



Executive Briefing

MANAGING KNOWLEDGE ASSETS FOR RESEARCH AND INNOVATION IN THE 21ST CENTURY

The problem

Research data is the new currency of the digital age. From sonnets to statistics, and genes to geodata, the amount of material being created and stored is growing exponentially. However, the [LERU Roadmap for Research Data](#) identifies a serious gap in the level of preparation amongst research performing organisations. This gap is prominent in areas such as policy development, awareness of current issues, skills development, training, costs, community building, governance, disciplinary/legal/terminological and geographical differences.

The solution

This **LEARN Executive Briefing** will help decision and policy makers identify sound solutions. In addition, stakeholders can follow the [LEARN Toolkit of Best Practice Case Studies](#), all of which will help organisations to grapple with the data deluge. LEARN also provides a self-assessment survey.¹

Research Data Policy

Every research performing organisation should have a research data policy, which lays down a framework for how research data is curated and managed. Research funders should also have a research data policy, stipulating the obligations that a researcher is expected to meet as a condition of the funding received. LEARN has created a model Research Data Management policy for research performing organisations, along with guidance for the implementation of this policy.² The model LEARN policy can be both adapted and adopted by individual research performing organisations, by regional, national and/or international consortia.

FAIR Data

Best practice indicates that research data should be FAIR³:

Findable – **A**ccessible – **I**nteroperable – **R**eusable

To be findable, the data should be adequately described, using standard taxonomies and ontologies where possible. To be accessible, research data should ideally be open data, available for sharing and reuse. Not all research data can be open, but best practice indicates that such data should be “as open as possible, as closed as necessary”⁴. Research data should also be interoperable, capable of being processed by machines using vocabularies which follow FAIR principles. To be reusable, metadata describing the data should meet domain-relevant community standards.

¹ All available at <http://learn-rdm.eu>; last accessed 16/12/16.

² As in n. 1 above.

³ See <https://www.force11.org/group/fairgroup/fairprinciples>; last accessed 12/12/16.

⁴ European Commission - Guidelines on FAIR Data Management in Horizon 2020 p.4 (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf); last accessed 12/12/16.

Research Data Management Stewardship

It is important that researchers plan the collection, curation, description and dissemination of their research data at the start of their research. This information is best captured in a Research Data Management plan, which provides a framework for research data stewardship.⁵

Infrastructure

To curate their research data, researchers and research performing organisations need access to the requisite digital eco-systems. These may be maintained locally; or they may be commercial services, subject domain offerings or regional/national/international platforms. Different subject communities and individual countries will want to provide such facilities in different ways. Commonly, the platform(s) will need to offer the following services:

- Storage, for researchers who are actively collecting data;
- A publication platform, where research data and related software can be made available for sharing and re-use;
- Archive facilities, to allow research data to be curated for the long term, often in response to the requirements of research funders;
- A discovery service, which will allow researchers and citizens to search for research data deposits both locally and across the Internet.

The European Commission is promoting the European Open Science Cloud.⁶ The EOSC is a metaphor to help convey both seamlessness and the idea of a commons based on scientific data. The EOSC will be a federated environment for the sharing and re-use of scientific data, based on existing and emerging elements in the Member States, with lightweight international guidance and governance and a large degree of freedom regarding practical implementation.

Training

The prevalence of research data requires all researchers, new and established, to equip themselves with the skills and tools to be confident in a data-driven environment. The lead needs to be taken by research performing organisations and, in many cases, by their institutional libraries.

Funding

Research data management comes with costs. There is no one method for assessing these costs, but a number of costing models exist to help, for example the 4C Project.⁷

Risks

There are dangers for stakeholders in the research data management landscape if Best Practice is not followed. Researchers may lose funding through lack of compliance with funder requirements. Important

⁵ For further information, see <http://www.dcc.ac.uk/resources/data-management-plans>; last accessed 12/12/16.

⁶ See <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; last accessed 12/12/16.

⁷ For further information, see <http://www.4cproject.eu/summary-of-cost-models>; last accessed 12/12/16.

research results may be lost through carelessness, making it difficult or impossible to validate research outcomes. Partnerships and collaborations cannot flourish where research results are not shared.

Benefits

The benefits of sound research data management are many. The integrity of research findings is enhanced where Best Practice is in place. Research performing organisations can join major global research initiatives such as the European Open Science Cloud. Research data can indeed become the new currency of research communication, alongside research publications, making a contribution to solving the grand challenges which face Society – poverty, disease, global warming.

Conclusion

Research data can drive innovation and stimulate new discoveries, to the great benefit of Society. All stakeholders in the research workflow have a role to play. This Executive Briefing highlights what researchers and research performing organisations need to do to rise to this exciting challenge.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.



Resumen ejecutivo – Spanish translation

GESTIONAR LOS ACTIVOS DEL CONOCIMIENTO PARA LA INVESTIGACIÓN Y LA INNOVACIÓN EN EL SIGLO XXI

El problema

Los datos de investigación son la nueva divisa de la era digital. Desde sonetos a datos estadísticos, y desde genes a datos geoespaciales, la cantidad de material que se crea y se almacena crece de manera exponencial. Sin embargo, la [Hoja de Ruta para los Datos de Investigación de la LERU](#) identifica una brecha significativa en relación al nivel de preparación entre las instituciones donde se realizan actividades de investigación. La brecha es considerable en áreas como el desarrollo de políticas, el conocimiento de los temas de actualidad, el desarrollo de habilidades, la formación, los costes, la creación de comunidades, la gobernanza, las diferencias disciplinarias/legales/terminológicas y geográficas.

