



La certification des entrepôts de données

Françoise Genova, CDS/Observatoire Astronomique de
Strasbourg, RDA France

Aude Chambodut, EOST, CNRS INSU

Olivier Rouchon, CINES, FAIRsFAIR

La certification des entrepôts de données

- Pourquoi?
 - Les cadres de certification ‘de base’
 - Exemple d’auto-évaluation
 - Les critères du CoreTrustSeal
 - Conclusions
-
- *Tout au long de l’exposé, l’exemple du Centre de Données astronomiques de Strasbourg*

La certification, pourquoi?



Certification

thanks to I. Dillo and H. L'Hours

research data sharing without barriers
rd-alliance.org



Perhaps the biggest challenge in sharing data is trust: how do you create a system robust enough for scientists to trust that, if they share, their data won't be lost, garbled, stolen or misused?

The Data Harvest:

How sharing research data can yield knowledge, jobs and growth

An RDA Europe Report

December 2014

Trust is at the very heart of storing and sharing data

- Users
- Depositors
- Funders

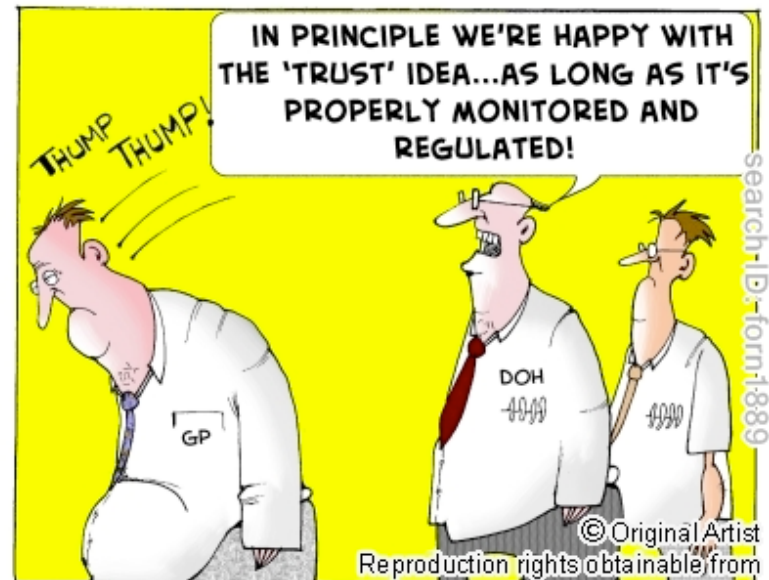


What is trust built on?

- Dedicate yourself (mission statement)
- Do what you promise (stable, sincere and competent reputation)
- Be transparent (peer review, get certified)



"They don't trust each other to share research."



What is a trustworthy repository?

- mission to provide reliable, long-term access to managed digital resources to its designated community, now and into the future
- constant monitoring, planning, and maintenance
- understand threats to and risks within its systems
- **regular cycle of audit and/or certification**



What is a trustworthy repository?

- mission to provide reliable, long-term access to managed digital resources to its designated community, now and into the future
- constant monitoring, planning, and maintenance
- understand threats to and risks within its systems
- **regular cycle of audit and/or certification**



Pourquoi une certification formelle?

- › Assurer que le centre est « de confiance »
- › Mais... il a peut-être déjà la confiance de ses utilisateurs...
- › L'exemple du Centre de Données astronomiques de Strasbourg (CDS)
 - › Créé en 1972
 - › Centre de données de référence pour la communauté astronomique internationale
 - › Infrastructure de Recherche sur la Feuille de Route nationale
 - › 1 000 000+ requêtes/jour sur les services

Oui, pourquoi?

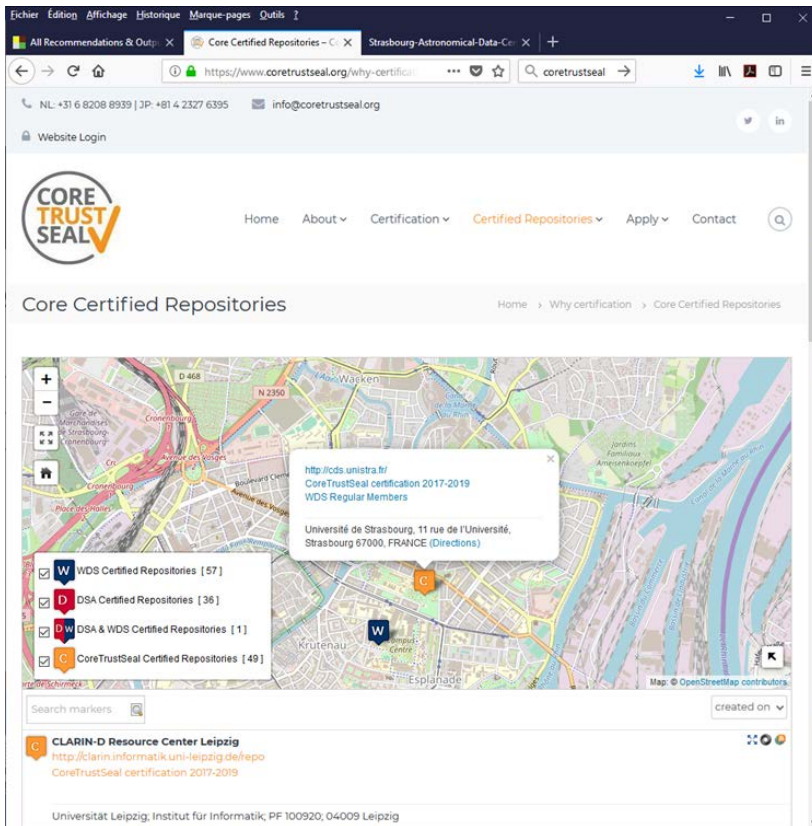
- Critères établis par des personnes compétentes et applicables quel que soit le cadre disciplinaire
- Au préalable, auto-évaluation selon les critères, qui permet de vérifier l'organisation et les process et d'identifier des améliorations possibles
- Evaluation externe par des personnes compétentes
- Le dépôt dans un centre de données certifié est un point important dans les Data Management Plans

Le CDS est certifié...

