



Rapport de synthèse du chantier DOI

Windpouire Esther Dzale Yeumo, Lydie Soler, Michael M. Alaux, Robert R. Bossy, Sebastien Carrere, Sylvie Cocaud, Stephane Dervaux, Marie-Françoise Fily, Nicolas N. Guinet, Célia Michotey, et al.

► To cite this version:

Windpouire Esther Dzale Yeumo, Lydie Soler, Michael M. Alaux, Robert R. Bossy, Sebastien Carrere, et al.. Rapport de synthèse du chantier DOI. [0] 2014, 47 p. hal-02801364

HAL Id: hal-02801364

<https://hal.inrae.fr/hal-02801364>

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Partage des données de la recherche

Chantier DOI



Juin / 2014

Ce document présente une synthèse des travaux du groupe de travail sur la gestion des identifiants numériques DOI.

Coordination : Dzale Yeumo Esther, Soler Lydie

Membres du groupe de travail : Alaux Michael, Bossy Robert, Carrere Sébastien, Cocaud Sylvie, Dervaux Stéphane, Fily Marie Françoise, Guinet Nicolas, Michotey Célia, Pommier Cyril, Steinbach Delphine.

1	LE CHANTIER DOI : INTRODUCTION	4
1.1	PERIMETRE DU CHANTIER	4
1.2	COMPOSITION DU GROUPE DE TRAVAIL	4
1.3	ORGANISATION DU DOCUMENT	4
2	DOI : DEFINITION, RESPONSABILITES ET OBLIGATIONS	5
2.1	DOI : DEFINITION	5
2.2	DOI : RESPONSABILITES ET OBLIGATIONS	6
3	MISE EN ŒUVRE DES DOI A L'INRA : PRINCIPALES RECOMMANDATIONS	7
3.1	LES OBJETS SCIENTIFIQUES IDENTIFIABLES PAR UN DOI	7
3.2	REGLE DE CONSTRUCTION DES DOI	9
3.3	ASSURER LA PERENNITE DES DONNEES	10
3.4	LES DEMANDEURS	10
3.5	LES METADONNEES	11
3.5.1	LES METADONNEES OBLIGATOIRES EXIGES PAR DATACITE	12
3.5.2	METADONNEES RECOMMANDEES PAR DATACITE ET AUTRES METADONNEES	13
3.5.3	DES METADONNEES « SPECIFIQUES » A L'INRA ?	16
4	OFFRE DE SERVICES PRECONISEE	17
4.1	ASPECTS ORGANISATIONNELS : UN GUICHET UNIQUE DE GESTION DES ATTRIBUTIONS DE DOI	17
4.2	ASPECTS TECHNIQUES	17
4.3	DESCRIPTION DES CAS D'UTILISATION	18
4.3.1	PRESENTATION GENERALE ET GLOSSAIRE	18
4.3.2	CAS D'UTILISATION : DEMANDE D'UN DOI	19
4.3.3	CAS D'UTILISATION : CHANGEMENT DE LA LANDING PAGE	23
4.3.4	CAS D'UTILISATION : MISE A JOUR DES METADONNEES	27
4.3.5	CAS D'UTILISATION : INACTIVATION D'UN DOI	31
5	PROJETS PILOTES ET PROCHAINES ETAPES	35
5.1	LES PROJETS PILOTES	35
5.1.1	PRESENTATION ET OBJECTIFS	35
5.1.2	PREMIERES ACTIONS / PREMIER BILAN	35
5.1.3	POURSUITE DES PROJETS PILOTES	36
5.2	PROCHAINES ETAPES	37
5.2.1	DECISIONS ET ACTIONS PREALABLES A L'OUVERTURE DU GUICHET DE GESTION ET D'ATTRIBUTION DE DOI	37
5.2.2	GESTION DES DEMANDES INDIVIDUELLES VIA DES FORMULAIRES	39
5.2.3	API POUR LA GENERATION ET LA GESTION DES DOI	40
	ANNEXES	41
1	DIGITAL OBJECT IDENTIFIER (DOI)	41
1.1	QU'EST-CE QU'UN DOI ?	41
1.2	UN DOI, POUR QUOI FAIRE / POUR QUELLES DONNEES ?	41

1.3	SYNTAXE D'UN DOI	41
1.4	CONDITIONS D'ATTRIBUTION DE DOI	42
1.4.1	QUI ATTRIBUE LES DOI ?	42
1.4.2	QUE PEUT-ON IDENTIFIER AVEC UN DOI ?	42
1.4.3	QUE COMPREND LE SERVICE D'ENREGISTREMENT DE DOI OFFERT PAR L'INIST-CNRS?	42
1.4.4	COUT D'ATTRIBUTION DES DOI	42
1.4.5	CONDITIONS D'ATTRIBUTION D'UN DOI	43
1.5	LES SERVICES PROPOSES PAR DATAcITE	43
1.6	LES SERVICES MIS A DISPOSITION AUTOUR DES DOI	44
2	<u>LE QUESTIONNAIRE DES PROJETS PILOTES</u>	45

1 Le chantier DOI : introduction

1.1 Périmètre du chantier

Afin de répondre aux enjeux de l'« Open » et du « Big » data, l'INRA souhaite se doter d'une offre de services en direction des équipes de recherche pour les aider à rentrer dans le mouvement de l'open science. Dans cette démarche de mise en place de services en faveur du partage des données, le séminaire du 2 et 3 décembre 2013 a mis en évidence la nécessité de mener une étude de faisabilité concernant les dispositifs suivant :

- un annuaire des sources de données de l'INRA,
- une infrastructure de dépôt (entrepôt de données),
- et la gestion des identifiants numériques (DOI).

Plusieurs manifestes internationaux dont le manifeste d'Amsterdam ont affirmé récemment l'importance des données de la recherche. Celles-ci doivent être considérées comme des produits de la recherche et doivent être citables au même titre que les publications scientifiques.

La généralisation de la citation des données rend nécessaire la mise en place d'infrastructures pour garantir un accès pérenne aux données citées. Les identifiants numériques tels que les DOI et le contrat social qu'ils impliquent sont des éléments de ces infrastructures.

Un groupe de travail a réfléchi aux enjeux de la mise en œuvre des DOI à l'INRA avec pour objectif de formuler des recommandations et proposer une offre de services. Ce document présente une synthèse des travaux de ce groupe de travail.

1.2 Composition du groupe de travail

Ce groupe de travail est composé de :

Coordination : Dzale Yeumo Esther, Soler Lydie

Membres du groupe de travail : Alaux Michael, Bossy Robert, Carrere Sébastien, Cocaud Sylvie, Dervaux Stéphane, Fily Marie Françoise, Guinet Nicolas, Michotey Célia, Pommier Cyril, Steinbach Delphine.

1.3 Organisation du document

Le chapitre 2 rappelle les grands principes des DOI ainsi que les responsabilités et obligations qui en découlent pour l'INRA. Ensuite, le chapitre 3 présente la synthèse des recommandations du groupe de travail pour la mise en œuvre des DOI à l'INRA. Le chapitre 4 décrit l'offre de service préconisée par le groupe de travail. Enfin, le chapitre 5 présente les premiers résultats des projets pilotes initiés ainsi que les prochaines étapes pour la mise en œuvre du service d'attribution de DOI à l'INRA.

2 DOI : définition, responsabilités et obligations

2.1 DOI : définition

Les tableaux ci-dessous présentent une synthèse de ce qu'est un DOI DataCite. Pour plus d'information, vous pouvez consulter l'annexe 1.

Digital Object Identifier (DOI) : la synthèse	
Qu'est-ce qu'un DOI ?	Un identifiant pérenne et unique permettant de référencer, citer et fournir un lien stable vers un objet numérique.
Pour quels types de données ?	Tout objet scientifique numérique que l'on souhaite rendre citable , qui respecte les exigences de DataCite, les règles de propriété intellectuelle et le cadre juridique.
Pourquoi un DOI ?	- Référencer de manière unique et pérenne un objet scientifique ; - Pouvoir le citer de façon fiable et y accéder de manière pérenne ; - Faciliter sa découverte, son accès et sa réutilisation.
Syntaxe	Un DOI est unique et composé de 2 parties :  - préfixe : le centre de données (INRA).
Qu'est-ce que DataCite ?	Consortium international, porté par des institutions nationales, dont le but est de proposer une infrastructure destinée à la promotion d'ensembles de données et d'informations non textuelles.
Qui attribue un DOI ?	L'INIST-CNRS (représentant le CNRS) pour la France est membre de DataCite. Il est doté depuis février 2010 d'un statut d'agence DOI en charge de la gestion et de l'attribution de DOI. Un contrat a été passé par la DV-IST avec l'INIST-CNRS permettant à l'INRA d'être centre de données.

Éléments requis pour la création d'un DOI

Des métadonnées	- Respecter le schéma de métadonnées de DataCite ; - Fournir les métadonnées gratuitement et librement ; - Sous une licence permettant leur disponibilité sans restriction et leur réutilisation (Creative Commons¹ CC0 par exemple).
Une page descriptive (ou landing page)	Cette page contient : - La citation complète de l'objet scientifique ; - Les métadonnées associées ; - Les informations concernant l'accès à l'objet scientifique (URL d'accès, conditions d'obtention, restrictions, etc.) ; - Les informations pour lire l'objet scientifique (logiciels, contexte, autres informations nécessaires à l'interprétation....).
Une URL valide	La page descriptive de l'objet scientifique doit être accessible grâce à une URL valide, pendant une durée minimale de 10 ans.
Une politique de curation et de conservation sur 10 ans	Tout objet scientifique identifié par un DOI doit être stocké pendant une durée minimale de 10 ans.

¹<http://creativecommons.fr/licences/les-6-licences/>

2.2 DOI : responsabilités et obligations

Le tableau ci-dessous présente la synthèse des responsabilités et obligations de l'INRA en tant que centre de données, déposant des DOI.

Responsabilités et obligations de l'INRA	
Respect des règles de propriété intellectuelle et du cadre juridique	Il revient à l'INRA de s'assurer du respect de ces règles (accords de projets, règles éthiques, respect de la licence d'origine).
Évaluation et contrôle de la qualité des objets scientifiques	Il revient à l'INRA de mettre en place une procédure de contrôle de tous les objets scientifiques enregistrés avec un DOI, et de veiller à la validité des données conformément : - à ses propres standards ; - aux règles générales de bonne pratique du domaine de recherche de l'objet scientifique.
Intégrité des métadonnées	Il revient à l'INRA de s'assurer que les métadonnées sont : - conformes au schéma de métadonnées de DataCite ; - fournies gratuitement, librement et sous une licence sans restriction de réutilisation.
Accessibilité, gestion	L'INRA en tant que centre de données doit assurer l'accessibilité via un URL et la gestion sur le long terme des objets scientifiques identifiés par des DOI. Le DOI doit passer par un résolveur en direction de la landing page donnant des informations suffisantes sur l'objet scientifique et les conditions d'accès.
Stockage	L'INRA est responsable du stockage des objets scientifiques identifiés par des DOI associés à son nom. Il doit mettre en œuvre des infrastructures de stockage adaptées (volumétrie, format ...).
Pérennité / persistance	Il revient à l'INRA de mettre en œuvre une politique et des moyens assurant la disponibilité, la pérennité et la persistance des objets scientifiques identifié par un DOI, cela pendant 10 ans minimum. Si un objet scientifique identifié par un DOI n'est plus considéré comme digne d'archivage et doit être supprimé, l'INRA doit s'assurer que l'URL associé au DOI a été changée et pointé vers une landing page expliquant la raison de cette indisponibilité.
Coût d'attribution des DOI	Un droit annuel de 180 euros par an doit être versé par l'INRA à l'INIST-CNRS. Ce droit est indépendant du nombre de DOI enregistrés.

