Choix d'une licence pour la publication de données bibliographiques en Linked Open Data

Février 2015

Sommaire

1 Introduction....2
2 Diagnostic....2
3 Situation juridique de départ....3
3.1 Propriété intellectuelle....3
3.1.1 Protection d'une base de données....3
3.1.2 Protection d'une notice isolée....3
3.2 Contrats....4
4 Les licences ouvertes....4
4.1 Domaine public....4
4.1.1 L'exemple DPLA....5
4.2 Attribution....5
4.2.1 L'exemple OCLC....5
4.3 Partage dans les mêmes conditions....5
4.4 Licences pas totalement ouvertes....6
5 Enjeux dans le contexte des Linked Open Data....6
6 Conclusion....6
7 Index des jeux de données bibliographiques et licences....8
8 Bibliographie....9

Ce document ne vaut pas comme document juridique.

1 Introduction

« Le modèle économique de la vente des données, c'est fini. » E. Bermès (1)

Le mouvement de l'Open Data prend de plus en plus d'importance, notamment dans les institutions publiques telles que les bibliothèques. Il s'agit de publier des données sous une licence qui, selon les modalités choisies, peut permettre la réutilisation, la modification et le partage de manière libre.

Le web sémantique est une autre tendance technologique en fort développement, initiée par Tim Berners-Lee au début des années 2000. Il vise à relier les données disponibles sur le web en se basant sur l'utilisation de standards communs tels que RDF.

Relier des données hétérogènes suggère donc une réutilisation de divers jeux de données externes à sa propre institution. Cet aspect a donc soulevé, plus récemment, la question de l'Open Data (2). C'est ainsi qu'est né le concept de Linked Open Data.

Les bibliothèques s'intéressent à cette mouvance. Néanmoins, elles font face à diverses questions quant à l'ouverture de leur données. Faut-il totalement libérer les données dans le domaine public ou choisir une licence ouverte plus restrictive ? Quels sont les enjeux que cela implique d'un point de vue pratique et technique dans un projet de Linked Data ?

Ce document expose tout d'abord les éléments du diagnostic à effectuer en vue du passage à l'Open Data. Il introduit ensuite brièvement la situation juridique sans licence (par défaut), puis présente les licences ouvertes possibles. Enfin, les enjeux de ce choix dans le contexte des Linked Data sont expliqués.

2 Diagnostic

Publier ses données en Open Data nécessite la réalisation d'un diagnostic préalable pour permettre le choix de la meilleure solution juridique. Ce diagnostic peut se résumer en quatre étapes (3) :

1. Que fait-on ou veut-on faire avec les données ?

Trois cas de figures sont possibles, qu'il s'agit d'identifier :

1. utiliser les données au sein de l'institution et de son site
2. mettre à disposition les données via des services web (Z39.50, SRU, SPARQL-Endpoint, etc.)
3. fournir les données (notamment en masse) à un tiers dans le cadre d'une relation établie ou d'un contrat.

2. Quelles sont les relations contractuelles en vigueur ?

Il s'agit ici de lister les partenaires, notamment les fournisseurs de données, et de déterminer les clauses contractuelles en lien avec la réutilisation ou la publication des données. Il s'agit par exemple aussi d'identifier si certaines données possèdent des contenus générés par les utilisateurs, auquel cas leur libre réutilisation n'est pas forcément possible.

3. Les relations contractuelles sont-elles en conformité avec l'activité envisagée ?

Cette analyse peut nécessiter l'aide d'un spécialiste des questions juridiques. Le but est de savoir si et dans quelle mesure les clauses contractuelles en vigueur peuvent restreindre l'activité envisagée avec les données (utiliser, mettre à disposition, fournir).

4. Répondre aux risques

Si les clauses contractuelles sont peu claires ou interdisent certaines activités envisagées, il faut évaluer les solutions possibles ainsi que la gravité du risque envisagé, et réagir en conséquence. Par exemple, on peut décider de ne rien faire, de négocier un nouveau contrat avec le fournisseur actuel, de renoncer à l'activité ou d'implémenter une solution technique permettant de respecter les clauses contractuelles.

5. Pérennisation des mesures de diagnostic

Pour être intégrée aux processus d'une institution, l'analyse de la situation juridique peut requérir du personnel qualifié. Des actions concrètes peuvent être mises en place pour le futur, par exemple la signature de nouveaux contrats uniquement s'ils sont en conformité avec les activités exercées.

3 Situation juridique de départ

3.1 Propriété intellectuelle

Cette section brosse un court aperçu de la situation juridique des données de bibliothèques lorsqu'elles ne possèdent pas de licence spécifique. La législation en vigueur peut donc être appliquée par défaut.

