



# La certification des entrepôts de données

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

Aude Chambodut, EOST, CNRS INSU, Board CTS

Olivier Rouchon, CNRS, FAIR-IMPACT, Board CTS

Gilles Ohanessian, GT Certification



# La certification des entrepôts de données

- Pourquoi?
- La procédure de certification
- Les critères du CoreTrustSeal
- Conclusions
- Annexes
  - Historique du CoreTrustSeal. Le rôle de la RDA
  - Le modèle OAIS
  - DataVerse et CoreTrustSeal



# La certification, pourquoi?

# Confiance (Trustworthiness)

- La confiance est au cœur de la Science Ouverte
- Les entrepôts sont une source majeure de données
- Il leur faut avoir la confiance
  - de leurs utilisateurs
    - Les personnes qui produisent et déposent les données
    - Les utilisateurs des données
  - de leurs autorités et de leurs financeurs

# Qu'est-ce qu'un entrepôt de confiance?

- Mission de fournir un accès fiable, long-terme à des ressources numériques contrôlées pour sa communauté cible, maintenant et dans le futur
- Supervision, planification et maintenance constantes
- Compréhension des menaces et risques dans les systèmes
- **Cycle d'audit et/ou de certification régulier**



*Credits: Ingrid Dillo & Hervé L'Hours*

# Qu'est-ce qu'un entrepôt de confiance?

- Mission de fournir un accès fiable, long-terme à des ressources numériques contrôlées pour sa communauté cible, maintenant et dans le futur
- Supervision, planification et maintenance constantes
- Compréhension des menaces et risques dans les systèmes
- **Cycle d'audit et/ou de certification régulier** ?



*Credits: Ingrid Dillo & Hervé L'Hours*



# 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 CDS
  - Créé en 1972
  - Centre de données de référence pour la communauté astronomique internationale
  - Infrastructure de Recherche sur la Feuille de Route nationale
  - 2 000 000+ requêtes/jour sur les services



# Pourquoi une certification formelle?

- 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 Plans de Gestion des Données (PGD)

# La stratégie Nationale

## Plan National pour la Science Ouverte

2018

### 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

2021

- Poursuivre le processus de **certification** (*Core trust seal*) des entrepôts de données français.

# La stratégie au CNRS



## 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 système de données SOLERIL a également mis en route une politique de gestion des données. Ces exemples sont multiples et devraient tendre à être généralisés. 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 disciplinaires ». Le CNRS continuera à soutenir ces infrastructures, et soutiendra le développement de nouveaux réservoirs 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.



p.8

- Le CNRS constituera à l'intention des chercheurs et des chercheuses un annuaire des entrepôts et des services de données existants, avec en particulier l'objectif d'aller vers la certification des entrepôts et services de données.

Un entrepôt doit avoir un rôle de curation et de préservation des données, et les principes FAIR sont un objectif dans le contexte Science Ouverte. La certification de base *CoreTrustSeal* explicite les critères pour un entrepôt « de confiance », ce qui permet de travailler à améliorer les pratiques en se basant sur les critères, sans nécessairement aller jusqu'à soumettre un dossier de certification.

p. 11

**Certification des dispositifs de prise en charge des données de la recherche (notamment le *CoreTrustSeal*<sup>5</sup>).** La certification des entrepôts et services de données, citée comme un objectif dans le Plan National pour la Science Ouverte, permet d'assurer qu'un centre de données est « de confiance », en examinant la manière dont il met en œuvre l'ensemble de la chaîne liée aux données, de leur ingestion à leur dissémination et à leur préservation. Elle peut aussi s'entendre dans le cadre de réseaux de centres de données, par exemple ceux des Pôles de données thématiques de l'IR Data Terra<sup>6</sup>, ou ceux de l'infrastructure européenne CLARIN<sup>7</sup>. Le CNRS pourra s'appuyer sur les activités de soutien à la certification mises en place par le Nœud National RDA France<sup>8</sup>.



# Entrepôts français actuellement certifiés CTS

- [CDS](#), Strasbourg
- [CDSP](#), Paris
- [HAL](#), Paris
- [IDOC/IDOC-DATA](#), Orsay
- [IFREMER SISMER](#), Plouzané
- [IPSL](#), Paris
- [ORTOLANG](#), Nancy
- [RESIF](#), Grenoble
- [SHOM](#), Brest
- [ESRF](#), Grenoble



# Les cadres de certification



Certification formelle ISO 16363:2012 > 2025  
(2 Centres ont été certifiés, 100+ critères)  
<https://www.iso.org/standard/87472.html>

Certification étendue DIN 31644  
(6 Centres certifiés, 34 critères)  
[www.langzeitarchivierung.de](http://www.langzeitarchivierung.de)

Certification CoreTrustSeal  
(200+ centres, 16 critères)  
[www.CoreTrustSeal.org](http://www.CoreTrustSeal.org)

Centres de données de confiance - Trustworthy Data  
Repositories

*Credits: Mustapha Mokrane*



# La procédure de certification



# La procédure d'évaluation (1)

- 🌐 Auto-évaluation de l'entrepôt
- 🌐 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





## La procédure d'évaluation (2)

- Les éléments du dossier sont publics (sauf exception légitime)
- Les évaluateurs sont membres de la communauté CoreTrustSeal « Assemblée des évaluateurs »
- Pour devenir membre du World Data System (WDS)
  - Il faut obtenir la certification CTS...
  - et remplir quelques conditions supplémentaires
- Renouvellement de la certification tous les 3 ans



# Les critères du CoreTrustSeal



# La certification CTS

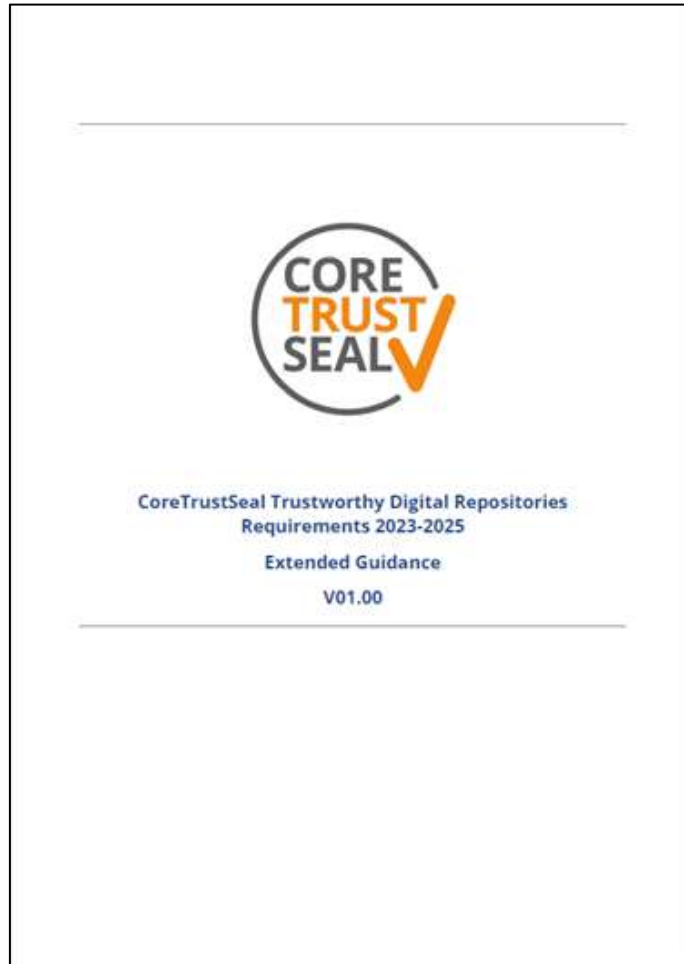
- Toute l'information est sur le site de CoreTrustSeal
  - <https://www.coretrustseal.org/>
- Contexte + 16 critères
  - <https://www.coretrustseal.org/why-certification/requirements/>
- Document pour guider les évaluateurs et les candidats
  - Extended Guidance 2020-2022 (V2.0) > 2023-2025 (V3.0)  
<https://doi.org/10.5281/zenodo.7051096>
  - Glossaire  
<https://doi.org/10.5281/zenodo.7051125>
  - Changements V2.0 > V3.0  
<https://doi.org/10.5281/zenodo.7051237>
- Lorsque les critères évoluent, les entrepôts qui ont soumis un dossier selon les anciens critères restent dans ce cadre
- « Administrative fee » 3000€



