



URI (Uniform Resource Identifier)

Recommandations pour la mise en œuvre d'URI pérennes à l'INRA



Septembre / 2016

Ce guide a été produit dans le cadre d'un projet inter-cati visant à promouvoir les technologies dites du Web sémantique ou Linked data à l'INRA.

CATI impliqués : ICAT (Ingénierie des Connaissances et Analyses textuelles), CODEX (Connaissances et Données Expérimentales), DIISCO (Développement Informatique pour l'Instrumentation, Suivi et Contrôle des propriétés et des processus) et CGI (Cati en Génomique-Info)

Contributeurs : Alaux Michael, Aubin Sophie, Buche Patrice, Dervaux Stéphane, Dibie Juliette, Durand Sophie, Dzalé Yeumo Esther, Neveu Pascal, Pommier Cyril, Tireau Anne

Sommaire

Introduction	3
Les recommandations générales.....	3
Les recommandations spécifiques	3
Identification des ressources publiées au format RDF (Resource Description Framework).....	3
Cas des données publiées au format RDF (ontologies, thésauri, bases de connaissances, etc.).....	3
Cas des datasets.....	3
Cas des ressources génétiques (accessions)	3
Cas des représentations Web de ressources physiques (exemple : les infrastructures de recherche).....	3

INTRODUCTION

Dans un contexte de données massives, hétérogènes et distribuées, la question de l'interopérabilité et de l'intégration des données est centrale. L'identification unique et pérenne des données qui sont partagées sur le Web participe de leur intégration et interopérabilité. En effet, les identifiants pérennes permettent d'éviter la confusion entre différentes versions d'une même ressource, facilitent la localisation, l'accès et l'utilisation des ressources, et permettent enfin de lier les données à l'échelle du Web, au-delà du système d'information dans lequel elles sont gérées. Pour garantir l'unicité et la pérennité des identifiants, il est indispensable de convenir d'une stratégie. Le W3C a édicté 4 principes pour le choix et l'implémentation des URI (<https://www.w3.org/TR/ld-bp/#HTTP-URIS>) : utiliser des URI http, faire pointer les URI vers une représentation web "machine-readable" de la ressource identifiée, ne pas inclure dans la syntaxe de l'URI des segments susceptibles de changer, utiliser ne URI opaques (les agents Web ne doivent pas pouvoir déduire les caractéristiques de l'objet identifié en analysant son URI). Plusieurs instances et organisations ont traduit ces bonnes pratiques en recommandations concrètes applicables à leur contexte. C'est par exemple le cas du gouvernement du Royaume Uni (<https://www.gov.uk/government/publications/designing-uri-sets-for-the-uk-public-sector>) ou de l'infrastructure INSPIRE pour les informations géospatiales (<http://inspire.ec.europa.eu/index.cfm/pageid/5120>). Le but de ce document est de proposer des recommandations pour l'identification des ressources numériques de l'INRA, en s'appuyant sur ces recommandations.

LES RECOMMANDATIONS GENERALES

- Se doter d'une stratégie institutionnelle pour l'identification pérenne des ressources INRA sur le Web
- Créer un comité d'experts inter-cati sur la question de l'identification pérenne des ressources INRA sur le Web. Ce comité sera chargé de faire évoluer la stratégie INRA décrite dans ce document en fonction des évolutions besoins, des technologies et du contexte de l'information scientifique et technique tout en garantissant la pérennité et l'unicité des identifiants déjà délivrés. Ce groupe Peut être constitué dans un premier temps du groupe qui a travaillé sur cette question.
- Dans le cadre du Web sémantique (et uniquement dans ce cadre-là), utiliser les URI pour identifier la représentation sur le Web des ressources et non les ressources en tant qu'objets physiques. Le cas échéant, les identifiants numériques pérennes réutiliseront les identifiants utilisés pour les objets physiques correspondants dans la mesure où ces derniers sont également pérennes.

LES RECOMMANDATIONS SPECIFIQUES

IDENTIFICATION DES RESSOURCES PUBLIEES AU FORMAT RDF (RESOURCE DESCRIPTION FRAMEWORK)

CAS DES DONNEES PUBLIEES AU FORMAT RDF (ONTOLOGIES, THESAURI, BASES DE CONNAISSANCES, ETC.)

- Se conformer à la syntaxe suivante pour les URI : <https://opendata.inra.fr/NomProjet/concept> dans laquelle NomProjet est le nom du projet (ontologie, thesaurus, base de connaissances, etc.), et concept la chaîne de caractères qui désigne un concept, une relation ou un littéral dans le projet en question.
- L'écriture des segments NomProjet et concept doit respecter les règles suivantes :
 - ◆ Ne rien contenir qui soit susceptible de changer avec le temps tel que le nom d'une structure
 - ◆ Ne rien contenir qui permette de déduire les caractéristiques techniques de l'objet identifié, tel que l'extension du fichier (.doc)
 - ◆ Utiliser des caractères alphanumériques et éviter les caractères spéciaux susceptibles d'être interprétés par un programme
 - ◆ Ne pas inclure de requête telle que <http://opendata.inra.fr/MonProjet/concept?nom=poussin>
- Si les données sont hébergées dans un entrepôt RDF accessible via un sparql endpoint, se rapprocher du comité URI (uri@versailles.inra.fr) pour mettre en oeuvre le déréférencement des URIs. Le déréférencement des URI permet d'assurer que chaque URI pointe vers une page descriptive de la ressource identifiée.

CAS DES DATASETS

- Nous recommandons l'utilisation des DOI accompagnés de métadonnées conformes à DataCite. Les demandes de DOI peuvent se faire de manière individuelle et ponctuelle via un formulaire Web ou de manière automatisée via un API. Se rapprocher du guichet DOI (doi@versailles.inra.fr) pour toute information ou support.

CAS DES RESSOURCES GENETIQUES (ACCESSIONS)

- Le Secrétariat du Traité international sur les ressources phytogénétiques pour l'alimentation et l'agriculture (RPGAA) met en oeuvre un service d'attribution de DOI (<http://www.planttreaty.org/doi>). Les métadonnées proposées par ce système pour décrire les ressources candidates sont mieux appropriées que celles de DataCite qui sont généralistes. Nous recommandons l'utilisation de ce service pour l'identification des accessions.

CAS DES REPRESENTATIONS WEB DE RESSOURCES PHYSIQUES (EXEMPLE : LES INFRASTRUCTURES DE RECHERCHE)

- Nous recommandons l'utilisation de DOI. Lors de la demande de DOI, nous recommandons de fournir uniquement les métadonnées minimales ou pertinentes pour la description de ces ressources. En complément, nous recommandons que chaque ressource identifiée dispose d'une page descriptive avec l'ensemble des métadonnées nécessaires pour les décrire de manière satisfaisante. Les DOI pointeront vers ces pages descriptives.