



Entrepôts de données de confiance : critères de conformité

Le présent document est la traduction française de « Core Trustworthy Data Repositories Extended Guidance », version 1.1 datée de juin 2018, accessible sur le site de CoreTrustSeal :

<https://www.coretrustseal.org/wp-content/uploads/2017/01/20180629-CTS-Extended-Guidance-v1.1.pdf>

CoreTrustSeal est un organisme communautaire sans but lucratif qui promeut le développement d'infrastructures de données durables et fiables.

« CoreTrustSeal Extended Guidance » a été conçu pour faciliter le travail des évaluateurs de demandes de certification CoreTrustSeal (CTS) mais également pour fournir des informations et conseils aux entrepôts de données désireux d'entreprendre une démarche de certification CTS.

La présente traduction a été produite par le nœud français de l'Alliance pour les Données de Recherche (Research Data Alliance), financé dans le cadre du projet RDA-Europe 4.0. Cette traduction ne se substitue pas au document original en langue anglaise qui seul fait autorité dans le processus de certification CTS.

Pour plus d'informations sur CoreTrustSeal et le processus de certification CTS voir le site :

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

Pour plus d'informations sur la RDA, voir le site : <https://rd-alliance.org/>

Pour plus d'informations sur le nœud français de RDA, voir le site :

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

Contenu

INTRODUCTION	3
CONTEXTE ET ORIENTATIONS GENERALES	3
GLOSSAIRE DES TERMES	4
PRINCIPES GENERAUX	5
Introduction : généralités	5
Informations/justifications manquantes	5
Intelligibilité de la documentation	6
Documentation en langue autre que l'anglais	6
Confidentialité des documents internes.....	6
Structure et longueur du document	7
SPECIFICATIONS	8
R0. Généralités	8
Contexte	8
Infrastructure organisationnelle	12
I. Mission/périmètre	12
II. Licences.....	13
III. Continuité de l'accès.....	15
IV. Confidentialité/éthique	17
V. Infrastructure organisationnelle.....	19
VI. Conseils d'experts.....	20
Gestion des objets numériques	21
VII. Intégrité et authenticité des données	21
VIII. Appréciation et sélection des données	23
IX. Procédures d'archivage documentées	25
X. Plan de préservation.....	26
XI. Qualité des données.....	27
XII. Processus de traitement.....	28
XIII. Découverte et identification des données.....	30
XIV. Réutilisation des données	31
Technologie	32
XV. Infrastructure technique	32
XVI. Sécurité.....	33
Feedback du demandeur	34
Commentaires	34
ANNEXE: GLOSSAIRE	35

INTRODUCTION

Le présent document décrit les **critères de conformité à satisfaire pour des entrepôts de données dignes de confiance** pour la période 2017-2019. Les informations sur les critères sont précédées d'une introduction sur le contexte et les orientations générales du processus de certification; ces critères ont été établis par le Comité de CoreTrustSeal et demeureront inchangés pour la période 2017-2019. Le texte qui restera stable figure dans le présent document sous forme d'encadrés.

Le document contient également **des recommandations détaillées** pour les candidats à une certification CoreTrustSeal et les évaluateurs. Ces informations pourront être modifiées et mises à jour pendant la période pour laquelle le certificat a été défini, en fonction des besoins de la communauté. Les modifications seront soumises à l'approbation du Comité de CoreTrustSeal.

Ce document pointe également vers un glossaire des termes employés (en anglais).

CONTEXTE ET ORIENTATIONS GENERALES

Les **critères de base relatifs aux entrepôts de données dignes de confiance** visent à décrire les caractéristiques des entrepôts de données utilisables en toute confiance. À ce titre, tous les critères doivent obligatoirement être renseignés. Ce sont des éléments d'appréciation indépendants les uns des autres, ayant la même pondération. Bien qu'un certain recouvrement soit inévitable, la redondance des informations à fournir pour chaque critère a été réduite au minimum dans la mesure du possible. Les options offertes dans les listes à choix (par exemple, type de dépôt, niveau de curation) ne doivent pas être considérées comme exhaustives, et un espace supplémentaire est prévu dans tous les cas pour permettre au demandeur de rajouter d'autres options qui manqueraient. Ces éventuels ajouts, ainsi que tous les commentaires reçus des candidats, pourront être utilisés ultérieurement pour affiner ces listes à choix.

Pour chaque critère figurant dans ce document, des conseils de rédaction sont fournis afin d'aider les demandeurs à produire suffisamment d'éléments pour démontrer que leur entrepôt satisfait les critères attendus. Le texte précise le type d'information qu'un lecteur est en droit d'attendre pour effectuer une évaluation objective. De plus, le demandeur doit indiquer un **niveau de conformité pour chaque critère selon le barème suivant** :

- 0 - Sans objet
- 1 - L'entrepôt n'en a pas encore tenu compte
- 2 - L'entrepôt maîtrise le concept de façon théorique

3 - L'entrepôt est en phase d'implémentation

4 - La spécification a été entièrement implémentée

Ces niveaux de conformité constituent une partie utile du processus d'auto-évaluation, mais les dossiers seront jugés sur la base des justifications et preuves apportées et non en fonction des niveaux de conformité proposés dans l'auto-évaluation. À cet égard, si un demandeur pense qu'une fonctionnalité attendue n'est pas applicable à son entrepôt de données, il devra en fournir les raisons en détail. Notez également qu'une conformité de niveau 1 ou 2 peut être suffisante pour les auto-évaluations menées en interne, alors que la certification peut être accordée avec certains critères au niveau 3 - phase d'implémentation – en supposant que les services de l'entrepôt continueront à progresser, ce qui est un postulat de base de la certification.

Les réponses doivent être rédigées en anglais. On tentera de faire correspondre au mieux la langue et la discipline des évaluateurs et celles des dossiers soumis, mais ce ne sera pas toujours possible. Si les justifications produites le sont dans une autre langue que l'anglais, un résumé en anglais devra être fourni dans l'auto-évaluation.

Étant donné que la certification de base fournie par CoreTrustSeal n'implique pas une visite sur site, les réponses aux critères devront s'appuyer sur des preuves accessibles publiquement. Néanmoins, il est entendu qu'il peut y avoir des raisons impérieuses, par exemple des raisons de sécurité, qui font qu'il n'est pas toujours possible de faire figurer toutes les informations nécessaires sur un site web institutionnel ; aussi des dispositions sont prises dans le cadre du processus de certification pour les entrepôts de données qui souhaitent que des parties sensibles de leur dossier demeurent confidentielles.

Les entrepôts doivent être réévalués tous les trois ans. Même si les systèmes et les fonctions de base évoluent continuellement en fonction de la technologie et des besoins des utilisateurs, il se peut que les entrepôts ne subissent pas de changements majeurs au cours de cette période. Cependant, la norme ISO sur les entrepôts dignes de confiance (ISO 16363) a un cycle de révision de cinq ans, et une période plus courte est jugée nécessaire pour asseoir une norme de confiance de base en permettant d'éventuelles modifications et corrections. Une organisation ayant mis en place une bonne gestion de ses processus opérationnels devrait raisonnablement pouvoir soumettre une demande de certification avec seulement des révisions minimales au bout des trois ans, à moins que les critères de conformité eux-mêmes n'aient été mis à jour dans l'intervalle.

GLOSSAIRE DES TERMES

Veillez-vous référer au Glossaire des spécifications des entrepôts de données dignes de confiance en annexe. La version initiale en langue anglaise de ce glossaire est accessible à : <https://goo.gl/rQK5RN>.

PRINCIPES GENERAUX

Ces instructions visent à élaborer des normes et des procédures communes pour l'examen des auto-évaluations en vue de la certification CoreTrustSeal des entrepôts de données. Elles sont principalement destinées à guider les évaluateurs lorsqu'ils analysent les dossiers d'auto-évaluation, mais elles sont aussi mises à la disposition de tous en consultation, car elles peuvent être utiles aux candidats lors de la rédaction de leur dossier d'auto-évaluation.

▪ Introduction : généralités

L'effort à faire en vue d'une première demande de certification devra s'appuyer sur des informations en libre accès, mises à jour de façon régulière. En effet, la mise à jour du système d'information de l'organisme au moment de la rédaction de la demande de certification n'est pas le but recherché. La maintenance (continue) de l'information nécessaire pour les opérations de votre système devrait être suffisante pour maintenir la certification.

Pour ce qui est du niveau de conformité, les évaluateurs peuvent décider d'un niveau de conformité différent de celui choisi par le candidat pour un critère donné. Si les évaluateurs modifient le niveau sélectionné pour une valeur inférieure, la raison en sera expliquée au candidat et discutée avec lui. Dans ce cas, une amélioration pour cette fonctionnalité sera attendue lors du renouvellement de la certification.

