

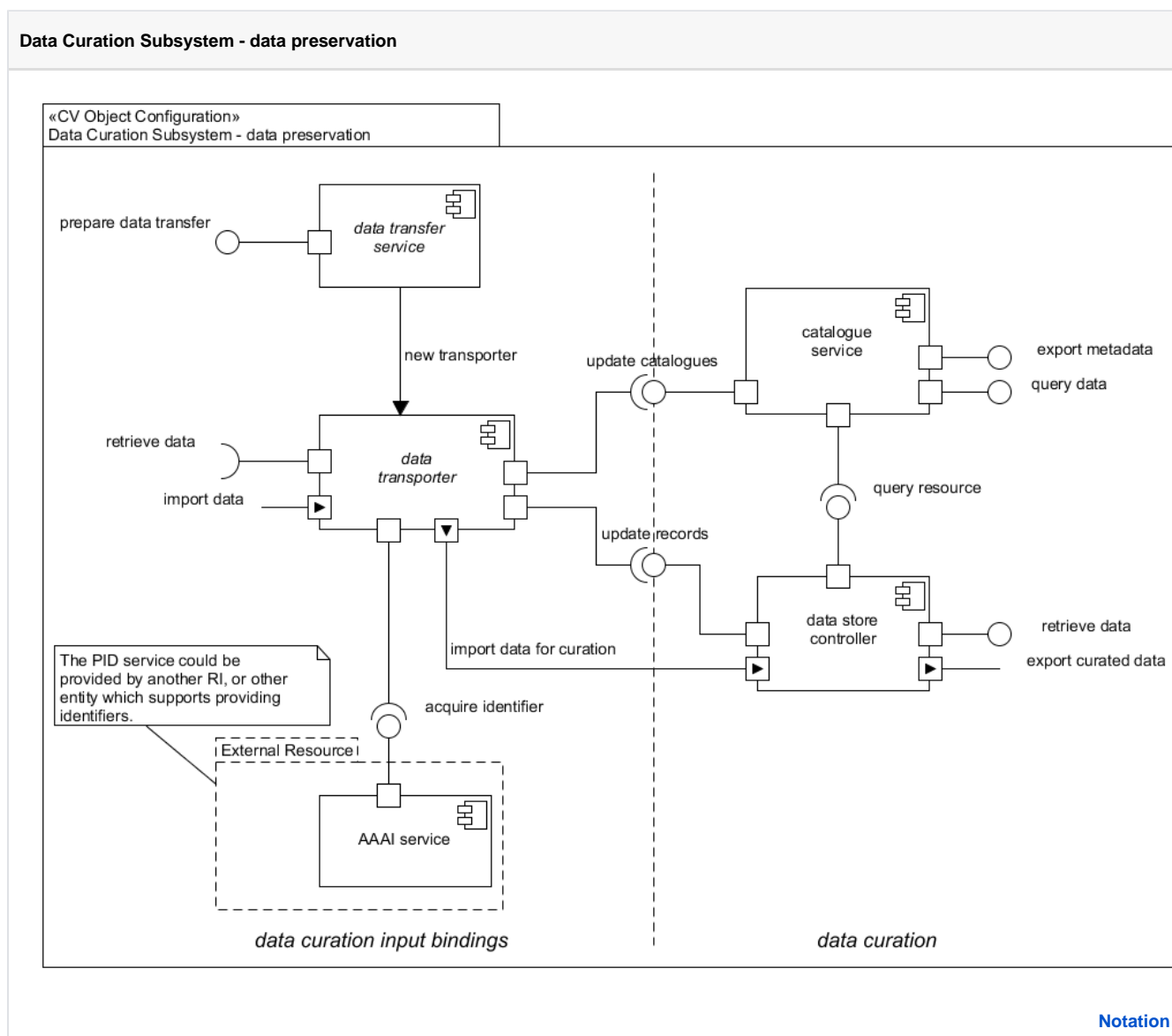
# CV Data Curation

One of the primary responsibilities of an environmental research infrastructure is the curation of the significant corpus of acquired data and derived results harvested from the data acquisition phase of the data lifecycle, data processing and community contributions. Scientific data must be collected, catalogued and made accessible to all authorised users. The accessibility requirement in particular dictates that infrastructures provide facilities to ensure easy availability of data, generally by replication (for optimised retrieval and failure-tolerance), publishing of persistent identifiers (to aid discovery) and cataloguing (aiding discovery and allowing more sophisticated requests to be made over the entirety of curated data). The following examples present two of the main functionalities of the data curation subsystem: data preservation and data annotation.

## Data Preservation

The diagram shows the organisation of five CV objects which participate in the preservation of research data. The **data transporter** in the diagram could be replaced by a **raw data collector** or a **data importer** object, and the change would not affect the integrity of the system. Consequently, this configuration supports both types of data acquisition described in the data acquisition subsystem section. In the example there are no **Presentation Objects** which implies that the data preservation process is automated.

