Funded by the Horizon 2020 Framework
Programme of the European Union

HEMERA

N° HEMERA/2021/9/RP/087 Issue 1/revision 2

Date: 07/12/2021

HEMERA-2020
User guide for online data providers

Draft version	
Final approved version	X

Work package applicability:

WP 1	WP 2	WP 3	WP 4	WP 5	WP 6	WP 7	WP 8	WP 9	WP 10	WP 11	WP12
	x										

DOCUMENT CREATION:

Type of document (Report RP, Note NT, Data DT, Minutes MN, List LT, Plan PL)	RP
Category (PU, CO)	PU
Subject	User guide for online data providers
Author(s)	Cathy Boonne, Sébastien Payan
Company	CNRS
Key words	Balloon, Data, Virtual Access

MODIFICATIONS:

Version	Date	Modifications	Observations
0.1	2021/05/10		First Draft
0.2	2021/12/03	Revised document based on version 0.1	First diffusion

HEMERA-2020 user guide for online data providers

The HEMERA-2020 Data Centre is hosted by the « Atmosphere and Service Data Pole » AERIS (CNRS, France). All data have been transferred and are physically stored on AERIS server in order to ensure sustainability of data access and long term archiving. The architecture of the Data Centre has been built following recommendations from the AERIS in order to benefit from a high level of expertise in the storage, the distribution techniques and the interoperability of the data.

A web interface has been developed for the Data Centre and the structure of the database has been implemented in order to make it user-friendly. The HEMERA-2020 Data Centre is available to the research community since July 2019 and reachable from this internet address: **https://data.hemera-2020.eu**. Currently, the two databases are operational:

- Database of atmospheric balloon-borne products,
- Database of astrophysical balloon-borne products.

Services offered to data providers and users have been assessed by the consortium and several improvements/developments have been proposed. This documents presents a user guide to help providers to submit their metadata and data in the HEMERA datacenter.

Automatic management of metadata and data tools

For each database, a provisional metadata interface has been developed and implemented within HEMERA-2020 DC site. This interface permits data providers to create, edit, modify or delete easily metadata sheets in order to populate and manage the metadata catalogue. It permits also to upload data files related to the new dataset described.

The data provision service has been improved all along the project to facilitate the integration of data from multiple sources and the generation of metadata. Several tools have been already specified and implemented.

For data providers, authentication with the ORCID account is required to get the rights to provide new data.

After registration, the ORCID site sends automatically a message to the user delivering the ORCID number. The ORCID number looks like this template: XXXX-XXXX-XXXX-XXXX.

This number has to be sent to HEMERA Data Centre administrator (currently: Marjorie Salvetat, Marjorie.Salvetat@ipsl.fr) in order to give rights to data providers to use the functionalities and tools developed.

- ***How to access to the functionalities tools developed***

The data providers have just to sign in the HEMERA-2020 Data Access by clicking on the item “LOGIN” (orange box in figure 1). A window appear, click on the button “ID” in green to sign into ORCID. The item “LOGIN” is refreshed and a “LOGOUT” item appears with the name of the data provider logged in (orange box in figure 2).

