

UNCLASSIFIED



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

***IC/DoD Content Discovery and Retrieval Reference  
Architecture***

***v1.1***

**25 February 2011**

UNCLASSIFIED

**REVISION/HISTORY**

<b>Doc Revision</b>	<b>Revised By</b>	<b>Revision Date</b>	<b>Revisions</b>
0.1	CDR IPT	24 September 2009	1 <sup>st</sup> draft for internal CDR IPT review
0.2	CDR IPT RA Subgroup	22 October 2009	Major updates across all sections
0.3	CDR IPT RA Subgroup	19 November 2009	Major updates/revisions to each section for clarification
0.4	CDR IPT RA Subgroup	16 December 2009	Updates based on 11/19 adjudicated comments. This is the pre-baseline draft.
0.5	CDR IPT	18 December 2009	Editing of document.
1.0	CDR IPT	19 December 2009	Internal Baseline
1.01	CDR IPT RA Subgroup	30 December 2010	Changes to reflect refinements to the architecture suggested by subsequent documents
1.02	CDR IPT RA Subgroup	05 January 2011	Changes to reflect addition of specification profiles and other changes from 03 JAN 2011 harmonization TEM
1.03	CDR IPT RA Subgroup	10 January 2011	Changes to reflect results of 10 JAN 2011 Harmonization TEM
1.04	CDR IPT RA Subgroup	20 January 2011	Changes to reflect the results of the 14 JAN 2011 Harmonization TEM
1.05	CDR IPT RA Subgroup	27 January 2011	Changes to reflect the results of the 24 JAN 2011 Harmonization TEM
1.06	CDR IPT RA Subgroup	04 February 2011	Changes to reflect the results of the 31 JAN 2011 Harmonization TEM
1.07	CDR IPT RA Subgroup	16 February 2011	Changes to reflect the results of the 16 FEB 2011 Harmonization TEM
1.08	Technical Edit	24 February 2011	Editing of document.
1.1	CDR IPT RA Subgroup	25 February 2011	Review of edits and incorporation of results of 22 FEB 2011 TEM

**TABLE OF CONTENTS**

<b>EXECUTIVE SUMMARY .....</b>	<b>v</b>
<b>1 Introduction.....</b>	<b>1</b>
1.1 Purpose.....	1
1.1.1 Relationship to Other CDR IPT Architecture Elements .....	1
1.1.2 Intended Use and Audience .....	4
1.2 Scope.....	4
1.3 Guiding Principles .....	4
1.4 Dependencies .....	5
1.5 Notational Convention .....	6
1.6 Document Overview .....	6
<b>2 Key Priorities.....</b>	<b>8</b>
2.1 Requirements Elicitation and Analysis .....	8
2.2 Content Discovery and Retrieval Key Priorities.....	8
2.2.1 Content Discovery .....	8
2.2.2 Retrieval.....	9
2.2.3 Mediation .....	9
<b>3 CDR Capabilities and Components .....</b>	<b>10</b>
3.1 Capability Descriptions.....	10
3.1.1 Content Discovery .....	10
3.1.2 Retrieval.....	11
3.2 Component Descriptions.....	11
3.2.1 Consumer Component .....	12
3.2.2 Provider Component .....	12
3.2.3 Core CDR Components .....	12
3.2.4 Key CDR Dependency Components .....	15
<b>4 Architecture Patterns .....</b>	<b>16</b>
4.1 Interaction Pattern 1: Search.....	17
4.2 Interaction Pattern 2a: Brokered Search .....	18
4.3 Interaction Pattern 2b: Brokered Search with Description Collection.....	19
4.4 Interaction Pattern 3: Content Retrieval .....	21
4.5 Interaction Pattern 4a: Generate Description .....	22
4.6 Interaction Pattern 4b: Generate Description for Collection .....	23
4.7 Interaction Pattern 5a: Deliver .....	24
4.8 Interaction Pattern 5b: Deliver Retrieved Resource .....	25
4.9 Interaction Pattern 6a: Save Query .....	26
4.10 Interaction Pattern 6b: Execute Saved Query .....	27
4.11 Interaction Pattern 6c: Persistent Search.....	28
<b>Appendix A – Acronyms and Abbreviations.....</b>	<b>29</b>
<b>Appendix B – Reference Documents .....</b>	<b>30</b>

**LIST OF FIGURES**

Figure 1 – CDR Architecture Model .....	2
Figure 2 – CDR RA Components .....	11
Figure 3 – CDR Interaction Pattern 1: Search .....	17
Figure 4 – CDR Interaction Pattern 2a: Brokered Search.....	18
Figure 5 – CDR Interaction Pattern 2b: Brokered Search .....	19
Figure 6 – CDR Interaction Pattern 3: Content Retrieval.....	21
Figure 7 – CDR Interaction Pattern 4a: Generate Description .....	22
Figure 8 – CDR Interaction Pattern 4b: Generate Description for Collection.....	23
Figure 9 – CDR Interaction Pattern 5a: Deliver .....	24
Figure 10 – CDR Interaction Pattern 5b: Deliver Retrieved Resource.....	25
Figure 11 – CDR Interaction Pattern 6a: Save Query.....	26
Figure 12 – CDR Interaction Pattern 6b: Execute Saved Query .....	27
Figure 13 – CDR Interaction Pattern 6c: Persistent Search.....	28

**LIST OF TABLES**

Table 1 – CDR RA Acronyms and Abbreviations.....	29
--	----

## EXECUTIVE SUMMARY

The Joint IC/DoD Content Discovery and Retrieval (CDR) Integrated Project Team (IPT) was established to develop and publish a set of architecturally driven standards and specifications to enable content discovery and retrieval from all IC and DoD data collections. The CDR Reference Architecture (RA) is the keystone artifact for the overall set of guidance artifacts. The primary content of the CDR RA is the definition of an extensible set of capabilities and components that are realized via service specifications. It also describes various architecture interaction patterns and the specific technical framework to enable robust, scalable, and repeatable implementations across varying business and mission boundaries.

The CDR RA defines two primary capabilities -- content discovery and content retrieval. Six core CDR components: Search, Brokered Search, Retrieve, Describe, Deliver, and Query Management provide these capabilities. The core components depend on externally defined Messaging and Security components, which are described in more detail in the CDR Specification Framework and CDR Specification documents. The CDR RA will iteratively evolve and mature in increments to maintain a clear scope and facilitate timely and actionable guidance.

## 1 Introduction

The Intelligence Community (IC) and Department of Defense (DoD) support an evolving and dynamic mission and business landscape, so their information systems must enable information sharing and interoperability by default. To achieve this objective, program managers, IT managers, and developers need guidance to plan, manage, model, and build interoperable information systems.

In July 2007, the DoD Chief Information Officer (CIO) and the Director of National Intelligence (DNI) CIO signed a memorandum committing the IC and DoD to a joint vision and shared oversight for realizing a common services-based environment. The key value proposition is to enable greater and more flexible information and capability sharing within and across the IC/DoD Enterprise. Under this vision, the Joint IC/DoD Content Discovery and Retrieval (CDR) Integrated Project Team (IPT) was established to develop and publish a set of architecturally-driven standards and specifications to enable content discovery and retrieval of all<sup>1</sup> IC and DoD data assets.

Organizations in the IC and DoD implement content discovery and retrieval in a variety of ways, based on their individual business and data requirements. A CDR capability must support a number of different query, content, and metadata formats, depending on the context. Rather than presenting one, monolithic schema for discovery and retrieval, the CDR RA introduces a number of concepts, components, and interactions in an abstract framework that supports flexibility.

### 1.1 Purpose

This CDR Reference Architecture (RA) serves as the keystone CDR IPT guidance artifact and describes an overall architecture to support Enterprise-wide content discovery and retrieval. The goal of the CDR RA is to establish a flexible framework for the IC/DoD environment that is extensible and scalable to support evolving mission/business requirements.

The following subsections describe this RA's relationship to other CDR IPT architectural elements, its intended uses, and its intended audience.

