Use Case Description and Specification

Annotated Use Case Model Diagrams

[Represent the work flow as a diagram showing the users’ interaction with the system components.]

Diagrams

Draw the pipeline as a graph of data nodes (as follows) connected by data flows.

Nodes:

- **Data Object**: The types of files/data store objects/... that are in a data repository
- **Data ingest**: The original source of data where data objects are produced.
- **Data repository**: A data repository has the capability to hold different types of data objects for a certain (possibly infinite) amount of time. A data repository is able to receive and provide data objects.
- **Processing station**: A processing station consumes data objects, processes them and typically produces new data objects, which are moved into data repositories.

Flows:

- **Data transport**: A data transport entity describes the connection between any of the previous entities for moving data.

![Diagram](image)

Figure 1: Example diagram.

Annotations

Characterize the nodes as follows:

<table>
<thead>
<tr>
<th>Data object</th>
<th>• General description of what data is stored</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Formats</td>
</tr>
<tr>
<td></td>
<td>• Metadata</td>
</tr>
<tr>
<td></td>
<td>• Database requirements</td>
</tr>
<tr>
<td>Data transport</td>
<td>• General description of what data is transported</td>
</tr>
<tr>
<td></td>
<td>• Maximum required bandwidth</td>
</tr>
<tr>
<td></td>
<td>• Average required bandwidth</td>
</tr>
</tbody>
</table>
- Data access patterns (request rate, transfer sizes)
- Interface requirements for attached entities

### Data ingest
- Description of input data source
- Description of data introduction (upload? scanner characteristics? simulation characteristics?)
- Characteristics of data: formats, loads, bandwidths, latencies, transports

### Data repository
- Classification of the data objects (see below)
- Access control requirements
- Access requirements
- Maximum and average capacity requirements
  - In case of repository for permanent data objects, i.e. repositories where data is accumulated, provide maximum capacity requirement as function over time.
  - In terms of size & file number
- Data availability requirements

### Processing station
- General description of data processing
  - Typical processing steps
  - Number of processing steps
- Data processing hardware architecture requirements
- Required software stacks (libraries, software frameworks etc.)
  - Version requirements and dependencies
  - Need for licences
- Ratio of data processing rate versus data consumption and production rate
  - In terms of injection rates, burst or constant, variability, availability, bandwidth and latency:
  - Data consumption access pattern
  - Data production access pattern

### Classifications
Classify data objects as follows:
- Transient (Temporary): Data discarded on simulation completion or when later processing steps are concluded.
- Short-term (Campaign): Data used throughout the execution of the scientific workflow.
- Permanent (Forever): Data outliving the machine used to generate it.

### Workflow Description
Describe the Use Case workflow in words, i.e. what the user does in the course of executing the workflow? What is the story?

### Current solution(s)
Give a detailed description of current and planned implementations of the workflow

### Goal(s)
What are the use cases’ requirements? What is the problem you are trying to solve with this pipeline?