# D3.1 Data Pipeline

| | |
|---|---|
| **Deliverable ID:** | **D3.1** |
| **Dissemination Level:** | **PU** |
| **Project Acronym:** | **FARO** |
| **Grant:** | **892542** |
| **Call:** | **H2020-SESAR-2019-02** |
| **Topic:** | **Safety and Resilience** |
| **Consortium Coordinator:** | **CRIDA** |
| **Edition date:** | **20/11/2020** |
| **Edition:** | **00.02.00-PUBLIC** |
| **Template Edition:** | **02.00.02** |

Founding Members
EUROPEAN UNION EUROCONTROL
SESAR JOINT UNDERTAKING

## Authoring & Approval

### Authors of the document

| Name/Beneficiary | Position/Title | Date |
|---|---|---|
| Miguel García Martínez / CRIDA | Project Contributor | 25/11/2020 |
| Icíar García-Ovies / CRIDA | WP2 Lead | 16/10/2020 |
| Christian Verdonk Gallego / CRIDA | Project Coordinator | 16/10/2020 |
| Javier Garcia Moreno / CRIDA | Project Contributor | 16/10/2020 |

### Reviewers internal to the project

| Name/Beneficiary | Position/Title | Date |
|---|---|---|
| Carlo Dambra / ZENABYTE | WP3 Lead | 25/11/2020 |
| Luca Oneto / ZENABYTE | Project Contributor | 16/10/2020 |
| Irene Buselli / ZENABYTE | Project Contributor | 16/10/2020 |
| Christian Verdonk Gallego / CRIDA | Project Coordinator | 26/11/2020 |
| Anthony Smoker / LUND UNIV. | WP5 Lead | 16/10/2020 |
| Fernando Gómez Comendador / UPM | Quality Manager | 16/10/2020 |

### Approved for submission to the SJU By - Representatives of beneficiaries involved in the project

| Name/Beneficiary | Position/Title | Date |
|---|---|---|
| Carlo Dambra / ZENABYTE | WP3 Lead | 25/11/2020 |
| Christian Verdonk Gallego / CRIDA | Project Coordinator | 26/11/2020 |
| Anthony Smoker / LUND UNIV. | WP5 Lead | 29/10/2020 |
| Fernando Gómez Comendador | Quality Manager | 29/10/2020 |

### Rejected By - Representatives of beneficiaries involved in the project

| Name/Beneficiary | Position/Title | Date |
|---|---|---|
| | | |
| | | |

## Document History

| Edition | Date | Status | Author | Justification |
|---|---|---|---|---|
| 00.00.01 | 28/08/2020 | Document Creation | Miguel García | TEMPLATE WITH TOC |
| 00.00.02 | 28/09/2020 | Draft | Miguel García | TOPIC 3 |
| 00.00.03 | 16/10/2020 | Final Draft | Miguel García | Draft completed |
| 00.01.00 | 27/10/2020 | Final Version | Miguel García | Comments reviewed |
| 00.02.00 | 20/11/2020 | Final Version | Miguel García | SJU Assessment reviewed |

# FARO

SAFETY AND RESILIENCE GUIDELINES FOR AVIATION

This deliverable is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 892542 under European Union's Horizon 2020 research and innovation programme.

## Abstract

This document describes the FARO Data Hub (FDH) pipeline created to support the research & development activities during the SESAR Joint Undertaking project FARO. This FDH is the pipe through which data will be collected, stored and transferred for all FARO research activities and this document describes its architecture, organisation and connectivity. Additionally, this document provides a detailed description of the data sources introduced into the FDH as raw materials for the research activities of FARO. After reading this document any member of the project should be able to upload/download datasets to/from the FDH and to understand the data good enough to be able to work with them on their researches. This pipeline is implemented in order to allow secure, efficient and reliable access to all the provided data sources, assuring the correct development of FARO tasks related to data.

# Table of Contents



# List of Tables



# List of Figures

# 1 Introduction

## 1.1 Purpose of the document

This document describes the FARO DATA HUB (FDH) to be used during the FARO project. The pipeline is implemented in order to allow a secure, efficient, scalable and reliable access to all the provided data sources. Its main goal is to maximise the data flow throughput to ease research and development of the use cases to be defined by FARO project members. This deliverable has been developed within the activities under "T3.1 Data Acquisition and Pipeline" [1].

This document covers how the data will be acquired, stored and transferred from a technical point of view as well as from the final user (FARO WPs) perspective. Additionally, this document includes a detailed description of some data sources used as raw materials of the FARO project. Moreover, the contents of this deliverable and results derived from the transformation of the ATM data are developed to comply with the Data Collection Objective (DCO) 3 and in compliance with DMO2 of [2].

Finally, it is worth noting that this document describes the data available at an early stage of the FARO project (M6) and that this description will be complemented with the issue of Deliverable D3.3 "Final data set" to be released at M15 (i.e. at a stage in which the data needs from WP4, WP5 and WP6 will be consolidated).

## 1.2 Intended readership

This document is intended to be used by FARO project members.