# Les critères sur le site CTS

A screenshot of the CoreTrustSeal website. The page title is 'Data Repositories Requirements'. The main heading is 'CoreTrustSeal Trustworthy Data Repositories Requirements'. The text explains that these requirements reflect the characteristics of trustworthy repositories and are mandatory. It also encourages repositories to explore these requirements and seek certification. A section titled 'CoreTrustSeal Requirements 2023-2025:' lists two documents: 'CoreTrustSeal Requirements 2023-2025' and 'CoreTrustSeal Extended Guidance 2023-2025'. The extended guidance is described as available to facilitate the work of reviewers and provide more guidance to repositories undergoing certification. The website header includes the CoreTrustSeal logo, navigation links (Home, About, Certification, Certified Repositories, Apply, Contact), and contact information (The Hague, phone number, email).

# Les critères de certification CTS

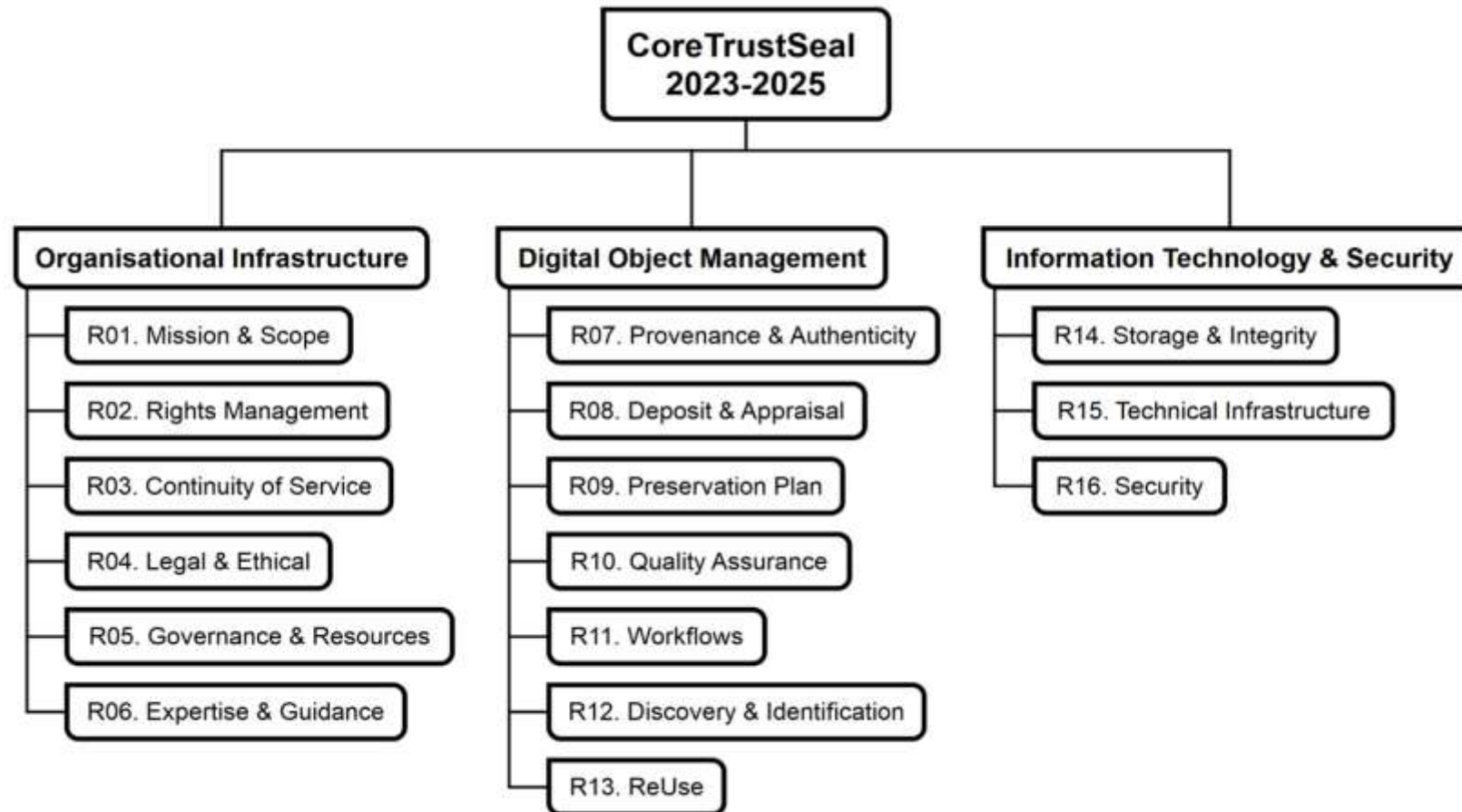


## 🌐 R0 - Le contexte

## 🌐 16 critères, 3 thèmes:

- Infrastructure organisationnelle
- Gestion des objets numériques (données et métadonnées)
- Technologie de l'information et sécurité

# Les critères 2023-2025





# Le contexte

- Identifiant Re3data
- 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 – Gestion des droits
- 🌐 R3 – Continuité de service
- 🌐 R4 – Légalité/éthique
- 🌐 R5 – Gouvernance et ressources
- 🌐 R6 – Conseils d'experts



# Gestion des objets numériques

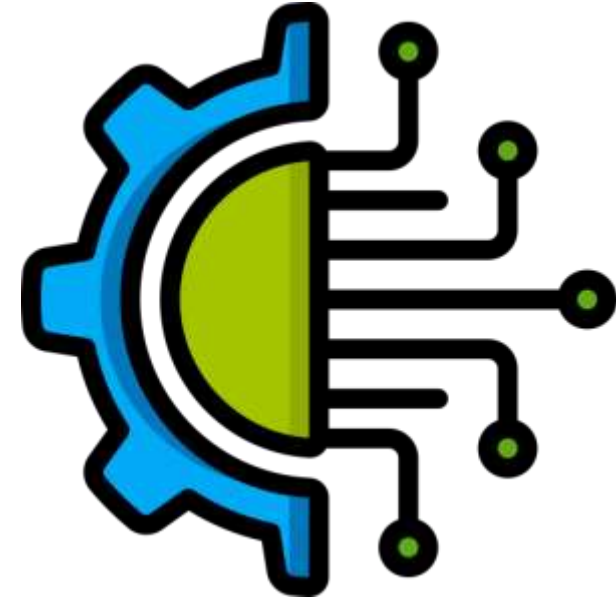
- 🌐 R7 – Provenance et authenticité des données
- 🌐 R8 – Appréciation et sélection des données
- 🌐 R9 – Assurance qualité
- 🌐 R10 – Plan de préservation
- 🌐 R11 – Processus de traitement (Workflows)
- 🌐 R12 – Découverte et identification des données
- 🌐 R13 – Réutilisation des données





# Technologie

- 🌐 R14 – Stockage et intégrité
- 🌐 R15 – Infrastructure technique
- 🌐 R16 – Sécurité