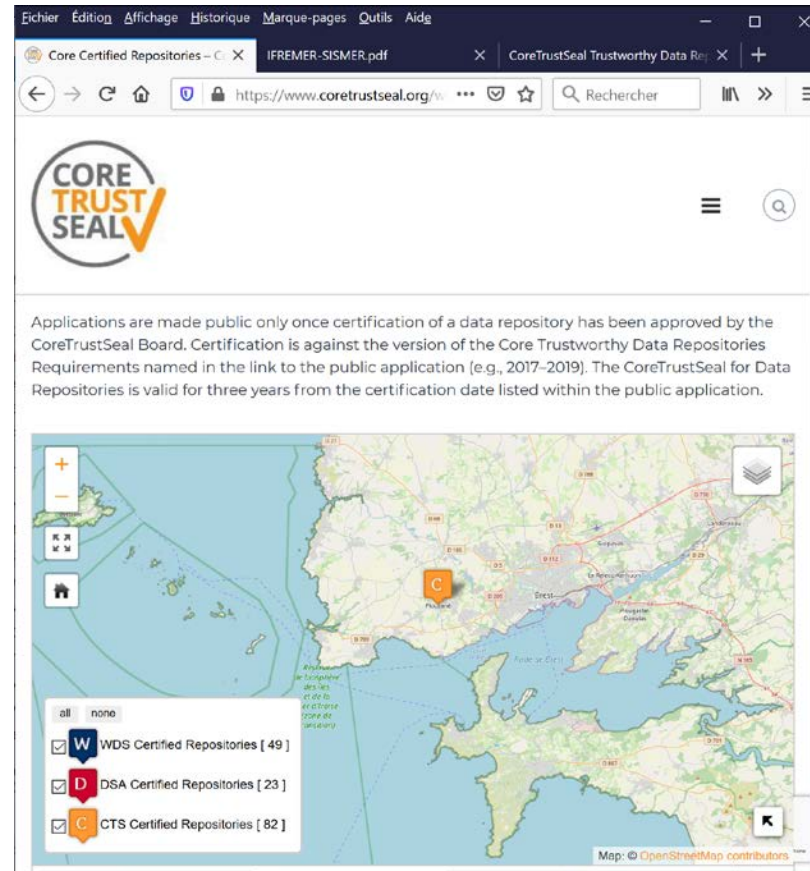
- *World Data System - WDS*
- *Data Seal of Approval - DSA*
- *CoreTrustSeal - CTS*

- Document produit pour la certification CoreTrustSeal:

<https://www.coretrustseal.org/wp-content/uploads/2019/02/Strasbourg-Astronomical-Data-Centre.pdf>



... ainsi que IFREMER-SISMER



Les cadres de certification 'de base'

Un peu d'histoire...

- 4 certification standards available



DIN 31644



ICSU
WORLD DATA SYSTEM



ISO 16363

Le World Data System (WDS)



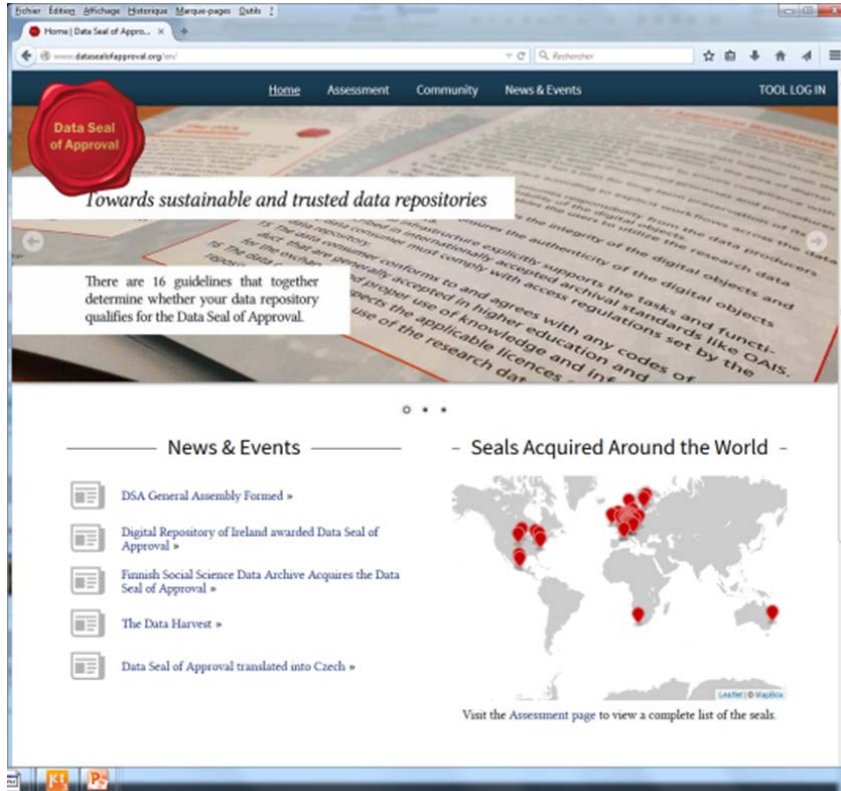
- Créé en 2008 par l'ICSU (=ISC)
- Essentiellement au départ données sur la planète (et astronomie) mais ouvert à tous
- *Promoting universal and equitable access to, and long-term stewardship of, quality-assured scientific data and data services, products, and information covering a broad range of disciplines from the natural and social sciences, and humanities.*
- *Coordinates trusted scientific data services for the provision, use, and preservation of relevant datasets*
- CDS membre du WDS en 2012

European certification framework (pre-RDA)

- **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 DIN 31644/nestorSeal
- **Formal Certification** is granted to repositories which *in addition to* Basic Certification obtain full external audit and certification based on ISO 16363

