



HAL
open science

Données ouvertes de la recherche : nouvelles pratiques de publication et de partage. Cours Enssib

Pascal Aventurier

► To cite this version:

Pascal Aventurier. Données ouvertes de la recherche : nouvelles pratiques de publication et de partage. Cours Enssib. Master. MASTERS ENSSIB. Publication numérique et Archives numériques (Cours Magistral), 2013, 99 p. hal-02810257

HAL Id: hal-02810257

<https://hal.inrae.fr/hal-02810257>

Submitted on 6 Jun 2020

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.

Les données de la recherche



Pascal Aventurier : données ouvertes de la recherche. Cours Enssib 2013

Remerciements / crédits

- Les exposés d'Odile Hologne directrice déléguée à l'IST

- Hologne, O. (2013). *Mise en perspective de l'évolution des métiers de l'IST dans un organisme de recherche*. Communication présentée à Frédoc 2013, AUSSOIS, France (2013-10-07 - 2013-10-10). <http://prodinra.inra.fr/?locale=en#!ConsultNotice:210664>
- Hologne, O. (2013). *Partage des données de la recherche : Politiques, Données, Publications, Compétences ... Inra* (Cours, CNAM-INTD). <http://prodinra.inra.fr/?locale=en#!ConsultNotice:192704>
- Hologne, O. (2013). *Données de la recherche : rôle des professionnels IST à l'Inra*. Communication présentée à 43. congrès de l'ADBU, Le Havre, France (2013-09-19) . <http://prodinra.inra.fr/?locale=en#!ConsultNotice:207958>

- Le groupe INRA. Données de la recherche **État de l'art, étude d'opportunité dans le domaine IST**

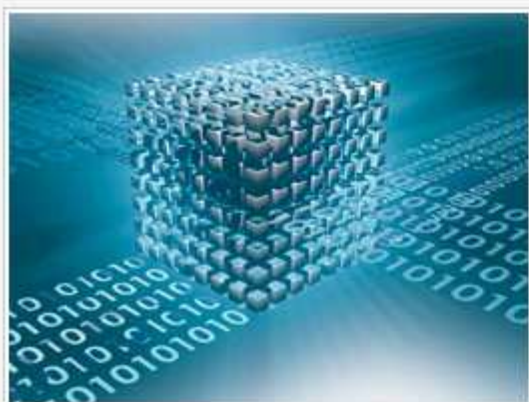
Esther Dzale, Dominique L'Hostis (coordinatrices), Pascal Aventurier, Véronique Garandel, Diane Le Hénaff, Virginie Lelièvre, Sylvie Cocard, Caroline Dandurand. [Wiki en cours de rédaction]

Plan

- Enjeux, définitions et exemples
- Les positions institutionnelles
- La publication des données : Entrepôts, DataJournals et Citation des données
- Gérer les données (producteur et gestionnaire) – qualité des données
- Sémantisation des données (Linked data)
- Data librarian et rôle de l'IST

Enjeux , définitions, exemples

Data et Big Data : le nouvel or noir ?



L'exploitation des données amène des défis technologiques importants auxquels la recherche scientifique a des solutions à apporter. Pour illustrer cela, trois équipes du centre Inria Lille - Nord Europe interviendront sur ce sujet le mercredi 13 mars à 14h, à EuraTechnologies.

Date : 13/03/2013

Lieu : EuraTechnologies

Organisateurs : Inria Lille - Nord Europe

Sur le site Inria

« Les données constituent les briques de base de la société de l'information. Leur quantité est en croissance exponentielle. Elles représentent déjà des masses considérables pour lesquelles on recourt à de nouvelles mesures. Ces données offrent un **potentiel extraordinaire** que l'on commence à exploiter. Elles permettent de **générer des connaissances**, qui étaient soit hors d'atteinte, soit inexistantes, parce que hors du domaine du pensable. » Le Monde, 07.01.2013

Average Cost of Hard Drive Storage

   Share This



Statistic Verification

Source: The Wayback Machine

Research Date: 6.18.2013

Over the past 30 years the cost per gigabyte has gone down by half every 14 months. The following list shows the change in cost since 1980 along with the actual price of specific brand hard drives over the years.

Year	Average Cost Per Gigabyte
2013	\$0.05
2010	\$0.09
2005	\$1.24
2000	\$11.00
1995	\$1,120
1990	\$11,200
1985	\$105,000
1980	\$437,500

<http://www.statisticbrain.com/average-cost-of-hard-drive-storage/>

Données / Données scientifiques

- Produites selon des critères établis par les institutions (gouvernement, villes)
- Statistiques
- Liées à des rapports
- Données d'observation ou d'expérimentation
- Pour démontrer un résultat scientifique
« Cahier de laboratoire »
- Liées à des publications

Données de la recherche

Enjeux

- Mettre à disposition des données pour reproduire les expériences
(éthique , qualité des recherches)
- Valorisation supplémentaire des travaux
- Réutiliser les données produites par d'autres laboratoires
- Conditions d'archivages et de stockage des données

Reproduction des expériences

Ethique

Réutilisation

Valorisation

*Bonnes conditions
de production*

Partage collaboration

Interopérabilité

Data mining

Archivage



SCIENCE & IMPACT

Pascal Aventurier : Les données de la recherche. Cours ENSSIB

Politiques scientifiques et technologiques

> L'innovation dans la science, la technologie et

Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics

Enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés **comme sources principales** pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour **valider des résultats de recherche**. Un ensemble de données de recherche constitue une représentation systématique et partielle du sujet faisant l'objet de la recherche.

<http://www.oecd.org/fr/sti/sci-tech/principesetlignesdirectricesdelocdepourlaccessauxdonneesdelarecherchefinanceesurfondspublics.htm>



open science

Open data

- Une **donnée ouverte** est une donnée [numérique](#) d'origine publique ou privée.
- Produite par une collectivité ou un service public
- Diffusée de manière structurée selon une méthodologie et une [licence ouverte](#) garantissant son libre accès et sa réutilisation par tous, sans restriction technique, juridique ou financière.
- L'ouverture des données s'inscrit dans une tendance qui considère l'information publique comme un [bien commun](#) (diffusion est d'[intérêt public](#) et [général](#)).

Source wikipédia



SCIENCE & IMPACT

Pascal Aventureur : Les données de la recherche. Cours ENSSIB



Premier ministre

PLATEFORME FRANÇAISE D'OUVERTURE DES DONNÉES PUBLIQUES (OPEN DATA)

universites

RECHERCHER

RECHERCHE AVANCÉE

Filtrer la recherche par : Données Idées Questions Articles Tout afficher

57 résultats

Ajouter des critères

Publication

Moins de 24 heures

LES UNIVERSITÉS : DISCIPLINES, CURSUS, ÂGE ET SEXE

Donnée D

Donnée publiée le : 05-12-2011

Producteur : Ministère de l'Education nationale, de la Jeunesse et de la Vie associative

Tableaux statistiques : répartition des étudiants en université selon le sexe, le cursus et la discipline ; répartition des nouveaux entrants à l'université selon le sexe, l'âge et la discipline.

Période : du 01/09/2009 au 31/08/2010

Date de production des données : 01/09/2010

Fréquence de mise à jour : annuelle

Licence : Licence Ouverte/Open Licence

Langue : français

Couverture géographique : France - Total France métropolitaine + DOM Public+Privé

Mots clés : éducation enseignement supérieur nouveaux entrants université étudiant université

Accéder aux données

FORMAT

XLS

Télécharger

Licence

LICENCE OUVERTE / OPEN LICENCE



LICENCE OUVERTE
OPEN LICENCE

PAS ENCORE DE DISCUSSIONS
sur ce jeu de donnée

POSER LA PREMIÈRE QUESTION

	A	B	C	D	E	F	G	H	I
1	fr								
2	[1] Répartition des étudiants en université par sexe,cursus et discipline en 2009-2010								
3	(France métropolitaine + DOM)								
4		Cursus licence		Cursus master		Cursus doctorat		Ensemble	
5	Disciplines	Effectifs	% femmes	Effectifs	% femmes	Effectifs	% femmes	Effectifs	% femmes
6	Droit sciences politiques	115 701	64,4	69 548	66,1	8 238	48,7	193 487	64,3
7	Sciences économiques, gestion	80 450	51,6	60 914	52,6	4 079	45,3	145 443	51,9
8	AES	33 773	59,8	6 780	60,5	18	33,3	40 571	59,9
9	Pluri droit, Sciences économiques, AES			33	45,5			33	45,5
10	Lettres sciences du langage	63 697	70,8	23 219	75,5	6 060	65,6	92 976	71,7
11	Langues	80 989	73,3	19 464	77,1	2 725	66,9	103 178	73,8
12	Sciences humaines sociales	123 486	68,4	66 442	68,3	14 323	53,6	204 251	67,3
13	Pluri Lettres, Langues, Sciences humaines	2 031	66,5	4 678	75,6	26	57,7	6 735	72,8
14	Sciences fondamentales et application	76 393	28,0	59 645	26,9	16 888	28,3	152 926	27,6
15	Sciences de la nature et de la vie	41 209	62,3	20 063	57,2	10 437	52,7	71 709	59,5
16	STAPS	25 436	31,4	6 266	33,8	492	36,8	32 194	32,0
17	Pluri sciences	21 323	44,5	1 899	47,3	123	30,9	23 345	44,6
18	Médecine-odontologie	59 456	65,7	109 589	59,0	1 183	50,6	170 228	61,3
19	Pharmacie	9 563	67,2	20 889	67,0	398	57,8	30 850	66,9
20	IUT	118 139	40,2					118 139	40,2
21	Total hors IUFM	851 646	56,8	469 429	58,1	64 990	47,2	1 386 065	56,8
22	Pourcentage par cursus hors IUFM	61,4		33,9		4,7		100	
23	IUFM			58 518	75,3			58 518	75,3
24	Total	851 646	56,8	527 947	60,0	64 990	47,2	1 444 583	57,6
25	Pourcentage par cursus	59,0		36,5		4,5		100	
26									
27	(1) Pour en savoir plus sur les DOM, les COM et la Nouvelle-Calédonie, voir 12.9								
28	Source : MESR-DGESIP-DGRI SIES / Système d'information SISE								

Format ?
Exploitabilité ?
Définition ?



de la recherche. Cours ENSSIB

12

- **Data set** (jeu de données)

- Un type de « contenant de données », publié et faisant partie d'une collection ou isolé avec un DOI (digital Object Identifier) (Green 2009)
- Un ensemble de données correspondant au contenu d'une table de base de données unique, ou d'une seule matrice, dans lequel chaque colonne du tableau, représente une variable particulière, et chaque ligne correspond à un élément donné de l'ensemble de données en question. L'ensemble de données répertorie les valeurs pour chacune des variables (wikipédia)

Exemple de jeu de données scientifiques



PANGAEA
Data Publisher for Earth & Environmental Science

<http://doi.pangaea.de/10.1594/PANGAEA.805269>

Not logged in (log in or sign up)

Always quote citation when using data

Data Description

[Show Map](#) [Google Earth](#)

Citation: **Van Wambeke, France (2013):** Bacterial production measured on water bottle samples during THALASSA cruise PROSOPE. doi:10.1594/PANGAEA.805269

Related to: **Claustre, Hervé; et al (1999):** PRODUCTIVITY OF PELAGIC OCEANIC SYSTEMS (PROSOPE). France-JGOFs/PROOF Data Management Office, May 2003, Observatoire Océanologique de Villefranche-sur-Mer (OOV), France, http://www.obs-vlfr.fr/cd_rom_dmtt/pr_main.htm

Further details: Bacterial production during PROSOPE cruise

Project(s): **Joint Global Ocean Flux Study (JGOFs)**

Biogeochemical Processes in the Oceans and Fluxes (PROOF)

Coverage: Median Latitude: 37.752616 * Median Longitude: 10.723848 * South-bound Latitude: 30.976000 * West-bound Longitude: -10.048683 * North-bound Latitude: 43.433083 * East-bound Longitude: 22.040017

Date/Time Start: 1999-09-09T13:35:00 * Date/Time End: 1999-10-03T12:03:00

Minimum DEPTH, water: 1.7 m * Maximum DEPTH, water: 781.1 m

Event(s): **PROSOPE_orp004** (UPW1) * Latitude: 31.015967 * Longitude: -10.023483 * Date/Time Start: 1999-09-09T13:35:00 * Date/Time End: 1999-09-09T13:48:00 * Campaign: PROSOPE * Basis: Thalassa * Device: Bottle, Niskin

PROSOPE_orp007 (UPW1) * Latitude: 30.976000 * Longitude: -10.048683 * Date/Time Start: 1999-09-09T22:09:00 * Date/Time End: 1999-09-09T22:22:00 * Campaign: PROSOPE * Basis: Thalassa * Device: Bottle, Niskin



