



La certification des entrepôts de données

Françoise Genova

CDS/Observatoire Astronomique de Strasbourg

RDA France

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

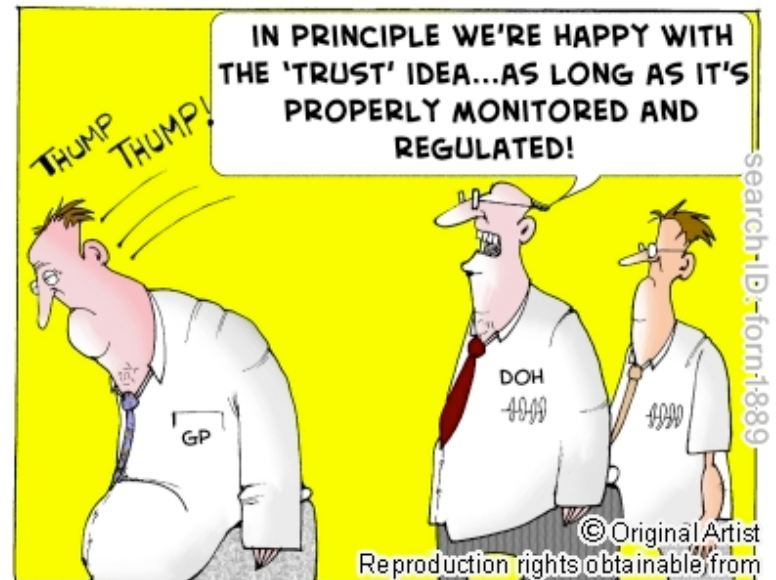


What is trust built on?

- 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
 - › 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?

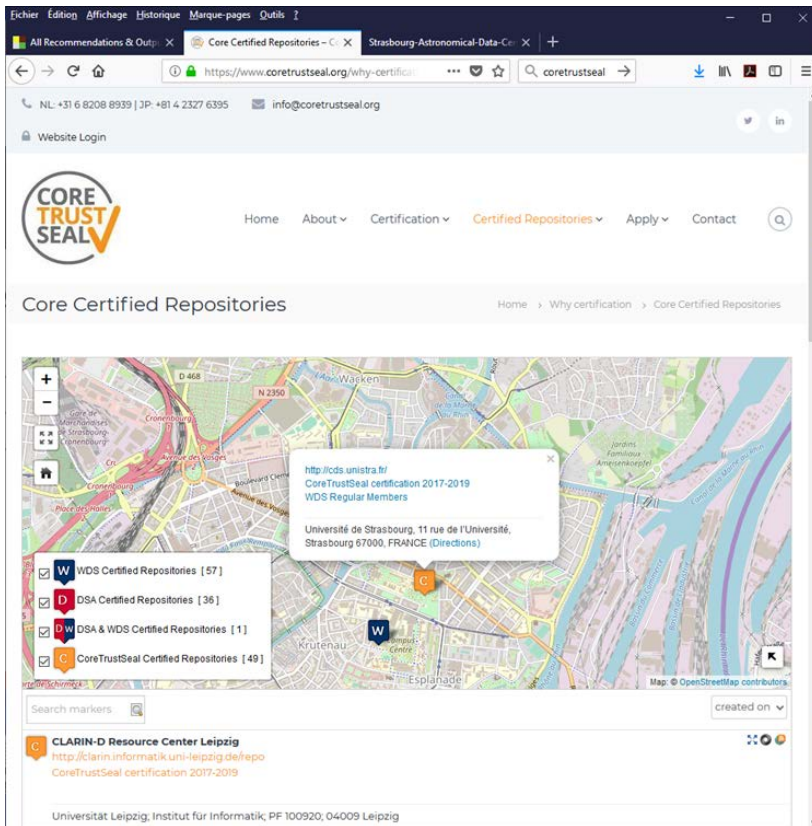
- Critères établis par des personnes compétentes et applicables quel que soit le cadre disciplinaire
- Evaluation externe par des personnes compétentes
- Au préalable, auto-évaluation selon les critères, qui permet de vérifier l'organisation et les process et d'identifier des améliorations possibles
- 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>



Les cadres de certification 'de base'

Un peu d'histoire...

- 4 certification standards available



DIN 31644



ISO 16363

Le World Data System (WDS)



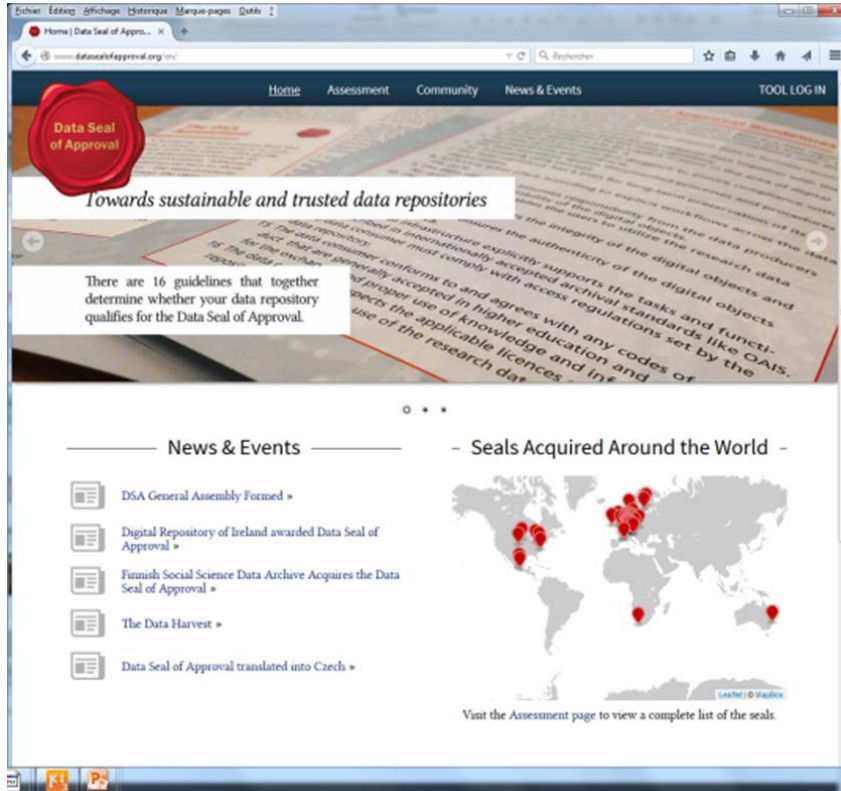
- Créé en 2008 par l'ICSU (=ISC)
- Essentiellement au départ données sur la planète (et astronomie) mais ouvert à tous
- *Promoting universal and equitable access to, and long-term stewardship of, quality-assured scientific data and data services, products, and information covering a broad range of disciplines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

