



Open-red Plan de Gestión de datos



M10-v2

Autor:

Francisco Sanz
(Fundación Ibercivis)

Fecha:

01/10/2024



Versión	Comentarios	Fecha	Autores
0.1	Estructura del documento		Francisco Sanz
0.2	Versión preliminar		Francisco Sanz
1.0	Versión maquetada		Ignacio Saenz
2.0	Actualización octubre 2024	01/10/2024	Francisco Sanz



Resumen ejecutivo

El proyecto Openred, fruto del convenio entre el Consejo de Seguridad Nuclear y la Fundación Ibercivis, se enfoca en desarrollar una red de ciencia ciudadana para la medición de radiación gamma ambiental en España. Los objetivos centrales incluyen mejorar la comunicación en protección radiológica y fomentar la participación ciudadana, educando sobre la radiactividad y sus principios de protección a través de actividades participativas que utilizan dispositivos e infraestructura validada, y la formación de una comunidad activa mediante la recopilación de datos.

El proyecto genera y recolecta una variedad de datos en múltiples formatos:

- Datos cuantitativos de medición de radiación gamma, recopilados automáticamente y manualmente, almacenados en SQL en la Universidad de Zaragoza.
- Contribuciones textuales de experimentos de inteligencia colectiva y datos cualitativos de encuestas de percepción, gestionadas a través de ThinkHub y LimeSurvey, respectivamente.
- Información de contacto para comunicaciones internas y externas, recolectada para newsletters y operación de herramientas, almacenada en bases de datos seguras y gestionada con estricto cumplimiento del RGPD.

El proyecto asegura que los datos sean localizables, accesibles, interoperables y reutilizables (FAIR):

- Implementa identificadores únicos y metadatos enriquecidos para facilitar la localización de datos.
- Ofrece datos accesibles a través de una API pública post-validación, asegurando el acceso mediante protocolos abiertos y condiciones claras de uso.
- Utiliza estándares y vocabularios controlados para la interoperabilidad.
- Proporciona documentación detallada y licencias claras para promover la reutilización de datos.

Se han establecido medidas de seguridad robustas, incluyendo conexiones seguras, sistemas de autenticación y estrategias de respaldo para proteger los datos. Además, se manejan cuestiones éticas relacionadas con la protección de datos personales conforme al GDPR y el consentimiento informado, asegurando una gestión de datos responsable y transparente.



Los costos asociados con la gestión de datos FAIR están cubiertos por el presupuesto del proyecto, lo cual incluye fondos para metadatos, almacenamiento seguro y tecnologías necesarias para la accesibilidad e interoperabilidad. La Fundación Ibercivis, bajo la dirección de Francisco Sanz y con el apoyo de un comité de expertos, es responsable de la gestión de datos, garantizando la adherencia a los estándares de calidad y los principios FAIR.

Este resumen refleja el compromiso de Openred con la gestión eficaz de datos, cumpliendo con altos estándares de calidad y ética, y promoviendo la transparencia y el acceso abierto a la información científica.



Índice

1. Resumen de datos	6
1.1 Propósito de la recolección/generación de datos	6
1.2 Tipos y formatos de datos generados y recolectados por el proyecto	6
2 Datos FAIR	9
2.1 Findable (Localizables):	9
2.2 Accessible (Accesibles):	9
2.3 Interoperable (Interoperables):	10
2.4 Reusable (Reutilizables):	10
2.5 Implementación y monitoreo:	10
3 Asignación de Recursos	11
4 Seguridad de los Datos	12
5 Aspectos Éticos	13
6 Otros aspectos	14
7 Referencias	14



1. Resumen de datos

1.1 Propósito de la recolección/generación de datos

El propósito de la recolección o generación de datos en el proyecto Openred es crucial para cumplir con los objetivos definidos en el marco del convenio entre el Consejo de Seguridad Nuclear y la Fundación Ibercivis. Este proyecto de innovación, centrado en el "desarrollo de una red de ciencia ciudadana para la medida de la radiación gamma ambiental en España", busca alcanzar dos metas fundamentales:

- **Mejorar la comunicación en protección radiológica con la ciudadanía**, educando y concienciando sobre la radiactividad y los principios de protección radiológica a través de la inteligencia colectiva. Esto se logra mediante la organización de actividades participativas y demostrativas que utilizan los dispositivos desarrollados y la infraestructura tecnológica validada. La recolección de datos en este contexto permite evaluar la efectividad de estas actividades y mejorar continuamente los métodos de comunicación y educación.
- **Incorporar a la ciudadanía en una red de medida**, promoviendo la creación de una comunidad amplia y comprometida. Esto implica no solo la participación activa en la medición de la radiación, sino también en la generación de bases de datos, el establecimiento de contactos y la formación de alianzas entre comunidades de usuarios. Los datos recogidos servirán para crear una base de datos de mediciones de radiación gamma de dispositivos low-cost.

