



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*




HOJA DE RUTA DEL CNRS PARA LA CIENCIA ABIERTA

18 de noviembre de 2019



ÍNDICE DE CONTENIDOS

Introducción	4
1. Publicaciones	6
2. Datos de la investigación	8
3. Minería y análisis de textos y datos	10
4. Evaluación individual de los investigadores y la ciencia abierta	11
5. La reconfiguración de la información científica y técnica para la ciencia abierta	12
6. Formación y competencias	13
7. Posicionamiento internacional	14



INTRODUCCIÓN

El movimiento internacional a favor de la ciencia abierta, que comenzó hace más de 30 años, está experimentando un crecimiento sin precedentes desde que la web ha hecho posible difundirlo a escala mundial a costos razonables. La difusión de la producción científica en Internet, su identificación y archivo, permiten eliminar las barreras de acceso permanente sin poner en tela de juicio ni la protección de los datos personales ni la de la propiedad intelectual. Se trata de «(un acceso) tan abierto como sea posible, cerrado en la medida de lo necesario».

La ciencia abierta no sólo promueve un enfoque transversal para compartir los resultados de la ciencia. Al abrir los datos, los procesos, los códigos, los métodos y los protocolos, también ofrece una nueva forma de hacer ciencia.

Existen varias razones científicas, cívicas y socioeconómicas por las que el desarrollo de la ciencia abierta es esencial hoy en día:

- Compartir los conocimientos científicos hace que la investigación sea más eficiente, más visible y menos redundante. El acceso abierto a los datos y resultados está revolucionando la forma de investigar y permite el uso de nuevas herramientas. Las herramientas resultantes de la reciente investigación en inteligencia artificial (minería de textos y datos, aprendizaje automático) también permiten una investigación más transversal, facilitando la interdisciplinariedad.
- La ciencia abierta está cambiando la manera en que la investigación se integra en la sociedad, ayudando a reconstruir la confianza entre el ciudadano y el científico. Su accesibilidad por parte de muchos actores mediante una variedad de disciplinas facilita la respuesta a los desafíos contemporáneos (por ejemplo, los desafíos multi y/o interdisciplinarios) y refuerza la integridad científica.
- Los(as) investigadores(as) a menudo pagan por publicar, suelen pagar por leer los artículos, hacen el trabajo de revisión de forma gratuita y los costos de suscripción son cada vez mayores. La publicación accesible de los resultados científicos devuelve a los científicos el control de un sistema que se nos ha escapado de las manos y que se está volvien-

do insostenible desde el punto de vista financiero debido a los costos cada vez mayores de la difusión científica.

Hace poco más de un año, nuestro país se incorporó en este vasto movimiento de transformación. Presentado el 4 de julio de 2018 por la ministra de la Educación Superior e Investigación, el «Plan Nacional de la Ciencia Abierta pretende, en palabras de Frédérique Vidal, que “los resultados de la investigación científica estén abiertos a todos, investigadores, empresas y ciudadanos sin trabas, sin demora, sin pago”.

En la misma dinámica, se creó un Comité para la Ciencia Abierta (CoSO según sus siglas en francés), presidido por el Director General de Investigación e Innovación (DGRI). Se divide en cuatro colegios: publicaciones, datos de investigación, competencias y formación, Europa e internacional. Los(as) investigadores(as) e ingenieros(as) del CNRS, así como los miembros de la DIST (Direction de l'Information Scientifique et Technique), están presentes en todos los grupos y codirigen tres de los cuatro colegios. El plan nacional de ciencia abierta incluye el compromiso de los operadores de la investigación a desarrollar una política de ciencia abierta.

Dado el carácter interrelacionado de los trabajos realizados por los investigadores y los profesores-investigadores en las unidades de investigación conjuntas, el desarrollo de la ciencia abierta debe pensarse a nivel nacional y en común con todos los actores de la investigación y, en particular, con otros organismos de investigación, así como con nuestros socios universitarios, de manera plenamente coherente con la acción del MESRI (Ministerio de Educación Superior, Investigación e Innovación) para la ciencia abierta.

La hoja de ruta del CNRS para la ciencia abierta se ejecutará mediante un plan de acción que se revisará periódicamente y que estará bajo la responsabilidad de la Dirección de Información Científica y Técnica (DIST), que movilizará en particular a las tres unidades de servicio de su ámbito:

el Inist (Instituto de la información científica y técnica), unidad de servicio propia del CNRS con sede en Nancy desde 1989, que acaba de ser reorganizada en tres departamentos vinculados a tres áreas de desarrollo de la ciencia abierta:

- El acceso a la información científica, uno de los ejes principales e históricos del Inist, con las negociaciones y el portal BibCNRS así como las nuevas plataformas ISTEEX y PANIST para las publicaciones;
- La puesta en valor de los datos de investigación, un área que se ampliará en el futuro y que se beneficia de la inversión pionera del Inist para el desarrollo de herramientas de datos de investigación como DMP OPIDoR (Optimización para Compartir y para la Interoperabilidad de los Datos de la Investigación) para hacer Planes de Gestión de Datos, etc., y también para hacer repositorios de datos para datos «pequeños»;
- El análisis y la minería de la información científica, un eje de futuro para el desarrollo de herramientas de Text and Data Mining.

El **CCSD** (Centro para la comunicación científica directa), unidad conjunta de servicios, principal impulsor de los HAL (HiperArtículos en Línea), los archivos abiertos, y que posteriormente desarrolló Episciences, una plataforma de publicaciones de revistas de acceso abierto y otros servicios; **PERSEE**, creada en 2013 para digitalizar y promover el patrimonio científico.

