



**HAL**  
open science

# Étude comparative des services nationaux de données de recherche Facteurs de réussite

Hugo Catherine

► **To cite this version:**

Hugo Catherine. Étude comparative des services nationaux de données de recherche Facteurs de réussite. [Rapport de recherche] Comité pour la science ouverte. 2021, 88 p. hal-03563136

**HAL Id: hal-03563136**

<https://hal-lara.archives-ouvertes.fr/hal-03563136>

Submitted on 9 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Etude comparative des services nationaux de données de recherche Facteurs de réussite

Hugo Catherine

# Préambule

Le Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI), par l'intermédiaire du Comité pour la science ouverte (CoSO), a demandé à une équipe d'experts pilotée par Jean-Christophe Desconnets de l'IRD, accompagnée par le cabinet Dataactivist, de mener une "étude de faisabilité d'un service générique d'accueil et de diffusion des données simples" afin de "concevoir un service générique destiné aux données de recherche pour lesquelles les entrepôts disciplinaires existants ou à venir ne constituent pas une solution adaptée".<sup>1</sup>  
Cette étude a été réalisée avec le concours de Jean-Christophe Desconnets et Pascal Aventurier.

Dans le cadre de cette étude, il est apparu nécessaire de s'intéresser aux initiatives nationales développées par les autres pays en examinant celles-ci au travers des aspects suivants :

- gouvernance,
- fonctionnalités proposées,
- modes d'organisation,
- modèles économiques.

Pour réaliser cet état de l'art bibliographique, nous avons consulté les informations disponibles sur les sites Internet dédiés, ainsi qu'une sélection de rapports et de présentations réalisés par des acteurs des différents services. Ces sources sont regroupées au sein d'une bibliographie sélective.

Par ailleurs, des échanges avec l'ensemble des responsables ou opérateurs des services nationaux étudiés ont permis d'apporter des précisions et des détails utiles.

A ce propos, nous tenons à remercier Mingfang Wu (ARDC), Peter Dorn, Chris Baars et Vyacheslav Tykhonov (DANS), Philipp Conzett (Dataverse Norvège), Lee Wilson (DFDR), Tamsin Burland et John Kaye (JORH), Kerstin Soltau (RADAR) et Matthew Woollard (UKDS).

Suite aux différentes recherches, la sélection porte sur les services suivants :

- Australian Research Data Cloud (ARDC), Australie,
- Dataverse Norway (DNO), Norvège,
- Dutch Archiving & Networking Structure (DANS), Pays-Bas,
- Dépôt Fédéré des Données de Recherche (DFDR), Canada,
- JISC Open Research Hub (JORH) Royaume-Uni,
- Research Data Repository (RADAR), Allemagne,
- UK Data Service (UKDS), Royaume-Uni.

Si tous ces services ont pour priorité d'accueillir des données de recherche, il faut noter que certains d'entre eux proposent des interfaces associant logiciels et/ou publications à leur cible prioritaire.

Le cœur de cible de ces services reste bien évidemment les acteurs de la recherche scientifique (chercheurs, enseignants-chercheurs, doctorants, personnels des services supports). Cependant, l'objectif d'un certain nombre de dispositifs nationaux est aussi de répondre aux besoins d'autres types de publics : journalistes, acteurs économiques, acteurs politiques

Tous ces entrepôts ont pour mission d'accueillir des données génériques en dehors de l'UKDS pour lequel seules des données relevant des domaines des sciences de l'homme et de la société sont acceptées à ce jour. Ce service a cependant retenu notre attention en raison de sa maturité (créé en 1967) et le rôle qu'il joue pour soutenir et faciliter la recherche pluridisciplinaire, objectif clé de l'[United Kingdom Research and Innovation](https://www.ukri.gov.uk/) (UKRI).

---

<sup>1</sup> <https://www.ouvrirlascience.fr/service-generique-daccueil-et-de-diffusion-de-donnees-simples/>



# Sommaire

<b>GLOSSAIRE</b> .....	<b>7</b>
<b>SIGLES</b> .....	<b>9</b>
<b>RÉSUMÉ</b> .....	<b>12</b>
<b>1 GOUVERNANCE</b> .....	<b>13</b>
1.1 AUSTRALIE - AUSTRALIAN RESEARCH DATA COMMONS (ARDC).....	13
Né de la fusion de trois infrastructures préexistantes, un service piloté par l'agence de financement national et des représentants d'université dont l'objectif est de rendre librement accessible un volume maximal de données à tout type de public.....	13
1.2 NORVÈGE - DATAVERSE NORWAY (DNO).....	14
Une infrastructure nationale confiée à l'université de Trömso, des instances consultatives regroupant l'ensemble des universités partenaires.....	14
1.3 PAYS-BAS - DUTCH ARCHIVED AND NETWORKED SERVICE (DANS).....	15
Un institut national relevant de l'Académie des sciences néerlandaises et de l'agence de financement national dont l'objectif est de proposer un accès pérenne aux productions scientifiques.....	15
1.4 CANADA - DÉPÔT FÉDÉRÉ DES DONNÉES DE RECHERCHE.....	16
« Fournir une plateforme unique à partir de laquelle les données de recherche peuvent être chargées, organisées, préservées, explorées, citées et partagées » : un projet mené conjointement par l'association des bibliothèques de recherche canadiennes et l'infrastructure nationale de calcul.....	16
1.5 GRANDE-BRETAGNE - JISC OPEN RESEARCH HUB (JORH).....	18
Pilotée par le consortium national des services d'information du Royaume-Uni, une plateforme de services dont l'objectif est d'apporter des solutions facilement adaptables aux environnements techniques des universités partenaires.....	18
1.6 ALLEMAGNE - RESEARCH DATA REPOSITORY (RADAR).....	19
Une infrastructure distribuée visant à promouvoir la découverte et la réutilisation de données de longue traîne en renforçant leur traçabilité, leur reproductibilité et leur transparence.....	19
1.7 GRANDE-BRETAGNE - UNITED KINGDOM DATA SERVICE (UKDS) .....	20
Service national des données de sciences sociales distribué entre une université hôte (université d'Essex) et des institutions partenaires.....	20
<b>2 OFFRE DE SERVICES</b> .....	<b>22</b>
2.1 PRODUCTEURS DE DONNÉES.....	22
2.1.1 <i>Dépôt de données</i> .....	22
2.1.2 <i>Archivage pérenne des données</i> .....	22
2.1.3 <i>Stockage des données</i> .....	23
2.1.4 <i>Aide à la gestion des données</i> .....	23
2.1.5 <i>Accès réservé aux relecteurs scientifiques</i> .....	24
2.1.6 <i>Soutien financier dédié à la gestion et à l'ouverture de données</i> .....	24
2.1.7 <i>Gestion des accès</i> .....	25
2.2 UTILISATEURS DES DONNÉES .....	25
2.2.1 <i>Outil de découverte interne versus portail de recherche fédéré</i> .....	25
2.2.2 <i>Services de calcul, d'analyse et de collaboration</i> .....	27
2.2.3 <i>Services d'accès aux données sensibles</i> .....	27
2.3 RÉSEAUX D'EXPERTISE ET D'ACCOMPAGNEMENT .....	28
<b>3 ORGANISATION ET MUTUALISATION DES SERVICES</b> .....	<b>29</b>
3.1 ARDC : FACILITER L'ÉMERGENCE DE COMMUNAUTÉS DE PRATIQUES ET LE DÉVELOPPEMENT D'OUTILS PARTAGÉS .....	29
3.2 DANS : DES SERVICES DISTRIBUÉS AU SEIN D'UN RÉSEAU DE PARTENAIRES .....	30
3.3 DATAVERSE NO : UN ENTREPÔT ADMINISTRÉ PAR LES SERVICES D'UNE UNIVERSITÉ, DES COLLECTIONS GÉRÉES PAR LES UNIVERSITÉS PARTENAIRES.....	30

3.4	DFDR : DES SERVICES FÉDÉRÉS ALLIANT INFRASTRUCTURES DE DONNÉES, UNIVERSITÉS ET RÉSEAU DE BIBLIOTHÉCAIRES .....	31
3.5	JORH : DES PRESTATIONS LOGICIELLES INTÉGRÉES AUX ENVIRONNEMENTS TECHNIQUES DES INSTITUTIONS CLIENTES .....	32
3.6	RADAR : DES SERVICES DISTRIBUÉS ENTRE ÉTABLISSEMENTS PARTENAIRES.....	32
3.7	UKDS : UN SERVICE DISTRIBUÉ SOUS LA FORME DE CONTRATS DE SOUS-TRAITANCE ENTRE L'UNIVERSITÉ D'ESSEX ET LES INSTITUTIONS PARTENAIRES.....	33
<b>4</b>	<b>MODÈLE ÉCONOMIQUE.....</b>	<b>34</b>
4.1	ARDC : UNE PLATEFORME DE SERVICES ENTIÈREMENT FINANCÉE PAR LE GOUVERNEMENT AUSTRALIEN .....	34
4.2	DANS : UN MODÈLE ALLIANT FINANCEMENT STRUCTUREL, SOUTIEN AUX PROJETS ET CONTRIBUTION DES ÉTABLISSEMENTS PARTENAIRES .....	34
4.3	DATAVERSE NO : UN SERVICE FINANCÉ PAR L'UNIVERSITÉ HÔTE ET LA CONTRIBUTION DES ÉTABLISSEMENTS PARTENAIRES .....	35
4.4	DFDR : UN MODÈLE EN COURS DE DÉFINITION .....	35
4.5	JORH : MODÈLE ÉCONOMIQUE BASÉ SUR L'ABONNEMENT À DES OFFRES DE SERVICE .....	35
4.6	RADAR : FINANCEMENT RÉPARTI À PARTS ÉGALES ENTRE L'ÉTABLISSEMENT D'ACCUEIL ET DES INSTITUTIONS PARTENAIRES .....	36
4.7	UKDS : FINANCEMENT STRUCTUREL DE L'ESRC ET DES ÉTABLISSEMENTS PARTENAIRES .....	36
<b>5</b>	<b>SYNTHÈSE.....</b>	<b>37</b>
5.1	DEGRÉ DE MATURITÉ DES SERVICES.....	37
5.2	PRIORITÉS AFFICHÉES PAR LES DISPOSITIFS .....	38
5.2.1	<i>Quels objectifs ?</i> .....	38
5.2.2	<i>Pour quels publics ?</i> .....	39
5.2.3	<i>Offre de services</i> .....	41
5.3	DES SERVICES CO-CONSTRUITS ET DÉVELOPPÉS AVEC L'ENSEMBLE DES ACTEURS .....	46
5.3.1	<i>Avec les partenaires institutionnels</i> .....	46
5.3.2	<i>Au sein de réseaux nationaux et internationaux</i> .....	46
5.3.3	<i>En sollicitant les communautés de recherche</i> .....	47
5.4	MODÈLES ÉCONOMIQUES.....	48
5.5	ECONOMIES D'ÉCHELLE .....	49
5.6	PÉRENNITÉ DU SERVICE.....	52
5.6.1	<i>Evaluation des services</i> .....	52
5.6.2	<i>Démarches de certification</i> .....	53
	<b>CONCLUSION.....</b>	<b>55</b>
	<b>S'INSÉRER DE MANIÈRE PRAGMATIQUE ET AGILE AU SEIN D'ÉCOSYSTÈMES NATIONAUX ET INTERNATIONAUX DYNAMIQUES.....</b>	<b>55</b>
	<b>SITES WEB UTILES.....</b>	<b>58</b>
	<b>BIBLIOGRAPHIE.....</b>	<b>59</b>
	<b>ANNEXES.....</b>	<b>64</b>
	ANNEXE 1. IDENTIFIANTS, MÉTADONNÉES ET INDICATEURS UTILISÉS PAR LES SERVICES NATIONAUX DE DONNÉES .....	65
	ANNEXE 2. FONCTIONNALITÉS DES SERVICES NATIONAUX DE DONNÉES.....	70
	ANNEXE 3. RESSOURCES DES SERVICES NATIONAUX.....	73
	ANNEXE 4. FAIR-AWARE, OUTIL D'AUTO-ÉVALUATION AUX PRINCIPES FAIR DÉVELOPPÉ PAR DANS NL, DIGITAL CURATION CENTER ET L'UNIVERSITÉ DE BRÊME .....	78
	ANNEXE 5. ARBRE DE DÉCISION DOI DÉVELOPPÉ PAR LE SERVICE DOI DE L'AUSTRALIAN RESEARCH DATA COMMONS (ARDC).....	79

ANNEXE 6. EXEMPLE DE RÉSULTAT PROPOSÉ PAR QUALIBANK, BANQUE DE DONNÉES QUALITATIVES PROPOSÉE PAR L'UKDS .....	80
ANNEXE 7. CAPTURES D'ÉCRAN EXTRAITES DU DATA COSTING TOOL DÉVELOPPÉ PAR L'UKDS .....	81
ANNEXE 8. QA MY DATA, OUTIL DE CURATION DÉVELOPPÉ PAR L'UKDS .....	82
ANNEXE 9. FORCES ET FAIBLESSES DU MODÈLE ÉCONOMIQUE DE DATAVERSE NORVÈGE (2020) .....	83
ANNEXE 10. ÉTAT DES LIEUX DE L'IMPLÉMENTATION DES PRINCIPES FAIR AU SEIN DE DATAVERSE NORVÈGE (2020) .....	84
<b>TABLE DES MATIÈRES .....</b>	<b>85</b>
<b>TABLE DES ILLUSTRATIONS .....</b>	<b>88</b>

# Glossaire

CoreTrustSeal	“Organisation de certification mise en place conjointement par le DSA (Data Seal of Approval) et le ICSU WDS (the International Council for Science’s World Data System). Cette certification garantit aux déposants que leurs données seront protégées et gérées de manière optimale.” <sup>2</sup>
Curation des données	Activités visant à maintenir, préserver, valoriser les données numériques de la recherche à travers leur cycle de vie. Le travail de curation porte sur la qualité des données (correction des erreurs) et leur pérennité.
Cycle de vie des données	“Ensemble des activités et opérations nécessaires à la gestion des données de recherche, tout au long de leur cycle de vie.” <sup>3</sup>  Le modèle proposé par UKDA définit les étapes suivantes <sup>4</sup> : <ul style="list-style-type: none"> <li>● Création ou collecte des données</li> <li>● Traitement des données</li> <li>● Analyse des données</li> <li>● Conservation des données</li> <li>● Accès aux données</li> <li>● Réutilisation des données</li> </ul>
Données de longue traîne	Ensembles de données faiblement structurées qui représentent la majeure partie des données produites par les communautés de recherche. Elles résultent d'une multitude d'activités de recherche, relevant de disciplines très diverses d'où leur caractère hétérogène. Pour ces données, l'utilisation de normes et de bonnes pratiques de gestion reste peu partagée ce qui explique leur dispersion. Si elles peuvent aussi être liées à un article, elles restent le plus souvent non publiées.
Easy	Système d'archivage pérenne des données développé par le réseau DANS
FAIRsFAIR Fostering Fair Data Practices in Europe	Projet européen visant à fournir des solutions pratiques pour l'utilisation des principes de données FAIR tout au long du cycle de vie des données de recherche, notamment l'élaboration de normes mondiales pour la certification FAIR des entrepôts et des données qu'ils contiennent.
Five Safes	Méthode modélisée par l'UKDS reposant sur un ensemble de principes qui permettent aux services de données de fournir un accès sûr aux données pour la recherche, répondant ainsi aux

<sup>2</sup> <https://doranum.fr/stockage-archivage/evaluation-labellisation-sites-archivage-perenne/>

<sup>3</sup> <https://doranum.fr/glossaire-donnees-recherche/>

<sup>4</sup> [https://www.inist.fr/wp-content/uploads/donnees/co/module Donnees recherche 7.html](https://www.inist.fr/wp-content/uploads/donnees/co/module_Donnees_recherche_7.html) )

	besoins de protection des données, mais aussi aux exigences d'ouverture scientifique et de transparence. Five Safes a été adopté par une série de laboratoires sécurisés, dont l'Office for National Statistics. <sup>5</sup>
Logiciel en tant que service Software as a service (SaaS)	Modèle de distribution de logiciels qui permet aux clients d'externaliser intégralement un aspect de leur système d'information auprès d'un prestataire. Les logiciels sont installés par le prestataire sur un serveur distant, l'accès est soumis au paiement d'un abonnement.
Narcis	Outil de découverte proposé par le réseau DANS qui donne accès à des publications, des données de recherche, des informations concernant les chercheurs, les institutions et les projets
Nectar Cloud	Service de stockage, calcul et collaboration de l'ARDC basé sur la suite logicielle OpenStack <sup>6</sup>
Research Data Management Gestion des données de recherche	La gestion des données fait référence au stockage, à l'accès et à la conservation des données produites lors d'activités de recherche. Les pratiques de gestion des données couvrent l'ensemble du cycle de vie des données, de la planification du projet de recherche à sa réalisation, de la sauvegarde des données au fur et à mesure de leur création et de leur utilisation à leur préservation à long terme. <sup>7</sup>
Reshare	Service d'auto-archivage de données proposé par UKDS
Service d'archivage pérenne	Service de conservation à moyen et long terme
Service de dépôt	Service de description et de préservation proposé par un entrepôt de données, à des fins de partage (accès et réutilisation)
Service de stockage sécurisé	Service de sauvegarde visant à assurer la continuité de l'exploitation sur du court terme
Vocabulaire contrôlé	"Un vocabulaire contrôlé est un ensemble organisé de mots et expressions utilisés pour indexer du contenu et/ou le retrouver par navigation ou recherche. Typiquement, il inclut des termes préférentiels et leurs variantes et opère dans un périmètre défini ou décrit un domaine spécifique". <sup>8</sup>

<sup>5</sup> <https://www.ukdataservice.ac.uk/use-data/secure-lab/security-philosophy>

<sup>6</sup> <https://www.ukdataservice.ac.uk/use-data/secure-lab/security-philosophy>

<sup>7</sup> Whyte, A., Tedds, J. (2011). 'Making the Case for Research Data Management'. DCC Briefing Papers. Edinburgh: Digital Curation Centre. Disponible à l'adresse suivante : : <https://www.dcc.ac.uk/guidance/briefing-papers/making-case-rdm> (consulté le 11 juillet 2020)

<sup>8</sup> [https://www.coar-repositories.org/files/coar-cv-infog-f\\_27052042-3.pdf](https://www.coar-repositories.org/files/coar-cv-infog-f_27052042-3.pdf)

# Sigles

AAF	Australian Access Federation
AARNET	Australia's academic and research network
ABRC	Association des bibliothèques de recherche du Canada
ACNC	Australian Charities and not for-profit commissions
ANDS	Australian National Data Service
ANR	Agence Nationale de la Recherche
ARDC	Australian Research Data Cloud
AUN	Artic University of Norway - Trömso
CMIST	Cathie Marsh Institute for Social Research - Université de Manchester
COSO	Comité pour la science ouverte
CRIS	Current Recent Information Systems
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CTS	Core TrustSeal
DANS	Dutch Archiving & Networking Structure
DFDR	Dépôt Fédéré des Données de Recherche
DNO	Dataverse Norway
DOI	Digital Object Identifier
DPS	Research Repositories Dynamic Purchasing System
EDINA	Edinburgh University Data Library
EOSC	European Open Science Cloud
ESDS	Economic and Social Data Service
ESRC	Economic and Social Research Council
FAIR	Findable Accessible Interoperable Reusable
FAIRsFAIR	Fostering Fair Data in Practices in Europe
FIZ-K	Fachinformationszentrum Karlsruhe
HUNI	Humanities Networked Infrastructure

HZI	Helmholtz Centre for Infection Research
IGSN	International Geo sample number
IPB	Institute for Plant Biochemistry
IRD	Institut de Recherche et de Développement
JISC	Joint Information Systems Committee
JORH	JISC Open Research Hub
KIT	Karlsruhe Institute of Technology
KNAW	Koninklijke Nederlandse Akademie van Wetenschappen (Académie des sciences néerlandaises)
MESRI	Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation
MVF	Minimum viable feature
MVP	Minimum viable product
NCI	National Computational Infrastructure - Australia
NCRIS	National collaborative research infrastructure strategy
NECTAR	National eResearch Collaboration Tools and Resources
NFDI	Nationale Forschungsdateninfrastruktur - Infrastructure nationale des données de recherche
NWO	Netherlands Organisation for Scientific Research
OCDE	Organisation de Coopération et de Développement Economique
QCIF	Queensland Cyber Infrastructure Foundation
RADAR	Research Data Repository
RAID	Research Activity Identifier
RDM	Research Data Management
RDNL	Research Data Netherlands
RDS	Research Data Services
SAAS	Software as a service
SCC	Steinbuch Center for Computing
TROLLing	The Trömsö Repository of Language and Linguistics
UKDA	United Kingdom Data Archive

UKDS	United Kingdom Data Service
UKRI	United Kingdom Research and Innovation
ZIH TU	Zentrum für Informations-dienste und Hochleis-tungsrechnen (ZIH) Technische Universität - Centre de services d'information et de calcul haute performance - Université technique de Dresde

# Résumé

Cette étude s'inscrit dans le cadre de travaux, commandés par le Comité pour la Science Ouverte/Ministère de l'enseignement supérieur, de la recherche et de l'innovation (COSO/MESRI), à l'intention de l'Institut de la recherche pour le développement (IRD), qui visent à évaluer la faisabilité d'un service générique d'accueil et de diffusion des données simples.

Ce travail a pour objectif de nourrir la réflexion en présentant un état des lieux de différents services nationaux de données.

Il porte sur sept infrastructures de services qui se sont développées dans des contextes différents mais qui partagent un grand nombre de points communs, à commencer par offrir un accès à un service de données à l'ensemble des chercheurs, à l'échelle nationale.

Que ce soit pour faciliter l'accès et le partage de l'information scientifique ou répondre aux exigences des bailleurs de fonds, les services nationaux de données qui font l'objet de cette étude se sont développés autour d'un noyau de fonctionnalités communes :

- un service de dépôt,
- un service d'archivage pérenne,
- un service d'assistance et d'accompagnement à la gestion des données.

La plupart d'entre eux ont aussi choisi de proposer un service de signalement et d'accès qui s'appuie soit sur un outil de découverte interne, soit sur un moteur de recherche fédérant l'accès à des données provenant de services externes à celles déposées dans leurs entrepôts.

Fruit de projets au long cours, les services nationaux de données restent des initiatives plutôt récentes, l'une d'entre elles doit ouvrir ses services au début de l'année 2021.

Compte tenu du contexte très dynamique qui voit se multiplier les projets d'entrepôts institutionnels, le "temps long" qui caractérise la mise en place de services nationaux incite les opérateurs à opter pour des solutions techniques et organisationnelles flexibles, de manière à pouvoir s'adapter à l'évolution des usages et des attentes des utilisateurs.

Dès le stade du projet, la consultation des acteurs clés et représentatifs de la diversité des écosystèmes scientifiques s'avère incontournable pour construire l'offre de services au plus près des attentes des publics visés.

Une fois le service mis en place, ce principe de co-construction prévaut tout autant : rôle déterminant des institutions partenaires au sein des comités consultatifs, appels à projets de service, implication des chercheurs pour développer de nouveaux outils (ex. retours exprimés par les utilisateurs, recueil des besoins, tests).

De la même manière, les services de données inscrivent leur activité au sein de réseaux nationaux et internationaux, ce qui participe à définir et renforcer leur rôle.

La majeure partie de ces infrastructures nationales s'appuie sur le financement structurel assuré par le ministère de la recherche ou un bailleur de fonds public. En s'appuyant sur la collaboration ou la fédération, l'objectif de ces initiatives nationales est aussi de réduire les coûts pour l'ensemble des partenaires.

Ces économies d'échelle, l'évaluation continue des services, l'interopérabilité des outils avec les écosystèmes existants et la fiabilité des solutions techniques représentent des points d'attention déterminants pour l'ensemble de ces dispositifs car ils conditionnent leur pérennité.

# 1 Gouvernance

Cette section est dédiée au contexte politique dans lequel chaque dispositif s'est construit ainsi qu'à la façon dont sont structurés les organes de décision et de mise en œuvre des différents dispositifs.

## 1.1 Australie - Australian Research Data Commons (ARDC)

Né de la fusion de trois infrastructures préexistantes, un service piloté par l'agence de financement national et des représentants d'université dont l'objectif est de rendre librement accessible un volume maximal de données à tout type de public.

A l'initiative du National collaborative research infrastructure strategy ([NCRIS](#)), l'ARDC est né en 2018 de la fusion des trois infrastructures suivantes :

- Australian National Data Service (ANDS - 2009), service chargé de favoriser la sensibilisation et la diffusion liées à la gestion des données
- National eResearch Collaboration Tools and Resources (Nectar - 2009) dont le rôle est de produire des outils et des services favorisant la collaboration
- Research Data Services (RDS) qui visent à mettre en place et développer des services de stockage et d'accès aux données

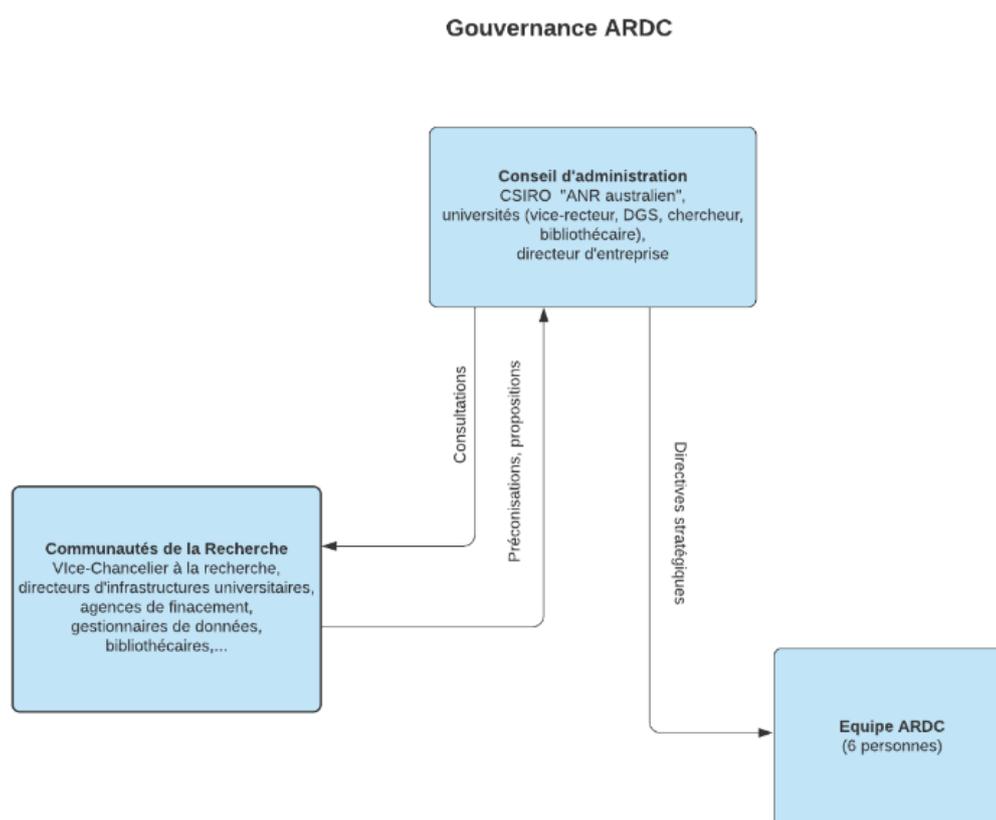
L'ARDC relève du statut de société à responsabilité limitée. Par ailleurs, elle est enregistrée comme organisation caritative auprès de l'Australian Charities and not for-profit commissions (ACNC).

Onze universités ainsi que l'agence de financement nationale Commonwealth Scientific and Industrial Research Organisation ([CSIRO](#)) la composent. La gouvernance s'opère à trois niveaux :

- un comité de gouvernance où se réunissent sept membres, représentants des agences gouvernementales et de différents domaines de l'université (service informatique, service de recherche, bibliothèque),
- l'équipe de l'ADRC (6 personnes) chargée de la mise en œuvre de la stratégie,
- des services fonctionnels auxquels sont rattachés les personnels.

Chaque plan stratégique quinquennal est le résultat de consultations avec la communauté de la recherche australienne : vice-Chancelier à la recherche, directeurs d'infrastructure universitaire, agences de financement, gestionnaires de données, bibliothécaires... Tous ces acteurs contribuent à la prise de décision, notamment pour déterminer les priorités de financement.

**Tableau 1. Schéma de gouvernance de l'ARDC (Australie)**



## 1.2 Norvège - Dataverse Norway (DNO)

Une infrastructure nationale confiée à l'université de Trömso, des instances consultatives regroupant l'ensemble des universités partenaires.

Créé en 2017, l'entrepôt national générique DNO a pour objectif de faciliter l'accès aux données publiées et leur (ré)utilisation dans une perspective à long terme. A ce jour, il regroupe neuf universités norvégiennes partenaires.

