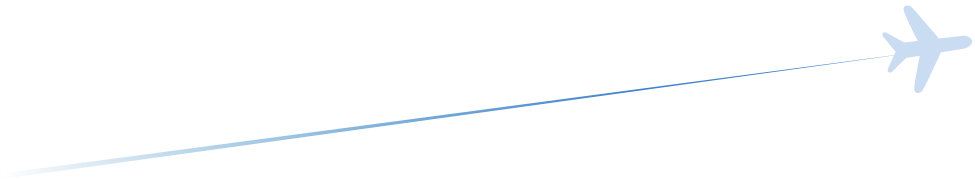

Data Transaction Pipeline Description

D1.2

DART

Grant:	**699299**
Call:	**H2020-SESAR-2015-1**
Topic:	**SESAR-02-2015 Data Science in ATM**
Consortium coordinator:	**University of Piraeus Research Center (UPRC)**
Edition date:	**25/11/2016**
Edition:	**00.01.00**

Founding Members

EUROPEAN UNION EUROCONTROL

SESAR JOINT UNDERTAKING

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Antonio Álvarez Fernández/CRIDA	Project Member	17/11/2016
Miguel García Martínez/CRIDA	Project Member	17/11/2016

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
George Vouros/UPRC	Project Coordinator	21/11/2016
Yannis Theodoridis /UPRC	Project Member	21/11/2016
David Scarlatti/BR&T-E	Project Member	21/11/2016
George Fuchs/FHFR	Project Member	21/11/2016
Jose Manuel Cordero/CRIDA	Project Member	21/11/2016

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

Name/Beneficiary	Position/Title	Date
George Vouros/UPRC	Project Coordinator	21/11/2016
David Scarlatti/BR&T-E	Project Member	21/11/2016
Georg Fuchs/FHFR	Project Member	21/11/2016
Jose Manuel Cordero/CRIDA	Project Member	21/11/2016

Rejected By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date

Document History

Edition	Date	Status	Author	Justification
0.1	02/11/16	Draft	A. Alvarez	New Document
0.2	10/11/16	Draft	A. Alvarez/M. Garcia	Release 1st draft, internal CRIDA review
0.5	21/11/16	Draft	A. Alvarez/M. Garcia	Inclusion of comments, format correction. Version for project review.
1.0	25/11/16	Final	A. Alvarez/M. Garcia	Reviewers comments inclusion. Version for official submission

DART

DATA-DRIVEN AIRCRAFT TRAJECTORY PREDICTION RESEARCH

This document is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 699299 under European Union's Horizon 2020 research and innovation programme.

Abstract

This document describes the Data Transaction Pipeline (DTP) to be used in DART Project. This DTP is the channel through which data will be gathered, stored and transferred for all DART research activities, and this document describes its architecture, organisation, and ways of been accessed. After reading this document any member of the project should be able to upload/download datasets to/from the DART data store. This pipeline is implemented in order to allow secure, efficient and reliable access to all the provided datasets, ensuring the data layer for DART activities.[1]

[1] The opinions expressed herein reflect the author's view only. Under no circumstances shall the SESAR Joint Undertaking be responsible for any use that may be made of the information contained herein.

Table of Contents

Executive Summary

This document contains everything related to gathering, storing, accessing and transferring data in the DART project, from a system perspective. It describes the architectural schema put in place, the network used, and also contains practical guides on how to access the data.

This document follows a practical, functional approach that allows its use within the project. It takes into account the fact that functionalities and capabilities are needed in a progressive way, being more demanding at the later stage of the project. So, the deployment of capabilities will be progressive, for which a scalable scheme has been used.

This way, the document contains a section describing the Data Transaction Pipeline (DTP) system description for reference, a Timeline description (making clear to users which capabilities will be available in each stage of DTP development), and a step-by-step guide to access data through DART Data Transaction Pipeline. This latter part is particularly relevant, as users, independently of their systems knowledge, should be able to access the data by following these guidelines, ensuring the main goal of this Data Transaction Pipeline. To allow a wide-scope access, guidance is provided for main Operating Systems.

