



COPiLOtE



Guide ANR COPiLOtE - Certification CTS

Guide ANR COPiLOtE pour la certification CoreTrustSeal (CTS)

Partenaires projet



CNRS - SORBONNE UNIVERSITÉ
Station Biologique
de Roscoff

Université
de BORDEAUX

Titre court

Guide ANR COPiLOtE - Certification CTS

Titre long

Guide ANR COPiLOtE pour la certification CoreTrustSeal (CTS)

Auteur

Cécile NYS, OCEANSCOPE ;
Joël SUDRE, CNRS UMS CPST ;
Valérie HARSCOAT, IFREMER IRSI/SISMER.

Approbation

Gilbert MAUDIRE, IFREMER IRSI/D

Dissémination

Publique

Copyright

© ODATIS, 2021

DOI

[10.13155/81272](https://doi.org/10.13155/81272)

Citation

Nys Cecile, Sudre Joel, Harscoat Valerie (2021). **Guide ANR COPiLOtE - Certification CTS. Guide ANR COPiLOtE pour la certification CoreTrustSeal (CTS)**. Version 2.0 du 24 juin 2021.
<https://doi.org/10.13155/81272>

Référence projet

Cette publication a été réalisée avec le soutien de l'ANR dans le cadre du projet ANR-19-DATA-0020

Historique

Version	Auteurs	Date	Commentaires
1.0	Cécile NYS, Joël SUDRE & Valérie HARSCOAT	27/04/2021	Reprise du chapitre CTS du guide Copilote en version draft (version 0.9) ; Organisation en 3 sections ; Révision §1 ; Correction inversion échangés sur R9-R10.
1.1	Cécile NYS	17/06/2021	Vérification des marques de révisions ; Anonymisation des échanges ; Ajout section « Références »
1.2	Cécile Nys	24/06/2021	Mise en page ; Mise à jour des références.
2.0	Cécile Nys	24/06/2021	Version diffusable en ligne

Table des matières

1	INFORMATIONS GENERALES	5
1.1	LIENS UTILES	5
1.2	CRUSOE	5
2	LECTURE DES CRITERES (« REQUIREMENTS ») DE LA CERTIFICATION CTS	6
2.1	R0 - LE CONTEXTE	6
2.2	INFRASTRUCTURE ORGANISATIONNELLE	8
	R1 – Mission/périmètre	8
	R2 – Licences	8
	R3 – Continuité de l'accès	8
	R4 – Confidentialité/éthique	9
	R5 – Infrastructure organisationnelle	10
	R6 – Conseils d'experts	10
2.3	GESTION DES OBJETS NUMÉRIQUES	11
	R7 – Intégrité et authenticité des données	11
	R8 – Appréciation et sélection des données	12
	R9 – Procédures d'archivage documentées	12
	R10 – Plan de préservation	12
	R11 – Qualité des données	14
	R12 – Processus de traitement (Workflows)	14
	R13 – Découverte et identification des données	15
	R14 – Réutilisation des données	16
2.4	TECHNOLOGIE	17
	R15 – Infrastructure technique	17
	R16 – Sécurité	17
3	FAQ	18
3.1	EMBARGO/MORATOIRE SUR LES DONNÉES ?	18
3.2	SAUVEGARDE DISTANTE ?	18
3.3	ACCÈS RESTREINT ?	18
4	REFERENCES	19

Abréviations

Abbreviations	Signification
CDS	Centre(s) de Données et Services
CTS	CoreTrustSeal (certification)
DOI	Digital Object Identifier (Identifiant unique d'un objet digital)
FAIR	Findable (Trouvable) Accessible (Accessible) Interoperable (Interopérable/Interopérabilité) Reusable (Ré-utilisable)
GT	Group de Travail
PIA3	Programme d'Investissements d'Avenir ¹ (numéro 3)
PID	Persistent Identifier (Identifiant permanent)
RDA	Research Data Alliance

¹ <https://www.gouvernement.fr/le-programme-d-investissements-d-avenir>

1 Informations générales

1.1 Liens utiles

Présentation du projet ANR COPiLOtE lors du Bureau Exécutif plénier du pôle de données ODATIS, le 11 juin 2020 matin : [ODATIS_BE_COPiLOtE_202006.pdf](#)².

