



**DEUTSCHE INITIATIVE
FÜR NETZWERKINFORMATION E.V.**

**Certificat DINI
Services de Dépôts
et de Publication en Libre Accès
2010**

Groupe de travail « Edition numérique »





**DEUTSCHE INITIATIVE
FÜR NETZWERKINFORMATION E.V.**

**Certificat DINI
Services de Dépôts
et de Publication en Libre Accès
2010**

Groupe de travail « Edition numérique »



Ce document est protégé par une licence Creative Commons CC-BY
<http://creativecommons.org/licenses/by/2.0/fr/>.

Le document est disponible en ligne à l'adresse
<http://nbn-resolving.de/urn:nbn:de:kobv:11-100200620>.

DINI – Deutsche Initiative für Netzwerkinformation e. V.

DINI-Geschäftsstelle [coordonnées]

c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen

Papendiek 14

37073 Göttingen

Allemagne

Tel.: +49 551 39-3857

Fax: +49 551 39-3856

E-Mail: gs@dini.de

www.dini.de

Sommaire

A propos de DINI	5
Synthèse	6
1 Objectifs	9
2 Critères	11
2.1 Visibilité du site	11
2.2 Charte (politique du site)	13
2.3 Aides aux auteurs et éditeurs	15
2.4 Aspects juridiques	17
2.5 Sécurité de l'information	21
2.6 Indexation et interfaces	24
2.7 Statistiques d'utilisation	26
2.8 Archivage pérenne	28
3 Certification et évaluation	30
4 Explications et exemples	32
4.1 Visibilité du site	32
4.2 Charte	32
4.3 Aide aux auteurs et éditeurs	32
4.4 Aspects juridiques	33
4.5 Sécurité de l'information	33
4.6 Indexation et interfaces	33
4.7 Statistiques d'utilisation	33
4.8 Archivage pérenne	33
Annexe A : Conditions pour l'interface OAI	35
A.1 Conformité avec le protocole	36
A.2 OAI-PMH: Conditions supplémentaires	39
A.3 Conditions pour métadonnées (Dublin Core simplifié)	49
Annexe B : Glossaire	52
Annexe C : Auteurs et traducteurs	55

Le certificat DINI est plus qu'un simple répertoire de « bonnes pratiques ». Mode d'élaboration, présentation et application le caractérisent très clairement comme (quasi) norme. Cela étant, le certificat distingue trois niveaux de recommandations dont seulement le premier niveau est obligatoire. De même, il contient des références à l'environnement allemand. La description des conditions et recommandations renvoient plus de vingt fois à des lois, établissements, formats, outils et sites allemands. Dans la traduction, nous avons marqué ces renvois par un astérisque (*) car l'utilisation en France devrait les transposer dans le contexte français.

A propos de DINI

Le développement des nouvelles technologies de l'information et de la communication engendre une mutation dans l'infrastructure des universités et organismes de recherche. Ce changement est l'une des priorités de l'enseignement supérieur allemand et nécessite plus que jamais concertation, coopération, recommandations et normes. L'initiative allemande pour l'information en réseau (Deutsche Initiative für Netzwerkinformation ou DINI) soutient cette évolution.

DINI a été créé pour faciliter l'amélioration des services d'information et de la communication et le développement des infrastructures informationnelles au sein des établissements de l'enseignement supérieur, ainsi qu'au niveau régional et supra-régional. La concertation et le partage du travail entre les services d'infrastructure doivent faire progresser les technologies de l'information et l'offre de service. Ceci implique aussi le développement de normes et recommandations.

DINI est une initiative de trois organisations partenaires :

- l'association des services audiovisuel et multimédia des universités (Arbeitsgemeinschaft der Medienzentren an Hochschulen, AMH),
- l'association des bibliothèques universitaires (section 4 du Deutscher Bibliotheksverband, dbv) et
- l'association des centres de calcul de l'enseignement supérieur et de la recherche (Zentren für Kommunikation und Informationsverarbeitung, ZKI).

L'objectif de DINI est

- de faire connaître et promouvoir des solutions et bonnes pratiques,
- d'encourager et de promouvoir le développement de normes et des recommandations pour leur mise en œuvre,
- de répertorier et faire connaître des centres de compétence à l'aide d'outils en réseau,
- d'organiser des conférences, ateliers, tables rondes etc. pour faciliter le retour d'expérience,
- d'informer sur des appels à projets et de favoriser le développement de nouveaux programmes.

Synthèse

Internet et les nouvelles technologies de l'information et de la communication ont bouleversé le système de la communication scientifique de fond en comble. De nouveaux dispositifs de distribution et de médiation se mettent en place, tel que le mouvement international vers le libre accès, dont l'objectif est la mise en ligne gratuite d'information scientifique.

Beaucoup d'institutions scientifiques se sont déjà investies dans le domaine des services de dépôts et de publication en libre accès. Elles se donnent les moyens de rendre leurs publications disponibles en ligne pour un public international, et de les conserver. En Allemagne, le Conseil Scientifique et la Conférence des Présidents des Universités soutiennent ce développement; le Ministère Fédéral de l'Education et de la Recherche et l'Agence Allemande pour la Recherche (DFG) subventionnent ces projets en priorité.

Ceci étant, les projets doivent s'appuyer sur une technologie confirmée et être compatibles avec des normes internationales pour optimiser la visibilité et l'impact de la production des chercheurs, pour refléter la performance scientifique d'une université ou d'un organisme de recherche.

Pour accompagner les projets et définir un cadre d'infrastructures, le groupe de travail « édition numérique » du DINI s'est saisi assez tôt de cette question et a publié début 2002 des recommandations pour l'édition universitaire¹. Par la suite, DINI a formalisé une série de critères dans le certificat DINI *Services de Dépôts et de Publication en Libre Accès* qui après deux éditions en 2004 et 2007 sort aujourd'hui dans une 3e version (2010); cette dernière version détaille tous les aspects techniques, organisationnels et juridiques nécessaires pour créer et faire fonctionner de tels services dans la durée.

La 1e version de 2004 avait mis l'accent sur les serveurs de publications universitaires, en particulier les plateformes pour la mise en ligne des thèses et habilitations. Le certificat DINI de 2007 insista sur les activités du mouvement international vers le libre accès, notamment la « voie verte », c'est-à-dire l'auto-archivage de documents publiés (articles, monographie, rapports, conférences etc.) sur des serveurs institutionnels ou thématiques (archives ouvertes). Par ailleurs, le certificat s'adressa davantage à des organismes de recherche et souligna la nature de « service » de ces sites.

¹ DINI: Elektronisches Publizieren an Hochschulen – Empfehlungen, 2002, voir <http://nbn-resolving.de/urn:nbn:de:kobv:11-10045992>

L'élaboration de la 3e version du certificat en 2010 prend en considération plusieurs aspects et développements :

- L'importance grandissante de la « voie dorée » du libre accès. Cette 2e stratégie concerne surtout la publication d'articles dans des revues en libre accès mais, dans un sens plus large, comprend également l'édition en libre accès d'autres types de documents comme par exemple les monographies ou les ouvrages collectifs. Ces documents sont normalement soumis à une procédure d'assurance qualité, le plus souvent sous forme d'une « peer review ». La DFG soutient l'édition en libre accès par une subvention structurante. Cette voie dorée vers le libre accès peut avoir par ailleurs une orientation institutionnelle (par exemple, éditeurs universitaires) ou thématique : elle est intégrée dans la nouvelle version du certificat DINI.
- Le besoin croissant d'interopérabilité avec des services fédératifs – services de référencement, moteurs de recherche et d'autres services à valeur ajoutée – qui sont particulièrement importants pour la visibilité des documents déposés : par exemple dans le réseau Open Access² en Allemagne ou DRIVER³ au plan européen. La qualité de ces services dépend de la qualité et de la normalisation des données mises à leur disposition. Pour cette raison, nous avons fait évoluer les recommandations pour l'interface OAI en les adaptant au manuel du projet DRIVER⁴.
- La virtualisation croissante de ces services facilite l'utilisation potentielle d'une infrastructure de base commune par plusieurs services. Pour autant, nous n'avons pas opté pour un certificat modulaire. Néanmoins il est possible de répartir les réponses au questionnaire du certificat sur plusieurs personnes, si le fonctionnement du service est assuré par plusieurs institutions avec des compétences différentes.
- Une approche holistique du processus de la recherche scientifique. A côté des publications textuelles comme produits « classiques » de la communication scientifique, cette approche inclut ce qui est à leur origine et les précède, les données brutes dont l'organisation et l'utilisation deviennent un enjeu majeur pour la recherche⁵. L'informatisation de la science ouvre de nouvelles pistes.

² Cf. <http://www.dini.de/projekte/oa-netzwerk/>

³ Cf. <http://www.driver-repository.eu/>

⁴ Cf. http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_2008-11-13.pdf

⁵ Cf. le document de DINI sur les données de recherche, Göttingen 2009, à <http://nbn-resolving.de/urn:nbn:de:kobv:11-10098082>

L'exploitation partagée des résultats de recherche dans le cadre de projets internationaux et interdisciplinaires n'est qu'un exemple parmi d'autres d'une gestion numérique des données scientifiques. Même dans sa version actuelle, le certificat DINI s'adresse d'abord aux services dédiés aux publications textuelles. Il est néanmoins possible d'adapter les conditions et recommandations, afin de certifier des services différents, avec d'autres formes d'information.

Du reste, la nouvelle version sert à consolider et développer les critères et leurs conditions. Les conditions et recommandations ont été reformulées et se présentent désormais sous forme d'une check-list, complétée par le questionnaire de la certification en parallèle.

La 3e édition du certificat DINI *Services de Dépôts et de Publication en Libre Accès* de 2010 tient compte des développements internationaux mentionnés ci-dessus et décrit comment de tels services peuvent contribuer à normaliser la publication en libre accès. Nous avons renoncé à un certificat modulaire afin de souligner la nécessité de normes et procédures partagées dans le processus scientifique. Mais nous sommes conscients que les spécificités disciplinaires peuvent limiter la portée d'une norme au sens strict.

Un groupe de travail vérifie et révisé les critères du certificat DINI à partir de normes et travaux internationaux. Pour cette raison, le certificat mentionne l'année de la version dans le titre.

1 Objectifs

Avec le *World Wide Web*, la science a créé un nouvel outil de communication qui est aujourd'hui largement commercialisé. Le web fait partie du quotidien des chercheurs mais son potentiel pour la communication scientifique est encore loin d'être épuisé. L'informatique facilite l'ensemble du processus scientifique, depuis l'émergence des idées, la préparation du projet, l'expérimentation, les enquêtes et mesures, jusqu'à la publication des résultats et leur valorisation sous forme de brevets ou d'enseignements.

Les médias numériques sont de plus en plus acceptés. Surtout dans les disciplines STM (sciences, technologie, médecine), la plupart des résultats sont publiés en format numérique. Néanmoins, le quasi-monopole de quelques grands éditeurs STM a entraîné une politique tarifaire exorbitante qui dépasse aujourd'hui les limites financières de la recherche publique. Dans le cadre du libre accès (voie verte ou voie dorée), la création de services de dépôts et de publication peut être une réponse à cette situation. Quand une majorité de publications sera accessible sur des serveurs à caractère non-commercial, ce nouveau réseau rendra la course effrénée aux bénéfices plus difficile.

