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

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

12 May 2011

REVISION/HISTORY

Doc Revision	Revised By	Revision Date	Revisions
0.1		29 January 2010	Initial draft for subgroup review.
0.2	Dave Lemen, Dan Wigglesworth	1 February 2010	Second draft, limiting specification to format-agnostic OpenSearch implementation.
0.3	Dave Lemen	16 February 2010	
0.4	Dave Lemen	23 February 2010	
0.5	Dave Lemen	9 March 2010	
0.5	Mary Farley	14 April 2010	Technical Edits

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Service Overview

The Search Component, as defined by the Intelligence Community/Department of Defense (IC/DoD) Content Discovery and Retrieval (CDR) Specification Framework [CDR-SF], serves as the primary content discovery mechanism to expose content collections for discovery and accessibility. This component provides a common interface and behavioral model for IC and DoD content collections, enabling content consumers to discover relevant content resources from disparate collections across the IC/DoD enterprise.

This specification defines requirements and provides guidelines for the realization of the CDR Search Component as a RESTful, OpenSearch [OS] web service, hereafter termed a Search service in this document. It describes a Search service's behavior, interface and other aspects in detail, providing enough information for Search service providers and implementers to create CDR-compliant Search services.

The Search service exposes a single Search operation that is responsible for three activities that underpin Content Discovery capabilities: search, result presentation, and results paging. As discussed in the CDR Specification Framework, a Search service's results are resource metadata rather than actual content resources. In the context of Search, resource metadata generally refers to a subset of a resource's available metadata, not the entire underlying record[1]. Some of the information contained within each Search result may provide the information necessary for a consumer to retrieve or otherwise use a resource.

1.2 Relationship to Other CDR Architecture Elements

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, discussed in detail within the CDR Reference Architecture [CDR-RA], illustrates this model.

[1] *The Search Component returns metadata about a resource, which may sometimes describe the underlying resource (e.g., an image), while at other times representing a sub-set of the data the makes up a resource (e.g., a collection of attributes). In some cases, the metadata returned from an instantiation of the Search function and the Retrieve function, which returns a resource itself, may happen to be the same, though this considered an edge condition.*

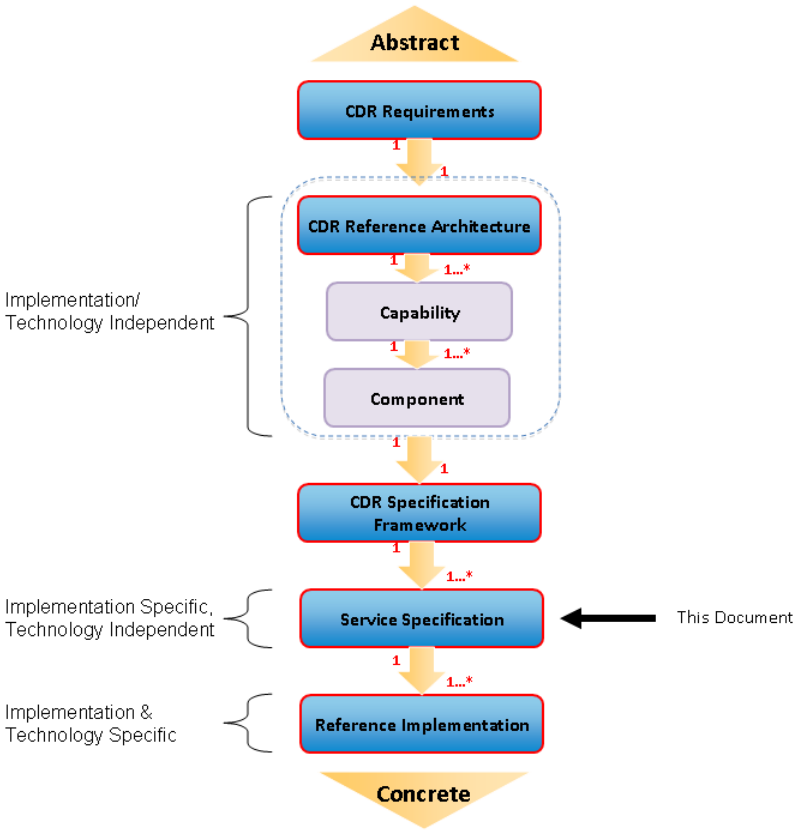


Figure 1 - CDR Architecture Model

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 allows multiple Service Specifications to provide consistent interfaces, both in terms of the structure and semantics of the exchanged information.

