Horizon 2020
European Union Funding
for Research & Innovation

CPaaS.io
City Platform as a Service – Integrated and Open

NICT

H2020-EUJ-02-2016
H2020 Grant Agreement Number 723076
NICT Management Number 18302

Deliverable D7.9

Data Management Plan (v2)

Version V2.0

December 31st, 2018

ABSTRACT:

The CPaaS.io project is articulated around on the one-hand a comprehensive set of real-world scenario where large amount of raw data is produced, and on the other hand, on reasoning and analytics techniques that allows cities to exploit that data – often referred to as the oil of the 21st Century"- through the generation of higher order meaningful knowledge.

This document covers EU- and Japan-originated applications and gives final details about all data either collected or generated by the CPaaS.io project and the strategies put in place in order to manage, sustain and share this large amount of information. In particular besides reminding the nature and purpose of the 5 CpaaS.io scenarios data-wise, each CpaaS.io partner provides statement about roles undertaken as far as managing the collected and generated data is concerned.

Disclaimer

Document Information

Editors	François Carrez (UNIS)
Authors	EU consortium
Reviewers	Stephan Haller (BFH)
Delivery Type	R
Dissemination Level	Public
Contractual Delivery Date	31st December 2018
Actual Delivery Date	31st December 2018
Keywords	Data management, Data openness, Privacy

Revision History

Rev.	Date	Description	Contributors
v0.1	10/12/2018	Extended ToC	François Carrez (UoS)
v0.2	27/12/2019	Final Statements	Stephan Haller (BFH)
v0.3	27/12/2018	Updated section 2.1 and 3.1	Martin Strohbach (AGT)
v0.4	28/12/2018	Final statements, checks and release for review	François Carrez (UoS)
v0.5	29/12/2018	Final version for delivery	François Carrez (UoS)
v0.6	29/12/2018	Contribution from JApan	Noboru Koshizuka, Toshihiko Yamakami, Chiaki Ishikawa
v1.0	30/12/2018	Final integration for delivery	François Carrez (UoS)

Table of Contents

List of Tables

List of Acronyms

Acronym	Definition
DMP	Data Management Plan
HDF	Hierarchical Data Format
XML	eXtensible Markup Language

1 Introduction

As described with the H2020 guidelines, Research funding organisations, as well as organisations undertaking publicly funded research, have an obligation to optimize the use the funds they have been granted. Part of this optimization is that data sets resulting from public funded research must be made available to other researchers to either verify the original results (which is integral part of proper scientific approach), or to build upon them.

In order to achieve this high-level objective, a data management policy has to be implemented and thoroughly followed by the CPaaS.io consortium as a whole, even if –per se- not all CPaaS.io partners will be involved in all aspects of those policies/principles.

The Data Management Plan (DMP) is a living document (with two formal versions of the same deliverable released in M6 and M30 respectively) that describes the data management policy (i.e. the management principles) and collected and generated data sets. It covers all aspects introduced in the "Guidelines on Data Management in Horizon 2020", which are:

1. Precise description of the collected and generated data (nature of data, related domain ontologies, standards and data formats used,…)
2. Detail about various aspects of the data management (how it is stored, by whom, under which responsibility, how it is secured, how it is sustained and backed up)
3. Sharing principles (licensing, access methods,…)
4. Detail about how the privacy is maintained

In this second iteration of the Data Management Plan we provide the current and final status on data generation and management and some final statements.

1.1 Delta with D7.2

This section highlight updates brought to version 2 of the Data Management Plan:

- Enhanced user Experience: Section**2.1** has been extended in order to include the Sapporo Snow Festival event;
- 3 applications from Japan in Section **2**;
- Final statements in Section **3** for EU partners and Data Management Plans for Japanese partners.

2 CPaaS.io Research Data

This section introduces the different EU-side use-cases as described in the CPaaS.io Description of Work document and the applications built upon them. It also describes the collected data (meaning the semantically annotated raw-data with no extra added value) and the generated data (meaning the semantic value-added information built from the annotated raw-data using various technics like analytics or reasoning). Part of the information described in this section can be found in a more complete form in CPaaS.io deliverable D2.1 [1].

The 3 use-cases considered in CPaaS.io are:

- Managing Fun and Sport events, which derives into 3 different applications at three different locations (and therefore based on 3 distinct data sets in Amsterdam, Sapporo and Tokyo)
- Waterproof Amsterdam
- Yokosuka Emergency Medical

The 5 derived applications are:

- Enhanced User Experience
- Sapporo Visitor Experience
- Tokyo Public Transportation (case 1 and case 2)
- Waterproof Amsterdam
- Yokosuka Emergency Medical Care

2.1 Data from Enhanced User Experience application

Short description
The core idea of this application is to use IoT sensors and analytics to enhance people's experience while visiting or participating at a fun or sports event. Wearables and mobile phones are used as sensors in order to learn about the activities of event participants. Event participants may include members of the audience, but also performing artists or athletes. For instance AGT has previously equipped referees and cheerleaders in basketball matches with wearable sensors and created content based on the analysed data for consumption on site and for distribution via TV broadcasting, social media and other digital distribution channels[1]. Furthermore the application uses sensor deployed at the venue to measure and analyse fan behaviour and engagement.