La rédaction d'un catalogue de critères pour le certificat DINI et la certification de services de dépôts et de publication poursuivent plusieurs objectifs :

- Le renforcement de l'infrastructure des services d'information dans le domaine du libre accès.
- La définition de conditions minimales pour ces services suivant une description détaillée.
- La mise en place d'un label de qualité pour faciliter la comparaison des services de dépôts et de publication par les utilisateurs, organismes et agences de moyens.
- L'identification de tendances actuelles et futures pour le développement des services de dépôts et de publication et pour l'échange d'information.
- Un positionnement visible des services de dépôts et de publication comme services à forte valeur ajoutée dans le cadre de leur institution ou discipline.

La 3e version du certificat DINI contient une description précise des critères de conformité d'un service de dépôts et de publication avec les normes et développements nationaux et internationaux. Pour les services qui remplissent les conditions du certificat DINI, ce label de qualité augmentera leur visibilité au sein de leur institution ou discipline mais aussi sur le plan international. Les services

certifiés constituent les maillons d'une chaîne croissante de fournisseurs non-commerciaux de contenu.

La certification est censée garantir l'interopérabilité globale des services au sein du réseau émergent de communication scientifique. Elle n'est pas limitée aux établissements de l'enseignement supérieur et de la recherche. Des éditeurs « open access », agrégateurs et centres de calcul sont également invités à contribuer à la communication scientifique de demain et à demander la certification DINI.

Par le biais de la certification, DINI propose un contrôle de qualité transparent des services de dépôts et de publication. Le certificat regroupe des conditions minimales indispensables pour une communication scientifique moderne. Des recommandations complémentaires anticipent des développements à venir et permettent d'aller plus loin.

2 Critères

Le certificat DINI inclut huit critères :

Critère 1 – *Visibilité du site* (section 2.1.)

Critère 2 – *Charte* (section 2.2.)

Critère 3 – *Aide aux auteurs et éditeurs* (section 2.3.)

Critère 4 – *Aspects juridiques* (section 2.4.)

Critère 5 – *Sécurité de l'information* (section 2.5.)

Critère 6 – *Indexation et interfaces* (section 2.6.)

Critère 7 – *Statistiques d'utilisation* (section 2.7.)

Critère 8 – *Archivage pérenne* (section 2.8.)

Les critères pour l'interface OAI qui sont aussi partie du certificat DINI sont écritent dans l'annexe A.

Chaque critère est présenté en deux parties. Dans un premier temps, chaque section décrit en détail les conditions minimales (M). Pour qu'un service de dépôt et de publication en libre accès obtienne le certificat DINI, il doit remplir toutes les conditions minimales. Ensuite, les sections proposent quelques recommandations supplémentaires (R) en termes de bonnes pratiques et quelques pistes d'évolution à venir. Pour la certification, il n'est pas nécessaire de satisfaire à toutes ces recommandations supplémentaires. Cependant, dans la mesure où le certificat évoluera avec le temps, il est probable que certaines de ces recommandations deviennent des conditions minimales dans une future version.

Chaque section commence par une courte introduction qui remet le critère dans le contexte et explique sa raison d'être. Les conditions minimales sont formulées de telle sorte qu'on puisse y répondre par un « oui » ou un « non », comme pour une check-list. Sur fond gris se trouvent d'autres définitions, interprétations, réalisations et exemples.

Pour mieux comprendre, le chapitre 4 contient quelques informations complémentaires (explications, exemples).

2.1 Visibilité du site

Parmi les avantages significatifs des publications numériques figurent une visibilité plus grande et un impact potentiel plus important, en particulier quand les publications sont mises en ligne en libre accès. Afin de réaliser au mieux ce potentiel, il faut promouvoir le plus largement possible l'offre globale du service. Le site doit être visible non seulement pour l'utilisateur direct (lecteur, auteur...) mais aussi pour d'autres services (moteurs de recherche, sites de référencement...). Mise à part la création d'interfaces techniques indispensables (cf. critère 6), il s'agit sur-

tout de faire enregistrer le site dans les bases et annuaires pertinents qui jouent un rôle de médiation entre les services de dépôt et de publication en libre accès et d'autres services de diffusion.

Conditions minimales

M.1-1 Le service est accessible à partir d'une page web.

⇒ Il s'agit de la page d'accueil ou page principale du site avec un lien direct vers le dépôt d'une publication (workflow) et vers l'accès aux documents déposés.

M.1-2 Le lien vers la page d'accueil du service est accessible à partir du site institutionnel de l'établissement (ou des établissements) responsable(s) du site.

⇒ Les utilisateurs potentiels doivent être guidés d'une manière intuitive depuis le site Web de l'université, du laboratoire ou de la bibliothèque etc. vers le service de dépôts et de publication.

M.1-3 Le service est enregistré auprès du DINI dans la liste des plateformes de publications.*

⇒ La liste du DINI est accessible à l'adresse <http://www.dini.de/dini-zertifikat/liste-der-repositorien/> L'enregistrement d'un nouveau site se fait à cette adresse.

M.1-4 Le service est enregistré auprès du répertoire *Directory of Open Access Repositories* (OpenDOAR).

⇒ OpenDOAR est accessible à <http://www.opendoar.org/> L'enregistrement d'un nouveau site se fait à cette adresse.

Recommandations complémentaires

R.1-1 Le service est enregistré auprès du répertoire *Registry of Open Access Repositories* (ROAR).

⇒ L'URL du site est <http://roar.eprints.org/>.

R.1-2 Le service est enregistré comme OAI-Data-Provider (fournisseur de métadonnées) auprès de l'*Open Archives Initiative* (OAI).

⇒ La liste des OAI-Data-Provider enregistrés se trouve à l'adresse <http://www.openarchives.org/Register/BrowseSites> Si les métadonnées d'une archive institutionnelles sont mises à disposition via un site agrégateur (par exemple, une fédération ou un réseau d'archives), il convient de faire enregistrer l'interface de ce dernier auprès de l'*Open Archives Initiative*. Cf. aussi le critère 6 – *Identification et interfaces*.

R.1-3 Le service est enregistré auprès de DRIVER.

⇒ DRIVER maintient un service européen d'accès aux publications déposées dans des archives institutionnelles. Cf. <http://www.driver-repository.eu/>

R.1-4 Tous les documents de la plateforme sont accessibles via des liens.

⇒ Ceci est important pour que les robots des moteurs de recherche puissent trouver les documents. Si un document n'est accessible qu'à travers une requête, sans adresse URL, un robot ne le trouvera pas.

2.2 Charte (politique du site)

Fiabilité et transparence sont essentielles pour la mise en place d'un service de dépôt et de publication en libre accès. L'organisme doit décrire en détail et d'une manière explicite sa politique en matière de contenu et fonctionnement technique (types de documents, publics ciblés, pérennité du service etc.). Cette charte engage l'organisme vis-à-vis des utilisateurs potentiels du service.

Conditions minimales

M.2-1 L'organisme dispose d'une charte publique pour le service.

⇒ Cette charte est formulée comme un engagement; elle est accessible à partir de la page d'accueil du site, sous forme d'un document.

La charte contient les points suivants :

M.2-2 Une définition des droits et obligations de l'institution.

⇒ Une description du service expliquant pour qui et sous quelles conditions il est mis à disposition.

M.2-3 Une définition des droits et obligations des auteurs et éditeurs qui utilisent ce service pour publier.

⇒ Avec par exemple une description des droits d'utilisation que l'auteur cède à l'institution.

M.2-4 Une description des documents publiés sur ce site (types, qualité technique, contenu).

⇒ Ce point correspond d'abord à une mission de collecte. Les critères de qualité concernant le contenu (peer review etc.) et l'aspect technique (format de fichier etc.) servent surtout d'orientation ou recommandations aux utilisateurs potentiels.

M.2-5 Une spécification de la durée minimale pour la mise à disposition des documents publiés sur la plateforme, avec la garantie correspondante.

⇒ Le délai de conservation minimale peut varier d'un type de document à l'autre ou peut dépendre du contenu ou de la qualité technique du document. Dans tous les cas de figure, ce délai doit être d'au moins cinq ans (cf. critère 8 – *conservation pérenne*, section 2.8.).

M.2-6 Une déclaration concernant la conservation pérenne des documents.

⇒ Cette déclaration précisera les mesures techniques pour assurer la conservation pérenne des documents, par exemple par une coopération avec d'autres institutions.

M.2-7 Une explication du fonctionnement technique du service.

⇒ Qui est l'opérateur technique du site ? Quels sont les paramètres de performance du serveur, en particulier concernant l'accès et la mise à disposition ?

M.2-8 Une déclaration solennelle concernant le libre accès (Open Access).

⇒ Cette déclaration doit préciser la position du service sur le libre accès et indiquer, le cas échéant, si une partie de l'offre n'est pas accessible au sens du libre accès.

⇒ D'une façon générale, la plupart des documents doit être mise à disposition en libre accès.

⇒ Si l'institution (par exemple, l'université) a signé une déclaration sur le libre accès à l'information, la charte doit y faire référence.

Recommandations complémentaires

En plus, la charte contient des paragraphes sur les points suivants :

R.2-1 Des instructions ou recommandations concernant le libre accès pour les auteurs.

⇒ Un tel paragraphe est indiqué surtout si l'institution poursuit une politique particulière en matière de libre accès, par exemple en faveur de l'auto-archivage des publications par les auteurs (« voie verte »). Les instructions peuvent varier en fonction du type de document.

R.2-2 L'énumération et la description des outils de développement du site.

⇒ Par exemple, le logiciel d'archive de la plateforme, les interfaces de dépôt, les procédures d'authentification des documents (le cas échéant en plusieurs versions), l'attribution automatique d'une licence en cas de publication primaire, etc.

2.3 Aides aux auteurs et éditeurs

L'ensemble du processus de publication doit être accompagné d'une aide technique. Pour les auteurs et éventuellement aussi les éditeurs qui veulent publier sur le site, une information visible et structurée est importante. Elle doit apporter des réponses claires aux questions relatives à l'édition numérique. Les pages d'aide doivent être accessibles à partir de la page d'accueil. L'information peut également être diffusée dans d'autres formats (plaquettes, brochures etc.) ou sur d'autres sites⁶.

Conditions minimales

M.3-1 Il existe un service d'aide et d'assistance accessible à partir du site web.

⇒ Les moyens de contact peuvent être une adresse électronique (e-mail), un numéro de téléphone, un formulaire de contact sur les pages web etc. Il n'est pas nécessaire de proposer tous ces moyens mais un contact est obligatoire.

M.3-2 Les auteurs peuvent déposer leurs documents de façon autonome, par l'intermédiaire d'un formulaire web ou par un autre biais.

⁶ Cf. par exemple l'information de DissOnline, à l'adresse <http://www.dissonline.de>

⇒ Le dépôt par l'auteur n'est pas nécessaire si le dépôt des documents est organisé d'une autre manière, par exemple s'il est à la charge du personnel de la bibliothèque.

M.3-3 Le site contient une information pertinente sur tous les aspects techniques du dépôt. Sinon, il peut y avoir un lien vers cette information sur un autre site.

⇒ L'information concerne surtout les formats acceptés et la procédure de dépôt.

M.3-4 Le site contient des liens vers des pages d'information sur les aspects juridiques, notamment pour le cas du dépôt d'un document publié ailleurs (politique du libre accès).

⇒ Il s'agit notamment du site SHERPA/RoMEO, aussi dans sa version allemande <http://www.dini.de/oap/>.*

Recommandations complémentaires

R.3-1 L'interface de dépôt intègre l'API du site SHERPA/RoMEO.

⇒ Ainsi, les auteurs peuvent vérifier leurs droits au moment du dépôt d'un document déjà publié par un éditeur. On trouve l'information à l'adresse suivante <http://www.sherpa.ac.uk/romeo/api.html>

R.3-2 En complément à l'auto-archivage (dépôt par l'auteur), il existe un service central pour prendre en charge le dépôt pour l'auteur ou l'éditeur.