This specification provides guidance for implementing the CDR Search Component using the RESTful OpenSearch [OS] standard. It is intended to provide minimal requirements for implementing OpenSearch. Additional sub-specifications will provide further guidance for implementation profiles that include specific query types and result formats.

1.3 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).

To make the text easier to read and understand, some examples include data types that are derived in auxiliary specifications from abstract types defined in this specification. To distinguish these derived types from those defined as part of this Search Specification, they are presented in green.

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>
```

1.4 Conformance

Search services must support OpenSearch 1.1 Draft 4 [OS].

This specification defines an interface to a Search 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 Search specification, it MUST adhere to all mandatory aspects of the specification.

1.5 Namespaces

Namespaces referenced in this document and the prefixes used to represent them are listed in the following table.

Table 1 – Referenced XML Namespaces

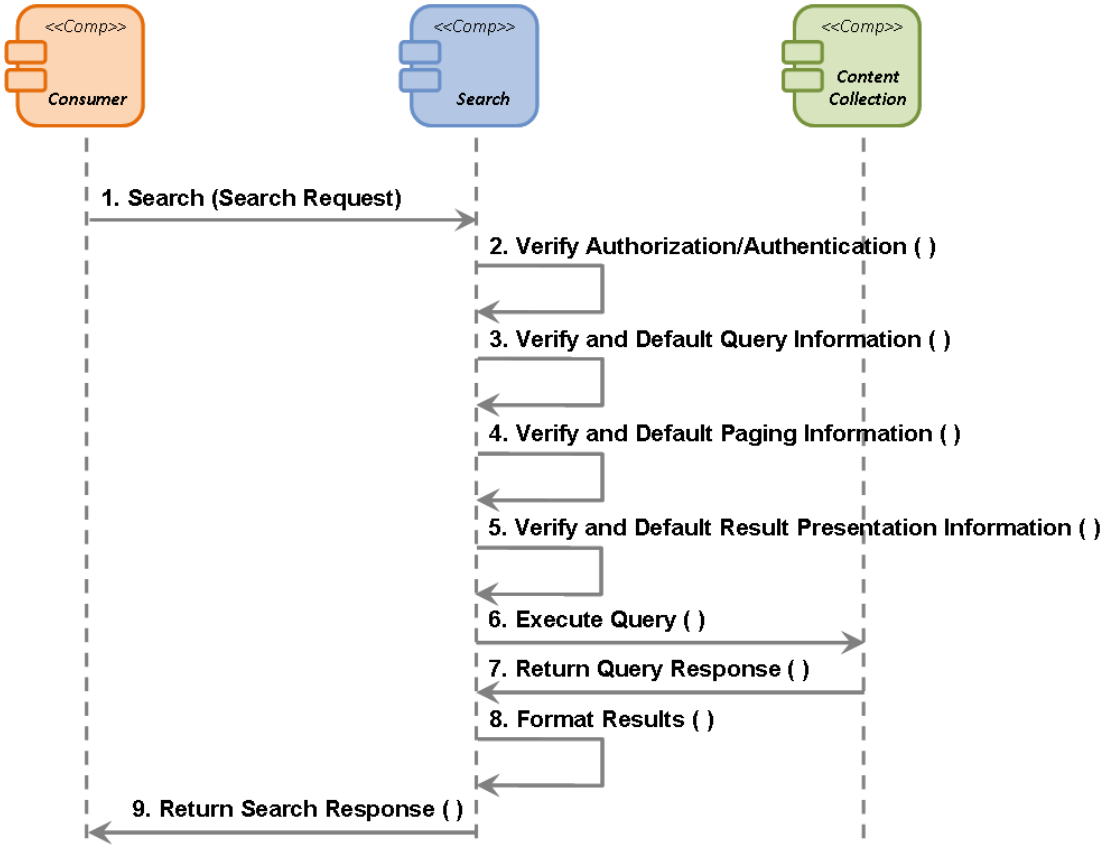
Prefix	URI	Description
opensearch	http://a9.com/-/spec/opensearch/1.1/	OpenSearch 1.1 (Draft 4)[2]
atom	http://www.w3.org/2005/Atom	Atom 1.0

1.6 License

2 *The OpenSearch specification can be found at http://www.opensearch.org/Specifications/OpenSearch/1.1/Draft_4.*

2 Search Service Behavior

2.1 Main Flow



1. The Consumer invokes the Search function on the Search Component with the parameters for a particular type of Search Request specified. The Search Request contains the query, paging information and any other information required by the particular query type being sent.
2. The Search Component leverages a set of security components to verify that the Consumer is authenticated and authorized to perform the search query. The DoD/IC Service-Oriented Architecture Security Reference Architecture [SEC-RA] defines the specific security components and interactions needed to perform this verification. See Section 4.5 of this document.
3. The Search Component checks the type of the incoming Search Request (e.g., keyword, temporal etc.) to verify that it supports that particular type and that its syntax is valid.
4. The Search Component checks the validity of the provided paging information and utilizes default values for any that are missing.
5. Because Search service result presentation is completely determined via the OpenSearch binding, this step of the main flow is not applicable to OpenSearch-based Search services.

