

A preservação digital num repositório de Dados de Pesquisa : o caso do IRD

Pascal Aventurier

Responsável do Serviço de Informação Científica e Tecnológica do IRD - Instituto Francês de Pesquisa para o Desenvolvimento

<https://orcid.org/0000-0003-0211-4549> @Pascal_AV

Blog <https://publicient.hypotheses.org/>



Attribution 4.0 International (CC BY 4.0)

**Palestra para a rede
Cariniana
8 de dezembro de 2020**



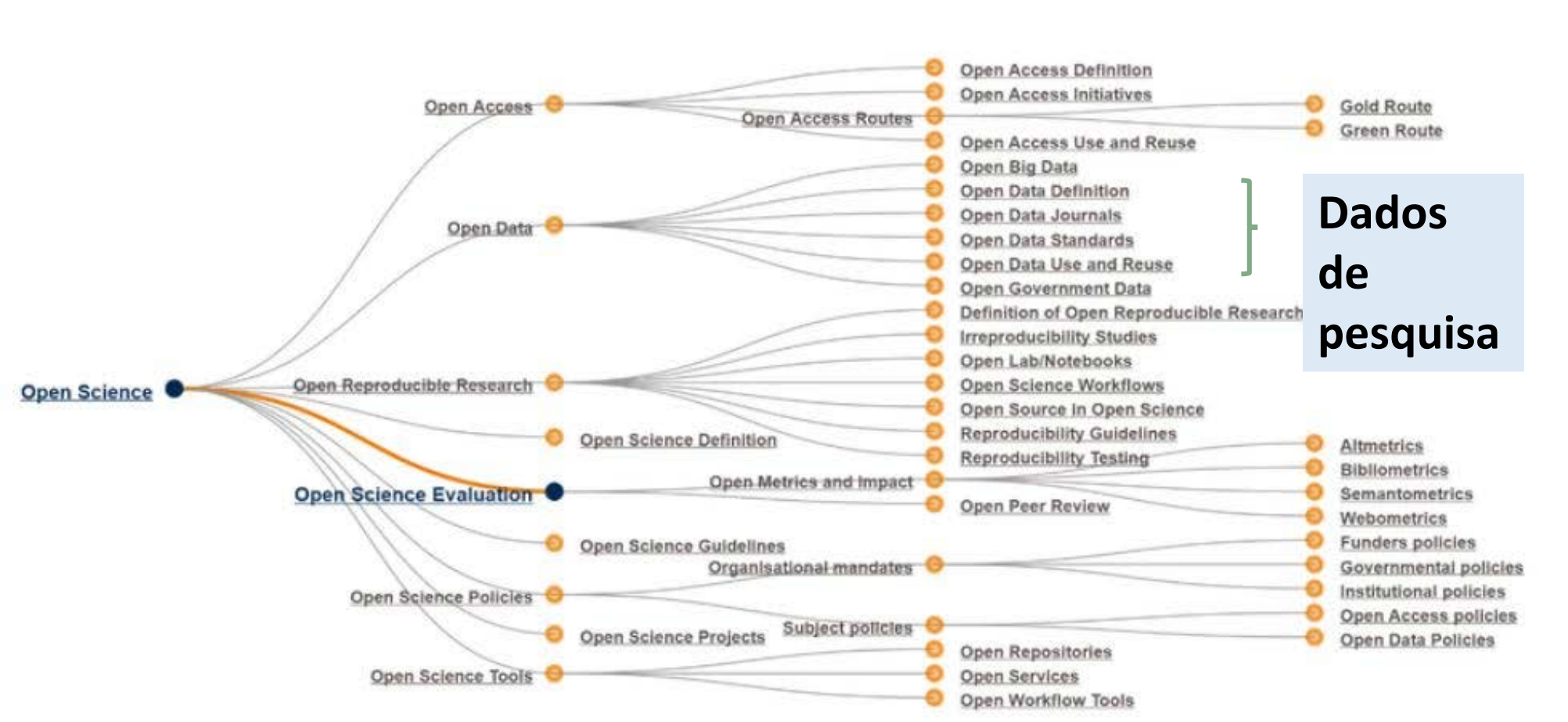
Cariniana

Rede Brasileira de Serviços de
Preservação Digital



ibict
Instituto Brasileiro de Informação
em Ciência e Tecnologia

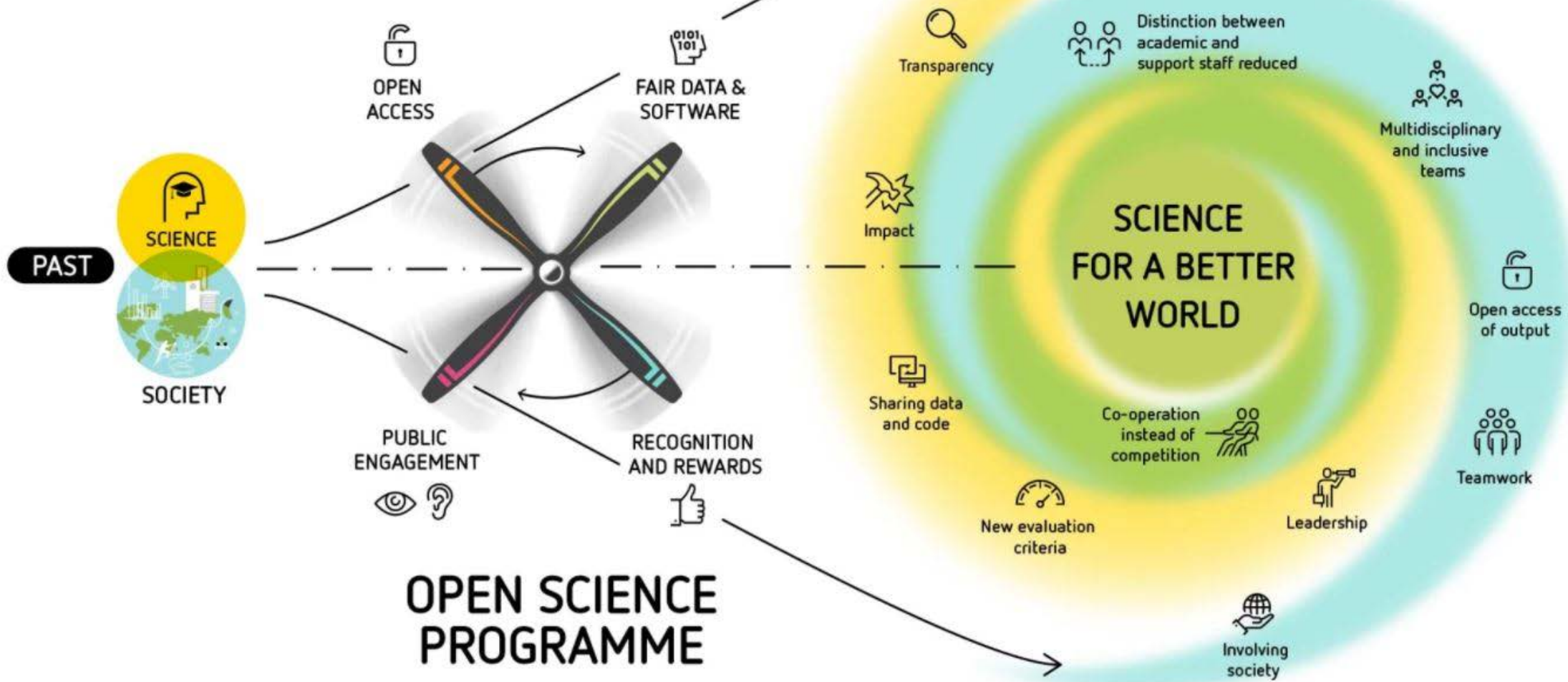
Ciência aberta : compartilhar conhecimento científico sem barreiras jurídicas ou tecnológicas