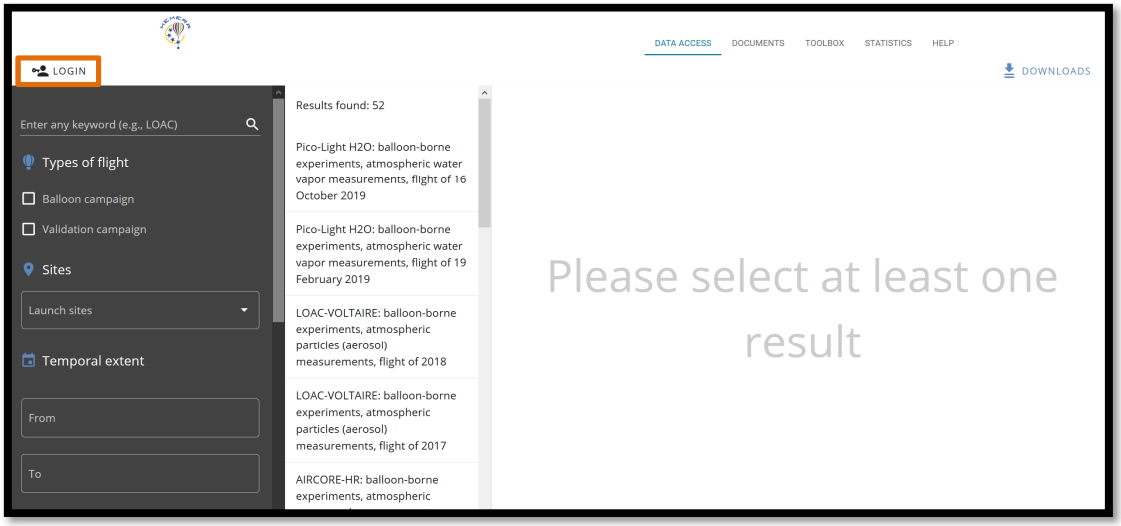


Figure 1: Snapshot of the data access page for the provisional metadata interface.

ORCID provides a persistent digital identifier that distinguishes a user from other researchers and supports the linkages between the researcher and his professional activity. This account is not dedicated to HEMERA-2020 Data Centre and can be used on others web sites sharing the same authentication method.
The way to create an ORCID is described in the architecture document [D2.3]. Data providers have to connect to the ORCID site (https://orcid.org).

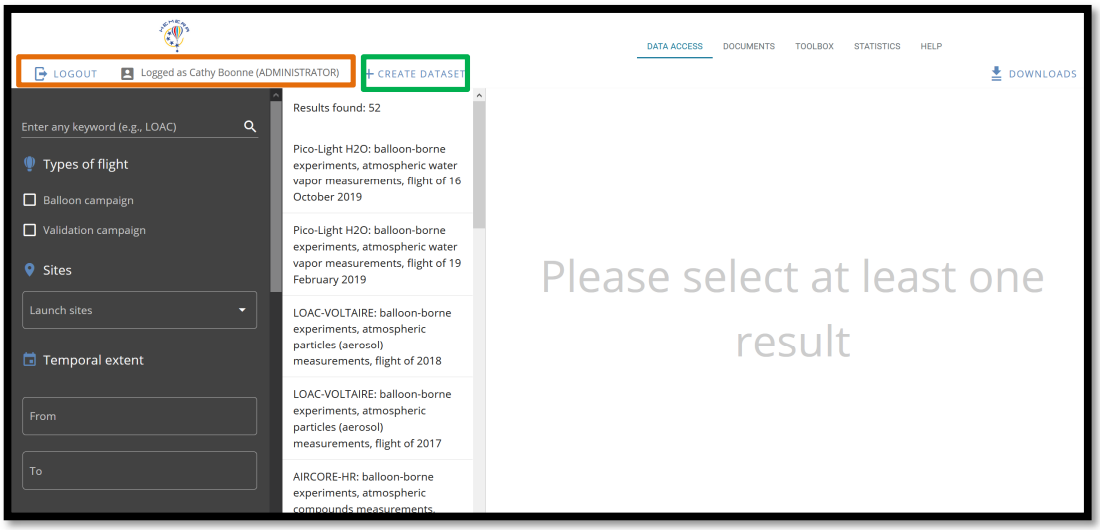
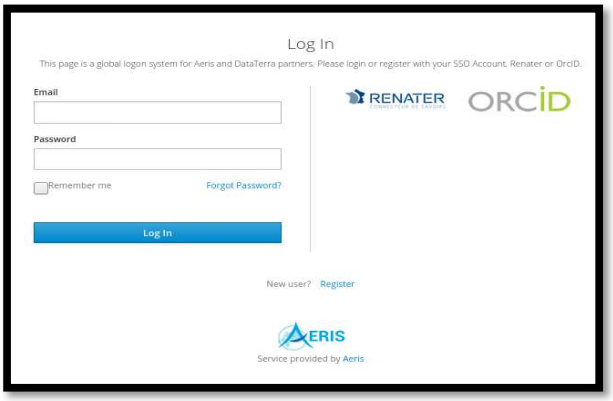


Figure 2: Snapshot of the data access page for metadata interface.



- ***The metadata provision service***

The creation of a new metadata sheet is available by clicking the button in blue "+ CREATES DATASET" (green box in figure 2).

a) <u>Generality</u>

Generally, the creation of metadata sheets is based on the type of instruments of measurement, all the data files produced will then be grouped in one dataset. However, if needed, it is possible to create metadata sheets by specific experiments.

