

Directrices para la Gestión de Datos en Horizonte 2020

Versión 1.0
11 de diciembre de 2013

Directrices para la Gestión de Datos en Horizon 2020

Versión original en inglés (Diciembre 2013)

Versión en español (Septiembre 2014)

¹Introducción

En Horizonte 2020 se implementará una acción piloto de alcance limitada en relación con el acceso abierto a los datos de investigación. Se requerirá a los proyectos participantes el desarrollo de un Plan de Gestión de Datos (PGD) en el que se deberá especificar qué datos estarán en acceso abierto.

Este documento proporciona unas directrices para la gestión de datos en Horizonte 2020 y está dirigido a los solicitantes y beneficiarios de proyectos incluidos en este programa marco². Estas directrices tienen por objeto ofrecer indicaciones sobre cómo cumplir con los requisitos exigidos en cuanto a calidad, intercambio y seguridad de los datos de investigación.

Presentación y evaluación de la propuesta

Cualquier propuesta de proyecto presentado como “acciones de Investigación e Innovación” así como “Acciones de Innovación” incluirá una sección en Gestión de Datos de Investigación que será evaluada bajo el criterio “Impacto”. Cuando sea pertinente, los solicitantes deberán aportar un breve esquema sobre su política de gestión de datos que contendrá los siguientes aspectos:

1 V. [Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020](#). El propósito del Piloto está indicado también en la introducción del Programa de Trabajo de Horizonte 2020. El piloto de Datos de Investigación en Abierto será supervisado a lo largo de todo el programa Horizonte 2020 con miras a un mayor desarrollo de la política de la Comisión Europea sobre acceso abierto a los resultados de investigación.

2 Información adicional sobre políticas disponible en “The G8 Science Ministers Statement on 12 June 2013”, <https://www.gov.uk/government/news/g8-science-ministers-statement>. La gestión de datos de investigación es relevante, por ejemplo, para el principio reflejado en la declaración: “Los datos de investigación en abierto deberán ser fácilmente descubribles, accesibles, calculables, inteligibles y usables, y siempre que sea posible interoperables conforme a unos estándares de calidad”.

- Qué tipos de datos generará y recopilará el proyecto.
- Qué estándares se utilizarán.
- Cómo serán explotados y/o compartidos/accesibles los datos para su verificación y reutilización. Si los datos no pudieran estar disponibles, es necesario explicar el motivo.
- Cómo se conservarán y preservarán los datos.

La política descrita debe reflejar el estado actual de los acuerdos de consorcio relacionados con la gestión de datos y ser coherente en lo referente a la explotación y protección de los resultados. Por favor, consulte la propuesta Plantilla B para ver la estructura e instrucciones adicionales.

La sección de gestión de datos puede considerarse también como un sistema de control y verificación en un futuro y como una referencia para la asignación de presupuesto y recursos relacionados con la gestión de datos.

Los PGDs se han introducido en el Programa de Trabajo de Horizonte 2020 para el período 2014-15:

La utilización de Planes de Gestión de Datos constituye una novedad en Horizonte 2020, detallando qué datos generará el proyecto, cómo serán explotados, si serán accesibles para su verificación y reutilización, y cómo serán conservados y preservados. El uso de un Plan de Gestión de Datos es obligatorio para los proyectos participantes en el Piloto de Datos de Investigación en Abierto. Otros proyectos podrán presentar un Plan de Gestión de Datos si es relevante para su proyecto de investigación.

En el Portal del Participante se proporcionará una descripción detallada y alcance sobre los requerimientos del Piloto de Datos de Investigación en Abierto.

Los proyectos que participen en el Piloto de Datos de Investigación en Abierto deberán proporcionar una primera versión del PGD en los seis primeros meses del proyecto. Tanto los proyectos participantes en este piloto, como los proyectos que presenten un PGD de forma voluntaria porque sea relevante para su investigación, deberán asegurarse que este documento se mencione en la propuesta. Durante la maduración del proyecto se pueden incluir diferentes versiones del PGD en fases posteriores. La finalidad del PGD consiste en prestar apoyo durante el ciclo de vida de los datos que sean recopilados, procesados o generados en el ámbito del proyecto.

