



EUNADICS-AV DELIVERABLE (D -N°: D41)

ICD between data providers and EUNADICS-AV Portal

File name: EUNADICS-AV-Deliverable_D41.pdf

Dissemination level: PU (public)

Author(s): Saskia Wagenaar
Wim Som de Cerff

Reviewer(s): Lucia Mona
Nicolas Theys
Matthieu Plu

Reporting period:

Release date for review: **06/10/2017**

Final date of issue: **24/11/2017**

Revision table			
Version	Date	Name	Comments
0.1	22-09-2017	Saskia Wagenaar	Initial version of the document
0.2	05-10-2017	Saskia Wagenaar	Included comments from Wim, created the Abstract and the Executive summary
0.3	06-10-2017	Saskia Wagenaar	Included comments from Wim.
0.4	12-10-2017	Saskia Wagenaar & Wim Som de Cerff	Included comments from Nicolas and Lucia
1.0	18-10-2017	Saskia Wagenaar & Wim Som de Cerff	Internal review ready version

Abstract

This deliverable formalizes the interface between EUNADICS-AV data providers and the data portal. It specifies interface requirements for the following types of data products:

- Routine data products, which are data products which are updated regularly, not triggered by special events
- Alert data products, which are data products which are created and updated when an alert is issued

The interface requirements address the ingestion location and procedure, the foreseen file formats, metadata standards and the (NRT) time requirements for certain datasets.

The EUNADICS-AV project has received funding from the European Union's Horizon 2020 research programme for Societal challenges - smart, green and integrated transport under grant agreement no 723986

Executive Summary

The goal of the EUNADICS-AV data portal is to disseminate the following data products to NMSs, VAACs, SESAR/SWIM portals and the wider public:

- Tailored and routine observation data products as provided by WP3
- Alerts based on observation data products as provided by WP5
- Air space specific model output prototype products as provided by WP6

This Interface Control Document (ICD) formalizes the interface between the data providers (WP3, WP5, WP6 partners) and the data portal (WP8).

The data portal will provide two options to ingest the offered data products:

- Ingestion of routine data products, which are data products updated regularly, not triggered by special events
- Ingestion of alert data products, which are data products created and updated when an alert is issued.

Interface requirements for both of these product types are specified, next to a few general interface requirements. The requirements address the ingestion location and procedure, the foreseen file formats, metadata standards and the (NRT) time requirements for certain datasets. It does not address the archiving requirements.

Since at the time of writing this document not all products are defined yet, it is possible that new interface requirements should be specified to accommodate for new product types. These new requirements will be defined in collaboration with the data providers and this document will be updated accordingly. When data providers are not able to comply to one or more of the interface requirements, an alternative solution should be discussed and documented.

Table of contents

1.	Introduction	4
1.1	Purpose of the document	4
1.2	Structure of the document	4
1.3	Intended readership	4
2.	Applicable and reference documents.....	5
2.1	Applicable documents	5
2.2	Reference documents	5
3.	Terms, definitions and abbreviated terms	6
3.1	Definitions	6
3.2	Acronyms and abbreviations	6
4.	Software overview.....	8
4.1	EUNADICS Data Portal architecture overview	8
a)	Frontend.....	9
b)	Backend	9
c)	Interfaces	10
4.2	Data product creation	10
a)	Routine data products	10
b)	Alert data products	11
c)	Relation between alert and routine products	11
4.3	File ingestion and archive process.....	12
a)	Happy flow routine products	13
b)	Happy flow alert products	13
5.	Requirements and design.....	15
5.1	Interface requirements	15
a)	General interface requirements.....	15
b)	Ingestion of routine data products	15
c)	Ingestion of alerts	15
5.2	Interface design	15
a)	FTP file upload process	16
b)	NetCDF CF compliancy	16
c)	Alert file format and filename structure	16
d)	NRT specifications	17

1. Introduction

1.1 Purpose of the document

The purpose of this Interface Control Document (ICD) is to formalize how data products are to be delivered to the EUNADICS-AV data portal. It will address physically ingesting files, the foreseen file formats, metadata standards and the (NRT) time requirements for certain datasets.

