
Le potentiel des infrastructures d'information pour les données de la recherche

Information Infrastructures and their Potential for Data Digital Research

Heinz Pampel et Maxi Kindling



Édition électronique

URL : <http://journals.openedition.org/edc/8549>

DOI : 10.4000/edc.8549

ISSN : 2101-0366

Éditeur

Université de Lille

Édition imprimée

Date de publication : 1 juin 2019

Pagination : 27-50

ISBN : 978-2-917562-21-5

ISSN : 1270-6841

Référence électronique

Heinz Pampel et Maxi Kindling, « Le potentiel des infrastructures d'information pour les données de la recherche », *Études de communication* [En ligne], 52 | 2019, mis en ligne le 01 janvier 2021, consulté le 21 janvier 2021. URL : <http://journals.openedition.org/edc/8549> ; DOI : <https://doi.org/10.4000/edc.8549>

© Tous droits réservés

Le potentiel des infrastructures d'information pour les données de la recherche¹

Information Infrastructures and their Potential for Data Digital Research

Heinz Pampel

Deutsches GeoForschungsZentrum GFZ, Bibliothek und Informationsdienste (LIS), Potsdam, Germany
heinz.pampel@gfz-potsdam.de

Maxi Kindling

Humboldt-Universität zu Berlin, Institut für Bibliotheks- und Informationswissenschaft (IBI), Berlin, Allemagne
maxi.kindling@hu-berlin.de

1 Cet article est la traduction française de Pampel H., Kindling M. (2017). "Informationsinfrastrukturangebote für digitale Forschungsdaten". In Jacob B., Kindling M., Müller U. T. (dir.), *Peter Schirnbacher sei Dank – E(hren)-Journal anlässlich seiner Emeritierung als Professor am Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin*, Berlin: Humboldt-Universität zu Berlin, p. 15-33. DOI 10.18452/2341.

Résumé / Abstract

Internet et l'expansion du numérique ont ouvert la voie à de nouvelles possibilités de diffusion et d'échanges des connaissances et d'information pour la science. Outre le concept de Science ouverte parallèle à celui d'Accès ouvert, l'accessibilité et la réutilisabilité des données de la recherche sont sources de débats. Les travaux de Peter Schirmbacher en recherche et dans son enseignement ont visé la pertinence, la conception et l'exploitation d'offres dans les infrastructures d'information pour les données de la recherche à l'ère du numérique et ont ainsi assuré des contributions pionnières pour le développement de ces services au profit des activités scientifiques. La présente contribution propose un panorama sur l'état de l'art de ces infrastructures d'information et de leurs offres et les situe dans le contexte actuel des débats sur la Science ouverte.

Mots-clés : données de recherche, entrepôts de données, infrastructures de recherche, science ouverte, libre accès.

The Internet and the expansion of digital technology have opened up new opportunities for the dissemination and exchange of knowledge and information for science. In addition to the concepts of Open Science and Open Access, the accessibility and reusability of research data have become sources of debate. Peter Schirmbacher has focused in his research and teaching on the relevance, design and use of information infrastructures for research data in the digital age. Schirmbacher has made pioneering contributions to development of these services for the benefit of scientific activities. This contribution provides an overview of these information infrastructures and their potential, and situates them in the current context of Open Science debates.

Keywords: research data, data repositories, research infrastructures, open science, open access.

Introduction – Débat chez les scientifiques²

Internet et l'expansion du numérique ont ouvert de nouvelles possibilités pour la diffusion et l'échange de connaissances et d'informations dans la science. Ce potentiel a été largement décrit en 2003 dans la *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*³ (Max-Planck-Gesellschaft, 2003)⁴. Cette déclaration, qui a également été signée par l'institution où Peter Schirmbacher a longuement exercé, l'Université Humboldt de Berlin, souligne l'importance de l'Accès ouvert aux publications sous forme textuelle (Suber, 2012) et l'élargit à l'ouverture et la réutilisation des données pour les activités scientifiques utilisant des outils numériques.

Les débats sur les opportunités et les défis du *data sharing* ont continué à prendre une ampleur significative au cours des dernières années et se sont révélés être une composante de la Science ouverte sur le plan de la politique scientifique (Kroes, 2012 ; Commission européenne, 2016). La recommandation faite par une vénérable société savante comme la *Royal Society* à tous les chercheurs, énonçant : « *Where data justify it, scientists should make them available in an appropriate data repository* »⁵ dans un rapport ayant reçu un large écho et intitulé *Science as an open enterprise*⁶, renforce l'exigence envers la science que les données de recherche soient librement accessibles. De nombreuses études corroborent ce souhait (Kuipers et Van der Hoeven, 2009 ; Schäfer *et al.*, 2011 ; Commission européenne, 2012 ; Tenopir *et al.*, 2011 ; Tenopir *et al.*, 2015). Toutefois, celles-ci attirent également l'attention sur les barrières juridiques, techniques et culturelles qui empêchent individuellement des chercheurs de rendre accessibles des données de recherche. Ces barrières sont souvent à mettre sur le compte d'une absence d'offres de service de la part des infrastructures d'information (Borgman, 2007 ; Tenopir *et al.*, 2011 ; van der Graaf *et al.*, 2011) et révèlent nettement combien les pratiques autour de l'ouverture de l'accès aux données de recherche et leur réutilisation sont encore à un stade tout à fait précoce de leur développement.

- 2 Le présent article est une version mise à jour et remaniée du chapitre « Outils de gestion des données de recherche » (*Werkzeuge des Forschungsdatenmanagements*) extrait du *Manuel de gestion des données de recherche (Lehrbuch Forschungsdatenmanagement)* : https://handbuch.tib.eu/w/Lehrbuch_Forschungsdatenmanagement (page consultée le 11 mars 2019).
- 3 Version française – *Déclaration de Berlin sur le Libre Accès à la Connaissance en sciences exactes, sciences de la vie, sciences humaines et sociales* : <http://openaccess.inist.fr/?Declaration-de-Berlin-sur-le-Libre> (page consultée le 11 mars 2019).
- 4 Voir la page des signataires de la *Déclaration de Berlin* : <https://openaccess.mpg.de/3883/Signatories> (page consultée le 11 mars 2019).
- 5 *Si les données le justifient, les scientifiques devraient les rendre disponibles dans un entrepôt de données adapté.*
- 6 *La science comme une entreprise ouverte.*

1.

Importance des infrastructures d'information

Pour ce qui est des données sous forme numérique, les besoins vont vers des infrastructures d'information qui, comme les définit Peter Schirnbacher, représentent « l'ensemble des ressources techniques, organisationnelles, financières et en personnels », qui contribuent à « la fourniture d'information à la science et à la recherche » (Schirnbacher, 2015), pour, par exemple, soutenir des chercheurs dans leur gestion d'un volume croissant de données numériques. Les dispositifs de ces infrastructures d'information sont en particulier des bibliothèques, des centres de calcul, des centres de données, des archives et des musées. En tant que composantes de l'infrastructure de recherche, ces intervenants se considèrent comme des prestataires de service pour un groupe d'utilisateurs défini. Dans la présente contribution, leurs prestations seront désignées par le terme d'« offres » des infrastructures de recherche. Durant la montée en puissance du numérique, ces acteurs se sont efforcés de rendre accessibles sur Internet les données qu'ils détenaient pour améliorer la transparence, la reproductibilité et l'efficacité de la recherche.