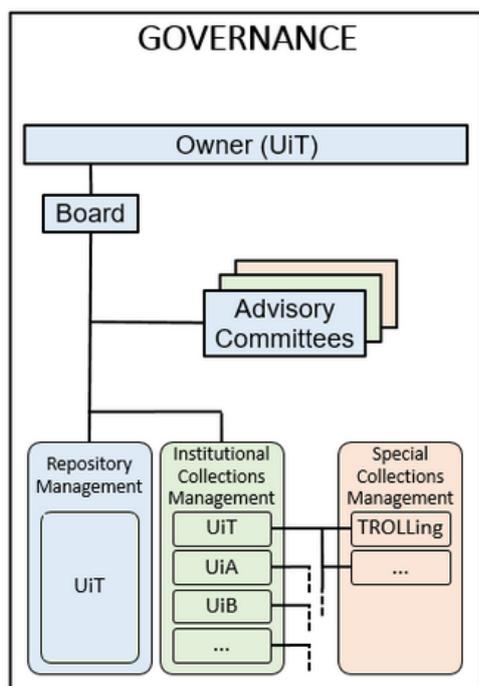
La responsabilité et la gestion de l'infrastructure DNO a été confiée à l'université de Trömso - Artic University of Norway (AUN).

Placé sous l'autorité du président de l'AUN, le comité de gouvernance est composé de personnels de cette même université, à savoir le vice-président recherche, le directeur de la bibliothèque universitaire, le directeur du service informatique, ainsi que le directeur du service de recherche et développement.

Un comité consultatif dédié à DNO vient apporter conseils, suggestions et assistance au comité de gouvernance ; ce comité consultatif est composé d'un représentant de DNO et d'un représentant pour chaque établissement partenaire.

Par ailleurs, chaque collection (collection dédiée à une institution ou collection thématique) dispose d'un comité consultatif spécifique. Composés de représentants de l'institution partenaire et de représentants de DNO, ces comités viennent apporter conseils et assistance aux gestionnaires de collections.

**Tableau 2. Schéma de gouvernance de Dataverse Norvège<sup>9</sup>**



La création d'un comité de pilotage dédié aux questions techniques est envisagée au cas où DNO venait à se développer de façon conséquente.

### 1.3 Pays-Bas - Dutch Archived and Networked Service (DANS)

Un institut national relevant de l'Académie des sciences néerlandaises et de l'agence de financement national dont l'objectif est de proposer un accès pérenne aux productions scientifiques.

Créé en 2005 par NWO (Netherlands Organisation for Scientific Research, agence de financement national) et KNAW (Académie des sciences néerlandaises), le réseau DANS se définit comme « l'institut néerlandais pour l'accès pérenne aux ressources de recherche numériques ».

En 2011, DANS a intégré les services Narcis (outil de découverte) et Easy (système d'archivage pérenne des données) puis l'entrepôt national [DataverseNL](https://site.uit.no/dataverseno/files/2020/06/DataverseNO_organisation_chart_english_v2.png) (DNL), à partir de 2014.

Animé par une équipe de 58 personnes, le réseau DANS est composé de trois services ("Projets et politique", "Archive & Support", "Recherche & Innovation"), dont les activités sont coordonnées par le directeur de DANS.

Le comité de pilotage de DANS supervise la gestion et le fonctionnement du réseau, des politiques menées par le directeur, ainsi que les résultats obtenus.

Il est assisté par un comité consultatif (advisory board) qui lui fait part de ses recommandations en matière de stratégie et de politique générale. Composée de quatre membres relevant de différentes

<sup>9</sup> [https://site.uit.no/dataverseno/files/2020/06/DataverseNO\\_organisation\\_chart\\_english\\_v2.png](https://site.uit.no/dataverseno/files/2020/06/DataverseNO_organisation_chart_english_v2.png)

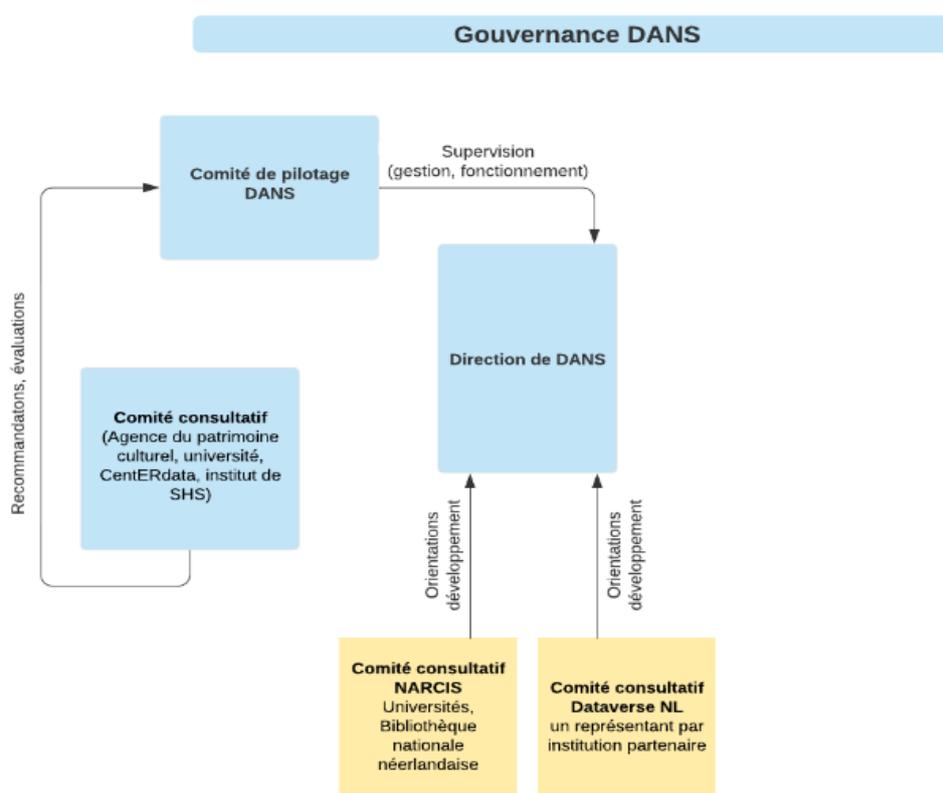
institutions (Agence du patrimoine culturel, université, CentERdata, institut de SHS), cette instance a aussi pour mission de répondre au rapport de la commission d'évaluation DANS liée à une évaluation externe semestrielle.

Par ailleurs, un comité consultatif propre à Narcis oriente les choix de la direction de DANS au sujet du développement et du fonctionnement de Narcis et de ses membres. Il est composé de 7 universités et de la bibliothèque nationale néerlandaise.

De la même façon, DNL dispose d'un comité consultatif dont le rôle est d'indiquer à la direction de DANS les axes stratégiques, ainsi que les développements souhaitables. Chacune des 13 institutions partenaires y est représentée.

Résultat des échanges entre ces différentes instances, la politique de DANS prend la forme de stratégies pluriannuelles (2015-2020).

**Tableau 3. Schéma de gouvernance du réseau DANS (Pays-Bas)**



## 1.4 Canada - Dépôt Fédéré des Données de Recherche

« Fournir une plateforme unique à partir de laquelle les données de recherche peuvent être chargées, organisées, préservées, explorées, citées et partagées »<sup>10</sup> : un projet mené conjointement par l'association des bibliothèques de recherche canadiennes et l'infrastructure nationale de calcul.

<sup>10</sup> [https://www.frdr-dfdr.ca/docs/fr/a\\_propos/#services-de-depot](https://www.frdr-dfdr.ca/docs/fr/a_propos/#services-de-depot) (consulté le 20/09/2020)

Depuis 2014, le [réseau Portage](#), [Calcul Canada](#) et l'Association des bibliothèques de recherche du Canada (ABRC) mènent des projets pilotes visant à évaluer la faisabilité de services nationaux de gestion des données de recherche.

Par ailleurs, la plupart des universités canadiennes se sont dotées d'un entrepôt Dataverse intégré à un portail commun [Scholars Portal Dataverse](#) qui est géré par le consortium des bibliothèques universitaires de l'Ontario (Ontario Council of Academic Libraries).

Le service DFDR vient compléter le portail national Dataverse en apportant une solution d'hébergement pour les très grands ensembles de données, et en proposant un accès unique à des collections de données provenant d'acteurs différents (scientifiques, politiques).

Actuellement, l'entrepôt fédéré de données de recherche ([DFDR](#)) est encore en phase de test. L'objectif est de lancer véritablement le service au début de l'année 2021.

Les modalités de gouvernance restent à définir en fonction de l'évolution de la plateforme.

A ce jour, DFDR est géré conjointement par deux organismes :

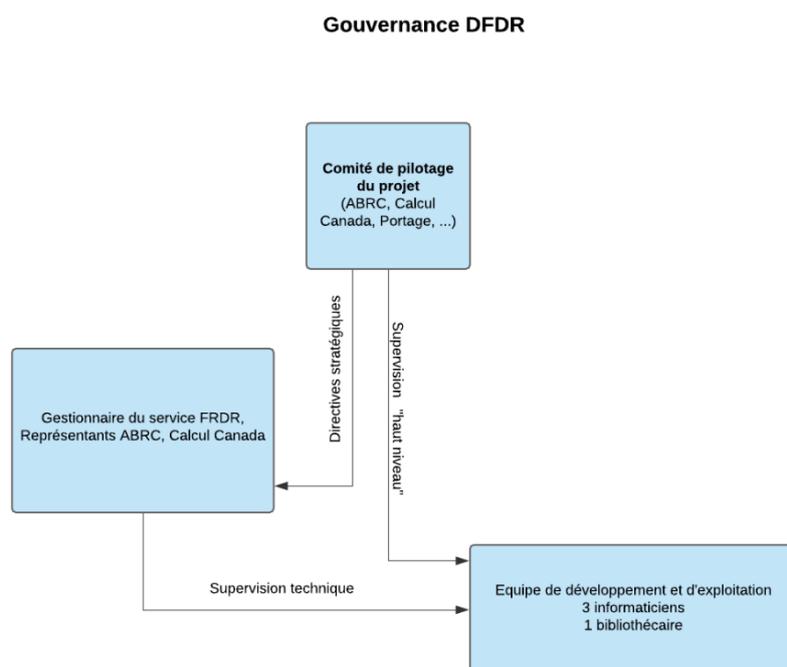
- le réseau Portage de l'ABRC, organisme national de gestion des données de recherche des bibliothèques
- et Calcul Canada, fournisseur national d'informatique de recherche de pointe

Le comité de pilotage du projet assure la direction du projet et veille à son développement. Il comprend des représentants du service Portage, de Calcul Canada, de l'ABRC, d'universités, des consortiums [ACEnet](#) et de [SHARCNET](#), de l'équipe de développement, de la Nouvelle organisation d'infrastructure de recherche numérique ([NOIRN](#)) ainsi que le président du comité consultatif des usagers.

La supervision du service DFDR (développement et opérations du service) est assurée par le gestionnaire du service en étroite collaboration avec des représentants des deux autres partenaires (Calcul Canada et ABRC).

L'équipe de développement et d'exploitation est composée de l'équipe centrale de l'Université de Saskatchewan (4 personnes) appuyée par un bibliothécaire de l'université Simon Fraser, chargé de la gestion des données et des systèmes de recherche.

**Tableau 4. Schéma de gouvernance du DFDR (Canada)**



## 1.5 Grande-Bretagne - JISC Open Research Hub (JORH)

Pilotée par le consortium national des services d'information du Royaume-Uni, une plateforme de services dont l'objectif est d'apporter des solutions facilement adaptables aux environnements techniques des universités partenaires.

JORH est le résultat du projet pilote "Research data shared service" mené par le le JISC ([Joint Information Systems Committee](#)) de 2015 à 2018.

L'objectif poursuivi était de fournir une "solution nationale basée sur les besoins de la recherche britannique dans son ensemble, en fournissant une infrastructure de recherche à ceux qui n'en disposaient pas et en améliorant l'interopérabilité entre les systèmes des universités qui en disposaient déjà."<sup>11</sup>

La plateforme proposée par le JISC vise notamment à permettre aux institutions de répondre aux exigences des bailleurs de fonds, en matière d'ouverture et de bonne gestion des données, ainsi qu'à fluidifier les échanges de données et de métadonnées entre systèmes de recherche. Du point de vue opérationnel, l'objectif est de proposer des services interopérables qui s'adaptent à des solutions techniques très différentes d'une université à l'autre. En connectant ses services à une très large gamme d'outils (ex. entrepôts institutionnels, Orcid, Datacite), le JORH doit simplifier au maximum le travail des chercheurs : ils déposent dans l'entrepôt de leur institution, et en fonction de l'offre de service souscrite par leur établissements, leurs productions sont archivées de façon pérenne, se voient attribuer un DOI, connecter à leur profil Orcid,...

Durant ce projet, plus de 70 universités britanniques ont été sollicitées pour recueillir les besoins et le JISC a travaillé plus étroitement avec un panel représentatif de 16 institutions.

<sup>11</sup> "Integrating the future of research outputs" : Dom Fripp  
<https://wonkhe.com/blogs/integrating-the-future-of-research-outputs/>

JORH est en production depuis 2019. La responsabilité et la gestion de l'infrastructure nationale sont du ressort du consortium national JISC.

Le comité d'administration de JISC est composé de 11 membres (représentants d'universités, de *college*, du JISC ou de fournisseurs de services numériques).

## 1.6 Allemagne - Research Data Repository (RADAR)

Une infrastructure distribuée visant à promouvoir la découverte et la réutilisation de données de longue traîne en renforçant leur traçabilité, leur reproductibilité et leur transparence.

Mis en service en 2017, l'entrepôt RADAR est le fruit d'un projet financé par l'agence de financement nationale allemande Deutsche Forschungsgemeinschaft (DFG) mené de 2013 à 2016, il s'est développé sur le thème "De la communauté - pour la communauté".

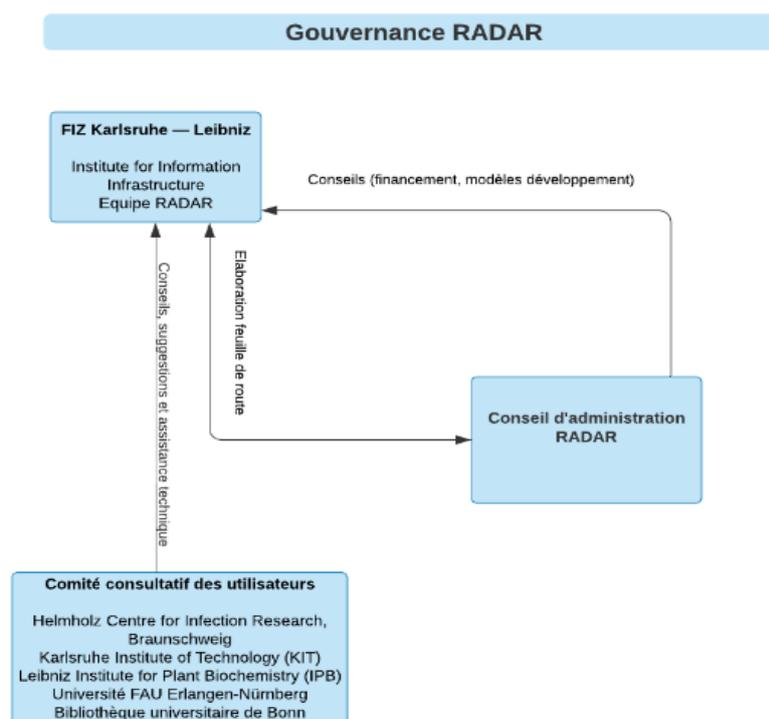
Service interdisciplinaire dédié à l'archivage et la publication des données de recherche issues de projets de recherche achevés, RADAR se concentre sur les données dites de "longue traîne", relevant de disciplines qui ne disposent pas de leur propre infrastructure de gestion des données de recherche. L'infrastructure technique est fournie à la fois par [Karlsruhe Institute of Technology \(KIT\)](#) via son Centre de calcul de Steinbuch (SCC) et l'Université technique de Dresde avec son Centre de services d'information et de calcul haute performance (ZIH).

Opérateur de l'infrastructure, le Fachinformationszentrum Karlsruhe (FIZ-Karlsruhe) agit également en tant que partenaire contractuel pour les institutions qui utilisent RADAR comme entrepôt de données. En collaboration avec l'équipe RADAR, le conseil d'administration établit chaque année la feuille de route RADAR et définit les priorités de développement. Il donne également des conseils sur les opportunités de financement et apporte son soutien au FIZ-Karlsruhe pour développer le modèle opérationnel RADAR.

Un comité consultatif des utilisateurs composé de représentants de cinq institutions ([Helmholtz Centre for Infection Research](#), Braunschweig ; [Karlsruhe Institute of Technology \(KIT\)](#) ; [The Leibniz Institute for Plant Biochemistry \(IPB\)](#), Halle/Saale ; [Universität FAU Erlangen-Nürnberg](#) ; [Universitäts und Landesbibliothek Bonn](#)) délivre des recommandations au FIZ Karlsruhe qui est chargé du développement du service RADAR : conseils techniques (feuille de route du produit et hiérarchisation des nouvelles fonctionnalités) comme organisationnels (contrats et modèles de facturation).

Par ailleurs, RADAR participe à deux consortiums dans le cadre de la Nationale Forschungsdateninfrastruktur (NFDI) infrastructure nationale allemande des données de la recherche : le Consortium pour les données de recherche sur le patrimoine culturel matériel et immatériel (NFDI4Culture) et le Consortium de chimie (NFDI4Chem). RADAR agit au sein de ces consortiums en tant qu'infrastructure générique ayant vocation à accueillir des données n'ayant pas trouvé de solutions adaptées.

**Tableau 5. Schéma de gouvernance de RADAR (Allemagne)**



## 1.7 Grande-Bretagne - United Kingdom Data Service (UKDS)

Service national des données de sciences sociales distribué entre une université hôte (université d'Essex) et des institutions partenaires.

Infrastructure de recherche du Royaume-Uni dédiée à la conservation et la fourniture d'accès aux données de sciences sociales, UKDS est née d'un contrat entre l'Economic and Social Research Council (ESRC, agence de financement britannique) et l'université d'Essex. Créé en 2012, UKDS a intégré United Kingdom Data Archive (UKDA) qui représente le plus grand réservoir britannique de données en science humaines et sociales, de manière à remplir l'objectif suivant : "Explore the UK's largest collection of social, economic and population data resources."

Organisé en tant que service distribué, UKDS fonctionne grâce à de multiples partenaires (UKDA, Cathie Marsh Institute for Social Research - Université de Manchester (CMIST), University College London, Université d'Edimbourg (service spécifique pour les produits de cartographie numérique)) avec lesquels l'Université d'Essex conclut des contrats de sous-traitance.

La responsabilité globale de l'ensemble des activités est assurée par le groupe de direction stratégique, composé de représentants des institutions partenaires.

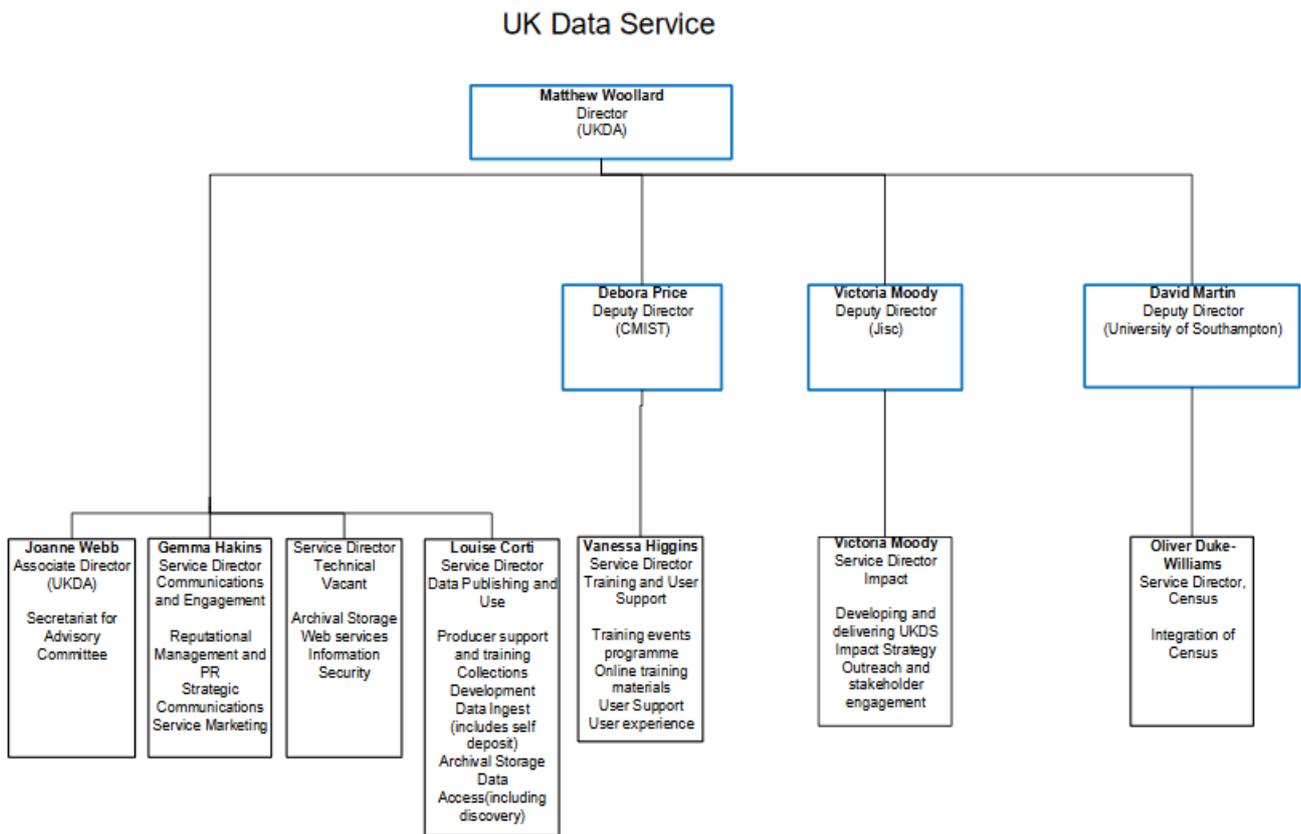
Un comité consultatif regroupant l'ensemble des parties prenantes assure un mécanisme de surveillance, à destination de l'ESRC.

Les responsabilités opérationnelles sont définies dans le cadre du comité réunissant les directeurs de service.

De son côté, l'UKDA fonctionne comme un département de l'université d'Essex. Son activité est supervisée au niveau de l'université (rapports financiers, opérationnels et de notoriété), le chef de département en assure la supervision et la gestion (responsabilités spécifiques en termes de finances, de personnels et de stratégie).

Commun à UKDA et UKDS, le directeur assure la responsabilité générale de la stratégie de service et des relations avec les principales parties prenantes et la direction des activités de conservation, d'archivage et de préservation des données.

**Tableau 6. Organigramme de l'UKDS (Grande-Bretagne)**



## 2 Offre de services

De manière à mettre en valeur les services proposés, il est apparu pertinent de structurer cette section autour des trois catégories d'utilisateurs potentiels, à savoir :

- les producteurs de données
- les utilisateurs de données
- les réseaux d'expertise et d'assistance

### 2.1 Producteurs de données

#### 2.1.1 Dépôt de données

A l'exception de l'ARDC, tous les services nationaux proposent aux chercheurs un entrepôt où déposer et décrire leurs données de manière à les publier et faciliter leur partage. Tous ces dispositifs proposent l'attribution de DOI, préalable à la publication des jeux de données.

De manière à faciliter et simplifier les opérations liées aux dépôts pour les chercheurs, l'entrepôt du JORH a aussi vocation à accueillir les publications et les thèses. L'objectif est de proposer un lieu de dépôt unique sur lequel les chercheurs titulaires de financement puissent s'appuyer pour remplir leurs obligations contractuelles.

Les durées de conservation des données proposées sont variables : le réseau DANS via DataverseNL comme DNO et le DFDR proposent de conserver les données au moins 10 ans. Le service allemand RADAR propose une durée de conservation illimitée des données associées aux publications (au moins 25 ans). La durée de conservation des données peut aussi varier en fonction de l'anonymisation ou non des données comme c'est le cas avec [ReShare](#), l'entrepôt de données d'UKDS.

#### 2.1.2 Archivage pérenne des données

A l'exception de l'ARDC qui considère que la mission d'archivage pérenne relève de la responsabilité de chaque institution, l'ensemble des infrastructures nationales intègre à leurs services cette fonctionnalité.

Afin d'offrir des garanties de sécurité, la plupart des différents entrepôts se sont engagés ou s'engagent dans des démarches de certification : DANS ([CoreTrustSeal](#), Data seal of approval, Nestor Seal [DIN 31644](#)), DNO (Core Trust Seal), UKDA (Core Trust Seal, [ISO 27001](#)), Radar (démarche Core Trust Seal en cours), JORH (ISO27001).

Chargé de cette mission d'archivage et d'accès pérenne au sein de l'UKDS, UKDA est notamment le dépositaire numérique de confiance de la plupart des grandes séries d'enquêtes nationales du Royaume-Uni.

Via le système [Easy](#), DANS propose aussi l'archivage de logiciels.

Seul le JISC Open Research Hub propose à ses utilisateurs de choisir parmi plusieurs solutions techniques d'archivage (Arkivum, Amazon Glacier, DataVault, Preservica, Archivematica) en se positionnant comme intermédiaire.

Il est intéressant de noter que RADAR conditionne la possibilité pour les producteurs d'archiver leurs données de façon pérenne, à la publication de celles-ci.

### 2.1.3 Stockage des données

Seuls l'ARDC et RADAR proposent aux chercheurs des solutions de sauvegarde sécurisée.

Le service Nectar de l'ARDC offre différents types de services de stockage couplés à des services de calcul. Comme l'ensemble des fonctionnalités proposées par l'ARDC, Nectar Cloud est accessible gratuitement à l'ensemble des chercheurs australiens et néo-zélandais.

De son côté, RADAR propose un service de stockage dit temporaire, le temps de documenter les données avant de les publier, ou bien de les archiver. La durée de sauvegarde ne peut excéder 6 mois.

### 2.1.4 Aide à la gestion des données

Tous les services nationaux accordent une place déterminante à l'accompagnement des déposants en proposant des outils d'aide et des services d'assistance.

Ces services visent à apporter des réponses à l'ensemble des questions en rapport avec la gestion des données : rédaction des plans de gestion de données, coût de stockage, choix des solutions de stockage/archivage, conformité avec les exigences contractuelles demandées par les agences de financement, gestion des données sensibles, attribution d'identifiants, ...

A titre d'exemple, le réseau DANS propose à ses utilisateurs divers formats de médiation (formations, présentations, webinaires, tutoriels, workshops) qui peuvent aussi prendre la forme de consultations personnalisées.

L'offre d'accompagnement développée par le JISC Open Research Hub prend notamment la forme d'une boîte à outils ([RDM Toolkit](#)) qui donne accès à des informations ciblées en fonction de la catégorie d'utilisateurs (chercheurs, services d'appui à la recherche, spécialistes des technologies de l'information); ce service renvoie aussi vers les pages des institutions partenaires dédiées à la gestion des données de manière à proposer une information contextualisée aux chercheurs.

L'ARDC propose aussi une gamme très large d'outils d'assistance<sup>12</sup> (plus de 80 supports), l'utilisateur peut affiner sa recherche via des filtres, notamment le domaine de recherche ou l'audience (chercheur, informaticien, bibliothécaire,...).

Du côté de DFDR Canada, les chercheurs peuvent s'appuyer sur Assistant PGD<sup>13</sup>, un outil d'aide à la rédaction de leur plan de gestion de données.

**L'un des objectifs communs à l'ensemble des services étudiés est de faire en sorte que les communautés d'utilisateurs intègrent les principes FAIR (Findable, Accessible, Interoperable, Reusable).** Pour accompagner les déposants, l'ARDC propose l'outil FAIR self assessment tool<sup>14</sup> qui permet d'évaluer rapidement le degré de "FAIRness" d'un jeu de données et d'identifier comment l'améliorer. De son côté, le JORH envisage d'ajouter sur le formulaire de dépôt des informations contextuelles ainsi qu'une **métrique FAIR** visant à inciter les déposants à adopter ces pratiques ouvertes.

Dans le cadre du projet européen Fostering Fair Data in Practices in Europe (FAIRsFAIR), le réseau DANS a collaboré avec le Digital Curation Center (DCC) et Middlesex University à la production de l'outil [FAIR-Aware](#) (voir annexe 4) qui vise à sensibiliser les chercheurs et gestionnaires de données à l'importance des principes FAIR.