1.2 Structure of the document

This document follows the structure of a Software Interface Control Document (ICD) conform the ECSS-E-ST-40C [R1] standard of the ECSS.

Chapter 1 describes the purpose of this document, its structure and the intended readership.

Chapter 2 lists all the documents referenced in this ICD.

Chapter 3 describes the definitions used in this document.

Chapter 4 gives an overview of the EUNADICS-AV data portal.

Chapter 5 describes the interface requirements, as well as the interface design.

1.3 Intended readership

This ICD is written for all data providers in the consortium and data providers outside the consortium, which will be delivering data products to the EUNADICS-AV data portal developed by WP8.

2. Applicable and reference documents

2.1 Applicable documents

- [A1] CF Conventions (22-09-2017). *CF Conventions and metadata*. Available at:
<http://cfconventions.org/>

2.2 Reference documents

- [R1] ECSS (06-03-2009). *ECSS-E-ST-40 – Software general requirements*. Available at:
<http://ecss.nl/standard/ecss-e-st-40c-software-general-requirements/>
- [R2] KNMI (22-09-2017). *KNMI Data center*. Available at:
<https://data.knmi.nl>
- [R3] EUNADICS-AV D43: *ICD between the EUNADICS-AV Portal and the
VAAC/NMA/SESAR-SWIM systems*

3. Terms, definitions and abbreviated terms

3.1 Definitions

Definition	Description
Alert	Alerts are given soon as an airborne hazard is detected. Alerts can contain information on where and when the hazard is detected, and can also be updated over time.
Alert data product	Data products which are created when an Alert is issued.
Data granule	A data file, part of a dataset.
Dataset	A group of data granules with a common set of metadata.
Dataset collection	A collection consisting of several datasets which can be previewed and downloaded together.
Event	An event is created in the data portal, when an Alert is provided by the data provider and ingested by the data portal. An event is named with a unique, human readable name. It is to be discussed within EUNADICS_AV how naming (and renaming) of events should occur. For the time being a generated name will be assigned, which can be altered.
Event Type	Data providers which want to offer alert products should create an event type for their alerts in the data portal. The following things should be specified per event type: Description of the event type, Relevant dataset collections, Alert file format, Ingest location (FTP server) from which the alert files can be ingested. Event types are flexible and can be created when needed. EUNADICS will start with providing event types with the highest priority.
Routine data product	Data products which are updated regularly, not triggered by special events. These products will not be archived by EUNADICS-AV (for most routine products this is already arranged, e.g. ceilometers by E-profile). But when an Event is triggered, the routine data product will be archived as part of this Event dataset.

3.2 Acronyms and abbreviations

Acronyms/ Abbreviations	Definition
API	Application Programming Interface
EUNADICS-AV	European Natural Airborne Disaster Information and Coordination System for Aviation
FTP	File Transfer Protocol
HTTP	Hypertext Transfer Protocol
ICD	Interface Control Document
INSPIRE	Infrastructure for Spatial Information in the European Community
ISO	International Organization for Standardization
KDC	KNMI Data Centre
KNMI	Royal Netherlands Meteorological Institute
NADC	KNMI Data ingestion, processing and distribution software
NADC	KNMI software for ingesting, processing and distribution of data
NetCDF	NetCDF is a set of software libraries and self-describing, machine-

	independent data formats that support the creation, access, and sharing of array-oriented scientific data. The project homepage is hosted by the Unidata program at the University Corporation for Atmospheric Research.
NetCDF-CF	NetCDF Climate and Forecast Metadata Conventions
NRT	NRT is a time delay introduced by automated data processing or network transmission between the occurrence of an event and the use of the processed data. Typically – for meteorological purposes – this is within 3 hours from sensing. [Definition from CEOS-WGISS Glossary of Acronyms and Terms, v1.0]
OPeNDAP	Open-source Project for a Network Data Access Protocol
VAAC	Volcanic Ash Advisory Center
WCS	Web Coverage Service
WMS	Web Map Service

4. Software overview

The goal of the EUNADICS data portal is to disseminate the following EUNADICS data products:

- Tailored and routine observation data products as provided by WP3
- Alerts based on observation data products as provided by WP5
- Air space specific model output prototype products as provided by WP6

The relevant products should be delivered to NMSs, VAACs, the relevant SESAR/SWIM portals and to the wider public. Note that (routine) products will only be disseminated if needed and if data policies allow for it.