3 Mise en œuvre des DOI à l'INRA : principales recommandations

Afin de permettre à l'INRA de respecter ses engagements vis à vis de DataCite, le groupe de travail a formulé un certain nombre de recommandations synthétisées dans ce paragraphe.

3.1 Les objets scientifiques identifiables par un DOI

L'attribution d'un DOI n'est pas une simple mise à disposition immédiate et évidente d'un objet scientifique. Une réflexion doit être menée portant sur des enjeux juridiques, scientifiques et techniques.

Critères d'éligibilité pour l'attribution d'un DOI à un objet scientifique

(Recommandations du groupe de travail)

Pour identifier un objet scientifique avec un DOI, il faut :

- Souhaiter le rendre citable ;
- Respecter les règles de propriété intellectuelle et le cadre juridique ;
- Mener une réflexion sur la maturité de l'objet scientifique et décider à quelle étape du cycle de vie de l'objet scientifique il convient de lui attribuer un DOI ;
- Mener une réflexion sur le niveau de granularité de l'objet scientifique et décider à quel le niveau de granularité il est pertinent d'attribuer un DOI ;
- Mener une réflexion sur les conditions d'accès à l'objet scientifique et décider d'éventuelles restrictions et/ou conditions de paiement
- Stocker l'objet scientifique dans une infrastructure de stockage assurant sa pérennité sur 10 ans minimum ;
- Fournir une landing page accessible pendant 10 ans minimum et contenant les informations exigées par DataCite :
 - La citation complète de l'objet scientifique ;
 - Les métadonnées associées ;
 - Les informations concernant l'accès à l'objet scientifique (URL d'accès, conditions d'obtention, restrictions, etc.) ;
 - Les informations pour lire l'objet scientifique (logiciels, contexte, autres informations nécessaires à l'interprétation....).

Il n'y a pas de restriction sur la nature de l'objet (jeux de données, images, document, document audio)

Les tableaux ci-dessous reprennent les critères d'éligibilité à l'attribution d'un DOI et les précisent.

Nature de l'objet scientifique

Recommandation	Tout objet scientifique numérique peut être identifié par un DOI
Contexte/Enjeux pour l'INRA	Par définition, un DOI est utilisé pour identifier des objets dans un environnement numérique.

Règles de propriété intellectuelle et cadre juridique	
Recommandation	L'INRA doit établir des recommandations claires sur les questions éthiques, juridiques, et de propriété intellectuelle. La mise en place d'un service d'accompagnement pour ces questions juridiques liées à l'ouverture des données nous semble indispensable.
Contexte/Enjeux pour l'INRA	- Il est de la responsabilité du déclarant de s'assurer que son objet scientifique respecte ces règles juridiques ; - Ces règles dépendent notamment de la nature des données, des méthodes d'obtention des données, de la discipline scientifique concernée (anonymisation statistique ...) ... - Ces règles sont complexes et vont au-delà de la durée du projet qui a vu naître/a financé les données ; - La question de l'attribution de DOI pourrait être intégrée dans le plan de gestion des données.

Niveau de granularité de l'objet scientifique	
Recommandation	Laisser la liberté sur le niveau de granularité. Le choix du niveau de granularité est une question complexe : un accompagnement doit être mis en place au sein du service d'attribution de DOI pour aider le déclarant à choisir le niveau de granularité.
Contexte/Enjeux pour l'INRA	- Attribuer les DOI aux grains les plus pertinents des objets scientifiques ; - DataCite n'impose aucune restriction sur le niveau de granularité ; - Le choix du niveau de granularité nécessite une connaissance et une compréhension précise de l'objet scientifique ; - La réflexion doit être menée par le titulaire de l'objet scientifique (le déclarant) et porter sur : • La partie du jeu de données à identifier (granularité spatiale) : identifier un savoir-faire, des données elles-mêmes ? • La version du jeu de données à identifier (granularité temporelle) : les changements sont-ils suffisants sur le plan scientifique pour être différenciés de la version précédente ? <u>Illustration de la problématique de la granularité temporelle d'un objet scientifique</u> : considérons des données collectées par un dispositif de mesure de température d'un milieu en quasi temps réel. Le jeu de données global est en constante évolution : de nouvelles données sont ajoutées quotidiennement, des corrections sont apportées si, par exemple, un défaut est détecté a posteriori sur un capteur. Il est possible d'attribuer un DOI au produit dans son ensemble (au sens, savoir-faire, procédure). Cependant, les données elles-mêmes (pas seulement le savoir-faire, la procédure) peuvent aussi être exploitées dans des travaux. Les auteurs de ces travaux doivent donc pouvoir citer un jeu de données au plus près de celui qu'ils ont réellement exploité dans leurs travaux. L'attribution d'un DOI sur une copie mensuelle du jeu de données global pourrait, par exemple, être alors plus appropriée.

Conditions d'accès à l'objet scientifique	
Recommandation	Laisser la liberté, au déclarant, dans la définition et l'attribution des conditions d'accès.
Contexte/Enjeux pour l'INRA	- Permettre l'attribution d'un DOI à un maximum d'objets scientifiques ; - DataCite impose uniquement un accès libre à la landing page et aux métadonnées ; - Les conditions d'accès sont à préciser sur la page descriptive : • Accès libre ou à un groupe restreint (partenaires d'un projet ANR par exemple) ; • Accès gratuit ou payant (en précisant les conditions de paiement).
Questions en suspens	- Est-ce que tous/certains objets scientifiques manipulés/créés par l'INRA tombent automatiquement, au bout d'un certain temps, dans le « domaine public » ?

3.2 Règle de construction des DOI

Règle de construction des DOI : les préfixes	
Recommandation	Associer un même préfixe à tous les objets scientifiques INRA identifiés par un DOI
Contexte/Enjeux pour l'INRA	- Établir l'INRA comme centre de données (un préfixe = un centre de données) ; - Avoir une vue d'ensemble des jeux de données de l'INRA (hors objets scientifiques identifiés par un partenaire de l'INRA). Les services de DataCite permettent de faire des recherches par centre de données ; - Gérer un seul contrat avec l'INIST-CNRS (celui établi par la DV-IST) au nom de tout l'INRA.

Règle de construction des DOI : les suffixes	
Recommandation	Attribuer des suffixes « opaques » sans une sémantique particulière susceptible de changer dans le temps.
Contexte/Enjeux pour l'INRA	L'utilisation de suffixe « opaques » simplifie la gestion des DOI.

Unicité des DOI	
Recommandation	Mettre en œuvre une API assurant l'unicité des suffixes pour les objets scientifiques de l'INRA.
Contexte/Enjeux pour l'INRA	En tant que centre de données, l'INRA est responsable de la génération de suffixes uniques.

3.3 Assurer la pérennité des données

Stockage, pérennité et persistance de l'objet scientifique	
Recommandation	Mettre en œuvre des infrastructures de stockage disposant des moyens adéquats ² (créer de nouvelles infrastructures ou s'appuyer sur des infrastructures existantes). Créer un label « Open Science » pour ces infrastructures et permettre ainsi leur identification par les utilisateurs.
Contexte/Enjeux pour l'INRA	- Les objets scientifiques identifiés par des DOI doivent être stockés/archivés pendant une durée minimale de 10 ans ; - L'INRA, en tant que centre de données, est responsable de la pérennité des données identifiées.

Accès et pérennité de la landing page	
Recommandation	Mettre en œuvre une gestion centralisée des landing pages.
Contexte/Enjeux pour l'INRA	- L'INRA doit s'assurer de la pérennité de l'URL vers la landing page : si l'URL vers la page descriptive était amenée à changer, il faudrait signaler ce changement dans un délai de 5 jours. (utilisation de l'API dataCite).

3.4 Les demandeurs

Qui peut demander un DOI ?	
Recommandation	Sous réserve que l'objet à identifier respecte les critères d'éligibilité : - Toute personne disposant d'un identifiant LDAP INRA sauf certains PEPS³ (voir tableau ci-dessous) ; - Toute plateforme de production ou de gestion de données INRA labellisée « Open Science ».
Contexte/Enjeux pour l'INRA	- L'Inra est responsable de la qualité et de la pérennité des données identifiées par des DOI associés à son nom ; - Suivant le statut précis du demandeur, une validation supplémentaire pourra être nécessaire (par exemple demander une confirmation à l'encadrant d'un thésard ou stagiaire).

² Voir avec les personnes responsables de ces structures pour connaître les moyens supplémentaires à mettre en œuvre si besoin

³ Personnels Et Partenaires Suivis

Précisions sur le périmètre des personnes disposant d'un LDAP		
Personnes autorisées à demander un DOI	- Personnels titulaires INRA ; - Personnels non titulaires INRA : - Main d'œuvre occasionnelle - Tout CDD - Boursier de thèse - Stagiaire ;	- PEPS⁴ ayant pour statut : - Enseignant chercheur - Chercheur ou post doc - Cadre EPIC ou privé - Ingénieur - Assistant ingénieur - Technicien - Tout doctorant - Personne en formation.
Personnes NON autorisées à demander un DOI	- PEPS ayant pour statut : - Prestataire de service - Tout missionnaire ULYSSE	

3.5 Les métadonnées

Licence des métadonnées	
Recommandation	Bien documenter l'objet scientifique à l'aide de métadonnées.
Contexte/Enjeux pour l'INRA	- Les métadonnées sont utilisées pour générer une citation pour l'objet scientifique. En général, elles permettent une meilleure exposition de l'objet scientifique. - Pour pouvoir être enregistré, un DOI doit être accompagné d'un fichier de métadonnées. - Toutes les métadonnées transmises à DataCite doivent être librement accessibles dans le cadre d'une licence <i>Creative Commons</i>⁵ <i>CCO</i>.

⁴ Personnels Et Partenaires Suivis

⁵<http://creativecommons.fr/licences/les-6-licences/>

3.5.1 Les métadonnées obligatoires exigées par DataCite

Le tableau ci-dessous résume les métadonnées exigées par DataCite. Le schéma complet des métadonnées est disponible à cette adresse :

http://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.0.pdf.

Le groupe a étudié ces métadonnées et formule des recommandations sur la manière de les remplir (colonne de droite).

ID	Propriété	Occurrence	Définition	Valeurs autorisées, exemples, recommandations INRA
1	Identifier	1	The Identifier is a unique string that identifies a resource	DOI enregistré par une agence de DataCite
1.1	IdentifierType	1	The type of the Identifier.	Valeur acceptée : DOI
2	Creator	1-n	The main researchers involved in producing the data, or the authors of the publication, in priority order.	Nom d'un organisme ou d'une personne. Le groupe recommande de mettre une information différente de « INRA, CNRS »... Celle-ci se retrouvera dans le champ Publisher
2.1	creatorName	1	The name of the creator.	Nom, prénom Exemples : Leblanc, Paul; Mader, Sylvie
2.2	nameIdentifier	0-1	Uniquely identifies an individual or legal entity, according to various schemes.	
2.2.1	nameIdentifierScheme	1	The name of the name identifier scheme.	Obligatoire si nameIdentifier est renseigné. Exemples de valeurs possibles : ORCID, ISNI
2.2.2	schemeURI	0-1	The URI of the name identifier scheme.	Exemple : http://www.isni.org
3	Title		A name or title by which a resource is known.	
3.1	titleType	0-1	The type of Title.	Valeurs acceptées : AlternativeTitle, Subtitle, TranslatedTitle
4	Publisher	1	The name of the entity that holds, archives, publishes prints, distributes, releases, issues, or produces the resource. This property will be used to formulate the citation, so consider the prominence of the role.	Valeurs acceptées : INRA
5	Publication Year	1	The year when the data was or will be made publicly available.	