## 1.3 Acronyms and Terminology

| Term | Definition |
|---|---|
| **AIP** | **Aeronautical information publication** |
| **ANS** | Air Navigation Service |
| **ANSP** | Air Navigation Service Provider |
| **API** | Application Programming Interface |
| **ATM** | Air Traffic Management |
| **ATC** | Air Traffic Control |
| **ATCo** | Air Traffic Controller |
| **AU** | Airspace User |
| **AUA** | Air Traffic Control Unit Airspace |
| **CEANITA** | Comisión de Estudio y Análisis de Notificaciones de Incidentes de Tránsito Aéreo |

| Term | Definition |
|---|---|
| **CSV** | comma-separated values |
| **DEN** | Diario de Novedades |
| **DWH** | Data WareHouse |
| **ETL** | Extract, Transform and Load |
| **FDH** | FARO Data Hub |
| **FIR** | Flight Information Region |
| **FTP** | File Transfer Protocol |
| **GIS** | Geographical Information System |
| **IFS** | Informe de Seguimiento |
| **IPV** | Informe de Plan de Vuelo |
| **NAS** | Network Attached Storage |
| **OCR** | Optical Character Recognition |
| **PDF** | Portable Document Format |
| **RADAR** | RAdio Detection And Ranging |
| **RAID** | Redundant Array of Independent Disks |
| **SACTA** | Sistema Automatizado de Control de Tráfico Aéreo |
| **SESAR** | Single European Sky ATM Research Programme |
| **SFTP** | SSH File Transfer Protocol |
| **SJU** | SESAR Joint Undertaking (Agency of the European Commission) |
| **SJU Work Programme** | The programme which addresses all activities of the SESAR Joint Undertaking Agency. |
| **SESAR Programme** | The programme which defines the Research and Development activities and Projects for the SJU. |
| **WEBDAV** | Web Distributed Authoring and Versioning |
| **WP** | Work Package |

# 2 Data Hub

## 2.1 Overview

The objective of the FARO Data Hub (FDH) is to provide constant bi-directional flows of data between the FARO data store and the different WPs. The system keeps high rates of security and speed of transactions to ensure the needs of the FARO project. At the same time, because of the nature of this big data project, goals and requirements may change and further adjustments may be required accordingly, especially when talking about scalability.

With all this in mind, the FDH design uses technologies that ease the scalability of the system keeping any possible change transparent to the user just in case that new, higher performance requirements arise. A first version of 1 TB will be established with the option to raise the storage to more than 10 TB.

## 2.2 Schema

This document explains the data access mechanism (FARO Data Hub - FDH) for secure and reliable access to data, ensuring both integrity of data and accessibility for all the WPs. The core of the FDH is the FARO data store where all data travelling through the pipeline will be hosted.

- Architecture:

The data store is mainly comprised by a NAS Server on CRIDA premises overlaid by an open source cloud solution software (OwnCloud) that provides a flexibility close to the best well-known cloud providers such as Amazon or Microsoft. This ensures the scalability and security of the FDH to store data without hosting confidential data in third parties infrastructures.

- Structure:

A first structure for the FARO data store is proposed in the following schema. This could be further refined in order to satisfy WPs needs.

***FDH ->***

- ***DATASOURCES:*** Raw materials. All the data sources acquired for the project. Organised by origin, for instance: CEANITA REPORTS; CRIDA DWH
- ***DATASETS:*** Processed and prepared datasets for a specific goal, i.e.: a use case, an algorithm or a substantive investigation. Organised by goal, for instance: UC1_DatasetX
- ***WPs:*** Any kind of output from a specific WP. Organised freely by WP.
- ***Results:*** Results of the project may include datasets, visualisations, reports, etc.

## 2.3 Access to data

Direct access to the data is provided by the cloud connection layer. This layer provides several ways to connect, as described in Section 4 - Usage. The FDH does not require an API, and it can be used just as a file system hierarchy with different access roles (2.4 Stakeholders Roles for roles and permissions).

### 2.3.1 Data Sources

The FDH will provide raw materials downloadable by separate files organised on folders identifying each data source. All sources will be located under the "\Data Sources" folder. The data sources can be any kind of files. Usually they will be tabular data as CSV (Comma Separated Values) files but others formats could be found such as images or PDF files. Sources can be or not be aligned on the space and the time to others, as unfortunately not all sources have the same availability. To keep sources manageable, the files will be split by ranges of time with limits depending on the size of the data source. Folders will be named by the period covered by data for instance "20160101-20160131-XXXX".

### 2.3.2 Data Sets

The FDH will make datasets downloadable by separate files organised on folders identifying each data set and the use case related. All datasets will be located under the "\Data Sets" folder. The datasets will be aligned on the space and the time by use case, and to keep sources manageable, the files will be split by ranges of time with limits depending on the size of the data source. Files and folders will be named by the period covered by data and the related use case, for instance "UC1_DATASETX_20160101-20160131-XXXX".

### 2.3.3 Work Packages

Machine learning inputs and outputs vectors coming from the WPs will be located on "\WPs" folder, as well as any kind of artefact that WPs may produce. Here the WPs will share their outputs and feedback to other WPs as input. Each WP will define the interfaces to communicate with the modules developed as part of their work. Any kind of artefact could be produced by WPs, for instance new datasets or machine learning trained models. All artefacts will be organised by WPs.

### 2.3.4 Results

In addition to the output of each WPs, further results could be produced at the final stages of the project as final conclusions of the FARO platform, including applications or visualizations to address the specified use cases and analysis to prove their efficiency. There is not a specific format defined for these products, just they will be located on the "\Results\XX" folder.

## 2.4 Stakeholders Roles

Key stakeholders that have been identified to be granted access to the FDH are as follows:

**Table 1 - Stakeholders roles and permissions**

| Stakeholder | Roles | Description |
|---|---|---|
| Data Providers | Data Sources: Read & Write<br>Data Sets: None<br>WPs: None<br>Results: None | Providers of data onFARO project, mainly CRIDA and ENAIRE |
| WPs | Data Sources: Read<br>Data Sets: Read & Write<br>WPs: Read & Write<br>Results: Read & Write | WPs will share results and feedback each other with visualizations and results of machine learning algorithms. |

## 2.5 Background

### 2.5.1 Network & Server

The FDH will be hosted on CRIDA premises over a CentOS 7 mounting a RAID 1 of 10 TB. The cloud system serving the FDH is connected to the internet by an optical fibre network of symmetrical 400 Mbit/s bandwidth.

