to distinguish a secondary resource using the hash URI syntax[3].

The other key aspect of URI governance is the stability of the http server. This was discussed above in section 3.1.

[3] Since a resource may have more than one identifier, there may always be an explicit non-hash URI to provide primary identification for a resource which is also a secondary resource in a particular context.

5 URI exploitation use-cases

There are two main use-cases using URI references. Both are supported by use of http URIs either as well as or better than URNs.

1. "Is the value P in a document equal to a value R?" – two modes are anticipated:
    - The simplest operation is string comparison on P and R; this is a lexical operation, so it can be done just as easily on a http URI as on a URN. This is expected to account for the largest usage of references in a mature SDI.
    - However, a negative result for string comparison does not necessarily mean that the underlying resources do not match. A single resource may have more than one identifier, or there may be different representations that are logically equivalent. In order to determine equivalence it is necessary to look deeper. The presence of the Semantic Web statements owl:sameAs or skos:exactMatch associated with the RDF representation of one of the resources may be enough. More complex logic or reasoning might be supported by inspection of representations of resource P and resource R. Any of these scenarios require that either or both URIs be resolved, which is easier if they are http URIs than URNs
2. "Tell me more about P" – there are two common applications
    - Within a dataset expressed in GML, xlink:href attributes[4] carry links to resources that may be logically embedded in the host document. In order to execute the implied transclusion and also maintain XML document validity a suitable representation of the resource must be obtained. Http content-negotiation provides a standard mechanism to support this.
    - More sophisticated processing of linked resources will typically be handled by a helper application. Dispatching the appropriate helper requires that the format of the resource representation be known. The http protocol supports this through the MIME-type reported in the response header.

[4] The semantics of GML's use of xlink:href is essentially the same as that of RDF's rdf:resource – a mechanism for using references in place of nesting. This is no coincidence: RDF provided the model for the original GML meta-model developed around 2000. It was only later re-interpreted in terms of the UML conceptual schema language favoured by the geographic information community organized through ISO/TC 211.

6 Resourcing implications

Http domains are easy to register, but temporary. Stability of a http domain and the hosted service is the responsibility of its owner. Superficially this appears to impose a greater and more specific obligation on the owner of a http domain, which makes http URIs less suitable for persistent identification compared with a URN NID which is perpetual.

Looking a bit deeper, however, while the 'resolver' for a URN is formally the document (RFC) that specifies the URN namespace, current practice for registration includes a requirement to identify a resolution mechanism, preferably a resolver service at an http address. For example, the OGC URN NID [URN-OGC] specifies that an operational resolver shall be provided at http://urn.opengis.net/. So in practice even with URN identifiers there is a requirement on OGC to maintain a resolver at a specific web location. Hence the resource implications for persistent http server registration and management are not significantly different between the URN or http URI route.

7 Summary and recommendations

Given that

- In order to support SDI deployment, resources provided by OGC for ongoing use should have well-governed, maintained, persistent identifiers
- A http URI or a URN is a text string, and may be used for persistent identification
- A http URI, unlike a URN, is implicitly bound to a resolver service known to all web-clients. Documents that include references expressed as http URIs can be consumed by generic web applications.
- Use of http URIs to identify OGC resources is likely to increase the impact of OGC in the mass market

1. **OGC should immediately mandate the use of http URI for identification of persistent OGC resources, in preference to URNs. (Approved June 17, 2010)**

Note that this recommendation applies to identifiers for resources that are governed by OGC, which are intended to be persistent. There is no obligation on any other organization, including those using spatial data infrastructures based on OGC technology. The use http URIs for transient resources may be inappropriate.

To support this:

2. **OGC should carefully manage (maintain for the long term) the http://www.opengis.net domain and identifiers in this domain**
3. **OGC should ensure that suitable representations of resources with identifiers in the domain http://www.opengis.net are provided through the http protocol. This includes redirection and content negotiation, where appropriate.**

8 References

8.1 URI and http standards

HTTP *Hypertext Transfer Protocol -- http/1.1* http://tools.ietf.org/html/rfc2616

URI *Uniform Resource Identifier (URI): Generic Syntax* http://tools.ietf.org/html/rfc3986

URN *URN Syntax* http://tools.ietf.org/html/rfc2141

W3C/IETF *Report from the Joint W3C/IETF URI Planning Interest Group: Uniform Resource Identifiers (URIs), URLs, and Uniform Resource Names (URNs): Clarifications and Recommendations* http://tools.ietf.org/html/rfc3305

8.2 IANA registries:

MIME Registry *MIME Media Types* http://www.iana.org/assignments/media-types/

URI Registry *Uniform Resource Identifer (URI) Schemes* http://www.iana.org/assignments/uri-schemes

URN Registry *Uniform Resource Names (URN) Namespaces* http://www.iana.org/assignments/urn-namespaces

8.3 OGC policies

URN-OGC *A Uniform Resource Name (URN) Namespace for the Open Geospatial Consortium (OGC)* http://tools.ietf.org/html/rfc5165

OGC-NA *OGC Naming Authority (OGC-NA) Policies & Procedures* http://portal.opengeospatial.org/files/?artifact_id=37800

OGC-def *OGC-NA Name type specification - definitions: Part 1 - basic name* http://portal.opengeospatial.org/files/?artifact_id=37801

8.4 Best practices in identifier design and management

CoolURIs *Cool URIs don't change* http://www.w3.org/Provider/Style/URI

CHIPS *Common http Implementation Problems* http://www.w3.org/TR/chips

IVOAid *International virtual observatory alliance identifiers* http://www.ivoa.net/Documents/latest/IDs.html

LD *Web Architecture (from How to Publish Linked Data on the Web* http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/#Terminology

RDFpub *Best Practice Recipes for Publishing RDF Vocabularies* http://www.w3.org/TR/swbp-vocab-pub

SemURIs *Cool URIs for the Semantic Web* http://www.w3.org/TR/cooluris

UKURI *Designing URI Sets for the UK Public Sector* http://www.cabinetoffice.gov.uk/media/308995/public_sector_uri.pdf

URIlife *The URI Lifecycle in Semantic Web Architecture* http://dbooth.org/2009/lifecycle

URI-URN *URNs, Namespaces and Registries* http://www.w3.org/2001/tag/doc/URNsAndRegistries-50

Webarch *Architecture of the World Wide Web, Volume One* http://www.w3.org/TR/webarch/

WebName *Dirk and Nadia design a naming scheme or Web naming schemes good practices* http://www.w3.org/2001/tag/doc/justSayhttp

Revision History

Date	Internal version	Editor	Sections modified	Description
18 May 2010	1	Simon Cox	N/A	Initial Draft Document.
19 May	2	Simon Cox, Sven Schade	All	Reorganized; nits; put into OGC document template
20 May	3	Simon Cox, Luis Bermudez	3, 4	Various nits throughout; clarify info- vs non-info-resource argument; hash URIs
27 May	4	Simon Cox, Ingo Simonis	3.4, 7	Reformatted references; simplified explanation of info vs non-info resources; clarified scope of recommendation
4 June	5	Simon Cox	3.4, 4, 5	Minor clarifications in response to comments from Stuart Williams

