H2020-ICT-2015

Grant Agreement No: 687591

Big Data Analytics for Time Critical Mobility Forecasting

datAcron

D8.5 Ethics Management Plan

Deliverable Form

Project Reference No.	H2020-ICT-2015 687591
Deliverable No.	8.5
Relevant Work Package:	WP 8
Nature:	R
Dissemination Level:	PU
Document version:	2.0
Due Date:	31/03/2016
Date of latest revision:	30/03/2016
Completion Date:	
Lead partner:	UPRC
Authors:	Gemma G. Clavell
Reviewers:	George Vouros
Document description:	This deliverable documents the datAcron Ethics deliverable plan
Document location:	Filestore : /datAcron/WP8/Deliverables

History of changes

Version	Date	Changes	Author	Remarks
1.0	28.03.2016		Gemma G. Glavell	1st draft
2.0	30.03.2016		Gemma G. Clavell	Final deliverable

EXECUTIVE SUMMARY

This deliverable details all ethical issues concerning datAcron and how these will be managed by consortium partners, including issues related to experimenting with humans, data protection, dual use and other ethical matters. It provides the necessary forms and instructions to get ethical clearance from the relevant authorities at the start of the project, as well as guidelines and information on how to identify and address legal, ethical and social concerns throughout the 36 months of the project.

TABLE OF CONTENTS

1. INTRODUCTION

There is an increasing awareness of the need to combine investment in new, innovative technologies with a deeper understanding of their legal, social and ethical impacts. Recent controversies surrounding the legality, acceptability and unexpected impact of technological developments in the field of security have led to an increased interest and concern in addressing societal issues in the field of EU research and innovation, with a specific emphasis on legal compliance, longer-term phenomena and the need to better govern and exploit, at an early/conception stage, the negative and positive externalities of technological innovation processes and products in the short, medium and long term.

In order to achieve this, several methodologies and definitions have been provided, linking the specific needs of security projects with broader principles of responsible research and innovation. In 2010, for instance, researchers from several EU-funded projects got together to compile a policy brief on 'Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields'. The group agreed on the following shared definition of responsible research and innovation,

> *Responsible Research and Innovation is a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view on the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society).*[1]

At around the same time, the European Commission's Directorate-General for Enterprise and Industry (DG ENTR) commissioned a report from an Expert Working Group on Societal Impact, mentioning how the Security Research Programme was at a 'key moment', as 'the agenda for security research and development in Horizon 2020 is gradually taking shape'.[2] The group, comprised of experts from the security industry, academia, the NGO and policy communities concluded, among other things, that

- Citizen rights should be a fundamental requirement which could and should lead to drawing boundaries of what is and what is not acceptable in EC funded security research.[3]

- The principles of research ethics should include accountability for scientific procedures, clarification of criteria and choice of research objects, disinterestedness, regard for conflicts of interest, consent of participants in research, confidentiality, transparency of methods and results, respect for data protection and ownership.

- EC-funded research should lead to enhancing the security of European citizens and show how it will affect the lives of citizens in doing so.

- Societal impact should be addressed in the following phases: work programme and annual calls, proposals, negotiation, project execution, and implementation of a completed product, system or techniques in different contexts.

In projects dealing with Big Data, this is a particularly sensitive issue, as data from different sources and of different times is generated, mined, combined, shared and used to forecast future events in

[1] 'Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields'. Available at http://ec.europa.eu/research/science-society/document_library/pdf_06/mep-rapport-2011_en.pdf. Pg. 9.

[2] Report of the Societal Impact Expert Working Group EC DG ENTR Report February 2012. Available at http://cies.ie/wp-content/uploads/2014/05/Report-of-the-Societal-Impact-Expert-Working-Group.pdf. Pg. 4.

[3] Íbid. Pg. 10.

ways that are often unaccountable or that render the information vulnerable. Big Data is defined as 'high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.'[4] The positive impact of these innovations can be maximized if the necessary legal, social and ethical precautions and mechanisms are put in place.

In this context, datAcron Ethics Management Plan, which will run throughout the life of the project, provides consortium partners with the necessary information to address the ethical risks of the project in a way that is consistent with the principles of responsible research and innovation, privacy and data protection rules and guidelines and ethical safeguards, specifically in the handling of data, development of experiments and potential future use of the technology.

1.1. Ethical and responsible innovation in datAcron

This deliverable details all ethical issues concerning datAcron and how these will be managed by consortium partners.

> *datAcron is a research and innovation collaborative project introducing novel methods for threat and abnormal activity detection in very large fleets of moving entities spread across large geographical areas. Specifically, datAcron aims to develop novel methods for real-time detection and prediction of trajectories and important events related to moving entities, together with advanced visual analytics methods, over multiple heterogeneous, voluminous, fluctuating, and noisy data streams from moving entities, correlating them with archived data expressing, among others, entities' characteristics, geographical information, mobility patterns, regulations and intentional data (e.g. planned routes), in a timely manner. Technological developments are validated and evaluated in user-defined challenges focusing on increasing the safety, efficiency and economy of operations concerning moving entities in the Air-Traffic Management and Maritime domains. The datAcron project brings together partners from academia and industry to develop the aforementioned novel methods, together with user and data-provision partners from the two domains, in close relation to user-interest groups, focusing on real-life, industrial and user-defined challenges concerning operations (e.g. surveillance, forecasting of trajectories, characterization, etc.) regarding moving entities in sea and air.*

Specifically, this deliverable will address ethical issues as described in section 5.1. of part B:

> *During deployment of the datAcron individual components and of the integrated prototype for validation/evaluation purposes, partners will comply with any Data Protection Act in Greece, Germany, Spain or France, which ensures that data is protected. The datAcron laboratory set-ups will run in servers hosted in Greece (both use cases), Spain (ATM use case) and France (Maritime use case) and will not be subject to hosting in third countries. During integration and preparation phase, UPRC, NARI and BRTE will consult once again with the national agencies for data and privacy protection to ensure the compliance.*

The report covers issues related to:

- Humans
  - Procedures and criteria that will be used to identify/recruit experiment participants
  - Informed consent procedures

- Data Protection
  - Potential privacy risks related to noise reduction processes

[4] Gartner. (2013). *IT Glossary*. http://www.gartner.com/it-glossary/big-data /

  - The integration of privacy by design within the process
  - Identification and documentation of the licensing options (creative commons, etc...)
  - Detailed information on the procedures that will be implemented for data collection, use and reuse, consent, storage, protection, retention, destruction, and how the professional users will be consented.
  - Measures to ensure continued compliance with national and EU legislation on data protection

- Dual use
  - Assessment, documentation and mitigation of mission/function creep risks

- Other ethical issues
  - Unexpected or controversial uses
  - Data management

These issues are addressed using the information from the datAcron Description of Action (DoA) and brief interactions with some of the project partners. In the cases where the information provides has been enough to establish procedural guidelines, these are provided below. In other cases, the information available at the moment is not sufficient to provide detailed instructions –in those cases, a general framework is sketched and it will be developed in follow-up reports later on in the project. Finally, there may be legal and ethical matters that are only identified as the project takes shape. Taking this into account, this Ethics Management Plan will run throughout the 36 months of the project and address those as they arise, including any new materials in the follow-up reports.

2. HUMANS

2.1. Procedures and criteria that will be used to identify and recruit experiment participants

As with all other research projects, which involve direct participation of human subjects, strong ethical and privacy management procedures must be put in place to ensure that the power differential between researcher and subject is not exploited in any way. To this end, stringent efforts need to be taken to guarantee the research of the project is not only ethical, but respects the privacy, dignity and welfare of research participants.

This chapter identifies specific data management and ethical issues that, if breached, may impact the privacy, dignity and welfare of research participants in the project.