The system that supports the Data Transaction Pipeline is part of CRIDA infrastructure, which has been exclusively dedicated to this purpose during the whole DART lifecycle. CRIDA is in charge of maintaining, scaling, upgrading and backing-up all the systems comprising the DART Data Transaction Pipeline.

1. Introduction

1.1. Purpose of the document

This document describes the Data Transaction Pipeline (DTP) proposed by DART; this pipeline is implemented in order to allow secure, efficient, scalable (according to project needs) and reliable access to all the provided datasets. Its main goal is to maximize data flow throughput to ease experimentation, satisfying all security requirements.

The DTP document covers how data will be transferred, stored and gathered from a technical point of view as well as from the user point of view. After reading this document any member of the project should be able to upload/download datasets to/from the DART data store. The DTP will be enhanced progressively to fit the needs of the partners and the project, especially needs regarding throughput and storage, in a scalable way.

1.2. Intended readership

This document is intended to be used by DART members.

1.3. Acronyms and Terminology

Term	Definition
ANS	Air Navigation Service
ANSP	Air Navigation Service Provider
ATM	Air Traffic Management
ATC	Air Traffic Control
AU	Airspace User
CFS	Certificate on the Financial Statements
DCB	Demand and Capacity Balancing
DTP	Data Transaction Pipeline
FTP	File Transfer Protocol
H	Humans
Horizon 2020	EU Research and Innovation programme implementing the Innovation Union, a Europe 2020 flagship initiative aimed at securing Europe's global competitiveness.
IPR	Intellectual Property Rights
KPI	Key Performance Indicator
PMP	Project Management Plan
POPD	Protection of Personal Data
PRC	Performance Review Commission
TRL	Technology Readiness Level

SESAR	Single European Sky ATM Research Programme
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.
WBS	Work Breakdown Structure
WP	Work Package

Table 1: Acronyms and Terminology

2. DESCRIPTION

2.1. OVERVIEW

The objective of the Data Transaction Pipeline (DTP) is to provide constant bi-directional flows of data between the DART data store and the different WPs. The system keeps high rates of security and speed of transactions to ensure the needs of the DART 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, especially when talking about scalability.

With all this in mind, DTP 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. While writing this document, there is already a planned upgrade of scalability regarding the size of the data store. After a first version of 3 TB a raise of storage with more than 20 TB will be implemented. This and other details of the scalability planning will be described on the Timeline section of this document (3 - DTP SCALABILITY TIMELINE).

2.2. SCHEMA

This document explains the data access mechanism (Data Transaction Pipeline - DTP) for secure and reliable access to data, ensuring both integrity of data and accessibility for all the WPs. This transaction pipeline is bi-directional, as it also receives the results from WP2 and WP3 processing for further visualization.

