



Certification CTS pour les entrepôts de données et gestion des risques

**RIP Data : quelle sélection, conservation et
suppression des données de recherche ?**

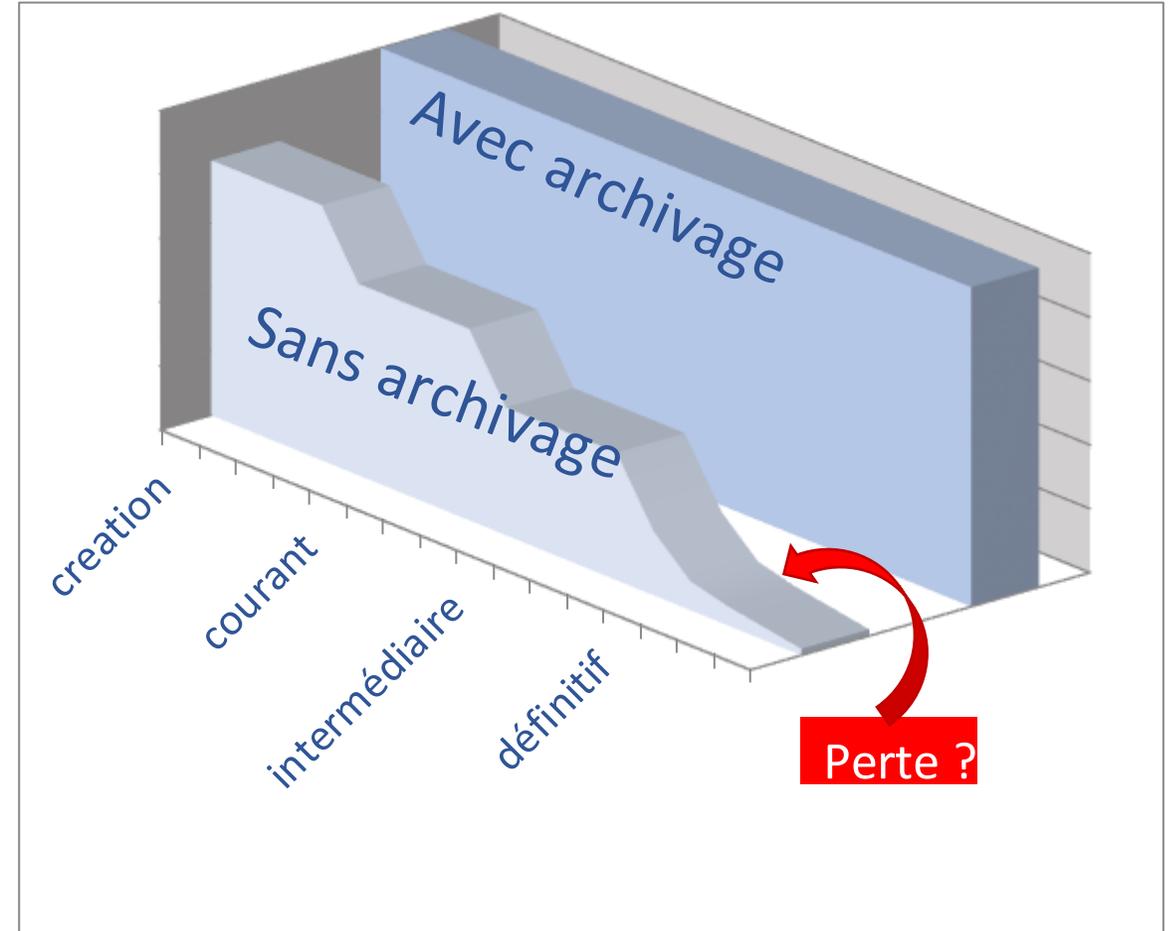
Aix-en-Provence, 01/10/2024

→ 01/10/2024

Les risques associés aux objets numériques

En informatique, 4 risques sont inéluctables si aucune mesure n'est prise :

- Connaissance perdue du contenu des fichiers
- Format de fichier inconnu
- Support physique détérioré
- Logiciel ou matériel de lecture disparu



Ce ne sont pas les seuls risques

L'opération d'un SAE est un ensemble de compromis

- Entre les impératifs immédiats et les besoins à long terme ;
- Entre les besoins et les moyens.