Les instructions de rédaction demeurent toujours indicatives. Les évaluateurs vérifieront que chaque élément des instructions est bien abordé dans le texte de justification. Comme il n'est pas possible de couvrir tous les scénarios dans ces instructions générales, on s'attend à ce que chaque candidat complète sa réponse pour décrire les conditions et le contexte local. Les évaluateurs cherchent à obtenir des justifications claires et accessibles publiquement, spécifiques au candidat. Tous les éléments de tous les critères ne sont pas obligatoirement à renseigner. La décision finale dépend de la complétude et de la qualité de la réponse dans l'auto-évaluation pour un critère donné.

▪ Informations/justifications manquantes

S'il manque des informations à un point tel qu'il est impossible de juger si le critère spécifique est satisfait ou non, une demande d'explication sera envoyée au demandeur. Aucun niveau de conformité ne sera renseigné à ce stade par l'évaluateur.

Cela découle du paragraphe Contexte et orientations générales, où il est indiqué : « Les niveaux de conformité constituent une partie utile du processus d'auto-évaluation, mais les dossiers seront jugés sur la base des justifications et preuves apportées et non en fonction des niveaux de conformité proposés dans l'auto-évaluation ».

Cela signifie que les affirmations spécifiques devront être étayées par des liens vers des preuves accessibles dans la mesure du possible. En d'autres termes, la preuve est essentielle. Il est problématique et/ou non acceptable que l'information soit (presque) complètement manquante, insuffisamment ou non clairement décrite, ou fasse référence à un autre critère de conformité du document et renvoie vers lui. Tout écart par rapport à cette règle devra faire l'objet d'un commentaire explicite.

Cela signifie également qu'une éventuelle familiarité ou connaissance préalable de l'entrepôt par l'évaluateur n'auront pas d'influence dans l'évaluation des éléments de preuve fournis.

▪ **Intelligibilité de la documentation**

Les évaluateurs et les lecteurs du document final d'évaluation, qui est rendu public à la fois par l'entrepôt de données et sur le site Web CoreTrustSeal, devront être en mesure de comprendre l'ensemble de la démarche sans avoir à lire en détail les justificatifs fournis en appui. Pour les longs documents présentés comme éléments justificatifs, le demandeur mentionnera les parties spécifiques de ces documents qui fournissent les éléments de preuve en appui et citera (ou résumera selon le cas) l'information dans le dossier. Une description claire et cohérente de l'organisation dans son ensemble est généralement utile.

▪ **Documentation en langue autre que l'anglais**

Pour ce qui concerne les éléments de réponse associés à la demande écrits dans des langues autres que l'anglais, il sera demandé un complément d'information dans le document d'auto-évaluation. La documentation dans des langues autres que l'anglais est acceptable si son contenu est suffisamment et clairement expliqué dans un résumé en anglais. Ce résumé peut être assez bref pour certains types de documents (p. ex. une liste des formats préférés), mais devra être plus long pour d'autres (p. ex. un document décrivant la politique de préservation).

Ceci découle directement de : « Les réponses doivent être rédigées en anglais. On tentera de faire correspondre au mieux la langue et la discipline des évaluateurs et celles des dossiers soumis, mais ce ne sera pas toujours possible. Si un élément de réponse est dans une autre langue que l'anglais, un résumé en anglais devra être fourni dans l'auto-évaluation. »

▪ **Confidentialité des documents internes**

Une information confidentielle, économiquement sensible ou qui présente un risque pour la sécurité, ne peut être considérée comme le genre de justificatif public attendu pour satisfaire aux critères de conformité des entrepôts de données dignes de confiance. Ceci s'applique également aux documents qui ne sont disponibles que sur l'intranet du gestionnaire de l'entrepôt. Les demandeurs peuvent produire des documents justificatifs contenant de tels renseignements confidentiels en sus des autres éléments justificatifs, mais ils doivent être alors soumis en toute confidentialité aux évaluateurs ; ces

documents confidentiels doivent être cités et décrits dans le dossier de certification¹. Les demandeurs devraient s'efforcer de réviser la documentation produite en en supprimant les renseignements confidentiels afin qu'ils puissent être rendus publics et pris en considération lors du prochain examen.

Si la documentation est strictement à usage interne, ou si elle n'existe pas ou est en cours de rédaction, une date de mise à disposition publique devra être indiquée. Cette documentation sera jugée de la même manière que l'information publique. Aucune référence explicite à ces documents internes n'apparaîtra dans les commentaires des évaluateurs, qui ainsi accepteront cette situation sur la base du fait que la documentation sera rendue publique dans les délais promis. Toutes les délais annoncés seront mentionnés dans les commentaires généraux de l'évaluateur et seront vérifiés lors d'un renouvellement ultérieur. Si, au renouvellement, les informations ne sont ou ne peuvent pas être rendues publiques par le demandeur, celui-ci devrait être invité à publier au moins un résumé du document manquant.

▪ **Structure et longueur du document**

Le lectorat des demandes de certification CoreTrustSeal (c.-à-d. des documents d'auto-évaluations) est initialement constitué des évaluateurs et du Comité de CoreTrustSeal, mais les demandes acceptées sont rendues publiques. Les candidats à la certification doivent avoir à l'esprit l'ensemble du lectorat. Les dossiers soumis ne devraient pas être rédigés en mode questions-réponses. Les demandes doivent comprendre des textes rédigés pour chaque élément, en y incluant des informations pertinentes conformes aux indications données dans l'aide à la rédaction (texte encadré pour chaque critère) ainsi que dans les notes additionnelles.

Le Comité de CoreTrustSeal a pleinement conscience que les candidats à la certification CoreTrustSeal sont issus d'un large éventail d'organisations dont les missions, la taille et la complexité varient (en termes de structure organisationnelle et de type de données collectées), et qu'il peut donc y avoir des sujets pertinents et des éléments d'information associés qui ne sont pas couverts par ce guide. Il est aussi conscient du fait qu'expliquer la pertinence des justificatifs fournis puisse nécessiter un certain développement, surtout s'ils ne sont pas disponibles en anglais. Le Comité ne fixe pas de longueur minimale ou maximale pour les réponses, mais d'après son expérience, même les énoncés justificatifs les plus complexes se situent plutôt dans la partie basse d'une fourchette de 500 à 800 mots. Dans la mesure du possible, les énoncés justificatifs devraient être étayés par des liens publics vers la documentation utilisée pour décrire votre organisation et gérer vos objets numériques. C'est cette preuve publique qui offre le plus d'assurance qu'une organisation gère ses collections comme un entrepôt de données de confiance.

¹ Pour protéger l'anonymat des évaluateurs, les documents confidentiels doivent toujours être soumis en utilisant l'adresse électronique de CoreTrustSeal.

SPECIFICATIONS

R0. Généralités

Contexte

R0. Veuillez fournir les éléments de contexte de votre entrepôt.

- *Type d'entrepôt.* Sélectionnez tous les types pertinents :

- Entrepôt disciplinaire ou thématique
- Entrepôt institutionnel
- Système national d'archivage, y compris les archives gouvernementales
- Dépôt de publications
- Bibliothèque/Musée/Archives
- Entrepôt de projet de recherche
- Autre (précisez)

- *Brève description de l'entrepôt de données*

- *Brève description de la communauté cible concernée*

- *Niveau de curation.* Sélectionnez tous les types pertinents :

- A. Contenu mis en accès tel que déposé
- B. Curation de base - p. ex. vérification rapide, ajout de métadonnées de base ou de documentation
- C. Curation avancée – p. ex. conversion vers de nouveaux formats, amélioration de la qualité de la documentation
- D. Curation au niveau des données - comme dans C ci-dessus, mais avec une édition supplémentaire des données déposées pour améliorer l'exactitude

Commentaires

- *Partenaires sous-traitants.* S'il y a lieu, veuillez les énumérer.

- *Autres renseignements pertinents*

Réponse

➔ Conseils de rédaction :

Pour évaluer un entrepôt, les évaluateurs ont besoin d'informations permettant de replacer l'entrepôt dans son contexte. Veuillez sélectionner parmi les options proposées et renseignez de façon détaillée les rubriques qui apparaissent dans ce volet relatif au contexte.

1) **Type d'entrepôt.**

Cette rubrique aidera les évaluateurs à comprendre la ou les fonctions remplies par votre entrepôt. Choisissez-le ou les types d'entrepôt qui correspondent le mieux à votre cas (sélectionnez tous les types pertinents). Si aucune des catégories n'est appropriée, n'hésitez pas à fournir un autre type pour décrire votre fonction. Vous pouvez également fournir plus de détails pour aider l'examineur à comprendre le type de votre entrepôt.