6. The Search Component executes the query against its Content Collection.
7. The Search Component receives the results for the query that was executed against its Content Collection.
8. The Search Component performs any necessary formatting of the results before returning a response to the Consumer.
9. The Search Component returns a formatted response back to the Consumer containing:
   - A result set containing the resource metadata for each search result, presented in the format associated with the service's interface.
   - A Timestamp containing the time that the search was executed.
   - The Start Index value that was used.
   - The Results Per Page value that was used.
   - The Input Search Request that was executed.
   - The Response Result Count containing the number of results within the current page.
   - The Total Result Count containing the number of total results matching the Input Search Request that was executed, if available.

2.2 Query

The OpenSearch specification does not define a syntax for its primary query parameter, searchTerms, but it is generally used to support simple keyword queries. There are a number of OpenSearch extensions that support query types beyond simple keyword queries as well. These include the OpenSearch Geo Extension [OS-GEO] for geospatial query terms, and the OpenSearch Time Extension [OS-TIME] for temporal query terms. Use of these extensions is discussed in more detail in section 3.1.3. Additional query types may be developed by the CDR in the future.

2.3 Paging

Paginated search results can be useful when the number of results is very large, infinite, or indeterminate. Service consumers can page through the result sets, accessing a subset of the overall result set as necessary. This capability will prevent search requests with very large result sets from overloading the server, network, or client.

Search result pages may be traversed using the information from the original Search service request combined with the endpoint information provided by the OpenSearch Description (OSD) document describing the Search service from which the current result set was generated. The Search service OSD allows a service consumer to issue a search request for the next "page" of data.

CDR implementations SHOULD support paging.

Description of significant parameters which can be included in the URL template within an OSD document:

/opensearch:count

The desired number of results to return per page

/opensearch:startIndex*

The starting index number of the returned results

/opensearch:startPage*

The starting page number of the returned results

* *A Search service SHOULD NOT use both startIndex and startPage, since their functions overlap.*

Example:

```
http://example.com/test?q={searchTerms}
&startIndex={startIndex?}&format=atom
```

Description of significant elements that a Search service SHOULD return in the response:

/opensearch:totalResults

The total number of results available for the query

/opensearch:startIndex

The start index of this result set

/opensearch:itemsPerPage

The maximum number of items to be returned in a result response

```
<atom:feed>
  …
  <opensearch:totalResults>492420</opensearch:totalResults>
  <opensearch:startIndex>1</opensearch:startIndex>
  <opensearch:itemsPerPage>10</opensearch:itemsPerPage>
  …
</atom:feed>
```

The response MAY include paging links to the previous and next results. In addition, the service MAY include the first results page, current results page, and last results page. For instance, if the response format is Atom, the paging links can be included as follows:

```
<atom:feed>
  …
  <atom:link rel="previous" href="http://example.com/test?q=tanks&page=2&format=atom"
type="application/atom+xml"/>
  <atom:link rel="next" href="http://example.com/test?q=tanks&page=4&format=atom"
type="application/atom+xml"/>
  <atom:link rel="self" href="http://example.com/test?q=tanks&page=3&format=atom"
type="application/atom+xml"/>
  <atom:link rel="first" href="http://example.com/test?q=tanks&page=1&format=atom"
type="application/atom+xml"/>
  <atom:link rel="last" href="http://example.com/test?q=tanks&page=447&format=atom"
type="application/atom+xml"/>
  …
</atom:feed>
```

The paging mechanism supported by the Search service does not guarantee continuity of search results while switching pages.Content resources may be added, updated, or

removed in the period of time between which the different pages of the result set are accessed -- without the consumer being aware of these changes. Therefore, service consumers SHOULD NOT present paged result sets as coherent or complete or make assumptions to that effect.

2.4 Result Presentation

The CDR Specification Framework identifies additional result presentation behaviors:

- Result Sorting Order
- Result Metadata Style

Support for sorting functionality is OPTIONAL, however Search services SHOULD provide results sorted by relevance by default, if possible. The base search request does not include any parameters to request or control sorting (i.e., the Result Sorting Order variable described by the CDR Specification framework is ignored). Instead, individual Query Type specifications MAY add sorting parameters and behavior requirements.

There will be a different endpoint for each supported response type (e.g., HTML, Atom feed).

2.5 Relevance of Search Results

The OpenSearch Relevance Extension [OS-RELV] provides a mechanism for communicating the relative importance of each result. Result relevance is generally a measure of how well a specific result matched the original query. Providing a result relevance measure allows better matched results to be prioritized relative to other results. A Search service implementation MAY provide relevance scores for individual search results with respect to the particular search with which it is identified. The OpenSearch Relevance extension defines a single element:

/relevance:score

The range of values allowed is any decimal between 0 to 1, inclusive, with 1 being the most relevant and 0 the least.

Usage:

```
<relevance:score>0.97</ relevance:score>
```

This scheme does not define the mechanism by which the relevance score is determined. In addition, comparing scores calculated under this scheme by different Search service instances may not provide a true comparison of relevancy.

An example Atom feed that includes opensearch:Relevance elements in its individual results follows:

```
<atom:feed>
  ...
  <atom:entry>
    result metadata
    ...
    <relevance:score>0.97</relevance:score>
  </atom:entry>
  <atom:entry>
    result metadata
    ...
    <relevance:score >0.42</ relevance:score>
  </atom:entry>
  ...
</atom:feed>
```

3 Search Service Interface

3.1 Input

The URL template within the OSD document defines the syntax for how to call the Search service. The URL element requires two attributes to be present:

/@type
Contains the MIME type of the query result format
/@template
Contains the query URL to be processed according to the syntax rules

The template syntax rules define how values can be placed into the URL to create a valid query. For example:

```
<Url type="application/atom+xml"
template="http://example.com/?q={searchTerms}&amp;pw={startPage?}&amp;format=atom"/>
```

The type parameter determines the return format of the query. Guidance on return formats is discussed in Section 4.3.

Both {searchTerms} and {startPage?} are variables that are to be replaced by a user or application using the URL. Variables without a question mark are required to be replaced with an actual value. Variables with question marks are required to be replaced by an empty string or by an actual value. An empty string indicates that the parameter should not be used for the query.

For further guidance regarding the URL element and the URL template syntax, refer to the OpenSearch specification.

3.1.1 Relation to Inputs Defined in the Specification Framework

The IC/DoD CDR Specification Framework defines a number of required (R) and optional (O) inputs to the Search operation. The following table relates the disposition of each variable defined in the Framework in this specification:

Table 2 – Framework Input Variable Disposition

Activity	*Framework Input Variable*	*Search Specification*
Search	Query (R)	The query is an aggregate of the parameters in the request (R).
	Query type (O)	Implicit – query type is implicit in the Open Search parameter.

	Query metadata (O)	Specialized – query type definitions MUST specify all possible information (including metadata) that can be used as input for its type. In essence, query metadata is moved from the common search request and down into the specific query types, to be defined and used as necessary on a per-type basis.
	Timeout (O)	*Not supported*
Results paging	Results per page (O)	opensearch:resultsPerPage (O)
	Start index (O)	opensearch:startIndex (O)
	Query identifier (O)	*Not supported*
Results presentation	Result metadata format (O)	Implicit – each Search service binding MUST be associated with a specific result metadata format. Therefore this input variable is not needed.
	Result sorting order (O)	Default sorting by relevancy is RECOMMENDED. Individual query types MAY define input variables to control custom sorting; otherwise sorting order input is not supported and no variable is defined.

The CDR Specification Framework allows for *Query Metadata* as an optional Search input. This Search Service Specification does not enable input of metadata independent of a particular type of query. Individual query types MAY define metadata inputs and other kinds of inputs, as desired, as part of their search syntax. In the RESTful implementation specified by this document, this information is implicit in the type of parameter(s) sent in the Search Service request, therefore a query type input is not necessary.

3.1.2 OpenSearch Paging Parameters

The Search service specification is REQUIRED to function as described by the Specification Framework with any input, behavior, output, and fault condition extensions listed below.

/startIndex
This OPTIONAL element indicates the offset of the first result in a result set. Its value, if provided, MUST be an integer. The base value SHOULD be 1.

/resultsPerPage
This OPTIONAL element describes the desired number of search results per page. Its value, if provided, MUST be greater than or equal to 1. The default value is 10. Support for this property is REQUIRED.

3.1.3 OpenSearch Query Type Extensions

The OpenSearch specification defines one parameter, searchTerms, for the query. OpenSearch does not specify a syntax for the contents of the searchTerms parameter. Extensions to OpenSearch have been developed to support query types that go beyond keyword-style queries. These include:

- **OpenSearch Geo Extension [OS-GEO]** – Defines parameters for supporting bounding box, point-radius, polygon, and other geospatial query types.

- **OpenSearch Time Extension [OS-TIME]** – Defines parameters for specifying a time range. This specification does not describe how to interpret the temporal query terms. For example, it does not specify that the temporal terms apply to the date a resource was created or posted, or that it may apply to a historical event described by the resource. Guidance related to interpretation can be found in implementation guidance, such as the DoD Discovery Metadata Specification Implementation Guide [DDMS].

3.2 Behavior

An implementation of the Search service MUST follow the behavior defined in the CDR Specification Framework. Additional requirements on behavior MAY be defined by query type specifications. Search services MUST implement the behaviors required by the query types they support.

3.3 Output

In addition to the requirements imposed by the CDR Specification Framework, the Search service Search function output is additionally constrained by the requirements specified in the OpenSearch Specification.

The following example illustrates the high level components of a response message (containing a result set of unspecified type) from a Search service:

```
<result-set>
   ...
   <result>...</result>
   <result>...</result>
   ...
</result-set>
```

3.3.1 Relation to Outputs Defined in the Specification Framework

The IC/DoD CDR Specification Framework [CDR-SF] defines a number of required (R) and optional (O) outputs from the Search operation. The following table relates the disposition of each variable defined in the Framework in this specification:

Table 3 – Framework Output Variable Disposition

Activity	*Framework Output Variable*	*Search Specification*
Search and Results Paging	Result set (R)	A ***Search*** service MUST return a formatted set of results. Search service implementations SHOULD support HTML and Atom response formats.
	Results metadata (R)	CDR Result Type specifications MAY require certain types of data to be returned as part of the result set or individual result entries. Those specifications MAY also allow other types of metadata to be included and describe the mechanism for doing so. A Search service that supports a particular result type MUST follow the syntax and processing rules defined by that type. (R)
	Result set retrieval properties (O)	Not Supported.
	Result relevancy value (O)	relevance:score in each result
	Result retrieval properties (O)	An element describing the linkage to the ***Retrieve*** service MUST be included in the results. (O)
	Timestamp (O)	CDR Result Type may require timestamp element (ie feed/updated).
	Query identifier (O)	Not Supported.
	Response result count (O)	opensearch:itemsPerPage SHOULD be included in the response.
	Total result count (O)	opensearch:totalResults SHOULD be included in the response.

3.3.2 Including the Search Request in the Response

To facilitate paging capabilities and to provide service consumers the ability to re-execute their queries, the search request that produced the output MAY be included in the results response. The mechanism for doing this will depend on the response format being returned. In the following example, the Search Request is inserted directly under the root of a notional response:

```
<atom:feed>
  ...
   <atom:link rel="self" href="http://example.com/test?q=tanks&amp;page=1&amp;format=atom"
type="application/atom+xml"/>
  ...
</atom:feed>
```

3.3.3 Including Metadata in the Results

Depending on the underlying data resources and the type of search request being executed, Search services MAY return metadata about each resource beyond that required by the Result Type specification. That specification controls the mechanism and

361 syntax for including any additional metadata and whether or not such inclusion is
362 permitted. Search services that support a particular Result Type in its response MUST
363 follow the requirements in the associated Result Type specification.
364

3.3.4 OpenSearch Paging Elements

Search service implementations SHOULD use the elements and parameters from the OpenSearch specification for paging purposes. The exact mechanism and syntax for including the OpenSearch extensions are defined by the specification for the Result Type being used. The requirements for the inclusion of OpenSearch extensions are listed in the table below.

Table 4 – OpenSearch Output Extensions

Element	Description
`opensearch:totalResults`	The actual or estimated number of resources that match the current query. In the absence of a totalResults element or a next link (//link[@rel='next']), the search client should consider the current page to be the last page of search results. • Restrictions: The value must be a non-negative integer. • Default: The default value is equal to the offset index of the last search result on the current page. • Requirements: The element may appear zero or one time. **Recommendation:** The Search Service implementation SHOULD include the opensearch:totalResults in the response. **Example:** <opensearch:totalResults> 492420 </opensearch:totalResults>
`opensearch:startIndex`	The offset of the first search result in the current set of search results. The OpenSearch specification allows for any integer offset for the first search result (0, 1, -5, etc.) The index of the first search result MAY be indicated in the OSD document's "Url" element, using the indexOffset attribute. If the startIndex element does not appear on the page then the search client should consider the current page to be the first page of search results. • Restrictions: The value MUST be an integer.

	• Requirements: The element MAY appear zero or one time. **Recommendation:** The Search Service implementation SHOULD include the opensearch:startIndex in the response. A Search implementation SHOULD use an indexOffset of 1. **Example:** <opensearch:startIndex> 11 </opensearch:startIndex>
opensearch:itemsPerPage	The number of search results returned per page. If the itemsPerPage element does not appear on the page then the search client SHOULD use the number of items of the current page as the default page size. • Restrictions: The value must a non-negative integer. • Default: The default value is equal to the number of search results on the current page. • Requirements: The element may appear zero or one time. **Recommendation:** The Search Service implementation SHOULD include the opensearch:itemsPerPage in the response. **Example:** <opensearch:itemsPerPage> 10 </opensearch:itemsPerPage>

373

The following XML illustrates one possible representation of these properties in a notional response (the Result Type specification will define the exact representation):

