



Open**data**Corsica

données publiques / dati publichi



Guide Technique





Guide Technique

Le présent guide technique a été créé dans le cadre du work package 5 du projet HOMER dont la Collectivité Territoriale de Corse est partenaire depuis le mois de juillet 2012 : <http://homerproject.eu/>

Version septembre 2014

Rédigé par : COLLECTIF

Licence : creative commons by nc



Les recommandations du présent guide technique ont été rédigées sur la base du référentiel des bonnes pratiques OPENDATA publié par le projet Opquast : (<http://opquast.com/fr/>)



GUIDE TECHNIQUE

Table des matières

Introduction	7
1. Contexte	8
2. Présentation des recommandations émises dans le document	9
2.1 Niveaux de priorité de la recommandation	10
2.2 Décomposition en composantes du modèle d'architecture technique OPENDATA	11
3. Liste des recommandations, des commentaires et des retours des partenaires du projet HOMER	12
3.1 Systèmes de gestion de contenu (cms)	12
3.1.1 Gestion du contenu (cont)	12
3.1.2 Animation (anim)	14
3.1.3 Applications (appl)	17
3.2 Indexation et recherche (rech)	18
3.2.1 API (api)	18
3.2.2 Recherche fédérée (fed)	20
3.3 Metadonnées (meta)	21
3.3.1 Catalogage (cat)	21
3.3.2 Identification (id)	23
3.3.3 Nommage (nom)	24
3.3.4 Licence (lic)	25
3.3.5 Transparence (tra)	26
3.4 Jeux de données (don)	27
3.4.1 Format (form)	27
3.4.2 Historique (hist)	29
3.4.3 Utilisabilité (util)	30
3.4.4 Vie privée (vp)	31
3.4.5 Publication (pub)	31
3.5 Vocabulaires (voc)	35
3.6 Le web des données - linked data (ld)	37
Conclusion	38
Tables des recommandations	39

GUIDE TECHNIQUE

Introduction

En septembre 2011, la Collectivité Territoriale de Corse s'associait au projet HOMER dans le cadre de l'appel à candidature du programme européen MED, objectif 3.2 « améliorer le soutien à l'utilisation des technologies de l'information pour une meilleure accessibilité et coopération territoriale ».

Par délibération N°12/107 du 5 juillet 2012, l'Assemblée de Corse autorisait officiellement la participation de la Collectivité Territoriale de Corse au projet HOMER.

Le projet HOMER regroupe 19 partenaires, acteurs du bassin méditerranéen impliqués dans des initiatives OPENDATA (ouverture des données publiques). La région du Piémont (Italie) étant coordinatrice du projet.

L'objectif est de partager des jeux de données publiques autour de cinq problématiques stratégiques en Méditerranée : l'agriculture, le tourisme, l'énergie, l'environnement et la culture.

Dans le cadre du projet HOMER, deux documents techniques ont été produits afin d'accompagner les travaux des partenaires, mais aussi de dresser un état des lieux et de formuler des recommandations sur les normes, standards, techniques et outils rattachés à un projet d'ouverture des données publiques (OPENDATA).

L'état des lieux a été dressé dans un premier document intitulé « Etat de l'Art ». Il est complété par ce « Guide Technique » qui intègre une série de recommandations techniques et fonctionnelles rattachées à une plateforme OPENDATA.

Ce document a pour objectif :

- D'améliorer la compréhension des normes et des solutions techniques évoquées dans l'Etat de l'Art.
- De proposer aux partenaires du projet HOMER une liste de recommandations techniques.
- De mesurer et d'évaluer l'importance de l'implémentation de chaque recommandation au sein du projet HOMER mais plus généralement d'un projet OPENDATA.

Il se structure en trois parties :

- **Contexte**
- **Présentation des recommandations émises dans le document**
- **Liste des recommandations, des commentaires et des retours des partenaires du projet HOMER**

1. Contexte

Au regard de l'évolution des normes et standards techniques liés à l'OPENDATA ce document n'a pas l'ambition d'imposer un point de vue figé sur les spécifications techniques qu'il mentionne. Il veut contribuer à construire un référentiel basé sur une série de recommandations techniques pour partager une vision d'une plateforme OPENDATA et de son environnement technique susceptible d'être commentée et discutée.

L'objectif initial de ce document était de garantir la cohérence technique et l'interopérabilité des plateformes des partenaires du projet HOMER. Il a évolué pour devenir un document de travail susceptible d'aider la communauté des chefs de projet engagés dans une stratégie d'ouverture des données publiques.

Afin de pouvoir répondre à ces ambitions, ce document s'articule autour de deux grandes catégories de recommandations :

1. Les recommandations qui s'imposent et qui doivent être considérées comme une exigence du projet

Ce type de recommandations cible les normes et standards désormais incontournables dans l'environnement technique de l'OPENDATA, qu'il s'agisse de normes de fait ou de juré. Ce type de recommandations s'impose pour tout projet OPENDATA.

