t

DELIVERABLE

D7.7 – Data Management Plan - DMP - Third release

Project Acronym: UNCAP

Grant Agreement number: 643555

Project Title: Ubiquitous iNteroperable Care for Ageing People

Revision:

Authors: Claudio Eccher (FBK)

Project co-funded by the the Horizon 2020 Framework Programme of the European Union		
Dissemination Level		
P	Public	**X**
C	Confidential, only for members of the consortium and the Commission Services	

1. Revision history and statement of originality

1.1. Revision history

Rev	Date	Author	Organization	Description
v0.0	30/06/2017	Claudio Eccher	FBK	
V1.1	30/06/2017	Irene Facchin	TRILOGIS	Quality check

1.2. Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

2. List of references

Number	Full Reference
1	http://www.openmhealth.org/documentation/#/schema-docs/schema-library
2	http://www.openstreetmap.org/copyright/
3	http://wiki.openstreetmap.org/wiki/OpenRouteService
4	*https://www.chino.io/*
5	*https://www.hetzner.de/en/*
6	*http://oauth.net/*

3. Table of Acronyms

Acronym	Description
AL	*Assisted Living*
CAP	*Clinical Assessment Protocol*
CHA	*Community Health Assessment*
GIS	*Geographical Information System*
HC	*Home Care*
LTCF	*Long Term Care facilities*
TLS	*Transport Layer Security*
ESB	*Enterprise Service Bus*
QoS	*Quality of Service*
SLA	*Service Level Agreement*

4. Executive Abstract

This document reports on how the data will be handled during the UNAP project. The goal is to consider the many aspects of data management, metadata generation, data preservation, and analysis, in order that data are well-managed in the present, and prepared for preservation in the future.

The present document is the third release of the DMP, and is divided in four main sections.

Section 7 presents the updated list of UNCAP data manged in the UNCAP project (reported from D7.6).

Section 8 (Data Storage and Backup) details how the data will be stored and backed up during the research, presenting also a simplified schema of the data storage solution (reported from D7.6).

Section 9 (Data Access and Sharing) presents the mechanism to access the data: authentication, authorization, data structure. And data access procedure.

Section 10 (Open Access to Data) presents the mechanisms to be implemented to enable open access to UNCAP data.

5. Table of Content

6. Table of Figures

7. UNCAP Data – Updated list

This section contains the updated set of health-related data collected, managed and stored by the UNCAP systems. Hence, we report here and update the set of data sketched in Section 9 of the Deliverable 7.5 DMP – first release. The section is subdivided into different groups to cover all the different types of information collected.

7.1. Personal Data/registry

The UNCAP system collects and stores information about the personal data of the patients enrolled in the study and that are using UNCAP.

The minimum required dataset that will be collected is as follows:

- Name
- Surname
- Date of birth
- Address
- UserID: unique identifier of the user, shared between all applications connected to UNCAP
- Password: password of the user's account, stored as a token
- Authorization profiles: list of the roles and services accessible by the user
- Others to be defined during the implementation phase (WP2)

The creation of a new user is managed using roles:

- Administrator: has full access to the system and can create new users with any role.
- Clinical staff: clinicians will be able to add a new patient to UNCAP.
- Patient: is the end user of the system and has access only to his own data. A patient cannot subscribe himself/herself autonomously to UNCAP but the subscription should be managed by a clinician or an administrator.

This list should be considered as the minimum required set but can be expanded or modified during the course of the project.

7.2. Health-related data

A number of health-related data will be collected during the project course, specifically during the piloting phase from M19 to M30. Data is collected from external devices that will be integrated during the implementation phase. Right now the devices that are planned to be integrated are:

- Pulse oximeter: it is a non-invasive device used for monitoring blood saturation and heart rate. O_2 saturation is expressed as a percentage with a measurable range going from about 70% to 99%. Heart rate (pulse) represents the number of times the heart beats per minute, with ranges typically going from 30 to 250 bpm.
- Blood pressure monitor: the output of the device is a value, expressed in mmHg (millimetres of mercury), representing the pressure exerted by circulating blood upon the walls of blood vessels. It is composed of two values that should always be

reported together: the maximum pressure (systolic) and minimum pressure (diastolic).