En Sciences de l'information, les infrastructures d'information et leurs offres font elles-mêmes l'objet de travaux de recherche et de développement. C'est ainsi que la pratique a suscité des initiatives inspirées sur le terrain par la chaire de gestion de l'information qu'occupait Peter Schirnbacher à l'Institut des Sciences de l'Information et de bibliothéconomie (*Institut für Bibliotheks- und Informationswissenschaft*, IBI) de l'Université Humboldt en coopération avec le groupe de travail de la bibliothèque de l'université et du service informatique et média de cette institution, destinées à déployer, analyser et faire évoluer des entrepôts et des services à valeur ajoutée, à mettre en place et poursuivre la mise en réseau de ces infrastructures sur les plans technique et de l'organisation, ainsi que la mise au point et la diffusion de normes de qualité. De plus, des modèles économiques et financiers ont été élaborés et l'utilisation de modèles juridiques de licences adaptés a été imaginée (Schirnbacher, 2015).

Les travaux accomplis dans ce domaine par Peter Schirnbacher ont largement œuvré ces dernières années à l'évolution et la professionnalisation de ces offres des infrastructures d'information qui portent l'activité de gestion des données de la recherche conçue en tant que « processus global [pour] l'affectation, la génération, le traitement, l'enregistrement, l'archivage et de la diffusion des données de recherche elles-mêmes ou résultant d'une production classique de textes associés » (Kindling et Schirnbacher, 2013). Pour ce qui est d'une distinction d'ordre thématique, Peter Schirnbacher propose en outre une approche des données de la recherche qui fait un distinguo entre

les dimensions relevant de politique de recherche, des aspects d'organisation et de la technique⁷.

Depuis 2008 au moins, année de lancement du plan d'action « Information numérique » (*Digitale Information*) par l'Alliance des organismes de recherche allemands (*Allianz der deutschen Wissenschaftsorganisationen*)⁸, on peut observer l'importance croissante prise par ces infrastructures dans la politique scientifique. Les travaux de la commission d'enquête parlementaire « Internet et société numérique » (*Internet und digitale Gesellschaft*) en 2012, lors de la dix-septième législature de la chambre basse du parlement fédéral, le Bundestag, ont aussi largement contribué à cette émergence. Un groupe de travail « Enseignement et recherche » (*Bildung und Forschung*) de cette commission s'est emparé de manière complète et exhaustive des conséquences de l'incursion du numérique dans la recherche et l'enseignement, et s'est exprimé dans un consensus transpartisan en faveur de « l'extension du principe du libre accès, autant que faire se peut, au domaine des données scientifiques issues de travaux de recherche financés sur fonds publics et l'adoption de mesures adaptées à cette fin ».

S'appuyant sur les principes sur l'emploi des données de recherche que l'Alliance des organismes de recherche avait publiés en 2010, l'État fédéral et les États-Régions ont réagi sous l'impulsion du monde scientifique et ont émis des recommandations pour la conception de ces infrastructures pour les années à venir, et qui seraient élaborées à un niveau national par la commission « Avenir de l'infrastructure d'information » (*Kommission Zukunft der Informationsinfrastruktur*, KII) et le Conseil pour la science (*Wissenschaftsrat*, WR) pour conduire à l'installation en 2013 du Conseil pour les infrastructures d'information (*Rat für Informationsinfrastrukturen*, RfII) par la Conférence commune pour la science (*Gemeinsame Wissenschaftskonferenz*, GWK). Cette politique avait pour priorités de « mieux coordonner la diversité des activités de ce domaine et de les articuler entre elles » (*Gemeinsame Wissenschaftskonferenz*, 2013). En 2016, le RfII présentait (*Rat für Informationsinfrastrukturen*, 2016) ses préconisations *Performance par la diversité* (*Leistung aus Vielfalt*), dont le contenu s'articulait autour de la création d'une infrastructure nationale des données de la recherche (*Nationale Forschungsdaten-Infrastruktur*, NFDI).

7 Ces dimensions ont été développées dans un cours intitulé *Quelques aspects de la fourniture numérique d'informations* (*Ausgewählte Aspekte digitaler Informationsversorgung*) à l'Institut des sciences des bibliothèques et de l'information dont la publication est actuellement en préparation.

8 Voir <http://www.allianzinitiative.de> (page consultée le 11 mars 2019).

2.

Fondements de l'accessibilité et de la réutilisation des données de la recherche

La mise à disposition en ligne de données de recherche peut s'opérer selon différentes conditions et modalités d'accès. À un premier stade, il peut s'avérer utile de s'intéresser sur ce point aux périmètres d'accès abordés par Treloar et Harboe-Ree (2008), qui font la distinction entre sphère privée, groupes ou public. Cette représentation comporte une division conceptuelle entre domaines d'accès qui en pratique ont des implications spécifiques, par exemple pour la préparation des données de recherche. Si des données sont partagées via un domaine public, celles-ci peuvent être citées dans l'idéal à l'aide d'un identifiant pérenne et enrichies à l'aide de métadonnées pertinentes tant pour leur création, que pour leurs éventuels traitement et réutilisation.

La mise à disposition de données de recherche selon les principes du Libre accès énoncés dans la *Déclaration de Berlin* dans sa version de 2003 implique que celles-ci soient rendues accessibles sans restrictions techniques, juridiques ou financières et puissent être réutilisées (Max-Planck-Gesellschaft, 2003).

Depuis quelques années le vocable *Open Research Data* désigne en fait des débats au sein de la communauté scientifique sur les stratégies et processus d'ouverture des données qui sont envisagés comme une composante d'une évolution vers la Science ouverte (Pampel et Dallmeier-Tiessen, 2014).

Pour susciter l'intérêt que les membres de la recherche pourraient trouver à ouvrir l'accès à leurs données, des stratégies de publication se sont installées au cours des années écoulées en interaction avec les scientifiques, les mondes des bibliothèques et de l'édition et qui mettent des données de tiers à la disposition des chercheurs et assurent une reconnaissance à l'intérieur du système de réputation. Pour cela, une condition est essentielle : que ces données puissent être citées, comme déjà indiqué, grâce à un adressage pérenne des jeux de données (Klump *et al.*, 2006 ; Neumann et Brase, 2014), par exemple, en leur attribuant un *Digital Object Identifier* (DOI). Dans la trajectoire de ce DOI, grâce aux possibilités de créer des liens avec d'autres identifiants pérennes de personnes (Pampel et Fenner, 2016), d'organismes (Cruse *et al.*, 2016), tout comme d'autres résultats des travaux de recherche, comme les échantillons (Lehnert *et al.*, 2011 ; Conze *et al.*, 2017), les fondations d'un réseau ouvert se mettent en place, pour assurer ainsi la visibilité des objets informationnels et permettre leur exploitation.

