arbido

2013/4 Linked Open Data Big Data - Tout en réseau

Estermann Beat,

Les institutions patrimoniales suisses sont favorables à l'open data

Depuis l'avènement de l'Internet, les institutions patrimoniales connaissent des changements successifs nombreux. On par exemple la numérisation systématique des objets patrimoniaux, la mise en réseau accrue des organisations, la coproduction de contenus par les utilisatrices et les utilisateurs ainsi que la mise à disposition de données et de contenus sur l'Internet sous des licences «libres». Une étude de la Haute école spécialisée bernoise montre où les institutions patrimoniales suisses se situent dans ce contexte.

1. En allemand: Estermann Beat (2013): Schweizer Gedächtnisinstitutionen im Internet-Zeitalter. Ergebnisse einer Pilotbefragung zu den Themenbereichen Open Data und Crowdsourcing, Berner Fachhochschule, E-Government-Institut. / En anglais: Estermann Beat (2013): Swiss Heritage Institutions in the Internet Era. Results of a pilot survey on open data and crowdsourcing, Bern University of Applied Sciences, E-Government Institute.

Les institutions patrimoniales en pleine mutation

Le Musée de Derby mise sur des codes QR pour enrichir via des articles de Wikipédia les explications concernant les objets d'exposition, et ce en plusieurs langues bien entendu. La Bibliothèque nationale australienne propose à des bénévoles des tâches de correction et de transcription sur l'Internet, offre qui rencontre un véritable succès. Le Tropenmuseum à Amsterdam met des centaines de milliers de photos haute résolution sur Wikimedia Commons, se félicite que ses fonds soient désormais plus faciles d'accès et en espère en tirer profit par l'augmentation du nombre de visiteurs. Le Brooklyn Museum s'en remet au jugement des visiteurs de son site web pour sélectionner des photographies destinées à une exposition. Un consortium de bibliothèques européennes utilise le crowdfunding pour financer la numérisation des œuvres tombant dans le domaine public. A Vienne, des chercheurs misent sur les données libres d'accès pour visualiser d'une nouvelle façon d'appréhender l'histoire de l'art.

Ces exemples montrent les profonds changements qui traversent le champ des institutions patrimoniales depuis l'avènement de l'Internet dans les années 1990 et les innovations qu'il a induites.

On peut, concrètement, identifier plusieurs tendances successives: Depuis le début des années 2000, les institutions patrimoniales en Europe conjuguent leurs efforts afin de numé-riser de manière coordonnée et la plus complète possible les objets patrimoniaux et leurs métadonnées. Outre une mise en réseau renforcée qui permet aux institutions d'échanger leur savoir faire et d'éviter les doublons, sont également proposées dans la foulée des offres de «single-point-of-access» sous la forme de systèmes de catalogues interorganisations et de bibliothèques virtuelles.

Dès 2005, on assiste à l'avènement du Web social qui offre encore plus de possibilités d'interaction. Cette tendance se caractérise par de nouvelles formes de personnalisation et par une implication plus forte des utilisatrices et des utilisateurs, ou des visiteuses et des visiteurs respectivement, qui sont appelés à coproduire.

Le développement spectaculaire de Wi kipédia, Flickr et d'offres analogues basées sur le principe de la commu nauté virtuelle focalise de plus en plus l'attention sur la constitution et la col lecte collaborative de contenus sur l'In ternet. De premières collaborations s'installent entre des institutions patri moniales et des communautés en ligne, tandis que certaines institutions lancent des projets spécifiques en faisant appel au crowdsourcing2.