- Glucometer: used to determine the concentration of sugar in blood. It is commonly represented in terms of millimoles per liter (mmol/L) or milligrams per decilitre (mg/dL).
- Scale: Wi-Fi/Bluetooth scales are now commercially available. Those devices report a number of measures related to the person health status such as:
    - Body weight: a value representing the mass of a person expressed in kilograms (or pounds).
    - Body fat percentage: it is the total mass of fat divided by the total body mass. It is expressed in percentage.
    - Total body water: expressed as a percentage represents the quantity of water contained in the body.
    - Body mass index: it is defined as the body mass divided by the square of the body height and is consequentially expressed as kg/m^2.
    - Bone density: is the amount of mineral matter per square centimetre in bones.
- Heart rate monitor: monitors the pulse. From the point of view of the data collected this does not differ from a pulse oximeter.

Depending on the sensor used and on the value measured, UNCAP will store measurements in two different formats: as single entries or as time series. As an example a glucose measure will be stored as a single entry while heart rate may be stored as a time series since it is a representation of a measured value collected repeatedly during a period of time.

Each entry in the database, in addition to the measured value, will report the user ID to which is associated the measurement, the device ID associated to the sensor used and a timestamp representing the time in which the measure was collected.

The project implements the standard OpenMHealth for all medical data[1].

7.3. GIS data

During the project course, a lot of information will be collected in the GIS (geographic information system) domain. What we address as GIS is basically every data needed to describe something that has some geographic features. In our case that data is composed of:

- Maps: those can be both outdoor and indoor maps.
    - Outdoor maps will be used mainly as background layers in the client frontends. The maps are acquired from open data repositories available online such as OpenStreetMap[2]. Maps will not be stored but will be requested directly from the client application when needed (caching on the client device may be enabled).
    - Indoor maps are used to represent the pilots spaces involved in the project. For each pilot a request for a detailed map of the structure will be done. If needed, the collected maps will then be converted in a digital format and

[1] http://www.openmhealth.org/documentation/#/schema-docs/schema-library
[2] http://www.openstreetmap.org/copyright/

orthocorreted to correctly represent the structure in world coordinates. The maps will be uploaded on the UNCAP Cloud and will be accessible from the exposed services of UNCAP both as raster images and as vector graphics.

- Routing graphs: routing graphs are lists of nodes (points) and connections (lines) used to represent paths between places. UNCAP will use online available repositories (from the OpenRouteService[3]) and will integrate indoor graphs needed for indoor routing inside the pilots. Indoor graphs are manually created starting from the pilot maps described before.
- Points of Interest: online available repositories will be integrated to show the nearest POI such as hospitals and stations. We will also give the ability to each user to store the position of his own personally preferred places (i.e. friend's house, rehabilitation centre) that will be stored in UNCAP and will be accessible only to who created them.
- Localization data: localization is one of the most important aspects in the project. The information on the position of specific users may be collected and stored in the system, both outdoor and indoor (for the pilots that plan to install an indoor location base service). Knowing the position of a user is a key aspect in case of emergency. Positions will be stored as geometries in UNCAP:
    - Points: used to store the position as a single entry (i.e. when the user calls for help)
    - Multipoint: to collect paths with multiple positions (i.e. to store a running session)

With respect to localization data is mandatory to notify the data treatment to the Data Protection Authority. For further details see "D1.2 Regulatory constraints".

7.4. Images and Videos

During the clinical study planned to evaluate the UNCAP solution, the pilot sites of Pergine and Città della Pieve use a video camera-based solution developed by Trilogis for fall detection. The application is able to automatically detect people by processing video streams through computer vision algorithms. Each video camera is directly connected to a PC that is in charge of the processing and forwards to UNCAP an alarming event that someone has fallen.

It has to be highlighted that the video stream is NOT forwarded to anyone/anything and that every frame, once processed, is dropped and deleted: neither the video is stored, nor are images. The video stream, in real time, can be retrieved only by accessing the PC directly connected to the video camera and it is restricted to system administrators for configuration purposes.