Il arrive que des données ne puissent pas être accessibles au public – pour des raisons de protection des données personnelles, comme c'est le cas en médecine ou en sciences sociales. Il est néanmoins conseillé de rendre au moins une partie de leurs métadonnées visibles pour garantir la transparence des travaux de recherche et offrir à des tiers la possibilité de s'y référer et

d'en connaître le contexte. Les conditions de mise en accès et d'utilisation des données de recherche doivent alors être réglées auprès des détenteurs/propriétaires des droits. Autre scénario : il peut arriver que l'accès aux données ne soit ouvert qu'après une période d'embargo. C'est ainsi que des chercheurs peuvent, pour commencer, charger leurs données dans un dépôt spécialisé à cet effet et prévoir de les doter d'un identifiant pérenne, qui sera par la suite cité dans un document destiné à la publication (par exemple, un article paraissant dans une revue scientifique). De la sorte, les chercheurs ont le temps d'exploiter largement leurs données avant de les rendre accessibles à la communauté des spécialistes.

En 2016 sont créés les standards FAIR – pour *Findable, Accessible, Interoperable* et *Reusable* (Wilkinson et al., 2016) – qui répondent aux attentes des différentes disciplines et qui peuvent être considérés comme fondamentaux pour l'utilisation des données de recherche aujourd'hui. Ils intègrent aussi de la sorte la nécessité d'assurer si possible un accès ouvert à ces données. Les principes FAIR proposent un cadre pour la gestion des données de recherche par les institutions scientifiques et c'est ainsi qu'ils ont été intégrés au programme de recherche Horizon 2020 par la Commission européenne, organisme de financement de la recherche (Commission européenne, 2016).

3. Préconisations et directives

La notion de *data policies* désigne les préconisations et les directives sur l'utilisation des données de la recherche. Leur caractère impératif varie selon l'acteur et le groupe qu'elles concernent (Pampel et Bertelmann, 2011). En Allemagne, les propositions de l'Agence allemande pour la recherche (*Deutsche Forschungsgemeinschaft*, DFG) énoncées pour la préservation des bonnes pratiques de la science (*Vorschläge zur Sicherung guter wissenschaftlicher Praxis*) en 1998 (puis actualisées : Deutsche Forschungsgemeinschaft, 2013) jouent un rôle central et revêtent un caractère obligatoire pour nombre d'institutions scientifiques. Elles prônent la conservation des données de recherche pendant au moins dix ans. Cependant, ces impératifs de « sauvegarde et préservation » ne garantissent pour autant ni l'accessibilité des données de recherche ni leur réutilisabilité. Ces notions sont donc à interpréter comme des prescriptions qui méritent d'être étoffées. Pour compléter tout cela, la DFG a publié en 2015 ses orientations sur le traitement des données de la recherche (*Leitlinien zum Umgang mit Forschungsdaten*) où il est prescrit à leurs utilisateurs intermédiaires de « [les] mettre [...] à disposition aussi rapidement que possible » et de veiller à « rendre possibles leur réutilisation et une nouvelle exploitation par des tiers » (Deutsche Forschungsgemeinschaft, 2015).

Au cours de ces dernières années, de nombreuses institutions scientifiques ont adopté des orientations et des directives de ce genre⁹. L'Université Humboldt elle-même, en s'appuyant sur les travaux menés par la chaire et l'équipe de recherche « Gestion de la recherche » de Peter Schirmbacher, a tracé la voie d'une « politique des données de la recherche » (Humboldt-Universität zu Berlin, 2014). Aux prescriptions des agences de financement et des institutions scientifiques s'ajoutent les consignes qu'imposent les revues scientifiques aux chercheurs. À titre d'exemple, les revues *PLoS* et *Nature* leur demandent au moment de soumettre un article une déclaration de disponibilité des données (*Data Availability Statement*) qui donne des indications sur l'accessibilité des données de recherche sur lesquelles s'appuie l'article en question (Bloom et al., 2014 ; *Nature*, 2016).

4.

Plans de gestion des données

Les plans de gestion des données de recherche (*Data Management Plan*, DMP) sont des instruments où sont consignées par écrit les données traitées lors d'un projet de recherche depuis le stade de sa programmation jusqu'à son achèvement. Les DMP sont des documents électroniques qui sont mis à jour tout au long du déroulement des projets de recherche. Ils offrent aux porteurs et partenaires d'un projet un cadre opérationnel pour le traitement des données produites et exploitées pendant toute sa durée. Ces plans de gestion précisent, par exemple, le type et la forme de ces données, leur degré de standardisation (par exemple, les métadonnées utilisées, les formats), des indications sur leur accessibilité et les possibilités de réutilisation, ainsi que sur les procédés employés pour leur conservation à long terme).

Les agences de financement de la recherche exigent de plus en plus fréquemment que les demandes de financement déposées comportent des indications sur l'accessibilité et la réutilisabilité des données. Ces prescriptions mentionnant un plan de gestion de données de recherche sont énoncées de façon concrète dans les consignes de dépôts de ces demandes.

Le guide que la DFG met à la disposition des déposants demandant un financement stipule un certain nombre de consignes¹⁰. C'est ainsi qu'il énonce : « Si grâce au financement du projet, des données de recherche ou des informations doivent être créées qui sont aptes à être réutilisées par d'autres

⁹ Voir un aperçu dans le wiki *Forschungsdaten-Wiki* : http://www.forschungsdaten.org/index.php/Data_Policies (page consultée le 11 mars 2019).

¹⁰ DFG, *Leitfaden für die Antragsstellung (Fils conducteurs pour le dépôt d'une demande)* : http://www.dfg.de/formulare/54_01/ (page consultée le 11 mars 2019).

scientifiques, veuillez le préciser et indiquez le cas échéant la manière dont elles seront mises à la disposition de tiers. Soyez attentifs ce faisant à prendre en compte les éventuels standards ou offres de dépôts de données ou d'archives proposées dans votre discipline ou domaine de spécialité » (Deutsche Forschungsgemeinschaft, 2014).

Si la DFG demande que des déclarations de DMP soient soumises au moment du dépôt d'une demande de financement, celles-ci sont exigées sous forme de « livrables » dans les six premiers mois de déroulement d'un projet soutenu par le programme de recherche européen Horizon 2020. D'autres justificatifs sur la gestion des données devront également être produits par la suite pendant son déroulement comme à son terme (Commission européenne, 2016).

Divers outils et documentations sont proposés pour pouvoir produire ces plans de gestion :

- les *checklists* produites par le programme WissGrid en Allemagne (Ludwig et Enke, 2013) ;
- *DMP Online* par le *Digital Curation Centre* au Royaume-Uni¹¹ ;
- *DMPTool* par l'Université de Californie aux États-Unis¹².

Ces plans de gestion des données servent également de base comptable pour calculer les coûts d'une telle gestion. Les agences de financement octroient des fonds qui couvrent en partie ces tâches. C'est ainsi qu'il est possible à l'intérieur d'un plan de gestion des données de recherche de décliner les financements demandés par poste de gestion dans une candidature.

5.

Infrastructures d'information : les offres

Nous allons dans ce qui suit présenter quelques infrastructures d'information centrales qui à la fois viennent soutenir la gestion des données de recherche et rendent possibles la transparence des travaux de recherche, comme la réutilisation de leurs données. Elles permettent de tracer ces données grâce à des métadonnées descriptives, de les rendre accessibles au public ou à des groupes d'utilisateurs sélectionnés (par exemple, en raison d'une protection juridique des données personnelles), tout comme d'assurer leur conservation à long terme.