2. Les recommandations qui relèvent de bonnes pratiques, mais qui ne constituent pas une exigence absolue

Ces recommandations sont rattachées à de bonnes pratiques identifiées. Pour autant, elles ne doivent pas être considérées comme des exigences absolues. Il s'agit parfois de normes ou de standards au fort potentiel qui s'imposeront sans doute à moyen terme. Ainsi ces recommandations appellent-elles une vigilance, mais ne réclament pas nécessairement une implémentation.

GUIDE TECHNIQUE

2. Présentation des recommandations émises dans le document

Les recommandations identifiées dans ce Guide Technique s'organisent en différents tableaux décomposés selon les champs ci-après :

- Code : il désigne la recommandation et l'index dans sa rubrique. Sachant que les recommandations sont regroupées selon la nature de la composante technique à laquelle elles appartiennent, déclinées elles-mêmes en sous rubriques.

Par exemple : CMS/CONT01 : désigne la première recommandation de la sous-rubrique « gestion de contenu » de la composante technique «Système de gestion de contenu»

- Titre : il s'agit du libellé de la recommandation.
- Niveau de priorité de la recommandation, selon la décomposition « MUST », « SHOULD », « MAY».
- Commentaires : commentaires des partenaires HOMER
- Niveau de pertinence de la recommandation exprimé par les partenaires HOMER : exprimé sur une échelle de 1 à 5, le niveau d'intérêt de la recommandation pour les partenaires du projet HOMER, à savoir :

- 1 - « peu d'intérêt »
- 2 - « assez intéressant »
- 3 - « intéressant »
- 4 - « très intéressant »
- 5 - « incontournable et essentiel »

- A l'usage des porteurs de projets - fonctionnalité présente dans le projet OPENDATA : cette colonne est proposée comme un outil d'aide à la décision pour les porteurs de projets OPENDATA. Elle permet au porteur de projet de pointer les fonctionnalités existantes dans son portail pour vérifier sa complétude ou pointer les fonctionnalités exigées dans le cahier des charges d'un futur portail.

2.1 Niveaux de priorité de la recommandation

Afin d'établir une décomposition du niveau de priorité de chaque recommandation, une grille de lecture a été constituée sur la base de trois mots clés : « MUST » (DOIT), « SHOULD » (DEVRAIT) et « MAY » (POURRAIT). Leur signification est décrite ci-dessous :

MUST : Ce mot qualifie une recommandation du présent document comme une exigence technique forte (voir absolue) qui doit être intégrée impérativement dans tout projet de plateforme OPENDATA.

SHOULD : Ce mot indique qu'une recommandation est fortement souhaitable sans pour autant l'exiger dans un projet de plateforme OPENDATA. Cette recommandation doit toutefois faire l'objet d'une réflexion particulière avant de décider son implémentation ou non dans les spécifications d'une plateforme OPENDATA. Le « SHOULD » est aussi utilisé pour les normes techniques émergentes à fort potentiel qui s'imposeront certainement à brève échéance.

MAY : Ce mot indique que la norme présente un intérêt mais ne se révèle pas comme essentielle. Elle reste optionnelle, son implémentation est facultative. « MAY » désigne aussi des normes et standards non matures à ce stade de l'étude ou en cours de développement.

Ces trois mots clés seront utilisés comme support dans les tableaux qui composent le présent document.

Ils ont été évalués à partir d'avis formulés par des experts ou par les partenaires HOMER. Dans tous les cas, les niveaux d'exigence spécifiés dans ce Guide Technique pourront être adaptés en fonction de la nature des projets.

GUIDE TECHNIQUE

2.2 Décomposition en composantes du modèle d'architecture technique OPENDATA

Le présent document se structure aussi autour du modèle de l'architecture technique OPENDATA décrite dans l'« Etat de l'Art » et illustré par la Figure 1. Cette représentation d'architecture permet de décomposer chaque recommandation selon les six composantes d'architecture suivantes :

1. CMS (Système de gestion de Contenu),
2. Indexation et Recherche,
3. Métadonnées,
4. Jeux de données,
5. Vocabulaires,
6. Le Web de données (Linked data)