3.5.2 Métadonnées recommandées par DataCite et autres métadonnées

Les métadonnées sur fond rose sont recommandées.

ID	Propriété	Occurrence	Définition	Valeurs autorisées, exemples, recommandations INRA
6	Subject	0-n	Subject, keyword, classification code, or key phrase describing the resource.	
6.1	subjectScheme	0-1	The name of the subject scheme or classification code or authority if one is used.	
6.2	schemeURI	0-1	The URI of the subject identifier scheme.	
7	Contributor	0-n	The institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource.	
7.1	contributorType	1	The type of contributor of the resource.	Obligatoire si Contributor est renseigné Valeurs acceptées : ContactPerson, DataCollector, DataManager, Distributor, Editor, Funder, HostingInstitution, Producer, ProjectLeader, ProjectMember, RegistrationAuthority, RelatedPerson, Researcher, ResearchGroup, RightsHolder, Sponsor, Supervisor, WorkPackageLeader, Other. Lire les annexes de la description du schéma de DataCite pour connaître la définition de chacune de ces valeurs.
7.2	contributorName	1	The name of the contributor.	Obligatoire si Contributor est renseigné
7.3	nameIdentifier	0-1	Uniquely identifies an individual or legal entity, according to various schemes.	Le groupe recommande l'utilisation des identifiants ORCID
7.3.1	nameIdentifierScheme	1	The name of the name identifier scheme.	Obligatoire si nameIdentifier est renseigné
7.3.2	schemeURI	0-1	The URI of the name identifier scheme	
8	Date	0-n	Different dates relevant to the work	

ID	Propriété	Occurrence	Définition	Valeurs autorisées, exemples, recommandations INRA
8.1	<i>dateType</i>	1	The type of date.	Valeurs acceptées: Accepted, Available, Copyrighted, Collected, Created, Issued, Submitted, Updated, Valid. Lire les annexes de la description du schéma de DataCite pour connaître la définition de chacune de ces valeurs.
9	<i>Language</i>	0-1	The primary language of the resource.	Valeurs acceptées : valeurs issues des codes IETF BCP 47, ISO 639-1 Exemples : en, de, fr
10	<i>ResourceType</i>	0-1	A description of the resource.	DataCite recommande de se baser sur la liste des types de ressources proposée dans CASRAI. Cette liste n'étant pas à ce jour assez riche pour les datasets, le groupe de travail recommande de suivre les travaux de projets ou groupes internationaux à même de proposer des vocabulaires de description des datasets. Exemples : COAR (https://www.coar-repositories.org/) et le groupe de travail Standardisation of Data Categories and Codes (https://rd-alliance.org/group/standardisation-data-categories-and-codes-wg.html) de RDA (Research Data Alliance)
10.1	<i>resourceTypeGeneral</i>	1	The general type of a resource	Obligatoire si <i>resourceTypeGeneral</i> est renseigné. Valeurs acceptées : Audiovisual, Collection, Dataset, Event, Image, InteractiveResource, Model, PhysicalObject, Service, Software, Sound, Text, Workflow, Other. Lire les annexes de la description du schéma de DataCite pour connaître la définition de chacune de ces valeurs.
11	<i>AlternateIdentifier</i>	0-n	An identifier or identifiers other than the primary Identifier applied to the resource being registered. This may be any alphanumeric string which is unique within its domain of issue. May be used for local identifiers. AlternateIdentifier should be used for another identifier of the same instance (same location, same file).	
11.1	<i>alternateIdentifierType</i>	1	The type of the AlternateIdentifier.	Obligatoire si <i>AlternateIdentifier</i> est utilisé.

ID	Propriété	Occurrence	Définition	Valeurs autorisées, exemples, recommandations INRA
12	<i>RelatedIdentifier</i>	0-n	Identifiers of related resources. These must be globally unique identifiers.	Dans le cas d'une donnée soutenant une publication, le groupe recommande d'utiliser cette métadonnée pour faire référence à la publication. Cette métadonnée peut également être utilisée pour référencer d'autres données qui sont en relation avec la donnée identifiée.
12.1	<i>relatedIdentifierType</i>	1	The type of the RelatedIdentifier	Obligatoire si <i>RelatedIdentifier</i> est renseigné. Valeurs acceptées : ARK, DOI, EAN13, EISSN, Handle, ISBN, ISSN, ISTC, LISSN, LSID, PMID, PURL, UPC, URL, URN Lire les annexes de la description du schéma de DataCite pour connaître la définition de chacune de ces valeurs.
12.2	<i>relationType</i>	1	Description of the relationship of the resource being registered (A) and the related resource (B).	Obligatoire si <i>RelatedIdentifier</i> est renseigné. Valeurs acceptées : IsCitedBy, Cites, IsSupplementTo, IsSupplementedBy, IsContinuedBy, Continues, HasMedadata, IsMedadataFor, IsNewversionOf, IsPreviousVersionOf, IsPartOf, HasPart, IsReferencedBy, References, IsDocumentedBy, Documents, IsCompiledBy, Compiles, IsVariantFormOf, IsOriginalFormOf, IsIdenticalTo Lire les annexes de la description du schéma de DataCite pour connaître la définition de chacune de ces valeurs.
12.3	<i>relatedMetadataScheme</i>	0-1	The name of the scheme	
12.4	<i>schemeURI</i>	0-1	The URI of the relatedMetadataScheme	
12.5	<i>schemeType</i>	0-1	The type of the relatedMetadataScheme, linked with the schemeURI.	
13	<i>Size</i>	0-n	Unstructured size information about the resource.	
14	<i>Format</i>	0-n	Technical format of the resource.	
15	<i>Version</i>	0-1	The version number of the resource.	

ID	Propriété	Occurrence	Définition	Valeurs autorisées, exemples, recommandations INRA
16	Rights	0-n	Any rights information for this resource.	L'existence d'un embargo peut être mentionnée en utilisant cette métadonnée. Cette métadonnée permet aussi de mentionner la licence attachée à la donnée identifiée (exemple : Creative Commons)
16.1	rightsURI	0-1	The URI of the license	
17	Description	0-n	All additional information that does not fit in any of the other categories. May be used for technical information.	
17.1	descriptionType	1	The type of the Description.	Obligatoire si Description est renseigné. Valeurs acceptées : Abstract, Methods, SeriesInformation, TableOfContents, Other Lire les annexes de la description du schéma de DataCite pour connaître la définition de chacune de ces valeurs.
18	Geolocation	0-n	Spatial region or named place where the data was gathered or about which the data is focused.	
18.1	geoLocationPoint	0-1	A point location in space	
18.2	geoLocationBox	0-1	The spatial limits of a place.	
18.3	geoLocationPlace	0-1	Description of a geographic location	

3.5.3 Des métadonnées « spécifiques » à l'INRA ?

En plus des métadonnées obligatoires définies par DataCite, le groupe recommande d'identifier les métadonnées spécifiques à chaque famille de données qu'il serait nécessaire d'exiger afin d'en garantir la qualité. Ce travail devra se faire en collaboration avec les différentes familles de données.

De plus, le groupe recommande que dans le cas de données soutenant une publication scientifique, le lien vers la publication scientifique (une fois acceptée) figure dans la liste des métadonnées requises.

4 Offre de services préconisée

4.1 Aspects organisationnels : un guichet unique de gestion des attributions de DOI

Guichet unique de gestion des attributions de DOI à l'INRA	
Recommandation	Mettre en place un guichet unique en charge du service d'attribution de DOI à l'INRA doté des moyens adéquats. Ce guichet serait porté par la DV-IST. Ce guichet fournirait : - Un formulaire de demande de création/mise à jour/inactivation de DOI - Une API « INRA » encapsulant l'API DataCite permettant l'attribution d'un DOI et la gestion des métadonnées associées à un DOI ; - Un accompagnement aux demandeurs : informer et assister les demandeurs de DOI (plateformes, demandeurs « isolés »).
Contexte/Enjeux pour l'INRA	- Assurer la mise en œuvre de la politique d'attribution de DOI ; - Accompagnement des demandeurs. Attribuer un DOI ne se résume pas à une mise à disposition d'un objet scientifique immédiate et évidente ; - Le guichet unique met en place un protocole d'accompagnement adapté à chaque type de demandeur (chercheur ou ingénieur faisant seul la démarche d'une part, les plateformes d'autre part) ; - Le guichet unique s'assure que les conditions d'attribution sont remplies quel que soit le type de demandeur.

4.2 Aspects techniques

Développement d'une API pour la génération et la gestion des DOI	
Recommandation	Développer une API pour la génération et la gestion des DOI à l'INRA.
Contexte/Enjeux pour l'INRA	- L'INRA dispose d'un compte unique pour la déclaration et la gestion des DOI. Le développement de l'API évitera de communiquer le compte à tous les demandeurs avec les risques que cela peut comporter (non-respect des règles de génération des DOI pouvant entraîner « l'écrasement » de DOI existants, etc.) ; - L'API s'assurera que chaque DOI attribué au sein de l'INRA est unique ; - L'API pourra embarquer des fonctions supplémentaires telles que la génération automatique de citation pour chaque DOI, des statistiques diverses, mise à jour d'un annuaire INRA, etc.

Développement de formulaires de gestion des demandes	
Recommandation	Développer des formulaires pour gérer les demandes individuelles de : - génération de nouveau DOI ; - changement de landing page ; - mise à jour des métadonnées ; - inactivation de DOI.
Contexte/Enjeux pour l'INRA	- Faciliter l'accès aux services d'attribution et de gestion des DOI.

Développement d'un service d'assistance	
Recommandation	- Mettre en place un service d'accompagnement des demandeurs de DOI - Mettre en place un service d'assistance, de gestion des incidents....
Contexte/Enjeux pour l'INRA	- Assurer une continuité de service ; - Assurer un service de qualité.

4.3 Description des cas d'utilisation

4.3.1 Présentation générale et glossaire

Le groupe de travail a réfléchi à 5 cas d'utilisation. Ils concernent les demandes de :

- génération d'un nouveau DOI ;
- changement de la landing page ;
- mise à jour des métadonnées ;
- inactivation de DOI ;
- assistance.

Dans les paragraphes suivants, nous allons détailler ces cas d'utilisation. Seul le cas de gestion d'une demande d'assistance ne sera pas détaillé. Celui-ci pourra être précisé lors de la mise en œuvre effective du guichet unique de gestion des attributions de DOI.