Présentation « La certification des entrepôts de données » de la RDA lors du webinar « Certification pour les débutants », le 13 mai 2020 : [La Certification des entrepôts de données_AteliersMai2020_V02.pdf](#)³.

Le « Guide étendu des critères CoreTrustSeal (CTS) » : [CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022](#)⁴.

Dossier de candidature de certification CTS de l'IFREMER/SISMER : [IFREMER-SISMER.pdf](#)⁵.

Template du projet ANR COPiLOtE pour les critères CTS (format .docx) : « ANR COPiLOtE - Critères CoreTrustSeal (CTS) - Template »⁶.

1.2 CRUSOE

Page d'accueil de l'application CRUSOE : <https://crusoe.omp.eu>.

Page d'authentification de l'application CRUSOE : <https://crusoe.omp.eu/application/#/login>.

Crusoe, est un outil qui permet de remplir un fichier pour chaque critère CTS (« requirements ») et qui permet de voir de façon très synthétique l'état/le statut de votre CDS vis-à-vis des différents critères. C'est un outil pour gérer l'auto-évaluation CTS et qui continue d'évoluer. Il permet de versionner le *Repository* (l'entrepôt de données). Il y a 2 rôles : lecteur et manager. Il faut un [ORCID](#)⁷ pour pouvoir accéder à l'application CRUSOE.

Guide d'utilisation de CRUSOE : « [CRUSOE_FR.pdf](#) »⁸.

Documentation de CRUSOE : <https://crusoe.omp.eu/documentation/>.

²https://www.odatis-ocean.fr/fileadmin/documents/activites/reunions_BE/202006/ODATIS_BE_COPiLOtE_202006.pdf

³https://www.rd-alliance.org/sites/default/files/La%20Certification%20des%20entrep%C3%B4ts%20de%20donn%C3%A9es_AteliersMai2020_V02.pdf

⁴ <https://doi.org/10.5281/zenodo.3632533>

⁵ <https://www.coretrustseal.org/wp-content/uploads/2019/11/IFREMER-SISMER.pdf>

⁶https://www.odatis-ocean.fr/fileadmin/documents/activites/projets/ANR_Copilote/Template_ANR_COPiLOtE_Criteres_CTS.docx

⁷ <https://orcid.org/>

⁸ https://www4.obs-mip.fr/wp-content-omp/uploads/sites/28/2020/01/CRUSOE_FR.pdf

2 Lecture des critères (« requirements ») de la certification CTS

Cette section reprend les échanges de la session de lecture croisée qui s'est déroulée le 11 Juin 2020 après-midi lors d'un Bureau Exécutif plénier du pôle Odatis (voir 1.1 ci-dessus).

Elle complète le Guide étendu des critères CTS de la RDA (voir 1.1 ci-dessus).

2.1 R0 - Le contexte

Type d'entrepôt

Brève description de l'entrepôt & différents jeux de données.

Brève description de la communauté concernée.

Niveau de curation

- A : Contenu 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.

Partenaires (lister & indiquer les accords)

Résumé des modifications depuis la candidature précédente (s'il y a lieu)

Question (Q) : Niveau de curation. Données générées par expert du domaine, quelle est la responsabilité de l'entrepôt en terme de qualité des données ?

Réponse (R) : Au niveau de la certification, ce n'est pas la donnée qui est certifiée, mais c'est l'entrepôt. Les évaluateurs ne vont pas vérifier la qualité de chaque donnée.

R : C'est plutôt la chaîne de traitement, le processus qui sera évalué.

R : Oui.