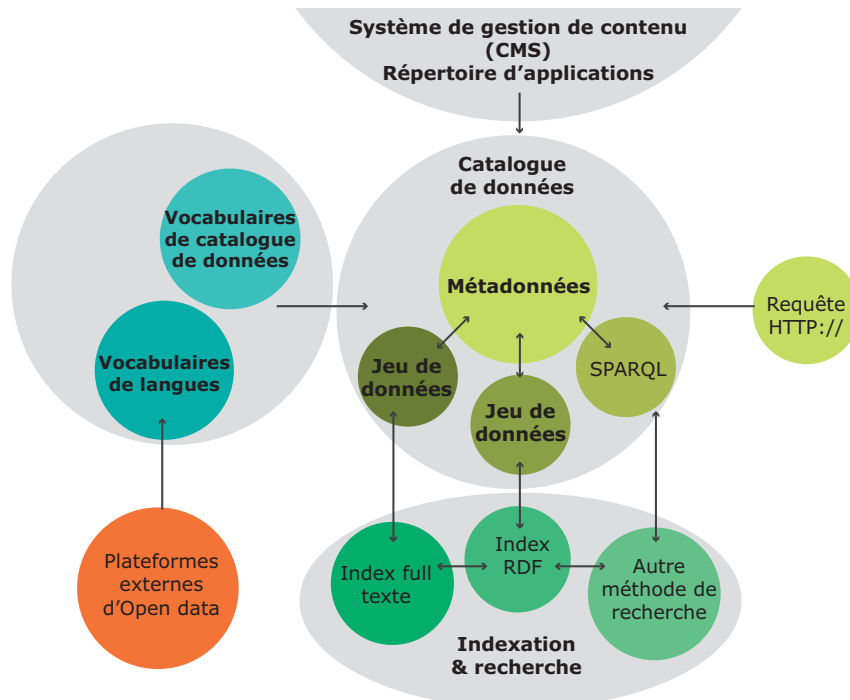


Figure 1. Modèle de l'architecture technique (OPENDATA)

OPENDATA

3. Liste des recommandations, des commentaires et des retours des partenaires du projet HOMER

3.1 Systèmes de gestion de contenu (cms)

3.1.1 Gestion du contenu (cont)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
CMS/CONT01	Publier des données éditoriales (news, approche OPENDATA)	X				4	
CMS/CONT02	Gérer une liste source (entités utilisant le portail pour publier leurs données: administrations ou partenaires)		X			5	
CMS/CONT03	Gérer une liste des différentes catégories de données (cette liste peut être commune pour toutes les sources)		X			5	
CMS/CONT04	Gérer la liste de formats des données (xls, doc, txt, xml...)	X				5	
CMS/CONT05	Créer un mot de passe/identifiant pour un partenaire		X			5	
CMS/CONT06	Gérer les profils utilisateurs pour la partie administration du CMS		X			4	

GUIDE TECHNIQUE

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
CMS/CONT07	Avoir les statistiques du trafic du site		X			4	
CMS/CONT08	Prévoir un espace pour les infos juridiques (contact, webmaster...)			X		4	
CMS/CONT09	Possibilité d'imprimer: licences, cartes de métadonnées et documents administratifs	X				5	
CMS/CONT10	Le producteur de données expose sa politique en matière de libération de données		X		Permet de décrire le contexte de production des données et les politiques publiques associées à leur gestion et à leur diffusion	3	

OPENDATA

3.1.2 Animation (anim)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
CMS/ANIM01	Le concept OPENDATA est expliqué	X			La page du portail intitulée «Approche et procédure» montre aux utilisateurs le concept et la procédure de l'ouverture des données. La section FAQ peut être utilisée pour énoncer les principes essentiels de l'OPENDATA	4	
CMS/ANIM02	Le site fournit aux utilisateurs la possibilité de connaître l'actualité du dépôt de données	X			Le portail peut fournir un flux Atom/RSS qui permet d'avoir la liste des jeux de données disponibles sur le site Internet. Le flux peut être divisé en 2 catégories : nouveaux jeux de données et jeux de données mis à jour	5	
CMS/ANIM03	Le site fournit un canal en ligne de remontée d'informations	X			Un formulaire de contact est disponible pour chaque jeu de données, ce qui permet aux utilisateurs de contacter directement le producteur des données	4	

GUIDE TECHNIQUE

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
CMS/ANIM04	Des exemples de réutilisations (fictifs ou réels) sont fournis	X			Fournir des tutoriels illustrés décrivant l'utilisation du code (sous forme de librairie côté client). Présenter les cas de réutilisation sous forme de copies d'écran des applications web ou mobiles. Dans le formulaire de métadonnées, donner les liens vers les applications utilisant les jeux de données en question. Donner des exemples de requêtes pour présenter les données issues de l'entrepôt de données	4	
CMS/ANIM05	Un canal d'information permet de connaître l'actualité de chaque jeu de données	X			Pour chaque jeu de données, la date de mise à jour doit être exacte et facilement identifiable. Plusieurs normes de métadonnées ont un élément spécifique qui permet de qualifier cette information	4	
CMS/ANIM06	Les jeux de données sont accompagnés de liens ou de ressources d'animation et de supports			X	La création d'un forum dédié à promouvoir la diffusion des jeux de données est cruciale. Cela permet de connecter les ré-utilisateurs entre eux	3	

OPENDATA

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
CMS/ANIM07	Il est possible d'accéder à un espace d'échange sur les jeux de données			X	La création d'un forum de discussions avec différentes rubriques pour l'échange d'opinions, d'exemples, de solutions techniques ou pour signaler les erreurs dans les jeux de données	3	
CMS/ANIM08	Il est possible de demander ou de proposer la mise en ligne de nouveaux jeux de données			X	Création d'un formulaire de contact permettant de demander de nouveaux jeux de données. Ce formulaire peut comprendre un champ indiquant la raison de la requête. Création d'une rubrique pour lister les requêtes reçues, avec la possibilité de voter pour les différentes propositions	4	
CMS/ANIM09	Le site fournit la possibilité de noter et/ou d'évaluer les jeux de données			X	L'affichage de chaque enregistrement décrivant les jeux de données offre un module d'évaluation accessible pour tout le monde. Ce module permet d'évaluer chaque jeu de données à l'échelle de 1 à 5. Un formulaire spécifique dédié aux utilisateurs qui veulent laisser leurs commentaires est possible	3	