TOWARDS OPEN SCIENCE

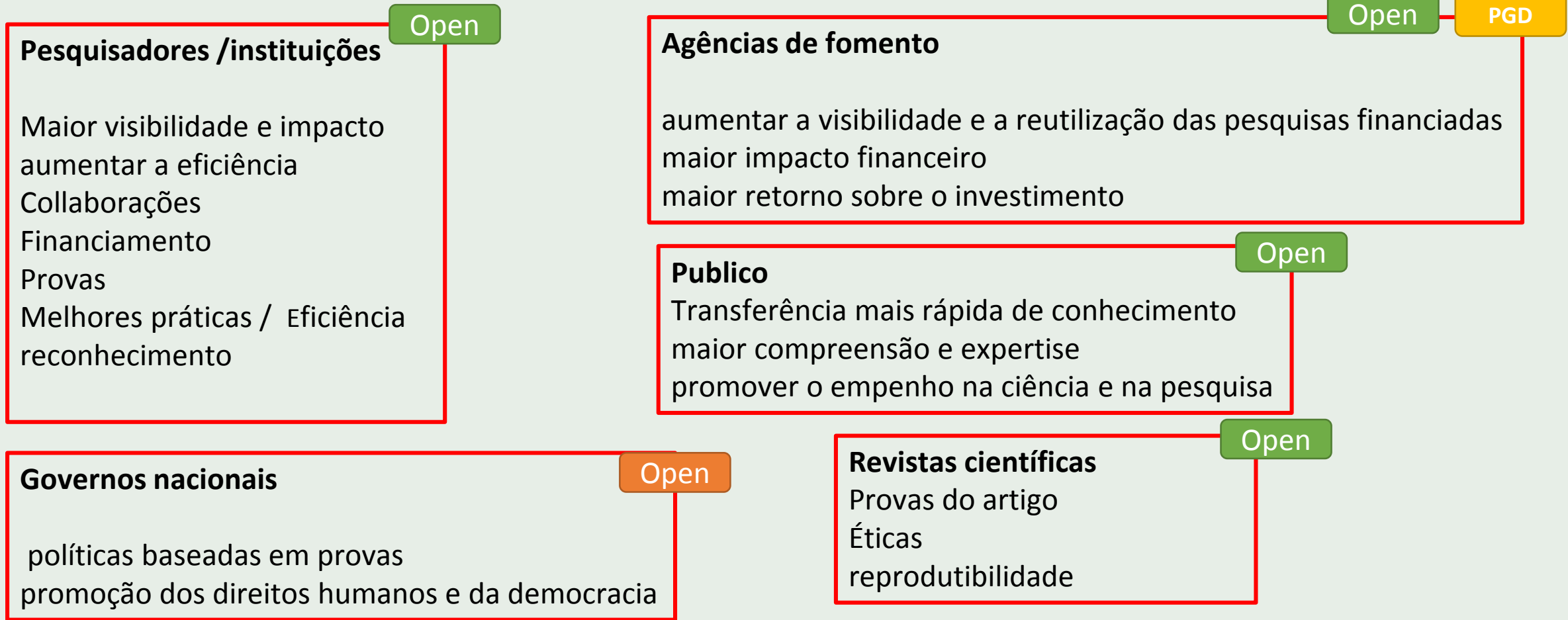
FUTURE





Porque compartilhar Dados ?

- Open** Exigência por compartilhar/disponibilizar
- Open** Incitação por compartilhar/disponibilizar
- PGD** Plano de gestão de dados



Open as possible , Close as necessary

Ciência dos Dados

PLOS NEGLECTED TROPICAL DISEASES

advanced search

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Spatio-temporal dynamics of dengue in Brazil: Seasonal travelling waves and determinants of regional synchrony

Mikhail Churakov, Christian J. Villabona-Arenas, Moritz U. G. Kraemer, Henrik Salje, Simon Cauchemez

Version 2 Published: April 22, 2019 • <https://doi.org/10.1371/journal.pntd.0007012>
• >> See the preprint

Article	Authors	Metrics	Comments	Media Coverage
---------	---------	---------	----------	----------------

Abstract

- Author summary
- Introduction
- Materials and methods
- Results
- Discussion
- Supporting information
- Acknowledgments
- References

Reader Comments (0)
Media Coverage (0)

Abstract

Dengue continues to be the most important vector-borne viral disease globally and in Brazil, where more than 1.4 million cases and over 500 deaths were reported in 2016. Mosquito control programmes and other interventions have not stopped the alarming trend of increasingly large epidemics in the past few years. Here, we analyzed monthly dengue cases reported in Brazil between 2001 and 2016 to better characterise the key drivers of dengue epidemics. Spatio-temporal analysis revealed recurring travelling waves of disease occurrence. Using wavelet methods, we characterised the average seasonal pattern of dengue in Brazil, which starts in the western states of Acre and Rondônia, then travels eastward to the coast before reaching the northeast of the country. Only two states in the north of Brazil (Roraima and Amapá) did not follow the countrywide pattern and had inconsistent timing of dengue epidemics throughout the study period. We also explored epidemic synchrony and timing of annual dengue cycles in Brazilian regions. Using gravity style models combined with climate factors, we showed that both human mobility and vector ecology contribute to spatial patterns of dengue occurrence. This study offers a characterization of the spatial dynamics of dengue in Brazil and its drivers, which could inform intervention strategies against dengue and other

Usar dados sem produzi-los !

Dados de SINAN , IBGE, DATASUS...

0 Save	3 Citation
4,097 View	0 Share

Download PDF

Print

Share

Check for updates

ADVERTISEMENT



Dengue cases data

Cases data were obtained from the Notifiable Diseases Information System (SINAN, Sistema de Informação de Agravos de Notificação) via the Health Information Department (DATASUS, Departamento de Informática do Sistema Único de Saúde) run by the Brazilian Ministry of Health. Monthly dengue data were extracted from January 2001 to August 2016 for each municipality (n = 5570). Cases were confirmed by clinical and epidemiological evidence, and approximately 30% of them were also laboratory-confirmed [24].

Population data

Municipality level human population data were retrieved from the Brazilian Institute of Geography and Statistics (IBGE, Instituto Brasileiro de Geografia e Estatística) for the year 2014 (http://www.ibge.gov.br/home/estatistica/populacao/estimativa2014/estimativa_dou.shtm)

Climate data

Precipitation data were obtained from the Climate Prediction Center's (CPC) rainfall data for the world (1979 to present, 50 km resolution) via raincpc R package [25] and aggregated for each Brazilian municipality over the study period. We also retrieved data on mean monthly surface temperature from Reanalysis CFSR model [26] (reference: CREATE-IP.reanalysis.NOAA-NCEP.CFSR.atmos.mon, data node: esgf.nccs.nasa.gov).

Vector suitability data

We obtained average monthly estimates of vector suitability for each Brazilian municipality using the modelling outputs from [27,28]. Mosquito occurrence data were fitted to annual data and covariates (i.e. time varying temperature-persistence suitability, relative humidity, precipitation and a static urban vs. rural covariate) and then subsequently re-applied to the same covariates at a monthly level. Maps were produced at a 5 x 5 kilometre resolution, aggregated to the municipality level and the mean value was used for our model. For consistency, we rescaled the monthly suitability values, so that the sum of all monthly maps equalled the annual mean map.



Definição de dados de pesquisas

- “registros factuais usados como fonte primária para a pesquisa científica e que são comumente aceitos pelos pesquisadores como necessários para validar os resultados do trabalho científico”. (OCDE 2007)
- As informações sobre a produção dos dados devem ser também disponibilizadas, isto é, devem ser explicitados a localização, a data, o protocolo, a ferramenta para a análise e o código computacional.

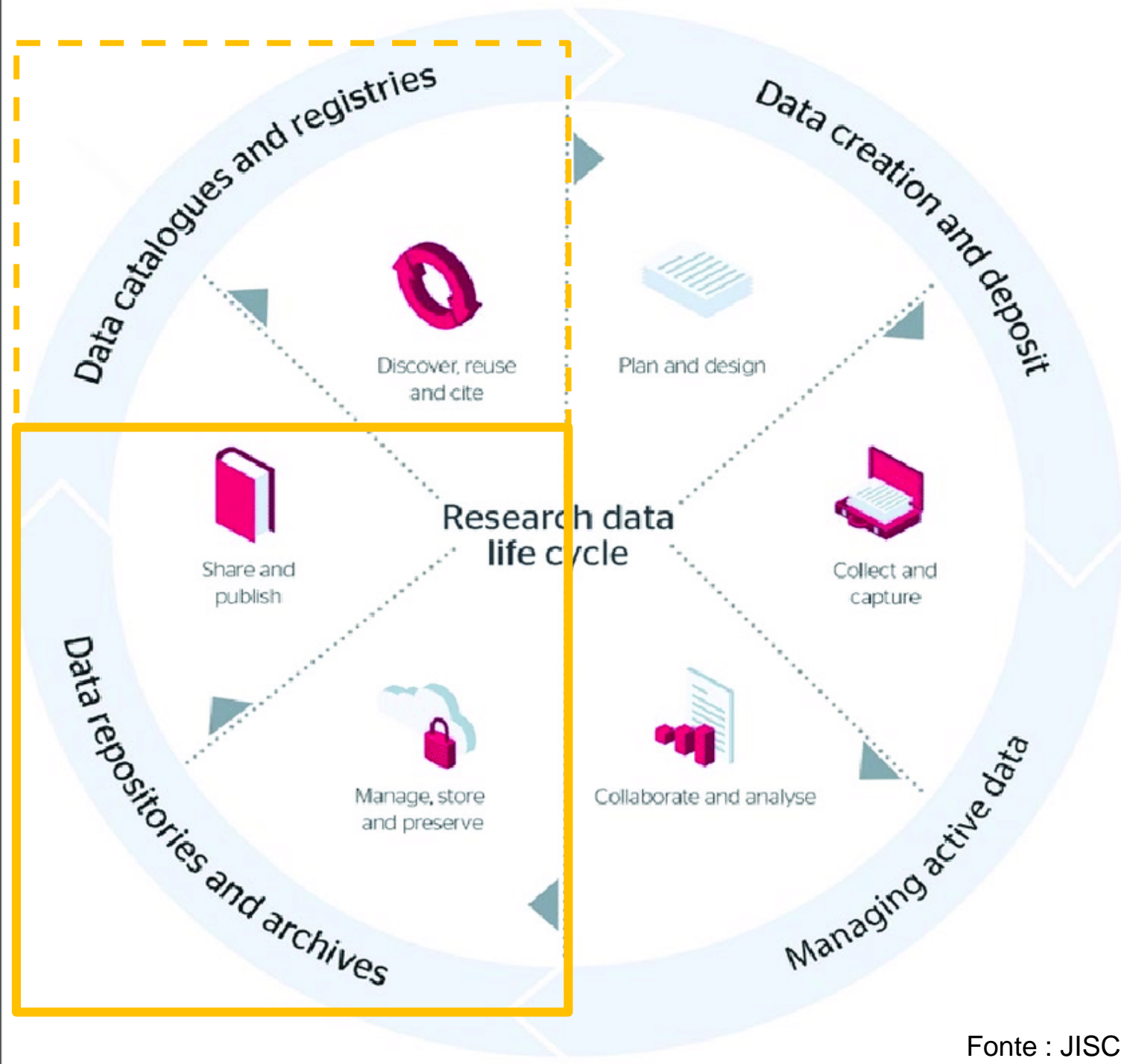
Definição de dados de pesquisas (2)

