

IV Lifecycle Overview

This section describes the alignment between data processing in the RI systems and the data lifecycle using **IV Information Objects** and **IV Information Action Types**. The description is framed against the phases of the **Model Overview**.

The diagram shown on the right provides a high level view of the data lifecycle. The rounded rectangles represent IV actions on data and the straight rectangles represent instances of IV objects at different states. The arrow lines link IV actions and IV objects as follows: arrows leaving an action connect to IV objects created by the action while arrows entering an action connect IV objects to actions applied on them. The black circle at the top of the diagram represents the starting point and the double circle at the bottom represents the end point. The types of diagrams used in this section are called activity diagrams (UML).



In the diagram each phase of the data lifecycle is represented as an action which produces a specific information object, in this case the main information object shown is persistent data. The diagram also adds a provenance tracking action. Provenance tracking is an action that can proceed in parallel during all phases of the data lifecycle. The overview of the data lifecycle phases is described as follows.

Data Acquisition: The data acquisition phase encompasses the actions defined for the observation/experimentation, storage, identification and storage of measurements/observations (raw data). In the diagram, the acquisition phase is represented by the "DataAcquisition" action which produces a measurement result data object with the state raw.

Data Curation: The data curation phase encompasses the actions that support the long term preservation and use of research data. The main product of this set of actions is persistent data in a stable state (curated data). In the diagram, the curation phase is represented by the "DataCuration" action which produces a persistent data object with the state curated.



Note

Data curation includes preservation which may require data transformation, for example media migration to a digital form.

Data Publishing: The data publishing phase encompasses the actions that guaranty data access and discovery for entities (people and systems) outside the RI. In the diagram, the publishing phase is represented by the "DataPublishing" action which produces a persistent data object with the state published.

Data Processing: The data processing phase encompasses the actions that support making use of the RI published data. In the diagram, the processing phase is represented by the "DataProcessing" action which produces a persistent data object with the state processed.

Data Use: The data use phase is a bridge phase which sits between processing and acquisition. In this phase, the data is used and may produce new data (raw data) which can in turn be persisted by an RI. In the diagram the usage phase is represented by the "DataUse" action which produces a data product object with the state raw.

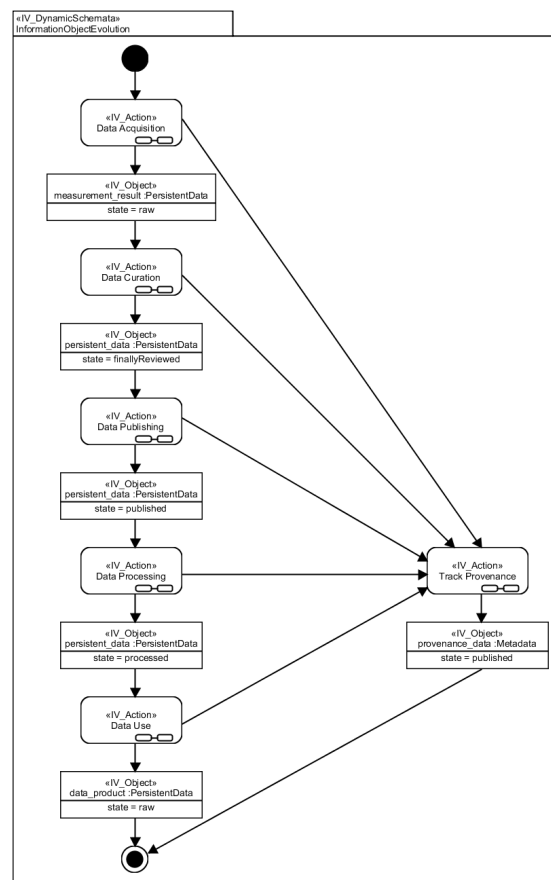
In the **IV Lifecycle in Detail** section, the actions in the diagram are expanded to present a more detailed view of the data lifecycle from the IV perspective.

Data Provenance Tracking

It is important to track state changes of information objects during their lifecycle. As illustrated in diagram above, the ProvenanceTracking action takes place in parallel to the phases of the lifecycle that change the state of persistent data.

Some of the states changes of information objects as effects of actions are summarised in the following table. As shown in the diagram, the outputs of each transition in which a new stable state is reached can be used to produce provenance data. For example, a provenance tracking service may record information objects being processed, action types applied and resulting objects, the timestamps for the actions, and some additional data and store that as provenance data.

Information Object Lifecycle



Notation

Simplified example of some provenance tracking points

Information Object	Applied Action Types	Resulting Information Objects
	Data Acquisition	persistent data (raw)
persistent data (raw)	Data Curation	persistent data (finallyReviewed) metadata (registered)
persistent data (FinallyReviewed) metadata(registered)	Data Publishing	persistent data (published) metadata (published)
persistent data (published)	Data Processing	persistent data (processed)
persistent data (processed)	Data Use	data product (new form of persistent data (raw))

The citation of data referencing the actors of involved in production of the data is an example of the use of data provenance

Correct interpretation of the data can also depend on reviewing the provenance, for instance to ensure origin of the data matches its intended use.