Intelligence Community and Department of Defense
Content Discovery & Retrieval Integrated Project Team

IC/DoD REST Interface Specification
for CDR Retrieve

Version 2.0

Approval Date: 3-OCT-2012

This document has been approved for Public Release by the
Office of the Director of National Intelligence. See ‘Distribution Notice’ for details

Distribution Notice

This document has been approved for Public Release and is available for use without restriction

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
2.0	Wigglesworth	17 February 2012	Harmonization

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Service Overview

The Retrieve Component, as defined by the “Intelligence Community/Department of Defense (IC/DoD) Content Discovery and Retrieval (CDR) Specification Framework” [CDR-SF], is the primary mechanism for content consumers to access one or more specific content resources from content collections. This component provides a common service interface and behavioral model for IC and DoD content collections, enabling consumers to retrieve and initiate delivery of content resources. Specifically, the Retrieve Component provides a means to retrieve the native content described in the Search Component query results.

This specification defines requirements and provides guidance for the realization of the CDR Retrieve Component as a web service using REST[1], hereafter termed a Retrieve Service in this document. The content of this specification describes the Retrieve Service’s behavior, interface and other aspects in detail, providing enough information for Retrieve Service providers and consumers to create and use CDR-conformant Retrieve Services.

The Retrieve service exposes a single Retrieve function. While the function is often used in concert with retrieving results of a search, it may be used in general to process any compliant retrieve instructions. As discussed in CDR Specification Framework, a Retrieve Service’s results are the content resource. In the context of Retrieve, the content resource generally refers to the entire underlying record.

1.2 Scope

The Retrieve Component as defined supports the retrieval of a specified resource from a content collection and returning that content to the requestor.

1.3 Artifact Overview

This specification is a part of the set of specifications that define the concrete, implementation-specific guidance for the services defined under the auspices of the Content Discovery & Retrieval (CDR) Integrated Project Team (IPT). The CDR Reference Architecture [CDR-RA] 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, discussed in detail within the CDR Reference Architecture, illustrates this model.

[1] *REST is an architectural style that encapsulates the design principles of the World Wide Web (WWW) [REST]*

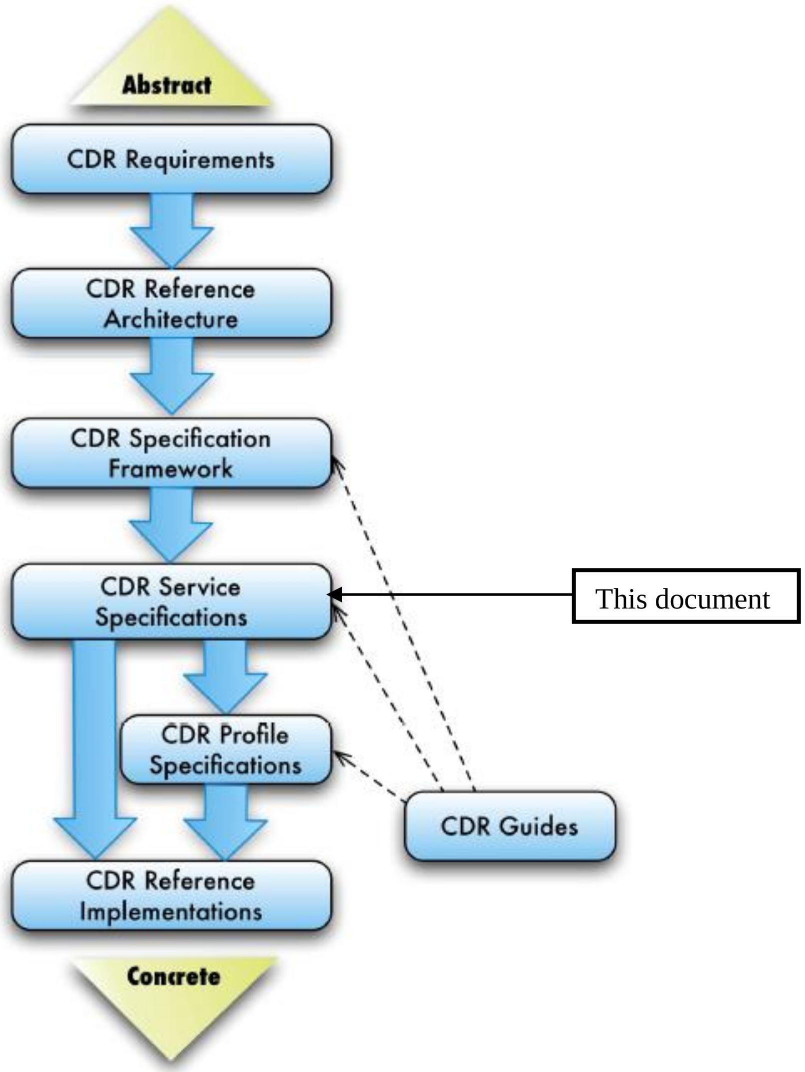
Abstract
CDR Requirements
CDR Reference Architecture
CDR Specification Framework
CDR Service Specifications
This document
CDR Profile Specifications
CDR Guides
CDR Reference Implementations
Concrete