### 2.5.2 Security

A comprehensive encryption architecture is provided with three levels of encryption: In transit, at rest and end-to-end. By default, In-transit encryption will be the only level active by using HTTPS connections and the latest TLS protocol. Files can be encrypted either server-wide or on a per-user basis. For further uses, the others levels of security could be activated, at rest where files can be encrypted either by master key or user specific keys and at End-to-End where files are encrypted End-to-End for instance using a hardware smart key. First level is considered secure enough for the FARO project to guarantee security at network transport level.

To guarantee security at application level, the cloud software provides built-in ransomware protection and each personal member of the project will require personal and non-transferable credentials provided by the FDH administrator (CRIDA) to access to the data. Brutal-force attacks are prevented by the built-in authentication software, requiring strong passwords and enabling two- or multi-factor authentication (e.g., a time-based one-time password), which can, theoretically, make brute-force attacks impossible.

In addition, apart from the data being saved on this restricted-access server on CRIDA's facilities, once the data is transferred to other members of the consortium, these shall provide adequate and commensurate measures to maintain the security of the data.

### 2.5.3 Backup & Integrity

The cloud software provides automatic backup services and RAID-1 mirroring provides guarantee of files integrity keeping two copies of the data on different physical hardware. Any write to the volume writes to both disks; a read can be satisfied from either, so if one drive fails, the data is still available on the other drive.

# 3 Data Pool

## 3.1 Data Life Cycle

The FARO Data Hub has a life cycle for the data that flows through it. This cycle includes from the first acquisition of data to the last result produced by the WPs. After the data is acquired, the cycle in the pipeline has an iterative part with these steps:

- Clean – Filter: Locate and remove data with errors or out of scope.
- Merge: Merge data of interest by business relationships
- Quality and Requirements Assessment: Check if the data provided is what use cases require to be solved.
- Modelling: Some model outputs could be inputs of other models so this stage is some times in the iteration.
- Feature Engineering: Calculation of new information from the available sources.

Finally, datasets for each use case are published to produce models and final results of the WPs.

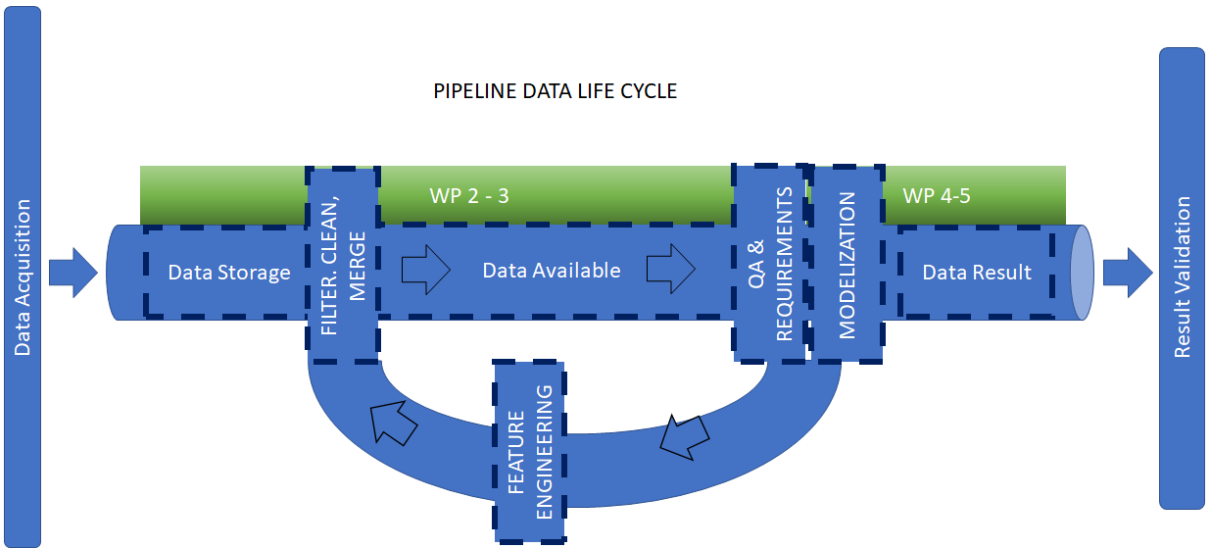


**Figure 1 - FARO DATA HUB PIPELINE**

After the end of the project the life cycle will stop to iterate and the FDH will be frozen cleaning temporal datasets but keeping the WPs datasets produced available for further works such as research papers of project members. To guarantee integrity and confidentiality of the data, the FDH will be completely stored in EnAire/CRIDA premises since they are the main data providers and the only stakeholder providing confidential data. To assure the security, the online interface of the FDH will be shut down and the data will be then stored on a non-internet accessible drive.

At any case, publications related to FARO project will have to guarantee meeting the provisions gathered in the DMP [2], including those related to the de-identification and de-localisation of raw, aggregated and derived data. Additionally, the consortium members remain subject to the provisions of the FARO Consortium Agreement in relationship with the Dissemination of Results (Section 8.3) and Non-disclosure of information (Section 10).