Acuerdo de subvención / Piloto de Datos de Investigación en Abierto en Horizonte 2020

Las referencias a la gestión de datos de investigación están incluidas en el artículo 29.3 del modelo de acuerdo de subvención (artículo aplicado a todos los proyectos participantes en el Piloto de Datos de Investigación en Abierto en Horizonte 2020).

Atendiendo a los datos de investigación digitales generados en la acción ('datos'), los beneficiarios deberán:

(a) depositarlos en un repositorio de datos de investigación y adoptar las medidas necesarias para posibilitar a terceras personas el acceso, la extracción, la explotación, la reproducción y la diseminación, libre de carga para cualquier usuario:

(i) depositar los datos, incluidos los metadatos asociados, necesarios para validar los resultados presentados en las publicaciones científicas, tan pronto como sea posible;

(ii) otros datos, incluidos los metadatos asociados, tal como se especifica, y dentro de los plazos establecidos en el plan de gestión de datos (V. Anexo I);

Un PGD, entendido como un documento en el que se describe cómo se gestionarán los datos de investigación durante y después del proyecto de investigación, es de suma importancia en todos los aspectos para los proyectos participantes en el Piloto de Datos de Investigación en Abierto de Horizonte 2020, así como para cualquier otro proyecto de investigación. Si el proyecto participa en el piloto, hay que incluir una descripción clara y razonada acerca del régimen de acceso que se prevea para los conjuntos de datos recopilados.

Excepcionalmente, los beneficiarios no estarán obligados a asegurar el acceso abierto a determinados datos de su investigación si la consecución del objetivo principal de la acción, tal y como se describe en el Anexo I, pudiera verse comprometida por este motivo. En este caso, el Plan de Gestión de Datos deberá reflejar las razones por las que no se proporciona acceso abierto a estos datos.

Revisiones del proyecto

Un PGD describe el ciclo de vida de la gestión de todos aquellos conjuntos de datos que sean recopilados, procesados o generados por el proyecto de investigación. Es un documento en el que se describe cómo serán manipulados los datos en el transcurso del proyecto de investigación e incluso tras su conclusión, describiendo qué datos serán recopilados, procesados o generados, qué metodología y estándares se utilizarán, cómo se compartirán y serán abiertos, y por último cómo se conservarán y preservarán. El PGD no

es un documento definitivo, sino que irá evolucionando y adquiriendo más precisión y entidad durante el período de vigencia del proyecto.

La primera versión del PGD deberá entregarse durante los primeros 6 meses de marcha del proyecto. Este PGD deberá ajustarse a la plantilla proporcionada por la Comisión (V. Anexo 1). En posteriores etapas del proyecto se podrán entregar versiones más elaboradas y completas. El PGD debería actualizarse al menos una vez hacia la mitad del transcurso del proyecto y otra al final para hacer los ajustes necesarios sobre los datos generados y los usos identificados por el consorcio, ya que no todos los datos o los usos potenciales que se pueda hacer de ellos están claros desde el principio. Se deberán crear nuevas versiones del PGD siempre que haya cambios importantes en el proyecto debidos a la inclusión de nuevos conjuntos de datos, a cambios en las políticas del consorcio o a factores externos. En el Anexo 2 encontrará información adicional acerca de estas versiones más elaboradas.

El soporte técnico en gestión de datos de investigación para los proyectos financiados por Horizonte 2020 se ha planificado a través de proyectos financiados bajo el Programa de Trabajo de Infraestructuras de Investigación 2014-15 (call 3 e-Infrastructures). Los planes exactos para la prestación de servicios estarán disponibles en el Portal del Participante a finales del 2014. Como muy tarde en 2015 estará disponible este servicio de soporte en coordinación con iniciativas nacionales similares. Estos servicios solo estarán disponibles para los proyectos de investigación de Horizonte 2020, dando preferencia a aquellos que participen en el Piloto de Datos de Investigación en Abierto.

Anexo 1: Plantilla del Plan de Gestión de Datos (PGD)

El Plan de Gestión de Datos (PGD) tiene como propósito proporcionar un análisis de los principales elementos de la política de gestión de datos, los cuáles constituirán las directrices de todos los conjuntos de datos generados por el proyecto.