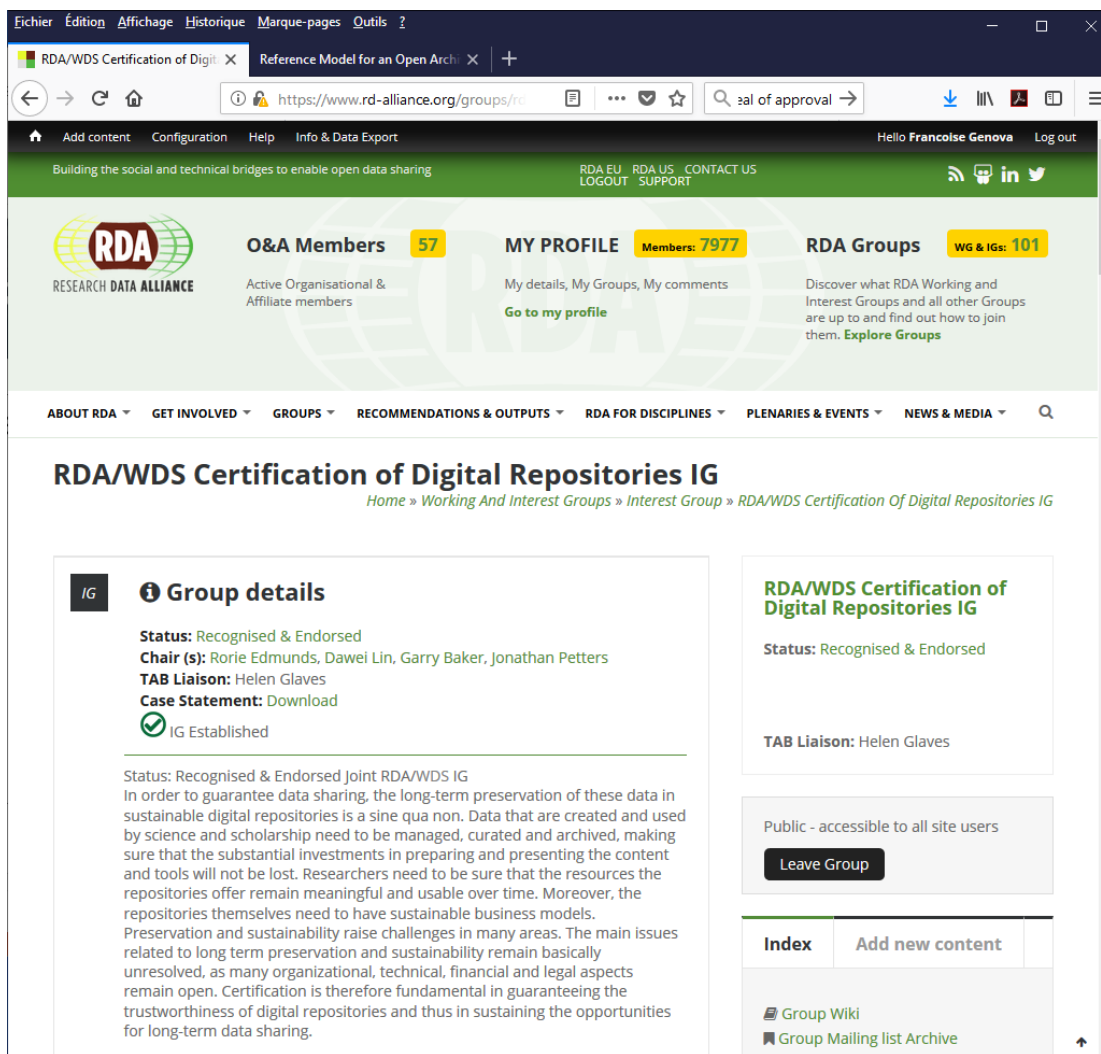


Le Data Seal of Approval



- Plutôt Humanités
- Plus dépôts de données que services
- Le CINES a été le premier centre français certifié DSA, le CDS a été le second en France et le premier centre certifié du domaine des sciences physiques (en 2014)

Dans la RDA, dès 2013



The screenshot shows a web browser window displaying the RDA website. The page title is "RDA/WDS Certification of Digital Repositories IG". The main content area includes a "Group details" section with the following information:

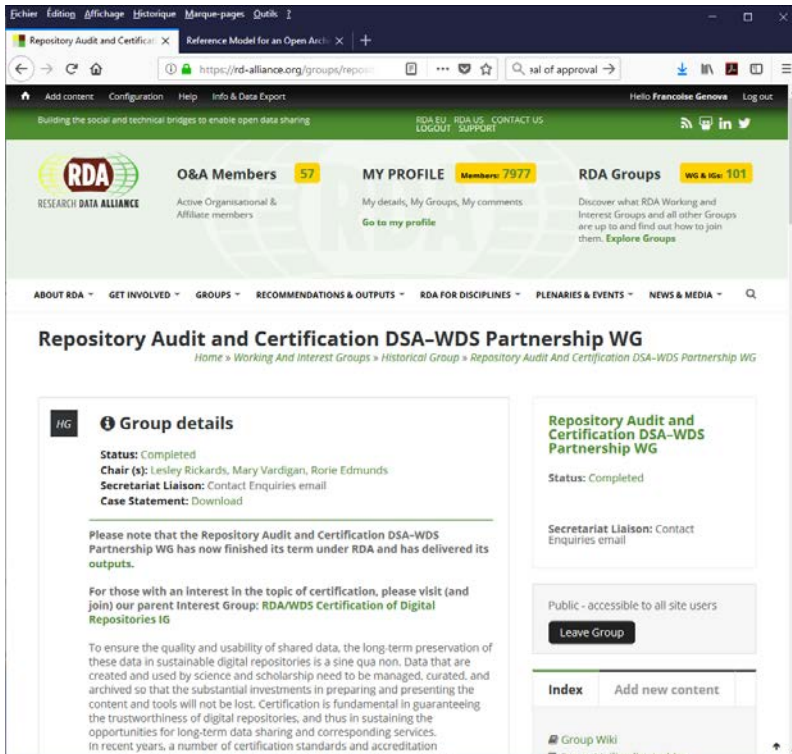
- Status:** Recognised & Endorsed
- Chair (s):** Rorie Edmunds, Dawel Lin, Garry Baker, Jonathan Petters
- TAB Liaison:** Helen Glaves
- Case Statement:** Download
- IG Established:** (indicated by a green checkmark icon)

The status description reads: "Status: Recognised & Endorsed Joint RDA/WDS IG. In order to guarantee data sharing, the long-term preservation of these data in sustainable digital repositories is a sine qua non. Data that are created and used by science and scholarship need to be managed, curated and archived, making sure that the substantial investments in preparing and presenting the content and tools will not be lost. Researchers need to be sure that the resources the repositories offer remain meaningful and usable over time. Moreover, the repositories themselves need to have sustainable business models. Preservation and sustainability raise challenges in many areas. The main issues related to long term preservation and sustainability remain basically unresolved, as many organizational, technical, financial and legal aspects remain open. Certification is therefore fundamental in guaranteeing the trustworthiness of digital repositories and thus in sustaining the opportunities for long-term data sharing."

On the right side of the page, there is a sidebar with the following elements:

- RDA/WDS Certification of Digital Repositories IG**
- Status:** Recognised & Endorsed
- TAB Liaison:** Helen Glaves
- Public - accessible to all site users**
- Leave Group** (button)
- Index** and **Add new content** (links)
- Group Wiki** and **Group Mailing list Archive** (links)

RDA: DSA + WDS (2016)



Repository Audit and Certification: Reference Model for an Open Arch...