Data collected for the Enhanced User Experience application (Color Run and Sappor Snow Festival)
Table 1 summarizes the data from the Enhanced User Experience application covering both events, the Color Run and the Sapporo Snow Festival.

Table 1: Data collected for Managing Fun and Sport events scenario

Biometric data	
Detailed Description	This data set includes a range of measurements from wearables such as wristbands, chest straps and smart sportswear that provides biometric measurements including heart rate, breathing rate and galvanic skin response, burned calories measurements and skin temperature.
OGD or private data	Private
Personal Data	Yes
Hosting	External, CPaaS.io
Data Provider	AGT, YRP
Format	JSON, RDF
Update Frequency	up to every 200ms
Update Size	~1 KB
Data Source	Sensor
Sensor	Wristband, chest strap, smart shirts

[1]http://www.euroleague.net/final-four/berlin-2016/news/i/6vokoibj5fsgqg4q/heed-the-event-platform-based-joint-venture-between-wme-img-and-agt-international-makes-first-official-foray-into-sports

Number of Sensors per person	~6

GPS Traces	
Detailed Description	GPS traces include positional data including altitude information as delivered by GPS devices.
OGD or private data	Private
Personal Data	Yes
Hosting	External, CPaaS.io
Data Provider	AGT, YRP
Format	common GPS formats (GPX, KML, CSV, NMEA)
Update Frequency	Up to 1s
Update Size	< 1KB
Data Source	Sensor
Sensor	GPS sensor in wristbands and mobile phones
Number of Sensors per person	1-2

Motion Data	
Detailed Description	Motion data that measures hand and body movements based on accelerometer and gyroscope sensors
OGD or private data	Private
Personal Data	Yes
Hosting	External, CPaaS.io
Data Provider	AGT, YRP
Format	JSON
Update Frequency	Up to every 16 ms
Update Size	~ 200 byte per sensor reading
Data Source	Sensors
Sensor	Accelerometer and gyroscope sensors of mobile phones, wristband and other wearables
Number of Sensors per person	2-3

Step Counts	
Detailed Description	This data set contains step counts.
OGD or private data	Private
Personal Data	Yes
Hosting	External, CPaaS.io
Data Provider	AGT, YRP
Format	JSON
Update Frequency	Up to 1Hz
Update Size	~ 200 byte per sensor reading
Data Source	Sensors
Sensor	Step count measurement of wristband
Number of Sensors per	1-2

person	

Environmental Data	
Detailed Description	This data set environmental data such light intensity and barometric pressure. The data is primarily collected from wearable sensors.
OGD or private data	Private
Personal Data	Yes
Hosting	External, CPaaS.io
Data Provider	AGT, YRP
Format	JSON
Update Frequency	Up to 1Hz
Update Size	~ 200 byte per sensor reading
Data Source	Sensors
Sensor	Sensors in wristband
Number of Sensors per person	1-2

Mobile Camera videos	
Detailed Description	This data set contains videos recorded by mobile cameras worn by Color Run participants.
OGD or private data	Private
Personal Data	Yes
Hosting	External, CPaaS.io
Data Provider	AGT, YRP
Format	MP4
Update Frequency	30fps
Update Size	(~45kbps)
Data Source	Mobile Camera
Sensor	GoPro Hero4 Camera
Number of Sensors per person	1

Data generated by the Enhanced User Experience application (Color Run and Sapporo Snow Festival)

The Enhanced User Experience application generates or uses the following types of data

1) **User Activity**. Mainly on motion data and therefore private information. A user activity is always linked to a user and therefore personal information. The re-use of the data is possible within the boundaries defined in the consent forms used to collect the data.
2) **Dominant Colors**. Provides information about the prevailing colour in a video feed and is used for detecting colour stations in the Color Run. The output is a colour value, duration and location. The generated can be provided in anonymised form, but requires further examination to what degree it can be opened.

3) **Clothing Analysis**. Clothing Analysis uses deep learning techniques to determine metrics based on clothing styles derived from images. By nature this metrics are linked to user and therefore reflect private data that can only be reused in the boundaries of the consent forms used to collect the data.
4) **Running Type.** This classifies whether the participant ran more like a fun or an ambitious runner during an event. This is personal data.
5) **Dance Energy.** Provides a measure of the accumulated energy overtime while dancing per user. This is personal data.
6) **Emotions.** Refers to people's emotion during the event. This is personal data, but also aggregated data is used.
7) **Tube Rider Classification.** Classifies the tube ride based on how intensely the tube rotated. This is personal data.
8) **Event Sentiment.** This is an aggregated measure of sentiments per event derived from Tweets.
9) **Ride Bumpiness.** Provides a metric for the bumpiness of a ride. This is personal dat.
10) **Throw Intensity.** This provides a measure for the intensity of throwing a ball such as a snow ball. This is personal dat.

