

18. Estrategia de gestión de datos del CITIC

Partimos de la premisa de que la política de gestión de datos de un centro de investigación 100% universitario no puede ser llevada a cabo de forma autónoma por el propio Centro por varios motivos. El primero, y más obvio, porque los datos son elementos sensibles y protegidos, por lo que debe ser una figura jurídica con capacidad de obrar la que ha de responder de su correcto tratamiento. En un caso como el nuestro esta figura sólo puede ser la institución matriz (Universidad da Coruña); dado que nuestro centro carece de personalidad jurídica propia.

A partir de esta premisa la siguiente consecuencia obvia es la necesidad de adecuarse a protocolos y estándares que afectarán no sólo a este Centro, sino a todos los proyectos de investigación de la propia universidad (otros Centros y unidades de I+D+i incluidos), lo cual confiere al objeto de trabajo una dimensión mucho mayor.

Como consecuencia lógica surgen unos requerimientos de gobernanza, gestión, almacenamiento, conservación y puesta a disposición pública muy elevados, de difícil implementación y alto coste por lo que el asunto habrá de tratarse en una escala colaborativa mayor.

A este respecto, la UDC ha empezado a trabajar este tema desde varias dimensiones:

- 1.- Iniciando consultas con el resto de universidades del SUG, vía consorcio universitario, sobre la conveniencia de adoptar modelos de soporte comunes que pudiesen minimizar costes y maximizar los recursos técnicos y financieros disponibles.
2. Buscando el soporte del gobierno autonómico que tiene una unidad, la Agencia para la modernización tecnológica de Galicia (AMTEGA) que puede dar soporte a iniciativas de este estilo.
- 3.- Verificando la capacidad del Centro de Supercomputación de Galicia (CESGA) para ofertar ese servicio en lo referente a protocolos, capacidad de almacenamiento, y una larga serie de elementos de soporte técnicos y operativos.
- 4.- Buscando la financiación extra necesaria para acometer la prestación de este servicio sin afectar a recursos existentes. Tema donde además de la propia capacidad de la universidad debería de haber una aportación estructural a mayores del principal organismo financiador (Xunta de Galicia).

No obstante, asumiendo que todas estas negociaciones tardarán tiempo en dar sus frutos, el Centro ya está trabajando en el diseño de cómo una política general afectaría a su mecánica de trabajo.

Preámbulo

El CITIC está en proceso de consolidación de su propio plan de gestión de datos mientras no se establezca una normativa marco para la red CIGUS.

Todas las áreas estratégicas del CITIC están basadas directa o indirectamente en la adquisición y procesamiento de datos. Se trabaja con datos de bases de datos abiertas y datos privados generados en los propios proyectos desarrollados, incluyendo ámbitos como la industria 5.0, la salud, el medio ambiente o movilidad personal.

Cada proyecto tiene su propio ciclo de vida de datos y cada proyecto tiene la especificidad de su propio plan. Generalmente se incluye el plan de datos en cada solicitud de proyecto o plan de trabajo de contratos con empresas. Se recomienda como guía de desarrollo el documento

Los datos que se utilizan en los proyectos, entendidos como aquellos empleados y/o generados en el transcurso de una investigación, son un elemento fundamental en el ecosistema actual de Ciencia Abierta. La Comisión Europea, en el programa Horizonte Europa correspondiente al MFP 2021-2027, aborda la evaluación de la Ciencia Abierta bajo el epígrafe "excelencia", ya que la considera parte de la metodología y asume el reto de adoptarla como el modo de trabajo del personal investigador. Como consecuencia, y en consonancia con la *Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación*, la *Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario* y la *Estrategia Nacional de Ciencia Abierta 2023-2027*, se establece esta política de gestión de datos. El principal objetivo del presente documento es elaborar un protocolo para tratar los datos de investigación siguiendo los principios FAIR a lo largo de todo su ciclo de vida, así como garantizar la transparencia, reproducibilidad y reusabilidad de la investigación.

Los principios FAIR son una serie de requisitos ampliamente aceptados que todo conjunto de datos debe seguir con el objetivo de cumplir cuatro condiciones básicas de la Ciencia Abierta: que sean fáciles de encontrar, accesibles, interoperables y reutilizables.

