How to be an USGIN node

USGIN Specification draft.   
Last edit: 2014-01-13 16:15

# History:

* 2014-01-13 Start with USGIN node definitions specification as a template

Contents

[Introduction 1](#_Toc369081022)

[USGIN Node requirements 1](#_Toc369081023)

[Details 2](#_Toc369081024)

[Discussion 3](#_Toc369081025)

[Functional USGIN node categories 4](#_Toc369081026)

[Files-only node 4](#_Toc369081027)

[Service node 4](#_Toc369081028)

[Catalog Node 4](#_Toc369081029)

[Data node 4](#_Toc369081030)

[USGIN Node Self-Description Document 4](#_Toc369081031)

[Standards Statements 5](#_Toc369081032)

[Requirements 5](#_Toc369081033)

[Status of this Memo 5](#_Toc369081034)

[References 5](#_Toc369081035)

# Introduction

This document proposes requirements that must be met in order to label an internet resource "USGIN node".

# USGIN Node requirements

An USGIN node is a web-accessible server that hosts at least one of the functional capabilities enumerated below to play a role in making geoscience information resources accessible.

To be considered a node in the US Geoscience Information Network, the following criteria must be met:

1. The capabilities offered [MUST](#_Requirements) play a role in making geoscience information resources accessible.
2. The capabilities offered [MUST](#_Requirements) be publicly Web-accessible.
3. A node [MUST](#_Requirements) offer at least one of the following capabilities:
   1. Host a Web-accessible repository of relevant data assets with metadata conforming to the USGIN ISO profile ([USGIN Standards and Protocols Drafting Team, 2010-11](http://repository.usgin.org/uri_gin/usgin/dlio/337)) that is published through a USGIN catalog node.
   2. Host a web-accessible folder that 1) contains USGIN-conformant metadata files and 2) is registered for harvesting by a USGIN catalog node.
   3. Host USGIN-conformant web services (WMS, WFS, WCS, etc.) that are registered in a USGIN catalog.
   4. Host a USGIN catalog node, which is a server operating a CSW 2.0.2 service that offers metadata conforming to the USGIN ISO profile.

USGIN nodes [SHOULD](#_Requirements) self-identify by providing an **USGIN node self-description document**. The USGIN Node Self-Description Document section (below) outlines proposed content that describes the node and enumerates the USGIN capabilities offered. The template is intended to be a hypermedia document that can be used by web applications to automate connection to node resources. If provided, the self-description document MUST be accessible at <http://hostName.xxx/.well-known/host-meta>, following the pattern specified by IETF RFC 6415 [http://tools.ietf.org/html/rfc6415].

# Details

**Hypermedia**: Extension of text as an encoding scheme to any media stream that includes information and controls through which the user (or automaton) obtains choices and selects actions. (R. Fielding, 2008, <http://roy.gbiv.com/untangled/2008/rest-apis-must-be-hypertext-driven> (and discussion))

**Metadata conforming to USGIN ISO profile**: See "Use of ISO metadata specifications to describe geoscience information resources" [(USGIN Standards and Protocols Drafting Team, 2010-11)](http://repository.usgin.org/uri_gin/usgin/dlio/337) for detailed guidance on this profile, and "Guidelines for implementation of USGIN metadata content recommendations in ISO 19139 XML metadata" [(USGIN Standards and Protocols Drafting Team, 2012)](http://repository.usgin.org/uri_gin/usgin/dlio/499) for additional implementation details.

**Metadata registered in an USGIN catalog node**: an arrangement is in place so that the registering party creates metadata records that are harvested into the target catalog node. Metadata may be created through a forms interface directly to the catalog, harvested from a non-USGIN catalog end point (e.g OAI-PMH, open-search, ATOM etc.), harvested from a Web Accessible directory, or harvested by a file transfer/upload workflow

**USGIN catalog node**: a server operating a CSW 2.0.2 service that offers metadata conforming to the USGIN ISO profile that describes resources intended as part of the USGIN. Such nodes MUST be registered with the USGIN core aggregating node, either directly or indirectly through another node, such that its metadata may be harvested.

**USGIN core aggregating node**: The central system administration node, includes a catalog repository that harvests metadata from all USGIN catalog nodes. The USGIN core portal accesses this repository for system-wide searches.

**USGIN conformant web service**: a web service that implements an USGIN data exchange, defined by a service protocol, interchange format, and exchange-specific profile.

**USGIN portal:** a web application (USGIN client) that provides a user interface for searching the entire USGIN catalog.

**USGIN home page:** a web site that is the public face of the USGIN, providing access to the portal, information about the system, tutorials, governance information, and specifications.

**USGIN roles:**

* **originator**: agent that created the intellectual property (work)
* **contributor**: agent that prepares a resource for publication and makes it available
* **submitter**: agent that generates metadata to register the resource
* **owner**: agent that holds the intellectual property rights for a resource. Copyright holder, legal role.
* **provider**: agent that operates the service endpoint from which a resource is obtained
* **steward**: agent that is responsible for the maintenance/update of a resource

# Discussion

A node is a server that provides USGIN resources, not a client that uses the offered resources. A server that hosts applications utilizing USGIN services is a vital part of the system, but is not considered a node. Such applications would be expected to be registered in the system by posting metadata to an USGIN catalog.