To create a new metadata sheet, fourteen information boxes listed below have to be filled; an orange star indicates the mandatory ones:

- *Access*
- *Institute*
- *Type of flight*
- *Balloon*
- *Station*
- *Temporal extent*
- *Main instrument*
- *Auxiliary instruments,*
- *Parameters*
- *Publications*
- *Images*
- *Contacts*
- *Summary*
- *Data*

All the providers have to follow each step in order to fill the information boxes. "NEXT" button guides the providers to a next step and expands the following information box to be filled, in the same way a "PREVIOUS" button helps them to go back to the previous step and permits them to correct the content of the previous box. A step by step example of the creation a new metadata sheet is given at the end of this document.

b) Type of access

We have defined a new strategy for the metadata provision service: two types of access are now possible for the data providers:

- Private access
- Public access

This functionality corresponds to the first step of the provisional interface (**Figure 3**).

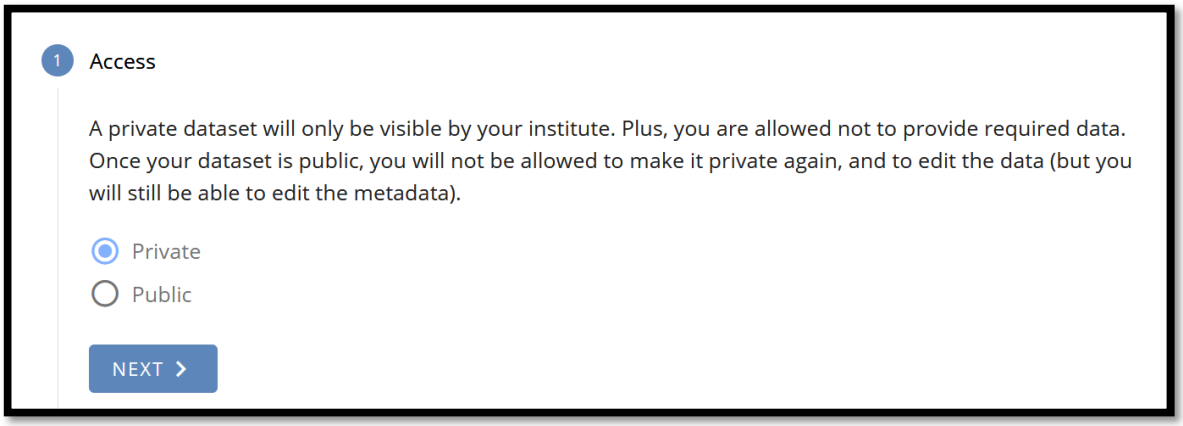


Figure 3: Snapshot of the access block of the data provision service.

The private access permits the data provider to modify as many times as he wants the data and metadata as they are not integrated to the operational metadata catalogue and will be visible by exclusively the members of his institute. It can be viewed as a draft to be improved before the data publication.

The public access permits the data provider to publish the final version of the data and metadata. The upload of data files and the completeness of the metadata sheet are therefore mandatory. The edition the data files uploaded

will not be feasible, but the modification of the metadata information entered is still possible.

c) Pre-defined fields

For the steps describing the "*Institute", "Type of flight", "Balloon", "Main instrument", "Auxiliary instrument", "Parameters", "station",* a pre-defined list of attributes is given to help the data providers to choose the relevant proposition.

As example, description of step 3: Station box

In this step we have to choose the name of the observatory station where the balloon was launched, we can choose one or more. It is also a mandatory field. A pre-defined list of observatory stations is given and available by clicking the dart circled in orange in figure 4. Currently, this list consists of 15 observatory stations:

- Kiruna, Sweden
- Teresina, Brazil
- Timmins, Canada
- Aire-sur-Adour, France
- Iceland, Iceland
- Alice Springs, Australia
- Bauru, Brazil
- Gap, France
- Leon, Spain
- Andoya, Norway
- Vanscoy, Canada
- Niamey, Niger
- Mc Murdo, USA
- Esrange, Sweden
- Latacunga, Equator

It is possible to add some more stations if necessary. At this stage, we have to select the right stations. Because this field is mandatory, the box title appears in red (figure 4) if you don't select a station. It will be the same for each mandatory field.

Figure 4: Snapshot of the stations box expanded.

d) Free fields

For the steps describing "*Temporal extent", "Publications", "Contacts", "Summary*", free texts are needed describing the relevant information.