La solución

Este **Resumen Ejecutivo de LEARN** ayudará a quien debe tomar decisiones y elaborar políticas a identificar buenas soluciones. Además, todas las partes interesadas pueden seguir el Documento de [LEARN](#) de Instrumentos de Buenas Prácticas creado a partir de casos prácticos, que ayudarán a las organizaciones a capear el diluvio de datos. LEARN también ofrece una herramienta de autoevaluación.¹

Política sobre datos de investigación

Cada institución que realiza actividades de investigación debería tener una política sobre datos de investigación, que establezca un marco para gestionar y conservar los datos de investigación. Las organizaciones que financian la investigación también deberían tener una política sobre datos, estipulando las obligaciones que deben cumplir los investigadores al recibir financiación. LEARN ha creado un modelo de Política de Gestión de Datos de Investigación para las organizaciones que realizan actividades de investigación, junto con una guía para su implementación.² El modelo de política de LEARN puede ser adaptado y adoptado de manera individual por instituciones que realizan investigación y por consorcios regionales, nacionales y/o internacionales.

Datos según los principios FAIR

Las buenas prácticas indican que los datos de investigación deberían seguir los principios FAIR³:

Findable (Encontrables) – **A**ccessible (Accesibles) – **I**nteroperable (Interoperables) – **R**eusable (Reutilizables)

¹ Disponibles en <http://learn-rdm.eu>; último acceso 07/01/17.

² Como en 1 anterior.

³ Consultar <https://www.force11.org/group/fairgroup/fairprinciples>; último acceso 07/01/17.

Para poder ser encontrados, los datos deberían ser descritos de una manera adecuada, utilizando taxonomías y ontologías estándar cuando fuera posible. Para ser accesibles, los datos de investigación, idealmente deberían ser datos abiertos, es decir estar disponibles para ser compartidos y reutilizados. No todos los datos de investigación pueden ser abiertos, pero las buenas prácticas muestran que los datos deberían ser “tan abiertos como sea posible, y tan cerrados como sea necesario”¹. Los datos de investigación deberían ser también interoperables, pudiendo ser procesados por máquinas utilizando vocabularios que sigan los principios FAIR. Para ser reutilizables, los metadatos que describen los datos deberían seguir los estándares de cada comunidad relevante en el dominio.

Administración de los datos de investigación

Es importante que los investigadores planifiquen la recopilación, conservación, descripción y difusión de sus datos al inicio de su actividad investigadora. La mejor manera de registrar esta información es mediante un Plan de Gestión de Datos, que ofrece una buena estructura para la administración de los datos de investigación.²

Infraestructuras

Para conservar los datos de investigación, los investigadores y las instituciones que desarrollan actividades de investigación, necesitan acceder a ecosistemas digitales adecuados. Pueden ser mantenidos localmente, o bien pueden ser servicios comerciales, estar dirigidos a dominios científicos específicos o ser plataformas regionales/nacionales/internacionales. Cada comunidad temática o cada territorio puede optar por ofrecer estas instalaciones de forma diferente. En principio, la(s) plataforma(s) deben ofrecer los siguientes servicios comunes:

- Almacenamiento, para los investigadores que están recopilando datos de manera activa;
- Una plataforma de publicación, que ofrezca al público los datos de investigación y el software relacionado para ser compartidos y reutilizados;
- Servicios de archivo, para poder conservar los datos a largo plazo, a menudo como respuesta a los requerimientos de los financiadores de la investigación;
- Un servicio de localización, que permita a los investigadores y a la ciudadanía buscar los repositorios de datos de investigación a nivel local y a través de Internet.

La Comisión Europea promueve la Nube Europea de Ciencia Abierta (EOSC).³ La EOSC es una metáfora para ayudar a transmitir tanto la fluidez, como la idea de un bien común creado a partir de los datos científicos. La EOSC será un entorno federado para compartir y reutilizar datos científicos, a partir de elementos ya existentes e incipientes en los estados miembro, que contará con pautas internacionales y gobernanza ligeras, así como un amplio grado de libertad en relación a su implementación práctica

Formación

El predominio de los datos de investigación requiere que todos los investigadores, los nuevos y los ya establecidos, se doten de las habilidades y las herramientas para sentirse seguros en este entorno orientado a los datos. Las instituciones donde se desarrolla la investigación deben encargarse de esta formación, y en muchos casos, desde sus bibliotecas institucionales.

⁴ European Commission - Guidelines on FAIR Data Management in Horizon 2020 p.4 (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf); último acceso 07/01/17.

⁵ Para más información, consultar <http://www.dcc.ac.uk/resources/data-management-plans>; último acceso 07/01/17.

⁶ Consultar <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; último acceso 07/01/17.

Financiación

Existen costes asociados a la gestión de datos de investigación. No hay un único método para analizar estos costes pero existen varios modelos que pueden ayudarnos, por ejemplo el que propone el Proyecto 4C.⁷

Riesgos

Si no se siguen las buenas prácticas establecidas para una gestión adecuada de los datos de investigación pueden surgir algunos riesgos para los participantes en esta gestión tales como: la pérdida del financiamiento debido al incumplimiento de los requerimientos establecidos por los financiadores; la pérdida, por negligencia, de resultados importantes del trabajo de investigación, dificultando o imposibilitando la validación de los frutos de la misma. Además, las alianzas y las colaboraciones que pudiesen establecerse, corren el riesgo de no prosperar cuando no se comparten los resultados de una investigación.