# Révision des critères 2026-2028

## Revue

- Revue des dossiers de candidatures des trois dernières années
- Proposition de changements et additions

## Consultation

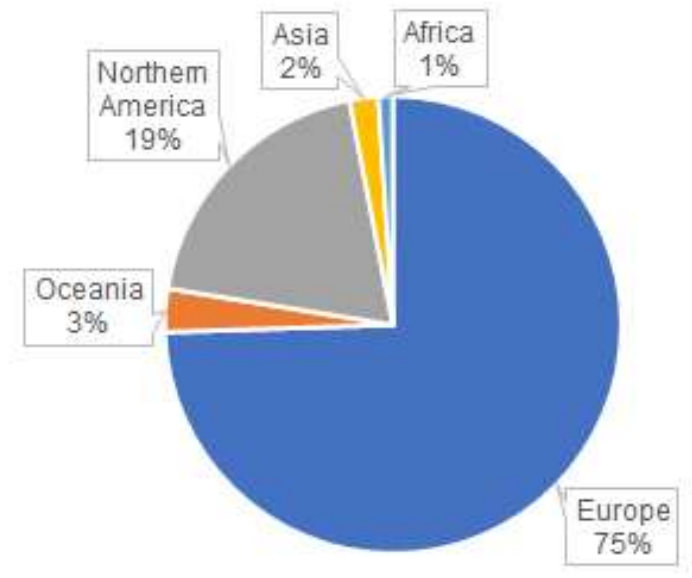
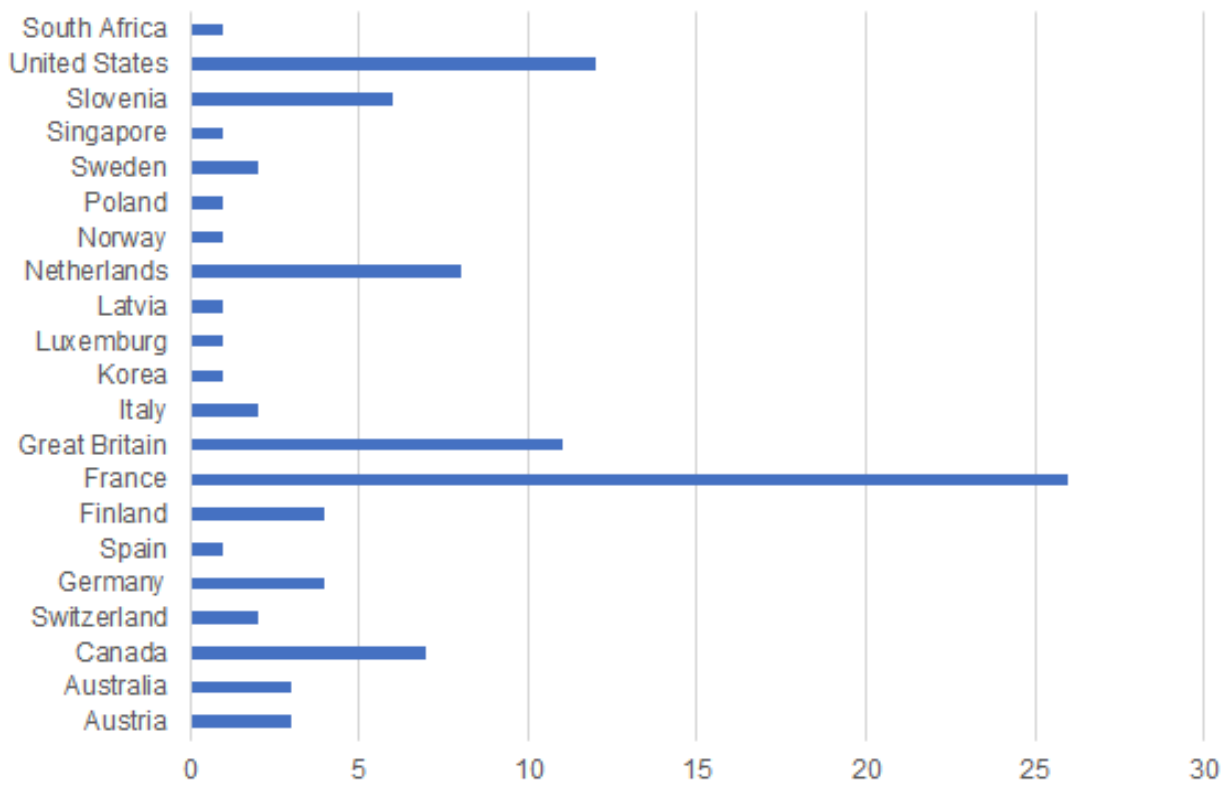
- Sondage de la communauté
- Revue des retours
- Incorporation des retours au projet de révision
- Revue par le Board et approbation de la nouvelle version

## Publication

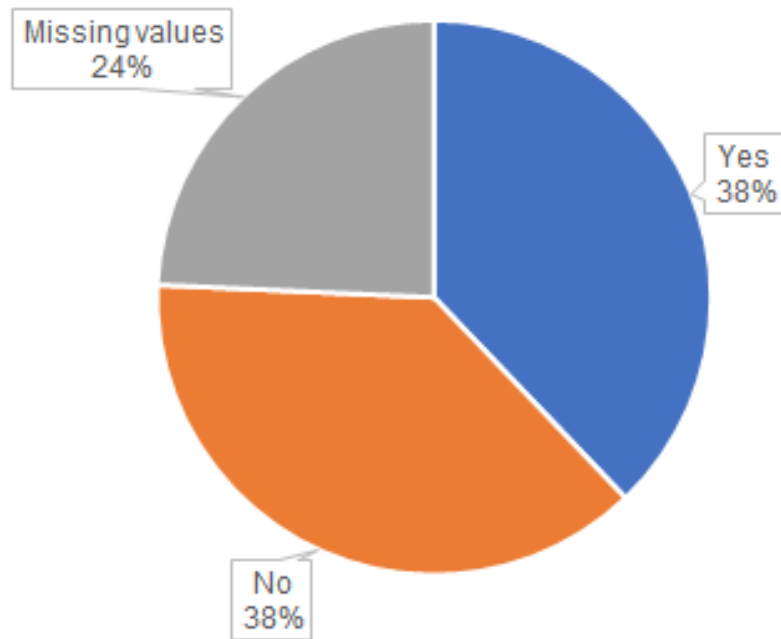
- Publication de la version finale
- Correction des dernières erreurs par la communauté et le Board
- Publication des critères 2026-2028