Este documento está alineado con las políticas de la Universidade da Coruña en relación con la gestión de datos, que se concretan, entre otras, en su participación en la European Open Science Cloud y en algunas de las acciones recogidas en su Planificación Estratégica 2022-2026, dentro del Objetivo Estratégico 2: *Impulsar una investigación de calidad que potencie la internacionalización, la multidisciplinariedad y el liderazgo de las unidades, agrupaciones, institutos y centros.*

Ámbito de aplicación

El CITIC ha creado el comité de ética, formado por 16 investigadores y presidido por Patricia Faraldo, investigadora del CITIC con perfil jurídico y con amplia experiencia en aspectos éticos, tanto en el ámbito de la IA como en temas generales sobre la adquisición y uso de datos. Es este comité el encargado de elaborar la guía de buenas prácticas sobre la investigación, y por consecuencia de los datos de las investigaciones.

El plan de datos y la guía de buenas prácticas deben ser conocida y aplicada por toda la comunidad CITIC: Consejo Rector y Comité de Dirección, personal investigador permanente y no permanente y personal de gestión. También será de aplicación para el alumnado de doctorado y el personal contratado con cargo a proyectos independientemente de la entidad de gestión del contrato: UDC, Fundación CITIC, Fundación Universidade da Coruña u otras.

En el caso de las actividades con financiación externa, se tendrán en cuenta y prevalecerán las condiciones y mandatos de las entidades financiadoras.

Esta información debe estar incluida en el manual de bienvenida y ser mencionada en todos los actos de acogimiento de nuevos investigadores o investigadores visitantes que desarrollen actividad de datos en el CITIC.

3.- Gobernanza de la estrategia de gestión de datos

La gobernanza de la gestión de datos, dada la pluralidad de elemento implicados en la misma, constituye una de las bases fundamentales para su implementación. El punto crítico es definir en qué manera la política de gestión de datos del Centro ha de relacionarse con la política de gestión de datos de la universidad (y viceversa). A este fin se ha creado una comisión de trabajo de la que participan:

- La Dirección del CITIC
- La dirección de la Red de Centros de Investigación de la UDC (REDCIDI)
- La Biblioteca universitaria
- El equipo de gobierno de la UDC

Esta comisión está trabajando para establecer las relaciones de primer nivel, dejando margen de actuación a cada centro de investigación para definir su propia política interna.

El plan de datos de cada proyecto

La estrategia seguida desde el CITIC para la elaboración de los planes de datos de los proyectos se apoya en los recursos disponibles a nivel de toda la UDC y a nivel técnico de los recursos del CITIC. La comunidad investigadora del CITIC contará, siempre que lo considere necesario y a lo largo de todo el ciclo de vida de los datos de investigación, con el apoyo del personal del Servicio de Biblioteca Universitaria de la UDC, especialmente en todas aquellas cuestiones relacionadas con el cumplimiento de los principios FAIR. El Comité de ética podrá resolver dudas concretas a través del procedimiento establecido para ello. Actualmente es a través de correo electrónico pero se está trabajando en definir el procedimiento estandarizado.

Siempre que existan datos sensibles o de carácter personal se hará especial hincapié en la observancia del Reglamento General de Protección de Datos (GDPR), pudiendo solicitar en todo caso la asistencia del [Comité de Ética de la Investigación y la Docencia de la UDC](#)

En todo momento los investigadores, bajo la responsabilidad del investigador principal o tutor si procede, deberán respetar el [código ético de la investigación de la Universidade da Coruña](#) aprobado por el consejo de

gobierno de la UDC con fecha 27 de febrero de 2029. Desde el 27 de noviembre de 2023 este comité ha conseguido la acreditación por parte del Servicio gallego de salud.

Antes de la investigación

Se elaborará un plan de gestión de datos (DMP) para cada proyecto en el que se recogerá y describirá cómo se llevarán a cabo cada una de las acciones clave de los procesos de recopilación, tratamiento, almacenamiento y difusión de los datos. Además, servirá como guía de buenas prácticas.

Durante la investigación

Se mantendrá el DMP actualizado, coherente con nuevas situaciones no previstas con anterioridad y que puedan surgir durante el desarrollo del proyecto de investigación.

Los datos de investigación se almacenarán de una manera correcta, completa, no adulterada, segura y responsable. En función de las características y el volumen de los datos, éstos podrán almacenarse en los servicios del CITIC y la UDC o en la infraestructura del Centro de Supercomputación de Galicia (CESGA). Se contemplará en todos los casos la realización de copias de seguridad.