Para asegurar que estos objetivos se alcanzan eficazmente, es imperativo implementar un plan de gestión de datos que esté alineado con los estándares de calidad de datos sugeridos por la Comisión Europea en proyectos de ciencia ciudadana. Este plan incluye estrategias detalladas para garantizar la calidad, accesibilidad, interoperabilidad y reutilización de los datos conforme a los principios FAIR.

Además, se toman medidas específicas para preservar las garantías de todos los participantes y asegurar el cumplimiento de la Regulación General de Protección de Datos (RGPD). Esto incluye el establecimiento de procedimientos rigurosos para el consentimiento informado, el manejo adecuado de datos sensibles y la seguridad en el almacenamiento y procesamiento de la información recopilada.



1.2 Tipos y formatos de datos generados y recolectados por el proyecto

El proyecto Openred genera y recolecta una variedad de datos a través de múltiples medios y formatos, enfocándose principalmente en los siguientes tipos:

- **B1 - Datos de medición de radiación gamma:**
 - Tipo de datos: Cuantitativos.
 - Método de recolección: Automático desde dispositivos de medición en posesión de los ciudadanos y manual a través de una aplicación móvil.
 - Formato de almacenamiento: SQL, utilizando bases de datos como MySQL o PostgreSQL alojadas en la Universidad de Zaragoza. Los datos serán validados conforme a metodologías FAIR antes de su disposición a la comunidad científica y al público general.

device-models

device-models_list

AUTHORIZATIONS: *Basic*

>

Responses

✓ 200

RESPONSE SCHEMA: *application/json*

Array [

id	integer (ID)
name <i>required</i>	string (Name) [1 .. 100] characters
manufacturer <i>required</i>	string (Manufacturer) [1 .. 100] characters
version	string or null (Version) <= 50 characters
technology	string or null (Technology) <= 100 characters
description	string or null (Description)
picture	string or null <uri> (Picture)
max_radiation_range <i>required</i>	number (Max radiation range) Maximum radiation range the device can measure (in appropriate units)

]



devices

devices_list

AUTHORIZATIONS: *Basic*

>

Responses

✓ 200

RESPONSE SCHEMA: application/json

Array [

id	integer (ID)
serial_number <i>required</i>	string (Serial number) [1 .. 100] characters
hash	string (Hash) <= 64 characters
purchase_date	string or null <date> (Purchase date)
calibration_date	string or null <date> (Calibration date)
is_active	boolean (Is active)
device_model <i>required</i>	integer (Device model)
owner	integer or null (Owner)

]



measurements

measurements_list

AUTHORIZATIONS: *Basic*

>

Responses

✓ 200

RESPONSE SCHEMA: application/json

Array [

id	integer (ID)
device <i>required</i>	integer (Device)
user <i>required</i>	integer or null (User)
latitude <i>required</i>	number (Latitude)
longitude <i>required</i>	number (Longitude)
altitude	number or null (Altitude)
values <i>required</i>	object (Values)
dateTime <i>required</i>	string <date-time> (DateTime)
accuracy	number or null (Accuracy) Accuracy of the measurement
unit	string or null (Unit) <= 20 characters Unit of measurement (e.g., Sieverts)
notes	string or null (Notes) Optional notes or comments on the measurement
weather	object or null (Weather) Weather conditions during the measurement (if available)
measurement_id	string <uuid> (Measurement id)

]



- **B2 - Contribuciones en los experimentos de inteligencia colectiva:**
 - Tipo de datos: Textuales (contribuciones y observaciones de los participantes).
 - Método de recolección: A través de la herramienta de inteligencia colectiva ThinkHub.
 - Formato de almacenamiento: Similar al de los datos de medición, almacenados en bases de datos SQL para facilitar el análisis posterior.
 - Esta base de datos podría ser publicada una vez anonimizada.