Building the social and technical bridges to enable open data sharing

O&A Members 57 MY PROFILE Members: 7977 RDA Groups 101

Repository Audit and Certification DSA-WDS Partnership WG

Status: Completed

Chair (s): Lesley Rickards, Mary Vardigan, Rorie Edmunds

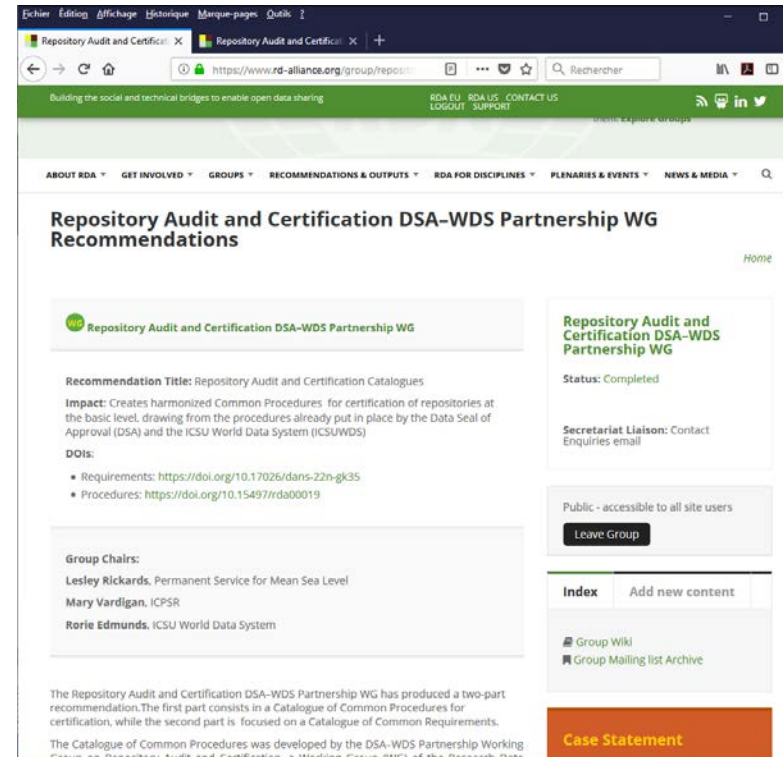
Secretariat Liaison: Contact Enquiries email

Case Statement: Download

Please note that the Repository Audit and Certification DSA-WDS Partnership WG has now finished its term under RDA and has delivered its outputs.

For those with an interest in the topic of certification, please visit (and join) our parent Interest Group: [RDA/WDS Certification of Digital Repositories IG](#)

To ensure the quality and usability of shared data, the long-term preservation of these data in sustainable digital repositories is a sine qua non. Data that are created and used by science and scholarship need to be managed, curated, and archived so that the substantial investments in preparing and presenting the content and tools will not be lost. Certification is fundamental in guaranteeing the trustworthiness of digital repositories, and thus in sustaining the opportunities for long-term data sharing and corresponding services. In recent years, a number of certification standards and accreditation



Repository Audit and Certification: Repository Audit and Certificat...

Building the social and technical bridges to enable open data sharing

Repository Audit and Certification DSA-WDS Partnership WG Recommendations

Recommendation Title: Repository Audit and Certification Catalogues

Impact: Creates harmonized Common Procedures for certification of repositories at the basic level, drawing from the procedures already put in place by the Data Seal of Approval (DSA) and the ICSU World Data System (ICSUWDS)

DOIs:

- Requirements: <https://doi.org/10.17026/dans-22n-gk35>
- Procedures: <https://doi.org/10.15497/rda00019>

Group Chairs:

Lesley Rickards, Permanent Service for Mean Sea Level

Mary Vardigan, ICPSR

Rorie Edmunds, ICSU World Data System

Status: Completed

Secretariat Liaison: Contact Enquiries email

Public - accessible to all site users

Leave Group

Index Add new content

Group Wiki

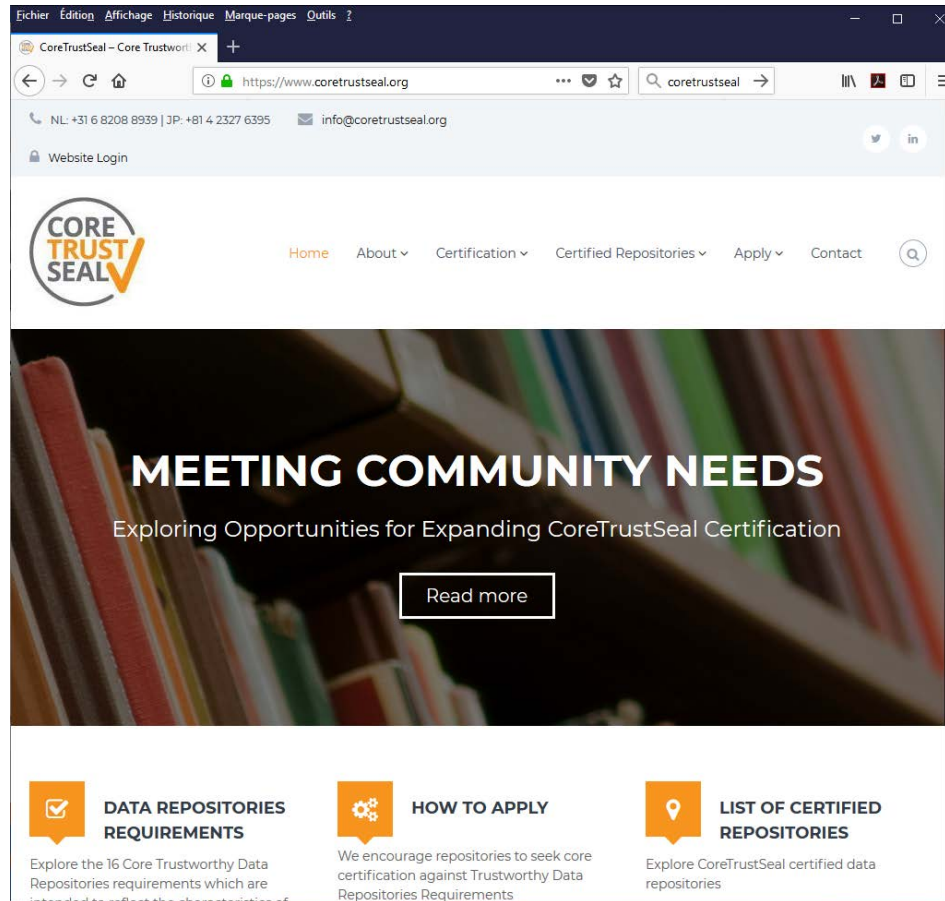
Group Mailing list Archive

The Repository Audit and Certification DSA-WDS Partnership WG has produced a two-part recommendation. The first part consists in a Catalogue of Common Procedures for certification, while the second part is focused on a Catalogue of Common Requirements.