Parameter(s)	Name	Short Name	Unit	Principal Investigator	Method	Comment
1	Event label	Event				Metadata
2	DATE/TIME	Date/Time				Geocode
3	LATITUDE	Latitude				Geocode
4	LONGITUDE	Longitude				Geocode
5	DEPTH, water	Depth water	m			Geocode
6	Pressure, water	Press	dbar	Van Wambeke, France		
7	Bacteria, production as carbon	Bact prod C	mg m ³ day	Van Wambeke, France		Bacterial production estimation (Smith & Azam, 1992, Mar Microb Food Webs 6:107)
8	Bacteria, production as carbon, standard deviation	Bact prod C std dev		Van Wambeke, France		Bacterial production estimation (Smith & Azam, 1992, Mar Microb Food Webs 6:107)
9	Bottle number	Bottle		Van Wambeke, France		

License: Creative Commons Attribution 3.0 Unported

Size: 1184 data points

Download Data

Download dataset as tab-delimited text (use the following character encoding: ISO-8859-1: ISO Western (PANGAEA default))

[View dataset as HTML](#)

Download dataset as tab-delimited text (use the following character encoding: ISO-8859-1: ISO Western (PANGAEA default))

1	2	3	4	5	6	7	8	9
Event	Date/Time	Latitude	Longitude	Depth water [m]	Press [d̄bar]	Bact prod C [mg·m ⁻³ ·day]	Bact prod C std dev [±]	Bottle
PROSOPE_orp004	1999-09-09T13:35	31.0160	-10.0235	1.73	1.742	12.4152	0.2616	20
PROSOPE_orp004	1999-09-09T13:35	31.0160	-10.0235	7.53	7.599	15.3888	1.3872	18
PROSOPE_orp004	1999-09-09T13:35	31.0160	-10.0235	12.37	12.481	12.3984	0.6120	16
PROSOPE_orp004	1999-09-09T13:35	31.0160	-10.0235	19.63	19.802	8.0136	0.1224	14
PROSOPE_orp004	1999-09-09T13:35	31.0160	-10.0235	30.60	30.880	3.0024	0.1872	12
PROSOPE_orp004	1999-09-09T13:35	31.0160	-10.0235	39.84	40.196	1.4592	0.0480	10
PROSOPE_orp004	1999-09-09T13:35	31.0160	-10.0235	48.41	48.853	1.3464	0.0960	8
PROSOPE_orp004	1999-09-09T13:35	31.0160	-10.0235	58.13	58.661	1.5864	0.0504	6
PROSOPE_orp004	1999-09-09T13:35	31.0160	-10.0235	65.84	66.440	1.3512	0.1248	4
PROSOPE_orp004	1999-09-09T13:35	31.0160	-10.0235	87.30	88.103	2.4192	0.0768	2
PROSOPE_orp007	1999-09-09T22:09	30.9760	-10.0487	2.08	2.103	9.8856	0.6048	20
PROSOPE_orp007	1999-09-09T22:09	30.9760	-10.0487	2.32	2.339	10.5936	0.1248	18
PROSOPE_orp007	1999-09-09T22:09	30.9760	-10.0487	4.73	4.769	10.1112	0.0672	16
PROSOPE_orp007	1999-09-09T22:09	30.9760	-10.0487	9.77	9.855	10.1544	0.1080	14
PROSOPE_orp007	1999-09-09T22:09	30.9760	-10.0487	14.78	14.916	7.6488	0.0648	12
PROSOPE_orp007	1999-09-09T22:09	30.9760	-10.0487	19.53	19.705	6.6048	0.4824	10
PROSOPE_orp007	1999-09-09T22:09	30.9760	-10.0487	24.66	24.887	5.1432	0.1416	8
PROSOPE_orp007	1999-09-09T22:09	30.9760	-10.0487	34.20	34.514	4.6584	0.1296	6
PROSOPE_orp007	1999-09-09T22:09	30.9760	-10.0487	44.35	44.754	4.7928	0.1200	4
PROSOPE_orp007	1999-09-09T22:09	30.9760	-10.0487	69.37	70.002	1.3584	0.0624	2
PROSOPE_orp011	1999-09-14T15:40	36.0987	-5.1975	4.70	4.931	2.3712	0.2040	20
PROSOPE_orp011	1999-09-14T15:40	36.0987	-5.1975	11.15	11.251	2.9616	0.2040	18
PROSOPE_orp011	1999-09-14T15:40	36.0987	-5.1975	20.50	20.682	3.3576	0.1560	16
PROSOPE_orp011	1999-09-14T15:40	36.0987	-5.1975	29.26	29.526	4.1064	0.0600	14
PROSOPE_orp011	1999-09-14T15:40	36.0987	-5.1975	38.24	38.585	5.4576	0.1320	12



http://commons.wikimedia.org/wiki/File:HMS_Endurance_patrolling_off_the_Antarctic_as_penguins_look_on_MOD_45147622.jpg



<http://archaeofacts.com/mesoamerican-highlands/archaeology-dissertation>

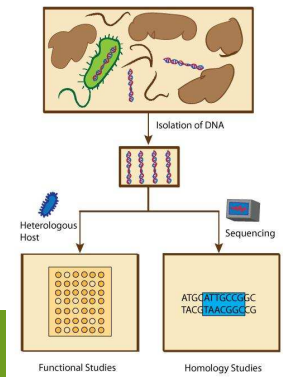
Observation

Enquêtes



http://commons.wikimedia.org/wiki/File:3Supernova_Bubble.jpg

Omiques (génomiques...)



Quel type de Données ?

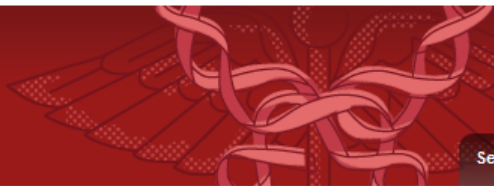
- **Données primaires ou brutes** : l'information enregistrée comme des notes, images, données d'observation, appartenant à un projet de recherche spécifique
- **Données issues d'un traitement** : Analyses, descriptions, conclusions arrangées sur un support ou dans une publication
- **Données publiés**: L'information distribuée aux personnes impliquées dans l'acquisition de données et leur administration

Partager les données de la recherche



Logo by [Greg Emmerich](#) / [CC-BY-SA](#)

open science



The delay in sharing research data is costing lives

Josh Sommer

Nature Medicine **16**, 744 (2010) | doi:10.1038/nm0710-744



It is not uncommon for potentially life-saving research data to be published years after being generated. But the setback to progress caused by the delay in releasing data is troublesome for people who selflessly participate in trials and desperately await new therapies. Scientists need to feel greater urgency to share their findings quickly, and they need additional avenues to facilitate this process.

Pourquoi partager les données de la recherche ?



<http://prezi.com/ytir00evayoj/the-jord-project-and-implications-for-repositories/>

Libre accès à l'information scientifique et technique

<http://openaccess.inist.fr/?Declaration-de-Berlin-sur-le-Libre>

Actualités, problématiques et perspectives

[Accueil](#) [Glossaire](#) [Problématiques et perspectives](#) [Textes de références](#) [Archives](#)

[> Accueil](#) [> Textes de références](#)

Déclaration de Berlin sur le Libre Accès à la Connaissance en sciences exactes, sciences de la vie, sciences humaines et sociales

A.A. [Partager](#) [Imprimer](#)

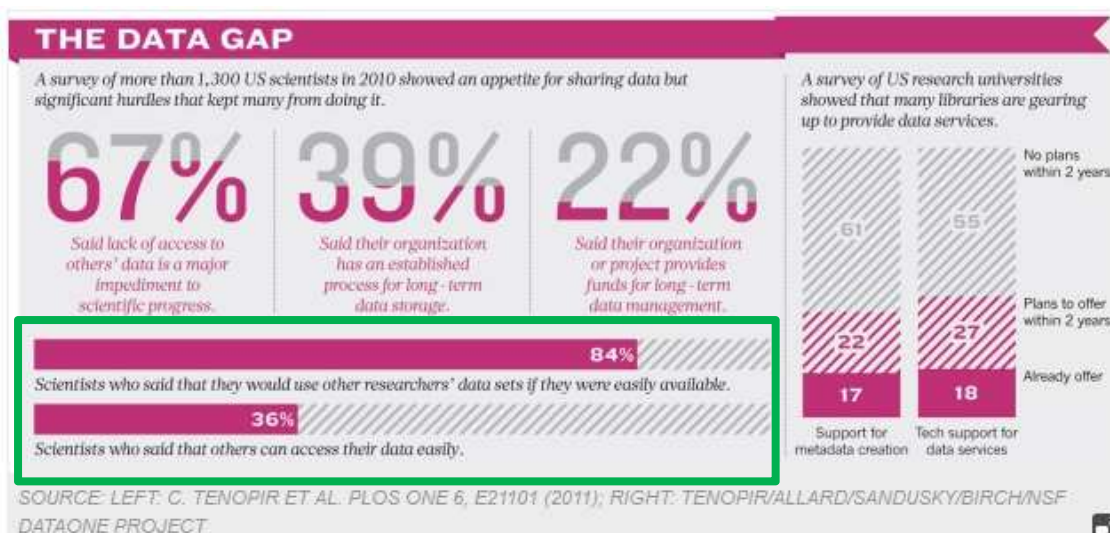
Déclaration de Berlin sur le Libre Accès à la
Connaissance en sciences exactes, sciences de la
vie, sciences humaines et sociales

Définition d'une contribution au libre accès (22 octobre 2003)

... le libre accès requiert l'engagement de tout un chacun en tant que producteur de connaissances scientifiques ou détenteur du patrimoine culturel. Les contributions au libre accès se composent de résultats originaux de recherches scientifiques, **de données brutes** et de [métadonnées](#), de documents sources, de représentations numériques de documents picturaux et graphiques, de documents scientifiques multimédia.



84% des chercheurs aimeraient utiliser des jeux de données d'autres chercheurs



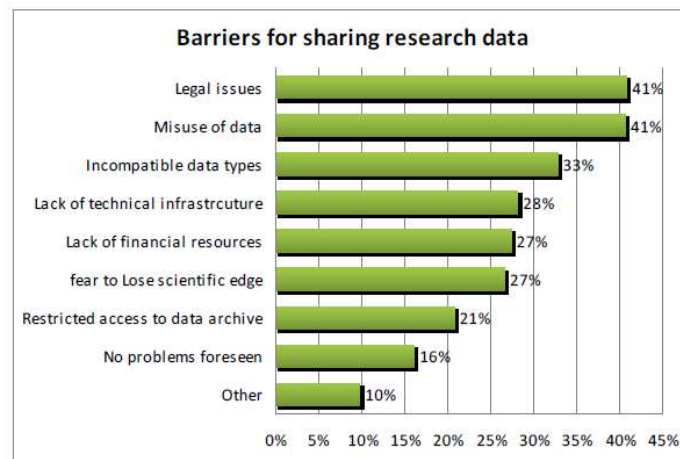
Source : <http://www.nature.com/news/publishing-frontiers-the-library-reboot-1.12664>

Barrières pour partager les données de la recherche

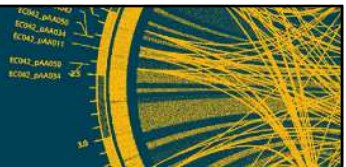
Report on Integration of Data and Publications

Grant Agreement no.: 261530

Researchers perceive legal and professional reasons for not sharing their data



Graph 9: source PARSE.Insight: Do you experience or foresee any of the following problems in sharing you data? N=1270(PARSE.Insight¹⁴)



6 domaines clés

- Les scientifiques doivent être plus ouverts entre eux, avec le public et les médias
- Une plus grande reconnaissance doit être donnée à la collecte de données, l'analyse et la communication
- Des normes et standards pour le partage de l'information sont nécessaires et doivent être utilisés
- Les données de publications publiées sous une forme réutilisable pour démontrer les nouveautés doivent être une obligation
- Plus d'experts dans le management et pour promouvoir l'utilisation des données digitales sont demandés
- De nouveaux outils sont nécessaires pour développer l'analyse des données collectées, en forte augmentation

Les positions institutionnelles

Recommendations of the participants of the workshop on
Open Access to Research Data to the European
Commission (21.2.2013)

We recommend that the European
Commission should introduce **mandatory
data management plans** which include a
description of the data sets, information on
data handling beyond project lifetime and
possible repositories for depositing the data
sets.