In principle EUNADICS-AV will not re-distribute routine data products but only value added products. Routine data products will be archived only when an event occurs and the routine datasets are of interest for this event.

The purpose of this ICD is to define the interface between the data providers and the EUNADICS data portal. In order to identify the interface requirements, this chapter will give a global overview of the data portal architecture and concepts. Furthermore it will provide a more in depth discussion of the data product creation and ingestion process, both of which are relevant processes for the interface between the data providers and the data portal.

Note if external data providers cannot comply to the proposed interfaces, separate agreements will need to be made.

4.1 EUNADICS Data Portal architecture overview

The EUNADICS Data Portal will be hosted by the KNMI data center facilities (KDC, data.knmi.nl). It will reuse existing KDC portal software to enable rapid development. Figure 1 illustrates the resulting combination of EUNADICS data portal components and existing KDC components. In the following paragraphs a short description of all the layers in the application will be given.

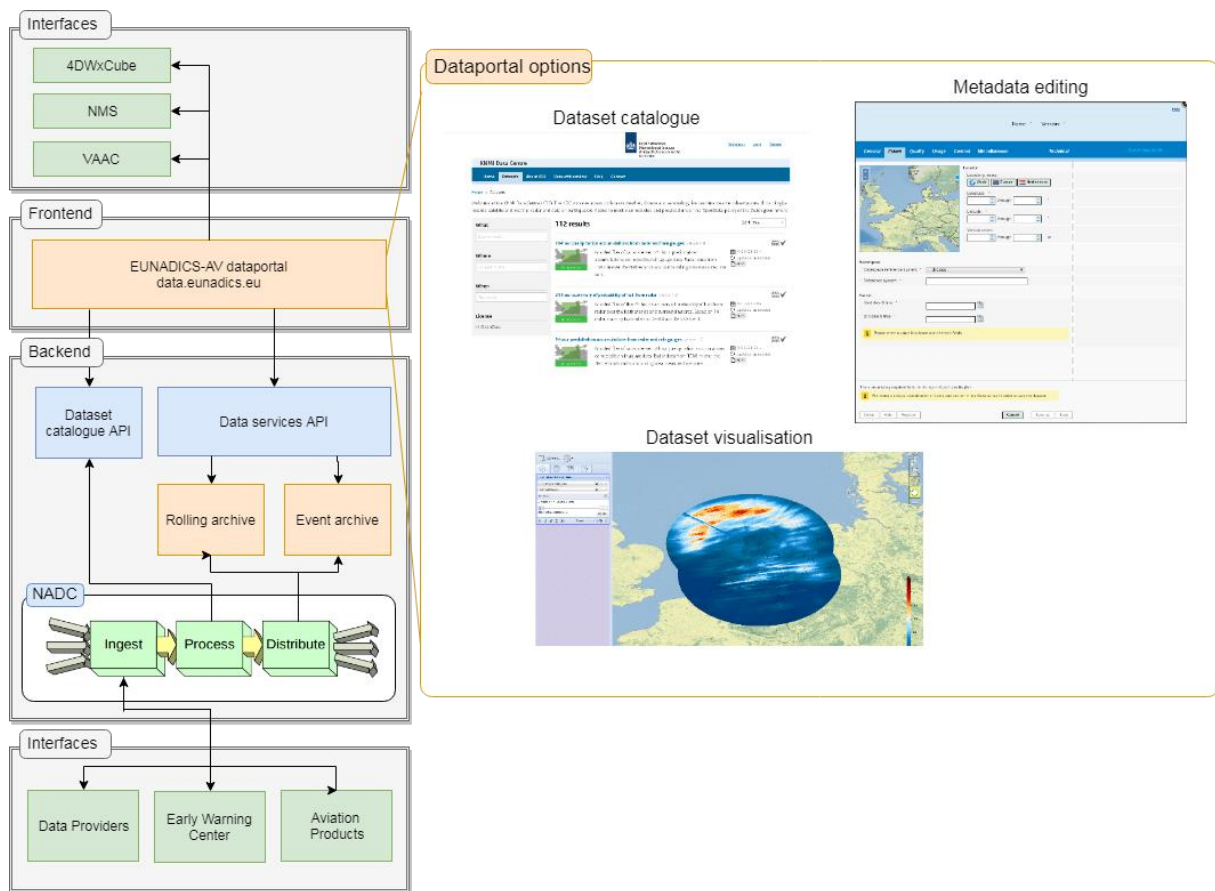


Figure 1: Architecture data portal

a) Frontend

A EUNADICS-AV specific frontend will be developed which will provide:

- A metadata editor portal
- A data dissemination portal, including a catalogue and data visualization.

It will be constructed combining existing components from the KDC frontend with newly developed EUNADICS-AV specific components.

b) Backend

The data portal will mostly reuse components of the KNMI Data Center for the backend. These components will provide the following functionality:

- **Data services**, like WMS, WCS, OPeNDAP, HTTP download and FTP download
- **Dataset catalogue API**, in order to make metadata searchable
- **Ingestion software (NADC)**, in order to ingest, process and distribute¹ files

¹ Only if needed and if data policy allows for it. In principle EUNADICS-AV will not re-distribute routine data products but only value added products. Routine data products will be archived only when an event occurs and the routine datasets are of interest for this event.

Furthermore, a EUNADICS-AV specific rolling archive and an event archive will be set up to archive data files.

c) [Interfaces](#)

The data portal has two types of interfaces:

- Interfaces to ingest new data from data providers
- Interfaces to expose the data to external institutes, like the VAAC