The Catalogue of Common Procedures was developed by the DSA-WDS Partnership Working Group on Repository Audit and Certification, a Working Group (WG) of the Research Data

Case Statement

DSA + WDS = CoreTrustSeal (CTS), 2017



Le modèle de certification européen



Centres de données de confiance - Trustworthy Data Repositories

thanks to Mustapha Mokrane (DANS)- NestorSeal and ISO numbers updated 22 January 2021

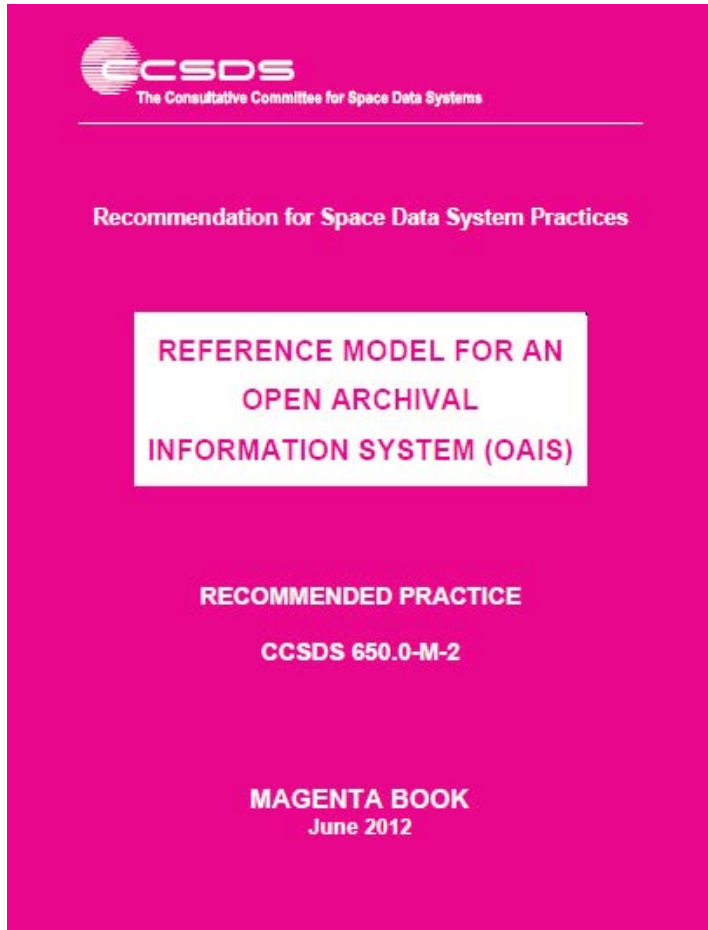
Exemple d'auto- évaluation

Le Centre de Données
astronomiques de Strasbourg (CDS)

Auto-évaluation

- Questionnaire à remplir
- Il faut les compétences
 - De la direction (mission, organisation, ...)
 - Des personnes en charge du contenu
 - Des personnes en charge de l'informatique
- Pour le CDS: un travail d'équipe qui a impliqué la direction, les documentalistes, l'informaticien en charge du service et l'ingénieur système

Description des process du CDS



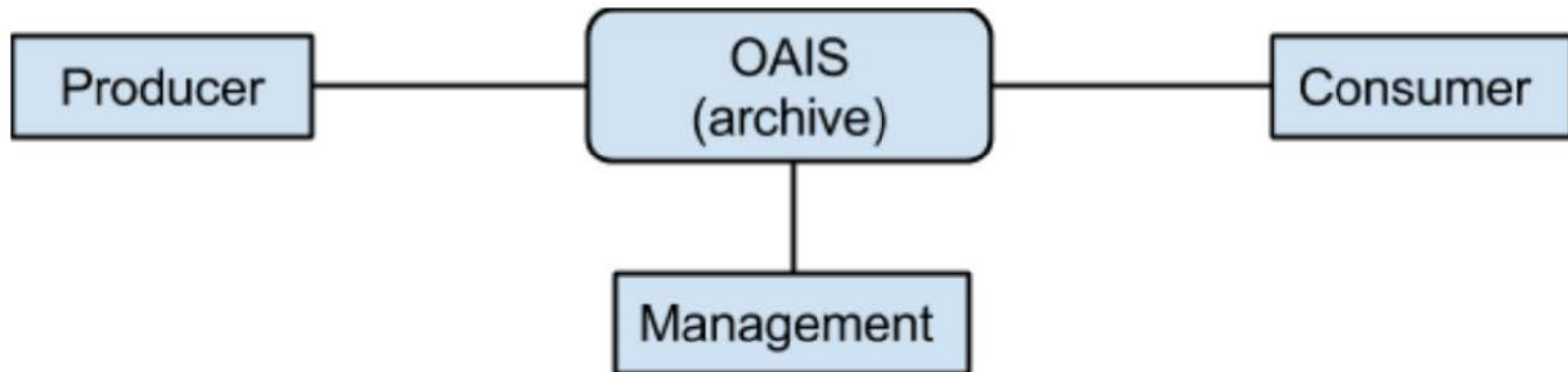
➤ Basé sur le modèle OAIS –
Open Archive Information
System

<https://public.ccsds.org/Pubs/650x0m2.pdf>

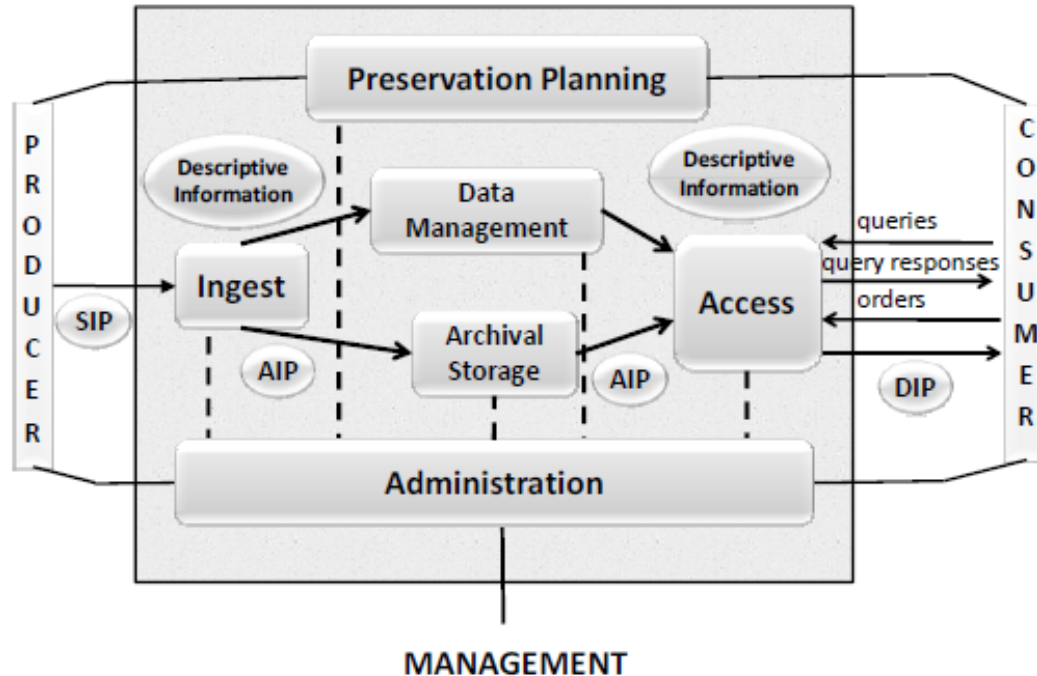
➤ Site en français

<https://www.cines.fr/archivage/un-concept-des-problematiques/le-modele-de-reference-loais/>