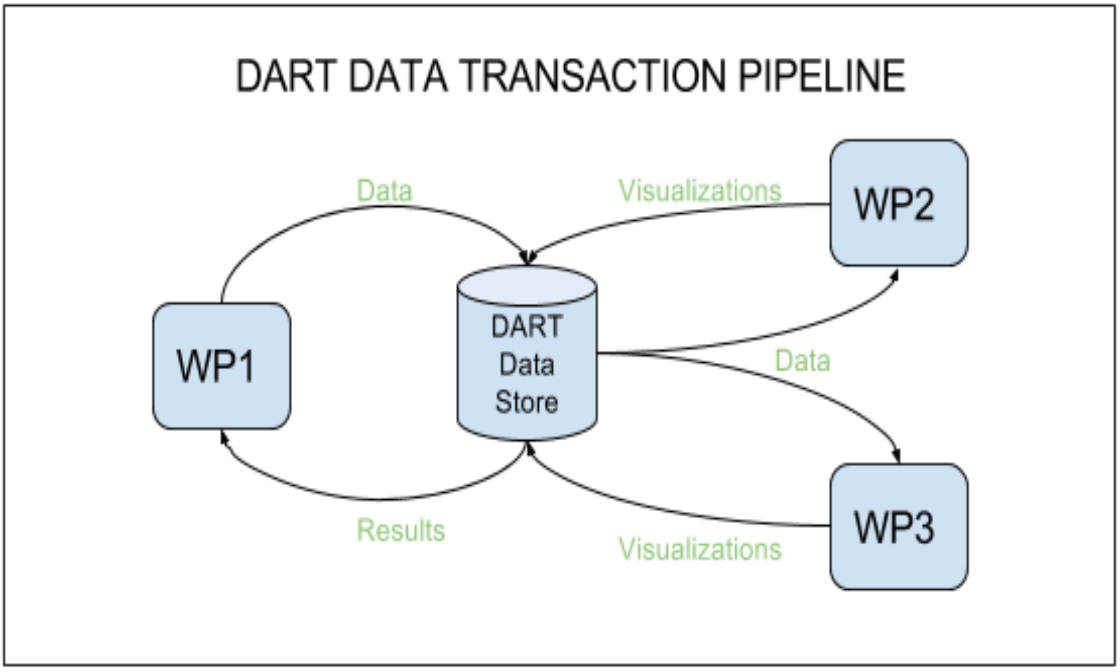

Fig. 1 - SCHEMA

The core of the DTP is the DART Data Store where all data travelling through the pipeline will be hosted.

- Architecture:

The data store is mainly compound by a Windows Server located on the Red Iris network that serves data through an FTP server (check 2.5 - Background).

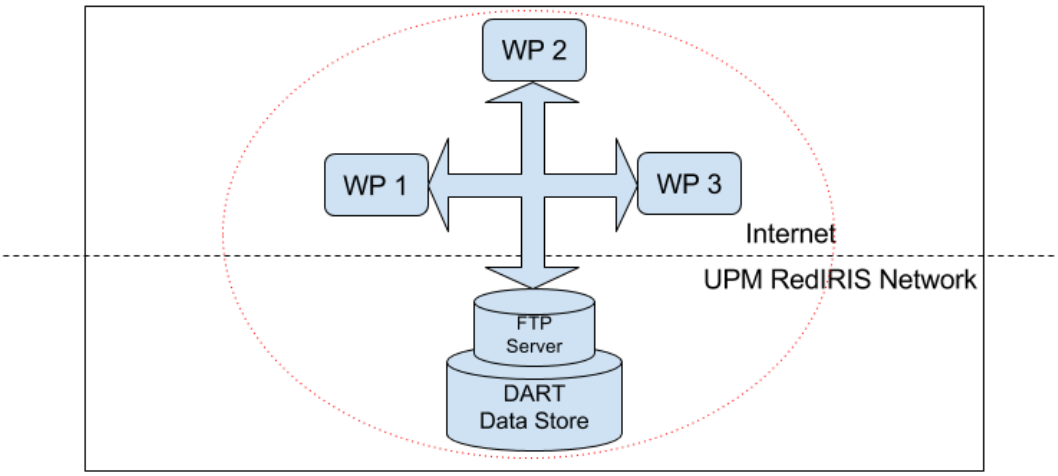

Fig 1 - Architecture

- Structure:

A first structure for the DART file store is proposed in the following schema; of course, this will be further refined is order to be finalized (and probably modified along the project).