Beneficios

Los beneficios de una buena gestión de los datos de investigación son múltiples como por ejemplo: la integridad de los hallazgos de la investigación mejora cuando se siguen las buenas prácticas; las organizaciones que desarrollan actividades de investigación pueden participar en importantes iniciativas globales como la European Open Science Cloud. De hecho, los datos de investigación pueden convertirse en la nueva divisa de la comunicación científica, junto con las publicaciones científicas, contribuyendo a solucionar los grandes retos que afronta la sociedad: pobreza, enfermedades, calentamiento global, etc.

Conclusión

Los datos de investigación pueden impulsar la innovación y estimular nuevos descubrimientos en beneficio de la sociedad. Todos los participantes en el proceso de la investigación tienen un papel que desempeñar. Este resumen ejecutivo resalta lo que los investigadores y las instituciones que desarrollan actividades de investigación necesitan para abordar este apasionante reto.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

Leitfaden für Führungskräfte – German Translation

MANAGEMENT VON WISSENSBESTÄNDEN – EIN ZENTRALES ANLIEGEN FÜR FORSCHUNG UND INNOVATION IM 21. JAHRHUNDERT

Die Problemstellung

Forschungsdaten sind der Rohstoff für die Wissenschaft im digitalen Zeitalter. Von Sonetten zu Statistiken, von Genfaktoren zu Geodaten – die Menge an Materialien, die generiert und gespeichert werden, wächst exponentiell. Dennoch wird durch die [LERU Roadmap for Research Data](#) evident, dass Forschungseinrichtungen in sehr unterschiedlichem Ausmaß für digitales Datenmanagement gerüstet sind. Diese Kluft tritt vor allem im Zusammenhang mit Policy-Fragen, dem Bewusstsein für aktuelle Themen und Problemstellungen, Kompetenzentwicklung, Ausbildung, Kosten, Netzwirkbildung und Governance sowie disziplinären/rechtlichen/terminologischen und geographischen Unterschieden besonders deutlich zutage.

Die Lösung

Der vorliegende Leitfaden **LEARN Executive Briefing** soll Führungskräften und EntscheidungsträgerInnen dabei helfen, tragfähige Lösungen zu ermitteln. Darüber hinaus können sich alle Beteiligten an den „[LEARN Toolkit of Best Practice](#)“ Case Studies orientieren, allesamt Studien, die dazu geeignet sind, Forschungsinstitutionen bei der Bewältigung der Datenflut zu unterstützen. LEARN bietet auch eine Umfrage zur Selbstbewertung im Hinblick auf Forschungsdatenmanagement an.¹

Policy für Forschungsdatenmanagement

Jede Forschungseinrichtung sollte über eine Policy verfügen, in der Regelungen für die Kuratierung und das Management von Forschungsdaten definiert sind. Forschungsförderer sollten ebenfalls über eine Forschungsdatenmanagement-Policy verfügen, in der Verpflichtungen festgehalten sind, die von den ForscherInnen als Leistung für die erhaltene Finanzierung erwartet werden. LEARN hat eine Vorlage für eine Forschungsdatenmanagement-Policy und einen Leitfaden zur Umsetzung erstellt.² Diese LEARN Modell-Policy lässt sich problemlos adaptieren und ist daher für unterschiedlichste Institutionen und Konsortien geeignet, sei es auf regionaler, nationaler und/oder internationaler Ebene.

Daten nach dem „FAIR“-Prinzip

Best Practice im Umgang mit Forschungsdaten bedeutet, dass diese FAIR³ sein sollten:

Findable (auffindbar) – **A**ccessible (zugänglich) – **I**nteroperable (interoperabel bzw. kompatibel) – **R**eusable (nachnutzbar)

¹ Verfügbar unter <http://learn-rdm.eu>; letzter Zugriff 16.12.2016.

² Wie in 1 oben.

³ Siehe <https://www.force11.org/group/fairgroup/fairprinciples>; letzter Zugriff 12.12.2016.

„Findable“ (also „auffindbar“), bedeutet, dass die Daten auf adäquate Weise beschrieben sein sollten, und zwar – wo dies möglich ist - unter Verwendung von Standardtaxonomien und -ontologien. Um „accessible“ (also „zugänglich“) zu sein, sollten Forschungsdaten idealerweise offene Daten sein, also für Sharing und Nachnutzung zur Verfügung stehen. Es können zwar nicht alle Forschungsdaten offen sein, aber Best Practice bedeutet in diesem Zusammenhang, dass der Zugang zu solchen Daten so offen wie möglich und so geschlossen wie nötig sein sollte. Forschungsdaten sollten auch „interoperabel“ bzw. kompatibel sein, in einem maschinenlesbaren Format zur Verfügung stehen und mit einem Vokabular versehen sein, das auf den „FAIR“-Grundregeln basiert. Durch die Beschreibung der Datensätze mit Metadaten, die den jeweiligen disziplinspezifischen Standards entsprechen, unterstützt man das Prinzip der Nachnutzung („reusable“).