- **Observacionais**, obtidos por meio de observação;
- **Computacionais**, resultantes de execução de modelos computacionais ou de simulação;
- **Experimentais**, provenientes de situações controladas em bancadas de laboratórios.
- **Genômicas** ...

Dados de pesquisa são diferente de dados abertos governamentais

Repositórios

no ciclo de vida dos Dados



Repositórios de dados

Repositório

- Depositar
- Descrever
- Preservar
- Pesquisar
- Disseminar
- Citar dados

Dados em elaboração

Cadernos de laboratorios

Hard Drive or Espaço compartilhado

Dados em em processamento

Github
R/Python
Excel

Hard Drive or Espaço compartilhado

Dados versão final

Descrição
DOI (Digital Object Identifier)
Arquivos em formato padrão

Repositórios

Princípios FAIR

Certificação



<https://www.coretrustseal.org/>

Tipo de repositório

+2200 repositórios
<https://www.re3data.org/>

Institucional



Temático



Editores



Nacional / Regional



Generalistas



Open source research data repository software



Researchers

Enjoy full control over your data. Receive *web visibility, academic credit, and increased citation counts*. A personal Dataverse collection is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. *Want to set up your personal Dataverse collection?*



Journals

Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data*. Participate in the open data movement by using a Dataverse collection as part of your journal data policy or list of repository recommendations. *Want to find out more about journal Dataverse collections?*



Institutions

Establish a research data management solution for your community. Federate with a growing list of Dataverse repositories worldwide for increased discoverability of your community's data. Participate in the drive to set norms for sharing, preserving, citing, exploring, and analyzing research data. *Want to install a Dataverse repository?*



Developers

Participate in a vibrant and growing community that is helping to drive the norms for sharing, preserving, citing, exploring, and analyzing research data. Contribute code extensions, documentation, testing, and/or standards. *Integrate research analysis, visualization and exploration tools, or other research and data archival systems with the Dataverse Project. Want to contribute?*

Ferramenta de código aberto
Desenvolvido pelo IQSS
(Institute for Quantitative
Social Science) com Harvard
University Info Techno,
Harvard Library

Comunidade ativa

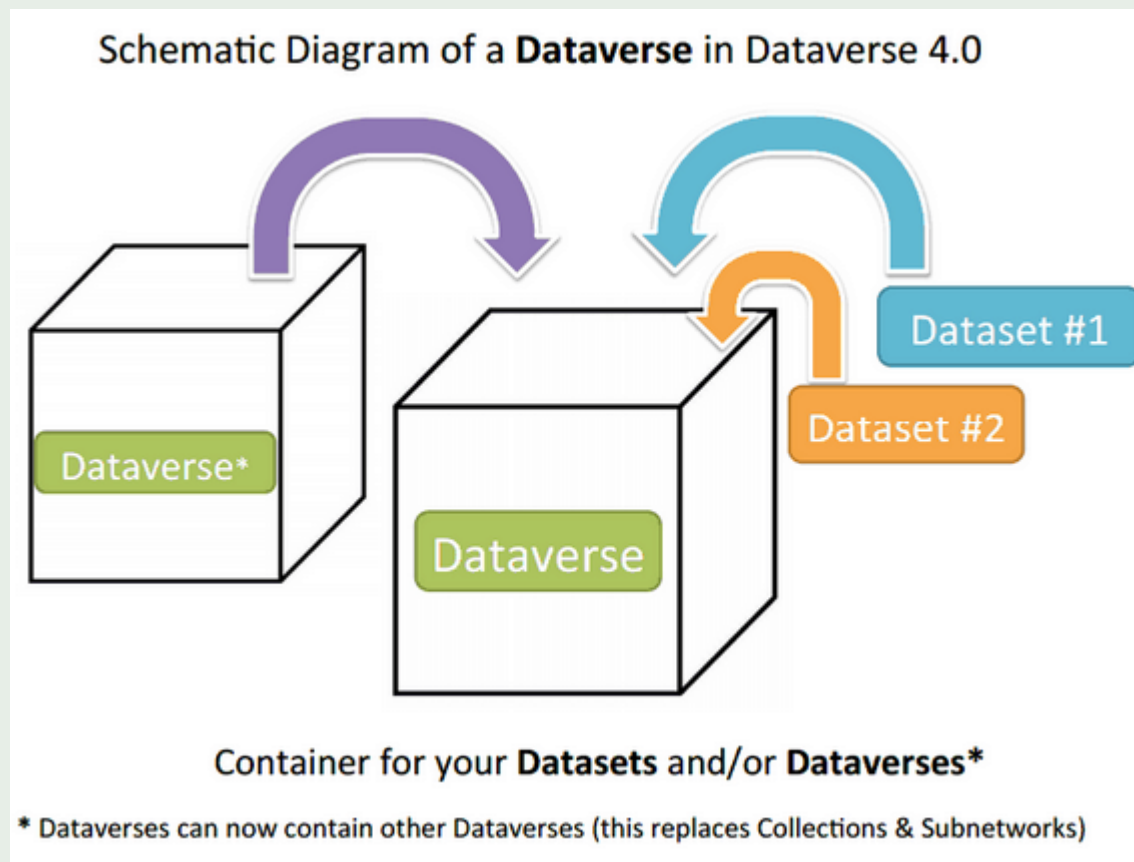
Global Dataverse Community
Consortium
<http://dataversecommunity.global/>

65 Installations



Dataverse ferramenta / Dataverse recipiente

Um dataverse é também um recipiente para conjuntos de dados (dados de pesquisa, código, documentação e metadados) e outros dataverse, que podem ser configurados para pesquisadores individuais, departamentos, periódicos e organizações.



- O IRD desenvolve atividades de pesquisa, formação e inovação em parceria com instituições brasileiras há mais de 60 anos.
- As pesquisas estão focadas nas relações entre o homem e o seu ambiente, na perspectiva de um desenvolvimento sustentável e dos **Objetivos de Desenvolvimento Sustentável**.



IRD Nacional 5 departamentos científicos

- DYNAMIQUES INTERNES ET DE SURFACE DES CONTINENTS (DISCO)
- ECOLOGIE, BIODIVERSITÉ ET FONCTIONNEMENT DES ECOSYSTÈMES CONTINENTaux (ECOBIO)
- OCÉANS, CLIMAT ET RESSOURCES (OCEANS)
- SANTÉ ET SOCIÉTÉS (SAS)
- SOCIÉTÉS ET MONDIALISATION (SOC)

2050

Pessoas
metropolitana

852

Resquisadores

30%

Pessoas fora da França



Dados de pesquisa no IRD

2015 - Observatórios de Dados Institucionais WG / Metadados / DOI

2016 - Treinamento de Conscientização sobre os Dados e PGD

avec



2018 - Missão Infraestrutura e dados digitais (**MIDN**) + serviço da informação científica (**IS**) + serviço da informática : Início do projeto Dataverse (Datasuds)

2018-2019 – Datathon 5 workshops : os pesquisadores vêm com os dados deles

Sept 2019 - Repositório DataSuds e site data.ird.fr + recrutamento de um administrador (cargo interno)

MIDN
+ IS

- Seminário de lançamento do DataSuds (+ 120 participantes)

- **Desde outubro 2019** – Treinamentos personalizados em cada laboratório

2019 - Criação de uma rede de pessoas referentes em cada laboratório

2020 – Lançamento do grupo PGD (Plano de Gestão de Dados)

- Treinamento certificado pelo Serviço de Treinamento de Recursos Humanos. Projeto de **e-learning**

- La MIDN tornou-se **Missão para a Ciência Aberta**

DataSuds (<https://dataverse.ird.fr/>)

Des données ouvertes pour une science durable au Sud

DataSuds (IRD) En savoir plus...

