1.1 Assumptions

This Functional Specification is written under the following assumptions.
1. The application programmer is presented with a set of Services that allows to trigger aggregated blocks of Data Interoperability functionality. Among these are Services to trigger Workflow Actions, Services to retrieve status information, and others.
2. The actual work for these steps is done by a set of Workflow Actions.
3. The Workflow Actions are executed as part of a workflow that is regarded as a Data Interoperability workflow, meaning that it mainly executes Workflow Actions from the Data Interoperability component and is set up orchestrate the Workflow Actions needed to perform the Services.

These assumptions indicate a division of the Data Interoperability component into two parts, that are not strongly coupled. The Workflow Actions exists independently from the Services.

1.2 Concepts

- *Ingestion*: The DI component provides Services that deal with external sources for objects that are to be stored in the eSciDoc system. This process is called ingestion.
- *Ingestion Source*: Information about external sources from which objects are retrieved, their supported protocols and the data formats they provide are called ingestion sources.
- *Ingestion Definition*: The concrete criteria which control which objects are to be retrieved and the rules how they are validated and stored is called ingestion definition.
- *Full Text Files*: Retrieved objects may include references to external files that contain the full text of the entities described as metadata in the objects. These external files are called full text files.

1.3 Services

For the first release, the Data Interoperability component (DI) offers Services in the following areas.

1.3.1 Ingestion

The Workflow Actions for an Ingestion definition are executed as part of a workflow. For Release 1 all ingestion definitions are used by a predefined workflow that can not be admistered. The workflow is triggered manually.
The Services provided by the DI component represent the workflow steps that can be triggered. Some steps may require user interaction (e. g. getting the user's approval to permanently store retrieved objects in the eSciDoc system). Services are provided to react to the resultst of this user interaction by triggering the workflow to continue and execute the appropriate Workflow Actions.
The Services provided can be grouped in two groups.
1. Triggering the retrieval of objects or full text files and storing them permanently in the eSciDoc system.

2. Getting information about ingestion sources, ingestion definitions, executed workflows, and their respective status.

The Services from the first group allow the application developer to trigger the retrieval of objects for an ingestion definition, to trigger the retrieval of referenced full text files, and to trigger the permanent storing of retrieved objects and full text files, possibly after manual approval of objects and full text files.

The Services from the second group allow the the application developer to get information about the definition of ingestion sources, ingestion definitions, and workflows. They also make available information about objects and full text files retrieved by the Services of the first group.

1.3.2 Export

A Service is provided to export objects contained in the eSciDoc system in their internal storage format. The calling client has to provide a set of object identifiers it wishes to receive.