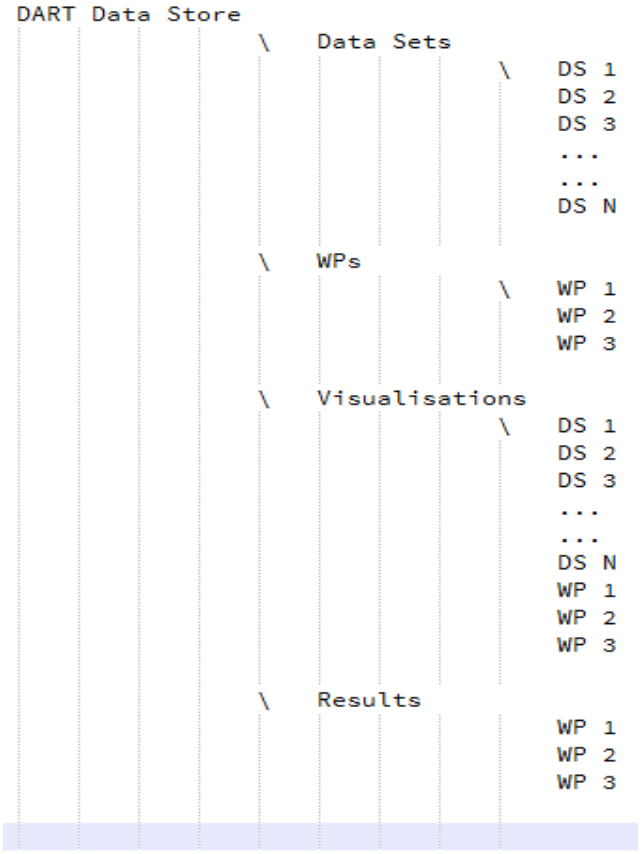

Fig. 2 - Data Store File Structure (proposal)

2.3. ACCESS TO DATA

Direct access to the data is provided by the DTP. No API, just a file-system hierarchy with different access roles (check 2.4 - STAKEHOLDERS for roles and permissions).

2.3.1. Data sets

The DTP will make datasets downloadable by separate files organized on folders identifying each data source. All data sets will be located under the "\Datasets" folder. The data sets will be aligned on the space and the time, to keep sources manageable the files will be separated 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". For further details check the DART D 1.1, Data Management Plan.

2.3.2. WPs

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 inputs. 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 organized by WPs.

2.3.3. Visualizations

Visualizations and internal results will be produced along the project to analyse the datasets and for WPs to provide feedback to each other.

In case of the visualizations, any kind of media file could be released as well as any kind of visualization tool output file (Geospatial Google Earth - kml format for instance). There is not a specific format defined for these products.

Visualizations will be located on the "\Visualizations\XX" folder.

2.3.4. Results

Further results will be produced at the final stages of the project as final conclusions of the DART platform, including applications to address the specified use cases and analysis to prove their efficiency. There is not a specific format defined for this products, just they will be located on the "\Results\XX" folder.

2.4. STAKEHOLDERS ROLES

Key stakeholders that have been identified to have access to the DTP are as follows:

Stakeholder	Roles	Description
Data Providers	Data Sets: Read & Write WPs: None Visualizations: None Results: None	Providers of data sets, on DART project mainly CRIDA and BR&T-E.
WPs	Data Sets: Read WPs: Read & Write Visualizations: Read & Write	WPs will share results and feedback each other with visualizations and results of machine learning algorithms.

	Results: Read & Write	
SJU	Data Sets: Read WPs: Read Visualizations: Read Results: Read	SJU will have interest on all products of the DTP.

2.5. Background

1.1.1. Network – RedIRIS

The DART data store as kernel of the DTP is located on the ETSIA (Aeronautical Engineering College) of the Polytechnic University of Madrid under the umbrella of the network RedIRIS. This network is the Spanish academic and research network that provides advanced communications services to the national scientific community and universities since 1988. It is funded by the Ministry of Science and Innovation and is included in the Ministry's map of Special Scientific and Technological Facilities. The use of advanced optical equipment means that the Spanish research community will have access to multiple circuits of up to 100 gigabytes per second from the main research centres, including the astronomical observatories in the Canary Islands.

The European nature of DART project makes this network ideal because of the direct link to the global pan-European research network, GÉANT. This network infrastructure has a direct link between ISP (Internet Service Providers) a.k.a. “Dark Network” with presence on 33 countries and their national research networks (see the figure that follows).