La política de ciencia abierta se llevará a cabo de forma diferente en las distintas disciplinas. Debe situarse en el núcleo de los institutos del CNRS. En particular, estos institutos gestionan muchas infraestructuras de investigación que producen datos. Algunas son infraestructuras temáticas de

gestión y de puesta en común de datos que figuran en la hoja de ruta nacional para las infraestructuras de investigación y que desempeñan un papel importante en el panorama nacional, europeo e internacional, como el Centro de Datos Astronómicos de Estrasburgo, la plataforma Huma-Num para las humanidades digitales o la infraestructura de investigación Data Terra (antes Sistema Tierra). Los institutos también han desarrollado iniciativas relacionadas con las publicaciones, en particular la plataforma «OpenEdition», para la publicación de revistas y libros en acceso abierto, el centro Mersenne, desarrollado para la publicación en acceso abierto de revistas de matemáticas, y actualmente en proceso de apertura a cualquier disciplina en la que los artículos estén compuestos en LaTeX. La DIST está en contacto permanente con los institutos temáticos a través de los corresponsales de Información Científica y Técnica (IST). Se ha creado una red de «corresponsales de datos» en el marco del nodo francés de la Research Data Alliance (RDA).

La aplicación de la hoja de ruta del CNRS para la ciencia abierta que aquí se presenta pretende acelerar el proceso hacia la ciencia abierta mediante acciones concretas estructuradas en torno a cuatro objetivos principales:

1. Mantener el control sobre nuestra producción científica y lograr el 100% de acceso abierto a las publicaciones del CNRS a escala de la hoja de ruta;
2. Desarrollar una cultura de gestión/intercambio de datos entre todos los actores del ciclo de vida de los datos: investigadores(as), ingenieros(as), informáticos(as), documentalistas, bibliotecarios(as), etc., basada en la aplicación de los principios FAIR (fáciles de encontrar, accesibles, interoperables y reutilizables);
3. Desarrollar y promover infraestructuras y herramientas que permitan la extracción y el análisis de contenidos científicos de manera totalmente independiente;
4. Transformar la evaluación individual de los(as) investigadores(as) para hacerla compatible con los objetivos de la ciencia abierta, por un lado, y, por otro, tener en cuenta la contribución de los(as) investigadores(as) a la ciencia abierta en la evaluación.

Todo este proceso se verá acompañado de actividades de formación y de una estrategia internacional.

1. bit.ly/nationalplanSite

1 | PUBLICACIONES

Objetivo: Las publicaciones científicas, producidas por el trabajo de los(as) investigadores(as) del CNRS, y financiadas en su mayor parte con fondos públicos, deben ser 100% accesibles (*) y reutilizables dentro del plazo de la hoja de ruta. No se deben ceder los derechos de autor.

La estrategia del CNRS para conseguir el 100% de las publicaciones científicas en acceso abierto es promover la «bibliodiversidad»: una diversidad de productos de investigación y una diversidad de métodos de publicación. De hecho, hay varias formas posibles de publicar en acceso abierto. Además de los archivos abiertos, se multiplican los modelos de publicación «virtuosos», en acceso abierto y sin ánimo de lucro. Las relaciones con las editoriales convencionales siempre se tienen en cuenta, pero en un contexto de negociaciones encaminadas a reducir los costos y promover el acceso abierto.

Acción 1: llevar a cabo una política de apoyo y desarrollo del archivo abierto de los HAL junto con una política de fomento del depósito de publicaciones científicas.

Esto incluye el desarrollo del CCSD, la Unidad Mixta de Servicio que dirige los HAL. En 2019, el CNRS puso a disposición recursos adicionales excepcionales por un valor de 650.000 euros, además de una plaza externa abierta al concurso de funcionarios. La perpetuación de este incremento de los medios necesarios para una nueva ambición estará garantizada en 2020 por un apoyo del fondo nacional para la ciencia abierta, luego por una contribución de los usuarios según un modelo económico para el desarrollo adaptado de HAL vinculado a sus numerosas alianzas y que fue validado por el Coso (comité para la ciencia abierta). Esto es coherente con el Plan Nacional, que ha optado por situar en el centro de su política el archivo abierto HAL, desarrollado por el CNRS y los demás organismos supervisores INRIA, INRA y la Universidad de Lyon. Esta infraestructura ha sido adoptada desde hace varios años por un gran número de universidades y organismos de investigación (se han desarrollado 137 portales institucionales).

Todos los estudios relativos a los archivos abiertos muestran que no basta con fomentar el depósito en archivos abiertos, sino que sólo se puede alcanzar un umbral del 80% de los depósitos mediante una política de incentivos más fuerte. Por lo tanto, el CNRS pide que todas las publicaciones científicas resultantes de la investigación financiada esencialmente con fondos públicos y que puedan depositarse en archivos abiertos sobre la base de la ley para una República Digital, sean accesibles en HAL.

En 2019 el CNRS adoptó una primera medida que consistió en que sólo se incluyeran en el Crac 2019 (informe anual de los investigadores) aquellas publicaciones presentes en HAL. La misma medida para el informe Ribac (actividad anual de los investigadores en Ciencias Humanas y Sociales) está prevista para 2020. Esta medida fue anunciada a los directores de las unidades en una carta de la Subdirección General de Ciencia (DGDS) fechada el 19 de abril de 2019.

El aumento de publicaciones depositadas en HAL debe ir acompañado de una mejora de la herramienta HAL para responder a las necesidades de las comunidades científicas francesas. Para ello, el CNRS se compromete a facilitar a los investigadores el proceso de depósito en HAL. Habrá que consolidar o desarrollar la interoperabilidad con otros archivos abiertos internacionales como ArXiv, PubMed Central y RePEc. Las cuestiones relacionadas con la moderación, con los repositorios, con la recolección, también se están considerando en el objetivo de mejorar la infraestructura de investigación de HAL.

Acción 2: Recomendar el uso de servidores de preprints, que alojen los manuscritos enviados a las revistas, para ofrecer soluciones rápidas de difusión de acceso abierto a través de plataformas sin ánimo de lucro.