GUIDE TECHNIQUE

3.1.3 Applications (appl)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
CMS/APPL01	Le site explique la façon de faire référencer une application			X	Un espace dédié aux développeurs permet de référencer leurs applications	3	
CMS/APPL02	Le site recense les applications éventuelles déjà développées à partir des jeux de données proposés			X	L'espace applications fournit la liste des applications existantes. Il est possible de faire une recherche et de filtrer les résultats de recherche	3	
CMS/APPL03	Il est possible d'être informé sur l'actualité des applications			X	Exemple : RSS, blog, twitter, newsletter	3	

OPENDATA

3.2 Indexation et recherche (rech)

3.2.1 API (api)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
RECH/API01	Les mécanismes d'accès et d'interaction avec les données sont documentés		X		Expliciter les conditions d'utilisation des jeux de données. Si un mécanisme de restriction d'accès est mis en place celui-ci est documenté	4	
RECH/API02	Si un jeu de données fait l'objet d'une restriction d'accès, la méthode de récupération des données est documentée		X		Des conditions spécifiques d'accès peuvent être définies pour l'utilisation d'un Web service ou de données en temps réel. Suivant le cas de figure, les raisons de la restriction mise en place et les moyens d'accès aux données sont documentés	4	
RECH/API03	Une commande spécifique de l'API permet de connaître les conditions de licence			X	Mettre en place un accès par HTTP des métadonnées (et des jeux de données si possible) en identifiant précisément le champ contenant les informations relatives aux conditions d'utilisation	4	
RECH/API04	Les versions dynamiques et statiques des données sont synchronisées			X	Le logiciel CKAN dispose d'une extension permettant d'interroger le site de publication initial du jeu de données dans le cas où une référence à cette source de données est disponible	3	

GUIDE TECHNIQUE

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
RECH/API05	Lorsque des données existent dans plusieurs formats, le serveur offre un mécanisme de négociation de contenu			X	Permettre l'accès immédiat aux données dans le format le plus approprié selon les besoins de l'utilisateur	3	
RECH/API06	Si des Web services d'accès aux données sont proposés, la version courante de chaque Web service est précisée			X	Dans la documentation des Web services mis à disposition il est décrit la version courante de chaque Web service et le moyen de l'utiliser	4	
RECH/API07	Aucun dispositif n'est mis en place pour empêcher la récupération automatique ou programmatique des données			X	Donner librement accès aux jeux de données sous une forme facilement téléchargeable	4	
RECH/API08	Chaque jeu de donnée est disponible via au moins un Web service			X	La mise à disposition d'un Web service sur les métadonnées permet de connaître automatiquement l'URL du jeu de données ou le point d'accès d'interrogation des données	3	
RECH/API09	Le catalogue des métadonnées est récupérable selon un protocole normalisé			X	Le catalogue de métadonnées mis à disposition sous la forme d'un fichier téléchargeable permet sa récupération automatique par HTTP GET	4	

OPENDATA

3.2.2 Recherche fédérée (fed)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinance	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
RECH/FED01	Disposer d'une structure de métadonnées compatible avec les recommandations techniques du projet de fédération de données	X			C'est le cas dans le projet HOMER avec notamment la définition des champs communs des métadonnées	5 Imposé dans le projet HOMER	
RECH/FED02	Disposer d'une ou de plusieurs méthodes d'accès à l'index fédéré	X			Pour le projet HOMER 3 méthodes d'accès existent : Interface de configuration d'API type CKAN (2 url et 2 fichiers json) OU Services de catalogues pour le protocole Web (2 url et 2 fichiers csw) OU Services Web équivalents à ceux mis à disposition sur le site www.dat1.piemonte.it (2 url et 2 fichiers xml) ». Pour HOMER : CKAN API PIEMONTE	5 Imposé dans le projet HOMER	
RECH/FED03	Faire référence aux labels EUROVOC pour intégrer les tags et les mot-clés dans les jeux de données	X			EUROVOC	5 Imposé dans le projet HOMER	
RECH/FED04	Assurer l'accès à l'index fédéré sur les différents portails	X			Pour le projet HOMER : Chaque portail offre le lien ou l'accès (Widgets) à l'index fédéré et au résultat de la recherche	5 Imposé dans le projet HOMER	
RECH/FED05	L'index fédéré est accessible en consultation à minima via une page Web	X			Pour le projet HOMER l'index fédéré est accessible sur la page spécifique Lien index fédéré d'HOMER : http://opendata-federation.csi.it/#pivot	5 Imposé dans le projet HOMER	