⇒ Ce service peut être assuré par l'institution du site ou par une bibliothèque. Il peut varier en fonction de la typologie de document.

R.3-3 Le site propose des modèles ou feuilles de style pour les publications scientifiques.

⇒ Ceci concerne des modèles pour des logiciels de traitement ou éditeur de textes (Open Office, Microsoft Word, LaTeX etc.).

R.3-4 Des formations sont régulièrement organisées pour les auteurs.

⇒ Ces formations porteront par exemple sur l'utilisation des feuilles de styles ou modèles, sur la citation des documents électroniques ou la conversion des documents selon les formats acceptés (PDF/A etc.).

R.3-5 Pour les éditeurs, des systèmes de production (workflow) sont proposés pour faciliter des projets éditoriaux d'envergure.

⇒ Ceci concerne surtout des systèmes d'évaluation dans le cadre d'une revue électronique ou d'une conférence scientifique.

R.3-6 Il existe un système d'assistance technique (helpdesk) pour répondre à des questions.

⇒ Il s'agit d'un service d'information qui gère les questions avec un « trouble-ticket system ».

R.3-7 Le site diffuse des conseils pour l'utilisation et la citation de documents électroniques.

⇒ Cette aide devrait par exemple expliquer pourquoi il vaut mieux citer un document électronique avec un identifiant pérenne (*persistent identifier*), ou comment citer des extraits d'une publication sans pagination.

R.3-8 L'information sur le site est au moins partiellement traduite en anglais.

⇒ Une telle traduction est recommandée surtout si le service souhaite s'adresser à des auteurs et/ou éditeurs dont la langue maternelle n'est pas l'allemand.*

2.4 Aspects juridiques

Pour diffuser et conserver les documents, l'institution qui gère le service a besoin de certains droits que les ayants-droits (le plus souvent, les auteurs ou éditeurs) doivent lui céder par un contrat appelé licence de dépôt (deposit licence). Par ailleurs, ce contrat doit exclure toute atteinte aux droits d'une tierce partie et dégager toute responsabilité de l'institution en cas de litige.

Le paragraphe concerne cet aspect et d'autres questions juridiques liées à la gestion d'un tel site. NB: Les explications ne sont pas à considérer comme conseil juridique avec valeur légale. Il est recommandé aux institutions de demander conseil à leur service juridique ou de chercher une assistance juridique externe.

Conditions minimales

M.4-1 La relation juridique entre les auteurs ou éditeurs (ayants-droits) et l'institution gestionnaire du site est réglée par un contrat formel (cession de droits).

⇒ Ce contrat contient la cession de droits et prend la forme d'une licence de dépôt (*deposit licence*). L'ayant-droit y cède des droits d'une manière non-exclusive à l'institution; celle-ci a besoin de ces droits pour gérer le service de dépôts et de publication en libre accès.

M.4-2 L'institution diffuse sa licence de dépôt en ligne en version allemande.*

⇒ La licence peut varier en fonction du type de document.

En signant la licence de dépôt, l'ayant-droits cède pour un document et ses métadonnées les droits suivants à l'institution :

M.4-3 Le droit de stockage électronique, en particulier dans des bases de données, le droit de mise à disposition, le droit de diffusion publique et d'envoi et, dans cette optique, le droit d'archivage et de reproduction.

⇒ Dans le cadre de cette mise à disposition, l'utilisateur est autorisé à se servir d'un document dans les limites du code de la propriété intellectuelle, sans frais; en particulier, il peut télécharger le document pour son usage privé, le stocker et l'imprimer en nombre limité (§53 UrhG).*

M.4-4 Le droit de signalement et de transfert à des tiers, par exemple dans le cadre d'une mission de collecte nationale, notamment dans le but d'un archivage pérenne.

⇒ Ceci concerne en particulier le transfert des métadonnées et documents à la Bibliothèque Nationale Allemande.*

M.4-5 Le droit de copie et de conversion vers d'autres formats numériques ou physiques pour assurer l'archivage pérenne, avec maintien de l'intégrité de l'œuvre.

⇒ Une telle conversion peut s'avérer nécessaire quand les formats de données du site deviennent obsolètes et ne peuvent plus être visualisés correctement par les logiciels de représentation ou de visualisation.

La licence de dépôt règle également les conditions de responsabilité, en particulier :

M.4-6 L'ayant-droit garantit à l'institution que le dépôt du document (ou de parties du document, par exemple images etc.) ne porte pas atteinte aux droits d'une tierce-partie (co-auteur, éditeur, organisme de financement etc.).

⇒ Quand il s'agit d'une copie d'auteur, cette question est normalement déjà réglée par le contrat d'auteur avec l'éditeur (cf. section 4.4.1).*

M.4-7 L'ayant-droit s'engage à informer l'institution immédiatement en cas de doute ou d'autres obstacles juridiques supposés ou réels.

⇒ Cette clause s'applique par exemple quand il s'avère à posteriori qu'un soi-disant ayant-droit n'a pas (plus) disposé des droits qu'il a cédés à l'institution, au moment de la cession.

M.4-8 L'ayant-droit dégage l'institution de toute revendication éventuelle de la part d'une tierce-partie.

⇒ Par exemple, de telles revendications peuvent résulter d'une utilisation illicite de matériel protégé dans le document déposé (images etc.).

D'autres conditions minimales :

M.4-9 Le site web du service contient un impressum en conformité avec la loi sur les médias électroniques allemande (TMG) et le cas échéant d'autres lois pertinentes.*

⇒ Ceci concerne surtout la législation des Länder.*

M.4-10 L'institution ajoute le statut légal d'un document publié aux métadonnées.

⇒ C'est-à-dire, pour chaque document l'organisme stocke les droits qui lui ont été cédés.

Recommandations complémentaires

- R.4-1 La licence de dépôt a été établie avec le service juridique de l'organisme ou avec l'aide d'un juriste.
- ⇒ Il s'agit d'une recommandation primordiale pour assurer la validité du contrat par rapport aux droits d'utilisation dont l'archive institutionnelle a besoin.
- R.4-2 L'organisme met en ligne une traduction anglaise de la licence de dépôt.
- ⇒ La plupart du temps, la version anglaise ne servira qu'à titre d'information tandis que la version allemande reste l'unique base légale du contrat.*
- R.4-3 Par ailleurs, le contrat autorise l'organisme à déléguer à une tierce partie la mise à disposition du document.
- ⇒ L'organisme a besoin de ce droit s'il veut un jour arrêter (partiellement ou entièrement) l'exploitation du service de dépôts et de publication mais souhaite continuer la mise à disposition des documents dont se chargerait une tierce partie, par exemple une institution spécialisée dans l'archivage pérenne.
- R.4-4 Lors du dépôt d'un document, l'auteur ou l'éditeur a le choix entre plusieurs licences.
- ⇒ Grâce à ces différentes licences, l'ayant-droit peut décider quels droits supplémentaires il souhaite céder à l'utilisateur, comme par exemple le droit à une exploitation commerciale. Ce choix devrait si possible s'appuyer sur des modèles standards, par exemple *Creative Commons* ou DPPL.*
- R.4-5 L'organisme contribue activement à la mise à jour de la base de données SHERPA/RoMEO.
- ⇒ Ceci concerne en particulier l'information sur la politique éditoriale de certains éditeurs dont l'organisme aurait eu connaissance.

2.5 Sécurité de l'information

Pour garantir le fonctionnement fiable d'une archive ouverte à la hauteur des exigences de la publication scientifique⁷, son système d'information⁸ et sa structure organisationnelle doivent répondre aux critères élémentaires de la sécurité d'information spécifiés par les *Critères Communs* de la norme internationale ISO/IEC 15408⁹, notamment la protection vis-à-vis de défaillances, la sécurité opérationnelle et la fiabilité de l'infrastructure technique ainsi que la disponibilité, l'intégrité et l'authenticité des documents publiés. Il faut donc protéger le site contre des attaques, abus et erreurs d'utilisation et contre des défaillances et erreurs techniques, par des mesures organisationnelles et techniques.

2.5.1 Système technique

Conditions minimales

M.5-1 Le système d'information du site est intégré dans le dispositif de sécurité de l'organisme.

⇒ Ce dispositif identifie et évalue les risques potentiels et décrit les mesures techniques, organisationnelles et personnelles pour y faire face d'une manière adéquate.

M.5-2 Un responsable de sécurité est nommé pour le système d'information.

⇒ Il s'agit d'une personne désignée par l'organisme; cet agent est l'interlocuteur central pour toutes les questions relatives à la sécurité du système d'information.

M.5-3 Il existe des spécifications fonctionnelles avec un plan de maintenance du système d'information.

⇒ Ces spécifications contiennent la description de l'ensemble des tâches, activités et processus nécessaires au bon fonctionnement du site, ainsi que les interfaces et rôles de chacun.

⁷ Disponibilité pérenne, non transformation et citabilité des documents ; fiabilité de l'ensemble du système technique.

⁸ Ceci inclut normalement le système de stockage, la base de données, les réseaux de communication, les serveurs web et toutes les autres applications qui font fonctionner le système.

⁹ Cf. <http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html>

M.5-4 Un document écrit fournit le détail des spécificités du système d'information avec toutes les composantes nécessaires à son bon fonctionnement.

⇒ Il n'est pas nécessaire de rendre cette documentation publique, en tout cas pas en intégralité. Les éléments en rapport avec la sécurité sont à gérer en interne.

M.5-5 Il existe des règles écrites sur l'accès au système d'information et à ses composantes.

⇒ Ces règles désignent les personnes autorisées à accéder au matériel informatique et/ou celles qui possèdent des droits d'utilisation (droits administratifs) élargis.

M.5-6 Tous les documents et données sont régulièrement sauvegardés (backup).

⇒ La périodicité des sauvegardes dépend surtout des mises à jour de la collection des données, par exemple, de la publication de nouveaux documents. Une sauvegarde quotidienne est conseillée.

M.5-7 Il existe une procédure en cas d'incident.

⇒ Cette procédure contient des préconisations en cas d'incident du système d'information.

M.5-8 La disponibilité des serveurs du site est régulièrement contrôlée par un logiciel de surveillance autonome.

⇒ Des procédures de documentation applicables pour des changements de configuration des logiciels et/ou matériels informatiques sont par exemple des systèmes logiciels autonomes de surveillance ou d'alarme ou des *ChangeLogs* ou journaux des modifications.

2.5.2 Traitement de documents

Conditions minimales

M.5-9 Les documents déposés sur le site ne sont plus modifiés.

⇒ C'est-à-dire, les changements de contenu sont réalisés dans une nouvelle version qui ne remplace pas l'originale et ne la rend pas inaccessible.

- M.5-10 Un identifiant pérenne est assigné à chaque document et à chaque nouvelle version.
- ⇒ Il peut s'agir par exemple d'un URN ou DOI.
- M.5-11 Les identifiants pérennes sont utilisés dans les métadonnées exportées et sur le site comme identifiant primaire sous forme d'une URL opérationnelle.
- ⇒ Pour cela, il faut ajouter comme préfixe à l'identifiant pérenne l'URL d'un service resolver. Concernant l'export des métadonnées, cf. aussi critère 6 – *Indexation et interfaces*, section 2.6.2., condition minimale M.6-7.
- M.5-12 La suppression d'un document reste exceptionnelle et doit être justifiée par écrit.
- ⇒ Une exception est par exemple quand la publication du document constitue un cas de litige au sens pénal.
- M.5-13 Lors de la mise en ligne d'un document déposé, le transfert de données se fait via le protocole SSL et sur la base d'un certificat digne de confiance (validé).
- ⇒ Cette condition est caduque si le site n'a pas d'interface de dépôt pour les auteurs, cf. critère 3 – *Aides aux auteurs et éditeurs*, section 2.3., condition minimale M.3.2.

