

# Data Management Plan – First Version

Public Report D8.7 Data Management Plan

This document contains preliminary information about the data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved.



JUNE 2020



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871477

# Foreword

The Deliverable 8.7, “Data Management Plan” contains preliminary information about the data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. The purpose of the Data Management Plan is to provide an analysis of the main elements of the data management policy that will be used by the consortium with regard to all the datasets that will be generated by the project. This deliverable also outlines how the research data collected, or generated, will be handled during and after the smashHit project, describes which methodology for data collection and generation will be followed, and whether and how data will be shared.

## Table of Contents

<b>1</b>	<b>Building a data management plan in the context of H2020 .....</b>	<b>5</b>
1.1	Purpose of the smashHit Data Management Plan .....	5
1.2	Background of the smashHit DMP .....	5
<b>2</b>	<b>Data Management Plan Overview .....</b>	<b>6</b>
<b>3</b>	<b>General Provisions .....</b>	<b>7</b>
3.1	Making Data Findable .....	7
3.1.1	Metadata .....	7
3.1.2	Persistent and unique identifier .....	7
3.1.3	Data discovery .....	7
3.1.4	Data identification .....	7
3.1.5	Keywords .....	7
3.2	Making Data Accessible .....	8
3.2.1	Projects datasets .....	8
3.2.2	Accessibility of files .....	8
3.2.3	Repositories .....	8
3.3	Making Data Interoperable .....	8
3.3.1	Data and metadata vocabularies .....	8
3.3.2	Mapping of non-common ontologies .....	8
3.4	Making Data Reusable .....	9
3.4.1	Data licensing .....	9
3.4.2	Third party access .....	9
3.4.3	Reasons for restriction .....	9
3.4.4	Long-term reusability .....	9
<b>4</b>	<b>Data generated in the smashHit Project .....</b>	<b>10</b>
<b>5</b>	<b>Ensuring Proper Commercialisation and IPR Protection of Generated Results .....</b>	<b>11</b>
<b>6</b>	<b>Datasets .....</b>	<b>12</b>
6.1	DS1 - Ontologies for Contracting and Consent Management .....	12
6.2	DS2 - Knowledge graphs with instance data representing the contracts .....	13
6.3	DS3 - Vehicle Data from ECUs .....	14
6.4	DS4 - Insurance customer facing consent flow .....	15
6.5	DS5 - Insurance customer facing vehicle data flow .....	16
6.6	DS6 - Vehicle information – static Vehicle Build .....	17
6.7	DS7 - Vehicle information – configurable Vehicle Build .....	18
6.8	DS8 - Vehicle information – Telematics .....	19

6.9	DS9 - Traffic information for the travellers .....	20
6.10	DS10 - Route suggestions .....	21
6.11	DS11 - Ferry arrival information.....	22
6.12	DS12 - End user mobility data .....	23
6.13	DS13 - Traffic light data .....	24
6.14	DS14 - City TMC - traffic information .....	25
6.15	DS15 - SmashHit Context Models .....	26

# 1 Building a data management plan in the context of H2020

---

## 1.1 Purpose of the smashHit Data Management Plan

smashHit is a Horizon 2020 project participating in the Open Research Data Pilot. This pilot is part of the Open Access to Scientific Publications and Research Data programme in H2020<sup>1</sup>. The goal of the programme is to foster access to data generated in H2020 projects.

Open Access to Research Data refers to the right to access and reuse digital research data under the terms and conditions set out in the Grant Agreement. In this way data become re-usable, and the benefit resulting from public investment in the H2020 research programme will be increased.

The EC provides a document with guidelines<sup>2</sup> for project that participate in the Open Research Data Pilot. The guidelines address aspects like research data quality, sharing and security. According to the guidelines, projects participating need to develop a Data Management Plan (DMP) within the first 6 months of the project.

The DMP describes the types of data that will be generated or gathered during the project, the standards that will be used, the ways how the data will be exploited and shared for verification or reuse, and how the data will be preserved.

This document has been produced following these guidelines and aims to provide a consolidated plan for smashHit partners that identifies the initial datasets and establishes the data management policies the project will follow. Therefore, the project partners intend to:

- Develop and keep up-to-date a Data Management Plan
- Deposit the project data in a research data repository
- Ensure third parties can freely access, mine, exploit, reproduce and disseminate published data
- Identify or provide the tools needed to validate the data

The industrial use cases apply to data and metadata needed to validate the scientific publications produced in the project and other data specified in the Data Management Plan.

