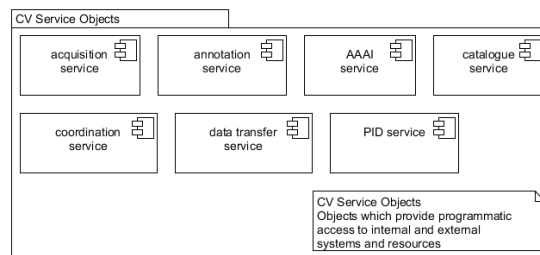


CV Service Objects

CV service objects offer programmatic access to distributed systems and resources (internal and external). This allows building RIs using both internal and external sourced components. The service layer includes the main services that enable data access, processing and transformation used in different phases of the research data lifecycle.

- The **acquisition services**, responsible for ensuring that any data is delivered into the infrastructure in accordance with current policies.
- The **annotation service**, concerned with the updating of records (such as datasets) and catalogues in response to user annotation requests.
- The **AAAI service** handles authorisation requests and authentication of users before they can proceed with any privileged activities.
- The **catalogue service**, concerned with the cataloguing of metadata and other characteristic data associated with datasets stored within the infrastructure.
- The **coordination service** delegates all processing tasks sent to particular execution resources, coordinates multi-stage workflows and initiates execution.
- The **data transfer service**, concerned with the movement of data into and out of the infrastructure.
- The **PID service** provides globally-readable persistent identifiers (PIDs) to infrastructure entities, mainly datasets, that may be cited by the community.

CV Service Objects

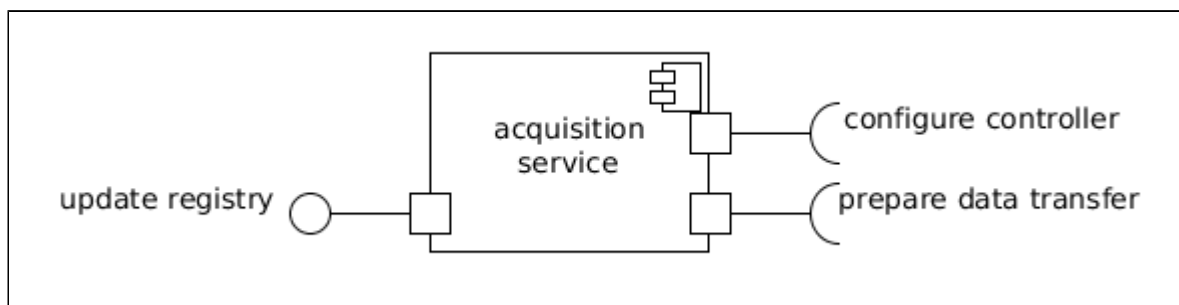


Notation

Acquisition service

Oversight service for integrated data acquisition.

An acquisition service object encapsulates the computational functions required to monitor and



manage a network of instruments. An acquisition service can translate acquisition requests into sets of individual instrument configuration operations as appropriate.

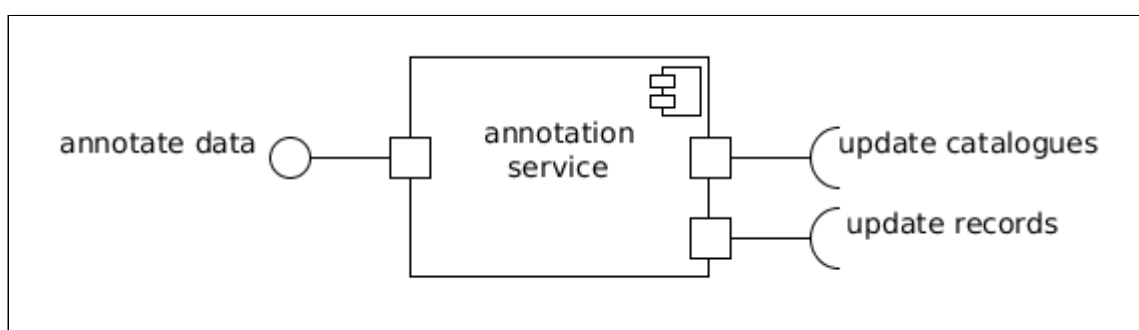
An acquisition service should provide at least three operational interfaces:

- **update registry (server)** provides functions for registering and deregistering instruments within the data acquisition phase.
- **configure controller (client)** is used to configure data collection (and other configurable factors) on individual instruments.
- **prepare data transfer (client)** is used to negotiate data transfers to data curation objects.