As example, <u>description of step 10:</u> Publications box
This step concerns the papers dedicated to the data produced and the way that it has been measured. Filling this box is not mandatory but recommended. Just open the pop window by clicking "+ ADD PUBLICATION" in figure 5.

Figure 5: Snapshot of the publications box.

The pop window (figure 6) permits to enter the five fields needed (title, authors, journal name, journal section, year of publication). For the field "Authors" a pre-defined list is given, this list consists of the name of the data providers.

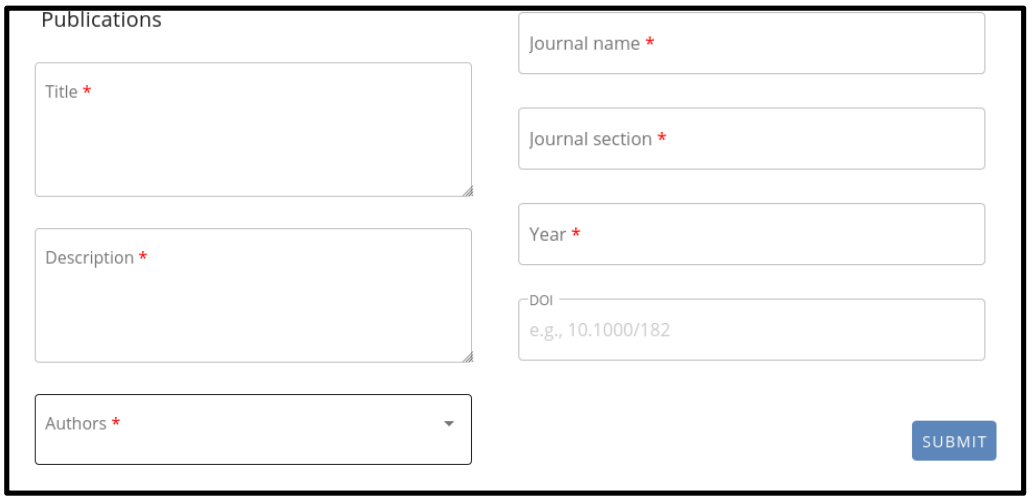


Figure 6: Snapshot of the publications fields

- ***The data upload service***

This functionality corresponds to the step 14 of the metadata provision service.

To upload a file, you have just to click on the "ADD FILE button" (figure 7), a popup window appears (figure 8).