2.1.1. Data Management Considerations

This section sets the limitations around the capture and use of research participants' data that partners agree to comply with.

- Data Collection & Data Minimisation

In gathering data, the project will record information that is necessary to addressing the central purpose of the research; that is, no unnecessary data will be gathered. The identities of those participating in the research will be protected at all times.

- Data protection

Any sharing of data across the consortium will be based on a) necessity and b) anonymity of data. Where partners require access to data to enable a synthesis of findings across studies, this will be provided in anonymised form only.

- Data Storage

All data will be stored on secured, password protected computers. Sensitive data will not be stored on unencrypted flash drives. In addition, the project will maintain back-ups of this data on a secure data service provided by the partners.

- Data Retention

Raw data from the experiments will only be retained for the lifetime of the research project, unless explicit permission is requested and given by the research participants for an extension period (which may necessitate an appropriate consent form amendment).

- Burden/Benefit

Research participants will have awareness and knowledge of the consequences of their agreement to participate. This will be achieved with the project informing potential participants of the likely benefits and burdens to their engagement with the project.

2.1.2. Experiments in datAcron

The research results (computational methods developed) and the integrated datAcron prototype will be validated and evaluated. According to the DoA,

> *Evaluation will be done in several scenarios such as evaluation of visual data analysis and reasoning, evaluation of user experience, evaluation of user performance, etc. Each scenario will define the exact goals and outputs and specify evaluation questions to be answered and evaluation methods to be used. Methods to be used incorporate state of the art behavioral analyses methods.*

> *The methods will be evaluated using state-of-the-art scenarios, in close collaboration with domain experts: Evaluation will be done in several scenarios, such as evaluation of visual data analysis and reasoning, evaluation of user experience, evaluation of user performance, evaluation of environment etc.*
>
> *Task 4.5 Evaluating VA methods in several scenarios and workflows*
> *VA methods developed in tasks T4.1-T4.4 will be evaluated in corresponding usage scenarios* ***with appropriate categories of professional users****. A part of evaluation will be done with computer science specialists focused on data processing and analysis, another part will be performed with domain experts from two use cases, aviation and sea traffic.*
> *The evaluation will address objective user performance (how fast/correct are users in problem solving) and subjective user satisfaction. For selected critical tasks, eye-tracking evaluation of problem solving activity will be performed for identifying common mistakes and suboptimal problem solving strategies. The results of this task will be reported in deliverable D4.9.[bold emphasis added]*

According to project participants, the professional users to take part in the validation and evaluation experiments will be adult, willing employees that will receive no specific compensation for taking part in these exercises. Moreover, personal data will not be used during the experiments, as the physical traits will not be linked to specific persons.

Users taking part in the validation and evaluation experiments will be professionals employed by datAcron beneficiaries, by organizations that are members of ATM and Maritime Use Case Interest Groups formed and expanded during the duration of the project, or adult students from the Ecole Nationale Supérieure de la Marine Marchande.

As part of the Ethics management Plan, project partners are requested to provide a grid to describe the characteristics of the expected validation and evaluation procedures. Example:

WP/ Task	Name	Description	Humans involved? (Number)	Personal data collected?	Personal data accessed?	Info sheets and Consent forms
WP4 /T4.5	VA Tests	Users of a visualization will be subjected to eye-tracking in controlled conditions. Objectives are to evaluate how visualizations are read on a basic functional level (fixations) and thus potentially improve the functional effectiveness of these visualizations as interactive tools and information source	X	no	no	yes
…	…	…	…	…	…	

Action point: Partners need to fill in the above grid to describe the characteristics of the experiments, validation and evaluation procedures.

2.2. Informed consent procedures

Informed consent is a key procedure for ensuring an ethical approach during research activities that involve investigation with humans. It also enables the legal-based participation of volunteers and the provision of information to the participants in order to guarantee their full awareness of the benefits, costs and risks that may involve the research.

The main features of the informed consent are:[5]

- **Disclosure**: subjects have adequate comprehension of the information provided.
- **Capacity**: subjects are able to understand the information provided as well as the potential consequences of their decision.
- **Voluntariness**: subjects are not participating under any kind of external pressure such as coercion, manipulation, or undue influence.

Subjects must be informed, as far as possible, about the details of their participation -actual and potential consequences of their contribution to the research (benefits, costs and risks). For this reason, the provided documentation should include the following elements:[6]

- A statement indicating that the study involves research, detailing the purposes of the research and the expected duration of the subject's participation, as well as a description of the procedures to be followed, and identification of any experimental products/procedures.
- Foreseeable risks or discomforts to the subject.
- Benefits to the subject or to others which may reasonably be expected from the research.
- A statement indicating the confidential nature of the records identifying the subject and the possibility that external agencies may inspect the records.
- Contact details for answers to pertinent questions about the research and research subjects' rights.
- A statement that participation is voluntary, that refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and that the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled.

Details of the procedure:

- Participants must read the information sheet and sign the consent form (provided in the Annex) **before** the experiments start.
- All signed Informed Consent Forms must be retained as part of the Project Documentation for inspection or ethical/societal audits.

2.2.1. Guidelines for project partners.

Step 1: Assess Your Potential Involvement

Identify whether your organization will conduct an experiment involving human participants at any time during the project. This can be done by viewing the table provided above, consulting the WP leader or coordinator, or seeking clarification as to your potential involvement from the project's ethical advisor.

[5] Faden, R. R.; Beauchamp, T. L. (1986). *A History and Theory of Informed Consent*. New York: Oxford University Press.

[6] UCI Office of Research (Undated).*Required Elements Of Informed Consent*. Available at: http://www.research.uci.edu/compliance/human-research-protections/irb-members/required-elements-of-informed-consent.html

Step 2: Provide Appropriate Documentation to Relevant Authorities

Modify the information sheet and consent form templates (provided in the Annex) for applicability and send for approval to the authorizing body where the activity will take place. The authorizing body will most probably be the Data Protection Authority for industry partners, and for universities this may be their internal ethical review board. However, it is the partners' responsibility to verify this information and identify the relevant authority and procedure at the national or regional level. For your convenience, but only as a reference, a list of national Data Protection Authorities in datAcron countries, complete with addresses, emails and phone numbers, is included below; please choose the one that is relevant for your organization. In the Annex, a sample letter to the DPAs in included.

French Data Protection Authority
Commission Nationale de l'Informatique et des Libertés
8, Rue Vivienne
CS 30223
75083 Paris cedex 02
Tel. +33 01 53 73 22 22
Fax +33 01 53 73 22 00
Electronic office: www.cnil.fr/cnil-direct

Greek Data Protection Authority
Hellenic Data Protection Authority
Kifissias 1-3
115 23 Athens
Tel. +30 210 6475600 or +30 210 6475696
Fax +30 210 6475628
E-mail: contact@dpa.gr

German Data Protection Authority
Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit
Husarenstraße 30
53117 Bonn
Tel. +49 228 997799 0 or +49 228 81995 0
Fax +49 228 997799 550 or +49 228 81995 550
E-mail: poststelle@bfdi.bund.de

Spanish Data Protection Authority
Agencia Española de Protección de Datos
C/ Jorge Juan, 6
28001 Madrid (Spain)
Tel: +34 901 100 099 or +34 91.266.35.17
Electronic office: https://sedeagpd.gob.es/sede-electronica-web/

Action point: Partners need to gain ethical clearance from the relevant authority in their country.

The content in the **information sheet** is designed to:

i. Explain the project & the activity
ii. Ensure that anyone who agrees to participate knows what they are doing
iii. Ensure they are informed about what will happen to any information they give you, including data about themselves

iv. Offer all options of anonymity about themselves and their data

v. Ensure they know their participation is completely voluntary and that they are free to leave at anytime, with no recrimination or penalty.