2) **Communauté cible de l'entrepôt.**

Cette rubrique sera utile pour évaluer la façon dont l'entrepôt interagit et communique avec sa communauté d'utilisateurs. Veuillez-vous assurer que la réponse est suffisamment spécifique - par exemple, " chercheurs et enseignants en sciences sociales quantitatives ".

3) **Niveau de curation.**

Cette rubrique vise à déterminer si l'entrepôt distribue son contenu aux utilisateurs sans aucune modification, ou si l'entrepôt apporte une valeur ajoutée en améliorant le contenu d'une façon ou d'une autre. Tous les niveaux de curation supposent que les dépôts initiaux sont conservés inchangés et que les modifications ne sont effectuées que sur des copies de ces originaux. Les annotations/révisions doivent s'inscrire dans les termes de la licence convenue avec le producteur de données et faire clairement partie du champ de compétences des curateurs. Ainsi, on s'attendra à ce que l'entrepôt de données démontre que de telles annotations/révisions sont entreprises et documentées par des experts aux compétences appropriés et que l'intégrité de toutes les copies originales est maintenue. La connaissance du niveau de curation facilitera le travail des évaluateurs lors de l'examen des autres critères de certification. D'autres détails peuvent être ajoutés pour aider à comprendre les niveaux de curation que vous mettez en œuvre.

4) **Partenaires sous-traitants.**

Veuillez fournir une liste des partenaires auprès desquels votre organisation externalise des fonctions de l'entrepôt, en décrivant la nature de la relation (organisationnelle, contractuelle, etc.), et indiquer si le partenaire a déjà entrepris une évaluation en tant qu'entrepôt de données de confiance. De telles coopérations extérieures peuvent inclure, mais sans s'y limiter : tout service fourni par une institution dont vous faites partie, le stockage fourni par d'autres dans le cadre d'une sauvegarde par redondance à copies multiples, ou l'adhésion à des organisations qui peuvent reprendre la gestion de vos données lorsqu'un problème de continuité d'activité survient. En outre, veuillez énumérer les critères de la certification pour lesquelles le sous-traitant fournit tout ou partie de la fonctionnalité/service concerné, y compris par un contrat ou accord de service. Étant donné que l'externalisation est presque toujours partielle, vous devrez fournir des informations pour les fonctionnalités requises qui ne sont pas sous-traitées et pour les parties du cycle de vie des données restant sous votre contrôle. Disposer de qualifications et certifications - y compris, mais sans s'y limiter, les certifications DSA ou WDS - serait un plus pour les partenaires

sous-traitants. Cependant, il n'est pas obligatoire qu'ils soient certifiés. Nous comprenons qu'il peut être compliqué de définir et de décrire les périmètres et les relations, mais de tels détails sont essentiels pour assurer un processus d'examen complet.

5) **Autres informations pertinentes.**

L'entrepôt peut souhaiter ajouter des informations contextuelles supplémentaires qui ne sont pas couvertes par les critères, mais qui peuvent être utiles aux évaluateurs. Par exemple, vous pourriez décrire :

- L'utilisation et l'impact des collections de données de l'entrepôt (citations, utilisation par d'autres projets, etc.).
- Un rôle national, régional ou mondial que votre entrepôt assure.
- Tout regroupement international ou organisation en réseau auquel votre entrepôt appartient.

Conseils de rédaction additionnels R0

1) **Type d'entrepôt**

La sélection de plus d'un type devrait être étayée par une description de l'entrepôt de données expliquant comment ces multiples rôles sont remplis. Idéalement, une telle description devrait fournir un contexte supplémentaire en décrivant les types de données, les formats, les disciplines, etc. gérés par l'entrepôt.

2) **Brève description de la communauté cible de l'entrepôt de données**

Le glossaire contient une définition formelle de " communauté cible " (« designated community »). Cependant, la description suivante est également fournie pour aider à comprendre la signification de l'expression " communauté cible " dans ce contexte :

Un exposé clair de ce qu'est la communauté concernée montre que le demandeur comprend le périmètre, le niveau de connaissances et les méthodologies utilisées par les utilisateurs - y compris les logiciels et les formats préférés -. Dans l'idéal, la compréhension du niveau de curation nécessaire (contexte supplémentaire, formats préférés, etc.) pour satisfaire aux besoins de cette communauté doit apparaître au cours de la démonstration, tout comme la preuve que le demandeur reste à l'écoute et répond à l'évolution des besoins de celle-ci.

Les éléments de réponse fournis pour les critères doivent montrer les interactions avec la communauté cible et la réactivité aux besoins de celle-ci, en particulier pour ce qui touche à leur évolution. Un entrepôt de données ayant une communauté d'utilisateurs très spécifique et limitée pourrait préciser le niveau de connaissances minimal requis pour les utilisateurs (p. ex. degré/niveau de compréhension minimum en statistiques ou en génétique). Une communauté d'utilisateurs très large (par exemple, le grand public) impliquerait que l'entrepôt de données fournit une plus grande variété de documentation contextuelle pour faire en sorte que ses données soient comprises par tout le monde.

3) Niveau de curation

Il est possible de choisir plus d'une option (A, B, C ou D) pour préciser le niveau (ou l'étendue) de la curation effectuée, selon le type de données et en fonction de ce qui a été convenu avec les déposants.

Dans le cas d'un choix D, les évaluateurs attendent un haut niveau de qualité des données de curation : provenance validée, intégrité, gestion des versions (journalisation des modifications, etc.).

4) Partenaires sous-traitants

Si plus d'un partenaire est impliqué, un diagramme indiquant l'étendue du processus de sous-traitance sera utile pour faciliter la compréhension des évaluateurs. Avoir plus d'un sous-traitant (p. ex., un pour le stockage, un pour la maintenance des sites Web) est acceptable pourvu que toutes les relations soient clairement indiquées, de préférence dans un schéma. Les évaluateurs demanderont que cette section soit corrigée si des entités qui ne sont pas mentionnées ici apparaissent plus avant dans le document d'auto-évaluation.

5) Autres renseignements pertinents

Un entrepôt peut fournir ici des informations complémentaires : sa notice dans la base re3data (<http://www.re3data.org/>), son nombre d'employés, la taille de ses collections, le nombre moyen de téléchargements, son évolution dans le temps, son modèle économique ou de financement, etc.

INFRASTRUCTURE ORGANISATIONNELLE

I. Mission/périmètre

R1. L'entrepôt de données a pour mission explicite de donner accès à des données dans son domaine et de les préserver.

Niveau de conformité :

Réponse

🔄 Conseils de rédaction :

Les entrepôts de données assument la responsabilité de la gestion d'objets numériques et veillent à ce qu'ils soient conservés dans un environnement et pendant une période de temps appropriés. Les déposants et les utilisateurs doivent avoir pleinement conscience que la préservation et l'accès continu aux données est une mission explicite de l'entrepôt.

Pour cette rubrique, veuillez décrire :

- La mission de votre organisation en matière de préservation et d'accès aux données, et inclure des liens vers l'exposé explicite de cette mission.
- Le degré d'officialisation que votre mission a reçu au sein de votre organisation (p. ex. déclaration publique approuvée, missions explicitées à la demande des financeurs, décision de principe prise par le conseil d'administration ou le comité de pilotage).

Conseils de rédaction additionnels R1

Si la gestion des données n'est pas mentionnée dans l'énoncé de mission, alors, en règle générale, ce critère ne peut recevoir un niveau de conformité supérieur ou égal à 3.

II. Licences

R2. L'entrepôt maintient toutes les licences applicables à l'accès et à l'utilisation des données et en surveille la conformité.

Niveau de conformité :

Réponse

↳ Conseils de rédaction :

Les entrepôts doivent maintenir toutes les licences applicables pour l'accès et l'utilisation des données, en informer les utilisateurs et contrôler que ces derniers agissent conformément à celles-ci. Ce critère concerne les règles d'accès et les licences applicables fixées par l'entrepôt de données lui-même, ainsi que tout code de conduite généralement accepté dans le domaine concerné pour l'échange et l'utilisation des informations et des connaissances. Les évaluateurs rechercheront les preuves que l'entrepôt a mis en place des contrôles suffisants pour l'application des critères d'accès à ses collections de données, ainsi que des preuves que les licences ou processus pertinents sont correctement gérés.