Table 2: Data generated for the Enhanced User Experience application

Types of generated data	Based on…	Anonymised Y/N	Open Y/N
User Activity	Motion Data	N	reusable, but not open
Dominant Colour	Mobile Camera Videos	Y	Reusable, but not fully open
Clothing Analysis	Mobile Camera Videos, Public Images	N	Reusable, but not open
Running Type	Motion Data	N	N
Dance Energy	Motion Data	N	N
Emotions	Video Data	N	N
Tube Rider Classification	Motion Data	N	N
Event Sentiment	Tweets	Y	N
Ride Bumpiness	Motion Data	N	N
Throw Intensity	Motion Data	N	N

2.2 Data from Waterproof Amsterdam

Short description

Extreme rainfall and periods of continued drought are occurring more and more often in urban areas. Because of the rainfall, peak pressure on a municipality's sewerage infrastructure needs to be load balanced to prevent flooding of streets and basements. With drought, smart water management is required to allow for optimal availability of water, both underground as well as above ground.

The Things Network develops the Amsterdam Waterproof application, which is a software tool creating a network of smart, connected rain buffers, be it rain barrels, retention rooftops or buffer otherwise, that can be both monitored and controlled centrally by the water management authority. Third party hardware providers will connect their buffers to this tool for uplink and downlink data transmission. External data such as weather data and sewerage capacity are added, in order to calculate the optimal filling degree of each buffer and so operate a pump or valve in the device. Waternet is the local water management company who will be the main user of the application.

Data collected for the Waterproof Amsterdam application

In the section below are the data sets used for the Waternet application. It consists of device data (rain buffer information), public weather data and government data about physical infrastructure. Device data will be stored in the application and could be stored in CPaaS, especially as it contains private data like name and address of device owner. As this stage however we cannot determine whether this privacy data will be shared by the vendors of the devices, who are also the ones maintaining them. They are the only actor who has direct contact with the end user and/or owner of the device. (Historical) weather data is publicly available on the web, so there is no need to store this data. It will be provided by a subscription data feed from the web. The third data set is already owned and stored by Waternet, so there is also no need for storage capabilities.

Table 3: Data collected for the Waterproof Amsterdam scenario

Weather data	
Detailed Description	Upcoming weather displaying periods of heavy rain or drought
OGD or private data	OGD
Personal Data	No
Hosting	Platform
Data Provider	KNMI – Dutch weather forecast agency
Format	HDF5/JSON
Update Frequency	Hourly
Update Size	20kb
Data Source	Sensors
Sensor	Water sensor
Number of Sensors	unknown

Rain buffer information	
Detailed Description	Specific information about each rainbuffer (rooftop, barrel, underground storage) • Buffer size and type • Filling degree • Temperature • Location • Battery status • Pump/valve capacity • Active pump/valve hours • Owner name, address, contact information
OGD or private data	Private
Personal Data	Yes – anonymised and not open
Hosting	Platform

Data Provider	Rain buffer hardware provider
Format	JSON
Update Frequency	Hourly
Update Size	10b
Data Source	Sensors
Sensor	Water sensor or infrared sensor
Number of Sensors	1 per buffer

Sewerage processing capacity	
Detailed Description	Geographical data on water infrastructure depicting remaining capacity of sewerage
OGD or private data	Private
Personal Data	No
Hosting	External
Data Provider	Waternet
Format	XML
Update Frequency	Hourly
Update Size	1kb
Data Source	Sensors, maps
Sensor	Water sensor
Number of Sensors	unknown

Data generated by the Waterproof Amsterdam application

The Waterproof Amsterdam generates different types of data.

1. Open/close command per buffer. This is the most important data generated, as it determines when an actuator inside a buffer should be operated (valve open or pump on). Based on all data sources available, an algorithm will determine which conditions are required to perform a certain command. The commands can be open and close, or a value in between as different water discharge mechanisms have different capacities (i.e. a percentage of full capacity)
2. Aggregated remaining buffer capacity per area. Waternet as the primary user of the application needs to monitor the total remaining capacity to buffer rain water, to understand whether there will be plenty capacity to catch up rain water in moments of heavy rainfall.
3. Aggregated litres of rain water processed per area. This is a metric to be used to show the impact the micro buffer network has generated over time. These insights may be used for PR and marketing purposes to stimulate individuals and companies to also buy and install such rain buffers.

The open data in the table below can be reused to perform analytics on historical data, and could be open data through a public (graphical or application) interface for third parties to interact with.

Table 4: Data generated for the Waterproof Amsterdam application

Types of generated data	Based on…	Anonymised Y/N	Open Y/N
Open/close command per buffer	All data sets	Y	N
Aggregated remaining buffer capacity (street, area, city level)	Individual rain buffers filling degree and location, map	Y	Y

Aggregated litres processed by the buffers	Individual rain buffer pump hours run and pump capacity, map	Y	Y

2.3 Data from Sapporo Visitor Experience application

Short description

The Sapporo *Visitor Experience application* is part of a set of applications that will use the CPaaS.io platform to manage fun and sports event. This application focuses on tourist information services including event information and public transportation information.

Data collected for the Sapporo Visitor Experience application

We established Sapporo Open Data Association whose members are data providers such as Sapporo City, companies and organizations in tourism, sports and transportations. This association collects data from those members, transforms them to Open Data and publishes them on the Sapporo Open Data Portal.