Statistiques 1 091 Téléchargements Contact Partager

L'entrepôt de données DataSuds propose aux scientifiques de l'IRD et à leurs partenaires un service pour diffuser, préserver et valoriser leurs données de recherche en facilitant leur identification et leur citation. Il constitue l'un des éléments du dispositif de la science ouverte pour le Sud que promeut l'IRD.

Comment déposer des données ?
→ service d'assistance (...demander la création d'un espace de dépôt dédié à un projet ou à une équipe de recherche) → mode d'emploi

Partnerships UMR iEES-Paris UMR NUTRIPASS UMR MIO

Chercher dans ce dataverse... Trouver Recherche avancée

1 à 10 de 143 Résultats

Informal waste collectors and aggregators space and activities in Ho Chi Minh City, Vietnam, COMPOSE project (2019-2021)
3 déc. 2020 - Informal waste collection system
Nguyen, Thai Huyen; Le, Thi Thao Trang; Nguyen, Thai Hoa; Nguyen, Thi Hai Yen; Dinh, Thien An; Dinh, Ngoc Anh; Le, Viet Anh; Hoang, Van Chien; Pham, Ngoc Tan; Le, Tuan Minh; Phan, Cong Hung; Nguyen, Tien Quoc Anh; Do, Khac Hoai Son; Nguyen, Duc Loc; Nguyen, Tien Tam, 2020, "Informal waste collectors and aggregators space and activities in Ho Chi Minh City, Vietnam, COMPOSE project (2019-2021)", <https://doi.org/10.23708/YRAASG>, DataSuds, V1, UNF:6:ftkKWINEti0QwvEIrN5BuA== [fileUNF]

- Dados em perigo
- Dados de cauda longa ([lien](#))
- Dados do IRD (UMR) + parceiros Sul
- A escolha do repositório não é imposta.
- Função de diretório planejada: coleta de metadados de outros repositórios (sem cópia de dados)

Princípio do depósito único

Dataverses (60)

Datasets (83)

Files (3,823)

Dataverse Category

Laboratory (30)

Research Project (19)

Organization or Institution (6)

Research Group (4)

Department (1)

Publication Year

2020 (101)

2019 (42)

Subject

Earth and Environmental Sciences (78)

Agricultural Sciences (40)

Medicine, Health and Life Sciences (25)

Social Sciences (10)

Computer and Information Science (9)

[More...](#)

Scientific Theme

Environmental sciences (46)

Soil science (26)

Health in sociocultural, economic and political aspects (9)

Plant science (9)

Urbanization and urban societies (7)

[More...](#)

Keyword Term

nitrogen (N) (14)

dataset (11)

Ghana (8)

microplastic (8)

1 to 10 of 143 Results

Sort ▾

Informal waste collectors and aggregators space and activities in Ho Chi Minh City, Vietnam, COMPOSE project (2019-2021)



Dec 3, 2020 - Informal waste collection system

Nguyen, Thai Huyen; Le, Thi Thao Trang; Nguyen, Thai Hoa; Nguyen, Thi Hai Yen; Dinh, Thien An; Dinh, Ngoc Anh; Le, Viet Anh; Hoang, Van Chien; Pham, Ngoc Tan; Le, Tuan Minh; Phan, Cong Hung; Nguyen, Tien Quoc Anh; Do, Khac Hoai Son; Nguyen, Duc Loc; Nguyen, Tien Tam, 2020, "Informal waste collectors and aggregators space and activities in Ho Chi Minh City, Vietnam, COMPOSE project (2019-2021)", <https://doi.org/10.23708/YRAASG>, DataSuds, V1, UNF:6:ftkWINeti0QwvEIrN5BuA== [fileUNF]

The system of informal waste collectors and aggregators in Ho Chi Minh city (Vietnam) was surveyed in the framework of the COMPOSE project (2019-2021). The social survey based on questionnaire was carried out in the entire area of Ho Chi Minh city to identify the location of the...

Informal waste collectors and aggregators space and activities in Da Nang city, Vietnam, COMPOSE project (2019-2021)



Nov 23, 2020 - Informal waste collection system

Nguyen, Thai Huyen; Nguyen, Thi Hai Yen; Le, Thi Thao Trang; Nguyen, Thai Hoa; Kieu, Thi Kinh; Nguyen, Minh Hien; Tran, Thanh Dang; Hoang, Minh Tieng; Le, Thi Huyen; Le, Viet Anh; Dinh, Ngoc Anh; Nguyen, Tien Tam, 2020, "Informal waste collectors and aggregators space and activities in Da Nang city, Vietnam, COMPOSE project (2019-2021)", <https://doi.org/10.23708/CGLFIL>, DataSuds, V1, UNF:6:ISB5tEJiQB+4e3tdWii/RQ== [fileUNF]

The system of informal waste collectors and aggregators in Da Nang city (Vietnam) was surveyed in the framework of the COMPOSE project (2019-2021). The social survey based on questionnaire was carried out in the entire area of Da Nang city to identify the location of the aggregat...

Equipe DIAL (Université Paris Dauphine, IRD, CNRS)



Nov 17, 2020 UMR LEDa

Développement, Institutions et Mondialisation DIAL est une équipe de l'UMR 260 LEDa qui réunit des enseignants de l'université Paris-Dauphine et des chercheurs de l'IRD et du CNRS. Cette équipe compte 19 enseignants chercheurs de Paris-Dauphine, 12 chercheurs de l'IRD et 1 cherch...

ADHI: African Database of Hydrometric Indices



Nov 17, 2020 - UMR HydroSciences Montpellier

Tramblay, Yves; Rouché, Nathalie, 2020, "ADHI: African Database of Hydrometric Indices", <https://doi.org/10.23708/LXGXQ9>, DataSuds, V2

This database contains different hydrometric indices, describing runoff characteristics, seasonality, floods and low flows, for about 1500 rivers covering all regions of Africa between 1950 and 2018.

UMR LEDa (Université Paris Dauphine, IRD, CNRS)



Nov 16, 2020



Contact Share

ADHI: African Database of Hydrometric Indices

Version 2.0

Tramblay, Yves; Rouché, Nathalie, 2020, "ADHI: African Database of Hydrometric Indices", <https://doi.org/10.23708/LXGXQ9>, DataSuds, V2

Cite Dataset

Learn about Data Citation Standards.

Dataset Metrics

148 Downloads

Description

This database contains different hydrometric indices, describing runoff characteristics, seasonality, floods and low flows, for about 1500 rivers covering all regions of Africa between 1950 and 2018. (2020-08-10)

Subject

Computer and Information Science; Earth and Environmental Sciences

Keyword

Africa, River, Discharge, Runoff, Flows, Afrique, Rivière, Débit, Ecoulement

Related Publication

Tramblay Y., Rouché N., Paturel J.-E., Mahé G., Boyer J.-F., Amoussou E., Bodian A., Dacosta H., Dakhlaoui H., Dezetter A., Hughes D., Hanich L., Peugeot C., Tshimanga R., Lachassagne P.: ADHI: The African Database of Hydrometric Indices (1950-2018). Earth System Science Data

Notes

For more details, please read the file "ADHI_README.txt"

Files Metadata Terms Versions

Change View Table Tree

Search this dataset... Find

Filter by File Type: All Access: All File Tag: All

Sort

1 to 8 of 8 Files

Download



ADHI_Annual7DayMin.zip

ADHI/
ZIP Archive - 474.3 KB - Sep 28, 2020 - 17 Downloads
MD5: ca6c934c736003ebf5587bc2a1dac72d
Time series for annual minimum of 7-day discharge, one file per station.