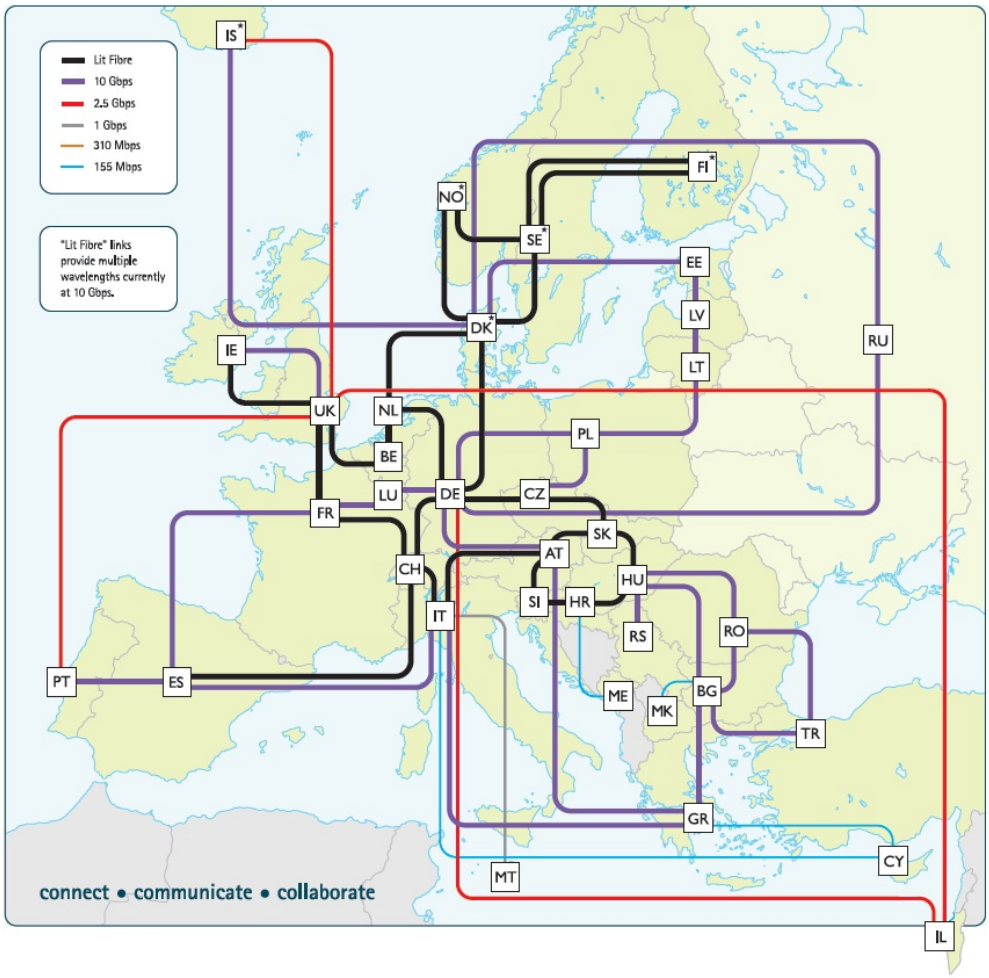

Fig. 3 - GÉANT SCHEMA

The physical location of the DART data store are the CRIDA premises located in the Universidad Politécnica de Madrid (UPM), where the access to RedIRIS is located. These premises rely on physical security services and restricted access areas that ensure reliability of the site for these purposes.

1.1.2. Server

The DART data store is designed as an FTP system with the following specifications:

- Windows Server 2012 R2
- Intel Xeon processor at 2.4GHz.
- Ram memory 8GB.
- Space of Disk (physical drives) 3TB**

The system is operating using CRIDA background (infrastructure).

Note:

** The space of disk of the system will grow up along with the project needs, having already the goal of a final system with no less than 20 TB. To host this huge amount of data, NAS technology has been selected to create an in-house cloud system for the DART project that, among others enhancements, provides automatic incremental backup systems.

1.1.3. Security

The access to the system is provided by username and password. Even if the DART project has no requirements about roles and permissions, the DTP provides the possibility to map directories to groups of users with specific roles, for instance Analyst, DataProvider, etc...

Each member of the project will have a specific username and password provided by the system administrator (CRIDA). Password restrictions avoid force-kind attacks with a policy of no less than ten characters, including at least one mandatory special characters ('.', ';', '%', etc...)