This ICD focuses on the interfaces between the data portal and the data providers. The interfaces necessary to expose the data will be described in deliverable D43 [R2].

Four different groups of data providers can be distinguished:

- The **data providers from WP3**, which will provide the data which is used to create warnings and aviation products.
- The **Early Warning Center from WP5**, which will provide alerts for several airborne hazards.
- The **Aviation products providers from WP6**, which will provide air space specific model output prototype products.
- Data providers not participating in EUNADICS-AV (e.g. EUMETSAT/ESA) providing operational products used in the projects.

The relevant interfaces for the data providers will be discussed in the following sections, covering both the creation of data products and the ingestion of routine data files and alert files.

4.2 Data product creation

Data providers can offer two types of data products to the data portal, routine products and alert products. This section will cover the concepts of these data products. Furthermore, it discusses how to create these products in the data portal.

a) [Routine data products](#)

Routine data products are products which are updated regularly, not triggered by special events. The routine data files will be grouped in datasets, which are groups of data files which have a common set of metadata. The dataset can grow by ingesting new data files. Note that through the provided metadata links to the data originators and datasets are available.

Data providers which want to offer routine products should create dataset metadata entries for their data granules using the metadata editor in the data portal, if necessary with help from WP8. The following things should be specified per dataset:

- Dataset metadata (standardized INSPIRE ISO 19115 profile). An example of this can be found in appendix A.
- Dataset file format, preferably NetCDF-CF compliant

- Ingest location (FTP server) from which the data granules can be ingested

After creation of the dataset, all data granules placed in the ingest location will be ingested and added to the dataset.

b) Alert data products

Alert data products are products which are created when an event is detected, for example a volcanic eruption. For this EUNADICS-AV will reuse the SACS (WP5) alerting mechanisms for volcanic ash warnings and expand on this for the other EUNADICS-AV

. Ingestion of alert files will trigger the creation of events of a specified event type in the data portal. The event can be updated with new alert data by ingesting alert update files with the same event name.

