

"Vers une archive ouverte des documents de marchés publics?"

Contexte

- ouverture des données des marchés publics : [démarche portée par Etalab](#) avec les Collectivités, s'appuyant sur des standards internationaux, elle porte sur les avis de marchés ; parallèlement [le BOAMP](#) a publié sous Licence Ouverte ses avis de marché et d'attribution
- [évolutions de la loi CADA](#) : les documents publics sont réutilisables et republiables librement

Objectif :

- moissonner les documents de consultation des entreprises (DCE) des appels d'offre de marchés
- créer un site d'archives permettant d'y accéder librement
- enrichir progressivement cette archive

Usages attendus

- aujourd'hui des sociétés commerciales font ce travail de moissonnage et classement des appels d'offre et proposent des services pour y accéder, essentiellement pour les fournisseurs, les entreprises qui répondent aux appels d'offre.
- l'usage des données publiques d'avis de consultations est plutôt de produire des indicateurs sur la dépense publique et éventuellement de détecter certaines irrégularités (lutte contre la corruption)
- l'usage attendu ici est **les acheteurs publics** eux-mêmes, en leur permettant d'améliorer leurs pratiques d'achat en capitalisant sur les DCE les plus aboutis dans un domaine métier particulier, et par ailleurs de permettre aux **usagers/citoyens** (associations, etc.) d'accéder à des documents techniques détaillés sur tel ou tel projet. Les **chercheurs** travaillant sur la gestion publique peuvent trouver dans une telle archive un matériau utile pour de nombreux travaux. Par ailleurs, les **entreprises** (consultant, prestataires, start-ups, industriels...) peuvent aussi bénéficier d'une telle archive ouverte, notamment pour proposer de nouveaux services aux collectivités.

Cette archive permettrait d'accéder à un DCE après la fin de la consultation, lorsqu'il n'est plus disponible sur la place de marché, et à la limite, le principal intérêt de l'archive ouverte n'est pas de permettre d'accéder au DCE pendant la consultation proprement dite.

3 étapes de traitement?

Le projet pourrait comprendre 3 étapes successives de traitement des documents :

- 1- collecter
- 2- classer et publier
- 2- enrichir et annoter

Techniquement, le contexte est favorable aussi : de manière générale, de la disponibilité de serveurs peu coûteux pour stocker et traiter

1) d'outils de moissonnage (web scraping) open source (ex. en python : scrapy, selenium, requests/bs4)

2) de logiciels de gestion d'archives ouvertes (ex. d-space, invenio...)

3) d'outils open source de text mining (cf. [exemples](#))

Dans le domaine de la recherche et de la science ouverte, des archives ouvertes de plusieurs centaines de milliers de documents existent depuis des années, le domaine a acquis une bonne maturité. On se propose de transposer ce type d'activités à la gestion des documents publics.

A terme, si le projet fonctionne et suscite un intérêt et un financement suffisant à son fonctionnement, il est bien sûr envisageable de compléter les documents archivés et de les relier entre eux : associations entre avis de marché et DCE, et surtout livrables produits dans le cadre des marchés (études en particulier, pour les marchés de service, qui sont également des documents administratifs au sens de la loi)

Comment commencer?

Quoiqu'il en soit il faut commencer. On peut se limiter au départ à un territoire, et/ou à un domaine, comme les transports. On peut se répartir la tâche à plusieurs, l'archive doit être conçue pour fonctionner de manière distribuée.

Pour la récupération des DCE, il y a 2 grands scénarios :

- demander (et obtenir) de la part d'un acheteur public l'accès aux DCE
- récupérer ces DCE sur la plateforme d'achat de la collectivités

Et ce, pour toutes les collectivités et établissements en France, en commençant par les plus importantes. Heureusement, il y a un nombre beaucoup plus limité de logiciels de plateforme d'achat. Par ailleurs, beaucoup de plates-formes permettent de récupérer les DCE anonymement, ce qui rend faisable l'écriture de scripts pour les récupérer. Certaines ont mis un captcha (ex. région PACA) interdisant à un robot de récupérer les DCE.

un bon point de départ : scraper-place

En cherchant sur github, je suis tombé sur ce repo : [Voir la suite...](#)

From:

<https://openmarches.frama.wiki/> - **Archives ouvertes des marchés publics?**

Permanent link:

<https://openmarches.frama.wiki/start>

Last update: **2018/06/17 19:35**