Dans certains domaines scientifiques, des infrastructures existent depuis de nombreuses années (c'est le cas, par exemple, des centres de données de

11 Voir *DMPonline* : <http://dmponline.dcc.ac.uk> (page consultée le 11 mars 2019).

12 Voir *DMPTool* : <http://dmp.cdlib.org> (page consultée le 11 mars 2019).

rang mondial pour les sciences de la nature depuis les années 1950)¹³, alors qu'on commence seulement à structurer une offre spécifique et la mettre en place dans d'autres.

Selon les disciplines scientifiques et en fonction des méthodes de collecte et de traitement, comme des types et formats de données, les attentes fonctionnelles sont différenciées. Il en résulte un paysage des infrastructures destinées aux données de la recherche d'une grande hétérogénéité (Pampel *et al.*, 2013). Depuis plusieurs années le vocable d'« entrepôts de données de recherche » fait son apparition dans les discussions sur la professionnalisation de ces infrastructures pour ce qui touche au stockage, l'archivage et la diffusion en libre accès. Vu sous l'angle des pratiques scientifiques, les stratégies de publication des données de recherche varient. Tantôt celles-ci sont publiées sous la forme d'un objet informationnel autonome, par exemple dans un entrepôt de données, tantôt comme « supplément » ou comme publication augmentée (*enhanced publication*) en étant directement reliées avec un article, ou carrément en tant que *data paper* dans une revue scientifique ou plus spécifiquement dans un *data journal* (Woutersen-Windhouwer et Brandsma, 2009 ; Pampel et Dallmeier-Tiessen, 2014). Le *framework Scholix* apparu en 2017 repose sur le concept de l'interaction entre entrepôts de données et articles de revues scientifiques (Burton *et al.*, 2017). Il existe par ailleurs des recommandations préconisant un adressage pérenne et une interconnexion des objets d'information dans les entrepôts de données (Fenner *et al.*, 2016) et d'articles de revues (Cousijn *et al.*, 2017), et qui contribuent de façon importante à la reconnaissance des données de recherche en tant qu'objets d'information en soi dans la communication scientifique.

6. Entrepôts de données

Dans le domaine des infrastructures, nous savons ce qu'est un entrepôt au travers de l'environnement des archives documentaires donnant accès à des publications (Lynch, 2003).

Le projet *re3data* (*Registry of Research Data Repositories*), auquel l'unité de recherche et d'enseignement de Peter Schirmbacher a participé en tant que partenaire, a élaboré un annuaire des entrepôts de données de recherche (Pampel *et al.*, 2013). Depuis sa création, celui-ci signale 1 805 entrepôts dans le monde¹⁴. Ceux-ci sont décrits à l'aide d'un schéma de métadonnées exhaustif

13 Repris depuis 2008 par le *ICSU World Data System*. Voir WDS : <http://www.icsu-wds.org> (page consultée le 11 mars 2019).

14 Chiffres du 26 janvier 2017. Le 9 juin 2019, *re3data* recense 2 343 dépôts (ndlr).

(Rücknagel et al., 2015). Les chercheurs, membres d'infrastructures et d'agences de financement de la recherche ou tout particulier intéressé peuvent rechercher des entrepôts les intéressant dans le but de rendre accessibles des données de la recherche, ou y trouver ceux qui proposent des données pouvant être réutilisées. Grâce à la richesse descriptive des métadonnées affectées à ces entrepôts, il est possible d'affiner les recherches opérées dans *re3data* avec des filtres.

L'annuaire *re3data* couvre tout le large spectre des entrepôts de données : il recense des entrepôts qui peuvent être disciplinaires, multidisciplinaires, institutionnels ou propres à des projets, etc.

Pour être signalé, un entrepôt doit correspondre à la définition suivante : « *A research data repository is a subtype of a sustainable information infrastructure which provides long-term storage and access to research data. Research data means information objects generated by scholarly projects for example through experiments, measurements, surveys or interviews* » (Rücknagel et al., 2015)¹⁵. Cette définition très large a été élaborée avec la motivation que cet annuaire propose à ses utilisateurs lors de leurs recherches un grand nombre d'entrepôts potentiellement pertinents pour eux.

Vus sous l'angle des Sciences de l'information, un entrepôt de données de la recherche se définit de manière plus spécifique. Il s'agit en l'espèce d'une ressource sur le web proposée par une infrastructure qui diffuse et rend accessibles des données de recherche et en permet l'utilisation. À cet effet, un tel entrepôt assure, en répondant à la demande d'un groupe d'utilisateurs que :

- les données de recherche sont préparées dans un format adapté pour une réutilisation ;
- ces données sont citables et décrites avec des schémas de métadonnées établis ;
- celles-ci sont assorties d'informations sur leurs conditions d'utilisation.

Un entrepôt de données est opéré par un établissement de confiance tel qu'une institution académique, qui a pour mission de préserver sa disponibilité à long terme. Le groupe d'utilisateurs concernés, les prescriptions attendues concernant les données à stocker, les conditions d'exploitation de l'entrepôt et des données sont présentés dans des documents accessibles au public. L'orientation d'un entrepôt va dépendre du traitement des données de recherche qu'il opère suivant les normes et standards en vigueur dans les communautés scientifiques concernées ainsi qu'en fonction des standards de

15 *Un entrepôt de données de recherche est un sous-ensemble d'une infrastructure d'information inscrite dans la durée qui offre un stockage et un accès à long terme à des données de recherche. Par données de recherche, on entend des objets créés par des activités scientifiques telles que des expérimentations, des mesures, des études ou des entretiens.*

qualité généraux pour les offres d'infrastructure comme les *Data Seal Approval*¹⁶, les normes DIN 31644¹⁷ ou ISO 16363:2012¹⁸ intitulée *Audit and certification of trustworthy digital repositories*¹⁹. Un entrepôt de données offre une recherche dans les contenus déposés par métadonnées et permet d'y accéder *via* une interface web. De plus, il fournit les métadonnées détectées au travers d'interfaces standardisées qui permettent ainsi l'agrégation des métadonnées par des services tiers (par ex. : moteurs, moissonneurs). En outre, dans leur conception, les entrepôts vont dépendre de facteurs comme :

- les modèles d'accès à l'entrepôt et ses données ;
- les fonctionnalités dont la population d'utilisateurs a besoin ;
- les logiciels utilisés ;
- les modèles économiques (coûts, financement).

On connaît des entrepôts de données spécialisés comme *PANGAEA* (Diepenbroek *et al.*, 2002) dans les géosciences et *GenBank* (Benson *et al.*, 2015) en biomédecine. Beaucoup d'institutions scientifiques créent actuellement des entrepôts pour accueillir les données de recherche de leurs membres. On a par exemple en Allemagne *heiDATA*²⁰ à l'Université de Heidelberg, *Open Data LMU*²¹ à l'Université Louis-et-Maximilien de Munich ou encore *panMetaDocs* (Ulbricht *et al.*, 2016) au GFZ, le centre de recherche allemand en géosciences. Pour certaines institutions, comme l'Université technique de Berlin²², un entrepôt commun est mis à disposition pour le dépôt de données de recherche, de documents publiés comme d'autres objets d'information. De même, l'Université Humboldt sous la direction et sous l'impulsion de Peter Schirmbacher en sa qualité de directeur du service informatique et média et du département de recherche et d'enseignement de l'IBI a vu la mise en place de différentes offres (comme l'entrepôt *LAUDATIO* accueillant des corpus de textes en histoire²³ et le *Medien Repositorium*²⁴) ou à son initiative (comme l'utilisation du serveur

16 Voir *Data Seal of Approval* : <http://www.datasealofapproval.org> (page consultée le 11 mars 2019).

17 Voir *Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644 (Archivage numérique à long terme digne de confiance selon DIN 31644)* : <https://www.beuth.de/de/publikation/vertrauenswuerdige-digitale-langzeitarchivierung/169654635> (page consultée le 11 mars 2019).