#### 1.1.1 Relationship to Other CDR IPT Architecture Elements

The CDR IPT 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 of the CDR IPT. The following graphic illustrates this model.

---

<sup>1</sup> The term "all" here implies that the CDR IPT will provide the technical guidance that any data provider can leverage (as identified during the creation of this document), but does not have the authority to enforce which data providers must implement the guidance.

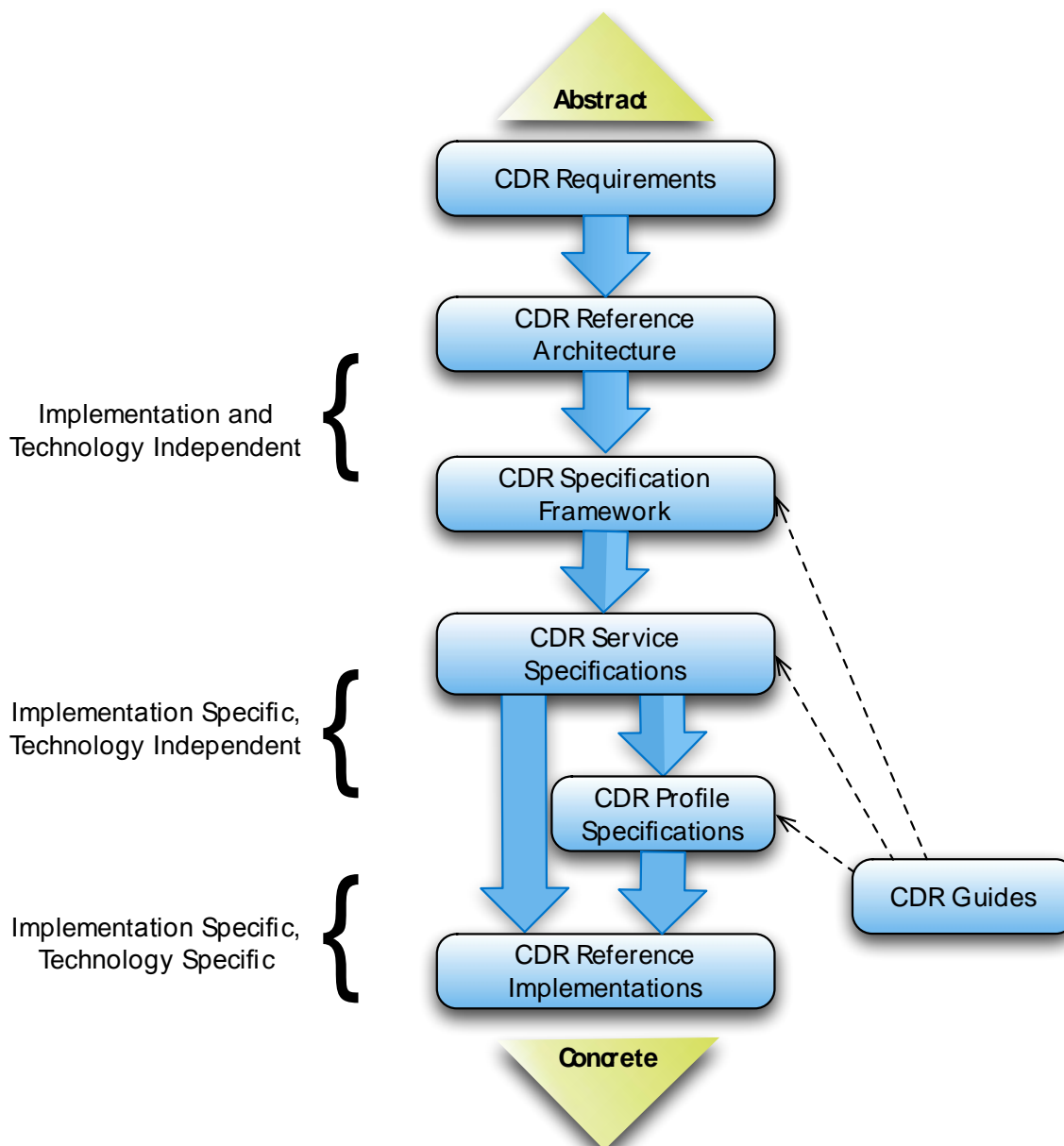


Figure 1 – CDR Architecture Model

The tiers illustrated in Figure 1 are described as follows:

- **CDR Requirements** – The critical driver of the CDR RA is the collection, analysis, and management of CDR requirements from disparate stakeholders across the IC/DoD Enterprise. Section 2 of this RA describes the requirements and the articulation of key priorities for the CDR RA. The consolidated CDR requirements are maintained in the CDR Requirements Master List.<sup>2</sup>
- **CDR Reference Architecture** – The CDR RA is the keystone artifact for the overall set of guidance artifacts. The CDR RA defines an extensible set of capabilities and components that are realized via service specifications.

<sup>2</sup> The CDR Requirements Master List is referenced in Appendix B, item 9.

Furthermore, the CDR RA describes various architecture interaction patterns and the specific technical framework to enable robust, scalable, and repeatable implementations across varying business and mission boundaries. The RA is described in terms of capabilities and components.

- **Capability** – A capability represents a logical grouping of requirements or prioritized needs. Each capability can be developed independently, as long as it follows the same constructs identified in the CDR RA. CDR Capabilities are described in Section 3 of this RA.
- **Component** – A component is an abstract, logical architectural element that is implementation and technology independent. Components identified in the CDR RA include key components to realize the capabilities and also address external dependencies (e.g., security). CDR Components are described in Section 3 of this RA.
- **CDR Specification Framework** – The CDR Specification Framework extends CDR RA. It provides the behavioral model for the CDR Components. It describes key properties of the model and decomposes key integration patterns.
- **CDR Service Specifications** – A service specification is a formal description of the interfaces, information models, and behaviors to which an implementation must conform for an abstract capability or component. A service specification is analogous to Application Programming Interface (API) documentation for software in that it provides the semantics and interfaces for a service capability. It provides service consumers with information to integrate with (i.e., consuming) a particular service, and service providers the information to implement (i.e., expose) the service. A service specification can be implemented by Commercial Off-the-Shelf (COTS) software, Government Off-the-Shelf (GOTS) software, custom software, or a combination of custom software wrappers interfacing to COTS/GOTS software, and represents the fundamental building block of services.
- **CDR Profile Specification**- A CDR Profile Specification is a formal description that establishes the compliance terms and options necessary to accomplish interoperability of CDR Service Specifications in the context of particular environments, implementation details, or policies. CDR Profile Specifications go beyond the generally applicable guidance of the CDR Service Specifications to provide the information necessary to apply them to the context addressed by the CDR Profile Specification. A single CDR Profile Specification may provide guidance on use of the interfaces defined across multiple CDR Service Specifications.
- **CDR Guides** – A CDR Guide is an informational (non-normative<sup>3</sup>) document intended to assist implementers of the specifications in effectively creating compliant implementations.
- **Reference Implementation** - A reference implementation is a concrete realization of a service specification. It serves as an example that can be used for testing consumers who wish to use the service or for providers who wish to expose the service.

---

<sup>3</sup> In standards terminology, “normative” content is considered to be a prescriptive part of the standard that must be adhered in order to be compliant. “Non-normative” content is meant to be informative, but does not require compliance.



## 1.1.2 Intended Use and Audience

This RA is intended to provide conceptual and contextual guidance to architects, engineers, and developers tasked to implement CDR specifications. The CDR RA describes key interaction patterns to guide implementers in meeting their specific needs. In this sense, the CDR RA supplements each of the CDR service specifications to provide the foundation for understanding their interactions, external dependencies, common assumptions and constraints.

## 1.2 Scope

Achieving a ubiquitous content discovery and retrieval solution presents a substantial challenge in the IC/DoD Enterprise where content exists in a large variety of structures<sup>4</sup>, are represented in diverse<sup>5</sup> semantics, and are exposed through many types of technical implementations. To meet this challenge, the CDR RA describes an architecturally driven approach for guiding the IC/DoD Enterprise towards enabling content discovery and retrieval.