GUIDE TECHNIQUE

3.3 Métadonnées (meta)

3.3.1 Catalogage (cat)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
META/CAT01	L'identité et la qualité du responsable de chaque jeu de données sont précisées	X			Sur le site, spécifier le nom de la personne ou de l'entité juridique responsable du jeu de données	5	
META/CAT02	Chaque jeu de données est accompagné d'une fiche descriptive	X			Simplifier l'accès à la documentation de chaque jeu de données et aux informations clés concernant son exploitation	5	
META/CAT03	Le jeu de données comporte au moins un titre et une description	X			Permettre la compréhension facile du contenu du jeu de données	5	
META/CAT04	Chaque jeu de données est accompagné d'une date de création	X			Expliciter la date de création du jeu de données, la date de publication ou de republication (après mises à jour)	5	
META/CAT05	Chaque jeu de données est accompagné d'une date de mise à jour	X			Indiquer pour chaque jeu de données, la date de mise à jour la plus récente	5	
META/CAT06	Les jeux de données sont catégorisés		X		Permettre la recherche multidimensionnelle par catégorie	5	

OPENDATA

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
META/CAT07	Les métadonnées associées avec chaque jeu de données sont proposées en format standard		X		Permettre la sauvegarde et l'import automatique des métadonnées venant d'autres catalogues de données	4	
META/CAT08	Chaque jeu de données est accompagné d'une mention à la langue		X		Différencier la langue du propriétaire de la donnée (Collectivité) et celle du jeu de données publié	4	
META/CAT09	Les métadonnées décrivant les jeux de données sont structurées de manière normative			X	Permettre l'exploitation des métadonnées conformément au protocole de requête	5	
META/CAT10	Si un jeu de données correspond à une zone géographique délimitée, il est possible de connaître sa localisation			X	La couverture géographique du jeu de données et ses coordonnées géographiques peuvent être explicitées	3	
META/CAT11	Les thématiques sont reliées à des listes d'autorités et des taxonomies publiques			X	En fonction du vocabulaire choisi il faut faire le lien entre le vocabulaire et les sujets génériques utilisés pour le catalogage des données. Cela permet d'associer facilement les connaissances liées aux mêmes sujets Référence technique : SKOS	3	

GUIDE TECHNIQUE

3.3.2 Identification (id)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
META/ID02	Chaque jeu de données est accompagné d'un identifiant unique	X			Pour les identifiants explicites, les règles d'URI définies dans la recommandation du W3C sont à utiliser. Pour les identifiants opaques on peut citer les systèmes DOI et ARK	5	
META/ID03	Un checksum et/ ou une signature sont proposés pour vérifier la validité de chaque fichier		X		Différentes bibliothèques permettent de calculer à la volée une somme de contrôles permettant d'identifier sans ambiguïté un objet numérique. Plusieurs fonctions de hachage sont disponibles mais MD5 est la plus courante	2	
META/ID04	Le ou les jeux de données sont accompagnés d'au moins un moyen de contact avec le producteur (ou le mainteneur)		X		Conformément aux recommandations de la directive INSPIRE qui vise à fournir un moyen d'identification des informations environnementales éventuellement nécessaires à la gestion d'une situation de crise, le schéma DCAT intègre différents éléments du schéma Dublin Core permettant de spécifier l'identité et le contact du mainteneur du jeu de données et/ ou de son créateur (respectivement dcterms:contributor et dcterms:creator)	5	

OPENDATA

3.3.3 Nommage (nom)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
META/NOM01	Les noms des fichiers de données contiennent exclusivement des caractères alphanumériques ou considérés comme sûrs		X			3	
META/NOM02	Si le nommage des jeux de données fait l'objet d'une charte, celle-ci est mise à disposition		X		Normaliser la production des URI des jeux de données permet de fournir un élément supplémentaire de confiance dans l'identification et la pérennité des ressources mises à disposition	3	

GUIDE TECHNIQUE

3.3.4 Licence (lic)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
META/LIC01	Les jeux de données sont accompagnés d'une licence d'utilisation	X			Le schéma DCAT ré-utilise l'élément dcterms qui spécifie la licence qui couvre les conditions d'utilisation du jeu de données. Cette information de licence s'applique à la fois aux métadonnées et aux jeux de données	5	
META/LIC02	La licence mentionne les conditions de paternité, de réutilisation, de redistribution et de commercialisation	X			Les licences OPENDATA décrivent les obligations qui engagent les ré-utilisateurs des jeux de données. Les partenaires du projet HOMER ayant eu principalement recours aux licences dites ouvertes ou ODBL Le portail OPENDATA de la Corse utilise la licence ouverte ETALAB : http://www.etalab.gouv.fr/licence-ouverte-open-licence	5	
META/LIC03	Les droits d'utilisation sont fournis pour une durée illimitée		X			3	
META/LIC04	Les jeux de données sont accompagnés d'un résumé et d'un lien vers la version complète de la licence			X		4	