In the case of publications related to FARO project that is using public data only, such as public safety reports processed with Natural Language Processing techniques, the requirement for publications will remain almost the same than for private data. That is, all members have to give their consent but no further anonymisation will be required from the level of anonymisation that the public original source had accomplished.

## 3.2 Data Acquisition

Data for the FARO project comes from several sources, both public and private. One of the main sources is the ENAIRE-CRIDA data warehouse (DWH), a private and confidential structured database that contains high-granularity structured ATM data such as flight plans, flight tracks and airspace volumes as well as ATM processed information such as environmental, flight performance or traffic flows information of the Spanish airspace. The DWH is composed of around 40.000.000.000 entries of structured data distributed in 539 tables. This data is exploited through the PERSEO web application that ENAIRE exploits for internal monitoring [5]. The acquisition process to fill the DWH includes the extraction of the information from different sources using complex ETLs jobs (out of the scope of FARO). Among them, and not limited to:

- The IFS (Informe de Seguimiento – Monitoring Report) and the IPV (Informe de Plan de Vuelo – Flight Plan Report) from the PALESTRA platform provided by ENAIRE. PALESTRA records data produced by the SACTA system used by air traffic controllers on the Spanish airspace.
- The INSIGNIA (GIS) web services provided by ENAIRE where all the AIP information has been digitised and published for the Spanish Airspace.
- The DEN (Diario de Novedades –Novelties Journal) web services provided by ENAIRE where all the sectorisations and sectors capacities (among other information) are notified for the Spanish Airspace.

To acquire the data for the FARO, simple ETL jobs are created to extract data from tables and dump them into the pipeline in CSV (comma-separated values) files.

About public sources, the FARO project will use CEANITA (Comisión de Estudio y Análisis de Notificaciones de Incidentes de Tránsito Aéreo) reports published on the internet [3]. In this case the source is a public web page of the Spanish administration (CEANITA Website) where reports are listed. A simple ETL job is created to scrap the web page and automatically download all the listed reports as PDFs files. Later, these PDFs files that are published on image mode are processed by OCR (Optical character recognition) algorithms and then introduced into the pipeline as plain text files.

Finally, EUROCONTROL has released a data archive for research and development purpose available just to direct download from their repository as compressed tabular data files and the FARO project will include them into the FDH data store.

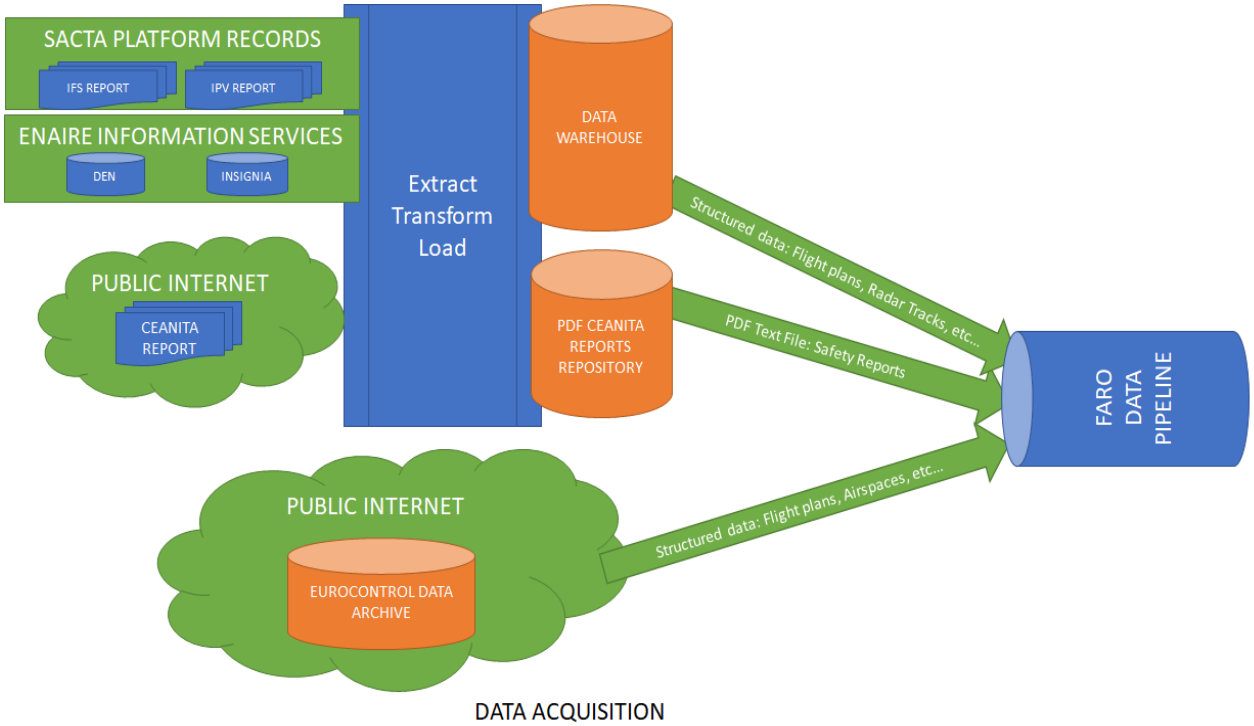


**Figure 2 - Data acquisition**

## 3.3 Data Sources

### 3.3.1 ENAIRE-CRIDA DWH

Here the structured entities provided by the ENAIRE-CRIDA DWH will be described. All entities are identified and linked between each other by internal unique keys. For confidentiality sake, the examples shown on this document are fake data illustrating the possible values of each field. Additionally, if any result involving this source is published the sensible data will be anonymised and reviewed for approval by the consortium before release.