Siguiendo el ejemplo de ArXiv, precursor en física y matemáticas, muchas comunidades han desarrollado o están empezando a desarrollar servidores de preprints diseñados para la rápida difusión de resultados científicos en paralelo a los procesos de certificación que realizan las revistas. Algunos ejemplos son ChemRxiv en química, bioRxiv en ciencias de la vida y SocArXiv en humanidades y ciencias sociales.

Estas plataformas permiten que cada manuscrito aún no evaluado garantice la atribución a sus autores, la fecha, la identificación perenne, el alojamiento y la atribución de una licencia que defina las condiciones de reutilización. En ocasiones, también permiten el envío directo del manuscrito a una serie de revistas y hacen el enlace con la publicación final si ésta tiene lugar.

(*) esto no pone en tela de juicio ni la protección de los datos personales, ni la protección de la propiedad intelectual, ni cualquier otra protección necesaria. Se trata de ser «tan abierto como sea posible, cerrado en la medida de lo necesario»

Para facilitar estos nuevos métodos de comunicación científica de acceso abierto, el CNRS seleccionará un cierto número de plataformas que serán apoyadas en función de su importancia para las comunidades en cuestión, en coordinación con el fondo nacional para la ciencia abierta. Por otro lado, se esforzará por avanzar en la interoperabilidad entre las distintas plataformas de depósitos. Movilizaremos recursos para que el CCSD proponga soluciones en coordinación con los responsables de estas plataformas y desarrolle las herramientas que permitan a los investigadores realizar un depósito único.

Acción 3: Brindar apoyo a las plataformas de publicación electrónica que albergan textos de acceso abierto y ofrecen facilidades de publicación y certificación

El objetivo es que cada comunidad científica, especialmente las disciplinares, puedan encontrar alternativas para orientar sus publicaciones hacia sistemas de publicación de acceso abierto, con modelos económicos que puedan ser apoyados por las instituciones para que no sea el autor quien pague, ni por publicar ni por leer.

Las plataformas de publicación de acceso abierto ofrecen diferentes tipos de objetos: revistas, libros, recomendaciones, informes de evaluación de los revisores, comentarios de los lectores, anotaciones u otras formas de discusión en torno a las publicaciones. Algunas plataformas son anfitrionas de facto, en el sentido de que alojan y mejoran los contenidos científicos producidos por otras organizaciones (revistas, editores, comunidades organizadas, archivos abiertos), mientras que otras seleccionan y producen directamente sus propios contenidos. En todos los casos, albergan etapas de revisión por pares, que pueden ser más o menos abiertas (informes abiertos, identidades abiertas). La gran mayoría de ellos publican documentos bajo una licencia creative commons, conservando la propiedad intelectual los autores de los textos.

Aunque algunas de ellas pretenden abarcar todas las disciplinas, están marcadas por una historia y unas opciones más restrictivas: por ejemplo, Copernicus está especializada principalmente en geociencias, Scipost en física, el centro Mersenne en matemáticas, OpenEdition en humanidades, E-life y F1000 en ciencias de la vida, Episciences en matemáticas, informática, humanidades, PCI (Peer community In) en ciencias medioambientales. Estas elecciones pueden estar relacionadas con el formato de escritura de los autores, en particular la elección de utilizar Latex u otro software de tratamiento de textos. Es el caso, por ejemplo, del Centro Mersenne y de OpenEdition, respectivamente. En cuanto a las publicaciones, se utilizan diferentes programas informáticos para producir documentos en HTML, PDF, Epub o XML (por ejemplo, Lodel, OJS).

El CNRS apoyará a las comunidades científicas, disciplina por disciplina, en el desarrollo de soluciones de publicación

alternativas con un contexto de publicación y revisión por pares comparable al de las editoriales tradicionales.

Acción 4: Apoyar una estrategia de documentación que tenga como objetivo reducir el precio de las suscripciones a las revistas y aumentar la accesibilidad de los manuscritos en la versión editada.

Cada año, el CNRS gasta más de 12 millones de euros (sin IVA) para el pago de unos cincuenta editores para que sus laboratorios tengan acceso a la documentación científica más allá de las barreras de acceso (portal BibCNRS implementado por Inist). Algunos accesos para las unidades mixtas del CNRS son financiados y propuestos por los socios. Los portales de las universidades y organizaciones pueden otorgar el acceso gratuito. Algunas bibliotecas temáticas financian sus propias suscripciones complementarias (por ejemplo, la red de matemáticas). Por ello, los investigadores dedican cada vez más tiempo a acceder a la producción científica. Las negociaciones llevadas a cabo en el marco de Couperin, de los contratos con los editores que acaparan la mayor cuota de mercado deben dar lugar a una reducción de los costos de las revistas y a una política de apertura de las publicaciones de acuerdo con la ley «Pour une République Numérique». (A favor de una República Digital). El pago de los costos de publicación para que un artículo esté disponible de forma gratuita en una revista que, de otro modo, se vende por suscripción, es un enfoque que debería desaconsejarse enérgicamente, ya que no está en consonancia con el plan nacional de ciencia abierta, y supone costos adicionales que son difíciles de supervisar y controlar para una institución. Se trata de las llamadas revistas híbridas, que no son a priori más virtuosas, ya que en realidad son de doble pago. La publicación en estas revistas, sin pagar derechos de publicación, acompañada de un depósito en HAL preserva la libertad de elección del canal de publicación por parte de los(as) investigadores(as).

Acción 5: Solicitar que todas las publicaciones resultantes de la investigación financiada por una convocatoria de proyectos del CNRS estén disponibles en acceso abierto.