```
<atom:feed>
  ...
  <opensearch:totalResults>12</opensearch:totalResults>
  <opensearch:startIndex>11</opensearch:startIndex>
  <opensearch:itemsPerPage>10</opensearch:itemsPerPage>
  ...
</atom:feed>
```

3.4 Fault Conditions

An implementation of the Search service MUST allow for the Fault Conditions defined in the CDR Specification Framework.

Query type specifications MAY create additional Fault Conditions, as necessary. Any new fault types SHOULD derive from existing fault types, if possible.

Table 5 maps the CDR Specification Framework fault conditions to the HTTP status that SHOULD be returned for each.

Table 5 – Fault Conditions and HTTP Responses

CDR Framework Fault Condition	HTTP Status	HTTP Description
Unauthorized Access	403	Forbidden
Unsupported Query Type	400	Bad Request
Unsupported Search Request Syntax	400	Bad Request
Unsupported Search Element	400	Bad Request
Invalid Paging Value	400	Bad Request
Paging Value Out of Range	404	Not Found
Service Execution Error	500	Internal Service Error

4 Search Service Implementation

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

4.1 Policy

This specification defines the technical requirements and guidelines for implementing a Search service. Policy for Search 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.

4.2 Query Types

The CDR Specification set includes a number of Query Type definitions that IC/DoD organizations can leverage in their Search service implementations, depending on the applicable policies and implementation profiles.

4.3 Result Types

The CDR Specification set includes a single predefined Result Type definition that IC/DoD organizations can leverage in their Search service implementations, the IC/DoD Content Discovery and Retrieval Atom 1.0 Result Set Specification [CDR-ATOM]. Implementers SHOULD consult appropriate policy and implementation guidance to determine requirements or recommendations concerning the use of particular Result Types.

4.4 Sorting of Search Results

Sorting is OPTIONAL for Search services. Search services that do implement sorting SHOULD return results sorted by relevance.

4.5 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.

5 Reference Documents

The documents in this section provide the foundation for the Search service. Each document is assigned a reference identifier, which is cited when the document is referenced within this Search Service Specification.

Ref.	Title	Version	Date
CDR-SF	IC/DoD Content Discovery and Retrieval Specification Framework	DRAFT 0.6.1	25 Jan 2010
CDR-RA	IC/DoD Content Discovery and Retrieval Reference Architecture	DRAFT 0.4	16 Dec 2009
ATOM	The Atom Syndication Format http://www.ietf.org/rfc/rfc4287	1.0	Dec 2005
CDR-ATOM	IC/DoD Content Discovery and Retrieval Atom 1.0 Result Set Specification	1.0	March 2010
OS	OpenSearch http://www.opensearch.org/Specifications/OpenSearch/1.1/Draft_4	1.1, Draft 4	2009
OS-RELV	OpenSearch Relevance Extension http://www.opensearch.org/Specifications/OpenSearch/Extensions/Relevance/1.0	1.0, Draft 1	2007

OS-GEO	OpenSearch Geo Extension http://www.opensearch.org/Specifications/OpenSearch/Extensions/Geo/1.0/Draft_1	1.0, Draft 1	2009
OS-TIME	OpenSearch Time Extension http://www.opensearch.org/Specifications/OpenSearch/Extensions/Time/1.0/Draft_1	1.0, Draft 1	2010
SDA	Joint IC/DoD Service Discovery Architecture	-	2010
SEC-RA	DoD/IC Service-Oriented Architecture Security Reference Architecture	-	2010
DDMS	DoD Discovery Metadata Specification Implementation Guide	-	2010

431