Content, in the scope of this architecture, refers to the information made available for discovery and retrieval. A Content Resource refers to a particular instance of Content that can be discovered or retrieved as a unit. A Content Collection is a component that makes a set of Content Resources available under an instance of the component. By addressing Content, the CDR RA thus addresses both Content Resources and Content Collections.

## 1.3 Guiding Principles

The core guiding principles of the CDR RA are as follows:

1. **Mission/Business Driven:** CDR requirements and statements of need will provide specific drivers or inputs to the CDR RA, which must be able to scale and adapt as needed. Improved information sharing requires a strategy that addresses discovery, access, context, trust, and interoperability. With a strategic eye on realizing these areas, and as constraints allow (including technology, standards, policy, culture, and resources), the CDR RA advances iteratively based on requirements collected.
2. **FEA Data Reference Model Compliant:** The Federal Enterprise Architecture (FEA) Data Reference Model (DRM) enables information sharing and reuse via the standard description and discovery of common data and the promotion of uniform data management practices. The CDR RA is intended to be consistent with the requirements of the FEA Data Sharing standardization area.
3. **Separation of Concerns:** The CDR RA prescribes technical guidance to enable an interoperable and repeatable CDR solution across a heterogeneous IC/DoD environment, independent of policy and other technical/data standards. It concentrates on the challenges related to Content Discovery and Retrieval and

---

<sup>4</sup> Content resources and/or collections include structural attributes that effect their discoverability and accessibility.

<sup>5</sup> In this sense, diversity also refers to the large percentage of IC/DoD information resources that lack any mature data models or adherence to any IC/DoD data standards

- describes the abstract components that can work together to form specific solutions. While cognizant of standards and implementation practice, the RA is independent of specific standards and products; it provides a foundation for specifying and enforcing policies, but it is independent of particular policy mandates. In this way, the concerns of the problems to be solved are separate from the concerns of particular solution techniques and are separate from policy and guidance that evolve in response to changing situations. This creates a more durable solution space, reducing the incidence and impact of conflicts.
4. **Open Standards:** The CDR RA prescribes an open standards-based approach. Open standards provide the basis for interoperability, extensibility, agility, vendor independence, and feasibility within and across IC and DoD mission/business domains.
  5. **Iterative/Agile Realization:** The CDR RA will iteratively evolve and mature in increments to maintain a clear scope to facilitate timely and actionable guidance.
  6. **Simplicity/Ease of Implementation:** Guidance prescribed within the CDR RA is intended to be both simple to use and non-intrusive. This accomplishes the overarching goal of lowering the barrier of entry for Enterprise members to interact with and to use the CDR RA services.
  7. **Supplement Existing Information Discovery and Access Approaches:** The CDR RA defines a supplemental architecture to existing information discovery and access approaches, but does not serve as a replacement. Organizations and programs may continue to develop and utilize Web portals and proprietary access mechanisms to meet their mission objectives. To facilitate information sharing for the IC/DoD Enterprise at large, the CDR RA defines the framework and constructs (e.g., components and service specifications) for achieving a common discovery and retrieval mechanism that spans all data assets within the Enterprise.

## 1.4 Dependencies

CDR services are loosely coupled and dependent on other IC/DoD Enterprise service efforts,<sup>6</sup> including, most notably, Security and Messaging.

- **Security:** The Security focus area provides a set of security-focused services to the IC and DoD to protect access to services, data, and interactions within the IC/DoD Enterprise. Integration of Security capabilities is advocated, both from the service discovery and the service access standpoint, to protect content providers and consumers from attack from any unknown entities. Security capabilities are responsible for authentication and authorization of consumers and consumer agents and enable cross-domain search and retrieval. Furthermore, Security capabilities provide integrity, confidentiality, and audit services that CDR providers can leverage. CDR providers should reference the IC/DoD Security Reference Architecture<sup>7</sup> for guidance on utilizing the security capabilities.

---

<sup>6</sup> For more information on these specific Enterprise activities (as well as others), reference the IC/DoD Enterprise Services Engineering Review Board (ES ERB) collaboration workspaces on both the Defense Knowledge Online (DKO) and Intellipedia.

<sup>7</sup> The Service Security Reference Architecture is referenced in Appendix B, item 4.

- **Messaging:**<sup>8</sup> The Messaging focus area provides the capabilities to handle the addressing and re-routing functions necessary for applications to connect to messaging systems and to use messages to exchange data and invoke behavior. Furthermore, Messaging services provide reliable messaging capabilities for large content, guaranteed delivery, message translation, and bridging.

Apart from the specific dependencies on other infrastructure type services, CDR services have non-technical dependencies. These include:

- Policy and Governance Related Dependencies:
  - Service-level enforcement is dependent on a governance process to enforce compliance and ensure quality-of-service requirements.
  - Certification and Accreditation of CDR services must provide efficient processes for getting service implementations operational.
- COTS Tools and Vendor Dependencies
  - Across the full Service-Oriented Architecture (SOA) spectrum, the state of standards and COTS vendors vary in maturity.
- Network Dependencies
  - Performance and quality of service for CDR RA services are extremely reliant on the network infrastructure on which they are implemented.

## 1.5 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.

Throughout this RA, examples are provided that cite specific IC, DoD, Joint IC/DoD, or industry standards. These examples provide a contextual understanding of a particular statement. To keep the CDR RA at an implementation and technology independent level and not subject to configuration management as standards evolve, the standards cited do not imply implementation-level requirements. Detailed guidance may be included in implementation-level specifications derived from this RA.

## 1.6 Document Overview

The main sections of this RA are organized as follows:

**Section 1 – Introduction:** Describes the purpose, scope, guiding principles, and dependencies.

**Section 2 – Key Priorities:** Provides an overview of the requirements elicitation and analysis process, as well as an overview of the key CDR features that drive the definition of the needed capabilities.

---

<sup>8</sup> The Messaging Reference Architecture is referenced in Appendix B. item 7.

230 **Section 3 – CDR Capabilities and Components:** Provides an overview of the  
231 capabilities and components needed to enable the key priorities for content discovery  
232 and retrieval.

233 **Section 4 – Architecture Patterns:** Illustrates how the CDR components can be  
234 leveraged within different business/mission contexts.

235 **Appendix A – Acronyms and Abbreviations:** Defines the acronyms and  
236 abbreviations used in this document.

237 **Appendix B – Reference Documents:** Lists the documents used in preparation of  
238 this document.

## 2 Key Priorities

One of the primary objectives of the CDR RA is to codify an architecture framework that is scalable and flexible to support a variety of implementations suitable to the individual business and mission needs of explicit environments (e.g., COIs, domains, enclaves, etc.) within the IC/DoD Enterprise. As such, the critical driver of the CDR RA is the collection, analysis, and management of CDR requirements from disparate stakeholders across the IC/DoD Enterprise. This section of the CDR RA describes the overall requirements elicitation, analyses, and prioritization that drive the specific CDR services capabilities and framework.

### 2.1 Requirements Elicitation and Analysis

Organizations within the IC and DoD have captured and analyzed discovery and retrieval requirements that have been, or are to be, used to build their own discovery and retrieval systems. To elicit specific requirements for the greater IC/DoD CDR solution, CDR IPT representatives from the IC and DoD provided their organizations' discovery and retrieval requirements as input to the development of the Master CDR IPT Discovery and Retrieval Requirements List.<sup>9</sup>

The requirements provided by the Enterprise CDR IPT representatives addressed the full breadth of discovery and retrieval topics. Although both discovery and retrieval capabilities have dependencies on processing, ingest, messaging, and security requirements, they were not included in the master requirements list. During the requirements analysis process, it was determined that a number of requirements were common across the greater IC/DoD Enterprise, while other requirements were determined to be unique to meet a particular business or mission function. Through analysis, some of the unique requirements were determined to be applicable to the Enterprise and were included in the CDR Requirements Master List. The list of requirements was normalized to provide a broader Enterprise context. From the Master List, key Content Discovery and Retrieval Features were extracted.