Este objetivo es más bien una acción mínima, concreta e inmediata que debe llevarse a cabo, en consonancia con el plan nacional de ciencia abierta, y con las exigencias de Europa y la ANR en sus respectivas convocatorias de proyectos. El acceso abierto de la publicación podría realizarse según la diversidad de formas posibles indicadas anteriormente: archivos abiertos, plataforma de publicación de revistas de acceso abierto. La vía de los archivos abiertos es una alternativa virtuosa a la exigencia de los APCs («articles processing charges» o gastos de procesamiento de artículos) muy elevados para las revistas llamadas «de prestigio».

2 | LOS DATOS DE LA INVESTIGACIÓN

Objetivo: Los datos (datos brutos, textos y documentos, códigos fuente y programas informáticos) producidos por los(as) investigadores(as) del CNRS o con medios implementados por el CNRS deben, en la medida de lo posible, ser accesibles y reutilizables según los principios FAIR para una consolidación del conocimiento esencial para el desarrollo de una ciencia más eficiente «Los datos deben ser tan abiertos como sea posible, y cerrados tanto como sea necesario»

Acción 1: Desarrollar una cultura de gestión/ intercambio de datos entre todos los actores del ciclo de vida de los datos: investigadores(as), ingenieros(as), informáticos(as), documentalistas, bibliotecarios(as), etc.

Con la actual masificación de los datos digitales producidos durante el proceso de investigación, es esencial desarrollar y compartir buenas prácticas (FAIR) entre los actores de la investigación. Por datos se entienden en su sentido amplio los datos brutos o procesados, los textos y documentos, los códigos fuente y los programas informáticos. Los servicios OPIDoR desarrollados por Inist permiten atribuir identificadores DOI (Digital Object Identifier) a los conjuntos de datos (a través de DataCite) así como un acompañamiento para la aplicación de un plan de gestión de datos (DMP o Data Management Plan). La Comisión Europea, y más recientemente la ANR, solicitan los DMP durante la elaboración de los proyectos de investigación. La oportunidad de identificar a las personas recurso de referencia se refleja en el concepto de «Data Stewardship» (administración de datos). La Misión para el pilotaje y las relaciones con las delegaciones regionales y los institutos (MPR) ha desarrollado una herramienta metodológica para la gestión de datos que puede ayudar a aplicar una política de intercambio y gestión de datos. Crearemos una red de personas recurso, responsables de los datos en las estructuras y proyectos de investigación, formados en el uso de servicios específicos.

Los principios se aplicarán de diferentes maneras en distintos contextos disciplinares, sin dogmatismo y respetando las prácticas establecidas. La propia definición de datos varía mucho de una comunidad científica a otra. Esto incluye un debate en profundidad dentro de las comunidades sobre el ciclo de vida de los datos, su almacenamiento, acceso, archivo, reutilización. Los institutos del CNRS y sus comunidades no están todos al mismo nivel de desarrollo de una cultura de gestión y puesta en común de datos. El reto es sensibilizar a todos los actores de la investigación para lograr la aplicación de los principios FAIR.

Acción 2: Desarrollar la publicación de datos (data papers o artículos de datos), el depósito conjunto de publicaciones y datos y acompañar a los investigadores en el uso de las herramientas de gestión de datos.

Este objetivo es una forma de comprometerse concretamente con la cuestión del intercambio de datos. Los datos en los que se basan las publicaciones deben ser accesibles y reutilizables lo antes posible. Este requisito ya se aplica en el marco del programa europeo H2020 y debe generalizarse a todos los datos producidos con fondos públicos. Por lo tanto, será necesario facilitar el depósito simultáneo de los conjuntos de datos en los que se basa la publicación a la publicación en el archivo abierto (esto no significa que los datos deban depositarse en el archivo abierto). Los conjuntos de datos asociados a las publicaciones se depositarán en un repositorio de datos adecuado, de ser posible un repositorio temático que garantice mejor la calidad de los datos y su difusión. De manera más general, el CNRS fomenta la difusión de todos los datos estructurados depositándolos en repositorios de datos temáticos o generalistas, posiblemente acompañados de la publicación de data papers. Para ello, los institutos del CNRS, en colaboración con el DIST, podrán identificar las estructuras de recursos de su entorno, como las bibliotecas universitarias y las MSH (Maison des Sciences et de l'Homme) en el ámbito de las ciencias sociales y humanas, así como las personas recurso capaces de apoyar a los científicos en estas nuevas prácticas. Estos «malabaristas» encargados de las actividades de conservación de datos prefiguran el reconocimiento efectivo de la «data stewardship (administración de datos)». Los institutos, que tienen una política de apoyo a las revistas, tienen una palanca adicional y en el futuro podrían incluir un capítulo sobre los datos vinculados a las publicaciones..

Acción 3: Apoyar y acompañar a las infraestructuras de investigación que producen datos en la definición y aplicación de políticas de datos.

El CNRS está muy implicado con sus socios en las Infraestructuras de Investigación (IR) nacionales e internacionales, que representan los lugares donde se crean y analizan los datos de la investigación: instrumentos analíticos, infraestructuras de cálculo, infraestructuras de datos, observatorios, etc. Para generalizar la aplicación de los principios FAIR a todas las disciplinas, el CNRS publicará una carta de las infraestructuras, mediante la cual se comprometen a respetar las prácticas FAIR y normas de calidad, mostrando las políticas de datos acordadas con las comunidades científicas que utilizan las infraestructuras en cuestión. Algunas infraestructuras (como Progedo y Humanum en el Instituto de las ciencias sociales y humanidades (INSHS)) ya están participando activamente en este proceso, mientras que otras están en proceso de acompañamiento, como las infraestructuras de investigación de Química. El sincrotrón SOLEIL también ha puesto en marcha una política de gestión de datos. Los ejemplos son numerosos y deben tender a generalizarse. Estos desarrollos deben estar correlacionados con certificaciones (del tipo CoreTrustSeal) en el caso de que las propias infraestructuras se encarguen de la distribución de sus datos.