Recommandations complémentaires

- R.5-1 L'intégrité des documents est assurée et prouvée par la création et publication d'une valeur de hash (somme de contrôle).
- ⇒ Il convient d'appliquer une méthode Hash sécurisée, par exemple SHA-1 ou RIPEMD 160.
- R.5-2 Le transfert de données et documents du serveur vers l'utilisateur se fait via le protocole SSL et sur la base d'un certificat digne de confiance (validé).
- ⇒ Ceci concerne l'accès pour la visualisation sur le serveur du site.

2.6 Indexation et interfaces

Pour trouver des documents publiés par voie électronique hors du système de production, il est crucial que ceux-ci soient indexés avec des métadonnées descriptives, et que ces métadonnées soient mises à disposition pour permettre des traitements automatiques. Ceci concerne surtout les services de référence ou de signalements ou encore d'autres services à valeur ajoutée pour lesquels une tierce partie exploite les données et documents de ces archives ouvertes. Par ailleurs, les options de recherche et les services supplémentaires proposées dans le système local font également partie intégrante des Services de Dépôts et de Publication en Libre Accès. Cette section décrit les conditions minimales en vue de répondre à ces exigences.

2.6.1 Indexation

Conditions minimales

- M.6-1 La politique et les règles d'indexation sont explicites et publiées sur le site à l'attention des auteurs, éditeurs et utilisateurs.
- ⇒ Par exemple, la procédure d'indexation – l'indexation est-elle réalisée par des indexeurs de la bibliothèque ou par les auteurs ou est-elle automatique etc. Ces règles d'indexation peuvent varier selon le type de document.
- M.6-2 Chaque document est représenté et indexé selon des moyens et méthodes du Dublin Core Element Set.
- ⇒ Par contre, il n'est pas nécessaire que les données soient aussi stockées en interne dans ce même format.
- M.6-3 Chaque document est indexé en fonction de son domaine (sujet ou thématique), soit avec des mots-clés libres, soit à l'aide d'un vocabulaire contrôlé ou système de classification.
- ⇒ Les mots-clés peuvent être assignés directement par les auteurs eux-mêmes.
- M.6-4 Chaque document est indexé selon la classification Dewey (DDC), au moins au niveau des groupes thématiques de la Bibliographie Nationale Allemande.*
- ⇒ Cf. <http://www.ddc-deutsch.de/anwendung/dnb.htm> et section A.2.2.*

M.6-5 A chaque document est assigné un type de document ou de publication en utilisant les recommandations du DINI mentionnées dans le vocabulaire commun pour les Types de Documents ou de Publications du DINI (DINI-Empfehlungen *Gemeinsames Vokabular für Publikations- und Dokumenttypen*).*

⇒ Cf. <http://nbn-resolving.de/urn:nbn:de:kobv:11-100109998> et section A.2.3.*

Recommandations complémentaires

R.6-1 D'autres systèmes d'indexation normalisés sont proposés, spécialisés.

⇒ Par exemple, SWD (Schlagwortnormdatei), LoC Subject Headings, CCS (ACM Computing Classification System), MSC (Mathematics Subject Classification) et PACS (Physics and Astronomy Classification Scheme).*

R.6-2 De plus, des mots clés en anglais sont assignés aux documents.

⇒ Ces mots-clés peuvent être assignés directement par les auteurs.

R.6-3 Par ailleurs sont proposés des synthèses ou des résumés en allemand et en anglais.

⇒ Les synthèses ou résumés peuvent être demandés aux auteurs ou extraits du texte intégral des documents.

2.6.2 Interfaces pour les métadonnées

Conditions minimales

M.6-6 Une interface Web, proposée aux utilisateurs finaux, donne accès aux documents et aux métadonnées.

⇒ Cette interface donne accès à la totalité de la collection de l'archive ouverte.

M.6-7 Il existe une interface OAI compatible avec le protocole OAI-PMH 2.0 et avec les recommandations OAI du réseau DINI.*

⇒ La description de l'interface OAI se trouve en annexe A de ce document.

Recommandations complémentaires

R.6-4 Les métadonnées sont mises à disposition dans d'autres formats via l'interface OAI, éventuellement uniquement pour une partie de la collection.

⇒ Ces formats peuvent être spécifiques pour un type de documents ou un domaine scientifique, destinés à faciliter des services complémentaires par une tierce partie. Un exemple : le format XMetaDissPlus pour le transfert des métadonnées de thèses électroniques à la Bibliothèque Nationale Allemande.*

R.6-5 Le site web permet l'export direct des métadonnées des documents ou des résultats de recherche dans différents formats.

⇒ Parmi des exemples de formats d'export, on peut citer : BibTex, EndNote, COinS etc. Cette option facilite l'importation des données dans des logiciels de gestion de références bibliographiques.*

R.6-6 Il existe des listes de liens vers tous les documents.

⇒ Cette fonction permet l'indexation complète des données par des robots de moteurs de recherche d'autres services.

R.6-7 Les métadonnées sont accessibles aux publics via d'autres interfaces.

⇒ Par exemple, SRU/W ou des API spécifiques.*

2.7 Statistiques d'utilisation

Les statistiques d'utilisation peuvent contribuer d'une manière quantitative, qualitative ou technologique à l'évaluation d'un site. Elles peuvent aussi refléter l'impact d'un document individuel, soit comme facteur d'usage complémentaire à d'autres facteurs d'impact, soit comme indice de prévision du taux de citation. Par ailleurs, ces statistiques peuvent contribuer à mesurer les cycles de vie de l'information scientifique, aussi par discipline, et alimenter les analyses scientométriques.

Conditions minimales

M.7-1 Le service de dépôts et de publication en libre accès collecte des statistiques d'utilisation du site d'une façon consistante et en conformité avec la loi.

⇒ Il s'agira normalement d'un fichier log du serveur.

M.7-2	Les fichiers log du site sont rendus anonymes (ou pseudonymes) pour un stockage à plus long terme.
	⇒ Ceci doit être conforme à la loi (§15 Abs. 3 et §13 Abs. 1 TMG).*
M.7-3	L'accès automatique au site, aux documents ou données du site est filtré.
	⇒ Ceci peut se faire de plusieurs manières : exploitation du champ utilisateur-agent du fichier log, comparaison des accès avec le robot.txt, utilisation de listes de robots connus, ou d'autres méthodes heuristiques.
	⇒ Cette condition s'applique seulement si les statistiques sont rendues publiques.
M.7-4	Une documentation publique renseigne les critères et normes appliqués lors de la création des statistiques d'utilisation.
	⇒ Ceci concerne notamment les normes COUNTER ¹⁰ , LogEC ¹¹ et IFABC ¹² . Si l'organisme publie d'autres statistiques d'utilisation non conformes avec ces normes, le site doit informer que ces statistiques ne sont pas comparables avec celles d'autres sites, notamment quand il s'agit de statistiques d'utilisation d'un document individuel.
	⇒ Cette condition s'applique seulement si les statistiques sont rendues publiques.

Recommandations complémentaires

R.7-1	Les statistiques d'utilisation sont ajoutées comme métadonnée dynamique à chaque document ; elles sont accessibles en ligne.
	⇒ Par exemple, les statistiques mensuelles pourraient figurer sur la page d'accueil de chaque document.
R.7-2	L'utilisation des documents est mesurée selon une des normes recommandées par DINI.*

¹⁰ COUNTER = Counting Online Usage of Electronic Resources, cf. <http://www.projectcounter.org/>

¹¹ Cf. <http://logec.repec.org/>

¹² Cf. <http://www.ifabc.org/>

⇒ Il s'agit de COUNTER, LogEC et des recommandations de l'IFABC. Cf. DFG-Projekt Open-Access-Statistik (OA-S) und DINI Arbeitsgruppe Elektronisches Publizieren: Nutzungsstatistiken elektronischer Publikationen, DINI Schriften. <http://nbn-resolving.de/urn:nbn:de:kobv:11-100101174> *

R.7-3 L'organisme soutient le transfert des statistiques vers un site fédérateur semblable à celui développé par le projet Open-Access-Statistik.*

⇒ Les données brutes sont mises à disposition via une interface OAI afin de rendre possible le moissonnage des statistiques d'utilisation par un service externe (cf. <http://www.dini.de/projekte/oa-statistik/>).*

2.8 Archivage pérenne

Ce certificat concerne des services de dépôts et de publication en libre accès et non pas des archives numériques à long terme qui font l'objet du *Catalogue de Critères d'Archives Numériques Fiables à Long Terme* de nestor¹³. Cependant, certaines questions sur l'archivage pérenne concernent aussi ces services dans la mesure où les documents déposés sont souvent transférés vers une archive numérique à long terme pour un archivage pérenne, et qu'il faut créer des conditions techniques appropriées.

Conditions minimales

M.8-1 Le service précise la durée minimale de l'archivage et de la mise à disposition des documents et métadonnées publiés sur ce site. Cette durée est d'au moins cinq ans.

⇒ Cette spécification doit faire partie de la charte du site, cf. critère 2 – *Charte*, section 2.2, condition minimale M.2-5. Les délais de conservation ne sont pas nécessairement les mêmes pour tous les types de documents.

M.8-2 Les fichiers déposés ainsi que leurs copies d'archivage ne sont pas protégés par des mesures techniques de protection.

¹³ Pour l'actuelle version 2 du catalogue cf. <http://nbn-resolving.de/urn:nbn:de:0008-2008021802>

⇒ Ceci concerne notamment les mesures telles que les DRM, une protection par mot de passe et des restrictions d'utilisation d'un document (copier-coller, imprimer etc.). Ces mesures de protection sont exclues car elles empêchent potentiellement les stratégies d'archivage pérenne (migration, émulation).

Recommandations complémentaires

R.8-1 La mise à disposition des documents à long terme est assurée.

⇒ Ceci peut se faire par exemple par une coopération avec une institution gestionnaire d'archives.

R.8-2 La conservation des documents se fait sous des formats libres et pertinents.

⇒ Il s'agit par exemple des formats PDF/A, ODF, TXT, TEX.

R.8-3 Il existe un règlement pour la suppression d'un document.

⇒ Ce règlement précise les conditions de suppression d'un document ainsi que la procédure technique et éventuellement les données à conserver au-delà de la date de suppression. Ce règlement fait partie de la charte du site, cf. critère 2, section 2.2.

3 Certification et évaluation

Le « Certificat DINI – Services de Dépôts et de Publication en Libre Accès » est décerné par le bureau de l'Initiative Allemande pour l'information en Réseau (Deutsche Initiative für Netzwerkinformation ou DINI) ou un groupe de travail mandaté par le DINI. Le certificat intègre la date (année) de la version. Il atteste de la conformité aux exigences minimales pour les service de dépôts et de publication en libre accès.

Des frais sont exigés pour la délivrance du certificat DINI :

1. Organismes sans but lucratif
 - Membres DINI 50,00 €
 - Autres 100,00 €
2. Établissements à caractère commercial
 - Membres DINI 150,00 €
 - Autres 250,00 €

Le prestataire de service fait sa demande de certification auprès du DINI en complétant un formulaire en ligne sur le site du DINI¹⁴. Ce formulaire a la structure d'une fiche de contrôle (checklist) et contient les exigences minimales ainsi que les recommandations prévues dans la section 2 du présent document. En remplissant le formulaire, le prestataire déclare que le service remplit les critères du certificat DINI et ce jusqu'à quel niveau. En outre, des explications et précisions peuvent être ajoutées dans des champs dédiés sur le formulaire ainsi que les adresses URL ou d'autres options sur la façon et le lieu d'obtention d'informations complémentaires.