Research Data Stewardship

Es ist wichtig, dass ForscherInnen bereits zu Beginn ihrer Forschungsarbeiten die Erfassung, Kuratierung, Beschreibung, Nachnutzung und Dissemination ihrer Forschungsdaten planen. Diese Informationen sind idealerweise in einem Datenmanagementplan erfasst, der somit das Gerüst für einen verantwortungsvollen Umgang mit Forschungsdaten darstellt.⁴

Infrastruktur

Für den adäquaten Umgang mit ihren Forschungsdaten benötigen ForscherInnen sowie Forschungseinrichtungen Zugang zu den erforderlichen digitalen Ökosystemen. Diese können auf lokaler Ebene betrieben werden, es kann sich dabei jedoch auch um kommerzielle Servicedienste, themen- oder disziplinbezogene Angebote, oder um regionale, nationale und internationale Plattformen handeln. Unterschiedliche Wissenschaftscommunities und unterschiedliche Länder werden derartige Einrichtungen möglicherweise auf unterschiedliche Art und Weise nutzen und bereitstellen wollen. Üblicherweise wird/ werden die Plattform(en) folgende Dienstleistungen anbieten:

Storage für ForscherInnen, die aktiv Daten sammeln;

- Eine Publikationsplattform, über die Forschungsdaten und dazugehörige Software geteilt und nachnutzbar gemacht werden;
- Archivfunktionen, die eine langfristige Kuratierung der Forschungsdaten ermöglichen, oftmals als Reaktion auf die von Forschungsförderern gestellten Anforderungen;
- Ein Suchdienst, der es ForscherInnen wie auch der Öffentlichkeit ermöglicht, Datenbestände sowohl auf lokaler Ebene als auch über das Internet zu durchforsten.

Die Europäische Kommission fördert die European Open Science Cloud.⁵ Diese EOSC ist keine Cloud-Lösung im eigentlichen Sinn, sondern eine Metapher für den reibungslosen Austausch von Daten und für den Gedanken von wissenschaftlichen Daten als Gemeinschaftsgut. Die EOSC wird ein gemeinsames Umfeld für die sichere Bereitstellung und Nachnutzung von Forschungsdaten sein, das sich aus bereits bestehenden und noch in Entwicklung befindlichen Infrastrukturen in den Mitgliedsstaaten zusammensetzt, mit einem Minimum an internationaler Führung und Steuerung und einem Maximum an Freiheit hinsichtlich der praktischen Umsetzung.

⁴ Für weitere Informationen siehe <http://www.dcc.ac.uk/resources/data-management-plans>; letzter Zugriff 12.12.16.

⁵ Siehe <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; letzter Zugriff 12.12.2016.

Training

Durch die neuen technischen Möglichkeiten und Arbeitsweisen ist es notwendig, dass sich alle ForscherInnen, etablierte genauso wie JungforscherInnen, all jene Fähigkeiten und Instrumente aneignen, die nötig sind, um sich in einer datengesteuerten Umwelt sicher zu bewegen. Die Führungsrolle sollte dabei von Forschungseinrichtungen und – in vielen Fällen – von deren institutseigenen Bibliotheken übernommen werden.

Finanzierung

Forschungsdatenmanagement ist immer mit Kosten verbunden. Es gibt keine bestimmte, vorgegebene Methode zur Berechnung dieser Kosten, jedoch existiert eine ganze Reihe von Kostenmodellen, die bei dieser Berechnung behilflich sein können, darunter das Projekt 4C.⁶

Risiken

Wenn die gute wissenschaftliche Praxis nicht eingehalten wird, ergeben sich für die beteiligten Akteure bestimmte Risiken. Forschende laufen Gefahr, Fördergelder zu verlieren, sollten sie die Vorgaben der Förderorganisationen nicht erfüllen können. Ein unachtsamer Umgang mit Forschungsdaten kann außerdem zu deren Verlust führen und so eine Validierung der Forschungsergebnisse verhindern. Werden Forschungsergebnisse nicht geteilt und zugänglich gemacht, können sich auch Partnerschaften und Kooperationen nicht entfalten.

Nutzen

Aus einem umsichtigen Umgang mit Forschungsdaten ergeben sich zahlreiche Vorteile. Die Integrität von wissenschaftlichen Erkenntnissen ist dann gegeben, wenn effiziente Richtlinien zum Forschungsdatenmanagement gelten. Forschungseinrichtungen können sich an globalen Initiativen wie der European Open Science Cloud beteiligen. Forschungsdaten dürfen durchaus als Rohstoff im Zusammenhang mit Forschungskommunikation und Publikationen gehandelt werden – bergen sie doch potenzielle Ansätze zur Lösung unserer großen gesellschaftlichen Herausforderungen: Armut, Krankheit und Klimaerwärmung.

Schlussfolgerung

Forschungsdaten sind der Motor für Innovationen, sie sind der Ausgangspunkt für neue Entdeckungen und von großem Nutzen für die gesamte Gesellschaft. Alle jene, die in die Arbeitsabläufe von Forschungsprojekten eingebunden sind, haben ihre ganz bestimmten Rollen zu erfüllen. Das vorliegende „Executive Briefing“ zeigt auf, was ForscherInnen und Forschungseinrichtungen leisten müssen, um sich dieser spannenden Herausforderung zu stellen.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁶ Für weitere Informationen siehe <http://www.4cproject.eu/summary-of-cost-models>; letzter Zugriff 12.12.2016.



Informe Executivo – Portuguese translation

GESTÃO DE ATIVOS DE CONHECIMENTO PARA PESQUISA E INOVAÇÃO NO SÉCULO XXI

O problema

Os dados de pesquisas são a nova moeda da era digital. De sonetos a estatísticas, e de genes a geodados, o volume de material que está sendo criado e armazenado cresce exponencialmente. Contudo, o [Roteiro LERU para dados de pesquisas](#) identifica uma grave lacuna no nível de preparação entre organizações de pesquisa. Essa lacuna é proeminente em áreas como formulação de políticas, conscientização acerca das questões atuais, desenvolvimento de habilidades, capacitação, custos, construção de comunidades, governança, diferenças disciplinares, jurídicas, terminológicas e geográficas.