Recommendations for an effective Open Data Pilot held by European Commission on 2 July 2013

•OpenAIRE

•Association of European
Research Libraries (LIBER)

•Confederation of Open Access
Repositories (COAR)

- Sensibiliser les chercheurs sur le besoins et les bénéfices de gérer et partager les données de la recherche
 - Encourager les bibliothèques, data centre, les éditeurs et les acteurs pertinents de commencer à travailler avec les chercheurs sur le cycle de gestion des données
 - Identifier ce qui est exigé pour tous les acteurs en terme d'expertise de connaissances et de compétences
 - Créer des e-infrastructures interoperables
 - Mobiliser les bibliothèques pour aider les communautés scientifiques dans le plan de gestion de leurs données
- Fournir à la Commission des éléments pour prendre des décisions pour augmenter le partage de données



Open Access Infrastructure for Research in Europe

- Accompagner l'obligation - CE et le Conseil Européen de la Recherche (ERC), de déposer en open access les publications et résultats des recherches financées par la CE
- Objectifs de OpenAIREplus (2011-2013) :
 - Étendre l'open access à toutes les publications
 - connexion entre publications et données
 - interopérabilité avec les infrastructures existantes majeures DataCite, Mendeley, ORCID, EUDAT, ..
 - Lance en mai 2013 Zenodo (avec le CERN)
 - Accord avec Re3data.org (nov 2013)

Open Access to data

Des obligations de dépôts

- [Wellcome Trust](#)
- [European Commission](#),
- [the Office of Science and Technology Policy \(OSTP\)](#) from the White House.

Fenner, Martin. "The Complete Article." Gobbledygook, October 20, 2013.



- La Mission pour l'interdisciplinarité
- Actualités interdisciplinaires
- Les défis
- PEPS de site
- PEPS en réseau
- Autres
- Espace communication
- Archives

Rechercher sur ce site

Sur le Web du CNRS

Mastodons - CNRS



Accueil du site > Les défis > MASTODONS

Grandes masses de données scientifiques MASTODONS

Année de démarrage : 2012

Contexte :
 Nos capacités de production, d'acquisition et d'archivage dépassent largement nos capacités de traitement, d'analyse et d'archivage. De nombreux domaines scientifiques (ex : cartographie du génome, observatoire virtuel en astronomie, simulation en physique et énergie), économiques (ex : e-commerce, systèmes bibliothèques numériques, distribution de contenu) de données considérables. Le CNRS, à travers ses projets, capitalise plusieurs centaines, voire milliers, de volumes croissant de façon exponentielle.

Objectifs :
 Le but est d'identifier et de soutenir des actions de recherche sans une fertilisation croisée des disciplines et sans que les verrous scientifiques posés par l'utilisation à grande échelle des données (sémantique, stockage, recherche, technologies proposées (protection de la vie privée))

Les projets sélectionnés couvrent les axes de recherche suivants :

1. Stockage, indexation et accès aux données
2. Extraction de connaissances, apprentissage
3. Qualité des données, confidentialité et sécurité
4. Problèmes de propriété, de droit d'usage, de droit
5. Préservation/archivage des données pour les

<http://www.cnrs.fr/mi/spip.php?article53&lang=fr>



Pascal Aventurier : Les données de la

Nos capacités de production, d'acquisition et d'accès aux données dépassent largement nos capacités de traitement, d'analyse et d'archivage

Soutenir des actions de recherche interdisciplinaires (synergie nécessaire entre chercheurs)

1. Stockage, indexation, accès
2. Extraction de connaissances, Apprentissage
3. Qualité, confidentialité et sécurité
4. propriété, droit (usage et oubli)
5. Préservation/archivage

2013 : Une politique de partage des données pour l'INRA

- ❖ L'Inra a une politique de partage des données
- ❖ Cette politique est mise en œuvre dans le cadre d'un projet qui implique une cinquantaine de personnes
- ❖ Les compétences mobilisées sont : scientifiques, informatique et IST
- ❖ L'IST est impliqué
 - ✓ Dans la mise en œuvre de la politique
 - ✓ Dans des projets avec les chercheurs
 - ✓ Dans des groupes de travail et des projets internationaux

Source : Hologne, O. (2013). Données de la recherche : rôle des professionnels IST à l'Inra. Communication présentée à 43. congrès de l'ADBU, Le Havre, France (2013-09-19) .

<http://prodinra.inra.fr/?locale=en#!ConsultNotice:207958>

Charte de déontologie de l'Inra

Charte de déontologie de l'INRA

Préambule

L'Institut national de la recherche agronomique (INRA) est un organisme public de recherche fondé, ouvert à l'innovation, conduisant des recherches, souvent en collaboration avec d'autres organismes publics ou privés. Ces recherches portent sur des enjeux sociétaux relevant de trois domaines principaux, l'agriculture, l'environnement et l'alimentation, interconnectés entre eux. Les domaines les plus importants sont l'énergie, le climat, la santé, les territoires. Les questions de recherche, complexes, pour répondre à la société ont en outre de multiples enjeux, à chaque instant, sous forme de mettre en œuvre une démarche scientifique rigoureuse, que de s'intégrer sur les aspects éthiques qui en découlent. Pour ce faire, le comité consultatif d'éthique INRA-CCEAD est en appui de la direction de l'Éthique et de l'ensemble des personnels pour servir le questionnement éthique sur les finalités des programmes de recherche et à évaluer un déontologie.

L'INRA réalise les missions de service public qui lui sont confiées : la production, la diffusion et la valorisation de connaissances nouvelles, considérées comme biens publics, l'expertise, notamment en appui des politiques publiques, la contribution à la formation et à l'éducation, la communication vers la société. En tant qu'établissement public qui mobilise des collections de recherche, d'appui et de soutien à la recherche, l'INRA a la responsabilité de servir l'intérêt général. Pour ce faire, l'Institut privilégie la recherche de la performance, de l'efficacité, de la qualité des résultats dans le respect mutuel, en mettant en œuvre l'impartialité et l'objectivité de ses comportements, et se fait particulièrement vigilant sur ces questions.

C'est de la responsabilité de l'établissement que de donner des repères déontologiques à l'ensemble de ses personnels pour que chacun puisse inscrire, au sein de ses missions et activités, l'ambition de servir cet intérêt général de façon impartiale et exemplaire.

La charte de déontologie exprime le sens général et les principes que chacun, personnellement, s'engage à respecter dans le cadre des missions qui lui sont confiées. Elle s'inscrit dans une démarche continue d'amélioration des pratiques professionnelles. Elle est portée par la direction générale de l'Institut et s'adresse à chaque agent qui y travaille, à compter de sa signature. L'établissement s'engage à ce que tous ont une connaissance. Elle est adossée à un ensemble de règles applicatives, réglementaires ou contractuelles, qui régissent les activités professionnelles de tous. La loi n° 2012-285 du 21 juillet 2012 portant droits et obligations des fonctionnaires, le Code de la recherche, le Code de l'éducation, la Charte européenne de chercheurs (11 mars 2002), la Déclaration de Singapour sur l'intégrité en recherche (2009) sont quelques-uns des principaux textes constituant le cadre qui promeut l'exercice de la responsabilité individuelle de chaque agent.

En étant pressé, à l'échelle individuelle et collective, d'un comportement qui met en pratique la charte de déontologie dans toutes ses activités, l'Institut exprime ainsi l'ambition de constituer le lieu de confiance qui lui est la société.

L'engagement de la direction générale de l'INRA

1. La direction générale de l'INRA définit et explicite les orientations, dans les champs scientifique et de l'appui à la recherche, qui seront mises en œuvre et les procédures qui seront suivies. Elle veille à prendre ses décisions avec impartialité et transparence et sera garante de l'application de ces principes à l'ensemble de l'Institut.

¹ <http://www.inra.fr/ethique/2012/06/04/le-ccead>
² <http://www.inra.fr/ethique/2012/06/04/le-ccead>

Les données, les savoir-faire et collections, la propriété industrielle : un patrimoine de qualité à préserver et valoriser

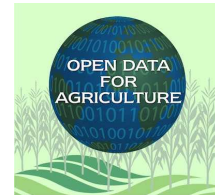
9. Les données de la recherche constituent l'une des composantes identifiables des productions de l'Institut, qu'elles soient issues des équipes de recherche, des infrastructures de recherche, des services administratifs. Elles sont la propriété de l'INRA, ou la copropriété, si elles sont obtenues dans le cadre d'accords de partenariat. La production, l'archivage, le traitement, la gestion des données de la recherche, obtenues par l'Institut ou transmises par des tiers, doivent obéir à des procédures explicites, respectant les règles juridiques et déontologiques qui prévalent, notamment dans l'expérimentation ou dans l'utilisation des données personnelles. Ces procédures garantissent leur qualité, leur sécurisation, en particulier informatique, leur traçabilité, et fixent les règles de leur mise à disposition.

Duée, P.-H. (2013). Charte de déontologie de l'INRA.
<http://prodinra.inra.fr/?locale=en#!ConsultNotice:188870>

Des catalyseurs récents

❖ G8+5 Open data for Agriculture (avril 13)

*Obtain commitment and action from nations and relevant stakeholders **to promote policies and invest in projects that open access to publicly funded global agriculturally relevant data streams**, making such data readily accessible to users in Africa and world-wide, and ultimately supporting a sustainable increase in food security in developed and developing countries*



❖ Research data Alliance (mars 13)

*The Research Data Alliance aims **to accelerate and facilitate research data sharing and exchange***



Source : Hologne, O. (2013). Données de la recherche : rôle des professionnels IST à l'Inra. Communication présentée à 43. congrès de l'ADBU, Le Havre, France (2013-09-19) .

<http://prodinra.inra.fr/?locale=en#!ConsultNotice:207958>



Research Data Sharing
without barriers

- Australian National Data Service; iCordi - European Commission (FP7); RDA/US activity (National Science Foundation)
- Acceleration internationale de l'innovation "data-driven" et de leur exploitation (discover) en facilitant le partage des données
- Utilisation et ré-utilisation des données, harmonisation des standards et "discoverability"

Livrables : infrastructure, règles, utilisation, standards...

Source : Hologne, O. (2013). Données de la recherche : rôle des professionnels IST à l'Inra. Communication présentée à 43. congrès de l'ADBU, Le Havre, France (2013-09-19) .

<http://prodinra.inra.fr/?locale=en#!ConsultNotice:207958>

34

ands
AUSTRALIAN NATIONAL DATA SERVICE

ANDS Home | Contact Us | Guides

Find Research Data: Search Search our site: Go

Australian National Data Service

Our Vision: More Australian researchers reusing research data more often

ANDS is enabling the transformation of:

Data that are:	to	Structured Collections that are:
Unmanaged	→	Managed
Disconnected	→	Connected
Invisible	→	Findable
Single-use	→	Reusable

[More>>](#)

Australian Research Data Commons
ANDS is building the [Australian Research Data Commons](#): a cohesive collection of research resources from all research institutions, to make better use of Australia's research data outputs.

Research Data Australia
[Research Data Australia](#), ANDS' flagship service, provides a comprehensive window into the Australian Research Data Commons. It is an Internet-based discovery service designed to provide rich connections between data, projects, researchers and institutions, and promote visibility of Australian research data collections in search engines.

Search for research data on researchdata.ands.org.au

ANDS News

Connecting Data to Services
ANDS is soliciting input into possible ways of connecting data to services. [Click here for details.](#)

Congratulations
Flinders University has completed its [Metadata Store Upgrade](#)

ANDS newsletter share
Issue 17 is a celebration of all completed ANDS-funded projects

Subscribe to [andsUP](#), our regular e-newsletter

ANDS Quick Links

- [Contact Us](#)
- [Guides](#)
- [Content Providers Guide](#)
- [ANDS Services](#)

<http://www.ands.org.au/>

INRA
SCIENCE & IMPACT
Pascal Aventurier : Les données de la recherche. Cours ENSSIB

35

How can the DCC help you?

About us

We are a world-leading centre of expertise in digital information curation...

What is digital curation?

Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle...

Working with HEIs

Under the institutional engagement programme we have been working with HEIs to provide tailored support...

Editor's choice

Latest news

Next events

Preserving Computer-Aided Design

A new Technology Watch Report from the Digital Preservation Coalition (DPC)...


Out of Cite, Out of Mind: The Current State of Practice, Policy and, Technology for the Citation of Data

A recently published CODATA report...

Economic Impact of the British Atmospheric Data Centre (BADC)

One in a series of studies on the value and impact of UK research data centres...

