

Services and Applications for a Healthy Ageing Servicios y Aplicaciones para Envejecimiento Saludable



Este proyecto ha sido concedido por la Agencia Estatal de Investigación del Ministerio de Ciencia e Innovación dentro del procedimiento de concesión de ayudas a 'Proyectos de I+D+i' en el marco de los programas estatales de generación de conocimiento y fortalecimiento científico y tecnológico del sistema de I+D+i y de I+D+i orientada a los retos de la sociedad, convocatoria 2020.



Proyecto de investigación financiado por
MCIN/AEI/10.13039/501100011033

Data Management Plan

WP4: Data analytics.



atlanTTic

research center
for Telecommunication Technologies



1- Introducción	3
2 - Resumen de los datos.	4
2.1 - Descripción del objetivo de los datos.	4
2.2 - Tipo y formato de los datos.	4
2.3 - Origen de los datos.	5
2.4 - Volumen esperado de los datos.	5
2.5 - Utilidad de los datos.	5
3 - FAIR data	6
3.1 - Encontrabilidad	6
3.2 - Accesibilidad	6
3.3 - Interoperabilidad	7
3.4 - Reutilizabilidad	7
4 - Asignación de recursos	8
5 - Seguridad de los datos	9
6 - Aspectos éticos	10
7 - Personal participante.	11
8 - Lista de acrónimos.	12

1- Introducción

A continuación se plantea las secciones y apartados que componen el plan de gestión de datos, un documento que se elabora por el equipo de investigadores del proyecto SAPIENS con el objetivo de describir todos los aspectos relacionados con la gestión de los datos, tanto durante el desarrollo del proyecto como después de concluido. En el PGD se describe el ciclo de vida de la gestión de los conjuntos de datos que se recopilan, procesan y generan en el proyecto de investigación, además, debe incluirse la metodología y estándares que se utilizarán, como se compartirán y serán abiertos los datos y de qué manera se conservarán y preservarán. El PGD es un documento que evoluciona y adquiere mayor precisión a lo largo del desarrollo del proyecto.

Los planes de gestión de datos persiguen un triple objetivo:

- garantizar que no se pierdan los datos, sobre todo al finalizar un proyecto de investigación.
- facilitar su correcta custodia desde su producción.
- permitir su preservación mediante documentación de todo su entorno tecnológico.

A su vez los elementos del plan de gestión de datos son:

- El tipo de datos que se producirá durante la investigación.
- Las normas que se utilizarán.
- Las políticas de acceso y reutilización de datos.
- Las medidas para proteger la privacidad, la seguridad, la confidencialidad y la propiedad intelectual.
- El archivo y la preservación de los datos.

2 - Resumen de los datos.

Los árboles de decisión son una técnica comúnmente utilizada en la implementación de agentes conversacionales inteligentes para procesar la entrada del usuario y generar respuestas automatizadas.

En términos generales, un árbol de decisión es una estructura de datos que representa una serie de decisiones y sus posibles consecuencias. En el contexto de los agentes conversacionales, un árbol de decisión se puede utilizar para modelar los diferentes caminos que puede tomar una conversación y las posibles respuestas que el agente puede dar en cada punto.

2.1 - Descripción del objetivo de los datos.

El objetivo de los datos usados (y del proyecto en sí) estará relacionado con la evaluación cognitiva (memoria a medio/largo plazo, atención, cálculo, fluidez verbal, orientación, etc.). Para esa tarea, se almacenará la información procedente del usuario bajo estudio que se capture a través de los altavoces inteligentes, excluyendo información personal. Todas las posibles respuestas del usuario estarán predeterminadas, serán genéricas y estarán solamente encaminadas a la detección temprana de enfermedades neuropsiquiátricas. Adicionalmente, se almacenarán encuestas de datos demográficos en formato físico.

2.2 - Tipo y formato de los datos.

Los archivos gestionados se corresponden con la información recolectada a través de dichos altavoces inteligentes. Estos datos serán almacenados en ficheros digitales de tipo JSON, del orden de 1 MB por participante. Estos datos serán observacionales.

Los datos almacenados en formato físico incluirán solamente el género, edad y años de estudios de los participantes.

Se estima que se almacenarán datos de un máximo de 100 usuarios.

No se reutilizarán datos de otros entornos o proyectos.

2.3 - Origen de los datos.

Los datos serán recopilados por los investigadores (mediante las encuestas) y capturados mediante los altavoces mencionados.

2.4 - Volumen esperado de los datos.