While the **consent form** is designed to act as a written record that demonstrates each participant understood the **information sheet** and consents to participate on the basis of the terms and information set out therein.

Included in the Annex is a letter template to send to the relevant authority requesting approval for data management processes and ethical research commitments.

Step 3: Request for Participation

Assuming permission is granted by the relevant authority – ethical review board for universities or data protection authority for industry partners– to proceed with the activity, the final step is to invite prospective participants, and ensure all documents are signed ahead and after the research.

Step 4: Retain Signed Information Sheets and Consent Forms

All signed informed consent forms must be securely retained as part of the project documentation and can be requested for inspection in any future ethical or societal audit.

In the Annex you will find:

- A template information sheet for datAcron
- A template consent form
- A template letter for DPAs

Action point: Partners need to adapt the information and consent forms and provide them to experiment participants, keeping the original signed forms in a secure location for ethical auditing.

3. DATA PROTECTION

3.1. Potential privacy risks related to noise reduction processes

Data protection and privacy risks in the context of the datAcron project depend on the specific features of the developed system. Privacy risks related to noise reduction, in this case, are low or inexistent, as noise relates to stream imperfections (i.e., the noise inherent in vessel positions due to sea drift, delayed arrival of messages, or discrepancies in GPS signals). However, even though the privacy risks associated with datAcron are low due to the scarce amount of personal data involved, a thorough assessment is still necessary. Privacy concerns like unintended revelations or access to personal data shall be understood in terms of probabilities, which means measuring the likeability of an undesired event and weighing the consequences (e.g. the potential identification of vessels/aircrafts that should remain anonymous and the potential outcomes of this situation).

In order to evaluate the privacy risks, there are three factors that should be taken into account in the context of this project:

- Type of moving entities covered
- Provenance and type of data exploited
- Actors and end-users involved

3.1.1. Type of moving entities involved.

A key factor to determine the potential privacy risks and issues depends directly on the kind of moving entities that are covered in the resulting system. It is important to clearly define the kind of aircrafts and vessels to be monitored. If only “very large fleets of moving entities” are being included in the scope of datAcron, the risks decrease. Nevertheless, it is still necessary to identify the features and functions of those aircrafts and vessels, as well as the potential consequences of a future modification of the moving entities covered (this is specifically covered in the section on function and mission creep, see below).

For instance, the consideration within the project of unmanned vessels and aircrafts (p. 10 DoA) is still unclear. The geolocalization of such entities could lead to the approximate localization in time and space (within a certain radius) of their owner or data that could lead to re-identification. During the design of the system, it is necessary to consider the potential (intended or unintended) identification of the moving entities:

- If the identification of a certain model of vessel or aircraft could provide information on its use beyond what the owner/s disclosed;
- If the combination of model and geolocalization data could facilitate the identification of a specific model or type of vessel or aircraft –information that, in its turn, could reveal information on use or other types of information that one may want to keep private;
- If determining the company in charge of a transportation would facilitate the identification of flights.[7]

Risks associated to potential identification and tracking (including past, present and future routes) are especially important in the case of private vessels/aircrafts or sensitive cases:

- Military and defense concerns. It may not be desirable or legal for entities linked to these fields to be monitored.

[7] The *Rendition Project*, for instance, managed to identify the ‘shell companies’ (entities created solely to have official ownership of certain aircraft) associated with rendition flights. http://www.therenditionproject.org.uk/flights/companies/index.html

- Members of the Government, dignitaries and/or public figures. The detailed tracking of moving entities linked to this type of actors could compromise their security. Private anchoring areas may need to be discarded.
- Route pattern inference. The possibility to infer common paths and travel routines through machine learning techniques and archival data.

These risks need to be taken into account, and specific protocols will need to be drawn during the project.

A remarking issue for the question of the type of moving entities covered is the scope of the AIS system (*Automatic Identification System*). For the specific case of vessels, it is relevant to point out that according to the Directive 2002/59/EC (art. 2)[8], traffic monitoring and information systems shall be established for a wide range of vessels. The scope of the Directive includes ships of 300 GT and upwards, with certain exceptions:

> *1. This Directive applies to ships of 300 gross tonnage and upwards,*
> *unless stated otherwise.*
> *2. Unless otherwise provided, this Directive shall not apply*
> *To:*
>
> > *(a) warships, naval auxiliaries and other ships owned or operated by a Member State and used for non-commercial public service;*
> > *(b) fishing vessels, traditional ships and recreational craft with a length of less than 45 metres;*
> > *(c) bunkers on ships below 1 000 gross tonnage and ships' stores and equipment for use on board all ships*

Nevertheless, this Directive has been amended, and a wider range of fishing vessels (15 meters and over) are be included.[9] datAcron protocols need to cover these issues.

3.1.2. Provenance and type of data

Another crucial factor to determine the risks related to privacy and personal data is the type of data processed and its provenance. Since big data exploitation implies the processing of several data sources, it is necessary to evaluate the potential consequences of its use. Regarding this question, several concerns have been identified:

- The existence of unique identifiers for aircrafts and vessels, and the potential development of 'reverse search' mechanisms
- The consequences of 'in-situ' data processing
- The use of open data and the possibilities of reutilization (both for the retrieved data as for the resulting datasets)
- Exploitation of messaging systems (e.g. CPDL**)**: content, meta-data, etc.
- Data revealed by monitoring and information systems like AIS and others.

A possible solution to ensure good practices is the development of a specific protocol for the selection of data sources. This would include a comprehensive list of all the data sources, and their assessment (if they include or allow the inference of personal data, among other relevant aspects). It would also be interesting to consider the possibility of monitoring the activities of vessels and aircrafts through anonymized codes.

[8] DIRECTIVE 2002/59/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 June 2002 establishing a Community vessel traffic monitoring and information system and repealing Council Directive 93/75/EEC http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:2002L0059:20110316:EN:PDF

[9] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32009L0017

As a preliminary assessment, datAcron's DoA has been reviewed to look for information on the types of data that will be used, and the instances in which monitoring and protocols will be needed. Two types of data sources have been identified:

- Streaming data sources
- Archival data sources

And two domains:

- Aviation
    - *"multiple, heterogeneous and disperse datasets are expected to be used in aviation use case validation"*
    - *"archival and real-time data available on the ground, including surveillance tracks, flight plans (updated at real-time), weather,* ***DataLink messages*** *and other contextual information such as airspace procedural models"* [bold emphasis added]

- Maritime
    - *"multiple, heterogeneous and disperse datasets are expected to be used in maritime use case validation"*
    - *"Although AIS data are only legally required for larger vessels, their use is growing (e.g., their use has been recently extended to smaller fishing vessels according to the European Commission regulations in the Mediterranean Sea)"*
    - *"disparate data sources (AIS data from coastal and satellite receivers, weather data, geographical data, environmental data, archival data concerning vessels routes and trajectories)"*

> *datAcron will address requirements from the ATM and maritime domains by developing advanced tools for detecting and visualizing threats and abnormal activity over heterogeneous, voluminous, fluctuating, and noisy data streams from thousands of moving entities in large geographic areas, correlating them with data expressing* ***entities' characteristics,*** *geographical information, weather data, patterns of mobility in specific areas, regulations, intentional data (e.g. planned routes) etc, in a timely manner. Towards these goals datAcron will deliver cross-streaming and archival data management solutions, advancing the processing of* ***data close to the data sources*** *-following the in-situ data processing paradigm - producing streaming data synopses at a high-rate of compression, without compromising detection and forecasting accuracy.* [bold emphasis added]
>
> *(...) In both domains,* ***there will be different scenarios of data growth****, also defining cases where data sources provide data at different levels of quality and veracity;* ***partners will specify the exact data sources to be used****,* ***will determine data sources,*** *while defining the detailed validation and evaluation methodologies per scenario of use, with the expected impact specified for the relevant metrics.* [bold emphasis added]
>
> *For the maritime domain*
> *(a) AIS data from both coastal and satellite receivers: AIS is a messaging system which is mandatory for many types of ships under the jurisdiction of the European Union Member States. datAcron, in addition to data from terrestrial networks of AIS receivers, whose performance is characterized by high persistence, but limited coverage, plans to exploit data from satellite-based systems that can pick up messages in the open sea, far away from the coastline.*

(b) Environmental data concerning Significant Wave Height (SWH).
(c) Contextual Information: Maritime regulations, protected areas, closed areas, anchoring areas, traffic separation schemes.
(d) Weather data.
(e) Archival data concerning vessels and ports (at least).