<http://www.dcc.ac.uk/>











Wednesday, November 06, 2013

search this site

- Home
- About
- CODATA Membership
- Resources
- Task and Working Groups
- Archives
- CODATA in the Press
- CODATA BLOG
- CODATA

International Council for Science : Committee on Data for Science and Technology

< home > < blog > < newsletter > < discussion list > < data science journal > < contact > < members area >


CODATA


At the 25th CODATA General Assembly on 26-27 October in Beijing, the following **Mission Statement** was approved:

The mission of CODATA is to strengthen international science for the benefit of society by promoting improved scientific and technical data management and use.


NEWS AND EVENTS (for past news and events, visit our Archives)

NEW! CODATA Newsletter 106

 **CODATA now has a BLOG!**
The blog will feature guest posts from the CODATA community and from Simon Hodson, Executive Director.

 **Release of Data Citation Report: "Out of Cite, Out of Mind"**
Directed by the staff of the US CODATA/Board on Research Data and Information (BRDI) the report was published by the CODATA Data Science Journal and is available freely and openly online at https://www.jstage.jst.go.jp/article/dsj/12/0/12_OSOM13-043/_article.

<http://www.codata.org/>

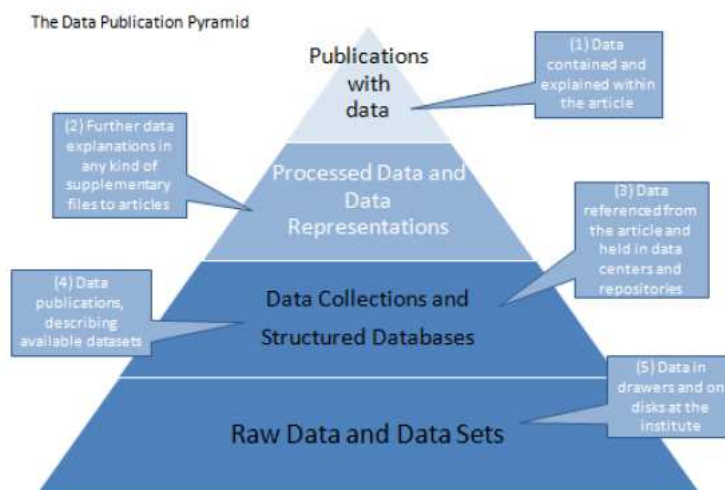


Pascal Aventurier : Les données de la recherche. Cours ENSSIB

37

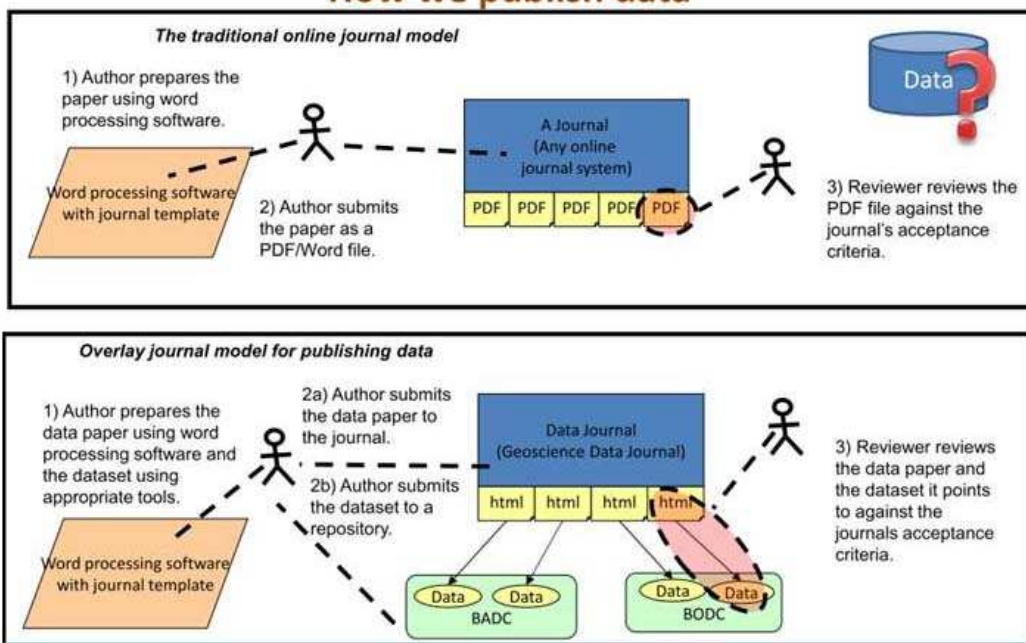
Publication des données : Entrepôts, Data Journals et Citation des données

Intégration des données et des publications

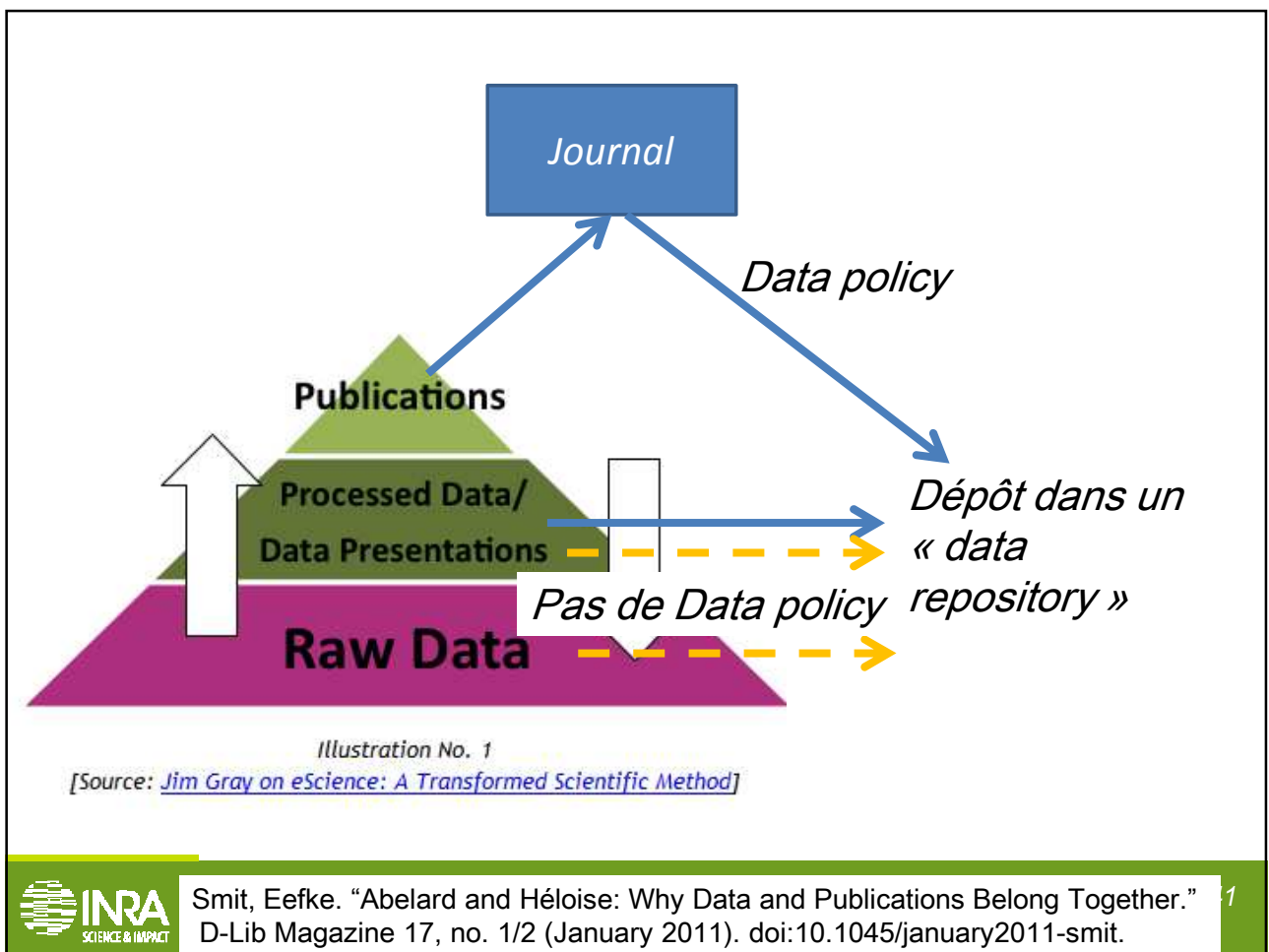


Graph 1 repeated, The Data Publication Pyramid, based on the Jim Gray Pyramid.

How we publish data



Whyte, A.; Callaghan, S., 2013. Perspectives on the Role of Trustworthy Repository Standards in Data Journal Publication. *IASSIST. Cologne - Allemagne: 2013/05/31. pp. 1-27.*
<http://fr.slideshare.net/angusawhyte/iassist-preparde-whyte>



Définitions

Data Paper :

Publication scientifique dont le but principal est de décrire un jeu de données ou un ensemble de jeux de données, publiés dans des **data journals** (ex Geoscience data journal)

Dater paper commenté en séance

- The GBS dataset: measurements of satellite site diversity at 20.7 GHz in the UK.

Geosciences Data journal

<http://onlinelibrary.wiley.com/doi/10.1002/gdj3.2/abstract>

Politique des revues scientifiques

The American Economic Review: Data Availability Policy

It is the policy of the American Economic Review to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Authors of accepted papers that contain empirical work, simulations, or experimental work must provide to the Review, prior to publication, the data, programs, and other details of the computations sufficient to permit replication. These will be posted on the AER Web site. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the requirements above cannot be met.

As soon as possible after acceptance, authors are expected to send their data, programs, and sufficient details to permit replication, in electronic form, to the AER office.

...

Authors are also asked to provide information on how the proprietary data can be obtained by others in their Readme PDF file. A copy of the programs used to create the final resu



D'après datacite. "2013 DataCite Summer Meeting - Elsevier's Program to Support Resear..." Technologies, October 7, 2013. <http://fr.slideshare.net/datacite/elseviers-program-to-support-research-data-hylke-koerseelsevier>

Site content

[Homepage](#)

Policies

- [Publication ethics](#)
- [Bioethics](#)
- [Availability of data & materials](#)
- [Peer-review policy](#)
- [Embargo](#)
- [Corrections](#)
- [License to publish](#)
- [Feedback](#)

Availability of data and materials

The policy outlined on this page applies to *Nature* journals (those with the word "Nature" in their title). NPG publishes many other journals, each of which has separate publication policies described on its website. A current list of these journals, with links to each journal's homepage [is available](#).

Reporting requirements for life sciences research

As of May, 2013, Nature journals require authors of life sciences research papers that are sent for external review to include in their manuscripts relevant details about several elements of experimental and analytical design. This initiative aims to improve the transparency of reporting and the reproducibility of published results. It focuses on [elements of methodological information](#) that are frequently poorly reported. During peer review, authors will be asked to confirm that these elements are included in the manuscript by filling out a [checklist](#) that will be made available to the editors and reviewers.

[SUBMIT NOW](#)

ADVERTISEMENT

**We've
started the
countdown!**

Availability of data and materials

A condition of publication in a Nature journal is that **authors are required to make materials, data and associated protocols promptly available to others without undue qualifications.**

...

Supporting data must be made available to editors and peer-reviewers at the time of submission for the purposes of evaluating the manuscript. Peer-reviewers may be asked to comment on the terms of access to materials, methods and/or data sets;

...

After publication, readers who encounter refusal by the authors to comply with these policies should contact the chief editor of the

Sharing materials

A condition of publication in a Nature journal is that authors are required to make materials, data and associated protocols promptly available to others without undue qualifications.

For biological materials such as mutant strains and cell lines, the Nature journals require authors to use **established public repositories when one exists**

Isaig 10100
we 1110

NOTICE TO CONTRIBUTORS

Revised February 2011

Manuscript Submission

Data Citation



Manuscript Submission Format: Manuscripts should meet the format guidelines specified in the Notice to Contributors published in the February and August issues of each volume. All text must be double-spaced and typed in Times New Roman, 12-point font size. Margins should be at least 1 inch on all four sides. You may cite your own work, but do not use wording that identifies you as the author.

Submission Requirements: Manuscripts submitted to *ASR* are processed electronically through SAGE track. Authors can create an account and log in to submit a manuscript at <http://mc.manuscriptcentral.com/asr>. As part of the blind peer review process, authors will need to upload a blinded manuscript without a title page, as well as a separate title page

5. References are presented in a separate section headed "REFERENCES." All references and data sets cited in the text must be listed in the reference section, and vice versa. Publication information for each must be complete and correct. List

References for data sets should include a persistent identifier, such as a Digital Object Identifier (DOI). Persistent identifiers ensure future access to unique published digital objects, such as a text or data set. Persistent identifiers are assigned to data sets by digital archives, such as institutional repositories and partners in the Data Preservation Alliance for the Social Sciences (Data-PASS).

