

CV Data Publishing

Aside from the curation of scientific data, a research infrastructure must provide means to access that data. Access can be provided in a number of ways, including the export of curated datasets and the querying of data catalogues. Beyond the actual mechanism of access however are the issues of discovery and interpretation. Specific datasets may be found via citation (the publication of persistent identifiers associated with data) or by browsing data catalogues (permitting queries over multiple datasets). Additionally, a functionality to allow identifying the location of specific datasets in data stores should exist. It should also be possible to identify the ontologies, taxonomies and other semantic metadata associated with datasets or data requests and provide some form of mapping between representations as necessary.

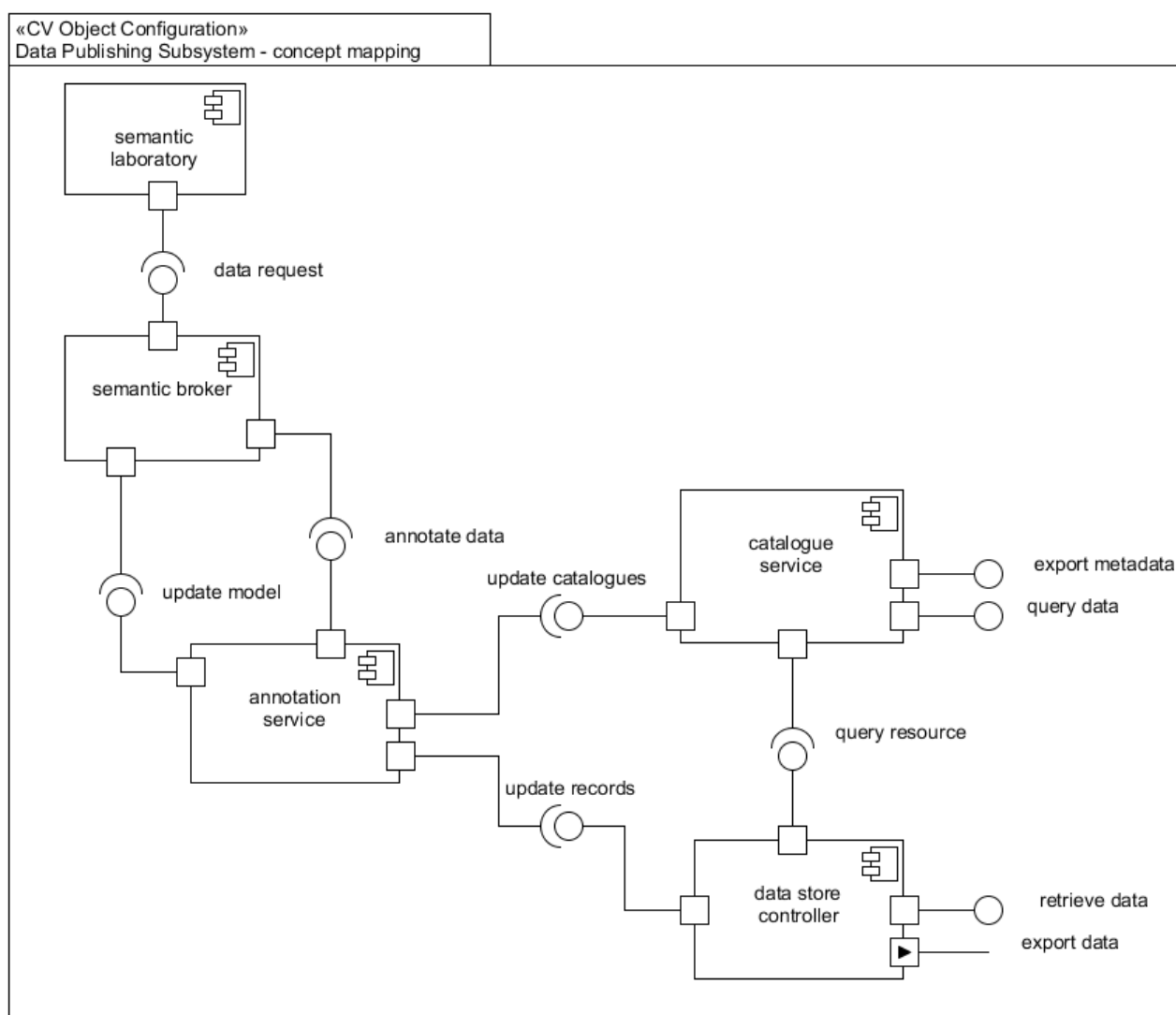
The data publishing objects provide **CV Broker Objects** which mediate between data stores and catalogues and presentation objects (**virtual laboratories**). **data broker** act as intermediaries for access to data held within the data store objects supporting **CV Data Curation**. **Semantic brokers** enable semantic interpretation. Brokers are responsible for verifying the agents making access requests and for validating those requests prior to sending them on to the relevant data curation service.