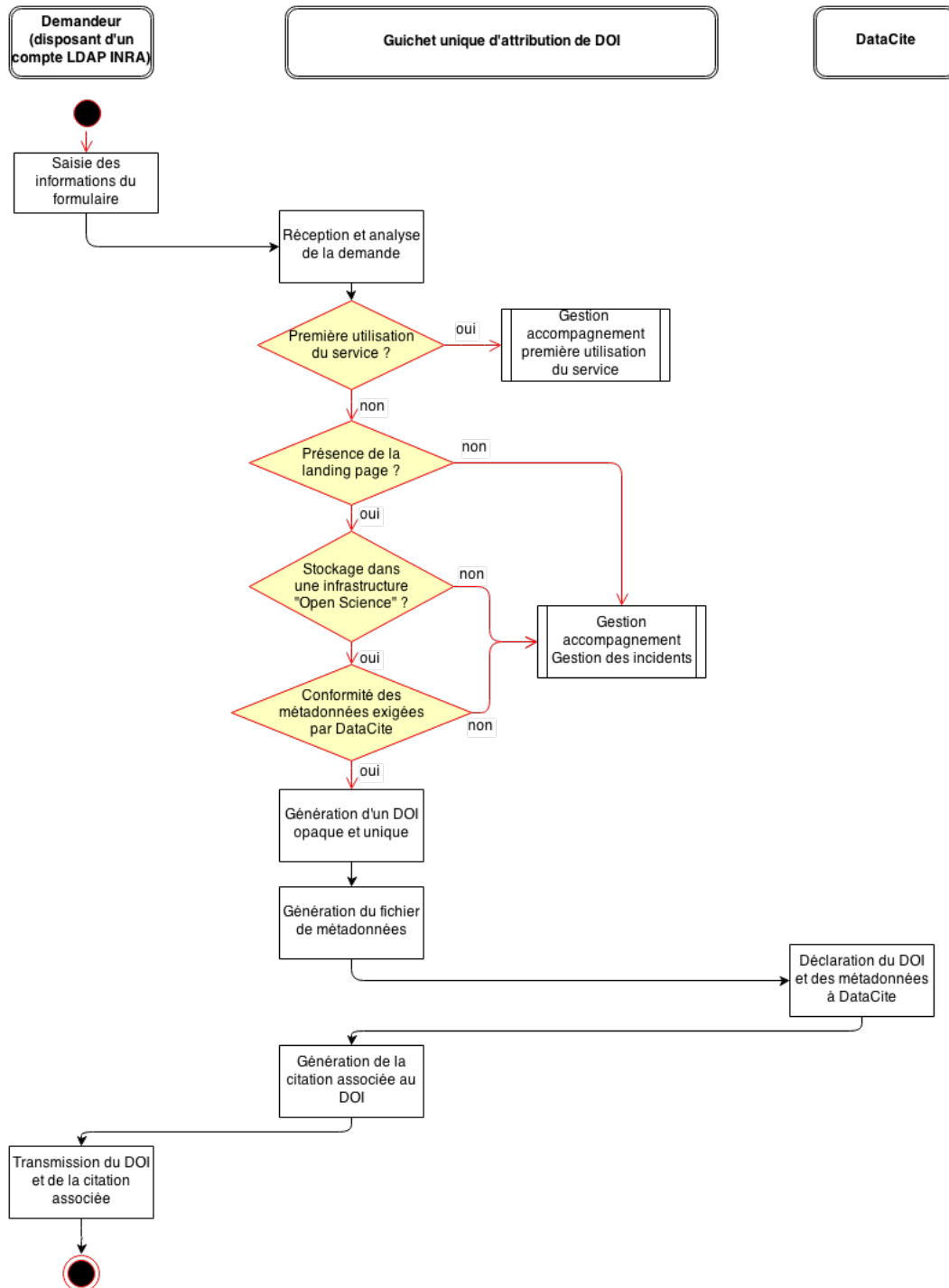
Glossaire	
Acteur	Les acteurs d'un scénario sont ceux qui sont sollicités pendant l'exécution du cas d'utilisation, ou pour qui l'exécution du scénario produit un effet observable.
Personnes disposant d'un compte LDAP INRA	- Personnels titulaires INRA - Personnels non titulaires INRA (MOO⁶, CDD, Boursier de thèse, stagiaire) - Personnels PEPS sauf : - Prestataire de service - Tout missionnaire ULYSSE
Pré-conditions	Les pré-conditions définissent ce qui doit être vrai en amont du cas d'utilisation pour que celui-ci puisse démarrer. Il n'y pas toujours de pré-conditions pertinentes.
Post-conditions	Les post-conditions définissent ce qui doit être vrai lorsque le cas d'utilisation se termine. Il n'y pas toujours de post-conditions pertinentes.
Scénarii	Les scénarii sont au nombre de 2 : - le scénario nominal qui satisfait les objectifs des acteurs par le chemin le plus direct de succès, - des alternatives qui couvrent les autres scénarii de fin normale ou d'échec.

⁶ Main d'œuvre occasionnelle

4.3.2 Cas d'utilisation : demande d'un DOI

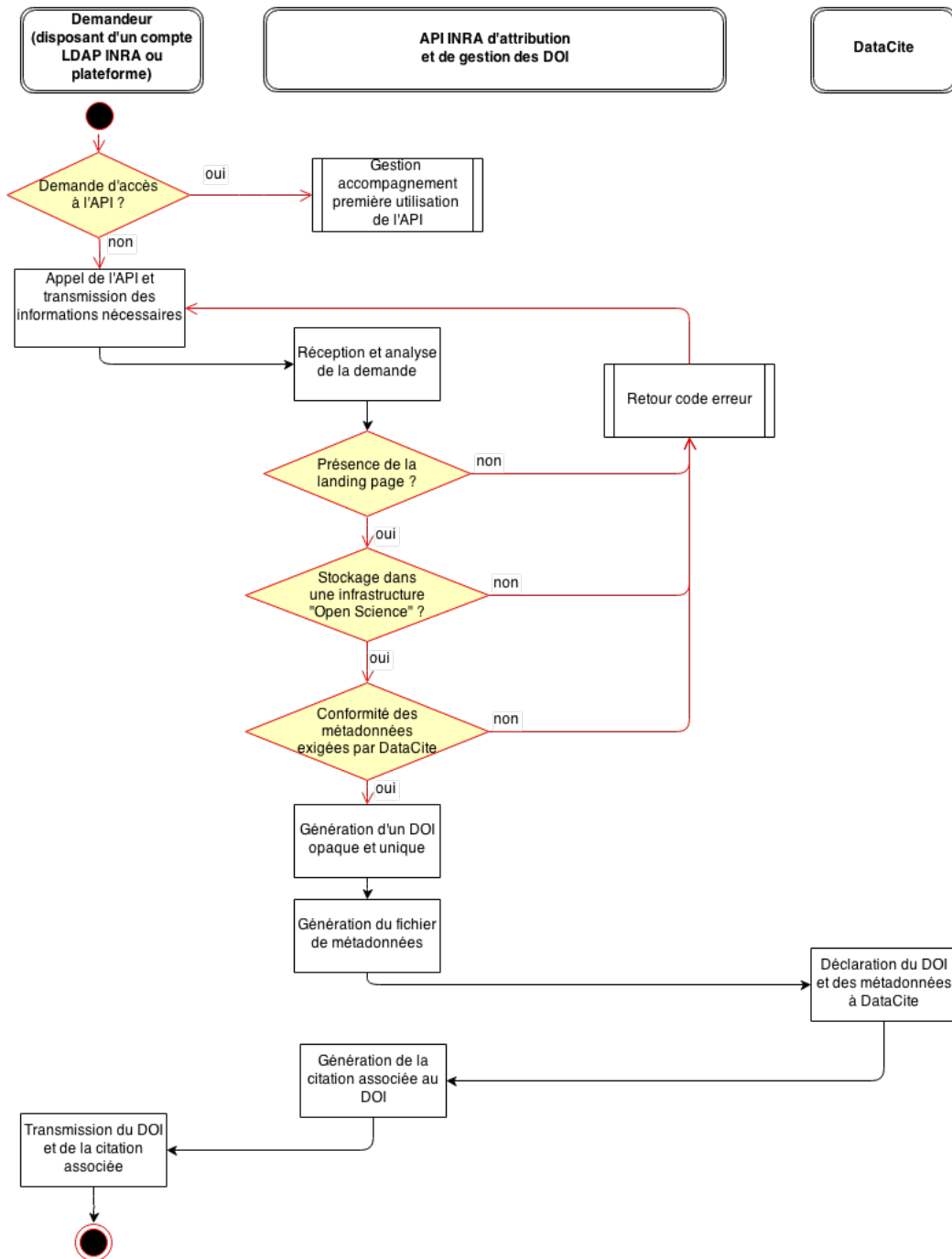
Cas 1	Demande d'un DOI via le formulaire de gestion des demandes
Acteurs principaux	- Le demandeur (une personne disposant d'un compte LDAP INRA) ; - Le guichet unique d'attribution de DOI ; - DataCite.
Pré-conditions	Le demandeur est une personne et doit disposer d'un compte LDAP INRA.
Post-conditions	
Données/informations en entrée	- L'URL vers la landing page (page descriptive) de l'objet scientifique ; - L'infrastructure de stockage de l'objet scientifique labellisée « Open Science » ; - Les métadonnées obligatoires requises par DataCite (Title, Creator(s), Publisher, Publication Year) ; - Les métadonnées obligatoires requises par l'INRA ; - Les métadonnées recommandées par DataCite si disponibles ; - Autres métadonnées si disponibles.
Données/informations en sortie	- DOI opaque ; - Citation associée au DOI.
Scénario nominal	1. Le demandeur remplit le formulaire de demande de DOI ; 2. Vérification de la conformité de la demande par le guichet unique : a. présence de la landing page ; b. stockage de l'objet scientifique dans une infrastructure labellisée « Open Science » ; c. conformité des métadonnées exigées par DataCite ; 3. Le guichet unique : a. génère un DOI opaque et unique ; b. génère le fichier des métadonnées associées conforme au schéma de DataCite ; c. déclare le DOI et les métadonnées associées à DataCite ; d. génère la citation associée au DOI ; 4. Le guichet unique transmet le DOI et la citation associée au demandeur.
Alternatives	
Commentaires	
Règles de gestion	

Cas d'utilisation 1 : demande d'un DOI via le formulaire de gestion des demandes



Cas 2	Demande d'un DOI via l'appel de l'API INRA
Acteurs principaux	- Le demandeur (une plateforme, une personne disposant d'un compte LDAP INRA) ; - L'API INRA d'attribution et de gestion des DOI ; - DataCite.
Pré-conditions	Le demandeur (plateforme ou personne) doit disposer d'une clé permettant de l'identifier et de l'authentifier.
Post-conditions	
Données/informations en entrée	- L'URL vers la landing page (page descriptive) de l'objet scientifique ; - L'infrastructure de stockage de l'objet scientifique labellisée « Open Science » ; - Un fichier conforme au schéma de DataCite et contenant : • les métadonnées obligatoires requises par DataCite (Title, Creator(s), Publisher, Publication Year) ; • les métadonnées obligatoires requises par l'INRA ; • les métadonnées recommandées par DataCite si disponibles ; • autres métadonnées si disponibles.
Données/informations en sortie	- DOI opaque ; - Citation associée au DOI.
Scénario nominal	1. Le demandeur appelle l'API INRA en transmettant les informations nécessaires ; 2. L'API INRA vérifie la conformité de la demande : a. présence de la landing page ; b. stockage de l'objet scientifique dans une infrastructure labellisée « Open Science » ; c. conformité des métadonnées exigées par DataCite ; 3. L'API INRA : d. génère un DOI opaque ; e. l'API INRA déclare le DOI et les métadonnées associées à DataCite ; f. l'API INRA génère la citation associée au DOI ; 4. L'API Inra transmet le DOI et la citation associée au demandeur.
Alternatives	
Commentaires	
Règles de gestion	