Data Sets: Deschenes, Elizabeth Piper, Susan Turner, and Joan Petersilia. *Intensive Community Supervision in Minnesota, 1991* Release [Computational Data] Political and Soc
dataciteplenaryv3-131007073432-phnapp01.pdf



Data Journals

- Revues sur les jeux de données
Data Journal

[*Scientific Data \(Nature\)*](#)
[*Geoscience Data Journal*](#)
[*CODATA Data Science*](#)
[*Journal Earth System*](#)
[*Science Data*](#)
[*Journal of Chemical &
Engineering Data \(ACS
Publications\)*](#)
[*Big Data*](#)
[*Biodiversity Data Journal*](#)
[*Genomics Data \(Elsevier\)*](#)
[*Journal of Open Archaeology
Data*](#)

Ne contiennent pas les données mais des articles de recherche décrivant les données




Les données doivent être accessibles dans des « répertoires de données »

Data journal (groupe Nature)

SCIENTIFIC DATA

[Home](#) | [About](#) | [For Authors](#) | [Advisory and Editorial Board](#) | [Open Access](#) | [Contacts](#) | [FAQ](#)

Helping you publish, discover,
and reuse research data

 Credit Credit, through a citable publication, for depositing & sharing your data	 Reuse Complete, curated & standardized descriptions enable the reuse of your data	 Quality Rigorous community based peer review
 Discovery Find datasets relevant to your research	 Open Promotes & endorses open science principles & available to all through a Creative Commons license	 Service In-house curation, rapid peer review & publication of your data descriptions

Welcome to *Scientific Data*



SCIENCE & IMPACT

Pascal Aventurier : Les données de la recherche. Cours ENSSIB

Geoscience Data Journal

© 2013 John Wiley & Sons Ltd.

Each article is made available under the terms of the [Creative Commons Attribution License](#)



Edited By: Dr Rob Allan, Met Office, UK

Online ISSN: 2049-6060

Article Publication Charges

All articles published in *Geoscience Data Journal* are fully open access: immediately freely available to read, download and share. To cover publication costs, *Geoscience Data Journal* charges a publication fee.

Geoscience Data Journal's standard publication charge for authors submitting directly to the journal is set out below.

Amount Payable by Author	Discount Reference	Discount Amount
\$1,500 / €1,200 / £1,000	Full Article Publication Charge	N/A
\$1,200 / €960 / £800	RMetS Members' Discount	20%

Dataset submission

To publish a Data Paper in *Geoscience Data Journal* about a dataset the authors must complete the following two-tiered process:

- The dataset, along with supporting metadata, must be formally archived in a *Geoscience Data Journal* approved repository or data centre (and preferably have been assigned a [digital object identifier \(DOI\)](#)). A list of approved institutions can be found below. If the one you have elected to work with does not appear please contact the Editor for consultation.
- A paper describing the dataset, giving details of its collection, processing, file formats etc. should be written and submitted using the *Geoscience Data Journal* online submission system (<https://mc.manuscriptcentral.com/geosciencedata>).

Subject to satisfactory reviews of both dataset and paper, *Geoscience Data Journal* will publish the data description paper, along with a link to the underlying dataset (usually by means of the dataset's DOI).

An approved repository is one that is commonly used by the scientific community it supports, has a formal data management policy in place, and can mint a DOI or provide a stable URL and unique identifier for the dataset. Current approved repositories are:

- 3TU.Datacentrum
- British Atmospheric Data Centre (BADC)
- British Oceanographic Data Centre (BODC)
- CSIRO Data Access Portal
- Environmental Information Data Centre (EIDC)
- Figshare
- IEDA:EarthChem
- IEDA:MGDS
- National Geoscience Data Centre (NGDC)
- NERC Earth Observation Data Centre (NEODC)
- PANGAEA
- Polar Data Centre (PDC)
- Zenodo



Pascal Aventurier : Les données de la recherche. Cours ENSSIB

The GBS dataset: measurements of satellite site diversity at 20.7 GHz in the UK

S. A. Callaghan*, J. Waight, J. L. Agnew, C. J. Walden, C. L. Wrench and S. Ventouras
STFC – Rutherford Appleton Laboratory, Chilton, UK

*Correspondence: S. A. Callaghan. *STFC – Rutherford Appleton Laboratory, Chilton, UK*

Dataset

The GBS (Global Broadcast Service) dataset comes as 3 separate data streams:

- Identifier: doi:10.5285/639A3714-BC74-46A6-9026-64931F355E07
Creator: Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [Callaghan, S. A., J. Waight, C. J. Walden, J. Agnew and S. Ventouras].
Title: GBS 20.7 GHz slant path radio propagation measurements, Chilbolton site
Publisher: NERC British Atmospheric Data Centre
Publication year: 2009
Resource type: Metadata document
Version: 1.0
- Identifier: doi:10.5285/db8d8981-1a51-4d6e-81c0-ccd9b921390
Creator: Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [Callaghan, S. A., J. Waight, C. J. Walden, J. Agnew and S. Ventouras].
Title: GBS 20.7 GHz slant path radio propagation measurements, Dundee site
Publisher: NERC British Atmospheric Data Centre
Publication year: 2009
Resource type: Metadata document
Version: 1.0
- Identifier: doi:10.5285/E8F43A51-0198-4323-A926-FE69225D57DD
Creator: Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [Callaghan, S. A., J. Waight, C. J. Walden, J. Agnew and S. Ventouras].
Title: GBS 20.7 GHz slant path radio propagation measurements, Sparsholt site
Publisher: NERC British Atmospheric Data Centre
Publication year: 2009
Resource type: Metadata document
Version: 1.0

Structure
Titre - Authors
Résumé
Dataset
Introduction
Experimental Set-Up
Acknowledgements
References

Le jeu de données de l'article



Centre for Environmental
Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

Search for in All

GBS 20.7GHz slant path radio propagation measurements, Chilbolton site

General Info

Title: GBS 20.7GHz slant path radio propagation measurements, Chilbolton site
Type: Activity
Sub-Type: Deployment
Publication State: Citable
URI: http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dep_11902119479621181

Summary

The GBS (Global Broadcast Service) dataset is a series of radio attenuation measurements made at three sites in the UK: Chilbolton and Sparsholt, both in southern UK, and Dundee in Scotland. The aim of the experiment was to make long term measurements of the signal strength received from a 20.7GHz beacon on the US Department of Defense satellite UFO-9 at multiple sites, in order to determine whether the use of site diversity as a fade mitigation technique would be effective. The dataset spans a period of 3 years, from August 2003 to August 2006 with signal attenuation sampled once per second.

Please cite this dataset as:

Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [S. A. Callaghan, J. Waight, C. J. Walden, J. Agnew and S. Ventouras]. GBS 20.7GHz slant path radio propagation measurements, Sparsholt site, [Internet]. British Atmospheric Data Centre, 2003-2005, 1st April 2011, doi:10.5285/639a3714-bc74-46a6-9026-64931f355e07

This dataset is cited in:

S. A. Callaghan, J. Waight, J.L. Agnew, C. J. Walden, C.L. Wrench, S. Ventouras "The GBS dataset: measurements of satellite site diversity at 20.7 GHz in the UK", Geoscience Data Journal, 17 March 2013, DOI: 10.1002/gdj3.2

http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dep_11902119479621181



SCIENCE & IMPACT

Pascal Aventurier : Les données de la recherche. Cours ENSSIB

53

Entrepôts de données

Définition : data repository

- **entrepôt de données= *data repository*,**
- Réservoir constitué majoritairement de données de recherche, brutes ou élaborées, qui sont décrites par des métadonnées de façon à pouvoir être retrouvées.
- Un identifiant est attribué à chaque ensemble déposé.
- analogie avec les entrepôts de publications (repository)
- Fonction de dépôt, classement,
- Optionnels : embargo, diffusion restreinte, statistiques de consultation
- outils de recherche, classement thématique
- Exposition en OAI-PMH, (open data repository)

Entrepôts de données

- Des annuaires de répertoires de données (res3data, datacite metadata search)
 - Plus de 400 répertoires de données
 - Thématique/généraliste
 - Organisation à but non lucratif/Organisation privée
- Points communs :
- Génère des identifiants pérennes
 - Permettent de diffuser les données en Open licence ou non
 - Métadonnées

Re3data.org

<http://www.re3data.org/> (20 mai 2013)



Home Search Suggest FAQ About Schema Contact Imprint

re3data.org launched

Posted on May 28, 2013 by re3data.org team

4

An increasing number of universities and research organisations are starting to build research data repositories to allow permanent access in a trustworthy environment to data sets resulting from research at their institutions. Due to varying disciplinary requirements, the landscape of research data repositories is very heterogeneous. This makes it difficult for researchers, funding bodies, publishers, and scholarly institutions to select an appropriate repository for storage of research data or to search for data.

The re3data.org registry allows the easy identification of appropriate research data repositories, both for data producers and users. The registry covers research data repositories from all academic disciplines. Information icons display the principal attributes

PARTNERS



INVA
SCIENCE & IMPACT

Pascal Aventurier : Les données de la recherche. Cours ENSIB

57

Search for repositories (605 repositories)

chemistry X Search

69 results for: chemistry

Subject	Content Type	Country (of the responsible institutions)
Add subjects	Add content types	Add countries

Certificates
 Open Access
 Persistent Identifier
 Repository reviewed by re3data.org

[remove filters](#)

69 results (1 - 25)

« 1 2 3 »

Atmospheric Science Data Center

NASA Langley Research Center, Atmospheric Science Data Center



Subjects: [Atmospheric Science](#) [Atmospheric Science and Oceanography](#) [Geosciences \(including Geography\)](#) [Natural Sciences](#)

Content types: [Archived data](#) [Audiovisual data](#) [Images](#) [Raw data](#) [Scientific and statistical data formats](#) [Standard office documents](#)

Countries: [United States](#)

The Atmospheric Science Data Center (ASDC) at NASA Langley Research Center is responsible for processing, archiving, and distribution of NASA Earth science data in the areas of radiation budget, clouds, aerosols, and tropospheric chemistry. The ASDC specializes in atmospheric data important to understanding the causes and processes of global climate change and the consequences of human activities on the climate.



DataCite

Recense plus de 480 sites de stockage de données

Fournit des DOI aux datasets

 **DataCite Statistics** Beta

Registrations by Allocators

Registrations by Datacentres

Registrations by Prefixes

Resolutions by Month

Allocator	DOI Registrations				Metadata Uploads				Metadata Ratio
	All Time	This Year	Last 30 Days	Last 7 Days	All Time	This Year	Last 30 Days	Last 7 Days	
ANDS - Australian National Data Service	2 617	102	7	0	1 711	102	7	0	65%
BL - The British Library	28 162	4 445	1 480	483	28 055	4 446	1 480	483	99%
CDL - California Digital Library	420 092	174 293	129 559	22 717	416 689	174 414	129 563	22 719	99%
CISTI - Canada Institute for Scientific and Technical Information	2 084	158	80	77	2 066	158	80	77	99%
CRUI - CRUI2011	7 697	4 596	4 107	0	4 570	1 469	960	0	59%
DATAcite - DataCite	7	0	0	0	7	0	0	0	100%
DELFT - TU Delft Library	7 190	4 498	123	30	6 914	4 500	125	32	96%
DK - Technical Information Center of Denmark	2 539	514	130	2	2 538	514	130	2	99%
ETHZ - ETH Zurich	411 106	32 404	1 270	440	411 103	79 837	17 044	15 665	99%
GESIS - GESIS - Leibniz Institute for the Social Sciences	7 455	1 400	1 034	21	7 194	1 534	1 034	21	96%
INIST - Institute for Scientific and Technical Information	394	294	29	16	394	294	29	16	100%
OSTI - Office of Scientific and Technical Information (OSTI), US Department of Energy	474	30	7	6	474	30	7	6	100%
PURDUE - Purdue University Library	2 162	1 724	39	3	2 161	1 724	39	3	99%
SND - Swedish National Data Service	2	1	0	0	2	1	0	0	100%



SCIENCE & IMPACT Pascal Aventurier : Les données de la recherche. Cours ENSSIB

59



Databib

[Find Repositories](#) | [Submit](#) | [Connect](#) | [About](#)

[Login/Register](#)

Featured Repository



UNdata

602 data repositories total in Databib.

Recently Added

USGS National Water Information System (NWIS)

CRYSTMET

ORGIDS - Open Rotterdam Glaucoma Imaging Data Sets

Databib is a searchable catalog / registry / directory / bibliography of **research data repositories**.

Search [Find](#) [Advanced Search](#)

Browse [[Subjects](#) | [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#) | [All](#)]

3

3TU.Datacentrum

A multidisciplinary data repository for a consortium of universities in the Netherlands housing over...