For the ATM domain
(a) Flight Plan related data: including flight plan creation, update and deletion messages, coming from airline dispatchers and Air Traffic controllers. This dataset contains relevant information like call sign, airline, DEP, DEST, planned route, ETO times, and equipment. As the Flight Plan is updated during flight, all this data is being continuously updated according to the real flight behavior and ATC events.
(b) Surveillance information: Aircraft position related data, is currently being tracked at high frequency by different types of individual sensors like Primary Radars, Secondary mode-s radars, multilateration radars, Surface Movement Radars and ADS-B/C systems. Individual sensor data and fused data will be exploited by datAcron.
(c) Weather data: Multiple sources provide weather data to Air traffic systems like satellite, met radar and the aircraft itself.
(d) Datalink information: Datalink capabilities allow the aircraft to downlink and uplink data with air traffic controllers and airlines.
(e) AirSpace information: This kind of information is highly dynamic in Air Traffic. So Routes, procedures, flow information and NOTAM messages have to be considered.
(f) ATC data: Air traffic Controller Clearances.
(g) ATFM data: Regulations, STAM measures and other information affecting ETOT/CTOT or other parameters of flights coming from the Network Managers (or local FMPs).

Crucial spatiotemporal measurements, like heading, speed, travel time, area of coverage, etc., can be calculated per trajectory at several time horizons and resolutions.

Many of these data are either from 3rd parties (provided under specific agreement or being freely available) or are provided by datAcron partners. Many of these data are either commercial (e.g. data from IMISG) or they are associated with intellectual property rights that do not allow sharing. Specifically, as far as the maritime domain is concerned, the intentions and performance of ship owners operations can represent the competitive edge of the company thus these data can be very sensitive. There are also service providers whose business models build on providing data in a condensed and quality way. Nevertheless, datAcron will develop a Data Management Plan that will refer to these issues in detail and will seek ways to make a large portion of the data available with respect to partners' agreements, without breaking any IPR or privacy laws.
Data shared will be provided as linked open data in RDF via specific endpoints, or be made downloadable, taking advantage of the developments in WP1.

datAcron, has scheduled specific tasks for data preparation, curation and preservation in WP1 (task 1.3.3), WP5 (task 5.2) and WP6 (task 6.2).

Reference datasets in datAcron (also relevant for licensing issues):

- Flight Plan related data: including flight plan creation, update and deletion messages, coming from airline dispatchers and Air Traffic controllers. This dataset contains relevant information like call sign, airline, DEP, DEST, planned route, ETO times, and equipment. As the Flight Plan is updated during flight, all this data is being continuously being updated according to the real flight behavior and ATC events.
- Surveillance information: Aircraft position related data, is currently being high frequency tracked by different types of individual sensors like Primary Radars, Secondary mode-s radars, multilateration radars, Surface Movement Radars and

ADS-B/C systems. All this data is fused and correlated with Flight plan information in ATM Systems. Individual sensor data and fused data will be input of the system.

- Weather Data: Multiple sources provide weather data to Air traffic systems like satellite, met radar and the aircraft itself. Weather forecast analysis can lead to predict turbulence, icing, convection, winds and temperatures on planned route.
- Datalink Information: Most of modern Aircraft are equipped with datalink capabilities that allow the aircraft downlink and uplink data information with air traffic controllers and airlines.
- AirSpace information: this kind of information is strongly dynamic in Air Traffic so Routes, procedures, flow information and NOTAM messages have to be considered.
- ATC data: Air traffic Controller Clearances.
- ATFM data: Regulations, STAM measures and other information affecting ETOT/CTOT or other parameters of flights coming from the Network Managers (or local FMPs).

"at a variable refresh rate, which depends on their motion (vessels at anchor transmit their position every two minutes and increase the broadcast rate up to two seconds when maneuvering or sailing at high speed; every five minutes, vessels transmit other data (static and voyage related information) containing identifiers, such as International Maritime Organization (IMO) number, call sign, ***ship name*** *and Maritime Mobile Service Identity (MMSI), used as a primary key to link the message to position information. Static information also includes size, type of vessel and cargo, whereas voyage related data, such as Estimated Time of Arrival (ETA) and destination, are manually set and not fully reliable"* [bold emphasis added]

- *Timestamp (reception time of the AIS frame in Unix epoch)*
- *MMSI_Number (the unique vessel ID)*
- *Longitude and Latitude (in WGS84 format)*
- *Speed over Ground (SOG)*
- *Course over Ground (COG)*
- *Ship Code (according to AIS specification; an 'ad hoc' routine has been developed to compute exact ship type from ship code)*

During the life-time of datAcron, the use of the data will be monitored, and any specific requirements or findings will be included in the final ethical report. In this initial assessment, partners are asked to provide information on the following in relation to data types and provenance:

- In page 15 of the DoA it is stated that data analysis methodologies "need to be validated operationally within security and defense constraints"→ what are these constraints?
- In page 213, "data mining techniques" are mentioned → what are these techniques?
- On page 73 "broadcast information on their location (positional, identification and other information)" is mentioned – what is meant by "other information"?

Action point: Partners need to clarify above questions and determine whether sensitive information can be derived from the data managed in datAcron.

3.1.3. Actors/End-users

Another factor that needs to be taken into account is the range of actors and end-users involved in the project. This is important due to the differential access to the information generated by the project. It is important to foresee who will have access to the resulting information and under which conditions and protocols -i.e. who will be able to monitor the moving entities, what technical and managerial capabilities will they have, and which is the responsibility chain in case of an event. It is also indispensable to describe the interfaces operated by the corresponding users as well as the utilization contexts. Depending on the characteristics of the collected, stored and shared data, confidentiality concerns could also arise. For this reason, once the data sources and resulting datasets are defined, it is necessary to define who will be able to access them and how will this information be recorded (logs to monitor the activities in the system). Potential issues related to the generation and sharing of open data (and thus, the open access to the information) are considered below.

3.2. The integration of privacy by design within the process

'Privacy by Design' (PbD) is a concept developed back in the 90's by Ann Cavoukian, Information & Privacy Commissioner for Ontario (Canada). It aims at establishing privacy assurance as the default mode of operation of an organization. Principles of Privacy by Design may be applied to all types of personal information, but should be applied with special attention to sensitive data. The objectives of Privacy by Design are ensuring privacy to achieve control over one's information, and gaining a sustainable competitive advantage for organizations.

The Privacy by Design approach is based on 7 foundational principles, which are:[10]

1. Proactive not Reactive; Preventative not Remedial
The Privacy by Design (PbD) approach is characterized by proactive rather than reactive measures. It anticipates and prevents privacy invasive events before they happen. PbD does not wait for privacy risks to materialize, nor does it offer remedies for resolving privacy infractions once they have occurred — it aims to prevent them from occurring. In short, Privacy by Design comes before-the-fact, not after.