**Confidential text. Removed for the public version of the deliverable.**

### 3.3.2 CEANITA REPORT

CEANITA reports are public reports published by the AESA (Agencia Estatal de Seguridad Aérea - Spanish Safety Aviation Agency) website under the commission CEANITA (Comisión de Estudio y Análisis de Notificaciones de Incidentes de Tránsito Aéreo – Commission for Study and Analysis of Notifications of Incidents of Air Traffic). This commission releases public studies over safety incidents that have been notified to the AESA. The content of these reports is in Spanish.

#### 3.3.2.1 CONTENT

This section describes the content of a report by sections.

## Location of the Incident (*Ubicación del Incidente)*

**2. UBICACIÓN DEL INCIDENTE**

1 Fecha: 28/03/2019 Hora (UTC): 10:54 2

3 Notificado por: ACC Sevilla / ACC Barcelona / Aeronave 1

4 Condiciones meteorológicas: VMC

5 Espacio aéreo: UIR Madrid Clase: C 6

7 Localización: Aproximadamente a 7 NM al SW del punto RESTU

8 Unidades ATS: 1. ACC Barcelona Sector LVS 2. ACC Sevilla Sector CEN 3. ACC Sevilla Sector SUR

**Figure 3 - Location of the incident (CEANITA)**

The report starts with an overview of the reported Safety-related Event. This information provides important data of the incident such as:

- Spatial Location: (7)
- Conflict Time: (1) and (2)
- Operational Sector Active: Three units (5), (6) and (8)
- Type of visual conditions: (4)
- Who reported the incident: (3)

## Description of the traffics (*Descripción de los tráficos)*

**3. DESCRIPCIÓN DE LOS TRÁFICOS**

1 **TRÁFICO 1** Aeronave Otro

2 Indicativo: Aeronave 1 Procedencia: LEBL 3 Destino: LEAM 4

5 Modelo de avión: B738 Reglas de vuelo: IFR 6

7 Tipo de operación: Comercial Tipo de vuelo: Pasajeros 8

**TRÁFICO 2** Aeronave Otro

Indicativo: Aeronave 2 Procedencia: LIRF Destino: SBGR

Modelo de avión: B763 Reglas de vuelo: IFR

Tipo de operación: Comercial Tipo de vuelo: Pasajeros

**TRÁFICO 3** Aeronave Otro

Indicativo: Aeronave 3 Procedencia: LEMO Destino: OETB

Modelo de avión: B744 Reglas de vuelo: IFR

Tipo de operación: Comercial Tipo de vuelo: Pasajeros

**Figure 4 - Traffic description (CEANITA)**

A list of the involved flights is provided in this section:

- Flight report identification (1)
- Callsign (de-identified) (2)
- Origin (3)
- Destination (4)
- Aircraft Type (5)
- Flight Rules (6)
- Type of Operation (7)
- Type of Flight (8)

## General Aspects (*Generalidades*)

A general description of the incident is then provided.

**4.1. Generalidades**

Se ha producido una situación a consecuencia de la cual ACC Sevilla, ACC Barcelona y la Aeronave 1 han notificado un acercamiento entre la Aeronave 1 y la Aeronave 2. Las aeronaves se encontraban aproximadamente a 7 NM al SW del punto RESTU, en el FIR Madrid, espacio aéreo clase C. Las condiciones meteorológicas eran VMC.

**Figure 5 - General aspects (CEANITA)**

## Initial Situation (*Situación inicial*)

In this section the start point of the incident is described using natural language and images (RADAR snapshots). The Figure above shows the radar display of the Controller Working Position (CWP) at the beginning of the scenario described in a CEANITA Report. The location and situation of the involved flights are detailed (i.e.: Flight level, heading, etc.)