Pour ce critère de conformité, veuillez décrire :

- les accords de licence actuellement appliqués.
- les conditions d'utilisation des données (diffusion, utilisation prévue, protection des données sensibles, etc.)
- la documentation sur les mesures prises en cas de non-respect des conditions d'accès et d'utilisation des données.

Il est à noter que si les collections de données sont en accès public et sans conditions particulières imposées aux utilisateurs - comme des exigences pour la citation des données ou la nécessité d'un accord pour rendre public les résultats d'une analyse secondaire des données - alors cette seule mention suffit.

Ce critère doit être mis en relation avec le critère R4 (Confidentialité/Ethique) dans la mesure où les dispositions en matière d'éthique et de protection de la vie privée ont un impact sur les licences. L'assurance que les licences applicables au dépôt fournissent suffisamment de droits pour que l'entrepôt puisse gérer les données, les préserver et leur y donner accès, est couverte par le critère R10 (Plan de préservation).

Conseils de rédaction additionnels R2

Les conditions d'accès et d'utilisation peuvent être fixées de plusieurs façons : soit sous la forme de conditions générales standards, soit sous la forme de conditions particulières pour certains déposants ou pour des ensembles de données particuliers. Ces conditions particulières peuvent concerner le niveau de curation, le niveau de responsabilité global, celui concernant spécifiquement les données, les restrictions d'utilisation, les conditions imposées à l'environnement de l'entrepôt (salle sécurisée, accès à distance sécurisé), et les restrictions d'usage applicables aux types d'utilisateurs (chercheur agréé, utilisateur ayant reçu une formation, etc.). Les licences Creative Commons (<https://creativecommons.org/>), y compris la licence CC 0 *Renonciation et transfert des données dans le domaine public*, peuvent être utilisées comme référence ici, mais des alternatives sont également possibles.

Bien qu'il puisse être difficile d'identifier les cas de non-conformité, il faut néanmoins tenir compte des conséquences si la non-conformité est détectée (p. ex. sanctions concernant l'accès ou l'utilisation présente ou future des données). Dans le cas d'une divulgation de données personnelles sensibles, il peut y avoir de graves conséquences pénales qui auraient un impact à la fois sur l'utilisateur et sur l'entrepôt de données. Idéalement, les entrepôts devraient afficher une politique en cas de non-conformité.

Le niveau de conformité minimal devrait être de 4, si l'entrepôt donne accès à des données personnelles.

III. Continuité de l'accès

R3. L'entrepôt dispose d'un plan de continuité pour assurer un accès continu à ses collections et leur préservation.

Niveau de conformité :

Réponse

➔ Conseils de rédaction :

Ce critère couvre les mesures prises pour assurer l'accès aux données et leur disponibilité, pour le présent et dans le futur. Les évaluateurs recherchent les preuves que les mesures permettant de faire face aux risques inhérents à l'évolution des circonstances externes sont en place.

Pour ce critère, veuillez décrire :

- Le niveau de responsabilité assuré à l'égard des collections de données, incluant notamment les durées garanties de conservation des données s'il y en a.
- Les plans à moyen terme (trois à cinq ans) et à long terme (> cinq ans) mis en place pour assurer la disponibilité et l'accessibilité continues aux données. En particulier, il faut décrire tant la réponse à des changements rapides de situation que la planification à long terme, en indiquant les options pour la relocalisation ou la migration du service vers une autre organisation, ou le retour des données à leurs propriétaires (c.-à-d. les producteurs de données). Par exemple, que se passerait-il en cas de cessation du financement, qui pourrait être un arrêt non prévu du financement, la fin planifiée du financement d'un entrepôt réalisé dans le cadre d'un projet à durée limitée, ou un changement d'orientation de l'organisme hébergeur?

Les éléments justificatifs pour ce critère R3 doivent porter davantage sur la gouvernance que sur les aspects techniques, qui sont à fournir pour les critères R10 (plan de préservation) et R14 (réutilisation des données), et devraient couvrir le cas où la mission décrite en R1 (Mission/Périmètre) est amenée à changer. Cette fonctionnalité R3 contraste avec R15 (infrastructure technique) et R16 (sécurité) en ce qu'elle couvre la continuité de l'ensemble des fonctions d'accès et de préservation des données.

Conseils de rédaction additionnels R3

L'évaluateur est en attente d'informations lui permettant de mesurer le niveau de responsabilité de l'entrepôt à l'égard des données. Par exemple, êtes-vous le dépositaire principal ou unique ? Le déposant est-il également responsable des données ? L'entrepôt promet-il de fournir l'accès, la préservation et/ou le stockage des données à un niveau de qualité minimum défini pendant une période de temps minimale ? Cette information aidera l'évaluateur à juger si l'entrepôt est pérenne en termes de ressources financières et de processus ; citons en particulier les informations sur la continuité des collections et des responsabilités associées, dans le cas d'une défaillance majeure de la

continuité d'exploitation. La responsabilité de la pérennité ne repose pas nécessairement sur l'entrepôt lui-même, mais peut relever plutôt d'une structure de tutelle ou de coordination. Si c'est le cas, cela devra être clairement indiqué. De plus, si l'entrepôt relève d'une telle tutelle ou structure de coordination, est-ce que cette tutelle, cette structure, ou toute autre institution (p. ex., les Archives Nationales) garantit assumer la responsabilité en cas de défaillance majeure ? S'il n'y a pas d'accord formalisé entre l'entrepôt et une telle organisation, le niveau de conformité ne peut excéder la valeur 3.

IV. Confidentialité/éthique

R4. L'entrepôt garantit, dans la mesure du possible, que les données sont créées, conservées, qu'on y accède et qu'elles sont utilisées dans le respect des normes disciplinaires et éthiques.

Niveau de conformité :

Réponse

➔ Conseils de rédaction :

Le respect des normes éthiques est crucial pour une science responsable. Le risque de divulgation - par exemple, le risque qu'une personne ayant participé à une enquête puisse être identifiée, ou que l'emplacement précis d'une espèce en voie de disparition puisse être localisé - est une préoccupation commune à de nombreux entrepôts. Les éléments attendus concernent non seulement les bonnes pratiques en matière de données présentant des risques de divulgation, mais aussi la nécessité de maintenir la confiance de ceux qui acceptent que leurs données personnelles/sensibles soient stockées dans l'entrepôt.

Pour satisfaire ce critère de conformité, les réponses devront inclure des éléments liés aux questions suivantes :

- Comment l'entrepôt se conforme-t-il aux normes disciplinaires applicables ?
- L'entrepôt demande-t-il confirmation que la collecte ou la création des données a été effectuée conformément à la législation et aux règles éthiques qui s'appliquent dans la zone géographique du producteur des données ou dans sa discipline ? (par exemple, avis d'un comité d'éthique ou d'un comité institutionnel d'évaluation, ou application d'une législation de protection des données).
- Des procédures spéciales sont-elles appliquées pour gérer les données dont la divulgation présente un risque ?
- Les données présentant un risque à la divulgation sont-elles stockées de manière appropriée pour en restreindre l'accès ?
- Les données présentant un risque à la divulgation sont-elles diffusées dans des conditions appropriées ?
- Des procédures sont-elles en place pour analyser les risques liés à la divulgation des données et prendre les mesures nécessaires en anonymisant les fichiers ou en ne les diffusant qu'à travers un accès sécurisé ?
- Le personnel est-il formé à la gestion des données dont la divulgation présente un risque ?
- Des mesures sont-elles en place si ces conditions ne sont pas respectées ?
- L'entrepôt fournit-il des conseils sur l'utilisation responsable de données protégées ou potentiellement protégées ?

Les éléments attestant de la conformité à ce critère devraient être en accord avec les dispositions décrites pour répondre aux critères R12 (Processus de traitement) et R2 (Licences).

Conseils de rédaction additionnels R4

Toutes les organisations ayant en charge des données ont un devoir éthique de les gérer dans le respect des pratiques scientifiques de leur communauté cible. Pour ce qui concerne les entrepôts contenant des données sur des personnes, des entreprises ou d'autres organisations, on s'attend en sus à ce que les droits des sujets concernés soient protégés, sur une base à la fois juridique et éthique.

La divulgation de ces données pourrait également présenter un risque de préjudice personnel, une violation de la confidentialité commerciale ou la divulgation de renseignements critiques (p. ex. l'emplacement d'espèces protégées, d'un site archéologique).

Le niveau minimum de conformité devrait être de 4 si l'entrepôt fournit au moment de la demande de certification l'accès à des données personnelles.

Les évaluateurs s'attendent à disposer d'éléments attestant que les candidats comprennent l'environnement légal et les pratiques éthiques applicables, et qu'ils ont des procédures documentées.