R : Ne pas hésiter à mettre toute la documentation des différents processus dans le dossier, d'autant plus s'ils sont en ligne. Documentation de logiciel, documentation de qualification, ...

R : Intégrer aussi les compétences des centres, ...

R : Dire que l'ensemble des données ont subi le protocole de "blind test" (ateliers à Résomar) effectué par des experts reconnus par la communauté nationale.

R : attention il faut être clair pour le contexte (et les autres critères!) car les reviewers ne sont pas forcément des experts de votre domaine.

Q : Faut-il traduire tous ces documents ?

R : Non pas obligatoire, mais conseillé pour certains documents. Cependant, les évaluateurs feront appel à des francophones (dans le pool des évaluateurs CTS) si besoins de traduction.

Q : Faut-il décrire tous les modèles de bancarisation ? Conceptualisation ?

R : Les modèles peuvent être décrits dans d'autres documents auxquels on fait référence dans le document de certification. Pas besoin d'entrer dans le détail du modèle relationnel. Par contre important d'expliquer pourquoi il y a ce mode de gestion de l'entrepôt. L'idée est d'expliquer dans quel contexte (national, international) l'entrepôt s'inscrit pour sa gestion.

Q : Périmètre ?

R : L'idée est de mutualiser tout ce qui est mutualisable. Il ne faut pas se restreindre à l'Océano/Odatis, mais réfléchir au niveau organisme qu'est-ce qui peut être certifié d'un bloc (ex. : même service d'archivage, ...) du moment que ça constitue un ensemble cohérent (Point de vue techno, mandat, méthode de gestion, équipe de gestion, ...).

2.2 Infrastructure organisationnelle

Responsabilité : Coordinateur/Coordination

R1 – Mission/périmètre

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Indiquer dans quel contexte et pourquoi, quel mandat on a reçu pour faire ce qu'on fait. Critère de périmètre de l'ensemble des services fournis. Derrière cela découlent tous les moyens attribués.

R2 – Licences

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Question (Q) : Contrôle d'accès. Surveillance de la conformité aussi pour les utilisateurs ?

Réponse (R) : Préciser si le centre dispose de licence.

Q : Licence sur toutes les données ? Contrainte sur le type de licence ?

R : Il faut que la licence ou l'absence de licence soit explicite et expliquer la politique mise en œuvre et la justifier. Les licences sont explicites pour les utilisateurs.

R3 – Continuité de l'accès

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Question (Q) : Le cache technique envisagé pour Odatis permettrait de rassembler les données pour un meilleur MCO (continuité de service) et une meilleure mise en accès. Cette partie peut potentiellement être identique et décrite pour plusieurs CDS au sein d'Odatis.

Réponse (R) : Certaines parties pourraient être déléguées à Odatis et/ou Data Terra. Idée de faire un état des lieux pour chacun des CDS si c'est un point bloquant pour certains CDS, voir si quelque chose ne pourra pas être implémenter.

Q : Comment ce point a-t-il été traité dans la certification d'Ifremer ?

R : Taux de disponibilité sur l'année. Il a également été spécifié dans le dossier de certification de l'Ifremer que ce n'était pas encore abouti, notamment pour l'aspect archive long-terme.

Information complémentaire (INFO) : Proposition d'un back-up dans un autre centre. Cette démarche permet de se mettre en contact éventuellement avec un autre CDS pour proposer un back-up croisé.

INFO : Il s'agit d'un critère en lien avec le mandat de l'entrepôt.

R4 – Confidentialité/éthique

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

- How does the repository comply with applicable disciplinary norms?
- Does the repository request confirmation that data collection or creation was carried out in accordance with legal and ethical criteria prevailing in the data producer's geographical location or discipline (e.g., Ethical Review Committee/Institutional Review Board or Data Protection legislation)?
- Are special procedures applied to manage data with disclosure risk?
- Are data with disclosure risk managed appropriately to limit access?
- Are data with disclosure risk distributed under appropriate conditions?
- Are procedures in place to review disclosure risk in data, and to take the necessary steps either to anonymize files or to provide access in a secure way?
- Are staff trained in the management of data with disclosure risk?
- Are there measures in place if conditions are not complied with?
- Does the repository provide guidance in the responsible deposit, download, and use of disclosive or potentially disclosive data?