OPENDATA

3.3.5 Transparence (tra)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
META/TRA01	Si un jeu de donnée comporte une date de péremption, celle-ci est mentionnée	X			Dans le schéma DCAT, les éléments Dublin Core dcterms:modified et dcterms:accrualPeriodicity permettent de savoir de manière claire si le fichier mis à disposition est bien à jour	4	
META/TRA02	Si une limite est fixée pour le nombre d'accès ou la quantité de données accessible par les Web services, cette limitation est précisée	X			Fournir une documentation indiquant les conditions d'utilisation de l'API	4	
META/TRA03	Les éventuelles restrictions d'accès à la donnée sont signalées	X			Permettre de s'assurer que les ré-utilisateurs ont bien pris connaissance des conditions d'accès au jeu de données	4	
META/TRA04	Si un jeu de données comporte des erreurs, des incertitudes ou est incomplet, cela doit être signalé		X			4	
META/TRA05	Chaque jeu de donnée est accompagné d'une fréquence de mise à jour		X		L'utilisation de l'élément dcterms:frequency dans le schéma DCAT permet de spécifier la date de dernière mise à jour du fichier	4	
META/TRA06	L'émetteur du jeu de données fournit des informations sur leur origine		X			4	
META/TRA07	Il est possible d'obtenir des informations concernant le niveau de confiance à accorder aux données			X	Permettre aux ré-utilisateurs d'apprécier la garantie de service associée à la ré-utilisation des données	2	

GUIDE TECHNIQUE

3.4 Jeux de données (don)

3.4.1 Format (form)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
DON/FORM01	Chaque jeu de données est accompagné d'une mention du jeu de caractères	X				3	
DON/FORM02	Le format des fichiers proposés en téléchargement est indiqué	X			Pour chaque ressource disponible, indiquer son type MIME. Plusieurs annuaires de formats permettent de relier les caractéristiques de chaque format à la documentation de référence le concernant	5	
DON/FORM03	Les dates sont présentées dans des formats documentés	X			Permettre l'utilisation et la conversion des données liées à la date compte tenu des besoins ou des exigences spécifiques	4	
DON/FORM04	Les dates sont proposées au format standard		X		Faciliter l'exploitation automatique des données liées à la date et leur conversion compte tenu des besoins ou des exigences spécifiques	4	

OPENDATA

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
DON/FORM05	Les données sont fournies dans au moins un format ouvert		X		Limiter les charges liées à l'exploitation de données	5	
DON/FORM06	Chaque jeu de données est accompagné de la norme d'encodage des caractères utilisés		X		Permettre les conversions nécessaires pour utiliser un jeu de données encodé différemment	3	
DON/FORM07	Si une donnée fait l'objet d'un format standard, celui-ci est mentionné et respecté			X	Permettre la réutilisation par les outils standards et la reconstitution d'outils pour accéder au jeu de données	4	
DON/FORM08	Il est possible de connaître la version ou la syntaxe (doctype) des formats de fichiers utilisés			X	Pour le producteur il peut être utile de présenter sur une page dédiée les formats et les versions des fichiers utilisés sur le site	2	
DON/FORM09	Les champs intégrés aux données tabulaires utilisent des noms significatifs		X			3	
DON/FORM10	Les données sont en format lisible par les machines (applications développées par les utilisateurs)	X			Il est fait directement référence ici au « modèle des 5 étoiles » qui offre une classification des critères techniques de mise à disposition des données. http://5stardata.info/	5	
DON/FORM11	Les données sont publiées dans un format qui encourage leur utilisation directe à travers des API		X		Les données peuvent être postées sur un serveur Web avec une url permanente, en format standard d'échange de données, comme JSON ou XML	4	

GUIDE TECHNIQUE

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
DON/FORM12	Les données sont fournies dans la plus large mesure possible, en format ouvert ou couramment utilisé, indépendant de la plateforme		X		L'utilisateur de données ne doit pas être obligé d'avoir accès à un logiciel propriétaire ou une bibliothèque de programmation afin de lire ou consommer les données	4	
DON/FORM13	Les données sont présentées sous leur forme brute et doivent inclure en plus, une version sans formatage de présentation		X		Par exemple : Si les données contiennent des marquages de présentation (ex : HTML comme <table>, <p>,), une version alternative sans marquage doit être incluse également Voir aussi http://wiki.data.gouv.fr/wiki/Transformer_des_donn%C3%A9es	4	
DON/FORM14	Les données temporelles doivent inclure le fuseau horaire		X		Si possible, le fuseau horaire doit être inclus avec les données. Il est préférable d'utiliser UTC pour toutes les indications de temps. Cela permet aux développeurs d'afficher le temps correct compte tenu de la localisation de l'utilisateur	2	

OPENDATA

3.4.2 Historique (hist)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
DON/HIST01	Chaque jeu de données est livré avec un « changelog »		X				
DON/HIST02	Il est possible d'accéder aux différentes versions des jeux de données			X			

GUIDE TECHNIQUE

3.4.3 Utilisabilité (util)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
DON/UTIL01	La taille des fichiers proposés en téléchargement est indiquée	X			Dans le schéma DCAT l'élément dcate:size permet de spécifier la taille des ressources associées aux jeux de données	4	
DON/UTIL02	Il est possible d'effectuer une recherche dans les métadonnées	X			Permettre de connaître le contenu du catalogue de données par l'interrogation des métadonnées	5	
DON/UTIL03	Des extraits des différents jeux de données sont disponibles en téléchargement		X		Faciliter l'accès aux jeux de données en fonction des besoins de réutilisation	5	
DON/UTIL04	Si le site propose plusieurs licences, le moteur de recherche permet de filtrer par licence			X	Faciliter l'accès aux jeux de données en fonction des besoins de réutilisation	4	
DON/UTIL05	Le catalogue des métadonnées est disponible sous la forme d'un jeu de données			X		3	
DON/UTIL06	Un extrait des données peut être pré-visualisé			X	Permet d'éviter le transfert systématique du jeu de données par les ré-utilisateurs	4	