### 2.2 Content Discovery and Retrieval Key Priorities

Content Discovery and Retrieval were the two high level categories initially applied against the requirements list. Requirements pertaining to metadata, query management, information sharing, ranking, and federated search were categorized as Content Discovery requirements. Retrieval requirements included delivery service and mediation and delivery management requirements. For access to the specific requirements statements, reference the CDR IPT Master Requirements List. The following subsections describe the key priorities for Content Discovery and Retrieval.

#### 2.2.1 Content Discovery

Across the IC and DoD, stakeholders have stated information systems must be searchable and discoverable. However, the ability to discover content is highly dependent on the

---

<sup>9</sup> The CDR Requirements Master List is referenced in Appendix B, item 9.

content provider's ability to expose the content and the degree of allowing access to authorized consumers.

The stakeholders desire flexibility when performing a search. The expectation is that the data will be searchable, available, and accessible. The implied requirement is to ensure the content provider exposes the data in a manner applicable to its structure and context. Stakeholders stated multiple search filter patterns should be supported and that the search parameters should be grouped and combined via logical operators.

Service consumers expressed the need for a capability to construct a query that includes instructions for when, where, and how (i.e. using what formats and data types) the results are to be delivered. Consumers have also expressed the need to build, store, edit, and share their query with other consumers.

The need to execute a query across multiple search engines and access data stores with multiple data types ranks high with service consumers. This federated or brokered search capability should inform the consumer of which data sources are available, which attributes may be searched, and the options the consumer may use to formulate the query. A key priority for brokered search includes the ability for consumers to request the delivery of results as they are available for immediate viewing and if desired, further processing. Several other key priorities include an extensible sorting/ranking mechanism upon delivery of results, as well as the ability for consumers to be able to distinguish between new results from those that have been previously reviewed. Additionally, the architecture must enable the consumer to re-sort the results upon receipt. Other desired features include: 1) results are to be de-coupled from presentation mechanisms; 2) results from multiple sources are to be aggregated upon delivery; 3) results are to be de-duplicated before delivery.

## 2.2.2 Retrieval

Many retrieval requirements contained within the CDR IPT Requirements Master List address stakeholders, including those who are well-connected, bandwidth limited or periodically disconnected. As such, the CDR architecture must account for this range of environmental conditions. The content provider should have available both synchronous and asynchronous delivery options, including support for such "use later" alternatives as FTP and caching results.

## 2.2.3 Mediation

Mediation is required to provide interoperability across divergent service implementations. Service providers and consumers may exist in different security domains, use different messaging protocols or query languages, or vary in other key aspects that prevent interoperability. Mediation is an aspect of brokered search, message delivery, and security, as defined by the CDR RA. It is a consideration for any component that must operate in a heterogeneous environment, but it is addressed separately, based upon which component of the architecture requires mediation.

## 3 CDR Capabilities and Components

As described in Section 2, the current CDR RA defines two distinct, but inter-related capabilities:

- Content Discovery
- Retrieval

### 3.1 Capability Descriptions

In addition to the CDR component interaction patterns illustrated in Section 4, the following sections provide an in-depth analysis of CDR capabilities.

#### 3.1.1 Content Discovery

Content Discovery provides consumers with the ability to discover the existence of potentially relevant content resources and content collections from across the IC/DoD Enterprise. Content Discovery also enables consumers to discover valuable metadata about content resources to support making them visible, accessible, understandable, and trusted.

To address the full range of requirements for Content Discovery, the CDR RA describes several critical assumptions to ensure the appropriate CDR components are defined. These assumptions include:

- The IC/DoD Enterprise employs multiple, independently developed content collections<sup>10</sup> that manage a large variety of content resource types. Across these content collections, common characteristics may have different implementation solutions and include:
  - The interface or interfaces for accepting queries
  - The supported query syntax types
  - The content collection type (e.g., database, document repository, etc.)
  - The metadata format describing the content collection
  - The types of content resources managed by the content collection (e.g., database records, imagery files, etc.)
  - The content format/structure for the content resources
  - The subject of the content
- Content consumers may submit a general search request to ascertain what information may be available from across the Enterprise based on a given set of search criteria, or content consumers may perform much more specific/refined searches based on specific attributes of content collections and/or content resources.
- Queries are expected to be save-able, discoverable, and reusable, both manually and automatically, through subscription.

---

<sup>10</sup> Content collections and content resources are provider components described in Section 3.2.2

### 3.1.2 Retrieval

Retrieval provides consumers the ability to retrieve and route specific content resources. To do so, content providers are to ensure their content is accessible to the Enterprise.

Additionally a consumer may submit a query from one location but have the content delivered at a specified time and/or to a different location. For this reason, the Retrieval capability should support both direct and indirect retrieval of content.

## 3.2 Component Descriptions

Components within the CDR RA constitute implementation and technology-independent concepts. Components may be used to define a logical set of properties, functions, or behaviors. Some components may lead to the realization of a specific service implementation, while other components may describe key characteristics or inform specific operations that are realized across many service implementations. The components are assumed to be composable, such that the components can be selected independently or assembled in combination to satisfy specific requirements

The CDR RA identifies four main component types including:

- Consumer Component
- Provider Component
- Core CDR Components
- Key CDR Dependency Components

The following diagram illustrates the CDR RA components.

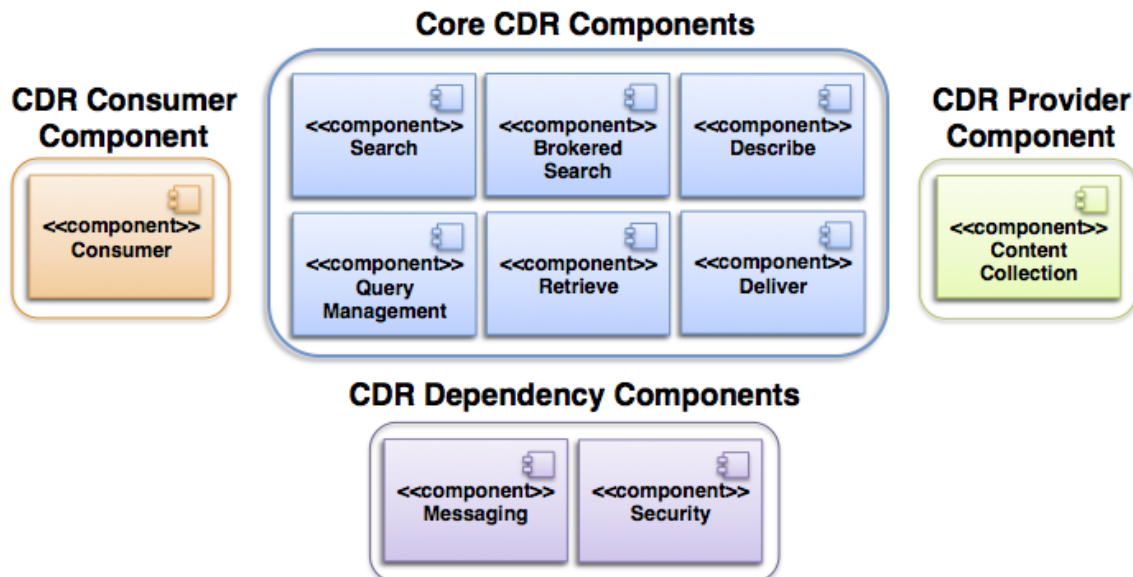


Figure 2 – CDR RA Components

The following subsections define each of these component types. Section 4 of this RA illustrates some key component interaction patterns.



### 3.2.1 Consumer Component

The Consumer Component represents any entity that initiates a content discovery or content retrieval interaction with any of the Core CDR Components. Existing CDR Components can serve as Consumer Components, allowing for service composability.