2. Privacy as the Default Setting
We can all be certain of one thing — the default rules! Privacy by Design seeks to deliver the maximum degree of privacy by ensuring that personal data are automatically protected in any given IT system or business practice. If an individual does nothing, their privacy still remains intact. No action is required on the part of the individual to protect their privacy — it is built into the system, by default.

3. Privacy Embedded into Design
Privacy by Design is embedded into the design and architecture of IT systems and business practices. It is not bolted on as an add-on, after the fact. The result is that privacy becomes an essential component of the core functionality being delivered. Privacy is integral to the system, without diminishing functionality.

4. Full Functionality — Positive-Sum, not Zero-Sum
Privacy by Design seeks to accommodate all legitimate interests and objectives in a positive-sum "win-win" manner, not through a dated, zero-sum approach, where unnecessary trade-offs are made. Privacy by Design avoids the pretense of false dichotomies, such as privacy vs. security, demonstrating that it is possible to have both.

5. End-to-End Security — Full Lifecycle Protection
Privacy by Design, having been embedded into the system prior to the first element of information being collected, extends securely throughout the entire lifecycle of the data involved — strong security measures are essential to privacy, from start to finish. This ensures that all data are securely retained, and then securely destroyed at the end of the process, in a timely fashion. Thus, Privacy by Design ensures cradle to grave, secure lifecycle management of information, end-to-end.

6. Visibility and Transparency — Keep it Open
Privacy by Design seeks to assure all stakeholders that whatever the business practice or technology involved, it is in fact, operating according to the stated promises and objectives, subject to

[10] Cavoukian, A. (2009). *Privacy by Design. The 7 Foundational Principles*. Information and Privacy Commissioner of Ontario (Revised Jan 2011). Available at: https://www.ipc.on.ca/images/Resources/7foundationalprinciples.pdf

independent verification. Its component parts and operations remain visible and transparent, to users and providers alike. Remember, trust but verify.

7. Respect for User Privacy — Keep it User-Centric
Above all, Privacy by Design requires architects and operators to keep the interests of the individual uppermost by offering such measures as strong privacy defaults, appropriate notice, and empowering user-friendly options. Keep it user-centric.

Privacy by Design includes both operational and organizational issues. Since the use of personal data in the datAcron project is minimal, and it is difficult to identify any privacy threats if all consent mechanism and protocols are put in place, a specific plan to implement PbD is not needed. Nevertheless, the following table can help partners identify the relevant operational issues highlighted in this dcument.

Application of Privacy by Design for operational issues:

Principle	**Application to the datAcron project**
Proactive not Reactive	• Consider the risks of the processed data by listing all the expected data sources and its characteristics • Consider the actors/end-users involved
Privacy as the Default Setting	• Consider the possibility of not being tracked if it is not compulsory by the law.
Privacy Embedded into Design	• Assess the privacy issues in various steps
Full Functionality	• Consider privacy as important as other factors
End-to-End Security	• A main concern regarding this question is that the procedures aimed at removing the collected data are still not clear, due to the archival data and machine learning needs. This should be addressed.
Visibility and Transparency	• Allow the inspection and verification of the resulting system and evaluation procedures
Respect for User Privacy	• Inform the affected users about the monitoring activities

Application of Privacy by Design principles for organizational issues is especially relevant in case of events or unexpected situations. In those cases, the pertinent staff responsible for the system should have protocols or handbooks that ensure good practice.

3.3. Identification and documentation of the licensing options

It is increasingly common for research funders to require that the data produced during a research be made available as broadly as possible, to foster further research and maximize the impact of the findings. Therefore, creating licensing protocols that include non-pay walled, non-confidential outputs are encouraged as part of responsible and ethical research and innovation. This, of course, requires for project partners to identify which parts of their research may include sensitive, confidential or personal information that needs to be kept confidential for legal or corporate purposes.

This Ethics Management Plan suggests datAcron uses a protocol to determine the license of each of its products that takes into account and renders explicit:

- The need to establish that all products should a priori be published using an open copyright license (such as creative commons) unless there are specific reasons to not do so.
- An assessment of the industry standard –how similar data is used by other actors or repositories, to make sure that the information released or secured by datAcron adheres to common practice.

There are currently several projects and commercial products aimed at making available datasets reflecting the activities of aircrafts and vessels. These projects give an interesting overview of data sharing procedures, as well as the related questions to be considered (privacy policies, licensing, provenance of datasets, etc.). These are some of the relevant initiatives that could be used as benchmarking for datAcron:

- **OpenFlights.org**: This website shares with the public airport, airline and route data since 2009. The databases are made available under the *Open Database License*[11] and any rights in individual contents of the database are licensed under the *Database Contents License*[12]. http://openflights.org/data.html

- **Federal Aviation Administration:** US' FAA analyses and shares data regarding airport operations and provides, among others, a database called *Air Traffic Activity Data System* that gathers the official NAS air traffic operations data available for public release. https://www.faa.gov/data_research/

- **FlightRadar24:** Flightradar24 is a flight tracking service that provides users with real-time info about thousands of aircraft around the world. Under the terms and conditions section, the group refers to the "non-transferable right to access and use the Services". This right is granted for personal, non-commercial use only. There are also references to frame the data collection and transmission procedures and the possible unauthorized use. https://www.flightradar24.com/data/

- **Flight Data Community:** This community, fostered by Flight Data Services (FDS) offers the POLARIS Suite, a set of tools for flight data analysis that facilitates data sharing in order to increase safety. The suite includes several tools for flight data exploitation -a converter, a plotter, an analyzer, a format transfer, a parameter tree, and data libraries. http://www.flightdatacommunity.com/

- **Internet Ships Register (ISR):** A database developed by IHS that includes information on commercial ships over 299 GT and their owners, operators, managers and builders.http://www.ihsfairplay.com/

- **Sea-web:** Sea-web combines comprehensive ships, owners, shipbuilders, fixtures, casualties, port state control, ISM and real-time ship movements data and ports information into a single application. Sea-web is an enhanced service that includes all information from the Internet Ships Register but provides many additional benefits. It covers over 180,000 ships down to 100 GT. http://www.ships-register.com/

[11] http://opendatacommons.org/licenses/odbl/1.0/

[12] Ibid.

- **FleetMon:** This database allows searches by name, IMO or MMSI numbers, flag state, length and vessel type; it also allows tracking specific vessels to follow their activity. The website grants a limited, revocable, non-exclusive, non-transferable and non-sublicensable license for beneficial use of the content accessible on the site. https://www.fleetmon.com/vessels/

- **Marine Traffic:** It is a vessel tracking service that covers monthly up to 800 million vessel positions and 18 million vessel and port related events; it also provides details of over 650 thousand marine assets available (vessels, ports, lights). Regarding utilization rights, it states that 'the User shall use the Information and Data for his/her own internal use only and shall have no other rights with respect to the Data, including without limitation, any right otherwise to use, distribute, furnish or resell the Data or any portion or derivative thereof'. http://www.marinetraffic.com

There are several standard licensing options that allow enough flexibility and control to be used by a project like datAcron:[13]

- **Creative Commons:** Creative Commons is a non-profit corporation set up in 2001 for the purpose of producing simple yet robust licenses for creative works. These licenses give the creators of such works finer-grained control over how they may be used than simply declaring them public domain or reserving all rights. As well as the legal text, the licenses all have quick clear summaries and a canonical URL for use in HTML, RDF and other code. A rights expression language is also provided for use with RDF. While originally aimed at works such as music, images and video, Creative Commons licenses have been used widely for most forms of original content, including data.