Information complémentaire (INFO) : Exemple des données liées à la directive/loi GDPR, données personnelles et des données d'échantillons biologiques visées par l'accord international APA (éviter le pillage des ressources, droit de la mer). Dans le cas d'Odatis seront concernées les données en terme de droit de la mer, directives environnementales, données APA.

R5 – Infrastructure organisationnelle

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Question (Q) : Sous-entend qu'il faut présenter le budget de fonctionnement ? Source de financement récurrent ?

Réponse (R) : Oui effectivement. Qui finance quoi ?

R : Pour le personnel, il faut distinguer le personnel permanent du personnel temporaire. Objectif de pérennité et durabilité du CDS.

R : Dépend aussi de la taille de l'entrepôt.

R : Tous les critères doivent aussi être discutés avec l'organisme car cela affiche aussi une forme d'engagement de l'organisme/tutelle.

R : Certaine forme d'assurance/de validation de la tutelle et/ou direction hiérarchique. Permet aussi "d'éviter" une fermeture du jour au lendemain du CDS. Manière de mettre le doigt sur certains points RH qui devront être pris en compte tôt ou tard par la direction.

R : Souvent appel à des prestataires extérieurs pour des compétences non présentes en interne au CDS.

R : Implique qu'il y a une demande récurrente d'une prestation.

R : Spécifier comment assurer le MCO dans ce contexte et montrer comment est contrôlée la prestation. Aspect organisationnel qui peut être variable et expliciter pourquoi et quelle continuité des services.

Information complémentaire (INFO) : Le formalisme peut être inspiré de dossiers déjà soumis.

R6 – Conseils d'experts

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).

Information complémentaire (INFO) : Expertise externe qui permet de consolider.

INFO : Expertise lié aux SNO, à des IR d'observations qui sont à formaliser. Expliquer comment cela est mis en œuvre, faire appel à des documents décrivant le processus.

2.3 Gestion des objets numériques

Responsabilité : Scientifique/Gestionnaire de données

R7 – Intégrité et authenticité des données

R7. The repository guarantees the integrity and authenticity of the data.

For this Requirement, responses on data integrity should include evidence related to the following:

- Description of checks to verify that a digital object has not been altered or corrupted (i.e., fixity checks) from deposit to use.
- Documentation of the completeness of the data and metadata.
- Details of how all changes to the data and metadata are logged.
- Description of version control strategy.
- Usage of appropriate international standards and conventions (which should be specified). Evidence of authenticity management should relate to the following questions:
- Does the repository have a strategy for data changes? Are data producers made aware of this strategy?
- Does the repository maintain provenance data and related audit trails?
- Does the repository maintain links to metadata and to other datasets? If so, how?
- Does the repository compare the essential properties of different versions of the same file? How?
- Does the repository check the identities of depositors?

Question (Q) : Derrière ce critère peut se trouver l'aspect des DOI et des identificateurs permanents (PID). Comment gérer les intégrités ? Comment un retour en arrière est possible pour un jeu de données spécifique ?

Réponse (R) : Couvre le cycle complet de la donnée de l'ingestion jusqu'à la diffusion.

R : Peut avoir des liens avec le matériel pour vérifier que les données ne sont pas corrompues (acte malveillant, ...). Code cyclique attaché au fichier, etc.

Q : checksum ? ou code d'autocorrection ?

R : A minima check-sum, mieux auto-correction.

Q : Critère couvre cycle complet de la donnée à partir de quand ?

R : A partir du moment où le CDS accueille la donnée. Donnée et métadonnée associée.