## 1.2 Background of the smashHit DMP

The smashHit DMP is written in reference to Article 29.3 in the Model Grant Agreement called “Open access to research data” (research data management). Project participants must deposit their open data in a research data repository and take measures to make the data available to third parties. The third parties should be able to access, mine, exploit, reproduce and disseminate the data. This should also help to validate the results presented in scientific publications. In addition, Article 29.3 suggests that participants will have to provide information, via the repository, about any tools and instruments that might be made available by the project for the validation of the project outcomes.

The DMP will be important for tracking all data produced during the smashHit project. Article 29 states that project beneficiaries do not have to ensure access to parts of research data if such

---

<sup>1</sup> [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

<sup>2</sup> [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

access would undermine the achievement of project goals, for example with respect to exploitation or commercialisation of project outcomes. In such cases, the DMP must contain the reasons for not providing access.

## 2 Data Management Plan Overview

The DMP covers the complete research data cycle of the smashHit project as described in Figure 1 below. In Step 1 of the DMP, smashHit will produce raw data (generated through measurements and simulations, collected through market researching, etc.). The data will then be processed and analysed into more usable forms; i.e. reports, publishable documents, data tables, codes, etc. In Step 2, the data will be preserved using appropriate naming rules and metadata schemes. The project's open access policy will be applied to determine which datasets shall be made accessible (shared) for re-use in Step 3. The publicly accessible datasets will then be re-used by the public for verification.

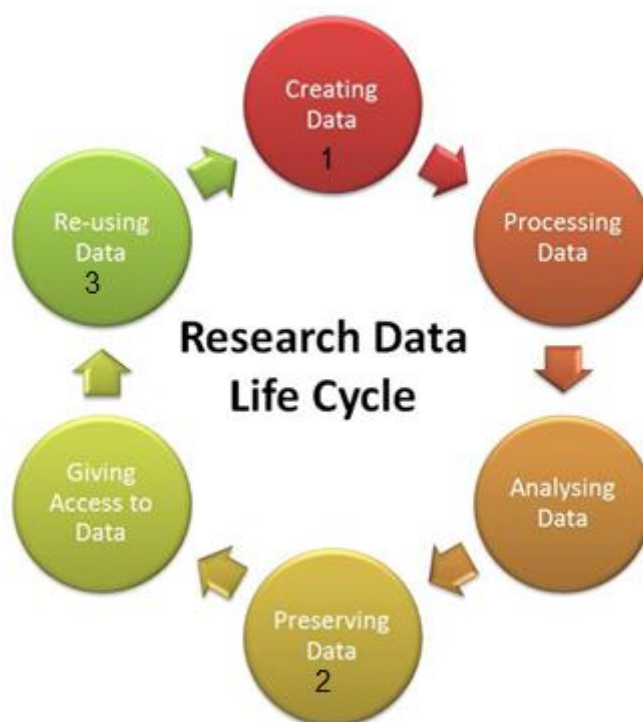


Figure 1 Research data life-cycle

This initial Data Management Plan is written during the first 6 months of the project, when little or no data have been produced. The overview therefore is a prediction of the datasets that are likely to be produced or utilised during the project lifetime. Information on their author, curator, access and relation to the workplan are given when possible.

### 3 General Provisions

---

A set of General Provisions have been established that provide the rules to ensure that the project datasets are **F**indable, **A**ccessible, **I**nteroperable, **R**eusable (FAIR). These are described in the following sections.

#### 3.1 Making Data Findable

##### 3.1.1 Metadata

Metadata are data on the research data themselves. They enable other researchers to find data in an online repository and is, as such, essential for the reusability of the dataset. By adding rich and detailed metadata, other researchers can better determine whether the dataset is relevant and useful for their own research. As described in the project Grant Agreement (Article 29.2), the bibliographic metadata should include all of the following:

- The terms “European Union (EU)” and “Horizon 2020”
- The name of the action, acronym and grant number
- The publication date, and length of embargo period, if applicable
- A persistent identifier

Note: All publications resulting from smashHit project must acknowledge the financial support of the EU by inclusion of the statement: “smashHit project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 871477.”

##### 3.1.2 Persistent and unique identifier

DOI and Creative Common’s license numbers will be used as persistent identifiers on open data repositories.

##### 3.1.3 Data discovery

All datasets will be referenced and indexed in the most common internet search engines. Datasets will be therefore immediately discoverable using simple web search capabilities. In addition, Open and Embargoed datasets will be discoverable on the designated open repositories for the project (see Section 3.2).

##### 3.1.4 Data identification

All Open and Embargoed datasets shall have a DOI. This DOI might be the same as the publication for which the data have been used for (e.g. technical publication), if the publisher requires data upload contextually to the publication.