In the background, RedIRIS provides an extra level of security to the DTP. The RedIRIS (IRIS-CERT) security team undertakes preventive measures and acts in a co-ordinated manner with the affiliated institutions in order to respond to any incidents that may arise on the network.

Further improvements on security may be implemented depending on the project needs. The DTP has been designed to be upgraded to another level of security at anytime, implementing certificates for secure connections. This raise on security would lower the performance rates because of the extra encryption and decryption tasks for the GB of data on each transaction. So the first approach does not include this extra level of security but it would still be kept in mind by the DART project members.

1.1.4. Backup

During the first stage of the DART DTP (until M06) automatic backups won't be launched. The content of the data store will be essentially inputs of data sets from data providers, in case of loss of data providers would restore them from their own backups. As mentioned before, when the NAS system will be online it would provide enough space to make automatic incremental backups including inputs and outputs of the WPs.

3. DTP SCALABILITY TIMELINE

3.1. Description

The disk space requirement for the DART DTP is huge due mainly to the NOAA data source that generates around 600 GB per month of data without curation or filtering. At least a couple of years of data has to be stored, this means no less than 15 TB of disk space. Adding to that the backup system, the required size is pretty much duplicated.

With this scenario, DART will implement a 2-stages delivery of the DTP in terms of space. First stage delivery provides enough capacity to host the first data quality set at least to assure that the project tasks depending on data can start without any delay. This first stage is planned to deliver around 5 TB of store (due M06). After the second stage, the DTP will be fully available in terms of disk space capacity, going until more than 20 TB (due M12).

Of course, the bigger is the need of space to host data, the bigger is the need of space to make backups. Therefore, the backup system will be ready only after the raise of the disk space during the first upgrade of scalability.

As mentioned before, there is a possible enhancement of the security system for the DTP that, at this stage, is not scheduled nor required.

3.2. Delivery timeline

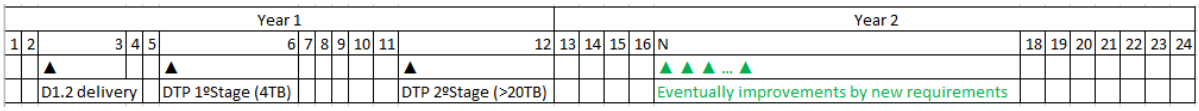

Fig. 4 - Delivery Timeline

The DART Transaction Pipeline will be in service until the end of DART Project.

4. USAGE

Along this document, several time-spread changes due to scalability are mentioned. Intention of DART DTP is that the usage does not change at all during all these stages described, creating a real scalable system transparent to the user. In this section, how-to kind tutorials are listed to connect to the DTP from the more well-known platforms, for further details on how to connect from a different platform contact with the DTP administrator (CRIDA).

The store Url is ***datastore.dart-research.eu*** and credentials for each user will be provided by direct email between the administrator and the partners.

4.1. General tutorial

The DTP system is implemented as an FTP system, so any generic client of FTP should be easy to set up with the parameters provided by the administrator. These would be an IP address and a user/password credentials. All parameters (ports, etc.) are the FTP default ones. It is also possible to connect to an FTP server via FTP user software. For this purpose, different software solutions exist, for example Filezilla. This software is one of the most popular FTP clients and is compatible with Windows, Linux and Mac systems.

4.2. Windows

- *Add a Windows Explorer Connection*

The simplest way to connect any FTP direction with Windows is through Windows Explorer. To do this access go to the next steps and observe if the connection will be stable.

1. Open a Windows Explorer window to connect to FTP, all you need is Windows Explorer. Open a new instance by clicking on the anchor in the taskbar or, if you prefer, using the shortcut Windows + E keyboard icon.

2. Enter the FTP server address In the navigation bar, type the full address of the FTP server. That is, you need to include the ftp:// protocol, the server (for example, ftp.server.com) and, where applicable, the internal path (the type / folder / subfolder). Press Enter and Windows will start to connect.