V. Infrastructure organisationnelle

R5. L'entrepôt dispose d'un financement adéquat et d'un nombre suffisant d'employés qualifiés gérés par un système de gouvernance clair pour mener à bien sa mission.

Niveau de conformité :

Réponse

➔ Conseils de rédaction :

Les entrepôts ont besoin d'un financement pour s'acquitter de leurs responsabilités, ainsi que d'un personnel compétent qui possède une expertise dans l'archivage des données. Toutefois, il est également entendu que la pérennité des financements est rarement garantie, ce qui doit être mis en balance avec la nécessité de stabilité.

Pour ce critère, les réponses doivent inclure des éléments probants les points suivants :

- L'entrepôt de données est hébergé par une institution reconnue (assurant la stabilité et viabilité à long terme) et pertinente pour sa communauté cible.
- L'entrepôt dispose d'un financement suffisant, y compris des ressources en personnel, des ressources informatiques et un budget permettant de se déplacer pour assister à des réunions lorsque cela est nécessaire, idéalement pour une période de trois à cinq ans.
- L'entrepôt fournit à son personnel un accès à la formation professionnelle continue.
- L'étendue et la profondeur de l'expertise, à la fois de l'organisation et de son personnel, incluant toute participation pertinente à des groupes nationaux ou internationaux, est appropriée à la mission.

La description complète des tâches réalisées par l'entrepôt de données - et celle des compétences nécessaires à leur exécution, peuvent être fournies, si elles sont disponibles. Ces descriptions ne sont toutefois pas obligatoires, car ce niveau de détail dépasse le périmètre de la certification de base.

Conseils de rédaction additionnels R5

La description de cette fonctionnalité de l'entrepôt devrait contenir des éléments attestant des processus décisionnels de gouvernance et de gestion de l'organisation et précisant les entités impliquées. Le personnel devrait avoir une formation appropriée à la gestion des données afin d'assurer des normes de qualité homogènes. Il est également important de connaître les proportions du personnel employé sur une base permanente et temporaire et comment cela pourrait affecter la qualité professionnelle de l'entrepôt, en particulier pour la préservation à long terme.

Dans quelle mesure le financement est-il structurel ou associé à des projets ? Cela peut-il être exprimé en nombre d'ETP ? Quelle est la fréquence de reconduction des financements ?

VI. Conseils d'experts

R6. L'entrepôt adopte un ou plusieurs mécanismes pour s'entourer de façon continue d'avis et commentaires d'experts (en interne ou en externe, incluant une expertise scientifique si le besoin en est pertinent).

Niveau de conformité :

Réponse

➔ Conseils de rédaction :

Un entrepôt de données performant s'efforce de s'adapter à l'évolution des types, des volumes et des débits de données, et d'adopter les nouvelles technologies les plus efficaces afin de conserver sa valeur pour sa communauté cible. Étant donné l'évolution rapide de l'environnement des données de recherche, il est recommandé qu'un entrepôt de données se procure des conseils et des retours d'utilisateurs experts sur une base régulière pour assurer la pertinence et l'amélioration continues des services offerts.

Pour ce critère, les réponses devraient fournir des éléments probants sur les points suivants :

- L'entrepôt dispose-t-il de conseillers en internes ou d'un comité consultatif externe auquel pourraient participer des experts techniques, des experts en sciences des données et des experts disciplinaires ?
- Comment l'entrepôt communique-t-il avec les experts pour obtenir des conseils ?
- Comment l'entrepôt communique-t-il avec sa communauté cible pour obtenir un retour ?

Ce critère cherche à confirmer que l'entrepôt a accès à des conseils objectifs d'experts, au-delà de ceux fournis par le personnel qualifié mentionné dans R5 (Infrastructure organisationnelle).

Conseils de rédaction additionnels R6

L'examineur attend des éléments montrant que l'entrepôt de données dispose d'un réseau d'expertise étendu lui prodiguant des avis et des conseils, tant pour ses activités au jour le jour que pour le suivi des nouveaux défis potentiels (veille scientifique et technologique). Il se peut qu'une partie de ces informations ait déjà été donnée sous "R0. Brève description de la communauté d'utilisateurs de l'entrepôt " et " Autres informations pertinentes ". Dans l'affirmative, il suffira d'y faire référence.

GESTION DES OBJETS NUMERIQUES

VII. Intégrité et authenticité des données

R7. L'entrepôt garantit l'intégrité et l'authenticité des données.

Niveau de conformité :

Réponse

➔ Conseils de rédaction :

L'entrepôt doit fournir la preuve qu'il exploite un système de gestion des données et des métadonnées approprié pour en assurer l'intégrité et l'authenticité lors des processus d'ingestion, de stockage, d'archivage et d'accès aux données.

L'intégrité garantit que les modifications apportées aux données et aux métadonnées sont documentées et tracées (raison d'être, auteur de la modification).

L'authenticité couvre le degré de fiabilité des données originales déposées et de leur provenance, y compris la relation entre données originales et données diffusées, et si les relations existant entre les jeux de données et/ou les métadonnées sont effectivement maintenues.

Pour ce critère, les réponses sur l'intégrité des données devront fournir des éléments probants relatifs aux points suivants :

- Description des contrôles visant à vérifier qu'un objet numérique n'a pas été altéré ou corrompu (p. ex., contrôles de fixité).
- Documentation de la complétude des données et métadonnées.
- Précisions sur la façon dont sont enregistrées toutes les modifications des données et métadonnées.
- Description de la stratégie de contrôle du versionage.
- Utilisation de normes et de conventions internationales appropriées (qui devront être spécifiées). La description de la gestion de l'authenticité devrait inclure les questions suivantes :
- L'entrepôt a-t-il une stratégie de modification des données ? Les producteurs de données sont-ils informés de cette stratégie ?
- L'entrepôt maintient-il les informations de provenance des données et effectue-t-il les vérifications associées?
- L'entrepôt maintient-il les liens vers les métadonnées et vers d'autres jeux de données ? Si oui, comment ?
- L'entrepôt compare-t-il les caractéristiques essentielles des différentes versions d'un même fichier ? Comment ?

- L'entrepôt vérifie-t-il l'identité des déposants ?

Ce critère couvre l'ensemble du cycle de vie des données dans l'entrepôt, et est donc lié aux étapes du processus de traitement décrites dans d'autres parties du document - par exemple, R8 (Appréciation et sélection des données) pour l'ingestion des données, R9 (Procédures de stockage) et R10 (Plan de préservation) pour l'archivage pérenne, et R12-R14 (Processus de traitement, Recherche et identification des données, Réutilisation des données) pour la diffusion. Au demeurant, la gestion de l'intégrité et de l'authenticité des données peut également être considérée comme un état d'esprit, engageant la responsabilité de chacun au sein de l'entrepôt de données.

Conseils de rédaction additionnels R7

Si une description claire et détaillée du contexte est importante pour tous les critères, c'est particulièrement le cas pour R7. L'organisation de la curation et les types de données guideront les attentes des évaluateurs. L'évaluateur tirera bénéfice d'une description générale claire des processus et des outils utilisés pour la curation des données, notamment la répartition entre traitements manuels et automatisés, et la façon dont ces processus, outils et pratiques sont documentés. Le demandeur pourra trouver utile de répondre séparément à chaque point des listes à puces et de traiter indépendamment l'intégrité et l'authenticité, tel que défini dans les conseils de rédaction de ce critère.

Les journaux des actions effectuées sur les données devront être décrits dans la réponse.

VIII. Appréciation et sélection des données

R8. L'entrepôt accepte des données et des métadonnées sur la base de critères définis garantissant pertinence et intelligibilité pour les utilisateurs.

Niveau de conformité :

Réponse

➔ Conseils de rédaction :

La fonction d'appréciation et de sélection (*appraisal*) est essentielle pour déterminer si les données répondent à tous les critères de versement dans les collections et pour établir une gestion appropriée pour leur préservation. Il faut veiller à ce que les données soient pertinentes et compréhensibles pour la communauté cible de l'entrepôt.