It is important to note that Consumer Components may be *anticipated* consumers (e.g., those resident on IC/DoD networks and infrastructures) or *unanticipated* consumers (e.g., U.S. and coalition partners). The Security Component (one of the Key Dependency Components described in Section 3.2.4.1) provides the necessary mechanism(s) to ensure only authorized consumers are able to perform the requested function(s).

### 3.2.2 Provider Component

The Provider Component describes the applicable construct related to the providers of content to support content discovery and retrieval use cases. It is embodied in the Content Collection Component, which represents a data source that holds one or more content resources. In this sense, a Content Collection Component may be a database, a document repository, an imagery repository, an authoring system, etc.

Content Collection Components are generally characterized by exposing certain information (e.g., metadata), including, but not limited to, the name of the content collection, the organization or element that serves as steward for the managed content, and a characterization of the type of content resources that are stored in the collection. For example, if a Content Collection Component was a database, the characterization might include a logical model for the database, in which the terms are expressed in COI-coordinated vocabulary and common representations.

A Content Resource is an individual data resource that is being made visible and accessible. Examples of Content Resources include a PDF file, a database record, a satellite image, a web page, etc. Information (i.e., metadata) about these resources may be used by Consumer Components to further inform the Consumer whether the resource should be accessed and what transformations may be needed to optimize its utility. For a specific instance, this can include the size of the resource, when it was created or last edited, its geospatial coverage scope, its subject coverage scope, whether it is releasable, and any alternate formats (e.g., MIME types) that might be available. To address conformity and broad adoption of these types of metadata-related concerns, the IC and DoD are engaged in joint activities to develop, promulgate, and enforce metadata standards.<sup>11</sup>

### 3.2.3 Core CDR Components

The Core CDR Components define a fundamental set of functionality that directly addresses specific requirements that enable Content Discovery and Retrieval capabilities. The following subsections provide a common interface and behavioral model for each core component: Search, Brokered Search, Describe, Query Management, Retrieve, and

---

<sup>11</sup> For example, relating to specific resource metadata, the DoD has adopted the Information Resource Metadata Technical Encoding Specification (IRM.XML) and DoD Discovery Metadata Specification (DDMS), respectively.

Deliver. Additionally, the CDR Specification Framework augments the CDR RA by elaborating on how these components interact to enable CDR capabilities. Neither the RA nor the Specification Framework address the implementation of the component, but, rather, focus on providing an overview of the collected interfaces provided by the components.

The CDR Components are intended to interact and build upon each other in a composable fashion. Composability is the property of the components that allows them to be combined in multiple ways to achieve more complex objectives. Many of the following component descriptions include non-exhaustive lists of examples of composability.

### **3.2.3.1 Search Component**

The Search Component serves as the primary content discovery mechanism to expose content collections for discovery and accessibility, enabling content consumers to discover disparate collections and relevant content resources from collections across the IC/DoD Enterprise. Specifically, the Search Component provides a means to accept requests that have a uniform syntax and semantics that can be transformed, as needed, and applied to newly-developed or existing content collections, unambiguously conveying a query without knowing or setting requirements on the implementation of the underlying content collections.

A Search Component generally does not return the actual content resources, but rather, metadata about the content resources, in the form of search results, contained in the response. The Search Component is intended to return information to help a consumer select which resources are to be retrieved. The Retrieve Component is typically used to return resources. However, the CDR RA does not rule out the possibility that the Search Component may return the entire content resources when it is practical to do so.

### **3.2.3.2 Brokered Search Component**

The Brokered Search Component serves as the primary mechanism to 1) facilitate the distribution of queries to applicable/relevant content collections (exposed as Search Components) and 2) process the returned results. To satisfy the basic “federated search” use case, a Consumer Component submits a query request to the Brokered Search Component, which then distributes the query to the applicable Search Components. The Search Components interpret the search request as best they can, compile a list of search results<sup>12</sup> and send them back to the Brokered Search Component.

In addition to the high-level functions of the Brokered Search Component described above, the Brokered Search Component may also perform other, more detailed functions in support of federated search-related processes. These may include, for example, the ability for the consumer to ascertain details regarding which content collections were queried, including any sources that were unavailable (if any).

---

<sup>12</sup> Results should be encoded in documented format, e.g. the DoD supported format of Atom augmented with DDMS.

### 3.2.3.3 Describe Component

The Describe Component serves as the primary mechanism for content collectors (i.e., providers) to expose information to describe the context, access constraints, and current inventory status of the underlying content resources to support static and dynamic discovery and accessibility of a content collection. Search and Brokered Search leverage the output of this component to determine whether the content collection may contain content resources that are relevant to the consumer's query. To support a wide array of use cases, the Describe Component should reflect both the static<sup>13</sup> and dynamic<sup>14</sup> information about the underlying content collection.

### 3.2.3.4 Query Management Component

The Query Management Component provides a coordinated set of activities that enable service consumers to store, retrieve, modify, delete, and execute Persistent Searches. Persistent Searches are searches that are saved into a data storage collection and can then run in the future to provide updated results to consumers. This component uses the interfaces defined by the Search Component, the Brokered Search Component, the Retrieve Component, the Deliver Component, and the Describe Component to create a comprehensive interface definition based on the composable services those components, and other components external to CDR, provide.

### 3.2.3.5 Retrieve Component

The Retrieve Component enables access to discovered content and is the primary mechanism for content consumers to access one or more specific content resources from a content collection. The content is typically addressed by a unique resource identifier. The Retrieve Component can return content to the requesting consumer or, in concert with the Deliver Component, can redirect the content to other consumers.

### 3.2.3.6 Deliver Component

The Deliver Component routes, processes, and conditionally delivers content, search results, or metadata to receiving consumers. This component provides a common interface for initiating consumers to request delivery and to specify the associated delivery properties. The Deliver Component may provide or coordinate additional processing services such as compression, encryption, or format translation. The Deliver Component may also utilize the Retrieve Component to enable initiating consumers to request delivery of content without providing the content themselves. The Deliver Component also enables conditional delivery, based on specified time or event criteria.

---

<sup>13</sup> Static information includes, but is not limited to, features of the exchange schemas (both discovery and content) that have been engineered into the collection. For example, the Provider may declare that it has implemented the DDMS 2.0 schema for discovery metadata, and that all metacards will reference "Unclassified" or "Secret" data resources.

<sup>14</sup> Dynamic information includes, but is not limited to, features of the exchange schemas (both discovery and content) that are present in the current collection. Issuing queries against the content collection produces this information. For example, the Provider may declare that for the DDMS 2.0 schema that has been implemented, and there are a total of 1200 metacards (instances of the DDMS schema).

### 3.2.4 Key CDR Dependency Components

These abstract components represent the key external dependencies needed to realize the CDR architecture. Specific guidance around the realization and implementation of these component types are out of scope for the CDR RA, but they are described here to illustrate the high-level dependencies between Core CDR Components and other applicable Enterprise service activities.

#### 3.2.4.1 Security Component

Each of the CDR Components depends upon and interacts with the Security Component to provide the authentication, authorization, access control, integrity, confidentiality, and non-repudiation functions necessary to ensure that the interaction patterns described by the CDR RA satisfy security policies. Interfaces between Security Components such as policy decision points, policy enforcement points, and the Core CDR Components are necessary to ensure that only authorized consumers gain access to content and metadata. For example, when a consumer searches for content collections, the Search Component requires that the Security Component be able to build and exchange consumer authentication and authorization data to determine if the consumer can discover specific metadata about a given content collection, and then enforces this determination. Furthermore, this function is used to determine if the searching consumer can have access to specific content resources and to enforce this determination. It is expected that the Security Component RA will provide guidance for implementers of the CDR components that identifies interface points for requesting security services. In some cases, (e.g. cross domain search and retrieval), the CDR specifications must be augmented to address processing queries, query results, retrieve requests, or retrieved content to ensure that they can be inspected by cross domain solutions, in accordance with applicable cross domain policies.