Desde el CITIC se presta apoyo para el almacenamiento digital, utilizando el CPD y todas las infraestructuras informáticas. El CDP dispone de sistemas de copia de respaldo de diferentes niveles, con accesos restringidos y registro de uso de los datos. Cada proyecto podrá definir su política de copias de respaldo que será configurada a través del servicio de sistemas del CITIC.

Después de la investigación

Se harán públicos mediante su depósito en la comunidad UDC del repositorio Zenodo y tan pronto como el desarrollo de la investigación lo permita:

- la versión definitiva del DMP
- los datos de investigación y/o sus metadatos, de manera obligatoria para aquellos de los que deriven publicaciones. También deberán depositarse los metadatos y/o los datos seleccionados para su preservación y difusión generados por proyectos de investigación cuyas condiciones así lo indiquen.

Salvo los casos en los que haya que restringir el acceso a los datos por cuestiones de propiedad intelectual o industrial, confidencialidad, seguridad, privacidad o debido a acuerdos con terceras partes, se publicarán los datos de manera abierta, es decir, con una licencia que permita una reutilización amplia, como por ejemplo las licencias CC0 o CC BY. Los datos que se publiquen cumplirán los principios FAIR.

Tanto los DMP como los datos podrán ser, además y complementariamente, depositados en otros repositorios temáticos o, en el caso de que sea posible, en el [repositorio institucional de la UDC](#) (RUC).

Para garantizar la reusabilidad, se fomentará el depósito y publicación de software, código, métodos o cualquier otro resultado de investigación creado a lo largo de un proyecto y relacionado con los datos que se publiquen en abierto.

El CITIC ofrece a su comunidad el gestor de código [Gitlab en servidor propio](#). Se ofrece sin ningún tipo de restricción o limitaciones de proyectos o espacios garantizando los datos y código dentro de los servidores del CITIC.

Responsabilidades

Responsabilidades del personal investigador

- Elaborar, mantener actualizado y cumplir el DMP del proyecto activo del que formen parte.
- Gestionar los datos obtenidos durante la investigación de acuerdo con esta Estrategia, la normativa y/o legislaciones aplicables, los principios éticos y cualquier otro requerimiento contractual.
- Seguir, durante todo el ciclo de vida de los datos de investigación, los principios FAIR.
- Obtener la aprobación del Comité de Ética de la Investigación y la Docencia de la UDC, en el supuesto de que existan datos sensibles o de carácter personal.
- Depositar en la comunidad UDC del repositorio Zenodo tanto el DMP como los datos de investigación de los que deriven publicaciones y/o los generados por proyectos donde así lo requieran las entidades financiadoras.

Responsabilidades del Consejo Rector y del Comité de Dirección del CITIC

- Ofrecer al personal investigador y de gestión del CITIC el apoyo y la formación necesarios para poder cumplir los requerimientos establecidos en esta Estrategia, así como en la legislación y en los mandatos de las entidades financiadoras.
- Velar por los derechos del personal investigador del CITIC y revisar que este cumpla con sus obligaciones en cuanto a los datos.
- En colaboración con la UDC, desarrollar y ofrecer la infraestructura necesaria para el almacenamiento interno y gestión de forma segura de los datos de investigación, asegurando su preservación.
- En colaboración con la UDC, ofrecer la infraestructura necesaria para la publicación y difusión de los datos de investigación, cumpliendo con los principios FAIR. En la actualidad la recomendación es publicarlos en la comunidad UDC del repositorio Zenodo, bien en acceso abierto, bien en acceso restringido, según la naturaleza de la investigación y de los datos en cada caso. El depósito en Zenodo garantiza su preservación gracias al respaldo de la Unión Europea.

Actuaciones realizadas

En el plano de la gobernanza:

Para implementar la política de gestión de datos del Centro se crea la figura del Gestor de datos del Centro. En nuestro caso recae en la subdirección del centro y cuenta con el apoyo puntual de investigadores de todas las áreas científicas.

Son funciones del Gestor de Datos del Centro: coordinar la política de gestión de datos del centro, enlazando los grupos de investigación, las políticas del centro, de la universidad y las de las entidades financiadoras. Sus responsabilidades implican el análisis de las necesidades de gestión de datos del colectivo al que sirven, proporcionar asesoramiento, coordinar los intereses de stakeholders en torno a los datos, formar y difundir buenas prácticas de gestión, desarrollar políticas de GDI del centro.