18 Voir *ISO 16363:2012* : http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510 (page consultée le 11 mars 2019).

19 NdT : *Audit et certification des référentiels numériques dignes de confiance*.

20 Voir *heiDATA* : <https://heidata.uni-heidelberg.de/> (page consultée le 11 mars 2019).

21 Voir *Open Data LMU* : <http://data.ub.uni-muenchen.de/> (page consultée le 11 mars 2019).

22 Voir *Deposit Once* : <https://depositonce.tu-berlin.de> (page consultée le 11 mars 2019).

23 Voir *LAUDATIO Repository* : <http://www.laudatio-repository.org/repository/> (page consultée le 11 mars 2019).

24 Voir *Medien-Repositoryum* : <https://medien.hu-berlin.de/> (page consultée le 11 mars 2019).

de publications *edoc* destiné aux données de recherche numériques dans le cadre du projet *eDissPlus*, projet proposé avec sa collaboration)²⁵.

S'il n'existe pas d'entrepôt pour une communauté spécialisée, ni institutionnel, il reste la possibilité de donner accès à ses données via un entrepôt générique, comme celui mis à disposition par le CERN, *Zenodo*²⁶.

7.

Data papers et data journals

Les *data papers* sont des documents qui décrivent exhaustivement des données de recherche. Ils sont particulièrement intéressants quand les métadonnées descriptives ne suffisent pas à donner une idée du contenu d'un jeu de données présent dans un entrepôt, notamment en exposant les possibilités de réutilisation qu'il offre. La publication d'un tel *data paper* s'effectue dans une revue scientifique, dans un *data journal* spécifique, tout comme dans d'autres types de publication (Pampel et Dallmeier-Tiessen, 2014)²⁷. C'est ainsi qu'a été créé en 2008 le *data journal Earth System Science Data* (ESSD) en géosciences (Pfeiffenberger et Carlson, 2011). Le taux élevé de citation des données qui y sont commentées a eu pour effet que ESSD a été classé troisième parmi les revues de géosciences les plus citées dans l'édition 2016 du *Journal Citation Report* (JCR)²⁸.

8.

Comment rechercher et trouver ces infrastructures ?

Quand on recherche des données à réutiliser, il est aisé de se diriger directement vers des offres de type entrepôts de données ou *data journal*, ou bien de rechercher des entrepôts de données dans son domaine scientifique propre à l'aide par exemple de *re3data* qui peut répertorier des données de recherche qui s'y rapportent. Outre les solutions locales ou spécialisées disciplinairement, il existe aussi des services génériques pour aller rechercher

25 Voir *eDissPlus* : <https://www2.hu-berlin.de/edissplus/> (page consultée le 11 mars 2019).

26 Voir *Zenodo* : <http://zenodo.org> (page consultée le 11 mars 2019).

27 Une liste de *data journals* est mise à jour dans le wiki *Forschungsdaten-Wiki : Data Journals* : http://www.forschungsdaten.org/index.php/Data_Journals (page consultée le 11 mars 2019).

28 Voir *First Journal Impact Factor for ESSD released* : https://www.earth-system-science-data.net/about/news_and_press/2016-06-17_first-impact-factor-for-essd.html (page consultée le 11 mars 2019).

des données. Les métadonnées étant pour la plupart spécialisées dans un domaine, ces outils bien souvent n'offrent guère de possibilités de recherche très détaillées. Dans ce qui suit nous allons citer trois exemples de services qui agrègent les métadonnées d'entrepôts de données et autres infrastructures et qui offrent des recherches multi-plates-formes (leurs périmètres de données se chevauchent en partie). Il faut tenir compte pour toutes ces offres du fait que la granularité des méta-informations qui peuvent être signalisées moyennant un identifiant comme un DOI peut être très différente. Cela veut dire que des métadonnées sont appliquées aussi bien à des données de recherche isolées sous forme de simples fichiers qu'à des collections entières de jeux de données.

8.1. *DataCite Metadata Search*²⁹

Ce service permet d'effectuer une recherche dans les métadonnées d'objets d'information, tels que des données de recherche dont les DOI ont été enregistrés chez *DataCite* et auxquelles des métadonnées ont été associées. En février 2017, il y avait environ 7,6 millions de jeux de données enregistrés et pouvant être recherchés, dont 3 178 328 de type ensemble de données (*dataset*)³⁰. Ces indications ne sont à prendre qu'à titre indicatif, car sous différents aspects elles peuvent être entachées de distorsions (qualité des métadonnées, doublons, classification abusive en tant que données de la recherche).

8.2. *BASE – Bielefeld Academic Search Engine*³¹

BASE permet d'effectuer des recherches dans les métadonnées des 5 135 sources actuellement recensées comme des entrepôts et d'autres types d'offres proposées sur l'infrastructure d'information et qui mettent leurs métadonnées à disposition via une interface OAI-PMH. En février 2017, 3 622 458 métadonnées ont été enregistrées pour le type de document « Données de recherche »³². Une offre comme un entrepôt de données peut également être répertoriée avec un autre type de document disponible. *BASE* intègre également les métadonnées de *Datacite* dans ses requêtes.

8.3. *EUDAT B2 Find*

Ce service résulte d'un développement issu du projet européen *EUDAT*³³ et permet d'effectuer des recherches dans des collections de données provenant de sources sélectionnées, dont celles des centres de données *EUDAT*³⁴.

29 Voir *DataCite Metadata Search* : <http://search.datacite.org/ui> (page consultée le 11 mars 2019).

30 Chiffres du 06.02.2017.

31 Voir *BASE* : <https://www.base-search.net> (page consultée le 11 mars 2019).

32 Chiffres du 06.02.2017.

33 Voir *EUDAT B2 Find* : <http://b2find.eudat.eu> (page consultée le 11 mars 2019).

34 Documentation disponible sous : *EUDAT User documentation* : <https://eudat.eu/services/userdoc/b2find> (page consultée le 11 mars 2019).

En février de cette année, il compte quelque 472 082 métadonnées renvoyant à des données de recherche³⁵. Parmi les métadonnées indexées, on compte celles de la *Common Language Resources and Technology Infrastructure*, ou *CLARIN*³⁶ ou de la *Global Biodiversity Information Facility*, ou *GBIF*³⁷.

9. Perspectives

Grâce à l'implication de Peter Schirmbacher, la Conférence des présidents des établissements d'enseignement supérieur (*Hochschulrektorenkonferenz*, HRK) a énoncé en 2014 la gestion des données de recherche comme une « mission stratégique des gouvernances d'établissements d'enseignement supérieur » (*Hochschulrektorenkonferenz*, 2014). C'est ainsi que les institutions de l'enseignement supérieur et les organismes de recherche non universitaires ont commencé à élaborer des orientations et des directives ainsi que des offres d'infrastructures destinées aux personnels de recherche pour développer l'accessibilité et la réutilisation des données de recherche.