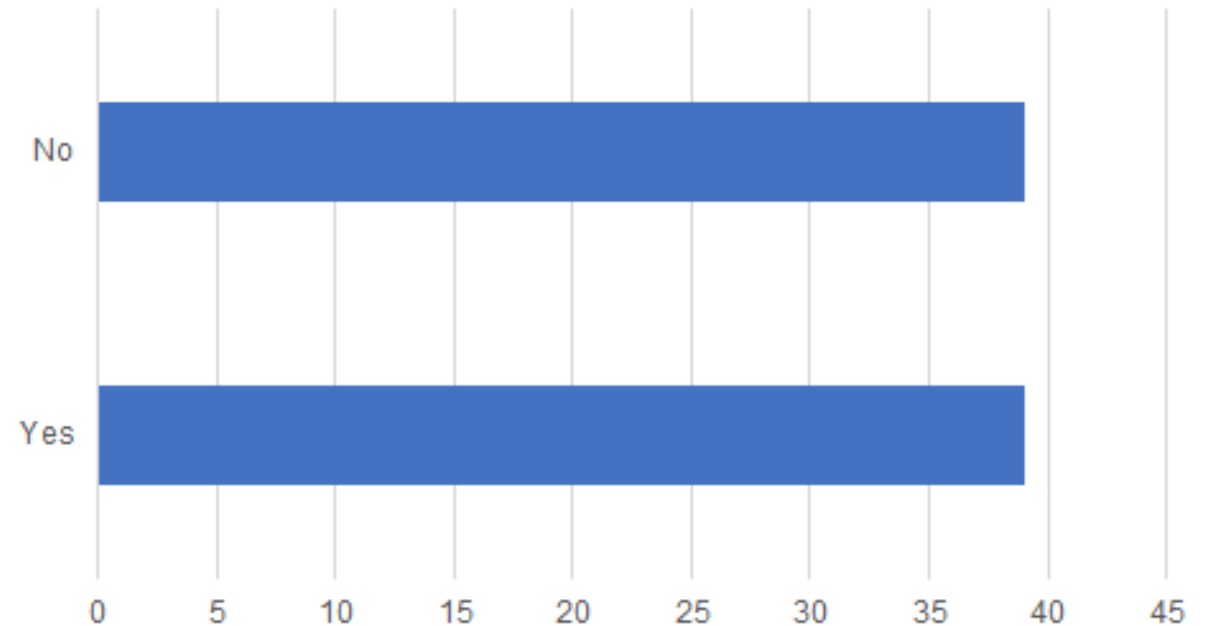
# Consultation de la communauté



# Origine des réponses

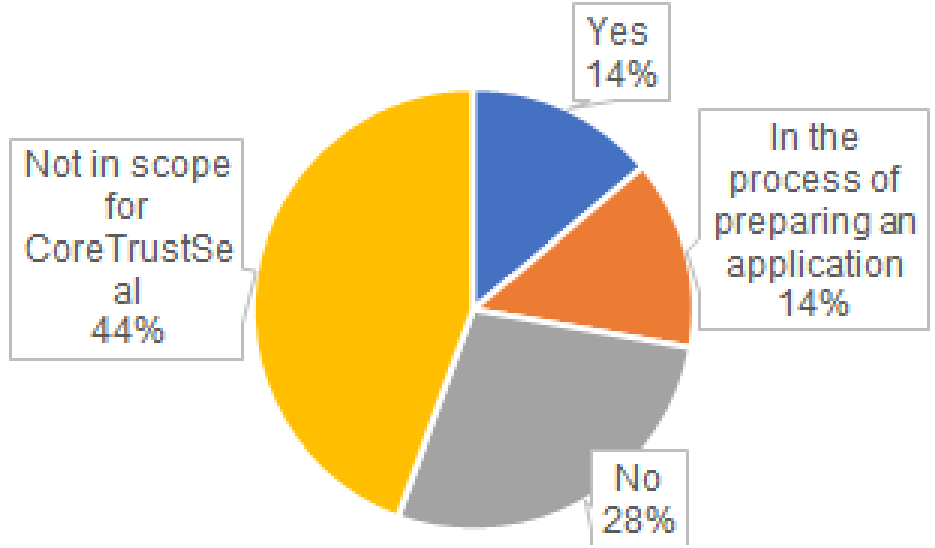
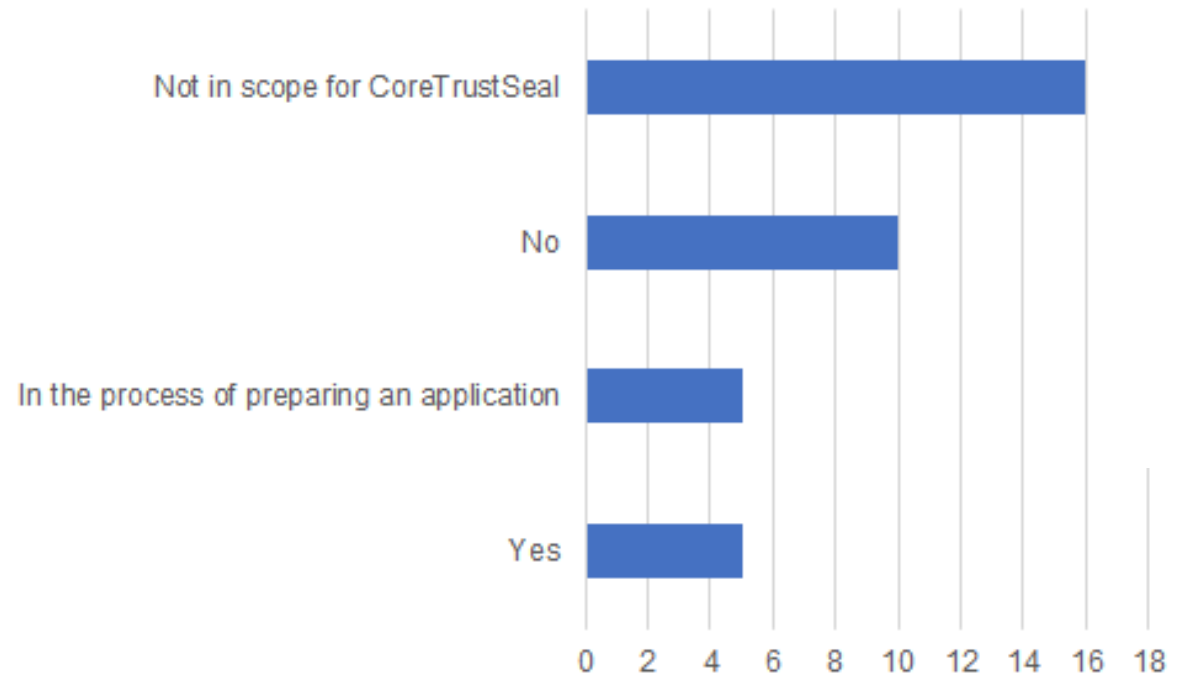


Please let us know if you are responding on behalf of a CoreTrustSeal Certified Repository?





If you are not from a CoreTrustSeal Certified Repository are you considering CoreTrustSeal?





# Background information

Proposals: The Board proposes to replace the name “Level of Curation” by “Level of Curation and Preservation” and change the levels with

- A. Active preservation
- B. Initial Curation
- C. Deposit Compliance
- D. Content distributed as deposited. Unattended deposit-storage-access”

as per the Position Paper (<https://zenodo.org/records/11476980>) published in 2024. Guidance will be amended accordingly, and additional explanations will be provided.

Yes, I agree with the changes proposed by the Board	20
I do NOT agree, I think other (additional) changes should be made (please suggest any changes or improvements below)	16
I do NOT agree, I think the text should stay the same as it was	1



# Organizational infrastructure

	Yes	No + other changes	No, no change
R1. Mission & scope	23	8	2
R2. Rights management	29	4	0
R3. <u>Continuity of service</u>	24	9	0
R4. <u>legal &amp; ethical</u>	24	9	0
R5. <u>Governance &amp; resources</u>	19	11	3
R6. Expertise & guidance	32	0	0

# Digital object management

	Yes	No + other changes	No, no change
R7. <u>Provenance &amp; authenticity</u>	23	7	3
R8. <u>Deposit &amp; appraisal</u>	17	12	4
R9. <u>Preservation plan</u>	29	4	0
R10. <u>Quality assurance</u>	32	1	0
R11. Workflows	31	2	0
R12. <u>Discovery &amp; identification</u>	30	2	0
R13. <u>Reuse</u>	30	2	0



# Information technology & security

	Yes	No + other changes	No, no change
R14. Storage & <u>integrity</u>	31	1	0
R15. <u>Technical</u> infrastructure	31	1	0
R16. Security	27	5	0



