Grant Agreement No.: 731677
Call: H2020-ICT-2016-2017
Topic: ICT-13-2016
Type of action: RIA

D7.4: Data Management Action Plan v2

Stephen C. Phillips (ITINNOV) | 6/3/2019

FLAME is operating a complex multi-stakeholder system with data, including personal data, being collected or generated in many components and needing to be shared to understand the system as a whole. This document explores the issues around managing these datasets including the legal and ethical operating framework, the various different stakeholder types and the contractual arrangements between them, licensing of data, data repositories and the expected data sets. Finally, a data management template is provided for project trials to promote both good data management and the opening of datasets.

Acknowledgements

The author would like to thank the following individuals for their expertise and assistance throughout all aspects of the analysis and for their help in writing the report.

- Bee Brown (Contracts Officer, Research and Innovation Services, University of Southampton)
- Professor Michael Boniface (Head of the IT Innovation Centre, University of Southampton and FLAME Project Coordinator)
- Dr Brian Pickering (Chair of Faculty of Engineering and Physical Sciences Ethics Committee, University of Southampton)
- Dr Alison Knight (Head of Research Integrity and Governance, University of Southampton)

Work package	WP7
Task	Task 7.1
Due date	31/10/2018
Submission date	06/03/2019
Deliverable lead	IT Innovation
Version	2.0 (second version of the data management plan, first version of "D7.4")
Authors	Stephen C. Phillips (IT Innovation)
Reviewers	Dirk Trossen (IDE)

DISCLAIMER

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731677.

This document reflects only the authors' views and the Commission is not responsible for any use that may be made of the information it contains.

Project co-funded by the European Commission in the H2020 Programme		
Nature of the deliverable:	**Report**	
Dissemination Level		
PU	Public, fully open, e.g. web	✓
CL	Classified, information as referred to in Commission Decision 2001/844/EC	
CO	Confidential to FLAME project and Commission Services	

Co-funded by the Horizon 2020 Framework Programme of the European Union

EXECUTIVE SUMMARY

As part of the Open research data pilot, the FLAME project is obliged to open and share datasets where possible and this document provides information on data licenses, recommends a data repository (Zenodo) and provides a data management plan template for open call trials to complete. However, as the FLAME project will be collecting personal data, the majority of this document deals with the legal and operational issues surrounding the management and sharing of data within the project consortium, regardless of whether the data is suitable for publication.

To achieve operational efficiency improvements and increase the quality of experience of users, FLAME is developing a platform which is able to combine data collected across architectural layers from the applications on user devices, through media services, platform services and right down into the network infrastructure and use this data to better manage the services provided to the end users. Personal data collected from both the application layer and network layer informs the management of the platform and as it is collecting personal data the project must be very aware of its obligations around information security and the rights of the user.

The major piece of legislation dealing with the collection and processing of personal data is the General Data Protection Regulation (GDPR) which came into force in May 2018. For a pan-European project such as FLAME, the GDPR is a benefit as it unifies the data protection law across countries in the European Community. The project does include partners based in Switzerland (outside of the EU) and in the UK (which is likely to leave the EU during the project) but both of these jurisdictions intend to follow the GDPR as well.

FLAME collects personal data, including some location data and some photographic media. FLAME will not be doing so, but these types of data may be used to infer sensitive personal data (such as religion) and therefore must be treated with the utmost caution. The analysis shows that FLAME should be seeking informed consent of all users of the system, both for legal and ethical reasons. FLAME will only keep personal data when necessary and where possible pseudonymise it as soon as practicable. The GDPR imposes various requirements on the design and implementation of the FLAME platform including that a "data protection by design and default" approach is taken as well as providing the means to delete any personal data which is stored upon the request of the data subject.

This is the second version of the FLAME Data Management Plan. The first version of this document was well received and has its content has not been disputed so in order to provide a single reference for data management, the first version has been added to (forming this document) and so the majority of the text is unchanged. The additions are: an updated ethics section describing the process now in place; some brief remarks on the consequences of Brexit; a commentary on the actual data sharing agreements being used by the replicas; an analysis of the situation where the infrastructure operator is not the same as the platform operator; an analysis of funded and unfunded replicator contracts an updated data management plan template to explicitly capture personal data for use in data sharing agreements.

Co-funded by the Horizon 2020 Framework Programme of the European Union

TABLE OF CONTENTS

Co-funded by the Horizon 2020 Framework Programme of the European Union

LIST OF FIGURES

Co-funded by the Horizon 2020 Framework Programme of the European Union

LIST OF TABLES

Co-funded by the Horizon 2020
Framework Programme of the European Union

ABBREVIATIONS

TCP	Transmission Control Protocol
3PPM	3rd Party Project Manager
B2B	Business to Business
B2C	Business to Customer
CA	Consortium Agreement
CC	Creative Commons
CLMC	Cross-Layer Management and Control
CSV	Comma-Separated Variable
DoA	Description of Action
DOI	Digital Object Identifier
DPA	Data Protection Agency
DPO	Data Protection Officer
EaaS	Experimentation as a Service
EC	European Commission
EM	Ethics Manager
EMB	Ethics Management Board
EU	European Union
FAIR	Findable, Accessible, Interoperable and Reuseable
FQDN	Full-Qualified Domain Name
GA	Grant Agreement
GDPR	General Data Protection Regulation
HTTP	Hypertext Transfer Protocol
ICO	Information Commissioner’s Office
IP	Intellectual Property
IP	Internet Protocol

Co-funded by the Horizon 2020 Framework Programme of the European Union

IPR	Intellectual Property Rights
MAC	Media Access Control
OC	Open Call
ODC	Open Data Commons
ORD	Open Research Data
PC	Project Coordinator
PM	Project Manager
PMB	Project Management Board
SF	Service Function
TLS	Transport Layer Security

Co-funded by the Horizon 2020
Framework Programme of the European Union

1 INTRODUCTION

The principles of the EC's Open Research Data pilot are to be "As open as possible, as closed as necessary". This is and related principles are expanded on in Section 3. The ORD pilot is primarily concerned with opening projects' research data to the wider world beyond each project consortium with a focus on data which underpins published research findings and/or has longer-term value. In FLAME we have a complex multi-stakeholder system with data being generated in many components and needing to be shared to understand the system as a whole. Some stakeholders are part of the FLAME consortium, others will be collaborating with the consortium through cascade funding to run trials on the FLAME platform and others will be trial participants whose personal data will be collected and processed (see Section 2). Therefore there are data sharing and management issues to be addressed even in the narrower context of the consortium and these third parties. This data management plan puts in place the processes to meet both the ORD objective and the particular FLAME issues.

The data management plan must take into account the operating framework, expanded on in Section 4, including intellectual property rights (IPR), the General Data Protection Regulation (GDPR), ethics, the Grant Agreement and Consortium Agreement. The GDPR in particular places strict constraints on the handling of personal data and defines the rights of the data subjects. The consortium must understand which datasets contain personal data and must have the necessary processes in place to deal with them carefully and in compliance with the law.

The right to share a dataset belongs (sometimes jointly) to the dataset owner. The owner may choose what terms and conditions the recipient(s) of a shared dataset must comply with and these terms and conditions are generally described using licences. Depending on how the dataset is created the owner may be constrained by a conditions (such as a licence) themselves. A discussion of common licences for open data may be found in Section 5.

To preserve data sets and make them readily available a managed data repository should be used. A discussion of data repository requirements and a review of existing repositories can be found in Section 6.

FLAME D3.3 (Architecture) makes some key points about the sharing and use of data in FLAME. For instance, Section 2.1 (FMI Ecosystem) states that (with added emphasis):

> "Irrespective of the delivery model, our approach needs to support the creation of knowledge about the structure and behaviour of interactive media systems. **The multi-stakeholder nature of the system is our priority** and therefore the knowledge created must play a key role in supporting the development of possible B2B and B2C relationship between stakeholders. To be as disruptive as possible and to avoid being constrained by current business models, **we are proposing that knowledge is created in an open way with the goal of driving best practice and standardisation efforts for ecosystem structures**. Policies related to access and rights to open knowledge will be fundamental to the development of the ecosystems."

In a multi-stakeholder system such as FLAME we must address who owns or otherwise has rights over what data, as it is the owner of the data (and potentially other joint owners or rights holders) who has some IPR in the data and must decide how to manage the data. We must agree rules for data sharing that retain the data owners' rights whilst providing for the necessary data sharing to meet the stakeholders' and project's objectives. The stakeholders and their motivations are discussed in Section

Co-funded by the Horizon 2020 Framework Programme of the European Union

7 along with the various contracts governing the various relationships and an analysis of the data controller and data processor roles.

A discussion of the primary data sets in FLAME, the stakeholders with an interest in the data and the manner and purpose for which they may be shared may be found in Section 8.

In common with the Grant Agreement, we take the view that the owner of data is the party or parties who generate the data. Different trials may have different requirements for data sharing (depending on the IPR involved). Therefore we will treat each trial separately, though within a common framework. A trial's participants are the recipient of trial cascade funding (or one of ETH, DRZ, NXW and VRT in the validation trials), the core consortium members providing the platform services and infrastructure that the trial uses (potentially a cascade funded replicator). As the objective of the project is to better understand these multi-stakeholder FMI systems we propose that, *at a minimum*, all the trial's participants have access to all the trial's non-personal data for the purpose of better understanding the operation of the platform and its components. The principles guiding data management in FLAME and a variety of FLAME scenarios are looked at in Section 9.

The issues of what is necessary for data management in the Cascading Funding Agreement are looked at in Sections 10. As noted above, each trial will be treated separately in terms of data management. To streamline the process and provide clarity, a menu of proposed data management plan templates, with most parts already completed, is provided in Section 11.

Finally, it is worth noting that this document relates to data management in the FLAME project. For any subsequent commercial version of the FLAME platform a similar analysis would be required and data would potentially need to be managed differently.

Co-funded by the Horizon 2020
Framework Programme of the European Union

2 FLAME DATA OVERVIEW

As described in D3.3 FLAME Platform Architecture and Infrastructure Specification [D3.3], the FLAME platform architecture comprises:

- Media Components deployed on the Platform, combining to create Media Services offered to Users;
- FLAME Platform Services providing the environment for Media Components to execute and the Cross-Layer Management and Control (CLMC) system; and
- the Infrastructure Slice, part of a larger managed infrastructure, hosting the FLAME Platform Services and Media Components.

An overview of the architecture is shown below in Figure 1.

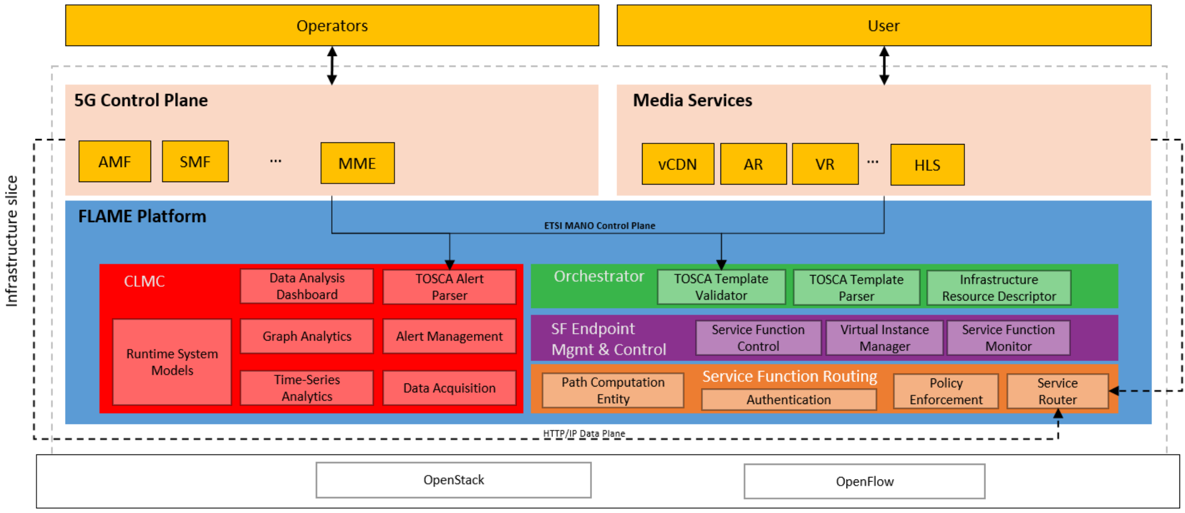


Figure 1. FLAME Platform Architecture.

The Platform itself is a multi-stakeholder concern. Conceptually:

- an infrastructure operator provides the infrastructure slice;
- a platform operator deploys the platform into the slice;
- media service providers create media services by combining media components (their own and those of others) and deploy them on the platform;
- trial participants use the media services via mobile apps (for instance).

Data, including personal data such as users' locations and activities, is gathered at all levels of this research infrastructure. Personal data is gathered by the applications (primarily mobile) used by the participants and the directly connected media components. Personal data (in the form of network identifiers) is also collected by the infrastructure slice for the operation of the infrastructure. Some of this data, such as location data, is potentially sensitive personal data. The project will use the data to understand the relationships between the demand from the users (in terms of their activity and

Co-funded by the Horizon 2020 Framework Programme of the European Union

expected quality of experience) and the load on the platform and infrastructure and the necessary quality of service and cost. The FLAME platform content, services and networking are highly configurable and the project will use the gathered data to optimise across the users' quality of experience and the arrangement of content, services and networking.

More succinctly, the global project purpose for processing personal data could be expressed as

> "To understand the relationships between participants' experiences and the platform's configuration and performance and to optimise the same".

The project consortium includes partners providing two locations for user trials (Bristol and Barcelona) as well as four partners leading initial trials. As the project progresses, three rounds of open call will bring in additional trial leaders and additional cities using cascade funding. Whilst the original consortium are signatories to the Grant Agreement and a Consortium Agreement, these open call partners do not accede to the Grant Agreement and Consortium Agreement and instead will be signatories to a Cascade Funding Agreement made with ITINNOV (as project coordinator and cascade funding controller).

Co-funded by the Horizon 2020
Framework Programme of the European Union

3 PRINCIPLES

3.1 OPEN RESEARCH DATA

As part of the Pilot on Open Research Data (ORD) in H2020, the FLAME project partners will endeavour to open the research data to the public where appropriate, by depositing it in a research data repository where it will be findable and accessible for others. As noted on the OpenAIRE website [OpenAIRE] "You do not need to deposit all the data you generate during the project either – only that which underpins published research findings and/or has longer-term value".

In the words of Jean-François Dechamp [Dechamp 2016]:

> Wider access to scientific facts and knowledge helps researchers, innovators and the public find and re-use data, and check research results.
>
> - Offers better value for research funds: a public benefit.
> - Encourages research across scientific fields: essential for solving today's complex societal challenges.
>
> Horizon 2020 grantees are required to:
>
> - deposit underlying research data and other research data of their choice in a repository; and
> - take measures to grant open access to this research data.

The key guiding phrase for ORD is that data should be "as open as possible, as closed as necessary". Grantees have the right to opt-out at any time but the opt-out must be justified, for instance if opening data might:

1. infringe on intellectual property rights;
2. infringe on someone's privacy; or
3. jeopardise the project's main objective.

All three of these opt-outs could be relevant in FLAME. IPR is discussed in the sub-section below, personal privacy is strictly controlled through law (see Section 4.2) and objective 6 of the FLAME DoA is "EaaS sustainability and business models that combine public-private investment in localised infrastructures with an exploitable software platform" and describes "formulating business models as well as exemplary business plans that are deemed suitable for a long-term sustainability with a drive to commercialization of both underlying infrastructure and the capability to provide large-scale EaaS offerings". It may be that the objective of commercialising the underlying infrastructure could be jeopardised by making public some parts of the project's research data.

Co-funded by the Horizon 2020 Framework Programme of the European Union

3.2 FAIR DATA

The ORD Pilot encourages grantees to share data using the FAIR principles: Findable, Accessible, Interoperable and Reusable. The explanation of FAIR below is copied from an article in Nature [Wilkinson 2016] under the CC-BY-4.0 licence.

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
  - A1.1 the protocol is open, free, and universally implementable
  - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
  - R1.1. (meta)data are released with a clear and accessible data usage license
  - R1.2. (meta)data are associated with detailed provenance
  - R1.3. (meta)data meet domain-relevant community standards

Co-funded by the Horizon 2020 Framework Programme of the European Union

3.3 OPEN DATA

With the ORD specifying the data be as open as possible, the meaning of "open" needs defining. The Open Definition [Open] states:

> "Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)."

We take this definition of "open" as the meaning of "as open as possible".

3.4 PUBLICATIONS

As stated in the project's Grant Agreement (Article 29.2), FLAME will provide open access to scientific publications. Two main routes for open access exist:

- **Gold**: is where an article is immediately provided in open access mode by the scientific publisher (also known as "open access publishing").

- **Green**: is where the published article or final peer-reviewed manuscript is archived by the researcher in an online repository before, after or alongside its publication. An embargo period may be used to provide exclusive access via the scientific publisher for a period. This is also known as "self-archiving".

FLAME targets "Gold" open access for scientific publications. Wherever "Gold" is not possible, "Green" open access will be pursued. The target is to maximize the impact on scientific excellence through result publication in open access yet highly appreciated journals. Furthermore, repositories for enabling "Green" open access to all project publications will be used, e.g. ITINNOV's institutional repository [ePrints] as well as Zenodo / OpenAIRE, which provide the means to promote and realise the widespread adoption of the Open Access Policy, as set out by the ERC Scientific Council Guidelines for Open Access and the EC.

3.5 IPR PROTECTION

FLAME is working in an area where significant knowledge, IP and innovation, is expected to be generated and the ORD principles absolutely permit keeping data closed to protect IPR and to protect the objectives of the project. The Consortium Agreement (CA) specifies the conditions under which either the consortium as a whole or its individual partners will share the rights for exploitation after the completion of the project. For the duration of the project these issues are governed by the CA under the auspices of the Project Coordinator and the Project Management Board (PMB).