Annotation service

Oversight service for adding and updating records attached to curated datasets.

An annotation service object collects the functions required to annotate datasets and collect observations that can be associated with the various types of data managed within a research infrastructure.



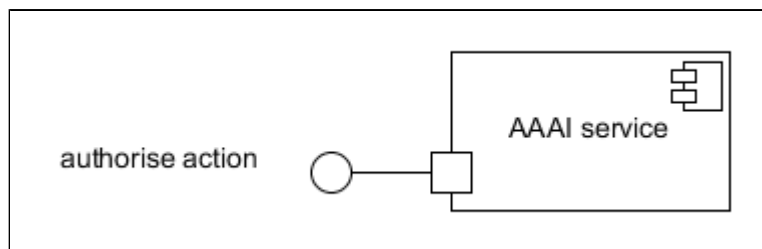
An annotation service should provide three operational interfaces:

- **annotate data (server)** provides functions for requesting the annotation of existing datasets or the creation of additional records (such as qualitative observations made by researchers).
- **update catalogues (client)** is used to update catalogues or catalogue information managed by a catalogue service.
- **update records (client)** is used to update annotation records of existing datasets curated within one or more data stores.

AAAI service

Oversight service for authentication, authorisation, and accounting of user requests to the infrastructure.

An AAAI service object encapsulates the functions required to authenticate agents, authorise any requests they make to services within a research infrastructure, and track their actions. Generally, any interaction occurring via a science gateway object or a virtual laboratory object will only proceed after a suitable transaction with an AAAI service object has been made.



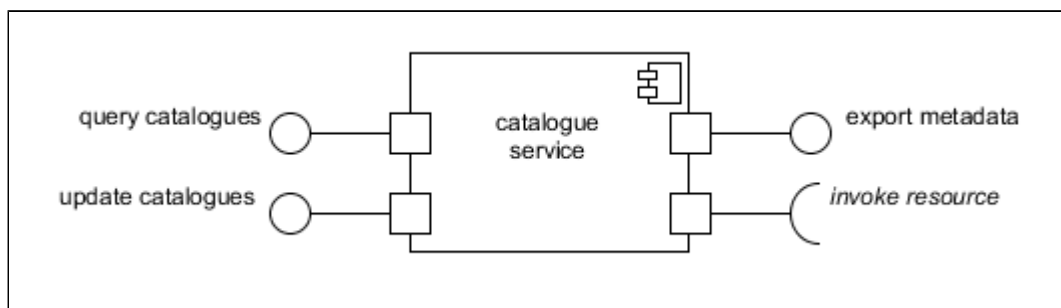
An AAAI service should provide at least one operational interface:

- **authorise action (server)** provides functions to verify and validate proposed actions, providing authorisation tokens (for example) where required

Catalogue service

Oversight service for cataloguing curated datasets.

A catalogue service object collects the functions required to manage the construction and maintenance of catalogues of metadata or other characteristic data associated with datasets (including provenance and persistent identifiers) stored within data stores registered.