Restricted and Closed datasets do not have a DOI, however, their metadata may have. The metadata shall clearly reference the dataset to which they refer.

##### 3.1.5 Keywords

All Open and Embargoed datasets shall be tagged with the keywords “H2020”, “smashHit”, “871477”, and at least one additional keyword indicative of the content of the dataset.



## 3.2 Making Data Accessible

### 3.2.1 Projects datasets

The smashHit project datasets will be first stored and organized in a database by the data owners (personal computer or on the organisation's secure server) and, where appropriate, on the project database. Some datasets, for which the Consortium declares no confidentiality or IPR issues, will be also stored in Zenodo, the open access repository of the Open Access Infrastructure for Research in Europe (OpenAIRE). In such cases, data access policy will be unrestricted. An embargo period may be incurred if collected datasets are linked to green open access publication. Access levels to datasets are specified as:

- O = Open
- E = Embargo, followed by expiry date
- R = Restricted
- C = Closed

### 3.2.2 Accessibility of files

Most data files are accessible with common and free software:

- .png, .tif, .jpg, .raw: Any image viewer such as XnView, IrfanView, GIMP etc.
- .pdf: Acrobat reader
- .xls, .xlsx: Google sheets, LibreOffice Calc
- .doc, .docx, .odt, .txt Google docs, LibreOffice Writer
- .xml, .csv, .sql: any text editor, or database import tools/viewers.

Additional filetypes that may be agreed during the progress of the project will be described accordingly within the smashHit services which produce or use them.

### 3.2.3 Repositories

All datasets that will be produced within the smashHit project, and for which the Consortium declares no confidentiality or IPR issues, will be uploaded to Zenodo, with their respective access rights.

## 3.3 Making Data Interoperable

### 3.3.1 Data and metadata vocabularies

The Consortium will seek to use standard vocabularies within the datasets that are widely accepted by the respective communities. Controlled vocabularies should be used in descriptive metadata fields to support consistent, accurate, and quick indexing and retrieval of relevant data. Keywords (see section 3.1.5) and their synonyms will be used for indexing and subject headings of the data and metadata.

### 3.3.2 Mapping of non-common ontologies

Non-common, non-obvious ontologies should be explained in the metadata of the individual datasets.



## **3.4 Making Data Reusable**

### **3.4.1 Data licensing**

Creative Common Licensing will be used to protect the ownership of the datasets. Both Share-Alike and NonCommercial-ShareAlike licenses will be considered for the parts of datasets that can be made publicly available.

### **3.4.2 Third party access**

Third parties must request access rights to Restricted and Closed data according to the procedure indicated in the respective metadata. Access rights to third parties will be evaluated by the data owner on a case-by-case basis. In general, third parties will have to state the purpose concerning why access rights are needed, and sign a Non-Disclosure Agreement.

### **3.4.3 Reasons for restriction**

Restricted and closed data are raw experimental results which still need to undergo rigorous evaluation, data which might undermine the beneficiaries' IP protection and commercialisation strategies if published, or data containing sensitive information.

### **3.4.4 Long-term reusability**

Unless otherwise indicated, all Open and Embargoed data will be available for indefinite time and curated until two years after project conclusion. Restricted and Closed data will be available and curated until two years after project conclusion.

## 4 Data generated in the smashHit Project

smashHit will produce data in a wide range of R&D activities and the expected datasets are summarized in Table 1 below. This list is considered as initial and will be modified (addition or removal of datasets) as the project evolves, however, any deviations will be described as required in the coming deliverables. Once generated (or collected), these datasets will be stored in several formats, which are: Documents, Images, Data and Numerical Codes.

Table 1: Data generated.

No.	Dataset reference and name	Main partners involved
1	DS1 - Ontologies for Contracting and Consent Management	UIBK
2	DS2 - Knowledge graphs with instance data representing the contracts	UIBK, L3S
3	DS3 - Vehicle Data from ECUs	VW
4	DS4 - Insurance customer facing consent flow	LexisNexis
5	DS5 - Insurance customer facing vehicle data flow	LexisNexis
6	DS6 - Vehicle information – static Vehicle Build	LexisNexis
7	DS7 - Vehicle information – configurable Vehicle Build	LexisNexis
8	DS8 - Vehicle information – Telematics [Mileage; ADAS sensor trigger/state; UBI; PAYD; PHYD]	LexisNexis
9	DS9 - Traffic information for the travellers	INFT
10	DS10 - Route suggestions	INFT
11	DS11 - Ferry arrival information	INFT
12	DS12 - End user mobility data	INFT
13	DS13 - Traffic light data	INFT
14	DS14 - City TMC - traffic information	INFT
15	DS15 - SmashHit Context Models	ATB

All datasets described in Table 1 above are described in detail under Section 6.