L'opération d'un SAE est une activité comme une autre, toute activité génère des risques ;

- La question n'est pas de supprimer les risques, mais de déterminer le niveau de risque acceptable ;
- La méthodologie de gestion de risques est ancienne dans le monde industriel, « à la mode » dans le monde du management et des services.

La gestion des risques est essentielle

- Elle fournit un référentiel sur lequel peut reposer l'élaboration de ces protocoles
- Elle fait en sorte que l'engagement de l'archive en termes de conservation et d'accès soit proportionné avec les risques qui pèsent sur les objets
- Elle définit les conditions d'une mission continue de veille et de surveillance de risques, définie dans le modèle OAIS comme « planification de la préservation ».

→ Plan de gestion des risques = outil de pilotage de la préservation des archives



La définition du risque



Un risque c'est....

- Un « danger éventuel, plus ou moins prévisible, inhérent à une situation ou à une activité »

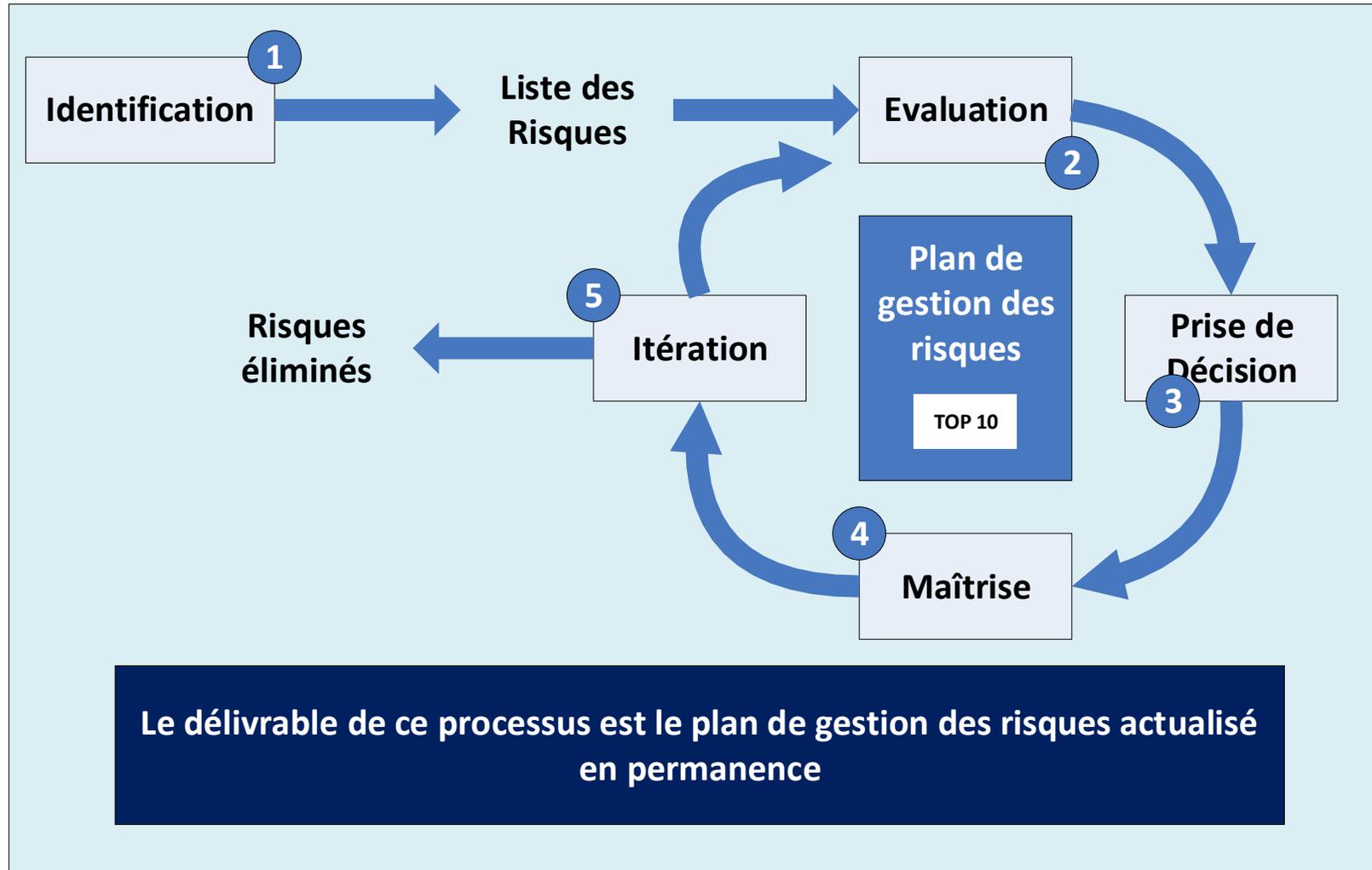
Le risque est défini par la combinaison de:

- Une vulnérabilité : un état de fait qui rend le risque possible
- Une menace : l'existence d'un potentiel d'exploitation de la vulnérabilité, avec une source et une action

Ex. : risque d'inondation par crue de la Seine

- Le risque n'existe que si on est vulnérable (la BnF est construite en zone inondable - vulnérabilité)
- Le risque n'existe que si la menace existe (des causes naturelles - source - peuvent provoquer la crue de la Seine - action)

La méthodologie de gestion des risques



L'évaluation et la prise de décision



Fixer un Indice de Priorité des Risques (IPR) à partir de

- Degré de probabilité,
- Impact

Effet \ Probabilité	Improbable	Rare	Occasionnel	Probable	Fréquent
Catastrophique	Red	Red	Red	Red	Red
Majeur	Orange	Orange	Orange	Orange	Red
Modéré	Light Green	Light Green	Light Green	Orange	Red
Mineur	Cyan	Cyan	Light Green	Light Green	Orange
Insignifiant	Cyan	Cyan	Cyan	Cyan	Light Green
	Faible	Modéré	Fort	Très Fort	

Les risques sont-ils acceptables ?

- Objectifs et activités de l'établissement
- Utilisateurs
- Coûts et bénéfices des opérations de maîtrise
- Obligation légale

Quelques exemples de risques

Risques naturels

- Catastrophes naturelles, etc.
- Maîtrise : plan d'urgence, duplication hors site.

Risques organisationnels

- Carence de ressources humaines compétentes, impossibilité de faire adhérer l'équipe au projet ;
- Maîtrise : visibilité du projet, formation, accompagnement au changement.

Risques financiers

- Coupes dans les budgets alloués au projet ;
- Maîtrise : plan pluri-annuel d'investissement.

Risques techniques

- Implémentation propriétaire du SAE ;
- Maîtrise : choix de normes, standards

L'assurance qualité



1. La qualité technique

- Qualité des métadonnées = capacité à garder la connaissance des contenus
- Qualité des formats de fichiers = capacité à convertir à de nouveaux formats
- Qualité du stockage = capacité à conserver le train de bits constituant les fichiers