A solução

Este **Informe executivo LEARN** ajudará os tomadores de decisões e formuladores de políticas a identificar boas soluções. Além disso, as partes interessadas podem seguir o kit [LEARN](#) de estudos de caso de boas práticas, que ajudará as organizações a lidar com o enorme volume de dados. LEARN também proporciona uma pesquisa de auto-avaliação.¹

Política de dados de pesquisas

Toda organização de pesquisa deve ter uma política de dados de pesquisas, que inclua um esquema de gestão dos dados de pesquisas. Os financiadores das pesquisas também devem ter uma política de dados de pesquisas, estipulando as obrigações que um pesquisador deve cumprir como condição do financiamento. LEARN criou uma política modelo de gestão dos dados de pesquisas para organizações, bem como orientações para a implementação da política.² A política modelo de LEARN pode ser adaptada e adotada por organizações de pesquisa ou por consórcios regionais, nacionais e internacionais.

Dados FAIR

A boa prática indica que os dados de pesquisas devem ser FAIR³:

Facilmente encontráveis – **A**cessíveis – **I**nteroperáveis – **R**eusáveis

Para serem facilmente encontráveis, os dados devem ser descritos adequadamente, usando taxonomias e ontologias padrão sempre que possível. Para serem acessíveis, os dados de pesquisas idealmente devem ser abertos, disponíveis para compartilhamento e reutilização. Nem todos os dados de pesquisas podem ser abertos, mas a boa prática indica que esses dados devem ser “tão abertos quanto possível, e tão protegidos quanto necessário”.⁴ Os dados de pesquisas também devem ser interoperáveis, capazes de serem processados por máquinas usando vocabulários que seguem os princípios FAIR. Para serem

¹ Disponíveis em <http://learn-rdm.eu>; acessado pela última vez em 16/12/16.

² Como em 1 acima.

³ Veja <https://www.force11.org/group/fairgroup/fairprinciples>; acessado pela última vez em 12/12/16.

reusáveis, os metadados que descrevem os dados devem cumprir padrões das comunidades relevantes ao domínio.

Gestão dos dados de pesquisas

É importante que os pesquisadores planejem a coleta, curadoria, descrição e disseminação dos dados no início da pesquisa. A melhor maneira de captar essa informação é mediante um plano de gestão dos dados de pesquisas, que proporciona um quadro para a administração dos dados de pesquisas.¹

Infraestrutura

Para fazer a curadoria dos dados de pesquisas, os pesquisadores e organizações de pesquisa precisam ter acesso a ecossistemas digitais. Esses ecossistemas podem ser mantidos localmente ou podem ser serviços comerciais, ofertas de domínios de temas ou plataformas regionais, nacionais e internacionais. Diferentes comunidades de temas e países proporcionarão esses meios de maneira diferente. Em geral, as plataformas devem oferecer os seguintes serviços:

- armazenamento, para pesquisadores que coletam dados;
- plataforma de publicação, onde os dados de pesquisas e respectivos softwares podem ser disponibilizados para compartilhamento e reutilização;
- serviços de arquivo, para permitir a curadoria dos dados de pesquisas no longo prazo, em geral como resposta às exigências dos financiadores da pesquisa;
- um serviço de descoberta, que permite que os pesquisadores e cidadãos busquem depósitos de dados de pesquisas localmente e em toda a Internet.

A Comissão Europeia está promovendo a Nuvem Europeia de Ciência Aberta (EOSC).² A EOSC é uma metáfora para ajudar a transmitir uniformidade e a ideia de um ambiente de compartilhamento voltado a dados científicos. A EOSC será um ambiente federado para compartilhamento e reutilização de dados científicos, baseado em elementos atuais e emergentes dos Estados membros, com leve orientação e governança internacional e alto grau de liberdade no tocante à implementação prática.

Capacitação

A prevalência de dados de pesquisas requer que todos os pesquisadores, novos e estabelecidos, se equipem com habilidades e ferramentas para serem confiantes num ambiente baseado em dados. A liderança deve ser assumida por organizações de pesquisa e, em muitos casos, por suas bibliotecas institucionais.

Financiamento

A gestão dos dados de pesquisas tem múltiplos custos. Não há um método único para avaliar esses custos, mas existem vários modelos de cálculo de custos que podem ajudar, por exemplo, o projeto 4C.⁷

Riscos

Há perigos para as partes interessadas na área de gerenciamento de dados de pesquisa se as boas práticas não são observadas. Pesquisadores podem perder apoio financeiro devido ao não cumprimento dos requisitos das agências de financiamento. Importantes resultados de pesquisa podem se perder pela falta de cuidado, tornando difícil ou impossível a validação dos resultados da pesquisa. Parcerias e colaborações não podem surgir quando resultados de pesquisas não são compartilhados.

Benefícios

Os benefícios da boa gestão de dados de investigação são muitos. A integridade dos resultados da pesquisa melhora quando as Boas Práticas foram observadas. Organizações de pesquisa podem se associar a importantes iniciativas globais de pesquisa como European Open Science Cloud. Os dados de pesquisa podem de fato se tornar a nova moeda da comunicação de pesquisa, juntamente com publicações científicas, contribuindo para a solução dos grandes desafios que a sociedade enfrenta - pobreza, doença, aquecimento global.