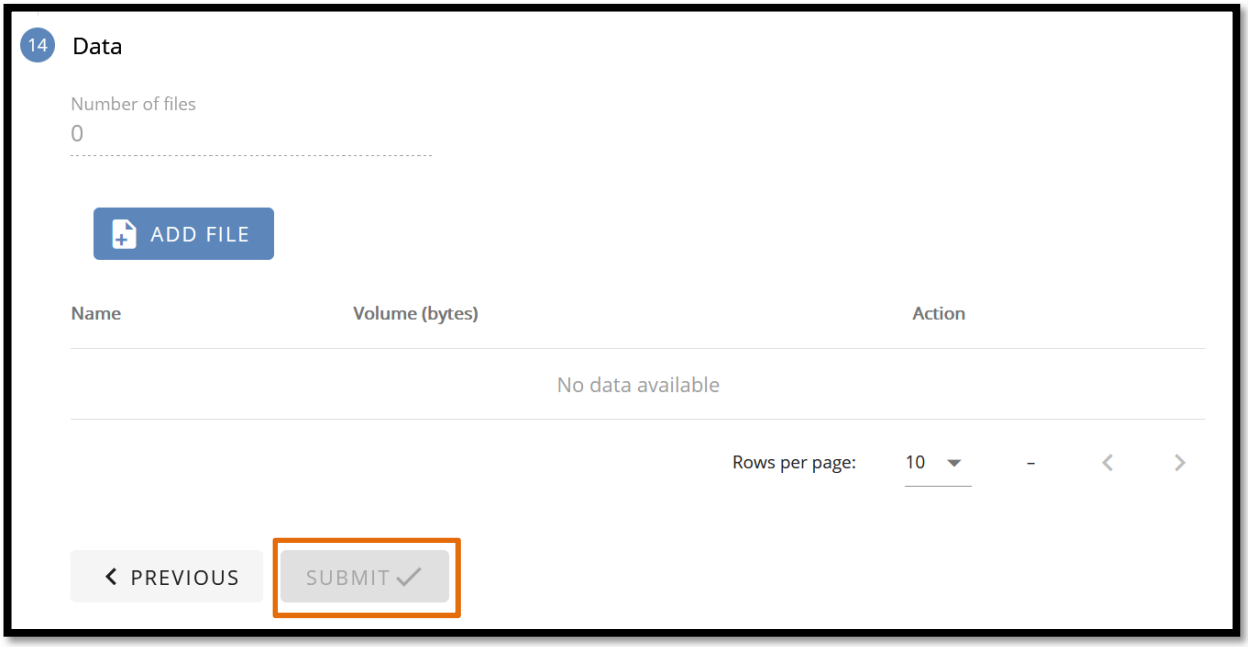


Figure 7: Snapshot of the data box.

You have just to select a file to upload it by a click on the "SUBMIT" button. Before uploading the data file, a check of the consistency of the data file format is done. If the data file format is wrong, the upload fails. You have just to fix the data file format before try again the upload.

This tool includes the automatic generation of metadata and the verification of completeness of the submission process by clicking on the "SUBMIT" button (orange box in figure 7). In order to prevent from incomplete data, new records cannot be saved if one of the two files is missing or if mandatory fields are not filled in.

Figure 8: Snapshot of the add file window.

Duplication of the metadata sheets

In order to ease the filling of the metadata form implemented for HEMERA databases, we have added a new button "Clone" (orange box in figure 9) to the existing functionalities of the data provision interface. This new functionality permits to duplicate an existing metadata sheet. This tool avoids the data providers to fill again each field, and provides a gain of time to populate the metadata catalogue.

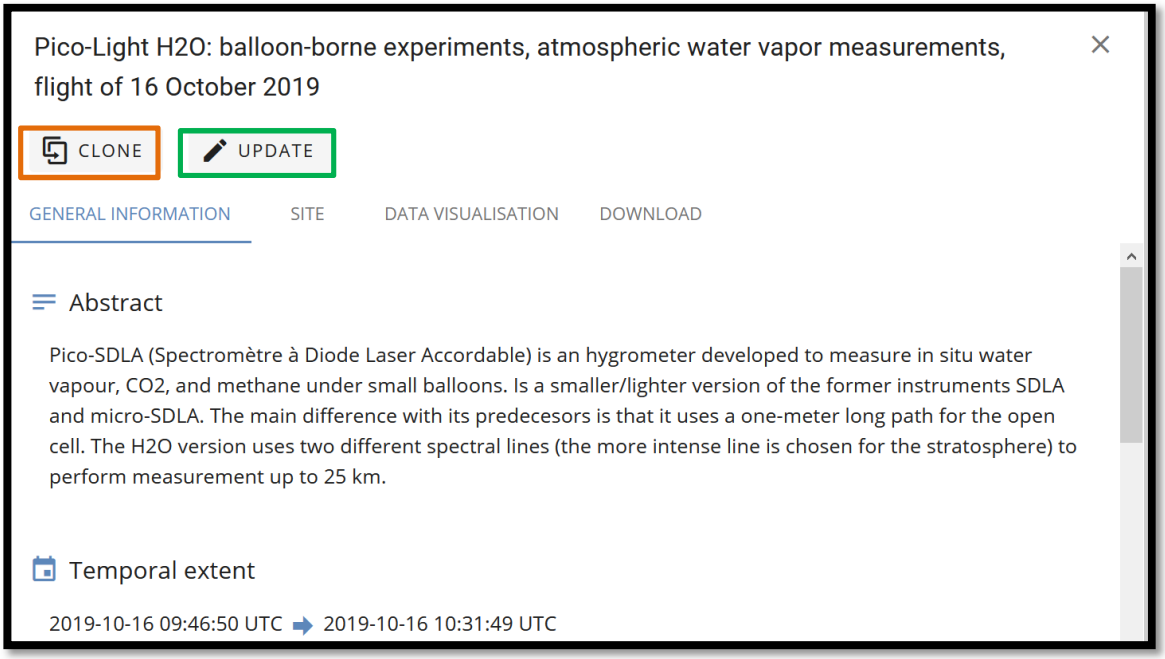


Figure 9: Snapshot of a metadata sheet.