The pilot sites of Baia Sprie, Thessaloniki, Simleu, ASL Ovest Vicentino use video cameras integrated with serious games. The images are processed inside the game console and neither recorded nor streamed to any external device.

[3] http://wiki.openstreetmap.org/wiki/OpenRouteService

8. Data storage and Backup

The data hosting service for all the pilot sites is provided by the company Chino.io[4], which offers secure backend and storage solutions complaint with the current EU data protection laws. The company will have the responsibility of ensuring storage, backup and control of access by third parties (data consumers). Chino is a Data Processor as defined in the Data protection EU directive 95/46/EC regulatory, entrusted by the pilot site organizations, acting as Data Controller.

To ensure confidentiality, Chino storage system deploys encryption (256-bit AES) to both the actual stored documents and to the data pathways leading to and out the system. Encrypted data and encryption keys are stored on separate servers.

To allow search operations over encrypted documents, Chino creates secure indexes, which provide an optimal trade-off between security and usability of data. This technology is being developed in collaboration with expert researchers from the Security Research Group of the University of Trento.

In addition to internal system security, Chino provides a flexible cloud infrastructure where API and data storage are deployed ensuring easy scalability. It applies also security safeguards at infrastructure level to protect applications and data against intrusions and cyber-attacks.

Chino relies on an Infrastructure provider located physically in Germany and owned by a German company [5]. The servers have biometric-controlled access. Chino manages redundancy, by replication on multiple servers (at least three), data backup, firewall setup, QoS (Quality of Service) and SLA (Service Level Agreement).

8.1. Specific solutions for UNCAP data storage

Two specific modifications respect to the standard storage solution have been implemented for the UNCAP data:

1. The UNCAP documents have been grouped by user and by data type, instead of saving them in the same repository according to a unique schema. Drawing a parallel with the SQL world, the documents of each user are saved in a different table space (repository), and in each repository there are tables for each data type according to predefined schema (see Figure 1). This approach ensures better scalability, and allows easier control of access permissions to single documents or group of documents. Specifically, each data consumer, e.g., the diabetes application, has its own access permissions for each single repository and data type.
2. The OAUTH authentication has been implemented for data consumers. OAuth[6] is an open standard for authorization, commonly used as a way for Internet users to log in to third party websites using their Microsoft, Google, Facebook, Twitter, One Network etc. accounts without exposing their password. OAuth provides to clients a "secure delegated access" to server resources on behalf of a resource owner. It specifies a process for resource owners to authorize third-party access to their server resources without sharing their credentials. OAuth guarantees a higher level of security respect to simpler token-based authentication. Every data consumer will have specific permissions on single data types for each user.