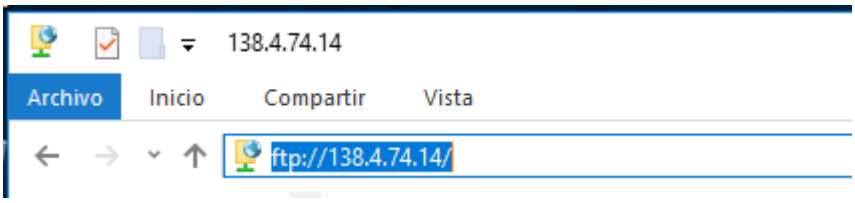

Fig. 5 – Windows Explorer connection

3. Enter the user and password.

- *Add a Network Location*

If you wish to add a Network Location, when you right-click on My PC (See first image), select **Add a network location**. From the Map FTP Drive box, you can also select the link at the bottom that reads, *Connect to a website that you can use to store your documents and pictures*. The Add Network Location wizard will open. Click Next and then Choose a custom network location. Click Next again. Now specify the Internet or network address or browser.

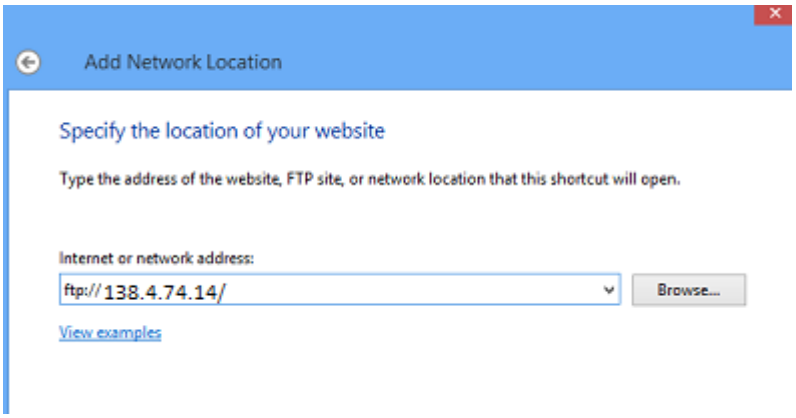

Fig. 6 - Windows Network Connection (Step 1)

Uncheck *Log for anonymously* and give the username & password. Click on Next. Give a name to the networked location, when asked. Click Next again. Now select *Open this network location when I click Finish.*

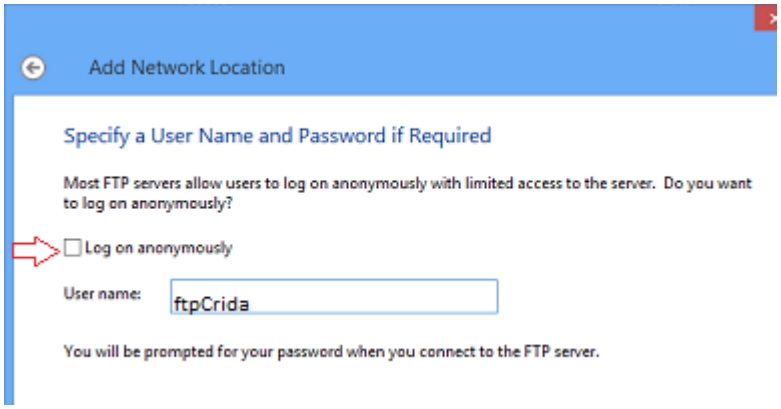

Fig. 7 - Windows Network Connection (Step 2)

You will be asked to enter your credentials, and once you do so, you will be connected to your network drive or your FTP drive or your website.

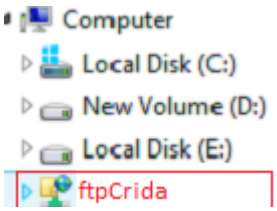

Fig. 8 - Windows Network Connection (Step 3)

This is quite useful if you need to connect your computers together for sharing files, store files online or run a website.