R : Peut inclure l'étape de changement de format suite à harmonisation. Ne décrit pas la technologie mise en œuvre, mais plutôt la méthode.

R8 – Appréciation et sélection des données

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Information complémentaire (INFO) : Montrer que dans les procédures mises en place seules les données "pertinentes" pour la communauté rentrent dans le CDS.

INFO : S'applique beaucoup quand on a des systèmes de dépôt de données en précisant la typologie de données particulière acceptée (cf. Mandat).

R9 – Procédures d'archivage documentées

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Question (Q) : Cela peut devenir intéressant de dire qu'on utilise un partenaire externe au CDS. C'est là qu'il faudra expliciter comment le CDS fait appel au CINES.

Réponse (R) : Difficulté avec les partenaires externes, est-ce qu'il y a une convention, la relation est-elle contractualisée ? Il faut expliciter le contexte dans lequel c'est fait. Odatis/Data Terra pourra peut-être se porter relais pour formaliser un tel accord/contrat.

Q : Agrément SIAF ?

R : Non.

R : Risque de rupture dans la chaîne de confiance sans un partenaire avec agrément SIAF.

R : Expert en France, BNF (Bibliothèque Nationale de France).

R10 – Plan de préservation

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Lié avec le critère R9.

Information complémentaire (INFO) : Peut faire référence à de la documentation mis en ligne.

Question (Q) : Est-ce qu'il y a différents niveaux de rédaction? Rédaction globale pour tous le CDS et puis silo de données par silo de données ?

Réponse (R) : Pour Ifremer, il y a 6 filières de données avec des troncs communs et en explicitant les spécificités des filières quand nécessaire (expliciter les différences).

Q : Archivage long-terme ? Ou back-up ?

R : Archivage long-terme. Se préoccupe de la pérennité des formats, modalités techniques, format.

Q : Sous-entend qu'il y a un modèle OAIS derrière ?

R : Oui qu'il y ait un modèle derrière, pas forcément OAIS, mais un modèle.

Q : Comment préserver sur le long-terme des BDD relationnelles ? Comment décrire ?

R : Export de toutes les BDD relationnelles avec un "format pivot".

Q : Coté Ifremer vous faites des exports de toutes les BDD relationnelles à des "formats pivots" ?

R : Oui en NetCDF.

Q : Un 'dump' de BDD ça ne suffit pas ?

R : Pas catégorique, dépend peut-être de la BDD. Pour BDD open source, un dump devrait suffire. Par contre pour BDD commercial, pas sûr à quelle « *sauce ça va être manger* » dans le futur si révision intégrale ou en profondeur de la structure commerciale. Le format ASCII – CSV est le plus pérenne.

INFO : L'idéal est d'utiliser comme format pivot un des formats reconnus par le CINES pour la pérennité. Pour cela il faut aller consulter les sites suivants du CINES qui répertorient tous les formats acceptés pour la pérennité : <https://facile.cines.fr/> & <https://www.cines.fr/archivage/nos-solutions-darchivage/pac/>. En ASCII-CSV et HDF ; en discussion sur NetCDF (notamment dans le cadre PIA3).

INFO : NetCDF est peu compatible avec de l'archivage pérenne.

INFO: Cela est accompagné de la conservation du code source des outils « readers » pour les formats.

INFO : A l'Ifremer « *formats d'export* » c'est la 1^{ière} approche, la 2^{ième} approche est un engagement à suivre les transitions technologiques (par exemple on passe de la version X à Y sur une base de données).

Q : Faut-il investiguer sur le format SIARD (SHS) de préservation de BDD relationnelles, utilisé pour les archives fédérales Suisse ? Partenariat externe possible ? Faut-il un agrément SIAD pour le partenaire (agrément de confiance des archives de France) comme a eu CINES ?

R : Peut être intéressant d'y réfléchir collectivement. Si partenariat, il faut une convention.

INFO : SIARD a été utilisé pour les archives fédérales Suisse pour archivage à long-terme des BDD relationnel.