Pour ce critère, les éléments fournis devront répondre aux questions suivantes :

- L'entrepôt possède-t-il une politique de développement des collections pour orienter la sélection des données à archiver ?
- L'entrepôt a-t-il mis en place des contrôles de qualité pour s'assurer de la complétude et de l'intelligibilité des données déposées ? Si oui, veuillez fournir les références aux normes utilisés dans le processus de contrôle qualité et aux mécanismes de reporting généralement acceptés par la communauté cible ; y inclure le détail de la façon dont les problèmes sont résolus (par exemple, les données sont-elles retournées aux déposants, corrigées par l'entrepôt, marquées par un indicateur de qualité directement dans le fichier de données et/ou inclus dans les métadonnées associées ?)
- L'entrepôt a-t-il mis en place des procédures pour s'assurer que les métadonnées nécessaires à l'interprétation et l'utilisation des données sont fournies?
- Quelle solution l'entrepôt met-il en œuvre dans le cas où les métadonnées disponibles sont insuffisantes pour assurer la préservation à long terme ?
- L'entrepôt publie-t-il une liste des formats de données préconisés?
- Des contrôles de qualité sont-ils en place pour s'assurer que les producteurs de données utilisent les formats préconisés?
- Quelle est l'approche retenue à l'égard des déposants qui n'utilisent pas les formats préconisés?

Ce critère traite de l'assurance qualité du point de vue de l'interaction entre le déposant des données/métadonnées et l'entrepôt. Il se distingue de R11 (Qualité des données), qui traite de la qualité des données/métadonnées du point de vue des utilisateurs.

Conseils de rédaction additionnels R8

Le demandeur doit montrer qu'il a mis en place des procédures assurant que seules les données conformes à la politique de développement des collections sont acceptées. Le personnel de l'entrepôt doit avoir à disposition toute l'information et toutes les procédures utiles, et avoir les compétences nécessaires pour assurer aux utilisateurs une préservation à long terme des données et une utilisation conforme aux pratiques de la communauté cible.

IX. Procédures d'archivage documentées

R9. L'entrepôt applique des processus et des procédures documentés pour l'archivage des données.

Niveau de conformité :

Réponse

➔ Conseils de rédaction :

Les entrepôts doivent stocker les données et les métadonnées depuis la phase de dépôt, par le processus d'ingestion, jusqu'au service d'accès. Les entrepôts ayant une mission de conservation doivent également offrir un « service d'archivage » (« archival storage ») en termes de norme OAIS.

Pour ce critère, les réponses devront fournir des éléments probants relatifs aux questions suivantes :

- Comment les processus et procédures de stockage/archivage sont-ils documentés et gérés ?
- Quels sont les niveaux de sécurité imposés et comment sont-ils mis en œuvre ?
- Comment le stockage des données est-il traité dans la politique de préservation ?
- L'entrepôt dispose-t-il d'une stratégie de sauvegarde/copies multiples ? Si oui, quelle est-elle ?
- Des dispositions pour la récupération des données sont-elles opérationnelles ? Quelles sont-elles ?
- Des techniques de gestion des risques sont-elles utilisées pour éclairer la stratégie ?
- Quels sont les contrôles opérationnels qui permettent de s'assurer de la cohérence des copies archivées ?
- Comment la détérioration des supports de stockage est-elle traitée et suivie ?

Cette partie du document porte sur les dispositions générales en matière de continuité d'activité opérationnelle. Veuillez également consulter R15 (Infrastructure technique) et R16 (Sécurité) pour plus de détails sur les dispositions spécifiques concernant la sauvegarde, la sécurité physique et logique, la reprise en cas de panne et la continuité des opérations.

Conseils de rédaction additionnels R9

L'évaluateur cherchera à localiser chacun des lieux de stockage des données où des processus de curation sont exécutés, à comprendre comment les données sont gérées de façon appropriée dans chaque environnement, et à identifier que les processus de contrôle et de gestion des mises à jour de la documentation sont effectifs. L'entrepôt peut-il se remettre de catastrophes à court terme ? Les procédures de restauration sont-elles documentées et normalisées de telle sorte que différents gestionnaires de données, tout en exécutant les mêmes tâches séparément, parviendront essentiellement au même résultat ?

X. Plan de préservation

R10. L'entrepôt de données assume la responsabilité de la préservation à long terme et gère cette fonction de façon planifiée et documentée.

Niveau de conformité :

Réponse

➔ Conseils de rédaction :

L'entrepôt de données, les déposants et la communauté cible doivent comprendre le niveau de responsabilité engagé pour chaque élément déposé dans l'entrepôt. Ce dernier doit être en capacité légale d'assumer ces responsabilités. Les procédures doivent être documentées et leur exécution doit être assurée.

Pour ce critère, les réponses doivent inclure des éléments probants relatifs aux questions suivantes :

- Un plan de préservation est-il en place ?
- Le " niveau de préservation " de chaque élément est-il compris? Comment est-il défini ?
- Le contrat entre le déposant et l'entrepôt prévoit-il toutes les actions que ce dernier peut avoir à entreprendre pour faire face à ses obligations ?
- Le transfert de la responsabilité du stockage et de la conservation des données est-il clair pour le déposant et pour l'entrepôt?
- L'entrepôt a-t-il les droits nécessaires pour copier, transformer et stocker les données, ainsi que pour en fournir l'accès ?
- Les actions relatives à la préservation sont-elles documentées, notamment le transfert de la responsabilité de conservation, et les normes en termes d'information sur l'archivage ?
- Existe-t-il des mesures pour s'assurer que ces actions sont mises en œuvre ?

Conseils de rédaction additionnels R10

L'évaluateur s'attend à trouver une documentation claire et bien gérée permettant de garantir : (1) un approche organisée de la préservation à long terme, (2) l'accès pérenne aux données malgré les changements de format et (3) qu'une documentation suffisante est disponible pour assurer que les données sont utilisables par la communauté cible. La demande doit indiquer si l'entrepôt a défini différents niveaux de préservation à long terme et, dans l'affirmative, comment ceux-ci sont mis en œuvre. Le plan de conservation devra être géré de manière à assurer que les évolutions des technologies et des besoins des utilisateurs sont pris en compte de façon stable et en temps utile.

Si le demandeur ne mentionne pas de plan ou de politique de préservation, il ne peut dépasser le niveau de conformité 3 et il devra en avoir mis un ou une en place avant la prochaine certification.

XI. Qualité des données

R11. L'entrepôt dispose de l'expertise appropriée pour traiter la question de la qualité technique des données et métadonnées et pour assurer qu'une information suffisante est accessible aux utilisateurs pour leur permettre d'évaluer la qualité des données.

Niveau de conformité :

Réponse

🔄 Conseils de rédaction :

Les entrepôts de données doivent travailler de concert avec les déposants pour s'assurer que l'information disponible sur les données est suffisante pour que la communauté cible puisse évaluer la qualité de fond des données. Une telle évaluation de la qualité devient de plus en plus cruciale lorsque la communauté des utilisateurs est multidisciplinaire et que les chercheurs n'ont pas nécessairement l'expérience personnelle nécessaire pour évaluer la qualité à partir des seules données. Les entrepôts doivent également être en mesure d'évaluer la qualité technique des données déposées en termes de complétude et de qualité du matériau fourni, ainsi que la qualité des métadonnées associées.

Les données, ou les métadonnées associées, peuvent présenter des problèmes de qualité vis-à-vis de leur valeur pour la recherche, mais cela n'interdit pas un usage scientifique si l'utilisateur peut décider en toute connaissance de cause de leur pertinence à l'aide de la documentation fournie.

Pour ce critère, veuillez décrire :

- L'approche adoptée par l'entrepôt pour traiter de la qualité des données et des métadonnées.
- Les éventuels mécanismes automatisés d'évaluation de la conformité aux schémas de métadonnées retenus.
- La capacité de la communauté cible à commenter et/ou évaluer les données et les métadonnées.
- Si des citations de travaux associés ou des liens vers des index de citations sont fournis.

Les dispositions relatives à la qualité des données sont également abordées par d'autres critères. Plus précisément, veuillez-vous référer à R8 (Appréciation et sélection des données), R12 (Workflows) et R7 (Intégrité et authenticité des données).

Conseils de rédaction additionnels R11

Les demandeurs doivent indiquer clairement dans leur réponse qu'ils comprennent les niveaux de qualité que l'on peut raisonnablement attendre de la part des déposants. Ils doivent décrire comment la qualité sera assurée tout au long du processus de curation, et les attentes des utilisateurs en matière de qualité, ce qui peut impliquer la production d'une documentation spécifique là où les seuils de qualité attendus n'ont pas été atteints.

XII. Processus de traitement

R12. L'archivage s'effectue selon des processus de traitement (*workflows*) précisément définis, de l'ingestion des données à leur diffusion.

Niveau de conformité :

Réponse

➔ Conseils de rédaction :

Afin d'assurer la cohérence des pratiques utilisées pour la prise en compte des données et les services et éviter des activités au cas par cas et sans réflexion préalable, les processus d'archivage devraient être documentés et des dispositions pour gérer les évolutions de ces processus de façon contrôlée devraient être mises en place. Les procédures devraient être adaptées à la mission et aux activités de l'entrepôt, et la documentation des procédures pour l'archivage des données devrait être claire.