Une fois le formulaire en ligne de demande de certification complété et soumis, les données seront vérifiées, généralement par deux examinateurs nommés à cet effet. L'accès au service pour lequel une demande de certification est faite, doit être autorisé pour ces deux examinateurs. Le prestataire de service doit être prêt à répondre aux questions des examinateurs. Des visites sur site peuvent être requises à titre exceptionnel. Les coûts supplémentaires qui pourraient apparaître au cours du processus de certification devra être couvert par le demandeur. DINI l'informerá au préalable de coûts additionnels possibles.

Le processus de certification doit être généralement achevé dans les deux mois qui suivent la demande. La durée du processus de certification dépend en partie de la rapidité avec laquelle le fournisseur répond aux questions que les examina-

¹⁴ Voir <http://www.dini.de/dini-zertifikat/fragebogen/>

teurs pourraient être amenés à poser. Le processus peut être plus long si l'un des critères ou plusieurs ne sont pas renseignés.

Un certificat délivré n'est pas limité dans le temps et conserve sa validité même après la publication d'une nouvelle version du certificat. A partir de la date de la version, on peut à tout moment vérifier avec quels critères le service en question est en conformité. Si jamais des violations de ces critères sont constatées ultérieurement, DINI peut être amené à révoquer le certificat.

Le prestataire d'un service certifié a le droit de l'appeler un « service de dépôts et de publication en libre accès certifié par le DINI » et peut l'indiquer par le logo officiel du certificat sur sa page d'accueil ou d'une autre façon. Toute utilisation abusive de l'appellation ou du logo sera poursuivie selon le Code de la Propriété Intellectuelle.

Après la publication du présent document, on peut demander une certification uniquement pour la version 2010.

4 Explications et exemples

Le quatrième chapitre contient des explications complémentaires avec des exemples d'application pour les huit critères décrits plus haut. Ces explications servent uniquement d'illustration sans constituer d'autres critères ou conditions et ne figurent donc pas dans le questionnaire.

Note des traducteurs : Le chapitre original compte plus de vingt pages, étroitement liées au contexte allemand, où le DINI joue son rôle de conseil technique pour la mise en place d'une infrastructure d'archives ouvertes et de plateformes de revues en libre accès en Allemagne. Cette partie est donc moins intéressante pour un lecteur français. Par conséquent, nous nous sommes contentés d'en résumer l'essentiel en quelques mots.

4.1 Visibilité du site

Le réseau DINI insiste sur l'intérêt d'une bonne visibilité du service au sein de l'institution ou de la communauté scientifique (auteurs) ainsi que par rapport au public cible (lecteurs). DINI conseille l'enregistrement dans des répertoires et sites comme DRIVER, Open Archives Initiative (OAI) ou OpenDOAR. Afin de faciliter l'indexation par des moteurs de recherche et d'autres services commerciaux, tous les documents et leurs pages d'accès devraient disposer de leur propre adresse URL (lien) de l'interface OAI.

4.2 Charte

Une charte devrait être un texte rédigé sous forme d'un auto-engagement explicite du prestataire. Une liste de questions-réponses (FAQ) ne suffit pas mais pourrait compléter la charte.

4.3 Aide aux auteurs et éditeurs

L'argument de DINI est simple et direct. L'objectif est d'inciter le plus grand nombre d'auteurs et d'éditeurs d'une institution à déposer et utiliser le service de dépôts et de publication en libre accès. En conséquence, il faut donc créer et offrir des services complémentaires qui répondent au mieux aux besoins et demandes de ces deux groupes. Ceci concerne en particulier les éditeurs qui d'après le DINI joueront de plus en plus un rôle d'intermédiaire entre le prestataire du service et les auteurs. Est ainsi proposé d'étendre les services intégrés au service de dépôts et de publication, en soutenant la création de publications ou l'organisation de projets de publication (structuration de documents numériques, modèles de documents, workflow,...).

4.4 Aspects juridiques

Le texte reprend trois aspects du certificat : les droits d'auteur par rapport aux contrats de cession avec les éditeurs, le risque d'une atteinte aux droits d'une tierce-partie et le cas particulier d'une publication hybride où le fichier en libre accès cohabite avec une version imprimée diffusée par un éditeur. Le texte fournit plusieurs exemples d'une licence de dépôt et finit par quelques réflexions sur l'évolution du droit d'auteur, accompagnées d'une bibliographie sélective.

4.5 Sécurité de l'information

Les auteurs de DINI insistent sur la documentation, les règles d'accès au système (accès aux serveurs, administration, responsabilités, gestion de l'authentification...), la sauvegarde (backup), l'anticipation (procédure) en cas d'un incident majeur, et l'utilisation d'un logiciel de surveillance. Concernant le traitement des documents, ils insistent sur la nécessité d'assigner un identifiant pérenne (persistant identifier) à chaque ressource et préconisent, sans le rendre obligatoire, le standard URN associé au système de résolveur de la bibliothèque nationale allemande. Ils reviennent sur les méthodes de hachage (Hash) pour assurer l'intégrité et l'authenticité pérenne des documents déposés.

4.6 Indexation et interfaces

DINI recommande fortement une indexation (mot-clés et résumé) transparente, explicite et contrôlée de tous les documents mis en ligne par ces services, que celle-ci soit réalisée par les auteurs ou le personnel de la bibliothèque, en conformité avec les normes et le cas échéant en lien avec les besoins d'une communauté spécifique. En ce qui concerne l'interface OAI, DINI suggère de proposer plusieurs formats, en complément au format du Dublin Core non-qualifié.

4.7 Statistiques d'utilisation

Ici, le document renvoie vers un projet financé par l'ANR allemande (DFG) qui avait pour but de créer une infrastructure pour les statistiques de plusieurs archives ouvertes et dont nous avons parlé ailleurs (Open-Access-Statistik).¹⁵

4.8 Archivage pérenne

Le minimum est une garantie de cinq ans, avec toutes les conditions de sécurité et sans aucune mesure technique de protection (type DRM). Les auteurs enchaînent

¹⁵ J. Schöpfel & H. Prost (2009). 'Les statistiques d'utilisation d'archives ouvertes. Etat de l'art'. In *Ressources électroniques académiques: mesures et usages. Colloque international. Lille, 26-27 novembre 2009.*

avec des conseils pour le choix des formats d'archivage et pour une définition claire quand un document peut être supprimé.

Annexe A Conditions pour l'interface OAI

Depuis sa publication en 2001, le protocole OAI s'est imposé comme la norme pour l'échange automatique et asynchrone de métadonnées bibliographiques entre archives ouvertes et fournisseurs de service (moissonneurs). Dans ce cadre, l'interface OAI désigne une composante logicielle qui fonctionne comme fournisseur de données au sens du protocole, c'est-à-dire elle donne accès aux métadonnées en réponse aux requêtes formulées par les fournisseurs de service. Une telle interface OAI permet aujourd'hui de gérer des archives ouvertes¹⁶ et d'autres systèmes de gestion de métadonnées¹⁷.

Le protocole OAI propose une interopérabilité assez simple d'utilisation ce qui a facilité sa large diffusion et l'adhésion de tous en très peu de temps. Cependant, l'absence d'exigences réduit l'intérêt pour les fournisseurs de service qui utilisent le protocole OAI pour la collecte de données puisque le protocole ne spécifie pas assez précisément la structure et la qualité des métadonnées.

Les métadonnées doivent être fournies au format standard *Dublin Core Simple* avec 15 éléments optionnels qui peuvent être omis ou répétés. Les recommandations concernant la structure interne des éléments¹⁸ sont elles aussi non-obligatoires. Le protocole prévoit un mécanisme appelé sets pour différencier ou structurer l'ensemble des données en lots ou regroupements ce qui permet entre autre un moissonnage sélectif. Néanmoins, la définition concrète et la dénomination de ces sets restent à la discrétion des fournisseurs de données.

Pour mettre en œuvre des services de qualité¹⁹ à partir du moissonnage des données au format du protocole OAI, il faudra des spécifications supplémentaires au protocole. Ces spécifications seront décrites dans les annexes A.2 et A.3 ; elles portent surtout sur la définition de la structure des sets et sur le contenu des différents éléments du Dublin Core. A ceci s'ajoutent quelques conditions en A.1 qui découlent directement des spécifications du protocole.

Comme pour les autres critères du certificat DINI, nous avons séparé conditions minimales et recommandations complémentaires qui ne sont pas nécessaires

¹⁶ Comme par exemple OPUS, MyCoRe, ePrints et DSpace.

¹⁷ Comme par exemple les SIGB ou les systèmes d'édition de revues (*Open Journal Systems OJS* etc.).

¹⁸ Pour le formatage de la date ou pour le codage de la langue, etc.

¹⁹ En premier lieu, des services de référencement avec des fonctionnalités de recherche et de feuilletage.

pour la certification DINI. Néanmoins, ces recommandations correspondent à de bonnes pratiques. La mise en place d'une interface OAI devrait en tenir compte afin d'améliorer la qualité des métadonnées et optimiser leur utilisation.

Cet annexe au certificat DINI est en même temps un développement des recommandations de 2005 sur la création de l'interface OAI²⁰ et les remplace. Par ailleurs, le certificat s'inspire des *guidelines*²¹ du projet européen DRIVER²²; les deux sont compatibles. Et comme l'ensemble du certificat DINI, l'annexe concerne en priorité des documents textuels et se limite aux métadonnées au format *Dublin Core Simple* (oai_dc).

A.1 Conformité avec le protocole

L'échange de données via l'interface OAI doit être conforme au protocole OAI, c'est-à-dire l'interface doit correspondre aux spécifications de l'*OAI Protocol for Metadata Harvesting* (OAI-PMH) dans sa version actuelle 2.0²³. Le contrôle de conformité peut se faire de plusieurs manières²⁴. Il est notamment effectué au moment où l'interface OAI est enregistrée auprès de l'OAI comme fournisseur de données (*data provider*).

Voici quelques critères qui concernent toutes les interfaces OAI. Comme ils posent régulièrement des problèmes, il faut leur prêter une attention particulière.

Conditions minimales

M.A.1-1 L'interface OAI fonctionne en conformité avec les spécifications de la version 2.0 du protocole.

⇒ Tout ce qui suit découle de cette première condition.

M.A.1-2 L'interface OAI est constamment accessible à l'URL de base et dispose d'une performance suffisante.

²⁰ Cf. <http://nbn-resolving.de/urn:nbn:de:kobv:11-10049220>

²¹ La version actuelle 2.0 se trouve ici http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_2008-11-13.pdf

²² DRIVER = Digital Repository Infrastructure Vision for European Research, cf. <http://www.driver-repository.eu/>

²³ Le détail des spécifications se trouve ici <http://www.openarchives.org/OAI/openarchivesprotocol.html>

²⁴ Il s'agit surtout du Repository Explorer (cf. <http://re.cs.ucl.ac.za/>) et du DRIVER Validator (cf. <http://validator.driver.research-infrastructures.eu/>). Ce dernier vérifie en plus le respect des DRIVER Guidelines (cf. http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_2008-11-13.pdf)

⇒ Ceci est indispensable pour une utilisation fiable de l'interface par les fournisseurs de service et limite les problèmes de communication, en particulier l'interruption des opérations de moissonnage.

M.A.1-3 L'interface OAI fournit toutes les réponses en format XML structuré et valide par rapport au schéma des spécifications OAI ou à d'autres schémas de métadonnées.

⇒ Les problèmes régulièrement rencontrés concernent l'encodage des caractères, s'il y a des caractères spéciaux dans les métadonnées. On peut également être confrontés à des messages d'erreurs dans le flux XML de la base de données ou de l'application.