Sightseeing Spot	
Detailed Description	Comments, pictures and locations of popular sightseeing spots
OGD or private data	OGD
Personal Data	No
Hosting	Sapporo Open Data Association
Data Provider	Sapporo City
Format	PDF, HTML, CSV, JPEG
Update Frequency	Yearly
Update Size	N/A
Data Source	Web
Sensor	Ucode marker
Number of Sensors	11

Meal	
Detailed Description	Restaurant name, location, opening hours, menu, service
OGD or private data	Private
Personal Data	No
Hosting	Sapporo Open Data Association
Data Provider	Hotel, Sapporo Station Shopping Mall
Format	PDF, HTML, CSV, JPEG
Update Frequency	Monthly
Update Size	N/A
Data Source	Web
Sensor	Ucode marker
Number of Sensors	11

Hotel	
Detailed Description	Hotel name, location, number of rooms, access
OGD or private data	Private
Personal Data	No
Hosting	Sapporo Open Data Association
Data Provider	Hotels
Format	PDF, HTML, CSV, JPEG
Update Frequency	Monthly
Update Size	N/A
Data Source	Paper
Sensor	N/A
Number of Sensors	N/A

Other Facilities	
Detailed Description	Location of ATM, Toilet, Rental Locker, etc.
OGD or private data	Private
Personal Data	No
Hosting	Sapporo Open Data Association
Data Provider	Sapporo Station Shopping Mall
Format	PDF, HTML, CSV, JPEG
Update Frequency	Monthly
Update Size	N/A
Data Source	Web
Sensor	Ucode marker
Number of Sensors	11

Snow Festival	
Detailed Description	Event information in Sapporo Snow Festival
OGD or private data	Private
Personal Data	No
Hosting	Sapporo Open Data Association
Data Provider	Operating Committee of Sapporo Snow Festival
Format	PDF, HTML, CSV, JPEG
Update Frequency	Yearly
Update Size	
Data Source	Web
Sensor	Ucode marker
Number of Sensors	11

Stadium Data	
Detailed Description	Location, facilities and drawing of stadiums
OGD or private data	OGD
Personal Data	No

Hosting	Sapporo Open Data Association
Data Provider	Sapporo City
Format	PDF, HTML, CSV, JPEG
Update Frequency	Yearly
Update Size	N/A
Data Source	Web, Paper
Sensor	N/A
Number of Sensors	N/A

Train & Bus Data	
Detailed Description	Time table, Operating status, Line, Real time location, bus stop, etc.
OGD or private data	OGD & private
Personal Data	No
Hosting	Sapporo Open Data Association
Data Provider	Sapporo City, Hokkaido Chuo Bus
Format	PDF, HTML, CSV, JPEG
Update Frequency	Monthly
Update Size	N/A
Data Source	File
Sensor	N/A
Number of Sensors	N/A

Express Way Data	
Detailed Description	Information of service area
OGD or private data	Private
Personal Data	No
Hosting	Sapporo Open Data Association
Data Provider	NEXCO East Japan
Format	PDF, HTML, CSV, JPEG
Update Frequency	Yearly
Update Size	N/A
Data Source	Web
Sensor	N/A
Number of Sensors	N/A

Underground Map Data	
Detailed Description	Drawing of underground map
OGD or private data	Private
Personal Data	No
Hosting	Sapporo Open Data Association
Data Provider	Sapporo Station Shopping Mall
Format	PDF, HTML, CSV, JPEG
Update Frequency	Yearly
Update Size	N/A

Data Source	File
Sensor	Ucode marker
Number of Sensors	11

Note: Japanese applications including Sapporo visitor experience are based on u2 architecture, ucode architecture Version 2. Data collected from hardware device is only ucode, a non-semantic 128-bit id umber. Only after a process called ucode resolution process is performed by an application, a meaningful data is provided to the application in a context-sensitive manner. Ucode itself has no notion of open vs proprietary / anonymous vs public.

Table 5: Data collected for the Sapporo Visitor Experience application

Devices	Types of data	Anonymised Y/N	Open Y/N
ucode marker	128 bit unique identifier, ucode.	N	n/a

Data generated by the Sapporo Visitor Experience application

Note: The data generated such as guidance uses open data map. However, particular guidance based on the user's current location is a private data and should be anonymised.

Table 6: Data generated for the Sapporo Visitor Experience application

Types of generated data	Based on…	Anonymised Y/N	Open Y/N
Sightseeing spot guidance	Location information (from ucode marker and GPS, ditto in rows below) and Sightseeing Spot data.	Y	N
Meal Guidance	Location information and Meal (restaurant, etc.) data.	Y	N
Hotel Guidance	Location information and hotel data.	Y	N
Facility Guidance (ATM, toilet, etc.)	Location information and facility data	Y	N
Snow Festival event guidance	Location information and Snow Festival data	Y	N
Stadium Guidance	Location information and Stadium data	Y	N
Train and Bus Guidance	Location information and Train and Bus data.	Y	N
Expressway Guidance	Location information and Expressway guidance	Y	N
Underground Route Guidance	Location information from ucode marker (GPS is not available!) and Underground Map data.	Y	N

2.4 Data from Tokyo Public Transportation: (a) Open Data Challenge contest

Short description

One prominent use case of Tokyo Public Transportation open data is the open data contest that uses data from public transportation operators in Tokyo area, i.e., Open Data Challenge for Public

Transportation in Tokyo.

