

SOUSCRIPTION NATIONALE

CONCEPTION ET REALISATION D'UNE PLATEFORME DE
QUALIFICATION DES DONNEES OUVERTES

PROJET VALIDATA



Portée par l'association OPENDATAFRANCE

Avec le soutien de l'Etat dans le cadre du Programme d'Investissement d'Avenir



Montant total du projet :	320 000 Euro
Cofinancement de l'Etat :	160 000 Euro
Montant de la souscription nationale :	160 000 Euro
Début du projet :	Janvier 2018
Durée :	18 mois

Partenaires techniques : La Fondation internet nouvelle génération, Dataactivi.st, Jaibreak

et 4 territoires pilotes : région Provence-Alpes-Cote d'Azur, Toulouse Métropole, Digne-les-bains, le GIP Territoire numérique Bourgogne-Franche-Comté.

Objet de la souscription :

La souscription consiste à financer le projet VALIDATA à hauteur de 50% par le concours des acteurs publics et des privés qui souhaitent pouvoir bénéficier d'un outil externe de validation des données produites par leurs services ou des tiers. Cette plateforme leur permet d'améliorer la **qualité** des gisements des **données** produites et des **traitements** qui peuvent en être fait (smartcity, observatoire, référentiels locaux ou nationaux, data science).

La souscription est ouverte jusqu'au 31 décembre 2018.

Synthèse du projet :

Le développement de l'ouverture des données publiques, engagé en France depuis près de sept ans, prend aujourd'hui une dimension incontournable dans les stratégies numériques des territoires, de l'Etat et de nombreux acteurs privés. La Loi pour une République Numérique sanctuarise le principe d'ouverture des données par défaut et de très nombreuses collectivités de petite et moyenne taille sont dorénavant concernées.

Le passage à l'échelle de l'ouverture des données présente un défi important : une plus grande quantité de données sera publiée mais la qualité sera-t-elle au rendez-vous ?

Dès à présent, la description, les formats et le contenu des données publiées par les collectivités pionnières de l'opendata pose un problème de consistance. Plus de 20 000 jeux de données sont ouverts par près de 200 collectivités de toute taille et de tout niveau hiérarchique. Si le mouvement est remarquable et mobilisateur, le manque de cohérence entre ces données engendre un gisement national de données difficile à exploiter et à comprendre.

Comment s'assurer que ces données sont dans un état de structuration facilitant au maximum leur réutilisation ? Comment s'assurer qu'elles sont conformes aux recommandations et aux standardisations ?

Cela vaut pour les collectivités qui souhaitent vérifier que leurs données correspondent bien aux recommandations : un outil de test et d'assurance qualité leur est nécessaire.

Cela vaut aussi pour les acteurs nationaux qui s'intéressent aux données locales dans leur ensemble, que ce soit pour constituer des référentiels nationaux (par ex. data.gouv.fr) ou pour des besoins d'analyses (observatoires, datascientists, entreprises...) : ils doivent se « nourrir » de centaines, demain de milliers, de données issues des niveaux locaux et l'acquisition d'information de mauvaise qualité détériorera la base globale des données traitées, ruinant les résultats ou demandant des efforts considérables de retraitement.

Les partenaires du présent projet proposent de mettre en place une plateforme nationale capable de tester la qualité des données publiées au niveau local, de qualifier les sources suivant plusieurs niveaux d'exigence et lorsque les données sont suffisamment fiables, de permettre des exploitations cohérentes au niveau national. Elle complète les dispositifs déjà existants sur data.gouv.fr (moissonnage) ou d'acteurs privés, et se positionne comme une couche dédiée à la qualification et le pré-traitement des données des territoires.

Objectifs du projet (finalités)

- Aider les collectivités à valider la qualité et la cohérence des données qu'elles publient grâce à des comptes-rendus de tests détaillés;
- Certifier les sources (un gisement de données plutôt qu'un portail dans son ensemble) ayant un niveau de qualité satisfaisant;
- Nourrir des référentiels permanents (ex data.gouv.fr) ou temporaires (une analyse nationale sur l'ensemble de jeux de données d'une thématique donnée)
- Fournir des indicateurs qualifiés en vue de l'évaluation de la situation de l'open data dans les territoires (observatoire dédié ou injecteur pour des observatoires-tiers)

Les impacts attendus

- Une montée en qualité des données publiques locales (déclaration, normalisation et contenus) ;
- La normalisation et l'interopérabilité des données favorisant ainsi la réutilisation (enrichissement de bases, développement d'applications ou travaux d'analyses)
- Des référentiels de données qualifiés ;
- Une aide opérationnelle pour les collectivités (réduction des coûts internes) ;
- L'émergence de gisement de données multi sources fiables.

Contenu du projet et livrables

- Plusieurs modules de validation, développés soit en spécifique, soit en s'appuyant sur des fonctions externalisées déjà existantes qui sont adaptées pour le projet (francisation, adaptation aux standards recommandés, etc.). Ces modules aborderont différents niveaux de normalisation :
 - Les ressources (contenants) : métadatas, formats des fichiers, ...
 - Les données : contenu et formats
 - Les syntaxes et structuration des champs
 - La cohérence de certains champs pivots
- Un ordonnanceur des modules de validation des données référencées ;
- Un outil de diagnostic et de certification des différentes sources ;
- Une base intermédiaire volatile permettant la consolidation des données à des fins de traitement et d'analyse.

Le projet est détaillé sur le site <http://opendatafrance.net> (lien raccourci <https://goo.gl/wiz6ku>)

Le cofinancement

La souscription consiste à solliciter les acteurs publics et privés pour le cofinancement du projet qui est soutenu par l'Etat à hauteur de 50%.

Cela se traduit par un appel sous forme de subvention ou de dons pour un projet d'investissement, collaboratif et innovant.

Une grille indicative est proposée pour les apports financiers :

Acteurs Publics

- Collectivités territoriales
 - Plus de 1 000 000 habitants : 10 000 Euro
 - Entre 200 000 et 1 000 000 habitants : 8 000 Euro
 - Entre 20 000 et 200 000 habitants : 4 000 Euro
 - Moins de 20 000 habitants : 1 000 Euro
- Ministères et agences nationales : 15 000 Euro

Acteurs Privés

- Grands Groupes : 15 000 Euro
- Entreprises de taille moyenne 8 000 Euro
- Start-Up, cabinet, Indépendant 2 000 Euro

(Réduction de 50% pour les partenaires associés au projet)

La contractualisation suit les formes habituelles d'intervention des acteurs : subvention, conventions, etc.

L'ensemble des travaux, documents et logiciels, sont publiés en Creative Common et versés sous forme de logiciel libre (github).

> La structure des coûts du projet est la suivante :

- Prestations intellectuelles :	255 000 Euro
○ <i>Dont prestataires extérieurs :</i>	<i>170 000 Euro</i>
○ <i>et ressources internes ODF</i>	<i>85 000 Euro</i>
- Logiciels et base et hébergement :	15 000 Euro
- Déplacements, communication	25 000 Euro
- Pilotage :	25 000 Euro
TOTAL	320 000 Euro

> La structure globale du financement :

- Etat (PIA2 – Open data Territoires) :	160 000 Euro
- Souscription nationale :	160 000 Euro