Grant Agreement No: 687591

Big Data Analytics for Time Critical Mobility Forecasting

datAcron

D8.2 datAcron Data Management Plan (1st Version)

Deliverable Form	
Project Reference No.	H2020-ICT-2015 687591
Deliverable No.	8.2
Relevant Work Package:	WP 8
Nature:	R
Dissemination Level:	PU
Document version:	2.0
Due Date:	30/06/2016
Date of latest revision:	30/06/2016
Completion Date:	30/06/2016
Lead partner:	UPRC
Authors:	George Vouros, Maria Halkidi
Reviewers:	
Document description:	The datAcron Data Management Plan (datAcron DMP) details what data the project will collect, generate, how will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. This is the 1st version of this deliverable.
Document location:	Filestore : /datAcron/WP8/Deliverables/Final

History of changes

Version	Date	Changes	Author	Remarks
0.1	15.05	ToC	George Vouros	
0.8	15.06	First draft	Maria Halkidi	
1.0	18.06	Complete draft	George Vouros	
1.1	22.06	Corrections and minor additions	Maria Halkidi	
1.5	23.06	Corrections and additions	George Vouros	
1.7	24.06	Corrections and addition of policies' details	George Vouros	
1.8	27.06	Update of data sources	Maria Halkidi	
1.9	29.06	Review comments	Elena Camossi & Anne-Laure Jousselme (CMRE) Enrie Batty (IMISG)	
2.0	30.06	Final with reviewers' comments	George Vouros	

EXECUTIVE SUMMARY

The datAcron Data Management Plan (datAcron DMP) details what data the project will collect, generate, how these will be exploited or made accessible to all stakeholders, how and what data sets will be made available for verification and re-use, and how they will be curated and preserved. In order for the document to be self-contained, a comprehensive view of the datAcron data lifecycle is provided with appropriate definitions of terms being used and stated assumptions under which the plan has been devised, with a succinct description of stakeholders' groups.

Data sources to be exploited per domain are described, also specifying information on existing metadata per data sources, size of data sets, modality, and provision methods. The deliverable specifies also a generic policy /methodology for associating data sources and datasets with metadata taking into account version control, licensing and distribution issues.

Furthermore, information and a generic policy on data sharing is provided, also taking into account, limitations of partners, IPR and legal issues, licensing and ethical issues.

The deliverable concludes with issues concerning archiving and preservation that should be addressed once the project has been completed.

This is the 1st version of this deliverable. As the project evolves, the DMP will be updated to reflect the changes in the data sources exploited in datAcron, updates on policies and methodologies to be used, given also further understanding of the necessity of additional data sources and decisions on specific datasets to be made available to stakeholders.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS & TERMS

Abbreviation	Term
ATM	Air Traffic Management
BRTE TPE	BRTE Trajectory Predictor Engine
CTOT	Calculated Time of Take-off
DCAT	Data catalogue vocabulary
DMP	datAcron Data Management Plan
FMS	Flight Management
NOAA	National Oceanic and Atmospheric Administration
TP	Trajectory Predictor
GA	Grant Agreement 687591
DDR	Demand Data Repository
AIXM	Aeronautical Information Exchange Model
AIS	Automatic Identification System
AIP	Aeronautical Information Publication
RAD	Route Availability Document

1. INTRODUCTION

1.1 Purpose and Scope

A Data Management Plan (DMP) is a formal project document which outlines the handling of the data sources at the different project stages. The H2020 guidelines [3] provide an outline that has to be addressed. The DMP covers how data will be handled within a project frame, during the research and development phase, but also details the intentions for the archiving and availability of the data once the project has been completed. As the project evolves, the DMP is updated to reflect the changes in the data situations and as the understanding of data sources and data created becomes more concrete.

datAcron project aims to develop novel methods for (a) **real-time detection and prediction of trajectories** and (b) **detection and prediction of important events** related to moving entities, together with (c) **advanced visual analytics methods**, over multiple heterogeneous, voluminous, fluctuating, and noisy data streams from moving entities, via the (d) **real-time in-situ processing of multiple data streams**, (e) the provision of **integrated views of streaming data with archival data** expressing entities' characteristics, geographical information, patterns of mobility in specific areas, regulations, intentional data (e.g. planned routes) etc, and (f) the provision of advanced solutions for **managing spatio-temporal data**.

Thus, data sources related to datAcron purposes include streaming data sources (mostly concerning surveillance data concerning moving entities), and archival data sources with historical, contextual and other data concerning the entities themselves, their positioning and movement.

Technological developments in datAcron will be validated and evaluated in user-defined challenges that aim at increasing the safety, efficiency and economy of operations concerning moving entities in the air-traffic management (ATM) and maritime domains: These use case, scenarios and challenges have been specified in deliverables D5.1 "Maritime use case detailed definition" and D6.1 "Aviation use case detailed definition".

Under this prism, the data management plan (DMP) identifies the data sources to be exploited for the purposes of the use cases detailed in D5.1 and D6.1, as these (sources) have been specified in D5.2 "Maritime data preparation and curation (interim)" and D6.2 "Aviation data preparation and curation (interim)" ; as well as results created, and related datasets. A dataset is any coherent subset of data, but in this document we use this term to denote any set of data from (individual or multiple) data sources, and/or data created from datAcron components. These datasets can be used for validation, reusability, dissemination or demonstration purposes, also according to GA articles 29.2 "Open access to scientific publications" and 29.3 "Open access to research data". Data sources provide input data to the datAcron components for realizing their functionality and computing their results, while datasets are subsets of data that are necessary and sufficient enough for validation, re-use and demonstration of computations.
The datAcron Data Management Plan (DMP) details what data the project will collect, generate, how will be exploited or made accessible for verification and re-use, and how it will be curated and preserved in the lifetime of the project. The datAcron Data Management Plan refers to these issues in detail and specifies ways to make a portion of the data available with respect to partners' agreements, with respect to IPR or privacy laws, and ethical issues specified in D8.5 "Ethics Management Plan".

The current version of the plan concerns the datAcron data sources that fit (are necessary) to the purposes of the use cases and scenarios, and have already recognized as such in D5.1 and D6.1. A detailed description of these sources is provided in D5.2 and D6.2

1.2 Relation to other Work Packages and Deliverables

This deliverable is related to WP1, WP5 and WP6. WP1 will provide the detailed description of how data are to be integrated and stored in the datAcron store, providing coherent, integrated views on data; as well as ways to access integrated data. WP5 and WP6 aim to identify the data sources that

should be used within the project and prepare datasets so that they are readily available for exploitation according to the datAcron value chain detailed in the next chapter of this document. More specifically DMP is related to D5.2 and D6.2 since many of the questions identified in DMP need to be answered as part of the data sources' preparation and curation from data providers. Furthermore, this deliverable is related to D8.5 "Ethics Management Plan", since it should consider licensing, IPR and ethical issues.

1.3 Approach Taken and Structure of the Deliverable

This 1st version of the DMP has been devised having the following issues in mind:

- As datAcron partners address their research objectives, driven by the use cases and scenarios already detailed in D5.1 and D6.1, and towards providing interesting results to the ATM and maritime domains, they may decide incorporating additional data sources than those already identified in this version of the DMP. These additional sources will be mentioned in subsequent versions of the DMP. This deliverable is written concurrently with deliverables D5.2 and D6.2, which are interim versions of deliverables providing the precise definition of the datasets to be used, including results of datasets quality assessments, as well as data curation techniques to guarantee continuous data consistency and availability. As such, these deliverables are connected to this deliverable: These have been taken into account for writing this deliverable.

- As WP1 aims to produce advanced data management components for integrating and managing data from heterogeneous and disparate data sources, these components are considered as key tools to the overall data management plan. Thus they should be incorporated into the data management plan, positioning them appropriately into the datAcron data value chain.

- Data sources provided from datAcron partners or from third parties, satisfy specific properties and are subject to constraints and limitations for access and exploitation and should be used with respect to specific ethical, IPR and legal restrictions. Such restrictions have been specified in the datAcron Consortium Agreement signed by partners and may be refined/revisited for particular subsets of data throughout the project.