4.3. Mac OSx

For writing and reading data connect using any free FTP client as Filezilla. For Read-Only purpose the FTP server can be mapped as a local volume:

1. Open on Finder context menu the connect to server option

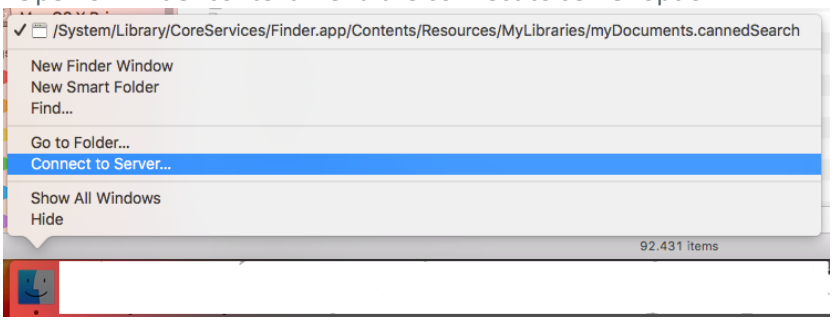

Fig. 9 Connect to Server (osx)

2. Then enter the server address

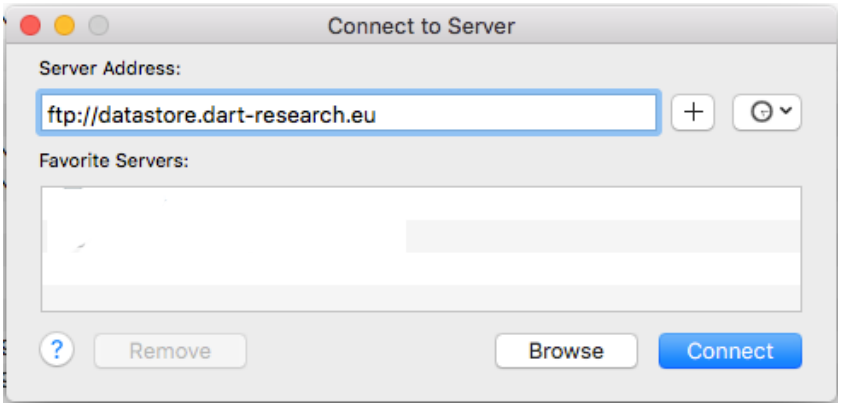

Fig. 10 Server address (OSX)

4.4. Linux

In order to complete the access to FTP files we include the description of how to use this tool in a terminal. As well, Ubuntu permits a graphical version more easy to use. We can use Filezilla for example or another graphical tool available for Linux operating systems

For the purposes of this document, we will use Ubuntu and explain the process to connect to one FTP server via command line:

To initialize the connection, we need to open a new terminal and write these commands:

1. The first step is write the direction of ftp and for this we will write: ftp datastore.dart-research.eu

   root@ubuntu:/etc/network# **ftp** datastore.dart-research.eu

   Connected to 138.4.74.14.

   220 Microsoft FTP Service

   Name (138.4.74.14:root): (**correct username**)

   331 Password required

   Password: (**correct password**)

   230 User logged in.

   Remote system type is Windows_NT.

2. Now we will see the directory:

   ftp> **ls**

   200 PORT command successful.

   In this sample we see the directory and the folders inside him.

   150 Opening ASCII mode data connection.

   11-11-16 09:23AM <DIR> Compartida

   11-08-16 11:49AM <DIR> Crida

   11-04-16 10:20AM <DIR> Dart

   11-04-16 12:00PM <DIR> Shares

   226 Transfer complete.

   ftp>

References

[1] RedIris, (www.rediris.es). National Research and Development Plan. Public Corporate Entity Red.es, 2016.

[2] GÉANT, (www.geant.org).The GÉANT network technology. GÉANT topology map. Pan-European network. Grant Agreement No. 731122 (GN4-2), 2016.

[3] DART, DART 699299 Consortium Agreement, v1.0

[4] European Comission, Guidelines on Data Management in Horizon 2020, Version 2.1, 15/02/2016

[5] DART, DART 699299 Grant Agreement, v1.0

[6] DART, DART D 1.1, Data Management Plan