Data providers which want to offer alert products should create an event type for their alerts in the data portal. To prevent that EUNADICS-AV reports the same event many times, new alert products need to be discussed in the appropriate work packages. The following things should be specified per event type:

- Description of the event type
- Relevant dataset collections
- Alert file format
- Ingest location (FTP server) from which the alert files can be ingested

c) Relation between alert and routine products

Routine products also play an important role during events, because they represent the data from which alerts are derived. Therefore the data portal will provide the possibility to present actual routine data as layers underneath alert data. It is likely that during a volcanic event other data is relevant than during a nuclear event. To be able to automatically couple relevant routine data to alert products, dataset collections can be coupled to event types. Figure 2 gives an overview of the relationship between alert and routine product relationships.

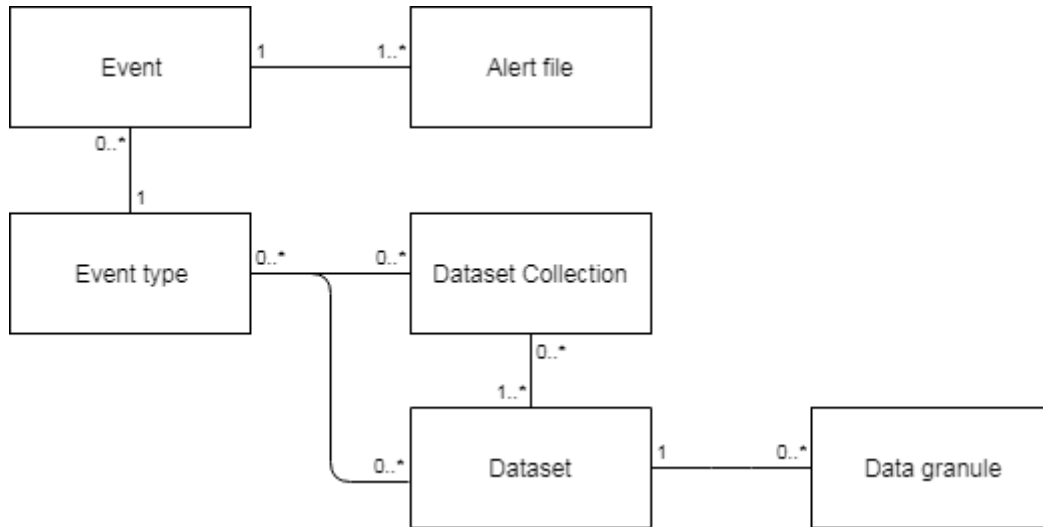


Figure 2: Relation between files and their collections

A dataset collection can be created and edited in the data portal. It will be possible to provide a description of the collection and to select existing datasets to add to the collection.

4.3 File ingestion and archive process

This section covers the ingestion and archiving of data files. The happy flow of this process for both routine products and alerts is visualized in Figure 3. The following paragraphs will describe the happy flow step by step.