Acción 4: Apoyar y acompañar a las infraestructuras de datos - Implementar un servicio coordinado con los institutos para promover el depósito de datos para todo el personal de las unidades del CNRS.

Las infraestructuras de datos temáticos desempeñan un papel nacional o internacional. Algunas están incluidas en la hoja de ruta nacional para las infraestructuras de investigación. Esto está en consonancia con la medida de estructuración del Plan Nacional de Ciencia Abierta, que recomienda «desarrollar centros de datos temáticos y disciplinarios». El CNRS seguirá apoyando estas infraestructuras y respal-

dará el desarrollo de nuevos repositorios y servicios de datos temáticos. Este apoyo estará condicionado a una evaluación de su impacto, su relevancia para las necesidades científicas y la calidad de su gestión. Se buscará la certificación CoreTrustSeal.

La puesta en común de datos ya está muy desarrollada en algunas disciplinas; por ejemplo, los almacenes disciplinarios, a menudo internacionales, se han desarrollado durante varias décadas, sobre todo en el ámbito de la astronomía. Este no es todavía el caso de otras disciplinas, para las que la estructuración internacional de los datos de la investigación es menos obvia y/o establecida, y porque la propia naturaleza de los datos producidos (tamaño, heterogeneidad, complejidad, etc.) requiere una reflexión previa sobre la propia naturaleza de los datos. Así, muchas unidades pueden crear una gran cantidad de «small data» (en el sentido de almacenamiento de datos, en contraposición a «big data»). Esto se llama los datos de cola larga. El CNRS estudiará la oportunidad de crear un almacén de datos generalista para albergar los llamados datos de cola larga para los que no se pueden identificar almacenes temáticos. Esta reflexión tendrá lugar en el marco de una reflexión nacional sobre el tratamiento de los datos de cola larga y la medida del Plan Nacional «para desarrollar un servicio genérico de recepción y difusión de datos simples». Dado que la reutilización de los conjuntos de datos recogidos de este modo sólo puede ser efectiva dentro de un marco de confianza reconocido, se aplicará un mecanismo de certificación CoreTrustSeal.

Acción 5: Crear y mostrar un directorio de repositorios y servicios de datos de los que es responsable el CNRS y en los que participa.

Este directorio sería un elemento importante de la contribución del CNRS a la ciencia abierta. Reuniría los servicios de datos de las mencionadas infraestructuras de investigación creadas en el marco de la acción 3, los centros y servicios de datos temáticos (acción 4) y el almacén de datos general posiblemente creado en el marco de la acción 4.

3 | MINERÍA Y ANÁLISIS DE TEXTOS Y DATOS

Objetivo: Facilitar la extracción de textos y datos con el desarrollo de infraestructuras, herramientas y habilidades para el análisis independiente de contenidos científicos.

En las últimas décadas, el volumen de información producido por la investigación ha ido en aumento. En un campo determinado, los investigadores no pueden «digerir» esta información producida, si es que cuentan con los recursos de sufragar su acceso. Los nuevos servicios de apoyo a los investigadores en el ámbito de la información científica y técnica deben evolucionar hacia la minería o análisis de contenidos (content mining) como complemento a la oferta de registro (bases de datos bibliográficas, catálogos) y acceso (portales) a la producción científica. Estos últimos servicios movilizan tecnologías ya probadas (indexación basada en ficheros de autoridad, motores de búsqueda, identificadores, etc.) y bien conocidas en el mundo de la Información Científica y Tecnológica. Por el contrario, el análisis de contenidos requiere el desarrollo y la aplicación de tecnologías que aún no están ampliamente dominadas por los científicos (ontologías, descripción semántica, lenguajes de representación del conocimiento, etc.). Así, la minería de contenidos se define como el acto de explorar textos y datos digitales con la ayuda de programas para extraer información. Este proceso, repetido en distintos corpus, puede producir nuevos conocimientos científicos. La minería de textos y datos es vista por los investigadores como una extensión del «derecho a la lectura».

Acción 1: Apoyar y desarrollar infraestructuras que permitan el análisis de contenidos.

Es importante invertir recursos en actividades de minería de contenidos (content mining). Recientemente, los métodos de aprendizaje profundo (Deep learning) han creado nuevas posibilidades de investigación para tratar datos masivos y de gran dimensión. El CNRS ha orientado la plataforma ISTEEX en esta dirección, llevando la idea de que la recopilación agregada de datos de diversas fuentes ha permitido el desarrollo de herramientas y servicios de exploración de contenidos, de acuerdo con las peticiones de los investigadores. El INIST, con su combinación de competencias en infraestructura técnica y terminológica y conocimientos científicos, es el operador elegido.

El proyecto VISATM (“Vers une Infrastructure de Services Avancés en Text Mining”) desarrollado en el Inist y apoyado por Mesri ha creado un vínculo entre la plataforma ISTEEX y la plataforma europea OpenMinted. Ofrece un panorama de las herramientas de minería de textos disponibles. El proyecto VisaTM pone de manifiesto la necesidad de contar con una infraestructura que proporcione a los investigadores el entorno técnico y científico necesario para llevar a cabo sus investigaciones mediante tecnologías de minería de contenidos.

Acción 2: Marco legislativo: acompañar, traducir e informar

En octubre de 2016, en su ley para una República Digital, Francia ha reconocido y tomado en consideración la necesidad de autorizar, a fines de investigación y educación, todo contenido accesible lícitamente. Sin embargo, esta voluntad no se ha concretado en un decreto de aplicación en la espera del voto de la directiva europea del derecho de autor. En Europa se ha votado recientemente, en marzo de 2019, el texto de la directiva del derecho de autor. En esta directiva se crea una excepción para la investigación al autorizar a practicar la minería de contenidos a fines de investigación en todos los contenidos accesibles lícitamente.