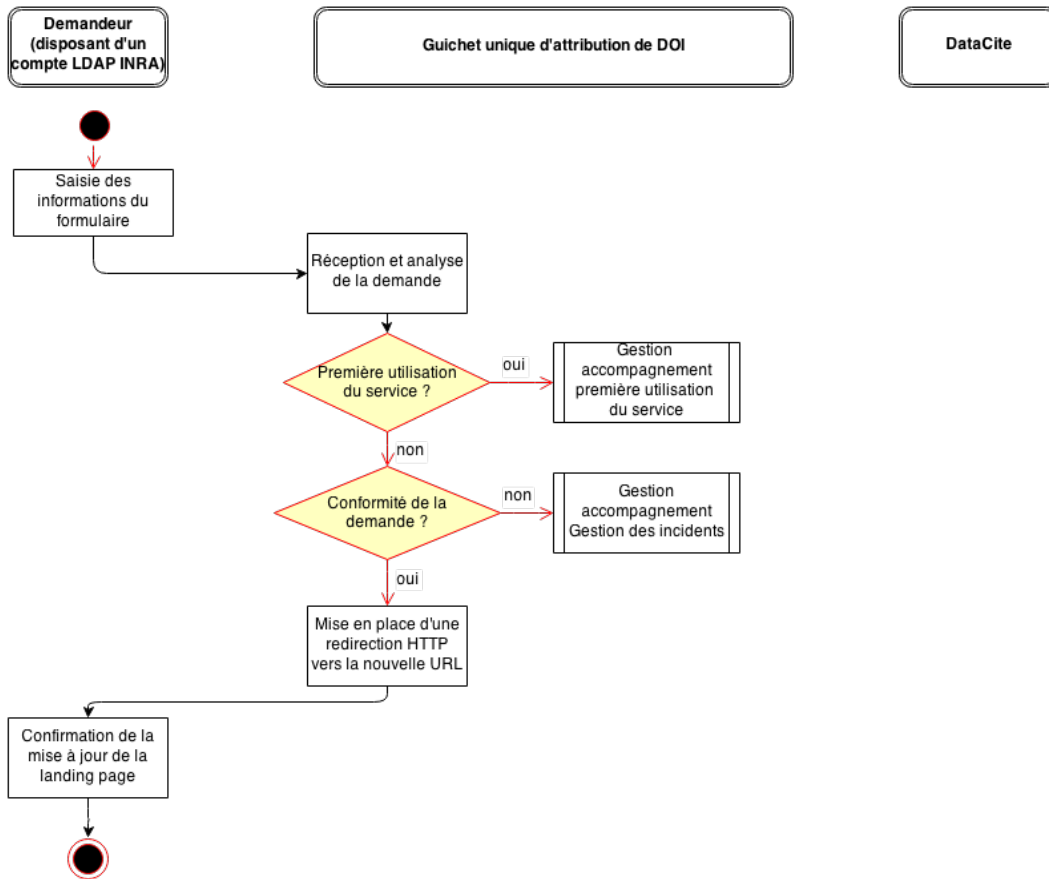
Cas d'utilisation 2 : demande d'un DOI via l'appel de l'API INRA



4.3.3 Cas d'utilisation : changement de la landing page

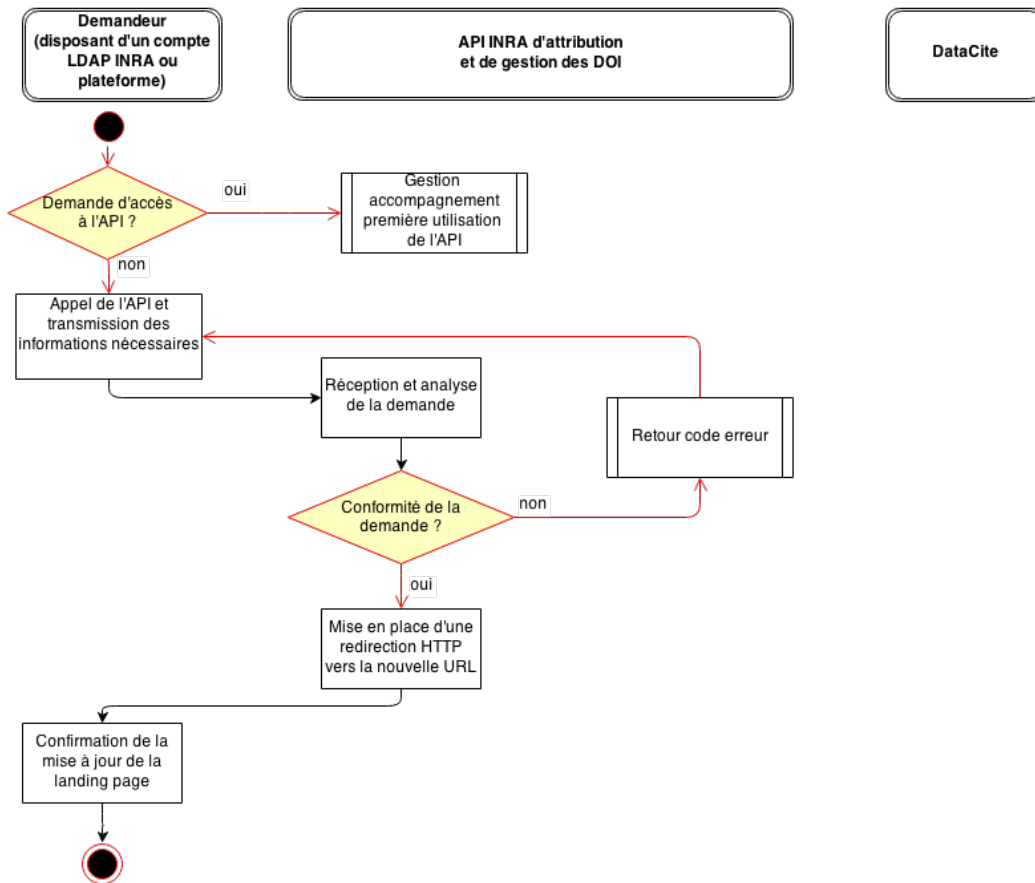
Cas 3	Changement de la landing page via le formulaire de gestion des demandes
Acteurs principaux	- Le demandeur (une personne disposant d'un compte LDAP INRA) ; - Le guichet unique d'attribution de DOI.
Pré-conditions	Le demandeur est une personne et doit disposer d'un compte LDAP INRA.
Post-conditions	
Données/informations en entrée	- Le DOI concerné par la demande de changement de la landing page ; - La nouvelle adresse URL de la landing page.
Données/informations en sortie	Confirmation du bon déroulement de la demande de changement.
Scénario nominal	1. Le demandeur remplit le formulaire de changement de la landing page ; 2. Vérification de la conformité de la demande par le guichet unique ; 3. Une redirection HTTP de l'ancienne URL vers la nouvelle adresse URL est mise en place ; 4. Une confirmation est transmise au demandeur.
Alternatives	
Commentaires	
Règles de gestion	

Cas d'utilisation 3 : changement de la landing page via le formulaire de gestion des demandes



Cas 4	Changement de la landing page via l'appel de l'API Inra
Acteurs principaux	- Le demandeur (une plateforme, une personne disposant d'un compte LDAP INRA) ; - L'API INRA d'attribution et de gestion des DOI.
Pré-conditions	Le demandeur (plateforme ou personne) doit disposer d'une clé permettant de l'identifier et de l'authentifier.
Post-conditions	
Données/informations en entrée	- Le DOI concerné par la demande de changement ; - La nouvelle adresse URL de la landing page.
Données/informations en sortie	Confirmation du bon déroulement de la demande de changement.
Scénario nominal	1. Le demandeur appelle l'API INRA en transmettant les informations nécessaires ; 2. Vérification de la conformité de la demande ; 3. Une redirection HTTP de l'ancienne URL vers la nouvelle adresse URL est mise en place ; 4. Une confirmation est transmise au demandeur.
Alternatives	
Commentaires	
Règles de gestion	

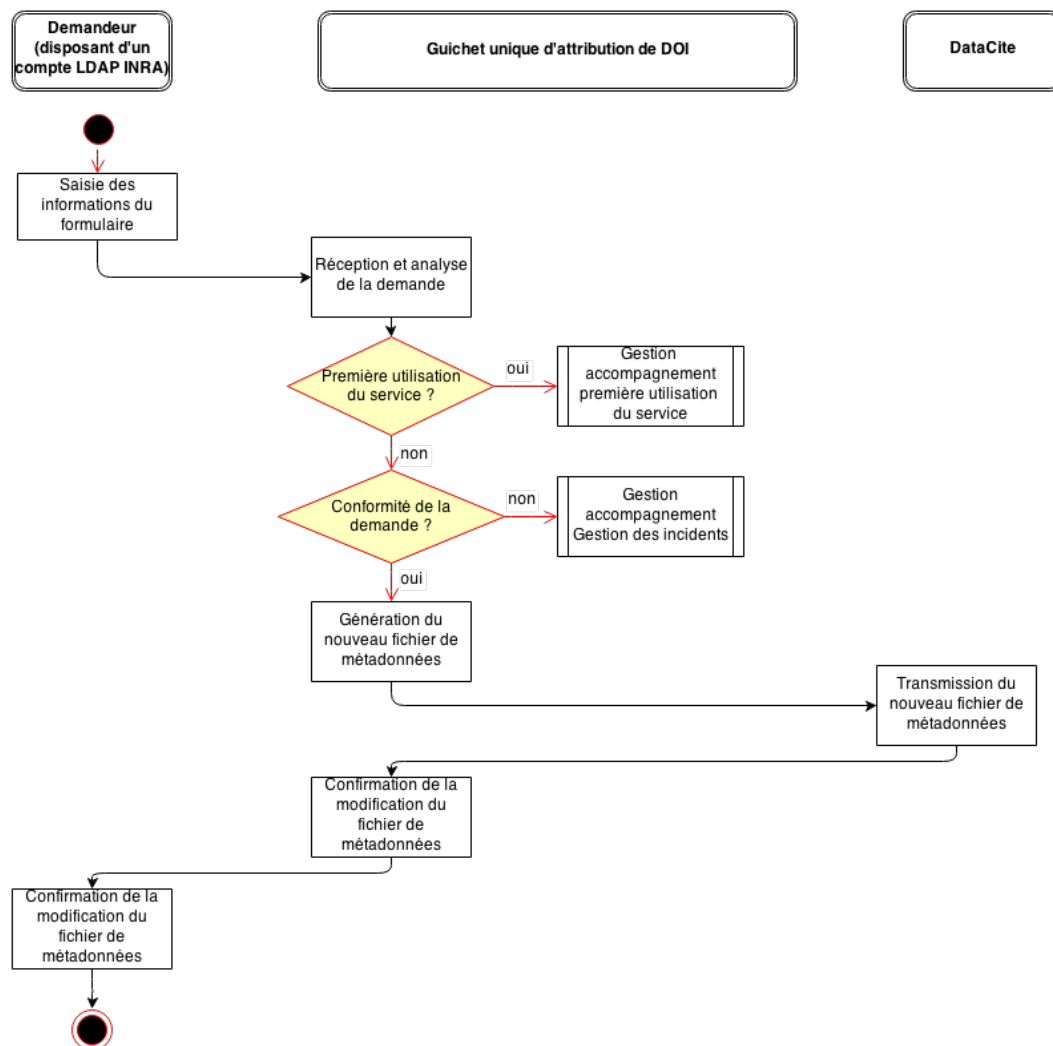
Cas d'utilisation 4 : changement de la landing page via l'appel de l'API Inra