One contest has already been performed and the second one is under way: From July 17, 2018 to March 31, 2019.
URL: *https://tokyochallenge.odpt.org/en/index.html*

Third one is already announced to start on January 16, 2019.

The contest is held by the Association for Open Data of Public Transportation (ODPT). See URL: *http://www.odpt.org/*It is chaired by Dr. Ken Sakamura, the director of YRP UNL, the Japanese coordinator of CPaaS.io. YRP and MSJ, both CPaaS.io partners, are members of ODPT.

The explanation of the open data provision framework within the context of Open Data Challenge for Public Transportation in Tokyo is in order.

The contest makes available many types of data from many raw data sources from public transportation operators in Tokyo area. However, they are all in proprietary formats and CPaaS.io can't mention them in a public document. They are covered by NDA.

From these diverse data sources, the contest developer site uses program modules that understand these formats and produces unique data format (JSON-LD) for all types of data so that ALL the information is provided in this single format.

So, the participants in the contest only need to know the API to access this JSON-LD data and the semantics of JSON-LD data. They do not have to deal with the idiosyncrasies of various data formats and network protocols used by the public transportation operators in the greater Tokyo metropolitan area.

There are many types of data available in the contest. There are dynamic data available, but the details change from time to time and is hard to keep track. So, the author refers the reader to the contest developer site. URL: https://ckan-tokyochallenge.odpt.org/dataset

As of now (December 28, 2018), there are 152 data sets made available from the contest site. (https://ckan-tokyochallenge.odpt.org/dataset)

The JSON-LD data set has the semantics which is explained by a document available at the developer site of the contest. (It is readable only after one registers itself at the website. See the contest site for details.) The following URL is the generic API and semantics initially used some ago for the contest framework (in Japanese). URL: https://docs.odpt.org/

However, the developer site has English translation of the more focused and updated document and so the interested reader is encouraged to register in the contest (there is NO OBLIGATION to submit the final contest work).

Data collected for the Open Data Challenge Contest

Due to the reasons mentioned above, we cannot really mention how and by means of which device the data is collected.
So instead, below, the generic description of data made available is listed (quoted from the URL of the contest site.)

Table 7: Data Collected for the two Tokyo Public Transportation applications

Devices	Types of data	Anonymised Y/N	Open Y/N
n/a	n/a	n/a	Not really in the original format.

- List: Published Data types.
  - Railway
    - ✧ Tokyo Metro
      - Static data, including train/station timetable, and dynamic data, such as train location information and status information
    - ✧ Bureau of Transportation, Tokyo Metropolitan Government
      - Static data, including train/station timetable of Toei Transportation, Tokyo Sakura Tram (Toden Arakawa Line), and Nippori-Toneri Liner, and dynamic data, such as status information, and train location information of Tokyo Sakura Tram (Toden Arakawa Line)
    - ✧ JR East
      - Static data, including train/station timetable, and dynamic data, such as train location information and status information, of multiple railway lines in Greater Tokyo
    - ✧ Odakyu
      - Static data and dynamic data, such as status information
    - ✧ Keio
      - Static data, including train/station timetable, and dynamic data, such as status information
    - ✧ Keisei
      - Static data, including station timetable, and dynamic data, such as status information
    - ✧ Keikyu
      - Static data, including station timetable, and dynamic data, such as status information
    - ✧ Seibu
      - Static data, including station timetable, and dynamic data, such as status information
    - ✧ Tokyu
      - Static data and dynamic data, such as status information
    - ✧ TWR Rinkai Line
      - Static data, including train/station timetable, and dynamic data, such as status information
    - ✧ Tobu
      - Static data, including station timetable
    - ✧ Yurikamome
      - Static data, including station timetable
  - Bus
    - ✧ Bureau of Transportation, Tokyo Metropolitan Government
      - Static data, including timetable of Toei Bus, and bus location information as dynamic data
    - ✧ Odakyu Bus
      - data, including timetable
    - ✧ Kanto Bus
      - data, including timetable
    - ✧ Keio Bus
      - Static data, including timetable

- ✧ Kokusai Kogyo
  - Static data, including timetable
- ✧ JR Bus Kanto
  - Static data, including timetable
- ✧ Seibu Bus
  - Static data, including timetable
- ✧ Tokyu Bus
  - Static data, including timetable
- ✧ Tobu Bus
  - Static data, including timetable
- ✧ Nishi Tokyo Bus
  - Static data, including timetable

➢ Airline

- ✧ All Nippon Airways
  - Static data, including flight timetable, and dynamic data, such as real-time arrival/departure information
- ✧ Tokyo International Air Terminal
  - Static data, including flight timetable, and dynamic data, such as real-time arrival/departure information
- ✧ Narita International Airport
  - data, such as real-time arrival/departure information
- ✧ Japan Airport Terminal
  - Flight timetable, and dynamic data, such as real-time arrival/departure information
- ✧ Japan Airlines
  - Static data, including flight timetable, and dynamic data, such as real-time arrival/departure information

➢ -station map and facility information of train stations

- The intra-station map and the facility information of the train stations around Shinjuku station and Tokyo station are being prepared by Ministry of Land Infrastructure, Transportation and Tourism anticipating the Tokyo Olympic and Paralympic Games in 2020. The data are made available in this challenge.