A l'image du [help desk de UK Data Service](#), chaque infrastructure propose un service support à l'adresse des chercheurs. Le guichet d'assistance [Nectar Cloud support](#) (ARDC) propose à la fois une

<sup>12</sup> <https://ardc.edu.au/resources/>

<sup>13</sup> <https://assistant.portagenetwork.ca/?locale=fr>

<sup>14</sup> <https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/>

aide immédiate pour les questions techniques, mais aussi des suites thématiques de tutoriels, une base de connaissances ainsi qu'un forum de discussion. Ces outils apportent des réponses sur des sujets variés : aide au stockage, à l'utilisation des clouds locaux, à l'utilisation de la suite OpenStack ; expertise et assistance en matière d'intégration ou de flux de travail complexes.

L'ensemble des services examinés garantissent aux déposants que leurs dépôts feront l'objet d'un travail de curation, préalable à leur validation. Ces activités sont prises en charge le plus souvent par les établissements partenaires du service national, à l'exception de l'UK Data Archive dont les équipes assurent ces opérations de traitement.

L'attribution d'identifiant numérique permanent fait systématiquement partie des services proposés. De manière manuelle ou via des API, tous les services proposent l'attribution de DOI, l'entrepôt national allemand RADAR propose de compléter le DOI d'un préfixe spécifique à l'institution d'affiliation du déposant.

De son côté, l'ARDC met à disposition des chercheurs un [arbre de décision](#) (voir annexe 5) qui leur permet d'orienter leur choix vers un DOI ou un identifiant Handle. Il faut noter que l'ARDC propose aussi l'attribution d'identifiants dédiés aux échantillons physiques ([International Geo sample number](#), IGSN) : identifiant réservé à ce jour aux échantillons géologiques et environnementaux. L'objectif de l'infrastructure nationale australienne est d'étendre ce service à d'autres types d'échantillons physiques tels que la végétation, les spécimens archéologiques et biologiques.

Parmi l'offre de services très développée en matière d'identifiants, l'ARDC propose aussi aux institutions d'attribuer un Research Activity Identifier ([RAiD](#)) aux projets et activités de recherche.

Offrir un accès aux statistiques d'usage (vues, téléchargements) fait là encore partie des services incontournables. RADAR se distingue en proposant une visualisation des statistiques sur les 6 derniers mois. De leur côté, le réseau DANS et DataverseNL ont pour projet d'implémenter des "badges alt-métriques" aux métadonnées.

### 2.1.5 Accès réservé aux relecteurs scientifiques

A ce jour, RADAR est la seule infrastructure nationale de données à proposer un service à destination des relecteurs scientifiques.

Une interface dédiée permet de partager les données de recherche associées à une publication avec les relecteurs désignés avant que celles-ci ne soient publiées. Une URL de relecture sécurisée est transmise au(x) relecteur(s), lien qui reste actif le temps du processus d'examen. Elle leur permet de prendre connaissance des données associées ce qui facilite le travail de validation. Pendant le processus d'examen, l'enregistrement des données en question est verrouillé pour la suite du traitement.

Le réseau néerlandais DANS a lui aussi pour objectif d'introduire un système réservant aux relecteurs un accès aux jeux de données, en amont de la publication.

### 2.1.6 Soutien financier dédié à la gestion et à l'ouverture de données

Parmi les divers services proposés, le réseau néerlandais DANS propose notamment de financer des "[Small data projects](#)", dans le but de décrire et de rendre accessible un ou plusieurs jeux de données dits d'importante valeur. D'un montant maximum de 10 000 euros, ces subventions sont principalement destinées aux domaines scientifiques où, jusqu'à présent, peu de données sont archivées et réutilisées. Les projets retenus doivent représenter une valeur stratégique pour DANS. La gestion des données doit répondre à des exigences de qualité (principes FAIR) pour que celles-ci soient stockées et accessibles de façon pérenne. Le ou les jeux de données, ainsi que les métadonnées associées, doivent être déposés sur la plateforme d'archivage de DANS (EASY) sous licence CC-BY.

De son côté, l'ARDC consacre une partie de son budget au financement de laboratoires virtuels qui sont développés par les différentes communautés.

## 2.1.7 Gestion des accès

Toutes les plateformes de service proposent aux utilisateurs de se connecter via un mécanisme de fédération d'identités. Les autres modalités d'accès proposées peuvent être un compte Orcid, un compte Google, un compte spécifique à la plateforme, un courriel ou encore un compte Dataverse.

La connexion via Orcid permet aux chercheurs néerlandais de voir leurs jeux de données reliés à leur fiche chercheur ([Narcis](#)).

## 2.2 Utilisateurs des données

### 2.2.1 Outil de découverte interne versus portail de recherche fédéré

Mis à part le JORH, l'ensemble des infrastructures nationales proposent un service d'accès aux données sous la forme d'un point d'accès unique à des ressources (données de la recherche, publications, projets) produites par diverses institutions.

Proposer un moteur de recherche qui fédère l'ensemble des ressources nationales, a minima en matière de données de la recherche, fait partie des missions du DFDR canadien comme du réseau DANS via Narcis, ou encore de DNO et de l'ARDC via Research Data Australia.

La mise en œuvre de cet objectif peut prendre deux formes différentes :

- un outil de découverte interne : DNO, RADAR
- un moteur de recherche fédéré moissonnant des sources externes : ARDC, DANS, DFDR, UKDS

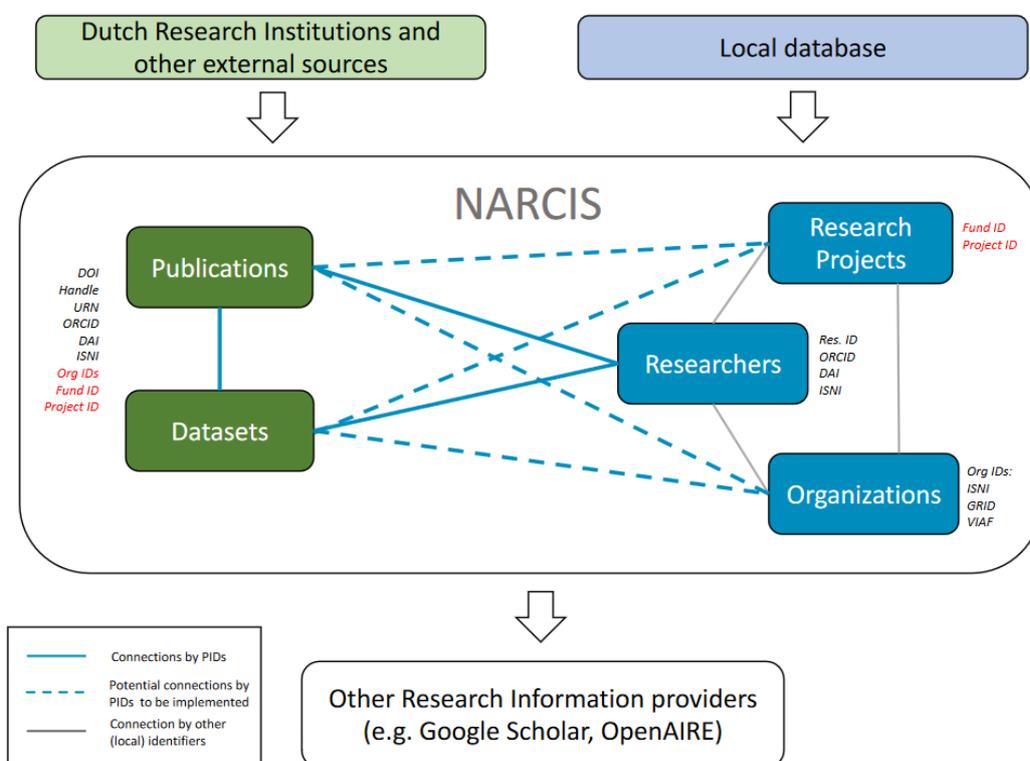
Au-delà des données de recherche déposées dans la plupart des entrepôts de données néerlandais, l'outil [Narcis](#) développé par DANS donne aussi accès à d'autres types d'informations : publications, institutions, chercheurs, projets de recherche. Un des objectifs de Narcis est de lier les chercheurs à l'ensemble de leurs productions de manière à développer les possibilités de **“research in context”** : liens entre données, publications, chercheurs, financement et organisation.

Le service national participe au projet européen **FREYA-PID graph**<sup>15</sup> qui vise à connecter et intégrer les différents systèmes d'identifiants persistants (PID) de manière à lier les données et renforcer leur accessibilité. Comme l'illustre le graphique ci-dessous, les DOI sont utilisés pour relier les publications aux données associées, les liens entre les publications et les auteurs sont facilités par les identifiants ORCID. Les différents identifiants auteurs sont connectés et comparés aux identifiants propres à NARCIS de manière à compléter les informations déjà présentes (ex. liens entre les chercheurs et les publications).

---

<sup>15</sup> <https://www.project-freya.eu/en/about/mission>

**Tableau 7. Narcis PID graph**



16

De son côté, [Research data Australia](#) (ARDC) propose d'étendre la recherche aux logiciels, aux subventions et projets de recherche, ou de rechercher par collections thématiques. A chaque jeu de données trouvé sont associées les publications liées, le ou les contributeurs ainsi que le ou les sites web du projet. L'interface propose des suggestions de jeux de données traitant de sujets similaires, elle renvoie aussi vers d'autres types d'outils et de services comme des simulateurs, des instruments et des logiciels de visualisation ainsi que d'autres outils de rapport, d'analyse et de génération.

Le DFDR canadien a ajouté au mois d'avril 2020 à ses services l'accès à [Geodisy](#), plateforme ouverte de découverte spatiale pour les données de recherche ouvertes canadiennes. Cet outil permet aux utilisateurs de rechercher des données et d'accéder à l'ensemble des métadonnées en fonction de leur emplacement, ce qui permet de prévisualiser les ensembles de données sous forme de superpositions sur une carte numérique.

Développer un service national de données peut aussi donner lieu à la création de collections de jeux de données spécifiques. Le programme Australian National Data Service qui a préfiguré le **lancement de l'ARDC a permis notamment la création d'"High value collections", collections de données dites "stratégiques"**, car liées à des investissements nationaux, et accessibles librement. De son côté, DNO a développé la collection [Trolling DNO](#), entrepôt de données en sciences du langage et en linguistique qui participe à l'infrastructure européenne [Clarin](#) et est indexée par le [Virtual Language Observatory](#).

Il faut noter que les fournisseurs de données ne se limitent pas forcément aux universités et aux établissements de recherche ; c'est le cas de l'ARDC, du DFDR et de l'UKDS qui comptent parmi leurs partenaires des agences gouvernementales, des ministères, des villes, voire des organisations intergouvernementales (UKDS).

Par ailleurs, tous les services décrivent leurs données selon le schéma Datacite, ce qui permet notamment aux utilisateurs d'être informés des conditions d'accès et de réutilisation (licences). Parmi

<sup>16</sup> <https://zenodo.org/record/2552138#.Xs0YY8Dgo2w>

les objectifs de RADAR, l'infrastructure allemande envisage de renforcer l'assistance aux utilisateurs, en matière de possibilité de réutilisation (licences, droits d'auteur et protection des données).

Le tableau « Identifiants, métadonnées et indicateurs » figure parmi les documents annexes (**voir annexe 1**). Il regroupe l'ensemble des solutions techniques retenues par chacun des dispositifs.

## 2.2.2 Services de calcul, d'analyse et de collaboration

Pilier de l'offre de services de l'ARDC, [Nectar research cloud](#) propose une infrastructure informatique, des logiciels et des services pour stocker, accéder et analyser des données, à distance, et de manière autonome.

Nectar propose notamment une offre de plates-formes de calcul informatique (cloud computing) définie par les besoins des communautés de recherche ([Science clouds](#)) pour leur favoriser l'accès aux données, outils, plates-formes et ressources informatiques.

Par ailleurs, Nectar research cloud a financé **15 laboratoires virtuels** relevant de divers champs académiques (ex. analyse génomique, climatologie, études maritimes, sciences de la communication), laboratoires qui sont accessibles à l'ensemble des chercheurs australiens. A titre d'exemple, le laboratoire virtuel HuNI ([Humanities Networked Infrastructure](#)) permet aux chercheurs en arts et humanités de créer, d'enregistrer et de publier des sélections de données, d'analyser et de manipuler les données, de partager leurs découvertes et d'exporter les données pour les réutiliser dans des environnements externes.

L'UKDS propose plusieurs interfaces d'accès et d'analyse des données comme :

- [Nesstar](#) qui permet d'accéder, d'analyser et de télécharger une sélection d'ensembles de données d'enquêtes clés
- [QualiBank](#), interface de recherche et de navigation pour les jeux de données qualitatives (interviews, essais, questions ouvertes et rapports) qui permet d'établir des hyperliens vers des objets connexes (**voir annexe 6**)

## 2.2.3 Services d'accès aux données sensibles

L'accompagnement en matière d'usage des données sensibles figure parmi les priorités de service de l'UKDS. Les équipes de l'UKDS produisent des documents de référence et délivrent des formations spécifiques dont la formation Five safes. Celle-ci s'adresse tout autant aux utilisateurs de données qu'aux déposants, et s'impose comme condition préalable à l'utilisation de laboratoires sécurisés. Ces "[UK Data Service Secure Lab](#)" offrent un **accès sécurisé à des données sensibles** ou confidentielles et permettent aux chercheurs d'analyser les données à distance depuis leur bureau ou en utilisant une salle sécurisée du UKDS.

Seule l'UKDS propose un véritable accompagnement pour les déposants, en les formant à la méthode dite des "[Five safes](#)" (safe data, safe projects, safe people, safe settings, safe outputs).

De son côté, l'ARDC propose une [page web de ressources dédiées](#) ainsi qu'une base de connaissances liée à réutilisation des données sensibles. Les autres services nationaux n'ont pas développé d'offre de service spécifique à la gestion des données sensibles. Le JORH propose quelques éléments d'information et renvoie vers les services de l'UKDS. Il faut aussi noter que les équipes du DFDR envisagent d'élaborer un entrepôt d'accès contrôlé ou restreint pour les jeux de données sensibles.

## 2.3 Réseaux d'expertise et d'accompagnement

Les bénéfices apportés par l'existence de réseaux d'experts, notamment lors de la définition de l'offre de services, représentent l'un des atouts majeurs de ces infrastructures nationales de données.

Les réseaux d'expertise en matière de curation jouent le plus souvent un rôle déterminant dans le développement des services nationaux de données ; c'est le cas notamment de [Research Data Netherlands](#) pour le réseau DANS ou de Portage Canada pour le DFDR.

En effet, les activités de Portage, créé en 2015 par l'ABRC, s'articulent autour de trois missions :

- le développement d'une culture de gestion des données de recherche,
- la structuration d'une communauté de pratique pour la gestion des données de recherche,
- la création de services et d'une infrastructure de gestion des données de recherche à l'échelle nationale.

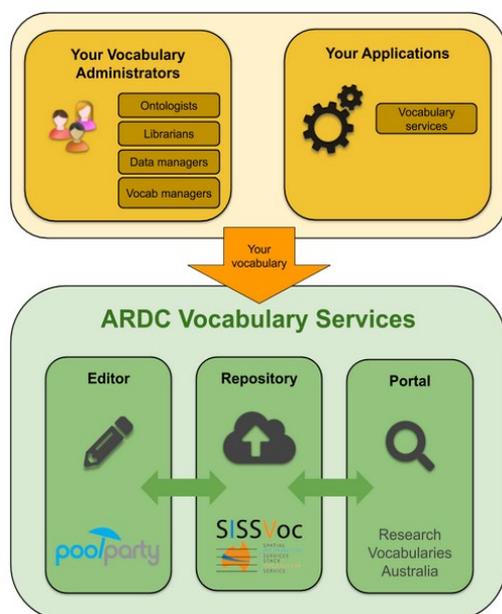
L'existence d'un réseau d'expertise national permet à chaque établissement de s'appuyer sur ce type d'organisation pour former ses référents ; à titre d'exemple, l'Université de Trömso organise des formations et des actions de sensibilisation, notamment à la préservation numérique, à l'intention du réseau de référents Dataverse NO.

Cette mise en commun des ressources favorise aussi les économies d'échelle, objectif partagé par l'ensemble des infrastructures étudiées.

Celle-ci permet aussi le développement d'outils comme le [Research Vocabularies Australia](#), **plateforme collaborative de vocabulaire contrôlés qui réunit plus de 150 organisations partenaires**. Ouverte à la contribution des gestionnaires de données-curateurs comme des utilisateurs, la plateforme donne accès aujourd'hui à près de 250 lexiques téléchargeables relevant de disciplines très diverses.

**Tableau 8. Fonctionnement du Research Vocabularies Australia<sup>17</sup>**

### Publish a vocabulary



<sup>17</sup> <https://vocabs.ardc.edu.au/vocabs/page/contribute>

Enfin, il est important de constater que ces réseaux jouent un rôle déterminant dans l'évolution des services nationaux. Les experts siègent le plus souvent au sein des comités consultatifs et participent, par là même, à la gouvernance du service par leurs retours d'expérience et les conseils qu'ils délivrent à l'intention des comités d'administration.

## 3 Organisation et mutualisation des services

### 3.1 ARDC : faciliter l'émergence de communautés de pratiques et le développement d'outils partagés

Investi d'une mission de coordination nationale, les activités de l'ARDC visent notamment à créer des communautés de pratique et encadrer le développement d'outils partagés. De ce fait, elle initie des projets de collaboration entre acteurs de la recherche australienne et développe de nombreux services dans le cadre de partenariats.

L'un des objectifs majeurs du "nuage australien" est de rafraîchir l'infrastructure de stockage et de calcul, Nectar cloud. Pour y parvenir, l'ARDC a proposé un appel à participation puis a réparti son investissement global (4,1M \$) entre les établissements retenus : l'université de Melbourne, l'université Monash, le Tasmanian Partnership for Advanced Computing (université de Tasmanie) et [Intersect Australia](#).

Depuis 2016, dans le cadre du [Nectar National Server Program](#), le service d'hébergement des serveurs locaux et le service support sont partagés entre les opérateurs suivants : université de Melbourne, université de Monash, National Computational Infrastructure (NCI), Queensland Cyber Infrastructure Foundation (QCIF), et Université de Tasmanie (TPAC). De la même façon, chacun des 15 laboratoires virtuels est géré et hébergé par l'établissement porteur du projet.

La création du Research Activity Identifier (RAID) résulte d'une collaboration avec les fournisseurs de services nationaux suivants : le réseau national d'éducation et de recherche australien ([AARNet](#)) et l'Australian Access Federation ([AAF](#)) pour les accès et authentifications. Identifiant de type handle, le RAID est attribué aux projets de recherche. Il permet notamment de renforcer la visibilité de l'ensemble des parties prenantes du projet et de refléter les pratiques de coopération entre les institutions. Sous cet identifiant sont rassemblées l'ensemble des informations liées au projet (financement, utilisation des infrastructures, résultats,...) ce qui permet d'améliorer leur traçabilité.

D'autres projets menés par les communautés scientifiques, et auxquels contribue l'ARDC, peuvent s'appuyer sur le soutien d'infrastructures nationales de calcul comme le NCI et Pawsey Supercomputing.

Les partenariats portent aussi sur la promotion des pratiques de Research Data Management (RDM), où le rôle de l'ARDC est d'initier le développement de communautés de pratique clés. C'est dans ce cadre que l'ARDC héberge et incite à l'alimentation du [Research Vocabularies Australia](#), **plateforme collaborative de vocabulaires contrôlés alimentée par les contributions de plus de 150 organisations partenaires**. Il en va de même pour le portail Research Data Australia, hébergé et

développé par l'ARDC qui se charge de récolter les métadonnées produites par près d'une centaine d'institutions (universités, organismes de recherche, agences gouvernementales).

Fournisseur national de DOI, l'ARDC propose aussi un service d'attribution d'IGSN (numéro international d'échantillon géographique), qui a été développé par l'Australian earth science research community.

## 3.2 DANS : des services distribués au sein d'un réseau de partenaires

Les équipes de DANS sont responsables de la gestion des plateformes Easy, Narcis et DNL, ainsi que de la fourniture des services idoines. Elles veillent aussi à l'interopérabilité entre ces trois systèmes. Par ailleurs, elles assurent la responsabilité de l'archivage des collections déposées dans Easy. Enfin, elles animent le réseau de partenaires impliqués.

La gestion de l'infrastructure technique de base qui assure notamment le stockage ou les opérations de sauvegarde est partagée entre plusieurs opérateurs : [SURFsara](#) (consortium d'institutions d'enseignement et de recherche néerlandaises), [Target](#) (projet de recherche collaboratif), des centres de calcul universitaires,...

Les services de back-office liés à l'accès, l'archivage, ou encore le service support et l'expertise qui sont proposés aux institutions sont assurés par plusieurs acteurs : le réseau DANS, SURFsara ou encore [4TU.Centre for Research Data](#) qui est un consortium dédié à l'archivage pérenne. Ces trois acteurs ont d'ailleurs conclu un partenariat stratégique [Research Data NL](#) qui porte sur l'archivage. Pour ce qui est de DNL, DANS se charge de la maintenance du serveur et des logiciels, il assure le support administratif ainsi que l'animation du réseau de partenaires. De leur côté, par le biais le plus souvent de leurs bibliothèques, les institutions partenaires se chargent de la gestion de leur collection des activités liées à la curation des données, et des services supports utilisateurs (formation, assistance,...).

## 3.3 Dataverse NO : un entrepôt administré par les services d'une université, des collections gérées par les universités partenaires

Le réseau DNO fédère l'ensemble des collections institutionnelles Dataverse développées par les universités norvégiennes. Le partage des activités entre les partenaires est défini dans le cadre de contrats signés entre l'Artic University of Norway et les universités partenaires.

Les services de stockage et de gestion back-end de DNO relèvent de la responsabilité de l'Artic University of Norway - Trømsø (AUN).

La gestion et le fonctionnement du Dataverse national norvégien sont assurés par une équipe DNO composée des acteurs suivants :

- équipe de la bibliothèque universitaire d'AUN,
- service informatique d'AUN,
- service de la recherche d'AUN

L'équipe DNO se charge d'accompagner les établissements en proposant des formations et des actions de sensibilisation (notamment à la préservation numérique), des guides, des consultations individuelles.

Par ailleurs, elle assure l'ensemble des activités liées à la gestion de la collection nationale [TROLLing](#) (The Trömsö Repository of Language and Linguistics).

Les activités liées à la curation des données et à la gestion de chaque collection relèvent de la responsabilité des établissements partenaires, tout comme les supports techniques.

Les opérations liées à la conservation à long terme sont partagées entre l'équipe DNO et les gestionnaires de collection, la préservation numérique relevant d'une responsabilité communautaire partagée. D'autre part, l'équipe DNO travaille sur ces questions d'archivage pérenne en collaboration avec la bibliothèque nationale de Norvège et le [Centre norvégien pour les données de recherche](#).

### 3.4 DFDR : des services fédérés alliant infrastructures de données, universités et réseau de bibliothécaires

Le projet DFDR s'articule autour de services fédérés qui viennent compléter les services existants chez les différents partenaires.

L'infrastructure nationale de données permet à tout chercheur qui le souhaite de déposer ses données, ce service s'adresse notamment à ceux qui relèvent d'une institution ne disposant pas d'entrepôt. Les données soumises au DFDR seront hébergées dans l'infrastructure gérée par Calcul Canada à l'Université de Victoria (Colombie-Britannique), ou à l'Université de Waterloo (Ontario). Ces entités assurent en particulier la fourniture d'un environnement de stockage et l'assistance technique.

Les données déposées puis stockées dans les entrepôts des différentes institutions sont connectés au DFDR.

L'outil de recherche fédéré donne ainsi accès aux collections des institutions canadiennes auxquelles seront ajoutées les données déposées directement dans le service national. La plupart des métadonnées qui alimentent l'outil de découverte proviennent de sources extérieures (universités, organismes de recherche, services administratifs, collectivités territoriales). Celles-ci sont partagées via le [service Globus](#) pour être collectées, indexées et regroupées par les équipes du DFDR. **Une grande partie des informations provient des portails Dataverse développées par la plupart des universités canadiennes dans le cadre de l'initiative [Scholars Portal Dataverse](#).**

Les activités de curation des ensembles de données soumis directement au DFDR sont effectuées par des bibliothécaires du réseau Portage (ABRC) ou par des gestionnaires relevant des services de données partenaires, en collaboration avec les chercheurs et les déposants. De leur côté, les établissements partenaires mènent ces activités de "*gestion active*" des données avec leurs propres équipes de curation et leurs propres flux de travail mais peuvent s'appuyer sur l'expertise du réseau Portage. Les services de téléchargement et de dépôt des fichiers volumineux seront pris en charge par Calcul Canada et le service [Globus connect](#).

Un projet d'archivage fédéré a pour but de confier à des fournisseurs de services de conservation, vraisemblablement des universités canadiennes, l'archivage à long terme de certains ensembles de données.

Le développement des logiciels utiles au fonctionnement des services du DFDR est assuré par l'équipe de développeurs de l'université de Saskatchewan. Le contrôle d'accès au service est assuré par les équipes du DFDR.

## 3.5 JORH : des prestations logicielles intégrées aux environnements techniques des institutions clientes

Conçu comme un nuage de services à disposition des institutions, le JORH est administré par le JISC qui héberge et gère l'infrastructure centrale.

Ouverts à l'ensemble des établissements membres du JISC, le JORH propose sous la forme de **software as a service (SaaS)**, quatre offres de service :

- entrepôt,
- archivage,
- entrepôt et archivage,
- interconnexion des systèmes de recherche (entrepôt, solution d'archivage, outils de gestion administrative et financière)

Les services proposés ont vocation à être intégrés aux environnements techniques propres à chaque institution.

Les institutions partenaires se positionnent comme des clients-abonnés ayant souscrit à une offre de service.

Le consortium propose un service d'accompagnement, Research Repositories Dynamic Purchasing System (DPS)<sup>18</sup>, qui a vocation à orienter les institutions vers l'entrepôt de stockage adapté à leurs besoins. La DPS définit les normes minimales que les fournisseurs doivent respecter pour que leur produit soit éligible. Les membres du Jisc peuvent utiliser le service pour organiser des mini-concours avec les fournisseurs, en utilisant les modèles standard fournis et en ajoutant si nécessaire leurs propres exigences. Jisc administre le processus au nom du membre.

Les premiers entrepôts intégrés au JORH sont DSpace et DSpace-CRIS (4science), Figshare (Digital Science), Mendeley Data (Elsevier) et Polaris (Mysciencework).

Le service d'archivage s'appuie sur les solutions [Archivematica](#) et [Preservica](#) qui peuvent être configurées pour utiliser un stockage local ou un stockage en nuage, voire une combinaison des deux.

## 3.6 RADAR : des services distribués entre établissements partenaires

Les services de stockage sont partagés entre le centre de calcul du FIZ Karlsruhe, Steinbuch Center for Computing (SCC), et le Center for Information Services and High Performance Calculation (ZIH) de l'Université technique de Dresde.

L'université Ludwig-Maximilians-Universität Munich, et l'Institut Leibniz de biochimie végétale (IPB) ont fourni les spécifications scientifiques utiles pour les services de gestion des données.

La gestion et l'attribution des DOI est confiée à la bibliothèque d'information technique (TIB) d'Hanovre.

Le service de données RADAR est basé sur le modèle contractuel suivant :

- les établissements partenaires s'engagent à assurer la curation des contenus
- le FIZ Karlsruhe assure le développement et le fonctionnement de l'infrastructure technique RADAR (stockage/archivage, publication, indicateurs)

---

<sup>18</sup> <https://scholarlycommunications.jiscinvolve.org/wp/2020/05/04/jisc-launches-research-repositories-dynamic-purchasing-system/>

### 3.7 UKDS : un service distribué sous la forme de contrats de sous-traitance entre l'université d'Essex et les institutions partenaires

L'opérateur du service UKDS est l'université d'Essex.

UKDS est un service distribué entre plusieurs partenaires dont UKDA, le CMIST Institute for Social Research (Université de Manchester), l'University College London et l'université d'Edimbourg (service spécifique pour les produits de cartographie numérique).

Chaque établissement partenaire agit en tant que sous-traitants de l'Université d'Essex :

- Membre de la [Digital Preservation Coalition](#) (Communauté UK), UKDA se charge des activités liées à l'archivage. UKDA joue aussi le rôle de lieu de dépôt pour les Archives nationales du Royaume-Uni. Ses équipes assurent la curation de l'ensemble des collections déposées et apportent leur expertise dans les domaines de l'acquisition, de la conservation et de l'accès aux données.
- Les services de l'[Edinburgh University Data Library](#) (EDINA) apportent leur assistance aux utilisateurs de l'UKDS en proposant d'accéder et d'utiliser les résultats géographiques des recensements.
- Spécialisé dans l'application de méthodes quantitatives avancées dans un contexte interdisciplinaire de sciences sociales, le CMIST de l'université de Manchester contribue au développement des microdonnées secondaires du Royaume-Uni, y compris l'échantillon de dossiers anonymisés du [Census Microdata Unit](#).