European certification framework (pre-RDA)

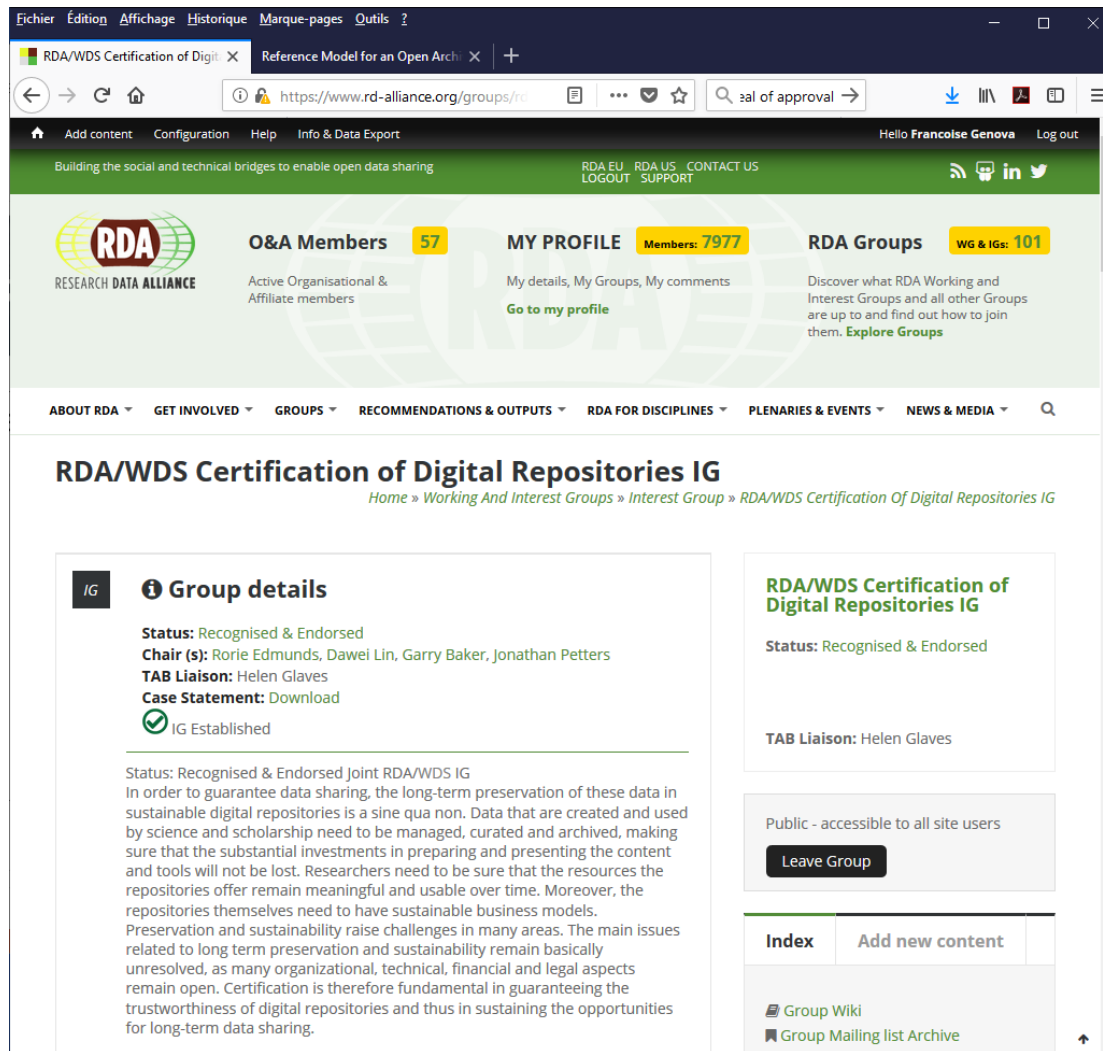
- **Basic Certification** is granted to repositories which obtain DSA certification
- **Extended Certification** is granted to Basic Certification repositories which *in addition* perform a structured, externally reviewed and publicly available self-audit based on DIN 31644/nestorSeal
- **Formal Certification** 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