Special attention has been given in knowledge management and protection issues from the beginning of the project. In particular, the following commercial issues are addressed in the CA:

i. confidentiality concerning the information disclosed by the parties during the project development;

ii. ownership of results resulting from the execution of the project;

iii. legal protection of results through patent rights;

Co-funded by the Horizon 2020
Framework Programme of the European Union

iv. commercial utilisation of results, also taking into account joint ownership of the results;

v. patents, know-how and information related to the use of knowledge owned by one of the parties, resulting from work carried out prior to the agreement; and

vi. sublicenses to third parties within clearly defined limits.

The PMB will ensure that IP rights are identified so that opportunities for gaining and exploiting patents are monitored, and to ensure that essential project know-how is protected. A standing item will be taken at all project meetings to discuss IPR expectations, and thereby ensure that necessary steps are taken to protect this intellectual property e.g., by the timely filing of patent applications. Ownership of IPR is defined in the CA, which will also govern the granting of licenses to FLAME consortium members and third party organisations wishing to benefit from FLAME IPR. A distinction between licenses for research purposes and licenses for exploitation is made. Protection for partners' background knowledge is also provided by use of confidentiality clauses in the CA. In addition, partners' relevant background IPR has been declared at the contract stage, and licenses for its free usage within FLAME has been arranged. Partners are obliged to license on fair and reasonable terms any background, which they own and which is necessary for exploitation of the foreground IPR developed in FLAME.

Co-funded by the Horizon 2020 Framework Programme of the European Union

4 OPERATING FRAMEWORK

4.1 INTELLECTUAL PROPERTY RIGHTS

4.1.1 Copyright

Copyright refers to the "right" to "copy" a "work". If you do not hold copyright then you cannot copy (or publish) a work unless you have permission from the copyright holder via a licence which permits copying.

In UK law [Copyright] for instance, copyright applies to literary, dramatic and musical works, artistic works, sound recordings and films, and more. Data comes under the category of "literary works" and therefore copyright law applies here.

In countries that are signatories to the Berne Convention (which is most), you do not actually have to assert or register copyright: you have the right automatically. To be clear, just because a work does not actually say "© Copyright XYZ" does not mean that you are allowed to copy it. Even so, many organisations do explicitly assert copyright because:

- it makes it clear and avoids any doubt;
- it lets readers know who to talk to if they want to ask to copy the work; and
- not having a copyright statement may reduce damages awarded in an infringement lawsuit as the defence of "innocent infringement" may be less successful.

Copyright must be asserted to belong to one or more legal entities, for instance one of the FLAME partners. As noted in FLAME D7.2 (Risk Identification and Management & Quality plan v1), a statement such as "© Copyright FLAME Consortium" does not make sense as the FLAME consortium is not a legal entity.

Copyright does expire after a number of years. How many years depends on the legal jurisdiction and on the type of work. Generally the length of copyright is at least 50 years beyond the time of the author's death.

Commentary

In FLAME, media content used in trials is likely to be copyrighted and trial leaders will have to ensure that they have the necessary rights from the copyright holder to use the content in their trials. New media content works (such as videos and photos) are likely to be created by participants in trials and each participant author will hold the copyright of their works. In the case of a participant's works being uploaded to a FLAME media service, for the purposes of copyright law the participant will need to agree to the trial leader being permitted to make copies of the work for the purpose of providing the service. Of course, in the case of photographs or video, it is quite possible that the data is personal data and therefore compliance with data protection legislation (see Section 4.2) would also be required.

Data files created during a trial (such as service log files) also attract copyright and sharing of such files between parties requires a licence or similar agreement.

Co-funded by the Horizon 2020 Framework Programme of the European Union

4.1.1.1 Derivative works

In copyright law, a derivative work is a separate creation which includes a substantial amount from a different copyright-protected piece of work. The derivative work must contain enough creative input to be protected in its own right. For instance, translations, screenplays and musical "cover versions" are all derivative works. A work which simply cites or links to another work is not generally counted as a derivative work.

Commentary

In the context of FLAME, derived works are likely to arise in the production of media content used in trials. As noted above, rights on the original and rights on the derived work will need to be managed and respected.

4.1.2 Database Rights

Information structured in a database may acquire copyright and a database right, independently of the rights of the content itself:

- If a substantial intellectual investment has gone into compiling the content in an original manner then a database right can be claimed, protecting the database against unauthorised extraction and content reuse for 15 years (in the EU).

- If the database structure is itself sufficiently complex then it can be categorised as a literary work and therefore attracts normal copyright protection.

Unfortunately, the exact requirements for protection vary across the world and across the EU. Directive 96/9/EC ('on the legal protection of databases') is the relevant EC directive.

Commentary

It may be that a database of sufficient complexity to attract database rights is created in FLAME. The exercising of such rights will be considered on a case by case basis.

4.2 GENERAL DATA PROTECTION REGULATION

One aspect of FLAME is to understand how the experience of the users of FMI systems is affected by the performance of the FMI system and as a result FLAME will be dealing with personal data.

Following Article 16 of the Treaty on the Functioning of the European Union, which is the legal basis for the adoption of data protection rules in the EU, the European Union legislator adopted Regulation 2016/679, the General Data Protection Regulation (GDPR) [GDPR], on 27 April 2016 to protect the fundamental right to data protection and to guarantee the free flow of personal data between Member States. The logic of the adoption of a GDPR was to prevent disparities between Member States in terms of procedures and sanctions, harmonizing the data protection in the EU. It replaces Directive 95/46/EC, which as a Directive, is transposed by each state into local law and therefore has some variation between states.

As the GDPR is a Regulation, it automatically passed into law in each member state around two years after publication. In this case, the GDPR became law on 25 May 2018. As the first FLAME trials begin in

Co-funded by the Horizon 2020 Framework Programme of the European Union

July 2018 the project must follow the GDPR. The UK Information Commissioner's Office has a useful online guide to the GDPR [ICO-GDPR] on which much of the following is based.

The sanctions for breaching the GDPR are significantly higher than under current laws. Fines of up to 4% of annual worldwide turnover or €20m, whichever is greater, may be levied on both controllers and processors.

4.2.1 Key Definitions

The GDPR Article 4 lists some definitions, the most relevant of which are summarised here.

Data subject: an identified or identifiable natural person.

Personal data: any information relating to an identifiable person where the person can be identified by a wide variety of means, including (but not limited to) name, email address, an identification number, location data or even one or more attributes of the person.

Sensitive data[1]: personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person data concerning health or data concerning a natural person's sex life or sexual orientation.

Pseudonymisation: the processing of *personal data* in such a manner that the *personal data* can no longer be attributed to a specific *data subject* without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the *personal data* are not attributed to an identified or identifiable natural person;

Data controller: the person or body which determines the purposes and means of processing *personal data*.

Data processor: the person (other than an employee of the data controller) or body which processes *personal data* on behalf of a *data controller*.

Processing: any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

Commentary

FLAME will be collecting and processing potentially sensitive personal data. The data collected may include location tracking data which may inadvertently reveal strong clues regarding sensitive personal data such as a Data Subject's religion (if they are tracked to a place of worship), health status (if they often visit a hospital), or sexual orientation for instance. As alluded to above, personal data does not just include data which is obviously relating to an identifiable individual (such as a questionnaire response which includes the participant's name). The GDPR recognises that an individual can be

[1] Known as "special categories of personal data" in GDPR Article 9.

Co-funded by the Horizon 2020 Framework Programme of the European Union

identified by more elaborate means: for instance if a questionnaire asked for a respondent's year of birth and employer then in many cases (by combining with other data sets) individuals could be identified and so the data is still "personal data" and subject to the GDPR.

Further information regarding the FLAME stakeholders, data controller and data processor can be found in Section 7 and a discussion on the nature of the data gathered in FLAME may be found in Section 8.

4.2.2 Principles

Article 5(1) of the GDPR requires that personal data shall be:

a) processed lawfully, fairly and in a transparent manner in relation to individuals;

b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall not be considered to be incompatible with the initial purposes;

c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed;

d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay;

e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes subject to implementation of the appropriate technical and organisational measures required by the GDPR in order to safeguard the rights and freedoms of individuals; and

f) processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures.

Commentary

Point (b) mentions that "further processing for … research purposes" is permitted in some cases even if the research purposes were not in the original purposes for collecting the data. Research will be the original purpose of collecting data in FLAME, but Article 5(1)(b) allows for further research in the case that some analysis becomes useful though not initially described to the participant.

Article 5(1)(c) enshrines the principle that an indiscriminate "data grab" must not occur: all person data, *and no more*, must be collected for the research purpose.

To comply with Article 5(1)(e) and the individual's rights (below) the project will pseudonymise any personal data, replacing any direct identifiers such as names or email addresses with pseudonyms as early as possible in the data collection process.

Co-funded by the Horizon 2020 Framework Programme of the European Union

The project will use data protection by design and data protection by default principles to create a system compliant with Article 5(1)(f) (see Section 4.2.5.3).

4.2.3 Lawfulness of Processing

A lawful basis must be identified for the processing of personal data. Article 6(1) of the GDPR states that processing for personal data shall be lawful only if and to the extent that at least one of the following applies:

a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;

b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;

c) processing is necessary for compliance with a legal obligation to which the controller is subject;

d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;

e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;

f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

Article 9(2) of the GDPR lists ten lawful bases for processing of sensitive personal data. Only two are of any relevance to FLAME:

a) explicit consent of the data subject, unless reliance on consent is prohibited by EU or Member State law;

e) processing relates to personal data manifestly made public by the data subject.

Commentary

Item (f) above suggests that, at least in some circumstances, research itself is a legitimate reason for processing personal data even in the absence of the data subject's consent. FLAME however has decided to use consent (a) as the lawful basis for data processing, not least because that also covers the processing of sensitive personal data.

4.2.4 Individual Rights

The GDPR Articles 12-23 gives individuals the following eight rights:

1. The right to be informed

   The data subject must be informed of, amongst other things, the identity and contact details of the data controller, the purpose and legal basis for processing the data and their rights.

Co-funded by the Horizon 2020 Framework Programme of the European Union

2. The right of access

   If a data subject requests a copy of the personal data being held on them, it must be provided (generally free of charge).

3. The right to rectification

   Requests from data subjects for their personal data to be rectified if it is incorrect or incomplete must be honoured.

4. The right to erase (otherwise known as "the right to be forgotten")

   In many circumstances, including the withdrawal of consent, an individual has the right to have their personal data erased and further processing prevented.

5. The right to restrict processing

   Individuals have the right to stop their data being processed (without it being deleted). If the data has been disclosed to third parties then they must also be informed so that they also stop processing the data.

6. The right to data portability

   This allows individuals to obtain their personal data in a form that can be reused for their own purposes or in a different service (for instance, financial transaction history held by a bank). This right only applies to personal data an individual has provided to a controller (including tracking data from a device), where the lawful basis for processing is consent (or performance of a contract) and where processing is automatic. Data must be provided in an open machine readable form such as CSV.

7. The right to object

   Individuals may object to processing based on the lawful basis 6(1)(e) or 6(1)(f) (FLAME uses consent, 6(1)(a)); to direct marketing; and to processing for purposes of scientific/historical research and statistics.

8. Rights in relation to automated decision making and profiling

   Individuals have the right not to be subject to an automated decision where it produces a legal or similarly significant effect on the individual. The right does not apply if the decision is based on explicit consent.

Commentary

Item (1) is dealt with in FLAME with by the participant information sheet for a trial and the consent form [D1.1]. FLAME systems must be designed to deal with (2), (3), (4), (5) and (6). Given FLAME's use of explicit consent for data processing, item (7) is not relevant. FLAME will make use of participant profiling but the effect on the individual will not be significant and explicit consent will have been obtained.

Co-funded by the Horizon 2020
Framework Programme of the European Union

4.2.5 Accountability and Governance

4.2.5.1 Contracts

Article 28.3 of the GDPR states that whenever a data controller uses a data processor (or a processor, with the permission of the controller, uses another processor) there must be a written contract in place to govern the processing.

The contract is there to:

- ensure that both parties understand their obligations, responsibilities and liabilities;
- help them to comply with the GDPR;
- help controllers to demonstrate their compliance with the GDPR; and
- potentially increase data subjects' confidence in the handling of their personal data.

There are many elements that the contract must contain and a good summary may be found on the UK ICO website [ICO-GRPR]. Very briefly, the contract covers:

- the subject matter and duration of the processing;
- the nature and purpose of the processing;
- the type of personal data and categories of data subject; and
- the obligations and rights of the controller.

The GDPR allows standard contractual clauses from the EU Commission or a Supervisory Authority (such as the ICO) to be used in contracts between controllers and processors. However, no standard clauses are currently available.

Commentary

A standard contract for the data controller / data processor relationship will be developed by the consortium and used. Further guidance on standard clauses may be forthcoming from the Commission, the ICO or other national supervisory authorities, in which case these standard clauses will be used.

4.2.5.2 Documentation

The GDPR Article 30 states that organisations with more than 250 employees must maintain internal documentation on data processing activities, namely:

1. Each controller and, where applicable, the controller's representative, shall maintain a record of processing activities under its responsibility. That record shall contain all of the following information:

   a) the name and contact details of the controller and, where applicable, the joint controller, the controller's representative and the data protection officer;

   b) the purposes of the processing;

Co-funded by the Horizon 2020 Framework Programme of the European Union

c) a description of the categories of data subjects and of the categories of personal data;

d) the categories of recipients to whom the personal data have been or will be disclosed including recipients in third countries or international organisations;

e) where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation and, in the case of transfers referred to in the second subparagraph of Article 49(1), the documentation of suitable safeguards;

f) where possible, the envisaged time limits for erasure of the different categories of data;

g) where possible, a general description of the technical and organisational security measures referred to in Article 32(1).

2. Each processor and, where applicable, the processor's representative shall maintain a record of all categories of processing activities carried out on behalf of a controller, containing:

a) the name and contact details of the processor or processors and of each controller on behalf of which the processor is acting, and, where applicable, of the controller's or the processor's representative, and the data protection officer;

b) the categories of processing carried out on behalf of each controller;

c) where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation and, in the case of transfers referred to in the second subparagraph of Article 49(1), the documentation of suitable safeguards;

d) where possible, a general description of the technical and organisational security measures referred to in Article 32(1).

Commentary

The University of Southampton and (independently) other FLAME partners are currently in the process of designing the necessary systems to document data processing activities in compliance with the GDPR and in time for the new regulations to come in. Each FLAME project partner handling personal data will use its own compliant in-house process.

4.2.5.3 Data Protection by Design and Default

Recital 78 of the GDPR states that the data controller "should adopt internal policies and implement measures which meet in particular the principles of data protection by design and data protection by default", such as "minimising the processing of personal data" and "pseudonymising personal data as soon as possible". It says that "producers of the products, services and applications should be encouraged to take into account the right to data protection when developing and designing such products, services and applications and, with due regard to the state of the art, to make sure that controllers and processors are able to fulfil their data protection obligations."

Co-funded by the Horizon 2020 Framework Programme of the European Union

Commentary

The design (and default nature) of the FLAME platform's data protection aspects have been described previously in "NEC – Requirement No. 1" Section 2.3 on "Confidentiality" [D1.1] and, more fully, in the "FLAME Platform Architecture and Infrastructure Specification v1" [D3.3] where Section 6.4 in particular describes the FLAME "information security and privacy" architecture.

In some circumstances, replacing a personal identifier such as a name with a pseudonym makes the associated data anonymous. The FLAME platform however is going to consume and process a wide variety of data types (not all known at this point) which will inevitably include photographic media. Such images, and the potential for someone to use multiple identifiers to re-identify a data subject means that the project must take the view that much of the data remains personal even after pseudonymisation. Therefore the project will use linked pseudonymisation where the link between the data subject and the pseudonym(s) is kept securely and separately from the rest of the data so that if (for instance) the data subject wished to exercise their right to have their data removed then their pseudonym(s) can be looked up and the data deleted. Wherever possible, data should be pseudonymised at source, otherwise it will be pseudonymised as early as possible.

Media services are likely to collect personal data, starting with a user's login (by email address for instance) and continuing with the collection of activity data and sometimes location data. Where possible, personal identifiers should be replaced by pseudonyms before storing in the media service. Only aggregate (non-personal) summary data will be reported to the CLMC.

Network services must deal with network identifiers (i.e. IP and MAC addresses) in order to work and these identifiers are regarded as personal identifiers as a user's MAC address does not often change (see Section 9.4.2). If a network service stores data for later analysis then it should pseudonymise the identifiers. Only aggregate (non-personal) summary data will be reported to the CLMC.

A useful review of issues around pseudonymisation and a proposed approach may be found in a paper by Aamot et al [Aamot 2013]. Use of linked pseudonymisation is one aspect of the privacy by design approach of the project. Other aspects include:

- the definition and use of security domains to clearly define responsibilities and to define the border controls on those domains;
- the platform's aim to use OAuth 2.0 for managing authorisations and the potential for use of the User Managed Access specification, building on OAuth 2.0, to provide a way for users to control access to their (distributed) data;
- the use of encryption for data in transit, using Transport Layer Security (TLS);
- the separation and clear identification of personal data to both minimise the risk of any data breach and minimise the management overhead of the personal data store(s); and
- the adherence to the standard practice found in ISO/IEC 27001.

4.2.5.4 Data Protection Impact Assessments

A data protection impact assessment is required by the GDPR when using new technologies where the processing is likely to result in high risk to the rights and freedoms of individuals (GDPR Article 35).

Co-funded by the Horizon 2020 Framework Programme of the European Union

Commentary

Although not strictly required, FLAME has defined a Privacy Impact Assessment in Appendix B of D1.1 that adopts the principle of proportionality via a three-part test:

1. The "suitability" test, which defines whether the measure is reasonably likely to achieve its objectives;
2. The "necessity" test, which evaluates whether there are other less restrictive means capable of producing the same result;
3. The "proportionality" test *stricto sensu*, which consists of a weighing of interests with which the consequences on fundamental rights are assessed against the importance of the objective pursued.