Data generated by Open Data Challenge Contest

As mentioned above, from the diverse data sources of many transportation operators, the contest developer site uses program modules that understand these formats and produces unique data format (JSON-LD) for all types of data so that ALL the information is provided in this single format. The generated data is stored in a ckan website.

From the data in the ckan website, many apps create their own data (guidance, visualization, etc.). But it is beyond the scope of this document to characterise the existing contest apps and their generated data.

Table 8: Data generated for the

Types of generated data	Based on…	Anonymised Y/N	Open Y/N
Many in JSON-LD format	Many	Varies	Yes, in JSON-LD format to contestants.

2.5 Data from Tokyo Public Transportation (b): Tokyo Management of Service Vehicles

Short description

Tokyo Management of Service Vehicles provides a platform mobility tracking infrastructure to enable management of service vehicles using monitoring, planning, coordination, analysing of things with mobility, such as service vehicles.

Data collected for the Tokyo Management of Service Vehicles application

(a) Time tables and real-time traffic information of public transportation such as railways, buses, and airplanes. Data are owned by transportation operators. For public transportation, it is public. For public service-related transportation, it depends on services.

(b) Geographical data for routes, locations of railway stations, bus stops, airports, etc. Data are owned by transportation operators.

(c) Data of origination and destination, such as maps, facility lists. Data are owned by transportation operators. Some are public, some (such as origination places) are company-confidential.

(d) Driver information. It is owned by transportation companies. It needs privacy protection.

Table 9: Data Collected for the Tokyo Management of Service Vehicles (* not yet defined)

Devices	Types of data	Anonymised Y/N	Open Y/N
Smartphone	Location, time	N	Y
Smartphone	Device-id	Y	N
Smartphone	Image	N	N
*	Route	N	N
*	Weather	N	Y
*	Road information	N	Y

Data generated by the Tokyo Management of Service Vehicles application

Route logs: generated from mobility tracking data. The data is owned by transportation providers. Stored inside transportation providers services on top of CPaaS.io

Table 10: Data generated for the two Tokyo Management of Service Vehicles application (* Not yet defined)

Types of generated data	Based on…	Anonymised Y/N	Open Y/N
Route logs	Time-series-mobility data	N	*

2.6 Data from Yokosuka Emergency Medical Care application

Short description

Yokosuka Emergency Medical Care Application provides a platform to improve quality and efficiency of medical information sharing situation of sick person and to reduce time to start initial treatment after ambulance arrival.

Data collected for the Yokosuka Emergency Medical Care application

Yokosuka Emergency Medical Care application collects IDs and location from smartphone in ambulance. This application also collects the video image in ambulances from camera(s) in the ambulances but not stored.

Summarised in the table below:

Table 11: Data collected for the Yokosuka Emergency Medical Care application

Devices	Types of data	Anonymised Y/N	Open Y/N
Smartphone	Location, time	N	Y
Smartphone	Device-id	N	N
camera	video image in ambulance (not stored)	N	N

Data generated by the Yokosuka Emergency Medical Care application

Yokosuka Emergency Medical Care application generated a location map of ambulances.

Summarised in the table below:

Table 12: Data generated for the Yokosuka Emergency Medical Care application

Types of generated data	Based on…	Anonymised Y/N	Open Y/N
Location map of ambulances	location and time from smartphone in ambulances	N	N

3 CPaaS.io Research Data management plan

CPaaS.io project follows the principle that research data will be handled and managed by those organisations/institutions that either collects or generates the research data. The CPaaS.io project comprise a number of partners that are involved directly in either:

- Producing the actual data during the trials, or
- Developing tools and enablers (e.g. analytics, reasoners, etc.) that are needed as core elements in the CPaaS.io system architecture, or

- Elaborating upon the produced data (using the aforementioned enablers) in order to produce new value-added knowledge.

The individual roles and duties of such partners and the research data management plans that are in place in the organisations taking part in CPaaS.io are described in the following sub-sections.

3.1 AGT International (AGT)

Data collection (from sensors)
The data collected by AGT has been described in Section 2.1 and is used for generated the data as described in Table 2 and for developing the Enhanced User Experience application. As described in D2.2 the collected data is enriched with additional metadata.

Data generation
The data generated by AGT has been described in Table 2 and is used in the Enhanced User Experience application.

Data Management

We have implemented appropriate technical and organizational measures to ensure generated data is protected from unauthorized or unlawful processing, accidental loss, destruction or damage. We review our information collection, storage and processing practices regularly, including physical security measures, to guard against unauthorized access to our systems. We restrict access to generated data to only those employees, contractors and agents who strictly need access to this information, and who are subject to strict contractual confidentiality obligations.