In addition to the datasets in Table 1, the following public outputs as software programs will be made “Open Access” (to be provided free of charge for public sharing) and, as such, the software code datasets are included in the Open Research Data Pilot, and will thus be managed according to the present Data Management Plan:

- D5.4 – Public optimised smashHit Framework (INFT)
- D6.5 – Demonstrator of Services Using Integrated CPP and Insurance Data (VW)
- D7.5 – Demonstrator of Services Using Integrated Traffic, Smart City and CPP Data (FVH)

Other technological datasets may be identified and made available for open access during the operation of the project.

## 5 Ensuring Proper Commercialisation and IPR Protection of Generated Results

For all data types, the Consortium will examine the aspects of potential conflicts against commercialisation and the IPR protection issues of the knowledge generated before deciding which information needs to be made public and when. The decision process, summarized in Figure 2 below, will be overseen by the Exploitation Manager.

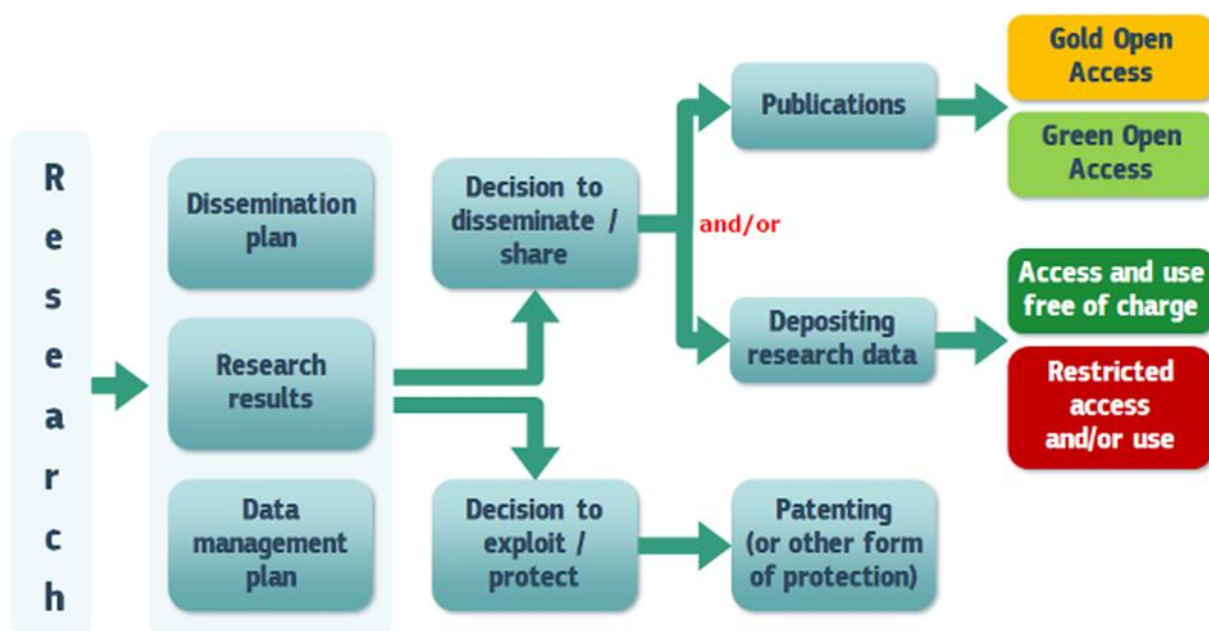


Figure 2: Open access to research data and publication decision diagram (from Guidelines to the Rules on Open Access to Scientific publications and Open Access to Research Data in Horizon 2020)

The Grant Agreement (Article 29.3) states “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 would be jeopardized by making those specific parts of the research data openly accessible.” Such an exception applies to the smashHit project when the project findings present high innovation levels (possibility of commercialisation). In this case, the consortium will consider two forms of protection:

- to withhold the data for internal use, or
- to apply for a patent or other IP protection measure in order to commercially exploit the invention and have in return financial gain.

In the former case, appropriate IPR protection measures (e.g. Non-Disclosure Agreement) must be taken for data sharing outside the consortium. In the latter case, publications will be delayed until the patent filing or other IP protection measures are completed. Otherwise, the results will be made “Open Access” by depositing the research data into the designated Zenodo repository service or by publishing in journals adhering to suitable “Open Access” (‘green’ or ‘gold’). In parallel, public deliverables will be stored on the following locations:

- the smashHit project website;
- the smashHit page on CORDIS website where all public deliverables submitted to the European Commission are hosted.