La mise en place de l'infrastructure nationale des données de la recherche (*Nationale Forschungsdaten-Infrastruktur*, NFDI) sous l'impulsion du Rfll représente une grande chance pour le dispositif scientifique allemand permettant de tirer parti de l'expansion du numérique dans l'enseignement et la recherche (*Rat für Informationsinfrastrukturen*, 2016) et de faire converger le paysage très différencié et hétérogène des offres en les faisant reposer sur des bases communes. Rien qu'au moment où nous écrivons ces lignes, il existe plus de 270 entrepôts de données de recherche, où des organismes allemands prennent part à leur fonctionnement (*Pampel et al.*, 2016). La science étant une activité de dimension internationale, il importe cependant de porter le regard au-delà des frontières nationales.

Le projet *GeRDI* de la DFG et la *Helmholtz Data Federation* (HDF) sont entrés dans ce cycle en tant qu'acteurs de premier plan et se sont signalés par leur volonté d'œuvrer activement à la création du NFDI (*Helmholtz-Gemeinschaft*, 2016 ; *Zentralbibliothek für Wirtschaftswissenschaften*, 2016). Ces deux acteurs soulignent l'importance de pouvoir s'interconnecter à l'EOSC, le réseau informatique scientifique européen en nuage (*European Open Science Cloud*) dont le déploiement est encouragé par la Commission européenne (*High Level Expert Group on the European Open Science Cloud*, 2016). Un premier programme de soutien du ministère allemand pour l'enseignement et la recherche (*Bundesmi-*

35 Chiffres du 06.02.2017.

36 Voir *CLARIN* : <http://www.clarin.eu> (page consultée le 11 mars 2019).

37 Voir *Global Biodiversity Information Facility* : <http://www.gbif.org> (page consultée le 11 mars 2019).

nisterium für Bildung und Forschung, BMBF) lancé au niveau national fédéral³⁸ et plusieurs activités de soutien au niveau des États-Régions témoignent certes de premiers engagements pour la réalisation de ce NFDI, bien que les signes d'une coordination globale restent encore peu visibles. C'est pour cette raison qu'encore récemment des voix comme celles de la Conférence des présidents des établissements d'enseignement supérieur en 2017 et celles de l'Initiative allemande pour l'information en réseau (*Deutsche Initiative für Netzwerkinformation*, DINI) en 2017, deux instances où Peter Schirmbacher s'est impliqué, se sont fait entendre pour appeler à un consensus entre les pouvoirs publics fédéraux et ceux des États-Régions.

Reste qu'il faut retenir que les synergies entre les institutions des infrastructures de la recherche et leurs offres en sont encore à leur début. À l'avenir, il importe de favoriser la standardisation des entrepôts de données de la recherche et d'harmoniser les orientations et préconisations données aux personnels de recherche. Dans ce domaine, un consensus entre les acteurs impliqués, que ce soit les agences de financement, les dispositifs d'infrastructure ou d'autres entités de service, revêt une grande importance. Sur ce point, il convient de renforcer la souveraineté de l'infrastructure de recherche publique. Pour cela, il est nécessaire, comme en appelle le Rfll, que des mécanismes de soutien soient créés durablement en plus des infrastructures de données de la recherche. Ce point est particulièrement important pour éviter des dépendances envers des infrastructures agissant sur le terrain commercial comme *Figshare*³⁹ ou *Mendeley Data*⁴⁰. Pour cela, la HRK et la DINI soulignent que des moyens financiers de la part de l'État fédéral comme des Régions sont nécessaires (Hochschulrektorenkonferenz, 2016 ; *Deutsche Initiative für Netzwerkinformation*, 2017). S'il convient de débattre de cette question au niveau politique, la pratique témoigne d'ores et déjà de résultats positifs qui se sont matérialisés dans l'organisation autonome des communautés scientifiques : c'est ainsi que des offres importantes ont été créées, comme le wiki *Forschungsdaten.org*⁴¹ ou le groupe de travail commun de DINI et du réseau de compétences Nestor grâce à l'action de Peter Schirmbacher. De plus, par ses travaux dans son champ disciplinaire à l'Université Humboldt de Berlin, Peter Schirmbacher a contribué à impulser l'initiative *Gestion des données de recherche*⁴², ce qui a eu aussi son influence sur d'autres organismes d'enseignement supérieur en Allemagne comme à l'étranger (Simukovic et al., 2012 ; Kindling et al., 2013).

38 Voir *Bekanntmachung* : <https://www.bmbf.de/foerderungen/bekanntmachung-1233.html> (page consultée le 11 mars 2019).

39 Voir *Figshare* : <https://figshare.com> (page consultée le 11 mars 2019).

40 Voir *Mendeley Data* : <https://data.mendeley.com> (page consultée le 11 mars 2019).

41 Voir *Forschungsdaten.org* : <http://www.forschungsdaten.org/> (page consultée le 11 mars 2019).

42 Voir *Forschungsdatenmanagement* : <https://www.cms.hu-berlin.de/dataman> (page consultée le 11 mars 2019).

Il importe que cette contribution à la science se perpétue dans le sens de cette orientation de service, et ce en se professionnalisant davantage. Enfin, il faut notamment encore espérer que ce domaine important des Sciences de l'information en tant que discipline de recherche et d'enseignement qui a pris part dans une large mesure au développement et à la création d'offres d'infrastructures d'information adéquates à l'intérieur des établissements de recherche, à l'interconnexion des acteurs, au renforcement des communautés, se poursuive encore à l'Université Humboldt de Berlin à l'issue de cette distinction par l'éméritat du professeur Peter Schirmbacher.

Bibliographie

Benson D. A., Clark K., Karsch-Mizrachi I., Lipman D. J., Ostell J., Sayers E. W. (2015). "GenBank". In *Nucleic Acids Research*, vol. 43, n° D1, p. D30-D35. DOI 10.1093/nar/gku1216.

Bloom T., Ganley E., Winker M. (2014). "Data Access for the Open Access Literature: PLOS's Data Policy". In *PLoS Biology*, vol. 12, n° 2, e1001797. DOI 10.1371/journal.pbio.1001797.

Borgman C. L. (2007). *Scholarship in the Digital Age. Information, Infrastructure, and the Internet*, Cambridge, Massachusetts, MIT Press.

Burton A., Aryani A., Koers H., Manghi P., La Bruzzo S., Stocker M., Diepenbroek M., Schindler U., Fenner M. (2017). "The Scholix Framework for Interoperability in Data-Literature Information Exchange". In *D-Lib Magazine*, vol. 23, n° 1/2. DOI 10.1045/january2017-burton.

Commission européenne (2012). *Online survey on scientific information in the digital age*. Rapport, Publications Office of the European Union, Luxembourg. DOI 10.2777/7549.

Commission européenne (2016). *Open Innovation, Open Science, Open to the World*. Rapport, Publications Office of the European Union, Luxembourg. DOI 10.2777/061652.

Conze R., Lorenz H., Ulbricht D., Elger K., Gorgas T. (2017). "Utilizing the International Geo Sample Number Concept in Continental Scientific Drilling During ICDP Expedition COSC-1". In *Data Science Journal*, vol. 16, n° 1, p. 1-8. DOI 10.5334/dsj-2017-002.