A catalogue service should provide four operational interfaces:

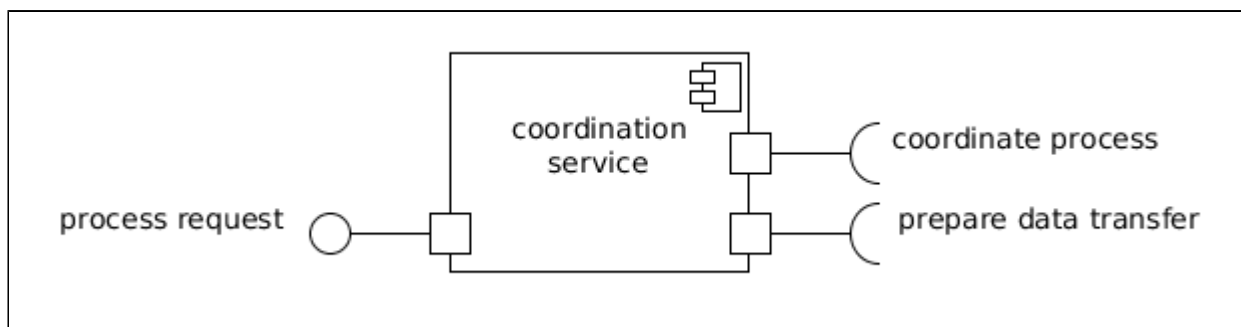
- **export metadata (server)** provides functions for gathering metadata to be exported with datasets extracted from the data curation store objects (data stores).
- **query catalogues (server)** provides functions for querying data held by the infrastructure, including the retrieval of datasets associated with a given persistent identifier.
- **update catalogues (server)** provided to update data catalogues metadata and their associated data assets.
- **invoke resource (client)** a generic interface provided for enabling the invocation of other services such as harvesting, exporting data, or automated update. This includes the communication with internal components such as the data store controller for retrieving data.



A data catalogue is itself a dataset, and can therefore be accessed and queried exactly as any other dataset.

Coordination service

Over sight service for data processes



g tasks deployed on infrastructure execution resources.

A coordination service should provide at least three operational interfaces:

- **process request (server)** provides functions for scheduling the execution of data processing tasks. This could require executing complex workflows involving many (parallel) sub-tasks.
- **coordinate process (client)** is used to coordinate the execution of data processing tasks on execution resources presented by process controllers.
- **prepare data transfer (client)** is used to move data into and out of the data store objects in order to register new results or in preparation for the generation of such results.

Data transfer service

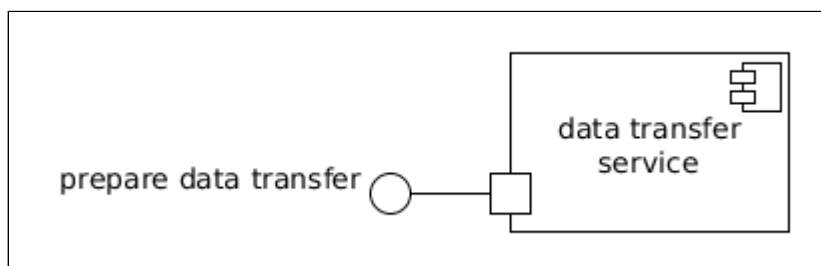
Oversight service for the transfer of data into and out of the data store objects.

A data transfer service object encapsulates the functions required to integrate new data into the RI and export that integrated data on demand. The data transfer service is responsible for setting up data transfers, including any repackaging of datasets necessary prior to delivery.

A data transfer object can create any number of new **data transporter** objects.

A data transfer service should provide one operational interface:

- **prepare data transfer (server)** provides functions for negotiating and scheduling a data transfer either into or out of the data stores of a RI.



The actual coordination of data transfers is handled by data transporter objects; the data transfer service is responsible for specifying the behaviour of a given transporter.

PID service

External service for persistent identifier assignment and resolution.

Persistent identifiers are

generated by a global service generally provided by an outside entity supported by the research community. A PID (persistent identifier) service object encapsulates this service and is responsible for providing identifiers for all entities that require them.

A PID service should provide at least two operational interfaces:

- **acquire identifier (server)** provides a persistent identifier for a given entity.
- **resolve identifier (server)** resolves identifiers, referring agents to the identified entity (in practice a science gateway providing access to the entity).



Different versions of artefacts, where maintained separately, are assumed to have different identifiers, but those identifiers can share a common root such that the family of versions of a given artefact can be retrieved in one transaction, or only the most recent (or otherwise dominant) version is returned.