R11 – Qualité des données

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality related evaluations.

Question (Q) : Quid des données brutes ? Peut-on à la fois avoir données brutes et ces mêmes données qualifiées avec une procédure de qualification mise en œuvre qui est décrite dans document ad-hoc.

Réponse (R) : Oui.

Q : La qualité des données commence déjà avant la récolte des données, faut-il dès lors décrire cette 'pré-qualification' des instruments de mesures (capteurs) ou de comment les scientifiques calibrent leur prise de donnée ?

R : Faire référence aux documents de ces ateliers de qualification "pré-récolte" des données. Il faut que l'utilisateur le sache.

Information complémentaire (INFO) : En lien avec les recommandations dans le « [cahier des charges ODATIS](#) »⁹, aller un peu plus haut que dublin core => ISO 19115 en terme de métadonnées, vocabulaires SeaDataNet, les EOVS ; au moins entre « nous ». Même si au-delà d'ODATIS tout ne sera opérable en claquant des doigts.

R12 – Processus de traitement (Workflows)

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Information complémentaire (INFO) : Du dépôt des données jusqu'à la diffusion. Cela peut être variable d'un CDS à un autre.

INFO : Workflow déjà décrit dans le R7. R7 très général, entre plus dans le détail dans ce R12.

⁹<https://doi.org/10.13155/81220>

R13 – Découverte et identification des données

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Proche du Findable de FAIR

Question (Q) : C'est en lien avec le DOI, ... ? Une fiche de metadata suffirait ?

Réponse (R) : Une fiche de metadata à minima par jeux de données.

R : Catalogue du niveau catalogue Odatis et SDN (catalogue à 2 niveaux de granulométrie).

Q : Pas d'obligation d'avoir un catalogue interne au CDS, il peut faire référence à un catalogue où il dépose ses metadata ?

R : Oui.

Information complémentaire (INFO) : Jointure avec R14. R13 concerne plutôt les metadata de découverte. Il s'agit de métadonnées plutôt standardisées de type ISO 19115. R14 concerne les metadata d'usage (permet traitement, analyse de donnée, ...), metadata souvent incluses dans le jeu de données.

INFO : Reprend/rejoint les 2 niveaux de catalogue : catalogue haut-niveau et catalogue avec metadata usage (ex. ERDDAP).

Q : Evoquer le moissonnage ? Si on répond à des standards de moissonnage.

R : C'est un plus.

R : Si on attend un standard de moissonnage, on risque d'attendre longtemps.

R : Il faudrait des recommandations de standard de moissonnage pour la communauté océano.

INFO : Les recommandations de standardisation/homogénéisation sont décrites pour partie dans le CdC des CDS d'Odatis. Il est conseillé de suivre ces recommandations afin d'être en accord avec ce qui se fait à Odatis et puis à plus haut niveau avec Data Terra.

INFO : Assez ISO19115 et OAI-PMH. Le minima serait de respecter l'ISO 19115 pour les metadata de découvertes.

Q : Il faut recommander des standards, mais pas que ça devienne des exigences immédiates.

Q : DOI, DublinCore.

R : En lien avec les recommandations dans le cahier des charges ODATIS, aller un peu plus haut que dublin core => ISO 19115 en terme de métadonnées, vocabulaires SeaDataNet, les EOVs (au moins entre « nous ».) Même si au-delà d'ODATIS tout ne sera opérable en claquant des doigts.

INFO: Pas mal de travaux en cours sur le sujet de metadata et standardisation et recommandations dans le pôle, au niveau de l'IR Data Terra, GT Interpôle et GT Tech Data Terra. Les metadata broker (le « liant ») sont plutôt effectués à un mode best effort pour le moment.

R14 – Réutilisation des données

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Information complémentaire (INFO) : Niveau plus précis : décrire les variables, unités, etc. Recommandation Odatis, c'est d'inclure ça dans les attributs dans les fichiers (cfr. <https://www.odatis-ocean.fr/ressources/bonnes-pratiques>).