In the same way, an "UPDATE" button (green box in figure 9) permits the data provider to correct the fields already filled.

Keycloack authentication

To get rights to provide new data, a data provider authentication is required using an ORCID account. We have improved the provider access security by implementing the Keycloack authentication solution. This authentication Keycloak is an open source software product that allows single sign-on with identity and access management. With this tool, we avoid any risk of compromising our data.

Example of a step by step metadata sheet creation

1 Access

A private dataset will only be visible by your institute. Plus, you are allowed not to provide required data. Once your dataset is public, you will not be allowed to make it private again, and to edit the data (but you will still be able to edit the metadata).

- (•) Private
- ( ) Public

NEXT >

2 Institute

Institute *

GSMA

< PREVIOUS NEXT >

3 Types of flight

4 Balloon

Balloon *

ZPB (Zero Pressure Balloon), volume: 402000 m3

< PREVIOUS NEXT >

5 Station

Station *

Aire-sur-Adour, FRANCE

< PREVIOUS NEXT >

6 Temporal extent

From *

2019-02-19 00:00:00 UTC

The date must be set in UTC

To *

2019-10-16 14:00:00 UTC

The date must be set in UTC

< PREVIOUS NEXT >

7 Main instrument

The name of this instrument and the type of measurement will be used to create the title of the dataset.

Group *
Temperature/Humidity Sensors

Name *
PICO-LIGHT

Type of measurement *
Atmospheric water vapor

Manufacturer

Model

PREVIOUS NEXT

8 Auxiliary instruments

9 Parameters

Number of parameters
1

ADD PARAMETER

Short name	Long name	Unit of measure	Comment	Actions
H2O	Humidity Mixing Ratio	ppmv		

Rows per page: 5 1-1 of 1

PREVIOUS NEXT

10 Publications

11 Images

12 Contacts

Number of points of contact: 1

Number of principal investigators: 1

Number of authors: 0

+ ADD CONTACT

Name	Orcid	Email	Point of contact	Principal investigator	Author	Actions
HEMERA contact		cathy.boonne@ipsl.fr	✓	✗	✗	
Georges Durry		georges.durry@univ-reims.fr	✗	✓	✗	

Rows per page: 5 1-2 of 2

< PREVIOUS NEXT >

13 Summary

Summary *

Pico-SDLA (Spectromètre à Diode Laser Accordable) is an hygrometer developed to measure in situ water vapour, CO2, and methane under small balloons. Is a smaller/lighter version of the former instruments SDLA and micro-SDLA. The main difference with its predecesors is that it uses a one-meter long path for the open cell. The H2O version uses two different spectral lines (the more intense line is chosen for the stratosphere) to perform measurement up to 25 km.

< PREVIOUS NEXT >

14 Data

Number of files: 0

ADD FILE

Name	Volume (bytes)	Action
No data available		

Rows per page: 10 -

< PREVIOUS SUBMIT ✓

Pico-Light H2O: balloon-borne experiments, atmospheric water vapor measurements, flight of 16 October 2019

GENERAL INFORMATION | SITE | DATA VISUALISATION | DOWNLOAD

Abstract

Pico-SDLA (Spectromètre à Diode Laser Accordable) is an hygrometer developed to measure in situ water vapour, CO2, and methane under small balloons. Is a smaller/lighter version of the former instruments SDLA and micro-SDLA. The main difference with its predecesors is that it uses a one-meter long path for the open cell. The H2O version uses two different spectral lines (the more intense line is chosen for the stratosphere) to perform measurement up to 25 km.

Temporal extent

2019-10-16 09:46:50 UTC → 2019-10-16 10:31:49 UTC

Parameters

Long name	Short name	Type
Water vapor	Water vapor	Atmospheric water vapor > Water vapor indicators > Water vapor

Instruments

Name	Type
Pico-Light H2O	Temperature/Humidity Sensors > Hygrometers > Frost Point Hygrometers

Contacts

Name	Email	Point of contact	Principal investigator	Author
HEMERA contact	cathy.boonne@ipsl.fr	✓	✗	✗
Mélanie Ghysels	melanie.ghysels-dubois@univ-reims.fr	✗	✓	✓

Formats

AMES

Other information

Unique identifier: dce3bb21-f0af-4101-8a99-436c04e4fa64
Data processing level: L2