- datAcron aims to the development of research components for advanced analytics and prediction / forecasting of trajectories and events. These, in conjunction to the data management components to be developed will be rigorously tested, and results will be disseminated using specific datasets, which will be agreed among datAcron partners. datAcron will seek ways to make these datasets and computed results available to the research community via a specific repository, registered to a widely-used repository registry.

- All data sources being used, datasets, and research results being produced and archived, will be described using a specific metadata schema. The schema/vocabulary to be used will be chosen among well-known alternatives, also detailed below.

Therefore datAcron DMP takes into account three important issues, distinguishing different needs of data management: (a) Preparation and curation of data sources by maritime and aviation data providers, (b) Data acquisition, integration and provision of data via the datAcron data management infrastructure to be developed, and (c) orthogonally to (a) and (b), preparation of specific datasets, as well as archiving created results, to be used for validation, reuse and dissemination of datAcron research results, also according to GA articles 29.2 "Open access to scientific publications" and 29.3 "Open access to research data".

As the project progresses the DMP will be refined by taking into account the data management techniques and the datAcron data management infrastructure developed, the data sources and datasets to be used and provided, with the consent of all partners.

The remainder of this deliverable is structured as follows:

- **Data life cycle** - section 2 presents the datAcron data life cycle, and relevant stakeholders.

- **Basic data information** - section 3 provides description of the data sources and datasets whose use in the datAcron project has been decided.

- **Metadata -** Data sources and datasets will be described with metadata. These metadata can be used on the one hand for automating the datAcron data ingestion, data reusability, but on the other hand for all stakeholders to have a concrete view of the data being used, independently on restrictions to access the data. This is further described in Section 4.

- **Access, sharing and re-use policies** – Data sources are associated with limitations for access, legal and IPR constraints. This presents challenges if datasets have to be shared with stakeholders that are not datAcron beneficiaries. An important challenge is the integration/interlinking of data from datasets having different usage and access policies. Interlinking data with certain constraints and requirements with data that are publicly and freely available, impacts the desired access policy. Section 5 addresses these issues.

- **Archiving and preservation** – section 6 describes the challenging issues that arise regarding the long term storage of data after project completion.

2. THE datAcron DATA LIFECYCLE.

2.1 The datAcron data value chain

Figure 1. datAcron **data value chain**

The datAcron data value chain (Figure 1) comprises the following stages:

Data preparation. Use case leaders provide the precise definition of the data sources to be used for research, evaluation and validation of research components, together with associated metadata, and detailed information about the origin, scale and provision method(s) per data source. Specifically, multiple, heterogeneous and disparate data sources are expected to be used in maritime and aviation use cases.

In situ-processing and data transformation / integration. The data sources are multiple streaming data sources, as well as archival data sources. The datAcron integrated system will incorporate in-situ processing methods that aim to provide data synopses and detect important recurring patterns in data, also integrating where appropriate data-in-motion from streaming sources, computing single and multi-streaming data at high rates of data compression, without affecting the quality of analytics results.
The data transformation components aim to covert data from (a) single and multiple streaming data synopses, (b) archival data sources, and (c) results computed by the datAcron analytics components, to RDF triples according to the datAcron RDF schema being devised. Bringing all data to such a common form aims to facilitate their integration/interlinking, also incorporating semantic information into the process.
The data integration component interlinks semantically annotated data using link discovery techniques for automatically computing correspondences between data from disparate sources. This produces integrated data views. Integrated data are provided to the analytics components, while they are also stored in the datAcron store.

Persistence storage. datAcron will develop a scalable data store to manage data from disparate sources, providing a coherent view on integrated data, ready for efficient utilization by the offline and real-time analytics components developed in the project.
Specifically, datAcron aims to develop a novel data processing framework for supporting the efficient distributed management and querying of RDF spatio-temporal data from disparate data sources (also concerning trajectories and events computed by datAcron analytics components), emphasizing on eager filtering of data using spatial and temporal predicates.

Data processing and analytics. The data analytics components include trajectory and complex event recognition and forecasting, as well as visual analytics. These consume the data provided by the data management component and compute results.
The data analytics components also use internal stores for frequent and fast data write/read, well-tuned to their requirements and to the rest of the overall datAcron architecture, so as to satisfy specific latency requirements for the computation of results.

The relation of the datAcron work packages to the stages of data value chain is described in Figure 2, starting the chain from the bottom of the figure. WP5 and WP6 provide detailed information about

available data sources (aviation and maritime data), as well as information about preparation, curation and accessibility of these data sources. WP1 describes the data that will be stored in the datAcron integrated system store while it will present ways to access integrated data. WP2, WP3 and WP4 provides description of analytics results generated and relevant datasets for validating, re-producing and disseminating their results, also according to GA articles 29.2 and 29.3. Computed results (e.g. trajectories and events detected/predicted) may also need to be integrated with data in the store and be stored in the datAcron store.

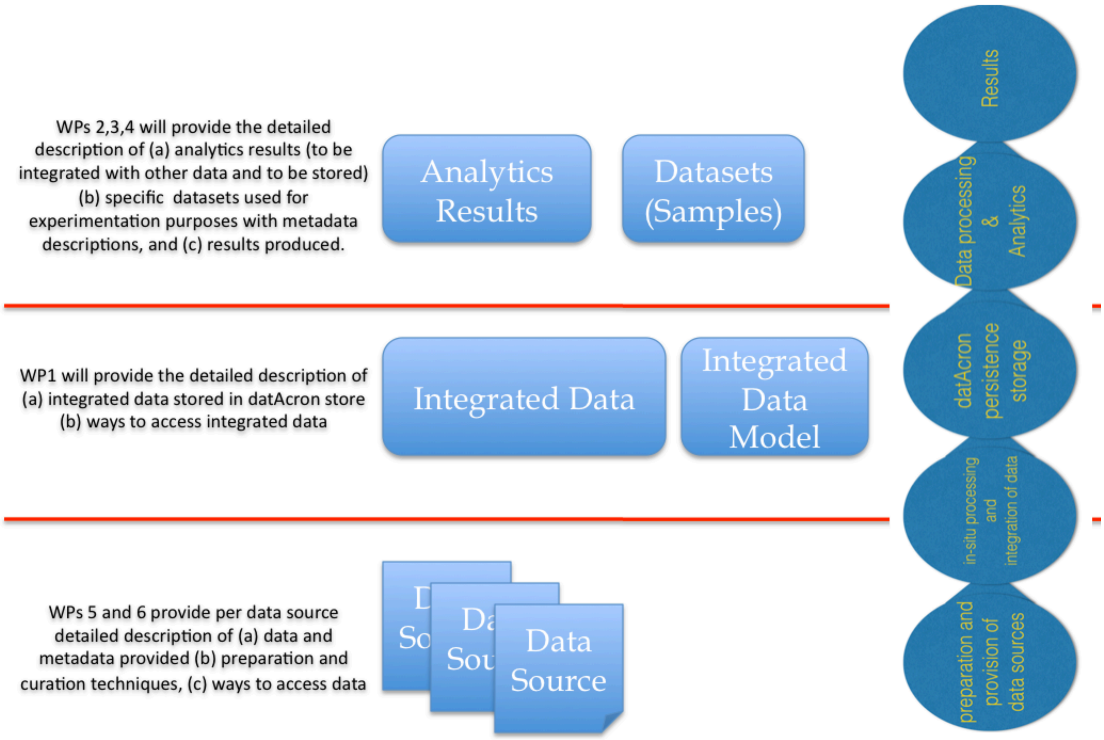

Figure 2. Relation of data value chain to datAcron work packages

It must be pointed out that WP1,2,3 and WP4 provide feedback to WP5 and WP6 concerning the data provided and their suitability for achieving their purposes. This may result to updating the data sources, reconsidering the quality of data gathered and/or provided to the project, or adding data sources.

2.2 Stakeholders and Constraints

Key stakeholders that have been identified and that influence the data management plan decisions as they have specific interest in data, research results, domain specific results (i.e. results connected to aviation/maritime operational concerns) are as follows:

Data providers. These represent the organizations that provide the data sources to be integrated into the datAcron store. These are datAcron beneficiaries, third parties connected to beneficiaries or to the project via the domain use case interest groups, or organizations providing (open) data related to the purposes of the maritime and aviation use cases. These parties are represented in the datAcron project by WP5 and WP6 leaders and participating partners. datAcron partners playing this role also play significant roles in defining the use cases per domain, specifying the data sources to be used and assist with the definition of requirements from the data management components.

datAcron partners. The datAcron partners provide data sources and define the use cases for validating results, collaborate to design and develop the architecture of the datAcron integrated system, specify the requirements for data management and data analytics according to the domain-

specific use cases defined and according to the project research objectives. The quality and the big-data characteristics of the data provided from disparate sources are important features, given also that the methods and algorithms developed in datAcron must be validated and evaluated in different scenarios of data growth and quality.