# Les critères les plus difficiles

🌐 R3 – Continuité de Service

🌐 R9 – Assurance qualité

🌐 R7 – R10

- R7 – Provenance et authenticité des données

- R10 – Plan de préservation

🌐 R2 – R6 – R13 – R14 – R15

- R2 – Gestion des droits

- R6 – Conseil d'experts

- R13 – Réutilisation des données

- R14 – Stockage et intégrité

- R15 – Infrastructure technique



# Conclusions



# Pourquoi la certification?

- 🌐 Quelques semaines de travail d'équipe dans la plupart des cas (tout compris)
  - Beaucoup plus pour d'autres mais ça continue à valoir le coup!
- 🌐 Evaluation – amélioration des process
  - Auto-évaluation
  - Evaluation externe
- 🌐 Importance croissante pour les financeurs des centres de données et des projets (PGD)
- 🌐 Priorité au niveau politique en France, intérêt des organismes (CNRS, Universités)



# L'impact de la RDA

- Fusion des deux cadres de certification 'de base', DSA (Data Seal of Approval) et WDS (World Data System)
- 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!
- Création de CoreTrustSeal
- Nombreux nouveaux candidats à la certification



# Le rôle de RDA France

- La certification est une priorité – aussi parce que **la réflexion sur les critères permet d'améliorer la manière dont on gère ses données, même si on ne va pas jusqu'à candidater à CoreTrustSeal**
- Groupe de Travail commun avec le collège Données du Comité pour la Science Ouverte depuis juillet 2021
  - Prise en charge des frais administratifs de certification
  - Outil de visualisation et de partage des dossiers de certification  
<https://crusoe.ouvrirlascience.fr/>
- Ateliers, présentations à la demande, etc.
  - Plus de 300 personnes ont suivi les ateliers
- Démarrage de la mise en réseau des entrepôts certifiés et « sur le chemin de la certification »
  - Ateliers pendant les Journées RDA France 2023 et 2024
- Liste de diffusion  
<https://listes.services.cnrs.fr/www/subscribe/rda-france-certification>



# Annexes

 Historique de CoreTrustSeal

 Le modèle OAIS

 Dataverse



# Historique de CoreTrustSeal



# Le paysage de la certification en 2015

4 certifications standards disponibles



DIN 31644

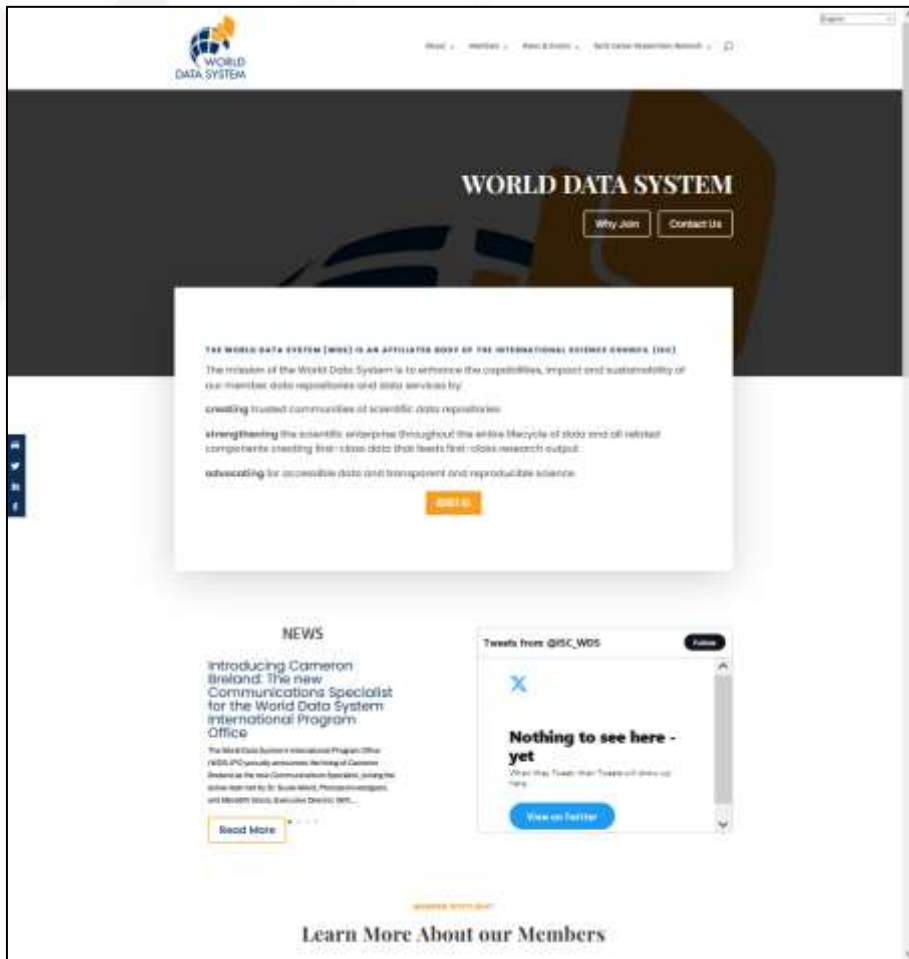


**ICSU**  
WORLD DATA SYSTEM



ISO 16363

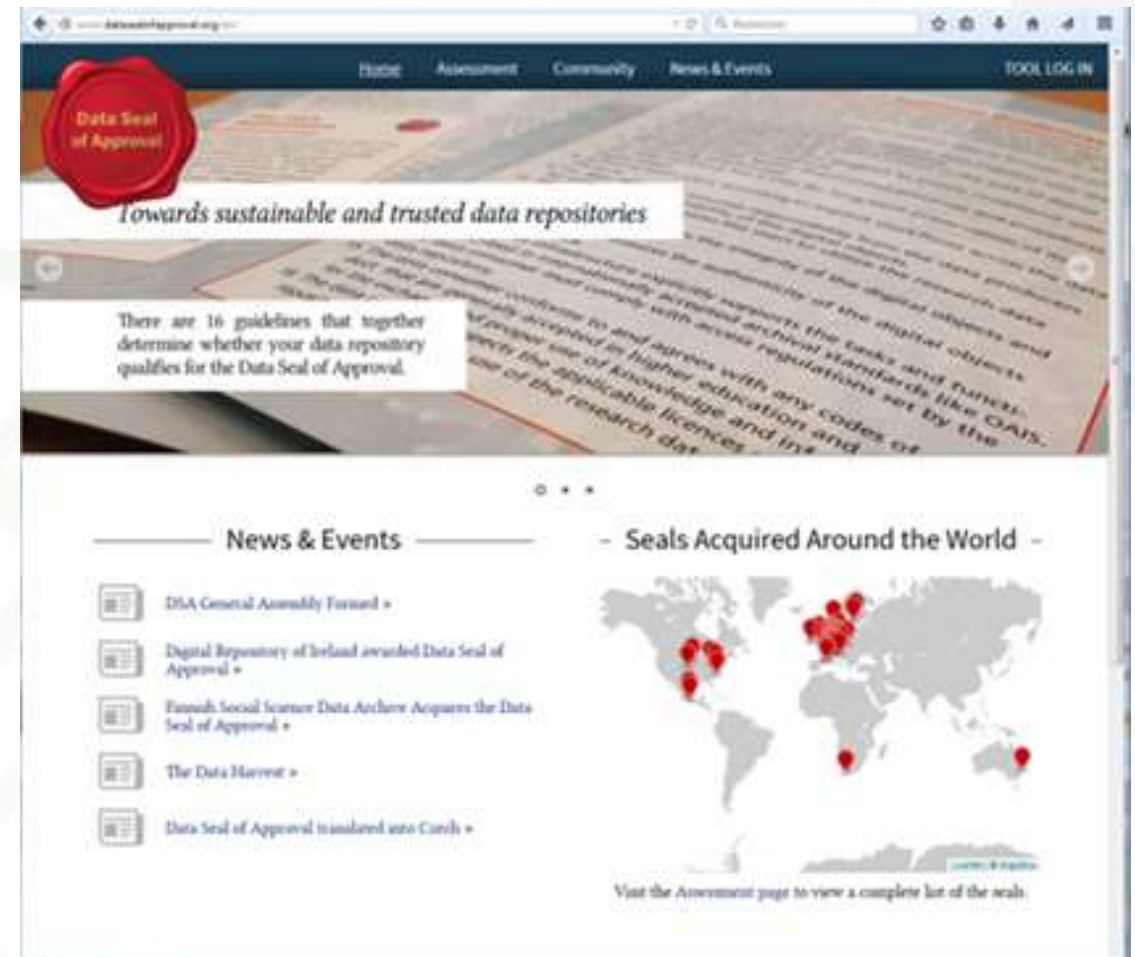
# Le World Data System (WDS)