Applications enabling search across the entire USGIN catalog are an essential part of the system. At least one such search application will be recognized as the official portal to the system, accessed from the USGIN home page on the web.

Each node is under the stewardship of a data **provider**. Individual organizations may host more than one node. For example different servers may host OGC web services, a CSW service, and a repository containing document resources, all under the stewardship of a single data provider. These would be considered three distinct nodes. A node may host data from more than one data **contributor**.

If a server is hosting resources that are described by metadata in an USGIN catalog, and those resources meet the 'geothermal-relevant' criteria, then the operator of that node MAY declare it to be an USGIN node, but doesn't have to.

Putting an USGIN self-description document on the server at the specified location constitutes a declaration that that server considers itself an USGIN node. One of the functions of the self-description document would be that a system registry of known USGIN nodes could have a web crawler application that searches for such self-description documents and starts a node-registration/approval/validation process when it finds new ones. This would enable a 'pull' model registration.

# Functional USGIN node categories

Both data resources and metadata can be made accessible either packaged in files or through services that offer capabilities extending HTTP. Thus USGIN nodes can be categorized according to whether they present information in files or through web services. An individual node my offer both kinds of information access.

## Files-only node

A publish-only node hosts web accessible files that are either information resources registered in an USGIN catalog, or USGIN-conformant metadata records in a web-accessible folder registered for harvest to an USGIN catalog.

## Service node

A service node offers at least one USGIN service that is registered in an USGIN catalog, either through the metadata for datasets exposed by the service, or as a catalog service.

### Catalog Node

A service node that offers a CSW service to search the metadata collection hosted by that node

### Data node

A service node that offers at least one registered data service (e.g. WFS, WMS, WCS, OpenDAP)

# USGIN Node Self-Description Document

Every USGIN node [SHOULD](#_Requirements) offer an USGIN Node self-description document available via http GET at standard path. Recommended practice is to use the IETF "/.well-known/host-meta" path (see http://tools.ietf.org/html/rfc6415). Because many web-server configurations are set up to hide dot files (files or folders with names beginning with a '.') from directory browsing or even direct access, it is also recommended to place a simple JSON-formatted manifest file in the root directory of the server. This file should contain the same information.

This document is a representation of the node resource. The purpose of this document is to make the node self-describing, both as a **Hypermedia** document for machine clients and as a web page for people to read. The document would be a publicly accessible HTML document, discoverable using standard web search technology (Google, Yahoo, Bing, etc.), with text description of the node, the node operator, node capabilities, expected lifetime, acknowledgements for support, etc. The template below defines an HTML profile for including information and links that a software agent could use to inspect the capabilities and access desired function. This might be encoded in HTML link elements in the head section, or using 'rel', 'class', and 'type' attributes on HTML anchors or other elements in the document. Some work needs to be done to define the vocabulary of 'rel' or 'class' attributes, or [microdata](http://schema.org/docs/gs.html) properties

The content for the self-description page would include the following:

1. Node Name: [Name to identify this node]
2. Supporting organization: [name, contact point to report issues]
3. Description: [is there any topical focus, group of contributors, other special characteristics of capabilities offered]
4. Node type: term from controlled vocabulary to indicate type of node. One of {filesOnly, catalog, dataService, catalogAndDataService}
5. Resources:
   1. Data services: List of service types, links to end points, like ArcGIS server REST page?
   2. Metadata publishing profile: how exposed (WAF, CSW, other, including specific profiles used?)
6. How can content of this node?
7. URL for RSS feed for service announcements and information broadcasts.

See the accompanying example document. When the content is established, this could be converted to a template to facilitate creation of new documents for USGIN nodes.

# Standards Statements

## Requirements

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in IETF RFC-2119 [[RFC2119](http://tools.ietf.org/html/rfc2119)].

## Status of this Memo

This USGIN Specification licensed under the CC BY license. USGIN Specification Drafts are working documents of the US Geoscience Information Network Working Group. Note that other groups may also distribute working documents as USGIN Specification Drafts. Current USGIN Specification Drafts can be accessed in the USGIN/usginspecs gitHub repository.

USGIN Specification Drafts are draft documents valid until rejected or adopted by the USGIN, and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use USGIN Specification Drafts as reference material or to cite them other than as "work in progress."

# References

USGIN Standards and Protocols Drafting Team, 2010, Metadata Recommendations for Geoscience Resources: U.S. Geoscience Information Network Best Practices Document, Doc ID gin2010-11, v. 1.0.3, available at <http://lab.usgin.org/profiles/doc/metadata-content-recommendations> (accessed 2010-11-18).

USGIN Standards and Protocols Drafting Team, 2012, Guidelines for implementation of USGIN metadata content recommendations in ISO 19139 XML metadata: U.S. Geoscience Information Network Best Practices Document, available at <http://repository.usgin.org/uri_gin/usgin/dlio/499>.

USGIN Standards and Protocols Drafting Team, 2010-11, Use of ISO metadata specifications to describe geoscience information resources, Doc ID gin2010-009, available at <http://repository.usgin.org/uri_gin/usgin/dlio/337>.