M.A.1-4 L'interface OAI supporte correctement le moissonnage incrémental.

⇒ L'élément *timestamp* de chaque ensemble de données doit indiquer la date de la création ou de la dernière mise à jour des métadonnées et non pas, la date de publication du document.

⇒ Ainsi, les fournisseurs de service peuvent régulièrement mettre à jour leurs données sans être obligés de moissonner toutes les métadonnées. Ceci implique aussi que le fournisseur de données supporte les paramètres *from* et *until* des requêtes OAI *ListRecords* et *ListIdentifiers* et qu'il fournit les lots de données correspondants, selon une périodicité au moins quotidienne (YYYY-MM-DD).

M.A.1-5 L'interface OAI utilise les informations des sets d'une forme cohérente.

⇒ Cela veut dire en particulier que tous les sets avec des notices soient effectivement fournis en réponse à la requête *ListSets*, et que toutes les notices fournies en réponse à une requête *ListRecords* ou *ListIdentifiers* qualifiées par le paramètre *set* appartiennent réellement au lot concerné du fait de leur *header information*.

Recommandations complémentaires

R.A.1-1 L'organisme procède régulièrement à des contrôles manuels et automatiques de l'interface OAI.

⇒ Cf. note 24 en bas de page 36.

R.A.1-2 Les modifications importantes de l'interface OAI sont communiquées aux répertoires dans lesquels l'interface ou le service ont été enregistrés auparavant.

⇒ Ainsi, un fournisseur de service a la possibilité de réagir efficacement face aux changements. Des modifications importantes dans ce contexte sont par exemple, l'installation d'une nouvelle version, le changement de l'URL de base ou le remplacement du système de gestion du service de dépôts et de publication en libre accès.

⇒ Pour les registres, voir aussi le critère 1 – *Visibilité de l'offre globale* dans la partie 2.1.

R.A.1-3 La réponse à la requête OAI *Identify* fournit une information exhaustive sur le service.

⇒ Il s'agit notamment d'une adresse électronique email valide de l'administrateur (élément *adminEmail*) et une courte description du service (élément *description*).

R.A.1-4 L'élément provenance est utilisé dans le container *about* pour les métadonnées fournies en réponse aux requêtes *ListRecords* et *GetRecord*.

⇒ Ceci permet de fournir une information complémentaire sur l'origine des métadonnées. Cf. aussi <http://www.openarchives.org/OAI/2.0/guidelines-provenance.htm>.

R.A.1-5 L'information descriptive des réponses OAI est en anglais.

⇒ Il s'agit par exemple des éléments de la réponse à la requête *Identify* ou de la description des sets avec l'élément *setName* dans la réponse à la requête *ListSets*.

A.2 OAI-PMH: Conditions supplémentaires

Les conditions supplémentaires concernent essentiellement la structure de l'ensemble (*set*) des métadonnées fournies (sections A.2.1 à A.2.4). Cette structure apporte de l'information normalisée supplémentaire aux différents documents et permet des requêtes sélectives, ce qui améliore l'interopérabilité avec les fournisseurs de services. Les autres sections contiennent des recommandations pour le traitement des documents ou données supprimés (A.2.5) et pour le contrôle des flux (A.2.6).

A.2.1 Données pour les documents en libre accès

Les documents en libre accès mis à part, les services de dépôts et de publication publient souvent aussi d'autres documents qui ne sont pas librement accessibles, qui sont par exemple réservés à un groupe d'utilisateurs restreint. Pour que les fournisseurs de service puissent assurer une prestation de service à valeur ajoutée, il est important de pouvoir distinguer les données des documents qui sont en libre accès de celles qui ne le sont pas et en plus, de permettre une sélection en amont.

Condition minimale

M.A.2-1 Il existe un ensemble de données (*set*) appelé (*setSpec*) 'open_access'. Cet ensemble regroupe toutes les métadonnées des documents en libre accès.

⇒ Cette recommandation vaut aussi pour les services où tous les documents sont publiés en libre accès. Dans ce cas, toutes les métadonnées feraient partie de cet ensemble.

A.2.2 Données pour les groupes DDC*

En Allemagne, pour effectuer l'indexation thématique générale des métadonnées ou documents, il est d'usage de s'appuyer sur les groupes thématiques de la bibliographie nationale allemande utilisés par la Bibliothèque Nationale d'Allemagne. Ces groupes s'inspirent de la *Dewey Decimal Classification* (DDC) dont ils reprennent surtout les deux premiers chiffres²⁵. Si un fournisseur de service qui utilise le protocole OAI souhaite effectuer une première sélection en fonction du sujet, il faut que les groupes thématiques de l'indexation des documents correspondent à la structure des données de l'interface OAI.

²⁵ Cf. <http://www.ddc-deutsch.de/anwendung/dnb.htm>*

Condition minimale

M.A.2-2 La structure de l'ensemble de données (*set*) est conforme au tableau 1. Toutes les données correspondant à un document sont attribuées à un *setSpec* selon le tableau suivant.

⇒ Une notice peut être caractérisée par une ou plusieurs classes DDC.

Tableau 1 : Intitulé et description des sets pour l'indexation thématique*

setSpec	setName	Description française
ddc:000	Generalities, Science	Généralités, science
ddc:004	Data processing Computer science	Traitement des données, informatique
ddc:010	Bibliography	Bibliographie
ddc:020	Library & information sciences	Bibliothéconomie et sciences de l'information
ddc:030	General encyclopedic works	Encyclopédies générales
ddc:050	General serials & their indexes	Publications en série d'ordre général
ddc:060	General organization & museology	Organisations générales et muséologie
ddc:070	News media, journa- lism, publishing	Médias documentaires, journalisme, édition
ddc:080	General collections	Recueils généraux
ddc:090	Manuscripts & rare books	Manuscrits, livres rares, autres documents imprimés rares
ddc:100	Philosophy	Philosophie
ddc:130	Paranormal pheno- mena	Parapsychologie et occultisme
ddc:150	Psychology	Psychologie
ddc:200	Religion	Généralités sur les religions
ddc:220	Bible	Bible
ddc:230	Christian theology	Christianisme, théologie chrétienne
ddc:290	Other & comparative religions	Autres religions

setSpec	setName	Description française
ddc:300	Social sciences	Sciences sociales
ddc:310	General statistics	Statistiques générales
ddc:320	Political science	Science politique
ddc:330	Economics	Economie
ddc:333.7	Natural resources, energy and environment	Ressources naturelles, énergie et environnement
ddc:340	Law	Droit
ddc:350	Public administration	Administration publique
ddc:355	Military science	Science militaire
ddc:360	Social services; association	Problèmes et services sociaux. Associations
ddc:370	Education	Education
ddc:380	Commerce, communications, transport	Commerce, communication, transport
ddc:390	Customs, etiquette, folklore	Coutumes, savoir-vivre, folklore
ddc:400	Language, Linguistics	Langues, Linguistique
ddc:420	English	Langues anglaises. Anglo-saxon
ddc:430	Germanic	Langues germaniques. Allemand
ddc:439	Other Germanic languages	Autres langues germaniques
ddc:440	Romance languages French	Langues romanes. Français
ddc:450	Italian, Romanian, Rhaeto-Romantic	Langues italienne, sarde, dalmate, roumain, langues rétho-romanes
ddc:460	Spanish & Portuguese languages	Langues espagnole et portugaise
ddc:470	Italic Latin	Langues italiennes. Latin
ddc:480	Hellenic languages Classical Greek	Langues helléniques. Grec
ddc:490	Other languages	Autres langues
ddc:500	Natural sciences & mathematics	Sciences de la nature et Mathématiques
ddc:510	Mathematics	Mathématiques

setSpec	setName	Description française
ddc:520	Astronomy & allied sciences	Astronomie et sciences connexes
ddc:530	Physics	Physique
ddc:540	Chemistry & allied sciences	Chimie et sciences connexes
ddc:550	Earth sciences	Sciences de la terre
ddc:560	Paleontology Paleozoology	Paléontologie, Paléozoologie
ddc:570	Life sciences	Sciences de la vie. Biologie
ddc:580	Botanical sciences	Plantes. Botanique
ddc:590	Zoological sciences	Animaux. Zoologie
ddc:600	Technology (Applied sciences)	Technologie (Sciences appliquées)
ddc:610	Medical sciences; Medicine	Sciences médicales
ddc:620	Engineering & allied operations	Art de l'ingénieur et activités connexes
ddc:630	Agriculture	Agronomie, agriculture et activités connexes
ddc:640	Home economics & family living	Économie domestique (arts ménagers). Vie familiale
ddc:650	Management & auxiliary services	Gestion et services auxiliaires
ddc:660	Chemical engineering	Génie chimique et techniques connexes
ddc:670	Manufacturing	Fabrication industrielle
ddc:690	Buildings	Bâtiments
ddc:700	The arts	Arts
ddc:710	Civic & landscape art	Urbanisme. Art du paysage
ddc:720	Architecture	Architecture
ddc:730	Plastic arts Sculpture	Arts plastiques. Sculpture
ddc:740	Drawing & decorative arts	Dessin. Arts décoratifs
ddc:741.5	Comics, Cartoons	Bandes dessinées, Dessins animés
ddc:750	Painting & paintings	La peinture et les peintures

setSpec	setName	Description française
ddc:760	Graphic arts Printmaking & prints	Arts graphiques. Gravures
ddc:770	Photography & photo- graphs	Photographie et photographies
ddc:780	Music	Musique
ddc:790	Recreational & performing arts	Loisirs et arts du spectacle
ddc:791	Public performances	Représentations scéniques, Spectacles
ddc:792	Stage presentations	Théâtre, Représentations sur scène
ddc:793	Indoor games & amu- sements	Jeux d'intérieur
ddc:796	Athletic & outdoor sports & games	Sports et jeux athlétiques en plein air
ddc:800	Literature & rhetoric	Littérature (Belles-Lettres) et techniques d'écriture
ddc:810	American literature in English	Littérature américaine en anglais
ddc:820	English & Old English literatures	Littérature anglaise et anglo-saxonne
ddc:830	Literatures of Germanic languages	Littérature des langues germaniques. Littérature allemande
ddc:839	Other Germanic literatures	Autres littératures germaniques
ddc:840	Literatures of Romance languages	Littérature des langues romanes. Littérature française
ddc:850	Italian, Romanian, Rhaeto-Romanic literatures	Littératures italienne, roumaine, rhéto-romane
ddc:860	Spanish & Portuguese literatures	Littératures espagnole et portugaise
ddc:870	Italic literatures Latin	Littératures des langues italiques. Littérature latine
ddc:880	Hellenic literatures Classical Greek	Littératures des langues helléniques. Littérature grecque
ddc:890	Literatures of other languages	Littératures des autres langues
ddc:900	Geography & history	Géographie, histoire et disciplines auxiliaires.

setSpec	setName	Description française
ddc:910	Geography & travel	Géographie et voyages
ddc:914.3	Geography & travel Germany	Géographie et voyages. Allemagne
ddc:920	Biography, genealogy, insignia	Biographie, généalogie, emblèmes, insignes
ddc:930	History of the ancient world	Histoire générale du monde ancien
ddc:940	General history of Europe	Histoire générale de l'Europe
ddc:943	General history of Europe Central Europe Germany	Histoire générale de l'Europe. Europe central. Allemagne
ddc:950	General history of Asia Far East	Histoire générale de l'Asie, Orient, Extrême-Orient
ddc:960	General history of Africa	Histoire générale de l'Afrique
ddc:970	General history of North America	Histoire générale de l'Amérique du Nord
ddc:980	General history of South America	Histoire générale de l'Amérique du Sud
ddc:990	General history of other areas	Histoire générale des autres parties du monde, des mondes extra-terrestres

A.2.3 Données pour les publications et autres types de documents

Le type de document est une métadonnée importante. Si un fournisseur de service souhaite moissonner les données de l'ensemble des documents d'un même type (par exemple, des thèses), le fournisseur de données doit proposer une telle structure de données. Cette structure s'appuie sur le vocabulaire développé pour le format de métadonnées XMetaDissPlus et dans le certificat DINI, il est publié sous la rubrique *Recommandations DINI Vocabulaire commun pour types de publication et document*²⁶.

