USGIN URIs

A Primer

Table of Contents

Section 1: Introduction

Millions of database records and numerous datasets are shared according to USGIN specifications in a distributed, interoperable network. Organizing and maintaining uniqueness among these records and datasets requires a robust system of identifiers that is well-suited for the digital medium.

 This document provides a general introduction to URIs and demonstrates URI applications within USGIN.

Section 2: What is a URI?

A Uniform Resource Identifier (URI) is a compact string of characters for identifying an abstract or physical resource (Berners-Lee ta al, 1998, rfc2396). Breaking this down into manageable chunks:

- URIs are *uniform:* they always use or conform to specific Internet Engineering Taskforce (IETF) URI syntax

- URIs identify a given resource.

Consequently: at their most basic level, URIs use a *consistent syntax* and *identify a resource*. Though multiple URIs can identify a given resource simultaneously, it is *imperative* that a single URI should not identify multiple different resources. Hence, each URI should be *globally unique.*

Generic IETF URI syntax may be found here; World Wide Web-specific IETF syntax may be found here.

Section 3: URNs and URLs

URIs come in two flavors: URNs and URLs. URNs and URLs are both URIs, and a URI can be a URN and a URL simultaneously.

- If a URI identifies a resource regardless of the location of that resource, then the URI is a **URN (Uniform Resource Name)**

- If a URI identifies a (web) location at which a bound resource may be found, then the URI is a **URL (Uniform Resource Locator)**

Though URIs (including URNs and URLs) must follow the appropriate IETF syntax to be considered URIs, most people are familiar with a number of URI analogs.

For example: Social Security numbers and car license plate numbers function as URNs, in that they use consistent syntax and identify resources without specifying that resource's location. Each individual Social Security number uses consistent syntax (NNN-NN-NNNN) and provides an identifier for a human resource (an individual person). Similarly, a car license plate number uses a combination of 6-7 letters and numbers and identifies a specific vehicular resources (a car). In both cases, the identifier does not indicate the location of the resource: a Social Security number does not indicate the location at which a

specific person may be found, nor does a license plate number indicate the location at which an individual car may be found.

Telephone numbers and street addresses are analogous to URLs, in that they follow a consistent syntax and identify a location at which a resource may be found. For example, street addresses use consistent syntax (NNNN [Street Name] [City] [State] [Zip Code] in the United States) and identify the location at which a resource (a shopping center, business, or office) may be found. Telephone numbers also function as URLs: they follow consistent syntax (NNN-NNN-NNNN) and identify a location (on the switchboard) at which a human resource (a person) may be found. In both cases, resources are associated with an identified location by means of a binding: a business resource is bound to a street address in the same way that a human resource is bound to a telephone number. If a business moves to another address or a person is no longer associated with a given phone number, then the binding is invalid.

All web addresses entered into your web browser are URLs. The web site http://www.google.com is a URL that identifies the location at which the Google search engine (a computing resource) may be found.

As with URIs, IETF-compliant syntax is available for both URNs and URLs.

Section 4: USGIN URIs

Like URLs, USGIN URIs can be entered into a web browser. Entering a USGIN URI into a web browser produces a *representation* of the resource identified by the URI. A representation is a symbolic proxy for something.

For example, a fault in the middle of the Mojave Desert can be represented by a symbol on a map, a digital photograph, or a text description of the fault. A USGIN URI identifying this fault would typically dereference to an XML representation of the fault (Figure 1). This XML document would provide structured information about the fault's attributes, including the fault's age, dip, and slip. These attributes can be used for analysis.
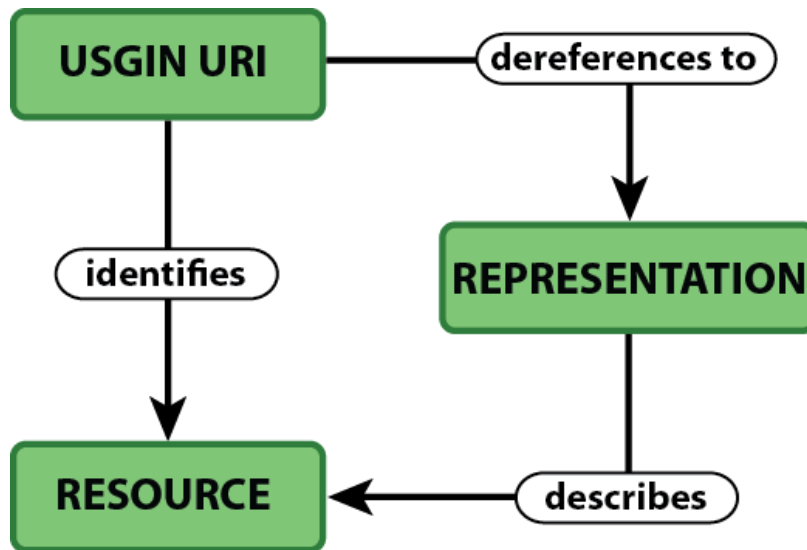
Figure 1: The relationship between USGIN URIs, XML documents, and the resource identified by the USGIN URI

However: even though they dereference to a web page when entered into a web browser, USGIN URIs are **not** URLs.