End-users. End-users represent the user-interest groups that focus on real-life, industrial and user-defined challenges concerning operations regarding moving entities in sea and air. These are represented in the datAcron project by WP5 and WP6 leaders, including also organizations and domain experts being members of the datAcron domain-specific use-case interest groups.

Big data analytics researchers and stakeholders of the big data value chain. These are interested in datAcron developments and research results: Algorithms and novel methods for the management of big data, and big data analytics. They are mainly interested to published results and to datasets for validating, evaluating, comparing, and testing methods. Devising such datasets and making them available to the research community is expected to increase the visibility and impact of datAcron scientific outcome.

3. THE datAcron DATA SOURCES.

This section presents data sources to be exploited per domain of interest: ATM and maritime. It further provides details on existing metadata per data source (i.e. metadata concerning the content of data sources, mostly according to domain – e.g. whether- standards), scale of data sources and data provision methods.
A fine description of the data in these sources is provided in D5.2 and D6.2, but here, for the purposes of the DMP, we enlist data sources with a succinct description of their contents, emphasizing mostly to the data providers (i.e. the origin of data), how data are acquired, when and where are these acquired and how often, also specifying the contact points for the data sources.
All data sources are to be available during the lifetime of the project (of course open data sources are expected to be available for a longer period) and documentation available to understand the available data is provided by datAcron deliverables D5.1, D5.2, D6.1 and D6.2, and subsequent versions of them.

3.1 Aviation Data Sources

Aeronautical data is heavily regulated, especially in Europe according to Eurocontrol Standards. For example, Flight Plan filling information follows ICAO FPL2012 format, Radar information is provided following ASTERIX standard (Asterix Cat62 for fused data), datalink between airlines dispatcher and aircraft follows A702-A format, Airspace information is mostly provided in AIXM format. The objective of the Aeronautical Information Exchange Model (**AIXM**) is to enable the provision in digital format of the aeronautical information that is in the scope of Aeronautical Information Services. That means that the research results can be applied nationwide in Europe.

Weather Data

A single data source (NOAA), will finally be considered for this category, as is the only European-wide weather data available.

NOAA (National Oceanic and Atmospheric Administration): This data source is used mainly to obtain the weather conditions at the position an aircraft is at any given time of the flight.
Weather models use a Grid with a specific resolution. For aviation in datAcron we'll work with NCEP Grid 4 which has a resolution of 0.5º. (see http://www.nco.ncep.noaa.gov/pmb/docs/on388/tableb.html)
Forecast models can be run several times a day, for aviation in datAcron we'll typically use the latest forecast available previous to the time we are interested in. Forecast models has too a time resolution, or "forecast step", which we expect to be 1 hour.

Metadata:
Data for weather models is typically distributed in "GRIB" format files. GRIB (**GRIdded Binary** or **General Regularly-distributed Information in Binary form**) format enables the compression of the weather data and includes metadata about the content of the file. Thus it is very convenient for transferring the data. The data can be extracted with many available tools (I.e. GRIB API from ECMW available at https://software.ecmwf.int/wiki/display/GRIB/Home).

Scale:
As a reference a global forecast, 6 hours step, for 24h, for 14 isobaric levels at .5º resolution expanded from .grib to .csv can amount about 3.2 Gb.

Provision methods:
The forecasts will be delivered in .grb files to datAcron. These files may be converted to .csv files using tools like "wgrib2".

Partner(s) responsible: CRIDA & BRTE

Radar

This data category comprises of three sources of surveillance information: IFS, ADS-B and DDR.

IFS: This data source provides radar tracks of the Spanish airspace controlled by the Spanish ATC provider EnAire. A radar track file consists on tabular data rows with a timestamp key and several rows of geospatial information for each one of these timestamps. The update interval is 5 seconds. The area provided is separated into 5 different regions delivered each one on a different plain text file (ifs files).

Metadata: Not available.

Scale:
IFS data is available for from 2013 till 2016 (complete years). The covered area is the Spanish airspace, as it also shown in D6.2.
On structured version, one day is around a hundreds millions of records.

Provision methods:
Structured version could be provided by web services or database direct connection.

Partner(s) responsible: CRIDA & BRTE

ADS-B Messages: This data source refers to the ADS-B messages broadcasted by many airplanes (practically all airliners) using their transponders. These messages are received by ground based receivers and can be used to reconstruct the trajectory of the flight. There are several types of messages that can be found but for datAcron the relevant ones are these about aircraft identification and position.
datAcron source of ADS-B messages is the ADSBHub network. This network is formed by 81 stations across the globe, 61 of them in continental Europe. The messages received by this network are stored in a human readable format know as "SBS-1 BaseStation port 30003".

Metadata:
Messages are contained in a single line of CSV files. ADS-B messages are captured in CSV format. An excerpt of such messages, describing a particular flight, look as follows:

```
MSG,3,,,34324E,,2015/08/03,01:05:03.844,2015/08/03,01:05:07.058,,30050,,,45.69032,5.54741,,,0,0,0,0
MSG,6,,,34324E,,2015/08/03,01:05:03.844,2015/08/03,01:05:07.058,,30050,,,,,1021,0,0,0,0
MSG,4,,,34324E,,2015/08/03,01:05:03.744,2015/08/03,01:05:07.058,,,428.0,51.0,,,-384,,,,
MSG,1,,,34324E,,2015/08/03,01:05:04.344,2015/08/03,01:05:07.058,IBE34CP,,,,,,,,
MSG,3,,,34324E,,2015/08/03,01:05:05.943,2015/08/03,01:05:15.337,,30025,,,45.69296,5.55218,,,0,0,0,0
MSG,6,,,34324E,,2015/08/03,01:05:05.943,2015/08/03,01:05:15.337,,30025,,,,,1021,0,0,0,0
MSG,4,,,34324E,,2015/08/03,01:05:10.643,2015/08/03,01:05:15.337,,,426.0,51.0,,,-128,,,,
```

ADS-B messages are formatted according to the SBS Station schema original from Kinetic's SBS-1 & SBS-3 Mode-S.

Scale:
ADSB data is available since late 2015 and is continuously recorded, however, not all the sensors are 100% of the time up and the recording system is not 100% of the time up. Power supply and/or network outages can create dates with less or even without data.
One day of messages tops about 2.3 Gb (not compressed).

Provision methods:
Historical data: Files for specific time periods can be assembled and delivered through Internet or physical media (I.e. DVD).
Real time data: Once a VPN connection is established to BR&TE Laboratory Network the client can issue a netcat command to receive the real time feed.

Partner(s) responsible: CRIDA & BRTE

DDR: In addition to the above mentioned sources, DDR, or Demand Data Repository (a European wide valuable data source) contains surveillance/radar data, which is embedded in the flight plan data. This data source, including the radar part, is described in the Flight Plan category subsequently, and is only mentioned here for reference as a key radar data source.

Airspace

DDR Sector Configuration: Air traffic control (ATC) is a service provided by ground-based controllers who direct aircraft on the ground through controlled airspace. The primary purpose of ATC worldwide is to prevent collisions, organize and expedite the flow of air traffic, and provide information and other support for pilots.
Airspace can be divided in a set of ways, with a different number of pieces (sectors). For instance, a sector configuration 9A means that a particular airspace (a region in Spain) is divided in 9 sectors, in a particular way. 9B also mean 9 sectors, but divided in a different way. Typically, due to low traffic at nights, the configuration set at those times is a 1A, meaning that a single sector (thus, a single controller) is in place.
This leads to the fact that configurations available are fixed, but configuration "in place" varies during day, adapting capacity resources (Air Traffic Controllers, mainly, as more sectors open mean more capacity, but also more controllers) to the expected demand.

Metadata: Not available.

Scale:
From 01/06/2011 to present. The range of available areas is the European airspace, 4KB per day.