- **B3 - Encuestas de percepción y participación:**
 - Tipo de datos: Cualitativos (feedback de los participantes).
 - Método de recolección: A través de la plataforma LimeSurvey, gestionada por la Fundación Ibercivis.
 - Formato de almacenamiento: En base de datos MySQL
 - Consideraciones de privacidad: Las encuestas serán anónimas para garantizar la privacidad y el cumplimiento de la RGPD.
 - Esta base de datos podría ser publicada una vez anonimizada.

- **B4 - Base de datos de contactos:**
 - Tipo de datos: Datos de contacto de:
 - Grupo de expertos
 - Entidades colaboradores en el hackathon
 - Otros grupos interesados
 - Método de recolección: A través de formulario en la web del proyecto. Formulario de contacto del proyectos. Contactos de Ibercivis y del grupo de expertos.
 - Formato de almacenamiento: Ficheros CSV.
 - Consideraciones de privacidad. Estricto cumplimiento de los principios de la RGPD, asegurando la protección y privacidad de los datos personales.
 - Esta base de datos será interna y no será publicada.

- **B5 - Datos de contacto para newsletter y comunicaciones:**
 - Tipo de datos: Datos de contacto para envío de newsletters.
 - Método de recolección: A través de la aplicación Sendy, con un formulario en la web del proyecto.
 - Formato de almacenamiento: MySQL en servidores de Ibercivis.
 - Consideraciones de privacidad: Será necesario recolectar el correo electrónico para enviar la newsletter, así pues habrá un estricto cumplimiento de los principios de la RGPD, asegurando la protección y privacidad de la información.
 - Esta base de datos será interna y no será publicada.

- **B6 - Datos personales relacionados con el uso de las herramientas:**
 - Tipo de datos: Datos personales mínimos necesarios para la operación y uso de la infraestructura de medición de radiación gamma.



- Método de recolección: Registro en las aplicaciones para recolección de datos.
 - Formato de almacenamiento: Alojados y gestionados por la Universidad de Zaragoza, cumpliendo rigurosamente con la RGPD.
 - Consideraciones de privacidad: En consonancia con el RGPD, se pedirán los datos mínimos que permitan un correcto uso de las aplicaciones e infraestructura.
 - Esta base de datos será interna y no será publicada.
- **B7 - Bases de datos internas para análisis de dispositivos de medición y proyectos e infraestructura similares:**
 - Tipo de datos: diversos, dependiendo de las necesidades específicas de análisis internos o colaboraciones con otros proyectos y artículos científicos.
 - Formato de almacenamiento: Principalmente CSV, destinado al uso interno y análisis por parte de los socios del proyecto.
 - Esta base de datos podría ser publicada por su carácter científico como parte de artículos de investigación.



2 Datos FAIR

Los principios FAIR son directrices que aseguran que los datos y metadatos son manejados de manera que sean fácilmente localizables, accesibles, interoperables y reutilizables, maximizando así su utilidad para la comunidad.

En el proyecto Openred, el mayor conjunto de datos que se hará público será el denominado "B1 - Datos de medición de radiación gamma". Estos datos serán alojados en servidores de la Universidad de Zaragoza, utilizando bases de datos SQL para su gestión y almacenamiento. Este formato de almacenamiento garantiza que los datos sean mantenidos de manera segura y estructurada, facilitando su acceso y análisis conforme a los estándares establecidos para la difusión de información científica y cumpliendo con los principios de los datos FAIR.

A continuación se describen los criterios específicos para cada principio FAIR:

2.1 Findable (Localizables):

- **Identificadores únicos:** Cada conjunto de datos recibirá un identificador único, como un DOI (Digital Object Identifier), que facilita su localización.
- **Metadatos enriquecidos:** Los datos estarán acompañados de metadatos detallados, siguiendo el estándar PPSR Core, que incluirán información clave como el tipo de sensor utilizado, la fecha y hora de la medición, y la localización geográfica.
- **Indexación en búsqueda:** Los metadatos serán indexables y se registrarán en bases de datos y plataformas de búsqueda relevantes para asegurar que puedan ser fácilmente encontrados por investigadores y el público.