Update at the end of the Project:

In cases where consent forms and legislation allow, the data will be maintained beyond the project mainly for presenting the results of the project.

3.2 University of Surrey (UoS)

ICS at the University of Surrey is not involved neither in the production of raw data nor in the exploitation or generation of higher-level information out of it. However, UoS is focussing on architecture work where particular attention is paid to ensuring that 1/ all privacy-related requirements are thoroughly taken into account 2/ important part of the data is publicly available following the project Open Data policy.

To this respect UoS is aiming at providing a bridge between CPaaS.io and another FIRE project called FIESTA-IoT, two projects where UoS is actively involved. UoS will in particular aim at involving CPaaS.io in either the 2nd Open Call of FIESTA-IoT or as a fellow contributor to that project via a cooperation agreement to be discussed between the two projects after both POs have been consulted on that matter. In both cases, CPaaS.io could play two non-exclusive distinct roles:

- Data-provider: playing this role the CPaaS.io project would inject its data or part of its data (either raw data or inferred data) to the FIESTA-IoT so that so-called experimenters can make use of it using the FIESTA-IoT enablers; or
- Experimenter: playing this role, CPaaS.io could reuse additional data sets produced by the FIESTA-IoT collaborators for testing our new own algorithms (e.g. Analytics) and techniques.

Data collection (from sensors)
UoS does not participate in any data collection

Data generation
UoS does not generate any new data from the project data sets

Data Management
UoS does not manage any gathered or generated data

Update at the end of the Project:

> The foreseen collaboration between CPaaS.io and FIESTA-IoT could not take place due to technical issues. It will be discussed –early 2019- whether the data collected internally as part of our smartBuilding test-bed can be ingested to the CPaaS.io platform. That data would be then open to students and used for the sake of machine learning experiments.

3.3 Bern University of Applied Sciences (BFH)

The BFH is not directly involved in the implementation of the envisaged use cases. Its main research focus is in the data management concepts – in particular the usage of Linked Data and Open Government Data as well as data quality annotations, the application of MyData approaches, and in the validation of the use cases. Hence it is not collecting, generating or storing any data.

However, as part of its exploitation, validation and knowledge transfer activities, BFH is planning to connect some sensors via the LoRa testbed network that another institute (Institute for Energy and Mobility Research in Biel) is currently setting up. What data will be collected and for what purposes exactly will be defined at a later stage; a related data management plan will be drawn up before any data collection starts.

Data collection
BFH is not collecting any data for the main use cases of CPaaS.io. It may collect and make available some sensor data through the LoRa network at BFH for testing and validation purposes; details will be determined at a later stage.

Data generation
BFH is not generating any data for the main use cases of CPaaS.io. It may link public data sources (e.g., from the Swiss Open Government Data portal at *www.opendata.swiss*) with the sensor data collected through the LoRa network at BFH for testing and validation purposes; potential use cases will be determined at a later stage.

Data Management
BFH is not managing any data for the main use cases of CPaaS.io. Data collected and generated for testing and validation purposes through the LoRa network at BFH will likely be made available publicly, in the spirit of open data research, unless the data could allow to infer any information about individuals. Details are to be determined at a later stage.

Update at the end of the Project:

> BFH used a single TTN sensor to generate some test data. This data however is not being persisted beyond the end of the project.

3.4 OdinS

OdinS as a partner involved on the security and privacy aspects, will check and support the project to check that data access and sharing activities will be implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and in the EU, as well as with the H2020 rules. Concerning the results of the project, these will become publicly available based on the IPRs as described in the Consortium Agreement.

Due to the nature of the data involved, some of the results that will be generated by each project phase will be restricted to authorized users, while other results will be publicly available. Data access and sharing activities will be rigorously implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and in the EU, as well as with the H2020 rules.

Data collection (from sensors)
OdinS will not be involved in the data generation of data from sensors, working exclusively in the architecture aspects of the data collections and its consequence over the security and privacy components.

Data generation
OdinS is not involved in the production of raw data, but as part of the Task 4.1 User Empowerment Component Definition and the definition of access control policies and use consent solution, OdinS will generate information associated to data for controlling access and sharing data between entities and components that will use the platform.

Data Management
As the raw data included in the data sources, will be gathered from sensor nodes and information management systems, those could be seen as highly sensitive. Therefore, access to raw data can only take place between the specific end users based on the policies associated and the partners involved in the analysis of the data. For the models to function correctly, the data will have to be included into the CPaaS.io repository. The results of the data analytics are set to be anonymised and made available to the subsequent layers of the framework, which will then allow the possibility for external industry stakeholders to use the results of the project for their own purposes.

3.5 NEC

NEC is not directly involved in the production or raw data. NEC's focuses are in the architecture (system integration including transferability and semantic interoperability) area and cloud-edge processing of the data. FIWARE's resources such as the Generic Enablers and NEC's IoT Platform can support storage and exploitation of data from use cases for generating higher-level analytical results. NEC pays particular attention to privacy related requirements as well as the Open Data policy of CPaaS.io.