²⁶ Cf. <http://nbn-resolving.de/urn:nbn:de:kobv:11-100109998> L'écriture incohérente des majuscules et minuscules des termes de sets (setSpec) est liée aux sources d'origine du vocabulaire (Dublin Core Type Vocabulary, Publication Type Vocabulary des DRIVER-Guidelines) et a été préservée pour des raisons de compatibilité.

Condition minimale

M.A.2-3 Une structure de données (set) conforme au tableau 2 regroupe toutes les métadonnées relatives à la typologie du document.

⇒ Suivant les explications des Recommandations DINI *Vocabulaire commun pour types de publication et document*, un document peut être caractérisé sous plusieurs types de document (cf. exemple 1).*

Tableau 2 : Intitulé et description des sets pour la structure de forme*

setSpec	setName	Description française
doc-type:preprint	Preprint	Prépublication
doc-type:workingPaper	WorkingPaper	Travail non publié
doc-type:article	Article	Article scientifique
doc-type:contributionToPeriodical	ContributionToPeriodical	Contribution dans une revue
doc-type:PeriodicalPart	PeriodicalPart	Section d'une périodique
doc-type:Periodical	Periodical	Périodique
doc-type:book	Book	Livre, monographie
doc-type:bookPart	BookPart	Chapître
doc-type:Manuscript	Manuscript	Manuscrit
doc-type:StudyThesis	StudyThesis	Thèse d'étude
doc-type:bachelorThesis	BachelorThesis	Mémoire de licence
doc-type:masterThesis	MasterThesis	Mémoire de Master
doc-type:doctoralThesis	DoctoralThesis	Thèse de doctorat
doc-type:conferenceObject	ConferenceObject	Conférence
doc-type:lecture	Lecture	Cours
doc-type:review	Review	Note de lecture
doc-type:annotation	Annotation	Annotation
doc-type:patent	Patent	Brevet
doc-type:report	Report	Rapport
doc-type:MusicalNotation	MusicalNotation	Partition
doc-type:Sound	Sound	Son
doc-type:Image	Image	Image

setSpec	setName	Description française
doc-type:MovingImage	MovingImage	Image animée
doc-type:StillImage	StillImage	Image fixe
doc-type:CourseMaterial	CourseMaterial	Ressource pédagogique
doc-type:Website	Website	Site web
doc-type:Software	Software	Logiciel
doc-type:CartographicMaterial	CartographicMaterial	Document cartographique
doc-type:ResearchData	ResearchData	Données de la recherche
doc-type:Other	Other	Autres
doc-type:Text	Text	Texte

A.2.4 Données concernant le statut d'une publication

Un service de dépôts et de publication en libre accès peut contenir plusieurs versions d'un document correspondant à différentes étapes du processus éditorial. Comme la qualité du contenu d'un document peut être liée à son statut éditorial, il paraît raisonnable de caractériser au moins d'une manière générale le statut ou la version dans le processus éditorial. Les procédures d'évaluation et de contrôle de qualité d'une publication diffèrent d'une discipline à l'autre. Par conséquent, la structure de ce set se limitera à une distinction très générale du statut d'évaluation qui inclut le *peer review* ainsi que d'autres procédures (*editorial review* etc.) ; elle suit la terminologie de DRIVER (*version vocabulary*, Guidelines 2.0).

Recommandation

R.A.2-1 Il existe une structure de set en conformité avec le tableau 3 qui regroupe les métadonnées concernant le statut éditorial des documents correspondants.

Tableau 3 : Intitulé et description des sets pour le statut d'évaluation

setSpec	setName	Description française
status-type:draft	draft version	Une version provisoire mise en circulation en tant que telle.
status-type:submittedVersion	submitted version	La version soumise à une revue pour une évaluation par des experts.
status-type:acceptedVersion	accepted version	La version qui a été rédigée par l'auteur, qui contient les commentaires des évaluateurs, et qui a été acceptée pour publication.

status-type:publishedVersion	published version	La version mise en forme et publiée par l'éditeur.
status-type:updatedVersion	updated version	Une version mise à jour après la publication.

L'exemple 1 montre les métadonnées d'un enregistrement fourni via l'interface OAI-PMH en conformité avec les critères A.2.1 à A.2.4. L'enregistrement décrit un article scientifique publié du domaine des mathématiques, mis à disposition en libre accès.

```
<header>
  <identifier>oai:MyRepository.de:423569</identifier>
  <timestamp>2010-05-15T12:45:01Z</timestamp>
  <setSpec>open_access</setSpec>
  <setSpec>doc-type:article</setSpec>
  <setSpec>doc-type:Text</setSpec>
  <setSpec>ddc:510</setSpec>
  <setSpec>status-type:publishedVersion</setSpec>
</header>
```

Exemple 1 : Informations possibles du header d'un enregistrement fourni sur une requête du type ListRecord, GetRecord ou ListIdentifier

A.2.5 Traitement des données supprimées

En principe, un document une fois déposé dans un service de dépôts et de publication ne devrait pas être supprimé ou retiré. Mais il y a des exceptions, cf. aussi le critère 5 – *Sécurité de l'information* dans la section 2.5. A cause du biais de moissonnage incrémental qui ne récupère pas toujours toutes les données, il est possible qu'un fournisseur de service OAI ne soit pas informé d'une telle suppression si toutes les métadonnées sont retirées en même temps que le document. La spécification du protocole OAI ne précise pas quelles métadonnées un fournisseur de service doit garder après la suppression d'un document mais propose plusieurs options. Chaque fournisseur de service peut déterminer une *Deleting Strategy* (stratégie de suppression) qu'il doit communiquer en réponse à la requête OAI *Identify*.

Condition minimale

M.A.2-4 Une des valeurs 'persistent' ou 'transient' a été choisie comme *Deleting Strategy* pour le fournisseur de service.

- ⇒ OAI-PMH propose trois options 'no', 'persistent' et 'transient'. Si on choisit 'no', aucune information sur les données supprimées n'est communiquée ce qui peut entraîner des données incohérentes du côté des fournisseurs de service.
- ⇒ Si on choisit 'transient', il faut mettre à disposition une notice de métadonnées qui indique la suppression du document en question, et ceci pendant au moins un mois après la date de suppression.

A.2.6 Contrôle du flux de données

Le protocole OAI propose un contrôle du flux de données pour éviter le transfert de grands volumes de résultats suite à d'éventuelles requêtes OAI. Le fournisseur de données peut définir une *Harvest Batch Size*, c'est-à-dire le nombre maximal de métadonnées à fournir en même temps suite à une requête OAI *ListRecords* ou *ListIdentifiers*. Si jamais un résultat contient davantage de données, le serveur communique un *Resumption Token* avec lequel la poursuite de la fourniture de données peut être réclamée. Le protocole OAI permet au fournisseur de données de choisir la taille des paquets de données, la durée de vie pour la poursuite du transfert, et s'il veut appliquer l'option du contrôle du flux de données.

Recommandations

- R.A.2-2 Le *Harvest Batch Size*, c'est-à-dire le nombre maximal de métadonnées à fournir suite à une requête OAI *ListRecords*, se situe entre 100 (minimum) et 500 (maximum).
 - ⇒ Des *batch* plus petits entraînent davantage de requêtes OAI et augmentent inutilement la durée et la probabilité d'erreur. Des *batch* plus grands augmentent le risque d'erreur de transmission.
- R.A.2-3 La durée de vie des *Resumption Tokens* est d'au moins 24 heures.
 - ⇒ La durée de vie indiquée dans l'attribut *lifeSpan* indique le laps de temps pendant lequel le fournisseur de données garantit de poursuivre un transfert de données incomplet. Un délai trop court peut provoquer l'interruption du processus de moissonnage puisqu'il est terminé avant que toutes les données soient complètement transférées.
- R.A.2-4 L'attribut *completeListSize* est utilisé.

⇒ L'attribut peut définir le volume global des résultats, une information importante pour le contrôle et le pilotage du processus du moissonnage. Cet attribut est facultatif.

A.3 Conditions pour métadonnées (Dublin Core simplifié)

Le protocole OAI préconise que les métadonnées soient fournies en format Dublin Core simplifié mais ne précise ni le contenu ni la structure interne des différents éléments. Afin d'assurer un minimum d'interopérabilité pour les métadonnées, voici quelques conditions et recommandations pour l'application du Dublin Core à l'interface OAI.

Conditions minimales

M.A.3-1 Les données fournies en Dublin Core (*oai_dc*) contiennent au moins les éléments **creator**, **title**, **date**, **type** et **identifiant**.

⇒ Ces éléments sont indispensables pour décrire des documents scientifiques numériques.

M.A.3-2 Chaque élément DC caractérise une seule et unique valeur.

⇒ Chaque élément DC peut être répété dans un set de métadonnées.

⇒ Par exemple, chaque auteur devrait apparaître dans un élément *creator* unique, chaque mot-clé dans son élément *subject*, chaque URL dans son propre élément *identifiant*, etc.

M.A.3-3 Chaque set contient au moins un élément **identifiant** avec une adresse URL opérationnelle, avec un lien direct vers le texte intégral du document.

⇒ Pour qu'un identifiant pérenne (URN, DOI etc.) devienne une adresse URL opérationnelle, il doit intégrer dans son préfixe la racine de l'URL d'un service de *resolving*.

⇒ D'autres éléments *identifiant* peuvent contenir des URL vers une page d'accueil (*jump-off-page*), vers d'autres versions du document (dans d'autres formats, par exemple) ou d'autres identifiants (ISBN, DOI etc.).

M.A.3-4 L'élément **creator** appliquera la structure interne suivante : nom, prénom.

⇒ Idem pour l'élément *contributor*, s'il désigne un nom de personne.

M.A.3-5 Tous les documents ont un ou plusieurs éléments type, suivant les recommandations DINI *Vocabulaire commun pour types de publication et document*.*

⇒ Cette recommandation préconise d'ajouter systématiquement une valeur du *Dublin Core Type Vocabulary*.

⇒ En ce qui concerne la terminologie à utiliser, cf. la première colonne du tableau 2 de la section A.2.3.

M.A.3-6 Chaque set contient au moins un élément **subject** avec un groupe thématique.

⇒ En ce qui concerne la terminologie à utiliser, cf. la première colonne du tableau 1 de la section A.2.2.

M.A.3-7 Le contenu de l'élément **language** est conforme avec à la norme ISO 639-3.

⇒ Par exemple, le code pour le français est 'fra', pour l'anglais 'eng'.

M.A.3-8 Le contenu de l'élément **date** est conforme avec à la norme ISO 8601.

⇒ Ceci correspond au format YYYY-MM-DD.

Recommandations complémentaires

R.A.3-1 L'ordre des éléments **identifier** dans un ensemble de métadonnées (set) place l'élément préférentiel en première position.

⇒ Beaucoup de fournisseurs de service considèrent la position comme indice de priorité pour l'utilisation des URL. Les archives ouvertes préfèrent souvent le lien vers la page d'accueil du document (Jump-Off-Page).