2.2 Accessible (Accesibles):

- **Protocolos de acceso:** Los datos se acceden a través de protocolos estándares abiertos, como HTTPS, garantizando que sean accesibles sin barreras tecnológicas innecesarias, pero encriptando el tránsito de los datos.
- **Condiciones de acceso:** Los datos serán accesibles a través del API y de la plataforma una vez validados. Las condiciones bajo las cuales los datos pueden ser accedidos estarán claramente definidas y comunicadas a través de la documentación de la API y los términos de uso.
- **Persistencia del acceso:** Aseguraremos la disponibilidad de los datos incluso después de finalizar el proyecto a través de su alojamiento en servidores de la Universidad de Zaragoza.



2.3 Interoperable (Interoperables):

- **Uso de estándares reconocidos:** Los datos y metadatos serán estructurados de acuerdo a estándares reconocidos que faciliten su integración con otros conjuntos de datos.
- **Vocabularios controlados:** Utilizaremos vocabularios controlados y esquemas para describir los datos (e.g., RDF, JSON-LD), facilitando su procesamiento y combinación con otros datos similares.

2.4 Reusable (Reutilizables):

- **Documentación completa:** Proveeremos documentación exhaustiva sobre el contexto, calidad y condición de los datos, lo que permite a otros usuarios entender y reutilizar los datos efectivamente.
- **Licencias claras:** Los datos estarán bajo licencias abiertas que permitan y promuevan su reutilización. Las licencias serán claras en términos de lo que está permitido y lo que no.
- **Uso de metadatos descriptivos:** Los metadatos incluirán información detallada sobre la metodología, los procedimientos y los criterios usados durante la recolección de datos, asegurando que los futuros usuarios puedan reproducir los experimentos o ajustar los datos para sus propios fines.

2.5 Implementación y monitoreo:

- **Validación por comité de expertos:** El comité de expertos revisará y validará tanto los datos como los metadatos para asegurar su calidad y adecuación a los principios FAIR.
- **API Pública y documentada:** La API a través de la cual se proporcionarán los datos será pública y estará bien documentada, asegurando que los interesados puedan acceder y utilizar los datos fácilmente.
- **Revisión y actualización periódica:** Las estrategias para hacer que los datos sean FAIR serán revisadas y actualizadas regularmente para reflejar los avances en tecnología, cambios en las normativas y el feedback de los usuarios.



3 Asignación de Recursos

Los costos asociados con la implementación de prácticas de datos FAIR son variados y cubren varios aspectos esenciales de la gestión de datos. Estos costos incluyen:

- **Creación y mantenimiento de metadatos:** Desarrollar metadatos ricos y detallados que cumplan con estándares internacionales para asegurar que los datos sean localizables, accesibles, interoperables y reutilizables.
- **Preparación de datos para compartir:** Procesos de validación, limpieza y formateo de datos para garantizar su calidad antes de compartirlos públicamente.
- **Almacenamiento a largo plazo:** Inversión en soluciones de almacenamiento seguras y sostenibles para garantizar la preservación de los datos a lo largo del tiempo incluyendo la creación de backups.
- **Tecnologías necesarias para asegurar la accesibilidad e interoperabilidad:** Inclusión de software y plataformas que faciliten el acceso y uso de los datos por parte de diversos usuarios y sistemas.

Los costos asociados al alojamiento de datos, así como los otros costos mencionados, estarán cubiertos por el presupuesto asociado al convenio entre el Consejo de Seguridad Nuclear y la Fundación Ibercivis.

En el proyecto Openred, la Fundación Ibercivis asumirá la responsabilidad de la gestión de datos. Esta tarea será liderada por Francisco Sanz, quien actuará como gestor de datos, coordinando todas las actividades relacionadas con la administración, el almacenamiento, la protección y la disposición de los datos recogidos en el marco del proyecto.

Además, el comité de expertos desempeñará un papel esencial en la validación de los datos antes de su publicación o uso más amplio. Este comité se encargará de asegurar que todos los datos cumplan con los estándares de calidad establecidos y los principios FAIR, proporcionando una capa adicional de supervisión y garantía de calidad para las actividades de gestión de datos llevadas a cabo por la Fundación Ibercivis y el gestor de datos.