Conclusão

Os dados de pesquisas podem orientar a inovação e estimular novas descobertas, trazendo grandes benefícios para a sociedade. Todas as partes no fluxo de trabalho de pesquisa têm um papel a desempenhar. Este Informe Executivo destaca o que os pesquisadores e organizações de pesquisa devem fazer para enfrentar esse desafio.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁴ European Commission - Guidelines on FAIR Data Management in Horizon 2020 p.4 (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).

⁵ Para mais informações, veja <http://www.dcc.ac.uk/resources/data-management-plans>; acessado pela última vez em 12/12/16.

⁶ Veja <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; acessado pela última vez em 12/12/16.

⁷ Para mais informações, veja <http://www.4cproject.eu/summary-of-cost-models>; acessado pela última vez em 12/12/16.

Synthèse – French Translation

LA GESTION DES DONNÉES DU SAVOIR POUR LA RECHERCHE ET L'INNOVATION AU 21^{ÈME} SIÈCLE

Le problème

Les données de la recherche représentent la nouvelle monnaie d'échange à l'ère numérique. Des sonnets aux statistiques, en passant par les gènes et les données géographiques, la quantité de matériel créé et stocké croît de façon exponentielle. Cependant la [Feuille de Route de LERU pour les Données de la Recherche](#) a identifié d'importantes inégalités entre les organismes de recherche en termes de préparation pour faire face à ces questions. L'écart est particulièrement sévère dans les domaines suivants : l'élaboration de politiques de gestion des données, la sensibilisation aux difficultés actuelles, le développement des compétences, la formation, les coûts, la création de communautés de pratique, la gouvernance, ainsi que les différences entre les disciplines, systèmes législatifs, terminologies et espaces géographiques.

La solution

La présente **Synthèse du projet LEARN** aidera les décideurs à prendre des décisions solides. En outre, le Kit de Bonnes Pratiques de [LEARN](#) est à disposition de tous les acteurs de la recherche; organisé autour de cas d'étude, ce kit permettra aux organismes de recherche de faire face au déluge de données. Le projet LEARN a également développé un outil d'auto-évaluation.¹

Politique des données de la recherche

Tout organisme de recherche devrait adopter une politique des données de la recherche énonçant les responsabilités dont sont investis les chercheurs lorsqu'ils reçoivent un financement. Le projet LEARN a créé un modèle de politique pour la gestion des données dans les organismes de recherche; ce modèle est accompagné de conseils pour mettre en place une telle politique.² Le modèle de politique proposé peut être adapté et adopté par chaque organisme mais aussi par des consortiums régionaux, nationaux et/ou internationaux.

Les données « FAIR »

Selon les bonnes pratiques, les données de la recherche doivent être « FAIR »³:

Faciles à trouver (Findable) – **A**ccessibles (Accessible) – **I**nteropérables (Interoperable) – **R**éutilisables (Reusable)

Pour être faciles à trouver, les données doivent être correctement décrites, si possible à l'aide de taxinomies et d'ontologies. Pour être accessibles, elles doivent, dans l'idéal, être ouvertes (open data) et disponibles

¹ Tous ces outils sont disponibles à l'adresse suivante: <http://learn-rdm.eu> (consulté le 12/01/17).

² Voir 1 ci-dessus.

³ Voir www.force11.org/group/fairgroup/fairprinciples (consulté le 12/01/17).

pour le partage et la réutilisation. Les données de la recherche ne peuvent pas toutes être ouvertes mais les bonnes pratiques indiquent que ces données doivent être «aussi ouvertes que possibles et aussi fermées que nécessaires » («as open as possible, as closed as necessary »)⁴. Ces données doivent aussi être interopérables et lisibles par des machines utilisant un vocabulaire conforme aux principes « FAIR ». Pour être réutilisables, les métadonnées qui décrivent les données doivent respecter les standards établis par le champ de recherche dont il est question.

Plans de gestion des données de la recherche

Il est recommandé aux chercheurs de planifier la collecte, le traitement, la description et la diffusion de leurs données dès le début de leur recherche. Ecrire un Plan de gestion des données de la recherche (Data Management Plan) permet de rassembler ces éléments et d'établir un programme de gestion durant la recherche.⁵

Infrastructures

Afin de traiter leurs données les chercheurs et organismes de recherche doivent pouvoir accéder à un écosystème numérique adéquat. Ces infrastructures peuvent être gérées par chaque institution; elles peuvent aussi être fournies par des services commerciaux ou par des entrepôts de données spécialisés dans un champ de recherche particulier; ou bien elles peuvent prendre la forme de plateformes régionales, nationales ou internationales. Chaque champ de recherche et chaque pays devra trouver le système qui lui convient. De manière générale, ces infrastructures doivent proposer les services suivants :

- Le stockage de données, pour les chercheurs qui en collectent activement;
- Une plateforme de publication, où les données de la recherche et les logiciels associés peuvent être partagés et réutilisés;
- Un système d'archivage, permettant que les données soient traitées et préservées sur le long terme, en conformité avec les exigences des financeurs;
- Un système permettant de chercher parmi les données entreposées, afin que chercheurs et citoyens puissent découvrir ces données (qu'elles soient accessibles in situ ou en ligne).

La Commission Européenne encourage l'usage de l'European Open Science Cloud (EOSC).⁶ L'EOSC est une métaphore qui entend exprimer un processus cohérent et sans obstacle, ainsi que l'idée que les données de la recherche constituent un bien commun (commons). L'EOSC sera un environnement fédérateur permettant de partager et réutiliser ces données. Cet environnement sera fondé sur un ensemble d'infrastructures déjà en place ou émergentes dans les Etats Membres et aura une supervision internationale allégée; une large part de liberté sera accordée aux questions pratiques d'utilisation.