4.2.5.5 Data Protection Officers

Under the GDPR, you must appoint a Data Protection Officer (DPO) if you:

- are a public authority (except for courts acting in their judicial capacity);
- carry out large scale systematic monitoring of individuals (for example, online behaviour tracking); or
- carry out large scale processing of special categories of data or data relating to criminal convictions and offences.

Commentary

FLAME data controllers and ITINNOV (as project Coordinator) will ensure that a DPO has been appointed in their organisation and notify them of the work being carried out in FLAME so that they may supervise and assist with the work.

4.2.6 Data Security and Breaches

The GDPR requires that personal data be processed securely and is protected against unauthorised or unlawful processing and accidental loss, destruction or damage. Appropriate measures (both technical and organisational) should be used to ensure security.

In the case of a breach of security leading to the loss, alteration or unauthorised access to personal data, if the breach is likely to result in a risk to the rights and freedoms of individuals then the relevant supervisory authority must be notified.

Commentary

The FLAME project software will use data protection by design and data protection by default principles. As data gathered in FLAME can only be traced back to an individual through the index held by ITINNOV, most possible data breaches would not result in any risk to the rights and freedoms of any individuals.

Co-funded by the Horizon 2020 Framework Programme of the European Union

4.2.7 Brexit and Non-EEA Countries

Several FLAME partners (ITINNOV, IDE and UNIVBRIS) are based in the UK which has elected to leave the EU on 29th March 2019 (before the end of the FLAME project but after the time of writing). As with Switzerland, organisations in the UK dealing with the data of EU data subjects after Brexit will be required to comply with the GDPR. The UK government has updated the Data Protection Act (now DPA 2018) to bring the GDPR into UK law and Switzerland is expected to update their data protection act in line with the GDPR in early 2019.

Under the GDPR, trial leaders must ensure that they are registered with their national data protection agency (DPA) as holding personal data for research purposes. Trial leaders based in non-EU countries performing trials in EU countries must also register with the DPA in the country of the trial (for instance, a UK partner performing a trial in Spain must register in the UK and in Spain). If the trial leader and the trial location are both outside of the EU then no double registration is required.

Data sharing agreements are required to facilitate data transfers between the trial leader and the platform operator of the trial location.

EEA to UK transfer:

The European Data Protection Board published an “Information note on data transfers under the GDPR in the event of a no-deal Brexit” on 12th February 2019 [EDPB] describing how to manage data transfers from the EEA to the UK in the case of a no-deal Brexit. The note describes options for legally transferring data to the UK, including standard approved data protection clauses. In the case of a no-deal Brexit, the project will follow this guidance if we have a UK trial leader operating in a replica based in the EEA.

UK to EEA transfer:

The UK government published a guidance note on 6th February 2019 [UKBrexit] which states that in the event of a no-deal Brexit, data transfers from the UK to EEA countries can continue as before.

4.2.8 Summary

This document highlights the most important parts of the GDPR for the FLAME project.

- FLAME will be collecting personal data, from which potentially sensitive personal data may on occasion be inferred.
- Personal data will only be collected where it is necessary and can be justified. There will be no indiscriminate data grab.
- The personal data will be pseudonymised as early in the processing as possible (and where possible at source).
- Each trial will:
    - complete a data protection impact assessment;
    - ask Data Subjects to provide explicit consent as the lawful basis for data processing; and

Co-funded by the Horizon 2020 Framework Programme of the European Union

  - provide Data Subjects with a participant information sheet to inform them of the purpose of the data collection, their rights and the data controller's contact details.

- ➔ The FLAME platform will:
  - be architected in such a way as to enable the Data Subject's rights, in particular, the "right to erase"; and
  - will use the data protection by design and default principles.

- ➔ Data Controllers and Data Processors will:
  - sign a standard contract to govern the data processing;
  - keep detailed documentation on data control and processing activities; and
  - will each appoint a Data Protection Officer and will inform them of the project.

4.3 EPRIVACY DIRECTIVE

The GDPR is complemented by Directive 2002/58/EC on privacy and electronic communications ("ePrivacy Directive"), amended by Directive 2009/136, which concerns the protection of privacy in the electronic communications sector and covers some data not classed as "personal" such as some communications meta-data. As a Directive, it is transposed into EU nations' laws rather than being imposed in a unified way as a Regulation is. In the UK for instance, the ePrivacy Directive is transposed into the Privacy and Electronic Communications Regulations.

The EC had planned to introduce a new ePrivacy *Regulation* to be introduced across the EU at the same time as the GDPR (25th May 2018). As anticipated, this deadline was missed: the ePrivacy Regulation is not finalised at the time of writing.

As FLAME will operate on a virtual slice of a city infrastructure and is purely in place for research purposes, we do not believe that it counts as a "public communications network". However, we will respect the spirit of the regulations pertaining to the protection of the users' data and their informed consent.

The Articles of the Directive pertinent to FLAME are discussed below.

Commentary

As and when the new ePrivacy Regulation is published, the advice to the consortium will be updated.

4.3.1 Article 3: Scope

The directive applies "to the processing of personal data in connection with the provision of publicly available electronic communications services in public communications networks in the Community."

4.3.2 Article 4: Security

"The provider of a publicly available electronic communications service must take appropriate technical and organisational measures to safeguard security of its services, if necessary in conjunction

Co-funded by the Horizon 2020 Framework Programme of the European Union

with the provider of the public communications network with respect to network security. Having regard to the state of the art and the cost of their implementation, these measures shall ensure a level of security appropriate to the risk presented."

4.3.3 Article 5: Confidentiality of the Communications

"Member States shall ensure the confidentiality of communications and the related traffic data … In particular, they shall prohibit listening, tapping, storage or other kinds of interception or surveillance of communications and the related traffic data … without the consent of the users concerned".

Paragraph 3 of Article 5 states that a communications network cannot be used to store data on a user's device ("terminal equipment") or gain access to that data unless the user has consented, unless the data is required for the service requested by the user. This is the law applied to cookies in websites.

4.3.4 Article 6: Traffic data

The directive describes the processing and storage of "traffic data": the communications meta-data required for the operation (including billing) of the communications network. Traffic data may be kept for a period for billing purposes, may (with consent) be processed for "for marketing electronic communications services or for the provision of value added services". The service provider must inform the subscriber or user of the data processing being done and its duration and must also erase or anonymise the data once the processes of all legitimate purposes are complete.

4.3.5 Article 9: Location Data other than Traffic Data

"1. Where location data other than traffic data, relating to users or subscribers of public communications networks or publicly available electronic communications services, can be processed, such data may only be processed when they are made anonymous, or with the consent of the users or subscribers to the extent and for the duration necessary for the provision of a value added service. The service provider must inform the users or subscribers, prior to obtaining their consent, of the type of location data other than traffic data which will be processed, of the purposes and duration of the processing and whether the data will be transmitted to a third party for the purpose of providing the value added service. Users or subscribers shall be given the possibility to withdraw their consent for the processing of location data other than traffic data at any time.

"2. Where consent of the users or subscribers has been obtained for the processing of location data other than traffic data, the user or subscriber must continue to have the possibility, using a simple means and free of charge, of temporarily refusing the processing of such data for each connection to the network or for each transmission of a communication.

"3. Processing of location data other than traffic data in accordance with paragraphs 1 and 2 must be restricted to persons acting under the authority of the provider of the public communications network or publicly available communications service or of the third party providing the value added service, and must be restricted to what is necessary for the purposes of providing the value added service."

4.3.6 Probable Changes Arising from the ePrivacy Regulation

It is not know at this point precisely how the new ePrivacy Regulation will differ from the existing rules. Changes in the rules regarding cookies, an expansion of its application to over-the-top communications channels (such as Skype) as well as traditional telecoms providers, and its application to organisations based outside of the EU if they do business with the EU are all expected.

Co-funded by the Horizon 2020
Framework Programme of the European Union

This document will be revised once the ePrivacy Regulation is ratified.

4.3.7 Summary

As FLAME is bound by the data protection rules of the GDPR, the ePrivacy directive (and upcoming regulation) do not impose significant additional constraints. In summary:

- The communications service must safeguard its security.
- Communications and related traffic must be kept confidential.
- Consent must be gained from the user to store or retrieve data from their device.
- Consent must be gained from the user to process their traffic data.
- Consent must be gained from the user to process their location data.
- The type of location data processed, the purpose and duration of processing, and whether it will be passed to a third party of location data must be described to the user.
- The user must simply be able to temporarily refuse processing of their location data.

4.4 ETHICS

Ethical oversight of the FLAME project is implemented through review by an Ethics Management Board (EMB). The EMB includes the Ethics Manager (EM), Project Coordinator (PC), Project Manager (PM), 3rd Party Project Manager (3PPM) and other partners as needed. The EMB ensures that trials are conducted in consideration of ethical requirements and adhere to legal and ethical requirements, while overseeing the application of the ethical management strategy of the consortium.

The consortium has created two documents in response to the ethical review conducted during the grant preparation phase. These documents ([D1.1] and [D1.2]) describe procedures for informed explicit consent; confidentiality and the minimisation (and proportionality) of data collection; participant information sheets; data collection and management procedures; and various issues to do with operating with countries outside of the EU.

The documents were written to comply with Directive 95/46/EC and Directive 2002/58/EC which, as noted above, are being replaced by the GDPR and ePrivacy Regulation. The procedures described in the documents have been revised in line with the new regulations (so for instance, consent forms will now need to also describe that trials may use automated decision making, including profiling).

A successful ethics application was made to the ethics board of the coordinator’s organisation (University of Southampton). The application included a description of the aims of the project, the way it expected trials to operate and the sorts of data expected to be gathered. As part of the ethics application a standard consent form and standard participant information sheet were created and submitted.

The application for this "blanket" consent has been provided to trials and any deviation from the activities described in the application must be checked with the project's Ethics Board and potentially a further ethics approval application would have to be made.

Co-funded by the Horizon 2020
Framework Programme of the European Union

Some organisations require researchers to obtain local ethics approval. If this is the case, then the local ethics board may be content to see the application made to the University of Southampton board and to know that it was approved. Evidence of a successful local application must be provided to the coordinator.

4.5 GRANT AGREEMENT

Article 26 of the FLAME Grant Agreement (the contract between the Beneficiaries and the EC) details the "Ownership of Results" where "Results" is defined in 26.1 as "any (tangible or intangible) output of the action such as data, knowledge or information - whatever its form or nature, whether it can be protected or not - that is generated in the action, as well as any rights attached to it, including intellectual property rights."

Article 26.1 states that "Results are owned by the beneficiary that generates them." as shown in Figure 2.

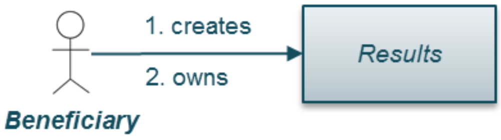


Figure 2. "Results are owned by the beneficiary who creates them."

Article 26.2 then describes "Joint ownership by several beneficiaries":

> Two or more beneficiaries own results jointly if:
>
> 1) they have jointly generated them and
>
> 2) it is not possible to:
>
>   (i) establish the respective contribution of each beneficiary, or
>
>   (ii) separate them for the purpose of applying for, obtaining or maintaining their protection (see Article 27).
>
> The joint owners must agree (in writing) on the allocation and terms of exercise of their joint ownership ('joint ownership agreement'), to ensure compliance with their obligations under this Agreement.

Article 29.3 of the FLAME Grant Agreement says:

> Regarding the digital research data generated in the action ('data'), the beneficiaries must:
>
> 1. deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:
>
>   i. the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;

Co-funded by the Horizon 2020 Framework Programme of the European Union

   ii. other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan' (see Annex 1);

2. provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).

> This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.
>
> As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the action's main objective, as described in Annex 1, would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access.

Commentary

Objective 6 of the FLAME DoA is "EaaS sustainability and business models that combine public-private investment in localised infrastructures with an exploitable software platform" and describes "formulating business models as well as exemplary business plans that are deemed suitable for a long-term sustainability with a drive to commercialization of both underlying infrastructure and the capability to provide large-scale EaaS offerings".

It may be that the objective of commercialising the underlying infrastructure could be jeopardised by making public some parts of the project's research data.

4.6 CONSORTIUM AGREEMENT

The FLAME Consortium Agreement which is the contract in place between the FLAME Beneficiaries (and does not cover any cascade funding partners) includes several clauses related to data management.

Data itself is covered by the term "results", defined (as in the Grant Agreement) as "any (tangible or intangible) output of the action such as data, knowledge or information - whatever its form or nature, whether it can be protected or not - that is generated in the action, as well as any rights attached to it, including intellectual property rights."

Partners are permitted access to other partner's *results* (including data) where they can show that access is "needed". If access is needed for the performance of the own work of a Party then it shall be granted on a royalty-free basis.

The Consortium Agreement also includes clauses regarding:

- joint ownership of *results* (adding to the terms in the Grant Agreement);
- regarding dissemination of *results*, and associated processes including notice period;
- regarding non-disclosure of confidential information shared between partners.

Co-funded by the Horizon 2020 Framework Programme of the European Union

These clauses must of course be followed along with the data management requirements arising from the other agreements and regulations.

The consortium agreement is being updated to harmonise it with the GDPR.

Co-funded by the Horizon 2020
Framework Programme of the European Union

5 DATA LICENCES

As discussed above, a dataset is a "work" and therefore the default position for a dataset is that the owner has copyright. Copyright tightly restricts what anyone else can do with the dataset. A licence for a dataset provides a licensee various rights over the dataset that they would otherwise not have. The owner of the work can choose whether or not to license a dataset to another party and may choose to license the dataset with different licences to different parties (for instance if a dataset has been licensed with a non-commercial licence such as CC-BY-NC it does not stop the owner also licensing the same dataset with a commercial licence to a specific party). The terms of the licence chosen by the owner may be constrained by other licences or contracts in place.

To recap from Section 3, the EC advocates the FAIR principles for research data management where "FAIR" is an acronym for Findable, Accessible, Interoperable and Reusable.

As noted above, the conditions for being "Reusable" are (with added emphasis):

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
    - R1.1. (meta)data are released with a **clear and accessible data usage license**
    - R1.2. (meta)data are associated with detailed provenance
    - R1.3. (meta)data meet domain-relevant community standards

Note, the FAIR principles themselves do not mandate an "open" licence, just that it should be clear what the licence is.

The ORD pilot follows the principle of data being "as open as possible, as closed as necessary" and the Open Definition [Open] states:

> "Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)."

Where possible, FLAME must apply licences to data which conform to the open definition. The Creative Commons [CC] and Open Data Commons [ODC] organisations both provide commonly used open licences, described below.

5.1 LICENCES CONFORMANT TO THE OPEN DEFINITION

As stated in [Open] the following licences are conformant with the principles set forth in the Open Definition.

The licenses found in Table 1 are widely used, current and conform to the Open Definition.

Co-funded by the Horizon 2020 Framework Programme of the European Union

Table 1. Recommended licences conformant with the Open Definition ordered from most permissive to most restrictive[2].

License	Requires attribution	Requires share-alike	Comments
Creative Commons CCZero (CC0)	N	N	Dedicate to the Public Domain (all rights waived)
Open Data Commons Public Domain Dedication and Licence (PDDL)	N	N	Dedicate to the Public Domain (all rights waived)
Creative Commons Attribution 4.0 (CC-BY-4.0)	Y	N	
Open Data Commons Attribution License (ODC-BY)	Y	N	Attribution for data(bases)
Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0)	Y	Y	
Open Data Commons Open Database License (ODbL)	Y	Y	Attribution-ShareAlike for data(bases)

FLAME should endeavour to use licences from this list for any open datasets.

5.1.1 CC-BY-4.0

The Creative Commons Attribution 4.0 International licence [CC-BY] is the most permissive of the Creative Commons licences.

The licensee is free to:

- Share — copy and redistribute the material in any medium or format
- Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as the licensee follows the following license terms:

- Attribution — the licensee must give appropriate credit, provide a link to the license, and indicate if changes were made. They may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- No additional restrictions — the licensee may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

[2] Adapted from http://opendefinition.org/licenses/, licensed under CC Attribution 4.0 International License.

Co-funded by the Horizon 2020 Framework Programme of the European Union

5.1.2 CC-BY-SA-4.0

The Creative Commons Attribution-ShareAlike 4.0 International licence [CC-BY-SA] builds on CC-BY-4.0 by adding the following term:

- ShareAlike — if the licensee remixes, transforms, or builds upon the material, they must distribute their contributions under the same license as the original.

This makes the licence "viral" to use a term often applied to the GPL software licence [GPL].

5.2 LICENCES NOT CONFORMING TO THE OPEN DEFINITION

Many licences do not conform to the Open Definition so it would not sever much purpose listing many here, but it is worth noting the following Creative Commons licences which do not conform:

- Those including the "no-derivatives" term: CC-BY-ND, CC-BY-NC-ND

  Not open as they restrict reuse / modification.

- Those including the "no-commercial" term: CC-BY-NC, CC-BY-NC-SA, CC-BY-NC-ND

  Not open as they restrict the purpose of reuse.

That does not mean that they cannot be used if there is justification. Indeed, under the ORD principle having no public licence at all is permissible if justified.

Co-funded by the Horizon 2020 Framework Programme of the European Union

6 REPOSITORY

6.1 REQUIREMENTS

This section describes the requirements for selecting a repository for any FLAME open research data and is adapted from a similar analysis done by Steve Taylor (ITINNOV) for the Fed4FIRE+ project.

Requirement 1 – DOIs as Unique Data Identifiers [DOI]. The repository must be an authorised issuer of DOIs, and must issue a DOI to identify each data deposition. DOIs are a well-established and proven mechanism to provide archival citations for digital resources, and DOIs are not vulnerable to link rot.

Requirement 2 – Evidence of Repository Longevity. The repository must provide evidence of its sustainable future existence over the long term. This could be evidence of funding for the foreseeable future, a sustainable business operation model. Further evidence of the potential for the repository's longevity e.g. that the repository is well-adopted and well-used is desirable. This requirement is necessary because the data needs to be stored for the long term, so the repository it is stored in must be sustainable over the long term.