Les équipes d'UKDS assurent l'hébergement et le développement du service de stockage via l'entrepôt de données [ReShare](#). Par ailleurs, elles ont pour rôle de développer des services de soutien et de formation à l'intention des utilisateurs de données, et d'élaborer des normes de bonnes pratiques en matière de curation et de partage des données.

L'infrastructure UKDS utilise la suite [Apache Hadoop](#).

## 4 Modèle économique

Les différents dispositifs nationaux reposent sur une ou plusieurs sources de financement, parmi lesquelles on retrouve le plus souvent un financement de type structurel (contrats signés avec le ministère de la recherche ou un bailleur de fonds public). Il faut noter que le modèle économique du DFDR est encore en cours de définition.

### 4.1 ARDC : une plateforme de services entièrement financée par le gouvernement australien

Le financement du gouvernement australien (via le programme NCRIS) est la seule source de financement de l'ARDC. Ce financement s'inscrit dans un cadre quinquennal fixé par le [NCRIS \(National collaborative research infrastructure strategy\)](#).

Ainsi, l'ARDC bénéficie d'un budget de 182 millions de dollars pour la période 2019-2023 réparti de la façon suivante :

- budget de fonctionnement = 110 millions de dollars australiens (67M d'euros);
- budget d'investissement = 72 millions de dollars australiens (44M d'euros) dédiés à des projets de gestion de données, d'application de données, de plate-forme de données, d'actifs de données, laboratoires virtuels,...

Le financement structurel dont dispose l'ARDC lui permet de proposer ses services gratuitement aux universités et institutions partenaires.

### 4.2 DANS : un modèle alliant financement structurel, soutien aux projets et contribution des établissements partenaires

Le réseau néerlandais est soutenu de façon structurelle par le financement annuel de KNAW (2,7 millions d'euros) et de NWO (700000 euros), soutien récurrent dont le montant est resté stable depuis 2012.

Le soutien financier aux développement de projets apporte une contribution importante, de l'ordre de 1,5 millions d'euros ces dernières années.

A ces deux principales sources de financement, il faut ajouter la contribution des établissements membres du Dataverse Nederland dont le rôle est de couvrir les coûts de l'entrepôt, environ 300000 euros par an. Pour chaque institution membre, l'adhésion est de l'ordre de 5000 euros (hors taxe) à laquelle vient s'ajouter la facturation des frais de stockage.

DANS précise que les activités de formation des personnels à la gestion des données restent à la charge des établissements.

Si les coûts spécifiques à chaque offre de service sont difficiles à déterminer, les frais liés au serveur du portail fédéré Narcis sont estimés à 500000 euros par an.

## 4.3 Dataverse NO : un service financé par l'université hôte et la contribution des établissements partenaires

Dataverse Norvège fonctionne grâce au financement de l'université de Trømsø et aux cotisations des institutions partenaires.

Ces cotisations sont composées de frais généraux fixes (gestion, exploitation et développement de l'entrepôt) et d'une tarification en fonction du volume de stockage.

Dataverse NO précise que chaque partenaire doit s'engager à garantir un financement et des ressources suffisantes ainsi qu'un personnel suffisamment qualifié pour gérer leurs collections. Cet engagement contractuel comprend, pour les partenaires, la prise en charge des frais liés à l'implémentation de leur archive institutionnelle, ainsi que des dépenses liées au renforcement des compétences et à la participation aux réunions.

## 4.4 DFDR : un modèle en cours de définition

Le DFDR s'appuie sur un financement structurel qui provient du gouvernement fédéral du Canada. Une partie du financement attribué à Réseau Portage par CANARIE, agence de financement national, sera réservée à DFDR.

Depuis le début du projet, de nombreux établissements partenaires apportent leur contribution en mettant à disposition des ressources humaines pour développer le service.

Les modalités de participation financière des établissements partenaires seront liées aux besoins exprimés. Si tous les chercheurs canadiens auront accès gratuitement à un espace limité de stockage dans des entrepôts, un modèle économique doit être mis en place pour les organismes de recherche qui auront de très grands besoins en matière de stockage.

Une combinaison de mécanismes financiers reste donc à définir :

- un financement national complété par des ressources attribuées à l'échelle régionale?
- un soutien financier et/ou des ressources humaines mises à disposition par les établissements partenaires?

## 4.5 JORH : modèle économique basé sur l'abonnement à des offres de service

La majeure partie du financement du JISC provient du soutien apporté par des agences de financement du Royaume-Uni (UKRI, Office for students, Scottish Funding Council, Higher Education Funding Council for Wales), par le ministère de l'éducation britannique, le ministère de l'économie d'Irlande du Nord et le gouvernement gallois. Les frais d'adhésion des membres représentent la seconde source de financement.

Le modèle économique du JORH repose largement sur les recettes générées par les abonnements aux différentes offres de service.

Organisé sous la forme d'un "logiciel en tant que service" (SaaS), le JORH propose aux institutions de souscrire à trois types d'offres de services qui peuvent comprendre tout ou partie des services proposés : dépôt, archivage, dépôt/archivage.

Les frais d'abonnement sont calculés en fonction de la taille de l'établissement : ils peuvent aller de moins de 10000 livres sterling pour le dépôt à plus de 50000 livres sterling pour des services couplant dépôt et archivage.

Des frais supplémentaires sont facturés aux universités clientes en cas de :

- besoin de stockage supérieur au premier Téraoctet(To) compris dans l'offre d'abonnement : 300 livres par To
- besoin d'intégration personnalisée nécessitant la production d'un développement

## 4.6 RADAR : financement réparti à parts égales entre l'établissement d'accueil et des institutions partenaires

L'estimation annuelle de fonctionnement du service est de l'ordre de 300000 euros.

Le financement de RADAR repose pour moitié sur le financement de l'Institut de Technologie de Karlsruhe (KIT) qui prend en charge la moitié des coûts fixes opérationnels à partir de son propre financement de base.

Le reste des coûts fixes et tous les coûts d'exploitation variables sont couverts par les frais d'utilisation. Les contrats entre chaque institution partenaire et KIT prévoient une durée minimale de trois ans.

L'adhésion annuelle va de 595 à 11900 euros si le volume de stockage demandé excède les 50 To.

La redevance annuelle varie en fonction de l'utilisation (archivage, publication ou stockage temporaire), elle est calculée selon :

- la taille de l'espace de stockage temporaire mis à disposition,
- le volume de toutes les données archivées depuis le début de la durée du contrat, moins les données qui ont été supprimées en raison de périodes de stockage expirées,
- la nature des données : archivées (0,46 euros/Gb), stockées temporairement (0,43) ou publiées (7,56).

## 4.7 UKDS : financement structurel de l'ESRC et des établissements partenaires

UKDS et UKDA bénéficient du financement structurel de l'Economic and Social Research Council (ESRC), défini lors de contrats quinquennaux.

Par ailleurs, les établissements partenaires (université de Manchester, le JISC, l'université d'Edimbourg, l'université de Southampton, l'University College of London et l'université d'Essex) contribuent financièrement au développement et à la pérennité de la plateforme UKDS.

UKDA est autorisé à facturer des frais de recouvrement des coûts pour l'accès à certaines données sécurisées. Cette redevance, qui n'est pas appliquée actuellement, ne concerne pas l'accès aux données, mais le recouvrement des coûts pour les services fournis par l'UKDA aux utilisateurs finaux (par exemple, les logiciels d'analyse des données).

## 5 Synthèse

### 5.1 Degré de maturité des services

Il est intéressant de noter que la **plupart des services nationaux dédiés aux données de la recherche qui font l'objet de cette étude ont été développés récemment.**

En effet, mis à part l'UKDS créé en 2012 et qui s'appuie sur l'UKDA fondé en 1968, ou le réseau néerlandais DANS qui héberge DNL depuis 2014, les différents services étudiés ont été mis en place ces toutes dernières années : JORH (2019), ARDC (2018), DNO (2017), RADAR (2017). En phase de test actuellement, le Dépôt fédéré des données de recherche canadien prévoit de passer en production à l'automne 2020.

De leur côté, les projets pilotes de National Data Services (USA)<sup>19</sup> et du National Institute of Health Data Commons (USA)<sup>20</sup> ont été menés respectivement de 2015 à 2018 et de 2017 à 2018; ils n'ont pas donné lieu jusqu'à maintenant à l'ouverture de plateformes de services.

Si le degré de maturité des services du réseau DANS permet à celui-ci d'afficher la volonté de figurer parmi les principales infrastructures de données européennes, la plupart des autres services commencent seulement à évaluer leurs activités.

**Tableau 9. Degré de maturité des dispositifs**

Dispositif	En production	Phase de test	Précisions
ARDC Australie	Depuis 2017		
DANS Pays-Bas	Depuis 2014 (ouverture de Dataverse NL)		
DFDR Canada		Ouverture prévue au début de l'année 2021	
DNO Norvège	Depuis 2017		
JORH GB	Depuis 2019		Encore en cours de développement  11/2020 : Jisc abandonne la « marque JORH » pour 4 marques propres à chaque service : <ul style="list-style-type: none"><li>- "Research repository"</li><li>- "Research repository plus"</li><li>- "Preservation"</li><li>- "Research system connect"</li></ul>
RADAR Allemagne	Depuis 2017		Réflexion en cours autour de l'évolution du service
UKDS GB	Depuis 2012		

<sup>19</sup> <http://www.nationaldataservice.org/>

<sup>20</sup> <https://commonfund.nih.gov/commons>

## 5.2 Priorités affichées par les dispositifs

L'ensemble des éléments d'information relatifs à cette partie sont extraits des sites web dédiés aux différents services étudiés.

### 5.2.1 Quels objectifs ?

La mission principale de l'ensemble des services nationaux s'articule d'abord autour du triple objectif suivant :

- offrir une solution de dépôt des données,
- accompagner les déposants et les services dédiés des établissements partenaires aux activités liées à la gestion des données : de la planification du projet à la réutilisation de données tierces
- préserver les données de façon pérenne.

Les enjeux liés à l'accès figurent parmi les points d'attention récurrents des infrastructures étudiées. Comme le prône le réseau DANS, l'accès aux informations doit être "Ouvert si possible, protégé si nécessaire", tandis que le DFDR propose de "trouver et partager des données de recherche canadiennes".

De la même façon, UKDS a pour mission de garantir un accès fiable à la plus grande collection de données sociales, économiques et démographiques du Royaume-Uni.

Accès ouvert aux données norvégiennes qui doit favoriser l'augmentation de leur réutilisation, tel est le mandat assigné à Dataverse Norvège par le MESRI norvégien.

L'infrastructure australienne ARDC vise à donner accès à une e-Infrastructure nationale constituée de plateformes, d'outils et de collections de données de haute valeur stratégique, ainsi qu'à un réseau de compétences.

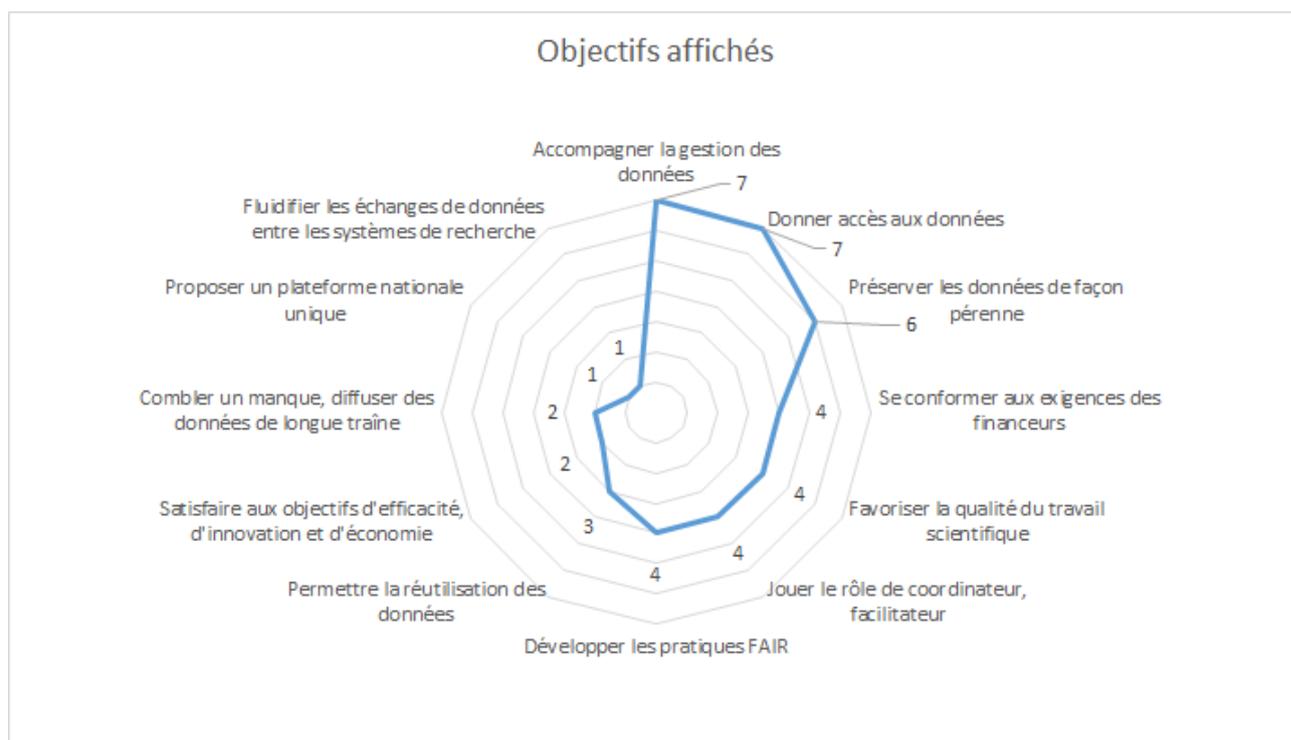
Faciliter le renforcement de la qualité des travaux scientifiques fait partie des axes stratégiques énoncés par DANS, l'ARDC, le DNO, RADAR et le JORH.

Par ailleurs, répondre aux exigences des financeurs figure parmi les missions affichées par le DFDR canadien, l'ARDC, le réseau DANS ainsi que le JORH.

Le rôle d'accompagnement, de facilitateur voire de coordinateur figure parmi les missions d'une majorité d'infrastructures. Ce rôle de soutien va souvent de pair avec l'objectif affiché de développer des pratiques de gestion de données FAIR. Il est notamment au cœur du projet de l'ARDC dont l'objectif est de faciliter le développement d'un écosystème cohérent. De la même façon, le JORH et ses services connexes se positionnent comme intermédiaires dont le but est d'accompagner, de simplifier la prise de décision des institutions (ex. [Research Repositories Dynamic Purchasing System](#)) comme le travail des chercheurs.

**Enfin, en connectant les systèmes de recherche existants (entrepôts, systèmes d'archivage, systèmes de gestion administrative et financière) chez les établissements clients, le service "Research system connect" a pour but de fluidifier les transferts de données et de métadonnées de manière à libérer le temps des personnels pour d'autres tâches.**

**Tableau 10. Représentation sous forme de radar des principaux objectifs affichés par les dispositifs analysés<sup>21</sup>**



## 5.2.2 Pour quels publics ?

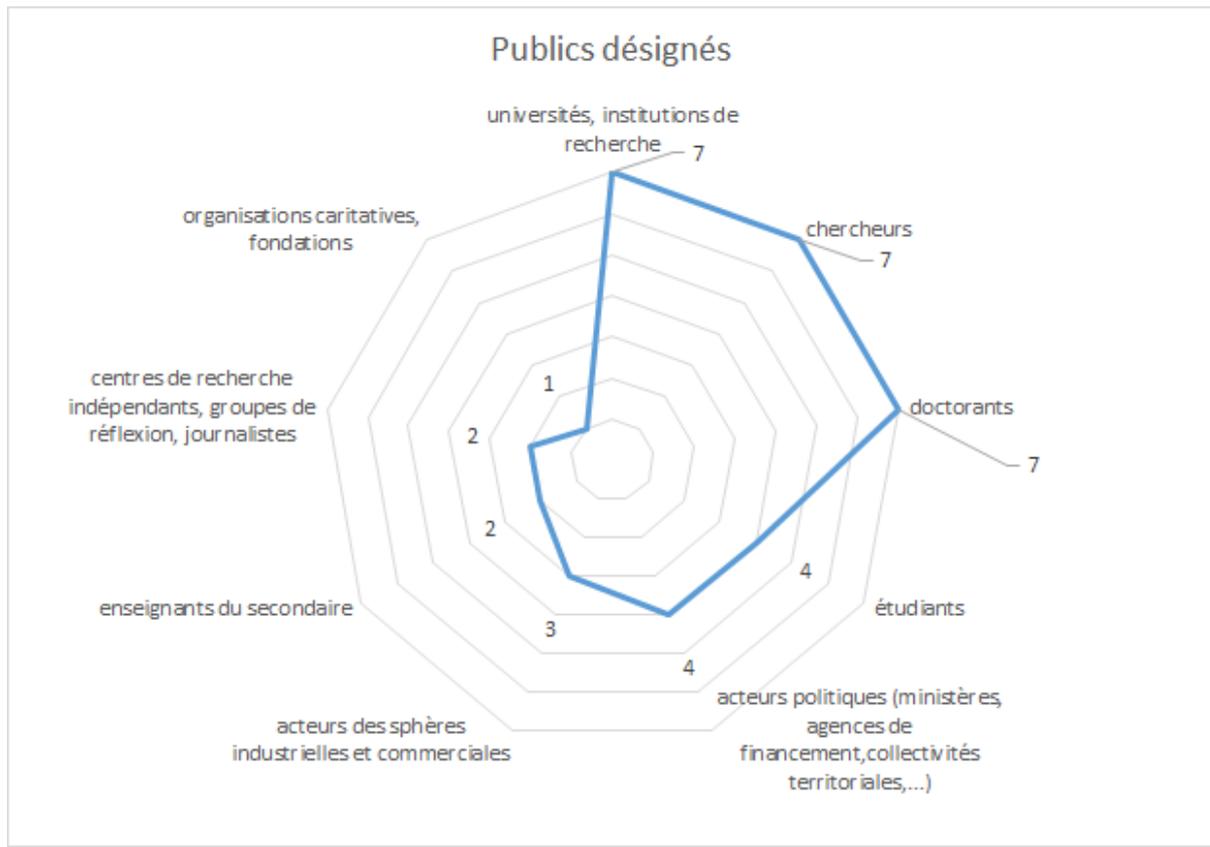
Toutes les infrastructures nationales se sont développées avec pour objectif de proposer leurs services tout d'abord aux chercheurs et aux doctorants, ainsi qu'aux universités et institutions de recherche. Plus de la moitié d'entre elles précisent dans leurs documents stratégiques que leurs activités ont aussi vocation à aider les acteurs politiques (ARDC, DANS, DFDR, JORH, UKDS), comme les étudiants (DANS, DFDR, RADAR, UKDS).

Il est intéressant de noter que peu de services considèrent aussi les acteurs économiques comme des publics cibles, c'est le cas de l'ARDC, l'UKDS ou le réseau DANS.

Il faut enfin noter que l'UKDS vise une gamme de publics très large dont font notamment partie les enseignants, les chercheurs indépendants, les journalistes ou encore des organisations caritatives.

<sup>21</sup> Graphique réalisé à partir des informations collectées sur les sites web dédiés des différents services nationaux

**Tableau 11. Représentation sous forme de radar des catégories de publics visés par les services nationaux étudiés<sup>22</sup>**

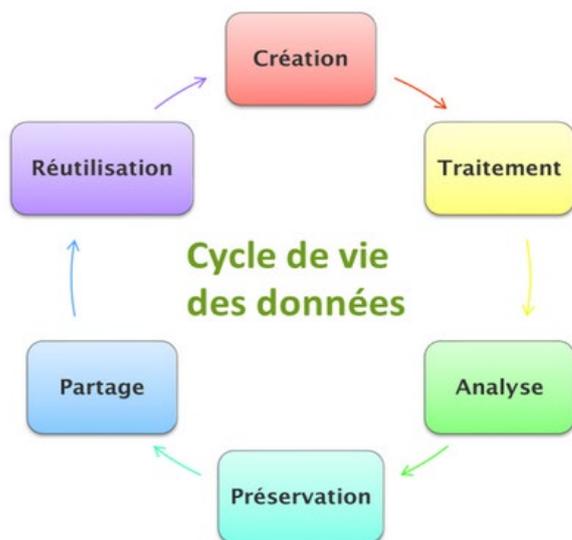


<sup>22</sup> Graphique réalisé à partir des informations collectées sur les sites web dédiés des différents services nationaux

### 5.2.3 Offre de services

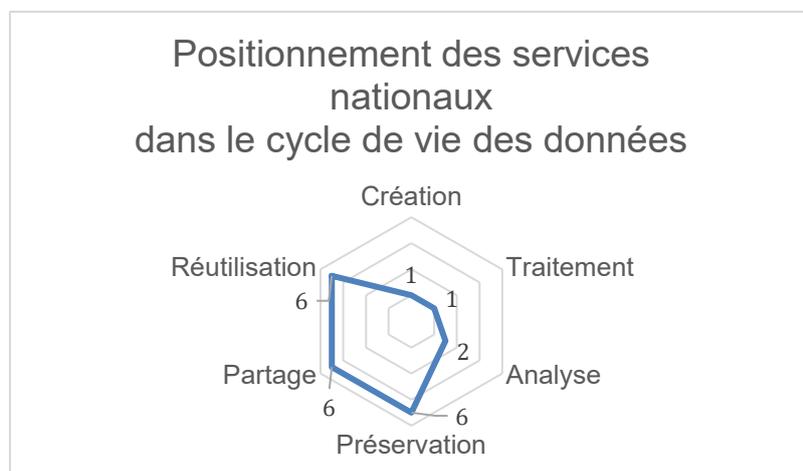
Dans un premier temps, il apparaît utile de positionner l'offre de services proposée par les différents dispositifs nationaux étudiés au sein du cycle de vie des données de la recherche. Pour ce faire, cette étude propose de retenir le découpage proposé le modèle de l'Institut de recherche pour le développement (voir image ci-dessous).

**Tableau 12. Cycle de vie des données, capture réalisée à partir du site IRD Data<sup>23</sup>**



Si l'on positionne l'ensemble des services proposés par les sept dispositifs au sein du cycle de vie des données, il apparaît clairement que ces services nationaux ont fait le choix de se concentrer sur la deuxième moitié du cycle de vie des données à savoir les activités de préservation, partage et réutilisation.

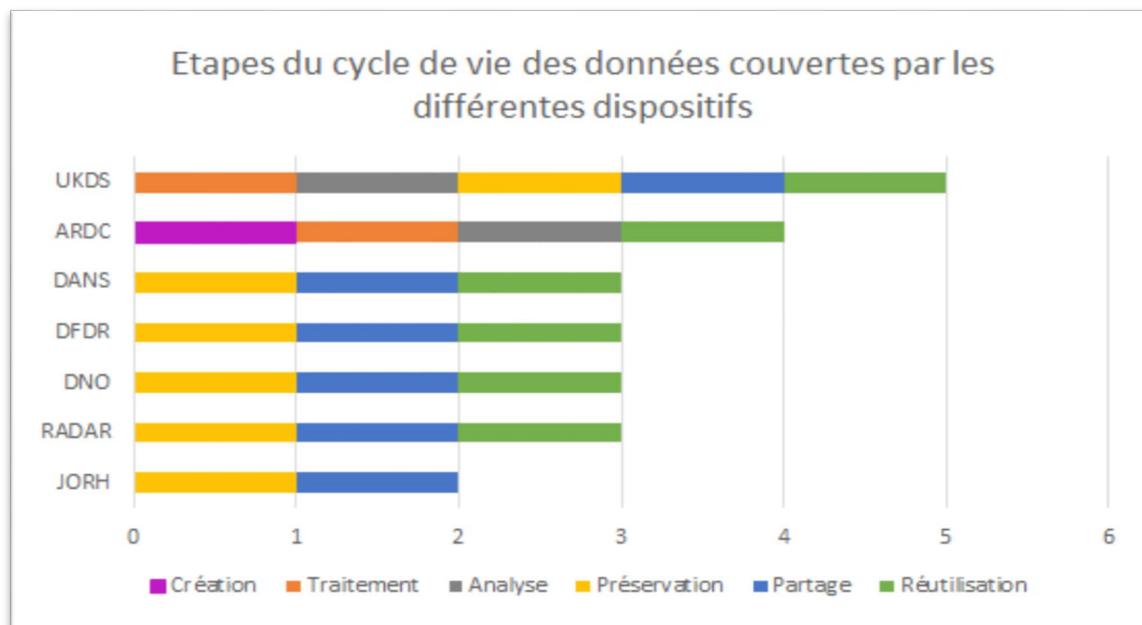
**Tableau 13. Etapes du cycle de vie des données couvertes par les services nationaux de données (cumul des différents services)**



<sup>23</sup> Image extraite de la page suivante : <https://data.ird.fr/plan-de-gestion-de-donnees/>

Aucun des services étudiés ne couvre l'ensemble des activités liées au cycle de vie des données. L'UKDS propose la plus large gamme de services tandis que l'offre de services proposée par l'ARDC revêt un caractère atypique puisqu'elle se concentre sur la production, l'analyse puis sur la réutilisation des données : les services de préservation et de partage ne sont pas couverts.

**Tableau 14. Etapes du cycle de vie des données par dispositifs**



A l'examen des différentes infrastructures de données, on peut constater que toutes proposent aux chercheurs un entrepôt de données, à l'exception de l'ARDC. Services de dépôt de données qui visent à faciliter l'accès et le partage des données, ainsi qu'à satisfaire aux exigences des financeurs.

L'accompagnement, l'assistance à la gestion des données sont elles aussi des activités incontournables des services nationaux de données. Elles s'appuient le plus souvent sur des réseaux d'expertise constitués en amont. Cet accompagnement porte le plus souvent sur l'ensemble des activités liées aux différentes étapes du cycle de vie des données. Les outils et services proposés ne s'adressent pas uniquement aux chercheurs déposants ou utilisateurs de données, mais aussi aux gestionnaires des services support (bibliothèques, services informatiques, administration de la recherche) avec comme objectifs affichés la qualité du service rendu et la montée en compétences de l'ensemble des acteurs concernés.

Seul l'ARDC australien ne propose pas de service d'archivage pérenne, offre de services qui reste du ressort des institutions partenaires australiennes à l'échelle locale.

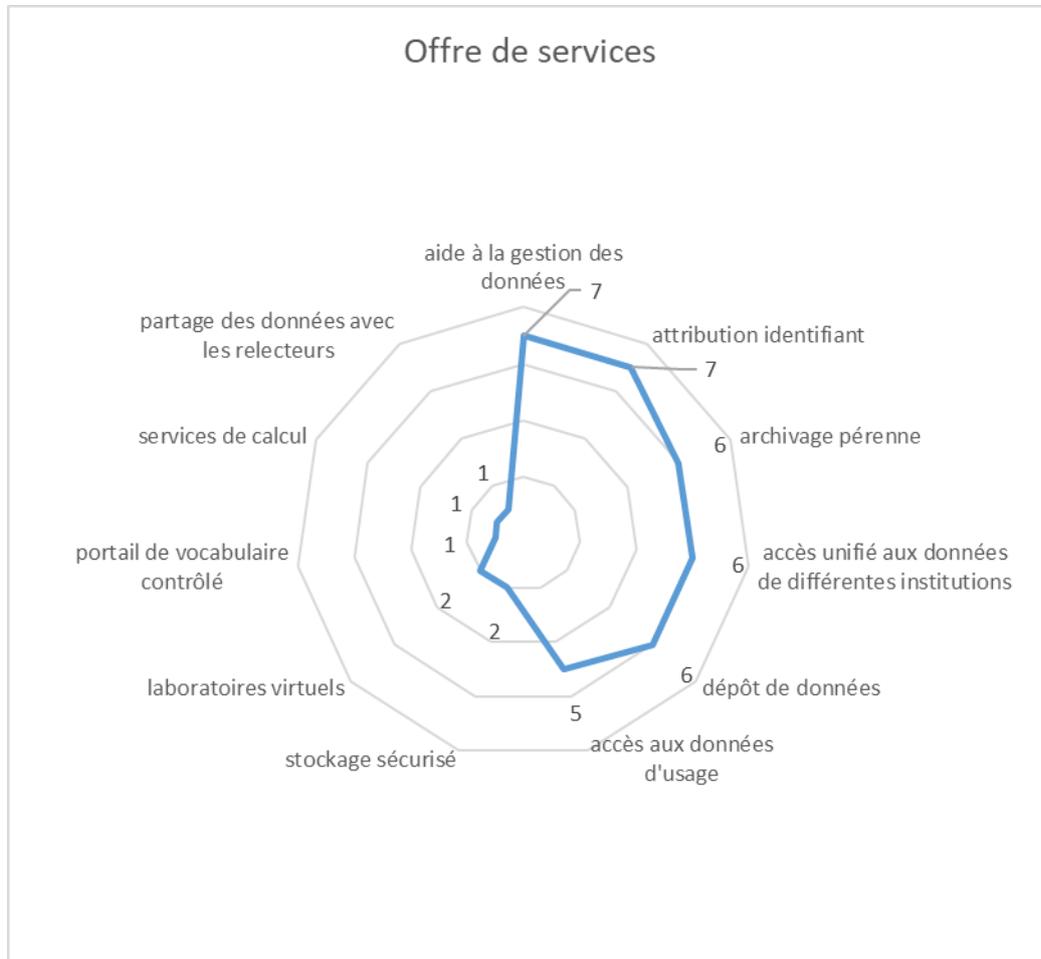
De la même façon, la quasi-totalité des infrastructures nationales ont pour objectif de donner un accès unique à un volume important de données produites à l'échelle nationale : ou en proposant un outil de découverte internalisé, ou en agrégeant des sources externes au sein d'un moteur de recherche fédéré. A l'inverse, le JORH n'a pas choisi de développer d'outils d'accès aux données, mais plutôt de proposer des services aux institutions clientes que celles-ci peuvent intégrer à leurs environnements via des échanges API, selon l'adage "les données visibles, [mais] l'infrastructure JORH invisible".