OPENDATA

3.4.4 Vie privée (vp)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
DON/VP01	Si les données contiennent des informations nominatives, leur diffusion doit est conforme à la législation dont dépend le producteur	X			Afin de garantir le respect de la vie privée le projet HOMER conseille l'utilisation de données anonymisées	4	
DON/VP02	Le site fournit un moyen de demander le retrait ou l'anonymisation d'une donnée		X		Dans le cas où des informations erronées ou nominatives auraient été publiées, un moyen doit être prévu de corriger les données et de republier le jeu de données modifié	4	

3.4.5 Publication (pub)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
DON/PUB01	Le projet OPENDATA doit définir précisément le cycle de transformation des données de la source au site OPENDATA ainsi que les acteurs engagés		X		Il s'agit ici d'une recommandation de Gestion de projet		

GUIDE TECHNIQUE

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
DON/PUB02	Chaque jeu de données a un propriétaire identifié qui est responsable de l'exactitude des données sources		X		Le propriétaire de la donnée doit être identifié comme le propriétaire du jeu de données OPENDATA	4	
DON/PUB03	Chaque jeu de données a une table de correspondance de champs avec les données présentes dans le système d'information de la collectivité		X		Il est important de pouvoir identifier d'où exactement provient un champ spécifique d'un jeu de données, pour permettre le dépannage et garantir l'intégrité du jeu de données	3	
DON/PUB04	Le propriétaire de la donnée au sein de la collectivité et le développeur de la plateforme OPENDATA travailleront ensemble pour identifier la méthode de publication la plus appropriée (automatique ou manuelle)	X			Il s'agit de définir les bonnes pratiques de mise à jour d'un jeu de données en choisissant le mode d'actualisation le plus adapté au jeu de données publié. Exemple : des données temps réel devront être mises à jour automatiquement, alors que des données historiques annuelles seront plutôt mises à jour manuellement	4	
DON/PUB05	Chaque jeu de données doit avoir une documentation relative au processus utilisé pour convertir le format original en format ouvert		X		Quand on fait la conversion des données du format propriétaire en format ouvert, la documentation du logiciel et les numéros de versions doivent être inclus. Cela permet de reconverter les données à l'avenir, quand le logiciel sera mis à jour	3	

OPENDATA

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
DON/PUB06	Avant de publier le jeu de données, le propriétaire doit examiner le tableau de correspondance des champs et les données qui seront publiées, pour s'assurer qu'aucune donnée sensible ne soit incluse dans le jeu de données ouvert		X		Une analyse des champs des données dans le système de la collectivité doit être faite afin de garantir qu'aucune donnée sensible n'est présente dans le jeu de données OPENDATA	4	
DON/PUB07	Si le jeu de données requis contient des données sensibles, le personnel doit analyser si ce jeu de données peut être agrégé, généralisé ou ré-identifié de manière à permettre la publication du maximum de données		X		Si un jeu de données contient une identification qui pourrait rendre les données sensibles (nom, adresse ou numéro du compte), les champs sensibles doivent être retirés. Les données peuvent être agrégées par le nom de la rue, le code postal ou le quartier, pour les rendre moins sensibles	3	
DON/PUB08	Les données sont mises à jour fréquemment pour assurer le maximum d'utilité		X		Les données dont la date de mise à jour est ancienne diminuent la valeur du jeu de données OPENDATA libéré en obligeant les utilisateurs à contacter la collectivité afin de s'assurer que les données sont toujours d'actualité	4	

GUIDE TECHNIQUE

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
DON/PUB09	Les données sont publiées de manière à pouvoir identifier facilement les nouveaux enregistrements et les nouvelles mises à jour		X		Les données doivent être horodatées pour permettre aux programmes informatiques de déterminer les enregistrements qui ont été ajoutés et/ou modifiés. Deux horodatages doivent être inclus, l'un pour la date d'enregistrement et l'autre pour la date de dernière modification	3	
DON/PUB10	Les données sont publiées automatiquement à chaque fois que c'est possible, pour réduire la charge du personnel des collectivités et augmenter l'utilité du jeu de données		X		Au lieu d'assigner une personne dédiée à la mise à jour journalière ou hebdomadaire d'un jeu de données, un programme d'interface devrait être écrit pour envoyer automatiquement les données dans un catalogue OPENDATA en suivant le calendrier des mises à jour	4	
DON/PUB11	Les données sont publiées de manière à permettre à tous les utilisateurs (techniciens ou non) de tirer profit de la valeur du jeu de données		X		Les données doivent être publiées dans différents formats (exemple : XML, JSON) mais doivent également utiliser des interfaces graphiques (exemple : cartes, graphiques, tableurs) pour rendre les données le plus digeste possible	4	