[4] https://www.chino.io/
[5] https://www.hetzner.de/en/
[6] http://oauth.net/

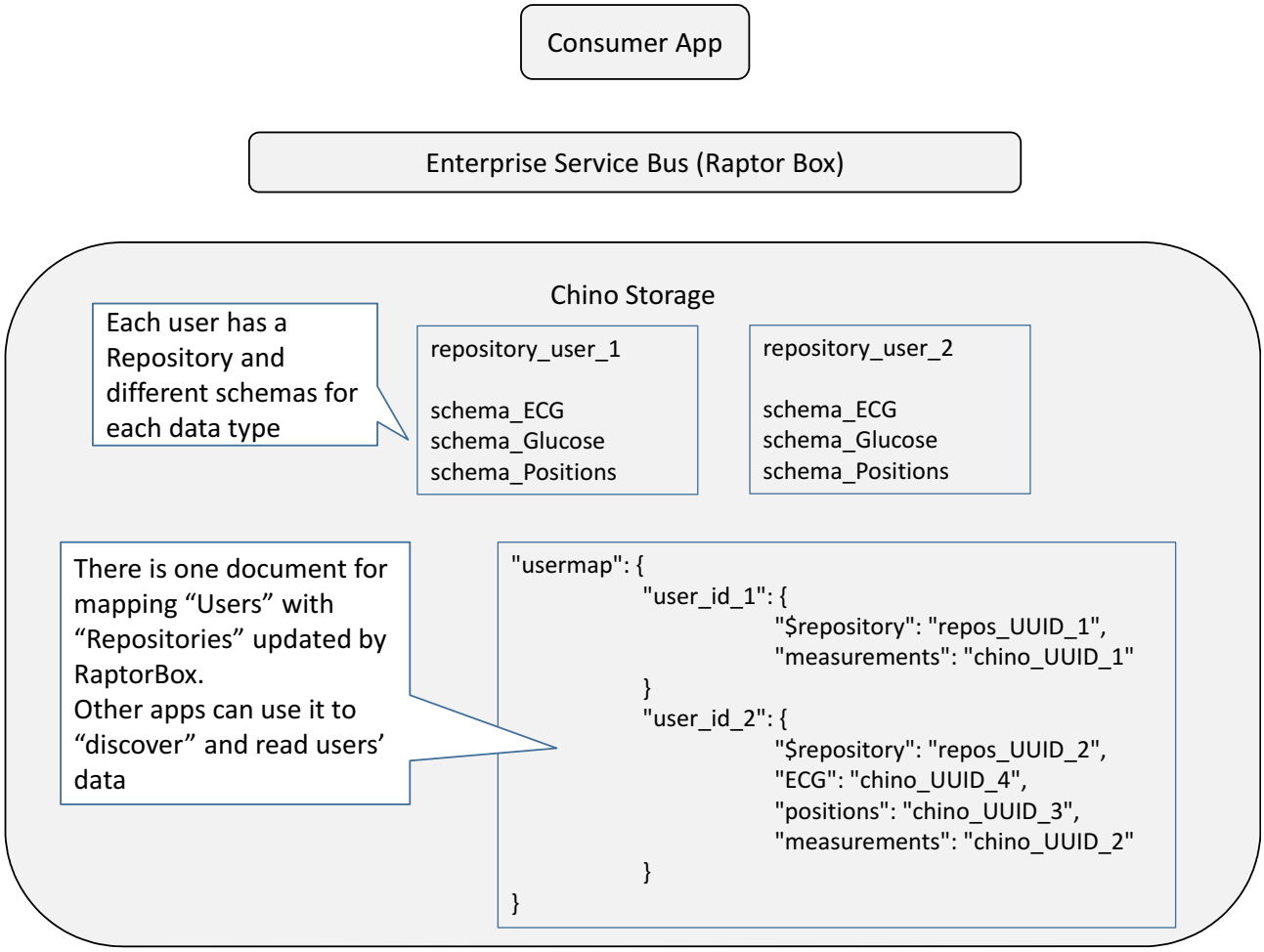

Figure 1: Chino data storage schema.

.

9. Data Access and Sharing

Data transmission security is ensured by channel level encryption using cryptographic protocol TLS (Transport Layer Security). Each operation over data is securely audited to ensure accountability.

The access to data is allowed to the Raptor Box and to authorized users according to a granular authentication and authorization schema.

9.1. Authentication and Authorization

9.1.1. Authentication

Raptor Box, is a custom Enterprise Service Bus acting as a broker between the Consumer App and the Chino storage (see Figure 1). Raptor Box uses access key with admin privilege, to save, delete, modify the data and define the access permissions of the users that access using OAuth.

Users, authenticated through OAuth2.0, can access only the data which allowed by the authorization profile assigned by the pilot sites. Access permissions can be defined at schema level or at document level allowing to define which data type a user can access. Full documentation of authentication process is described here: https://docs.chino.io/#user-authentication.

Each application accessing to data on Chino needs either to user access Keys or impersonate a User and use the OAuth2.0 protocol. The User creation process and setup is described here in the documentation here: https://docs.chino.io/#user-schema.

9.1.2. Authorization

The access to data is governed by granular access control policies, also called Permissions on Chino. Each User accessing to data via OAuth2.0 protocol has assigned permission policies which follow the CRUDA schema: Create, Read, Update, Delete, Administer.

A Permission policy can be defined for a User on a single Document or on a group of documents in the same Schema, or to a Repository (next section describes the data structure).

For more details on Permissions refer to the documentation section here https://docs.chino.io/#permissions.

9.2. Data Structure

Chino API implements a noSql document-oriented database such as MongoDB. It reuses concepts that are familiar to developers to facilitate its adoption.

