

Data Fabric Interest Group Goals

Focusing on the data creation and consumption circle as it happens daily in the scientific and industrial labs and on the identification of ways to make this work more efficiently and thus more cost-effective. Identifying Core Components and defining their characteristics and services that can be combined and used across boundaries to solve a variety of data scenarios.

Brokering Interest Group Goals

Providing a cross-disciplinary, global forum for data providers, cyberinfrastructure developers, and data users to discuss short and long term steps that could be taken to make data more available and interoperable through the services of brokering frameworks.

Definitions

A **broker** is a middleware component in a cyberinfrastructure that connects resources by mediating interactions between those resources without requiring either party to adapt new conventions or install software components¹. A **brokering framework** can consist of multiple brokering components to support capabilities such as discovery, semantic and natural language mediation, data access services, workflow processes, and publishing.

A **PID System** consists of at least one PID Resolver, a name schema and a defined mechanism for issuing PIDs that conform to the name schema.

PID Resolution is the process of resolving a PID to useful state information about a DO by using a globally available system.

A **PID Information Type** is a set of metadata, in the form of type value/pairs, returned by resolving a PID.

A **Collection** is a digital object which is identified by a PID and consists of a set or a list of PIDs/Ids and a set of additional pointers/links and metadata together with each PID/Id.

A **Data Type Registry** (DTR) is a registry that records the implicit details, such as structure, of data in the form of Data Types and associates those Types (and links various data types) with the executable data processing functions that can be useful for working with a specific data type.

Source for above definitions: RDA Data Foundation and Terminology Wiki
<http://smw-rda.esc.rzg.mpg.de/index.php>

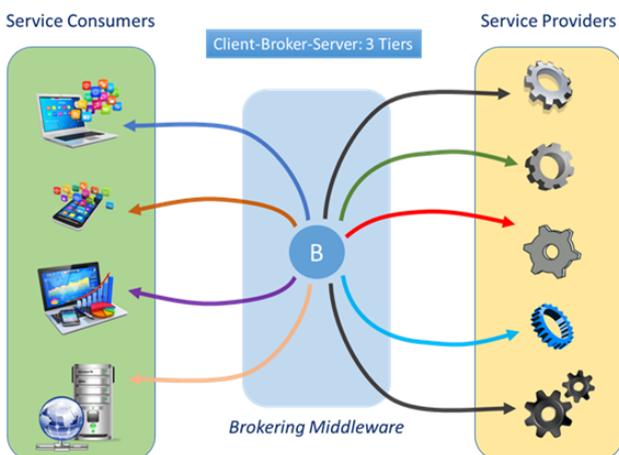
Core Components and Roles

The **PIT Registry** defines standard core PID information types to enable simplified verification of data identity and integrity.
(*Status: RDA Recommendation*)

The **Data Type Registry** enables data producers to classify their data sets in standard data types, allowing data users to automatically identify instruments to process and visualise the data.
(*Status: RDA Recommendation*)

The **Collections API** is a common API for management operations (e.g. Create, Read, Update, Delete, List) on a collections of of data objects.
(*Status: RDA Working Group in Progress*)

A **Broker** translates and harmonizes the Collections API functionality across disciplines.



Source: Stefano Nativi

¹ Source: SiriJodha Khalsa, CIRES University of Colorado „Brokering for Multi-Disciplinary Interoperability: An Earth Cube Perspective.“

Sample Use Case:

A Data Consumer wants to search multiple repositories for individual data objects and collections of data objects of a specific type, normalize and transform these data objects according to domain-specific needs and then create a new Collection of data objects from the results. The original data objects may have been created by multiple projects across different domains, but are all referenced by PIDs with associated PID Information Type metadata that identifies the data types of the underlying data objects using registered types. The role of broker in this workflow is that of a controller, driving the workflow, interacting with Collection Services, PIT Registries, Data Type Registries and Data Providers, using standard APIs, to query for, retrieve, filter and transform data.