Formation

La prévalence des données de la recherche oblige tous les chercheurs, qu'ils soient débutants ou expérimentés, à s'armer de compétences et d'outils leur permettant de travailler en confiance dans cet environnement où les données abondent. Les organismes de recherche doivent prendre l'initiative en termes d'offre de formation; souvent, la tâche en revient aux bibliothèques universitaires ou de recherche.

⁴ European Commission - Guidelines on FAIR Data Management in Horizon 2020 p.4 (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf). (Consulté le 12/01/17).

⁵ Pour plus d'informations, voir www.inist.fr/donnees/co/module_Donnees_recherche_26.html ou www.dcc.ac.uk/resources/data-management-plans (consulté le 12/01/17).

⁶ Voir <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud> (consulté le 12/01/17).

Financement

La gestion des données a un coût. Il n'y a pas de méthode unique pour estimer ce coût mais plusieurs modèles de calcul, tels que le 4C Project, sont disponibles.⁷

Les risques

Les acteurs de la recherche s'exposent aux dangers suivants si les Bonnes Pratiques en matière de gestion des données ne sont pas respectées. Les chercheurs risquent de perdre des financements s'ils ne se conforment pas aux exigences des financeurs. La négligence peut causer la perte d'importants résultats de la recherche, sous peine de rendre difficile voire impossible la validation des conclusions du projet. Enfin, les partenariats et collaborations ne peuvent pas se développer si les résultats de la recherche ne sont pas partagés.

Financement

La gestion intelligente des données de la recherche offre de nombreux avantages. La mise en place de Bonnes Pratiques améliore la rigueur et la transparence des résultats. Les organismes de recherche peuvent rejoindre des initiatives internationales majeures telles que l'European Open Science Cloud. Les données de la recherche peuvent en effet devenir le nouvel élément clé de la communication scientifique et, tout comme les publications, contribuer à résoudre les grands défis auxquels la Société est confrontée : la pauvreté, les maladies et le réchauffement climatique.

Conclusion

Les données de la recherche peuvent favoriser l'innovation et stimuler de nouvelles découvertes pour le plus grand bénéfice de la Société. Tous les acteurs du cycle de la recherche ont un rôle à jouer. Cette Synthèse met en valeur ce que les chercheurs et organismes de recherche doivent faire pour être à la hauteur de ces défis passionnants.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁷ Pour plus d'informations, voir www.4cproject.eu/summary-of-cost-models (consulté le 12/01/17).



Briefing ejecutivo– Italian Translation

COME VALORIZZARE LA GESTIONE DELLA CONOSCENZA PER MIGLIORARE LA RICERCA E L'INNOVAZIONE NEL 21° SECOLO

La problematica

Nell'era digitale i dati della ricerca valgono oro. Dai sonetti alle statistiche, dai dati genetici ai geodati: la quantità di materiali creati e memorizzati in formato digitale cresce in maniera esponenziale. La [LERU Roadmap for Research Data](#) (tabella di marcia per i dati della ricerca) identifica tuttavia gravi lacune per quanto riguarda il livello di preparazione delle organizzazioni attive nel campo della ricerca. Questo divario emerge in modo evidente per quanto riguarda lo sviluppo di politiche di gestione dei dati, la consapevolezza di problemi ancora aperti, lo sviluppo di competenze, la formazione, i costi, la creazione di comunità, la *governance* e le differenze disciplinari/legali/terminologiche e geografiche.

La soluzione

Il presente vademecum **LEARN Executive Briefing** vuole essere una guida a disposizione degli attori coinvolti nei processi decisionali e politici al fine di individuare soluzioni sostenibili. Inoltre, i soggetti interessati possono beneficiare del [LEARN Toolkit of Best Practice Case Studies](#), e dei suoi strumenti risultanti da studi casistici di buone pratiche. Ogni singolo strumento aiuterà le istituzioni che operano nell'ambito della ricerca ad affrontare meglio le questioni riguardanti il cosiddetto "diluvio di dati" (data deluge). LEARN offre anche uno strumento quale il questionario di autovalutazione.¹

Research Data Policy (Policy in materia di dati della ricerca)

Ogni organizzazione che svolge attività di ricerca dovrebbe dotarsi di una propria politica dei dati, per poter definire un quadro di riferimento su come curare e gestire i dati della ricerca. Anche i finanziatori della ricerca dovrebbero dotarsi di una tale *policy*, per stabilire gli obblighi che i ricercatori sono tenuti ad adempiere come premessa per la concessione dei finanziamenti. LEARN ha creato un modello di "policy" per la politica di gestione dei dati della ricerca (Research Data Management) per le organizzazioni che svolgono attività di ricerca, oltre a delle linee guida per l'attuazione di questa policy.² Il modello fornito da LEARN può essere adattato e adottato sia da singole organizzazioni di ricerca, che da associazioni attive nella ricerca a livello regionale, nazionale e/o internazionale.

¹ Il tutto è disponibile su <http://learn-rdm.eu>; ultimo accesso: 16/12/16.

² Come sopra1.