⇒ Même si l'ordre des éléments n'a pas d'importance formelle dans le Dublin Core, le respect de cette convention s'est révélé un moyen pragmatique pour « orienter » les fournisseurs de service vers l'URL préférentiel.

R.A.3-2 L'élément **contributor** contient par occurrence le nom d'une personne ou d'une institution ayant contribué à la création du document en question.

⇒ Il s'agit par exemple d'un directeur de thèse ou de l'éditeur scientifique d'un ouvrage collectif, etc.

R.A.3-3 L'élément **source** applique les règles des Guidelines for Encoding Bibliographic Citation Information in Dublin Core Metadata.

⇒ L'élément sert de référence pour la citation de la version numérique. Cf. <http://dublincore.org/documents/dc-citation-guidelines/>

R.A.3-4 L'élément **relation** désigne des objets en relation avec le document en question.

⇒ Il s'agit par exemple d'une relation hiérarchique (*isPartOf*) ou d'une mise à jour (*isVersionOf*).

R.A.3-5 L'élément **subject** sert à indiquer la thématique du document.

⇒ Normalement, la thématique est décrite à l'aide de mots-clés ou d'un code issu d'un plan de classement.

R.A.3-6 L'élément **date** n'a qu'une seule occurrence par set.

⇒ On privilégie la date de publication à d'autres dates (dépôt, création etc.) car c'est la date la plus importante pour l'utilisateur final.

Annexe B Glossaire

Cette section identifie et définit les termes les plus importants de ce document.

Déclaration sur le Libre Accès. Des recommandations pour la mise en œuvre du Libre Accès, en particulier par rapport à un établissement scientifique. Par exemple, une telle déclaration peut désigner le libre accès comme modèle éditorial préférentiel de l'établissement et inciter les auteurs à publier leurs propres publications en libre accès.

Document. La plus petite unité logique publiée et mise à disposition par un service de dépôts et de publication en libre accès. En règle générale un mémoire scientifique cohérent sous une forme textuelle, avec des auteurs clairement identifiés. Synonymes : document numérique, publication, œuvre, ouvrage. Pour d'autres services d'édition (images, données), remplacer « document » par le terme *objet*.

Droits d'utilisation. Ici : des droits cédés à l'utilisateur final concernant l'utilisation de documents et métadonnées publiés par un service de dépôts et de publication. Suivant le code de propriété intellectuelle allemand, ces droits appartiennent à l'origine au créateur (auteur) et doivent être transférés par un procédé adapté.

Fournisseur de données. Serveur de documents au sens du protocole OAI. Il fournit via l'interface OAI uniquement les métadonnées des documents en question. Synonyme : *data provider*

Fournisseur de service. Prestataire de services fédérés à partir de données réparties agrégées via le protocole OAI. Synonyme : *moissonneur, harvester, service provider*.

Identifiant pérenne. Nom unique, universel et permanent d'un objet (ressource) numérique d'information, ici en particulier d'un document numérique. Il facilite la citation d'une publication numérique car il est, à la différence de l'URL, pérenne. Il y a différents systèmes d'identifiants pérennes, comme les URN, DOI ou PURL. La structure syntaxique d'un identifiant pérenne est définie par une description formelle de structure. Les identifiants pérennes et les URLs correspondants doivent être enregistrés (le plus souvent) d'une manière centralisée, pour permettre un service de résolveur (de liens) de transférer des requêtes d'identifiants pérennes vers les adresses physiques réelles.

Libre accès. Libre accès à l'information scientifique à l'échelle mondiale, surtout aux publications scientifiques, en format numérique et en ligne, tel que défini

entre autre par la déclaration de Berlin en 2003²⁷. Un mouvement international avec de nombreuses initiatives nationales et internationales travaille à la promotion et réalisation de cet objectif. D'habitude, on distingue deux formes de libre accès – la voie verte et la voie dorée. La voie verte désigne la mise à disposition d'une version en libre accès (publication parallèle) d'un document qui a été ou qui sera publié ailleurs (normalement par un éditeur), par exemple dans une archive ouverte (*repository*). Quant à la voie dorée, il s'agit d'une publication originale en libre accès, par exemple dans une revue en libre accès.

Licence de dépôt. Accord formel par lequel les ayants-droits (en règle général, les auteur ou éditeurs) cèdent certains droits d'utilisation à l'organisme du service de dépôts et de publication afin qu'il puisse mettre à disposition les documents en libre accès et pour qu'il en assure la conservation à long terme. Cet accord doit aussi exclure que les droits d'un tiers soient violés ou que l'organisme puisse en être tenu responsable. Synonymes : accord de licence, cession de droits.

Métadonnées. Des données qui caractérisent des objets, ici surtout des documents. Normalement, on distingue des métadonnées descriptives, techniques et administratives. Les métadonnées descriptives contiennent des informations sur la forme et le contenu (thématique). Les métadonnées peuvent être codées, stockées et échangées dans différents formats. Des normes pertinentes pour les publications scientifiques sont par exemple Dublin Core, MARC, MODS et, pour l'échange avec la bibliothèque nationale allemande, XMetaDissPlus.

Organisme. L'institution responsable du fonctionnement et de la mise à disposition (ici) du service de dépôts et publication. Il est prestataire de services pour différents groupes d'utilisateurs et porte la responsabilité globale vis-à-vis de l'extérieur, même si en interne certaines responsabilités peuvent être partagées voire externalisées. Synonymes : prestataire, prestataire de service.

Page d'accueil (d'un document). Page d'accès au texte intégral du document intégrant métadonnées et liens vers les fichiers du texte intégral et éventuellement d'autres fonctionnalités et informations (comme des données biographiques téléchargeables, des liens vers des systèmes de marque-page social et d'impression à la demande, des statistiques d'utilisation des documents). Normalement, la page d'accueil est générée automatiquement car ses informations sont stockées le plus souvent dans des bases de données. Synonymes : page d'accès, *splash page*, *front door*.

²⁷ Déclaration de Berlin sur le Libre Accès à la Connaissance en sciences exactes, sciences de la vie, sciences humaines et sociales, cf. <http://openaccess.inist.fr/spip.php?article38>

Serveur de document. Infrastructure technique d'un service de dépôts et de publication, caractérisée par les composants de base d'une infrastructure (réseaux, serveur, systèmes d'exploitation, systèmes de fichiers, bases de données, système de communication) et surtout par son logiciel de serveur de documents (OPUS, MyCoRe, ePrints, DSpace etc.). Synonymes : serveur de publication, *repository*, archive, serveur de dépôt.

Service de dépôts et de publication. Service complet pour la publication et mise à disposition en ligne de publications scientifiques en format numérique à destination aussi bien de leurs producteurs que de leurs consommateurs ; comprend aussi bien l'infrastructure technique (y compris le *serveur de documents*) que le cadre organisationnel et juridique.

Utilisateurs. Personnes physiques utilisant l'offre du service de dépôts et de publication, en particulier comme producteurs (auteurs, éditeurs) ou consommateurs (lecteurs, chercheurs) de documents.

Annexe C Auteurs et traducteurs

Ce document est protégé par une licence Creative Commons CC-BY
<http://creativecommons.org/licenses/by/2.0/fr/>.

Voici la liste des auteurs de la version actuelle :

Kim Braun. Bibliotheks- und Informationssystem der Carl von Ossietzky
Universität Oldenburg, kim.braun@uni-oldenburg.de.

Stefan Buddenbohm. Niedersächsische Staats- und Universitätsbibliothek
Göttingen, buddenbohm@sub.uni-goettingen.de.

Susanne Dobratz. Humboldt-Universität zu Berlin, Universitätsbibliothek,
Arbeitsgruppe Elektronisches Publizieren, dobratz@cms.hu-berlin.de.

Ulrich Herb. Saarländische Universitäts- und Landesbibliothek, Saarbrücken,
u.herb@sulb.uni-saarland.de.

Uwe Müller. Humboldt-Universität zu Berlin, Institut für Bibliotheks- und
Informationswissenschaft, u.mueller@ibi.hu-berlin.de.

Heinz Pampel. Helmholtz-Gemeinschaft, Potsdam, pampel@gfz-potsdam.de.

Birgit Schmidt. Niedersächsische Staats- und Universitätsbibliothek Göttingen,
bschmidt@sub.uni-goettingen.de.

Frank Scholze. Karlsruher Institut für Technologie, KIT-Bibliothek,
frank.scholze@kit.edu.

Matthias Schulze. Universität Stuttgart, Universitätsbibliothek,
matthias.schulze@ub.uni-stuttgart.de.

Thomas Severiens. Universität Osnabrück, Fachbereich Mathematik/
Informatik, tseverie@mathematik.uni-osnabrueck.de.

Friedrich Summann. Universität Bielefeld, Universitätsbibliothek,
friedrich.summann@uni-bielefeld.de.

Tobias Steinke. Deutsche Nationalbibliothek, Frankfurt, t.steinke@dnb.de.

Stefan Wolf. Bibliotheksservice-Zentrum Baden-Württemberg, Konstanz,
stefan.wolf@bsz-bw.de.

Ont contribué aux versions antérieures :

Margo Bargheer.

Stefan Gradmann.

Eberhard Hilf.

Wolfram Horstmann.

Elmar Mittler.

Peter Schirmbacher.

Silke Schomburg.

Heinrich Stamerjohanns.

Bert Wendland.

Christoph Ziegler.

Remerciements à d'autres personnes dont l'aide et l'engagement ont facilité la réussite du projet :

Maren Brodersen. Deutsche Nationalbibliothek, Frankfurt,
m.brodersen@dnb.de.

Nina Gerlach. Niedersächsische Staats- und Universitätsbibliothek Göttingen,
gerlach@sub.uni-goettingen.de.

Sarah Hartmann. Deutsche Nationalbibliothek, Frankfurt,
s.hartmann@dnb.de.

Maxi Kindling. Humboldt-Universität zu Berlin, Computer- und Medienservice,
Arbeitsgruppe Elektronisches Publizieren, maxi.kindling@ibi.hu-berlin.de.

Nikola Korb. Universitätsbibliothek Siegen, korb@ub.uni-siegen.de.

Berthold Weiß. Hochschul- und Landesbibliothek Fulda,
berthold.weiss@hllb.hs-fulda.de.

Pour la traduction :

Joachim Schöpfel. Université Charles-de-Gaulle Lille 3,
joachim.schopfel@univ-lille3.fr.

Sylvie Dalbin. ATD, Paris, sylvieATD@aol.com.

Uwe Müller. Deutsche Nationalbibliothek, Frankfurt, u.mueller@dnb.de.

Remerciements pour leur aide et soutien :

Stefan Gradmann. Humboldt-Universität zu Berlin, Institut für Bibliotheks- und Informationswissenschaft, s.gradmann@ibi.hu-berlin.de.

Kristina Hanig. Universitätsbibliothek, Universität Würzburg,
kristina.hanig@bibliothek.uni-wuerzburg.de.

Ulrich Herb. Saarländische Universitäts- und Landesbibliothek, Saarbrücken,
u.herb@sulb.uni-saarland.de.

Jean-François Lutz. Université Henri-Poincaré Nancy 1,
Jean-Francois.Lutz@scd.uhp-nancy.fr.

Jean-François Nominé. INIST-CNRS, Nancy, jean-francois.nomine@inist.fr

Hélène Prost. INIST-CNRS, Nancy, helene.prost@inist.fr.

DINI – Deutsche Initiative für Netzwerkinformation e.V.

DINI-Geschäftsstelle [coordonnées]

c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen

Papendiek 14

37073 Göttingen

Allemagne

Tel.: +49 551 39-3857

Fax: +49 551 39-3856

E-Mail: gs@dini.de

www.dini.de