Specifically, the Core CDR Components depend upon and interface with the Security Component to:

- Provide a framework for the exchange of authentication credentials and authorization information for systems and people.
- Protect the confidentiality and integrity of the content exchanged between consumers and providers.
- Enable cross-domain content discovery and retrieval.
- Enforce access and policy rules to allow anticipated and unanticipated consumers to be associated with search requests so that authorization decisions can be made in a distributed manner.
- Audit security relevant decisions made during content search and retrieval functions.

#### 3.2.4.2 Messaging Component

The Messaging Component provides support for the transport of search requests and query results. Furthermore, the Messaging Component provides reliable messaging capabilities, guaranteed delivery, message translation, and bridging. The definitions of elements of CDR systems that use messaging are provided by the Messaging Component, and not by the CDR component that relies on messaging. This allows the CDR

architecture to be defined as independent from the underlying communication and transport layers. The Messaging Component assumes the existence of a reliable communication network layer upon which its services can run.

## 4 Architecture Patterns

The following architecture patterns describe how the components in the CDR RA may be used in various contexts. Using this information, a Consumer or Content Collection providers may determine which patterns work best to meet their needs. These patterns are not an exhaustive list of all possible integrations; however, they provide several example usages that build upon the basic interaction patterns to demonstrate the components' composability.

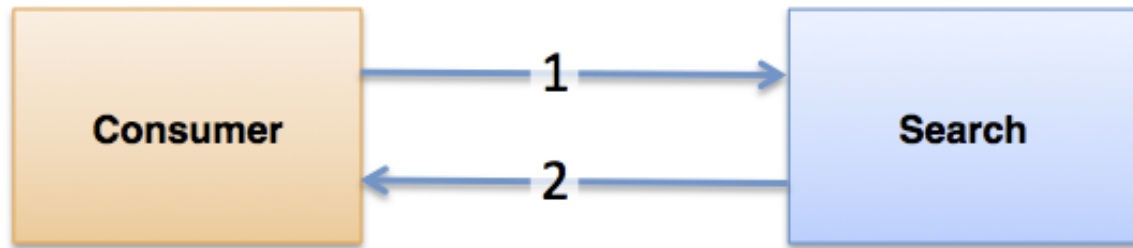
The current CDR RA discusses core interactions that are used as a baseline for more advanced architecture patterns. For each pattern, the interactions among the components are shown as a sequence of steps, with each step listing any corresponding key architectural information that must be conveyed by that interaction. The key architectural information includes the formats used to exchange data via requests and responses between components. For example, a Search Component receiving a query must understand the query type being sent and the semantics of the query parameters being used. While the target values assigned to the query parameters convey the current needs of the requester, these values are part of the usage of the RA and are not an architectural concern.

The specific details of what happens within the bounds of Core CDR Dependency Components are not discussed in this section, though references to these patterns are provided. For instance, security concerns crosscut nearly every aspect of the following interaction patterns. However, these concerns are excluded from the diagrams in the interest of clarity and to focus on the core interactions between components introduced in this document or those external components that are not crosscutting in nature. This does not imply that security is not pertinent to these interactions.<sup>15</sup>

In addition, the interactions between components assume that a common communication mechanism has been established. The definition and specification of the methods by which the components establish communications is not in the scope of this architecture. The components, patterns, and interactions described are specifically designed to be flexibly independent of the communication channel.

---

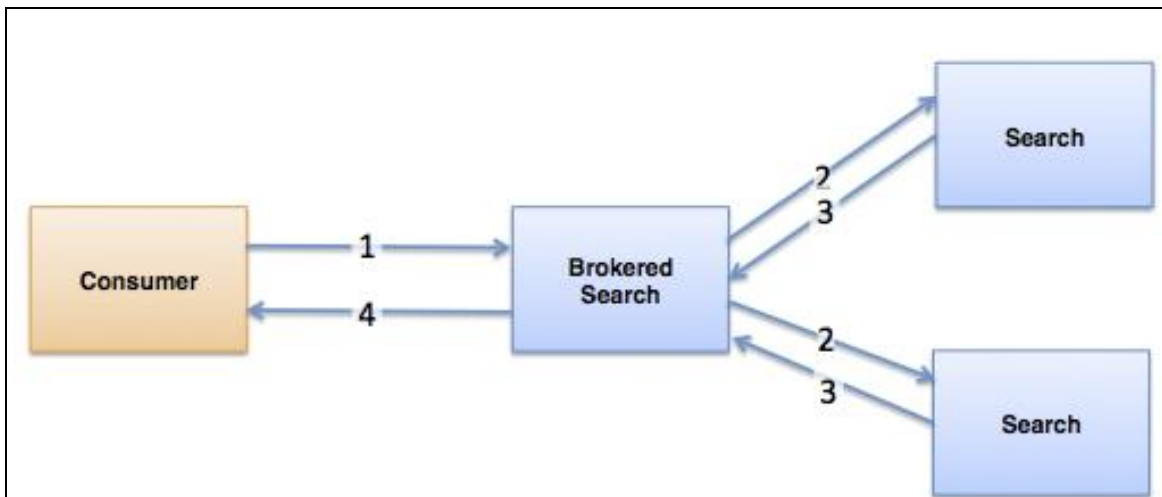
<sup>15</sup> For a full discussion on security patterns, implications, and details, please refer to the Service Security Reference Architecture, referenced in Appendix B item #4

579 **4.1 Interaction Pattern 1: Search****Figure 3 – CDR Interaction Pattern 1: Search**

Step	Description	Key Architectural Information
1	A Consumer Component sends a search request to a Search Component.	Query information, including well-defined query type and query parameters
2	The Search Component performs the search <sup>16</sup> and return metadata describing the data resources that fit the criteria outlined by the search request to the Consumer Component.	Format used in returning metadata results

580

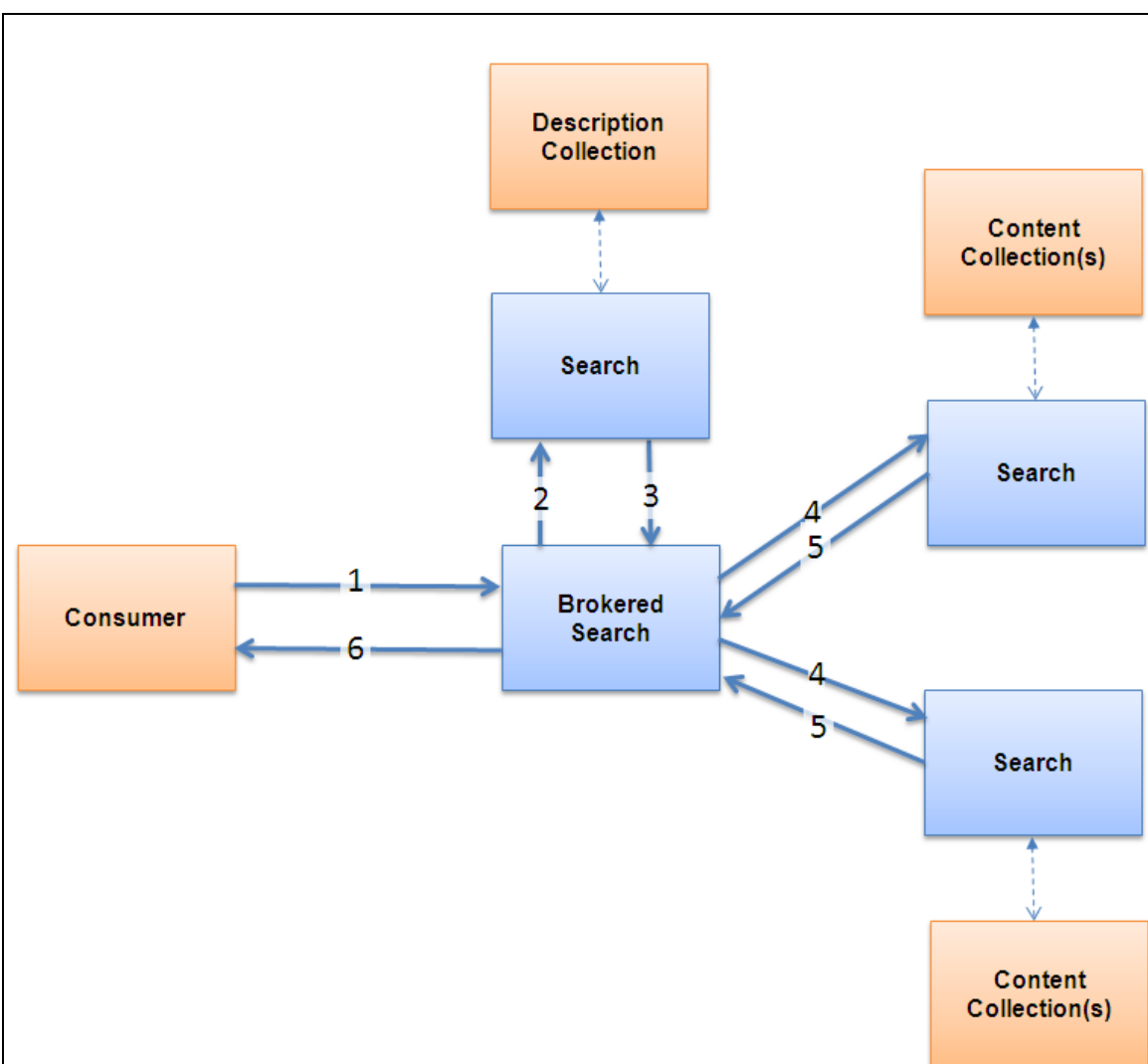
<sup>16</sup> In this case, the Search Component could be a search engine that understands CDR operations and performs the Search directly on an associated index or an adapter translating CDR operations to an existing search engine. The key element of interest in this interaction is the definition of the interface that is provided by the Search Component.

