Intelligence Community and Department of Defense Content Discovery & Retrieval Integrated Project Team (CDR IPT)

IC/DoD REST Interface Encoding Specification for CDR Retrieve v1.1

12 May 2011

REVISION/HISTORY

Doc Revision	Revised By	Revision Date	Revisions
0.1		15 February 2010	Initial draft for subgroup review.
0.2	CDR-IPT	01 March 2010	Revision incorporating discussion from 18 Feb.
1.0-20100312	CDR-IPT	12 March 2010	Revision incorporating discussion from 11 Mar.
1.0-20100319	CDR-IPT	19 March 2010	Revision incorporating discussion from 18 Mar.
1.0-20100331	CDR-IPT	31 March 2010	Revision incorporating discussion from 23 Mar.
1.0-20100331	CDR-IPT	15 April	Farley

TABLE OF CONTENTS

LIST OF FIGURES

1 Introduction

1.1 Service Overview

This specification defines requirements and provides guidelines for the realization of the Content Discovery and Retrieval (CDR) Retrieve Component as a web service using the REST style binding, hereafter termed a Retrieve service in this document. It describes a Retrieve service's interface and other aspects in detail, providing enough information for Retrieve service providers and implementers to create CDR-compliant Retrieve services.

The Retrieve Component, as defined by the Intelligence Community/Department of Defense (IC/DoD) CDR Specification Framework, serves as a "pull" mechanism to access the information resource. The Retrieve Component relies on mechanisms that are already well established in the internet infrastructure:

- HTTP GET method [5]
- HTTP Result Codes (maintained by Internet Assigned Numbers Authority – IANA) [8]
- MIME Types (maintained by IANA) [9]

1.1.1 Scope

The Retrieve Component as defined *supports* the retrieval for a specified resource from a Content Collection. The Retrieve Component, as defined, can only support returning a resource directly to the requestor. It cannot redirect output to a component other than the requestor. In addition, no special handling instructions (e.g., routing) may be specified.

1.1.2 Relationship to Other CDR Architecture Elements[1]

The CDR Architecture prescribes an abstract-to-concrete model for the development of architecture elements and guidance for content discovery and retrieval. Each layer or tier of the model is intended to provide key aspects of the overall guidance to achieve the goals and objectives for joint DoD/IC content discovery and retrieval. The following graphic in Figure 1, discussed in detail within the CDR Reference Architecture (RA), illustrates this model.

1 *For a detailed description of each of the layers, please reference the CDR RA Section 1.*

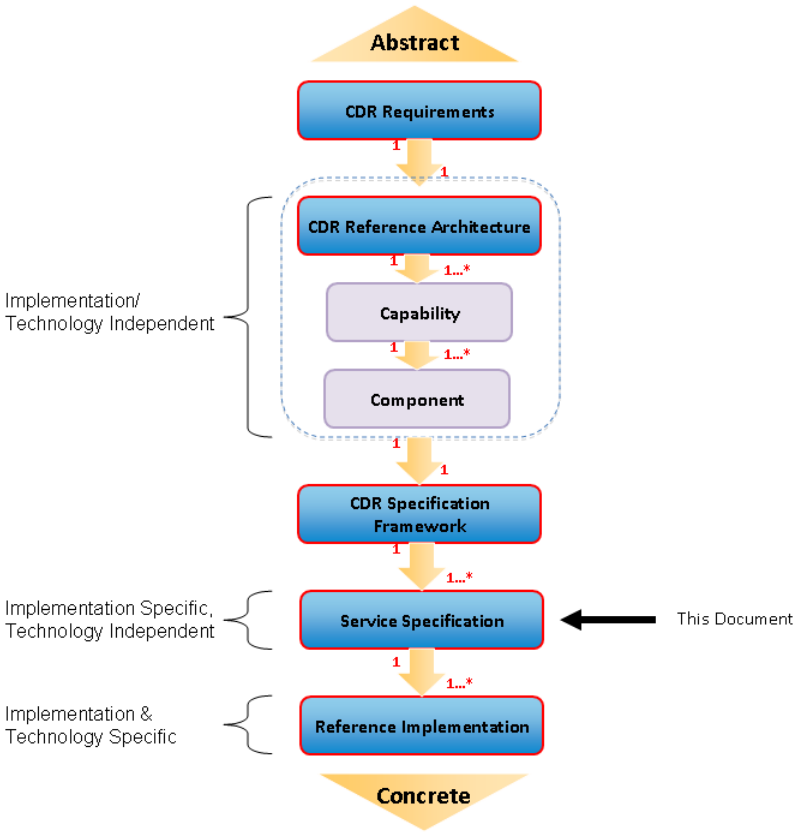

Figure 1 - CDR Architecture Model

As illustrated in Figure 1, the Specification Framework derives from the Reference Architecture (RA) and can describe behavior in terms of the capabilities, components, and usage patterns defined in the RA. The Specification Framework then expands on the details of information flows and the information conveyed in those flows to provide a consistent basis for multiple Service Specifications to provide consistent interfaces both in terms of the structure and the semantics of the exchanged information. Service Specifications, such as this one, provide implementation-specific guidance. More specifically this Retrieve Specification defines the specific guidance for implementing the Retrieve Component.