Pour ce critère, les réponses devront inclure des informations probantes relatives aux éléments suivants :

- Description des traitements et des processus opérationnels
- Information claire donnée aux déposants et aux utilisateurs sur la façon dont les données sont traitées.
- Niveaux de sécurité et leur impact sur les processus de traitement (protection des données personnelles, etc.).
- Contrôle qualitatif et quantitatif des résultats des traitements.
- Appréciation et sélection des données.
- Politique à l'égard des données qui n'entrent pas dans le périmètre de la mission de l'entrepôt ou ne correspondent pas au profil des collections.
- Les types de données gérées et leur impact sur les processus de traitement.
- Mécanismes de prise de décision au sein des processus de traitement (p. ex. reformatage des données pour archivage).
- Gestion des évolutions des processus de traitement.

Ce critère permet de confirmer que tous les processus de traitements sont documentés. Des éléments descriptifs partiels de ces workflows peuvent avoir été fournis dans d'autres critères, comme le processus d'ingestion dans R8 (Appréciation et sélection), les processus de stockage dans R9 (Procédures de stockage documentées), les dispositions de sécurité dans R16 (Sécurité) et la confidentialité dans R4 (Confidentialité/éthique).

Conseils de rédaction additionnels R12

L'évaluateur s'attend à trouver des informations montrant que le demandeur a adopté une approche cohérente, rigoureuse et documentée pour gérer ses activités tout au long des divers processus de traitement, et que les modifications apportées à ces processus sont mis en œuvre, évaluées, consignées et gérées de façon appropriée.

XIII. Découverte et identification des données

R13. L'entrepôt permet aux utilisateurs de découvrir les données et d'y faire référence de façon pérenne par un mécanisme de citation approprié.

Niveau de conformité :

Réponse

➔ Conseils de rédaction :

L'accès effectif aux données est un élément crucial du partage des données, et la plupart des entrepôts fournissent des catalogues interrogeables qui décrivent leurs fonds, de sorte que les utilisateurs potentiels peuvent évaluer les données pour voir si elles répondent à leurs besoins. Une fois repérés, les jeux de données doivent pouvoir être citables par une référence complète aux données, incluant des identifiants pérennes afin d'assurer une accessibilité sur la durée. Les citations permettent également de donner crédit aux personnes qui ont contribué à la création du jeu de données et de le leur attribuer correctement.

Pour ce critère, les réponses devront inclure des éléments probants sur les questions suivantes :

- L'entrepôt offre-t-il des fonctionnalités de recherche ?
- L'entrepôt offre-t-il un catalogue de métadonnées interrogeable selon les normes (internationales) appropriées ?
- L'entrepôt permet-il le moissonnage des métadonnées ?
- L'entrepôt est-il signalé dans un ou plusieurs registres de ressources disciplinaires ou génériques ?
- L'entrepôt offre-t-il des recommandations pour la citation des données ?
- L'entrepôt utilise-t-il des identifiants pérennes ?

Conseils de rédaction additionnels R13

La réponse démontre que le processus de curation des données et des métadonnées est conçu pour permettre la découverte d'objets numériques clairement définis et identifiés. La manière dont les données doivent être citées doit être claire pour les utilisateurs de façon à permettre de créditer les auteurs de façon appropriée dans le contexte académique et de lier entre elles des recherches apparentées.

XIV. Réutilisation des données

R14. L'entrepôt permet de réutiliser les données sur la durée, en veillant à ce que des métadonnées appropriées soient disponibles pour permettre la compréhension et l'utilisation des données.

Niveau de conformité :

Réponse

➔ Conseils de rédaction :

Les entrepôts doivent assurer que les données puissent être comprises et utilisées efficacement dans le futur, malgré des évolutions technologiques. Ce critère évalue les mesures prises pour assurer que les données sont réutilisables.

Pour ce critère, les réponses doivent inclure des informations probantes sur les points suivants :

- Quelles métadonnées sont exigées par l'entrepôt au moment du dépôt (p. ex. métadonnées Dublin Core ou métadonnées orientées contenu) ?
- Les données sont-elles présentées dans des formats utilisés par la communauté cible ? De quels formats s'agit-il ?
- Des mesures spécifiques sont-elles prises pour tenir compte de l'évolution possible des formats ?
- Des plans relatifs aux migrations futures sont-ils en place ?
- Comment l'entrepôt assure-t-il l'intelligibilité des données ?

Le concept de " réutilisation " est crucial dans les situations où les résultats d'une analyse ultérieure des données sont eux-mêmes redéposés dans un réservoir à côté des données primaires, entraînant alors des filiations de données et des problèmes de droits associés d'une grande complexité.

La réutilisation des données dépend des licences applicables couvertes par le critère R2 (Licences).

Conseils de rédaction additionnels R14

Le candidat doit comprendre les besoins de la communauté cible en termes de pratiques de recherche, d'environnement technique et de normes applicables. Les évolutions technologiques sont un point important, mais la disponibilité de métadonnées appropriées de bonne qualité joue également un rôle essentiel et doit être mentionnée dans l'information fournie. Cette dernière information est essentielle pour concevoir des processus de curation aboutissant à des objets numériques répondant aux besoins de l'utilisateur final, et conformes aux normes génériques ou disciplinaires.

TECHNOLOGIE

XV. Infrastructure technique

R15. L'entrepôt fonctionne avec des systèmes d'exploitation et d'autres systèmes de base bien maintenus et utilise des technologies matérielles et logicielles adaptées aux services qu'il fournit à sa communauté cible.

Niveau de conformité :

Réponse

🔄 Conseils de rédaction :

Les entrepôts doivent fonctionner sur des infrastructures de base fiables et stables qui maximisent la disponibilité du service. De plus, le matériel et les logiciels utilisés doivent être pertinents et appropriés pour leur communauté cible et adaptés aux fonctions attendues d'un entrepôt. Certaines normes telles que le modèle de référence OAIS décrivent les fonctions d'un entrepôt répondant aux besoins des utilisateurs.

Pour ce critère, les réponses doivent inclure des éléments probants sur les points suivants :

- Quelles normes l'entrepôt utilise-t-il comme référence ? S'agit-il de normes internationales et/ou de normes de communautés (p. ex., normes de l'Infrastructure de Données Spatiales (IDS), OGC, W3C, ou ISO 19115) ? À quelle fréquence sont-elles révisées ?
- Comment les normes sont-elles mises en œuvre ? Existe-t-il des déviations importantes par rapport à l'implémentation standard ? Si c'est le cas, veuillez préciser.
- L'entrepôt a-t-il un plan de développement de son infrastructure ? Si oui, quel est-il ?
- Un inventaire des logiciels utilisés est-il tenu à jour et une documentation du système est-elle disponible ?
- Des logiciels maintenus par des communautés sont-ils utilisés ? Si oui, veuillez préciser.
- Concernant les flux de données en temps réel ou semi temps réel, la disponibilité de la bande passante (en connectivité 24/24) des réseaux publics et privés est-elle suffisante pour satisfaire aux exigences du périmètre mondial et/ou régional de l'entrepôt ?

Conseils de rédaction additionnels R15

Les traitements de données (et les personnels qui opèrent les services de l'entrepôt) s'appuient sur une infrastructure technique. L'évaluateur attend des éléments montrant que l'entrepôt candidat maîtrise l'écosystème des normes, des outils et des technologies disponibles pour la gestion et la curation des données (de recherche) et qu'il a choisi des options qui satisfont les exigences locales. Si possible, cela devra être démontré en s'appuyant sur un modèle de référence.

R16. L'infrastructure technique de l'entrepôt assure la protection de l'installation et de ses données, de ses produits, de ses services et de ses utilisateurs.

Niveau de conformité :

Réponse

➔ Conseils de rédaction :

L'entrepôt devrait analyser les menaces potentielles, évaluer les risques et créer un système de sécurité cohérent. Il devrait décrire des scénarios de dégradation suite à des actions malveillantes, à des erreurs humaines ou à des défaillances techniques qui constituent une menace pour l'entrepôt et ses données, ses produits, ses services et ses utilisateurs. Il devrait mesurer la probabilité et l'impact de tels scénarios, décider quels sont les niveaux de risque acceptables et déterminer quelles mesures devraient être prises pour contrer les menaces qui pèsent sur l'entrepôt et sa communauté cible.

La sécurité est un processus continu.

Pour ce critère, veuillez décrire :

- Les procédures et dispositions en place pour assurer en cas de panne un rétablissement rapide des services ou une solution de backup des services essentiels.
- Votre système de sécurité informatique, votre plan de reprise sur sinistre et votre plan de continuité d'activités ; les employés impliqués dans la sécurité (p. ex. responsable de sécurité) et tout outil d'analyse des risques que vous utilisez (p. ex., DRAMBORA).