La plupart des services permettent aux utilisateurs, aux institutions notamment, d'extraire des données d'utilisation des services et de mesurer l'impact de leurs productions. C'est notamment le cas du JORH qui place l'activité de reporting au cœur des services proposés aux établissements partenaires en leur proposant d'interroger et de récupérer des informations via des API. Celles-ci proviennent de la

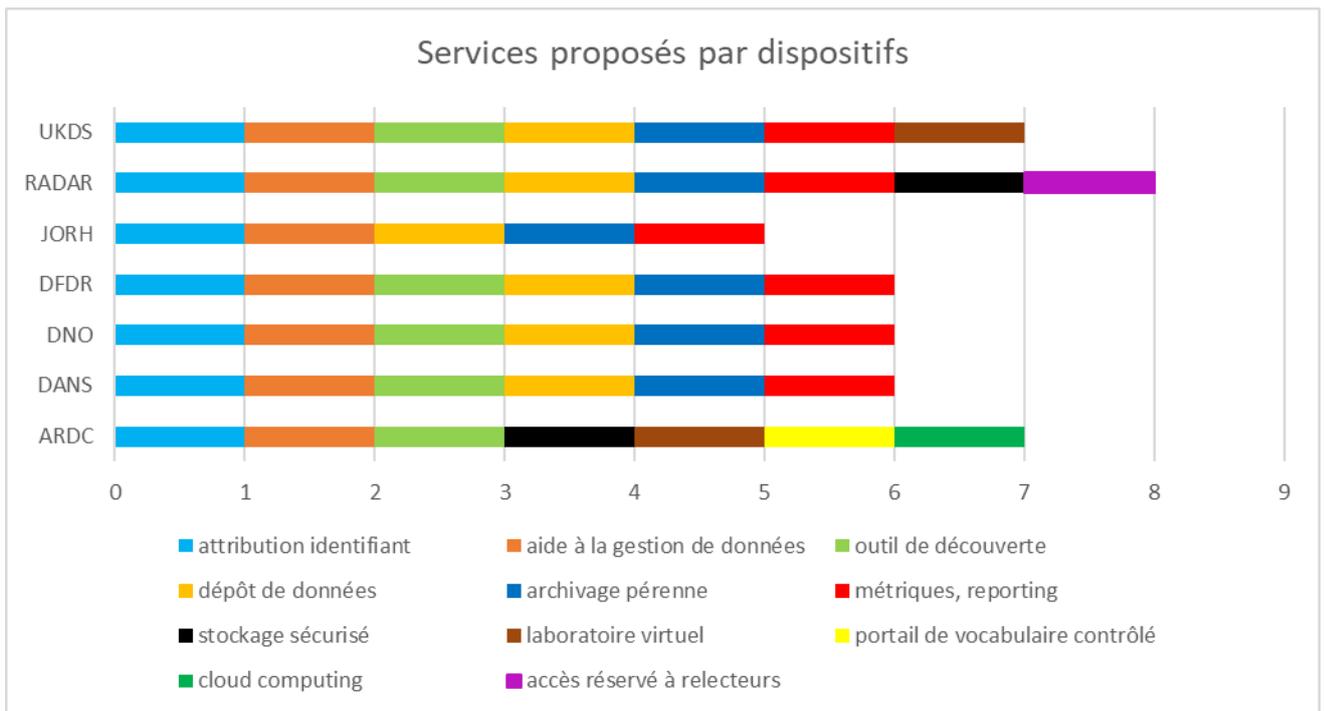
combinaison de métadonnées hébergées dans l'infrastructure de base de JORH et d'autres systèmes ou services intégrés.

Certaines infrastructures vont se distinguer par une offre de services plus spécifique : accès à des laboratoires virtuels (UKDS, ARDC), services de calcul (ARDC), validation scientifique des jeux de données (RADAR), portail de vocabulaires contrôlés (ARDC).

**Tableau 15. Services proposés (cumul des différents services)**

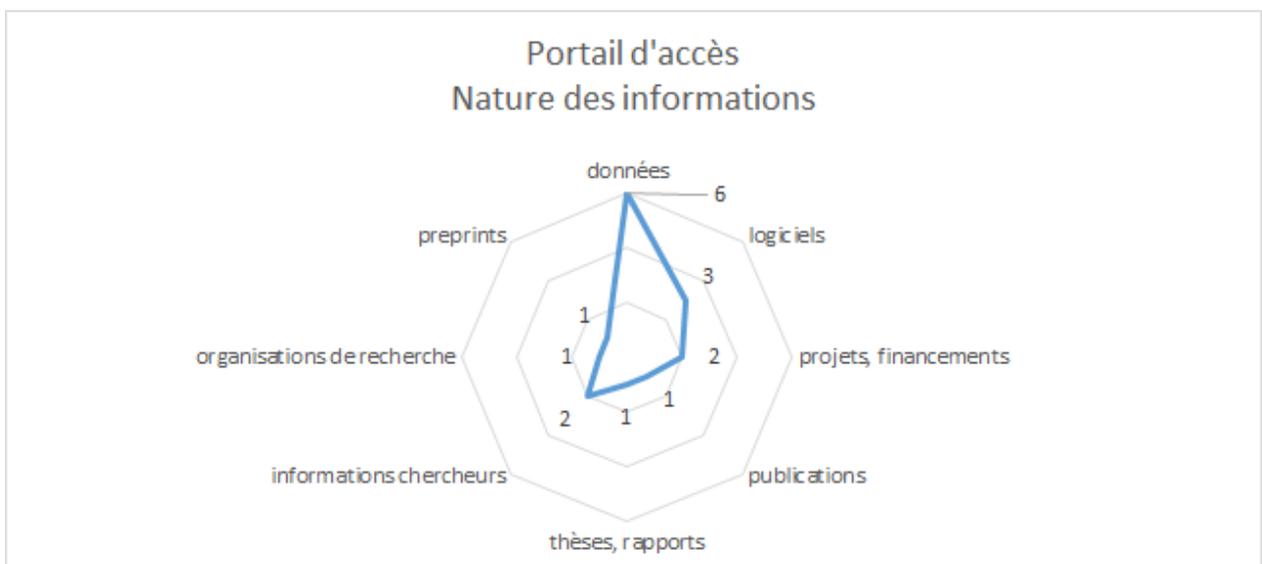


**Tableau 16. Services proposés par chaque service national**



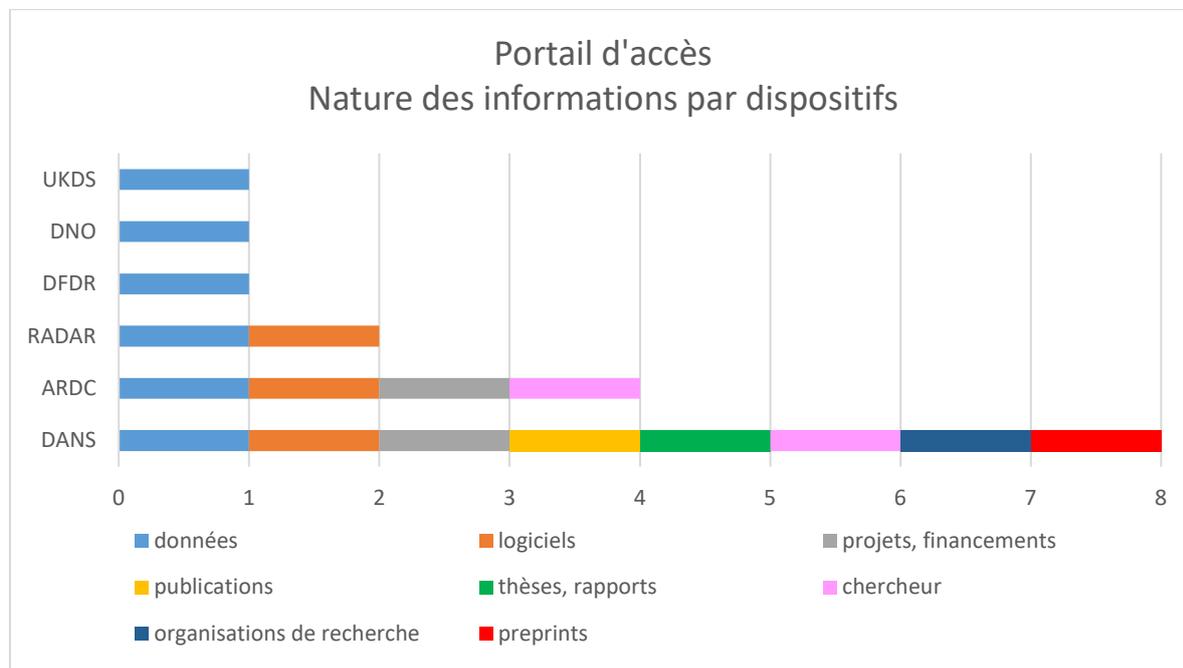
Parmi les six infrastructures qui proposent un portail d'accès aux données de la recherche, la moitié d'entre elles ont fait le choix de proposer un accès aux logiciels (DANS, Research Data Australia (ARDC), RADAR).

**Tableau 17. Nature des informations contenues dans les portails d'accès (cumul des différents services)**



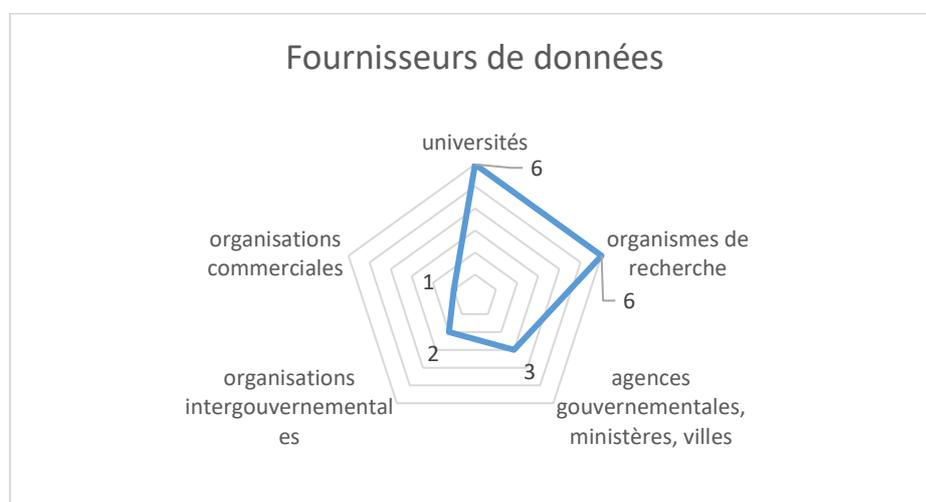
Il est intéressant de noter que le portail NARCIS du réseau DANS propose un accès unique à une gamme d'informations très large qu'il s'est donné pour objectif de lier entre elles : données, publications, thèses, rapports, prépublications, informations portant sur les chercheurs ou les organisations.

**Tableau 18. Nature des informations contenues dans les portails d'accès des différents services**



Si les universités comme les instituts de recherche restent les fournisseurs "naturels" des services nationaux de données, **près de la moitié d'entre eux donnent aussi accès à des données provenant d'établissements à caractère administratif, politique ou économique.**

**Tableau 19. Sources des données contenues dans les portails**



Le tableau « Fonctionnalités des différents services » figure parmi les documents annexes (**voir annexe 2**). Il donne notamment accès à des éléments d'information liés à l'activité des services (volume de dépôts par an, volume total de l'entrepôt).

## 5.3 Des services co-construits et développés avec l'ensemble des acteurs

Au vu du contexte actuel qui voit un nombre croissant d'institutions développer leurs entrepôts de données de recherche se pose la question du positionnement et du rôle d'un service national au sein du paysage des infrastructures préexistantes.

**Tous les services nationaux étudiés se sont développés suite à des projets menés en collaboration avec un large panel de partenaires représentatifs des écosystèmes de recherche.**

Il en va de même pour l'évolution de l'offre de services qui résulte de la consultation, voire de la participation des établissements ou des utilisateurs.

### 5.3.1 Avec les partenaires institutionnels

De 2016 à 2019, le projet pilote intitulé Research Data Shared Service<sup>24</sup>, a réuni de nombreux partenaires (universités, Figshare, Islandora, [digital preservation coalition](#), [open preservation foundation](#), [software sustainability institute](#),...) pour concevoir ensemble les activités et l'organisation du JORH.

Cette dimension partenariale structure la gouvernance des différentes infrastructures. A titre d'exemple, l'organisation du réseau DANS repose sur des accords de partenariat signés avec 12 établissements. Pour chaque service, **le rôle des conseils consultatifs est central dans le développement des services**. Celui de l'UKDS réunit les représentants de l'ensemble des parties, dont le financeur ESRC, il apporte ses conseils en matière d'orientation stratégique et assure un mécanisme de surveillance à destination de l'ESRC. Pour les composantes de DANS (DNL, Easy et Narcis) comme pour DNO ou RADAR, chaque établissement partenaire est représenté au sein des comités consultatifs, instances qui délivrent des conseils et des requêtes à leurs comités d'administration respectifs.

De son côté, lorsqu'il s'agit de "rafraîchir" l'infrastructure de stockage Nectar cloud, l'ARDC sollicite les communautés académiques par le biais d'[appel à propositions](#). Pour initier la création de [nouvelles plateformes disciplinaires](#), l'ARDC a soumis là encore un appel à projets en 2019 auprès des communautés. Les dix projets de plateformes retenus relèvent d'un large panel de disciplines, chaque projet est porté par une institution et s'appuie sur un co-financement partagé entre l'ARDC, l'institution porteuse et d'autres organisations relevant du domaine concerné.

### 5.3.2 Au sein de réseaux nationaux et internationaux

**Tous ces services nationaux inscrivent leurs activités dans le cadre des réseaux nationaux et internationaux, ce qui participe à définir leur identité et l'offre à proposer.**

Ainsi, DNO collabore avec la bibliothèque nationale de Norvège et le Centre norvégien pour les données de recherche<sup>25</sup> autour des activités liées à l'archivage numérique, pour former une communauté de réflexion et de mise en pratique où les responsabilités et les champs d'action sont répartis entre les membres.

<sup>24</sup> <https://research-data-network.readme.io/docs/research-data-shared-service-overview>

<sup>25</sup> <https://nsd.no/nsd/english/index.html>

Il en va de même pour DANS qui est à l'origine de la création du réseau Research Data Netherlands ([RDNL](#)), coalition nationale d'entrepôts de données dans le domaine de l'archivage à long terme des données. Par ailleurs, à l'image des autres infrastructures nationales, le réseau DANS collabore à de nombreux projets dans le cadre de partenariats nationaux (ex. RNDL, Clariah, Bibliothèque nationale néerlandaise) ou internationaux (European Open Science Cloud, Council of European Social Science Data Archives, Datacite, Darjah-EU, CoreTrustSeal, Dryad, Eurocris, Mendeley, Research Data Alliance Europe, Organisation de Coopération et de Développement Economique (OCDE), Science Europe,...).

L'un des autres objectifs majeurs du travail mené au sein de ces réseaux porte sur les bonnes pratiques en matière de gestion des données, notamment en matière d'interopérabilité (ex. standards, normes, protocoles). A ce titre, le réseau DANS coordonne le projet européen [FAIRsFAIR](#) dont l'UKDS, via UKDA, est partenaire et pour lequel DNO et le DFDR ont été choisis comme entrepôts tests. L'objectif majeur de ce projet est d'élaborer des normes internationales pour la certification FAIR des entrepôts et des données qu'ils contiennent, de manière à faciliter la mise en œuvre du programme European Open Science Cloud (EOSC).<sup>26</sup>

Créé pour favoriser une communauté de pratiques, le Réseau Portage Canada réunit les experts nationaux en matière de gestion des données de recherche, il a aussi pour mission de co-piloter le projet du DFDR. La production d'outils résultant des activités du réseau a vocation à être intégrée à l'entrepôt national : outil d'aide à la rédaction de plan de gestion de données, et prochainement des [modèles de plan de gestion spécifiques aux disciplines](#).

**De son côté, le service national norvégien DNO envisage lui aussi de créer un réseau de professionnels experts en matière de curation de données.**

### 5.3.3 En sollicitant les communautés de recherche

Toutes les infrastructures nationales accordent une grande importance aux retours de leurs utilisateurs que ce soit en recueillant leurs besoins via des enquêtes d'utilisateurs (DANS), ou par le biais des retours d'expériences communiqués par les gestionnaires d'archives des institutions partenaires.

Afin de concevoir l'offre de services, les acteurs du projet pilote du JORH ont adopté une approche centrée sur l'utilisateur, acteurs clés sollicités dans le cadre des nombreux cycles d'acceptation<sup>27</sup>.

Ces consultations peuvent donner lieu à la conception de nouveaux outils comme le [Data costing tool](#) proposé par l'UKDS (**voir annexe 7**), outil de calcul des coûts de gestion des données, qui a été développé avec des chercheurs dans le cadre du projet Jisc Data Management Planning<sup>28</sup>. De la même manière, l'UKDS a sollicité l'appui de chercheurs pour définir les besoins lors du développement de [QAMydata](#) (**voir annexe 8**) outil de vérification, de nettoyage et de documentation des données, disponible sous licence CC-BY-NC.

De son côté, l'ARDC propose à l'ensemble des acteurs de la recherche, et notamment aux chercheurs de participer à l'alimentation [Research Vocabularies Australia](#), plateforme collaborative de vocabulaires contrôlés. Elle permet aux organismes de recherche de publier, de réutiliser, de créer et de gérer leurs propres vocabulaires contrôlés, ou encore de gérer les nouvelles versions tout en conservant les versions remplacées. D'autre part, l'ARDC sollicite les différentes communautés scientifiques pour

<sup>26</sup> <https://www.ouvrirlascience.fr/portail-web-de-leosc/>

<sup>27</sup> <https://researchdata.jiscinvolve.org/wp/2019/05/22/going-off-road/>

<sup>28</sup> <https://www.data-archive.ac.uk/about/grants-and-projects/data-management/data-management-planning-for-esrc-data-rich-investments/>

développer son offre de services dans le cadre notamment des appels à projets concernant la mise en place de laboratoires virtuels thématiques.

C'est aussi le cas du Réseau Portage qui a lancé au mois d'avril 2020 un appel à financement auprès des experts des différents domaines scientifiques en vue d'élaborer des modèles de plans de gestion de données spécifiques à chaque discipline.

## 5.4 Modèles économiques

Le rapport "[Business models for sustainable research data repositories](#)"<sup>29</sup> de l'OCDE identifie plusieurs sources de financement qui, combinées ou non, donnent lieu à divers modèles économiques :

- financement structurel : financement central ou contrat d'un bailleur de fonds de la recherche ou de l'infrastructure, sous la forme d'un contrat pluriannuel,
- financement et soutien de l'institution d'accueil : soutien direct ou indirect de la part d'une institution d'accueil,
- frais de stockage : sous la forme de contrats annuels avec les institutions ou de frais liés au volume de dépôts,
- frais d'accès : facturation de l'accès aux données standards ou aux services à valeur ajoutée et les installations,
- services contractuels ou financement de projets : contrats et projets à court terme pour toute autre type d'activités non couvertes ci-dessus.

Les modèles économiques des services étudiés reposent sur une ou plusieurs de ces sources de financement. Si l'on exclut le DFDR dont le modèle économique est en cours de définition, on constate que **le nombre moyen de sources de financements sur lesquelles s'appuient les infrastructures est compris entre 2 et 3.**

**La majorité d'entre elles (5 sur 7) s'appuient sur un financement structurel qui prend la forme d'un contrat pluriannuel signé avec le ministère de la recherche ou un bailleur de fonds public.**

Seule l'ARDC propose à ses établissements partenaires un accès gratuit à l'ensemble de ses services, les autres services nationaux demandent une contribution financière aux institutions membres. Un modèle fréquent consiste à demander aux membres des frais annuels d'adhésion ainsi que des frais liés au volume de stockage (DANS, DNO, RADAR).

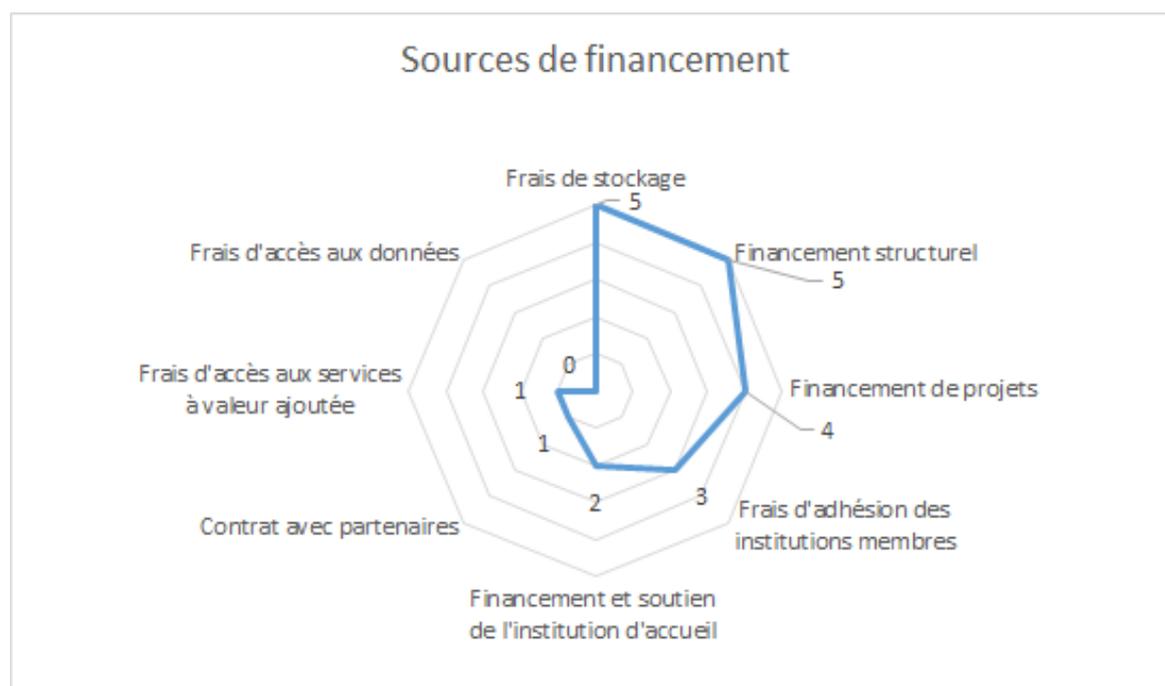
De son côté, le JORH se place dans une logique de fournisseur de services payants en proposant 3 offres distinctes : service d'archivage, service de dépôt, service couplant archivage et dépôt.

**La gratuité d'accès aux données est un principe central, partagé par l'ensemble des services nationaux étudiés.**

---

<sup>29</sup> OCDE (2017), « Business models for sustainable research data repositories », OECD Science, Technology and Industry Policy Papers, no 47, Éditions OCDE, Paris. Disponible au <http://dx.doi.org/10.1787/302b12bb-en>

**Tableau 20. Sources de financement des services nationaux**



Le réseau DNO s'est inspiré des catégories et des points d'attention relevés dans ce même rapport de l'OCDE pour concevoir un modèle d'évaluation des forces et faiblesses de son modèle économique (voir annexe 9).

## 5.5 Economies d'échelle

Différents leviers d'optimisation des coûts sont activés par les infrastructures, à commencer par les gains d'échelle liés à la mutualisation des services, la collaboration entre institutions, ou encore les bénéfices inhérents à un service unique partagé ; c'est le cas du JORH qui affiche **l'objectif suivant "Save institutions time and money as a national, fully managed shared service"**<sup>30</sup> comme l'une de ses priorités. Par ailleurs, JORH pilote actuellement un projet "Financial X-Ray for Research service" dont le but sera de fournir des informations utiles pour diagnostiquer le coût des différents services dont celui du stockage et de l'archivage des données. Ce projet s'appuie sur le modèle "[Financial X-Ray](#)".

Parmi les recommandations pour un service Dataverse national canadien produites par le groupe de travail [Dataverse Nord](#), les économies d'échelle sont citées en premier lieu parmi les avantages liés à un entrepôt national. Ce rapport conclut qu'"**un service national aura un impact considérable sur la capacité des institutions de petite et moyenne taille à développer des services de dépôts institutionnels de GDR.**"<sup>31</sup>

<sup>30</sup> <https://www.jisc.ac.uk/open-research-hub>

<sup>31</sup>

<https://open.library.ubc.ca/cIRcle/collections/facultyresearchandpublications/52383/items/1.0385836>

Le coût des opérations liées à la gestion des données, notamment la **curation**, est un enjeu déterminant pour la plupart des services nationaux. Ces activités sont le plus souvent assurées par les gestionnaires des établissements partenaires à qui il est déterminant d'offrir la possibilité de s'appuyer sur le réseau national pour recueillir l'expertise nécessaire, faciliter leur montée en compétence, ou encore mutualiser certaines activités. **De nombreuses infrastructures nationales affirment faire des économies d'échelle, générées par la mutualisation de ce type d'activités, un de leurs objectifs.**

Dataverse Norvège explique notamment aux partenaires que la montée en compétence de leurs agents permettra de limiter la hausse des coûts de gestion liée à l'augmentation continue des volumes de données à gérer. Partant de ce constat, l'enjeu de la formation est de faire **baisser le temps moyen utilisé pour la curation de chaque jeu de données.**

D'autres possibilités d'optimisation des coûts peuvent prendre la forme du **recours à des solutions externalisées**, c'est notamment le cas du **JORH qui agit en tant qu'intermédiaire** auprès des établissements partenaires en leur proposant différentes solutions de stockage et d'archivage chez des prestataires privés (Amazon Web Services, Microsoft,...).

Il n'a pas été possible de recueillir d'éléments chiffrés concernant les éventuelles économies d'échelle dégagées par les services nationaux étudiés.

Cependant, Peter Dorn (DANS) estime que l'utilisation de Dataverse NL coûte aux établissements partenaires dix fois moins cher que de développer seuls leur propre entrepôt : "You gain substantially by joining forces, not having each university buy their own data repository system but to share that."<sup>32</sup> Il souligne notamment **l'importance d'adopter des solutions communautaires de manière à ne pas se rendre captif de son propre système**, comme cela est actuellement le cas du réseau DANS avec leur solution d'archivage Easy qui ne propose pas de couche d'archivage et dont les coûts de maintenance deviennent particulièrement élevés.