This specification covers the following aspects of a REST-based Retrieve Component:

- **Service Interface** defines the base REST constructs to expressing inputs, outputs and faults.
- **Implementation**provides additional implementation guidance beyond the behavior and interface guidance.
- **Reference Documentation**provides references to other CDR and community artifacts (i.e., Service Security RA).

1.2 Notational Convention

The key words "MUST," "MUST NOT," "REQUIRED," "SHALL," "SHALL NOT," "SHOULD," "SHOULD NOT," "RECOMMENDED," "MAY," and "OPTIONAL" in this specification are to be interpreted as described in the IETF RFC 2119. When these words are not capitalized, they are meant in their natural-language sense.

1.3 Conformance

This specification defines an interface to a Retrieve service to which an implementation MUST conform. For an implementation to conform to this Retrieve specification, it MUST adhere to all mandatory aspects of the specification.

2 Retrieve Service Interface

2.1.1 Behavior

The Retrieve Component is the application of an HTTP/HTTPS GET method (request) to a single information resource, as identified by a Uniform Resource Locator (URL). A URL is the address of an information resource. The data provider dispatching the information resource constitutes the service response. Note the following about the URL, it:

- MAY refer to any type of information resource. The data provider SHOULD ensure that the content type (i.e., MIME type) and content encoding are available in the HTTP response header so that the consumer may anticipate how it should be processed and to inform any user agents (outside the scope of CDR) that may be available in the future to assist in mediating the referenced information resource to a preferred format. However, there are no guarantees that the consumer will have the necessary software to use the information resource that is retrieved.
- MAY reference content that is constructed dynamically or retrieved from a separate or internal storage location.
- MAY include a fragment identifier (i.e., "#") to denote an information object within a larger information resource.
- SHOULD be permanent to allow bookmarking, emailing, or sharing it in other ways. If a resource is replaced by a new version, the new version SHOULD be given a new URL and the old URL SHOULD return an HTTP 301 (Moved Permanently) status. Some exceptions to this rule are noted in Section 3.3.
- SHOULD NOT be tied to the technology that is used to deliver the information resource (e.g., .JSP, .PHP). This independence further extends the life of the URL as new technologies replace old ones.
- SHOULD be unclassified so that a potential recipient may request access to the resource without the URL leaking information about the content of the resource. Some content management systems have the option of using some or all of the resource title in the URL. This should be avoided for classified content, especially if the title is classified.

HTTP GET requests MUST be safe (i.e., no side-effects), and SHOULD be idempotent (i.e., repeated requests for the same URL yield the same result, as long as a separate activity does not alter the resource) [7]. The HTTP/HTTPS GET method SHOULD be used for information retrieval only and SHOULD NOT cause a change to the information resource beyond relatively harmless effects such as logging, caching, the serving of banner advertisements, or incrementing a web counter. Accordingly, the data provider MUST ensure against unintended effects to the information resource.

2.1.2 Fault Conditions

An implementation of the Retrieve service MUST return the appropriate HTTP status code (based on values from the HTTP Status Code Registry maintained by IANA). The data provider MUST also consult relevant governing policy if the consumer is not properly authorized to access an information resource.[2] See 3.2 for Security Considerations.

3 REST Retrieve Specification Usage

This section provides additional implementation guidance beyond the behavior and interface guidance provided in the previous sections.

3.1 Policy

This specification defines the technical requirements and guidelines for implementing a Retrieve service. Policy for Retrieve service implementations is described in auxiliary documents. See the Reference Documents section for a listing of relevant policy documents. Implementers MUST follow the guidance in those policy documents.

3.2 Security Considerations

Any resource may have associated policies for use, especially as applies to authentication and authorization. These policies may be asserted by both the resource owner and those responsible for governance and management of the enterprise. The implementation of policies related to security considerations SHOULD leverage the specific security components and interactions defined by the Joint IC/DoD Security Reference Architecture (SRA), and MUST be in compliance with requirements and guidance for security outcomes as specified in the SRA and its associated specifications. Retrieve implementations MUST follow the guidance in those documents.