4 Seguridad de los Datos

Para garantizar la seguridad de los datos durante su recopilación, procesamiento, almacenamiento y transferencia, el proyecto Openred ha implementado una serie de medidas de seguridad robustas:

- **Conexiones seguras:** La transferencia de datos se realizará a través de conexiones seguras, utilizando protocolos como HTTPS, para garantizar que toda comunicación entre dispositivos y servidores sea segura.
- **Sistemas de autenticación y autorización:** Se implementarán sistemas robustos de autenticación y autorización para controlar el acceso a los datos. Esto incluye mecanismos de control de acceso basado en roles, asegurando que solo el personal autorizado tenga acceso a información sensible.
- **Estrategias de respaldo y recuperación de datos:** Se realizan respaldos regulares de todos los datos para garantizar su disponibilidad continua y para proteger contra la pérdida de datos. La frecuencia de los respaldos está alineada con la importancia y la sensibilidad de los datos, y los almacenamientos de respaldo se encuentran en ubicaciones seguras y geográficamente dispersas para mayor protección.

Los datos recopilados en el proyecto Openred se almacenan en servidores de la Universidad de Zaragoza y de la Fundación Ibercivis. Estos servidores están equipados con medidas de seguridad adecuadas y son administrados bajo metodologías que aseguran su certificación y cumplimiento para la preservación de datos a largo plazo.



5 Aspectos Éticos

En el proyecto Openred, varias cuestiones éticas y legales pueden influir en la capacidad de compartir datos. Estos problemas están principalmente relacionados con la protección de datos personales y la sensibilidad de la información recogida:

- **Protección de datos personales:** La naturaleza de los datos recopilados en OpenRed, particularmente aquellos obtenidos a través de dispositivos de medición en posesión de los ciudadanos, puede incluir información geográfica que podría ser considerada sensible. Por lo tanto, es crucial asegurar que todos los datos personales se manejen conforme al Reglamento General de Protección de Datos (GDPR) de la UE.
- **Información sensible:** Dado que el proyecto implica la medición de radiación gamma, los datos podrían ser utilizados para inferir información sobre actividades humanas o industriales específicas, lo que requiere una gestión cuidadosa para evitar la divulgación de información potencialmente sensible.
- **Revisión ética del proyecto:** Todos estos aspectos se discuten en el contexto de la revisión ética del proyecto para asegurar que se tomen medidas adecuadas para mitigar cualquier riesgo asociado. ¿Está incluido el consentimiento informado para la compartición de datos y la preservación a largo plazo en los cuestionarios que tratan datos personales?

El consentimiento informado es un componente crucial del proyecto Openred, especialmente cuando se trata de la recopilación de datos personales a través de cuestionarios:

- **Proceso de consentimiento informado:** Todos los participantes de quienes se recolectan datos personales deben proporcionar su consentimiento informado. Esto se realiza asegurándose de que los participantes estén plenamente informados sobre cómo se utilizarán, compartirán y preservarán sus datos a largo plazo.
- **Documentación y almacenamiento del consentimiento:** El consentimiento informado se documenta de manera clara y se almacena de forma segura para cumplir con las regulaciones éticas y legales pertinentes. Esto incluye detalles sobre la naturaleza de la información recogida, el propósito del recogido de datos, y las medidas de seguridad implementadas para proteger la información.
- **Comunicación clara y transparente:** Se proporcionará a los participantes información detallada sobre el uso de sus datos, incluyendo cómo pueden acceder a sus datos o retirar su consentimiento en cualquier momento. Esta información se presentará de manera comprensible y accesible para garantizar que todos los participantes puedan tomar decisiones informadas.



6 Otros aspectos

Es importante destacar que este documento no es estático sino que será revisado y actualizado periódicamente para reflejar los avances del proyecto, los cambios tecnológicos, normativos y las retroalimentaciones recibidas. Cada nueva revisión del plan será debidamente documentada y comunicada al Consejo de Seguridad Nuclear (CSN) para asegurar transparencia y conformidad continua con las expectativas y requisitos del proyecto. Esta práctica asegura que el manejo de los datos se mantenga relevante y efectivo a lo largo de toda la duración del proyecto.

7 Referencias

European Commission. (n.d.). Data management - H2020 Online Manual. Retrieved from https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