Question (Q) : Cache de donnée niveau Odatis : Est-ce dans la rubrique Recommandations ?

Réponse (R) : Le cache Odatis va juste être une copie locale au niveau d'un "centre" pour éviter les ruptures de liaison. Ce cache de donnée serait qqch du genre « data lake ».

R : C'est les CDS qui doivent générer les metadata de données dans les fichiers de leur CDS pour permettre aux utilisateurs une meilleure utilisation/compréhension des données que l'utilisateur va utiliser sans besoin de devoir aller à la pêche aux infos.

2.4 Technologie

Responsabilité : ASR/Informaticien

R15 – Infrastructure technique

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

Information complémentaire (INFO) : Cela concerne toute la partie serveur, hardware et software. Comment sauvegarder les software, capacité de reconstruire après un crash ? Concerne le hard- et software permettant de rendre disponible la donnée.

INFO : Il faut décrire serveur, hardware, software, communication réseau, comment la partie logicielle est sauvegardée. On peut revenir ici sur les parties archivages de l'entrepôt si ce n'est pas fait de manière externalisée.

Question (Q) : Peut-on parler technologie de façon très détaillée ?

R : Dossier relu par 3 experts dont un expert "technique". Ecrire de façon compréhensible, mais peut renvoyer vers un document beaucoup plus détaillé. Il est important de mettre en évidence que l'infrastructure technique est pérenne, bien suivie, qu'on a tous les contrats de maintenance (machine & logiciel) pour que ça ne tombe pas en panne. Qu'on a une configuration spéciale dans une salle machine avec gestion de serveur, connexion serveur, baies de stockage, ... Ce ne doit pas être ultra-technique, mais plutôt démontrer que c'est une infrastructure suivie et qui a les moyens de fonctionner.

Q : Peux-t-on faire référence aussi aux contraintes ?

R : Oui et non. S'il y a c'est bien d'expliquer, de même que s'il n'y a pas faut expliquer aussi pourquoi et comment.

R16 – Sécurité

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Information complémentaire (INFO) : Il y a plusieurs volets pour ce critère :

- Sécurité physique : salle machine protégée, seule une personne habilitée peut entrer, extinction "automatique" d'incendie, protection contre événements extérieurs (ex. inondations), ...
- Sécurité logique : quelles sont les personnes habilitées à gérer les données. Comment s'assurer que c'est bien la personne en charge derrière l'écran. Politique de l'annuaire, de gestion des MdP et des comptes.

3 FAQ

Cette section reprend les questions et réponses au fil des réunions régulières de support aux Centres de données et Services (CDS) pendant la phase d'auto-évaluation et de renseignement des critères CTS.

3.1 Embargo/moratoire sur les données ?

Réponse : Possible mais non normalisé dans CTS. Il faut juste que la durée soit limitée dans le temps, expliquée (dans un doc si ce n'est pas un truc standard comme ETALAB ou autre) et explicite/connue des utilisateurs. L'embargo est donc recevable pour une durée limitée. La licence française EtatLab par exemple prévoit un embargo de 6 mois. La pratique à Ifremer/SISMER est d'un an renouvelable (normalement une fois). En pratique autant que nécessaire, surtout en biologie, car l'interprétation initiale des données par l'équipe embarquant peut-être très longue pour de très bonnes raisons : par exemple disponibilité des taxinomistes spécialistes de l'océan profond).

3.2 Sauvegarde distante ?

Réponse : Les 250km ne sont pas actuellement un critère CTS, c'est un critère pour les archives selon la certification ISO. Il faut simplement détailler les dispositifs pour assurer une bonne conservation physique des données (pour Ifremer double archive, dans deux bâtiments séparés par exemple). Cela reste un point de vigilance pour l'avenir malgré tout.