Figure 1: CDR Architecture Model

As illustrated in Figure 1, the CDR Specification Framework [CDR-SF] derives from the CDR Reference Architecture [CDR-RA] and describes behavior in terms of the capabilities, components, and usage patterns defined in the RA. Multiple CDR Service Specifications are derived from the CDR-SF, with separate specifications associated with the components of the architecture (e.g., Retrieve) and, for each service, separate specifications to address Representational State Transfer (REST) and SOAP implementations.

This specification provides guidance for implementing the CDR Retrieve Service as a REST Web Service. It is intended to parallel the corresponding SOAP specification, the IC/DoD Content Discovery & Retrieval SOAP Interface Specification for CDR Retrieve [CDR-SR], as closely as possible, to minimize the difficulties in interoperating. Additional CDR Guides, Profile Specifications, or Reference Implementations may provide additional guidance on implementing this specification in a particular context.

1.4 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.

When describing concrete XML schemas and example XML documents, this specification uses XPath as the notational convention. Each member of an XML schema is described using an XPath notation (e.g., `/x:RootElement/x:ChildElement/@Attribute`). The use of {any} indicates the presence of an element wildcard (<xs:any/>). The use of @{any} indicates the presence of an attribute wildcard (<xs:anyAttribute/>).

Items contained in curly braces ({item}) are meant to indicate template or notional values to be replaced by actual values (without the use of curly braces) when in actual use.

Examples in this text are distinguished by a black border. These are meant to be illustrative and only one way that the described syntax can be used.

```
<atom:entry>
    <atom:title>This is an example.</atom:title>
</atom:entry>
```

1.5 Conformance

This specification defines an interface to a Retrieve Service to which an implementation and a subsequent deployment MUST conform. A deployment is an instance of an implementation. For an implementation to conform to this Retrieve specification, it MUST adhere to all mandatory aspects of the specification.

1.6 Namespaces

No namespaces are used in this specification with the possible exception of examples.

1.7 Security

This specification does not directly address security concerns. It will be necessary for any implementation of this specification to address security concerns relevant to the systems with which they interact and the governance bodies. Several aspects of retrieve, to include returning only the content resource for which the requesting entity is authorized, should be addressed in the detailed security plan of an implementation, but are out of scope for this document.

2 Retrieve Service Interface

2.1 *Retrieve Function*

2.1.1 Preconditions

The following preconditions MUST be satisfied if the retrieve function is to correctly process input and generate results and post-conditions as described.

1. The requester is authenticated and authorized according to applicable policy requirements for this function.
2. The resource exists and can be retrieved.

2.1.2 Input

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).

2.1.2.1 HTTP Method

The Retrieve function MUST use the HTTP GET method.