- 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 depuis 2012

# Le Data Seal of Approval (DSA)

- 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

A screenshot of a web browser displaying the RDA website. The browser's address bar shows the URL 'https://www.rd-alliance.org/group...'. The page header includes navigation links like 'Add content', 'Configuration', 'Help', 'Info &amp; Data Export', and 'Hello Françoise Genève'. The main content area features a navigation menu with 'ABOUT RDA', 'GET INVOLVED', 'GROUPS', 'RECOMMENDATIONS &amp; OUTPUTS', 'RDA FOR DISCIPLINES', 'PLENARIES &amp; EVENTS', and 'NEWS &amp; MEDIA'. The central heading is 'RDA/WDS Certification of Digital Repositories IG'. Below this, there is a 'Group details' section with fields for 'Status: Recognised &amp; Endorsed', 'Chair (s):', 'TAB Liaison: Helen Graves', and 'Case Statement: Download'. A 'Leave Group' button is visible in the right-hand sidebar. The page also includes a 'Public - accessible to all site users' notice and an 'Index' section with an 'Add new content' link.



# RDA: DSA + WDS (2016)

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 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)

DOI:

- Requirements: <https://doi.org/10.17026/dans-22n-gk35>
- Procedures: <https://doi.org/10.15497/rd00019>

Group Chairs:

Lesley Rickards, Permanent Service for Mean Sea Level  
 Mary Vardigan, ICPSR  
 Rorie Edmunds, ICSU World Data System

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

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





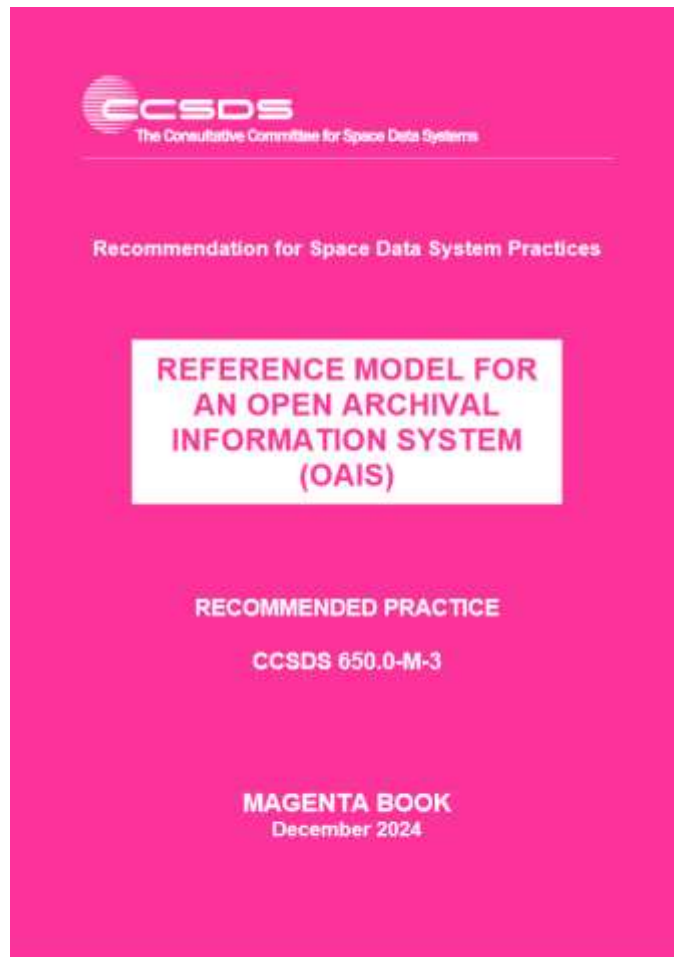
# DSA + WDS = CoreTrustSeal (CTS), 2017





# Le modèle OAIS

# Le modèle OAIS



● Modèle de référence sous-jacent dans CoreTrustSeal OAIS – Open Archive Information System Version 2

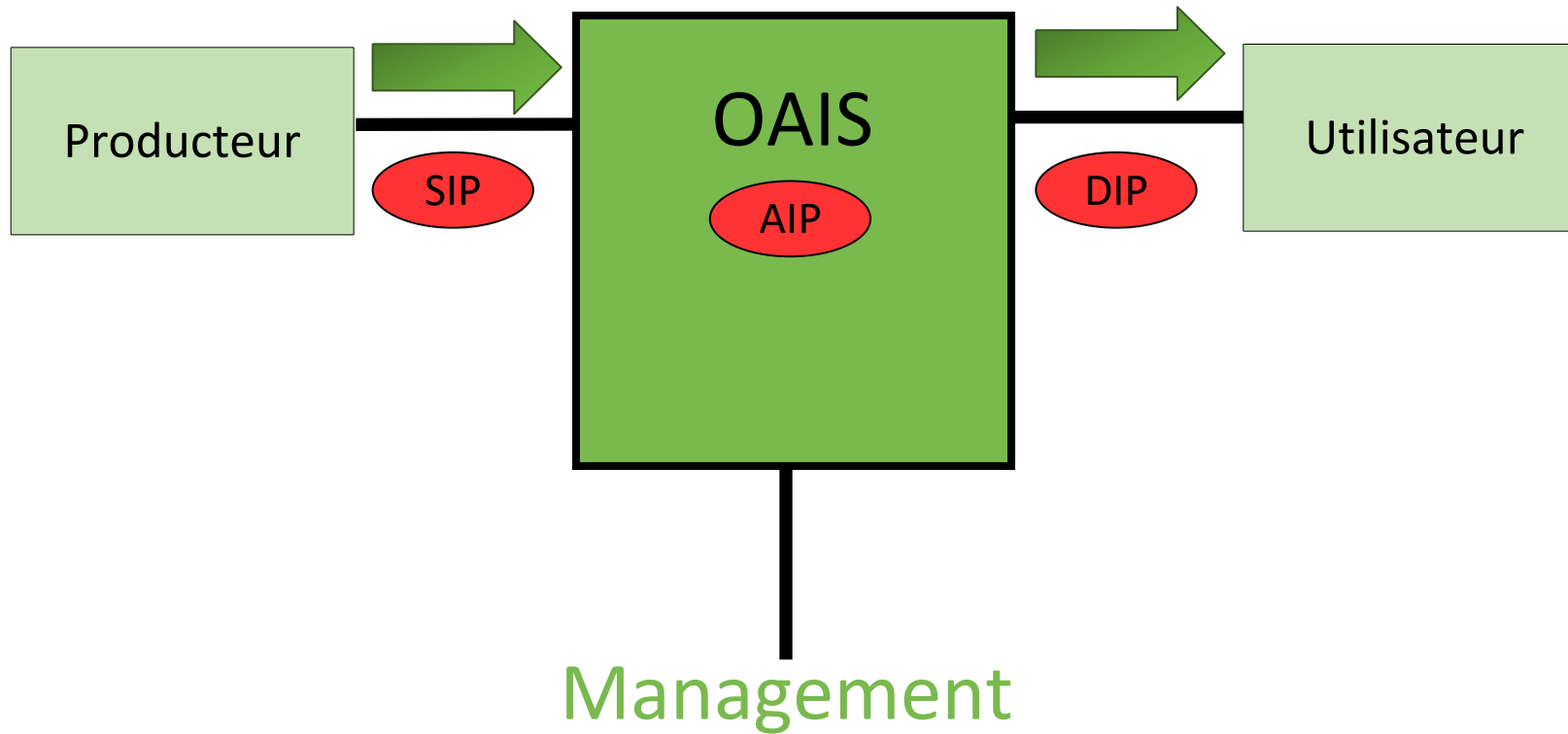
● La version actuelle: Version 3, décembre 2024