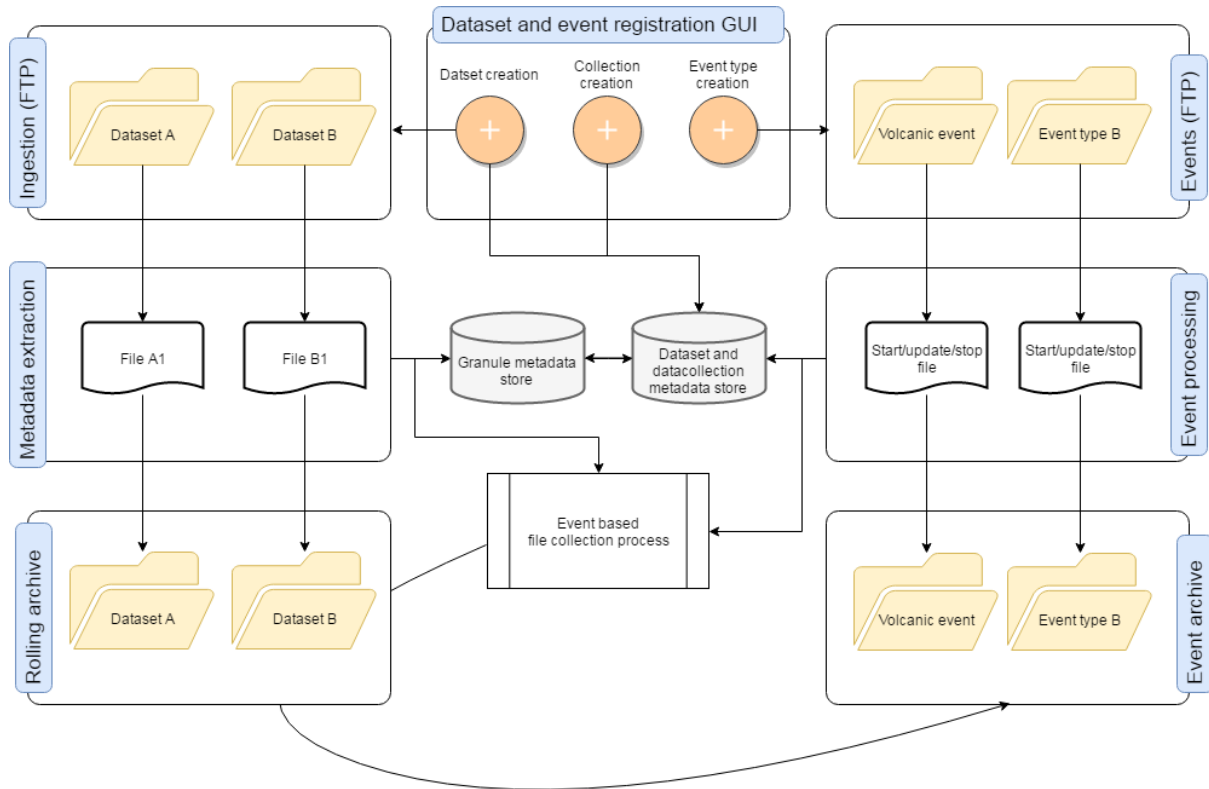


Figure 3: Dataset and alert ingestion and archiving process

a) [Happy flow routine products](#)

In order to be able to ingest data granules it is necessary to create a dataset with an FTP ingest location. After this, the data portal will scan the FTP directory for new files. As soon as a file is detected and completed, the following steps will be performed by the data portal:

1. The data granule will be downloaded to KNMI
2. Granule metadata will be extracted from the file
3. The granule metadata will be added to the granule metadata store
4. The file will be added to a rolling archive
5. The file will be removed from the rolling archive after a time window has passed

As long as the file is available in the rolling archive and it is allowed by the data policy of the dataset, it can be found, previewed and downloaded in the data portal. From the metadata the data originator is referred (and its data archive, if available).

b) [Happy flow alert products](#)

Alerts are given soon as an airborne hazard is detected. Alerts can contain information on where and when the hazard is detected, and can also be updated over time. Alerts will also trigger the distribution of relevant routine products to a permanent event archive²; this collection of data is archived for research purposes and future studies of the event. As soon as

² If allowed by the data policy of the dataset

the threat of the detected hazard has subsided, the distribution of data granules to the event archive should be stopped.

In order to facilitate the process described above, the following alert ingestion happy flow will be followed by the Early Warning Center and the data portal:

1. An alert-start file is added to the ingestion location for the corresponding event type by the Early Warning Center.
2. The data portal scans the ingestion locations and starts to ingest the alert-start file as soon as it is completed:
 - a. An event of the specified event type will be created in the data portal. This event will be identifiable with a unique name as specified in the start file, such that alert-update and alert-stop files can be linked to this event. Furthermore, this name can be used to link to data in the data portal. The name will be automatically generated at first and can be altered (see chapter 3.1, Event definition).
 - b. All current data granules in relevant dataset collections will be added to this events event archive.
 - c. All future data granules in relevant dataset collections will be added to this events event archive.
3. Zero or more alert-update files are added to the ingest location by the Early Warning Center.
 - a. The event with the same unique name will be updated
4. An alert-stop file is added to the ingestion location by the Early Warning Center.
 - a. The data portal will stop adding data granules to the corresponding events event archive.