4.3.4 Cas d'utilisation : mise à jour des métadonnées

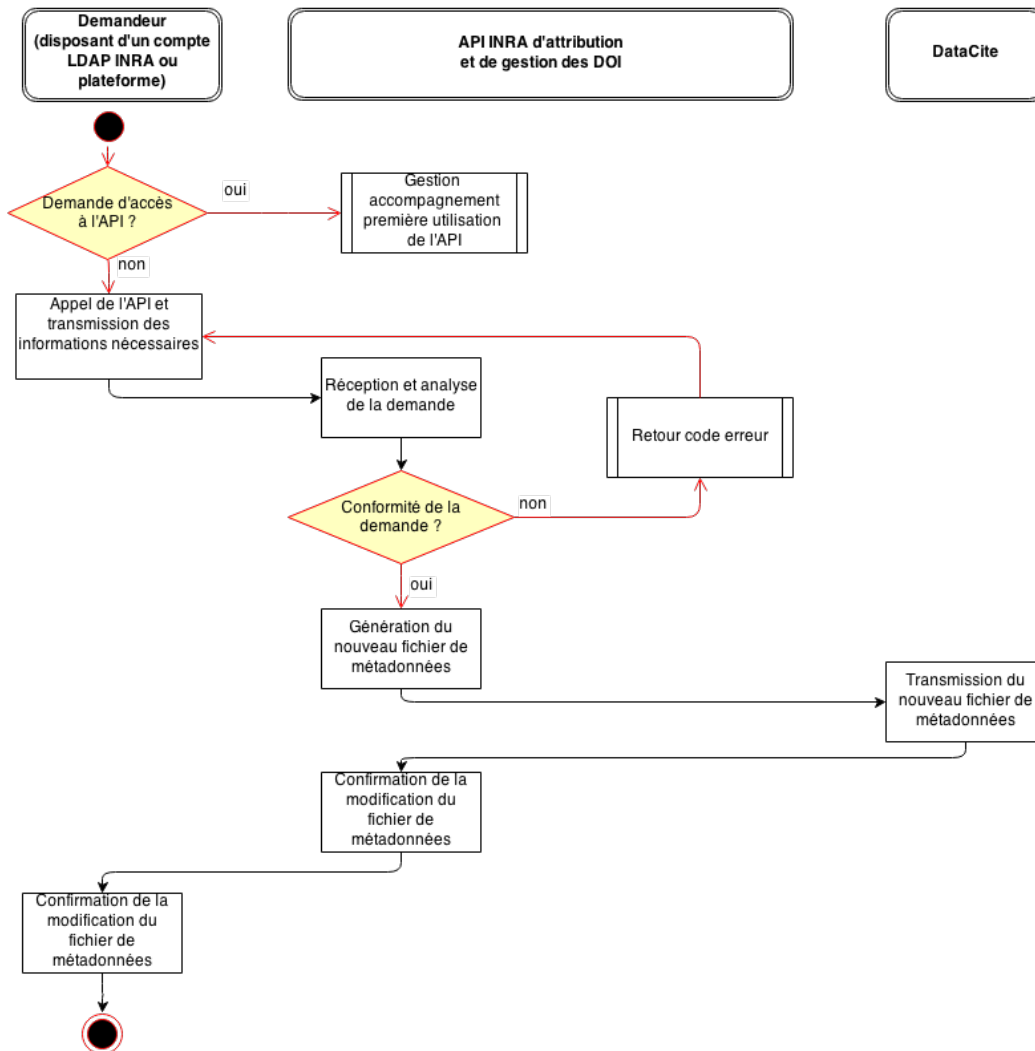
Cas 5	Mise à jour (ajout/modification) des métadonnées associées au DOI via le formulaire de gestion des demandes
Acteurs principaux	- Le demandeur (une personne disposant d'un compte LDAP INRA) ; - Le guichet unique d'attribution de DOI ; - DataCite.
Pré-conditions	Le demandeur est une personne et doit disposer d'un compte LDAP INRA.
Post-conditions	
Données/informations en entrée	- Le DOI concerné par la demande de mise à jour des métadonnées ; - Les nouvelles métadonnées.
Données/informations en sortie	Confirmation du bon déroulement de la demande de changement.
Scénario nominal	1. Le demandeur remplit le formulaire de mise à jour des métadonnées ; 2. Vérification de la conformité de la demande par le guichet unique ; 3. Le guichet unique : a. génère le nouveau fichier de métadonnées associées conforme au schéma de DataCite ; b. transmet le nouveau fichier de métadonnées à DataCite ; 4. Une confirmation est transmise au demandeur.
Alternatives	
Commentaires	Le changement des métadonnées obligatoires requises par DataCite (Title, Creator(s), Publisher, Publication Year) induit forcément un changement dans la citation. Par défaut, le service n'autorisera pas leur changement. Un traitement au cas par cas sera à envisager.
Règles de gestion	

Cas d'utilisation 5 : mise à jour (ajout/modification) des métadonnées associées au DOI via le formulaire de gestion des demandes



Cas 6	Mise à jour (ajout/modification) des métadonnées associées au DOI via l'appel de l'API INRA
Acteurs principaux	- Le demandeur (une plateforme, une personne disposant d'un compte LDAP INRA) ; - L'API INRA d'attribution et de gestion des DOI ; - DataCite.
Pré-conditions	Le demandeur (plateforme ou personne) doit disposer d'une clé permettant de l'identifier et de l'authentifier.
Post-conditions	
Données/informations en entrée	- Le DOI concerné par la demande de mise à jour des métadonnées ; - Les nouvelles métadonnées.
Données/informations en sortie	Confirmation du bon déroulement de la demande de changement.
Scénario nominal	1. Le demandeur appelle l'API INRA en transmettant les informations nécessaires ; 2. Vérification de la conformité de la demande ; 3. L'API INRA : a. génère le nouveau fichier de métadonnées associées conforme au schéma de DataCite ; b. transmet le nouveau fichier de métadonnées à DataCite ; 4. Une confirmation est transmise au demandeur.
Alternatives	
Commentaires	Le changement des métadonnées obligatoires requises par DataCite (Title, Creator(s), Publisher, Publication Year) induit forcément un changement dans la citation. Par défaut, le service n'autorisera pas leur changement. Un traitement au cas par cas sera à envisager.
Règles de gestion	

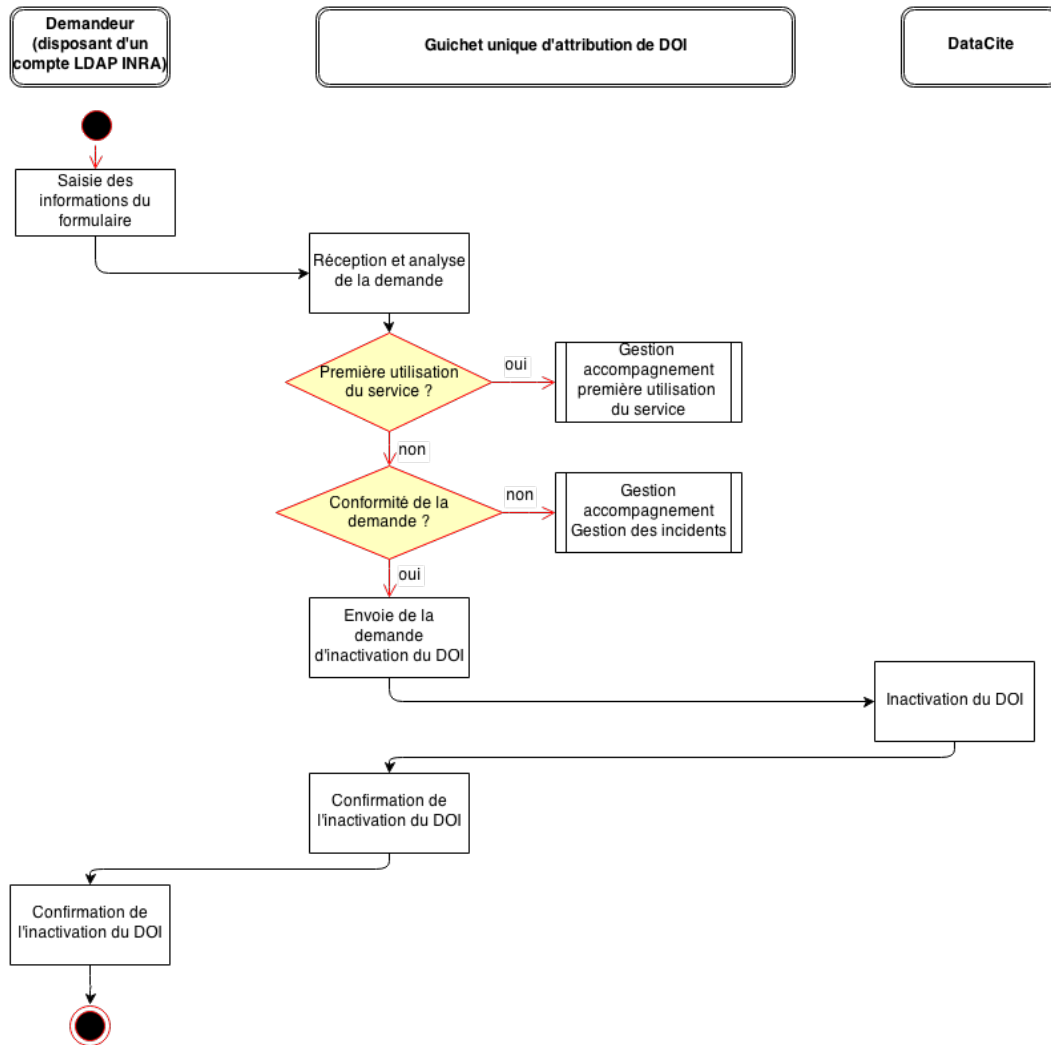
Cas d'utilisation 6 : mise à jour (ajout/modification) des métadonnées associées au DOI via l'appel de l'API INRA



4.3.5 Cas d'utilisation : inactivation d'un DOI

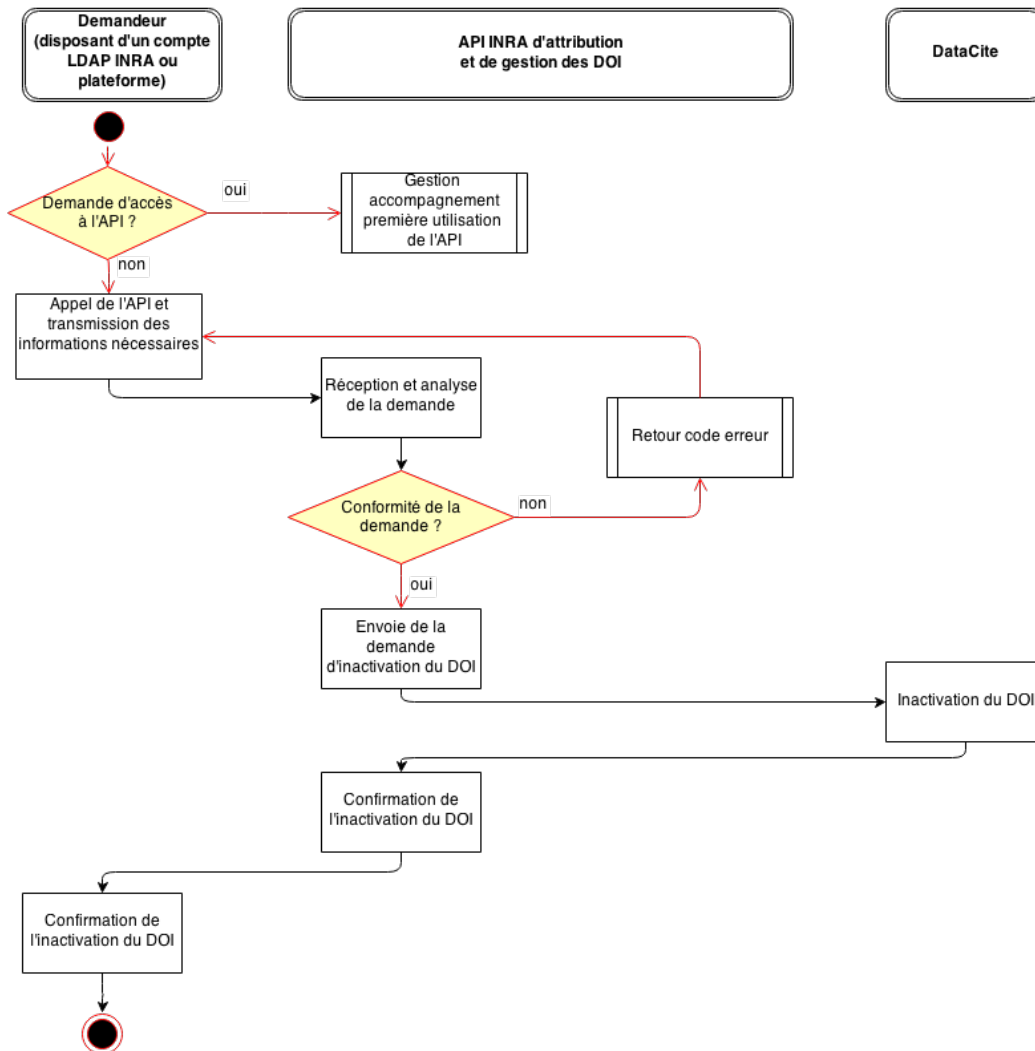
Cas 7	Inactivation d'un DOI via le formulaire de gestion des demandes
Acteurs principaux	- Le demandeur (une personne disposant d'un compte LDAP INRA) ; - Le guichet unique d'attribution de DOI ; - DataCite.
Pré-conditions	Le demandeur est une personne et doit disposer d'un compte LDAP INRA.
Post-conditions	
Données/informations en entrée	Le DOI concerné par la demande de mise à jour.
Données/informations en sortie	Confirmation du bon déroulement de la demande de changement.
Scénario nominal	1. Le demandeur remplit le formulaire de demande de DOI ; 2. Vérification de la conformité de la demande par le guichet unique ; 3. Le guichet unique envoie la demande d'inactivation à DataCite ; 4. Une confirmation est transmise au demandeur.
Alternatives	
Commentaires	
Règles de gestion	