## 6 Datasets

The following overview of the Data Generated in the smashHit Project is a prediction of those datasets that are likely to be produced during the project lifetime. Information on their author, curator, access, relation to the workplan etc. are given when possible.

### 6.1 DS1 - Ontologies for Contracting and Consent Management

<b>Dataset reference and name</b>	DS1 - Ontologies for Contracting and Consent Management
<b>Purpose and relation to the objectives of the project</b>	The dataset will comprise the schemas (ontologies, vocabularies) that will model and represent the contractual and consent data management aspects, particularly, for sensor data.
<b>Data types</b>	.rdf, .rdfs, .owl, .json
<b>File formats</b>	files, query end points (e.g. for SPARQL)
<b>Reuse of existing data</b>	Other ontologies and vocabularies will be reused.
<b>Data production methods</b>	Manual, semi-automatic
<b>Data utility</b>	Servers, file systems, graph database management
<b>Potential for reuse</b>	Very high, in the project and beyond the project
<b>Usage rights</b>	Open data
<b>Data retention</b>	t.b.d.
<b>Curator</b>	UIBK

## 6.2 DS2 - Knowledge graphs with instance data representing the contracts

<b>Dataset reference and name</b>	DS2 - Knowledge graphs with instance data representing the contracts
<b>Purpose and relation to the objectives of the project</b>	Instance data produced as a result of automated contracting environment, including the details on consent, traceability, etc.
<b>Data types</b>	.rdf, .rdfs, .owl, .json
<b>File formats</b>	files, query end points (e.g. for SPARQL).
<b>Reuse of existing data</b>	Other ontologies and vocabularies will be reused.
<b>Data production methods</b>	Automatic, semi-automatic, manual
<b>Data utility</b>	Servers, file systems, graph database management
<b>Potential for reuse</b>	Very high, useful and usable in many applications
<b>Data retention</b>	t.b.d.
<b>Usage rights</b>	Closed data, may be made available also for third-parties upon permissions
<b>Curator</b>	UIBK, L3S

### 6.3 DS3 - Vehicle Data from ECUs

<b>Dataset reference and name</b>	DS3 - Vehicle Data from ECUs
<b>Purpose and relation to the objectives of the project</b>	Gathered vehicle and driving data from several electronic controls units to compute scores for vehicle insurance quotes (BC1)
<b>Data types</b>	.json, .csv
<b>File formats</b>	files
<b>Reuse of existing data</b>	Existing data might serve as sandbox data for test rack emulation
<b>Data production methods</b>	Manual, semi-automatic
<b>Data utility</b>	Servers, file systems
<b>Potential for reuse</b>	May not be reused
<b>Data retention</b>	t.b.d.
<b>Usage rights</b>	Restricted and / or closed
<b>Curator</b>	VW

#### 6.4 DS4 - Insurance customer facing consent flow

<b>Dataset reference and name</b>	DS4 - Insurance customer facing consent flow
<b>Purpose and relation to the objectives of the project</b>	Consent authentication by Nat. ID or Passport or if not possible by fName, IName, DoB, Current Address. Validation of consent.
<b>Data types</b>	Text string, captured on Insurers website and passed to LexisNexis.
<b>File formats</b>	Webservice between insurers and LexisNexis.
<b>Reuse of existing data</b>	This could be a new enquiry or an insurance renewal.
<b>Data production methods</b>	Data is provided by the end customer, on the insurers website* and sent through the insurer to LexisNexis (*Terms and conditions plus related consent material available within the site)
<b>Data utility</b>	Requires consent, to be collected by the insurer for onward validation.
<b>Potential for reuse</b>	No reuse – this is a one-time transactional event that requires validation.
<b>Data retention</b>	Data is not stored, only transactional
<b>Usage rights</b>	Closed data
<b>Curator</b>	LexisNexis



