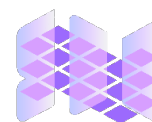


Document title

DATA PRODUCER TO SDC METADATA TEMPLATE

Prepared by	Hector Gonzalez
Reference	SDC-TN-PROC001
Issue and revision	i2, r1
Date of issue	15/09/2022
Status	Baseline
Distribution	ESA USOC Management, USOCs, SDC data producers



SIGNATURES SHEET

Title: Data Producer to SDC Metadata Template	Issue: i2, r1
---	---------------

Prepared by: _____ Signature and date

Hector Gonzalez (SDC)

Reviewed by: _____ Signature and date

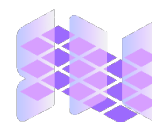
Jacobo Rodriguez (SDC)

Approved by: _____ Signature and date

Ana Laveron Simavilla (SDC)

Concurred by: _____ Signature and date

Alvaro Soria-Salinas (ESA)



CHANGE LOG

Reason for change	Issue	Date
Initial release	i1, r0	27/06/2019
Modifications on the metadata for data producers in line with HREDA v2.1	i2, r0	11/02/2022
Updates on the metadata schema for HREDA v2.2	i2, r1	15/09/2022

CHANGE RECORD

Issue: i2, r1

Description of change	Page(s)	Paragraph(s)
Corrected the RD-2 and RD-3 document reference codes	5	2.2
Changes in the file metadata fields: new element data author, specify accepted integrity methods	7	3.1
File Metadata schema updated	9	Annex 1
File Metadata example updated	11	Annex 2

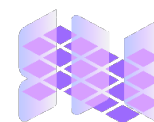
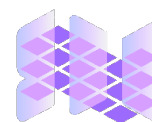


TABLE OF CONTENTS

1. Introduction.....	5
2. References.....	5
2.1. Applicable Documents	5
2.2. Reference Documents.....	5
2.3. Acronyms and abbreviations	5
3. Metadata definition	5
3.1. Metadata file for scientifically relevant data products	6
3.2. Checksum file for ancillary data	7
Annex 1 – File Metadata schema	9
Annex 2 – File Metadata example	11

LIST OF FIGURES

<i>Figure 1. Investigation data and metadata files.....</i>	<i>6</i>
---	----------



1. Introduction

This document provides a template for metadata relevant to ESA HRE investigations data, using XML files. It describes in detail the metadata files' format and content.

The template is applicable to investigation data to be transferred to the Science Data Centre (SDC), to standardise the data handling and exchange as much as possible.

2. References

2.1. Applicable Documents

Document	Document title	Issue	Source
[AD1] SDC-RP-PROC001	SDC processes: ISS data to SDC	i2, r2	SDC

2.2. Reference Documents

Document	Document title	Issue	Source
[RD1] ESA-ISS-COL-USOC-PL-0001	USOCs Data Management Policy	2.0	ESA
[RD2] SDC-TN-PROC002	Investigation Data Blank Book	i3, r0	SDC
[RD3] SDC-TN-PROC004	Investigation Data Blank Book (Multi-experiment)	i1, r0	SDC

2.3. Acronyms and abbreviations

AD	Applicable document
BDC	Baseline Data Collection
EM	Engineering Model
ESA	European Space Agency
FM	Flight Model
HRE	Human and Robotic Exploration
HREDA	HRE Data Archive
IETF	Internet Engineering Task Force
ISS	International Space Station
MD5	Message-digest algorithm 5
RD	Reference document
RFC	Request For Comments (Internet standards by the IETF)
SDC	Science Data Centre
SHA	Secure Hash Algorithms
TN	Technical Note
USOC	User Support and Operations Centre
XML	Extensible Markup Language

3. Metadata definition

As a general rule, a metadata file must be generated, in XML format, for each data file to be delivered to the SDC.

The metadata files must have the same filename as the associated data file, adding the .xml extension: [data file name].[data file extension].xml

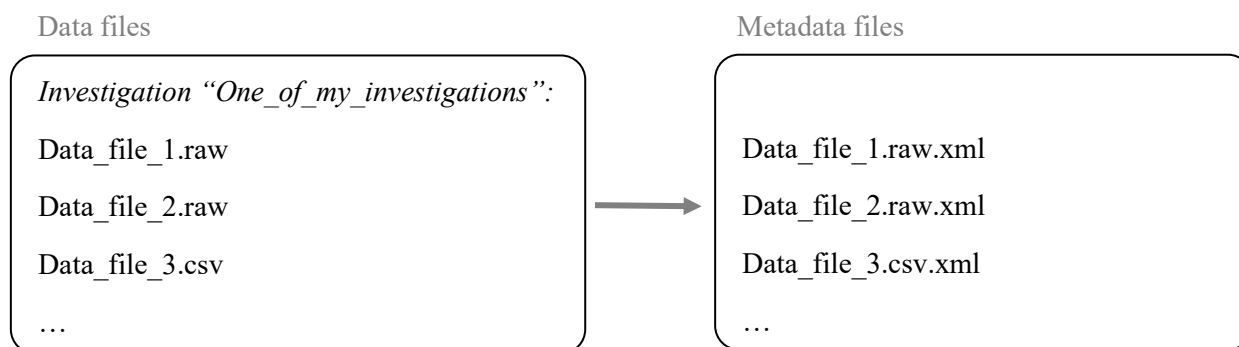
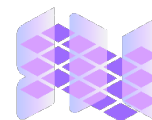
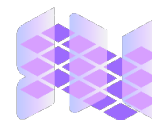