Requirement 3 – Commitment to Data Integrity Preservation. The repository must publish policies as to the level and type of their efforts concerning protection from data decay, alteration or loss, including what will happen to data the event of the repository going out of business. This is so that users of the repository can judge the repository's methods and commitment to data integrity preservation. The policies should provide evidence that at least best efforts are made to protect the integrity of the data.

Requirement 4 – Descriptive Metadata Specification. The repository must determine or choose a metadata specification that is adequate to enable the data to be found (e.g. provide keywords, title, description). The metadata specification must also unambiguously identify the creator of the data set (e.g. name, affiliated organisation) and the creation date. The metadata specification should provide descriptions of data types or formats, especially if they are proprietary. Support for Dublin Core [Dublin Core] or DataCite [DataCite] is considered advantageous.

Requirement 5 – Storage & Export of Descriptive Metadata. The repository must store descriptive metadata immutably bound to the data it describes. The metadata must be made available to open research data search engines. The metadata exported to search engines should be in a standardised format (e.g. Dublin Core or DataCite) and exported automatically using standardised protocols (e.g. OAI-PMH). Storing and exposing descriptive metadata is necessary to enable the data it describes to be found easily.

Requirement 6 – Flexibility of Licensing. The repository must not mandate a single license for all its content. This is because the data creators need some say in deciding the license terms of their data. It is acceptable to provide a set of licenses from which the data creator can choose. Any license stipulations by the repository should be compatible with the Open Definition. This is because the overall objective of storing data in a repository is to open it.

Co-funded by the Horizon 2020 Framework Programme of the European Union

6.2 REVIEW

This section contains an evaluation of two candidate repositories with a view to providing at least one repository as a starting point for FLAME. More repositories may be evaluated and added as is deemed necessary.

6.2.1 Dryad

The Dryad Digital Repository is:

> “… a curated resource that makes the data underlying scientific publications discoverable, freely reusable, and citable. Dryad provides a general-purpose home for a wide diversity of datatypes.” [Dryad]

Requirements Assessment

Dryad’s policies [Dryad Policies] provide information to asses Dryad against the repository requirements. All quotations in this assessment, unless otherwise attributed, are from [Dryad Policies].

Requirement 1 – DOIs as Unique Data Identifiers. The preamble to Dryad’s policy concerning Dryad’s objectives states:

- “Guiding these Terms of Service are Dryad's aims to:
    - […]
    - assign and provide Digital Object Identifiers (DOIs) to repository content;”

This is adequate evidence that DOIs are used within Dryad.

Requirement 2 – Evidence of Repository Longevity. Dryad’s policies concerning sustainability state:

- “… Dryad's governance and business model are designed to provide for long-term organizational stability and viability, ensuring that revenues from Data Publishing Charges (DPCs) cover the Repository's core operating costs (including curation, storage, and maintenance). Thus, long-term funding for Dryad is not dependent on a small number of grants, or the continued largesse of a single host institution.”

This is adequate evidence of the sustainability of Dryad.

Requirement 3 – Commitment to Data Integrity Preservation. Dryad’s policies concerning its procedures for preservation state:

- “Dryad aims to preserve the originally submitted Content indefinitely. Steps to ensure long-term availability include:
    - Persistent identification in the form of DOIs.
    - Replication of data and metadata.
    - Periodic verification that stored content remains uncorrupted. Upon ingest, an MD5 checksum is calculated and recorded for each submitted data file. Integrity checks are run nightly.

Co-funded by the Horizon 2020 Framework Programme of the European Union

  - If data objects are found to be corrupted, affected data will be restored from known-good copies as appropriate.
  - To prevent data loss, multiple copies of content are kept at different sites.
  - Prior to release, files are manually inspected by a Dryad curator. Any issues that affect preservation (i.e., file corruption, etc.) are addressed before release.
  - Ingest, curation, and publication actions are documented by provenance metadata.
  - Dryad content is replicated through participation in the DataONE network[3].
  - To the greatest extent possible, personnel create documentation to reflect their activities. We strive to ensure that more than one employee can perform each critical function of the repository."
- In addition, Dryad's policy on sustainability states: "… Dryad's participation in the DataONE network ensures that all its data will be available through other institutions if the Dryad organization ever dissolves."

This is adequate evidence that data stored within Dryad is protected for the long term or at risk from the Dryad going out of business.

Requirement 4 – Descriptive Metadata Specification. Dryad is not specific about the metadata necessary, because its assumption is that data is bound to a publication stored in a different publisher's repository, and the publication will reference the data. It states in its FAQ[4]:

- "Dryad welcomes data files associated with any published article in the sciences or medicine, as well as software scripts and other files important to the article."
- "Dryad works with journals to integrate article and data submission, streamlining the submission process."
- "[submitters need to] Provide titles, descriptions and keywords for your datafiles, to make the data more discoverable and to assist in understanding the relationship of the datafile to the publication."

FLAME needs to cater for data being findable independent of publications, so the fact that Dryad relies on data bound to publications is problematic.

Requirement 5 – Storage & Export of Descriptive Metadata. It is clear that Dryad stores some metadata. Dryad provides a search within its website based on keywords from the publications, but the metadata does not appear to be exported to external open data search engines. This is most likely due to the operating model of Dryad where data is bound to a publication, which can be found externally. For FLAME, this approach is problematic.

[3] https://www.dataone.org/

[4] https://datadryad.org/pages/faq

Co-funded by the Horizon 2020 Framework Programme of the European Union

Requirement 6 – Flexibility of Licensing. Dryad uses the Creative Commons CC0 rights waiver for all data it hosts meaning that the data is explicitly public domain and the creator has waived all their rights to it. This policy is directly contradictory to the requirement that the repository must not mandate a single license.

Costs: Dryad charges a fixed fee of $US 120 for each submission.

Given the problems described above. Dryad cannot be considered as a repository for FLAME

6.2.2 Zenodo

Zenodo [Zenodo], hosted by CERN, is recommended as the *de facto* OpenAIRE-compliant data repository, but other repositories are permissible provided they are compliant with the OpenAIRE standards:

"Researchers working for European funded projects can participate by depositing their research output in a repository of their choice, publish in a participating Open Access journal, or deposit directly in the OpenAIRE repository ZENODO – and indicating the project it belongs to in the metadata. Dedicated pages per project are visible on the OpenAIRE portal."[5]

Zenodo's implementation of the FAIR principles is described in [Zenodo Principles]. Zenodo provides Digital Object Identifiers (DOIs) [DOI] for all uploaded data bundles.

Requirements Assessment

This assessment uses [Zenodo Principles] and [Zenodo Policies] as its primary sources.

Requirement 1 – DOIs as Unique Data Identifiers. Zenodo mints and assigns DOIs to data submissions. From [Zenodo Principles] on FAIR Principles:

- "F1: (meta)data are assigned a globally unique and persistent identifier
    - A DOI is issued to every published record on Zenodo."

This is clearly adequate evidence that Zenodo issues DOIs for the data it stores.

Requirement 2 – Evidence of Repository Longevity. The Longevity section of [Zenodo Policies] state:

- "Retention period: Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least."

In the "Accessible" section, [Zenodo Principles] state:

- "A2: metadata are accessible, even when the data are no longer available

[5] https://www.openaire.eu/support/faq/openaire-faq

Co-funded by the Horizon 2020 Framework Programme of the European Union

- Data and metadata will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least."

Metadata are stored in high-availability database servers at CERN, which are separate to the data itself.

Given that Zenodo is hosted by CERN, which is large, well-funded and has a research program planned for at least the next 20 years, Zenodo is deemed to have sufficient chances of longevity for the long term.

Requirement 3 – Commitment to Data Integrity Preservation. In the section concerning Longevity, [Zenodo Policies] states:

- "Versions: Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and record are preserved.
- Replicas: All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- Retention period: Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.
- Functional preservation: Zenodo makes no promises of usability and understandability of deposited objects over time.
- File preservation: Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
- Fixity and authenticity: All data files are stored along with a MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant.
- Succession plans: In case of closure of the repository, best efforts will be made to integrate all content into suitable alternative institutional and/or subject based repositories."

The [Zenodo FAQ] has a specific question on data preservation:

- "Is my data safe with you / What will happen to my uploads in the unlikely event that Zenodo has to close?
- Yes, your data is stored in CERN Data Center. Both data files and metadata are kept in multiple online and independent replicas. CERN has considerable knowledge and experience in building and operating large scale digital repositories and a commitment to maintain this data centre to collect and store 100s of PBs of LHC data as it grows over the next 20 years. In the highly unlikely event that Zenodo will have to close operations, we guarantee that we will migrate all

Co-funded by the Horizon 2020 Framework Programme of the European Union

content to other suitable repositories, and since all uploads have DOIs, all citations and links to Zenodo resources (such as your data) will not be affected."

Files are redundantly stored in multiple locations, are also backed up and have frequent integrity tests in the form of checksums. In addition, Zenodo has a succession plan in the event of its closure. These factors provide adequate evidence of Zenodo's commitment to the integrity of the data it hosts.

Requirement 4 – Descriptive Metadata Specification & Requirement 5 – Storage & Export of Descriptive Metadata. The Content section of [Zenodo Policies] states:

- "Metadata types and sources: All metadata is stored internally in JSON-format according to a defined JSON schema. Metadata is exported in several standard formats such as MARCXML, Dublin Core, and DataCite Metadata Schema (according to the OpenAIRE Guidelines)."

The Access and Reuse section of [Zenodo Policies] states:

- "Access to data objects: Files may be deposited under closed, open, or embargoed access. Files deposited under closed access are protected against unauthorized access at all levels. Access to metadata and data files is provided over standard protocols such as HTTP and OAI-PMH."

Zenodo exports metadata in Fed4FIRE+'s preferred standards, Dublin Core and DataCite. OAI-PMH is used to export metadata, which can be harvested by open research data search engines. Therefore both requirements 4 and 5 are deemed satisfied.

Requirement 6 – Flexibility of Licensing. The Content section of [Zenodo Policies] states:

- Licenses: Users must specify a license for all publicly available files. Licenses for closed access files may be specified in the description field.

In the Reusability section, [Zenodo Principles] state:

- "R1.1: (meta)data are released with a clear and accessible data usage license
    - License is one of the mandatory terms in Zenodo's metadata, and is referring to a Open Definition license.
    - Data downloaded by the users is subject to the license specified in the metadata by the uploader."

The metadata that describes any data stored in Zenodo is CC0 license, forcing it to be public domain and non-copyright. The Access and Reuse section in [Zenodo Policies] states:

- "Metadata access and reuse: Metadata is licensed under CC0, except for email addresses. All metadata is exported via OAI-PMH and can be harvested."

Licensing only metadata as CC0 is acceptable because the metadata's purpose is to advertise and describe the type and purpose of the data so that it may be reused. The owner of the data is still free to choose a license for the actual data they upload.

That the user specifies a license for their data is mandatory. Zenodo recommends that the license is compliant with [Open Definition] (i.e. using one of the licenses in [Open Definition Licenses]). Therefore, Zenodo gives the user a choice of license, so requirements 6 is deemed satisfied.

Co-funded by the Horizon 2020 Framework Programme of the European Union

Costs: Zenodo is free at the point of use for most submissions. [Zenodo Terms] state:

- "Content may be uploaded free of charge by those without ready access to an organized data centre."

Other Factors:

Zenodo also provides automated reporting to the EC for open data stored within it, so evidence of the commitment from FLAME and its trials to provide open data can be easily tested.[6] Zenodo also supports collections, and it is possible that a collection could be created for the FLAME trial community.

6.2.3 Recommendation

Clearly Zenodo is the obvious first choice for an approved data repository. Others may follow as appropriate, provided they are compliant with the requirements in Section 6.1.

[6] http://help.zenodo.org/features/

Co-funded by the Horizon 2020 Framework Programme of the European Union

7 STAKEHOLDERS

There are many ways to subdivide the FLAME stakeholders. We present two lists here designed to be appropriate to the following discussion of data management. Firstly by funding type (and hence contract type):

- Core Partner
  - One of the direct Beneficiaries of the FLAME Grant Agreement.
- Open Call (Trial / Replica) Partner
  - A recipient of FLAME cascade funding either leading for running a trial or operating a FLAME replica.
- Unfunded Partner
  - An organisation running an unfunded trial.
- Project Partner
  - A Core Partner, Open Call Partner or Unfunded Partner.
- Member of the Public
  - None of the above.

It is also useful to define:

- Trial Member:
  - Project Partner who has an interest in a trial because of the use of their software component (or other IP) or infrastructure in the operation of the trial.

Secondly, more by role in the project along with motivation and dataset concerns:

- Trial Participant ("user")
  - Most likely a member of the public (N.B. if a Trial Participant is a member of one of the funded parties then they should not be treated any differently in terms of personal data).
  - Is interested in helping research and in new media experiences. Takes part in a trial.
  - Requires that their data is protected and only used for the stated purposes of the trial.
- Publication Consumer
  - A Member of the Public.
  - Is interested in reading about future media internet research.

Co-funded by the Horizon 2020 Framework Programme of the European Union

  - Wants publications (and their supporting data) to be open.

- Open Data Consumer
  - Most likely a Member of the Public (at another research organisation).
  - Is interested in better understanding the project's data (often backing up a publication) and potentially reanalysing the data, sometimes to build on the work.
  - Would like the data to be as open as possible and the FAIR principles to be adhered to.
- Trial Leader
  - One of DRZ, ETH, NXW and VRT (for the validation trials), an Open Call Trial Partner or Unfunded Partner.
  - Has a novel media service and/or associated mobile app and wants to explore and understand the new opportunities provided by FLAME. Deploys a Media Service (comprised of one or more Media Service Components) on the FLAME Platform.
  - Collects personal data about Trial Participants. Wants their own IPR protected and wants to ensure that the trial participants' rights are protected[7]. Wants access to (pseudonymised) activity data associated with their trial in order to improve their software and services.
- Media Component Developer
  - Core Partner, Open Call Trial partner or Unfunded Partner.
  - Wants to verify and validate their component, understand its usage and performance and how the performance can be optimised through novel resourcing strategies.
  - Wants access to (pseudonymised) activity data associated with trials using the Media Component in order to improve it.
- Platform Service Developer
  - A Project Partner, primarily one of IDE, ITINNOV, Martel, Atos.
  - Wants to verify and validate their component, understand its usage and performance and how it reacted to the inputs from the trial.
  - Wants access to (pseudonymised) activity data associated with all trials.

[7] Of course, all stakeholders want their IPR protected and want to uphold the rights of Trial Participants. They are both noted here as (a) the IPR issue is key for Open Call Trial Partners working with the Core Partners, (b) the Trial Leader has the direct relationship to the Trial Participants.

Co-funded by the Horizon 2020 Framework Programme of the European Union

- Platform Operator
    - A Core Partner, Open Call Replica Partner or Unfunded Partner. Potentially the same as the Infrastructure Operator.
    - Deploys the FLAME Platform (comprised of Platform Services) in an infrastructure slice provided by the Infrastructure Operator. Is interested in exploring and understanding new business models enabled through FLAME.
    - Wants access to (pseudonymised) activity data associated with all trials executing on the Platform instance. Collects personal (traffic) data and wants to ensure that the Trial Participants' rights are respected.
- Infrastructure Operator
    - A Core Partner, Open Call Replica Partner or Unfunded Partner.
    - Manages the physical infrastructure in a city upon which the FLAME Platform is deployed.
    - Wants to understand the resource requirements for the FLAME Platform and new business models enabled by the platform.

7.1 CONTRACTS

The rights, obligations and liabilities around sharing data vary between the different types of party involved: Members of the Public, Core Partners, Open Call Partners and Unfunded Partners. An analysis of Unfunded Partners will be added in the second version of this document.

7.1.1 Core Partners

If a Core Partner generates a Result (such as data) and a second Core Partner needs access to the result to carry out their project work then this is permitted by the Consortium Agreement (Figure 3). If the data is personal data then an additional agreement is required (see below).

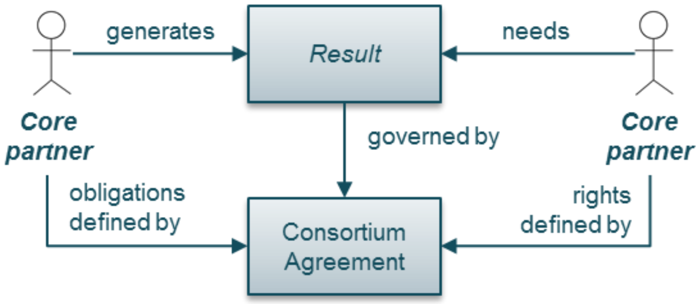


Figure 3. Core Partner rights are governed by the Consortium Agreement.

If two or more Core Partners jointly generate a Result then the rules are primarily set out in the Grant Agreement (with some additional terms in the Consortium Agreement), see Figure 4. As discussed

Co-funded by the Horizon 2020 Framework Programme of the European Union

above in Section 4.5 the Grant Agreement requires a written "join ownership agreement" between the Parties.

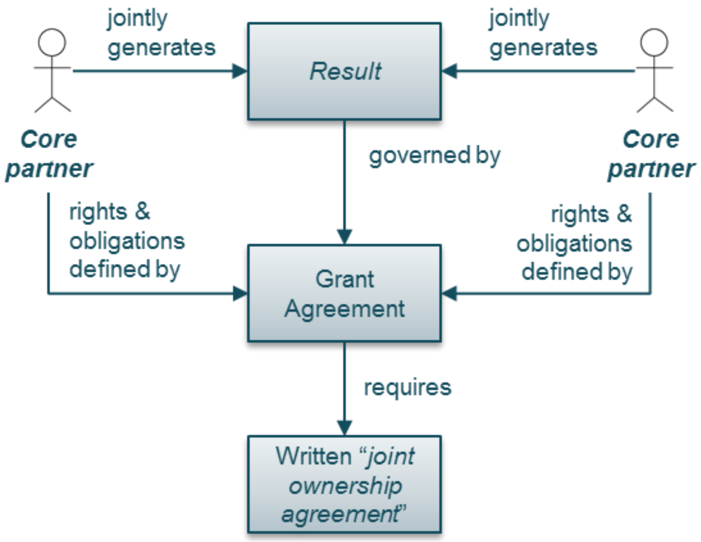


Figure 4. Core partners jointly owning results must refer to the Grant Agreement.