## 6.5 DS5 - Insurance customer facing vehicle data flow

<b>Dataset reference and name</b>	DS5 - Insurance customer facing vehicle data flow
<b>Purpose and relation to the objectives of the project</b>	Vehicle VIN or VRN relating to the insurance policy. Plus details of other 'named drivers' for the vehicle. Linking of VIN/VRN to existing vehicle records, and if no records exist vehicle data retrieved from third party data providers. Return of data would be subject to valid consent.
<b>Data types</b>	Alphanumeric string
<b>File formats</b>	Web service between insurers and LexisNexis
<b>Reuse of existing data</b>	This could be a new enquiry or an insurance renewal.
<b>Data production methods</b>	Data is obtained from a third party data provider and provided to the insurer by LexisNexis. The origin of the data may be the OEM itself (vehicle build data) or data generated by the driver (mileage & telematics).
<b>Data utility</b>	Legitimate interest of the insurer may apply depending on the dataset requested. Where consent is required, this needs to be collected from the end-customer.
<b>Potential for reuse</b>	Data may be stored and reused for insurance use cases
<b>Data retention</b>	Data may be retained based on the legal regulation of the insurance sector of each market (audit purposes)
<b>Usage rights</b>	Closed data
<b>Curator</b>	LexisNexis

## 6.6 DS6 - Vehicle information – static Vehicle Build

<b>Dataset reference and name</b>	DS6 - Vehicle information – static Vehicle Build
<b>Purpose and relation to the objectives of the project</b>	The exact vehicle specification (especially ADAS features) allows the insurer to provide a more precise and accurate insurance quotation that takes account of the reduced risk.
<b>Data types</b>	Vehicle build information linked to the unique VIN. Raw data retrieved from the OEM systems and converted into a normalized set of ADAS Features and a numeric ADAS rating by LexisNexis.
<b>File formats</b>	Various – mainly JSON; CSV
<b>Reuse of existing data</b>	Subject to data contract, raw data or derived product data retained by LexisNexis.
<b>Data production methods</b>	The 'raw data' is ingested by LexisNexis and run through a process of normalisation to create a consistent output file listing the Vehicle Build features and a Rating Indicator.
<b>Data utility</b>	Creation of an actionable insight data product from the raw build data.
<b>Potential for reuse</b>	Where the data is 'static' it is possible to reuse it subject to appropriate consent.
<b>Data retention</b>	Default to 12-years but can be set per data source / local market.
<b>Usage rights</b>	Closed data
<b>Curator</b>	LexisNexis – proprietary product containing I.P. / Trade Secret information.

## 6.7 DS7 - Vehicle information – configurable Vehicle Build

<b>Dataset reference and name</b>	DS7 - Vehicle information – configurable Vehicle Build
<b>Purpose and relation to the objectives of the project</b>	The exact vehicle specification (especially ADAS features) allows the insurer to provide a more precise and accurate insurance quotation that takes account of the reduces risk. Dynamic data can confirm any over-the-air updates and remote deployment of features.
<b>Data types</b>	Vehicle build information linked to the unique VIN. Raw data retrieved from the OEM systems and converted into a normalized set of ADAS Features and a numeric ADAS rating by LexisNexis. Data requested from OEM source in real-time.
<b>File formats</b>	Various – mainly JSON
<b>Reuse of existing data</b>	Due to the dynamic nature of this information, it must be re-freshed.
<b>Data production methods</b>	The 'raw data' is ingested by LexisNexis and run through a process of normalisation to create a consistent output file listing the Vehicle Build features and a Rating Indicator – accurate in real-time.
<b>Data utility</b>	Creation of an actionable insight data product from the raw build data.
<b>Potential for reuse</b>	None.
<b>Data retention</b>	Retained in log files for support and analysis only.
<b>Usage rights</b>	Closed data
<b>Curator</b>	LexisNexis – proprietary product containing I.P. / Trade Secret information.

## 6.8 DS8 - Vehicle information – Telematics

<b>Dataset reference and name</b>	DS8 - Vehicle information – Telematics [Mileage; ADAS sensor trigger/state; UBI; PAYD; PHYD]
<b>Purpose and relation to the objectives of the project</b>	<p>Data points are collected from the vehicle at the start, during and at completion of a journey, and transmitted to LNRS.</p> <p>LNRS transform and enrich the data points to derive products relevant to their B2B markets. Customers may use the products in the pricing and management of insurance policies as well as the evaluation of extreme events such as accidents and harsh driving behaviour.</p>
<b>Data types</b>	Vehicle safety and ADAS sensor data, GPS data, CANBus data and unique identifiers applied by the OEM to the file headers to ensure the integrity of the data. Generally transmitted to LNRS as ASCII based JSON, XML or .CSV
<b>File formats</b>	Dependant on the OEM, primarily JSON or XML
<b>Reuse of existing data</b>	The data may be stored by LNRS and used to process scores that are evaluated over agreed periods of time
<b>Data production methods</b>	Vehicle telemetry data is transmitted from the data provider (OEM or OEM's platform provider) and ingested by LexisNexis. The data is then passed through a process of normalisation before being enriched and derived into digitised data products. The products are either requested by the Customer at the point of need, or supplied to the customer as a scheduled batch. This processes are recorded for billing. The format of this data may vary and may include JSON, XML or .CSV files
<b>Data utility</b>	t.b.d.
<b>Potential for reuse</b>	The data owner may request to have their data re-used as part of the insurance renewal process or as part of an insurance claim
<b>Data retention</b>	<p>Saleable (PII linked) data as part of a policy based product - retained for up to 7 years as part of legislative procedures</p> <p>Anonymized data – retained indefinitely for internal R&amp;D / evaluation purposes</p>
<b>Usage rights</b>	Closed data
<b>Curator</b>	LexisNexis – derived proprietary products containing I.P. / Trade secret information.