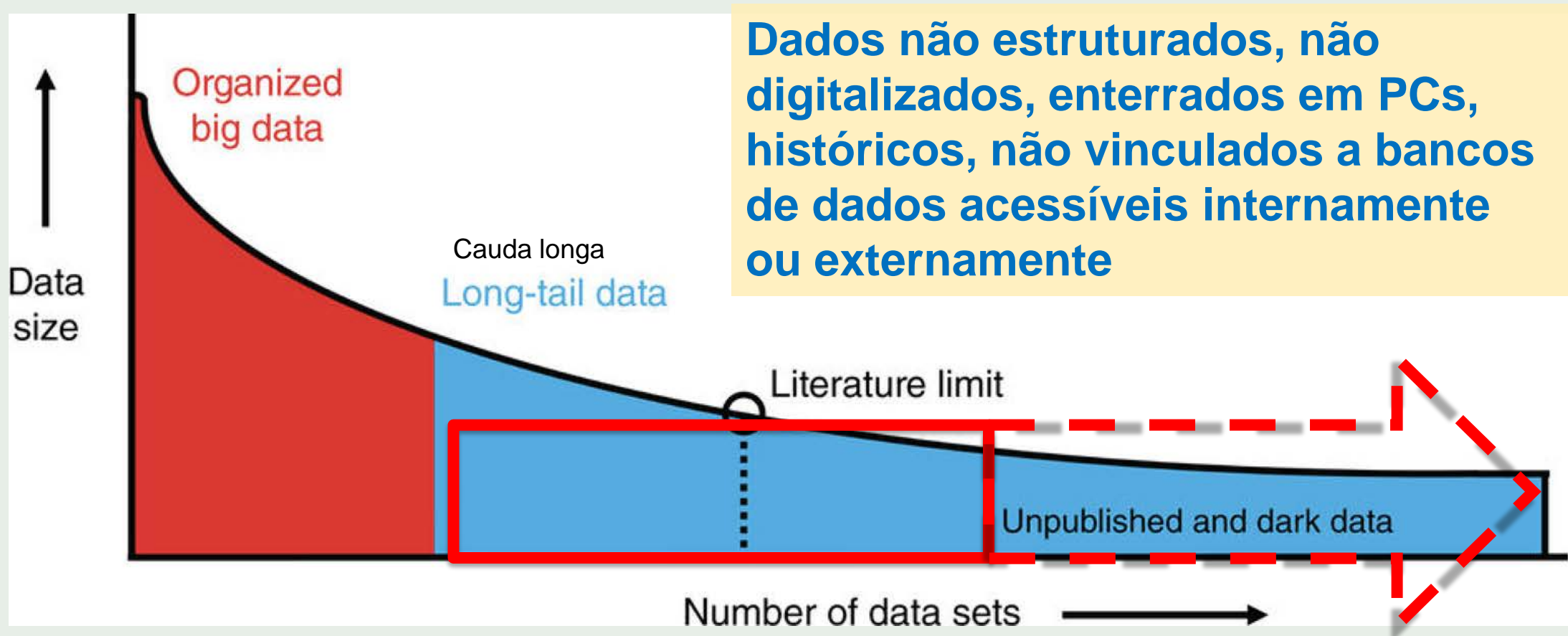
Data

Download

Conjunto de dados

Posicionamento do repositório DataSuds

Observatórios (infraestrutura), centros de dados,
bases de dados acessíveis via web



Source : (Ferguson et al., 2014)

Treinamentos e informação



Bienvenue sur le site support IRD Data,
pour des données ouvertes et une science durable au Sud

Vous trouverez sur ce site un ensemble d'informations et de recommandations, ainsi que des ressources et services sélectionnés pour la gestion des données de la recherche à l'IRD et, plus généralement, la promotion de la Science Ouverte pour le développement.

Parmi les outils recommandés, vous pouvez désormais utiliser l'entrepôt de données DataSuds, qui a été lancé par l'IRD le 6/9/2019.
Pour en savoir plus, vous pouvez parcourir notre menu déroulant « L'Entrepôt DataSuds » et visionner une courte vidéo :



- Para informações sobre boas práticas
 - Com base no existente (DORANUM , FOSTER, INRAE, CIRAD...)
 - Workshop
- constatações
- pouco conhecimento por parte dos pesquisadores
 - Falta de tempo para a formatação dos dados
 - Confusão entre os dados durante o projeto e o arquivamento em Repositório
 - Precisam de ajuda para Escolher um Repositório

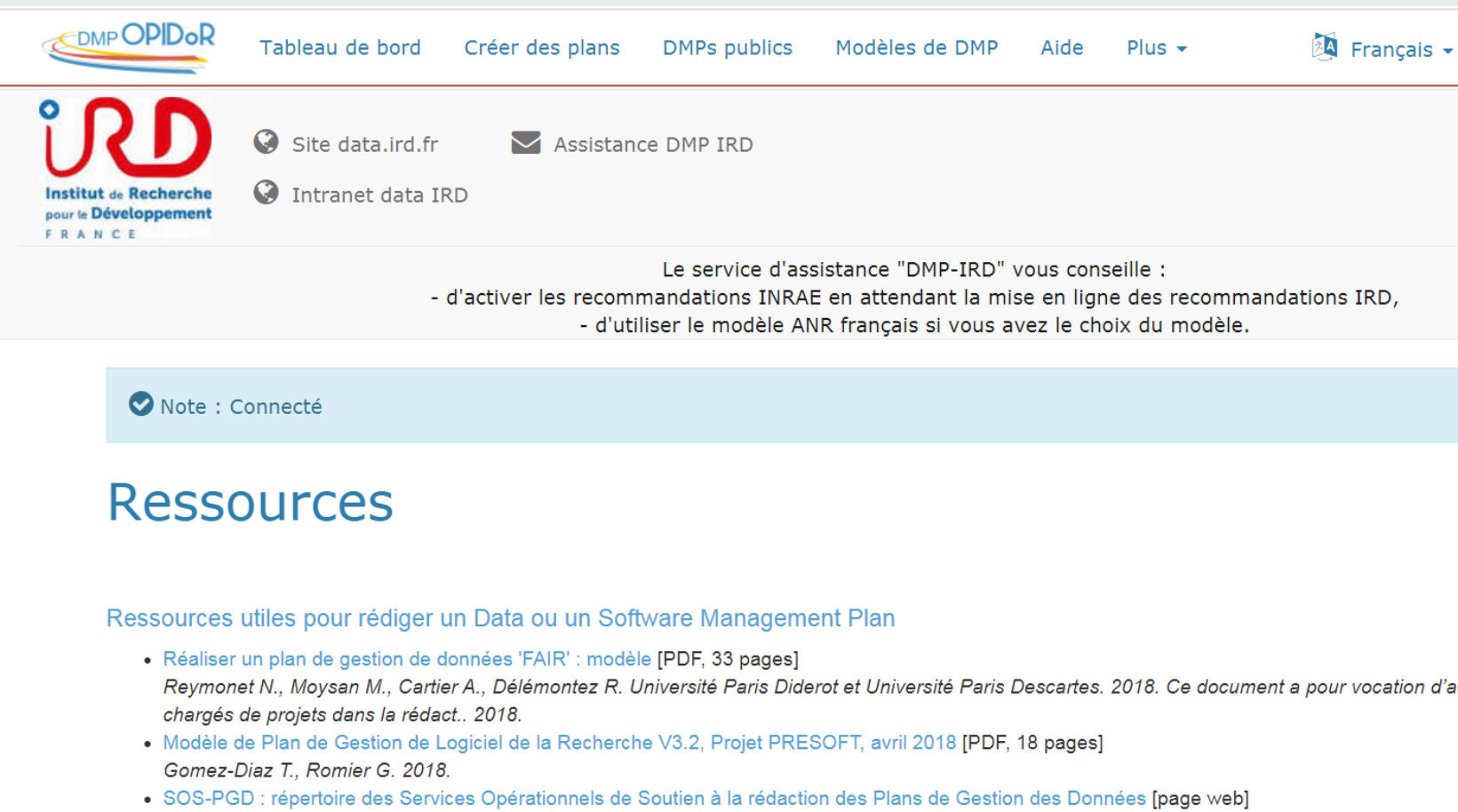
Plano de Gestão de Dados (PGD)

Documento que descreve que dados serão coletados ou gerados; quais as metodologias e padrões que serão utilizados nesses processos; se, como e sob que condições esses dados serão compartilhados e/ou tornados abertos para a comunidade de pesquisa; e como eles serão curados e preservados.

Tem Modelos Padrão e ferramentas (com exemplo) <https://dmptool.org/> e <https://dmponline.dcc.ac.uk/>

Exigindo para agência de fomentos - exemplo FAPESP

Plano de Gestão de Dados



The screenshot shows the DMP OPIDoR website interface. At the top, there is a navigation bar with links: "Tableau de bord", "Créer des plans", "DMPs publics", "Modèles de DMP", "Aide", and "Plus". A language selector is set to "Français". Below the navigation bar is the IRD logo and the text "Institut de Recherche pour le Développement FRANCE". There are two links: "Site data.ird.fr" and "Assistance DMP IRD". Below this, there is a message from the DMP-IRD service: "Le service d'assistance 'DMP-IRD' vous conseille : - d'activer les recommandations INRAE en attendant la mise en ligne des recommandations IRD, - d'utiliser le modèle ANR français si vous avez le choix du modèle." A blue bar indicates the user is "Connecté". The main section is titled "Ressources" and contains a sub-section "Ressources utiles pour rédiger un Data ou un Software Management Plan" with three bullet points:

- [Réaliser un plan de gestion de données 'FAIR' : modèle](#) [PDF, 33 pages] *Reymonet N., Moysan M., Cartier A., Délémontez R. Université Paris Diderot et Université Paris Descartes. 2018. Ce document a pour vocation d'accueillir les recommandations de projets dans la rédaction. 2018.*
- [Modèle de Plan de Gestion de Logiciel de la Recherche V3.2, Projet PRESOFIT, avril 2018](#) [PDF, 18 pages] *Gomez-Diaz T., Romier G. 2018.*
- [SOS-PGD : répertoire des Services Opérationnels de Soutien à la rédaction des Plans de Gestion des Données](#) [page web]

Ferramenta DMP Opidor (adaptação francesa do DMP Tool)

Obrigatório para cada projeto financiado grupo multidisciplinar

- Serviços Jurídicos
- Avaliação
- Informação científica
- Tecnologia da Informação
- Serviços
- Elaboração de recomendações em andamento
- "revisão" de planos de gestão de dados

Quais competências para a gestão de dados?