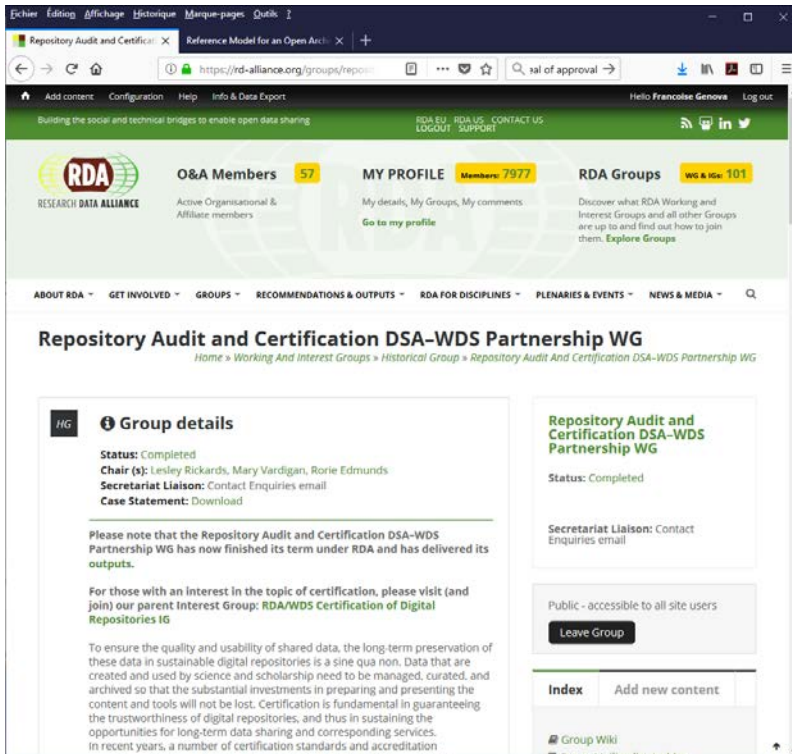


- Plutôt Humanités
- Plus dépôts de données que services
- Le CDS a été le premier centre certifié DSA du domaine des sciences physiques (en 2014)



The screenshot shows a web browser window displaying the RDA website. The address bar shows the URL: <https://www.rd-alliance.org/groups/rda-wds-certification-of-digital-repositories-ig>. The page header includes navigation links like 'Add content', 'Configuration', 'Help', and 'Info & Data Export'. The main content area features the RDA logo and statistics: 'O&A Members 57', 'MY PROFILE Members: 7977', and 'RDA Groups WG & IGs: 101'. Below this is a navigation menu with categories like 'ABOUT RDA', 'GET INVOLVED', 'GROUPS', etc. The main heading is 'RDA/WDS Certification of Digital Repositories IG'. The page content is divided into two columns. The left column, titled 'Group details', lists the status as 'Recognised & Endorsed', the chair(s) as 'Rorie Edmunds, Dawel Lin, Garry Baker, Jonathan Petters', the TAB Liaison as 'Helen Glaves', and the case statement as 'Download'. It also features a green checkmark icon and the text 'IG Established'. Below this is a paragraph of text explaining the importance of digital repository certification. The right column contains a summary of the group's status and the TAB Liaison's name. At the bottom right, there is a 'Leave Group' button and a section for 'Index' and 'Add new content', with links for 'Group Wiki' and 'Group Mailing list Archive'.

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 Wds & Icsu: 101

Repository Audit and Certification DSA-WDS Partnership WG

Home » Working And Interest Groups » Historical Group » Repository Audit And Certification DSA-WDS Partnership WG

Group details

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 DSA-WDS Partnership WG

Status: Completed

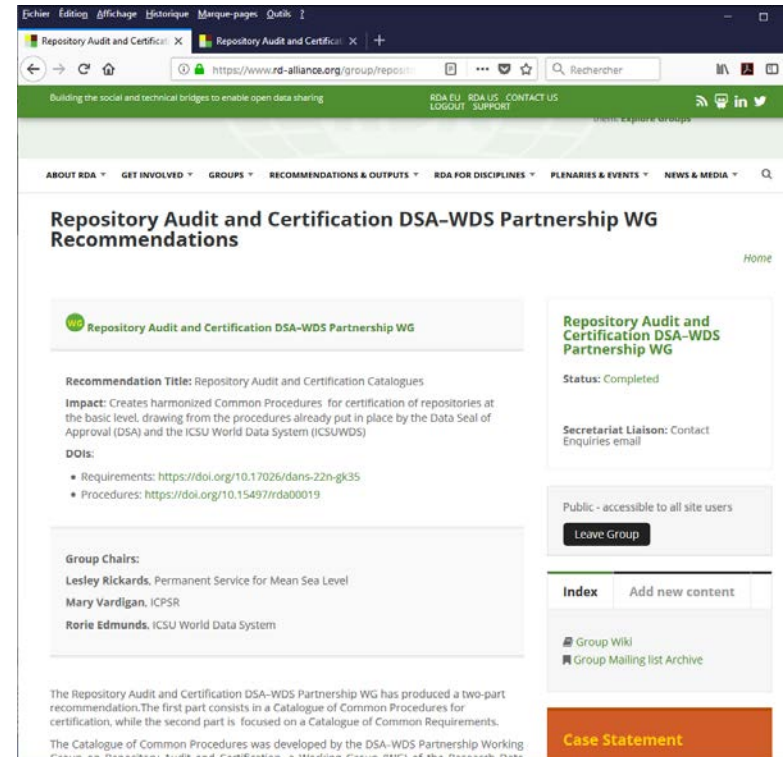
Secretariat Liaison: Contact Enquiries email