En el plano de soporte:

El personal técnico de gestión del Centro está trabajando en el desarrollo de plantillas de DMP adaptadas a nuestras condiciones de trabajo más habituales; así como en los procedimientos en cómo se han de migrar los DMP a Zenodo.

El CPD permite dar soporte tecnológico a las necesidades de almacenamiento y control de acceso así como las capacidades de procesamiento.

En el plano de formación:

Desde la Comisión de formación se deben organizar actividades formativas o de divulgación sobre cómo deben gestionarse los datos utilizados desde el centro. Estas actuaciones ya están en marcha.

A través de FEUGA como servicio ofertado desde la Secretaría Xeral de Universidades ya se han iniciado diversas actividades relacionadas con este ámbito.

Dentro del servicio de “*ventanilla abierta*” desde el CITIC se han mantenido conversaciones con expertos sobre el plan de datos.

Según se indica desde FEUGA, se está trabajando en el desarrollo de un plan de gestión de datos, dentro del marco de actividades de la Red CIGUS. Está trabajando en ello Ignasi Labastida para presentar una propuesta de política de gestión de datos para discutir con todos los centros.

Se ha participado en las actividades formativas

- **¿Cómo elaborar un plan de gestión de datos?** Destinatarios: Personal investigador de los centros CIGUS. 10 y 11 de noviembre. Duración: 4 horas
- **Cómo apoyar la elaboración de planes de gestión de datos desde un centro de investigación** Destinatarios: Personal de apoyo a la investigación de los centros CIGUS. Fecha: 22 de noviembre. Duración: 4 horas

A través de la intranet del CITIC los investigadores pueden acceder a los recursos que se ofrecen de apoyo a necesidades de almacenamiento y cómputo de GPUs, solicitud de máquinas virtuales VMWare, VPNs y apoyo para securizar contenidos y accesos: <https://citic.udc.es/intranet/>

Asesoramiento

La comunidad del CITIC podrá solicitar asesoramiento sobre cuestiones relacionadas con los datos de investigación.

El Comité Asesor Externo asesorará al Comité de Dirección y al Consejo Rector del CITIC para la revisión periódica de esta Estrategia y sobre la adopción de nuevas políticas relacionadas.

El Servicio de Biblioteca de la UDC podrá prestar asesoramiento y apoyo al personal investigador del CITIC que lo solicite en cuestiones como la creación de DMP y el uso de herramientas relacionadas (Argos, DMPonline...), depósito de datasets (incluyendo el depósito delegado), selección de repositorios temáticos y de licencias, estándares de metadatos, formatos de los datos para publicación y preservación, nomenclatura, etc.

Cuando en una investigación se trate con datos personales, sensibles o de experimentación con animales, se consultará con el [Comité de Ética de la Investigación y la Docencia de la UDC](#) o con el [Comité de ética de la investigación de Galicia](#), tal y como se establece en la [información complementaria](#) del Comité de ética de la investigación y la docencia de la UDC. En el caso de que se recopilen datos personales también podrá ser consultada la Delegada de Protección de Datos de la UDC.

Se pone a disposición de todos los investigadores el apoyo y las [guías para el desarrollo y ejecución de este plan desde los servicios de la biblioteca de la Universidad](#) y las [“Guidelines on FAIR Data Management in Horizon 2020”](#)

El responsable del tratamiento de los datos es quien determina los fines y los medios del tratamiento de los datos personales. Así, la UDC es la responsable de las actividades de tratamiento de datos que se llevan a cabo en el CITIC y sus datos de contacto son:

Universidad de A Coruña Secretaría General CIF: Q6550005J Dirección postal: Rúa de la Maestranza, 9. 15001 A Coruña Teléfono: 881 011 161 Correo electrónico: rpd@udc.gal

La delegada de protección de datos es la persona designada por el responsable del tratamiento que tiene como función supervisar el cumplimiento de la normativa de protección de datos en la UDC. Sus datos de contacto son los siguientes:

Universidad de A Coruña Luz María Punte Ala (Adjunta a la Secretaría General) Dirección postal: Rúa de la Maestranza, 9. 15001 A Coruña Teléfonos: 881 011 605 y 88 101 161 Correo electrónico: dpd@udc.gal
--