URLs identify a location at which a bound resource may be found. When entered into a web browser, a URL instructs the web browser to go to the indicated location and retrieve whatever resource is there.

By contrast: USGIN URIs are not URLs because the resource they identify is neither the representation found at the web location to which the USGIN URI dereferences, nor the web location at which said representation may be found

Returning to the example of a fault in the middle of the Mojave Desert: the *fault* is the resource identified by the USGIN URI. The USGIN URI does not identify the *representation* of said fault, nor the location at which said representation may be found. The USGIN URI identifies only the fault; the representation of the fault can be considered an intentional incidental.

Section 5: USGIN URI Syntax

As described above, USGIN URIs identify resources and use HTTP-compatible syntax to produce representations of said resources in web browsers. A USGIN URI can be broken down into a number of component tokens, which are represented in simplified form below:

http://host/uri-gin/authority/resource-type/resource-specific/

The tokens that comprise USGIN URIs are as follows:

- **http://** This token specifies the HTTP protocol for the URI.

- **host/** This [token](#) is a standard World-Wide Web domain name that is used to locate a particular server that has the necessary software to [dereference](#) the URI and return a representation of the [resource](#) identified by the URI. Currently, all USGIN URIs are dereferenced through the host located by:

  resources.usgin.org

- **uri-gin/** By USGIN convention, this [token](#) indicates that the URI is a USGIN URI. All URI components following this token are terminal components of the USGIN URI. The terminal components of the USGIN URI often reflect or incorporate existing [resource](#) [identifiers](#) that are not URIs, such as American Petroleum Institute (API) numbers. This is only practical, as existing resource identifiers are often sufficient for the purposes of USGIN - why reinvent the wheel?

- **authority/** This [token](#) identifies the local naming agent responsible for issuing and maintaining any [resource](#) [identifiers](#) that are present within the the terminal components of the USGIN URI. For example, if a USGIN URI identifies an oil well with an API number, and if the URI incorporates that API number into the URI, then the **authority/** token would reflect the authority responsible for issuing and maintaining API numbers: the American Petroleum Institute (or their agent in a particular state), abbreviated API. Examples of authority abbreviations in the context of USGIN URIs are as follows:

  - **American Petroleum Institute (API)**: http://host/uri-gin/api/

  - **United States Geological Survey (USGS)**: http://host/uri-gin/usgs/

  - **Arizona Geological Survey (AZGS)**: http://host/uri-gin/azgs/

  - **Department of the Interior (DOI)**: http://host/uri-gin/doi/

  - **Department of Energy (DOE)**: http://host/uri_gin/doe/

- **resource-type/** This [token](#) indicates the type of [resource](#) identified by the URI. Note that this token may itself be hierarchical, with resource subtypes delimited by the path separator character '/' Sample USGIN URI conventions can be found below:

  - Active or quaternary fault resource: http://host/uri-gin/authority/fault/

  - Borehole temperature resource: http://host/uri-gin/authority/bhtemp/

  - Thermal spring resource: http://host/uri-gin/authority/thermalspring/

  - Well header resource: http://host/uri-gin/authority/well/

- **resource-specific/** This [token](#) uniquely identifies a [resource](#). Literally anything is possible within this component of the USGIN URI, although USGIN recommends constructing the [string](#) systematically, using existing resource identifiers and URL-safe characters. A list of URL-safe characters for resource-specific strings is as follows:

- o 0-9
- o a-z
- o A-Z
- o /
- o - (dash)
- o _ (underscore)

Consequently, a finished URI that follows USGIN specifications and identifies a specific car (vehicular resource) might appear as follows:

http://resources.cars.org/uri-gin/azdmv/vehicle/ABC123

Likewise, a finished URI that follows USGIN specifications and identifies a military aircraft (military resource) might appear as follows:

http://resources.airforce.gov/uri-gin/usaf/vehicle/aircraft/f22/1234567

This concludes the USGIN URI tutorial.