  There are six main Creative Commons licenses. While the spirit behind them has remained constant, the wording of their legal deeds has been revised over time, resulting in different *versions*, and adapted to different legal jurisdictions, resulting in different *ports*.
  Each license includes the *Attribution* condition. In the version 3 licenses and earlier, it is left up to the licensor to specify the way in which credit is given. Recognising the difficulties this may cause in the context of attribution stacking, the version 4 licenses can be satisfied by a link to a Web page containing attribution information, though licensors can specify additional, alternative mechanisms.

  There are three other conditions that licensors can add, and the various possible combinations produce the six licenses. Using just the attribution condition is known as the CC BY license.

  There is a *Non-Commercial* condition, where commercial is defined as 'primarily intended for or directed toward commercial advantage or monetary compensation'.

  The *Share Alike* condition inserts a strong copyleft clause into the license. The version 1 licenses are very strict: derivations may only use the exact same version 1 license. The version 2 licenses onwards, however, allow derivations to use a later version or a different port of the same license. Nevertheless, derivations may not use a Creative Commons license with a different set of conditions.

  Finally, including the *No Derivatives* condition in the version 3 licenses and earlier means that the licensee is forbidden from altering, transforming or building upon the work. The version 4 condition is more flexible: it allows these things for private use, but prevents the licensee from sharing the derivations. It and the Share Alike condition are mutually exclusive.

[13] Ball, A. (2014). 'How to License Research Data'. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Available online at: http://www.dcc.ac.uk/resources/how-guides/license-research-data.

The versions of the licenses prior to version 4 were not specifically aimed at data, so using them for such presents some problems. The most significant is that they do not explicitly cover *sui generis* database rights such as the one in force in the European Union. This means, for example, that use of substantial portions of a database licensed using the unported terms of version 3 or earlier may constitute a rights infringement in such jurisdictions. The version 4 licenses, however, do explicitly include *sui generis* database rights unless the licensor specifically reserves them.

All versions of the licenses treat datasets and databases as a whole: they do not treat the individual data themselves differently from the collection/database. This might be considered an advantage in terms of simplicity, but means they cannot be used without difficulty in certain complex cases such as collections of variously copyrighted works.

Similarly, the licenses do not distinguish using data as part of a new collection/database from using them to generate content (graphs, models, maps, etc.). This means the Share Alike and No Derivatives conditions might have further reaching consequences than intended. Indeed, the No Derivatives condition would likely disallow most substantive types of reuse, leaving only such cases as checking that data within the set derive from each other as claimed. It should therefore be avoided.

- **Open Data Commons:** The Open Data Commons Project was set up in 2007 to develop a successor to the Talis Community License (TCL). The first license to be produced was a public domain dedication for databases. The project transferred to the Open Knowledge Foundation in 2009 and has produced two further licenses having some of the character of the Creative Commons licenses, but designed specifically for databases. All three follow the Creative Commons model of providing a clear summary and canonical URL alongside the full legal text.

  The Open Data Commons Attribution License (ODC-By) allows licensees to copy, distribute and use the database, to produce works from it and to modify, transform and build upon it for any purpose. If content is generated from the data, that content should include or accompany a notice explaining that the database was used in its creation.[40] If the database is used substantially to create a new database or collection of databases, the licence URL or text and copyright/database right notices must be distributed with the new database or collection.

  The Open Data Commons Open Database Licence (ODC-ODbL) is the same as ODC-By but for a couple of additional conditions. It adds a copyleft condition that applies to new databases derived from the database (but not collections of databases or non-database content produced directly from it); this condition would be satisfied by future versions of the same license or a compatible one as judged by the licensor. The other condition is that technological restrictions such as Digital Rights Management (DRM) mechanisms can only be applied to the database or a new database derived from it if an alternative copy without the restrictions is made equally available.

  Being written in database terms, these licenses are suited to a wider range of research data than the Creative Commons equivalents. The ODC-ODbL copyleft condition is also slightly more flexible than Creative Commons' Share Alike, though the ODC attribution requirement is slightly less flexible.

Both these licenses suit the datAcron project and the demands of the funder. Taking into account the licensing issues for a project like datAcron implies considering a wide range of questions. Among the concerns detected, we highlight the following:

- *Licensing of the source data*. Once the provenance of the integrated data is detailed, it is necessary to ensure that all utilization permissions are granted (including all types of data, like ship pictures). Mapping systems may be based on different alternatives like OpenStreetMap (OSM) or Google Maps. Product and company names may be the trademarks of their respective owners. Open data sources also have conditions and limitations for their use, which depend on the type of license upon which they were published.
- *Licensing of the outcoming data*. The integrated data should be subject to clear licensing options. It is necessary to foresee the use scenarios and the scope of end-users. Storage, property rights, reuse, reutilization and commercialization.
- *Licensing of the resulting tools*. The resulting tools (interfaces, analysis software, etc.) will be subject to specific distribution, utilization and, in certain cases, modification rights.
- *Transparency and open data issues*. In case of publicly accessible data, transparency regulations are subject to limitations related to privacy and security issues.
- *Protocols.* It is recommendable to designate a contact person and define protocols in case of misuse or conflict.

A useful solution to keep control of the licensing issues is to use the same table where source data and output data are listed, and adding a column for the licensing concerns as they arise during the project. This will be reflected in the follow-up of this Ethics Management Plan.

Action point: Partners need to establish the licensing protocol for datAcron.

3.4. Detailed information on the procedures that will be implemented for data collection, use and reuse, consent, storage, protection, retention and destruction

A priori, datAcron will not gather personal data outside of the experiments described above, in which information and consent procedures need to be implemented. Detailed information on data collection, use and reuse, consent, storage, protection, retention and destruction is therefore not necessary outside of the scope of such experiments.

Nonetheless, and as part of the Ethics Management Plan, data management procedures will be monitored in project deliverables and any instance of use that could lead to reidentification or raise ethical concerns will be highlighted and addressed accordingly, especially during the validation and evaluation experiments.

Moreover, pieces of data that a priori are not considered personal data, but which could lead to re-identification (unique identifiers of any kind, or geolicalisation) will be assessed from an ethical and data protection perspective.

Finally, as mentioned above, as part of Privacy by Design and basic responsible innovation principles, End-to-End Security will be enforced in datAcron. This includes intra-project communication and the storage of data, where encryption will be the norm in order to ensure:

- Authentication: The origin of the message/information/data set can be verified.
- Integrity: Proof that the contents of a message/information/data set have not been changed.
- Non-repudiation: The sender/creator of a message/information/data set cannot deny their role.

The communication and storage systems used in datAcron will be audited from a security perspective and detailed in the Ethics Management Plan follow-up, including:

- E-mail services
- Cloud services
- Server location and encryption
- Pen drives and other peripherals
- Data Bases
- Intranet

Throughout the project, data over-collection and usage outside of the potential perimeter of the planned work will be prevented.

Action point: Partners need to provide information of their data management plans as soon as they take shape.

3.5. Measures to ensure continued compliance with national and EU legislation on data protection.

As mentioned in the DoA,

> *The datAcron laboratory set-ups will run in servers hosted in Greece (both use cases), Spain (ATM use case) and France (Maritime use case) and will not be subject to hosting in third countries. During integration and preparation phase, UPRC, NARI and BRTE will consult once again with the national agencies for data and privacy protection to ensure the compliance.*

The current data protection regulation at the EU level is the Directive 95/46/EC (*Data Protection Directive*). Nevertheless, there is currently a changing regulatory scenario and from 2017 on, all European organizations shall be compliant with the General Data Protection Regulation (GDPR). As part of this Ethics Management Plan, the evolving regulatory context will be monitored and any necessary updates will be incorporated if the new GDPR comes into effect before the end of the research.