L'environnement d'une archive OAIS



Les entités fonctionnelles de l'OAIS

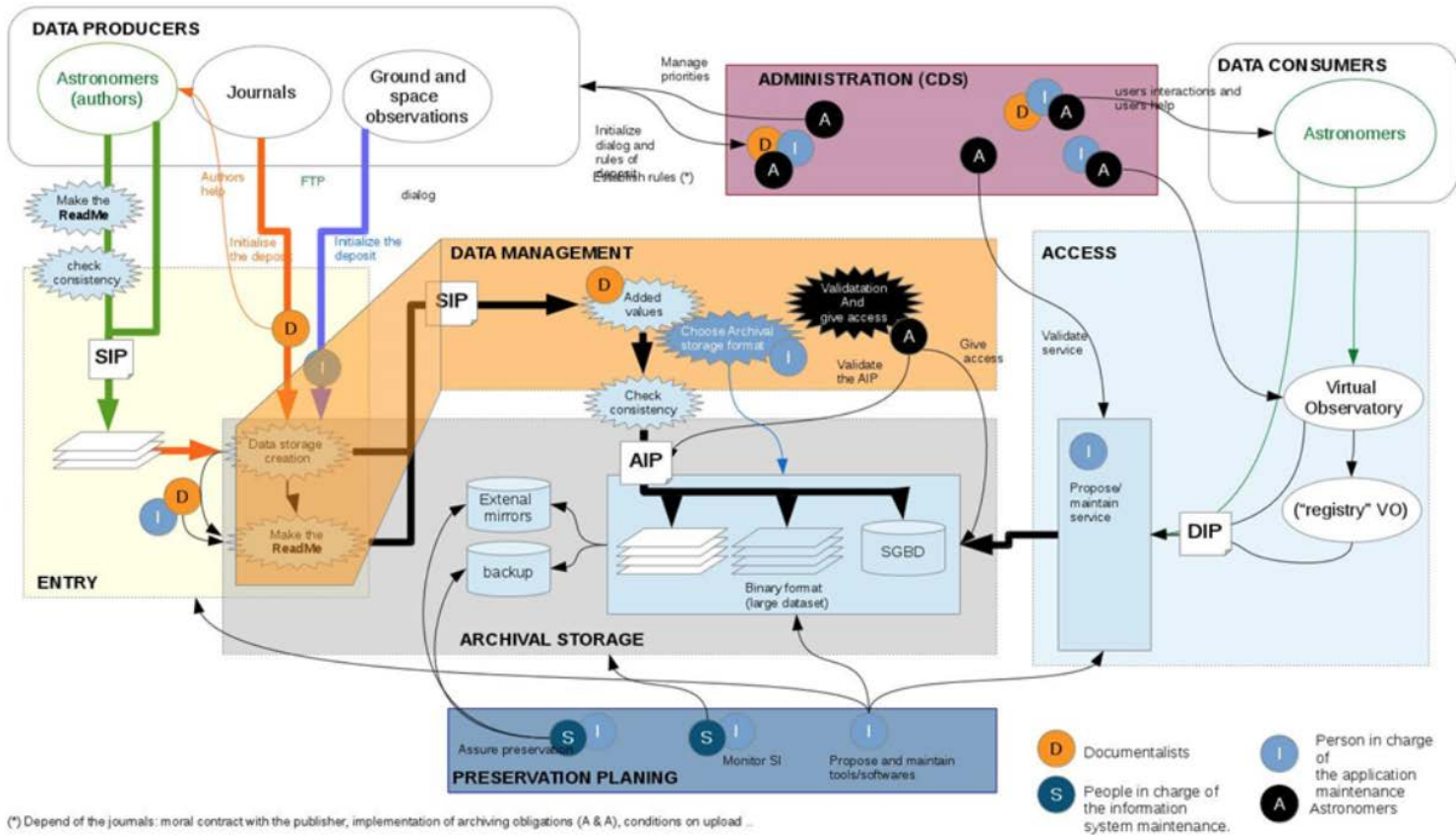


SIP: Submission Information Package

AIP: Archival Information Package

DIP: Dissemination Information Package

Le pipeline de données du CDS dans le modèle OAIS



Les conséquences pour le CDS

- Description de bout en bout des process et des rôles
- Pas de modification majeure
- Des améliorations suite à l'auto-évaluation pour le DSA
 - Clarification des licences
 - Checksums des fichiers
- Le document soumis à CTS en 2018 a été accepté sans modification majeure
- Réaction très positive de nos autorités

La procédure de certification

- › Soumission d'un formulaire en ligne
- › Examen par deux évaluateurs
- › Examen des évaluations par le Board
 - › Renvoi des évaluations des évaluateurs et des commentaires du Board aux candidats
 - › Acceptation du dossier

- › Les éléments du dossier sont publics (sauf exception légitime)
- › Les évaluateurs sont membres de la communauté CoretrustSeal