<https://public.ccsds.org/Pubs/650x0m3.pdf>

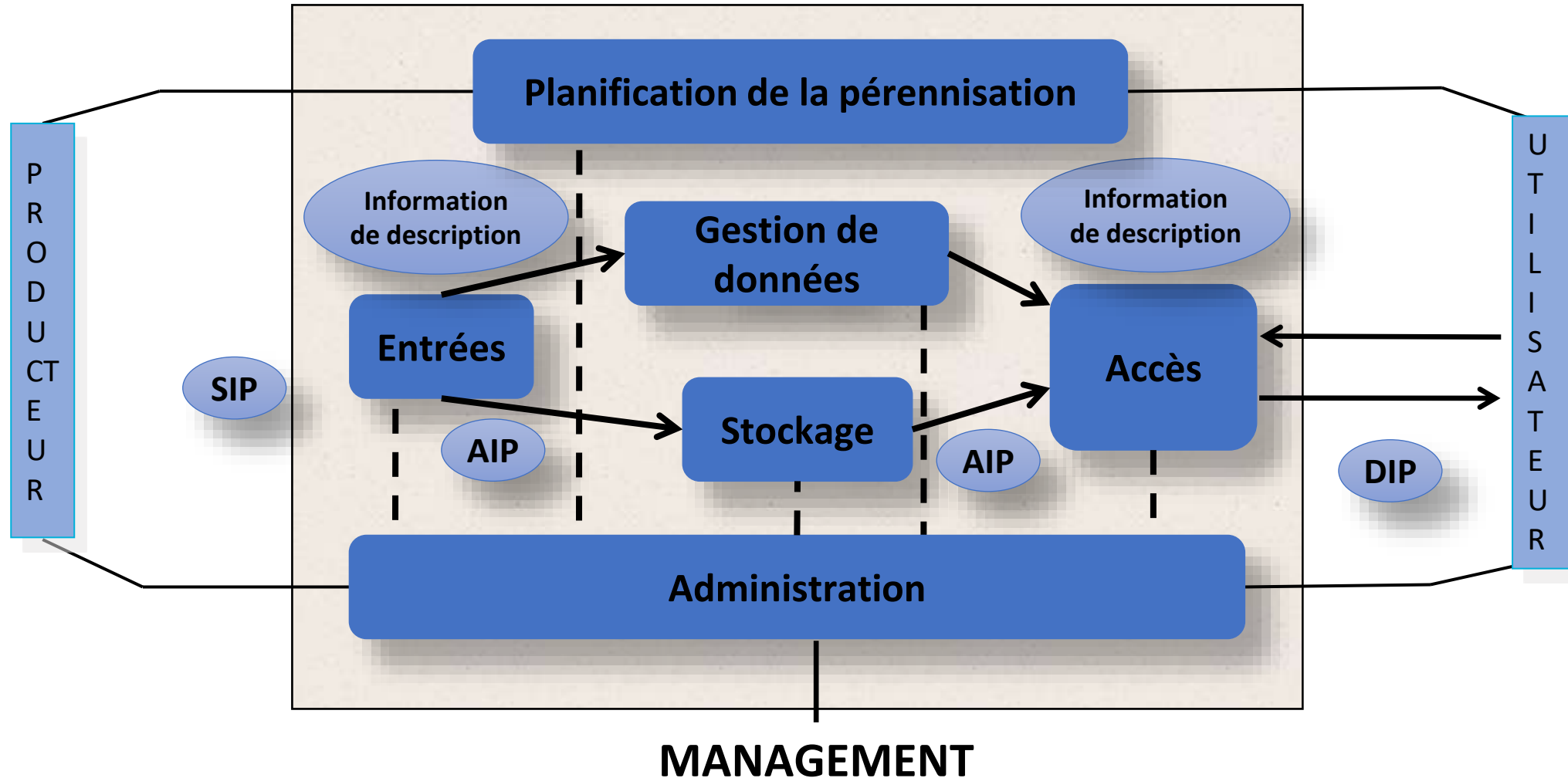
● Site en français

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

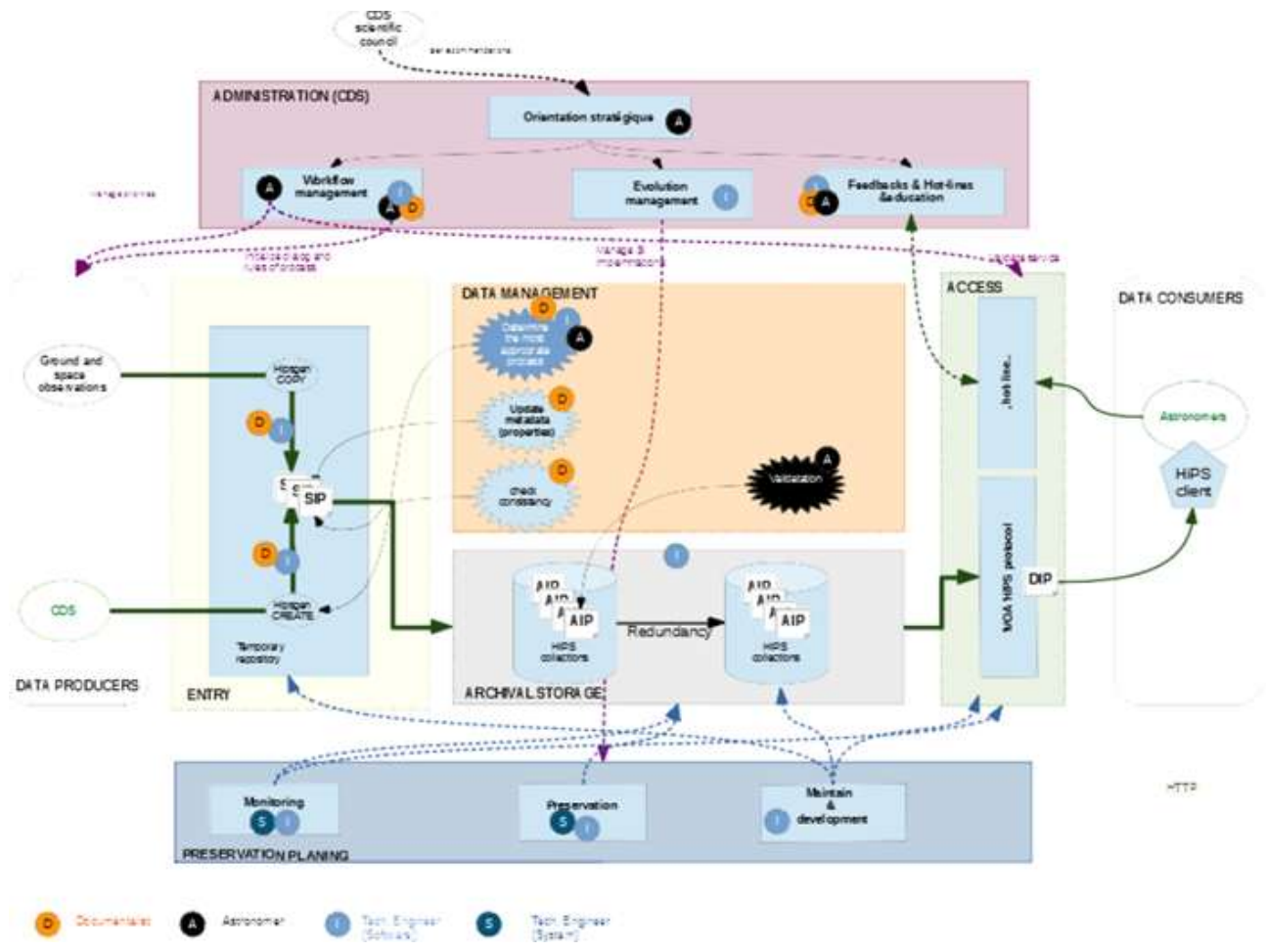
# L'environnement d'une archive OAIS



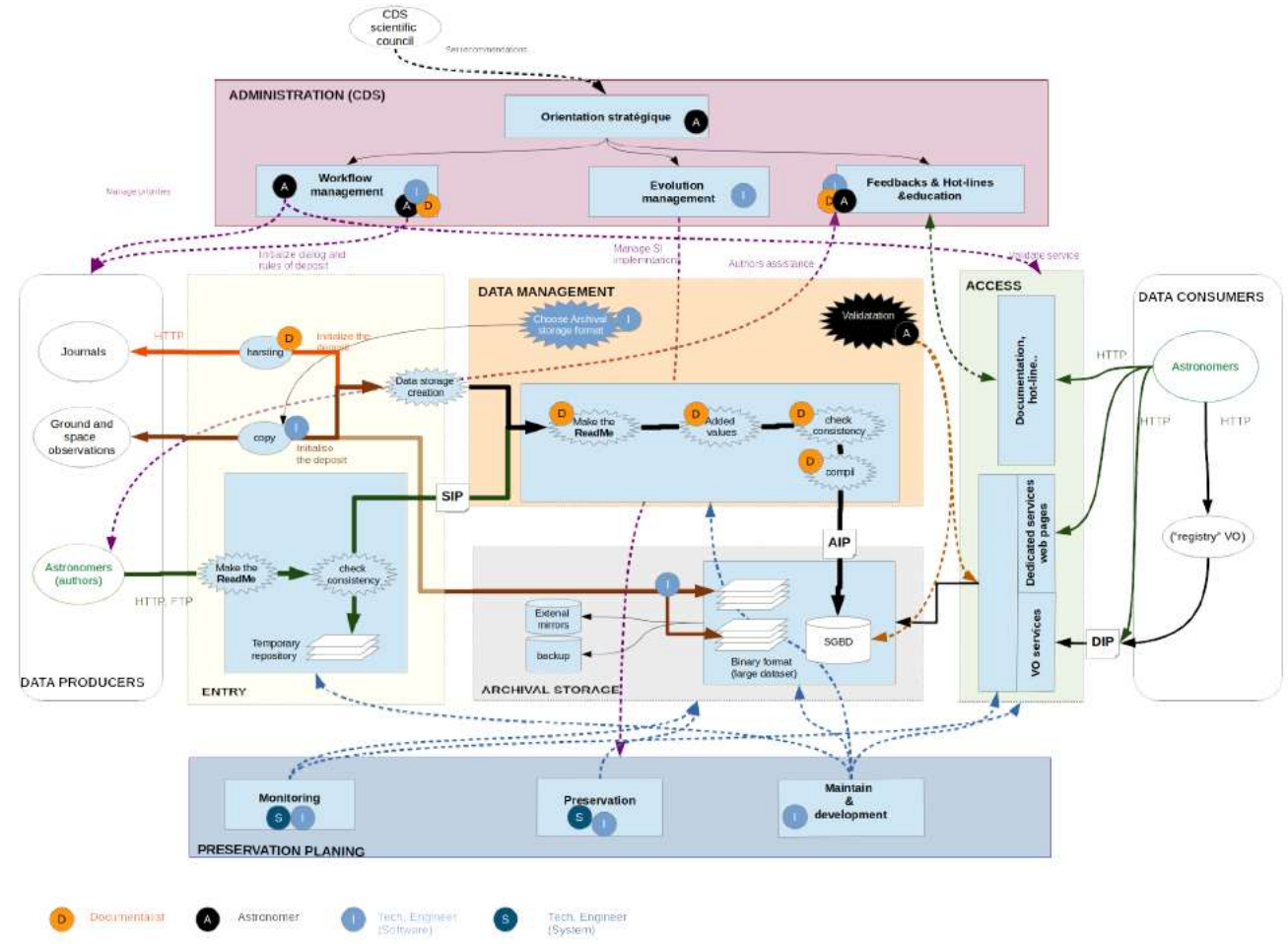
# Les entités fonctionnelles de l'OAIS



# Le pipeline de données pour Aladin/CDS



# Le pipeline de données pour VizieR/CDS



# La nouvelle version de l'OAIS

## Version 3 – publiée en Décembre 2024

- Version 1 Janvier 2002, version 2 Juin 2012

## Principaux changements

- Modèle conceptuel
  - Information de représentation,
  - Information de pérennisation,
  - Objectifs de préservation
- Modèle fonctionnel
  - Planification de la préservation: veille
- Perspectives de préservation
- Interopérabilité de l'archive





# Dataverse et CoreTrustSeal

# Le cas de Dataverse

## Guide Dataverse pour la certification CTS

- <https://dataverse.org/cts-guide>
- Publié en Mars 2021
- Basé sur la version des critères 2020-2022
- Combine
  - Éléments de réponse basé sur la technologie et le processus de l'outil
  - Exemples de réponses d'entrepôts Dataverse certifiés (Tilburg, QDR, dataverseNO, etc.)
- Certains points organisationnels ne sont pas détaillés
  - Mission
  - Continuité de service
- Pertinent pour la partie technique