Se calcula que el volumen de información no superará los 10GB en todo el ciclo del proyecto.

2.5 - Utilidad de los datos.

Estos datos se usarán para la evaluación/detección de deterioros cognitivos en los participantes por parte de los investigadores. Una vez anonimizados y liberados los datos, estos podrán ser usados por la comunidad de científicos para fines similares.

3 - FAIR data

Se proporcionará una URL dentro de la propia institución Universidad de Vigo en la que estarán disponibles todos estos datos, una vez hayan sido anonimizados.

El nombre de cada archivo tendrá la siguiente estructura:

ID_NUM.json

Donde ID es el identificador del usuario y NUM el número de la interacción.

La ubicación de los datos será publicada en la web oficial del proyecto y su acceso permitido a la comunidad científica.

Dado que la información es puramente observacional, no da lugar a la existencia de diferentes versiones de los datos.

Se generará un documento describiendo el significado de las etiquetas que componen los JSON. Para ello, se tomará como referencia el estándar para el intercambio de datos de salud FHIR, con el objetivo de maximizar la interoperabilidad del sistema.

3.1 - Encontrabilidad

Los datos estarán alojados en una URL publicada en la web oficial del proyecto.

3.2 - Accesibilidad

Se encontrarán bajo el formato JSON. Por lo tanto, para su uso e manipulación no se imponen restricciones derivadas de ninguna plataforma o soporte. Para su acceso bastará con cualquier navegador (incluso basados en texto) y software de cualquier origen capaz de gestionar ficheros en formato JSON.

3.3 - Interoperabilidad

Se generará un documento describiendo el significado de las etiquetas que componen los JSON. Para ello, se tomará como referencia el estándar para el intercambio de datos de salud FHIR, con el objetivo de maximizar la interoperabilidad del sistema.

3.4 - Reutilizabilidad

No se ponen limitaciones al uso de estos datos, una vez debidamente anonimizados. Esto será materializado 3 meses antes del fin del proyecto.

4 - Asignación de recursos

Para su almacenamiento en la nube, se hará uso de los recursos de la propia institución por lo que no tendrá coste económico para el proyecto.

Los investigadores principales del proyecto serán los responsables de esta tarea.

5 - Seguridad de los datos

Dado que el volumen de información a manejar se considera limitado y acotado, la tarea de realizar copias de seguridad se prevé sencilla. Para ello se hará uso de DynamoDB. En este soporte se realizan copias de seguridad protegidas por login/contraseña.

Estas operaciones se realizan en equipos de la propia universidad, que es la responsable de la seguridad de las instalaciones.

6 - Aspectos éticos

Los procesos para la recogida de datos están avalados mediante la aprobación del protocolo por parte del Comité Ético del SERGAS y la obtención del consentimiento informado de los participantes.

A cada participante se le asignará un número identificador aleatorio, con el único objetivo de hacer el enlace con sus datos demográficos correspondientes. En ningún caso se almacenará ni en la nube ni en local, datos personales que permitan identificar al sujeto.

El consentimiento informado indica que los datos anonimizados se compartirán a través de un repositorio de datos accesible libremente, para otros investigadores.

7 - Personal participante.

Equipo de investigación:

- Investigadores Principales (IPs)
 - ❖ Luis Eulogio Anido Rifón (Uvigo)
 - ❖ Luis Modesto Álvarez Sabucedo (Uvigo)
- Equipo de Investigación (EI)
 - ❖ Manuel Fernández Iglesias (Uvigo)
 - ❖ Martín Llamas Nistal (Uvigo)
 - ❖ Manuel Gandoy Crego (USC)
 - ❖ Marcos Pazos Couselo (USC)
 - ❖ Manuel Caeiro Rodríguez (Uvigo)
 - ❖ Juan Manuel Santos Gago (Uvigo)
 - ❖ Fernando Ariel Mikic Fonte (Uvigo)
- Equipo de Trabajo (ET)
 - ❖ Alba Elena Martínez Santos (USC)
 - ❖ César Bugallo Carrera (USC)
 - ❖ Sonia María Valladares Rodríguez (USC)
 - ❖ Martín Liz Domínguez (Uvigo)
 - ❖ Moises Ruben Pacheco Lorenzo (Uvigo)
- Personal Contratado (PC)
 - ❖ Iván Otero González

8 - Lista de acrónimos.

DMP	Data Management Plan
FHIR	Fast Healthcare Interoperability Resources
JSON	JavaScript Object Notation
PGD	Plan de Gestión de Datos
URL	Uniform Resource Locator