Data collection
NEC is not planning to collect any raw data for the use cases of CPaaS.io.

Data generation
NEC is not generating data for the main use cases, NEC may exploit shared data from use cases and generate higher level data as a result. Potential use cases will be determined at a later stage.

Data management

While NEC is not directly involved with the use cases, it will take part in data transferability and management via the provided IoT Platform. NEC has implemented necessary organizational and technical measures for the usage of the data and its protection from unauthorized persons.

3.6 The Things Network

Data collection (from sensors)
The data collected by The Things Network has been described in Section 2.2 and is used for generated the data as described in table 2 and for developing the Waterproof Amsterdam application. As described in D2.4 the collected data is enriched with additional metadata.

Data generation
The data generated by The Things Network has been described in and is used in the Waterproof Amsterdam application. Private data from owners of a rain buffer is anonymised. Based on an algorithm, data from various sources is processed by the application to determine the optimal filling degree for each individual rain buffer. The results may be used for automated control of buffers, or push notifications to trigger manual control.

Data Management
Open data such as weather data will be streamed into the application and not stored locally.

Private data from external sources such as device location will be stored in the application and only released in an anonymised and aggregated manner. Personal details about a device, such as name, address and contact details will also be stored in the application in a secure account server. These data may be transferred to CPaaS.io at some time, easing security and privacy demands on the application end and transferring those to CPaaS.io

Parts of the personal data, such as buffer location, size and processed litres, will be released in an aggregated, anonymised manner (e.g. on a heat map) per area of a city or the city as a whole.

Readily available data from Waternet about sewerage capacity will abide by the policies of Waternet. These policies are not yet clear at the moment.

We restrict access to generated data to only those employees, contractors and agents who strictly need access to this information, and who are subject to strict contractual confidentiality obligations.

3.7 YRP

Data collection (from sensors)

The data collected by YRP has been described in Section 2.4 and is used for generated the data as described in table 6 for developing the Tokyo Public Transportation application. It also has been described in Section 2.6 and is used generated the data as described in table 12 for developing the Yokosuka Emergency Medical Care application. As described in D2.4 the collected data is enriched with additional metadata.

Data generation

The data generated by YRP has been described in Section 2.4 and is used in the Tokyo Public Transportation application. It also has been described in 2.6 and is used in Yokosuka Emergency Medical

Care application. Neither Tokyo Public Transportation nor Yokosuka Emergency Medical Care application generates private data.

Data Management

Tokyo Public Transportation application coverts data such as timetables, location of trains and provides converted data. This application does not manage private data.

Yokosuka Emergency Medical Care application manages the location of ambulances and provides a map of the location. This application streams private data, such as video image in ambulance and does not store locally.

Explicitly private data that was gathered during Sapporo Snow Festival in February 2018 was collected with explicit user agreement. (The English translation of the agreement is part of D2.12 Final Ethics Report.). That data was exchanged with AGT under a mutually signed data exchange agreement.

3.8 Microsoft Japan

Data collection (from sensors)

MSJ is not directly involved in the production or raw data.

Data generation

MSJ is not generating data for the main use cases.

Data Management

While MSJ is not directly involved with the data management of each CPaaS.io use cases, MSJ has implemented necessary organizational and technical measures for the usage of private data and its protection from unauthorized persons in its cloud service.

In the case of collecting data from visitors to Sapporo Snow Festival in February 2018, MSJ personnel acted as an agent of YRP which was responsible for the gathering and storage of personal data in Japan.

3.9 ACCESS

Data collection (from sensors)

ACC collects data from smartphones. The collected data are location, time, device ID of the smartphone and the images taken by the camera of the smart phone.

Data generation

ACC generates route logs: generated from mobility tracking data in Tokyo Management of Service Vehicles application. The data is owned by transportation providers. This data is stored inside transportation provider's services on top of CPaaS.io.

Data Management

ACC manages the data with the usual access control system so that they are not leaked to non-owners. Also, the identity of smartphones are crucial for the application in so far as the route log is generated, and are not anomyzed within the application domain.

3.10 Ubiquitous Computing Technology

Data collection (from sensors)

Ubiquitous Computing Technology (UCT) is not planning to collect any raw data for the use cases of CPaaS.io.

Data generation

UCT does not generate any new data from the project data sets

Data Management

UCT does not manage any gathered or generated data

3.11 University of Tokyo

Data collection (from sensors)

University of Tokyo (UoT) was not planning to collect any raw data for the use cases of CPaaS.io in 2016. However, as of 2018, it has now facilities for monitoring indoor temperature, lighting, human movement, humidity and other data from environmental sensors.

Data generation

UoT does not generate any new data from the project data sets.

Data Management

UoT does not manage any gathered or generated data other than the pure sensor data (non-private data).

4 Conclusions

In this second version of the Data Management Plan deliverable we presented the CPaaS.io approach towards data management as handled by the EU CPaaS.io consortium and final statements made at the end of the project as far as management of data is concerned.

5 References

[1] *CPaaS.io Deliverable D2.1: "Requirements Specification".*