Provision methods:
Raw data: Text plain files. ('.cfg', '.cos', '.ncap', '.spc', '.gsl/sls', '.gar/are').

Partner(s) responsible: CRIDA & BRTE

Network Management

This category covers the data sources that contain Network Management (also known as Flow Management) information, thus the regulations put in place to ensure a proper Demand Capacity balance in a tactical way. There is a single source considered: CFMU, coming from the Network Management organization (Eurocontrol), thus covering European airspace.

CFMU: This data source provides list of flights and regulations that these flights may have. The source is separated in two files, one for flights and other for regulations. When a flight has a regulation, the code of the regulation applied is provided on the row. When a regulation is applied to a flight, a CTOT is set for the flight and a delay over this time without a window [-15;+5] minutes is recorded.

Metadata: Not available

Scale:
IFS data is available for from 2013 till 2016 (complete years).
datAcron will exploit data concerning the Spanish airspace regulations.
On structured version, one day of information is some MBs on compressed files.

Provision methods:
Structured version could be provided by web services, database direct connection or CVS files.

Partner(s) responsible: CRIDA & BRTE

Synthetic trajectories

This data source represents trajectories generated by a Trajectory Predictor (TP). A TP is a software/routine that is included in any software or tool that needs to forecast the future state of the

aircraft to perform its tasks. Depending on the particular application that the TP serves, the level of detail (i.e., number of variables and number of aircraft states) that needs to be included in, the aircraft state may vary. The trajectory generated by flight management system (FMS) contains multiple aircraft states (e.g., an aircraft state at least each 30 seconds) and each state multiple variables, such as latitude, longitude, altitude, time, calibrated airspeed or mass. These variables are used by other FMS subsystems to generate guidance modes, or monitor the aircraft performances.
The format of a synthetic trajectory depends on the particular TP model and software design and implementation of that model.
datAcron source of synthetic trajectories messages is a stand-alone model-based TP engine developed by BRTE (BRTE TPE) to generate trajectories for a set of given input information (flight plan, weather, aircraft model, operational context). For a given particular flight, BRTE TPE can be used to generate different alternative synthetic trajectories representing all the possible conditions that the flight may encounter of for what-if analysis. Under the aviation use case, BRTE TPE will be used to generate reference synthetic trajectories that will serve as benchmark.
In datAcron we provide in D6.2 the format generated by a stand-alone TP engine developed by BRTE to generate trajectories.

Metadata: Not available.

Scale:
Synthetic trajectory data will be available under demand for the particular scenario that is going to be studied. In principle, the range of available dates should coincide with the range of available flight plans and surveillance data. Ideally, for a particular data set of flight plans and/or corresponding surveillance data, there should be a set containing *n* synthetic trajectories per flight, where *n* is driven by the particular use case scenario that is studied.
One synthetic trajectory for one aircraft containing 300 aircraft states, each of them with 56 different variables would be around 1 MB (XML format; text file of near 20000 lines). This size can be easily reduced either by decreasing the number of variables and/or the sample rate (number of aircraft states)

Provision methods:
Historical data: Files processed for specific time periods can be assembled and delivered through Internet or physical media (i.e. DVD).

Partner(s) responsible: BRTE

Aircraft Identification

This data source provides details concerning aircrafts in a trajectory. In ADSB sources the aircraft is identified by ICAO 24-bit address or (informally) Mode-S "hex code". The ICAO 24-bit address can be represented in three digital formats: hexadecimal, octal, and binary, and typically in ADSB sources will be represented in hexadecimal. One of the most important data to obtain given the ICAO address is the model of the aircraft, or more specifically, the ICAO Type Designator, according to DOC 8643.

Metadata:
A csv file will be used to distribute the list of known aircrafts for datAcron.
The file will contain the following fields:

- icao character varying (6) - ICAO address in hexadecimal.
- regid character varying - Unique alphanumeric string issued by a National Aviation Authority to identify an aircraft.
- mdl character varying – Aircraft Type according to ICAO DOC 8643
- type character varying- Aircraft model
- operator character varying – last known operator

Scale:
The identification will be available for all aircrafts in the trajectories.

About 25Mb.

Provision methods:
A csv file will be used to distribute the list of known aircrafts for datAcron.

Partner(s) responsible: BRTE

Flight Plans

Flight Plan is a essential category as contains the information that triggers a lot of operational decision, both in planning and execution phase, and both on the Air Navigation Service Provision side, and in the airline one.
Two sources of information are considered: Network manager and (again) DDR. They have similar information but each one of them may be preferable for different Aviation scenarios.

Network Manager Flight Plans: The Flight Plan is the specified information provided to air traffic services units, relative to an intended flight or portion of a flight of an aircraft.

Metadata:
Flights Plan data is compliant with ICAO 4444 Flight Plan 2012 and is a direct translation to XML format.
Detailed field explanation available at:
http://www.eurocontrol.int/sites/default/files/content/documents/nm/network-operations/HANDBOOK/ifps-users-manual-current.pdf

Scale:
The size of the XML file for one flight with all updates can be around 400kb. A file that refers to a flight with a subset of changes can be about 150Kb.

Provision methods:
A xml file for every Flight Plan update and per flight will be used to distribute Flight related information.

Partner(s) responsible: CRIDA & BRTE

DDR Flight Plans: This dataset is focused on the historical traffic data stored in ALLFT+, which contains the flight plans information.
The format of the files is plain text, where each line contains all information of a single flight. Details on this source are provided in D6.2.

Metadata: Not available

Scale:
From 01/06/2011 to present. About 1,30 GB per day.

Provision methods: Daily Files for European airspace.

Partner(s) responsible: CRIDA & BRTE

Contextual information data sources

This contextual information category is the complementary one to the airspace data. This category contains a single data source and it includes purely static data, describing the operation environment: It describes the existing airspace organization, with no gaps or overlaps, and all the possible ways of combining volumes to generate different operational sector configurations, also with the associated sector capacities, or flights that a sector can manage in a period of time.

Network Manager Contextual Information: The Contextual Information provided at European level by Eurocontrol is intended to provide services related to the management and sharing of airspace data (e.g. airspaces, routes, aerodromes, etc.)
The airspace data consists of two types of information:

- Airspace Structure Information for retrieving up-to-date airspace data from the CACD database. The CACD database is the repository for the environment data (a.k.a. airspace data) used in the network management systems to perform Flight Planning and Flow Management. This data includes AIP (Aeronautical Information Publication) concepts (such as Routes, Points and Aerodromes), and non-AIP concepts (such as Flows, RAD (Route Availability Document) Restrictions and Traffic Volumes).
- Airspace Availability Information for querying and modifying the airspace availability information; this includes the Flexible Use of Airspace.

The Airspace services make use of AIXM 5.1/ADR-E (http://www.aixm.aero) types when possible (ADR-E stands for ADR Extension). The following main information areas are in the scope of AIXM:

- Aerodrome/Heliport including movement areas, services, facilities, etc.
- Airspace structures
- Organizations and units, including services
- Points and NavAids
- Procedures
- Routes
- Flying restrictions

Metadata:
A xml file for every AIXM feature or set of AIXM features update for will be used to distribute context information.
The structured Airspace Information uses a subset of AIXM 5.1, and includes three tables about Airports, as also detailed in D6.2.

TABLE1

Attributes and Associations	**Item - Description**
name	Airport name in plain text
locationIndicatorICAO	ICAO airport Identifier
designatorIATA	IATA airport identifier
controlType	Authority controlling the airport (CIVIL / MILITARY)
defaultTaxiTime	Default time elapsed from aircraft off-block time from the gate to lineup in the runway
servedCity	City served by the airport
ARP	Point corresponding to the geometrical center of the airport

TABLE2

Attributes and Associations	**Item - Description**
hostAirport	Link to Airport feature containing the host airport of the airport set
dependentAirport	Link to Airport feature containing a dependent airport of the airport set

TABLE3

Attributes and Associations	**Item - Description**
airportHeliport	Link to Airport feature containing an airport that belongs to the set
airportHeliportSetPattern	To implicitly add groups of aerodromes to an AirportHeliportSet based on a pattern in the designator. The value is a string of alphabetic characters and represents the first letters of the ICAO identifier. The semantic is therefore the following: "include all aerodromes whose ICAO identifier starts with the pattern". For example a pattern such as "EB" includes all aerodromes whose ICAO designator starts with 'EB'.