En su transposición en el derecho francés, es necesario prestar especial atención a:

1. La imposición de exigencias para los medios de acceso autorizados para la minería de textos.
2. Los medios puestos en práctica por las plataformas para preservar la integridad de las redes y servidores.
3. El conjunto de los datos de uso (interrogación y descarga) recabados por las plataformas y conservados y que pueden ser eventualmente difundidos.
4. Las licencias negociadas con los editores y la manera en la que sus cláusulas toman en cuenta el marco legislativo del TDM.

Acción 3: desarrollar el uso de herramientas y técnicas de tratamiento de datos y de visualización.

Las diferentes disciplinas comunican entre ellas mediante textos y datos, ya sea que estos últimos estén asociados a textos o no. Esta información es muy importante y voluminosa y es difícil que el ser humano pueda procesarla en su conjunto. Mediante la utilización de estas herramientas, se multiplican las posibilidades de extracción de la información. Las unidades adscritas al CNRS ya se han dedicado a estas problemáticas. Por ejemplo, el Inist del CNRS ha desarrollado técnicas y herramientas de tratamiento de corpus. De esta manera, el Inist propone servicios de extracción de contenidos. Humanum, la TGIR (très grande infrastructure de recherches), las Ciencias Humanas y Sociales, proponen una lista de herramientas para extraer informaciones de los textos. El Instituto de los sistemas complejos de París Ile de France pone a disposición una plataforma que permite, entre otros, la extracción de contenidos y su visualización. Estas técnicas y herramientas deben ser conocidas por los(as) investigadores(as) y los(as) ingenieros(as) en los laboratorios y debe facilitarse su utilización. Estas herramientas son capaces de tratar todo tipo de cuestionamiento: cuestiones científicas reuniendo corpus de ciencias al responder a una consulta en particular o preguntas sobre orientaciones mediante el estudio de todo un conjunto de informes.

4 | LA EVALUACIÓN INDIVIDUAL DE LOS INVESTIGADORES Y LA CIENCIA ABIERTA

Objetivo: Repensar la evaluación individual de los investigadores utilizando una evaluación compatible con los objetivos de la ciencia abierta y teniendo en cuenta en la evaluación la contribución de los investigadores a la ciencia abierta.

Numerosos estudios y publicaciones indican que el actual sistema de evaluación de la investigación es el principal obstáculo para la transición hacia la ciencia abierta. Esto tiene dos aspectos:

1. Las evaluaciones se basan ahora en gran medida en las bibliografías. Evaluar la calidad de las producciones en función del prestigio de la revista o del editor equivale a delegar la responsabilidad de la evaluación en los «revisores o árbitros» movilizados por los editores. En algunas disciplinas, este prestigio se basa en indicadores «cerrados y no verificables», como el factor de impacto de las revistas. En este caso, la responsabilidad de la evaluación se delega en algoritmos «propios» basados en datos «cerrados»;
2. La energía y los recursos que deben desplegar los investigadores para garantizar que su trabajo sea lo más abierto posible no son valorados en absoluto por los procedimientos de evaluación.

Se han tomado varias iniciativas desde la base. Entre ellas se encuentra el llamamiento de Jussieu¹, la Declaración de San Francisco² sobre la evaluación de la investigación (DORA por sus siglas en inglés), redactada en 2012 por un grupo de directores y editores de revistas académicas, firmada por numerosas organizaciones e individuos.

El CNRS la firmó el 14 de julio de 2018. Se trata de un compromiso para evitar el uso de la bibliometría y dar la preferencia a una evaluación más cualitativa, así como para tener en cuenta toda la variedad de tipos de productos de investigación.

Las secciones y comisiones del Comité Nacional deben definir y publicar los criterios que utilizan para evaluar a los investigadores. Al mismo tiempo, el CNRS desea que todos respeten ciertos criterios comunes. Algunos de estos criterios son relevantes en el contexto de la aplicación de las políticas de ciencia abierta.

1. jussieucall.org
2. sfдора.org/read

En este contexto, cada sección y cada comité deberían incluir los siguientes cuatro principios en sus criterios de evaluación:

1. Lo que hay que evaluar son los resultados en sí mismos, no el hecho de que hayan sido publicados en una revista prestigiosa o en otro medio de comunicación de renombre: los miembros del CoNRS (Comité Nacional de la Investigación Científica) deben asumir la responsabilidad de su juicio y no confiar en las evaluaciones anónimas de los editores o en los algoritmos. Esto debe reflejarse en los informes de evaluación.
2. Para cada una de las producciones citadas en los expedientes de evaluación, los investigadores deben explicar el alcance, el impacto y la contribución personal que han hecho a la misma: La lista exhaustiva de las producciones es innecesaria.
3. Todos los tipos de producción deben poder ser elementos de la evaluación³: En particular, siempre que tenga sentido, deben estar disponibles los datos que sustentan a la publicación, así como el código fuente necesario para producir los resultados. Los preprints y otros documentos de trabajo son productos aceptables para la evaluación. Lo mismo ocurre con los artículos de datos («data papers»).
4. Todas las producciones citadas en los expedientes de evaluación deben ser accesibles en HAL o eventualmente en otro archivo abierto⁴: Se trata de las propias producciones y no de sus referencias. Normalmente no es necesario proporcionarlos en el archivo, el enlace activo al archivo debería ser suficiente.

3. Véase las guías de los productos de la investigación del HCERES : www.hceres.fr

4. Se admiten tres excepciones a esta regla:

- Los resultados demasiado recientes pueden estar bajo embargo. En este caso, igual deben haber sido depositados en HAL, con una duración del embargo que no exceda de los previstos por la ley (6 meses en el caso de las STM, y 12 meses en el caso de las ciencias humanas y sociales). Y en ese caso, se facilitan a través de un enlace privado en HAL (o en el expediente).
- Para las contrataciones, esta norma no puede ser absoluta para los candidatos que trabajan en el extranjero en instituciones extranjeras o internacionales o en instituciones privadas.
- Cuando el tipo de producción puede no ser aceptado en HAL.