Figure 1. Investigation data and metadata files

3.1. Metadata file for scientifically relevant data products

Each data file from the products identified as scientifically relevant in the Investigation Data Blank Book [RD2] must be accompanied with a metadata file with the following elements (fields):

- *investigationName*: name of the investigation to which the data belongs to.
- *experimentName*: name of the experiment to which the data belongs to (optional, in case the investigation is divided in different experiments).
- *model*: element that indicates if the data is coming from a Flight Model (FM), Engineering Model (EM) or other payload models.
- *dataSource*: element that specifies which type of activity originated the data:
 - On-board execution
 - Ground reference
 - Post-flight
 - BDC
 - Other (specify)
- *dataOwner*: entity (or entities) owner of the data. In most cases it will be ESA.
- *dataAuthors*: list with the names of the persons who produced the data, as acknowledgement.
 - *authorName*: name and surname of the data author.
 - *authorAffiliation*: affiliation or company of the data author.
- *acquisitionTime*: acquisition time of the data (optional, in case this information is known). For files containing data acquired along a time interval, this field indicates the start time of the data acquisition interval.
- *acquisitionEndTime*: end time of the data acquisition interval (optional, only for files containing data acquired along a time interval and not at a specific timepoint).
- *creationTime*: creation time of the data.
- *subjects*: list of subjects of the investigation that have information belonging to them in the associated data file (optional). Note: as a general rule, a file will only contain data related to one subject. Exceptions to this rule will be handled case-by-case.
 - *subject*: Name and surname of a subject.
- *processingLevel*: processing level (0, 1, 2 or higher) of the data:
 - 0: raw data
 - 1: calibrated data
 - 2: post-processed (also defined as analysed data)



- *productType*: product category (e.g. science image, operational image, telemetry, video, ...) to which the data belongs to. It must be consistent with the information provided in the Investigation Data Blank Book [RD2].
- *fileFormat*: information about the type/encoding of the data file and other information needed to read the file. It must be consistent with the information provided in the Investigation Data Blank Book [RD2].
- *relativePath*: path to the directory in which the file is located (relative to the investigation directory), using slash (/) as separator. For example, if the file “example.txt” from investigation “INV” is located in the following path:
/home/experiments/INV/telemetry/day/example.txt, then the *relativePath* is “telemetry/day”.
- *integrity*: metadata that provides information about the integrity of the file.
 - *method*: integrity method used. Accepted algorithms are SHA-256, SHA-1 or MD5.
 - *value*: integrity value.
- *investigationSpecificMetadata*: additional metadata fields that are specific for the investigation. For example: run number, container identification, sample identification, etc. They must be consistent with the information provided in the Investigation Data Blank Book [RD2].
 - *parameter*: metadata that provides information about one specific parameter.
 - *name*: parameter name.
 - *value*: parameter value.

The document root (in each XML metadata file) needs to include the following information for processing at the SDC:

```
<metadata xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://sdc.upm.es/SDC/Metadata/1"
xsi:schemaLocation="file_metadata_schema.xsd">
```

The XML schema (*file_metadata_schema.xsd*) and an example metadata file for an investigation data file (*data_file_example.fits.xml*) are provided in Annex 1 and 2, respectively.

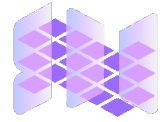
3.2. Checksum file for ancillary data

Ancillary data (or non-scientifically relevant data) should be transferred to the SDC accompanied with an associated list of checksums to help verify data integrity. The proposed format consists of a CSV file (as per RFC 4180) with three values/columns:

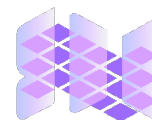
- The checksum algorithm used. Accepted algorithms are SHA-256, SHA-1 or MD5.
- The checksum value (hexadecimal, lower-case, zero-padded).
- The relative path to the investigation directory (similar to the *relativePath* field described in section 3.1), including the data file name.

As in the following example checksum list:

```
SHA-256,7928ae4eafbc8eb93cf0ecd3879fcb68ed3f65986de8ac8621e2bc6d4161609,FDs/TAC-1.05-6.4.8-QS/DATA/dir_sample.txt
SHA-256,2505c95d9f207320c1f224645aeec16218c1a9109c6373b4ef98e4511c11d3c,FDs/TAC-1.05-6.4.8-QS/DATA/d_listing_2.txt
SHA-256,f7e1afc9f3581869a77f9423bdd9463cd984f0f0e1a51671edbce3537155a8fc,FDs/TAC-1.05-6.4.8-QS/DATA/d_listing.txt
```



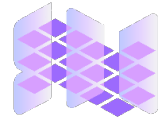
SHA-256,b7db6140d7be9bb76634485e349ac8512644914cc307e6300ac888d7ab5dfcf9,FDs/TAC-1.05-6.4.8-QS/DATA/img/09_N_210414_063312_3419@0016_00.tif