## 6.9 DS9 - Traffic information for the travellers

<b>Dataset reference and name</b>	DS9 - Traffic information for the travellers
<b>Purpose and relation to the objectives of the project</b>	The recommended routes for personal cars as well as to bigger commercial vehicles and buses. Traveller can also see if there is any congestion, incidents, warnings or roadworks on one's route options. The location of the nearest public transport stops as well as the next departures (departure time, line number, destination) on each stop can be seen. Also the possible disruptions of the relevant lines are shown. The nearest locations of the City bike stations and availability of the bikes on each station can also be seen.
<b>Data types</b>	.json, .csv
<b>File formats</b>	files
<b>Reuse of existing data</b>	Historical data can be used for statistics
<b>Data production methods</b>	Integrating existing data sources, from e.g. public transport, Helsinki City open data, Finnish Transport Infrastructure Agency open data.
<b>Data utility</b>	Can be used to provide real time information and statistics
<b>Potential for reuse</b>	City may use for statistical purposes
<b>Data retention</b>	t.b.d. Data will be stored in INFT server at least to the end of the project
<b>Usage rights</b>	t.b.d. Probably open data (CC BY 4.0)
<b>Curator</b>	INFT

## 6.10 DS10 - Route suggestions

<b>Dataset reference and name</b>	DS10 - Route suggestions
<b>Purpose and relation to the objectives of the project</b>	To help drivers make their trips safely, easily and without unnecessary delays. When approaching the harbour area drivers want to choose the best timing as well as route for the destination. The residents of Jätkäsaari area know the traffic can be jammed. Residents would get information on the potential traffic jams conveniently before they hit the road by car or bike.
<b>Data types</b>	.json, .csv
<b>File formats</b>	files
<b>Reuse of existing data</b>	No reuse
<b>Data production methods</b>	Software
<b>Data utility</b>	Usable for routing purposes
<b>Potential for reuse</b>	No reuse.
<b>Data retention</b>	Data is not stored, only transactional
<b>Usage rights</b>	Closed data
<b>Curator</b>	INFT

## 6.11 DS11 - Ferry arrival information

<b>Dataset reference and name</b>	DS11 - Ferry arrival information
<b>Purpose and relation to the objectives of the project</b>	The city TMC needs to be aware when the ferry will arrive on quayside and the vehicles start to enter the street network. As exact prediction as possible is needed to prepare e.g. the traffic management processes to minimize the negative impact for the overall traffic.
<b>Data types</b>	.json, .csv
<b>File formats</b>	files
<b>Reuse of existing data</b>	No reuse.
<b>Data production methods</b>	Sensors / Software
<b>Data utility</b>	Restricted, industrial data
<b>Potential for reuse</b>	No reuse.
<b>Data retention</b>	t.b.d. May be used for statistical purposes
<b>Usage rights</b>	Restricted
<b>Curator</b>	INFT



## 6.12 DS12 - End user mobility data

<b>Dataset reference and name</b>	DS12 - End user mobility data
<b>Purpose and relation to the objectives of the project</b>	All stakeholders need more and better traffic information on the Jätkäsaari area. The information quality can be improved by taking new data sources into use: data from the cars moving in the area as well as end users can provide valuable observations and mobility data when equipped with appropriate mobile crowdsourcing applications. Residents who are familiar with the area and know where and when something exceptional probably happens would be a very valuable source of information for other stakeholders.
<b>Data types</b>	.json, .csv
<b>File formats</b>	files
<b>Reuse of existing data</b>	Historical data can be used for statistics
<b>Data production methods</b>	e.g. mobile application
<b>Data utility</b>	Can be used to provide real time information and statistics
<b>Potential for reuse</b>	City may use for statistical purposes
<b>Data retention</b>	t.b.d. Data will be stored in INFT server at least to the end of the project
<b>Usage rights</b>	t.b.d. Probably open data (CC BY 4.0)
<b>Curator</b>	t.b.d INFT or city