5 | LA RECONFIGURACIÓN DE LA INFORMACIÓN CIENTÍFICA Y TÉCNICA PARA LA CIENCIA ABIERTA

Durante mucho tiempo, la información científica y técnica se basaba en dispositivos privados, propiedad de empresas con ánimo de lucro que cobraban caro sus servicios. Además de su costo económico, esta configuración limitaba el control de los datos y la capacidad de los operadores públicos para enriquecerlos, fusionarlos y reutilizarlos.

La ciencia abierta, característica, por ejemplo, de la iniciativa Open Citation, debería permitir cambiar esta situación y hacer más visible la investigación realizada en el CNRS, al tiempo que se racionalizan* los procesos de adquisición de información para facilitar la tarea de los investigadores y del personal de investigación (limitando el número de encuestas y las múltiples solicitudes a las unidades) y ayudar a la gestión.

Acción 1: Desarrollar una mayor adhesión de los investigadores a Orcid

El CNRS se unió a Orcid en mayo de 2019 como institución y desea que se adopte ampliamente el uso de este identificador único para el personal científico. Por ello, el CNRS tomará medidas para alentar a su personal científico a obtener un número ORCID y crear su perfil para ganar visibilidad internacional. A cambio, se compromete a aclarar los procesos de intercambio de información entre Orcid y los sistemas de información del CNRS «para simplificarles la vida».

Acción 2: Trabajar en nuevos indicadores bibliométricos

El mercado de las bases de datos sobre publicaciones, patentes u otras producciones científicas está experimentando una importante transformación con, por un lado, la aparición de competidores con actores bien establecidos (Web of science Scopus) y, por otro, el deseo de los financiadores de la investigación, las instituciones de investigación y los editores de hacer más accesibles y utilizables los metadatos de estos objetos, además del acceso abierto a las producciones propiamente dichas. Además, estos nuevos actores, en lugar de presentar una selección de producciones en función del «prestigio» de sus lugares de certificación, pretenden ofrecer imágenes lo más exhaustivas posibles del estado de la producción, sea cual sea el tipo de objeto considerado (artículo, libro, conferencia, etc.) y el lenguaje de expresión. En este contexto, el CNRS reevaluará periódicamente sus necesidades en materia de herramientas y promoverá la adopción y difusión de herramientas coherentes con su política general de ciencia abierta (firma del DORA, llamamiento de Jussieu sobre bibliodiversidad).

(*) Entered by the information holders, no double entry...

6 | LA FORMACIÓN Y LAS COMPETENCIAS

Objetivo: Apoyar y formar a todo el personal de investigación, en particular a los estudiantes de doctorado, y también ofrecer formación a los profesionales de apoyo a la investigación en relación con las ramas de actividades profesionales y las redes de profesionales.

El objetivo de adoptar y apropiarse de los principios de la Ciencia Abierta y de las nuevas prácticas vinculadas al ejercicio de la investigación se logrará si la comunidad científica es acompañada y asistida en este proceso. La comprensión de los nuevos usos y, en definitiva, la ayuda prestada en el acompañamiento de las nuevas prácticas requiere la comprensión de los nuevos retos de la investigación científica abierta.

Las herramientas, servicios e infraestructuras desarrollados por el CNRS, en particular, son aceleradores esenciales de la ciencia abierta. Sin embargo, la mera presencia de estos elementos no significa que la ciencia abierta se convertirá en la norma y se pondrá en práctica. Para ello, los investigadores y el personal de apoyo deben tener los conocimientos y las competencias necesarias para adoptar las nuevas prácticas científicas, conocer y utilizar las herramientas, los servicios y la infraestructura, y acompañar a sus colegas durante la fase de transición.

Competencias en materia de ciencia abierta « **Open Science skills** »¹

Acción 1: Desarrollar las competencias y los conocimientos necesarios para la publicación en acceso abierto

- Plataformas;
- Archivos abiertos (HAL y temáticas);
- Licencias.

Acción 2: Desarrollar las competencias en la gestión de datos de investigación

- Servicios de acompañamiento relativos a la producción, uso/reutilización de datos, difusión de datos según el principio «(datos) abiertos en la medida de lo posible,

cerrados si es necesario»;

- Respeto de los principios FAIR;
- Ayuda para la elaboración de un plan de gestión de datos (data management plan);
- Trabajo sobre los metadatos, normas de interoperabilidad, etc.

Competencias técnicas, en particular en ingeniería de datos, ciencias de datos y gestión de datos

Acción 3: Desarrollar competencias «científicas» para realizar una investigación abierta, incluyendo competencias en materia de integridad de la investigación, ética y derecho

Desarrollar las competencias para actuar dentro y fuera de la propia comunidad científica y disciplinar.

Ya existe una amplia oferta de cursos de formación destinados a desarrollar las competencias en la ciencia abierta. El plan de formación futura debe centrarse en mejorar la calidad y la pertinencia de las competencias. Para facilitararlo, será necesario ofrecer y promover cursos de formación tradicionales y/o en línea para científicos adecuados a su nivel de carrera.

Acción 4: Desarrollar las competencias de acompañamiento en los laboratorios para el análisis y la explotación de los resultados

El CNRS participa en el desarrollo de las competencias necesarias para utilizar las herramientas de análisis de textos y datos y la visualización de los resultados de estos análisis. El objetivo es dar a conocer las nuevas prácticas documentales y reforzar a la comunidad de usuarios de estas herramientas y tecnologías.