El PGD no es un documento definitivo, sino que se desarrollará a lo largo del ciclo de vida del proyecto.

El PGD debería considerar los puntos abajo indicados que constituirían un *dataset* básico y debería reflejar el estado del conjunto de datos que se producirán.

- **Referencia y nombre del conjunto de datos**

Se creará un identificador para el conjunto de datos.

- **Descripción del conjunto de datos**

Descripción de los datos que serán generados o recopilados, su origen (en caso de que sean recogidos), naturaleza y escala, a quién podrían resultar útiles, si sustentan una publicación científica. Información de si existen (o no) datos similares y las posibilidades de su integración y reutilización.

- **Estándares y metadatos**

Referencia a los estándares existentes para la disciplina. Si no existiesen, un esquema sobre cómo y qué metadatos se crearán.

- **Datos compartidos**

Descripción sobre cómo se compartirán los datos, incluyendo procedimientos de acceso, períodos de embargo en su caso, sumarios de los mecanismos técnicos para su diseminación y software necesario y otras herramientas que permitan su reutilización. Definición de si el acceso será abierto o restringido a determinados grupos. Identificación del repositorio en el que se archivarán los datos. Si ya existen y están identificados, indicación del tipo de repositorio (institucional, temático, etc.) En caso de que los datos no se puedan compartir, han de mencionarse las razones (por ej. éticas, jurídicas, protección de datos de carácter personal, propiedad intelectual, comercial, privacidad y seguridad...).

- **Archivo y preservación (incluido el almacenamiento y copias de seguridad)**

Descripción de los procedimientos que se establecerán para la preservación a largo plazo de los datos. Indicación del tiempo que deberían ser preservados los datos, cuál es su volumen final aproximado, qué costes asociados conlleva y como éstos se planificarán y cubrirán.

Anexo 2: Orientación adicional para Planes de Gestión de Datos

Las indicaciones de este anexo pueden aplicarse a cualquier proyecto que produzca, recoja o procese datos de investigación, y se contemplan como una referencia para la elaboración de Planes de Gestión de Datos en los proyectos Horizonte 2020. Esta guía se estructura en una serie de cuestiones comunes clarificadoras para el conjunto de datos producidos en el marco del proyecto.

Los datos científicos de investigación deberían ser fácilmente

1. Descubribles

a. pregunta PGD: ¿los datos y el software asociados producidos y utilizados en el proyecto son descubribles e identificables por medio de un mecanismo de identificación normalizado (ej. DOI)?

2. Accesibles

a. pregunta PGD: ¿los datos y el software asociados producidos y utilizados en el proyecto son accesibles?, ¿a través de qué modalidades, ámbito, licencias (ej. marco de licencias para la investigación y educación, períodos de embargo, explotación comercial, etc.)?

3. Evaluables e inteligibles

a. pregunta PGD: ¿los datos y el software asociados producidos y utilizados en el proyecto son evaluables e inteligibles para terceros en contextos tales como un examen científico o una revisión por pares (ej. aparece un conjunto mínimo de datos adjunto al *paper* cuando se haga la revisión por pares, se facilitarán los datos de manera que pueda juzgarse su fiabilidad y la competencia de quién los creó)?

4. Utilizables más allá del propósito original para el que fueron recopilados

a. pregunta PGD: ¿los datos y el software asociados producidos y utilizados en el proyecto son utilizables por terceros incluso mucho tiempo después de la recopilación de los datos (ej. están los datos almacenados de forma segura en repositorios certificados para su preservación y conservación a largo plazo; están almacenados junto al software, los metadatos y la documentación necesaria para que sean útiles; los datos son útiles para las

necesidades del público en general; son utilizables para los probables propósitos de usuarios no especializados)?

5. Interoperables conforme a estándares de calidad específicos

a. pregunta PGD: ¿los datos y el software asociados producidos y utilizados en el proyecto son interoperables permitiendo su intercambio entre investigadores, instituciones, organizaciones, países, etc. (ej. adhiriéndose a las normas para la anotación de datos, intercambio de datos, siendo compatibles con aplicaciones de software disponibles, y permitiendo combinaciones con diferentes conjuntos de datos de diferentes fuentes)?