Cousijn H., Kenall A., Ganley E., Harrison M., Kernohan D., Murphy F., Polischuk P., Martone M., Clark T. (2017). "A Data Citation Roadmap for Scientific Publishers". In *bioRxiv*, 100784. DOI 10.1101/100784.

Cruse P., Haak L., Pentz E. (2016). *Organization Identifier Project: A Way Forward*. Rapport. DOI: 10.5438/2906.

Deutsche Forschungsgemeinschaft (2013). "Sicherung guter wissenschaftlicher Praxis". In *Sicherung Guter Wissenschaftlicher Praxis*, Weinheim, Germany, Wiley-VCH Verlag GmbH, p. 1-109. DOI: 10.1002/9783527679188.oth1.

Deutsche Forschungsgemeinschaft (2014). *Leitfaden für die Antragstellung. Projektanträge*. DFG-Vordruck 54.01 – 06/14, Rapport. Disponible sur https://www.dfg.de/formulare/54_01/ (page consultée le 11 mars 2019).

Deutsche Forschungsgemeinschaft (2015). *Leitlinien zum Umgang mit Forschungsdaten*. Rapport. Disponible sur https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf (page consultée le 11 mars 2019).

- Deutsche Initiative für Netzwerkinformation (2017). *Koordiniertes Förderprogramm von Bund und Ländern zum Forschungsdaten nötig*. Rapport. Disponible sur <https://idw-online.de/de/news666601> (page consultée le 11 mars 2019).
- Diepenbroek M., Grobe H., Reinke M., Schindler U., Schlitzer R., Sieger R., Wefer G. (2002). "PANGAEA – an information system for environmental sciences". In *Computers & Geosciences*, vol. 28, n° 10, p. 1201-1210.
- Fenner M., Crosas M., Grethe J., Kennedy D., Hermjakob H., Rocca-Serra P., Berjon R., Karcher S., Martone M., Clark T. (2016). "A Data Citation Roadmap for Scholarly Data Repositories". In *bioRxiv*, 097196. DOI 10.1101/097196.
- Gemeinsame Wissenschaftskonferenz (2013). *Ergebnisse der Sitzung der Gemeinsamen Wissenschaftskonferenz (GWK) am 22. November 2013*. Rapport. Disponible sur <http://www.gwk-bonn.de/fileadmin/Pressemitteilungen/pm2013-16.pdf> (page consultée le 11 mars 2019).
- Helmholtz-Gemeinschaft (2016). *Digitale Forschungsdaten offen zugänglich machen*. Rapport. Disponible sur https://www.helmholtz.de/aktuell/presseinformationen/artikel/artikeldetail/digitale_forschungsdaten_offen_zugaenglich_machen/ (page consultée le 11 mars 2019).
- High Level Expert Group on the European Open Science Cloud (2016). *Realising the European Open Science Cloud*. Rapport. Disponible sur http://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf (page consultée le 11 mars 2019).
- Hochschulrektorenkonferenz (2014). "Management von Forschungsdaten als strategische Aufgabe der Hochschulleitungen". Rapport. Disponible sur https://www.hrk.de/fileadmin/redaktion/hrk/news/migrated/downloads/HRK_PM_Forschungsdaten_14052014.pdf (page consultée le 11 mars 2019).
- Hochschulrektorenkonferenz (2016). *Hochschulen sind wichtige Akteure innerhalb der Infrastrukturen für das Forschungsdatenmanagement. Spürbare Impulse von Bund und Ländern sind unverzichtbar*. Rapport. Disponible sur <https://www.hrk.de/themen/forschung/forschungsdaten-management/> (page consultée le 11 mars 2019).
- Humboldt-Universität zu Berlin (2014). *Grundsätze zum Umgang mit Forschungsdaten an der Humboldt-Universität zu Berlin*. Rapport. Disponible sur <https://www.cms.hu-berlin.de/de/dl/dataman/policy> (page consultée le 11 mars 2019).
- Kindling M., Schirmbacher P. (2013). "'Die digitale Forschungswelt' als Gegenstand der Forschung / Research on Digital Research / Recherche dans la domaine de la recherche numérique". In *Information – Wissenschaft & Praxis*, vol. 64, n° 2-3, p. 127-136. DOI 10.1515/iwp-2013-0017.
- Kindling M., Schirmbacher P., Simukovic E. (2013). "Forschungsdaten-

management an Hochschulen. Das Beispiel der Humboldt-Universität zu Berlin". In *Libreas*, vol. 23, p. 43-63. Disponible sur <http://nbn-resolving.de/urn:nbn:de:kobv:11-100212700> (page consultée le 11 mars 2019).

Klump J., Bertelmann R., Brase J., Diepenbroek M., Grob H., Höck H., Wächter J. (2006). "Data publication in the open access initiative". In *Data Science Journal*, vol. 5, p. 79-83. DOI <https://doi.org/10.2481/dsj.5.79>.

Kroes N. (2012). *Opening Science Through e-Infrastructures*. SPEECH/12/258, Rapport. Disponible sur <http://commentneelie.eu/speech.php?sp=SPEECH/12/258> (page consultée le 11 mars 2019).

Kuipers T., Van Der Hoeven J. (2009). *Insight into digital preservation of research output in Europe*. Rapport, PARSE.Insight. Disponible sur <https://libereurope.eu/wp-content/uploads/2010/01/PARSE.Insight.-Deliverable-D3.4-Survey-Report.-of-research-output-Europe-Title-of-Deliverable-Survey-Report.pdf> (page consultée le 11 mars 2019).

Lehnert K. A., Klump J., Arko R. A., Bristol S., Buczkowski B., Chan C., Chan S., Conze R., Cox S. J. D., Habermann T., Hangsterfer A., Hsu L., Milan A., Miller S. P., Noren A. J., Richard S. M., Valentine D. W., Whitenack T., Wyborn L. A., Zaslavsky I. (2011). "IGSN e.V.: Registration and Identification Services for Physical Samples in the Digital Universe". In *AGU Fall Meeting 2011*, IN13B-1324. Disponible

sur <http://abstractsearch.agu.org/meetings/2011/FM/IN13B-1324.html> (page consultée le 11 mars 2019).

Ludwig J., Enke H. (dir.) (2013). *Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt*, Glückstadt, Verlag Werner Hülsbusch.

Lynch C. (2003). "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age". In *ARL: A Bimonthly Report*, vol. 226. Disponible sur <http://old.arl.org/resources/pubs/br/br226/br226ir-print.shtml> (page consultée le 8 mars 2019).

Max-Planck-Gesellschaft (2003). *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. Rapport. Disponible sur <https://openaccess.mpg.de/Berlin-Declaration> (page consultée le 11 mars 2019).

Nature (2016). "Announcement: Where are the data?". In *Nature*, vol. 537, n° 7619, p. 138-138. DOI 10.1038/537138a.

Neumann J., Brase J. (2014). "DataCite and DOI names for research data". In *Journal of Computer-Aided Molecular Design*, vol. 28, n° 10, p. 1035-1041. DOI 10.1007/s10822-014-9776-5.