4 pessoas " Dados " (2 FTE) no serviço de informação científica do IRD de 13 pessoas :

1 administrador do repositório , 1 gerente de projeto

2 bibliotecários

Dentro de um grupo de trabalho IRD com competências complementares (TI, jurídico, avaliação, P...)

Competências essenciais:

Comunicação científica / ciência Aberta

Gerenciamento de projetos

Metadados e o papel das normas

Competências chave à dominar

Princípios FAIR

Conhecimento jurídico

Conhecimentos técnicos

Plano de gestão de dados

Recursos para ajudar



<https://doranum.fr>



<https://www.ouvrirlascience.fr>

GTSO -Couperin

Abertura em 2020 : Diploma
Universitario Scientific Data
Management - l'Univ. de Montpellier

<https://sdm.edu.umontpellier.fr/>



· Discipline-specific skills needed to practice open science (does not include generic computer skills, wider librarianship skills and personal competencies)

· Mapped to LIBER OS Roadmap 7 focus areas, Digcomp 2.0 framework and FOSTER learning resources

Preservação de Dados de Pesquisa



[Digital Preservation Coalition](#)

[Preservation issues - Digital Preservation Handbook](#) Créateur : www.digitalbevaring.dk

Droits d'auteur : Published under a Creative Commons Attribution 2.5 Denmark

Preservação : Dados de pesquisa em Perigo ?

Pêrdidas de 17 % / año !

Current Biology

Volume 24, Issue 1, 6 January 2014, Pages 94-97

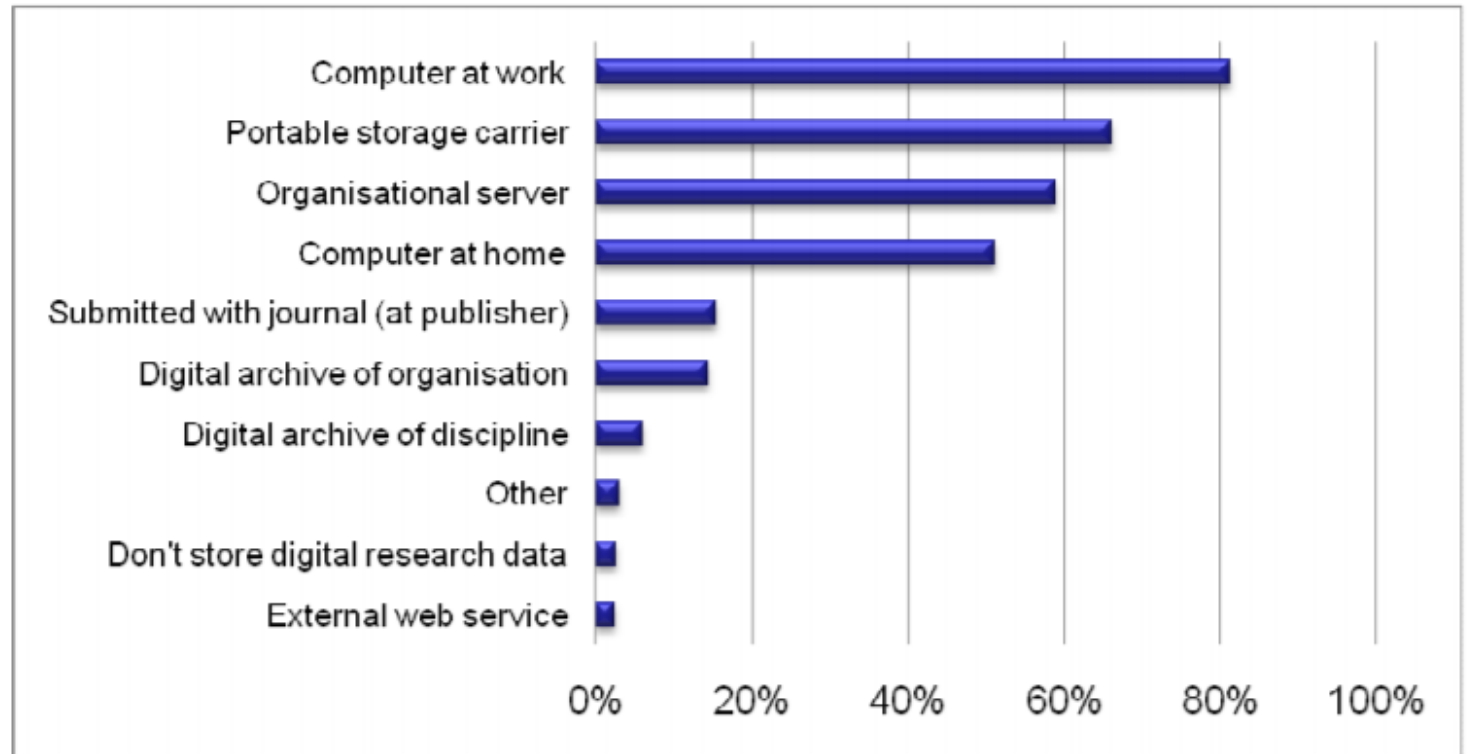


Report

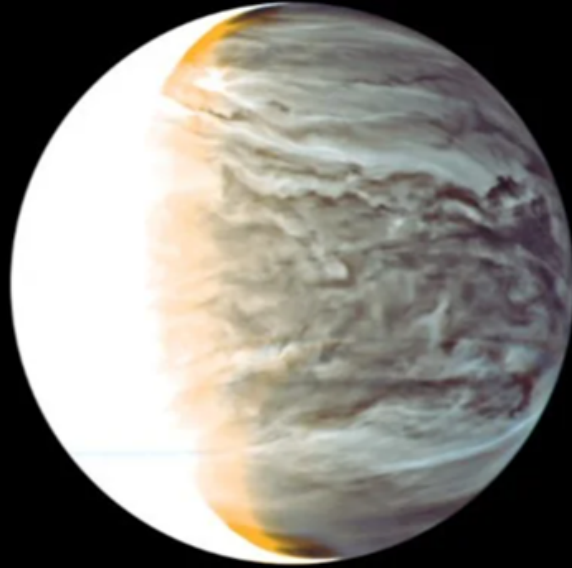
The Availability of Research Data Declines Rapidly with Article Age

Timothy H. Vines^{1,2} ✉, Arianne Y.K. Albert³, Rose L. Andrew¹, Florence Débarre^{1,4}, Dan G. Bock¹, Michelle T. Franklin^{1,5}, Kimberly J. Gilbert¹, Jean-Sébastien Moore^{1,6}, Sébastien Renaut¹, Diana J. Rennison¹

Where do you currently store your research data ? (multiple answers possible)




Graph 2: Source: [PARSE.Insight² survey](#), held among researchers internationally, N = 1202 researchers



False-colour image of the nightside of Venus. (JAXA/PLANET-C Project Team)

SPACE

NASA Detected Phosphine on Venus Decades Ago And Didn't Realise, Scientists Claim

 RAFI LETZTER, LIVE SCIENCE
1 OCTOBER 2020

If life does exist on [Venus](#), NASA may have first detected it back in 1978. But the finding went unnoticed for 42 years.