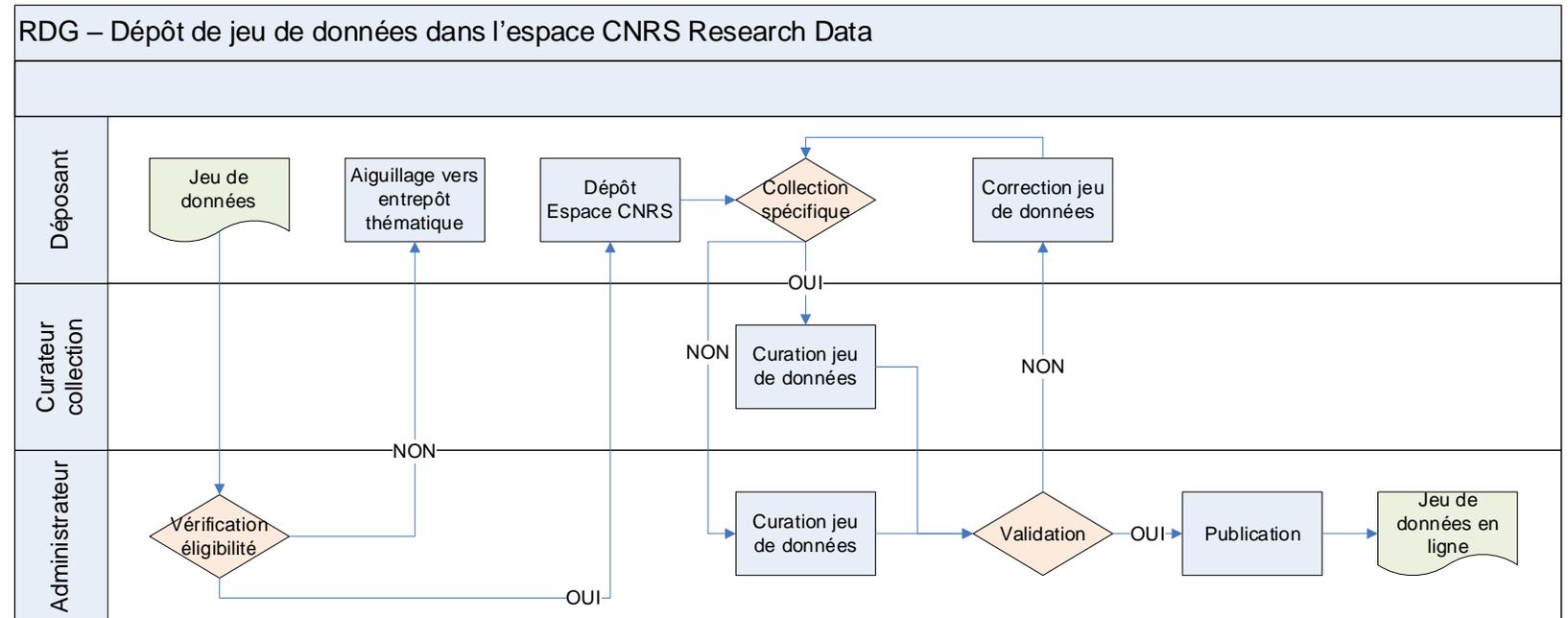
2. La qualité organisationnelle

- Documentation des processus métiers = répétabilité et amélioration des mécanismes de conservation
- Gestion des risques = maintien d'un niveau de qualité acceptable en identifiant de façon proactive les événements pouvant impacter la conservation et les plans d'actions à mettre en place
- Démarche de certification = validation des actions entreprises et constitue un levier pour l'obtention de budgets auprès des décideurs

L'adoption de normes/standards facilite la démarche qualité

La formalisation des processus métiers

- Décomposition des fonctions métier de l'archivage en processus, sous-processus, activités
- Mise au point d'une cartographie générale des processus
- Détail de chaque processus identifié – à rapprocher des groupes fonctionnels OAIS



La certification



« Processus de vérification qu'un produit ou service est conforme à un référentiel (norme, standard) de gestion de la qualité, de sécurité. »

Intérêts :

- Label de confiance & qualité pour les autres acteurs (clients, utilisateurs, management...);
- Valorisation de la qualité / fiabilité des prestations ;
- Utilisation de ces critères lors du développement et de la mise en fonctionnement de l'Archive.

Les référentiels de certification existants

- TRAC – Trustworthy Repositories Audit & Certification (OCLC, 2007)
- Levels of preservation matrix (NDSA, 2017)
- RAM – Rapid Assessment Method (DPC, 2019, 2024)
- Marque NF 461 : référentiel AFNOR base sur NF Z42013
- CTS : CoreTrust Seal
- Nestor : DIN 31644
- ISO 16363