The main concepts defining the data structure are:

- **Repository** is used to organize data. A Repository is conceptually similar to a DB Schema in SQL jargon. In Uncap each end-user's data are stored in different Repositories, helping to organize easily the data and permissions.
- **Schema** is used to describe the Document structure or content (i.e. list of fields). Chino requires explicit Schema definition in order to implement content verification

and data indexing. If a field is indexed then search operations can be over Documents in an extremely fast way, even tough Documents are always encrypted.

- **Document** contains the actual data in JSON format. Each Document is associated to a Schema. Document fields can be Integer, Float, String, Text, Boolean, Date, Time, DateTime (ISO 8601), JSON, Base64, BLOB. A BLOB field is a link to an attachment that can contain any data representation (e.g. XML, HL7 or binary data).

Each Document that is saved on Chino for Uncap project has the following structure:

```
{
    "description": "Positions",
    "structure": {
        "fields": [
            {
                "type": "string",
                "name": "type"
            },
            {
                "type": "integer",
                "name": "timestamp"
            },
            {
                "type": "string",
                "name": "user_id"
            },
            {
                "type": "string",
                "name": "device_id"
            },
            {
                "type": "text",
                "name": "payload"
            }
        ]
    }
}
```

A Document example:

```
{
    "repository_id": "b52231ef-xxxxx-xx-xxx-xxxxxxx",
    "schema_id": "eff65107-0729-4xxx-xxx-xxxxxxx",
    "document_id": "ad4d3aaa-3120-xx-xxxx- eff65107",
    "insert_date": "2016-10-06T08:06:42.317Z",
    "is_active": true,
    "last_update": "2016-10-06T08:06:42.317Z",
    "content": {
        "user_id": "2",
        "timestamp_begin": 1475740581,
        "doctor_id": "2",
        "timestamp_end": 1475740581,
        "type": "Alarm",
        "payload": [
            {
                "asset_id": "042e868c-2fdf-4f2a-a48f-c649d3637f86",
                "direction": "IN",
                "user_id": "2",
                "description": "GeofencingModule event triggered.",
                "fence_id": "TRI_Reception"
            }
        ],
        "device_id": "1448371967817d7d76133ba5c4c38a5a8ba5607279d81"
    }
}
```


All personal data will be held until the end of the project and then destroyed.

9.3. Procedure to Retrieve Data

The procedure to start accessing to the data stored on Chino is:

- Obtain access (either access Keys or OAuth2.0 username and password)
- Get the access to the UserMap document, which contains the mapping of the Uncap user-id and the Chino Repository where the data are stored. The Repository contains also the Schemas and descriptions about their content.
- Once the Schema containing the data is identified, a data consumer can:
- Search specific documents using the Search API: https://docs.chino.io/#search-api
- Retrieve all documents using the Get Document API: https://docs.chino.io/#documents

10. Open Access to Data

The data content currently stored on Chino and managed by RaptorBox and third party applications is to be considered privacy sensitive content which fully describes the health status of a citizen.

With the following data structure Uncap can easily engage with any application and data consumer for the **primary use of data**, which is typically related to the provision of assistance services to citizens.

Regarding the **secondary use of data**, which is typically associated with open access, statistics, data mining, analytics etc., the current data content needs to be anonymized by ensuring that:

- all **identifiers** (e.g. name, surname, SSID etc) and **pseudo-identifiers** (e.g. unique numeric identifiers) are removed from each data point
- all **quasi-identifiers** (zip, geolocation, addresses, device identifiers, hospital locations etc) are removed from a specific dataset that needs to be analyzed. This is particularly difficult practice since it depends on the actual dataset that needs to be analyzed, and the bigger is the dataset the bigger is the probability to gather private information. For more info consider: https://en.wikipedia.org/wiki/Quasi-identifier

As a summary, before using Uncap data for the secondary usage, the data must be protected by applying state of the art anonymization techniques and recommendations from the Article 29 Working Party, the EU advisory body on privacy topics: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

The most practically feasible approach for Uncap project to provide information in Open Access format is to extract directly statistical data according to predefined KPIs and distribute relevant statistics to interested stakeholders. This approach would reduce the risks of privacy violations and misuse of data.