Public - accessible to all site users

Leave Group

Index Add new content

Group Wiki



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

Home

Repository Audit and Certification DSA-WDS Partnership WG

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

Repository Audit and Certification DSA-WDS Partnership WG

Status: Completed

Secretariat Liaison: Contact Enquiries email

Public - accessible to all site users

Leave Group

Index Add new content

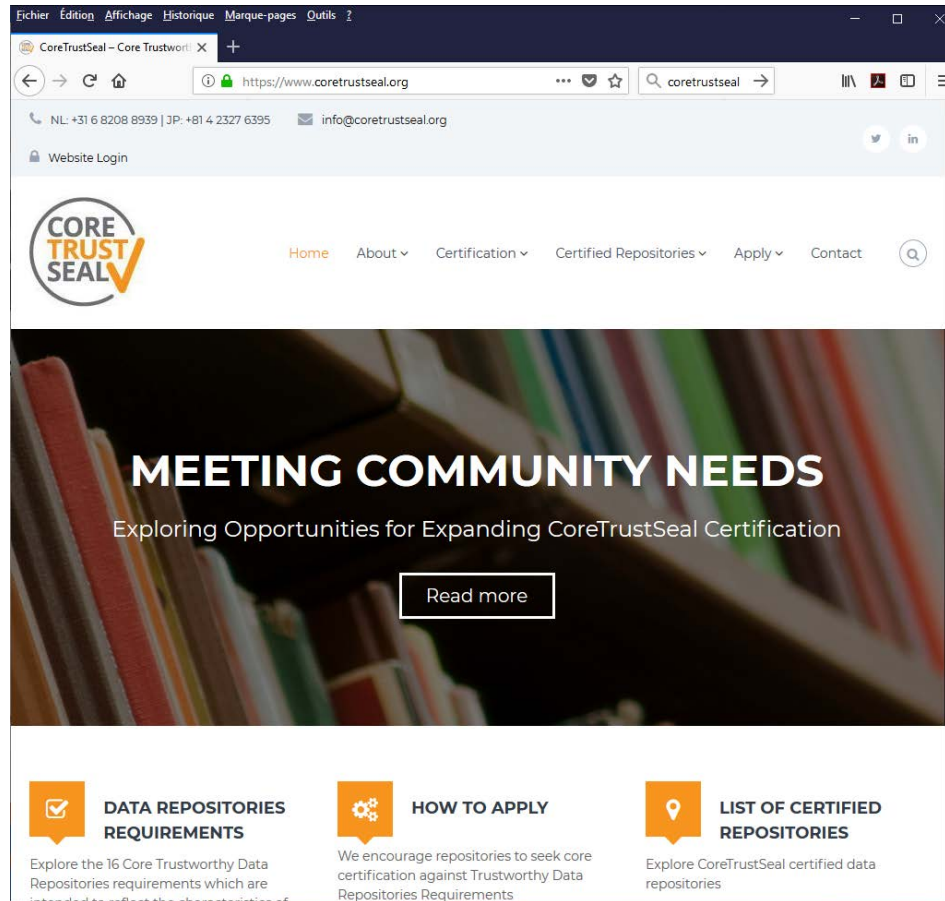
Group Wiki

Group Mailing list Archive

Case Statement

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

DSA + WDS = CoreTrustSeal (CTS)



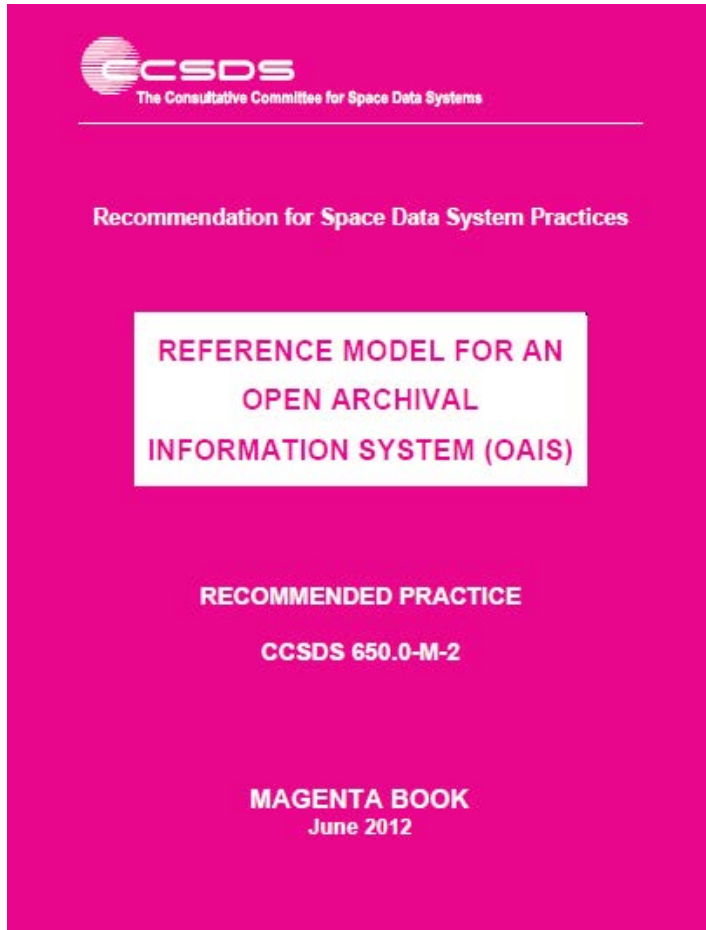
Exemple d'auto- évaluation

Le Centre de Données
astronomiques de Strasbourg (CDS)