### 6.13 DS13 - Traffic light data

<b>Dataset reference and name</b>	DS13 - Traffic light data
<b>Purpose and relation to the objectives of the project</b>	The data from the traffic signal system will be investigated to improve the understanding of traffic situation in the Jätkäsaari area.
<b>Data types</b>	.json, .csv
<b>File formats</b>	files
<b>Reuse of existing data</b>	Historical data can be used for statistics
<b>Data production methods</b>	Traffic light control system
<b>Data utility</b>	Restricted, industrial data
<b>Potential for reuse</b>	City may use for statistical purposes
<b>Data retention</b>	Raw data is not stored, only transactional. Refined data is stored in INFT server
<b>Usage rights</b>	Closed data
<b>Curator</b>	INFT

#### 6.14 DS14 - City TMC - traffic information

<b>Dataset reference and name</b>	DS14 - City TMC - traffic information
<b>Purpose and relation to the objectives of the project</b>	The city TMC needs to inform various stakeholders if some incidents are detected or expected in the traffic. The incidents may be planned or unplanned. City TMC will use all the data sources and information that they have access to, when they are making the decision whether to publish the information or not.
<b>Data types</b>	.json
<b>File formats</b>	DATEXII
<b>Reuse of existing data</b>	Historical data can be used for statistics
<b>Data production methods</b>	Professionally operated TIC (traffic information collector) tool
<b>Data utility</b>	Can be used to provide real time information and statistics
<b>Potential for reuse</b>	City may use for statistical purposes
<b>Data retention</b>	Data will be stored in INFT server
<b>Usage rights</b>	Open data (CC BY 4.0)
<b>Curator</b>	INFT

## 6.15 DS15 - SmashHit Context Models

<b>Dataset reference and name</b>	DS15 - SmashHit Context Models
<b>Purpose and relation to the objectives of the project</b>	The smashHit context models will describe context information for contracts, contract compliance and data traceability as well as relations between this information. The context models will form the basis for the context sensitive security and monitoring & extraction modules.
<b>Data types</b>	Context models are ontologies based on standards defined by the W3C.
<b>File formats</b>	Context models will be in the formats: RDF/XML, OWL/XML
<b>Reuse of existing data</b>	Cross-CPP context models
<b>Data production methods</b>	Software
<b>Data utility</b>	The context models are usable to any individual using the smashHit platform
<b>Potential for reuse</b>	The smashHit context models will be used in future research projects
<b>Data retention</b>	Retained in the ATB data storage centre
<b>Usage rights</b>	Open Data - will be published as open source under Apache 2.0 license
<b>Curator</b>	ATB



## About smashHit

The objective of smashHit is to assure trusted and secure sharing of data streams from both personal and industrial platforms, needed to build sectorial and cross-sectorial services, by establishing a Framework for processing of data owner consent and legal rules and effective contracting, as well as joint security and privacy preserving mechanisms. The vision of smashHit is to overcome obstacles in the rapidly growing Data Economy which is characterized by heterogeneous technical designs and proprietary implementations, locking business opportunities due to the inconsistent consent and legal rules among different data-sharing platforms actors and operators. The Framework will provide methods and tools, such as smashHit Dispatcher, to assure common consent over data shared using semantic models of consent and legal rules. The new tools include traceability of use of data, data fingerprinting and automatic contracting among the data owners, data providers, service providers and users. These tools are specifically critical for enormous volumes on data streaming from the usage of mass products with cyber physical features (e.g. vehicles). These data streams offer new opportunities to build innovative services, but their combination with other personal and industrial data is subject to complex ownership and consent aspects, as the data streaming from these products belong to persons or organizations who are owners or users of the products. The project will be based on the solutions developed or under development in previous and current projects (AutoMat, Cross-CPP, CAMPANEO, DALICC etc.). smashHit is driven by 2 industrial Business Cases involving several existing industrial and personal data platforms owned by the leading data providers in three diverse sectors (automotive industry, insurance, smart city), and will provide 3 demonstrators of various applications of the developed solutions. More information is available at [www.smashHit.eu](http://www.smashHit.eu)



Funded by the Horizon 2020  
Framework Programme of the  
European Union

Every effort has been made to ensure that all statements and information contained herein are accurate, however the smashHit Project Partners accept no liability for any error or omission in the same.

© 2020 Copyright in this document remains vested in the smashHit Project Partners.