Scale:
TABLE 1: AirportHeliport Feature size is about 10Mb.
TABLE 2: AirportHeliportCollocation Feature size is about 15Kb.
TABLE 3: AirportHeliportSet Feature size is about 5Mb.

Provision methods:
XML files will be distributed with current airports data.

Partner(s) responsible: CRIDA & BRTE

3.2 Maritime data sources

The maritime use case should be supported by data sources including:

- Automatic Identification System (AIS, www.navcen.uscg.gov/?pageName=AISmain) messages broadcasted by ships;
- Maritime regulations, specifying the legislation and the rules for navigation and fishing;
- Marine protected/closed areas, where fishing and sea traffic may be (temporarily) forbidden;
- Traffic separation schemes and Nautical charts, useful to define vessel routes;
- Vessel routes and Fishing areas estimated from historical traffic data;
- Registry data on vessels and ports;
- Records of past events, such as incidents and illegal activities reports;
- Meteorological and oceanographic data (METOC) on atmospheric and sea state conditions and currents;

Surveillance information

Vessel position reports are coming from multiple sensors and sensor networks and in the context of datAcron concern AIS (Automatic Identification System) data:

AIS (Automatic Identification System) data.

According to the European Commission regulations several types of ships are obliged to broadcast AIS messages, including: ships of 300 gross tonnage and upwards in international voyages; 500 and upwards for cargoes not in international waters and passenger vessels; and, more recently, smaller fishing vessels.
Two main classes of messages are distinguished:

- Kinematic messages from which 2D vessel routes can be derived.
- Static messages providing ship meta-information such as ship identifiers (MMSI and IMO number), name, type, and dimension of vessel, and route-based information, such as destination (Port of Call), danger, Estimated Time of Arrival (ETA), draught.

The AIS data for the datAcron project has been sourced from a range of terrestrial AIS (T-AIS) and satellite AIS (S-AIS) sources. The terrestrial sources are collated from various sources in Europe and decimated to limit the amount of data. The satellite data is obtained from the ORBCOMM constellation of satellites of various generations and include an AIS receiver on the International Space Station (ISS), a range of older generation satellites and 11 new generation satellites that go to make up 19 sources of satellite data.
Each of the satellites can only download AIS data when there is a ground station within their coverage footprint. There are three satellite ground stations in Morocco, Italy and Norway, servicing the area of interest for which the datAcron project is consuming AIS data. The overlap of the satellite footprint, the orbit of the satellite and the coverage of a ground station primarily affects the delay between when an AIS signal is received on the satellite to when it is available to the datAcron partners via the satellite data collection and processing network.
datAcron has defined the following area of interest for AIS data to be obtained and provided:

North West Coordinates: 52 Degrees North, 12 Degrees West
South East Coordinates: 30 Degrees North, 30 Degrees East (which will be expanded further east, including all the Mediterranean Sea).

Metadata:
AIS messages provided by IMISG include a comment or TAG block which provides metadata and additional information for the IEC 61162-1 message, with the following format:

Identifier	Description
s:	The source of the message.
c:	The unix timestamp of the message when received (seconds since midnight, January 1st, 1970)
T:	The human readable timestamp of the message when received in yyyy-mm-dd hh.nn.ss
e:	The message error flag. Bits in this identifier are set to '1' if any of the 15 errors described in Annex A are true. If no error is present in the data, is this field excluded from the message output
i:	Proprietary data and contains the following fields, separated by a '|' character:

Identifier	Description
X=	Data source RX / TX capability = always set to '0'
D=	Data source 'delayed data flag' = always set to '1'
T=	Proprietary timestamp of the message
P=	The IP address and port where the message was received by the MSA
R=	The direction of the message

Scale:
The data growth on the IMISG existing satellite AIS systems is about 7GB per day and this is expected to increase to about 12GB per day based on the increased number of satellites collecting AIS data and additional data available within the message structure.
Limiting the data to the Mediterranean area only the AIS data per day is expected to be about 15% of the above for the initial phase (1.8GB per day). As and when the metadata is included, the data served is expected to increase by a further 30% on top of what is available in the initial phase (2.3GB per day).
datAcron also plans to include some terrestrial AIS data that could add a further 3GB per day only as archival data.

Provision methods:
The AIS data, once having been tagged and analysed, is served as a data stream via a data server. The AIS data is stored along with ownership and filter parameters.
To extract historical data, a query can be executed via the Human Machine Interface (HMI) or can be extracted using a REST web service.
Real time data is served within 1,000ms of being available. Connection to the AIS data stream is made by using a Secure proxy and the TCP/IP protocol.
Both the S-AIS and T-AIS data is provided in the international IEC 61162-1 format. The metadata is attached to each message in the NMEA 0183 version 4 TAG block method.

Partner(s) responsible: IMISG, CMRE, NARI

METOC Data

Weather data and ocean data from forecast models and from observations (e.g. sensor data), which are openly available from several providers, can help validate analysis results and reduce false alarm rate, for example identifying sea and weather conditions that force vessels to change direction or modify their normal behaviour. They can also be used to characterize seasonal trends in traffic routes, and to further contextualize movement parameters such as speed of vessels.

The reference source of harmonized oceanographic data in Europe is by far the Copernicus Marine Environment and Monitoring Service (CMEMS)[1], developed by the EU as part of the European Programme for the establishment of a European capacity for Earth Observation and Monitoring. This operative service provides an interactive catalogue of updated oceanographic products produced by the network of oceanographic centres in Europe.

In particular, weather and ocean datasets that can be used in the maritime use case include:

The Mediterranean Sea Physics Analysis and Forecast (MFS) model: This is a hydrodynamic wave model for the Mediterranean basin that forecast physical ocean variables including the sea currents (sea water velocity) and the sea surface height above the sea level. MFS is one of the products available in the CMEMS catalogue. (see also the MFS data product documentation [2]).

Metadata:
ISO19115 and ISO19115-2

Scale:
MFS is updated daily, historical data are available from 2013-01-01. Variable values are available as daily mean and hourly mean. Its geographical coverage is Latitude North 45.937, Latitude South 30.187, Longitude East 36.25, Longitude West -15, with a resolution of 1/16 degree (i.e. , ca. 6-7 km).

Provision methods:
Updated forecast is freely available for download from CMEMS catalogue[3] using FTP download and through Python scripts using MOTU client.
OGC Web Map Service (WMS) and Catalogue Service for the Web (CSW) services are also available.
MFS is distributed in netCDF binary multidimensional array format. Full documentation and APIs for accessing netCDF are available[4].

Global Ocean Wind observations available through CMEMS service, with horizontal resolution of 0.25x0.25 degrees and 6 hours in time[5], as well as daily and monthly mean from 2007;

Ocean Wave model data from the European Centre for Medium-Range Weather Forecasts (ECMWF) for example from the ERA-Interim model[6], which offers many other atmospheric variables, and is available daily until 2015; the ERA-20C model[7] available until 2010. Other open ECMWF datasets are available[8].

Model data for the Adriatic from the Adriatic Forecasting System[9], which makes available one week of data.

Model data for the Ionian Sea, from the Ionian Forecasting System[10,] which makes available three months of data.

---

1 Copernicus Marine Environment and Monitoring Service: marine.copernicus.eu

2 http://marine.copernicus.eu/documents/PUM/CMEMS-MED-PUM-006-001.pdf

3 MFS: from marine.copernicus.eu/web/69-interactive-catalogue.php?option=com_csw&view=details&product_id=MEDSEA_ANALYSIS_FORECAST_PHYS_006_001_a

4 http://www.unidata.ucar.edu/software/netcdf/docs/netcdf_data_set_components.html

5 marine.copernicus.eu/web/69-interactive-catalogue.php?option=com_csw&view=details&product_id=WIND_GLO_WIND_L4_NRT_OBSERVATIONS_012_004

6 apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/

7 ERA-20C apps.ecmwf.int/datasets/data/era20c-daily/levtype=sfc/type=an/)

8 apps.ecmwf.int/datasets/

9 www.ionioproject.eu

10 ionioproject.hcmr.gr

In-situ Observation on the Adriatic Sea, from the Ionian in-situ database[11], with historical data from 1986.