Though an ‘alert-stop’ file is simple in concept, it will be difficult to apply in practice. Relying on ‘no detection’ in the data might cause mixing events, particularly for desert dust and forest fire events. The criterion for alert-stop will need more discussion within EUNADICS-AV.

5. Requirements and design

From the software overview in Chapter 4 it becomes clear that the data portal should be able to ingest routine data products and alerts. This chapter describes the interface requirements for this ingestion process as well as the design of the interface.

5.1 Interface requirements

The following interface requirements are split in general requirements and data product specific requirements. All data providers should comply to these requirements or explain why it is not possible. In the latter case, it is necessary to discuss a workaround.

a) General interface requirements

- [IR1] Data providers shall provide their products on a FTP server in a predefined, unique directory per dataset which can be accessed by the data portal ingestion facility.
- [IR2] Data providers and the data portal shall agree on a NRT delivery timespan for each product.
- [IR3] Data providers shall indicate that a file is uploading to the FTP directory by using a .tmp extension, as further specified in the interface design.

b) Ingestion of routine data products

- [IR4] Data providers shall create a dataset in the data portal, providing all necessary dataset metadata via the data portal metadata form.
- [IR5] Data providers shall provide data files in a NetCDF CF-compliant format as further specified in the interface design.
- [IR6] Data providers shall provide data products with a unique filename per data granule.

c) Ingestion of alerts

- [IR7] Data providers³ shall provide alerts in the format as specified in the interface design.
- [IR8] Data providers shall provide alerts with a filename structure as specified in the interface design.
- [IR9] Data providers shall provide at least a file marking the start of an alert and the end of an alert. Data providers may provide zero to more updates to alerts.
- [IR10] Data providers shall provide a unique identifier for each alert. The data portal shall provide a link to relevant alert data based on this unique identifier.

5.2 Interface design

In this section all the interface requirements which require extra explanation are discussed in more details. Note these interface requirements are applicable for EUNADICS-AV data providers. If external data providers (e.g. EUMETSAT) cannot comply to the proposed interfaces, separate agreements will need to be made.

³ These will be provided by WP5 (based on observations)

a) [FTP file upload process](#)

The data portal will regularly scan the FTP ingest directory for new files to ingest. To prevent ingestion of files which are not completely uploaded yet, it is necessary that the data provider signals that a file is still uploading. This is accomplished by using a .tmp extension for a file which is still uploading, resulting in the following happy flow:

1. A data provider wants to upload the file *fileX.nc*
 - a. The data provider renames the file to *fileX.nc.tmp*
 - b. The data provider adds the .tmp file to the FTP ingest location
 - c. The data portal will not start to ingest this file due to the .tmp extension
2. The data provider notices the file upload is completed
 - a. The data provider renames the file back to *fileX.nc*
 - b. The data portal will start to ingest the file

b) [NetCDF CF compliancy](#)

To be able to automatically extract granule metadata and provide data services like WMS and OPeNDAP, it is necessary that files are delivered in a NetCDF-CF compliant file format. The NetCDF-CF conventions can be found on the cf-conventions website [A1]. Both the description of the format as file format checkers can be found there.

c) [Alert file format and filename structure](#)

In order to be able to correctly ingest alert files, the following format should be followed for alert start files, update files and stop files:

Start file:

Filename format: start_<event_type>_<unique_notification_name>_<time>

File type: NetCDF CF-compliant

Contents:

- Contours of affected area
- Min/Max/Average location
- All pixels of the affected area

Update file:

Filename format: update_<event_type>_<unique_notification_name>_<time>

File type: NetCDF CF-compliant

Contents:

- Contours of affected area
- Min/Max/Average location
- All pixels of the affected area

Stop file:

Filename format: stop_<event_type>_<unique_notification_name>_<time>

File type: irrelevant

Contents: None

d) NRT specifications

The NRT delivery and processing of products is an important aspect of providing timely and useful alerts. *NRT (Near Real Time) is a time delay introduced by automated data processing or network transmission between the occurrence of an event and the use of the processed data. Typically – for meteorological purposes – this is within 3 hours from sensing.* [Definition from CEOS-WGISS Glossary of Acronyms and Terms, v1.0]