Cas d'utilisation 7 : inactivation d'un DOI via le formulaire de gestion des demandes



Cas 8	Inactivation d'un DOI via l'appel de l'API INRA
Acteurs principaux	- Le demandeur (une plateforme, une personne disposant d'un compte LDAP INRA) ; - L'API INRA d'attribution et de gestion des DOI ; - DataCite.
Pré-conditions	Le demandeur (plateforme ou personne) doit disposer d'une clé permettant de l'identifier et de l'authentifier.
Post-conditions	
Données/informations en entrée	Le DOI concerné par la demande de mise à jour.
Données/informations en sortie	Confirmation du bon déroulement de la demande de changement.
Scénario nominal	1. Le demandeur appelle l'API INRA en transmettant les informations nécessaires ; 2. Vérification de la conformité de la demande ; 3. L'API INRA envoie la demande d'inactivation à DataCite ; 4. Une confirmation est transmise au demandeur.
Alternatives	
Commentaires	
Règles de gestion	

Cas d'utilisation 8 : inactivation d'un DOI via l'appel de l'API INRA



5 Projets pilotes et prochaines étapes

5.1 Les projets pilotes

5.1.1 Présentation et objectifs

Le groupe de travail a identifié 7 projets pilotes pour expérimenter et valider le service d'attribution des DOI. Ces projets pilotes ont plusieurs objectifs :

- Valider les recommandations faites par le groupe de travail concernant l'attribution de DOI ;
- Valider l'offre de service préconisée et en préciser les contours ;
- Identifier les prochaines actions et décisions pour la mise en œuvre prochaine du service d'attribution de DOI à l'INRA.

Ces projets concernent plusieurs familles de données et sont de volumétries différentes. Le tableau ci-dessous décrit brièvement ces projets.

Plateforme/équipe	Famille de données	Nature des données	Volumétrie
1. Archive SUNRISE 2. Archive BBRIC	Omiques et ressources génétiques	- Données brutes issues de séquençage à haut débit	Plusieurs To
3. GnpIS	Expérimentation et d'observation	- Données de phénotypage	Une centaine d'expérimentations
4. DASI 5. OQALI	Enquête et de cohortes	- Données brutes - Données retravaillées	15 Go
6. UMR MIA 7. AnaEE-France	Textuelle	- Ontologie - Thésaurus	Qq Mo

Un questionnaire (fourni en annexe 2) a été envoyé aux contacts des différents projets pour mieux connaître les données candidates et accompagner les projets pilotes.

5.1.2 Premières actions / premier bilan

- Le projet pilote de l'UMR MIA (famille de données « textuelle ») a soulevé des problématiques juridiques de datamining. Celles-ci sont en cours d'étude. L'attribution de DOI pour ces données n'est donc pas possible pour le moment. Le guichet unique de gestion des attributions de DOI bénéficiera des conclusions de ce travail.
- Un test complet d'attribution de DOI a pu être réalisé pour un projet pilote de la famille de données « omiques ». Des DOI ont été attribué à 4 jeux de données de l'archive BBRIC. Ces DOI ont été déclarés avec succès sur la plateforme de test de DataCite. Ce test a été l'occasion de :
 - Réaliser un script de génération de suffixes uniques pour les DOI. Celui-ci est opérationnel et pourra être réutilisé par le service d'attribution de DOI ;
 - Valider la procédure de génération de métadonnées de la plateforme BBRIC. Ces métadonnées sont conformes au schéma de métadonnées de DataCite. Cette plateforme est donc un bon candidat à l'utilisation de la future API INRA d'attribution et de gestion de DOI.

5.1.3 Poursuite des projets pilotes

Développement de formulaires de gestion des demandes individuelles de DOI :

- Une première version du formulaire de demande d'attribution de DOI est déjà disponible. Il permet de saisir et de générer les métadonnées conformément au schéma de DataCite. Le service d'attribution de DOI pourra réutiliser ce formulaire et poursuivre son développement afin de l'intégrer dans le site web ou le portail dédié à la gestion des DOI à l'INRA ;
- Les 2 projets pilotes de la famille de données « textuelle » de l'équipe AnaEE-Franceet de l'UMR MIA permettront d'expérimenter le déroulement des demandes individuelles d'attribution de DOI. Ces projets testeront et valideront le formulaire de demande d'attribution de DOI (aspects techniques, ergonomiques etc...).

Développement d'une API pour la génération et la gestion des DOI :

- Un premier script de génération de suffixes uniques est déjà disponible. D'autres développements sont nécessaires afin de mettre en œuvre l'API INRA de génération et de gestion des DOI.
- Les projets pilotes des plateformes BBRIC, SUNRISE, DASI et OQALI vont permettre de tester l'API INRA une fois que celle-ci sera disponible. Ces projets permettront de s'assurer qu'une attribution de DOI peut être gérée de manière automatique de bout en bout (de la génération d'un DOI unique à l'enregistrement dans la plateforme de DataCite).

5.2 Prochaines étapes

Le tableau ci-dessous liste les prochaines étapes pour mettre en œuvre le service d'attribution de DOI à l'INRA. Il précise leur état d'avancement.

5.2.1 Décisions et actions préalables à l'ouverture du guichet de gestion et d'attribution de DOI

À court terme, il est impératif que l'INRA :

- Labellise des infrastructures « Open Science » pour le stockage des données à identifier avec des DOI ;
- Crée le guichet unique d'attribution de DOI et missionne les personnes en charge de ce guichet. Le nombre de personnes à missionner dépendra de la volumétrie des demandes à traiter⁷.

Étape	Description	État d'avancement	Acteurs responsables
1. Signature d'un contrat d'attribution de DOI	Signer un contrat avec l'INIST-CNRS permettant à l'INRA d'être centre de données.	 Réalisé	DV-IST
2. Labellisation d'infrastructures de stockage « Open Science »	- Désigner ou créer des infrastructures de stockage « Open Science » ; - Doter ces infrastructures de moyens adéquats⁸ ; - Labelliser ces infrastructures pour les rendre facilement identifiables.	 À faire	Management Inra
3. Création du guichet unique d'attribution de DOI	- Valider les propositions du groupe de travail ; - Missionner les personnes en charge du guichet unique d'attribution de DOI⁷ : - Un responsable pilotant le guichet ; - Une ou plusieurs personnes en charge du traitement des demandes ;	 À faire	Management Inra
4. Mise en place de l'accompagnement des demandes	- Mettre en œuvre un réseau de personnes et de compétences permettant d'accompagner au mieux les demandeurs par la désignation de correspondants « Open Science »⁷.	 À faire	Management Inra + Guichet unique

⁷ Voir le tableau « Profil des personnes en charge du guichet unique d'attribution de DOI »

⁸ Voir avec les personnes responsables de ces structures pour connaître les moyens supplémentaires à mettre en œuvre si besoin

Profil des personnes en charge du guichet unique d'attribution de DOI	
Un responsable	Missions : Piloter le guichet unique d'attribution de DOI. Cette personne est responsable de la réalisation de l'ensemble des services délivrés par le guichet, de leurs mises en œuvre et de leurs adéquations aux besoins des utilisateurs. Elle veille au maintien en conditions opérationnelles des services délivrés par le guichet. Profil : Cette personne connaît et maîtrise les enjeux des DOI et de l'Open Science en général. Elle dispose des connaissances informatiques nécessaires pour conduire la maîtrise d'œuvre des projets de développement et peut réaliser tout ou partie des développements informatiques nécessaires.
Une ou plusieurs personnes en charge du traitement des demandes	Missions : Prendre en charge et traiter les demandes adressées au guichet unique d'attribution de DOI, notamment les demandes individuelles. Profil : Professionnels de l'IST connaissant les enjeux du DOI et de l'Open Science en général. Ces personnes disposent de connaissances sur les standards de métadonnées en général et sur le schéma des métadonnées de DataCite en particulier.
Des correspondants « Open Science »	Missions : Chaque correspondant « Open Science » doit permettre d'aider les demandeurs dans leur démarche. Ils viennent en appui au guichet unique d'attribution et répondent à leurs éventuelles sollicitations. Profils : - Expert juriste spécialiste des questions de propriétés intellectuelles, des contrats de recherches, de l'ouverture des données publiques.... - Expert des problématiques et bonnes pratiques associées à chaque famille de données.

5.2.2 Gestion des demandes individuelles via des formulaires

Étape	Description	État d'avancement	Acteurs responsables
1. Création d'un site web ou d'un portail	Créer un site Web ou un portail pour le guichet unique d'attribution de DOI : - pour informer les demandeurs ; - pour donner accès aux formulaires de demande (cf. cas d'utilisation) ;	 À faire	Guichet unique
2. Création d'un script de génération de DOI	Créer un script permettant de générer des DOI dont : - Le préfixe est le code de l'INRA en tant que centre de données ; - Le suffixe est opaque et unique.	 Réalisé Le script pourra être intégré à l'API INRA de gestion des DOI	Guichet unique
3. Création des formulaires	Créer les formulaires de demande de : - génération de nouveau DOI ; - changement de landing page ; - mise à jour des métadonnées ; - inactivation de DOI.	 Commencé Un formulaire permettant de saisir les métadonnées et de générer un fichier conforme au schéma de DataCite est déjà disponible.	Guichet unique
4. Rédaction de documentations	Réaliser l'ensemble des documentations nécessaires à l'accompagnement des utilisateurs des formulaires. Notamment : - Un guide de bonnes pratiques pour l'attribution de DOI ; - Un guide d'utilisation des formulaires.	 À faire	Guichet unique
5. Validation du service à l'aide de projets pilotes	Tester et valider les développements informatiques réalisés ainsi que l'accompagnement mis en œuvre.	 Commencé Les projets pilotes de l'équipe AnaEE-Franceet de l'UMR MIA permettront de valider cette partie du service.	Guichet unique + plateformes/équipes concernées

5.2.3 API pour la génération et la gestion des DOI

Étape	Description	État d'avancement	Acteurs responsables
1. Création d'une API INRA	Créer une API « INRA » encapsulant l'API DataCite permettant l'attribution d'un DOI et la gestion des métadonnées associées à un DOI.	 À faire	Guichet unique
2. Rédaction de documentations	Réaliser l'ensemble des documentations nécessaires à l'accompagnement des utilisateurs de l'API INRA. Notamment : - Un guide de bonnes pratiques pour l'attribution de DOI ; - Un guide d'utilisation de l'API.	 À faire	Guichet unique
3. Validation du service à l'aide de projets pilotes	Tester et valider les développements informatiques réalisés ainsi que l'accompagnement mis en œuvre.	 En cours Les données des plateformes BBRIC, SUNRISE, DASI et OQALI serviront de pilotes pour valider cette partie du service.	Guichet unique + plateformes concernées

Annexes

1 Digital Object Identifier (DOI)

Ce paragraphe rappelle les grands principes des DOI. Il a pour objectif de faciliter la compréhension de l'ensemble du document mais il ne se veut pas exhaustif sur le sujet. Nous encourageons le lecteur souhaitant en savoir plus à consulter les références décrites dans ce document.