Auto-evaluation

- 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



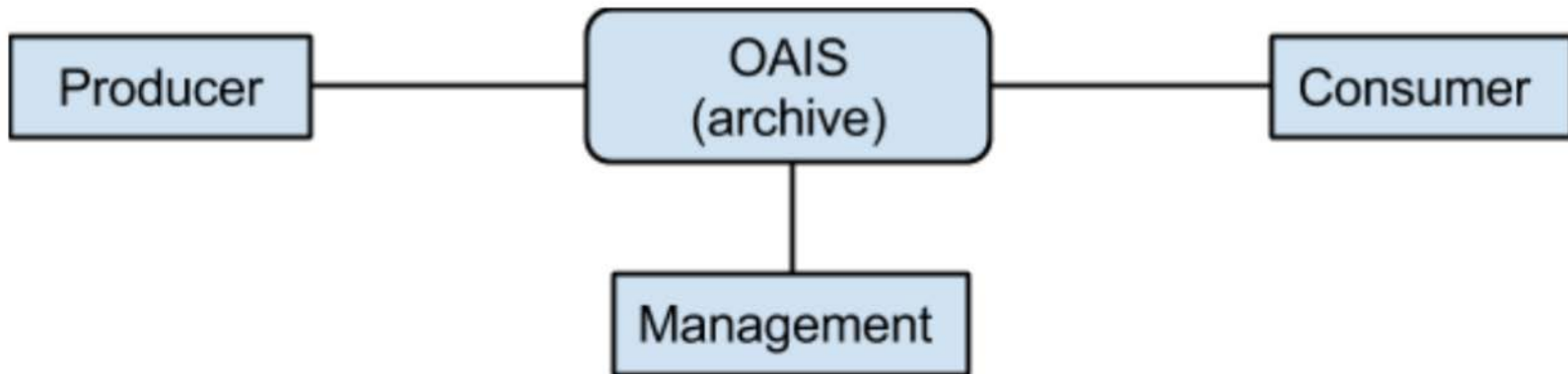
➤ Basé sur le modèle OAIS –
Open Archive Information
System

