



# Instructions pour la mise à disposition de données pour Hackathons

Version 2, 11 décembre 2015

## Choix des jeux de données, des données visuelles, sonores et vidéo

Les fondements du succès d'un Cultural Hackathon reposent sur la mise à disposition de données appropriées. Pour cela, il vous faut identifier, parmi vos fonds/ collections, des jeux de données dont vous possédez les droits d'utilisation ou, le cas échéant, pour lesquels les droits nécessaires peuvent être obtenus.

Merci de ne pas limiter les jeux de données à l'avance, mais de vérifier à l'interne combien de temps vous pouvez et voulez consacrer au Hackathon. Cherchez le dialogue avec des utilisateurs potentiels afin de faire ressortir quels sont les jeux de données qui vous paraissent particulièrement intéressants du point de vue des utilisateurs.

En principe, les jeux de données suivants présentent un intérêt pour le Hackathon:

- les métadonnées d'objets patrimoniaux
- images digitales, vidéos, enregistrements sonores / reproductions digitales d'objets patrimoniaux
- lexiques / indexes de mots-clés
- fichiers d'autorités
- données du domaine de l'économie culturelle ou économie de la création: par ex. données relatives à des événements culturels (représentations théâtrales et autres), statistiques qui éclairent l'utilisation des biens culturels, etc.

Considérez que, du point de vue des utilisateurs, la visualisation des métadonnées est souvent tout aussi importante que les métadonnées elles-mêmes. Dans le cas idéal, mettez à disposition des métadonnées avec le matériel correspondant, que celui-ci soit visuel, sonore ou vidéo (en particulier s'il n'y a pas de restrictions de droits d'auteur).

## **Libération des droits / choix de la licence**

Vérifiez qui détient les droits d'utilisation sur les données choisies et choisissez une licence adaptée en vue de leur publication. Nous sommes à disposition si vous avez besoin d'aide en matière d'argumentation ou si vous avez des questions quant aux licences. Pour des questions concernant le "Public Domain", l'Institut Fédéral de la Propriété Intellectuelle a publié un "FAQ" avec les questions les plus courantes.

### **Idéalement:**

- Creative Commons - Marque du Domaine public  
La marque "Public Domain" est conseillée pour toutes les oeuvres dont les droits d'auteurs sont échus.
- Creative Commons - Transfert dans le domaine public - "no rights reserved"  
Avec cette licence la personne renonce - dans les limites du droit en vigueur - à l'ensemble de ses droits d'auteurs sur une œuvre. L'utilisation de cette licence est conseillée par exemple lors de la publication de métadonnées, pour signaler aux utilisateurs qu'aucun droit d'auteur ne sera revendiqué sur ces métadonnées.
- Creative Commons - Attribution  
Cette licence autorise toutes les utilisations d'une œuvre, pour autant que l'auteur soit nommé. Lors de l'utilisation de cette licence, l'auteur ou l'institution doit être nommé. Si vous utilisez cette licence, merci de préciser l'intitulé précis selon lequel l'auteur ou l'institution doit être cité. De même, pour pouvoir utiliser cette licence, il faut que les œuvres concernées soit protégées par le droit d'auteur et que l'institution en possède les droits d'utilisation.

### **Est également accepté:**

- Creative Commons - Attribution, partage dans les mêmes conditions  
Cette licence oblige les utilisateurs d'une œuvre à publier leurs travaux à venir avec une licence identique à celle de l'œuvre qu'ils ont utilisée. Pour pouvoir utiliser cette licence, il faut que les œuvres concernées soit protégées par le droit d'auteur et que l'institution en possède les droits d'utilisation.

### **Merci de ne pas limiter l'utilisation commerciale des données**

En règle générale, les licences qui excluent toute utilisation à des fins commerciales sont particulièrement problématiques de par le fait qu'elles excluent également des scénarios légitimes de réutilisation: comme par exemple la plupart des comptes-rendus journalistiques ou l'intégration dans Wikipédia.

## Choix du format et mise à disposition

Pour pouvoir être utilisées dans le cadre du Hackathon, les données, en particulier les métadonnées liées aux d'objets patrimoniaux, doivent être lisibles par une machine. Les formats réputés aisément lisibles sont JSON (JavaScript Object Notation) et CSV (Comma Separated Values). Une migration des données vers ces formats par le moyen d'Excel ou d'une fonction d'export de base de données ne pose en général pas de problème. Vous ne devriez pas avoir de souci si vous disposez dans votre organisation d'un interlocuteur spécifique pour les questions techniques. Vous pouvez, bien entendu, nous contacter pour des questions concrètes. Nous pourrions, le cas échéant, vous aider à rendre vos données lisibles par machine.

En plus des formats sus-mentionnés, les formats suivants peuvent également dans certains cas de figure s'avérer adéquats pour les données: GeoJSON, KMZ, XML, RDF ou Shapefile. La communauté des développeurs apprécie que les données soient mises à disposition par le moyen de Web Services tels que des RESTful API, Web Feature Service (WFS) ou Web Map Service (WMS).

Des données images devraient avoir au moins 300 dpi et être mises à disposition dans un des formats ouverts courants tel que PNG, JPEG, SVG (et non pas en tant que GIF, PDF ou de document Word). Il est souvent judicieux de mettre directement à disposition des TIFF haute résolution (p.ex. pour l'upload sur Wikimedia Commons ou s'il est attendu que l'on utilise seulement certaines parties de l'image ou encore si l'on veut permettre que certaines images soient retouchées - dans tous ces cas-là, il est utile d'avoir des données images de la meilleure qualité possible).

Pour les données audio ou vidéo, les formats suivants sont adéquats: Ogg (en lien avec les codecs FLAC, Speex, Opus ou Vorbis), FLAC, WAVE ainsi que WebM.

Il est important de veiller à l'homogénéité des données et également à ce qu'elles puissent subir un traitement automatique. Par exemple, les ordinateurs et les développeurs préfèrent des années sous forme de millésimes plutôt que de la prose du genre "au temps de nos ancêtres les gaulois" ! La notation unifiée des noms aide à l'exploitation statistique des données.

Les données seront présentées avant l'événement sur le site web du Hackathon et proposées au téléchargement. Ainsi les participants auront l'occasion se pencher sur leur potentiel et pourront déjà chercher à explorer l'une ou l'autre piste.

Vous pouvez mettre à disposition vos données primaires ou audio-visuelles soit par le biais du site web de votre institution ou via un service externe. Pour la mise à disposition des données primaires, datahub.io ou le Government Data Portal sont des options adaptées. Wikimedia Commons ou Flickr se prêtent particulièrement bien pour la mise à disposition des données audio-visuelles. Contactez nous si vous avez des questions concernant l'un ou l'autre des services externes.

## **Continuité**

A moyen terme, il serait souhaitable que les données et les données audio-visuelles soient trouvables et restent accessibles à des tiers après le Hackathon. Par conséquent, nous conseillons de mettre à disposition vos données par le moyen du futur Open Government Data Portal ou Europeana ainsi qu'une mise à disposition des données audiovisuelles (métadonnées incluses) sur Wikimedia Commons. Noter bien que la publication des données sur ces plateformes exigent un certain temps de préparation et ne serait ainsi accessibles qu'après le Hackathon.

## **Contacts**

Vos interlocuteurs pour les questions relatives à la mise à disposition des données:

Beat Estermann  
031 848 34 38  
beat.estermann@openglam.ch

Oliver Waddell  
043 259 69 74  
oliver@waddell.ch