Ce critère décrit certains des aspects couverts sur un plan général par d'autres critères - par exemple, R12 (Processus de traitement) - et est complémentaire de R9 (Procédures de stockage documentées).

Conseils de rédaction additionnels R16

L'évaluateur attend des éléments justifiant que le demandeur comprend les risques techniques applicables à l'entrepôt pour les utilisateurs des données et l'environnement physique, et qu'il a mis en place des mécanismes pour répondre aux incidents de sécurité. Les informations fournies doivent se concentrer sur l'infrastructure technique plutôt que sur les aspects de gestion et de procédure de la continuité d'activité. Comment l'infrastructure technique est-elle contrôlée par l'entrepôt ou par l'institution hôte/sous-traitante ? Qui en est responsable ? L'entrepôt peut-il spécifier d'une manière ou d'une autre l'infrastructure technique si celle-ci est externalisée ? Les dispositions sont-elles suffisantes pour garantir la préservation à long terme des collections de données et/ou l'accès à long terme à celles-ci ?

Feedback du demandeur

Commentaires

Ces critères ne sont pas considérés comme définitifs, et nous apprécierions votre contribution à l'amélioration de cette procédure de certification. Pour cela, veuillez nous laisser vos commentaires sur la qualité de ce document d'évaluation et sur sa pertinence pour votre organisation, ainsi que toute autre réflexion additionnelle.

Réponse

Annexe : Glossaire

Note du traducteur :

Pour chaque définition, le terme original anglais est conservé. Lorsque celui-ci comporte une astérisque, c'est la définition de l'OAIS qui a été retenue.

Le texte anglais de référence de l'OAIS utilisé est :

Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS). Recommended Practice -- CCSDS 650.0-M-2. Magenta Book, June 2012.

<http://public.ccsds.org/publications/archive/650x0m2.pdf>

Le texte français de référence de l'OAIS utilisé est : Modèle de référence pour un système ouvert d'archivage d'information (OAIS) : Recommandations de pratiques pour les systèmes de données spatiales

[https://public.ccsds.org/Pubs/650x0m\(F\).pdf](https://public.ccsds.org/Pubs/650x0m(F).pdf)

Information de droits d'accès (Access Rights Information*) : information qui identifie les restrictions d'accès portant sur l'Information de contenu, en particulier : cadre réglementaire, licences et contrôle d'accès. Elle comprend les conditions d'accès et de diffusion établies dans le Protocole de versement, concernant à la fois la pérennisation (par l'OAIS) et l'utilisation (par l'Utilisateur). Elle inclut également les spécifications pour la mise en œuvre des mesures d'application des droits.

Archive (Archive*) : organisation chargée de pérenniser l'information pour permettre à une communauté cible d'y accéder et de l'utiliser.

Authenticité (Authenticity*) : degré de confiance qu'une personne (ou un système) peut attribuer à un objet au regard de ce qu'il prétend être. Le degré d'Authenticité est évalué à partir d'éléments probants.

Utilisateur (Consumer*) : toute personne ou système client en relation avec les services de l'OAIS pour trouver des informations pérennisées présentant un intérêt, et pour accéder au détail de ces informations. Il peut s'agir aussi d'autres OAIS, ou de personnes ou systèmes internes à l'OAIS

Données (Data*) : une représentation réinterprétable formalisée de l'information, adaptée à la communication, à l'interprétation ou au traitement.

Exemple : une séquence de bits, un tableau de nombres, les caractères d'une page, un enregistrement de paroles ou un échantillon de roche lunaire.

Communauté cible (Designated Community*) : groupe identifié d'utilisateurs potentiels, susceptibles de comprendre un ensemble donné d'informations. La Communauté cible peut être constituée de plusieurs communautés d'utilisateurs. La Communauté cible est définie par l'Archive et sa définition peut évoluer au cours du temps.

Migration numérique (Digital Migration)* : transfert d'informations numériques au sein de l'OAIS dans le but de les pérenniser. La Migration numérique se distingue des autres types de transfert sur trois points :

- on vise la pérennisation de l'intégralité du contenu qui doit être pérennisé,
- on présuppose que l'information résultant de ce transfert est destinée à remplacer l'ancienne,
- on admet que l'OAIS dispose de la pleine maîtrise et responsabilité de tous les aspects du transfert.

Objet numérique (Digital Object)* : objet constitué de séquences de bits.

Entrepôt disciplinaire ou thématique (*Domain or subject-based repository*): spécialisé dans un champ de recherche ou un type de données.

Ingestion (*Ingest*) : Processus d'alimentation d'un entrepôt par des données et métadonnées associées.

Intégrité (*Integrity*) : Cohérence interne ou absence de corruption des objets numériques. L'intégrité peut être compromise par des erreurs matérielles même lorsque les objets numériques ne sont pas touchés, ou par des erreurs logicielles ou humaines lorsque les objets numériques sont transférés ou traités.

Entrepôt institutionnel (*Institutional repository*): Entrepôt générique, multidisciplinaire au service d'un organisme de recherche.

Long terme (*Long Term) :** période de temps suffisamment longue pour qu'il soit nécessaire de prendre en compte les changements technologiques qui ont un impact sur l'information détenue par un OAIS, et notamment la gestion des nouveaux supports et formats de données ainsi que l'évolution de la Communauté cible. Cette période n'est pas limitée dans le temps.

Pérennisation (*Long Term Preservation) :** fait de maintenir sur le Long terme l'information immédiatement compréhensible par la Communauté cible, avec les preuves de son authenticité.

Système national de dépôt, incluant le niveau gouvernemental (*National repository system, including governmental*): Infrastructure nationale, multidisciplinaire. Possède une mission de dépôt légal pour certains types de données (publiques ou gouvernementales).

Système ouvert d'archivage d'information (*Open Archival Information System – OAIS) :**

Archive, c'est-à-dire une organisation (qui peut faire partie d'une organisation plus large) constituée d'une équipe et de systèmes dont la responsabilité est de pérenniser des informations et de les rendre accessibles à une Communauté cible. L'OAIS assume un ensemble de responsabilités, définies à la section 4, qui permettent de distinguer une Archive OAIS des autres acceptions du mot « archive ». Le terme « ouvert » dans OAIS signifie que cette recommandation et les recommandations et normes ultérieures liées à l'OAIS sont élaborées de façon ouverte et collaborative, et en aucun cas que l'accès à l'Archive est libre.

Formats préférés (*Preferred Formats*) : Formats dont un entrepôt peut raisonnablement garantir qu'ils resteront lisibles et utilisables. Il s'agit généralement des normes *de facto* utilisées par une discipline particulière.

Producteur (*Producer) :** toute personne ou système client qui fournit des informations à pérenniser. Il peut s'agir d'autres OAIS, ou de personnes ou systèmes internes à l'OAIS.

Information de provenance (*Provenance Information) :** information qui documente l'historique de l'Information de contenu. Cette information renseigne sur l'origine ou la source de l'Information de contenu, sur toute modification intervenue depuis sa création et sur ceux qui en ont eu la responsabilité. L'Archive est responsable de la création et de la pérennisation de l'Information de provenance dès le moment du versement ; cependant, le Producteur est censé fournir l'Information de Provenance antérieure au versement. L'Information de provenance est un élément de preuve qui renforce l'authenticité.

Entrepôt de données de publication (*Publication Repository*) : Entrepôt générique, multidisciplinaire, dédié aux données liées aux publications.

Modèle de référence (*Reference Model)** : cadre permettant la compréhension des relations essentielles entre entités d'un environnement donné, ainsi que le développement de normes ou de spécifications cohérentes liées à cet environnement.

Un Modèle de référence s'appuie sur un nombre restreint de concepts unificateurs et peut servir de support à la formation et à la sensibilisation des non-spécialistes à ces normes.

Entrepôt de projet de recherche (*Research project repository*) : Collecte les résultats d'un projet de recherche soumis à cette obligation de dépôt par un financeur ou une organisation comme un « enregistrement de la science ». Souvent lié à un projet (pluri)disciplinaire spécifique.

Réutilisation (*Reuse*): L'utilisation de données collectées dans le seul but d'étudier un nouveau problème ou de vérifier les conclusions du producteur initial de données.

Plan de succession (*Succession Plan)** : document qui détermine quand et comment la gestion, la propriété et/ou le contrôle des fonds de l'OAIS seront transférés à un autre OAIS afin d'assurer la continuité de leur pérennisation.