A key question to stress is that the data protection regulations only apply to the data of natural persons, not to organization data. According to the Data Protection Directive, ‘personal data’ shall mean,

> *“any information relating to an identified or identifiable natural person (“data subject”); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity”.*[14]

When personal data is involved, data protection principles apply. There is a set of data protection principles which are transversal to any regulation and are formulated as good practices regarding personal data which serve also as obligations of the data controller:

- **Fairness:** Data must be fairly and lawfully processed.
- **Finality:** Data must be processed for limited purposes.

[14] Applicability: Opinion 4/2007 on the concept of personal data http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf

- **Data minimization:** Data collection, use and retention must be adequate, relevant and not excessive.
- **Data quality:** Data collection, use and retention must be accurate.
- **Conservation:** Data must not be kept longer than necessary.
- **Confidentiality:** Data must be processed in accordance with the data subject's rights (detailed below). They shall be communicated to the data subject together with the corresponding data processing details.
- **Security:** Data must be proportionately secure and must not be transferred to countries without adequate consent and protection.
- **Notification to the Supervisory Authority:** The data controller must notify the supervisory authority the details of existing personal data filing systems.

If an organization is processing personal data, it will have to observe the data subject's rights[15] and allocate the necessary resources to ensure that the data subject is able to exert them appropriately. The data subject's rights are the following:

- **Information on collection and awareness:** The data subject should be given notice of the data controller's information practices before any personal information is collected from him. Without notice, an individual cannot make an informed decision as to whether and to what extend his personal information is disclosed.
- **Choice and consent:** The data subject has the right to choose how any personal information collected for him may be used. This choice relates to the secondary uses of the information as well.
- **Access, correction and deletion:** The data subject has the right to challenge the accuracy of the data and to provide corrected information. The access process should be timely, inexpensive, simple, providing a mechanism for verification of the data and a means by which corrections and objections may be recorded. The data subject can also ask for the erasure or blocking of the data.
- **Integrity and security:** The data subject has the right to know the extent to which the data will be secured. To ensure data integrity, collectors must take reasonable steps by using reputable sources and cross-checking data against multiple sources, providing individuals access to data, and destroying access data. Security of the data would include both the management of and the technical measures to protect against loss, unauthorised access, use, and disclosure of the data.
- **Enforcement and redress:** The data subject has the right to seek legal relief to protect his privacy rights".

[15] Hugelier, S. Janssen, K. and Dumortier, J. (July 2013). *Legal and IPR Management Framework Specification*. (Draft v1.0). Retrieved from http://opensciencelink.eu/wp-content/uploads/2013/06/OSL_D3_2_LegalAndIPRManagementFrameworkSpecification.pdf

4. DUAL USE

4.1. Assessment, documentation and mitigation of mission/function creep risks

The concept of 'function creep' refers to the "gradual widening of the use of a technology or system beyond the purpose for which it was originally intended, especially when this leads to potential invasion of privacy."[16] Different aspects of a project or system can vary with time - functions, scope and objectives, geographical reach, end-users, etc. This may have ethical implications or cause unforeseen or unintended privacy harms, or even constitute criminal offences. Therefore, while not each and every future practice or use can be anticipated, it is important to assess what the possibilities for function or mission creep for any given project may be.

The steps to be taken in this direction include:

1. Describing the initial definition of the expected application of datAcron:

    - Geographical areas/countries included
        - Areas covered
        - Open sea / coastline (indicate potential function differences)
    - Entities and end-users that are expected to have access to the information and make use of the system
    - Utilization guidelines and protocols (especially in case of abnormal events).

2. Considering the later incorporation of data sources through the scalability options.

    - Broadening of the scope of the system
    - Moving entities involved
    - Utilization possibilities (e.g. for law enforcement purposes)

3. Updating and reviewing the system and its function/mission creep risks

4. Avoiding the use of the resulting systems and tools for other contexts/other moving entities (e.g. vehicle traffic management)

5. Identifying unexpected actors that may make use of the technologies developed (e.g. criminals attempting to act in open seas)

Function/mission creep is particularly relevant when the technologies or procedures being developed can be used in different contexts. In the context of datAcron, the need to specify users and actors is a clear priority, as the contexts of utilization of datAcron-related products include:

- Military/Security
- ATM and maritime management
    - Transportation of passengers
    - Transportation of goods
    - Private use
- Government - Management of infrastructures
- Industrial/commercial

[16] Collins English Dictionary - Complete & Unabridged 2012 Digital Edition http://www.dictionary.com/browse/function-creep

Action point: Partners need to specify the mission limits of datAcron.

5. Other ethical issues

Other ethical issues may be identified during the project. These will be related mainly to unexpected or controversial uses and data management.

5.1. Unexpected or controversial uses

In order to establish the potential ethical issues that could arise while implementing the tools and systems resulting from the datAcron project, it is useful to take into account past events regarding the ATM and maritime domains that could help to foresee potential risks. These events and conflicts shed light on potential situations that may cause ethical issues:

- Management of 'abnormal' events and emergency protocols: adequate and responsible interpretation of deviations and uncommon patterns.
- The implementation of anti-terrorist protocols and alert systems (for instance, the correct differentiation of cases).
- Differences on information access (sources, real-time and archival data).
- Legal considerations of the processed information: who and under which circumstances is legally entitled to access the processed information (e.g. for judicial investigations).
- The actual utilization (legal and legitimate) of the information processed. Including the future development of technologies (like remote control systems) and information resources (like PNR databases).

During such events, datAcron could play a key role in what concerns the timely reactions and decisions taken. Some of these cases affect more than one country, as well as both public and private actors. Therefore, different accesses to information could alter the consequences of an event and imply different power positions (e.g. during the Prestige Oil Spill, all the affected countries wanted to send away the vessel from their coastline). Moreover, the outcomes of datAcron could also make a difference while clarifying legal responsibilities. What follows are examples picked up in order to provide partner with an idea of what happens when things go wrong or there is abnormal activity. While the datAcron tool would only be one amongst the many involved in such events and accidents, taking such examples into account can increase the resilience of the datAcron solution.

5.1.1. Rendition Project

The Rendition Project is a collaborative research initiative run by Prof Ruth Blakeley at the University of Kent and Dr Sam Raphael at the University of Westminster. The main goal of the project is to "investigate and understand the use of rendition, secret detention and torture by the CIA and its allies"[17]. For this reason, the project provides a public database of rendition flights by CIA aircraft and reveals a complex web of companies that helped operate a network of rendition aircraft and secret prisons. This kind of projects could suppose a source of conflict between secrecy and transparency regarding air traffic data, since it could compromise the security of specific operations. Intelligence and law enforcement agencies would probably encourage avoiding activities that involve the tracking of those flights, aircrafts and companies. It would be therefore necessary to take into account the potential role of datAcron both as an information source and as an information node and archive.

5.1.2. Accidents

In 2002, a single-hulled oil tanker, the *MV Prestige*, faced technical problems during a storm off Galicia (Spain). The captain called for help but the French, Spanish and Portuguese governments refused to allow the Prestige to dock in their ports. Finally, the vessel split in half on six days later and

[17] Rendition Project website: http://www.therenditionproject.org.uk

76,000 m^3 of oil were spilled. The environmental devastation has been compared to that of the *Exxon Valdez Oil Spill*.

The Costa Concordia disaster is another recent episode regarding the maritime domain. In 2012, this cruise ship capsized and sank after striking an underwater rock obstruction off Isola del Giglio (Italy). Thirty-two people lost their lives as consequence of the disaster. The sentence against the captain included ten years for multiple manslaughter, five years for causing the shipwreck, one year for abandoning the passengers, and one month for providing false information to port authorities.