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

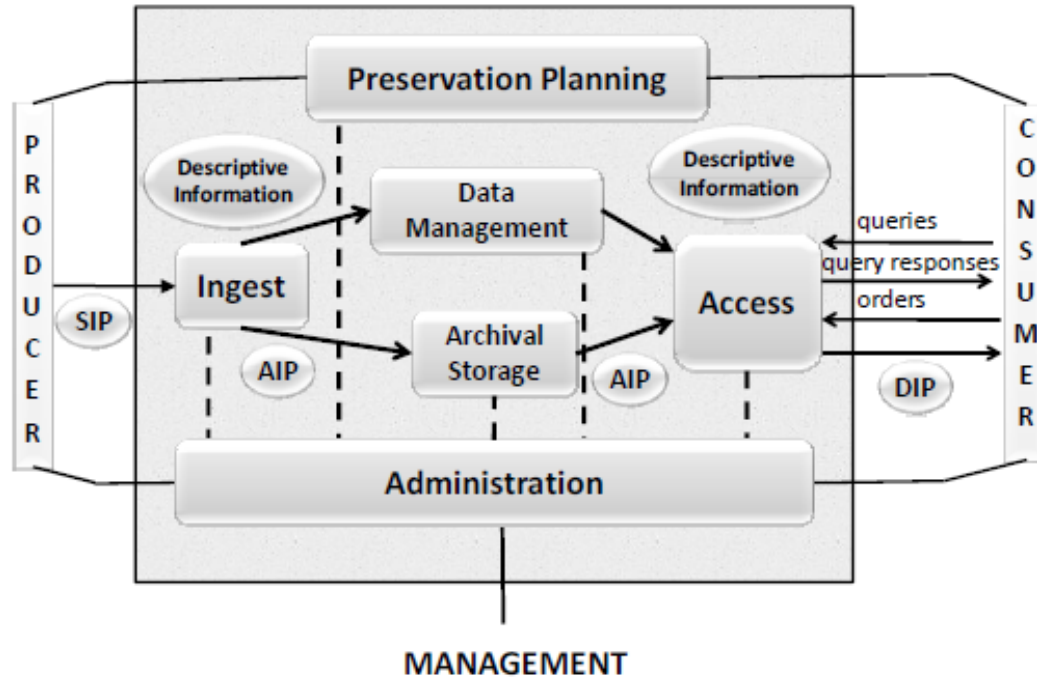
➤ Site en français

<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

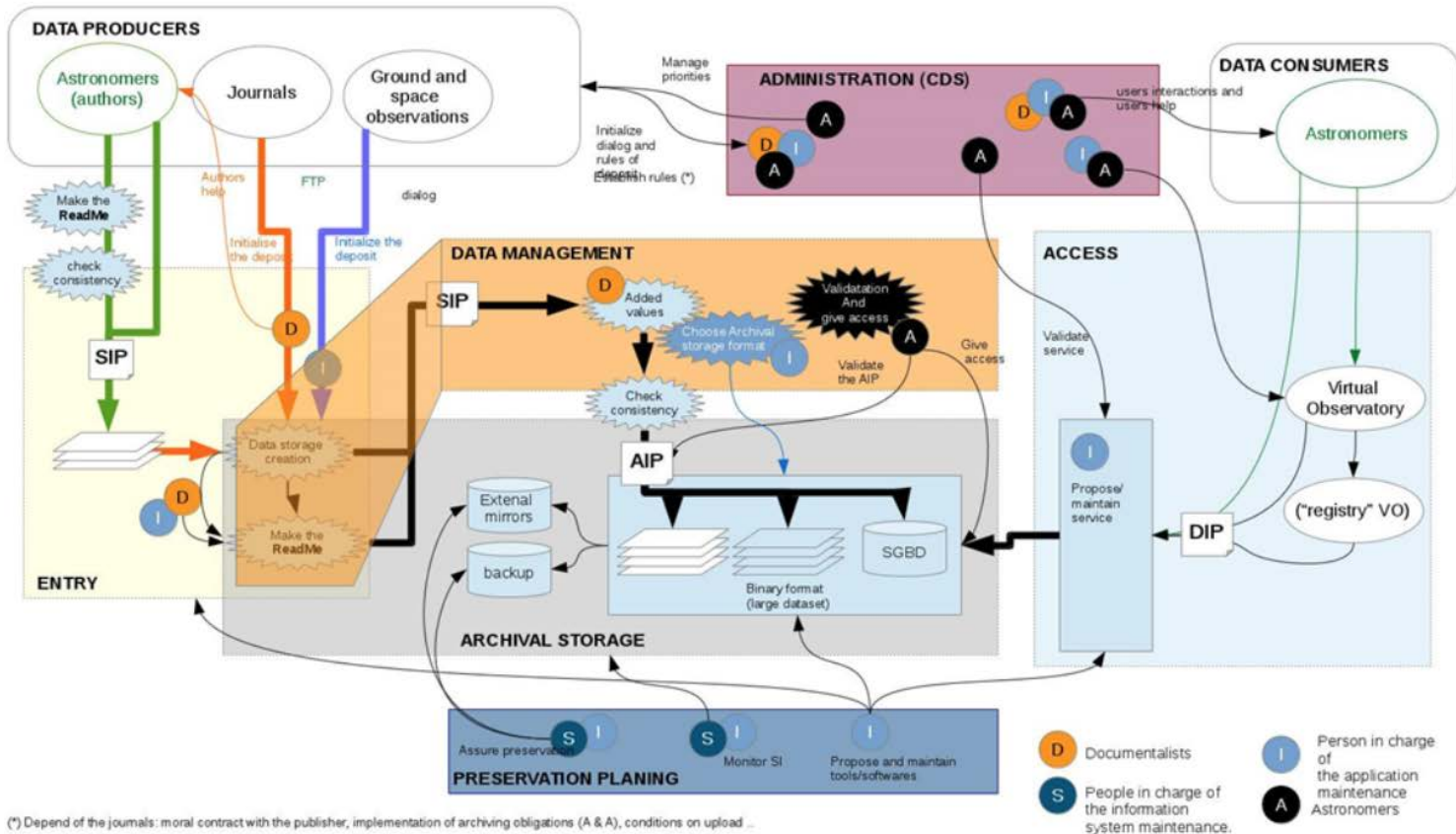


SIP: Submission Information Package

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

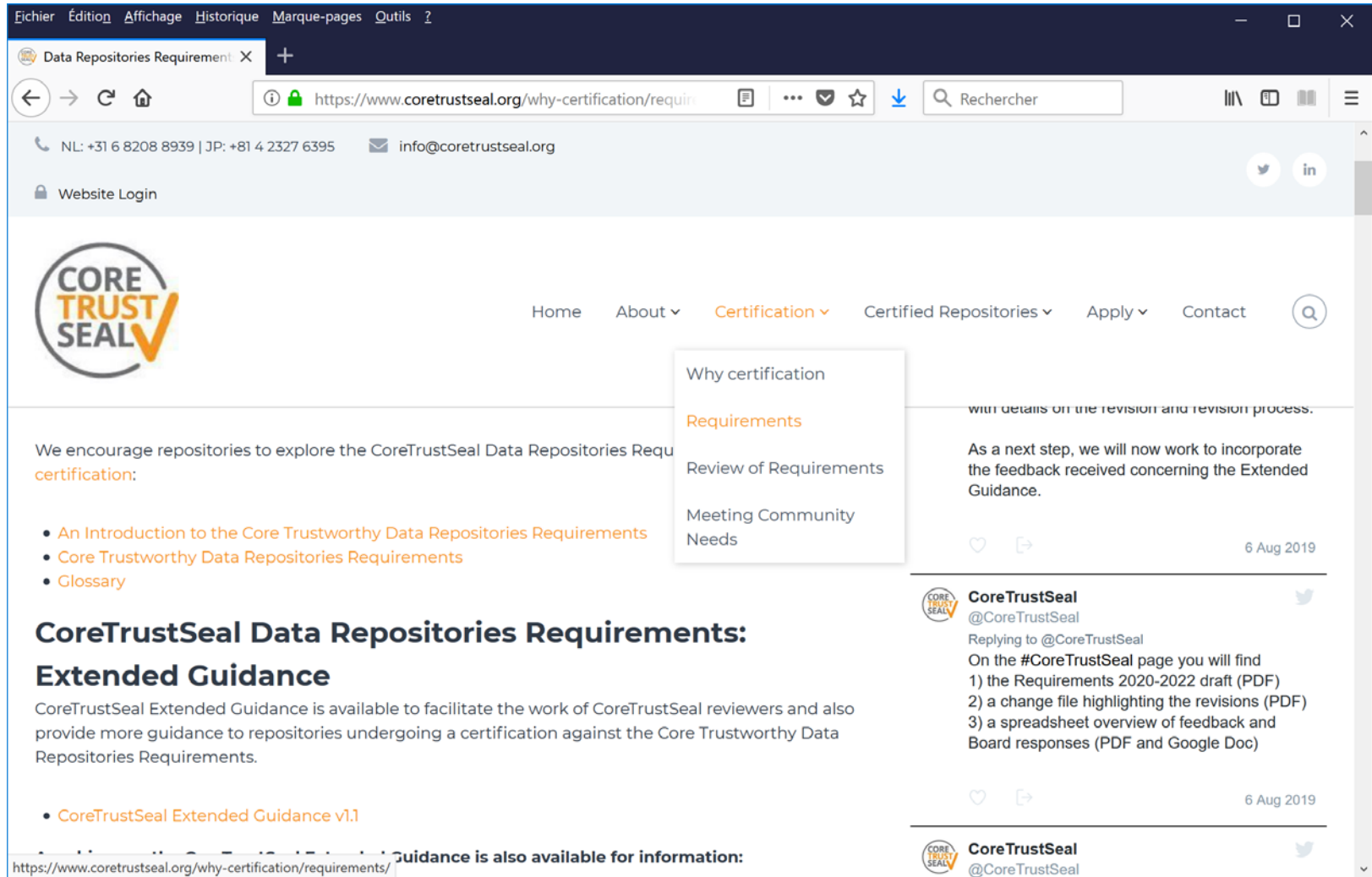
Les critères du CoreTrustSeal

La certification CTS

- › Toute l'information est sur le site de CoreTrustSeal
 - › <https://www.coretrustseal.org/>
- › Contexte + 16 critères
- › Document pour guider les évaluateurs et les candidats
 - › En cours, V1.1 2017-2019
 - <https://www.coretrustseal.org/wp-content/uploads/2017/01/20180629-CTS-Extended-Guidance-v1.1.pdf>
 - › Traduction française par RDA France
 - <https://www.rd-alliance.org/coretrustseal-criteres-de-conformite>
- › Guide en cours de révision pour 2020-2022 - Les critères ne changent pas!
 - › Version préliminaire et modifications
 - <https://www.coretrustseal.org/why-certification/review-of-requirements/>
- › « Administrative fee » 1000€

Transition

- Les candidatures sont acceptées dans l'ancien système jusqu'au 31 octobre, 23:59 UTC
