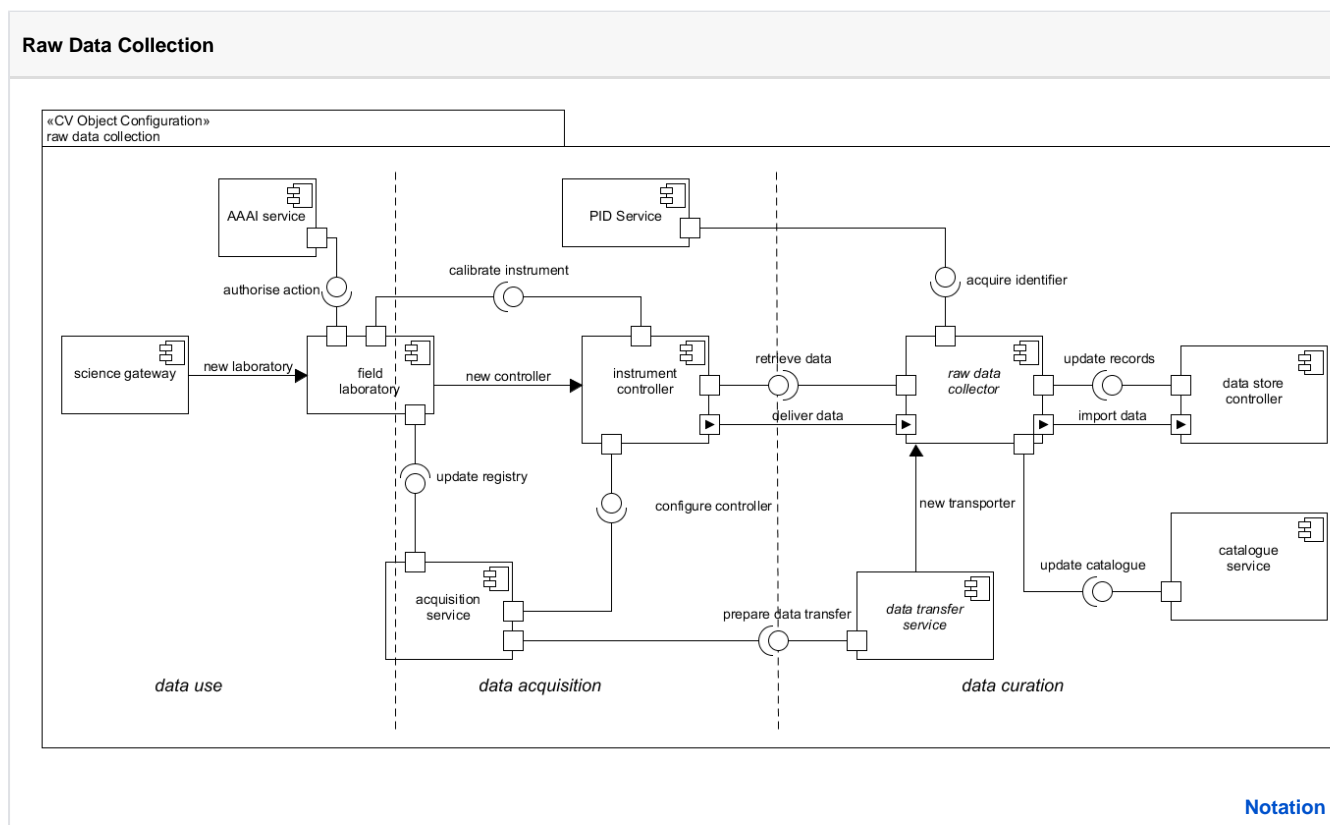


CV Raw Data Collection

The collection of raw scientific data requires coordination between the [CV Data Acquisition](#) phase (which extracts the raw data from instruments) and the [CV Data Curation](#) phase (which packages and stores the data).



The delivery of raw data into a research infrastructure is driven by collaboration between an [acquisition service](#) and a [data transfer service](#). This process can be configured using a [field laboratory](#) subject to an [AAAI service](#) authorisation, via the [AAAI service](#)'s *authorise action* interface. Regardless, the acquisition service identifies the instruments that act as data sources and provides information on their output behaviour, whilst the data transfer service provides a [data transporter](#) that can establish (multiple, persistent) data channels between instruments and data stores. The data transporter (a [raw data collector](#)) can initiate data transfer by requesting data from one or more [instrument controller](#) and preparing one or more [data store controller](#) to receive the data.

The raw data collector is considered responsible for packaging any raw data obtained into a format suitable for curation - this may entail chunking data streams, assigning persistent identifiers and associating metadata to the resulting datasets. To assist in this, a raw data collector may acquire identifiers from a [PID service](#). It may also want to register the presence of new data and any immediately apparent data characteristics in infrastructure data catalogues - this is done by invoking an update operation on the [catalogue service](#).