The following examples present two important groups of functionalities provided by the data publishing subsystem: concept mapping and data publishing.

Concept Mapping

The **semantic laboratory** facilitates three activities which support linking data and metadata to one or more global models: (1) Build Global Conceptual Model, (2) Setup Mapping Rule, and (3) Perform Mapping. The **semantic broker** will facilitate updating the data and internal concept model to preserve the mappings by invoking the **catalogue service** and the **data store controller**.

Data Publishing Subsystem - concept mapping

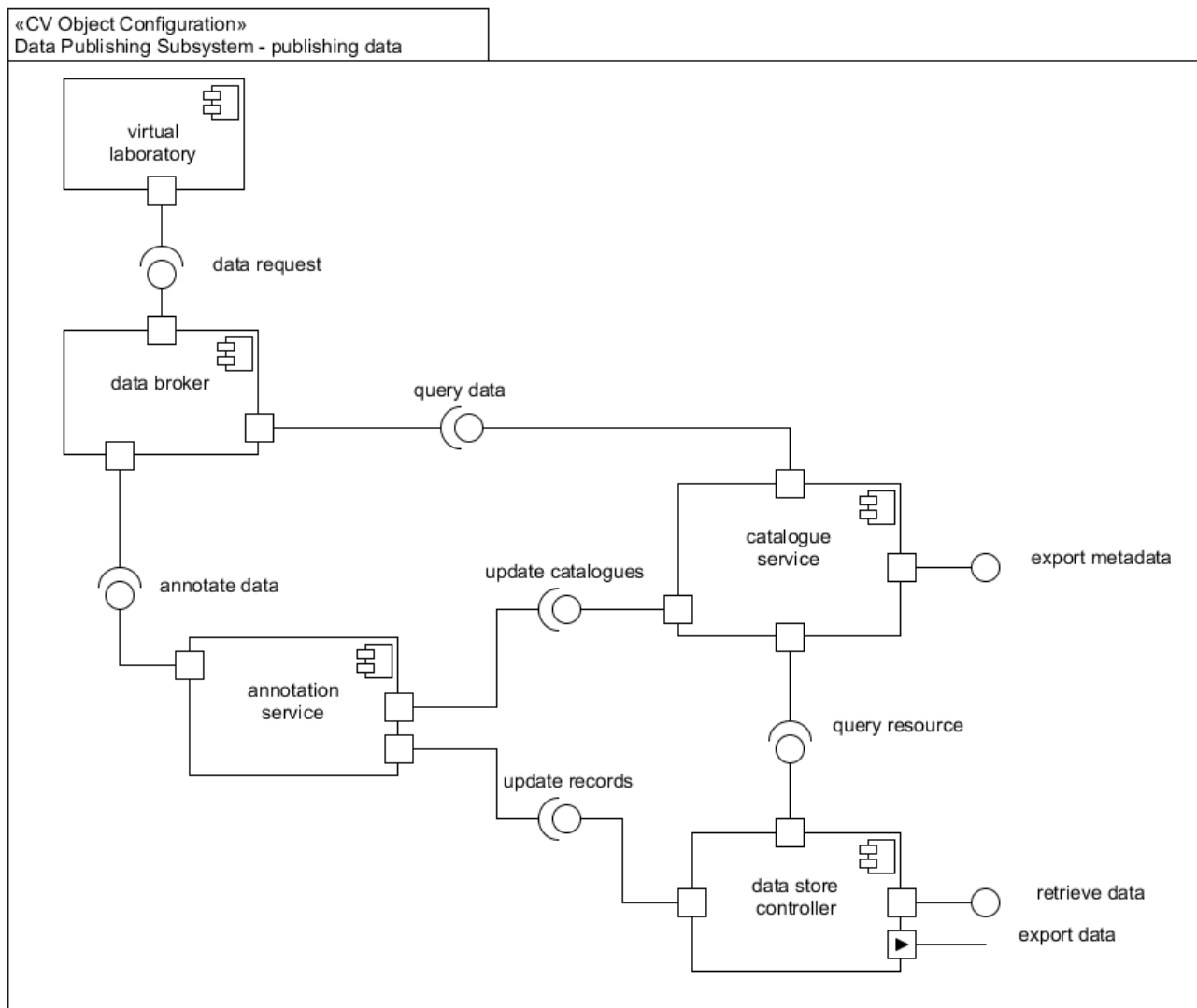


Notation

Publishing Data and Metadata

The **virtual laboratory** facilitates the activities which support publishing data and metadata: (1) reviewing the data to be published, (2) publishing data, and (3) publishing metadata. The **data broker** will facilitate updating the data and metadata catalogues to establish that the data has been reviewed, as well as storing the new links that make data and metadata publicly accessible. These actions are performed by invoking the appropriate **catalogue service** and the **data store controller**.

Data Publishing Subsystem - publishing data



Notation