- Réouverture en janvier
- En fait la mise à jour est essentiellement une clarification: on peut travailler sur la nouvelle version (et faire une petite vérification avant de soumettre)



The screenshot shows a web browser window displaying the CoreTrustSeal website. The address bar shows the URL: <https://www.coretrustseal.org/why-certification/requirements/>. The website header includes the CoreTrustSeal logo and a navigation menu with items: Home, About, Certification, Certified Repositories, Apply, and Contact. The 'Certification' menu is open, showing a sub-menu with: Why certification, Requirements (highlighted), Review of Requirements, Meeting Community Needs, and Meeting Community Needs.

The main content area features a section titled 'CoreTrustSeal Data Repositories Requirements: Extended Guidance'. Below this title, there is a paragraph of text and a list of links:

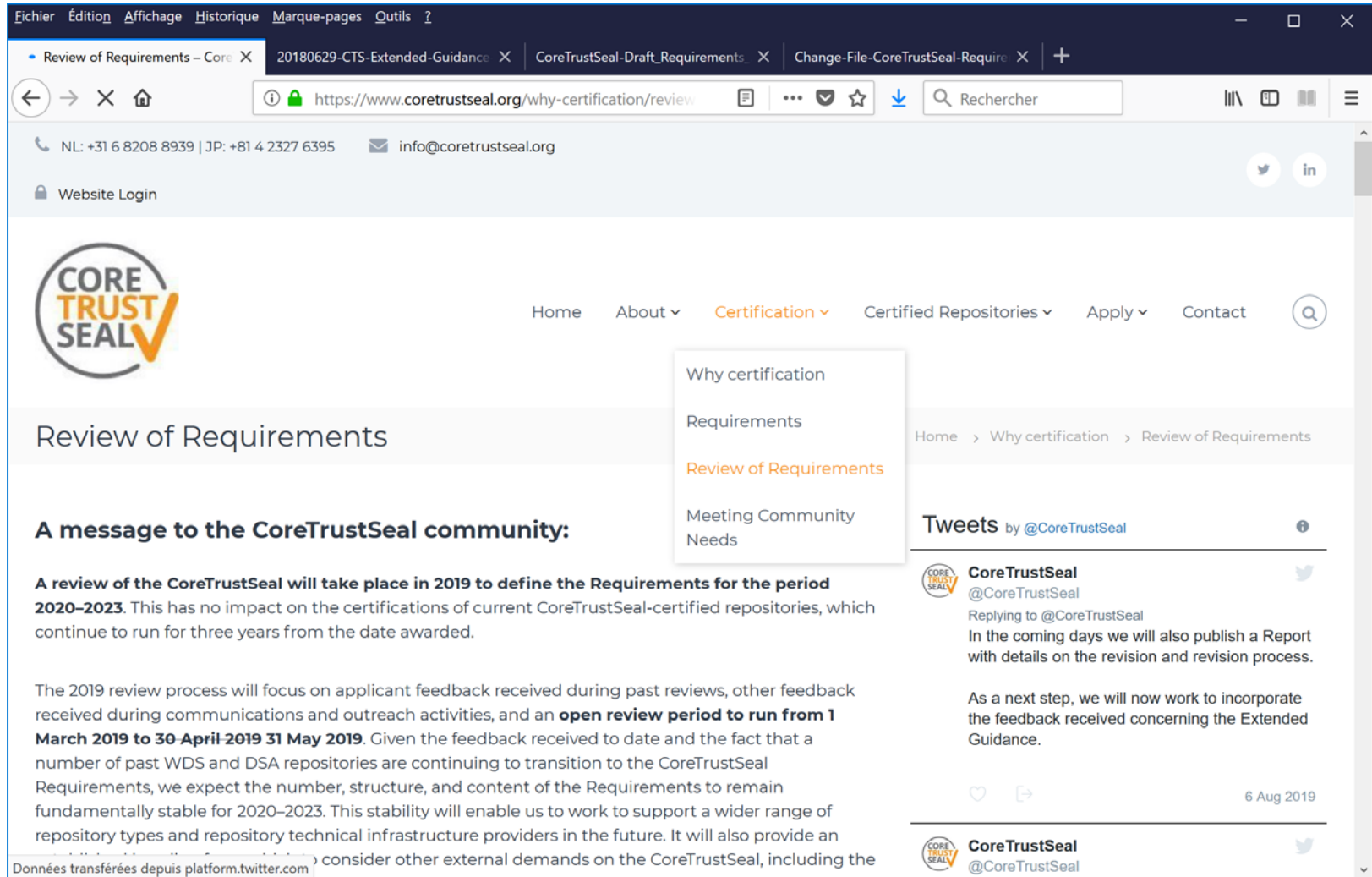
- An Introduction to the Core Trustworthy Data Repositories Requirements
- Core Trustworthy Data Repositories Requirements
- Glossary

Below the list, there is another link: [CoreTrustSeal Extended Guidance v1.1](#).