7.1.2 Open Call Partners

The contractual terms defining the relationship between two or more Open Call Partners ("OC Partner" in the Figures), an Open Call Partner and a Core Partner or a joint work involving at least one Open Call Partner will be within the Cascade Funding Agreement.

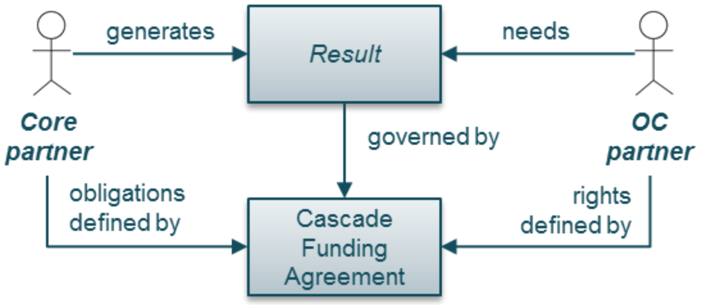


Figure 5. Results involving at least one Open Call Partner will be governed by the Cascade Funding Agreement.

Figure 5 shows an example of a Core Partner generating a Result which is needed by an Open Call partner. Access rights to the Result will be defined in the Cascade Funding Agreement. Similar diagrams could be drawn for the situations of the Result being generated by the Open Call Partner and needed by either a Core Partner or another Open Call Partner.

The Cascade Funding Agreement will also govern results which are jointly generated by either:

a) two or more Open Call Partners; or

b) one or more Core Partners and one or more Open Call Partners.

Co-funded by the Horizon 2020 Framework Programme of the European Union

7.1.3 Personal Data

Any sharing of Personal Data is governed by the GDPR. Figure 6 sketches the relationship between the Data Subject and the Data Controller.

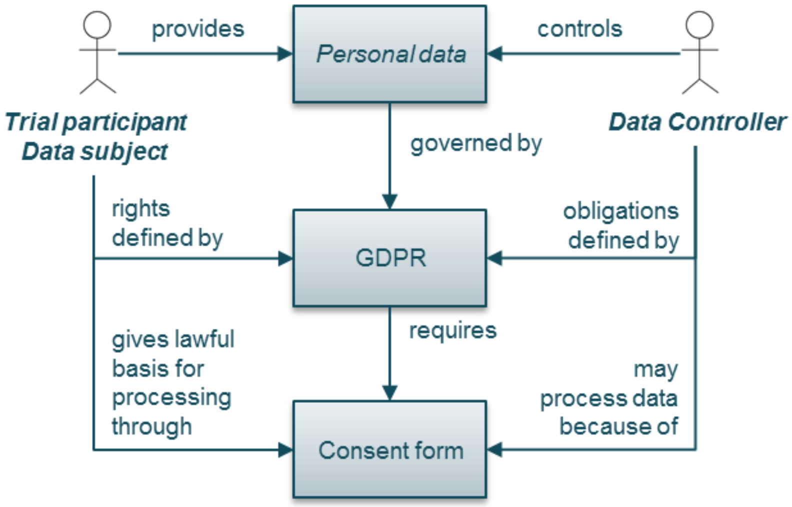


Figure 6. How the GDPR governs the relationship between the Data Subject and the Data Controller.

Once Personal Data has been gathered by the Data Controller, if it is processed by a Data Processor then the GDPR once again comes into force and mandates a separate contract for the processing of the data (see Figure 7).

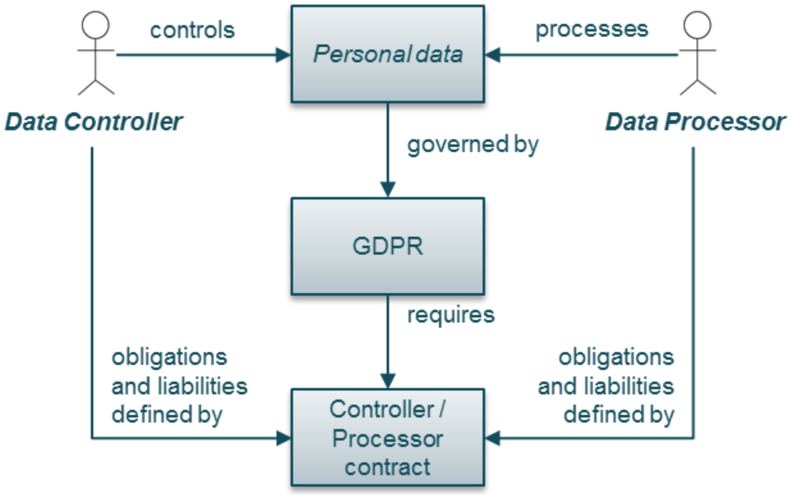


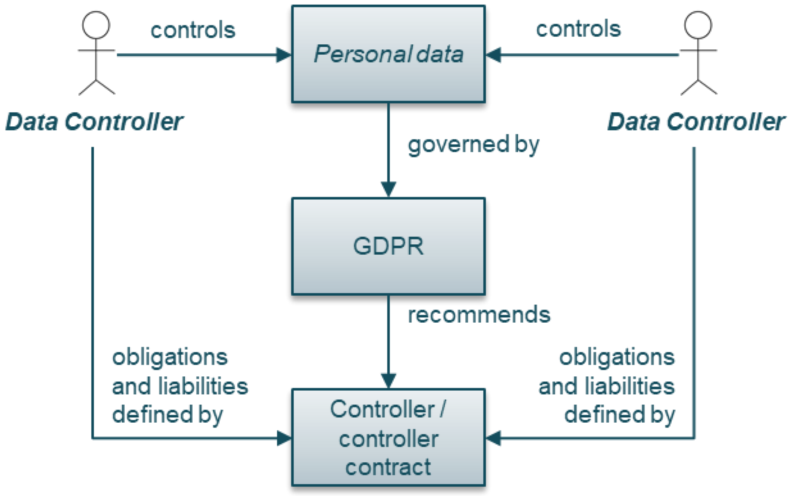
Figure 7. A separate contract is required between a Data Controller and a Data Processor for the processing of a dataset.

Note, it will not be necessary to have distinct contracts for each dataset: a single contract between any two parties can be designed to apply to the whole of the FLAME project work.

The replica partners in FLAME feel that their involvement in the trials is strong enough for them to be considered Data Controllers alongside the Trial Leaders. Therefore we have a Controller-Controller

Co-funded by the Horizon 2020 Framework Programme of the European Union

situation for which the GDPR does not mandate any particular contract. Best practice certainly recommends a contract however. The situation is sketched in Figure 8.



controls
Personal data
controls
Data Controller
Data Controller
governed by
GDPR
recommends
obligations and liabilities defined by
obligations and liabilities defined by
Controller / controller contract

Figure 8. Controller-Controller contract.

7.1.4 Published Data

To be able to publish a dataset resulting from work in FLAME the Project Partner (or partners in the case of a jointly generated result) must be careful to take into account all of:

- copyright law:
    - if the result is a derived work then the rights holder of the original work may restrict publication;
- the Grant Agreement:
    - governing the Core Partners and encouraging the open publication of data but not at the expense of confidentiality, the objectives of the project or the obligations to protect personal data;
    - defining jointly generated results and the need for join agreement in such cases;
- the Consortium Agreement:
    - which adds terms to the Grant Agreement regarding jointly generated results;
    - which defines a notice period for publication and a publication dispute resolution process;
- the Cascade Funding Agreement:
    - which will come into play if a Cascade Funding Partner is involved singly or jointly in generating the Result;
- the GDPR:

Co-funded by the Horizon 2020 Framework Programme of the European Union

- if personal data is involved then Data Subjects' rights must be honoured.

If all these agreements and laws permit the publication of the Result then the partner(s) who own the Result may choose a licence defining the rights of the Member of the Public who would access the Result. The Open Research Data scheme encourages an "open" licence but it may be that there are constraints on the licence imposed by one or more of the governing agreements, see Figure 9.

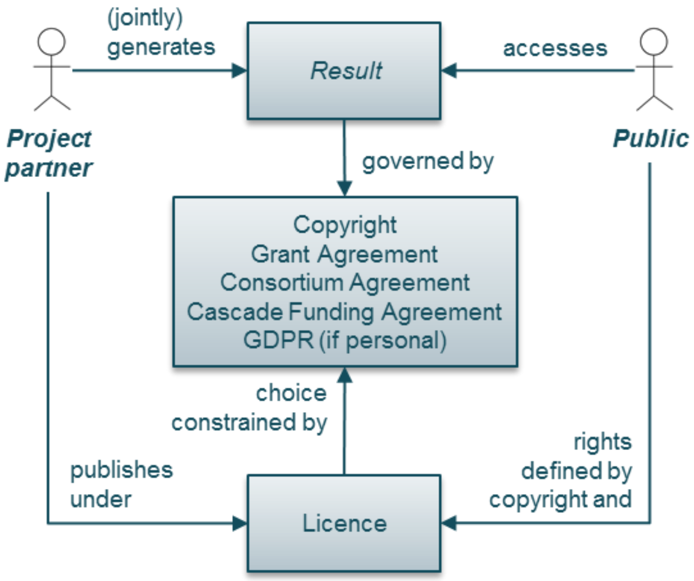


Figure 9. If a result is to be published then many contracts must be respected and a compliant licence chosen.

7.2 DATA CONTROLLER AND PROCESSOR

As described in the FLAME ethics deliverable [D1.1]:

> FLAME will adopt a general Data Controller/Data Processor architecture for the handling and management of the corresponding risks, in accordance with the provisions of the EU Data Protection Directive and local country implementations. The Data Controller and Data Processors each have certain responsibilities under the relevant legislation, but the Data Controller should be a representative of the organisation that has the direct relationship with the data subjects. For FLAME studies, any human subjects are likely to be citizens within the cities, so Data Controllers for trials will be the institutions responsible for operation of the city infrastructure. In most other cases, the party who perform the trial (either platform partners, media service providers or third parties via an open call) will be the lead the trial.

As discussed briefly in Section 2, personal data will be gathered at two points: at the app and media component(s) used by the participant and at the infrastructure slice which the participant's device connects to. Personal data will be processed by various services in the platform and non-personal aggregate data generated.

As alluded to in the stakeholder analysis of Section 7, there are two scenarios for operating the FLAME Platform: where the Infrastructure Operator and the Platform Operator are the same entity and where they are different. Figure 10 shows the infrastructure, key processes and key data owned and controlled by each of the Infrastructure Operator, Platform Operator and Trial Leader. In this scenario

Co-funded by the Horizon 2020 Framework Programme of the European Union

of fully-separated roles, the location and control of the personal data and the hosting relationships indicate that the Platform Operator and Infrastructure Operator would need a data sharing agreement (with the Platform Operator as Controller) and the Trial Leader would need a data sharing agreement with the Platform Operator (with the Trial Leader as the Controller).

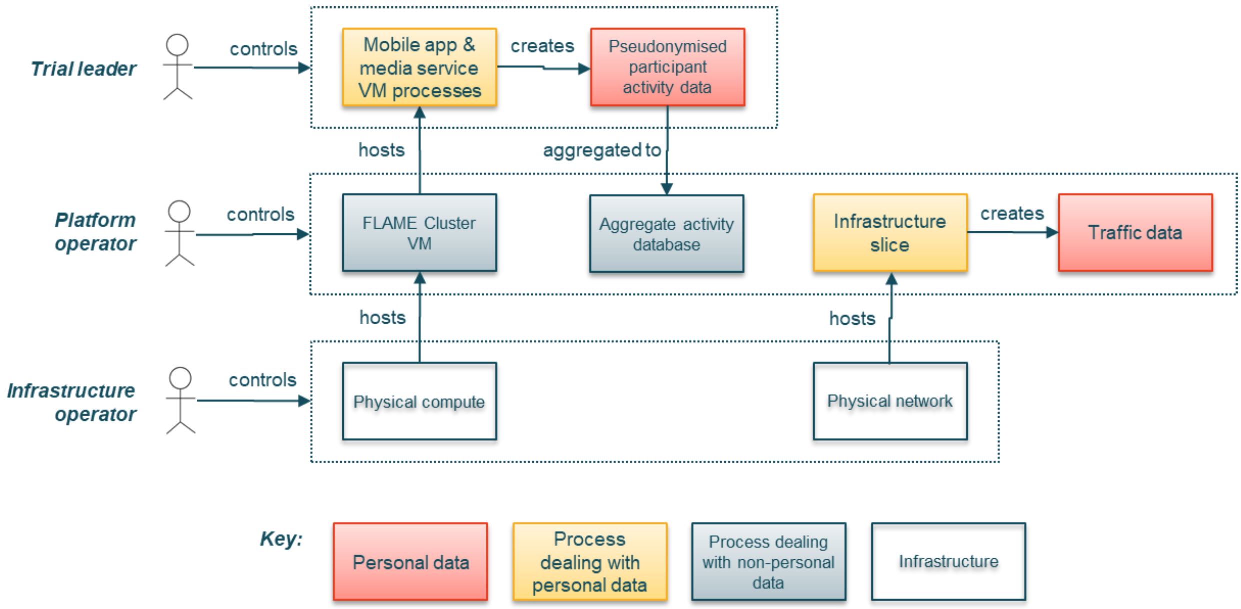


Figure 10. Relations between the three stakeholders, the processes and data.

In the case of the Platform Operator and the Infrastructure Operator being the same entity, the obvious adjustment to Figure 10 of combining the controlled assets in can be made and just a single data sharing agreement would be required (as there are just two entities). Figure 11 below shows the relationships between the stakeholders, data and processes in this reduced situation (omitting the infrastructure operator and physical infrastructure for clarity).

The form of the data sharing agreement is for the two parties to negotiate. Where one party is a data controller and the other party is a data processor, the GDPR mandates that a controller/processor agreement be signed to define the rights and responsibilities of each party. If both parties are data controllers ("joint controllers") then the GDPR does not directly mandate anything [ICO-GDPR-Contract].

The FLAME coordinator (ITINNOV) has worked with the core Bristol and Barcelona replicas (UNIVBRIS and i2CAT) to create standard contracts for the replicas to use with Trial Leaders. Both UNIVBRIS and i2CAT have decided that they are data controllers (not just data processors) and as such are both using Controller-Controller agreements vetted by ITINNOV as coordinator.

In the case of UNIVBRIS there are two starting points for negotiating a Controller-Controller agreement depending on the data to be gathered by the trial and the user equipment to be used. The University makes available up to 20 Android phones for use in trials, and by (a) the University handling the process of distributing and collecting the phones and (b) the trial ensuring that no personal data such as real names or existing logins is collected by their app then the parties can ensure that no personal data is stored in the trial's media services: thus minimising the collection of personal data as required by law. The University thus has a template contract for this situation and another template contract for when

Co-funded by the Horizon 2020 Framework Programme of the European Union

the University's phones are not being used. The contracts must also be adjusted depending on whether the other party is a core partner or 3rd-party as different background contracts are in place.

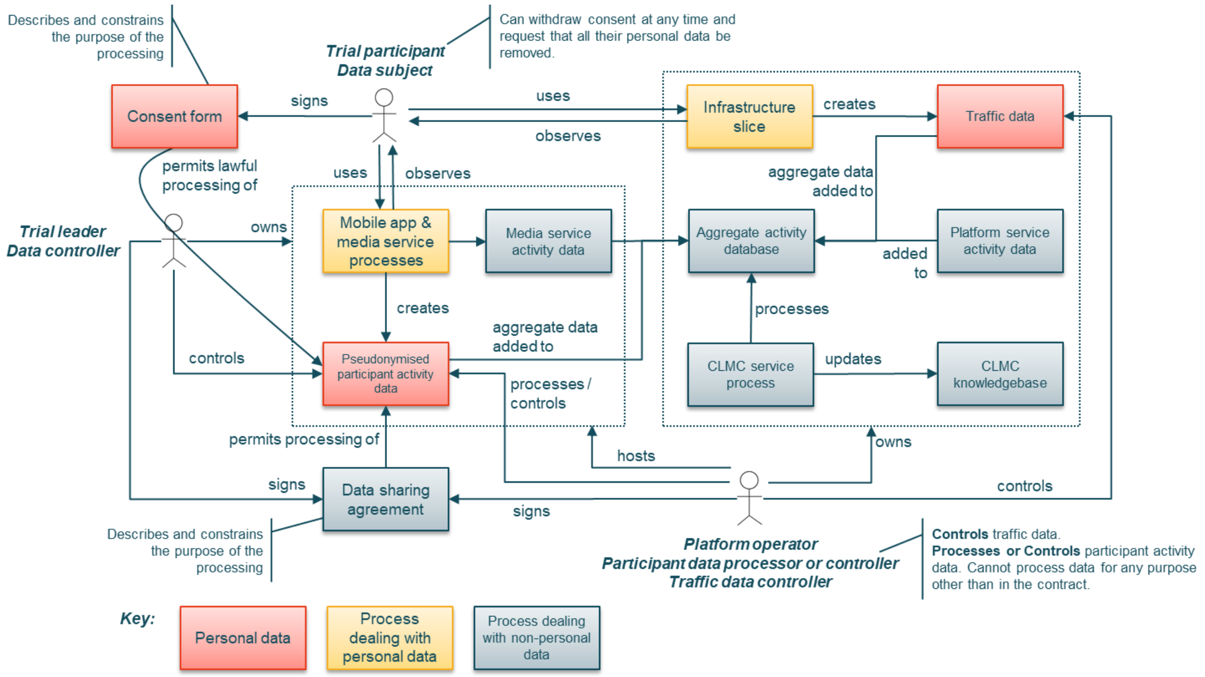

Describes and constrains the purpose of the processing
Trial participant
Data subject
Can withdraw consent at any time and request that all their personal data be removed.
Consent form
signs
uses
observes
Infrastructure slice
creates
Traffic data
permits lawful processing of
uses
observes
aggregate data added to
Trial leader
Data controller
owns
Mobile app & media service processes
Media service activity data
Aggregate activity database
added to
Platform service activity data
creates
processes
aggregate data added to
controls
Pseudonymised participant activity data
processes / controls
CLMC service process
updates
CLMC knowledgebase
permits processing of
hosts
owns
signs
Data sharing agreement
signs
controls
Describes and constrains the purpose of the processing
Platform operator
Participant data processor or controller
Traffic data controller
Controls traffic data. Processes or Controls participant activity data. Cannot process data for any purpose other than in the contract.
Key:
Personal data
Process dealing with personal data
Process dealing with non-personal data

Figure 11. Overview of personal data control and processing.