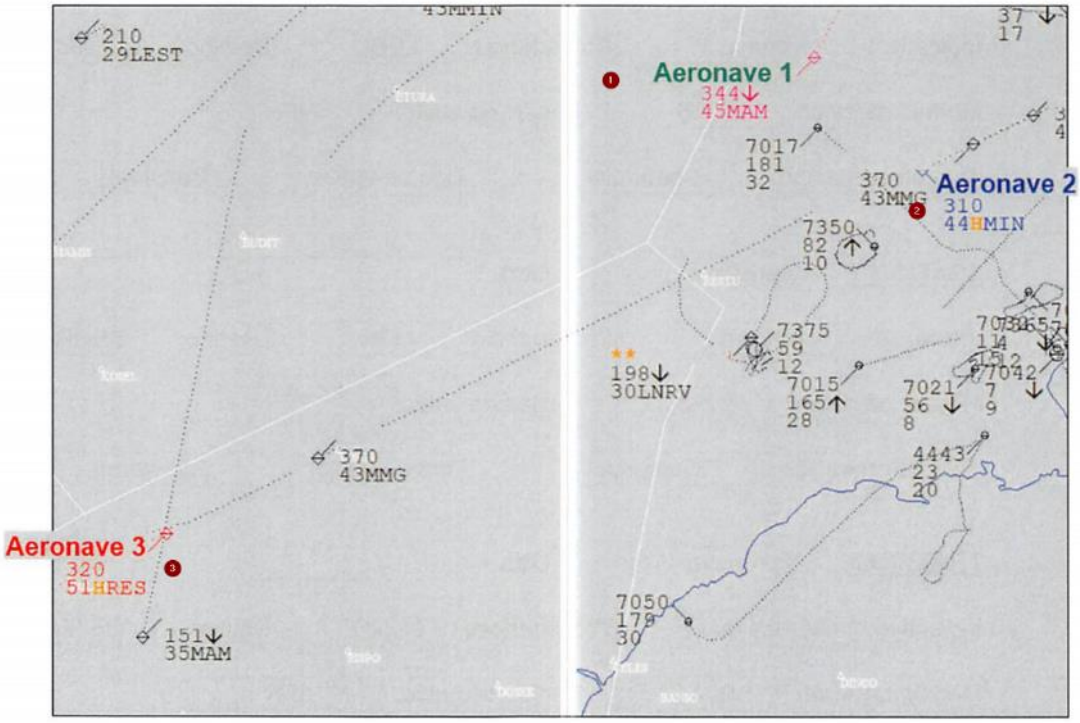


*Fig. 1 –Posición de las aeronaves en la situación inicial.*

**Figure 6 - RADAR Snapshot of the beginning of the incident**

## Communications and Radar Tracks (*Comunicaciones y trazas radar*)

In this section the communications of interest between ATCos and Pilots are transcribed (not literally) using natural language. They are supported by RADAR snapshots to describe the evolution of the incident and the interactions between all the parts.

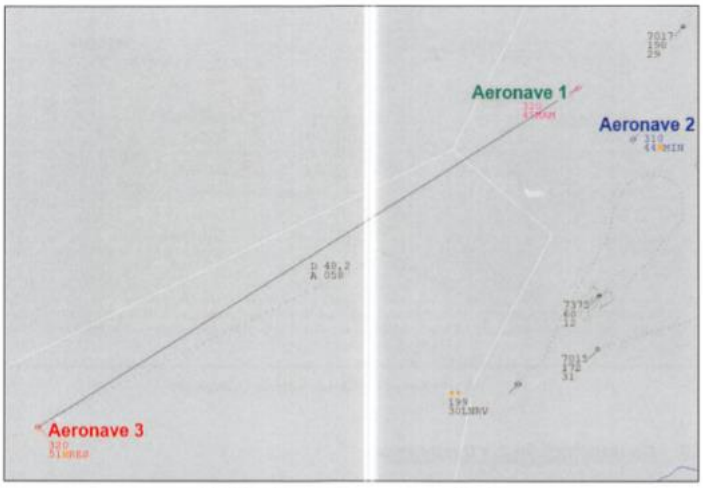


*Fig. 3 –Posición de las aeronaves a las 10:44:59.*

**Figure 7 - RADAR Snapshot of the evolution of the incident**

## Extract from received reports. (*Extracto de informes recibidos*)

This section describes using natural language the facts and statements of interest extracted from different reports received by the commission during the investigations. Examples of these reports could be reports from the pilot, the co-pilot or the executive controller.

## Events and factors of interest (*Eventos y factores contribuyentes*)

This section contains a summary of the events and facts discovered during the report previous sections. Is usually provided as a compact table with a vertical chronologic axis.

## Incident Rating (*Calificación del incidente*)

This section rates the severity and the repeatability of the incident using safety ATM metrics and indicators such as the rate of closure or the risk of collision. As summary it indicates:

- Incident type (1)
- ATM contribution (2)
- Pilot contribution (3)
- Minima Separation (4)
- Severity (5)
- Repeatability (6)

6. CALIFICACIÓN DEL INCIDENTE

| | | | | | |
|---|---|---|---|---|---|
| 1 | TIPO DE INCIDENTE: | Autorización errónea de ATC | Detección y resolución de conflictos | Procedimientos ATM | |
| | | Vulneración separación mínima | | | |
| 2 | CONTRIBUCIÓN DEL ATM: | Directa | | | |
| 3 | CONTRIBUCIÓN DEL PILOTO: | Sin contribución | | | |
| 4 | SEPARACIÓN MÍNIMA PRESCRITA: | V (ft) 1000 | H (NM) 5 | | |
| | SEPARACIÓN MÍNIMA ALCANZADA: | V (ft) 400 | H (NM) 2,4 | | |
| 5 | SEVERIDAD: | C-Incidente Significativo | | | |
| 6 | REPETITIVIDAD: | 4 | | | |

Reference number: 04/10/2018 931
Occurrence type: More than one aircraft
Description: AIRPROX
Occurrence date: 04/10/2018
Occurrence time: 11:32:00

| | | | | | |
|---|---|---|---|---|---|
| A1 | B1 | C1 | E1 | D1 | N1 |
| A2 | B2 | C2 | E2 | D2 | N2 |
| A3 | B3 | C3 | E3 | D3 | N3 |
| A4 | B4 | C4 | E4 | D4 | N4 |
| A5 | B5 | C5 | E5 | D5 | N5 |
| A | B | C | E | D | N |

ATM | ATM Ground | ATM Final | ATM Ground Final

RELIABILITY FACTOR:
OVERALL: 100%
OVERALL SEVERITY: 100%
OVERALL REPEATABILITY: 100%
ATM GROUND: 100%
ATM GROUND SEVERITY: 100%
ATM GROUND REPEATABILITY: 100%