At the bottom of the main content area, there is a section titled 'Guidance is also available for information:' followed by a URL: <https://www.coretrustseal.org/why-certification/requirements/>.

On the right side of the screenshot, there is a Twitter feed. One tweet is visible, dated 6 Aug 2019, from @CoreTrustSeal. The tweet text reads: 'Replying to @CoreTrustSeal On the #CoreTrustSeal page you will find 1) the Requirements 2020-2022 draft (PDF) 2) a change file highlighting the revisions (PDF) 3) a spreadsheet overview of feedback and Board responses (PDF and Google Doc)'. There are also icons for liking and retweeting the tweet.

Information sur la révision en cours



Review of Requirements – Core X | 20180629-CTS-Extended-Guidance X | CoreTrustSeal-Draft_Requirements X | Change-File-CoreTrustSeal-Require X

https://www.coretrustseal.org/why-certification/review

NL: +31 6 8208 8939 | JP: +81 4 2327 6395 | info@coretrustseal.org

Website Login

CORE TRUST SEAL

Home About Certification Certified Repositories Apply Contact

Home > Why certification > Review of Requirements

Review of Requirements

A message to the CoreTrustSeal community:

A review of the CoreTrustSeal will take place in 2019 to define the Requirements for the period 2020–2023. This has no impact on the certifications of current CoreTrustSeal-certified repositories, which continue to run for three years from the date awarded.

The 2019 review process will focus on applicant feedback received during past reviews, other feedback received during communications and outreach activities, and an **open review period to run from 1 March 2019 to 30 April 2019 31 May 2019**. Given the feedback received to date and the fact that a number of past WDS and DSA repositories are continuing to transition to the CoreTrustSeal Requirements, we expect the number, structure, and content of the Requirements to remain fundamentally stable for 2020–2023. This stability will enable us to work to support a wider range of repository types and repository technical infrastructure providers in the future. It will also provide an opportunity to consider other external demands on the CoreTrustSeal, including the

Tweets by @CoreTrustSeal

CoreTrustSeal @CoreTrustSeal
Replying to @CoreTrustSeal
In the coming days we will also publish a Report with details on the revision and revision process.

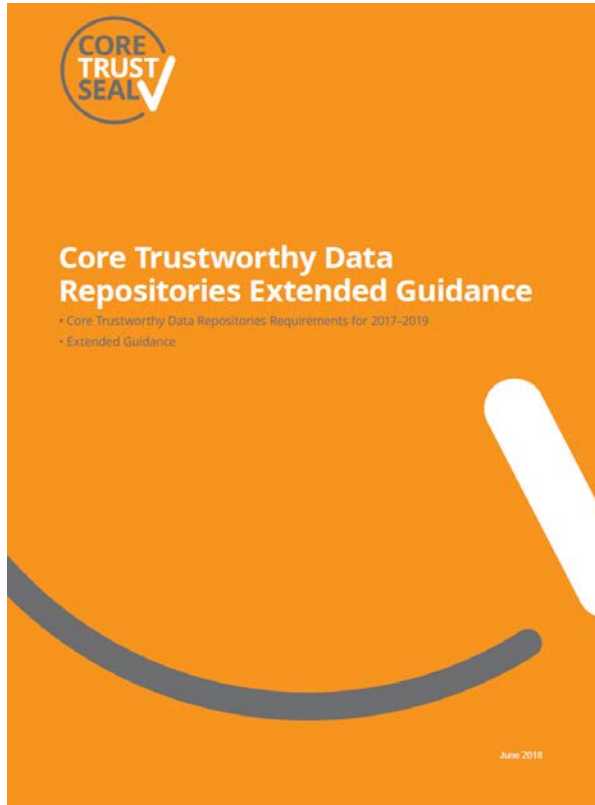
As a next step, we will now work to incorporate the feedback received concerning the Extended Guidance.

6 Aug 2019

CoreTrustSeal @CoreTrustSeal

Données transférées depuis platform.twitter.com

Les critères de certification CTS



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

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

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 (tout compris)
- › Evaluation interne
- › Evaluation externe
- › Importance croissante pour les financeurs des centres de données et des projets (DMP)
- › Priorité au niveau politique en France

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

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 quand la version 2020-2022 sera finalisée)
- › Ateliers, présentations dans des réunions à la demande, discussions au téléphone, etc
- › Une question récurrente: granularité pour les réseaux de dépôts de données/services. Sera discuté en plénière le 13.
- › Relayer les initiatives du projet FAIRsFAIR et les appels d'offre, les informations sur les activités de la RDA, etc
- › Inscrivez-vous sur la liste de diffusion

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



RDA in a Nutshell/ RDA en Bref

WWW.RD-ALLIANCE.ORG/
[@RESDATALL](https://twitter.com/RESDATALL)



RDA Global

Email - enquiries@rd-alliance.org

Web - www.rd-alliance.org

Twitter - [@resdatall](https://twitter.com/resdatall)

LinkedIn - www.linkedin.com/in/ResearchDataAlliance

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

RDA Europe

Email - info@europe.rd-alliance.org

Twitter - [@RDA_Europe](https://twitter.com/RDA_Europe)

RDA FRANCE

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

Email - contact-rdafrance@services.cnrs.fr

43