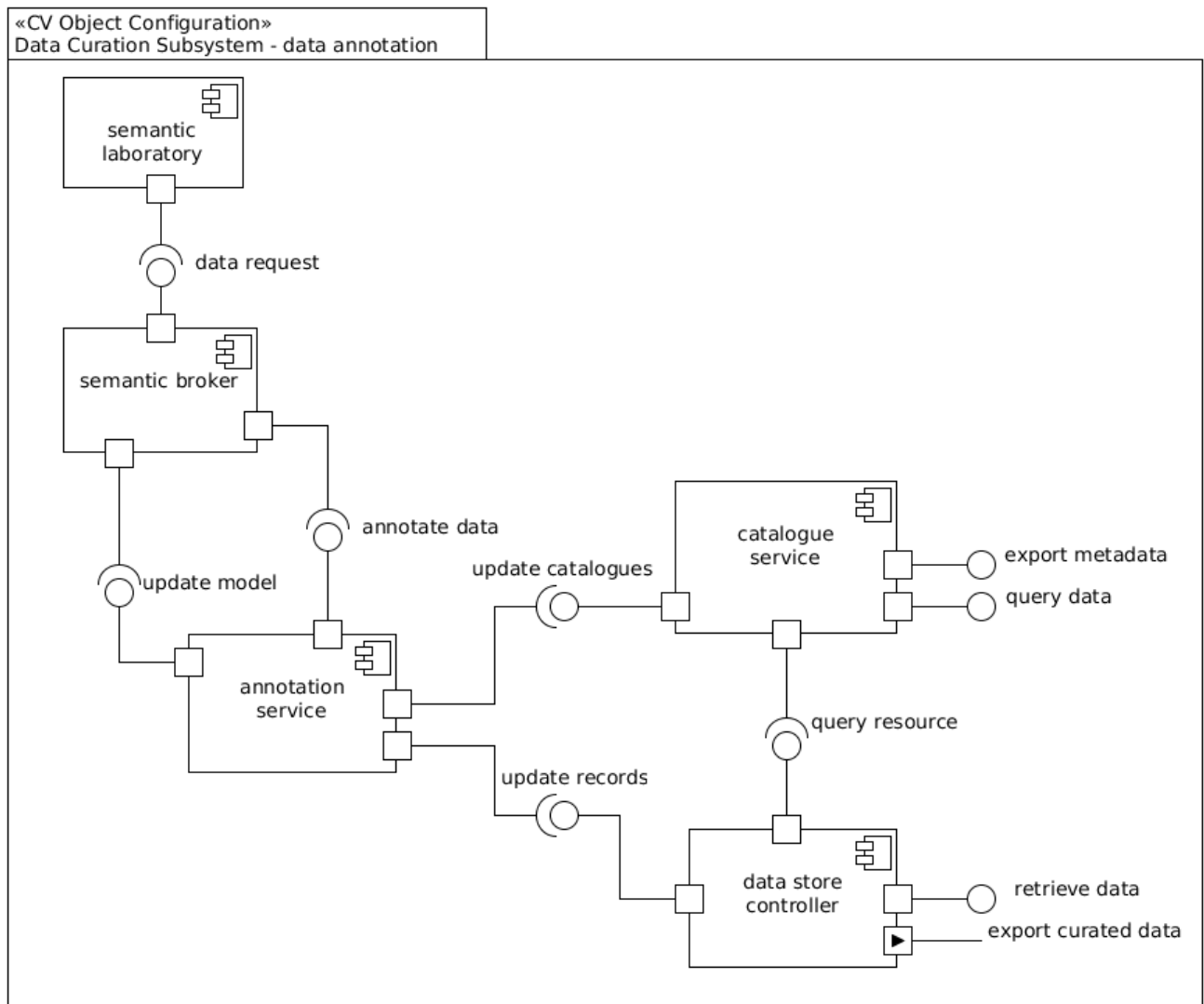
The **data transporter** modelled in the diagram is a complex device which at the same time invokes the **PID service**, the **catalogue service**, and the **data store controller**. The **PID service** is invoked to acquire a unique identifier for the incoming data set. The **catalogue service** is invoked to store the metadata associated with the incoming data set. The **data store controller** is invoked to store the incoming data set, along with its persistent identifier and linked to its associated metadata.



## Data annotation

The diagram shows the organisation of five CV objects which participate in the annotation of research data. This task is carried with the oversight of a user or on request from a user, this is why the presentation object **semantic laboratory** is included. The **semantic laboratory** invokes a **semantic broker** which in turn invokes the **annotation service**. The **annotation service** provides two functionalities annotation and updating of the conceptual model, both the annotation and conceptual model are special types of metadata which are stored in the RI's catalogues and linked to a specific dataset, for this the annotation service invokes the **catalogue service** and the **data store controller**.

#### Data Curation Subsystem - data annotation



Notation