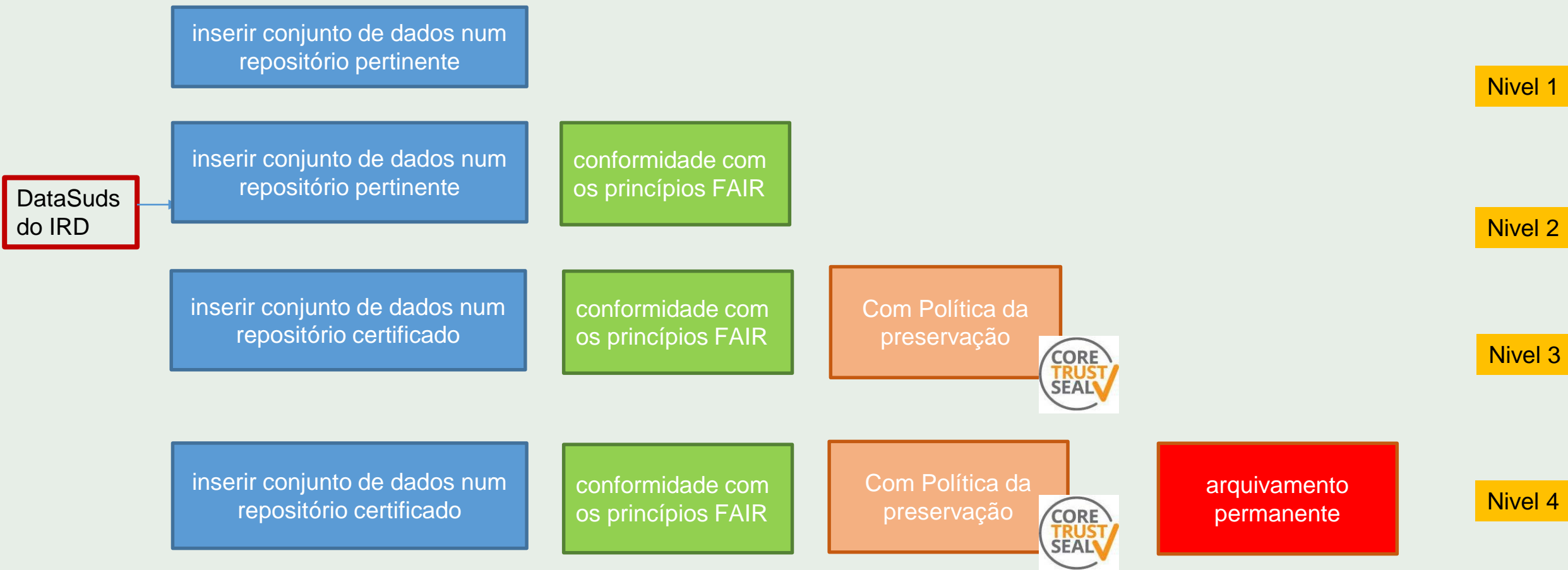
Life on Venus is still a long shot. But there's reason to take the idea seriously. On September 14, a team of scientists made a bombshell announcement in

phosphine, a toxic gas long proposed as a possible sign of alien microbial life

NASA may have first detected it back in 1978. But the finding went unnoticed for 42 years

The 1978 data comes from the Large Probe Neutral Mass Spectrometer (LNMS), one of several instruments that descended into Venus' atmosphere as part of the Pioneer 13 mission.

Preservação para Dados de pesquisa



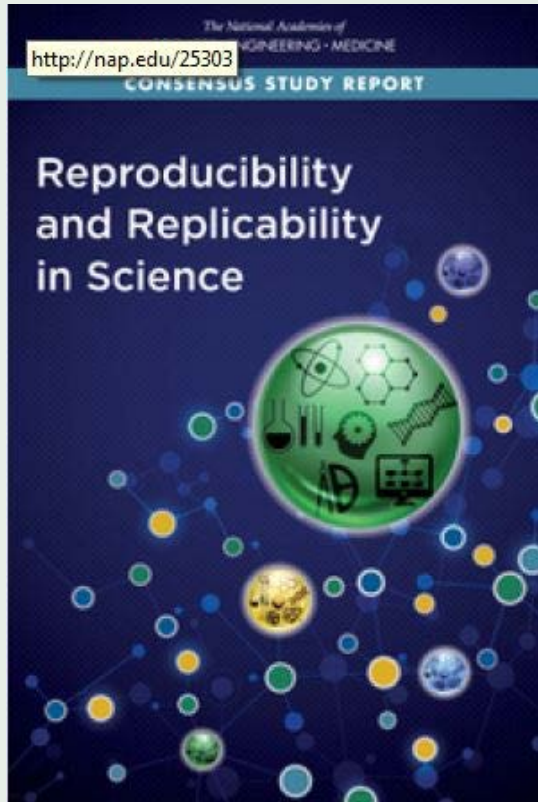
Verificar a Reprodutibilidade

Authors generally do not provide sufficient information to enable others to access their associated research codes (Stodden)

reproducibility to mean computational reproducibility—obtaining consistent computational results using the same input data, computational steps, methods, code, and conditions of analysis (involve data and code)

replicability
obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data.

Replicability involves new data collection and similar methods used by previous studies



National Academies of Sciences, Engineering, and Medicine.
(2019). *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25303>.

Stodden, Victoria, Peixuan Guo, and Zhaokun Ma. 2013. 'Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals'. Edited by Dmitri Zaykin. *PLoS ONE* 8 (6): e67111. <https://doi.org/10.1371/journal.pone.0067111>.

Plano de Gestão de dados

1. Data Collection and Documentation



What kind of data are generated

How will data be generated

What metadata are needed

2. Ethics, legal and security Issues



How will ethical issues be handled

How are the data accessed

Are there copyright issues

Are there sensitive data

What about intellectual property rights

3. Data Storage and Preservation



How are the data stored?

Are there back up systems

How are data safely preserved

4. Data Sharing and reuse



How and where will the data be shared?

How are sensitive data protected

How can data be accessed

Princípios FAIR: Melhorar a qualidade dos dados e metadados

<https://publicient.hypotheses.org/1456>



- Identificadores permanentes (DOI)
- Descrever os dados dentro de um repositório

**Findable
(Achavel)**

- administrar o acesso aos dados a longo prazo
- Obtenção do acordo dos produtores

Acessível

Interoperável

Reutilizável




- Formatos abertos
- Vocabulários compartilhados
- Metadados normalizados

- Licencia de uso apropriado





FAIR questions


FINDABLE

1. Are you aware that a dataset should be assigned a globally unique and persistent identifier when deposited with a data repository?  Yes No
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?  Yes No
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?  Yes No





ACCESSIBLE

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?  Yes No
5. Are you aware that metadata should remain available over time, even if the data is no longer accessible?  Yes No

INTEROPERABLE

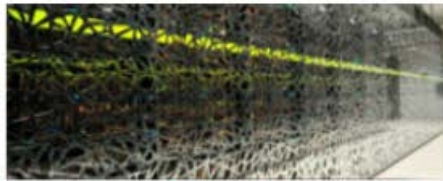
6. Are you aware that the metadata describing your datasets should use controlled vocabularies?  Yes No

REUSABLE

7. Are you aware that provenance information about the collection and/or generation of data should be included in the metadata?  Yes No
8. Are you aware that metadata describing your data should follow the specifications of a community-endorsed standard?  Yes No
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?  Yes No
10. Are you aware that maintaining your dataset FAIR over time requires professional data curation and preservation?  Yes No



CINES



Calcul Intensif



Hébergement

Accueil

Présentation

Calcul

Archivage

Hébergement

Services

Actualités

L'Europe

WEBINAR | 27 June 2020 - 11:00am (CEST)

GENCI, le CEA, la CPU et Atos arment les scientifiques européens pour lutter contre le COVID-19 grâce au
Calcul Haute Performance and processing

Paris, le 27 avril 2020 - Deux des supercalculateurs les plus puissants de France, JoliotCurie, opéré au TGCC (Très Grand Centre de Calcul du CEA) et Occigen au CINES (centre national de calcul de la région Occitanie) ont été utilisés pour la simulation de la propagation de la pandémie de COVID-19.

Le CINES





Le C.I.N.E.S. (Centre Informatique National de l'Enseignement Supérieur) est un Etablissement Public à caractère Administratif national (EPA), basé à Montpellier, et placé sous la tutelle du Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation.

Le CINES & L'Europe



Dans le cadre de ses missions, le CINES, grâce à son expertise, est devenu un acteur reconnu dans la communauté scientifique. Que ce soit pour la partie archivage pérenne ou le calcul intensif. Il participe ...

-  Recrutement et stages
-  Documentation technique
-  Ressources
-  Comment calculer au CINES?
-  Comment archiver au CINES/
-  Assistance en ligne

ARTICLES



Le HPC XXL



Reporté à une date ultérieure : AAF - Groupe PIN : Les données numériques, une espèce en voie d'extinction ?



Nouvelles Formations : DU « Scientific Data Management » Université de Montpellier

CINES = Centre Informatique National de l'Enseignement Supérieur (instituição pública nacional)

Missões estratégicas nacionais

Computação de alta performance

Preservação de Longo prazo

hospedagem de aplicações

CINES

Planejamento de preservação.