1. "Providing researchers with the skills and abilities they need to practice Open Science"- European Commission Open Science skills Working Group – July 2017

7 | POSICIONAMIENTO INTERNACIONAL

Objetivo: Ocupar un lugar en las organizaciones internacionales para definir y compartir estrategias así como buenas prácticas de ciencia abierta (tanto en las publicaciones científicas como en los datos de investigación).

La política del CNRS se inscribe en un entorno internacional. La DIST participa en numerosas actividades y colaboraciones internacionales, ya sea como tal o a través de unidades o individuos. Dentro del CoSO, el Colegio de Europa e Internacional, se encarga de coordinar a los actores franceses en todas las acciones, el monitoreo y la prospectiva transversales en los demás colegios. El CNRS desempeña un papel muy activo en ello.

Acción 1: Alinear las posiciones del CNRS con el marco europeo e internacional de la ciencia abierta en materia de datos.

La Ciencia Abierta es una reflexión muy amplia a nivel internacional llevada a cabo por los organismos de política científica (DG Connect, DG Investigación de la Unión Europea) y por las estructuras de financiación (Science Europe, etc.), y desde la base por las propias comunidades científicas (RDA, GO FAIR, Codata, etc.). La Research Data Alliance (RDA) cuenta con casi 9.000 miembros de 133 países agrupados en 88 grupos de trabajo o de interés. En 2019, expertos del CNRS fundaron el «Nodo francés».

La implementación nacional de estos esfuerzos se organiza en torno al MESRI (CoSO, SPSO (secretaría permanente para la ciencia abierta) y sus diversos órganos colegiados) en enlace con el departamento de gestión de infraestructuras de investigación. El CNRS ya está participando plenamente en estos mecanismos para aportar un punto de vista multidisciplinario y reforzará el compromiso de los institutos que están llevando a cabo iniciativas de apertura de los datos producidos por sus comunidades científicas.

Participar en la implementación del EOSC y de los servicios de datos que se están creando en él es un reto importante para el CNRS.

Acción 2: Apoyar las iniciativas que trabajan para definir los elementos de la FAIRización de datos

La FAIRización de los datos requiere un trabajo a nivel de las disciplinas para definir los formatos de los datos y la manera de describirlos. También es necesario un trabajo más general sobre los aspectos tecnológicos y sociológicos del intercambio de datos. Este trabajo debe llevarse a cabo a nivel internacional. Como se señala en el Plan Nacional de Ciencia Abierta, la RDA describe las buenas prácticas para compartir datos. Puede dar cabida a los debates dentro de las diversas disciplinas cuando no existe un marco internacional para ellos. El CNRS apoyará la participación de sus equipos en la RDA y en las

iniciativas de las disciplinas.

Acción 3: Comunicar con nuestros socios europeos e internacionales sobre las estrategias de publicación científica de acceso abierto.

El intercambio de información científica, la puesta en común de los resultados de la investigación a través de las publicaciones científicas sólo tiene sentido si se lleva a cabo a nivel internacional. La ciencia abierta sólo puede avanzar de forma correlativa al nivel internacional. El CNRS ocupa plenamente su lugar en este entorno global.

Knowledge-exchange es una colaboración original entre seis socios de seis países europeos que valoran la ciencia abierta y tienen un papel importante -aunque diferente- en la construcción, el apoyo o la financiación de infraestructuras y servicios digitales para la investigación y la educación superior. Junto con sus cinco socios (JISC [Reino Unido], SURF [Países Bajos], DFG [Alemania], CSC [Finlandia] y DAF-SHE [Dinamarca]), el CNRS, a través de la DIST, puede así coproducir estudios y recomendaciones, intercambiar prácticas, e incluso coordinar, implicando a expertos del CNRS o de otras instituciones. Por ejemplo, Knowledge-Exchange acaba de elaborar un informe sobre el uso de preprints.

El CNRS, a través de la DIST y del CCSD, es miembro activo de la Confederación Internacional de Repositorios de Archivos Abiertos (COAR), un importante foro de coordinación de iniciativas destinadas a garantizar que la investigación conserve el control de la información que produce. En su seno, se elaboran normas para la interoperabilidad de los futuros repositorios de publicaciones, lo que permitirá desarrollar nuevos servicios. La Asamblea General de 2019 fue organizada por el CCSD en Lyon.

El CNRS participa regularmente en las conferencias de Berlín sobre la alineación de estrategias para fomentar el acceso abierto, que en diciembre de 2018 reunió a participantes de 37 naciones y 5 continentes (China, Sudáfrica, California, Alemania, etc.). Se elaboró un texto de convergencia entre el llamamiento de Jussieu, que aboga por la bibliodiversidad, y el plan alemán OA2020. El reto es alinear los criterios de apoyo a las revistas de acceso abierto, las plataformas de publicación y las infraestructuras, y coordinar los esfuerzos de inversión a nivel estratégico. Cabe señalar que el Plan S ha modificado sus recomendaciones para tener en cuenta los distintos canales de publicación de acceso abierto, especialmente gracias a la presión francesa, y organizando una amplia consulta internacional.

Referencias:

Libro blanco de una ciencia abierta en una república digital, DIST, CNRS (2016):
bit.ly/booksopenedition

Plan Nacional de la Ciencia abierta (4 de julio 2018):
bit.ly/plannational

Turning FAIR into reality, Final report and action plan from the European Commission expert group on FAIR data, European Commission (2018):
bit.ly/turningFAIR

La declaración de San Francisco DORA:
bit.ly/SanFransiscoDOR

3, rue Michel-Ange 75016 Paris
www.cnrs.fr



Publisher: Antoine Petit, Chairman and CEO of the CNRS

Magnetic tape storage library at the National Institute of Nuclear and Particle Physics' Computing Centre (CC- IN2P3). © Cyril Frésillon/CC-IN2P3/CNRS Photothèque