Key points:

- Trial Leaders are data controller for the data coming from the data subject via the mobile app and associated media component(s).
- The Platform Operator deploys the platform services and is therefore responsible for them.
- Trial Leaders will need a contract with the Platform Operator so that the Platform Operator can process the personal data (at a minimum, "process" meaning host the services using and storing the data).
- The Platform Operator controls the slice of the infrastructure in which the platform is deployed and therefore is data controller for the infrastructure logs.
- Parties who which to examine any of the personal data (e.g. to debug or optimise a component) will also need contracts with the relevant data controller.

Co-funded by the Horizon 2020 Framework Programme of the European Union

8 DATA SETS

Here we enumerate the various datasets used and generated during the lifecycle of a FLAME trial and note their primary owner (often also the data controller), who requires access to the data, whether the data is personal and if there is any potential for the data becoming Open Research Data.

Pre-existing data is listed in Table 2.

Table 2. Data assets pre-existing a trial.

Asset	Owner	Access required by	Personal data	ORD potential
Trial plan template	Martel	Project Partners	No	Yes
Participant information sheet template	Martel	Project Partners	No	Yes
Ethical consent form template	Martel	Project Partners	No	Yes
Service description templates	Platform service developers	Project Partners	No	Yes

Data related to trial planning and set-up is listed in Table 3.

Table 3. Data assets created during trial planning and set-up.

Asset	Owner	Access required by	Personal data	ORD potential
Trial plan	Trial Leader	Project Partners	No	Yes
Participant information sheet	Trial Leader	Project Partners	No	Yes
Ethical consent form	Trial Leader	Project Partners	No	Yes
Completed ethical consent forms	Trial Leader		Yes	No
Service description	Platform service developers & Trial Leader	Project Partners	No	Yes
Media content	Trial Leader	Trial Partners	Possible	Yes

Co-funded by the Horizon 2020
Framework Programme of the European Union

As the situation during the execution of a trial is quite complex, we first have a diagram of the data assets and their relationships (Figure 12).

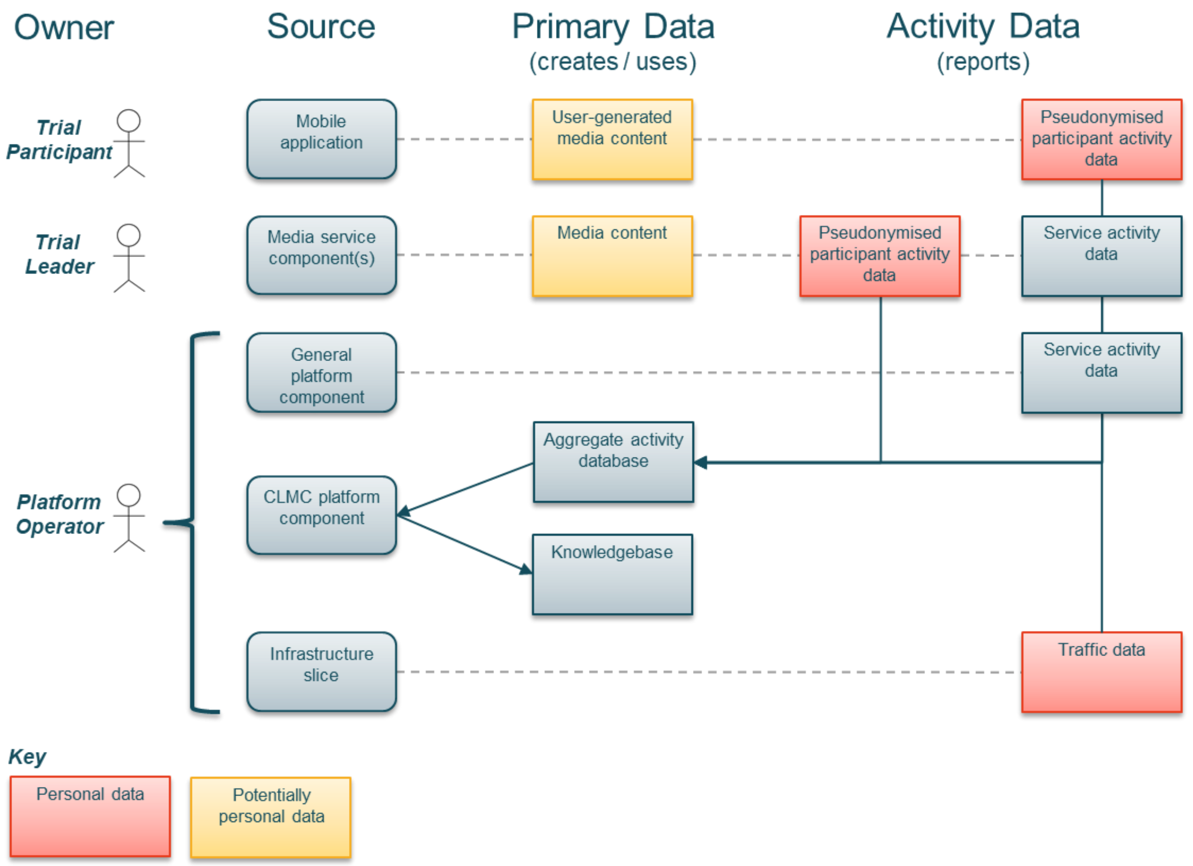


Figure 12. Overview of data generated during trial execution.

The Figure can be read in a similar way to a table with the data related to each source depicted on the same line. The solid lines with arrows show just the data flow for the CLMC. Data is aggregated before transmission to the CLMC which removes any personal identifiers. Some platform components may generate service activity data useful to the CLMC. This interaction is represented by the "General platform component" element.

The data related to trial execution is listed in Table 4.

Table 4. Data assets generated during trial execution.

Source	Asset	Owner	Access required by	Personal data	ORD potential
Mobile application	Pseudonymised participant activity data (measured in app)	Trial Participant	Trial Members	Yes (potentially sensitive)	No

Co-funded by the Horizon 2020 Framework Programme of the European Union

Source	Asset	Owner	Access required by	Personal data	ORD potential
Mobile application	User-generated media content created or modified in app	Trial Participant	Trial Members	Potentially	Yes
Media service component	Pseudonymised participant activity data (measured in service)	Trial Leader	Trial Members	Yes	No
Media service component	Service activity data	Trial Leader	Trial Members	No	Yes
Media service component	Media content created or modified in service	Trial Leader	Trial Members	Potentially	Yes
General platform component	Service activity data	Platform Operator	Trial Members	No	Yes
CLMC platform component	Aggregate activity database	Platform Operator	Trial Members	Yes	No
CLMC platform component	Knowledgebase	Platform Operator	Trial Members	No	Yes
Infrastructure slice	Traffic data	Platform Operator		Yes	No
Infrastructure monitor platform service	Pseudonymised traffic data	Platform Operator	Trial Members	Yes	No

Post-trial data is listed in Table 5.

Table 5. Data generated after a trial.

Asset	Owner	Access required by	Personal data	ORD potential
Participant feedback	Trial Participant	Trial Members	Yes	No
Aggregate and/or anonymised	Trial Leader	Trial Members	No	Yes

Co-funded by the Horizon 2020 Framework Programme of the European Union

Asset	Owner	Access required by	Personal data	ORD potential
participant feedback				
Analysis of trial data	Trial Members	Trial Members	No	Yes (likely)
Trial write-up	Trial Leader		No	Yes
Publications	Trial Members		No	Yes (open publication)

Co-funded by the Horizon 2020
Framework Programme of the European Union

9 SCENARIOS

9.1 PRINCIPLES

To aid discussion of the various datasets and their associated rights, we define a general abstract domain model for the stakeholders, datasets and contracts (Figure 13).

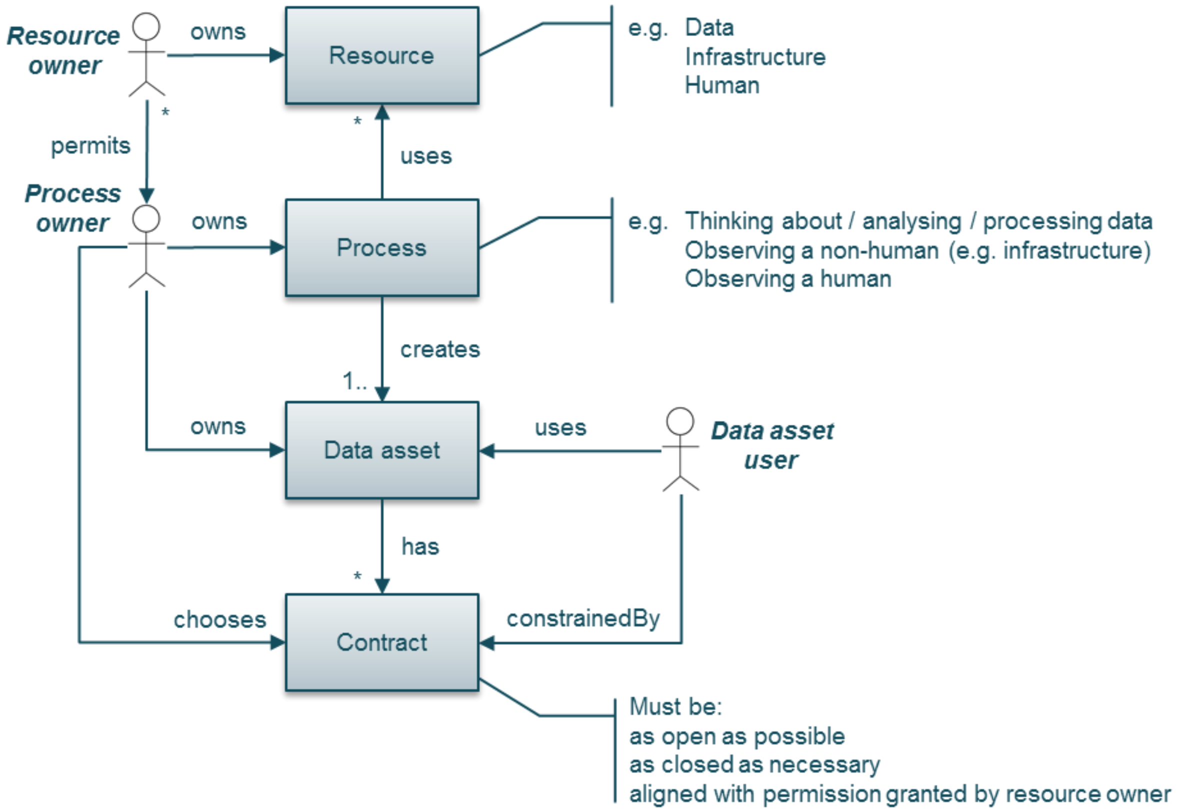


Figure 13. Abstract model relating stakeholders, datasets and contracts.

Key points:

- In the simple case, a data asset is owned by the owner of the process (cognitive or computing) which creates the data asset. However, joint ownership can arise through the combination of data and process.
- If the process uses another resource in the creation of the data asset then the owner of that resource may need to provide permission for it to be used and for the data created to be processed. The resource owner may also have some rights to determine how the data asset it shared with a third party.
- The form of permission granted by the *Resource* owner varies according to the resource (and may not be required).
- The *Contract* for the *data asset user* to use the data asset may take many forms (see below).

Co-funded by the Horizon 2020 Framework Programme of the European Union

Different trials may have different requirements for data sharing (depending on the IPR involved). Therefore we will treat each trial separately, though within a common framework. As the objective of the project is to better understand these multi-stakeholder FMI systems we propose that, *at a minimum*, all the Trial Members have access to all the trial's non-personal data for the purpose of better understanding the operation of the platform and its components. The Trial Members are the (Validation) Trial Leader, the Core Partners providing the platform services and the Platform Operator of the platform instance used by the trial (potentially an Open Call Replica Partner).

9.2 DERIVED DATA

In this section, a case where the data is derived from existing data is presented. Figure 14 shows the abstract model of derived data.

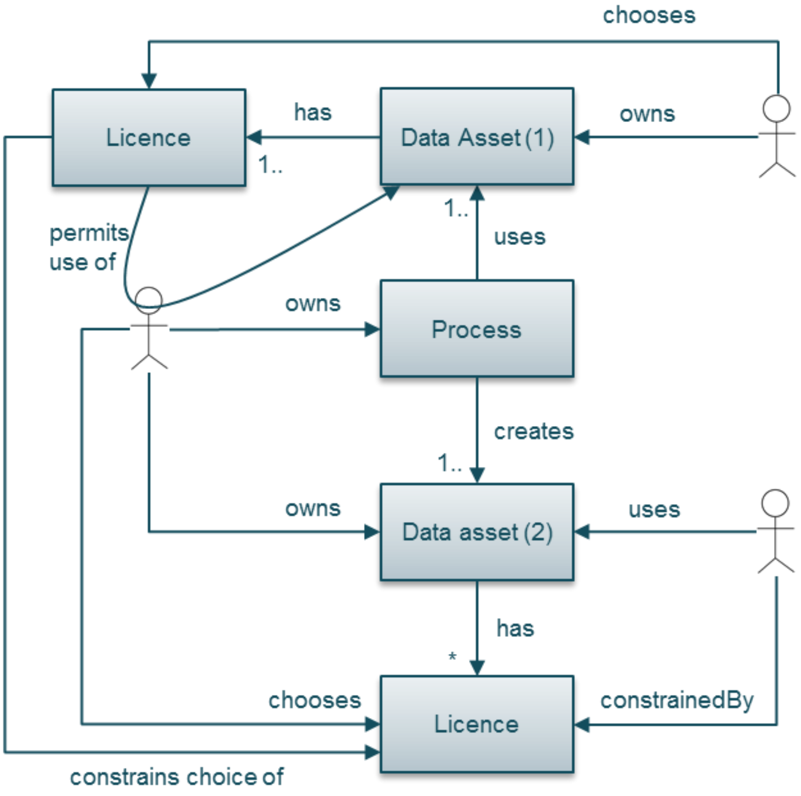


Figure 14. Abstract model of derived data.

Here:

- Data Asset (2) is derived from Data Asset (1)
- Licence terms of Data Asset (1) must permit this
- Licence of Data Asset (2) may depend on the licence of Data Asset (1)

A concrete example of this pattern is where some media content created for a trial is remixed by a trial participant into some new media content (Figure 15).

Co-funded by the Horizon 2020 Framework Programme of the European Union

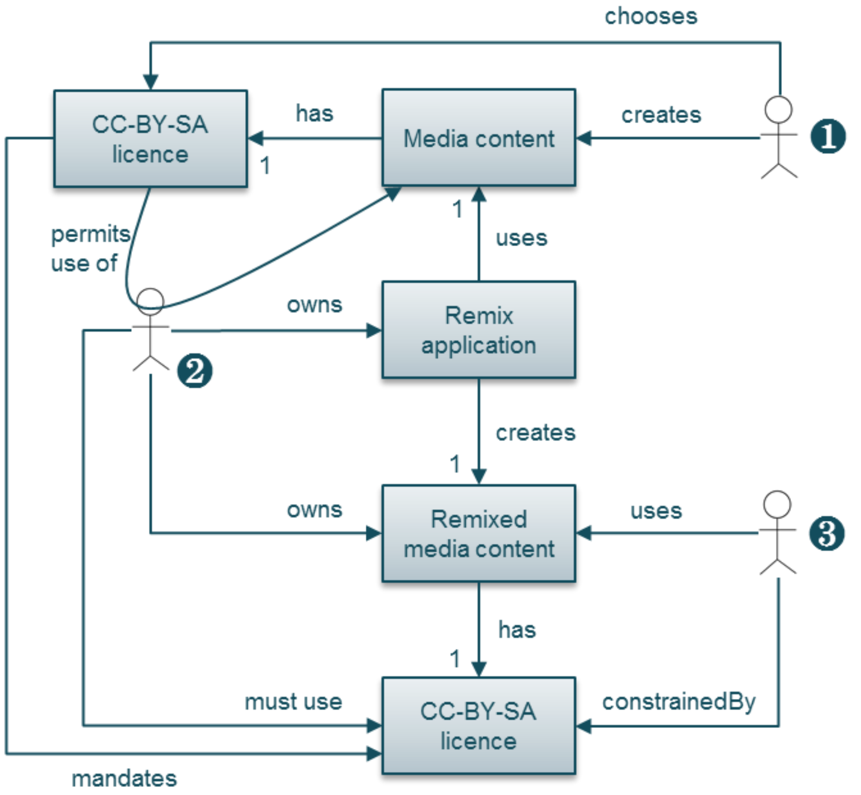


Figure 15. Concrete example of derived data.

In detail:

1. Actor (1), the Trial Leader, creates some media content and chooses to license it using CC-BY-SA.
2. Actor (2), the Trial Participant, takes the media content and remixes it using the mobile application of the trial.
3. Actor (2) is constrained by the CC-BY-SA terms and must license the remixed media content using the same licence.
4. Actor (3) is permitted to use the new remixed media content under the CC-BY-SA terms

9.3 OBSERVING AN IT PROCESS

To manage and optimise the operation of the FLAME platform there are many monitoring systems. In the most general abstract form, one process may monitor another process and the two process owners may be different actors (Figure 16).