581 **4.2 Interaction Pattern 2a: Brokered Search****Figure 4 – CDR Interaction Pattern 2a: Brokered Search**

Step	Description	Key Architectural Information
1	A Consumer Component sends a search request to a Brokered Search Component.	Query information, including well-defined query type and query parameters
2	The Brokered Search Component distributes the search request to the appropriate Search Components.	Query information, including well-defined query type and query parameters
3	The Search Components perform the search, and return metadata describing matching data to the Brokered Search Component.	Format used in returning metadata results
4	The Brokered Search Component returns the processed results from the invoked Search Components to the Consumer Component.	Format used in returning metadata results, including identifier of associated Search Component

582

583 **4.3 Interaction Pattern 2b: Brokered Search with Description**  
 584 **Collection**



**Figure 5 – CDR Interaction Pattern 2b: Brokered Search**

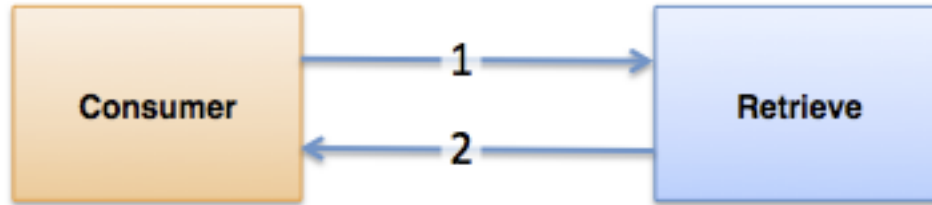
Step	Description	Key Architectural Information
1	A Consumer Component sends a search request to a Brokered Search Component.	Query information, including well-defined query type and query parameters
2	The Brokered Search Component sends a request to a Search Component that searches a Description Collection to request description information for relevant resource collections .	Format of request used to request information from a Description Collection
3	The Description Collection returns relevant description information to the Brokered Search Component.	Format used in returning description information



4	The Brokered Search Component distributes the search request to the appropriate Search Components based on the information obtained from the Description Collection.	Query information, including well-defined query type and query parameters
5	The Search Components perform the search, and return metadata describing matching data to the Brokered Search Component.	Format used in returning metadata results
6	The Brokered Search Component returns the processed results from the invoked Search Components to the Consumer Component.	Format used in returning metadata results, including identifier of associated Search Component

585

586

587 **4.4 Interaction Pattern 3: Content Retrieval****Figure 6 – CDR Interaction Pattern 3: Content Retrieval**

Step	Description	Key Architectural Information
1	A Consumer Component sends a retrieve request to a Retrieve Component.	Resource Identifier
2	The Retrieve Component executes the retrieve request and returns the Content Resource identified by Resource Identifier to the Consumer Component.	Format used in returning retrieved Resource

588

589

#### 590 4.5 Interaction Pattern 4a: Generate Description

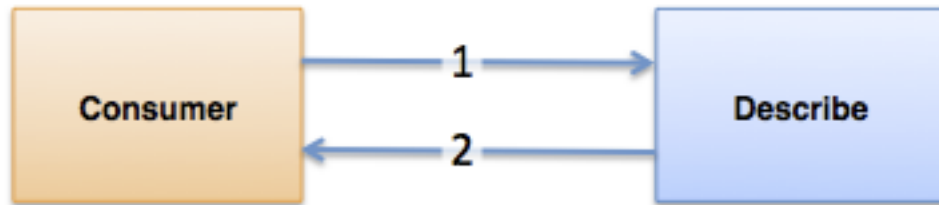


Figure 7 – CDR Interaction Pattern 4a: Generate Description

Step	Description	Key Architectural Information
1	A Consumer Component sends a describe request to a Describe Component.	Information needed to generate a description
2	The Describe Component generates a description and returns it to the Consumer Component.	Format used in returning generated description

591

592

#### 4.6 Interaction Pattern 4b: Generate Description for Collection

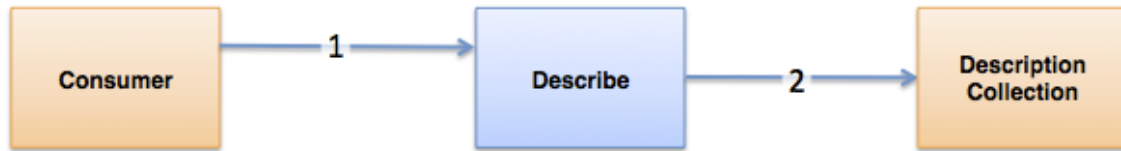
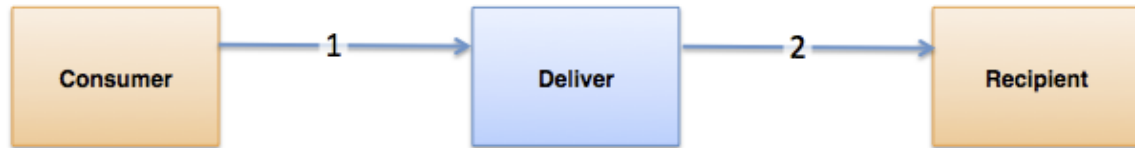


Figure 8 – CDR Interaction Pattern 4b: Generate Description for Collection

Step	Description	Key Architectural Information
1	A Consumer Component sends a describe request to a Describe Component.	Information needed to generate a description
2	The Describe Component generates a description and sends it to a Description Collection.	Format used in sending generated description