A

Access to Archival Databases (AAD)

The AAD is a database through the U.S. National Archives and Records Administration that allows user...

ACEpepDB: Peptide Database

ACEpepDB is a database ran by the Central Food Technological Research Institute. It contains records...

Addgene Plasmid Database

Addgene is a non-profit organization dedicated to making it easier for scientists to share plasmids...

Adult Blood Lead Epidemiology and Surveillance (ABLES) Interactive Database

<http://www.databib.org>



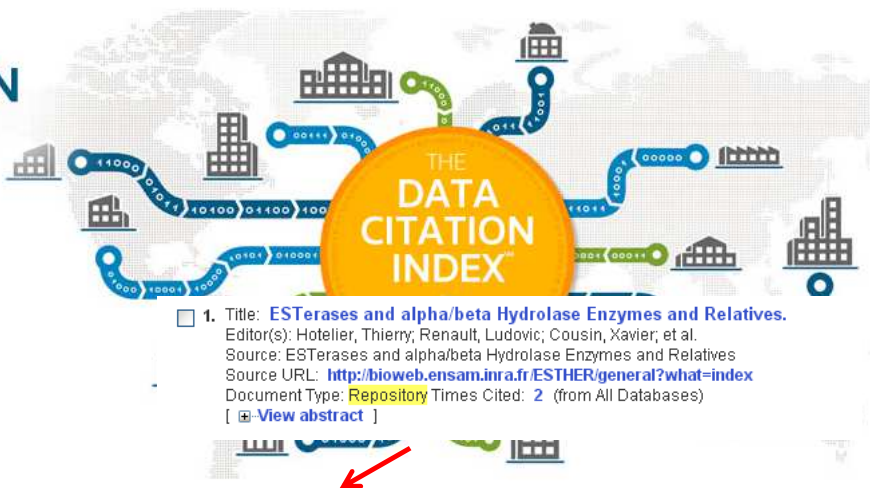
SCIENCE & IMPACT

Pascal Aventurier : Les données de la recherche. Cours ENSSIB

THE DATA CITATION INDEX™

CONNECTING THE DATA TO THE RESEARCH IT INFORMS

What is it? 
VIEW VIDEO



1. Title: **ESTerases and alpha/beta Hydrolase Enzymes and Relatives.**
 Editor(s): Hotellier, Thierry; Renault, Ludovic; Cousin, Xavier; et al.
 Source: ESTerases and alpha/beta Hydrolase Enzymes and Relatives
 Source URL: <http://bioweb.ensam.inra.fr/ESTHER/general?what=index>
 Document Type: **Repository** Times Cited: **2** (from All Databases)
 [[View abstract](#)]

1. Title: **Enzymatic Activity and Protein Interactions in Alpha/Beta Hydrolase Fold Proteins: Moonlighting Versus Promiscuity**
 Author(s): Marchot, Pascale; Chatonnet, Arnaud
 Source: PROTEIN AND PEPTIDE LETTERS Volume: **19** Issue: **2** Pages: **132-143** Published: **FEB 2012**
 Times Cited: **3** (from All Databases)
 [[View abstract](#)]

2. Title: **ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins**
 Author(s): Hotellier, T; Renault, L; Cousin, X; et al.
 Source: NUCLEIC ACIDS RESEARCH Volume: **32** Special Issue: **SI** Pages: **D145-D147** DOI: **10.1093/nar/gkh141** Published: **JAN 1 2004**
 Times Cited: **79** (from All Databases)
 [Full Text](#) [[View abstract](#)]



Data Citation Index



- Indexe les métadonnées de plus de 80 entrepôts de données **sélectionnés**
- Points forts
 - Pas d'autres outils équivalents
 - Intégration dans le Web of Knowledge (en supplément)
 - Une seule interface pour rechercher les données et les publications, utilisation des facettes
 - Lien vers les documents citants

Data Repositories

- [Dryad](#)
- [Figshare](#)
- [Pangea](#) (données sur la terre)
- [ICPSR](#) (Inter-university Consortium for Political and Social Research)
-
- Dryad : [Joint Data Archiving Policy \(JDAP\)](#)





Data Publishing Charges to help sustain open data at Dryad

Learn More »

Submit data now

[How and why?](#)

Search for data

Enter keyword, author, title, DOI, etc.

[Advanced search](#)

Browse for data

- Recently published
- Popular
- By Author
- By Journal

Recently Published Data

Lloyd MJ, Metaxas A, deYoung B (2012) Data from: Patterns in vertical distribution and their potential effects on transport of larval benthic invertebrates. *Marine Ecology Progress Series* doi:10.5061/dryad.20

Be part of Dryad



Publishers, societies, universities, libraries, funders, and other stakeholder organizations are invited to become



- 1 DEPOSIT DATA**
- 2 GET PERMANENT IDENTIFIER**
- 3 WATCH YOUR CITATIONS GROW!**
- 4 RELAX, YOUR DATA ARE DISCOVERABLE AND SECURE**

<http://dx.doi.org/10.5061/dryad.20>

Dryad [Joint Data Archiving Policy \(JDAP\)](#)

- Accord avec des journaux “classiques” qui exigent le dépôt des données dans Dryad
- Au minimum les données ayant servi à l'article

Journaux - 30

environ dont

BMC Ecology

BMC Evolutionary

Genetics

Journal of Heredity

Nature

Nucleic Acids

Research

Paleobiology

PLOS

Science

Métadonnées - Licence CCO

DOI

titre

Journal (vol Issue), Année

Auteurs – auteur contact

Résumé - Mots clés


Nom taxonomique

Localisation géo

Titre et description des données

Embargo




Research. Shared.

Search Communities Browse Upload Get started

Search 1019 records for

Filter by types ▼

Recent Uploads

31 October 2013 Dataset Open access

Raw Data for Studies 1 and 2 in False-Positive Psychology

Simmons, Joseph ; Nelson, Leif ; Simonsohn, Uri

Excel file containing data for experiments 1 and 2 reported <http://pss.sagepub.com/content/22/11/1359.short>

Uploaded by [uri](#) on 01 November 2013.

27 October 2013 Journal article Open access


EBI metagenomics—a new resource for the analysis and archiving of metagenomic data

Hunter, Sarah ; Corbett, Matthew ; Hubert, Denise ; Fraser, Matthew; *et al*

Metagenomics is a relatively recently established but rapidly expanding field that uses high-throughput next-generation sequencing technologies to characterize the microbial communities inhabiting different ecosystems (including oceans, lakes, soil, ...

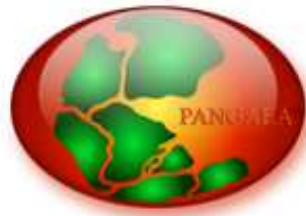
New to ZENODO?

- Research. Shared.** — all research outputs from across all fields of science are welcome!
- Citeable. Discoverable.** — uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
- Community Collections** — accept or reject uploads to your own community collections (e.g workshops, EU projects or your complete own digital repository).
- Funding** — integrated in reporting lines for research funded by the European Commission via OpenAIRE.
- Flexible licensing** — because not everything is under Creative Commons.
- Safe** — your research output is stored safely for the future in same cloud infrastructure as research data from CERN's Large Hadron Collider.
- DropBox integration** — upload files straight


Pascal Aventurier : Les données de la recherche. Cours ENSSIB
66

PANGAEA®

Data Publisher for Earth & Environmental Science



[Help](#) [Advanced Search](#) [Preferences](#) [more...](#)

[About](#) – [Submit Data](#) – [Projects](#) – [Software](#) – [Contact](#)



Pascal Aventurier : Une introduction aux données ouvertes de la recherche. Journée d'étude « Les enjeux de l'Open Access pour la recherche ». Agropolis international 17/06/2013

Certifier les entrepôts de données



Data seal of approval DSA – autoévaluation

(<http://assessment.datasealofapproval.org/apply>)

- Cinq principes

- Les données sont disponibles sur Internet,
- Les données sont accessibles, compte tenu de la législation applicable concernant les informations personnelles et la propriété intellectuelle,
- Les données sont disponibles dans un format utilisable,
- Les données sont fiables,
- Les données sont citables.

Le producteur (qualité des données), l'entrepôt (conservation et disponibilité) et l'utilisateur (usage) sont concernées dans la procédure.

European Framework for Audit and Certification of Digital Repositories



Trusted event

On July 8, 2010 a memorandum of understanding (MoU) was signed between three groups which are working on standards for Trusted Digital Repositories being David Giarretta in capacity as Chair of the [CCSDS](#) (Consultative Committee for Space Data Systems) ISO Repository Audit and Certification Working Group ([RAC](#)), Henk Harmsen in his capacity of Chair of the Data Seal of Approval ([DSA](#)) Board and Christian Keitel in his capacity as Chair of the [DIN](#) Working Group "Trusted Archives - Certification".

The parties to this Memorandum of Understanding all lead separate groups aiming at certifying digital repositories. They wish to put in place mechanisms to ensure that the groups can collaborate in setting up an *integrated framework* for auditing and certifying digital repositories.

The framework will consist of a sequence of *three* levels, in increasing trustworthiness:

- Basic Certification** is granted to repositories which obtain DSA certification;
- Extended Certification** is granted to Basic Certification repositories which *in addition* perform a structured, externally reviewed and publicly available self-audit based on ISO 16363 or DIN 31644;
- Formal Certification** is granted to repositories which *in addition* to Basic Certification obtain full external audit and certification based on ISO 16363 or equivalent DIN 31644.

Granting of these certificates will allow repositories to show one of three symbols (to be agreed) on their web pages and other documentation, in addition to any other DSA, DIN or ISO certification marks.

Trusted links

- [Data Seal of Approval](#)
- [DIN](#)
- [RAC](#)
- [CCSDS](#)
- [DANS](#)
- [Landesarchiv BW](#)
- [CINES](#)
- [DNE](#)
- [UKDA](#)

[Signed MoU](#)

Trusted repositories



The MoU was signed as part of a series of EC sponsored initiatives on the audit and certification of trusted digital repositories.

The EC have been hugely supportive of this initiative and will be examining the results of at least three trial audits against the DSA and the DIN/ISO standards early 2011.

<http://trusteddigitalrepository.eu>

herche. Cours ENSSIB

Gérer les données

Citation des données métadonnées plan de gestion des données

Citation des données

Pourquoi citer les données ?

- Reconnaissance des « producteurs » → incite à la publication ouverte des données
- Lien article et jeu de données : augmentation de la citation de l'article
- Nécessite un mécanisme robuste de citation
- Collaboration, Financement
- Permet d'établir des métriques sur la citation

Ball, A. & Duke, M. (2012). 'How to Cite Datasets and Link to Publications'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides> - See more at: <http://www.dcc.ac.uk/resources/how-guides/cite-datasets#x1-12000>



Pascal Aventurier :

Éléments pour la citation de données

(Ball & Duke 2012, DCC)

Analyse de plusieurs sources bibliographiques

Author. Le createur du jeu de données

Publication date. Date ou le jeu de données est disponible ou la date de la période d'embargo ou la date ou la procédure d'assurance qualité est complète

Title. Nom de la ressource en elle même mais peut inclure le nom de l'organisme ou le nom de la collection si il s'agit d'une sous collection

Edition. niveau d'état de traitement des données

Version.

Feature name and URI. Nom d'une caractéristique ISO 19101:2002¹ 'feature' (e.g. GridSeries, ProfileSeries)

Resource type. Exemples: 'database', 'dataset'.

Publisher. organisation qui détient les données ou en assure la qualité

Unique numeric fingerprint (UNF). Cryptographe pour assurer qu'aucune modification n'a eu lieu **ou un identifiant selon un schéma perenne**

Location. A persistent URL or URI



Ball & Duke 2012, DCC

APA

Cool, H. E. M., & Bell, M. (2011). Excavations at St Peter's Church, Barton-upon-Humber [Data set]. doi:10.5284/1000389

Chicago (notes)

2. H. E. M. Cool and Mark Bell, Excavations at St Peter's Church, Barton-upon-Humber (accessed May 1, 2011), doi:10.5284/1000389.

Cool, H. E. M., and Mark Bell. Excavations at St Peter's Church, Barton-upon-Humber (accessed May 1, 2011). doi:10.5284/1000389.

MLA

Cool, H. E. M., and Mark Bell. "Excavations at St Peter's Church, Barton-upon-Humber." Archaeology Data Service, 2001. Web. 1 May 2011. <<http://dx.doi.org/10.5284/1000389>>.

Oxford

Cool, H. E. M. & Bell, M. (2011) Excavations at St Peter's Church, Barton-upon-Humber [data-set]. York: Archaeology Data Service [distributor] <DOI:10.5284/1000389>.