Prédécesseur de l'UKDS, l'Economic and Social Data Service (ESDS) a fait l'objet d'un rapport réalisé en 2012 à la demande de l'ESRC intitulé "Economic Impact Evaluation of the Economic and Social Data Service"(UKRI)<sup>33</sup>. Parmi ses conclusions, celui-ci avance que "la valeur pour la communauté des utilisateurs au sens large, mesurée en termes de gains d'efficacité et d'économies de coûts, est de l'ordre de 100 millions de livres sterling par an ou plus, et nous estimons que l'utilisation accrue, facilitée par l'ESDS, des données hébergées entraîne une augmentation du retour sur investissement annuel dans la création et la collecte des données allant jusqu'à 233 millions de livres sterling sur 30 ans (exprimée en valeur actuelle nette)<sup>34</sup>." Si la méthodologie de cette étude a fait l'objet de critiques, elle a démontré que l'investissement dans l'infrastructure offre un retour sur investissement considérable : **la valeur économique du service national est estimée cinq fois supérieure à ses coûts opérationnels, les retours sur investissement dans les données hébergées seraient dix fois supérieure aux coûts.**

Au-delà de l'impact économique, ce rapport souligne la diversité des opportunités générées par un service national de données, pour l'ensemble des communautés impliquées (scientifiques, services des

<sup>32</sup> Extrait d'un entretien avec Peter Dorn tenu le 02 octobre 2020

<sup>33</sup> "Economic Impact Evaluation of the Economic and Social Data Service": <https://esrc.ukri.org/files/research/research-and-impact-evaluation/economic-impact-evaluation-of-the-economic-and-social-data-service/>

<sup>34</sup> trad. de : "Economic Impact Evaluation of the Economic and Social Data Service"

institutions) mais aussi pour les publics non académiques. Les auteurs du rapport proposent le tableau ci-après listant l'ensemble des bénéfices identifiés.<sup>35</sup>

**Tableau 21. Impact directs et indirects de l'Economic and social data service**

<b>Direct Benefits</b>	<b>Indirect Benefits (Costs Avoided)</b>
Time and resource savings for researchers and teachers	No re-creation of data -Trusted Digital
Verification of research through increased data citation thanks to relevant citation information and tools	Repository status eliminates re-ingest costs
Access to data provides new research opportunities by increasing use of data within collections thanks to proper Collections Development Policy	Lower future archiving costs increase likelihood of data being available, earlier in the lifecycle
Re-purposing and re-use of data maximises the value of data holdings when the data is available for third parties	Re-purposing data for new audiences
Increasing research productivity	Use by new audiences
Assisting in provision of a skills base	Protecting returns on earlier investments
(-Customer service ethos of ESDS staff	Reduces potential duplication of effort
Near Term Benefits	Long-Term Benefits
Value to current researcher and students	Data preserved for the long-term
Single point of access Increasing speed of access to data	Secures value of high quality data for future researchers and students
Ease of access for researchers and students	Value added over time as collection grows and develops critical mass
Quality of documentation and contextual information, and resources for teaching	Input for future research by maximising use across data user community ( ie including students)
No data lost from post-doc turnover thanks to a reliable preservation system	Impact on wider profession as a centre of excellence

<sup>35</sup> extrait du rapport "Economic impact evaluation of the economic and social data service"  
<https://esrc.ukri.org/files/research/research-and-impact-evaluation/economic-impact-evaluation-of-the-economic-and-social-data-service/>

Secure storage for data intensive research	Promoting quantitative methods and skills
Availability of data underpinning journal articles	Fostering innovation in research and data management practice
Research data integrity since data is of high quality	
Private Benefits	Public Benefits
Benefits to sponsor of research	Source of high-quality and often unique data
Benefits to sponsor of data service provider	Motivating new research
Benefits to researcher	Enables research that otherwise could not be undertaken
Fulfil grant obligations	Research integrity since others can check the outcomes of research
Increased visibility/citation	Service targeted at academic community and supports their needs
Aggregator of data for data providers – providing them with one point of access to UK customers	Fostering transferable skills in data analysis
Higher usage (and profile) of their data for depositors	Articulating user needs to data providers
Removes user burden from depositors	

Le tableau « Ressources des différents dispositifs » figure parmi les annexes de ce document (**voir annexe 3**). Lorsque les informations ont été rendues publiques, ce document donne notamment accès à des éléments de chiffrage.

## 5.6 Pérennité du service

### 5.6.1 Evaluation des services

Comme le souligne le rapport de l'OCDE, la pérennité du modèle économique passe notamment par la compréhension de la proposition de valeur qu'offre l'entrepôt de données. Cette bonne compréhension détermine l'engagement et le maintien des bailleurs de fonds structurels, institutionnels, philanthropiques ou autres et suppose que des analyses portant sur le rapport entre les coûts et les bénéfices, la valeur et l'impact des services soient menées.

**Il n'est donc pas étonnant de constater que l'ensemble des services accorde une place importante aux indicateurs d'usage ainsi qu'aux retours des utilisateurs.**

Le JORH fait du service de reporting un des piliers de son offre de service. Interrogeable via API, ce service offre aux établissements partenaires un accès à des ensembles de métadonnées internes (stockées dans l'infrastructure nationale) ainsi qu'à des informations provenant d'autres systèmes et services intégrés. Ces informations peuvent également être exportées dans les systèmes d'information des établissements.

L'entrepôt DNL propose un accès aux statistiques d'usage et de dépôt pour l'ensemble du réseau comme pour chaque institution membre.

Dans son contrat stratégique (2015-2020), le réseau néerlandais DANS s'est fixé des objectifs annuels en termes d'utilisation des services (+50% en 5 ans) et fait de l'organisation d'enquête qualité l'une de ses priorités (voir infra). Celle-ci repose sur 7 indicateurs qui font l'objet d'un suivi trimestriel et donnent lieu à des échanges avec l'ensemble de l'équipe.

**Tableau 22. Tableau 20. Indicateurs liés aux objectifs stratégiques du réseau DANS <sup>36</sup>**

Strategic priority	Strategic objective	2015	2016	2017	2018	2019	2020
<b>1. Leading building block</b>	<b>1.</b> Prominent in federative data infrastructure	Participation in RI projects	Participation in RI projects	Consulting peers	Participation in RI projects	Participation in RI projects	Consulting peers
<b>2. Improved services for more users</b>	<b>2.</b> Growth in use of services	Index 100	Index 110	Index 120	Index 130	Index 140	Index 150
	<b>3.</b> User satisfaction	User Satisfaction Survey (GTO)		GTO / External evaluation		GTO	Mid-term evaluation

L'évaluation des services peut aussi porter sur des objectifs en termes de mise en œuvre de projets comme c'est le cas avec le service DNO qui a dressé en 2020 un « **état des lieux de l'implémentation des principes FAIR** » au sein de l'entrepôt national norvégien (voir annexe 10).

## 5.6.2 Démarches de certification

La fiabilité de l'entrepôt étant une condition nécessaire pour un accès pérenne aux données hébergées, **la plupart des services nationaux de données ont entrepris des démarches de certification ou de labellisation.** C'est notamment le cas du service d'archivage Easy du réseau DANS et de Dataverse Norvège qui ont reçu la certification [Core TrustSeal](#).

A l'image de nombreux autres services proposés par le JISC, l'infrastructure centrale du JORH dispose de la certification [ISO 27001](#), norme internationale de gestion de la sécurité de l'information (confidentialité, intégrité et disponibilité).

<sup>36</sup> Tableau extrait de « Data sharing together : summary of DANS strategy » : <https://dans.knaw.nl/en/about/organisation-and-policy/information-material/DANSstrategienota20152020UK.pdf>

Lui aussi certifié ISO 27001, l'UKDS propose les mêmes garanties de sécurité. Par ailleurs, l'UKDS participe activement à la structuration de réseaux d'entrepôts de confiance à la fois en tant que membre de [Digital Preservation Coalition](#) et en occupant la vice-présidence de CoreTrustSeal (CTS).

De la même manière, le réseau DANS contribue activement aux activités de l'organisation internationale de certification CTS puisque plusieurs de ses membres participent au conseil d'administration et qu'il met des ressources humaines à disposition de cette organisation internationale de certification.

Après avoir accompagné des établissements lors de leur démarche de certification (CSIRO, le National Imaging Facility et l'Australian Data Archive), l'ARDC coordonne actuellement les activités du réseau "Trusted Data Repositories Community of Practice<sup>37</sup>" dont l'un des objectifs est de mettre en place des procédures visant à réduire les efforts à déployer lors d'une démarche de certification CTS.

---

37

<https://docs.google.com/document/d/1uDuaAunIS9QCxKRKWhXvDAL7UCtoo6jDvJ14Qw0Xjkc/edit#heading=h.r45gaecuqh44>

## Conclusion.

# S'insérer de manière pragmatique et agile au sein d'écosystèmes nationaux et internationaux dynamiques.

Pour conclure, il semble important de faire un nouveau détour par le rapport de l'OCDE qui recommande notamment aux opérateurs d'entrepôts de données de faire preuve de pragmatisme.

Si l'objectif évident est de réfléchir à l'ensemble des conditions à réunir pour proposer un outil qui puisse offrir des services de manière pérenne, il est tout aussi important de savoir se détacher d'un modèle qui ne donne pas ou plus satisfaction.

Partant de ce constat, recueillir l'adhésion des partenaires suppose aussi de donner à ces derniers des garanties en termes de continuité de service, en sachant anticiper un éventuel transfert vers un autre service. A titre d'exemple, dans ses documents contractuels, DNO précise à ses partenaires que le service s'engage à transférer les données vers un autre entrepôt en cas d'interruption de service.

Par ailleurs, à l'examen de la mise en place des différents dispositifs, il ressort que les services ou les **projets d'infrastructures nationales de données demandent plusieurs années avant de se développer**. A titre d'exemple, le réseau DANS créé en 2005 a intégré l'outil de découverte Narcis et la plateforme d'archivage Easy en 2011, puis DNL en 2014. Certaines initiatives comme le NIH Data Commons (Data Commons Pilot Phase Consortium 2017-2018) ou le National Data US (pilot 2015-2018) n'ont pas encore dépassé le stade du projet. Actuellement composé de cinq universités membres, le service national allemand RADAR voit son extension freiner par la durée des négociations qu'il mène avec de potentiels nouveaux partenaires. Le projet pilote du JORH, [Research Data Shared Service](#), a duré près de 4 ans (2016-2019).

Compte tenu du contexte très dynamique qui caractérise l'environnement des entrepôts de recherche (institutionnels comme disciplinaires), les services nationaux optent le plus souvent pour des choix techniques et organisationnels qui leur offrent une forme de flexibilité, de manière à adapter leur offre de services aux usages et attentes des utilisateurs.

A titre d'exemple, RADAR dont la devise est "De la communauté pour la communauté" étend actuellement son champ d'actions vers des solutions disciplinaires (contrairement à l'offre générique publiée en 2017) et réfléchit au développement de modèles d'exploitation alternatifs en vue de proposer un archivage local des données (par opposition à la solution cloud actuelle).

De la même manière, suite aux attentes de ses tutelles et aux retours de ses utilisateurs, le réseau néerlandais DANS s'est fixé comme objectif d'orienter davantage leurs services vers d'autres disciplines que les sciences humaines et sociales. Pour ce faire, **un projet de création de dataverses dédiés à certaines communautés disciplinaires et proposant des services spécifiques est en cours**.

La démarche retenue par le Research Shared Data Service, projet pilote du JORH, illustre plus encore la nécessaire prise en compte de l'évolution constante des attentes des institutions. Afin de fournir un produit fonctionnel composé de caractéristiques suffisantes pour à la fois satisfaire les premiers clients et pour fournir un retour d'information en vue du développement futur du produit, l'équipe a développé

une approche basée sur le concept du minimum viable product (MVP). De cette façon, **le JORH s'est développé de façon incrémentale selon un processus itératif**. Ce choix leur permet de s'adapter à l'évolution des besoins et des priorités mais suppose de consulter régulièrement les partenaires pour hiérarchiser les priorités.

Les évolutions à apporter peuvent aller jusqu'à l'intitulé même du service national. Partant du constat que la formulation "open research" ne reflétait pas l'ensemble des services proposés, le Jisc a choisi d'abandonner, en novembre 2020, la marque "JORH" pour quatre intitulés propres à chaque service :

- "Research repository"
- "Research repository plus"
- "Preservation"
- "Research system connect"

Par ailleurs, s'il est difficile de pouvoir anticiper l'ensemble des évolutions à venir, il apparaît déterminant pour sa réussite qu'un service national de données s'insère au mieux au sein du paysage national et international.

L'ensemble des acteurs impliqués dans le développement des services nationaux étudiés insiste en effet sur l'importance de connaître précisément les services et systèmes existants au sein des établissements relevant de leur périmètre national, de relever leurs attentes ainsi que les difficultés rencontrées. Comme nous l'a exprimé Peter Dorn (DANS) lors d'un entretien, la question à se poser est la suivante : "If you set up something new, how does it fit in what already exists?"

Ce travail préliminaire doit permettre de réfléchir à la meilleure façon d'inclure le service national au sein d'un écosystème préexistant, de manière à compléter l'offre de services proposée par les dispositifs des établissements. Cependant, John Kaye (JORH) suggère d'impliquer un nombre limité d'établissements "testeurs" dans la mise en œuvre de ce travail de co-construction, pour éviter tout risque de dispersion.

De la même façon, tous les services étudiés contribuent à l'activité d'infrastructures, de réseaux ou de projets internationaux. Ces activités de coopération et de coordination leur permettent par exemple de s'aligner sur des normes et des pratiques communes, mais aussi de s'appuyer sur les acquis et la dynamique d'une communauté (ex. Dataverse project).

Actuellement, une attention particulière est portée au développement de l'European Open Science Cloud (EOSC) au sein duquel la plupart des services étudiés ont vocation à s'intégrer. **A l'heure où des acteurs privés renforcent leur positionnement en tant que services de données, l'EOSC pourrait offrir un cadre de collaboration internationale entre services de données nationaux.**

Par ailleurs, comme le soulève le rapport de l'OCDE "Coordination and support of international research data network"<sup>38</sup>, le succès des réseaux internationaux liés aux données de la recherche dépend étroitement des politiques d'incitation à l'ouverture des données. Il suggère notamment aux gouvernements de fournir un cadre politique cohérent autour des données ouvertes.<sup>39</sup> Ce constat rejoint celui exprimé par plusieurs acteurs des services étudiés qui insistent sur **le rôle clé que peuvent**

<sup>38</sup> [https://www.oecd-ilibrary.org/science-and-technology/co-ordination-and-support-of-international-research-data-networks\\_e92fa89e-en](https://www.oecd-ilibrary.org/science-and-technology/co-ordination-and-support-of-international-research-data-networks_e92fa89e-en)

<sup>39</sup> "But the top issue faced by data networks in open sharing of data is the varying attitudes and policies across countries. As with interoperability, fostering openness is a trust building exercise. It is a long process, but governments can do much to accelerate the process by providing a consistent policy framework around appropriately open data."

**notamment jouer les financeurs** : leurs incitations à l'ouverture des données déterminent en partie le succès des services nationaux.

## Sites web utiles

- « Australian Research Data Commons (ARDC) ». Consulté le 21 juin 2020. <https://ardc.edu.au/>.
- Borulko, Stefan Ramaker, Sergey. « RDNL ». TU Delft - Research Data Netherlands. Consulté le 21 juin 2020. <https://researchdata.nl/>.
- CoreTrustSeal. « CoreTrustSeal ». Consulté le 21 juin 2020. <https://www.coretrustseal.org/>.
- « Dataverse Norvège ». Consulté le 21 juin 2020. <https://site.uit.no/dataverseno/european-dataverse-workshop-2020/>.
- « Dépôt Fédéré de Données de Recherche (DFDR) ». Consulté le 21 juin 2020. <https://www.frdr-dfdr.ca/repo/?locale=fr>.
- « DORANum ». Consulté le 21 juin 2020 <https://doranum.fr/>
- « Dutch Archiving and Networked Services (DANS) — Nederlands ». Consulté le 21 juin 2020. <https://dans.knaw.nl/nl>.
- « European Open Science Cloud (EOSC) Portal ». Consulté le 21 juin 2020. <https://www.eosc-portal.eu/>.
- « FAIRsFAIR ». Consulté le 21 juin 2020. <https://www.fairsfair.eu/>.
- Réseau Portage. « Groupe de travail Dataverse Nord ». Consulté le 21 juin 2020. <https://portagenetwork.ca/fr/reseau-dexperts/groupe-de-travail-dataverse-nord/>.
- Jisc. « JISC Open Research Hub (JORH) ». Consulté le 21 juin 2020. <https://www.jisc.ac.uk/open-research-hub>.
- « Ouvrir la science ». Consulté le 21 juin 2020. <https://www.ouvrirlascience.fr/>.
- « Research Data Management Toolkit | Jisc ». Consulté le 21 juin 2020. <https://rdmtoolkit.jisc.ac.uk/>.
- « Research Data Repository (RADAR) - Ein Repository für die Wissenschaft ». Consulté le 3 avril 2020. <https://www.radar-service.eu/en/home>.
- RVA. « Research Vocabularies Australia ». Consulté le 21 juin 2020. <https://vocabs.ardc.edu.au/>.
- « UK Data Service ». Consulté le 21 juin 2020. <https://ukdataservice.ac.uk/>.

# Bibliographie

- « Abschlussbericht\_DFG-Projekt\_RADAR\_Veröffentlichung.pdf ». Consulté le 12 mai 2020. [https://www.radar-service.eu/sites/default/files/publications/Abschlussbericht\\_DFG-Projekt\\_RADAR\\_Vero%CC%88ffentlichung.pdf](https://www.radar-service.eu/sites/default/files/publications/Abschlussbericht_DFG-Projekt_RADAR_Vero%CC%88ffentlichung.pdf).
- Andreassen, Helene N., Philipp Konzett, Stein Høydalsvik, Laura Alexis Janda, Leif Longva, Tore Nettet, et Obiajulu Odu. « TROLLing: Scope and Operation of an Open Respository for Linguistic Data », janvier 2017. <https://munin.uit.no/handle/10037/11868>.
- « ARDC-Strategic-Plan-2019-2023 ». Consulté le 5 janvier 2020. <https://ardc.edu.au/wp-content/uploads/2019/05/ARDC-Strategic-Plan-2019-2023.pdf>.
- Baars, Chris, et Maaike De Jong. « NARCIS & FREYA Can PIDs Contribute to Research in Context? », 29 janvier 2019. <https://doi.org/10.5281/zenodo.2552138>.
- Barker, Michelle, Ross Wilkinson, et Andrew Treloar. « The Australian Research Data Commons ». *Data Science Journal* 18, n° 1 (5 septembre 2019): 44. <https://doi.org/10.5334/dsj-2019-044>.
- Barsky, Eugene, Corey Davis, Alan Darnell, Jason Flynn, Lisa Goddard, Meghan Goodchild, Amber Leahey, et al. « Recommendations pour un Service dataverse national », 14 février 2018. <https://doi.org/10.14288/1.0385836>.
- Borgman, Christine L. « Big Data and the Long Tail: Use and Reuse of Little Data », 12 mars 2013. <https://escholarship.org/uc/item/7740w0pg>.
- Buchhorn, Dr Markus. « Surveying the Scale of the Research-IT Support Workforce. », 2019. <https://ardc.edu.au/project/skilled-workforce-summit-and-projects/>.
- Burland, Tamsin, John Kaye, Dom Fripp, Tom Davey, Paul Stokes, et Caroline Ingram. « Introducing the Jisc Open Research Hub ». 5 juin 2019. <https://doi.org/10.5281/zenodo.3239361>.
- Burton, Adrian, David Groenewegen, Cynthia Love, Andrew Treloar, et Ross Wilkinson. « Making Research Data Available in Australia ». *IEEE Intelligent Systems* 27, n° 3 (mai 2012): 40-43. <https://doi.org/10.1109/MIS.2012.57>.
- « Business Models for Sustainable Research Data Repositories ». OECD Science, Technology and Industry Policy Papers, 6 décembre 2017. <https://doi.org/10.1787/302b12bb-en>.
- Carter, Ian. « Current Practices in Research Data Management and Sharing ». Zenodo, 28 février 2020. <https://doi.org/10.5281/zenodo.3824605>.
- « Certification CoreTrustSeal DataverseNO.pdf ». Consulté le 6 novembre 2020. <https://www.coretrustseal.org/wp-content/uploads/2020/03/DataverseNO.pdf>.

- Conzett, Philipp. « DataverseNO: A National, Generic Repository and Its Contribution to the Increased FAIRness of Data from the Long Tail of Research ». *Ravnetrykk*, n° 39 (31 mai 2020): 74–113-74–113. <https://doi.org/10.7557/15.5514>.
- « Co-Ordination and Support of International Research Data Networks ». OECD Science, Technology and Industry Policy Papers, 8 décembre 2017. <https://doi.org/10.1787/e92fa89e-en>.
- CoreTrustSeal Board. « CoreTrustSeal Foundation Statutes and Rules of Procedure », 14 février 2018. <https://doi.org/10.5281/zenodo.1142960>.
- Corti, Louise, Veerle Van den Eynden, Libby Bishop, et Matthew Woollard. *Managing and Sharing Research Data: A Guide to Good Practice*. SAGE, 2019. [https://books.google.fr/books?hl=fr&lr=&id=PkSxDwAAQBAJ&oi=fnd&pg=PP1&dq=research+data+service&ots=xZOzisfL00&sig=-JnQVrMy3WpFj6Fx-luzm9UymL0&redir\\_esc=y#v=onepage&q=research%20data%20service&f=false](https://books.google.fr/books?hl=fr&lr=&id=PkSxDwAAQBAJ&oi=fnd&pg=PP1&dq=research+data+service&ots=xZOzisfL00&sig=-JnQVrMy3WpFj6Fx-luzm9UymL0&redir_esc=y#v=onepage&q=research%20data%20service&f=false).
- Corti, Louise, Myles Offord, et Christina Magder. « Quality Assessment of Numeric Data: QAMyData ». s. d. [https://ukdataservice.ac.uk/media/622448/qamydata\\_overview.pdf](https://ukdataservice.ac.uk/media/622448/qamydata_overview.pdf).
- Corti, Louise. « UK Data Service (UKDS) Institutional Partnership Project ». 1 mai 2020. <https://doi.org/10.5281/zenodo.3781009>.
- « CRIS2018\_Jisc\_Research\_Data\_Shared\_Service\_CERIFication.pdf ». Consulté le 11 mai 2020. [https://dspacecris.eurocris.org/bitstream/11366/675/4/CRIS2018\\_Jisc\\_Research\\_Data\\_Shared\\_Service\\_CERIFication.pdf](https://dspacecris.eurocris.org/bitstream/11366/675/4/CRIS2018_Jisc_Research_Data_Shared_Service_CERIFication.pdf).
- « DANS deposit-agreement ». Consulté le 19 janvier 2020. <https://dans.knaw.nl/en/about/organisation-and-policy/legal-information/deposit-agreement.pdf>.
- « DANS strategie 2015-2020 ». Consulté le 22 janvier 2020. <https://dans.knaw.nl/en/about/organisation-and-policy/information-material/DANSstrategienota20152020UK.pdf>.
- « DANS strategy policy 2011-2015 ». Consulté le 22 janvier 2020. <https://dans.knaw.nl/en/about/organisation-and-policy/information-material/DANS-strategy-policy-20112015-summaryUK.pdf>.
- « Data costing tool UKDS ». Consulté le 21 juin 2020. <https://www.ukdataservice.ac.uk/media/622368/costingtool.pdf>.
- « Dataverse NL General Terms of Use Dataverse NL ». Consulté le 19 janvier 2020. [https://dans.knaw.nl/en/about/services/DataverseNL/GeneralTermsofUseDataverseNL\\_DEF.pdf](https://dans.knaw.nl/en/about/services/DataverseNL/GeneralTermsofUseDataverseNL_DEF.pdf).
- Davey, Tom, Dom Fripp, et John Kaye. « Jisc Open Research Hub – Supporting Open Scholarship ». 26 novembre 2019. <https://doi.org/10.5281/zenodo.3554342>.
- Dom Fripp, et Tom Davey. « A Proposal to Add a FAIR Metric to the Jisc Open Research Hub ». 31 mars 2019. <https://doi.org/10.5281/zenodo.2619357>.

- Doorn, P.K., L.S. Bommeljé, J.A.C. Vroom, H. Van Wijngaarden, et Y.B. Bommeljé. « The Aetolian Studies Project ». Data Archiving and Networked Services (DANS), 2016. <https://doi.org/10.17026/DANS-XXU-6UTQ>.
- « EASY General Terms and Conditions of Use ». Consulté le 19 janvier 2020. <https://dans.knaw.nl/en/about/organisation-and-policy/legal-information/EASYGeneralTermsandConditionsofUse.pdf>.
- « Évaluation et labellisation des sites d'archivage pérenne – DoRANum ». Consulté le 24 mai 2020. <https://doranum.fr/stockage-archivage/evaluation-labellisation-sites-archivage-perenne/>.
- Frapp, Dom. « Going Off-Road ». *Research Infrastructure and Data* (blog), 22 mai 2019. <https://researchdata.jiscinvolve.org/wp/2019/05/22/going-off-road/>.
- « Integrating the Future of Research Outputs ». *Wonkhe* (blog). Consulté le 19 avril 2020. <https://wonkhe.com/blogs/integrating-the-future-of-research-outputs/>.
- Goldstein, Stephane. « The evolving landscape of Federated Research Data Infrastructures ». Zenodo, 27 novembre 2017. <https://doi.org/10.5281/zenodo.1064730>.
- Heeley, Melanie. « Jisc Launches Research Repositories Dynamic Purchasing System | Jisc Scholarly Communications ». Consulté le 10 mai 2020. <https://scholarlycommunications.jiscinvolve.org/wp/2020/05/04/jisc-launches-research-repositories-dynamic-purchasing-system/>.
- Ingram, Caroline. « Piloting Institutional Digital Preservation ». *Research Infrastructure and Data* (blog), 7 novembre 2019. <https://researchdata.jiscinvolve.org/wp/2019/11/07/piloting-institutional-digital-preservation/>.
- Jensen, Eric A., et Mark Reed. « Investigating the Link Between Research Data and Impact ». Zenodo, 24 octobre 2019. <https://doi.org/10.5281/zenodo.3543505>
- JiscSD/Canonical-data-model*. 2016. Reprint, JiscSD, 2020. <https://github.com/JiscSD/Canonical-data-model>.
- Kaye, John, Rachel Bruce, et Dom Frapp. « Establishing a Shared Research Data Service for UK Universities ». *Insights* 30, n° 1 (10 mars 2017): 59-70. <https://doi.org/10.1629/uksg.346>.
- Kraft, Angelina, Matthias Razum, Jan Potthoff, Andrea Porzel, Thomas Engel, Frank Lange, Karina Van den Broek, et Filipe Furtado. « The RADAR Project—A Service for Research Data Archival and Publication ». *ISPRS International Journal of Geo-Information* 5, n° 3 (mars 2016): 28. <https://doi.org/10.3390/ijgi5030028>.
- Lamprecht, Anna-Lena, Leyla Garcia, Mateusz Kuzak, Carlos Martinez, Ricardo Arcila, Eva Martin Del Pico, Victoria Dominguez Del Angel, et al. « Towards FAIR Principles for Research Software ». *Data Science* 3, n° 1 (1 janvier 2020): 37-59. <https://doi.org/10.3233/DS-190026>.

- Ltd, Charles Beagrie. « Economic Impact Evaluation of the Economic and Social Data Service », mars 2012, 97. <https://esrc.ukri.org/files/research/research-and-impact-evaluation/economic-impact-evaluation-of-the-economic-and-social-data-service/>
- Maegaard, Bente, et Riccardo Pozzo, éd. *Stay Tuned to the Future: Impact of the Research Infrastructures for Social Sciences and Humanities*. Lessico Intellettuale Europeo, CXXVIII. Firenze: Leo S. Olschki editore, 2019. [https://www.clarin.eu/sites/default/files/STAY\\_TUNED\\_TO\\_THE\\_FUTURE.pdf](https://www.clarin.eu/sites/default/files/STAY_TUNED_TO_THE_FUTURE.pdf)
- Mokrane, Mustapha, et Jonas Recker. « CoreTrustSeal–Certified Repositories: Enabling Findable, Accessible, Interoperable, and Reusable (FAIR) », 27 septembre 2019. <https://doi.org/10.17605/OSF.IO/9DA2X>.
- « The Jisc Open Research Hub and Its Role in Open Research Infrastructure », 20 novembre 2019. <https://dspacecris.eurocris.org/handle/11366/1240>.
- « Optimising the Operation and Use of National Research Infrastructures ». OECD Science, Technology and Industry Policy Papers, 3 août 2020. <https://doi.org/10.1787/7cc876f7-en>.
- « Processing agreement Easy (DANS) ». Consulté le 19 janvier 2020. <https://dans.knaw.nl/en/about/organisation-and-policy/legal-information/processingagreementeasy.pdf>.
- CASRAI. « Research Data Management ». Consulté le 21 juin 2020. <https://casrai.org/term/research-data-management/>.
- Research Data Network. « Research Data Shared Service Pilot ». Consulté le 9 mai 2020. <https://research-data-network.readme.io/docs/research-data-shared-service-overview>.
- Russell, Kate. « JISC Research Data Shared Services and JORH Project End Blog Post - Research Kaleidoscope ». Consulté le 11 mai 2020. <http://blogs.plymouth.ac.uk/research/2019/09/30/jisc-research-data-shared-services-and-jorh-project-end-blog-post/>.
- Sloyan, Victoria. « Overview of a Born-Digital Archives Access Workshop Held at Wellcome Collection », 2018, 591528 Bytes. <https://doi.org/10.6084/M9.FIGSHARE.6087194.V1>.
- Soltau, Kerstin, Matthias Razum, et Dorothea Strecker. « Management digitaler Forschungsdaten im akademischen Umfeld – Lessons learned aus der Einführung von RADAR », 2019, 17.
- Stillwell, John. *The Routledge Handbook of Census Resources, Methods and Applications: Unlocking the UK 2011 Census*. Routledge, 2017.
- « Survey about Depositing via EASY ». Consulté le 18 avril 2020. <https://www.surveymonkey.com/r/9CGQ8YF>.
- Sykes, Jean. « The UK Research Data Service Project », 2009, 20.
- Turp, Clara, Lee Wilson, Julianne Pascoe, et Alex Garnett. « The Fast and the FRDR: Improving Metadata for Data Discovery in Canada ». *Publications* 8, n° 2 (juin 2020): 25. <https://doi.org/10.3390/publications8020025>.