FINAL VALUES
Risk ATM : C4
Justification :
Risk ATM Ground : C4
Justification :

| Severity | | | | | | |
|---|---|---|---|---|---|---|
| Criteria | Recorded Value ATM Ground | | Recorded Value ATM Airborne | | Value ATM | Description |
| Risk of collision | | | | | | |
| Separation | 26-50% | 5 | | | 5 | (1) |
| Rate of Closure | Low (<=85knots, >=1000ft min) | 1 | | | 1 | (2) |
| Controllability | | | | | | |
| Conflict Detection | Potential Conflict detected | 0 | | | 0 | (3) |
| Plan | Plan INADEQUATE | 1 | | | 1 | (4) |
| Execution | Execution INADEQUATE | 1 | Execution CORRECT | 0 | 1 | (5) |
| STCA | Not applicable | 0 | | | 0 | (6) |
| Recovery | Recovery INADEQUATE | 5 | Recovery CORRECT | 0 | 5 | (7) |
| TCAS / Own Initiative see and avoid | Not applicable | 0 | Not applicable | 0 | 0 | (8) |
| Pilot Reaction | | | Not applicable | 0 | 0 | (9) |
| Total | | 17 | | 0 | 17 | |

| Repeatability | | | | | | |
|---|---|---|---|---|---|---|
| Criteria | Recorded Value ATM Ground | | Recorded Value ATM Airborne | | Value ATM | Description |
| Systemic Issues | | | | | | |
| Procedures | Not applicable | 0 | Not applicable | 0 | 0 | (10) |
| Equipment | Not applicable | 0 | Not applicable | 0 | 0 | (11) |
| Human Resources Management | Not applicable | 0 | Not applicable | 0 | 0 | (12) |
| Non-systemic / Human Involvement Issues | | | | | | |
| Other Contributing Factors | Non-Systemic / Human Involvement Issues without Contextual Conditions | 8 | Not applicable | 0 | 8 | (13) |
| Window of Opportunity | | | | | | |
| Method | Normal | | | | | |
| Situation | Daily Routine | 7 | | | 7 | (14) |
| Total | | 15 | | 0 | 15 | |

150/18- Pág.14 de 16

**Figure 8 - Rating of the incident**

## Conclusions (*Conclusiones*)

This section describes the conclusions of the investigation of the agency using natural language.

## Regulation and applicable procedures (*Normativa y procedimientos aplicables*)

This section lists the regulations of interest for the process of the incident conclusions and responsibilities.

## Stakeholders (*Partes afectadas*)

This section lists the stakeholders who should be notified by the result of the investigation.

Recommendations (*Recomendaciones*)

This section includes special recommendations from the agency to stakeholders.

#### 3.3.2.2 Range of available dates and areas

Available for the Spanish airspace, from 2017 to 2019.

### 3.3.3 EUROCONTROL R&D DATA ARCHIVE

EUROCONTROL releases data sets for research and development purposes. They contain traffic datasets covering all historic commercial flights in four fixed, sample months of specific years, while remaining within the restrictions of use placed by EUROCONTROL. The sample months – March, June, September, December – give broad access to seasonal patterns and large-scale data (several million flights). This meets research needs while remaining consistent with EUROCONTROL's data policy. The data are provided with no quality guarantees, and no support beyond the provision of the metadata and description of the dataset fields [4].

FARO members will ensure the compliance of EUROCONTROL terms and conditions summarized in:

- To use the ATM Dataset only for research & development
- To not share or distribute the ATM Dataset
- To acknowledge EUROCONTROL as the source of the ATM Dataset
- To be aware of and accept the fact that EUROCONTROL provides the ATM Dataset as-is, without warranties of any kind.

#### 3.3.3.1 FLIGHTS

This entity describes a flight and its details from EUROCONTROL Network Manager Flight plans in PRISME Data Warehouse (DWH)

**Table 2 - Flight (EUROCONTROL)**

| Field | Description |
|---|---|
| ECTL_ID | Unique numeric identifier for each flight in EUROCONTROL PRISME DWH |
| ADEP | ICAO airport code for the departure airport of the flight. The ICAO airport code or location indicator is a four-letter alphanumeric code designating each airport around the world. These codes are defined by the International Civil Aviation Organization, and published in ICAO Document 7910: Location Indicators. |
| ADEP Latitude | Latitude of departure airport in decimal degrees. |
| ADEP Longitude | Longitude of departure airport in decimal degrees. |

| Field | Description |
|---|---|
| ADES | ICAO airport code for the destination airport of the flight. The ICAO airport code or location indicator is a four-letter alphanumeric code designating each airport around the world. These codes are defined by the International Civil Aviation Organization, and published in ICAO Document 7910: Location Indicators. |
| ADES Latitude | Latitude of destination airport in decimal degrees. |
| ADES Latitude | Latitude of destination airport in decimal degrees. |
| Filed Arrival Time | Time of arrival (UTC) based on the last filed flight plan. It is the time at which the aircraft will land at the aerodrome according to the planned profile calculated for the flight. |
| Actual Off-Block Time | Off-Block Time (UTC) based on the ATFM-updated flight plan. The time that an aircraft departs from its parking position. This time may be known from flight data updates received by NM, or in the absence of such updates may be calculated from the known take-off time minus a standard taxi time value for the |
| Actual Arrival Time | Time of arrival (UTC) based on the ATFM-updated flight plan. It is the time at which the aircraft lands at the aerodrome, reflecting the best picture that NM has based on available radar updates, ATFM messages received etc. |
| AC Type | The ICAO aircraft type designator is a two-, three- or four-character alphanumeric code designating every aircraft type that may appear in flight |
| AC Operator | Three-letter ICAO operator code. Aircraft operator codes are defined by ICAO and published in Document 8585. If the operator is unknown, not provided in |
| AC Registration | Aircraft registration. In accordance with the Convention on International Civil Aviation, all civil aircraft must be registered with a national aviation authority |
| ICAO Flight Type | ICAO Flight Type: S – Scheduled, N - Non-scheduled commercial operation |
| STATFOR Market Segment | Market segment definitions can be found in http://www.eurocontrol.int/sites/default/files/content/documents/official-documents/facts-and-figures/statfor/statfor-market-segments-rules-for-sid-2016- definition.xls |
| Requested FL | Requested cruising flight level from the flight plan. |
| Actual Distance Flown (nm) | Distance flown in nautical miles, corresponding to the 'actual' profile below. |

### 3.3.3.2 FLIGHT POINTS

This entity lists the flight flown points in planned and actual version.