Les communautés Wikipédia et Flickr découvrent de leur côté que les institu tions partimoniales sont des parte naires importants et cherchent depuis quelques années à collaborer active ment avec elles. A partir de 2009, c'est le mouvement pour l'ouverture des données publiques qui, depuis les Etats-Unis et la GrandeBretagne, se propage rapidement pour toucher quelque 50 Etats de par le monde. Son exigence: les données des autorités publiques doivent être rendues accessibles sous une forme lisible par ordinateur afin qu'elles puissent être utilisées par des tiers. Si le mouvement pour l'ouverture des données publiques trouve son origine dans une vision de la politique étatique que l'on veut plus transparente et plus participative, il est également motivé par une vision technico-économique: la mise en réseau du plus grand nombre possible de données «ouvertes» doit permettre l'émergence de ce qu'il est convenu d'appeler le «Web sémantique». Les données des autorités publiques ainsi que, le cas échéant, celles d'entreprises privées, deviennent ainsi des ressources infrastructurelles libres d'accès que des tiers peuvent utiliser pour offrir des services à valeur ajoutée: pour la valorisation des données, la création et la visualisation de combinaisons de données d'un nouveau genre ou pour d'autres prestations basées sur des données. L'utilisation de licences standardisées «libres», qui permettent également de modifier et de diffuser plus loin les œuvres à des fins commerciales, est primordiale dans ce contexte, étant donné qu'elles réduisent sensiblement les coûts de transaction pour l'utilisation d'œuvres par des tiers[P. ex. les licences de droits d'auteur de Creative-Commons, telles que CC-by ou CC-by-sa (pour plus d'informations à ce sujet, voir: www.creativecommons.org)].

Où se situent aujourd'hui les institutions patrimoniales suisses par rapport à l'open data/open content et au crowdsourcing?

Afin de montrer où se situent actuellement les institutions patrimoniales suisses par rapport à l'open data /open content et au crowdsourcing, comment elles évaluent les chances et les risques de ces tendances et où elles voient leurs bénéfices potentiels, la Haute école spécialisée bernoise a réalisé une enquête pilote à la fin 2012. Quelque 200 institutions de Suisse alémanique d'importance nationale ont été sollicitées dans ce contexte[L'inventaire suisse des biens culturels d'importance nationale a servi de base pour la sélection des institutions.], dont 72 ont répondu au questionnaire.

Environ 60% des institutions interrogées mettent, à tout le moins en partie, à disposition sur l'Internet des objets patrimoniaux et des métadonnées. Quelque 40% échangent des métadonnées dans le cadre d'alliances. Ces deux pratiques, qui représentent des conditions importantes du point de vue de l'open data / open content et du crowd sourcing, sont donc déjà très répandues. Les institutions qui ne sont pas encore parvenues à numériser leurs fonds les plus importants, pour des raisons de manque de moyens ou de savoir-faire, peuvent toutefois causer quelque souci.

On constate deux dynamiques différentes en ce qui concerne l'open data/ open content et le crowdsourcing. Pour ce dernier, on constate un pourcentage certes relativement élevé (10%) des institutions interrogées qui disposent déjà de collaborateurs qui apportent leur contribution à Wikipédia dans le cadre de leur travail, ainsi qu'un pourcentage tout aussi élevé de participants à l'enquête qui estiment que le travail bénévole sur l'Internet joue un rôle plutôt important pour leur organisation. Cela étant, la plupart des institutions interrogées estiment que le crowdsourcing recèle davantage de risques que de chances. Il n'en va pas de même pour ce qui est de l'open data/open content: à la fin 2012, il n'y avait certes que très peu d'institutions qui étaient disposées à mettre des contenus disponibles en ligne libres de droit, mais les institutions interrogées estimaient néanmoins que les chances qu'offre une telle approche étaient bien réelles. Le feu, pour l'open data, est donc pour ainsi dire au vert! Comme l'on s'y attendait, les linked data constituaient également déjà un thème pour quelques-unes des institutions interrogées, mais aucune d'elles n'avait toutefois de projet dans ce contexte à la fin 2012.

Open data/open content: besoin d'information concernant les licences «libres»

Les résultats de l'enquête permettent de conclure qu'à ce jour moins de 10% des institutions interrogées mettent «librement» à disposition sur l'Internet des illustrations de leurs objets patrimoniaux, mise à disposition qui permet, autrement dit, à des tiers de les modifier et de les utiliser gratuitement. Plus de la moitié mettent certes des illustrations à disposition sur le Net, mais en limitent toutefois leur utilisation. Le plus souvent, les institutions sont prêtes à mettre gratuitement à disposition des contenus pour la formation et la recherche, pour une utilisation privée ou des projets d'intérêt public. Elles sont nettement plus réticentes lorsqu'il s'agit d'autoriser la mise en ligne de contenus qui peuvent être utilisés à des fins commerciales ou faire l'objet de modifications.