« UKDS Annual report18 ». Consulté le 10 mai 2020.

<https://ukdataservice.ac.uk/media/622467/ukdsannualreport18.pdf>.

Van Horik, René. « Dutch Data Delights with DANS ». EUDAT, 27 août 2015.

<https://eudat.eu/news/dutch-data-delights-dans>.

# Annexes

## Annexe 1. Identifiants, métadonnées et indicateurs utilisés par les services nationaux de données

Services	Identifiants	Standards de métadonnées et vocabulaire	API	Indicateurs Forme et méthodes	Conformité COUNTER Code of Practice	Moissonnage par services tiers
ARDC Research Data Australia	<p><u>Research Data Australia</u> Jeux de données :</p> <ul style="list-style-type: none"> <li>- DOI</li> <li>- identifiant RDA</li> <li>- identifiant institutionnel</li> <li>- URI</li> </ul> <p>Auteurs :</p> <ul style="list-style-type: none"> <li>- identifiant institutionnel</li> <li>- ISNI, Orcid, URI</li> </ul> <p>URL des sites web associés URI des projets</p> <p>Soutien consultatif d'experts aux organismes de recherche australiens qui intègrent les PID dans leurs processus opérationnels</p> <p><u>Identifiants requis pour les projets financés</u> : ORCID, Handle, DOI, ROR, RAID, PURL, Grant</p>	Métadonnées : norme ISO 19115 , Repository-Developed Metadata Schemas, FundRef Registry	OAI-PMH REST	<p><u>Research Data Australia</u> Nombre de vues</p> <p>Pour tout projet financé, rapport à livrer entre 12 et 24 mois après la fin du projet :</p> <ul style="list-style-type: none"> <li>- résultats : publications, financements, projets, brevets, rapports, licences, usage (ventilé par catégories), citations, taux de satisfaction, formation délivrée, réutilisation</li> <li>- impact : quels publics, type, étendue, période</li> </ul> <p>Recommandation pour plan de suivi : <a href="https://www.oecd-ilibrary.org/docserver/3ffee43b-en.pdf?expires=1599211044&amp;id=id&amp;accname=quest&amp;checksum=E57A9DCB1B0B73FFECC85C27F8AE3D7E">https://www.oecd-ilibrary.org/docserver/3ffee43b-en.pdf?expires=1599211044&amp;id=id&amp;accname=quest&amp;checksum=E57A9DCB1B0B73FFECC85C27F8AE3D7E</a></p> <p>WP pour la mise en œuvre de l'évaluation Intégration automatisée des identifiants</p> <p>Communications permanentes automatisées avec les utilisateurs finaux concernant leur utilisation Autres systèmes de suivi des publications de recherche, des subventions, des</p>	Oui pour projets financés	Open Aire

	ID, IGSN, Crossref Funder			projets, des prix, des financements industriels, des brevets, etc. rendus possibles par l'infrastructure  Collaboration avec les éditeurs de revues afin qu'ils saisissent les citations des données  Collaboration avec les sociétés savantes et les bailleurs de fonds afin de mettre en œuvre des politiques communautaires pour la citation des données		
DANS	Jeux de données : DOI, Handle, URN ou ISBN  Auteurs : ORCID, ISNI et Digital Author Id (DAI), Researcher Id (ISNI a vocation à remplacer DAI)  Organisations : ROR, ISNI et GRID.  Participation au projet FREYA-PID graph : lier des données par l'enrichissement d'identifiants persistants (PID)	Métadonnées : Data Documentation Initiative (DDI), Dublin Core, DataCite Metadata Schema  Classification Narcis basée sur SKOS RDF (codes disciplines et codes scientifiques) interdisciplinaires)  Echange de données avec le résolveur national et la bibliothèque nationale : MODS et DIDL, Dublin Core possible  Vocabulaires et ontologies : <a href="#">VOResource Schema format</a> (International Virtual Observatory Alliance), <a href="#">ISA-Tab Specification</a> , <a href="#">OBI Ontology</a> , <a href="#">NCBI</a>	OAI-PMH  SWORD  REST	<a href="#">Evaluation services voir annexes rapport 2017</a>  <a href="#">Narcis</a> ( <a href="https://www.narcis.nl/metrics/Language/en">https://www.narcis.nl/metrics/Language/en</a> ) portent sur publications (OA, closed, restricted, type)  <a href="#">Dataverse NL</a> Indicateurs généraux ou par institutions Nombre total de dataverses - publiés ou non - par domaines - créés par trimestre  Nombre total de datasets - publiés ou non - par domaines - créés par trimestre  Fichiers ajoutés par trimestre Nombre total de téléchargement Nombre de téléchargement par trimestre	Projet d'intégration à Dataverse NL	Open Aire, ARDC, Clarivate Citation Index, CORE, DART-Europe, EBSCO, EUDAT B2FIND, BOLAS, Google Scholar, Isidore, Microsoft Academic Research, HU Berlin IBI, OCLC, ProQuest LLC, TDNet, Ex Libris

		<a href="#">Taxonomy for Organisms (see .tsv version)</a> .				
DFDR	Jeux de données : DOI (Datacite)  Auteurs : ORCID  Organisations : ROR  Financeurs : Crossref funder	Métadonnées : Dublin Core, DataCite Metadata Schema, FundRef Registry  Vocabulaire : FAST (Faceted application of subject terminology)	OAI-PMH CKAN	Pas encore en production  Projet d'associer des indicateurs internes à des outils tiers comme Google analytics	Oui	Clarivate Data Citation Index
DNO	Jeux de données : DOI (Datacite), Handle  Auteurs : Orcid	Métadonnées : Schema.org, Dublin Core, DataCite Metadata Schema, DDI - Data Documentation, PROV  Vocabulaires et ontologies : International Virtual Observatory Alliance (IVOA), ISA-Tab  Pas d'ontologies et vocabulaires intégrés (non supportés par défaut par Dataverse)	SWORD OAI-PMH REST	Nombre de téléchargement Nombre de vues  Données d'usage : Google analytics  17101 téléchargement (20200901) soit une moyenne de 22 téléchargements par datasets	Intégration à Dataverse en projet	B2FIND, BASE, DataCite Search, Google Dataset Search, ExLibris Primo CLARIN Virtual Language Observatory pour la collection TROLLing .
JORH	Jeux de données : DOI (Datacite)  Publications : DOI (Crossref)	Métadonnées : JISC Canonical Data Model basé sur Datacite, Premis, Cerif, Dublin-Core	OAI-PMH REST ORH	Nombre de vues Nombre de téléchargements  Intégration des altmétriques sur les notices (projet)	En projet	????

	Auteurs : Orcid					
RADAR	Jeux de données : Handle, DOI  Auteur : Orcid  Financeurs : ISNI, GRID, CrossRef Funder  Archivage : Identifiant Radar	Métadonnées : DataCite Metadata Schema, Dublin Core, ABCD - Access to Biological Collection Data FundRef Registry  Vocabulaires Domaines : liste Gepri/DFG <a href="https://gepris.dfg.de/gepris/OCTOPUS?task=showSearchSimple">https://gepris.dfg.de/gepris/OCTOPUS?task=showSearchSimple</a> Noms géographiques : ISO 3166 Langage : ISO 639-3 FundRef Registry	OAI-PMH REST	Statistiques sur les 6 derniers mois pour chaque notice : nombre de vues et téléchargements par mois	????	DataCite, Google et EUDAT B2FIND
UKDS	Jeux de données : DOI  Auteur : ORCID	Métadonnées : <u>Reshare</u> Dublin Core, Datacite metadata schema, Data Documentation Initiative (DDI)  <u>UKDA</u> Dublin Core, Datacite metadata schema, Data Documentation Initiative (DDI), GEMINI, ISO 19115 for geographic information, Statistical data and metadata exchange (SDMX), Metadata encoding and	OAI-PMH	Nombre de téléchargements (datasets, fichiers, fichiers de données, documentation) Nombre de vues Palmarès des données les plus téléchargées Palmarès des déposants (data creators) Lieux d'origine des requêtes	????	????

		transmission standard (METS), General International standard archival description (ISAD5G))  Vocabulaire : Thésaurus Humanities and Social Science Electronic Thesaurus (HASSET) <a href="https://hasset.ukdataservice.ac.uk/">https://hasset.ukdataservice.ac.uk/</a>				
--	--	--	--	--	--	--

## Annexe 2. Fonctionnalités des services nationaux de données

Services de données nationaux	Service d'accès aux données	Entrepôt	Nombre de dépôts par an	Volume total de l'entrepôt	Taille maximale pour un jeu de données
Australian Research Data Commons (ARDC)	<p><b>Research Data Australia</b> Moteur de recherche fédéré Accès à plus de 160000 datasets Métadonnées envoyées par près de 100 établissements contributeurs (organismes de recherche, agences gouvernementales et institutions culturelles) via logiciel RDA Registry Moissonnage des formats : XML, JSON, XSLT, http get, oai-pmh, ogc catalogue services for the web, CKAN Action API Données pluridisciplinaires</p> <ul style="list-style-type: none"> <li>- 39000 jeux de données</li> <li>- 276 logiciels</li> <li>- plus de 62000 projets</li> </ul>	<p><b>Pas de services d'entrepôt national</b></p> <p>Solutions de stockage et de calcul déployées à la demande des communautés scientifiques</p>	Non concerné	Non concerné	Services de calcul (instances Nectar cloud computing) : 480 Go
Data and Networked Services (DANS)	<p><b>Narcis</b> Moteur de recherche fédéré Accès à publication, données et logiciels 60 sources moissonnées dont 23 entrepôts de données (21 nationaux et 2 internationaux) Accès à plus de 300000 datasets</p> <p>Exemple d'entrepôts de données moissonnés : SURF Data, Cancer Data, NIOZ Royal Netherlands Institute for Sea Research, Université d'Utrecht, et UvA/HvA, entrepôt de logiciels (<a href="https://www.esciencecenter.nl/">https://www.esciencecenter.nl/</a>)</p>	<p><b>Dataverse NL</b> Entrepôt générique de données de recherche regroupant 16 dataverses institutionnels</p> <p><b>Easy</b> Plateforme d'archivage</p>	<p>Dataverse NL (2020/11/15)</p> <ul style="list-style-type: none"> <li>- 2020 : 306</li> <li>- 2019 : 651</li> <li>- 2018 : 202</li> </ul>	<p><b>Dataverse NL</b> <u>Dépôts (2020/11/15)</u> 394 collections, 1525 datasets, 14512 fichiers</p> <p><u>Téléchargements</u> 2019 : 48126 soit une moyenne de 33 téléchargements</p> <p><a href="https://dans.knaw.nl/en/current/news/dans-annual-report-2019">https://dans.knaw.nl/en/current/news/dans-annual-report-2019</a></p> <p><b>Easy</b></p>	Pas de limite en termes de volume Taille max pour chaque fichier : 10Go

				Plus de 150000 jeux de données archivés contre 45000 en 2018. Augmentation liée à la connexion avec les entrepôts institutionnels via interface machine à machine entre le système <a href="#">Questasy de CentERdata</a> et EASY (2019)	
Dataverse Norvège (DNO)	Outil de découverte interne Accès aux données des 10 dataverses hébergés par Dataverse Norvège	<b>Dataverse NO</b> Entrepôt générique de données de recherche regroupant 9 dataverses institutionnels et 1 dataverse dédié à une collection thématique	Dataverse NO (2020/11/15) - 2020 : 167 - 2019 : 371 - 2018 : 201 - 2017 : 37 Près de 500 comptes créés (septembre 2020)	Dataverse NO 794 jeux de données (202060901) Volume total de stockage inférieur à 1 To	Pas de limite en termes de volumes Contacter support si le jeu de données excède les 50Go Téléchargement par fichier de 8Go max
DFDR	Moteur de recherche fédéré Accès à des données produites par des établissements de recherche et des administrations publiques 84 sources moissonnées dont Scholar Dataverse Portal (47 dataverses institutionnels) et Open Data Canada  Total : 117.000 datasets (10/2020) dont 85000 sont moissonnées depuis Open Data Canada, 3000 depuis Scholars Dataverse	Entrepôt générique de données de recherche destiné à accueillir prioritairement des grands ensembles de données. Services de dépôt accessibles aux chercheurs ne disposant d'entrepôt institutionnel	Pas encore en production	Pas encore en production	Taille maximale : 25 000 fichiers ou 300 Go
JORH	Pas d'outil de découverte	Entrepôt générique de données de recherche, de publications et de thèses	Non concerné	Service d'archivage : Moins de 1 téraoctet	Pas de limite en termes de volume

	Service assuré par Core pour les publications et Google Datasets Search pour les données de recherche			(à ce jour, 3 universités clientes)	Taille maximale des fichiers : 2Go
RADAR	Outil de découverte interne	Entrepôt générique de données de recherche	Pas d'informations sur le nombre total de jeux de données	Environ 70 jeux de données dont la moitié sont en accès ouvert	Taille maximale : 600 Go
UKDS/UKDA	<p>Moteur de recherche fédéré</p> <p>Sources :</p> <ul style="list-style-type: none"> <li>- organismes officiels, en particulier le gouvernement central</li> <li>- agences statistiques internationales</li> <li>- projets financés par l'ESRC</li> <li>- agences d'études de marché</li> <li>- sources historiques</li> <li>- archives de données issues de pays étrangers</li> </ul> <p>Accès à plus de 8000 jeux de données 180000 téléchargements par an Mars 2020 : 56500 utilisateurs (déposants et utilisateurs) sur la base d'une connexion lors des deux dernière années</p>	<p>Entrepôt de données relevant des SHS</p> <p>Reshare : entrepôt pour auto-archivage</p> <p>UKDA : entrepôt de données de confiance, gestion sécurisée des accès aux données</p>	Reshare : moyenne de 300 dépôts par an	<u>Reshare</u> : 1828 jeux de données (11/2020) Soit 1,3 Tb	Pas de taille maximale en théorie, uniquement si difficulté de téléchargement (pas d'indication précise sur le volume)

### Annexe 3. Ressources des services nationaux

Services	Nombre ETP dédié à l'ensemble des activités	Budget initial de construction	Budget annuel opérationnel Modèle économique
Australian Research Data Commons (ARDC)	Equipe de 61 personnes soit 58 ETP	Chiffrage impossible car l'ARDC est né de la fusion de différents services préexistants	Financement quinquennal 2019-2023 : plus de 110 millions d'euros, soit 22 millions d'euros par an. Postes de dépenses répartis ainsi : <ul style="list-style-type: none"> <li>- fonctionnement = plus de 67 millions d'euros soit 13 millions par an</li> <li>- investissement = 44 millions d'euros soit 9 millions par an</li> </ul> Financement structurel assuré par le programme national NCRIS Services entièrement gratuits pour les utilisateurs
Data and Networked Services (DANS)	Equipe de 50 personnes soit 40 ETP	Pas évident de chiffrer compte tenu de l'historique du réseau (création successive des 3 services entre 2005 et 2014)	Budget 2017 = 5 millions d'euros (2011 = 3,8 millions d'euros) Rapport d'activité 2017 : <a href="https://bit.ly/32Yo5O6">https://bit.ly/32Yo5O6</a> Budget fluctue en fonction des projets (personnel + infrastructure) Financements structurels = 3,5 millions d'euros <ul style="list-style-type: none"> <li>- 2,7 millions d'euros par KNAW</li> <li>- 0,7 millions d'euros par NWO</li> </ul> Financements projets = 1,5 million d'euros Répartition par sources de financements (2017) <ul style="list-style-type: none"> <li>- Financement KNAW/NWO : 65%</li> <li>- Fonds externes, services payants : 35%</li> <li>- 2011 : 77%/23%</li> </ul> Recherche de financements supplémentaires pour faire face à l'inflation des coûts générés par un volume croissant de données et le développement des services. Besoin de renouveler infrastructure estimé à 2 millions d'euros sur les 4-5 prochaines années, soit 200000 euros/an <u>Coûts ventilés par services :</u> <ul style="list-style-type: none"> <li>- Narcis = 500000 millions d'euros</li> <li>- DNL = 300000 millions d'euros</li> <li>- Easy + coûts généraux = 4,2 millions d'euros</li> </ul> Coûts communs : formation, communication

			<p><u>Dataverse NL</u>  Adhésion annuelle forfaitaire : 5000 euros HT  Facturation des frais de stockage  Objectif : couvrir les coûts liés à l'administration et la gestion avec les frais facturés aux institutions membres</p>
<p>Dataverse Norvège (DNO)</p>	<p>Equipe de 9 personnes soit 3 ETP répartis ainsi :  - 3 personnes dédiées à l'administration fonctionnelle et la gestion de l'entrepôt, soit 2 ETP  - 4 personnes dédiées à l'exploitation technique et la maintenance du dépôt, soit 0,75 ETP  - 1 personne dédiée à l'alignement sur la stratégie de l'UiT, soit 0,1 ETP  - 1 personne dédiée au développement stratégique de l'infrastructure informatique, soit 0,1 ETP</p> <p>Pas de création de poste UiT mais modification de tâches de certains employés</p> <p><u>Moyens mobilisés par les institutions membres de DNO</u>  <b>Gestion de collection</b>  - 0.2 ETP pour la gestion de chaque collection  - 0.3 ETP pour la gestion de la collection thématique TROLLing  <b>Curation</b>  - entre 2 et 4 personnes pour chaque collection</p>	<p>Pas de visibilité précise des coûts.</p> <p>Service DNO s'est construit par étapes successives : ajout d'une couche supérieure, d'un sub-dataverse à l'entrepôt institutionnel Dataverse de l'université de Trömso</p> <p>Majeure partie des coûts sont liés aux ressources humaines nécessaires aux activités de curation, ces dépenses restent à la charge des établissements.</p>	<p>Financement Université de Trömso (pas d'informations sur le montant)</p> <p>Contribution des institutions partenaires : frais généraux fixes (gestion, exploitation et développement de l'entrepôt) et d'une tarification en fonction du volume de stockage</p> <p>Revenus :  - frais unique au démarrage = 2300 euros  - attribution de DOI = entre 450 et 2300 euros en fonction du nombre de DOI  - redevance annuelle = 3700 euros, couvrant les dépenses de UiT pour l'exploitation de Dataverse NO, y compris le serveur/stockage, la maintenance, le développement, etc. Cette redevance comprend le stockage jusqu'à 1 To.  - Pour le stockage supplémentaire, un accord séparé doit être conclu. 92 euros par an pour 1 To supplémentaire, avec une période minimale de 3 ans.</p>

	<ul style="list-style-type: none"> <li>- 8 pour la collection UiT et la collection top level</li> <li>- 2 pour la collection TROLLing</li> </ul> <p>Au total <b>52 personnes soit 11 ETP</b></p>		
Dépôt Fédéré des Données de la Recherche (DFDR)	<p>2016 : projet pilote</p> <ul style="list-style-type: none"> <li>- équipe de développement (3 ETP)</li> </ul> <p>2020 :</p> <ul style="list-style-type: none"> <li>- 6 développeurs (5,5 ETP)</li> <li>- 8 postes pour RDM et curation, 1 chef de projet et 1 chef de service (??? ETP)</li> </ul>	<p>Pas de visibilité sur la totalité des frais de lancement</p> <p>Achat de 1,6Pb lors du lancement + dépenses liés à achat de matériels (serveurs dédiés)</p>	<p>Modèle économique en cours de validation</p> <p>Financement structurel qui provient du gouvernement fédéral du Canada</p>
JORH	<p>Equipe de 9 personnes soit 9 ETP répartis ainsi :</p> <ul style="list-style-type: none"> <li>- 1 responsable de produit</li> <li>- 1 manager</li> <li>- 3 personnes pour la gestion des produits (entrepôt, archivage)</li> <li>- 1 responsable de service (contrat, helpdesk, satisfaction des clients)</li> <li>- 3 développeurs, descendant à 1,5 plus tard dans l'année pour la maintenance de ce système</li> </ul>	<p>2015-2018 : programme de 5,5 millions d'euros financé par les UK Research Councils</p> <p>Coût du système Archivemata estimé à 1,1 million d'euros</p>	<p>Budget annuel : Information confidentielle</p> <p>Depuis 2019, plus de soutien financier institutionnel JORH géré comme un service commercial : budget dépend des revenus versés par les utilisateurs</p> <p>3 offres de services (dépôt, archivage, dépôt/archivage)</p> <p><u>Tarification des services</u> (en fonction de la taille de l'université)</p> <p>A minima 7000 livres pour l'abonnement à un service de dépôt (1To)</p> <p>Au maximum 50000 livres pour un couplage dépôt (1To) / système de préservation</p> <p>Frais supplémentaires :</p> <ul style="list-style-type: none"> <li>· Stockage : 300 livres/ To supplémentaires / an (facturés par Amazon)</li> </ul>

			<ul style="list-style-type: none"> <li>Intégration personnalisée si nécessite un développement</li> </ul> <p>Objectif : dégager des profits d'ici 3 ans</p>
RADAR	<p>2 ETP répartis ainsi :</p> <ul style="list-style-type: none"> <li>- 1 personne dédiée à la gestion de produits à plein temps, soit 1 ETP</li> <li>- 0,25 ETP pour l'administration et la maintenance du système</li> <li>- 0,75 ETP pour le développement et la maintenance de logiciels.</li> </ul>	Information confidentielle	<p>Estimation annuelle de fonctionnement du service est de l'ordre de 300000 euros</p> <p>Financement de RADAR repose pour moitié sur le financement de l'Institut de Technologie de Karlsruhe (KIT) qui prend en charge la moitié des coûts fixes opérationnels à partir de son propre financement de base.</p> <p>Financement de 2 consortiums NFDI auxquels appartient Radar par le gouvernement fédéral et les régions</p> <p>Le reste des coûts fixes et tous les coûts d'exploitation variables sont couverts par les frais d'utilisation.</p> <p>Les contrats entre chaque institution partenaire et KIT prévoient une durée minimale de trois ans.</p> <p>L'adhésion annuelle va de 595 euros à 11900 euros si le volume de stockage demandé excède les 50 To.</p> <p>Redevance annuelle varie en fonction de l'utilisation (archivage, publication ou stockage temporaire), elle est calculée selon :</p> <ul style="list-style-type: none"> <li>- la taille de l'espace de stockage temporaire mis à disposition</li> <li>- le volume de toutes les données archivées depuis le début de la durée du contrat, moins les données qui ont été supprimées en raison de périodes de stockage expirées</li> <li>- la nature des données : archivées (0,46 euros/Gb), stockées temporairement (0,43) ou publiées (7,56)</li> </ul> <p>Facturation de services liés au développement de fonctionnalités spécifiques commandées par une institution utilisatrice (ex. prise en charge de métadonnées spécifiques à une discipline)</p>

UKDS/UKDA	Equipe regroupant plus de 50 personnes, soit 35 ETP	Pas possible de chiffrer car UKDS s'est construit sur les actifs des précédents services de données (1967-...) Dès son lancement, l'UKDS a disposé des moyens mis en place par l'Economic and social data service (personnel, processus, procédures et infrastructure technique)	3 millions de Livres sterling par an Coûts économiques complets qui comprennent le reversement financier aux universités hôtes  Financement structurel de l'Economic and Social Research Council (ESRC), défini lors de contrats quinquennaux.  Revenus indirects provenant de départements universitaires auxquels l'UKDS est associé  Gratuité d'accès aux services
-----------	---	---	--

## Annexe 4. Fair-Aware, outil d'auto-évaluation aux principes FAIR développé par DANS NL, Digital Curation Center et l'université de Brême<sup>40</sup>

<p><b>FINDABLE</b></p> <p>1. Are you aware that a dataset should be assigned a globally unique and persistent identifier when deposited with a data repository? <input type="radio"/> Yes <input type="radio"/> No ⓘ</p> <p>2. Are you aware that when you deposit a dataset with a repository, you will need to provide some details (known as discovery metadata) in order to make the data findable, understandable and reusable to others? <input type="radio"/> Yes <input type="radio"/> No ⓘ</p> <p>3. Are you aware that the repository providing access to your dataset should make the metadata describing your datasets available in a format readable by machines as well as humans? <input type="radio"/> Yes <input type="radio"/> No ⓘ</p>	<p><b>Feedback</b> ↶</p> <p>Please answer the four questions below to help us improve the tool and make it even more relevant for the community.</p> <ul style="list-style-type: none"> <li>Are any of the issues covered in this assessment difficult to understand and/or answer?             <ol style="list-style-type: none"> <li>Globally unique persistent identifier (PID) ⓘ <input type="checkbox"/></li> <li>Metadata for citation and discovery including PID ⓘ <input type="checkbox"/></li> <li>Metadata available online is readable by humans and machines ⓘ <input type="checkbox"/></li> <li>Metadata includes licence, level of access and conditions to access the data ⓘ <input type="checkbox"/></li> <li>Persistence of metadata ⓘ <input type="checkbox"/></li> <li>Use of controlled vocabularies in metadata ⓘ <input type="checkbox"/></li> <li>Metadata includes provenance ⓘ <input type="checkbox"/></li> <li>Community-endorsed metadata ⓘ <input type="checkbox"/></li> <li>Data in a preferred format for reuse &amp; preservation ⓘ <input type="checkbox"/></li> <li>Digital curation and preservation ⓘ <input type="checkbox"/></li> </ol> </li> <li>Are there any issues relevant to your discipline and/or needed for enabling reusability of your dataset that are missing from this assessment? (Please do not include any personal data, such as your name and email address in your response.)             <div style="border: 1px solid #ccc; height: 30px; margin-top: 5px;"></div> </li> <li>Please submit any other feedback on how we might improve the FAIR-Aware assessment tool, including possible additional guidance? (Please do not include any personal data, such as your name and email address in your response.)             <div style="border: 1px solid #ccc; height: 30px; margin-top: 5px;"></div> </li> <li>Undertaking this assessment helped raise my awareness and my understanding of the FAIR data Principles.             <p><input type="radio"/> Strongly Agree <input type="radio"/> Agree <input type="radio"/> Neutral <input type="radio"/> Disagree <input type="radio"/> Strongly Disagree</p> </li> </ul> <p style="text-align: right;"></p>
<p><b>ACCESSIBLE</b></p> <p>4. Are you aware that access to your dataset may need to be controlled and that metadata should include licence information under which the data can be reused? <input type="radio"/> Yes <input type="radio"/> No ⓘ</p> <p>5. Are you aware that metadata should remain available over time, even if the data is no longer accessible? <input type="radio"/> Yes <input type="radio"/> No ⓘ</p>	
<p><b>INTEROPERABLE</b></p> <p>6. Are you aware that the metadata describing your datasets should use controlled vocabularies? <input type="radio"/> Yes <input type="radio"/> No ⓘ</p>	
<p><b>REUSABLE</b></p> <p>7. Are you aware that provenance information about the collection and/or generation of data should be included in the metadata? ⓘ <input type="radio"/> Yes <input type="radio"/> No</p> <p>8. Are you aware that metadata describing your data should follow the specifications of a community-endorsed standard? ⓘ <input type="radio"/> Yes <input type="radio"/> No</p> <p>9. Are you aware that data should be deposited preferably in a file format that is open – to support reuse – and supported by the repository for long-term preservation? ⓘ <input type="radio"/> Yes <input type="radio"/> No</p> <p>10. Are you aware that maintaining your dataset FAIR over time requires professional data curation and preservation? ⓘ <input type="radio"/> Yes <input type="radio"/> No</p>	

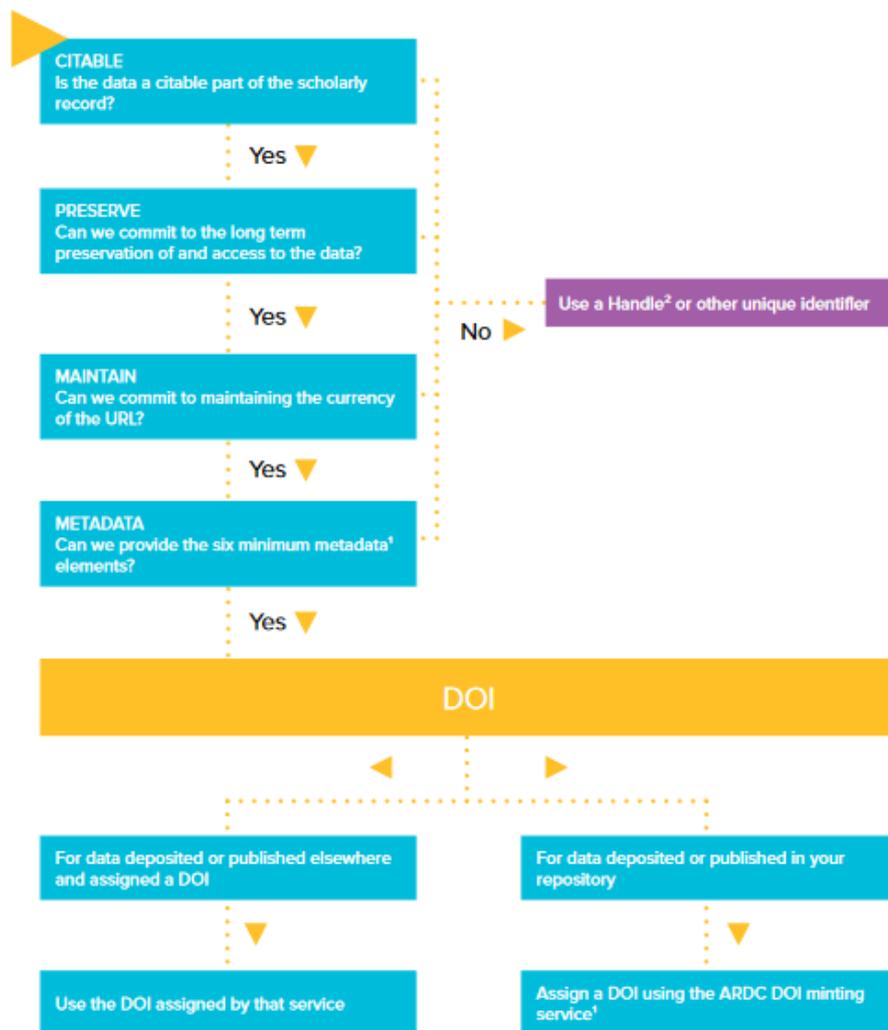
<sup>40</sup> <https://fairaware.dans.knaw.nl/>

# Annexe 5. Arbre de décision DOI développé par le service DOI de l’Australian Research Data Commons (ARDC)<sup>41</sup>

## Digital Object Identifiers (DOIs)

### A pathfinder for data managers

Australian publicly funded research organisations and government agencies may use the ARDC DOI service<sup>1</sup> to mint DOIs for data. Data includes data and associated workflows, software, models, and grey literature.