**Table 3 - Flight point (EUROCONTROL)**

| Field | Description |
|---|---|
| ECTL_ID | As in Flights file above |
| Sequence Number | Numeric sequence number of the points crossed by the flight in chronological order.<br><br>(Points can be not only known named waypoints, navaids, etc. but also intermediate |
| Time Over | Time (UTC) at which the point was crossed |
| Flight Level | Altitude in flight levels at which the point was crossed |
| Latitude | Latitude in decimal degrees |
| Longitude | Longitude in decimal degrees |

### 3.3.3.3 FLIGHT AIRSPACES

This entity lists the flight flown airspaces in planned and actual mode. Two types of airspaces are listed FIRs and AUAs.

#### FIR

**Table 4 - FIRs (EUROCONTROL)**

| Field | Description |
|---|---|
| ECTL_ID | As in Flights file above |
| Sequence | Numeric sequence number of the airspace entered by the flight in chronological |
| FIR ID | The identifier of the FIR |
| Entry Time | Time (UTC) the flight entered the airspace |
| Exit Time | Time (UTC) the flight exited the airspace |

#### AUA

**Table 5 - AUAs (EUROCONTROL)**

| Field | Description |
|---|---|
| ECTL_ID | As in Flights file above |
| Sequence | Numeric sequence number of the airspace entered by the flight in chronological |
| AUA ID | The identifier of the AUA |

| | |
|---|---|
| Entry Time | Time (UTC) the flight entered the airspace |
| Exit Time | Time (UTC) the flight exited the airspace |

#### 3.3.3.4 AIRAC CYCLE ENVIRONMENT DATA

This entity lists the AIRAC (Aeronautical Information Regulation And Control) cycle providing the valid dates for each one of them.

**Table 6 - AIRAC (EUROCONTROL)**

| Field | Description |
|---|---|
| External ID | Unique Id |
| Date From | Numeric sequence number of the airspace entered by the flight in chronological |
| Date To | The identifier of the AUA |

#### 3.3.3.5 ROUTES ENVIRONMENT DATA

This entity lists the waypoints that conforms routes and airways.

**Table 7 - Routes (EUROCONTROL)**

| Field | Description |
|---|---|
| Route ID | Unique route identifier. According to ICAO Annex 11 basic designators for ATS routes shall consist of a maximum of five, in no case exceed six, alpha/numeric characters in order to be usable by both ground and airborne automation systems. The designator shall indicate the type of the route: high/low altitude, specific airborne navigation equipment requirements (RNAV), aircraft type using the route primarily or exclusively. |
| Sequence Number | Numeric sequence number of a point on the route |
| Latitude | Latitude in decimal degrees of a point on the route |
| Longitude | Longitude in decimal degrees of a point on the route |

#### 3.3.3.6 FIRs ENVIRONMENT DATA

This entity lists and describes geographically the flight information regions.

**Table 8 - FIR Environment (EUROCONTROL)**

| Field | Description |
|---|---|
| Airspace ID | Unique identifier of the FIR (could also be a UIR, Upper Information Region) |
| Min Flight Level | Minimum vertical boundary of the airspace volume expressed as a flight level, repeated for each point |
| Max Flight Level | Maximum vertical boundary of the airspace volume expressed as a flight level, |

| | |
|---|---|
| Sequence Number | Numeric sequence number of a boundary point of the FIR's shape |
| Latitude | Latitude in decimal degrees of a point on the route |
| Longitude | Longitude in decimal degrees of a point on the route |

#### 3.3.3.7 Range of available dates and areas

Available for the European airspace, from 2015 to 2020 but only 4 non-consecutive months of each year.

### 3.3.4 Others Sources Disclaimer

Due to the nature of research and development of the FARO project, new sources could appear resulting from other WPs' work and outcomes as well as from the use case requirement tuning. In all projects involving data, some iterations may be conducted along the process of selecting and acquiring the data of interest. Any new incoming source will be described in a similar way to this document and included in the deliverable D3.3 – Final Dataset delivered on M15.

The D3.3 document will have a long preparation period starting from the initial point that this document provides defining raw data sources. Several internal versions reflecting the progress of the data tuning phase will be available for the project members before producing the final delivered version to guarantee the correct development of the pipeline iterations.

# 4 Usage

In this section, how-to kind tutorials are listed to connect to the Faro Data Hub (FDH) from the most well-known platforms, for further details on how to connect from a different platform contact with the FDH administrator. For security credentials, all URLs shown on this document are examples and the valid ones will be provided directly to each member of the FARO project by an automatic email generated by the platform.

## Confidential text. Removed for the public version of the deliverable.

# 5 References


[1] FARO Consortium, “D1.1 Project Management Plan”. Brussels, 05/10/2020

[2] FARO Consortium, “D1.2 Data Management Plan”. Brussels, 05/10/2020

[3] CEANITA Website - https://www.seguridadaerea.gob.es/lang_castellano/g_r_seguridad/ceanita/informes_definitivos/default.aspx

[4] EUROCONTROL – R&D DATA ARCHIVE https://www.eurocontrol.int/dashboard/rnd-data-archive

[5] ENAIRE Commercial products - https://www.enaire.es/services/commercial_products_and_services/ air_traffic_management_systems_atm