Co-funded by the Horizon 2020 Framework Programme of the European Union

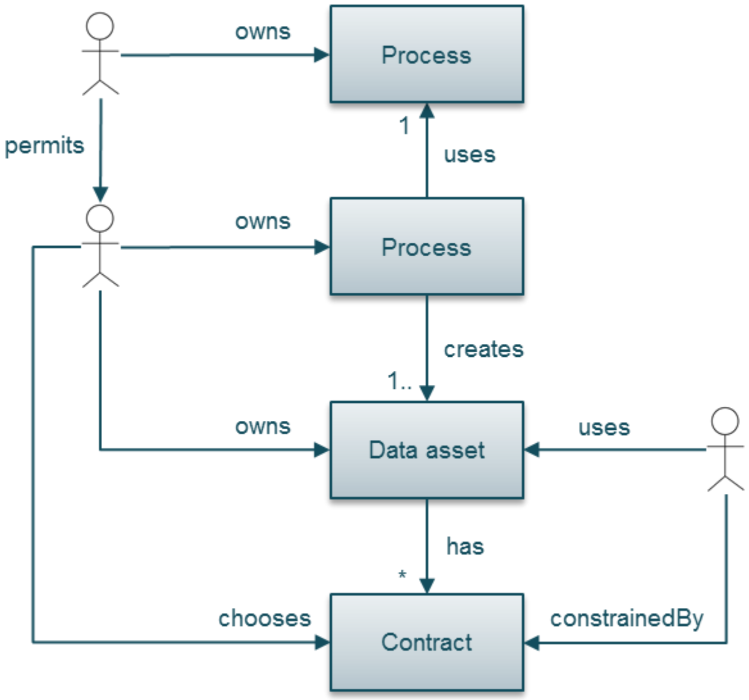


Figure 16. Abstract model of a process monitoring a process.

A concrete example of this situation would be where a Validation Trial Partner deploys a media service on the FLAME platform alongside a monitoring component which creates a log file (see Figure 17). In this case, the activity data is needed by the Platform Operator (2) to run the FLAME platform and so the Consortium Agreement requires the Validation Trial Partner (1) to share the data and governs the terms of that sharing. Similarly, if the media service was deployed by an Open Call Trial Partner, the data sharing would be governed by the Cascade Funding Agreement.

Co-funded by the Horizon 2020
Framework Programme of the European Union

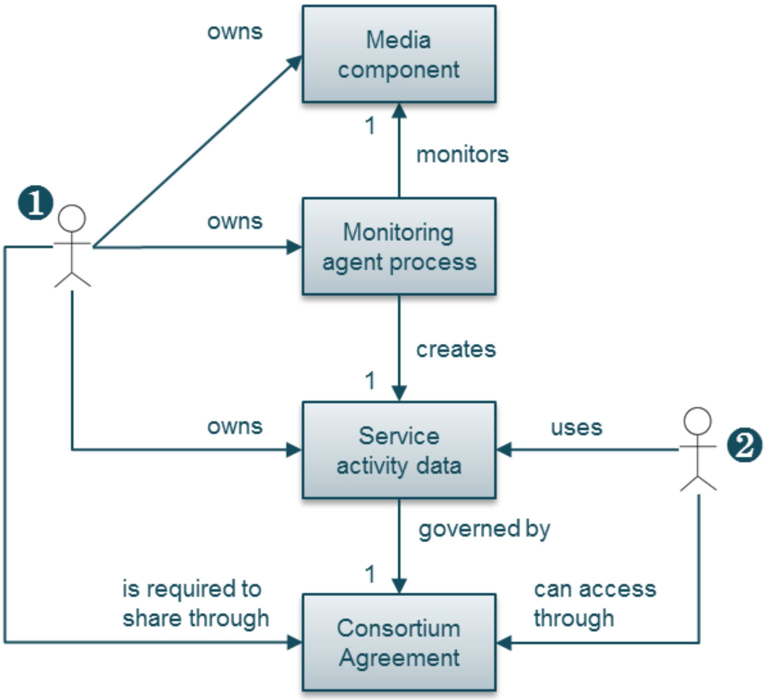


Figure 17. Concrete example of a process monitoring a process.

Building on the example above, Figure 18 below shows how the Platform Operator can add the service activity data arising from a media component deployed by an Open Call Trial Partner to the CLMC and updates the CLMC knowledgebase.

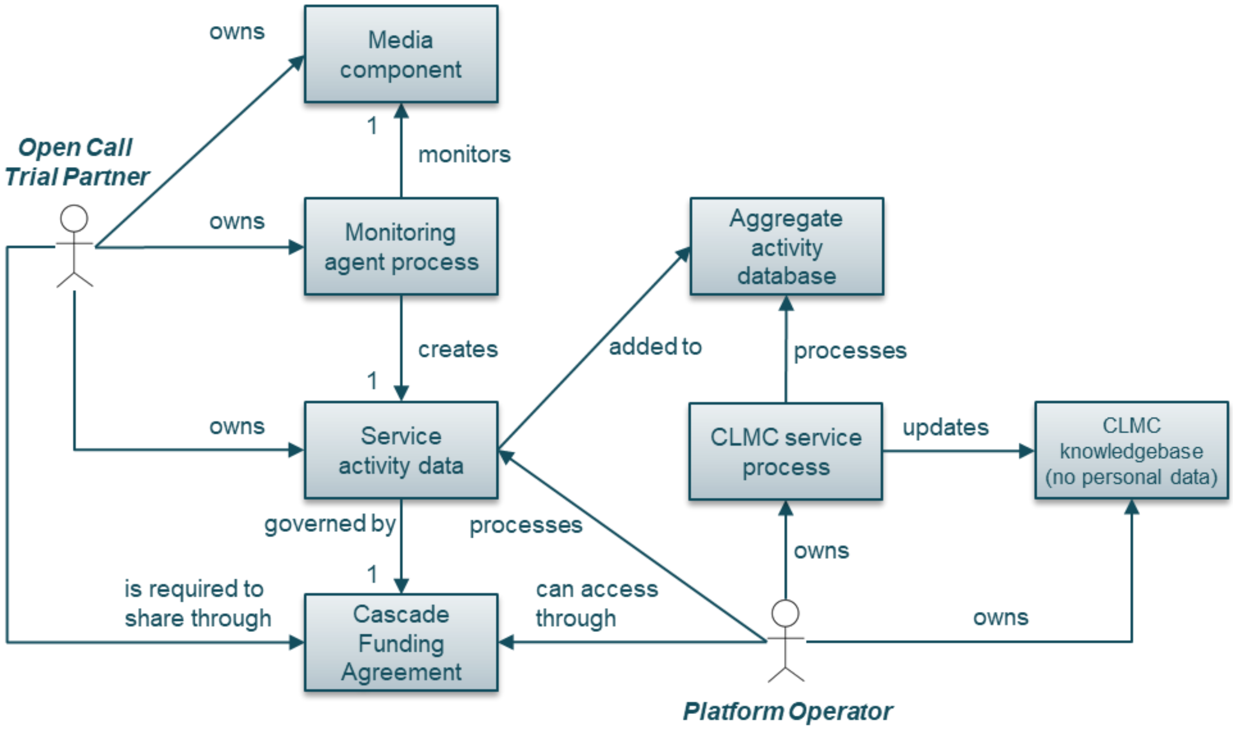


Figure 18. Sharing a media component's service activity log to update the CLMC knowledgebase.

Co-funded by the Horizon 2020 Framework Programme of the European Union

In this example we take the *service activity data* to be a log from a media component that does not include any personal data (so could be CPU and memory usage over time). The *service activity data* is added to all the other data flowing into the CLMC's aggregate activity database. The CLMC analyses all the data to create a knowledgebase which informs decisions.

Partners contributing data to the CLMC database will have joint ownership over the CLMC knowledgebase.

9.4 COLLECTING PERSONAL DATA

When collecting and processing personal data we are concerned with the stakeholders' responsibilities to protect and otherwise manage the data. In the abstract model in Figure 19 below we include the Data Subject (being observed by some process, either human or electronic), the Data Controller who owns the observation process and hence owns the resulting data asset and the Data Processor who is contracted by the Data Controller to process the data asset.

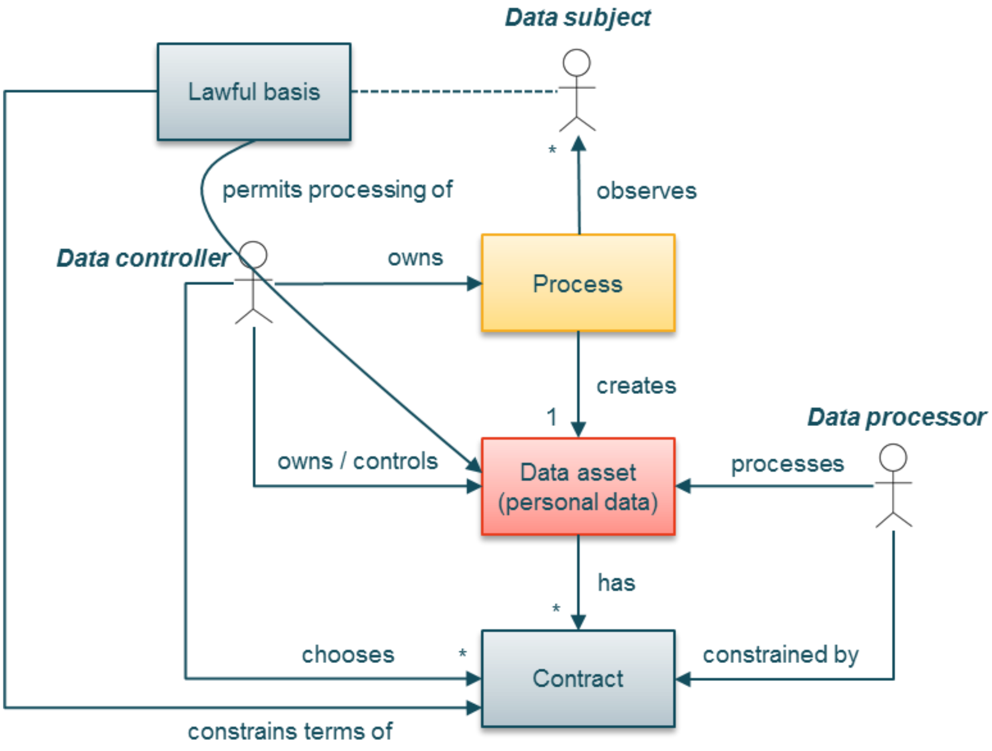


Figure 19. Abstract model of a process collecting personal data.

Key points:

- The *data controller* creates a *data asset* containing personal data using a *process* which observes the *data subject*. The process could be an IT process (such as location tracking in a mobile phone) or it could be the data controller manually observing the data subject: it does not matter.
- The *data controller* must have some *lawful basis* for processing the personal data of the *data subject* (i.e. one of the justifications given in Section 4.2.3). This lawful basis has some link to the *data subject*.

Co-funded by the Horizon 2020 Framework Programme of the European Union

- ➔ If another party needs to process the *data asset* then they are a *data processor* and must have a *contract* describing the data processing with the *data controller*.
- ➔ In the Figure, the personal data is highlighted in red and the process interacting with the personal data in yellow to draw attention to them as key assets to which regulations apply.

9.4.1 Data from the Application

In this concrete example depicted in Figure 20, the lawful basis for processing the Data Subject's data is that the Data Subject has signed a consent form. The Data Subject's interactions with a mobile application and its directly associated media service are recorded in an interaction log. The data is pseudonymised to reduce the risks of data breaches. More concretely, the Data Controller here could be a Trial Leader and the Data Processor could be the Platform Operator with the data being ingested into the CLMC (in a similar way to Figure 18 above).

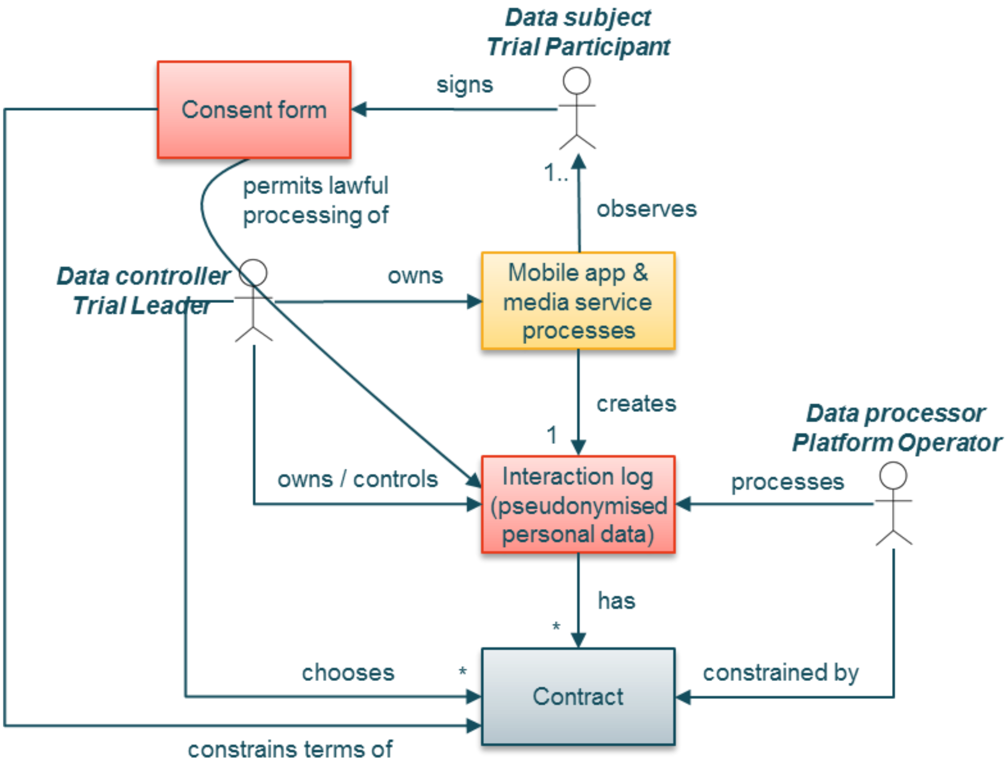


Figure 20. Concrete example of the collection of an interaction log containing personal data.

9.4.2 Data from the Network

Data from the infrastructure slice in which the platform operates, namely MAC addresses and IP addresses (described previously as "traffic data"), are personal data as noted in Recital 30 of the GDPR:

> Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them

Co-funded by the Horizon 2020 Framework Programme of the European Union

The UK's Information Commissioner's Office (ICO) has a relevant guide to Wi-Fi location analytics [ICO-WiFi] along with a concrete example of good practice from a study done by Transport for London [TFL-WiFi].

In FLAME, the Platform Operator operates the platform in an infrastructure slice obtained from the Infrastructure Operator. In the normal functioning of the infrastructure slice, users' MAC addresses or routing identifiers (uniquely identifying their devices) are transmitted over the network and logged. The Platform Operator has access to this data and can potentially use it data for additional purposes beyond the normal functioning of the network, for instance to measure and predict demand.

In a similar way to Figure 20 above, the Trial Participant will need to provide their consent for their traffic data to be processed for purposes beyond the normal functioning of the network infrastructure (see also the ePrivacy Directive, Section 4.3). In this case the data controller is the Platform Operator who is also the operator of the processes processing the data (Figure 21).

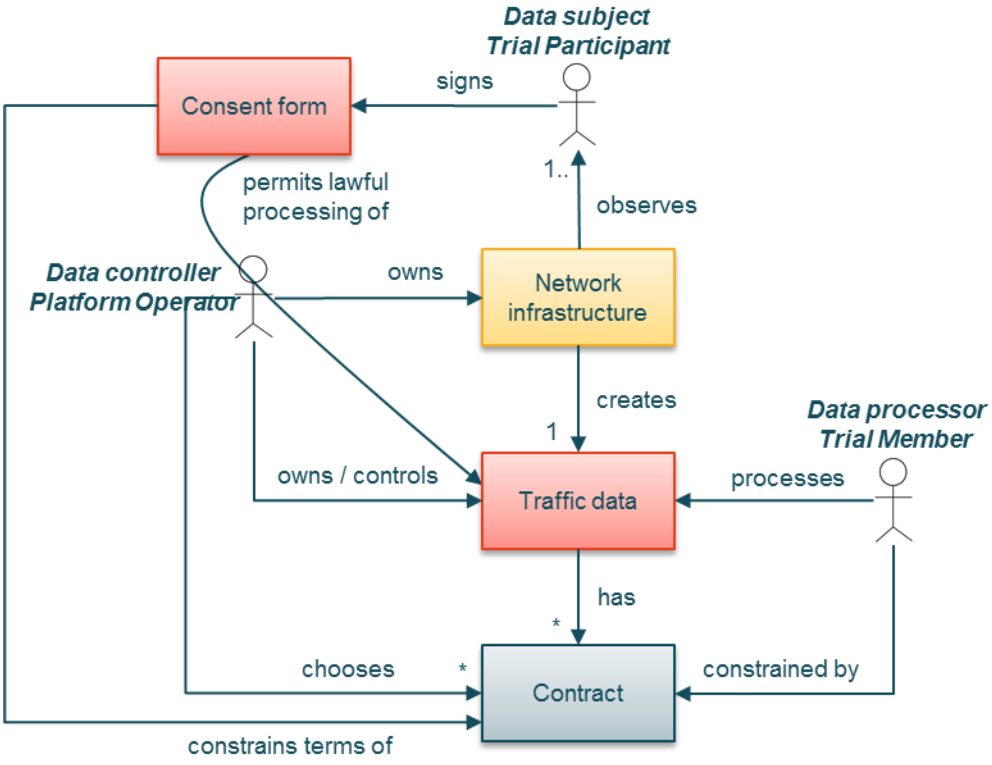


Figure 21. Concrete example of the collection and processing of traffic data.