Figure 1: Data citations in common styles[12]

PANGAEA

Willmes, S et al. (2009): Onset dates of annual snowmelt on Antarctic sea ice in 2007/2008. doi:10.1594/PANGAEA.701380

Dryad

Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hill CE, Hoang A, Gibert P, Beerli P (2001) Data from: The strength of phenotypic selection in natural populations. Dryad Digital Repository. doi:10.5061/dryad.166

Dataverse

Frederico Giroi; Gary King, 2006, 'Cause of Death Data', <http://hdl.handle.net/1902.1/UOVMCPSWOL>
UNF:3:9JU+SmVyHgwRhAKclQ85Cg== IQSS Dataverse Network [Distributor] V3 [Version].

Figure 2: Data citation formats suggested by repositories



Pascal Aventurier : Les données de la recherche. Cour

Pré-requis pour la citation de données

- La citation doit être capable d'identifier de manière unique l'objet cité Mais différentes citations peuvent utiliser différentes méthodes pour le faire
- Elle doit être capable d'identifier des sous-ensembles des données aussi bien que l'ensemble complet
- Elle doit donner au lecteur assez d'information pour accéder aux jeux de données et fournir un mécanisme pour accéder aux jeux de données à travers l'infrastructure du web
- Doit pouvoir être utilisé par des humains /des machines pour construire de nouveaux services.

Voir : <http://www.dcc.ac.uk/resources/how-guides/cite-datasets#x1-4000>



Duke, M. (2011, Aug. 22). Requirements for data citation: The prequel [Blog post]. Retrieved 22 Aug. 2011, from the SageCite blog: <http://blogs.ukoln.ac.uk/sagecite/2011/08/22/requirements-for-data-citation-the-prequel/>. - See more at: <http://www.dcc.ac.uk/resources/how-guides/cite-datasets#fn6x0>

Les recommandations Datacite

(<http://www.datacite.org/whycitedata>)

5 éléments obligatoires de son schéma de métadonnées (<http://schema.datacite.org/>)

de la manière suivante : **Creator (PublicationYear): Title. Publisher. Identifier.**

Dans le cas où il est souhaitable d'inclure des informations des propriétés optionnelles Version et ResourceType : **Creator (PublicationYear): Title. Version. Publisher. ResourceType. Identifier.**

Métadonnées des jeux de données

document commenté en séance

**Vocabulary to describe Research Data
Repositories online**

<http://www.re3data.org/2012/07/vocabulary-to-describe-research-data-repositories-online/>

DataCite Metadata Schema v 3.0

DataCite

[DataCite Metadata Schema 3.0 XML Schema](#)
doi:10.5438/0009

Documentation

[DataCite Metadata Schema 3.0](#)
doi:10.5438/0008

Examples

- [Example for a simple dataset](#)
- [Example with complicated values](#)
- [Example with DataCollector as Contributor and a geoLocation box](#)
- [Example with GeoLocation](#)
- [Example with HasMetadata as related resource](#)
- [Example with IsIdenticalTo as related resource](#)
- [Example with ResearchGroup as Contributor and Methods as Description](#)
- [Example with Collection as ResourceType](#)
- [Example with Video as ResourceType](#)
- [Example for a workflow ResourceType](#)



INRA
SCIENCE & IMPACT

<http://schema.datacite.org/meta/kernel-3/index.html>

Pascal Avenirier : Les données de la recherche. Cours ENSSIB

79

In this section

- Briefing Papers
- How-to Guides
- Developing RDM Services
- Curation Lifecycle Model
- Curation Reference Manual
- Policy and legal
- Data Management Plans
- Tools
- Case studies

Disciplinary Metadata

While data curators, and increasingly researchers, know that good metadata is key for research data access and re-use, figuring out precisely what metadata to capture and how to capture it is a complex task. Fortunately, many

Search by Discipline



Biology



Earth Science



General Research Data



Physical Science



Social Science & Humanities

Plan de gestion des données Data Management Plan

Plan de gestion des données

- **MIT** <http://libraries.mit.edu/guides/subjects/data-management/>
- **Digital Curation Center**
<http://www.dcc.ac.uk/resources/data-management-plans>
- [DMP Online](#) un outil pour créer des plans personnalisés selon le contexte du financer de la recherche. Pour les chercheurs : exemples, guides et bonnes pratiques
- [Funders' data plan requirements](#) : résumés des attentes des financeurs
Summary of funders' expectations for data management and sharing plans
- [Checklist for a Data Management Plan](#)
- **Monash Univ**
<http://monash.edu/library/researchdata/index.html>

Outil pour la création d'un plan de gestion des données <https://dmp.cdlib.org>



[Contact Us](#) | [Get Started](#) | [Login](#)

[Home](#) | [About DMP Tool](#) | [DMP News](#) | [My Plans](#) | [Funder Requirements](#) | [Help](#)



The DMP Tool allows you to: **1** **2** **3** **4**

[Get Started!](#)

Data Management Plan: Sample Plan Created at the DataONE Best Practices Workshop - Santa Fe NM 7/2011 Atmospheric CO2 Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013

1. Types of data produced

An archive at Mauna Loa Observatory will be created continuously from an archive located at the level of a central data center for several decades of continuous operation. This data will include atmospheric CO2 concentrations, carbon dioxide isotopes, methane concentration, and other greenhouse gas data. The program will include data quality control to ensure the reliability of data which is essential for research. In addition to the CO2 data, we will report major data level based on monthly, quarterly, yearly, semi-annual, and annual cycles. The archive will be a long-term data repository and will be used for research. The first data product will consist of 5-minute, 15-minute, hourly, daily, and monthly average atmospheric concentration of

[See a plan created with the DMP Tool](#)

Recent DMP News

[DMPTool's Top 10](#)

[Data Management Resources: Libguides](#)

[DMPTool2 Project - May 2013 Report](#)

[More news >](#)



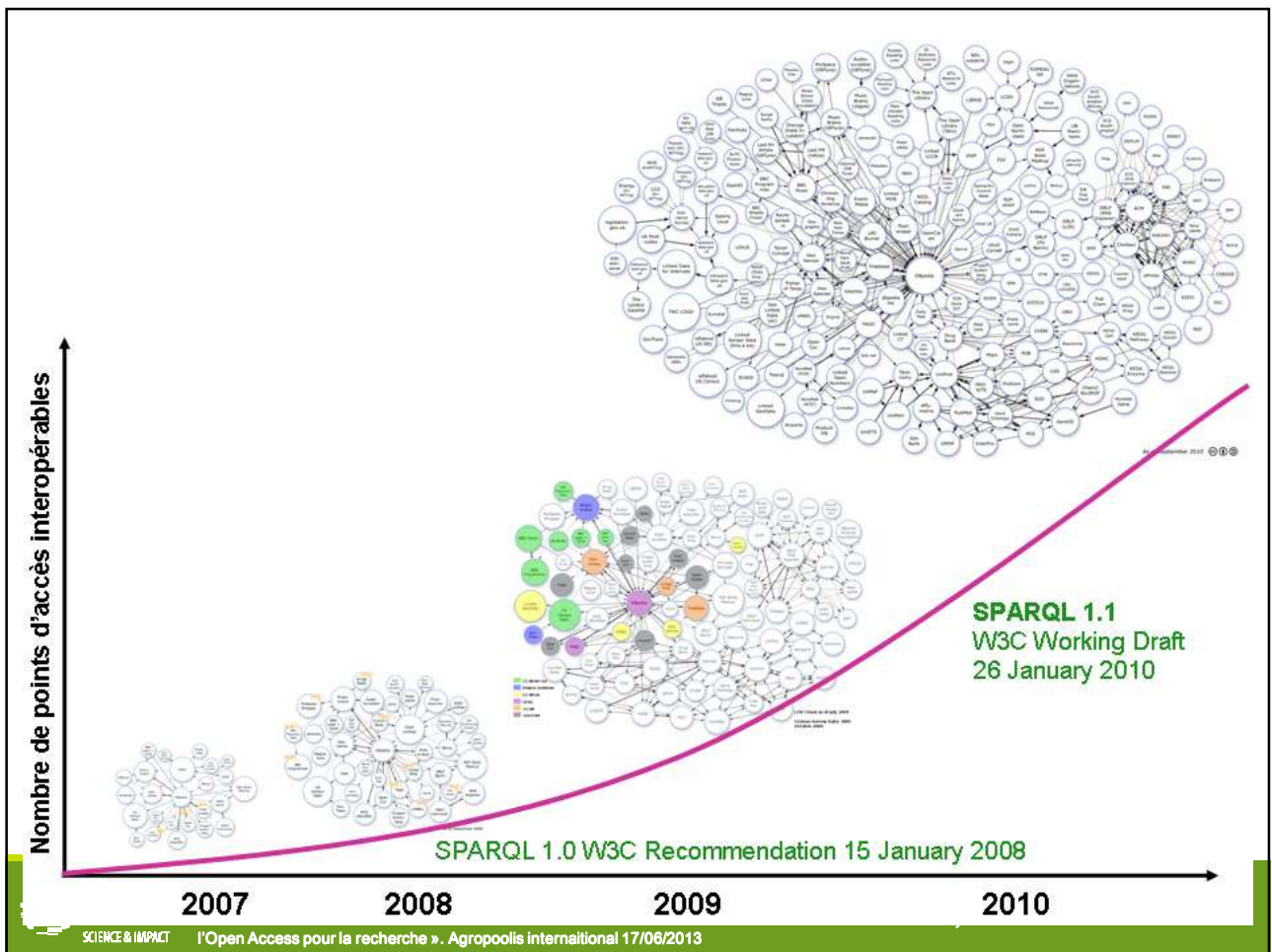
[University of California Curation Center](#) of the [California Digital Library](#)

Que contient un plan de gestion des données ?

<http://monash.edu/library/researchdata/advice/data-managment-brochure.pdf>

1. Copyright, Propriété intellectuelle (IP), éthique : comment sont-ils reliés à vos données de recherche ?
2. Où stocker les données numériques et les sauvegarder ?
3. Méthodes pour organiser les données et la documentation les décrivant
4. Partager vos données: au bon moment et avec les bonnes personnes
5. Optimiser l'impact de vos recherches en diffusant vos données et des informations de base
6. Prendre des dispositions pour les données après la fin de votre projet de recherche

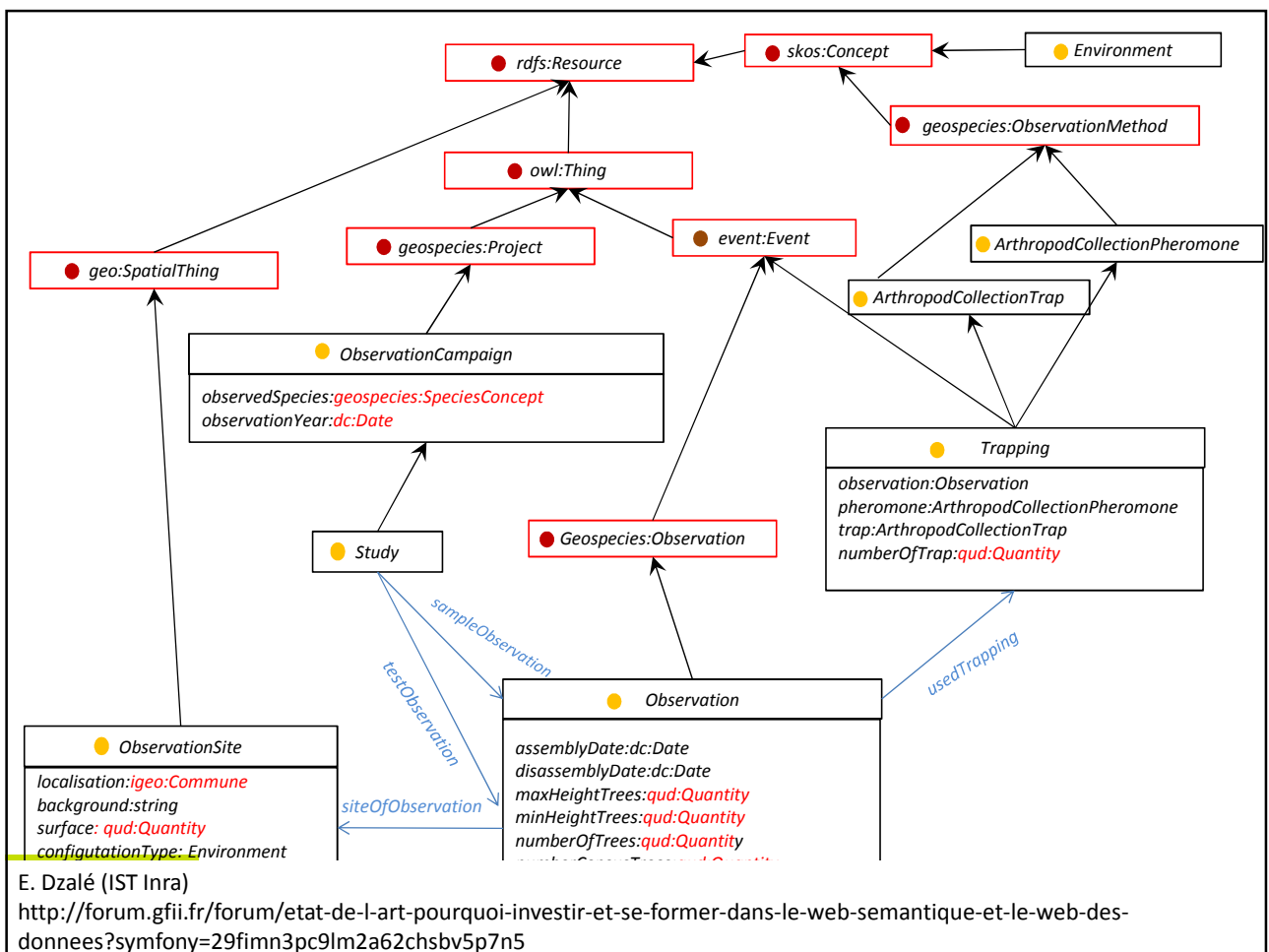
Données sémantiques et liés



Lier les données

- Utiliser des formats qui permettent de lier ses données avec les données d'autres sources
- Exemple : Données de recherche et thesaurus, lui-même relier à d'autres ressources (catalogue de documents)
- Pour
 - Enrichir son jeu de données
 - Construire de nouvelles interfaces intelligentes
 - Nouveaux modes de recherche
 - Mieux documenter les données, à travers des ontologies