Le modèle de certification Européen



Centres de données de confiance - Trustworthy Data Repositories

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/>

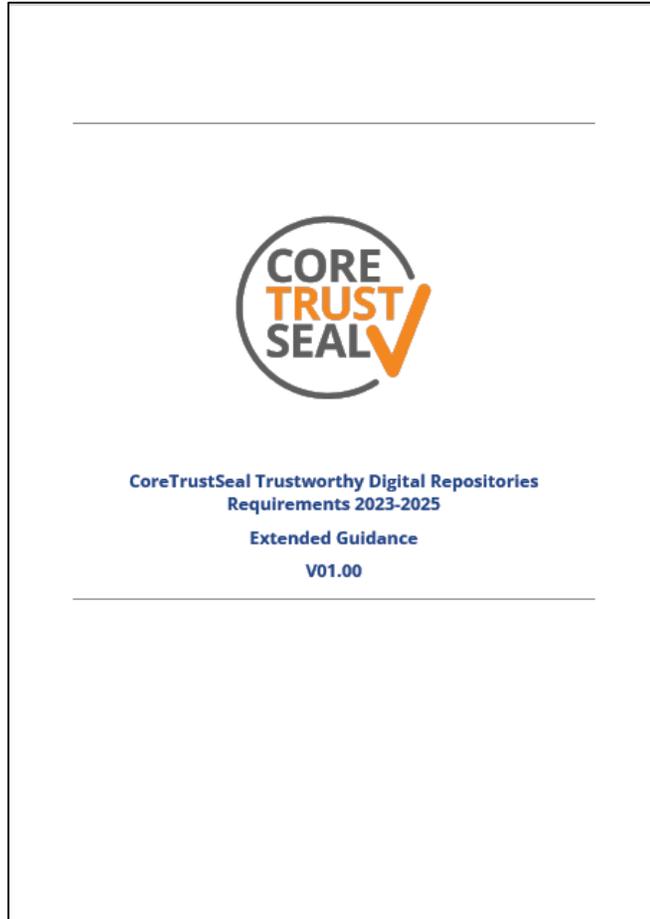
Document pour guider les évaluateurs et les candidats

- Extended Guidance 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 » 3000€



La certification CTS



R0 - Le contexte

16 critères, 3 thèmes:

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

Pourquoi la certification ?

Quelques semaines de travail d'équipe (tout compris)

Evaluation – amélioration des processus

- 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)



Plan national pour la Science Ouverte

ANR Flash 2019

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.

2021

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

Axe Structurer et Ouvrir les données de la recherche

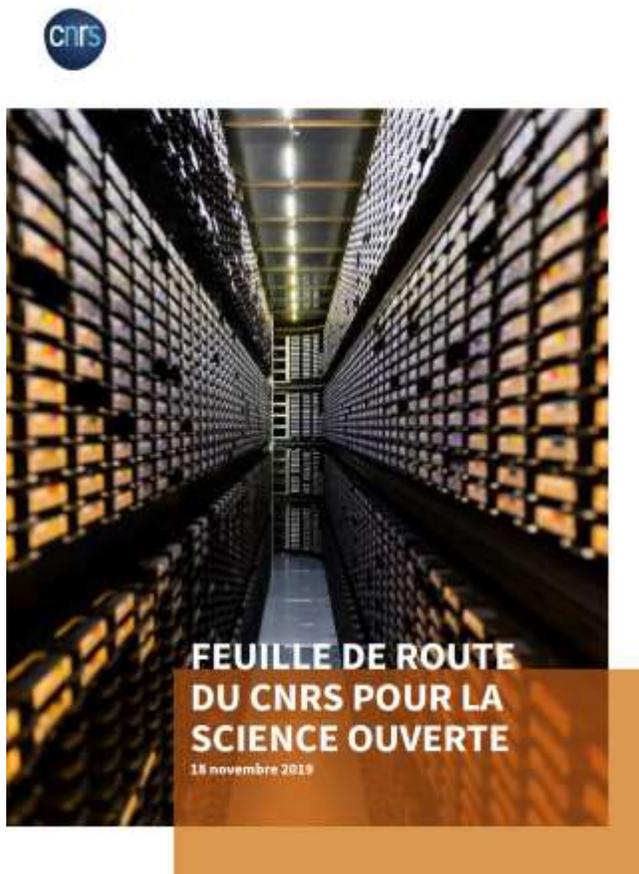
Structurer

- 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



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 et synthèse SOLEIL a également mis en route une politique de gestion des données. Ces 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 disciplinaires ». 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.

Au niveau des organismes: le CNRS



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*⁶). 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 Terra⁶, ou ceux de l'infrastructure européenne CLARIN⁷. Le CNRS pourra s'appuyer sur les activités de soutien à la certification mises en place par le Nœud National RDA France⁸.

Les entrepôts français actuellement certifiés CTS

[IDOC/IDOC-DATA](#), Orsay

[IFREMER SISMER](#), Plouzané

[ORTOLANG](#), Nancy

[ESRF](#), Grenoble

[CDSP](#), Paris

[IPSL](#), Paris

[CDS](#), Strasbourg

[RESIF](#), Grenoble