OPENDATA

3.5 Vocabulaires (voc)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
VOC01	Les vocabulaires sont documentés	X			Le vocabulaire doit être documenté. Cela comprend les labels, les commentaires et les mots clés de la langue utilisée. Les pages lisibles par l'utilisateur doivent être fournies par l'éditeur et décrire les classes et les propriétés, avec des exemples d'utilisation de préférence	4	
VOC02	Les vocabulaires sont auto-descriptifs		X		Chaque propriété ou terme du vocabulaire doit avoir un label, une définition et un commentaire définis. Les données auto-descriptives suggèrent l'information sur les encodages utilisés pour chaque représentation	3	
VOC03	Les vocabulaires sont décrits dans différentes langues		X		Les fonctionnalités multi-langues doivent être supportées par le vocabulaire, c'est-à-dire tous les éléments du vocabulaire doivent avoir des labels, des définitions et des commentaires disponibles dans la langue de la collectivité	3	

GUIDE TECHNIQUE

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
VOC04	Les vocabulaires sont utilisés par d'autres jeux de données		X		Si le vocabulaire est utilisé par d'autres jeux de données « LINKED DATA » cela peut être utile. C'est en réutilisant les vocabulaires qu'on voit tous les avantages des fichiers « LINKED OPENDATA »	3	
VOC05	Les vocabulaires sont accessibles pour une période longue		X		Le vocabulaire sélectionné doit fournir une garantie du maintien pour une période spécifiée, idéalement pour une durée indéterminée	2	
VOC06	Les vocabulaires sont publiés par un groupe ou une organisation de confiance	X			Tout le monde étant capable de créer un vocabulaire, il est cependant préférable de vérifier qui est en charge de la publication et de la maintenance du vocabulaire utilisé	3	
VOC07	Les vocabulaires ont des URLs permanentes		X		L'accès permanent au serveur hébergeant le vocabulaire est nécessaire, pour permettre leur réutilisation	3	
VOC08	Les vocabulaires doivent gérer l'évolution de leurs différentes versions		X		Idéalement, l'éditeur traitera la compatibilité des versions avec le temps. Les modifications majeures des vocabulaires doivent être intégrées dans la documentation	3	

OPENDATA

3.6 Le web des données - linked data (ld)

Code	Titre	Must	Should	May	Commentaires	Niveau de pertinence	A l'usage des porteurs de projets : fonctionnalité présente dans le projet OPENDATA
LD01	Chaque jeu de données fait référence à des vocabulaires identifiés et documentés		X		Permettre par exemple d'intégrer des thésaurus existants sous la forme de fichiers RDF-SKOS et de ce fait, augmenter les possibilités de rapprochement de jeux de données provenant de portails différents	3	
LD02	Les données respectent la syntaxe définie dans le ou les vocabulaires choisis			X	Permettre la validation automatique des données	3	
LD03	Il est possible d'interroger les données et métadonnées conformément aux standards du Web de données (Linked OPENDATA)			X	Permettre le croisement des données avec les données du Web de données	4	
LD04	Les données liées (Linked OPENDATA) sont signalées			X	La localisation de l'entrepôt des données liées permet de connaître les liens qui peuvent être faits à partir de l'importation de ces données au sein d'une application	3	

GUIDE TECHNIQUE

Conclusion

Ce document entend donner quelques clés de lecture opérationnelles des normes et techniques évoquées dans l'Etat de l'Art. Il apporte une vision technique partagée par les partenaires du projet HOMER d'une plateforme OPENDATA bâtie autour d'un référentiel qui peut s'adapter à l'environnement de chacun.

Ce document peut être très utile dans deux cas de figure :

- Comme un document d'aide à l'évolution d'un portail OPENDATA : si vous possédez déjà un portail opendata, il permet de pointer les recommandations que vous avez déjà implémentées et de vérifier ensuite celles qui potentiellement pourraient enrichir votre plateforme.
- Comme un document d'aide à l'évaluation d'un nouveau portail : si vous travaillez à la rédaction d'un cahier des charges de votre portail, les recommandations peuvent figurer en annexe pour faciliter l'analyse des offres. Le tableau pourrait être soumis aux candidats afin qu'ils précisent les recommandations qu'ils intègrent et celles qu'ils ne prennent pas en compte.

Nous invitons le lecteur à nous faire part de ses retours d'expériences et propositions d'amendement sur le site du projet HOMER <http://homerproject.eu> ainsi que sur celui de la Collectivité Territoriale de Corse http://numerique.corse.fr/Open-data-HOMER_r56.html

OPENDATA

Tables des recommandations

CMS	Système de gestion de contenu	CONT	Gestion de contenu	p.12
		ANIM	Animation	p.14
		APPL	Applications	p.17
RECH	Index et recherche	API	API	p.18
		F&D	Recherche fédérée	p.20
META	Métadonnées	CAT	Catalogage	p.21
		ID	Identification	p.23
		NOM	Nommage	p.24
		LIC	Licence	p.25
		TRA	Transparence	p.26
DON	Jeux de données	FORM	Format	p.27
		HIST	Historique	p.29
		UTIL	Utilisabilité	p.30
		VP	Vie privée	p.31
		PUB	Publication	p.31
VOC	Vocabulaires			p.35
LD	Web des données – Linked data			p.37

GUIDE TECHNIQUE

OPENDATA