The Sea Conditions dataset[12], which includes Significant Wave Height (SWH).

The National Oceanic and Atmospheric Administration (NOAA) datasets[13], including global meteorological and oceanographic datasets from cooperating networks of ships and buoys.

Energy data from Seabed Habitat from EMODnet include wind, waves and currents harmonized at European Level[14].

Physics data from EMODnet, including Sea water temperature, salinity or density, water currents, level, Waves and winds, Light attenuation, Atmospheric parameters at sea level, HF radar data[15].

Concluding the above, there are many data sources concerning whether data for different purposes (i.e. different attributes) and different areas. As it is said, the reference source of harmonized oceanographic data in Europe is the Copernicus Marine Environment and Monitoring Service (CMEMS).

datAcron will refine/update the sources of contextual data to be used for maritime operational purposes and for achieving research objectives.

Partner(s) responsible: NARI and CMRE

Contextual Information data sources.

Several data sources, mainly European sources or sources provided from third parties, can be used to contextualize the Maritime Sustainable Development scenario, as well as to further characterize the other scenarios included in the fishing activities monitoring use case, as it has been detailed in D5.1, including:

EU regulated fishing areas included in the recent European proposal on the conservation of fishery resources and the protection of marine ecosystems through technical measures. These are given as thematic as well as spatial information, together with the coordinates of the regulated fishing areas.

FAO[16] **and ICES**[17] **Fishery Statistical Areas, and Fish Catches by FAO**. The Community Fishing fleet register, a European register of official fishing vessels maintained by the European Commission.

The Community Fishing fleet register, a European register of official fishing vessels maintained by the European Commission[18].

The European Marine Observation and Data Network (EMODnet) datasets, that include a series of standard and harmonized datasets including European coastal maps, human activities such as ports and fishing areas, biological datasets, a digital terrain model for bathymetry, etc. The EMODnet portal is an entry point for harmonized marine data generated by 100 organisations, free of restrictions on use. These datasets can help contextualizing the illegal fishing use case, as well as other use cases that will be developed.

[11] www.mediterraneanmarinedata.eu/ionio/home.htm
[12] sea-conditions.com
[13] www.ndbc.noaa.gov/data
[14] http://www.emodnet-seabedhabitats.eu/default.aspx?page=1934
[15] http://www.emodnet-physics.eu/Map/service/Catalogue.aspx
[16] (http://www.emodnet-humanactivities.eu/search-results.php?dataname= FAO+ Fishery+ Statistical+Areas)
[17] (http://www.emodnet-humanactivities.eu/search-results.php?dataname=ICES+Statistical+Area)
[18] http://ec.europa.eu/fisheries/fleet/index.cfm

The marine protected areas in Europe defined by the NATURA2000 ecological network of protected areas and freely downloadable from the European Environmental Agency (EEA) website.

Environmental biodiversity datasets (e.g. e, marine biodiversity, waste, sediments) harmonized at European level and freely downloadable in different formats from the EEA website that can be useful to further develop the maritime sustainable development scenario.

A high resolution coast line map from EEA, created for highly detailed analysis, e.g. 1:100 000, for geographical Europe. The criteria for defining the coastline is the line separating water from land. The EEA coastline is a product derived from two sources: EU-Hydro and GSHHG. In the 2015 version of the dataset, several corrections were made in the Kalogeroi Islands (coordinates 38.169, 25.287) and two other Greek little islets (coordinates 36.766264, 23.604318), as well as in the peninsula of Porkkala (around coordinates 59.99, 24.42).

The World Port Index (WPI), an open, freely available and distributable port database maintained by the National Geospatial-Intelligence Agency that contains the locations, the physical characteristics and the facilities and services offered by major ports and terminals world-wide. It contains approximately 3700 entries. It is distributed as a PDF report, as an access database and as ESRI shape file.

The Open Sea Map, an open and free nautical chart, including beacons, buoys and other navigation aids as well as port information, repair shops and chandlers. It can integrate the World Port Index supporting the maritime use case with information on port facilities and provide also information on vessel routes close to ports. It is an Open Street Map project and data are freely available, usable and distributable according to the ODBL open data common license.

Official nautical charts provide cartographic information that could help understand sea traffic (IHO-S-57, format for Electronic Navigation Charts, ENC).

datAcron will further refine the sources to be used for the operational purposes described in the maritime use cases and towards achieving its research objectives. For each source to be exploited, there should be information concerning:

- Metadata accompanying each source.
- Scale of data source, subject on the areas to be covered and growth rate.
- Provision methods subject to the different data modalities being provided (historical/archival, real time).

Partner(s) responsible: NARI and CMRE

3.3 Data to be created

Data to be created during the project and by means of the project-specific computing components are as follows:

Analytics components results

These concern trajectories detected and predicted, computed by the components to be developed in WP2, events, either low-level events computed by the in-situ components (WP1) or high-level events computing by the event recognition and forecasting components (WP3). The methodology for creating these datasets are within the core of datAcron research objectives and are described in the description of action for the project.

Metadata:
There are not metadata for these data sources / datasets. However, datAcron has developed an RDFs Schema for describing the data for trajectories and events, their interlinking, as well as their interlinking with other data sources (contextual and weather). The metadata to be used for describing such datasets will be according to the schema for describing any data source and data set (Section 4).

Scale:
Trajectory and events data will be available under demand for the particular scenario that is going to be studied. In principle, the range of available dates should coincide with the range of available surveillance data. Ideally, for a particular data set of surveillance data should be a set containing *n* synthetic trajectories per moving entity, and m specific events detected or forecasted, where *n* and m are driven by the particular use case scenario that is studied.
The typical size of each trajectory and event is difficult to be estimated at this stage of the project, but their size depends on the compression achieved (e.g. number of moving entity states), as well as by the number of variables in each state (e.g. contextual and weather variables associated).

Provision methods:
Trajectories and/or events should be retrieved by specific queries addressed to the datAcron store, with respect to spatiotemporal constraints and specific properties of moving entities, contextual and weather data.
However, specific datasets – to be decided among partners responsible for data sources and for the computation of analytics results – may be devised and stored in a specific form (e.g. as RDF or csv files) accompanied with specific metadata.

Partner(s) responsible: UPRC (WP1 & WP2), NCSD'D' (WP3), FRHF (WP4).

Surveillance data synopses

These concern synopses of moving entities trajectories, aiming to reduce the velocity and volume of surveillance data without compromising the accuracy of analytics components to be developed. One of the key project objectives and key performance indicators is to achieve a rate of compression that is greater than 95%, subject to the accuracy of detection and forecasting components (both for trajectories and events).

Metadata:
There are not metadata for these data sources / datasets created. However, datAcron has developed an RDFs Schema for describing the data for synopses of trajectories, and their interlinking to other data sources (contextual and weather). The metadata to be used for describing such datasets will be according to the schema for describing any data source and data set (Section 4).

Scale:
Trajectory synopses will be available as streaming data and will also be stored in the datAcron store. In addition to that, such synopses may be made available under demand for any particular scenario that is going to be studied. Ideally, for a particular data set of surveillance data should be a set containing a synoptic trajectory data per moving entity.
The typical size of each trajectory synopses depends on the compression achieved (e.g. number of moving entity states), and if such a trajectory has been linked to other data, it also depends by the number of variables in each state (e.g. contextual and weather variables associated).

Provision methods:
Synopses of trajectories should be made available either as streaming data to any other components of the datAcron integrated system, or to third parties, if this is agreed among partners. These can be retrieved by specific queries addressed to the datAcron store, with respect to spatiotemporal constraints and specific properties of moving entities, contextual and weather data. Specific datasets – to be decided among partners responsible for data sources and for the computation of synopses – may be devised and stored in a specific form (e.g. as RDF or csv files) accompanied with specific metadata.

Partner(s) responsible: UPRC, FRHF (WP1 & WP2).

4. STANDARDS AND METADATA

4.1 Metadata (Standards and Methodology for Capturing/Creating Metadata)

Data sources to be exploited, as described in the previous section and detailed in deliverables D5.2 and D6.2, may be accompanied with domain-specific metadata descriptions. However, there is a diversity of the metadata being used, either at the syntactic or at the semantic levels, while we need metadata for registering sources and datasets in a common repository for locating, accessing, and sharing data.
In datAcron there are two complimentary ways to locate, access and share data:

First, an implicit, but quite natural and direct way – based on the datAcron data value chain: This is based to the fact that all data, from any data source, together with results created (computed), will be integrated/interlinked and be stored in the datAcron store. Thus, in principle all stakeholders may access the data store and fetch data under a coherent schema, according to specific spatiotemporal constraints and constraints on other qualities represented by means of RDF property values and semantic types.