To reduce the risk of any data breach, traffic data containing personal identifiers should be deleted or pseudonymised as early as possible. We recognise though that such pseudonymisation cannot occur at the operational level of the network however. It may be that keeping individual data points containing personal identifiers is not required. If only aggregate data is needed by other components and Trial Members then the situation is simpler as sharing and storing of personal data is then not required.

Co-funded by the Horizon 2020
Framework Programme of the European Union

9.5 CLMC OPERATION

Finally, we return to the overview of the data and processes surrounding the CLMC showing how personal data and non-personal data are combined in the CLMC (Figure 22).

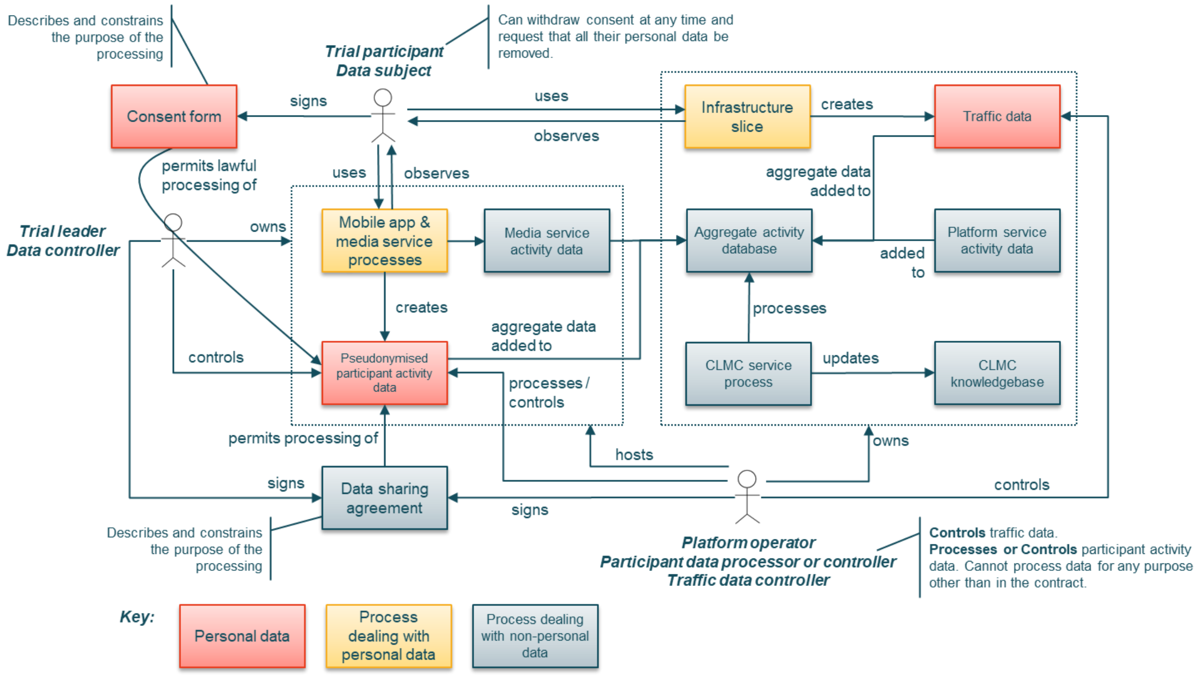

Describes and constrains the purpose of the processing
Trial participant
Data subject
Can withdraw consent at any time and request that all their personal data be removed.
Consent form
signs
uses
observes
Infrastructure slice
creates
Traffic data
permits lawful processing of
uses
observes
aggregate data added to
Trial leader
Data controller
owns
Mobile app & media service processes
Media service activity data
Aggregate activity database
added to
Platform service activity data
creates
aggregate data added to
processes
controls
Pseudonymised participant activity data
processes / controls
CLMC service process
updates
CLMC knowledgebase
permits processing of
hosts
owns
signs
Data sharing agreement
signs
controls
Describes and constrains the purpose of the processing
Platform operator
Participant data processor or controller
Traffic data controller
Controls traffic data.
Processes or Controls participant activity data. Cannot process data for any purpose other than in the contract.
Key:
Personal data
Process dealing with personal data
Process dealing with non-personal data

Figure 22. Overview of data, processes and stakeholders surrounding the CLMC.

Key points:

- The consent form signed by the Data Subject / Trial Participant permits the lawful processing of participant activity data collected through the application layer and traffic data collected through the network layer (this second link is not shown in the Figure).
- If possible, the participant's application will identify them using a pseudonym and this pseudonym will be used in any communication between the participant's device and the linked media service.
- The MAC address and IP address of the participant's device(s) will be observed by the infrastructure slice for the normal operation of the network. If these identifiers are required to be stored or shared they will be replaced with pseudonyms as early as possible in the platform services.
- The Trial Leader, as initiator of the media service owns the media service components used by the participant and therefore controls the participant activity data recorded.
- The Platform Operator owns the platform services (such as the CLMC) responsible for providing and managing the media service.
- The Platform Operator hosts the media services and is therefore a data processor for the personal data captured in the media services.

- The Trial Leader will have a data sharing agreement with the Platform Operator so that the Platform Operator may process the participant activity data as a data processor.

- The Platform Operator is the data controller for the traffic data recorded by the infrastructure.

- Aggregate participant activity data and aggregate traffic data will be added to the CLMC aggregate activity database along with (non-personal) activity data from the platform services and media services.

- The CLMC will process the data in the aggregate activity database and update the knowledgebase (with non-personal data).

- All the data from a trial will be available to all Trial Members for the purpose of verifying and validating the platform (who will sign a data controller/processor contract if they require access to personal data).

Co-funded by the Horizon 2020
Framework Programme of the European Union

10 CASCADE FUNDING AGREEMENT

The “Cascade Funding Agreement” for third parties joining the FLAME project through open calls has been referred to several times above. Each open call winner will be a single organisation and will sign a contract with ITINNOV (as project coordinator and cascade funding holder) to access the cascade funding. The discussion here is limited to just the data management aspects of the contract.

Constraints arising from this data management plan form some of the clauses in the Cascade Funding Agreement.

As stated in DoA:

- The standard contract will have to protect the intellectual property of third parties and beneficiaries involved, to guarantee the access to the FLAME platforms if a product must be designed.
- This standard contract will also protect the background IPR of the beneficiaries.
- Results are owned by the beneficiaries or third parties that generate them.
- Detailed IPR terms and conditions will be stated in the Consortium Agreement.

Regarding data management, the Cascade Funding Agreement defines:

- “Experimentation Results” to mean any tangible or intangible outputs of the Experimentation Project such as data knowledge or information, in whatever form or nature, whether it can be protected or not, that are generated by the Company in the Experimentation Project, as well as any IP Rights attached to it.
    - That the IP rights of the Experimentation Results shall be owned by the 3rd party.
- “FLAME Results” to mean any tangible or intangible outputs of hosting the Experimentation Project on the FLAME Platform such as data knowledge or information, in whatever form or nature, whether it can be protected or not, that are generated by the FLAME Platform as a result of hosting the Experimentation Project, as well as any IP Rights attached to it.
    - That the IP rights of the FLAME Results shall be owned by the replicator partner.
    - Access to the FLAME Results by other consortium partners is via the replicator partner through the terms of the Consortium Agreement, e.g. permitted where accesses is “needed”.
- Joint ownership of results, and how his situation will be governed using reasonable and non-discriminatory terms.
- Open research Data obligations on the 3rd Party (similar to Section 29.3 of the Grant Agreement).
- Data sharing (access) between cascade funding partners and core partners.
    - Rights for data generated.

Co-funded by the Horizon 2020 Framework Programme of the European Union

  - Rights over background data brought in to the project by Open Call Partner (such as media content).

- Confidentiality terms.

- Dissemination / publication terms.

10.1 FUNDED REPLICATORS

The discussion above applies to all cascade funding recipients but a variation on these terms is required for funded replicators in the second open call.

The primary variation is to ensure that the FLAME partners who have developed FLAME Platform services (IDE and ITINNOV) along with the developer of the Foundation Media Services (Atos with Martel) can have access to the "FLAME Results" (as described above).

10.2 UNFUNDED REPLICATORS

Another variation of the contract is required for unfunded replicators.

In this case, the changes are primarily to do with the changes to the funding and obligations of the unfunded replicator so are not relevant to this document.

Co-funded by the Horizon 2020
Framework Programme of the European Union

11 DATA MANAGEMENT PLAN TEMPLATES

The data management plan has two main purposes:

1) Documenting personal data to be captured to inform the necessary data sharing agreement with a replica.

2) Documenting experimental datasets (i.e. data generated in running an experiment or trial) for the purpose of providing "open data" where possible.

Each experimental project will need to complete a data management plan, initially in bare-bones form (just the mandatory questions shaded red) and then in complete form by the end of the trial. Information about other material created in a project (such as promotional material, reports or deliverables) is not relevant to this plan.

Three classes of data management plan are provided: Gold, Silver and Bronze. The distinction being that:

- for Gold all data is open,
- for Silver some data is open and
- for Bronze no data is open.

Those projects choosing Silver or Bronze must justify why the data will not be open.

The tables below indicate the questions to be answered along with whether an answer is mandatory (through red shading) for each data management plan class and propose answers along with guidance notes (in italics). Some questions for some DMP classes are marked "N/A" for "not applicable".

The Trial Leader (i.e. the 3rd-party organisation for open calls or the validation experiment partner) is responsible for writing and updating the data management plan and it will be reviewed by the 3rd Party Project Manager.

For clarity, once Gold, Silver or Bronze is chosen, the other columns should be deleted. Existing italicised comments should be deleted.

Project Summary	
Question	
Project name	
Lead partner	
Project replica	*e.g. Bristol, Barcelona, …*
Project start date	
Project end date	

Co-funded by the Horizon 2020 Framework Programme of the European Union

Organisations involved	*e.g. you, the replica partner, any others deeply participating*

Data Summary	
Question	
What is the purpose of the data collection/generation and its relation to the objectives of the project?	To understand the experience of the end users, the performance of the media, platform services and infrastructure and the relationships between them. *Plus any further purposes.*
What datasets (types and formats of data) will the project's trials generate/collect? Which datasets contain personal data?	Datasets with no personal data: • *e.g. Aggregate latencies* • *e.g. Other aggregate data from the CLMC* Datasets containing personal data: • *e.g. User activity logs in app linked to Google account* • *e.g. Photo or video media involving participants*
Will you re-use any existing data and how?	*If any external data is anticipated before the experiment starts, state it here. If any external data has been used during an experiment, it must be stated, along with any license terms or stipulations.*
What is the origin of the data?	
What is the expected size of the data?	*An estimate may be given initially and the actual size at the end.*
To whom might it be useful ('data utility')?	
Which data management plan class has been chosen?	*Gold, Silver, or Bronze*

FAIR: Making data findable, including provisions for metadata

Questions	**Gold**	**Silver**	**Bronze**
Are the data produced and/or used in the project's trials discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?	Yes, using DOI	Yes, using DOI	No
What naming conventions do you follow?	*Thinking about this at the start is recommended but optional.*		
Will search keywords be provided that optimize possibilities for re-use?			N/A

Co-funded by the Horizon 2020 Framework Programme of the European Union

Questions	Gold	Silver	Bronze
Do you provide clear version numbers?			
What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.	🟥	🟥	

Co-funded by the Horizon 2020 Framework Programme of the European Union

FAIR: Making data openly accessible

Question	Gold	Silver	Bronze
Which data produced and/or used in the project's trials will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.	All data	*State open datasets and which data will be closed with reasons*	*State reasons why all data will be closed*
How will the data be made accessible (e.g. by deposition in a repository)?	Deposition in Zenodo	Deposition in Zenodo	N/A
What methods or software tools are needed to access the data?			N/A
Is documentation about the software needed to access the data included?			N/A
Is it possible to include the relevant software (e.g. in open source code)?			N/A
Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.	Zenodo	Zenodo	N/A
Have you explored appropriate arrangements with the identified repository?			N/A
If there are restrictions on use, how will access be provided?	No restrictions	No restrictions	N/A
Is there a need for a data access committee?			N/A
Are there well described conditions for access (i.e. a machine readable license)?			N/A
How will the identity of the person accessing the data be ascertained?			N/A

The following sections are equally relevant for Gold and Silver classes but we would advise even those Trials adopting Bronze to consider the questions for two reasons (1) to better manage their data within their organisation and (2) in case the trial decides to publish some data at a later stage.

FAIR: Making data interoperable	
Question	**Gold & Silver**
Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to	Yes

Co-funded by the Horizon 2020 Framework Programme of the European Union

FAIR: Making data interoperable	
Question	**Gold & Silver**
standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?	
What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?	
Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?	
In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?	

FAIR: Increase data re-use (through clarifying licences)	
Question	**Gold & Silver**
How will the data be licensed to permit the widest re-use possible?	
When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.	
Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.	
How long is it intended that the data remains re-usable?	
Are data quality assurance processes described?	

Allocation of resources	
Question	**Gold & Silver**
What are the costs for making data FAIR in your project?	
How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).	

Co-funded by the Horizon 2020 Framework Programme of the European Union

Allocation of resources	
Question	**Gold & Silver**
Who will be responsible for data management in your project?	
Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?	

Data security	
Question	**Gold & Silver**
What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?	
Is the data safely stored in certified repositories for long term preservation and curation?	

Ethical aspects	
Question	**Gold & Silver**
Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).	
Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?	Yes

Other issues	
Question	**Gold & Silver**
Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?	

Co-funded by the Horizon 2020
Framework Programme of the European Union

12 CONCLUSIONS

This document broadly tackles two issues: that of the management of data within the FLAME project (including cascade funding partners) and the management of open data.

The principles of data management and of open data and detail regarding the (changing) legal and ethical framework in which the project must operate have all been described. FLAME will collect and process data which indirectly may be sensitive personal data and will therefore put in place the necessary strong controls on data management. FLAME will require trial participants to give their informed consent for the data collection and processing activities.

The project's stakeholders and their roles have been listed in order to describe the various legal contracts affecting data sharing between different partner types and the ownership and interest in the various datasets which will be produced in the project. The licences, stakeholders and datasets are all brought together in the scenarios section.

Regarding datasets which are suitable for publishing: a data management plan template has been presented for use by FLAME trials and a summary of suitable open data licences has been provided as well as a recommendation (Zenodo) for an open data repository.

Co-funded by the Horizon 2020 Framework Programme of the European Union

13 REFERENCES

[Aamot 2013] Aamot H, Kohl CD, Richter D, Knaup-Gregori P. Pseudonymization of patient identifiers for translational research. *BMC Medical Informatics and Decision Making*. 2013;13:75. doi:10.1186/1472-6947-13-75.

[CC] Creative Commons https://creativecommons.org/

[CC-BY] Creative Commons CC-BY licence https://creativecommons.org/licenses/by/4.0/

[CC-BY-SA] Creative Commons CC-BY-SA licence https://creativecommons.org/licenses/by-sa/4.0/

[Copyright] UK copyright legislation http://www.legislation.gov.uk/ukpga/1988/48/contents

[D1.1] FLAME confidential deliverable "NEC – Requirement No. 1"

[D1.2] FLAME confidential deliverable "NEC – Requirement No. 2"

[D3.3] FLAME public deliverable "FLAME Platform Architecture and Infrastructure Specification v1.0", https://www.ict-flame.eu/deliverables/

[DataCite] DataCite https://www.datacite.org/

[Dechamp 2016] Jean-François Dechamp, Open by default: the challenges of research data in Europe, European Commission, Directorate-General for Research & Innovation, @OpenAccessEC, 3rd LEARN Workshop, 28 June 2016, Helsinki

[DOI] Digital Object Identifier, https://www.doi.org/index.html

[Dryad Policies] Dryad Digital Repository – Policies. http://datadryad.org/pages/policies/

[Dryad] The Dryad Digital Repository. http://datadryad.org/

[Dublin Core] The Dublin Core Metadata Initiative (DCMI). http://dublincore.org/specifications/

[EDPB] Information note on data transfers under the GDPR in the event of a no-deal Brexit. https://edpb.europa.eu/our-work-tools/our-documents/other/information-note-data-transfers-under-gdpr-event-no-deal-brexit_en

[ePrints] University of Southampton ePrints http://eprints.soton.ac.uk/

[GDPR] GDPR law http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

[GPL] GNU Public Licence https://www.gnu.org/licenses/gpl-3.0.en.html

[ICO-GDPR] ICO GDPR guidelines https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr

Co-funded by the Horizon 2020 Framework Programme of the European Union

[ICO-GDPR-Contract] ICO GDPR Contract guidelines https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/contracts/

[ICO-WiFi] ICO Wi-Fi analytics guidelines https://ico.org.uk/media/for-organisations/documents/1560691/wi-fi-location-analytics-guidance.pdf

[ODC] Open Data Commons https://opendatacommons.org/

[Open] The open definition http://opendefinition.org/

[OpenAIRE] Open research Data Pilot https://www.openaire.eu/what-is-the-open-research-data-pilot

[TFL-WiFi] Transport for London Wi-Fi pilot http://content.tfl.gov.uk/review-tfl-wifi-pilot.pdf

[UKBrexit] Using personal data after Brexit https://www.gov.uk/guidance/using-personal-data-after-brexit

[UMA] User-Managed Access profile https://docs.kantarainitiative.org/uma/rec-uma-core.html

[Wilkinson 2016] FAIR data https://www.nature.com/articles/sdata201618

[Zenodo FAQ] The Zenodo FAQ. http://help.zenodo.org/

[Zenodo Policies] Zenodo Polices. http://about.zenodo.org/policies/

[Zenodo Principles] Zenodo Principles http://about.zenodo.org/principles/

[Zenodo Terms] Zenodo Terms of Use v1.0. http://about.zenodo.org/terms/

[Zenodo] The Zenodo open research data repository, https://www.zenodo.org/

Co-funded by the Horizon 2020
Framework Programme of the European Union