# La suite documentaire Borealis

- Développée par Borealis et le [groupe d'experts sur la préservation de l'ARN du Canada](#)
- Soutien des partenaires de Borealis et les dépôts de données canadiens et internationaux avec la certification CoreTrustSeal et les efforts de préservation des données de recherche de manière plus générale
- Modèles de politique/documentation, lignes directrices, diagrammes pour soutenir la gestion des collections institutionnelles dans Borealis
- Comprend également un modèle de demande pour aider les institutions intéressées à demander la certification CoreTrustSeal
- [https://docs.scholarsportal.info/view/Main/SP/Borealis/Preservation/Projects/Borealis\\_CTS\\_Certification\\_Documentation\\_Suite\\_2024-2025/](https://docs.scholarsportal.info/view/Main/SP/Borealis/Preservation/Projects/Borealis_CTS_Certification_Documentation_Suite_2024-2025/)





### RDA Global

Contact - <https://www.rd-alliance.org/contact/>

Web - [www.rd-alliance.org](http://www.rd-alliance.org)

Twitter - @resdatall

LinkedIn - [www.linkedin.com/in/ResearchDataAlliance](http://www.linkedin.com/in/ResearchDataAlliance)

Slideshare -

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

### RDA FRANCE

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

Email - [contact-rdafrance@services.cnrs.fr](mailto:contact-rdafrance@services.cnrs.fr)