Second, an explicit one, where any coherent subset of data throughout the datAcron data chain is provided as a downloadable file in a specific form, or via an API. Such a subset may concern integrated views of data stored in the datAcron store, or subset of data in its raw form (i.e. in the form provided by data source providers) or being processed in a specific way (e.g. for getting just a subset of the variables provided, in a required form, etc), or data created by any of the datAcron research components. In this case, each such dataset will be described by metadata properties. For this purpose, we will use a widely used schema such as DCAT (Data catalogue vocabulary).

DCAT [4] is a widely used standard that is RDF designed to facilitate interoperability between data catalogs published on the Web, although initially intended for governmental data only. An application profile of the W3C standard DCAT called DCAT-AP (DCAT Application Profile) [5] is used within Europe.
A dataset in DCAT is defined as a "collection of data, published or curated by a single agent, and available for access or download in one or more formats".
A dataset does not have to be available as a downloadable file. For example, a dataset that is available via an API can be defined as an instance of dcat:Dataset and the API can be defined as an instance of dcat:Distribution. DCAT itself does not define properties specific to APIs description.

An overview of metadata description based on DCAT is given in Figure 3.

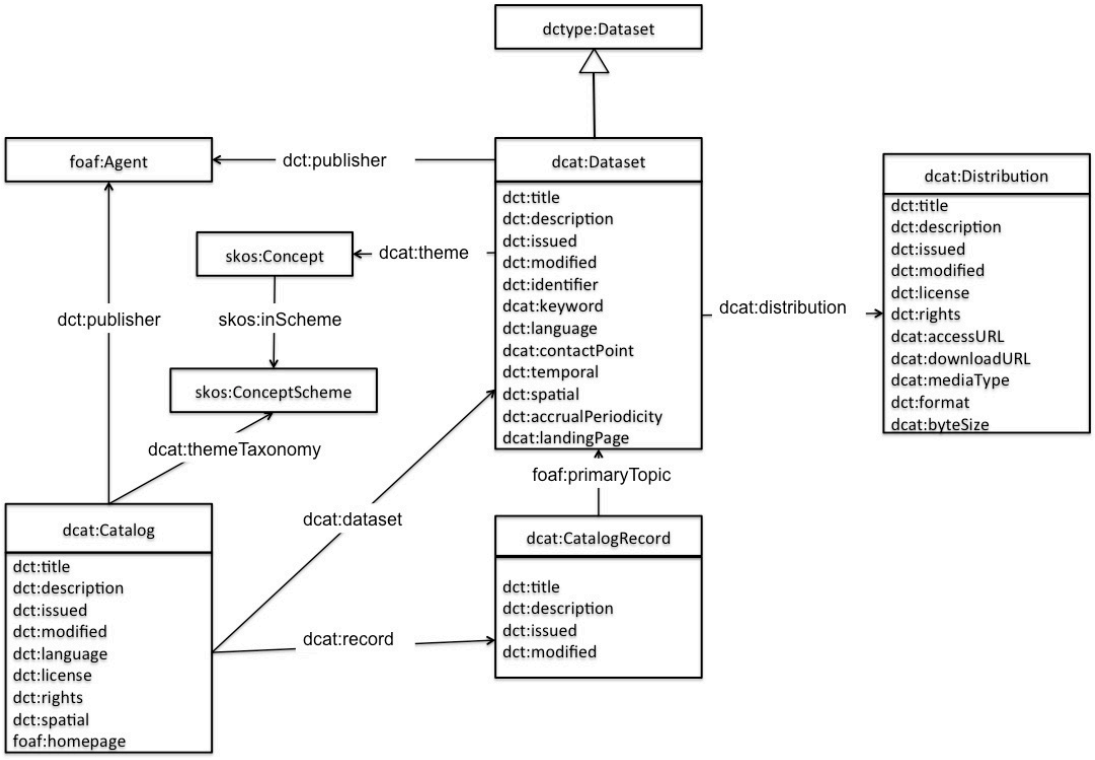

Figure 3. DCAT overview

DCAT provides properties for datasets version control (e.g. via dct: issued and dct:modified), spatiotemporal characteristics of data sets (i.e. area dct:spatial and period dct:temporal covered), as well as licensing via the dct:licence property for each dataset distribution.

Although there are several tools for DCAT, these do not seem to be suitable for extracting metadata descriptions from datAcron datasets. This imposes a workload to creating metadata that we aim to keep at a minimum level, although clearly to the level of quality that allows managing and sharing datasets.

The use of a data management tool such as DCAN[19], CKAN[20] and/or widely used content management systems (CMSs) will be examined for providing search and presentation functionality for datasets.

4.2 Naming conventions, organization.

Datasets to be provided will be stored together with their metadata descriptions providing a link to a dataset file or to an API, according to the metadata schema to be used.
Files will be organized according to (a) the type of data source/dataset provided or combinations of these, (b) the geographical coverage, (c) the temporal coverage, (d) their origin and (e) date issued, in corresponding folder structures according to the above-mentioned characteristics and in the order specified above.
The name of each file will be constructed using the above-mentioned characteristics.

[19] http://docs.getdkan.com/dkan-documentation/dkan-overview
[20] http://ckan.org/

5. DATA SHARING

datAcron has to consider the possibility to share datasets with respect to the legal, IPR and ethical constraints applied to the data sources to be used. Specifically, we have to consider where, how, and to whom the data could be made available.

Datasets will be shared either by opening a specific API for third parties to fetch data from the datAcron store, or via making such datasets downloadable as single files.

The methods used to share data provided by beneficiaries will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. Regarding IPR and legal restrictions, these have already specified in the background section of the datAcron consortium agreement.

BRTE specific background information:

Describe Background	Specific limitations and/or conditions for implementation (Article 25.2 Grant Agreement)	Specific limitations and/or conditions for exploitation (Article 25.3 Grant Agreement)
European Air Traffic Data (Flight Plans, Airspace information, Airspace use plan, AIXM + ADR, operational airspace data, AIP's, routes, notams, restrictions, regulations)	BRTE can use European Air Traffic Data in datAcron activities. The transfer of data to other consortium members would require prior approval from the third party data owner.	BRTE has to explicitly mention the third party data owner copyright of the data in any dissemination involving data publishing.
Weather data	BRTE can use public domain and third party weather data in datAcron activities. The transfer of the third party data to other consortium members would requires prior approval from the third party data owner.	Public data can be disseminated stating that such material is not subject to copyright protection Third party data can be disseminated following license rules (i.e. not modifying it and with the right attribution)
ADSB messages collected by receivers in Europe	There are no commercial restrictions on how to use this data. Everybody can publish the data for free or to use it for commercial purposes.	Receiver network may require credits on the published dissemination.

CRIDA specific background information:
CRIDA is a research center, not a data-generator, which entrusted to use, exploit and maintain the datasets generated by ENAIRE, owner of that data. CRIDA is entitled to use the data for research purposes such as the datAcron project (where it acts as data provider), and can share them with the consortium only for research purposes.

The following general conditions for accessing and using the data provided by CRIDA apply:

- The Dataset provided will never be the raw data obtained from operational sources, but the result of the fusion and processing performed by CRIDA.
- The Dataset can only be used to achieve the research purposes stated in the datAcron Technical Annex. No other use is allowed.
- The user is not allowed to create or derive new datasets from the original one.
- The Dataset will be anonymized through its “delocalization” (either in place or time) except in specific cases where the specific confidentiality agreement allows data concerning places and time to appear. Delocalization will be done to disable its traceability to operational events and situations. The user might not change or process the Dataset in any way to remove the anonymity of the data.

- The Dataset will be stored physically in the premises of CRIDA. No physical copy of the Dataset will be provided. The users will not store, copy or otherwise move the Dataset (physically or logically) to databases or systems outside the premises of CRIDA.
- Access to the Dataset will be granted as Needed by the research members of the consortium either for implementing their own tasks under the datAcron action or for exploiting their own results.
- Access to the Dataset will be granted under a specific confidentiality agreement that will clearly identify the Dataset users, purpose, usage timeframe and any specific clauses that might be needed.