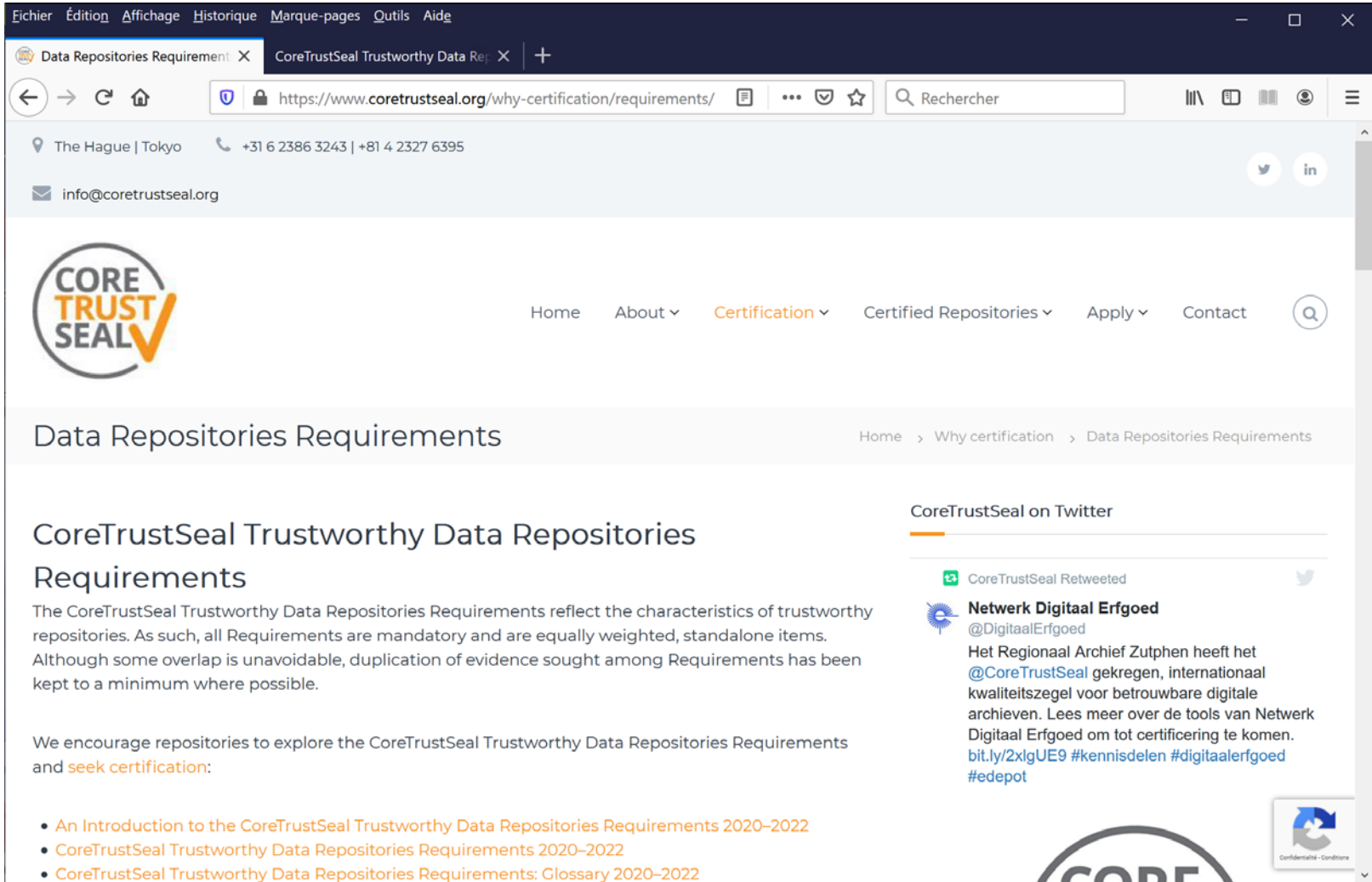
Les critères du CoreTrustSeal

La certification CTS

- Toute l'information est sur le site de CoreTrustSeal
 - <https://www.coretrustseal.org/>
- Contexte + 16 critères
 - <https://www.coretrustseal.org/why-certification/requirements/>
- Document pour guider les évaluateurs et les candidats
 - En cours, Extended Guidance V0.2
 - <https://doi.org/10.5281/zenodo.3632533>
 - Traduction française par RDA France **de la version précédente**
 - <https://www.rd-alliance.org/coretrustseal-criteres-de-conformite>

- « Administrative fee » 1000€

Les critères sur le site CTS



The screenshot shows a web browser window displaying the CoreTrustSeal website. The browser's address bar shows the URL <https://www.coretrustseal.org/why-certification/requirements/>. The website header includes the CoreTrustSeal logo, contact information (The Hague | Tokyo, phone numbers, and email), and a navigation menu with links for Home, About, Certification, Certified Repositories, Apply, and Contact. The main content area is titled 'Data Repositories Requirements' and features the heading 'CoreTrustSeal Trustworthy Data Repositories Requirements'. The text explains that the requirements are mandatory and equally weighted. A sidebar on the right shows a tweet from 'Network Digitaal Erfgoed' mentioning the Regional Archive of Zutphen and the CoreTrustSeal certification. At the bottom, there are three links: 'An Introduction to the CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022', 'CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022', and 'CoreTrustSeal Trustworthy Data Repositories Requirements: Glossary 2020-2022'. A 'Confidential - Conditions' icon is visible in the bottom right corner of the page content.

Fichier Édition Affichage Historique Marque-pages Outils Aide

Data Repositories Requirement X CoreTrustSeal Trustworthy Data Re: X +

https://www.coretrustseal.org/why-certification/requirements/ Rechercher

The Hague | Tokyo +31 6 2386 3243 | +81 4 2327 6395

info@coretrustseal.org

CORE TRUST SEAL

Home About Certification Certified Repositories Apply Contact

Data Repositories Requirements Home > Why certification > Data Repositories Requirements

CoreTrustSeal Trustworthy Data Repositories Requirements

The CoreTrustSeal Trustworthy Data Repositories Requirements reflect the characteristics of trustworthy repositories. As such, all Requirements are mandatory and are equally weighted, standalone items. Although some overlap is unavoidable, duplication of evidence sought among Requirements has been kept to a minimum where possible.

We encourage repositories to explore the CoreTrustSeal Trustworthy Data Repositories Requirements and [seek certification](#):

- [An Introduction to the CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022](#)
- [CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022](#)
- [CoreTrustSeal Trustworthy Data Repositories Requirements: Glossary 2020-2022](#)

CoreTrustSeal on Twitter

CoreTrustSeal Retweeted

Network Digitaal Erfgoed
@DigitaalErfgoed
Het Regionaal Archief Zutphen heeft het @CoreTrustSeal gekregen, internationaal kwaliteitszegel voor betrouwbare digitale archieven. Lees meer over de tools van Network Digitaal Erfgoed om tot certificering te komen. bit.ly/2xlgUE9 #kennisdelen #digitaal erfgoed #edepot

Confidential - Conditions

Les critères de certification CTS



- R0 - Le contexte
- 16 critères, 3 thèmes:
 - Infrastructure organisationnelle
 - Gestion des objets numériques (données et des métadonnées)
 - Technologie

Le contexte

- › Type d'entrepôt
- › Brève description de l'entrepôt
- › Brève description de la communauté concernée
- › Niveau de curation
 - › Contenu en accès tel que déposé
 - › Curation de base (p. ex. vérification rapide, ajout de métadonnées de base ou de documentation)
 - › Curation avancée (p. ex. conversion vers de nouveaux formats, amélioration de la qualité de la documentation)
 - › Curation au niveau des données
- › Partenaires
- › *Résumé des modifications depuis la candidature précédente (s'il y a lieu)*
- › Autres informations pertinentes

Infrastructure organisationnelle

- R1 – Mission/périmètre
- R2 – Licenses
- R3 – Continuité de l'accès
- R4 – Confidentialité/éthique
- R5 – Infrastructure organisationnelle
- R6 – Conseils d'experts

Gestion des objets numériques

- › R7 – Intégrité et authenticité des données
- › R8 – Appréciation et sélection des données
- › R9 – Procédures d'archivage documentées
- › R10 – Plan de préservation
- › R11 – Qualité des données
- › R12 – Processus de traitement (Workflows)
- › R13 – Découverte et identification des données
- › R14 – Réutilisation des données

Technologie

- R15 – Infrastructure technique
- R16 – Sécurité

Conclusions

Pourquoi la certification?

