

Projet de fin de semestre
Réutilisation des données de l'Open Data français

Modalités :

- Travail en **groupe** (2 à 4 personnes).
- Restitution des **réalisations techniques** avant le **jeudi 07 janvier 2016** : données XML, programmes, résultats, commentaires, etc.
- **Présentation du projet** lors de la **dernière séance** du cours (07/01/2016) : oral (entre 15 et 30 minutes) suivi de questions.

Consignes :

Le site <http://www.data.gouv.fr/> est une **plateforme ouverte de diffusion de données publiques** (*Open Data*) mise à disposition par l'État français. Toutes les données produites ou reçues dans le cadre d'une **mission de service public** y sont mises en ligne, ainsi que des **données dites d'intérêt public** émanant de sources non-gouvernementales (citoyens, associations, entreprises privées, etc.).

Les données sont accessibles via un **moteur de recherche**, permettant une navigation par facettes. Il est ainsi possible de trier les jeux de données en fonction de leur **format** : CSV, JSON, HTML, ou même **XML** (→ <https://www.data.gouv.fr/fr/search/?q=&format=xml>). Les utilisateurs de la plateforme sont invités à **réutiliser ces données** et proposer des applications innovantes les exploitant.

L'objectif du projet est de **sélectionner** au **minimum deux jeux de données** différents, qui seront croisés pour en proposer une **interprétation** (graphiques, cartes, ou tout autre type de visualisation / d'utilisation). Pourquoi ne pas croiser, par exemple, les listes des musées, cinémas et bibliothèques de France (→ spatialité de l'offre culturelle en France, statistiques sur le nombre de lieux culturels / habitant et leur diversité, en fonction des régions) ? Ou celles des maires et autres élus de France, associées aux retranscriptions des débats de l'Assemblée Nationale (→ visualisation des élus les plus loquaces, avec leur lieu d'origine, ou de ceux qui cumulent le plus de fonctions et ne brillent pas par leurs interventions à l'Assemblée) ?

Des **exemples de réutilisations** des données publiques sont disponibles, pour inspiration, à cette adresse : <https://www.data.gouv.fr/fr/reuses/>.

Les **données sources** devront explicitement être **mentionnées**. Elles peuvent être dans **n'importe quel format** (XML ou non).

Vous devez, dans tous les cas, **créer un format XML pivot** associant vos différentes sources de données. Ce **format** devra être **décrit** au moyen d'un **langage de schéma** de votre choix (*DTD* ou *Relax-NG*) et l'**intégrité des données** des instances XML en découlant contrôlée à l'aide d'un **Schematron (optionnel)**. Ne sélectionnez que les **données** qui vous seront **utiles**, croisez les informations et n'hésitez pas à vous **détacher des modèles d'origine** pour proposer la **modélisation la plus appropriée**.

La ou les **instance(s) XML** de votre modèle pivot peuvent être créées **à la main** ou **via un programme** informatique (par exemple, des feuilles de style XSLT si les données sources sont en XML). Ce sont elles qui seront ensuite exploitées par vos programmes de visualisation / d'interprétation.

Si les **données d'origine** ne sont **pas en XML** (ex. : données tabulaires CSV) mais que vous souhaitez générer l'instance pivot à l'aide de XSLT, **vous pouvez les modifier** pour les passer en XML "à moindre frais" (→ structure XML basique puisque temporaire).

Vous pouvez **intégrer** dans votre **pivot** des espaces de noms et des **langages XML standardisés** (ex. : KML si vous structurez des données spatiales).

Si les données d'origine sont **trop volumineuses**, vous pouvez **les réduire** en sélectionnant une aire géographique particulière ou en jouant sur la dimension temporelle.

Dans un deuxième temps, **exploitez les données** de votre modèle pivot à l'aide de **XSLT**, afin d'en proposer **plusieurs interprétations**. Vous pouvez générer des **sorties** qui pourront être exploitées par différents services proposant un **affichage sous forme de cartes** ou de **graphiques** (voir des exemples d'outils en ligne sur [le site data.gouv.fr](http://le.site.data.gouv.fr) ou [ici](#), mais vous pouvez en utiliser d'autres ou créer vos propres outils). Vous pouvez également générer **une ou plusieurs pages HTML** à partir du XML pivot, présentant une analyse synthétique des données (ex. : tableaux croisant les informations avec ou sans fonction de tri, statistiques chiffrées, etc.). Vous êtes libres d'imaginer **toutes sortes d'exploitations de vos données** : le but étant d'obtenir **différentes sorties** à partir d'un **même XML pivot**.

Résumez l'intégralité du projet (différentes étapes, choix techniques, contraintes, prises de décisions, difficultés rencontrées, résultats obtenus, etc.) afin de le présenter à l'**oral** lors de la **dernière séance** du cours (le jeudi 07 janvier 2016). Le choix du **support de présentation** est laissé **libre** (diapositives, site Web, vidéo...).

Dans tous les cas, vous devez **produire un site Web** (même très simple - voir Google Sites) **centralisant et décrivant les différents fichiers produits** par vos programmes XSLT. Celui-ci pourra être **soumis** comme nouvelle contribution (réutilisation de données) **sur le site data.gouv.fr**.