The Dataset access confidentiality agreement will be signed before the start of the work to set the conditions of this access. CRIDA is committed to provide access in a secure way, royalty-free, in order to achieve the research goals of the project in an efficient way.

- Access to the Dataset will be provided only for the specific purposes and scheduled time windows described in each one of the user confidentiality agreements.
- For security and confidentiality reasons no permanent access to the Dataset will be granted.
- The Dataset will be available through a secure channel as specified by CRIDA.
- Subsets of the Dataset may be stored, copied or otherwise be moved (physically or logically) to databases or systems outside the premises of CRIDA, as Needed by the research members of the consortium either for implementing their own tasks under the datAcron action or for exploiting their own results. These subsets will be treated as "confidential" and will be provided only under a specific confidentiality agreement that will clearly identify the characteristics of the subset, users, purpose, usage timeframe and any specific clauses that might be needed. This subset will be labelled specifically for the purpose needed.

The above mentioned restrictions apply to the following data sources:

- **Spanish ATC Platform On-line Flight Plan Data**
  Including Creation, update and deletion messages for the flight plans in Spanish airspace
  Dataset containing information from 2013-2015 (3 years)

- **Spanish ATC Platform Off-line Flight Plan Data**
  Including relevant flight messages for all the flights in Spanish airspace (Flight plan creation, deletion and major updates, sector enter, sector leave, …)
  Dataset containing information from 2009-2015 (7 years), with focus on 2013-2015

- **Spanish ATC Radar Data**
  Including actual radar tracks for all the flights in Spanish airspace
  Dataset containing information from 2009-2015 (7 years), with focus on 2013-2015

- **Spanish Sector configurations**
  Including actual sector configuration put in place for all the Spanish airspace
  Dataset containing information from 2009-2015 (7 years), with focus on 2013-2015

IMISG specific background information:
As far as the AIS related data sources from IMISG are concerned, the following background is identified and agreed upon for the Project.

Describe Background	*Specific limitations and/or conditions for implementation (Article 25.2 Grant Agreement)*	*Specific limitations and/or conditions for exploitation (Article 25.3 Grant Agreement)*
Terrestrial & Satellite AIS data. These will be enriched with metadata, and include archival and real-time data made available via REST query services and CSV data files.	Access is on a royalty-free basis to background, for beneficiaries needed to implement their own tasks under the action.	Access is under fair and reasonable conditions to background needed for beneficiaries to exploit their own results. The raw data cannot be published although samples can be provided.

To limit restrictions, datAcron partners will aim to gain the consent of data providers in limited datasets to be shared, maybe for specific periods and under specific licenses.
Concerning licensing, the datAcron Ethics Management Plan D.8.5 (section 3.3, Identification and documentation of licensing options) makes a clear identification of all options that the project must consider for making datasets available to 3rd parties.

An important challenge is the integration/interlinking of data from data sources and datasets having different usage and access policies. Interlinking data with certain constraints and requirements with data that are publicly and freely available impacts the access policy applied. This is not problematic when the aggregated data is subject to the same or more restrictive access, usage and dissemination conditions as the source data themselves.
There should be great care to avoid problematic situations where data are being distributed throughout channels to stakeholders that do not satisfy the conditions stipulated by one of the sources. To prevent incorrect usage, managing the access, usage and dissemination conditions of the newly created datasets is important. That information will form the cornerstone of the correct implementation of the required access, usage and dissemination policies.

datAcron will apply the following strategy:

- The data providers ensure that for each dataset, the access, sharing and reuse policy is known. In case the dataset integrates data from multiple data sources, the more restrictive access, sharing or reuse conditions apply. This information is reported to the datAcron Executive Board, together with a request to apply a specific access/share/reuse policy to a dataset.
- The datAcron Executive Board decides on the specific policy to be applied for each dataset, and decides on the license to be applied. The decision must be taken in 3 weeks from the date of request.
- The exposure of the dataset is made according to the above-made decision. Metadata are updated accordingly.
- The metadata of the created outcome is always public. This ensures transparency of the knowledge that is gathered.

The workflow to be followed for deciding on granting open access to specific parts of the research data (datasets) and under which restrictions, license etc is as follows:

Given a specific publication or digital research data generated in the action by a beneficiary, the beneficiary must aim to deposit at the latest on publication time of the manuscript, or at the latest one month after the generated dataset (if this is not associated to a publication), the research data (datasets) in the deposited scientific publication, together with the metadata. To do so

- Metadata descriptions of the datasets are specified to the datAcron repository, together with either a link to a dataset file, or to an API for fetching the associated data from the datAcron store.
- The beneficiary ensures that for each dataset, the access, sharing and reuse policy information is known. In case the dataset integrates data from multiple data sources, the more restrictive access, sharing or reuse conditions apply.
  This information, together with the (link to the) dataset is provided to the Executive Board at least 4 weeks prior to the deadline for registering the dataset.
- The datAcron Executive Board decides on the specific policy to be applied for each dataset, and decides on the license to be applied. The decision must be taken in 3 weeks from the date of request.
- The exposure of the dataset is made according to the above-made decision. Metadata are updated accordingly.
- The metadata of the created outcome is always public. This ensures transparency of the knowledge that is gathered.

datAcron will register its repository to a widely used registry such as re3data.org [6].

6. ARCHIVING AND PRESERVATION

Data sources concerning moving entities' behavior in the air and in the sea, are growing in size due to market and service demands. datAcron will exploit these data sources, together with other geospatial, environmental and weather data sources, and develop scalable methods for processing these data. There is a growing interest in archiving, sensing and performing analytics over mobility and behavioral data. In datAcron, we aim to address issues concerning long term storage of these data as well as research generated data.

These issues related to safety and long-tem data storage are complicated when the data evolves or/and they could be merged with other data coming from other sources. So, in this project we need to consider:

- What is the volume of the data to be maintained?
- What is considered long-term (2-3 years, 10 years, etc.)?
- Identification of archive for long-term preservation of data.
- Which datasets will need to be preserved in the archive?
- What about relevant dependent datasets?

Preserved datasets will need to be updated and this means a data preservation policy and process will need to be defined.

A central consideration for any long-term DMP is the cost of preserving that data and what will happen after the completion of the project. Preservation costs may be considerable depending on the exploitation of the project after its finalization. Examples include:

- Personnel time for data preparation, management, documentation, and preservation,
- Hardware and/or software needed for data management, backing up, security, documentation, and preservation,
- Costs associated with submitting the data to an archive,
- Costs of maintaining the physical backup copies (disks age and need to be replaced).

7. CONCLUSIONS

The datAcron Data Management Plan (datAcron DMP) details what data the project will collect, generate, how will be exploited or made accessible to all stakeholders, how and what data sets will be made available for verification and re-use, and how it will be curated and preserved during the lifetime of the project, in accordance to the datAcron data value chain. In order for the document to be self-contained, a comprehensive view of the datAcron data lifecycle is provided with appropriate definitions of terms being used and stated assumptions under which the plan has been devised, with a succinct description of stakeholders' groups.

Data sources to be exploited per domain are described, also specifying information on existing domain-specific metadata per data source, size of data sets, modality, and provision methods. The deliverable specifies how data sources and datasets are to be associated with metadata and be archived in a repository.

Furthermore, information and an initial policy on data sharing is provided, also taking into account, limitations, IPR and legal issues applying to the data sources, together with licensing and ethical issues.

The deliverable concludes with issues concerning archiving and preservation.

8. REFERENCES

[1] Resource Description Format, http://www.w3.org/RDF/ & http://www.w3.org/standards/techs/rdf#w3c_all

[2] UK DMP checklist [http://ukdataservice.ac.uk/manage-data/plan/checklist.aspx]

[3] Guidelines on Data Management in H2020, Version 2.1, 15 Feb. 2016 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[4] DCAT Vocabulary [http://www.w3.org/TR/vocab-dcat/]

[5] DCAT-AP [https://joinup.ec.europa.eu/asset/dcat_application_profile/description]

[6] r3data.org Schema Definition: http://service.re3data.org/schema