1.1 Qu'est-ce qu'un DOI ?

Un **D**igital **O**bject Identifier est un identifiant pérenne et unique permettant de référencer, citer et fournir un lien stable vers un objet numérique. Il fournit un lien stable à des ressources en ligne (données, documents, images, documents audio...). Les DOI reposent sur le système Handle⁹.

1.2 Un DOI, pour quoi faire / pour quelles données ?

Un DOI permet :

- d'identifier de façon pérenne un objet numérique ;
- de citer de façon fiable et pérenne un objet numérique ;
- de faciliter l'accès, le partage et la réutilisation de l'objet numérique ;
- le crosslinking à travers des objets variés, notamment entre publications et données.

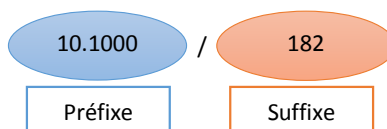
Les DOI permettent d'identifier une grande variété d'objets scientifiques : jeux de données, images, données audiovisuelles, etc...

Les DOI sont depuis longtemps associés aux articles de périodique à travers CrossRef¹⁰.

Important : L'objectif premier de l'identification d'un objet scientifique à l'aide d'un DOI est de le rendre **citable**. Tous les choix qu'il faudra faire concernant cet objet seront fait dans le respect de ce principe.

1.3 Syntaxe d'un DOI

Chaque DOI est unique et est composé de deux parties :



a. Préfixe

C'est un numéro à 6 chiffres attribué par l'agence DataCite (en France l'INIST-CNRS) qui délivre le DOI. Il identifie le déclarant (l'INRA). Par exemple, « 1000 » identifie l'International DOI Foundation.

b. Suffixe

Il peut être choisi par le déclarant (l'INRA) mais il repose néanmoins sur des règles communes. Chaque suffixe doit être unique et utiliser uniquement une liste de caractères autorisés :

A-Z a-z 0-9 - . _ : /

⁹<http://www.handle.net/>

¹⁰<http://www.alliancepermanentaccess.org/index.php/2012/08/13/crossref-joins-stm-datacite-statement/>

1.4 Conditions d'attribution de DOI

1.4.1 Qui attribue les DOI ?

L'attribution et la gestion des DOI est sous la responsabilité d'agences externes internationales (International DOI Foundation - IDF). DataCite est un consortium international, porté par des institutions nationales, dont le but est de proposer une infrastructure destinée à la promotion d'ensembles de données et d'informations non textuelles. DataCite est habilité à attribuer des identifiants DOI à diverses ressources numériques. En France, DataCite est représenté par l'INIST-CNRS ayant ainsi le statut d'agence DOI.

À l'INRA, la DV-IST est le guichet interne de gestion de l'attribution des identifiants numériques. Un contrat d'attribution entre l'INIST-CNRS et la DV-IST a été passé. Il contient les modalités et conditions relatives à l'attribution de DOI à l'INRA.

1.4.2 Que peut-on identifier avec un DOI ?

Tout objet scientifique produit par l'INRA ou par d'autres producteurs de données pour lesquels l'INRA est mandaté peut être identifié.

1.4.3 Que comprend le service d'enregistrement de DOI offert par l'INIST-CNRS?

Le service offert par l'INIST-CNRS comprend :

- L'attribution de DOI aux objets scientifiques de l'INRA ou de ses partenaires pour que ceux-ci soient identifiés à titre permanent,
- L'enregistrement des métadonnées associées, rendues accessibles au public via la plateforme MDS (MetadataStore)¹¹ de DataCite.

La plateforme MDS est un service proposé par DataCite, destiné aux producteurs de données, permettant la création de DOI et l'enregistrement des métadonnées associées. Pour utiliser cette plateforme, il faut disposer d'un compte. L'INIST-CNRS a fourni un identifiant/mot de passe à l'INRA pour la création de DOI et la gestion de ses métadonnées sur la plateforme MDS de DataCite. À l'aide de ce compte et de l'API¹² DataCite, il est possible d'enregistrer les DOI. Grâce à ce compte, l'INRA est identifié comme datacenter par les services fournis par DataCite.

1.4.4 Coût d'attribution des DOI

Un droit annuel de 180 euros par an doit être versé par l'INRA. Ce droit couvre :

- L'enregistrement des noms de DOI pour une année ;
- Le stockage des métadonnées associées dans le système de l'INIST-CNRS et/ou le système central de DataCite.

Ce droit est indépendant du nombre de DOI enregistrés.

¹¹Création de DOI et enregistrement de métadonnées associées : <https://mds.datacite.org/>

¹²<https://mds.datacite.org/static/apidoc>

1.4.5 Conditions d'attribution d'un DOI

Tout objet scientifique peut potentiellement être identifié par un DOI. Cependant, avant d'attribuer des DOI, DataCite exige que le déclarant s'engage à :

- fournir les **métadonnées** obligatoires pour chaque objet scientifique. Ces métadonnées doivent respecter le schéma de métadonnées de DataCite¹³ et être fournies gratuitement. Les métadonnées doivent être librement accessibles dans le cadre d'une licence *Creative Commons*¹⁴ CCO ou tout autre licence permettant la disponibilité sans restriction et la réutilisation de ces métadonnées ;
- fournir une **page descriptive** ou "landing page" pour chaque objet scientifique. Tout objet scientifique enregistré avec un DOI doit être accessible via une URL. Il est recommandé que cette page contienne : la citation complète de l'objet scientifique, les métadonnées, toute information (logiciel, contexte ...) permettant de lire l'objet scientifique ainsi qu'une description des conditions d'accès à la donnée :
 - o Si la ressource identifiée est en accès ouvert, la page descriptive doit indiquer un lien (URL) d'accès direct à la ressource ;
 - o Si la ressource n'est pas en accès ouvert, la page descriptive doit expliquer les conditions d'accès et fournir un lien vers les services de paiement ou autres outils à partir desquels l'accès à la ressource est possible ;
- maintenir des **URL valides** vers les pages descriptives de l'objet scientifique ;
- avoir une **politique de curation et de conservation** pour les objets scientifiques : les objets scientifiques identifiés par des DOI doivent être stockés pendant une durée minimale de 10 ans.

La pérennité des DOI relève de la responsabilité du titulaire des données ou de l'entrepôt hébergeant les données qui doit assurer leur accessibilité et leur gestion sur le long terme. De même, la qualité des données identifiées par un DOI relève de l'entrepôt ou du titulaire des données, ainsi que leur curation et leur conservation.

1.5 Les services proposés par DataCite

Le consortium DataCite propose un certain nombre de services :

- DataCite Metadata Store (MDS) : Création de DOI et enregistrement de métadonnées associées : <https://mds.datacite.org/>
- DataCite Metadata Search : Recherche des métadonnées associées à l'objet scientifique enregistré dans DataCite : <http://search.datacite.org/ui>
- DataCite Statistics : Statistiques d'enregistrement et de résolution de DOI : <http://stats.datacite.org/>
- DOI Citation Formatter : En collaboration avec CrossRef, création de différents formats de citation pour les DOI DataCite et CrossRef : <http://crosscite.org/citeproc/>
- Content Negotiation : Possibilité d'obtenir des métadonnées dans des formats divers et/ou d'accéder automatiquement et directement à un objet plutôt que par la « landing page » : <http://data.datacite.org/static/index.html>
- DataCite Test Environment : Environnement de test DataCite : <http://test.datacite.org/>
- DataCite OAI Provider : Exposition des métadonnées en OAI-PMH : <http://oai.datacite.org/>

¹³<http://schema.datacite.org/meta/kernel-3/index.html>

¹⁴<http://creativecommons.fr/licences/les-6-licences/>

1.6 Les services mis à disposition autour des DOI

Un certain nombre de service sont disponibles autour des DOI. Pour en savoir plus, vous pouvez vous référer au wiki INRA :

<https://wiki.inra.fr/wiki/donneesrechercheist/Main/Services+autour+des+DOI?ticket=ST-58374-5TH9CFriQIneAqy9iVrF-cas>

2 Le questionnaire des projets pilotes



Partage des données de la recherche
Chantier DOI – Questionnaire projets pilotes



Mai / 2014

Questionnaire DOI – Projets pilotes

Afin de nous permettre de mieux vous accompagner, nous avons besoin d'en savoir plus sur votre plateforme et les données que vous envisagez d'identifier.

La plateforme

Nom de la plateforme

Description de la plateforme

Contact(s)

Famille(s) des données gérées par la plateforme

Nombre approximatif des données concernées par le projet d'identification

La plateforme dispose-t-elle des moyens techniques et humains pour préserver les données identifiées sur le long terme (10 ans minimum) ?

Avez-vous prévu une procédure pour assurer la continuité de service en cas d'évolution administrative (restructuration, fermeture de la plateforme, départ d'agents..) ?

Les données

Quels types d'objets scientifiques (données brutes, données élaborées, méthodes, etc.) sont dans le périmètre de votre projet d'attribution de DOIs?

S'agit-il uniquement de données liées à des publications scientifiques (articles, data pares, etc.)?

Comment sont actuellement stockées ces données ?

Les données stockées sur la plateforme sont-elles documentées?

Comment les données sont-elles documentées

- *Extraction automatique de métadonnées ?*
- *Documentation manuelle par le déposant ?*

Y a-t-il une liste minimale de métadonnées disponibles pour les données stockées sur la plateforme?

Quel(s) schéma(s) de métadonnées sont recommandés/imposés par la plateforme?

Dans quel(s) format(s) sont les données? Jugez-vous ces formats préservables sur le long terme?

Les données que vous envisagez d'identifier sont-elles dynamiques? Expliquez le caractère dynamique des données.

Dans le cas de données dynamiques, envisagez-vous d'identifier des archives régulières de ces données?

Ou au contraire, envisagez-vous d'identifier le jeu de données dynamique dans son ensemble en tant que produit défini par un savoir faire et une procédure, etc.?

Les données seront-elles accessibles publiquement et gratuitement sans restriction?

Les données seront-elles téléchargeables publiquement et gratuitement sans restriction?

Quelles restrictions d'accès aux données envisagez-vous?

Toutes les métadonnées seront-elles accessibles publiquement et gratuitement?

Disposez-vous d'un site web pour héberger les pages descriptives (landing pages) des données qui seront identifiées ?