Pampel H., Bertelmann R. (2011). "'Data Policies' im Spannungsfeld zwischen Empfehlung und Verpflichtung". In Büttner S., Hobohm H.-C., Müller L. (dir.), *Handbuch For-*

- schungsdatenmanagement*, Bad Honnef, Bock und Herchen, p. 49-61. Disponible sur <http://nbn-resolving.de/urn:nbn:de:kobv:525-opus-2287> (page consultée le 11 mars 2019).
- Pampel H., Dallmeier-Tiessen S. (2014). "Open Research Data: From Vision to Practice". In Bartling S., Friesike S. (dir.), *Opening Science. The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*, Heidelberg, Springer, p. 213-224. DOI 10.1007/978-3-319-00026-8_14.
- Pampel H., Fenner M. (2016). "ORCID – Offener Standard zur Vernetzung von Forschenden". In *Nachrichten aus der Chemie*, vol. 64, n° 1, p. 57-58. DOI 10.1002/nadc.20164042239.
- Pampel H., Reuter E., Scholze F., Semrau A., Fritze F., Van De Sandt S., Goebelbecker H.-J., Schnepf E., Schirnbacher P., Kloska G., Vierkant P., Kindling M., Ulrich R., Neuberger N., Bertelmann R., Rücknagel J. (2016). *Forschungsdaten-Repositoryn mit Beteiligung deutscher Institutionen. Eine re3data.org-Analyse*. Rapport. DOI 10.5281/zenodo.50148.
- Pampel H., Vierkant P., Scholze F., Bertelmann R., Kindling M., Klump J., Goebelbecker H.-J., Gundlach J., Schirnbacher P., Dierolf U. (2013). "Making Research Data Repositories Visible: The re3data.org Registry". In *PLoS ONE*, vol. 8, n° 11, e78080. DOI 10.1371/journal.pone.0078080.
- Pfeiffenberger H., Carlson D. (2011). *Earth System Science Data (ESSD) – A Peer Reviewed Journal for Publication of Data*, vol. 17. DOI 10.1045/january2011-pfeiffenberger.
- Rat für Informationsinfrastrukturen (2016). *Leistung aus Vielfalt*. Rapport. Disponible sur <http://www.rfii.de/?wpdmdl=1998> (page consultée le 11 mars 2019).
- Rücknagel J., Vierkant P., Ulrich R., Kloska G., Schnepf E., Fichtmüller D., Reuter E., Semrau A., Kindling M., Pampel H., Witt M., Fritze F., Van De Sandt S., Klump J., Goebelbecker H.-J., Skarupianski M., Bertelmann R., Schirnbacher P., Scholze F., Kramer C., Fuchs C., Spier S., Kirchhoff A. (2015). *Metadata Schema for the Description of Research Data Repositories: version 3.0*. Rapport. DOI 10.2312/re3.008.
- Schäfer A., Pampel H., Pfeiffenberger H., Dallmeier-Tiessen S., Tissari S., Darby R., Giarretta K., Giarretta D., Gitmans K., Helin H., Lambert S., Mele S., Reilly S., Ruiz S., Sandberg M., Schallier W., Schrimpf S., Smit E., Wilkinson M., Wilson M. (2011). *Baseline Report on Drivers and Barriers in Data Sharing*. Rapport. DOI 10.5281/zenodo.8306.
- Schirnbacher P. (2015). "Informationsmanagement als Bestandteil der bibliotheks- und informationswissenschaftlichen Ausbildung". In *Bibliothek Forschung und Praxis*, vol. 39, n° 2, p. 143-149.

Simukovic E., Kindling M., Schirmbacher P. (2012). "Unveiling Research Data Stocks: A Case of Humboldt-Universität zu Berlin". In *iConference 2014 Proceedings*, p. 742-748. DOI 10.9776/14351.

Suber P. (2012). *Open Access*, Cambridge, Massachusetts, The MIT Press. Disponible sur <http://mitpress.mit.edu/books/open-access> (page consultée le 11 mars 2019).

Tenopir C., Allard S., Douglass K., Aydinoglu A. U., Wu L., Read E., Manoff M., Frame M. (2011). "Data Sharing by Scientists: Practices and Perceptions". In *PLoS ONE*, vol. 6, n° 6, e21101. DOI 10.1371/journal.pone.0021101.

Tenopir C., Dalton E. D., Allard S., Frame M., Pjesivac I., Birch B., Pollock D., Dorsett K. (2015). "Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide". In *PLoS ONE*, vol. 10, n° 8, e0134826. DOI 10.1371/journal.pone.0134826.

Treloar A., Harboe-Ree C. (2008). "Data management and the curation continuum: How the Monash experience is informing repository relationships". In *Proceedings of VALA 2008*, Melbourne. Disponible sur <http://www.vala.org.au/vala2008-proceedings/vala2008-session-6-treloar/> (page consultée le 11 mars 2019).

Ulbricht D., Elger K., Bertelmann R., Klump J. (2016). "panMetaDocs, eSciDoc, and DOIDB—An Infrastructure for the Curation and Publication of File-Based

Datasets for GFZ Data Services". In *ISPRS International Journal of Geo-Information*, vol. 5, n° 3, 25. DOI 10.3390/ijgi5030025.

Van Der Graaf M., Waaijers L., Davidson J., Hodson S., Christoffersen M., Heller A., Doove J., Grim R., Sessink L., Regner F., Pfeiffenberger H., Winkler-Neefs S. (2011). *A Surfboard for Riding the Wave. Towards a four country action programme on research data*. Disponible sur <https://pure.uvt.nl/ws/portalfiles/portal/1427340/Surfboard.pdf> (page consultée le 8 mars 2019).

Wilkinson M. D., Dumontier M., Aalbersberg IJ. J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.-W., Da Silva Santos L. B., Bourne P. E., Bouwman J., Brookes A. J., Clark T., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C. T., Finkers R., Gonzalez-Beltran A., Gray A. J. G., Groth P., Goble C., Grethe J. S., Heringa J., 't Hoen P. E. C., Hooft R., Kuhn T., Kok R., Kok J., Lusher S. J., Martone M. E., Mons A., Packer A. L., Persson B., Rocca-Serra P., Roos M., Van Schaik R., Sansone S.-A., Schultes E., Sengstag T., Slater T., Strawn G., Swertz M. A., Thompson M., Van Der Lei J., Van Mulligen E., Velterop J., Waagmeester A., Wittenburg P., Wolstencroft K., Zhao J., Mons B. (2016). "The FAIR Guiding Principles for scientific data management and stewardship". In *Scientific Data*, vol. 3, 160018. DOI 10.1038/sdata.2016.18.

Woutersen-Windhouwer S., Brandsma R. (2009). "Enhanced

Publications. State of the Art". In Vernooy-Gerritsen M. (dir.), *Enhanced Publications. Linking Publications and Research Data in Digital Repositories*, Amsterdam, Amsterdam University Press. Handle 11245/1.311760.

Zentralbibliothek für Wirtschaftswissenschaften (2016). *GeRDI wird Modell für vernetzte Forschungsdaten-Infrastruktur*. ZBW. Disponible sur <https://www.zbw.eu/de/ueber-uns/aktuelles/meldung/news/gerdi-wird-modell-fuer-vernetzte-forschungsdaten-infrastruktur/> (page consultée le 11 mars 2019).