3.3 Implementation Considerations

Smart caching [6] of resources from previous Retrieve requests can significantly improve server and network performance. To maximize quality of service and reduce use of computing and network resources, data providers should implement HTTP caching. This is done by sending appropriate cache control headers, such as “Etag” and “Last-modified,” and by proper handling of conditional request headers, such as “If-modified-since” and “If-none-match.”

As noted in Section 2, URLs should be permanent. Exceptions to this general rule include home-pages, blogs, and the common practice of using a rotating URL to identify the latest version of a specification. Home-pages and blogs serve as portals to potentially dynamic information. The rotating URL supplements the authoritative URL, which

[2] *CJCSI 6212.01E stipulates that the consumer must be presented with instructions within two mouse clicks for petitioning access to a denied information request [1]*

includes versioning information and reflects the principle of permanence and immutability. Here are some examples:

- http://www.w3.org/TR/xml/ (rotating) vs http://www.w3.org/TR/2008/REC-xml-20081126/ (authoritative)
- https://metadata.dod.mil/mdr/ns/DDMS/current/ (rotating) vs http://metadata.dod.mil/mdr/ns/DDMS/3.0/ (authoritative)

As noted in Section 2, a URL MAY include a fragment identifier (i.e., "#") to denote an information object within a much larger information resource. The data provider should facilitate this type of granular addressing within larger information resources to facilitate human and machine processing. Additionally, a number of semantic web technologies (e.g., RDF [10], OWL [11], SKOS [12]) use the fragment to denote a semantic object that can be referenced in an ontology or other knowledge representation. Note: it is the responsibility of the consumer to process the internal reference.

4 Reference Documents

The documents in this section provide the foundation for, define extensions to, and include implementation guidance for the Retrieve service. They include additional specifications, including those provided as part of the greater CDR specification set, and guidance documents that communicate current policy or implementation details. Each document is assigned a reference identifier, which is cited when the document is referenced within this Retrieve Service Specification.

In some cases, documents have been referenced with a version and date of "future" in order to track the iterative development of some of the extensions.

4.1 Specifications

4.1.1 Content Discovery and Retrieval Specifications

The following documents provide a foundation and guidance for the development of this Retrieve Specification document. Retrieve service implementers should have a thorough understanding of the concepts and guidance in these documents; this Retrieve Specification represents a realization of the Retrieve Component defined therein

Ref.	Title	Version	Date
SF	IC/DoD Content Discovery and Retrieval Specification Framework	DRAFT 0.6.2	29 Jan 2010
RA	IC/DoD Content Discovery and Retrieval Reference Architecture	DRAFT 0.4	16 Dec 2009

4.1.2 Other Specifications

4.1.2.1 Security Specifications

Ref.	Title	Version	Date
S1	Joint IC/DoD Security Reference Architecture	1.0	25 Jul 2008

4.1.2.2 Service Discovery Specifications

Ref.	Title	Version	Date
D1	Joint IC/DoD Service Discovery Architecture	DRAFT 1.2	28 Sep 2007

4.2 Policy and Guidance

4.2.1 Content Discovery and Retrieval Policy and Guidance

This specification primarily addresses the behavioral and interface aspects common to all CDR Retrieve service implementations. The following documents provide additional requirements and expectations set by policy or as a result of

Ref.	Title	Version	Date
P1	IC/DoD Content Discovery and Retrieval Retrieve Service Policy for REST Implementations	Future	Future

4.3 Additional References

[1] Chairman of the Joint Chiefs of Staff Instruction 6212.01E – http://www.dtic.mil/cjcs_directives/cdata/unlimit/6212_01.pdf
[2] DoD Discovery Metadata Specification (DDMS) – http://metadata.dod.mil/mdr/irs/DDMS
[3] Fragment Identifier – http://en.wikipedia.org/wiki/Fragment_identifier
[4] HTML – http://en.wikipedia.org/wiki/HTML
[5] HTTP/HTTPS – http://www.w3.org/Protocols/
[6] HTTP Caching - http://www.w3.org/Protocols/rfc2616/rfc2616-sec13.html
[7] HTTP Method Definitions - http://www.w3.org/Protocols/rfc2616/rfc2616-sec9.html
[8] Hypertext Transfer Protocol (HTTP) Status Code Registry – http://www.iana.org/assignments/http-status-codes
[9] MIME type – http://www.ietf.org/rfc/rfc2045.txt
[10] RDF – http://www.w3.org/RDF/
[11] Semantic Web – http://www.w3.org/standards/semanticweb/
[12] SKOS – http://www.w3.org/2004/02/skos/