3.3 Accès restreint ?

Réponse : Possible, mais accès en ligne impératif pour faire la demande (donc a minima un formulaire) et ensuite accès en ligne aux données mises à disposition de façon restreinte. L'accès restreint est également possible et justifié dans certains cas : respects de restrictions tels qu'accord avec pays tiers, convention public/privé, etc. La liste des exceptions est cependant limitée (grosso modo : protection des intérêts économiques ou géopolitiques par exemple sur la pêche). Normalement, hors période d'embargo initiale, les données environnementales doivent être rendues publiques. Accès restreint veut aussi dire complexité technique, au-delà de l'identification : gestion de droit d'accès (catégories d'utilisateurs, ...), protections des services d'accès, des données extraites, etc. Marin Dacos (Science Ouverte au Ministère) l'exprime ainsi « *Les données doivent être aussi ouvertes que possibles, aussi fermées que nécessaire* ». Donc on a une part d'appréciation, mais il faut la justifier. Ce qui ne peut pas l'être pour des motifs tels que ce sont MES données.

4 Références

- CINES (2017). **PAC – Plateforme d’Archivage au CINES**. Available at: <https://www.cines.fr/archivage/nos-solutions-darchivage/pac/> [Accessed June 11, 2020].
- CINES (2021). **FACILE - Service de validation de formats**. Available at: <https://facile.cines.fr/> [Accessed June 11, 2020].
- CoreTrustSeal Standards and Certification Board (2019). **CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022**. 1–26. doi:10.5281/ZENODO.3632533.
- CoreTrustSeal Standards and Certification Board (2019). **Core Trustworthy Data Repositories Requirements**. 1–14.
- Genova, F., & Rouchon, O. (2020). **La certification des entrepôts de données**. 1–44. Available at: https://www.rd-alliance.org/sites/default/files/La_Certification_des_entrepôts_de_données_AteliersMai2020_V02.pdf.
- Harscoat, V., Sudre, J., & Nys, C. (2020). **Présentation du projet ANR COPiLOtE**. 1–8. Available at: https://www.odatis-ocean.fr/fileadmin/documents/activites/reunions_BE/202006/ODATIS_BE_COPiLOtE_202006.pdf.
- IFREMER, & IFREMER-IRSI-SISMER (2019). **IFREMER-SISMER - CoreTrustSeal Application**. Available at: <https://www.coretrustseal.org/wp-content/uploads/2019/11/IFREMER-SISMER.pdf>.
- Maudire, G., Schmidt, S., & Sudre, J. (2020). **Cahier des charges des Centres de Données du Pôle de Données Océan – ODATIS de l’Infrastructure de Recherche Data Terra**. 1–23. doi:10.13155/81220.
- Nys, C., & ODATIS (2021). **ANR COPiLOtE - Critères CoreTrustSeal (CTS) - Template**. 1–8. Available at: https://www.odatis-ocean.fr/fileadmin/documents/activites/projets/ANR_Copilote/Template_ANR_COPiLOtE_Criteres_CTS.docx.
- ODATIS (2021). **Bonnes pratiques pour la gestion de données**. www.odatis-ocean.fr. Available at: <https://www.odatis-ocean.fr/ressources/bonnes-pratiques> [Accessed June 24, 2020].
- SEDOO (2021). **Crusöe – Documentation & FAQ**. Available at: <https://crusoe.omp.eu/documentation/> [Accessed June 11, 2020].
- SEDOO (2021). **Crusöe – Monitor your certification**. Available at: <https://crusoe.omp.eu/> [Accessed June 11, 2020].
- SEDOO (2020). **Crusöe - User’s guide (French)**. 1–19. Available at: https://www4.obs-mip.fr/wp-content-omp/uploads/sites/28/2020/01/CRUSOE_FR.pdf.
- SUDRE, J., & Nys, C. (2020). **Compte rendu de l’Atelier Technique ODATIS du 8 et 9 octobre 2019**. doi:10.13155/81249.