Par rapport aux licences «libres», il existe en outre un besoin d'information: près de la moitié des institutions qui rendent des objets patrimoniaux accessibles en ligne ne semblent pas être conscientes du fait qu'il n'est pas possible de libérer des œuvres pour leur utilisation dans Wikipédia et en même temps d'interdire à des tiers de modifier des images ou de les utiliser à des fins commerciales.

Plus de la moitié des institutions interrogées considèrent l’open data comme important. Les groupes cibles les plus importants dans ce contexte sont les institutions de recherche et de formation, les personnes privées et d’autres institutions culturelles. Les institutions interrogées estiment que les plus grandes chances de l’open data/open content résident dans la meilleure visibilité et accessibilité des fonds, dans la meilleure visibilité des institutions elles-mêmes ainsi que dans leur meilleure mise en réseau. Concernant les risques de l’open data, les institutions interrogées craignent surtout des dépenses supplémentaires en ressources et une certaine perte de contrôle sur les données, en particulier pour ce qui est des violations des droits d’auteur, de la protection des données et des infractions contre les règles de confidentialité.

Crowdsourcing: important mais porteur de risques

Près de 40% des institutions interrogées considèrent le crowdsourcing comme étant important. La mise en œuvre du crowdsourcing, c’est-à-dire la possibilité de créer ou de compléter des contenus sur l’Internet par des bénévoles qui collaborent entre eux, est considérée par ces mêmes institutions comme envisageable surtout dans les domaines de la classification et de la complétion des métadonnées ainsi que pour des tâches de correction et de transcription. Près de la moitié des institutions constatent par ailleurs des besoins d’amélioration dans le domaine des métadonnées, en particulier pour ce qui est de l’intégralité, de la disponibilité et de la numérisation de ces dernières.

En revanche, les institutions interrogées estiment que le crowdsourcing présente lui aussi une série de risques. Parmi ces derniers figurent la charge importante qui y est liée au niveau de la préparation et du suivi. Autres risques: la continuité incertaine de l’entretien des données, de la difficulté d’influer sur les résultats et d’estimer correctement la charge temporelle ainsi que la faible sécurité en matière de planification.

Conclusions

Les résultats de l’enquête montrent que miser sur des approches de type crowdsourcing implique une certaine charge et des impondérables. Cela s’explique également par le fait que les institutions doivent participer activement aux communautés de bénévoles et de partenaires impliquées, ce qui ne manque souvent pas de poser des problèmes compte-tenu des différentes cultures organisationnelles. Les processus d’intégration qui y sont liés peuvent certes, en fin de compte, déboucher sur un bénéfice pour les deux parties, mais cela ne va en général pas sans un investissement temporel considérable.

En contrepartie, opter pour une stratégie open data/open content est nettement plus simple pour les institutions. Il est donc conseillé de miser, à court terme et à grande échelle, surtout sur des approches qui permettent aux institutions la «libre» mise à disposition des données et des contenus, sans qu'elles aient à assumer elles-mêmes les dépenses qu'impliquent de longs processus de coopération. Néanmoins, une certain niveau de standardisation des métadonnées, de la qualité des données et des interfaces techniques pourrait malgré tout être utile pour réduire au minimum les coûts de transaction. Il s'agit également de mettre l'offre en rapport avec la demande, car si tel n'est pas le cas, on court le risque que les données et les contenus mis à disposition restent inutilisés.

Compte tenu du fait que de nombreuses institutions considèrent la thématique du crowdsourcing et de la création collaborative de contenus comme étant aussi importante que l'open data/open content, il serait judicieux de mener quelques projets-pilotes dans ce domaine et d'entamer un processus d'apprentissage collectif parmi les institutions partimoniales et les communautés en ligne concernées. Les conditions pour ce faire sont une culture positive de l'erreur, la documentation et l'évaluation systématique des expériences faites ainsi que la création de plateformes d'échange appropriées, tels que conférences, ateliers ou offres de formation continue.

Beat Estermann

collaborateur scientifique, Institut pour le gouvernement e?lectronique de la Haute e?cole spe?cialise?e bernoise