<https://www.cines.fr/archivage/nos-solutions-darchivage/pac/strategies-de-perennisation/>

No CINES, ele está centralizado em três ações:

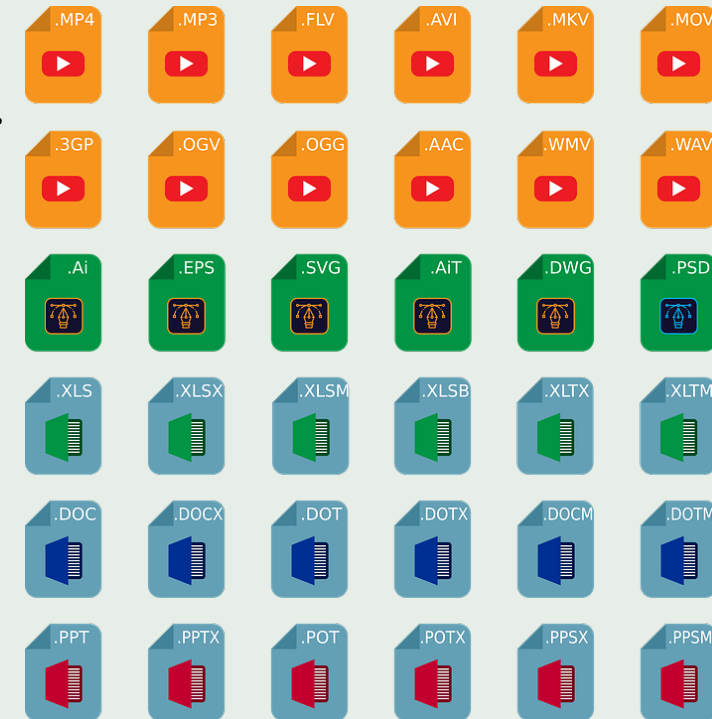
- **O uso de metadados**, para garantir a compreensão dos documentos ao longo do tempo.
- **A preservação das informações de representação**, para informar como os formatos de arquivo são constituídos.
- A seleção de **formatos de arquivo** para arquivamento e a conversão de formatos de arquivo obsoletos, para garantir a leitura do arquivo ao longo do tempo.

Formatos de arquivos

Formato aceitados FACILE (<https://facile.cines.fr/>)

Um formato de arquivo é uma convenção sobre a representação de dados em um meio digital. Pode ser :

- **Especificado:** há uma descrição da convenção utilizada para representar os dados, e é suficientemente descrita para desenvolver uma implementação completa.
- **Aberto:** a convenção é pública (caso contrário o formato é dito fechado). É, portanto, sem restrição de acesso ou de implementação.
- **Normalizado:** a convenção é adotada pelos órgãos de padronização (ISO, W3C). Exemplo: PDF/A.
- **Standardizado:** não existe um padrão para este formato, mas seu uso é tão difundido que é considerado um padrão. Exemplo: PDF. AVISO: em inglês "standard" significa "norma"
- **Proprietário:** se o uso do formato se enquadra no âmbito do direito privado, então depende da existência do proprietário. Pode ser publicado. Exemplo: PD



<https://www.cines.fr/archivage/des-expertises/les-formats-de-fichier/>



FACILE - Service de validation de formats

Vérifier l'éligibilité de vos documents à un archivage sur la plateforme PAC du CINES.

<https://facile.cines.fr/>

Validation

Correction PDF

Tutoriels

Web Service

Archivage de la TEI

+ Choisissez des fichiers - Taille max 2,5 Go

Valider

Annuler

[Cliquez ici pour demander l'aide d'un expert du CINES](#)

Liste des formats validables

⚠ Attention : le validateur de formats permet de valider certains formats qui ne sont pas pris en charge par la plateforme d'archivage du CINES.

Format	Nom	PRONOM PIUD	Type MIME	Commentaire	Archivable dans PAC
AAC AAC	Advanced Audio Codings	[fmt/199]		Format Mpeg-4 contenant uniquement un flux audio au format AAC.	✓
AIFF PCM	Audio Interchange File Format	[fmt/414]	[audio/x-aif, audio/x-aiff]	Format audio contenant uniquement un flux PCM.	✓
APNG	Animated Portable Network Graphics	[fmt/935]	[image/vnd.mozilla.apng, image/apng]	L'APNG est une extension du format PNG permettant de réaliser des animations graphiques.	✗
DAE UTF-8 1.4.1	Collada		[application/xml]	Format permettant de stocker des données géométriques sous forme de scènes (plusieurs objets combinés dans le même référentiel), et d'y ajouter des informations supplémentaires pour décrire la scène et les objets (matériaux, environnement lumineux, animations, ...) ou pour ajouter des notions sémantiques (relations entre les objets, découpage d'un objet en plusieurs éléments fonctionnels, etc...).	✗
FLAC FLAC 1.2.1	Free Lossless Audio Codec	[fmt/279]	[audio/ogg, audio/x-flac]	Format audio compressé sans perte.	✓

Preservação de dados no Cines

- Decidir o que preservar
- Usar as normas e os princípios FAIR para Dados e metadados
- Arquivamento : definir o material e os metadados para archivar
- Transferir o Material na Ferramenta de arquivamento (VITAM) com um conector



<http://www.programmevitam.fr/pages/documentation/vademecums/>

Perfil de arquivamento

Um **perfil de arquivamento** é uma ferramenta de controle para o boletim de transferência, ele não permite alimentar automaticamente o boletim de transferência.

(http://www.programmevitam.fr/ressources/DocCourante/autres/fonctionnel/Vademecum_Profil_archivage.pdf)

Um perfil de arquivamento requer análise prévia e modelagem dos dados a serem arquivados :

- a organização estrutural dos dados a serem arquivados;
os metadados associados a esses dados, permitindo identificá-los, gerenciá-los e recuperá-los após sua transferência para o sistema de arquivamento eletrônico ;
- o conteúdo dos dados, ou seja, a natureza dos dados a serem arquivados (presença ou não de um arquivo associado e tipo de arquivo)

esquema imposto pela SEDA Padrão de intercâmbio de dados para arquivamento

https://francearchives.fr/seda/documentation/SEDA_description_standard_v2_1.pdf

(norme [norme ISO 20614](#) DEPIP » Data exchange protocol for interoperability et preservation nov 2017)

Ferramenta para preservação digital
Open Source

Desenvolvido por três ministérios
(defesa, cultura, pesquisa e educação superior)

Custo 27 TO= 100 00 RS (22 000 euros)



<https://www.coretrustseal.org/why-certification/requirements/>

Organizational Infrastructure

1. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

Response

Guidance:

Repositories take responsibility for stewardship of digital objects, and for ensuring that materials are held in the appropriate environment for appropriate periods of time. Depositors and users must be clear that preservation of and continued access to the data is an explicit role of the repository.

For this Requirement, please describe:

- Your organization's mission in preserving and providing access to data, and include links to explicit statements of this mission.
- The level of approval that the mission has received within the organization.

Evidence for this Requirement could take the form of an approved public mission statement, roles mandated by funders, policy statement signed off by governing board.



<https://www.coretrustseal.org/why-certification/requirements>

3. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

Response

Guidance:

This Requirement covers the governance related to continued operation of the repository over time and during disasters, as well as evidence in relation to succession planning; namely, the measures in place to ensure access to and availability of data holdings, both currently and in the future. Reviewers are seeking evidence that preparations are in place to address the risks inherent in changing circumstances, including in mission and/or scope.

For this Requirement, please describe:

- The level of responsibility undertaken for data holdings, including any guaranteed preservation periods.
- The medium-term (three- to five-year) and long-term (> five years) plans in place to ensure the continued availability and accessibility of the data. In particular, both the response to rapid changes of circumstance and long-term planning should be described, indicating options for relocation or transition of the activity to another body or return of the data holdings to their owners (i.e., data producers). For example, what will happen in the case of cessation of funding, which could be through an unexpected withdrawal of funding, a planned ending of funding for a time-limited project repository, or a shift of host institution interests?

Evidence for this Requirement should relate specifically to governance. The technical aspects of business continuity, and disaster and succession planning should be covered in R15 (Technical infrastructure).

Software Heritage : Preservação do Código

<https://www.softwareheritage.org/>



[Mission](#) ▾ [Archive](#) ▾ [Community](#) ▾ [Support](#) ▾ [About](#) ▾ [Donate](#)  ▾



Software is the key

to access information and
a fundamental part of
human heritage

We are building the universal software archive



Conclusão

Na França , uma solução de preservação centralizada com um operador nacional (CINES),

Estudo de oportunidade para um repositório de dados nacional

Boas práticas em preservação mais também em gestão de dados

- Plano de Gestão de Dados
- Dados FAIR
- Repositório certificado (CoreTrustSeal)
- Necessidade de serviço de assistência de dados
- Necessidade de Infraestructuras (Regional/ Nacional)

Obrigado ! Perguntas ?

Pascal Aventurier

Pascal.aventurier@ird.fr

IRD - Instituto Francês de Pesquisa para o
Desenvolvimento

<https://orcid.org/0000-0003-0211-4549>

Twitter @Pascal_AV

Blog <https://publicient.hypotheses.org/>



Obrigado por Nicolas Cazenave - CINES

[Digital Preservation Coalition](#)

[Preservation issues - Digital Preservation Handbook](#) Créateur : www.digitalbevaring.dk

Droits d'auteur : Published under a Creative Commons Attribution 2.5 Denmark