2.1.2.2 URL Template

`{ContentResourceID}?{RetrieveProperties}`

> `{ContentResourceID}` - REQUIRED - The value that represents the content resource. This value MAY include a fragment identifier (i.e., "#") to denote an information object within a larger information resource.
>
> `{RetrieveProperties}` - OPTIONAL - Parameters through which the consumer may specify and configure optional behavior supported by the Retrieve function implementation.

Consideration should be given in the construction of the URL such that a potential recipient may request access to the resource without the URL leaking restricted information about the content of the resource.

2.1.2.3 HTTP Message Header

There is no request message header specified for this function, however this specification encourages the use of HTTP/1.1[HTTP11] content negotiation[2] for requesting the content in a specified format.

[2] *Content negotiation is a mechanism used by a provider to have different format representations of the same content to the consumer in the same request URL, allowing the consumer to choose the format representation that best fits their needs.*

2.1.2.4 HTTP Message Body

There is no request message body for this function.

2.1.2.5 Retrieve Request - Message Example

```
GET /1225c695-cfb8-4ebb-aaaa-6fda344efa6a HTTP/1.1
```

Figure 2: Example Retrieve Input

2.1.3 Output

The output of the CDR Retrieve Service is the content resource specified by the identifier provided in the service request. For requests that result in an error, a HTTP Error code will be returned.

2.1.3.1 HTTP Status Code

If the GET request is successful, the service will respond with a '200 OK' Status Code and the content resource.

2.1.3.2 HTTP Message Header

The response header SHOULD include the content type (i.e., MIME[3] type) and content encoding of the content resource 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.

2.1.3.3 HTTP Message Body

The body of the HTTP message MUST consist of the retrieved content resource and MUST be present in the HTTP `Message`.

2.1.3.4 Output Example

This example represents sample output embedding the content resource (binary data) into the body of the HTTP message.

```
HTTP/1.1 200 OK
Content-Length: 300
Content-Type: application/pdf

IkdpdmUgbWUgYSBsZXZlciBsb25nIGVub3VnaCBhbmQgYSBmdWxjcnVtI
G9uIHdoaWNoIHRvIHBsYWNlIGl0LCBhbmQgSSBzaGFsbCBtb3ZlIHRoZS
B3b3JsZC4iIC0gIEFyY2hpbWVkZXM=
```

Figure 3: Example Retrieve Output

2.1.4 Post-Conditions

The following conditions MUST be met upon successful completion of the function.

[3] *Internet media types (MIME types) are identifiers for file formats on the internet [MIME-TYPES].*

1. The results returned are the content resource identified by the request.
2. The use this function has been audited according to applicable policy.[4]

2.1.5 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.

3 References

[CDR-RA]

"CDR IPT Reference Architecture ", 1.1, 25 Feb 2011.

[CDR-SF]

"IC/DoD Content Discovery & Retrieval Specification Framework 2.0." 2011.

[CDR-SR]

"IC/DoD Content Discovery & Retrieval SOAP Interface Specification for CDR Retrieve 2.0." 2012.

[HTTP11]

"RFC2616 Hypertext Transfer Protocol -- HTTP/1.1." June 1999.

[MIME-TYPES]

"Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types." 1996.

[REST]

"Principled Design of the Modern Web Architecture", R Fielding, R Taylor. ACM Transactions on Internet Technology (New York: Association for Computing Machinery), May 2002. Available at http://www.ics.uci.edu/~taylor/documents/2002-REST-TOIT.pdf

[4] *The use of this function may be audited according to applicable policy and may include auditing of the success or failure of the function.*

Appendix A. Mapping to Specification Framework

This section explicitly ties the items in this specification to the requirements of the CDR-SF. The CDR-SF identifies the requirements for creating specifications, while the implementation details are outlined here.

A.1. Retrieve Request

Table 1: Retrieve Request Mapping to Specification Framework

Specification Framework Variable	REST Retrieve Specification
Identifier	`{ContentResourceID}`
Retrieve Properties	`{RetrieveProperties}`

A.2. Retrieve Request

Table 2: Retrieve Response Mapping to Specification Framework

Specification Framework Variable	REST Retrieve Specification
Content Resource	HTTP Message Body