A	B	C	D	E	F	G	H	I	J	K	L
	Communes	Saint-Fargeau-Ponthierry	Obernai	Angers							Saint-Médard-de-Rouvray
	Nombre de pins et/ou de cèdres	96 pins	20 pins	97 pins 3 cèdres	18 pins	66 pins	22 pins	11 pins 3 cèdres	6 pins	31 pins	261 pins
	Hauteur des arbres	15 à 20 mètres	non relevé	10 à 15 mètres	10 à 25 mètres (avec un seul à 5 mètres)	non relevé	3 mètres à 20 mètres	5 mètres à 15 mètres	15 mètres à 20 mètres	4 mètres à 15 mètres	2 mètres à 10 mètres
Pièges à phéromones été 2011	Relevés des nids réalisés à l'hiver 2011	non réalisés	non réalisés	non réalisés							non réalisés témoignage
	Nombre de pièges installés	10	15	14 a) 1 b)	9	12 a) 1 b)	12 a) 2 b)	8 a) 1 b)	6 a) 1 b)	8	18
	Type de pièges	Procerex	Mastrap	a) Mastrap b) Nufarm	Mastrap	a) Mastrap b) Nufarm	a) Mastrap b) Nufarm	a) Mastrap b) Nufarm	a) Mastrap b) Nufarm	Mastrap	Mastrap
	Type de phéromones	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex
	Date de la pose des pièges	25 juillet (sem 30)	01 juin (sem 22)	01 juillet (sem 26)			01 juillet (sem 26) et 06 juillet (sem 27)		31 mai (sem 22) et 19 juillet (sem 29)	19 juillet (sem 29)	19 juillet (sem 29)
Premiers vols de papillons constatés	non communiqué	Sem 26 et 27	sem 25							non relevé	

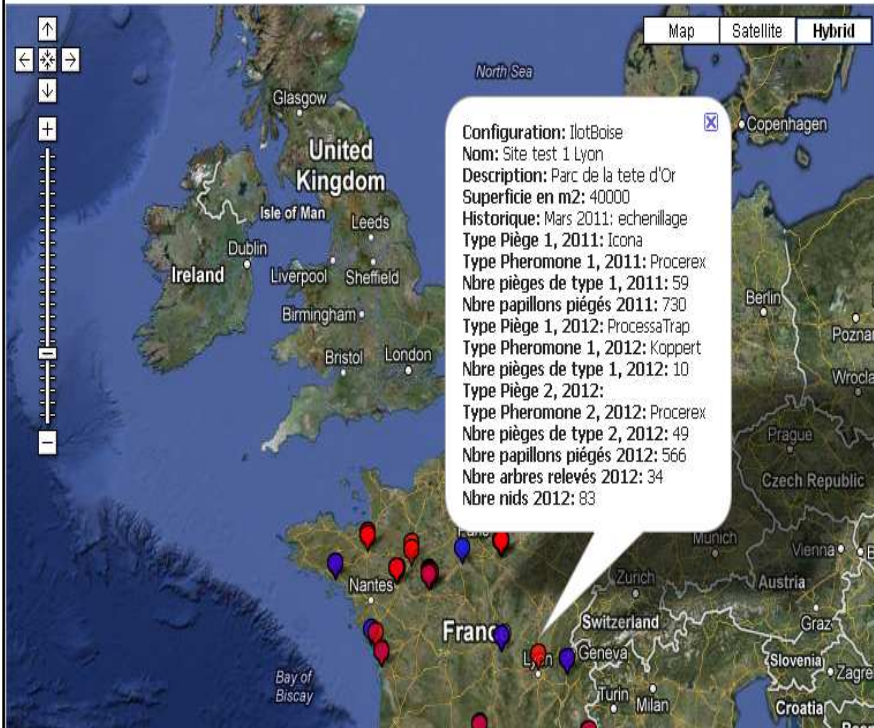


Sites d'observation

[CARTE](#) • [DETAILS](#)

69 SiteObservation

6 résultats sur 69 ne peuvent pas être tracés.



Search

Localisation

- 2 Aix-les-Bains
- 8 Angers
- 2 Arzon
- 4 Avignon
- 3 Bayonne
- 2 Entraunes

Configuration

- 1 (missing this field)
- 3
- 14 Alignement
- 3 ArbreSeul
- 48 IlotBoise

Type Site

- 23 Temoin
- 46 Test

[get data file](#)

Data librarian et rôle de l'IST



LIGUE DES BIBLIOTHÈQUES
EUROPÉENNES DE RECHERCHE
ASSOCIATION OF EUROPEAN
RESEARCH LIBRARIES

Login

Re-inventing the Lib

HOME ABOUT **NEWS** BLOG ACTIVITIES EVENTS JOIN PARTNERS SPONSORS

24
Aug
2016

Ten recommendations for libraries to get started with research data management

...

Offrir un plan de gestion des données

- S'engager dans le développement de standards de métadonnées
- Participer activement aux plans de gestion des données de votre institut
- Supporter le cycle de vie des données en fournissant des services pour le stockage, la « découverte » et l'accès permanent
- Mettre en place un catalogue des données
- Mettre en place des identifiants pérennes

...

<http://www.libereurope.eu/news/ten-recommendations-for-libraries-to-get-started-with-research-data-management>

Data Scientist Training for Librarians

HOME | COURSE DETAILS | PARTICIPANTS | PREVIOUS COURSE | BLOG | CONTACT

“ The ability to take data — to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it — that’s going to be a hugely important skill in the next decades.”

Dates/Times:

July 26 – [Intro to DST4L](#)

Aug 23-24 – [Software Carpentry Bootcamp](#)

Aug 28 – Nov 26 – [Classes by T. Morris, R. Dave, L. Cherny](#)

Dec 14 – Projects/Class Ends

Mainly Tues 10-1 & 11-2 also Wed, Thurs (see syllabus)

General Course Outline:

Overview, Data Sources, Data Extraction, Data Cleansing, Sharing, Visualization, Presentation, Advanced Topics

Technologies

Unix Shell, Git, Python, OpenRefine, Excel

* D3, R, Tableau

PDF extraction, JSON/XML wrangling, HTML page scraping, pandas and viz, NLTK/scikit text extraction, text-tools/RegEx, recommender system, basic statistics, simple ml/naive bayes, static viz, dynamic viz, OpenRefine and tools, Excel spreadsheet usage, basic NumPy/matplotlib

The Data Scientist Training for Librarians Course will provide you with the skills and knowledge that they can better serve the data needs of your library.

The class will be provided with real world use cases and hands-on experience with data extraction, cleansing, sharing and presentation. You will also look into telling a story with data through visualization. A variety of tools currently used in the industry will be introduced to you.

At the end of the course, participants will be able to:


<http://altbibl.io/dst4l>



SCIENCE & IMPACT

Pascal Aventurier : Les données de la recherche. Cours ENSSIB

93

 because good research needs good data [Contact us](#)

[Home](#) | [Digital curation](#) | [About us](#) | [News](#) | [Events](#) | [Resources](#) | [Training](#) | [Projects](#) | [Community](#)

Home > Digital curation training for all > Data management and curation education and training > The role of data management for various professions

In this section

- Tools of the Trade training
- Digital Curation 101
- Materials for Trainers
 - Career profiles and related data management skills**
 - DC 101 training materials
 - Disciplinary RDM training
- Data management courses and training
- DaMSSI
- RDM for librarians
- Skills frameworks

The role of data management for various professions

As part of our work on the JISC/RIN funded [Data Management Skills Support Initiative \(DaMSSI\)](#), the DCC and RIN have produced a series of career profiles that aim to demonstrate how data management skills contribute to and underpin high-quality performance in a number of professions. These profiles can be helpful for:

- illustrating potential career paths for both undergraduate and graduate programmes
- promoting professional development training courses
- engaging with professional bodies

Professions covered by the series so far include:

- [Librarians](#)
- [Conservators](#)
- [Social science researchers](#)
- [Archaeologists](#)
- [Clinical psychologists](#)
- [Data managers](#)

Useful links

- [DaMSSI](#)
- [RIN](#)
- [JISC MRD programme](#)

Lien vers la plaquette
http://www.dcc.ac.uk/webfm_send/1345

Conclusion

Conclusion

- Données n'appartiennent pas aux éditeurs : contexte différent des publications
- Des compétences à développer
- Droit et licence : un élément clé pour les données
- However, libraries are particularly well positioned to fill the data-management gap, says Michener. They have long experience helping faculty members and they are not likely to disappear, he says. "Libraries have a significant amount of trust capital."*



* <http://www.nature.com/news/publishing-frontiers-the-library-reboot-1.12664>

SCIENCE & TECHNOLOGY Pascal Aventurier : Les données de la recherche. Cours ENSSIB

96

Site d'information sur les

Données de la Recherche

Rechercher

OK

ACCUEIL NORMES POLITIQUES PROJETS MÉTIERS WEBOGRAPHIE ARCHIVES

> Accueil

Partager Imprimer

Ce site, créé à l'initiative du ministère de l'Enseignement supérieur et de la Recherche, est une source d'informations sur les données de la recherche.

Lire la suite

Actualités

Call for expressions of interest : EC public consultation on open research data

2 juillet 2013 - Bruxelles, Belgique

Dear Sir, dear Madam, Open enquiry is at the heart of scientific endeavour, and rapid technological change has profound implications for the way that science is both conducted and communicated. (...)

Lire la suite

Charte Éthique & Big Data

La charte Éthique & Big Data est une co-construction d'acteurs majeurs, académiques et industriels, pour faciliter la création, la diffusion et l'utilisation des grands volumes de données (Big Data) (...)

Événements

International Information Conference on Search, Data Mining and Visualization
>14 et 15 Avril 2014 - Nice, France

Commodity, catalyst or change-agent ?
Data-driven transformations in research, education, business & society
>Du 24 au 27 février 2014 - San Francisco, États-Unis

APA International Conference on Digital Preservation and Development of Trusted Digital Repositories
>5 et 6 février 2014 - New Delhi, Inde

MTSR 2013 : 7th Metadata and Semantics Research Conference
>Du 19 au 22 novembre 2013 - Thessalonique, Grèce



Inist



Prêt pour un Data Cake ?

Merci de votre attention

Pascal.Aventurier@paca.inra.fr



<http://epicgraphic.com/data-cake/>

- OECD Principles and Guidelines for Access to Research Data from Public Funding <http://www.oecd.org/science/sci-tech/38500813.pdf>
- Dossier Nature <http://www.nature.com/news/specials/scipublishing/index.html>
- Royal Society **Final report**, case studies of **data use** and **data repositories** and **the launch event** published June 2012
<http://royalsociety.org/policy/projects/science-public-enterprise/report/>
- Research Data Curation Bibliography . Charles W. Bailey, Jr. <http://digital-scholarship.org/rdc/rdc.htm>
-