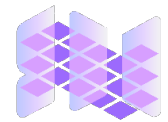
Annex 1 – File Metadata schema

File name: file_metadata_schema.xsd

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns="http://sdc.upm.es/SDC/Metadata/1"
targetNamespace="http://sdc.upm.es/SDC/Metadata/1" elementFormDefault="qualified"
version="2.0.0">
  <xs:annotation>
    <xs:appinfo>Data Producer file metadata</xs:appinfo>
    <xs:documentation xml:lang="en">
      This Schema defines the format for the File Metadata file, which contains additional
information on the referred file.
    </xs:documentation>
  </xs:annotation>
  <xs:element name="metadata">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="investigationName" type="xs:token"/>
        <xs:element name="experimentName" type="xs:token" minOccurs="0"/>
        <xs:element name="model" type="xs:token" minOccurs="0"/>
        <xs:element name="dataSource" type="xs:token"/>
        <xs:element name="dataOwner" type="xs:token" maxOccurs="unbounded"/>
        <xs:element name="dataAuthors" type="DataAuthors" minOccurs="0"/>
        <xs:element name="acquisitionTime" type="xs:dateTime" minOccurs="0"/>
        <xs:element name="acquisitionEndTime" type="xs:dateTime" minOccurs="0"/>
        <xs:element name="creationTime" type="xs:dateTime"/>
        <xs:element name="subjects" type="Subjects" minOccurs="0"/>
        <xs:element name="processingLevel" type="xs:token"/>
        <xs:element name="productType" type="xs:token"/>
        <xs:element name="fileFormat" type="xs:token"/>
        <xs:element name="relativePath" type="xs:token"/>
        <xs:element name="integrity" type="Integrity"/>
        <xs:element name="investigationSpecificMetadata"
type="InvestigationSpecificMetadata" minOccurs="0"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:complexType name="DataAuthors">
    <xs:sequence>
      <xs:element name="dataAuthor" type="DataAuthor" maxOccurs="unbounded"/>
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="DataAuthor">
    <xs:sequence>
      <xs:element name="authorName" type="xs:token"/>
      <xs:element name="authorAffiliation" type="xs:token"/>
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="Subjects">
    <xs:sequence>
      <xs:element name="subject" type="xs:token" maxOccurs="unbounded"/>
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="Integrity">
    <xs:sequence>
      <xs:element name="method" type="xs:token"/>
      <xs:element name="value" type="xs:token"/>
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="InvestigationSpecificMetadata">
    <xs:sequence>
      <xs:element name="parameter" type="Parameter" minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
  </xs:complexType>
</xs:schema>
```



```
</xs:complexType>  
<xs:complexType name="Parameter">  
  <xs:sequence>  
    <xs:element name="name" type="xs:token"/>  
    <xs:element name="value" type="xs:token"/>  
  </xs:sequence>  
</xs:complexType>  
</xs:schema>
```



Annex 2 – File Metadata example

File name: data_file_example.fits.xml

File content:

```
<?xml version="1.0" encoding="utf-8"?>
<metadata xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://sdc.upm.es/SDC/Metadata/1" xsi:schemaLocation="file_metadata_schema.xsd">
  <investigationName>DCMIX</investigationName>
  <experimentName>DCMIX-2</experimentName>
  <model>FM</model>
  <dataSource>On-board facility</dataSource>
  <dataOwner>ESA</dataOwner>
  <dataOwner>NASA</dataOwner>
  <dataAuthors>
    <dataAuthor>
      <authorName>First author</authorName>
      <authorAffiliation>University of First</authorAffiliation>
    </dataAuthor>
    <dataAuthor>
      <authorName>Second author</authorName>
      <authorAffiliation>Second Institute</authorAffiliation>
    </dataAuthor>
    <dataAuthor>
      <authorName>Third author</authorName>
      <authorAffiliation>Triple Industries, Limited</authorAffiliation>
    </dataAuthor>
  </dataAuthors>
  <acquisitionTime>2016-10-17T01:24:14</acquisitionTime>
  <acquisitionEndTime>2016-10-18T01:24:14</acquisitionEndTime>
  <creationTime>2016-10-22T15:06:19</creationTime>
  <subjects>
    <subject>Jeanne Doe</subject>
  </subjects>
  <processingLevel>1</processingLevel>
  <productType>Science image</productType>
  <fileFormat>FITS data. Image data file containing a metadata header and a collection of
8-bit linear grayscale, 1920 pixels wide by 1080 pixels high</fileFormat>
  <relativePath>images/day1</relativePath>
  <integrity>
    <method>SHA-256</method>
    <value>029658b0741d408857d0ec765fbaf9981662c2038501dc1c78f1c557c5f1c812</value>
  </integrity>
  <investigationSpecificMetadata>
    <parameter>
      <name>runName</name>
      <value>1</value>
    </parameter>
    <parameter>
      <name>phaseName</name>
      <value>thermalisation</value>
    </parameter>
  </investigationSpecificMetadata>
</metadata>
```