I principi FAIR

Secondo le buone pratiche, i dati della ricerca devono essere FAIR³:

Findable – **A**ccessible – **I**nteroperable – **R**eusable (reperibili – accessibili – interoperabili – riutilizzabili)

Per essere “findable”, i dati devono essere adeguatamente descritti, utilizzando, ove possibile, ontologie e tassonomie standardizzate. Per essere “accessible”, i dati della ricerca dovrebbero essere idealmente “open data” (ad accesso aperto), disponibili per la condivisione e il riutilizzo. Non tutti i dati della ricerca possono essere “open data”, ma secondo le buone pratiche l’accesso dovrebbe essere “il più aperto possibile e chiuso solo quanto necessario”. Inoltre, i dati della ricerca dovrebbero essere “interoperable”, ossia in grado di essere elaborati da sistemi operativi con linguaggi conformi ai principi FAIR. Infine, per essere “reusable”, i metadati descrittivi devono conformarsi alle norme comunitarie rilevanti in materia.

Research Data Management Stewardship (Gestione responsabile dei dati della ricerca)

È importante che i ricercatori provvedano a pianificare la raccolta, conservazione, descrizione e divulgazione dei propri dati della ricerca sin dall’inizio dell’attività. Queste informazioni dovrebbero essere integrate in un Research Data Management Plan che fornisca un quadro di riferimento per una gestione responsabile dei dati della ricerca.⁴

Infrastruttura

Per curare i dati della propria ricerca, i ricercatori e le organizzazioni attive nella ricerca devono poter accedere agli appositi ecosistemi digitali; questi sistemi possono essere mantenuti a livello locale; in alternativa è possibile ricorrere a servizi commerciali, a offerte specifiche a seconda del settore, oppure a piattaforme regionali/nazionali/internazionali. Le diverse comunità scientifiche e/o disciplinari e i singoli Paesi adotteranno soluzioni differenti per rendere fruibili tali strutture. Di norma, la piattaforma, o le piattaforme dovranno essere in grado di offrire i seguenti servizi:

- archiviazione dei dati prodotti e raccolti dai ricercatori;
- una piattaforma di pubblicazione che consenta la condivisione e riuso dei dati della ricerca e relativi software;
- funzioni di archiviazione, affinché i dati della ricerca possano essere conservati nel lungo termine, spesso in risposta ai requisiti imposti dai finanziatori della ricerca;
- un *discovery service* (servizio di ricerca) che consenta ai ricercatori e alla comunità, di esplorare gli archivi dei dati della ricerca sia localmente che attraverso Internet.

La Commissione Europea ha deciso di promuovere la European Open Science Cloud.⁵ L’EOSC è una metafora ideata per trasmettere sia il concetto di un’integrazione di sistemi in modo continuo, che l’idea di un bene comune basato su dati scientifici. L’EOSC sarà un ambiente federato per la condivisione e il riutilizzo di dati scientifici, basato su elementi esistenti ed emergenti negli Stati membri, con un minimo di dirigenza e di forme di *governance* internazionali e un massimo di libertà in termini di attuazione pratica.

Formazione

³ Cfr. <https://www.force11.org/group/fairgroup/fairprinciples>; ultimo accesso: 12/12/16.

⁴ Per ulteriori informazioni: <http://www.dcc.ac.uk/resources/data-management-plans>; ultimo accesso 12/12/16.

⁵ Cfr. <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; ultimo accesso 12/12/16.

Per un’adeguata diffusione dei dati della ricerca è imprescindibile che tutti i ricercatori, dalle nuove leve a quelli già affermati, si dotino delle competenze e degli strumenti necessari per muoversi con sicurezza in un ambiente caratterizzato da dati. Un ruolo di capofila in questo senso dovrebbe essere svolto dalle organizzazioni attive nella ricerca e, in molti casi, dalle biblioteche istituzionali ivi presenti.

Finanziamento

La gestione dei dati della ricerca comporta costi. Non esiste un unico metodo specifico per valutarli, bensì una serie di preziosi modelli di calcolo, ad esempio il progetto 4C.⁶

Rischi

La mancanza di conformità con le regole previste dai codici di buona pratica può comportare dei rischi per tutti quei soggetti che lavorano negli ambiti della gestione dei dati della ricerca. Se i ricercatori non rispettano i requisiti previsti dal finanziamento ottenuto potranno incorrere in sanzioni da parte di chi li promuove. In seguito a negligenze nel processo di gestione dei dati, alcuni risultati di un certo valore potrebbero risultare o introvabili o persi, il che significherebbe non poter validare gli esiti della ricerca. Infine le collaborazioni non potranno prosperare se i risultati della ricerca non saranno condivisi.

Benefici

La corretta gestione dei dati comporta diversi benefici. L’uso di codici di buona pratica aumenta l’integrità degli esiti della ricerca. Gli istituti di ricerca potranno partecipare con più facilità ad iniziative globali di ricerca, come la European Open Science Cloud. I dati della ricerca che accompagnano le pubblicazioni, potranno essere considerati come una nuova moneta di scambio in ambito di comunicazione della ricerca, contribuendo a loro modo alla soluzione delle grandi sfide che riguardano la nostra società – povertà, malattie, effetto serra.

Conclusioni

I dati della ricerca possono dare una spinta verso l’innovazione e offrire lo spunto verso nuove scoperte, a grande beneficio della società. Tutti coloro che sono coinvolti nei processi scientifici sono chiamati a svolgere la propria parte. Il presente Executive Briefing mette in evidenza ciò che i ricercatori e le istituzioni attive nella ricerca sono invitati a fare per raccogliere questa sfida coinvolgente e stimolante.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁶ Per ulteriori informazioni: <http://www.4cproject.eu/summary-of-cost-models>; ultimo accesso 12/12/16.