The IOM estimates that up to 700 migrants died in the Mediterranean in 2013; 3,072 in 2014 and 3,692 in 2015. According to this organization, since 2000 more than 22,000 migrants have lost their lives trying to reach Europe[18]. This so-called migrant and refugee crisis is connected to several issues implying political intervention, humanitarian action, coastguard strategies and cross-national management of critical events and conflicts that may occur both at open sea and at coastline (drowns, boats capsizes, and even violent episodes).

The Germanwings Flight 9525 crashed on March 2015 in the French Alps after a constant descent that began one minute after the last routine contact with air traffic control. There were no survivors. The co-pilot Andreas Lubitz had previously been treated for suicidal tendencies and been declared "unfit to work" by a doctor. During the flight, he locked the pilot out of the cockpit before initiating deliberately a descent that caused the crash.

5.1.3. Conflicts

Other events worth mentioning are related to conflicts and attacks to aircrafts and vessels. The environmental NGO Greenpeace suffered the sinking of the *Rainbow Warrior* in 1985 by the action of the French DGSE ("Opération Satanique"); in 2008 the *Artic Sunrise* was surrounded by Turkish tuna fishing vessels and attacked by the crew; in 2014, Spanish Navy rigid-hulled inflatable boats repeatedly slammed into Greenpeace protesters off the Canary Islands. In 2010, the raid against the 'Gaza Freedom Flotilla' targeted a total of 6 civilian ships carrying humanitarian aid and construction materials in international waters in the Mediterranean Sea. Moreover, numerous episodes of piracy have been reported in the Horn of Africa, jeopardizing the fishing vessels that operate in that area. Regarding the air traffic domain, it is also possible to list several dramatic events involving both the actual hijack of flights and false alarms.

5.2. Data management

5.2.1. Subjective thresholds

Data management concerns can go beyond data protection to include broader ethical matters. datAcron includes among its aims the detection of 'threatening' or 'abnormal' activity, and 'important' events. There are culturally-charged definitions that require that specific protocols to determine what is 'threatening', 'abnormal' and 'important' are defined to ensure the consistency of the decisions made by the relevant actors or automatic processes. This will be ensured and covered in the follow-up of this Ethics Management Plan.

5.2.2. Algorithms and automatic decision-making

datAcron will use algorithms to make decisions. In most cases, and according to the DoA, these algorithms will only make simple decisions (identify speed gating and displacement of vessels, for instance). However, if at any point algorithms are expected to make complex decisions (beyond database matching, for instance), the potential societal impact and risk of these will need to be assessed and established. This will be ensured and covered in the follow-up Ethics Management Plan.

[18] http://www.theguardian.com/world/2015/apr/14/400-drowned-libya-italy-migrant-boat-capsizes

5.2.3. Factuality of data

Data can be contradictory or imprecise. Making decisions on the basis of non-factual data can have important consequences for a project of technology and its decision-making processes. The steps taken to minimize the use of such data will be specified in the Ethics Management Plan follow-up.

5.2.4. Undesirable reuse and threat modeling

Once data is created, it cannot be protected against all risks, as the possibility of a threat gaining access to it is never zero. Therefore, it is important to define who or what would constitute the threat model of datAcron –could other entities want unwarranted access to datAcron data? Could datAcron data benefit any category of entities (insurance companies, for instance). Would the data developed by datAcron have any commercial value that would make it attractive to third parties? All partners should compile a list of threats and design their systems against them.

5.2.5. Technological divide and discrimination

The use of technology is not evenly distributed geographically or socially. Therefore, any data-based initiative needs to take steps to ensure that it does not reproduce or reinforce existing discrimination. In the case of datAcron, for instance, only 60% of all passenger aircraft around the world are equipped with an Automatic Dependent Surveillance – Broadcast (ADS-B) transponders. The project will need to establish whether those entities using less or different technologies will be left out of the benefits datAcron envisages, whether this is due to geographical differences that can limit the scope of the solutions developed, if this will translate into different access to security for different actors, and to what extent this can be minimized.

ANNEX

Experiment Information Sheet

datAcron is a 36-month EU-funded project looking to improve operations concerning moving entities in the Air-Traffic Management and Maritime domains.

The datAcron project requires that those who, like you, participate in the experiments give explicit consent to do so and are aware of what this means, and entails, as well as to the right to withdrawal.

Please take time to read and understand the following, and if you agree with the content sign the consent form provided together with this Information Sheet. You can keep this form but will need to leave the signed form with the person responsible for the experiment and who made contact with you. In the consent form, you sign that you freely and voluntarily consent to be an experiment participant in the datAcron project to be conducted at ______. If, at any point, you feel unable or unwilling to continue, you are free to leave without negative consequences. That is, your participation in this experiment is completely voluntary, you may withdraw from this project at any time and you will not receive any gratification for taking part in it.

The datAcron project assures you that any data or information you provide will be kept strictly confidential. In gathering our data, we will only record information that is necessary to address the central purpose of our research, and ensure it is anonymised. This information will be securely stored and retained for the lifetime of the project and finally deleted.

Furthermore, your name will not be linked with the research materials, as the researchers are interested in the content in general, and not in any individual values, choices or traits. Therefore, there are no risks to your participation, and no potentially harmful procedures will be conducted.

The specific procedures that will be performed during the experiments will consist of ____.

The burdens of participating in this research are your opportunity cost, the risk of entrusting your data in the hands of others, and the potential harm for misuse of those sensitive, identifiable data. The reassurances around strict data governance, given by the datAcron team, are designed to alleviate potential participation burdens.

On the other hand, the benefits of participating in this research reside in the opportunity to be involved in a significant piece of research, and the role you play in improving the efficiency of Management and Maritime operations via the validation/evaluation of computational methods for moving entities (increasing the safety, efficiency and economy of operations concerning moving entities in the Air-Traffic Management and Maritime domains).

By signing the consent form you declare that you have been informed that if you have any questions seeking further clarification or assurances about the ethical issues relating to the project, you are free to contact datAcron Project Coordinator George Vouros at georgev@unipi.gr, affiliated with University of Piraeus Research Center.

Informed Consent Form

I …………………………………… agree to participate in this datAcron experiments.

The purpose of the experiments has been explained to me in writing (in the information sheet).

I am participating voluntarily and understand that I can withdraw from the experiment without repercussions, at any time, before it starts or while I am participating.

I am satisfied that the assurances of responsible and strict data governance, given by the datAcron project, will be upheld.

I understand that anonymity, by disguising my identity, will be ensured at each research stage in the project.

A copy of the information sheet has been given to me.

Signed……………………………………. Date……………….

Data Protection Authority Letter Template

Dear DPA,

I write to you on behalf of a European research project, datAcron, seeking privacy & ethical approval for research to be hosted in [___].

datAcron is a 36-month EU-funded project looking to improve operations concerning moving entities in the Air-Traffic Management and Maritime domains.

An important part of the project's work will involve experiments with human participants in order to ____.
datAcron's independent ethics expert has drafted an Ethics Management Plan including template information sheet and consent forms (see attached), which, we believe, comply with the requirements for appropriate ethical behaviour, data protection & privacy within research.

The project respectfully requests your approval to undertake, with due regard to legal and regulatory compliance requirements, this research.

I attach the process that will be followed by the project as it undertakes this research; it includes an information sheet and informed consent form, which outline the processes & procedures the project will follow to ensure privacy and data protection of individual participants.

Yours Sincerely,

Name:
Position:
Institution:

Attached:

(1) Information sheet;
(2) Informed consent form;