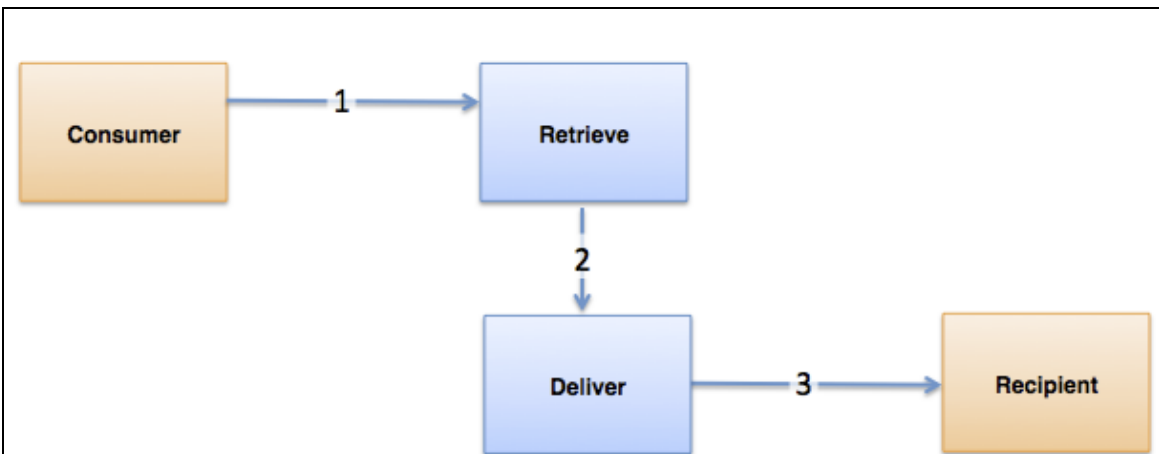
596

597 **4.7 Interaction Pattern 5a: Deliver****Figure 9 – CDR Interaction Pattern 5a: Deliver**

Step	Description	Key Architectural Information
1	A Consumer Component sends a deliver request for a resource to a Deliver Component.	Resource to be delivered, delivery destination
2	The Deliver Component delivers the resource to Recipient.	Method used by Recipient to receive resource

598

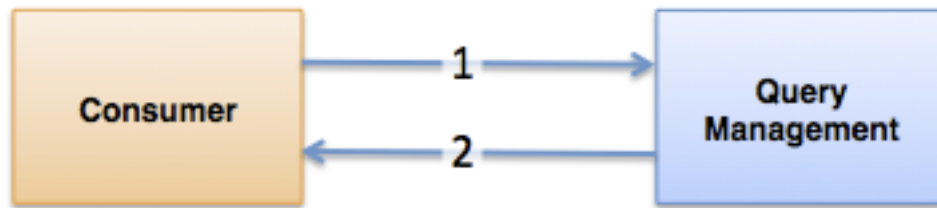
599

**4.8 Interaction Pattern 5b: Deliver Retrieved Resource****Figure 10 – CDR Interaction Pattern 5b: Deliver Retrieved Resource**

Step	Description	Key Architectural Information
1	A Consumer Component sends a retrieve request for a resource identifier with Deliver Properties to a Retrieve Component.	Format of Deliver Properties and Resource Identifier for resource to be delivered.
2	The Retrieve Component retrieves the resource and sends it to the Deliver Component with the Deliver Properties.	Format of retrieved Resource, format of Deliver Properties
3	The Deliver Component delivers the resource to Recipient Consumer Component.	Method used by the recipient component to receive deliveries, format of delivered resource

600

601

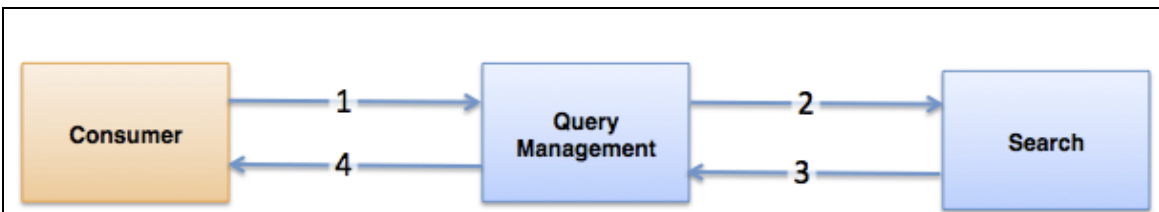
602 **4.9 Interaction Pattern 6a: Save Query****Figure 11 – CDR Interaction Pattern 6a: Save Query**

Step	Description	Key Architectural Information
1	A Consumer Component sends a Persistent Search to the Query Management Component.	Format of data to be saved and communications
2	The Query Management Component saves the Persistent Search and returns a resource identifier to the Consumer Component.	Structure of resource identifier

603

604

605

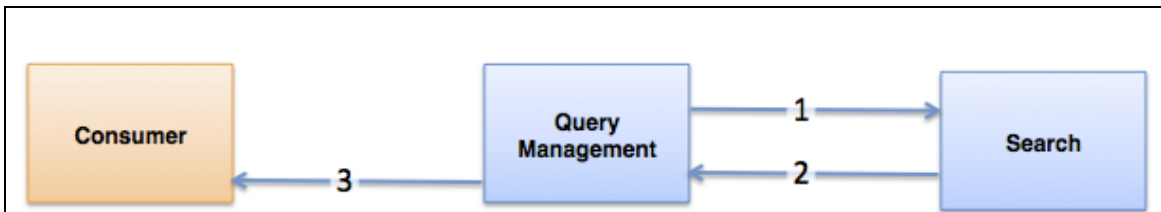
**4.10 Interaction Pattern 6b: Execute Saved Query****Figure 12 – CDR Interaction Pattern 6b: Execute Saved Query**

Step	Description	Key Architectural Information
1	A Consumer Component makes a request to invoke the Persistent Search by providing resource identifier of the Persistent Search to be invoked to the Query Management Component.	Message exchange standards between core architectural components
2	The Query Management Component retrieves the Persistent Search and sends a Search Request to the Search Component.	Format used to convey the search request to the Search Component
3	A Search Component executes the Search and returns the results to the Query Management Component. <sup>17</sup>	Format used to return search results
4	Query Management Component delivers the search results to the recipient Consumer Component.	Format used in returning metadata results

606

<sup>17</sup> This activity could be configured to have the search component deliver the results directly to the consumer, rather than through the Query Management component, if the implementation of the Search Component has that capability.



607 **4.11 Interaction Pattern 6c: Persistent Search****Figure 13 – CDR Interaction Pattern 6c: Persistent Search**

Step	Description	Key Architectural Information
1	The Query Management Component may use a Search Component to execute the Persistent Search on a schedule.	Format used to execute a search request
2	The Search Component may return search results to the Query Management Component <sup>18</sup> when delivery conditions are met.	Format used in returning search results
3	The Query Management Component may send updated search results containing new, changed, or deleted search results to the Consumer Component.	Format used in returning search results

608

609

<sup>18</sup> This activity could be configured to have the search component deliver the results directly to the consumer, rather than through the Query Management component, if the implementation of the Search Component has that capability

## Appendix A – Acronyms and Abbreviations

Table 1 – CDR RA Acronyms and Abbreviations

Acronym	Definition
CDR	Content Discovery and Retrieval
COI	Community Of Interest
COTS	Commercial-Off-The-Shelf
DDMS	DoD Discovery Metadata Specification
DNI	Director of National Intelligence
DoD	Department of Defense
FTP	File Transfer Protocol
GOTS	Government-Off-The-Shelf
IC	Intelligence Community
IETF	Internet Engineering Task Force
IPT	Integrated Project Team
MIME	Multipurpose Internet Mail Extensions
OASIS	Organization for the Advancement of Structured Information Standards
PDF	Portable Document Format
RA	Reference Architecture
RFC	Request For Comment
SOA	Service-Oriented Architecture
XML	Extensible Markup Language

## Appendix B – Reference Documents

The following references were used in the development of the CDR Reference Architecture:

1. IC Services Strategy, v3.2, February 2008
2. DoD Net-centric Services Strategy, v1.0, February 2007
3. OASIS SOA Reference Model, v1.0, October 2006
4. Joint IC/DoD Security Reference Architecture, v1.2, July 2008 (v2.0 set for release in early 2010)
5. Internet Engineering Task Force (IETF) RFC 2119, March 1997
6. IC Service Discovery Reference Architecture, v1.2, September 2007 (Joint IC/DoD v1.0 set for release in early 2010)
7. IC SOA Messaging Reference Architecture, v1.1, December 2006
8. Joint IC/DoD Memorandum of Agreement for Services-Based Environment, July 2007
9. CDR Requirements Master List, October 2009
10. IC/DoD Unified Authorization and Attribute Service, Concept of Operations, v1.11, December 2008