<sup>41</sup> Capture d’écran réalisée à partir de : [https://ardc.edu.au/wp-content/uploads/2019/05/DOIs\\_decision\\_tree.pdf](https://ardc.edu.au/wp-content/uploads/2019/05/DOIs_decision_tree.pdf)

# Annexe 6. Exemple de résultat proposé par Qualibank, banque de données qualitatives proposée par l'UKDS<sup>42</sup>

## Interview with Sara Arber

---

[External resources](#) | [Texts](#) | [Files](#)
[Access this collection from Discover](#)

---

### DETAILS ^

**Collection ID (SN):** 6226

**Title:** [Pioneers of Social Research, 1996-2018](#)

**Principal investigator:** Thompson, P., University of Essex, Department of Sociology

**Sex:** Female

**Age group:** 65-74

**Socio-economic status:** Higher managerial/admin/professional

**Region:** South East

**Licence:** [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Licence.](#)



**Additional metadata**

**Idno:** 6226int045

**No. pages:** 142

**Placename:** University of Surrey, Surrey

**I forename:** Blanche

**I surname:** Girouard

**R title:** Ms.

**R forename:** Sara

**R surname:** Arber

---

### INTERVIEW TRANSCRIPT ^

Interview conducted 26 April 2017 at Department of Sociology, University of Surrey

**I:** Can you just give your name and your date of birth.

**R:** My name is Sara Lynne Arber, and I was born on the 19th March, 1949.

**I:** And where were you born?

**R:** I was born in Chingford, Essex. I'm an Essex girl!

**I:** In a hospital?

**R:** No, I was born, I came - I've always been awkward throughout my life, so I came rather too quickly, and my mum said, "Oh, I think something's happening", and apparently, in the bathroom, I appeared! And so my dad and grandmother did whatever, and I was booked into a Nursing Home, so my father carried me into the Nursing Home the next day, because I'd been booked in for ten days with my mother. But I wasn't born in there, I was born in my grandmother's house in Chingford.

**I:** In the bathroom.

**R:** Yes!

**I:** How amazing!

**R:** Well, whatever! (LAUGHS)

**I:** Tell me about your parents. I'm just going to take this off you. And also, can we take your ring off because it's going -

**R:** I can't.

**I:** Oh, can't you? Okay. Fine.

### RELATED RESOURCES

**Related external resources**

- [XML TEI Transcript](#)
- [XML QuDEx metadata](#)
- [Web resource: Yoube list for Sara Arber](#)
- [Web Resource: Wikipedia page for Sara Arber](#)
- [Web Resource: Book Reference for "Gender and Later Life "](#)
- [Web Resource: Book Reference for "Ageing, Independence, and the Life Course"](#)
- [Web Resource: Book Reference for "Gender and Ageing: Changing Roles and Relationships"](#)

**Related texts**

- [Summary of the interview with Sara Arber](#)

**Related files**

- [Thematic Highlights: Interview Extracts for Sara Arber](#)

<sup>42</sup> <https://discover.ukdataservice.ac.uk//QualiBank/Document/?id=q-ddda86f6-b4c3-4e37-ba8c-6273b87eee9d&q=arber#relatedexternalresources>

## Annexe 7. Captures d'écran extraites du Data costing tool développé par l'UKDS<sup>43</sup>

ACTIVITY	COMMENTS AND SUGGESTIONS	√	COST
<b>Anonymisation</b> <ul style="list-style-type: none"> <li>Do you need to remove identifying information or conceal the identity of participants (e.g. using pseudonyms) before data can be shared?</li> <li>Anonymisation needs to be consistent throughout a data collection.</li> </ul>	<ul style="list-style-type: none"> <li>for quantitative data (e.g. survey data) – low cost if identifiers are a priori excluded from data files, are easy to remove, or identifiable variables are coded to avoid disclosure; cost may be higher if variables need recoding afterwards to avoid disclosure</li> <li>for qualitative textual data (e.g. interview transcripts) – may be high cost as entire texts will need to be read and checked for identifying information; costs can be reduced if anonymisation is carried out during transcription (or at least highlighted during transcription)</li> <li>for audio-visual data – anonymising/editing voices or faces can be very costly and reduces the usefulness of data</li> <li>cost depends on how sensitive or complex data are and how much identifying information is recorded in the data – if only removal of names is required, cost is low; pseudonymisation will require more time</li> <li>if anonymisation is planned before data collection or transcription/digitisation – cost can be lowered</li> <li>for all files, check file properties and edit to remove disclosive information such as editor/author name</li> </ul>		
<b>Copyright</b> <ul style="list-style-type: none"> <li>Do other parties hold copyright in the data?</li> <li>Do you need to seek copyright clearance before sharing data?</li> </ul>	<ul style="list-style-type: none"> <li>is time required to seek copyright clearance?</li> <li>is legal advice required?</li> </ul>		
<b>Data sharing</b> <ul style="list-style-type: none"> <li>Will your data be deposited with a data centre or institutional repository?</li> <li>Which requirements exist to prepare data to particular standards e.g. regarding documentation or format?</li> <li>Will journal publishers require deposit of data supporting article findings?</li> </ul>	<ul style="list-style-type: none"> <li>consider the cost of data deposit and/or longer-term storage – find out from data centre/repository/journal whether charges apply</li> <li>cost in time and effort needed to prepare the data for sharing and preservation</li> </ul>		
<b>Roles and responsibilities</b> <ul style="list-style-type: none"> <li>Do you need to allocate roles and responsibilities for various data management activities?</li> </ul>	<ul style="list-style-type: none"> <li>if multiple partner institutions, researchers or funders are involved in research – consider cost of data management planning meetings or discussions</li> </ul>		
<b>Operationalising data management</b> <ul style="list-style-type: none"> <li>What measures are needed to implement and operationalise data management throughout the research lifecycle?</li> </ul>	<ul style="list-style-type: none"> <li>do you need extra time and resources to implement data management throughout your research, e.g. regular team meetings, setting up a collaborative research environment?</li> <li>if staff training is required - higher cost</li> <li>do you need a dedicated data manager?</li> </ul>		

<sup>43</sup> <https://ukdataservice.ac.uk/media/622368/costingtool.pdf>

# Annexe 8. QA My Data, outil de curation développé par l'UKDS<sup>44</sup>

## Basic file checks

File opens	Checks whether acceptable format
Bad filename check, regular expression via RegEx pattern	Regex requires quotes "[a-z]". To use a special characters, e.g. a backslash (\) a backslash before is required e.g. \\

## Metadata checks

Report on number of cases and variables	Always run
Count of grouping variables	
Missing variable labels	Must be set to true . If set to false the test will not run
No label for user defined missing values e.g. - 9 not labelled	SPSS only
'Odd' characters in variable names and labels	User specifies the characters
'Odd' characters in value labels	User specifies the characters
Maximum length of variable labels, e.g. >79 characters	User specifies the length
Maximum length of value labels, e.g. >39 characters	User specifies the length
Spelling mistakes (non-dictionary words) in variable labels using a dictionary file	User specifies a dictionary file.
Spelling mistakes (non-dictionary words) in value labels using a dictionary file	User specifies a dictionary file

## Data integrity checks

Report number of numeric and string variables	
Check for duplicate IDs	User specifies the variables. Multiple variables can be added on new lines e.g. - Caseno - AnotherVariableHere
'Odd' characters in string data	User specifies the characters
Spelling mistakes (non-dictionary words) in string data using a dictionary file (can check if date format set correctly!)	User specifies a dictionary file
Percentage of values missing ('Sys miss' and undefined missing)	User sets the threshold, e.g. more than 25%

## Disclosure control checks

Identifying disclosure risk from unique values or low thresholds (frequencies of categorical vars or minimum values)	User sets the threshold value, e.g. 5
Direct identifiers using a RegEx pattern search	User runs separately for postcodes, telephone numbers etc. Advise tests are separately as may be resource intensive
Direct identifiers/named entities in string data using a dictionary file (to be added)	Specify a dictionary file containing lists of stop words or named entities e.g. for places, names etc. Advise tests are separately as may be resource intensive

## Health check report

QAMyData

**teaching-data%set.sav**

Raw Case Count: 10210  
Aggregated Case Count: 0  
Total Variables: 188  
Data Type Occurrences: Numeric: 186, String: 2  
Created At: 2019-02-18 13:37:39  
Last modified at: 2019-02-18 13:37:39  
File Label:  
File Format Version: 2  
File Encoding: WINDOWS-1252  
Compression type: Rows

**Basic File Checks**

Name	Status (N)	Description
Bad file name	failed (1)	File name should match the user specified pattern

**Metadata Checks**

Name	Status (N)	Description
Missing variable labels	failed (8)	Variables should have a label
Variable odd characters	failed (2)	Variable names and labels should not contain the specified characters ["&","%","'","@","!","'","'","'","'","'"]
Variable label max length	failed (6)	Variable labels should not exceed the defined number of characters (79 characters)



# (limited to 1000)	Variable	Row number
1	OwinTV	-
2	V137	-

<sup>44</sup> <https://www.ukdataservice.ac.uk/about-us/our-rd/qamydata.aspx>

## Annexe 9. Forces et faiblesses du modèle économique de Dataverse Norvège (2020)<sup>45</sup>

Table 7: Main strengths and weaknesses of the DataverseNO business model

Funding source		Applicable to DataverseNO	
Institutional funding	Pros	Compatible with open data principles	Yes
		Longer-term stability	Yes
		Convergence of interest between host and repository	Yes
	Cons	Cost optimisation	Yes
		Close to researchers	Yes
		Fragmentation of domain data	To some extent; but considerable support for domain data
Deposit-side contract	Pros	Lower levels of curation	No
		Lower interoperability	To some extent; but continuous improvement
		Compatible with open data principles	Yes
Project funding	Pros	Economies of scale	Not relevant
		Cost optimisation	Yes
		Unpredictable funding	To some extent; but no threat to sustainability
	Cons	High administrative costs	No
		Limited engagement with users	No
		Support for innovation and development	Yes
Project funding	Pros	Compatible with open data principles	Yes
		Short-term nature	No
	Cons	Lack of flexibility	No
		Diversion of staff and effort	To some extent; but may be prevented by hiring of temporary staff

<sup>45</sup> Graphiques extraits de “DataverseNO: A National, Generic Repository and its Contribution to the Increased FAIRness of Data from the Long Tail of Research” par Philipp Konzett, <https://doi.org/10.7557/15.5514>

# Annexe 10. Etat des lieux de l'implémentation des principes FAIR au sein de Dataverse Norvège (2020)<sup>46</sup>

Table 2: The implementation of Findability in Dataverse and its adoption in DataverseNO. Adapted from Crosas (2020).

Principle	Implementation in Dataverse	Applied in DataverseNO
F1	Support for DOI and Handle	Yes (DOI)
	Always at the dataset level	Yes
	Optionally at file level	Yes
F2	Metadata standards in human- and machine-readable formats: Dublin Core; Documentation Data Initiative (DDI); DataCite; Schema.org	Yes
	Optional custom metadata	No
F3	Dataset PID is part of metadata record presented on Dataset landing page.	Yes
	File PID is part of metadata record presented on File landing page.	Yes
	PIDs are included in exported metadata files.	Yes
F4	DataCite metadata is harvested and indexed by DataCite Search.	Yes. In addition: B2FIND and VLO.
	Schema.org metadata is indexed by Google Dataset Search.	Yes

Table 3: The implementation of Accessibility in Dataverse and its adoption in DataverseNO. Adapted from Crosas (2020).

Principle	Implementation in Dataverse	Applied in DataverseNO
A1	Yes	Yes
A1.1	Support for HTTP (W3C), Rsync over ssh (GNU General Public license)	Yes
	RESTful API (e.g., access through cURL)	
A1.2	Authentication API Tokens	Yes
	Authorization service	Yes, but only for embargo. By default, DataverseNO only accepts open data.
A2	A deaccessioned dataset (data not available) is still findable and citable.	Yes
	Metadata includes information about why the data are not available.	Yes

Table 4: The implementation of Interoperability in Dataverse and its adoption in DataverseNO. Adapted from Crosas (2020).

Principle	Implementation in Dataverse	Applied in DataverseNO
I1	Linked data support with JSON-LD for Schema.org	Partially, for general attributes such as file type
	DDI (XML) as a rich schema to support extensive variable metadata	Partially/in some datasets
I2	FAIR controlled vocabularies and data models may be deployed manually, e.g. in well-curated datasets	Partially/in some datasets
	Custom metadata and metadata template can help.	Partially/in some datasets (metadata template)
	Controlled vocabularies and ontologies not supported by default. But, cf. ongoing work on support for some domains (e.g. SSHOC).	No
I3	DDI schema supports references to other data.	Yes, where applicable
	Not yet supported: structured metadata about related objects included in exported DataCite metadata (coming soon)	No

Table 5: The implementation of Reusability in Dataverse and its adoption in DataverseNO. Adapted from Crosas (2020).

Principle	Implementation in Dataverse	Applied in DataverseNO
R1		
R1.1	Included in metadata: data use license/waiver; data access and use terms. But, licenses other than CC0 are not predefined and by default not machine-readable.	Yes. Almost all datasets are published under default license CC0.
	By default no support for explicit information about metadata license	Terms for reuse of metadata described on website
R1.2	Rich citation metadata including information about data authors and other contributors, providers, distributors, related data (input data)	Yes
	Versions with changes documented automatically	Yes
	W3C PROV support	No
R1.3	DDI for social science data	Partially/in some datasets
	Metadata blocks for other community standards	Partially/in some datasets
	Ongoing work on support for more domains.	No
	Custom metadata	No
	FITS for astronomy data	N/A (so far)
	File format conversion to reusable formats (tabular)	Partially/in some datasets
		Data in preferred file formats Datasets include ReadMe file.

Table 6: Summary of current FAIR implementation or support in Dataverse and DataverseNO

	F				A			I			R		
	1	2	3	4	1.1	1.2	2	1	2	3	1.1	1.2	1.3
Dataverse	Green	Yellow	Green	Green	Green	Green							
DataverseNO	Green	Yellow	Green	Green	Green	Green							

- = (more or less) full implementation or support
- = partial implementation or support
- = (more or less) lacking implementation or support
- = not applicable

<sup>46</sup> Graphiques extraits de "DataverseNO: A National, Generic Repository and its Contribution to the Increased FAIRness of Data from the Long Tail of Research" par Philipp Conzett, <https://doi.org/10.7557/15.5514>

# Table des matières

<b>PRÉAMBULE</b> .....	<b>2</b>
<b>SOMMAIRE</b> .....	<b>4</b>
<b>GLOSSAIRE</b> .....	<b>7</b>
<b>SIGLES</b> .....	<b>9</b>
<b>RÉSUMÉ</b> .....	<b>12</b>
<b>1 GOUVERNANCE</b> .....	<b>13</b>
1.1 AUSTRALIE - AUSTRALIAN RESEARCH DATA COMMONS (ARDC).....	13
Né de la fusion de trois infrastructures préexistantes, un service piloté par l'agence de financement national et des représentants d'université dont l'objectif est de rendre librement accessible un volume maximal de données à tout type de public.....	13
1.2 NORVÈGE - DATAVERSE NORWAY (DNO).....	14
Une infrastructure nationale confiée à l'université de Trömso, des instances consultatives regroupant l'ensemble des universités partenaires.....	14
1.3 PAYS-BAS - DUTCH ARCHIVED AND NETWORKED SERVICE (DANS).....	15
Un institut national relevant de l'Académie des sciences néerlandaises et de l'agence de financement national dont l'objectif est de proposer un accès pérenne aux productions scientifiques.....	15
1.4 CANADA - DÉPÔT FÉDÉRÉ DES DONNÉES DE RECHERCHE.....	16
« Fournir une plateforme unique à partir de laquelle les données de recherche peuvent être chargées, organisées, préservées, explorées, citées et partagées » : un projet mené conjointement par l'association des bibliothèques de recherche canadiennes et l'infrastructure nationale de calcul.....	16
1.5 GRANDE-BRETAGNE - JISC OPEN RESEARCH HUB (JORH).....	18
Pilotée par le consortium national des services d'information du Royaume-Uni, une plateforme de services dont l'objectif est d'apporter des solutions facilement adaptables aux environnements techniques des universités partenaires.....	18
1.6 ALLEMAGNE - RESEARCH DATA REPOSITORY (RADAR).....	19
Une infrastructure distribuée visant à promouvoir la découverte et la réutilisation de données de longue traîne en renforçant leur traçabilité, leur reproductibilité et leur transparence.....	19
1.7 GRANDE-BRETAGNE - UNITED KINGDOM DATA SERVICE (UKDS) .....	20
Service national des données de sciences sociales distribué entre une université hôte (université d'Essex) et des institutions partenaires.....	20
<b>2 OFFRE DE SERVICES</b> .....	<b>22</b>
2.1 PRODUCTEURS DE DONNÉES.....	22
2.1.1 <i>Dépôt de données</i> .....	22
2.1.2 <i>Archivage pérenne des données</i> .....	22
2.1.3 <i>Stockage des données</i> .....	23
2.1.4 <i>Aide à la gestion des données</i> .....	23
2.1.5 <i>Accès réservé aux relecteurs scientifiques</i> .....	24
2.1.6 <i>Soutien financier dédié à la gestion et à l'ouverture de données</i> .....	24
2.1.7 <i>Gestion des accès</i> .....	25
2.2 UTILISATEURS DES DONNÉES .....	25
2.2.1 <i>Outil de découverte interne versus portail de recherche fédéré</i> .....	25
2.2.2 <i>Services de calcul, d'analyse et de collaboration</i> .....	27
2.2.3 <i>Services d'accès aux données sensibles</i> .....	27
2.3 RÉSEAUX D'EXPERTISE ET D'ACCOMPAGNEMENT.....	28
<b>3 ORGANISATION ET MUTUALISATION DES SERVICES</b> .....	<b>29</b>

3.1	ARDC : FACILITER L'ÉMERGENCE DE COMMUNAUTÉS DE PRATIQUES ET LE DÉVELOPPEMENT D'OUTILS PARTAGÉS .....	29
3.2	DANS : DES SERVICES DISTRIBUÉS AU SEIN D'UN RÉSEAU DE PARTENAIRES .....	30
3.3	DATAVERSE NO : UN ENTREPÔT ADMINISTRÉ PAR LES SERVICES D'UNE UNIVERSITÉ, DES COLLECTIONS GÉRÉES PAR LES UNIVERSITÉS PARTENAIRES.....	30
3.4	DFDR : DES SERVICES FÉDÉRÉS ALLIANT INFRASTRUCTURES DE DONNÉES, UNIVERSITÉS ET RÉSEAU DE BIBLIOTHÉCAIRES .....	31
3.5	JORH : DES PRESTATIONS LOGICIELLES INTÉGRÉES AUX ENVIRONNEMENTS TECHNIQUES DES INSTITUTIONS CLIENTES .....	32
3.6	RADAR : DES SERVICES DISTRIBUÉS ENTRE ÉTABLISSEMENTS PARTENAIRES.....	32
3.7	UKDS : UN SERVICE DISTRIBUÉ SOUS LA FORME DE CONTRATS DE SOUS-TRAITANCE ENTRE L'UNIVERSITÉ D'ESSEX ET LES INSTITUTIONS PARTENAIRES.....	33
<b>4</b>	<b>MODÈLE ÉCONOMIQUE.....</b>	<b>34</b>
4.1	ARDC : UNE PLATEFORME DE SERVICES ENTIÈREMENT FINANCÉE PAR LE GOUVERNEMENT AUSTRALIEN .....	34
4.2	DANS : UN MODÈLE ALLIANT FINANCEMENT STRUCTUREL, SOUTIEN AUX PROJETS ET CONTRIBUTION DES ÉTABLISSEMENTS PARTENAIRES .....	34
4.3	DATAVERSE NO : UN SERVICE FINANCÉ PAR L'UNIVERSITÉ HÔTE ET LA CONTRIBUTION DES ÉTABLISSEMENTS PARTENAIRES .....	35
4.4	DFDR : UN MODÈLE EN COURS DE DÉFINITION .....	35
4.5	JORH : MODÈLE ÉCONOMIQUE BASÉ SUR L'ABONNEMENT À DES OFFRES DE SERVICE .....	35
4.6	RADAR : FINANCEMENT RÉPARTI À PARTS ÉGALES ENTRE L'ÉTABLISSEMENT D'ACCUEIL ET DES INSTITUTIONS PARTENAIRES .....	36
4.7	UKDS : FINANCEMENT STRUCTUREL DE L'ESRC ET DES ÉTABLISSEMENTS PARTENAIRES .....	36
<b>5</b>	<b>SYNTHÈSE.....</b>	<b>37</b>
5.1	DEGRÉ DE MATURITÉ DES SERVICES.....	37
5.2	PRIORITÉS AFFICHÉES PAR LES DISPOSITIFS .....	38
5.2.1	<i>Quels objectifs ?</i> .....	38
5.2.2	<i>Pour quels publics ?</i> .....	39
5.2.3	<i>Offre de services</i> .....	41
5.3	DES SERVICES CO-CONSTRUITS ET DÉVELOPPÉS AVEC L'ENSEMBLE DES ACTEURS .....	46
5.3.1	<i>Avec les partenaires institutionnels</i> .....	46
5.3.2	<i>Au sein de réseaux nationaux et internationaux</i> .....	46
5.3.3	<i>En sollicitant les communautés de recherche</i> .....	47
5.4	MODÈLES ÉCONOMIQUES.....	48
5.5	ECONOMIES D'ÉCHELLE .....	49
5.6	PÉRENNITÉ DU SERVICE.....	52
5.6.1	<i>Evaluation des services</i> .....	52
5.6.2	<i>Démarches de certification</i> .....	53
	<b>CONCLUSION.....</b>	<b>55</b>
	<b>S'INSÉRER DE MANIÈRE PRAGMATIQUE ET AGILE AU SEIN D'ÉCOSYSTÈMES NATIONAUX ET INTERNATIONAUX DYNAMIQUES.....</b>	<b>55</b>
	<b>SITES WEB UTILES.....</b>	<b>58</b>
	<b>BIBLIOGRAPHIE.....</b>	<b>59</b>
	<b>ANNEXES .....</b>	<b>64</b>
	ANNEXE 1. IDENTIFIANTS, MÉTADONNÉES ET INDICATEURS UTILISÉS PAR LES SERVICES NATIONAUX DE DONNÉES .....	65
	ANNEXE 2. FONCTIONNALITÉS DES SERVICES NATIONAUX DE DONNÉES.....	70

ANNEXE 3. RESSOURCES DES SERVICES NATIONAUX.....	73
ANNEXE 4. FAIR-AWARE, OUTIL D'AUTO-ÉVALUATION AUX PRINCIPES FAIR DÉVELOPPÉ PAR DANS NL, DIGITAL CURATION CENTER ET L'UNIVERSITÉ DE BRÊME .....	78
ANNEXE 5. ARBRE DE DÉCISION DOI DÉVELOPPÉ PAR LE SERVICE DOI DE L'AUSTRALIAN RESEARCH DATA COMMONS (ARDC) .....	79
ANNEXE 6. EXEMPLE DE RÉSULTAT PROPOSÉ PAR QUALIBANK, BANQUE DE DONNÉES QUALITATIVES PROPOSÉE PAR L'UKDS .....	80
ANNEXE 7. CAPTURES D'ÉCRAN EXTRAITES DU DATA COSTING TOOL DÉVELOPPÉ PAR L'UKDS .....	81
ANNEXE 8. QA MY DATA, OUTIL DE CURATION DÉVELOPPÉ PAR L'UKDS .....	82
ANNEXE 9. FORCES ET FAIBLESSES DU MODÈLE ÉCONOMIQUE DE DATAVERSE NORVÈGE (2020) .....	83
ANNEXE 10. ÉTAT DES LIEUX DE L'IMPLÉMENTATION DES PRINCIPES FAIR AU SEIN DE DATAVERSE NORVÈGE (2020) .....	84
<b>TABLE DES MATIÈRES .....</b>	<b>85</b>
<b>TABLE DES ILLUSTRATIONS .....</b>	<b>88</b>

# Table des illustrations

TABLEAU 1. SCHÉMA DE GOUVERNANCE DE L'ARDC (AUSTRALIE) .....	14
TABLEAU 2. SCHÉMA DE GOUVERNANCE DE DATAVERSE NORVÈGE .....	15
TABLEAU 3. SCHÉMA DE GOUVERNANCE DU RÉSEAU DANS (PAYS-BAS) .....	16
TABLEAU 4. SCHÉMA DE GOUVERNANCE DU DFDR (CANADA) .....	18
TABLEAU 5. SCHÉMA DE GOUVERNANCE DE RADAR (ALLEMAGNE) .....	20
TABLEAU 6. ORGANIGRAMME DE L'UKDS (GRANDE-BRETAGNE) .....	21
TABLEAU 7. NARCIS PID GRAPH.....	26
TABLEAU 8. FONCTIONNEMENT DU RESEARCH VOCABULARIES AUSTRALIA.....	28
TABLEAU 9. DEGRÉ DE MATURITÉ DES DISPOSITIFS .....	37
TABLEAU 10. REPRÉSENTATION SOUS FORME DE RADAR DES PRINCIPAUX OBJECTIFS AFFICHÉS PAR LES DISPOSITIFS ANALYSÉS .....	39
TABLEAU 11. REPRÉSENTATION SOUS FORME DE RADAR DES CATÉGORIES DE PUBLICS VISÉS PAR LES SERVICES NATIONAUX ÉTUDIÉS.....	40
TABLEAU 12. CYCLE DE VIE DES DONNÉES, CAPTURE RÉALISÉE À PARTIR DU SITE IRD DATA.....	41
TABLEAU 13. ÉTAPES DU CYCLE DE VIE DES DONNÉES COUVERTES PAR LES SERVICES NATIONAUX DE DONNÉES (CUMUL DES DIFFÉRENTS SERVICES) .....	41
TABLEAU 14. ÉTAPES DU CYCLE DE VIE DES DONNÉES PAR DISPOSITIFS .....	42
TABLEAU 15. SERVICES PROPOSÉS (CUMUL DES DIFFÉRENTS SERVICES) .....	43
TABLEAU 16. SERVICES PROPOSÉS PAR CHAQUE SERVICE NATIONAL .....	44
TABLEAU 17. NATURE DES INFORMATIONS CONTENUES DANS LES PORTAILS D'ACCÈS (CUMUL DES DIFFÉRENTS SERVICES).....	44
TABLEAU 18. NATURE DES INFORMATIONS CONTENUES DANS LES PORTAILS D'ACCÈS DES DIFFÉRENTS SERVICES.....	45
TABLEAU 19. SOURCES DES DONNÉES CONTENUES DANS LES PORTAILS .....	45
TABLEAU 20. SOURCES DE FINANCEMENT DES SERVICES NATIONAUX.....	49
TABLEAU 21. IMPACT DIRECTS ET INDIRECTS DE L'ECONOMIC AND SOCIAL DATA SERVICE.....	51
TABLEAU 22. TABLEAU 20. INDICATEURS LIÉS AUX OBJECTIFS STRATÉGIQUES DU RÉSEAU DANS .....	53