- Quelques semaines de travail d'équipe dans le cas du CDS (tout compris)
- Evaluation – amélioration des process
 - Evaluation interne
 - Evaluation externe
- Importance croissante pour les financeurs des centres de données et des projets (DMP)
- Priorité au niveau politique en France, intérêt des organismes (CNRS, Universités)

Au niveau national: Plan national pour la Science Ouverte

Axe Structurer et Ouvrir les données de la recherche

Les chercheurs seront invités à déposer les données dans des entrepôts de données certifiés, dont la gouvernance et les règles de propriété intellectuelle seront conformes aux bonnes pratiques. À ce titre, les infrastructures nationales et européennes de recherche seront privilégiées, notamment via des centres de données thématiques et disciplinaires.

ANR Flash 2019

Structurer

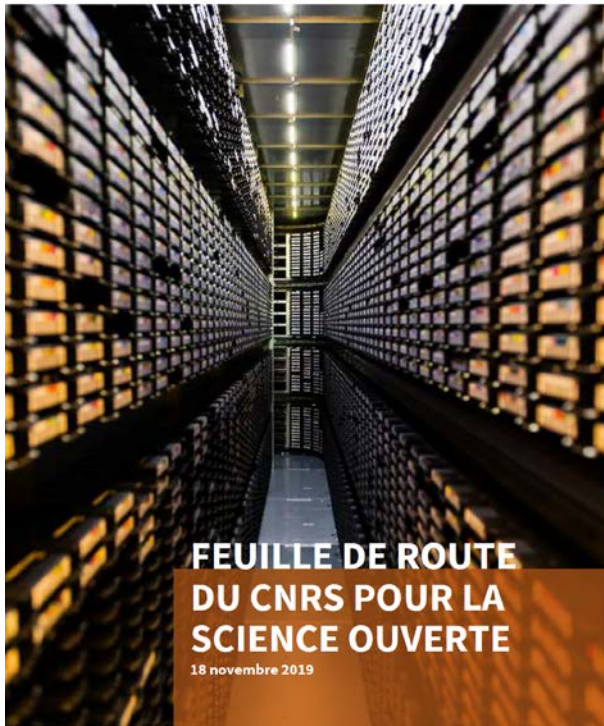
- Généraliser la mise en place de plans de gestion des données dans les appels à projets de recherche
- Développer des centres de données thématiques et disciplinaires.
- Développer un service générique d'accueil et de diffusion des données simples.
- Engager un processus de certification des infrastructures de données.



Organiser

- Soutenir la **Research data alliance** (RDA) et créer le chapitre français de l'alliance (RDA France).
- Soutenir **Software heritage**, la bibliothèque des codes sources

Au niveau des organismes: le CNRS



**FEUILLE DE ROUTE
DU CNRS POUR LA
SCIENCE OUVERTE**

18 novembre 2019

Action 3 : soutenir et accompagner les infrastructures de recherche, productrices de données, dans la définition et la mise en œuvre de politiques de données.

Le CNRS est largement engagé avec ses partenaires dans les Infrastructures de Recherche (IR) nationales et internationales, qui représentent les lieux où se créent et s'analysent les données de la recherche : instruments analytiques, infrastructures de calcul, infrastructures de données, observatoires, etc. Pour généraliser l'application des principes FAIR à toutes les disciplines, le CNRS publiera une charte des infrastructures, engageant celles-ci à respecter les pratiques FAIR et des standards de qualité, en affichant des politiques de données concertées avec les communautés scientifiques utilisatrices des infrastructures concernées. Certaines infrastructures (telles que Progedo et Humanum à l'institut des SHS (INSHS)) sont déjà bien engagées dans ce processus. D'autres sont en cours d'accompagnement telles que les IR de Chimie. Le synchrotron SOLEIL a également mis en route une politique de gestion des données. Les exemples sont multiples et devraient tendre à être généraliser. Ces développements doivent être corrélés avec les certifications (de type CoreTrustSeal) dans le cas où les infrastructures prennent elles-mêmes en charge la distribution de leurs données.

Action 4 : soutenir et accompagner des Infrastructures de données - Mettre en œuvre un service coordonné avec les instituts pour favoriser le dépôt des données pour tous les personnels des unités du CNRS

Les infrastructures de données thématiques jouent un rôle national ou international. Certaines sont inscrites sur la feuille de route nationale des infrastructures de recherche. Cela s'inscrit dans la mesure de structuration du Plan national pour la Science ouverte qui préconise de « développer des centres de données thématiques et usagers ». Le CNRS continuera à soutenir ces infrastructures, et soutiendra le développement de nouveaux réservoirs et services de données thématiques. Ce soutien sera conditionné à une évaluation de leur impact, de leur adéquation aux besoins scientifiques, et de la qualité de leur gestion. Une certification CoreTrustSeal sera recherchée.

L'impact de la RDA

- Fusion des deux cadres de certification 'de base'
- Clarification du paysage pour les centres de données et les agences de financement
- Deux cadres complémentaires au départ: le résultat est meilleur que chacun des originaux!
- Nombreux nouveaux candidats à la certification

Le rôle de RDA France

- La certification est une priorité
- Traduction du Guide, à reprendre pour la version 2020-2022
- Ateliers, présentations dans des réunions à la demande, discussions au téléphone, etc
- Inscrivez-vous sur la liste de diffusion

<https://listes.services.cnrs.fr/wws/subscribe/rda-france>

- Groupe de Travail commun avec le collège Données du Comité pour la Science Ouverte

RDA Global

Email - enquiries@rd-alliance.org

Web - www.rd-alliance.org

Twitter - @resdatall

LinkedIn - www.linkedin.com/in/ResearchDataAlliance

Slideshare - <http://www.slideshare.net/ResearchDataAlliance>

RDA FRANCE

<https://rd-alliance.org/groups/rda-france>

Email - contact-rdafrance@services.cnrs.fr