3.1.1 Protection d'une base de données

Les métadonnées des bibliothèques peuvent être protégées par le droit d'auteur, car elles forment des bases de données considérées comme des recueils en Suisse et en Europe.

> *« Les bases de données peuvent être protégées par le droit d'auteur à titre de recueils, indépendamment du fait que les différentes données qu'elles recèlent soient protégées ou non, si elles possèdent un caractère individuel en raison du choix et de la disposition des données. »*
>
> Institut fédéral de la propriété intellectuelle (4)

En Suisse, la *Loi fédérale sur le droit d'auteur et les droits voisins* définit dans l'article 4 la notion de recueil :

> *1 Les recueils sont protégés pour eux-mêmes, s'ils constituent des créations de l'esprit qui ont un caractère individuel en raison du choix ou de la disposition de leur contenu.*
> *2 La protection des œuvres réunies dans les recueils est réservée.*

Ainsi, une base de données dans son ensemble est protégée, mais pas les données prises individuellement. Dans le contexte des bibliothèques, une donnée correspond à une notice, une indication de titre, d'auteur, etc.

Les données provenant de fournisseurs externes (OCLC ou autres bibliothèques) sont protégées car elles forment un ensemble (un recueil). Elles ne peuvent donc pas être republiées sous une licence contraire aux dispositions juridiques originelles.

3.1.2 Protection d'une notice isolée

Une notice bibliographique prise individuellement se compose généralement de deux types de données :

1. Les données centrales, purement descriptives. Exemples : auteur, titre, éditeur, année.
2. Les données secondaires. Exemples : liens, table des matières, mots-clés, couvertures, données d'exemplaires

Les données centrales prises individuellement ne sont en principe pas protégées. En revanche, les données secondaires peuvent l'être (5), d'une part parce qu'elles peuvent provenir de fournisseurs externes revendiquant certains droits, d'autre part parce qu'elles peuvent prétendre à une certaine originalité les soumettant à la loi sur le droit d'auteur (par exemple un résumé rédigé par un catalogueur).

3.2 Contrats

Un contrat conclu avec des partenaires peut limiter ou contraindre l'accès à certaines données plus que ne le ferait la législation par défaut.

4 Les licences ouvertes

Publier sans licence signifie laisser les lois actuelles protéger par défaut ses données. Choisir une licence pour ses données permet d'une part de clarifier la situation juridique quant à leur réutilisation (qui peut être opaque) et d'autre part de modifier cette situation, par exemple libérant les données dans le domaine public. En 2014, moins de 10 % des fournisseurs indiquaient la licence à appliquer à leurs données en Linked Data (6).

Il est possible d'adopter une licence développée « maison » selon les besoins de l'institution ou d'adopter une licence préexistante. Néanmoins, cette seconde solution est fortement recommandée dans l'optique d'une réutilisation des données :

> *« [...] the use of standard licensing terms can lead to greater interoperability of data as well as increased user awareness of the licence terms, thereby enabling better compliance. »*
>
> Korn & Oppenheim, 2011 (7)

Deux standards dans le domaine existent :

1. Les licences Creative Commons (CC)
   Le standard le plus répandu, tous domaines confondus.
2. Les licences Open Data Commons (ODC)
   Le standard prévu en particulier pour les données et les jeux de données.

Le choix de l'une de ces licences est une décision politique importante. En effet, une fois attribuées, ces licences ne peuvent être révoquées (sauf pour une licence moins restrictive). Ces standards disposent d'un certain choix de clauses plus ou moins restrictives et librement combinables, notamment Creative Commons.

4.1 Domaine public

Le type de licence le plus ouvert et celui qui permet de renoncer tous les droits (copie, modification, distribution, représentation, y compris à des fins commerciales) – dans la limite du

droit en vigueur – en dédiant l'objet au domaine public.

Exemples de licences :

- Licence *Creative Commons Zero 1.0* (CC0)[1]
- Licence *ODC Public Domain Dedication and Licence* (PDDL)[2]

Avantages :

- Aucun contrôle n'est nécessaire ou à envisager.
- Le jeu de données est juridiquement compatible avec tous les autres.
- La réutilisation des données est favorisée, car elle n'est entravée par aucune contrainte, telle que mention d'attribution ou partage dans les mêmes conditions (voir chapitres suivants), qui peut nécessiter une analyse juridique avant la réutilisation.

4.1.1 L'exemple DPLA

La Digital Public Library of America (DPLA) (8) a publié toutes ses données en licence CC0. Pour ce faire, elle a dû conclure un accord avec tous ses fournisseurs de données, afin qu'ils renoncent à leurs droits, au moins dans le cadre d'une publication sur le site web de la DPLA.

Par ailleurs, la DPLA a émis un document incitant les réutilisateurs à mentionner la source de leurs données ainsi qu'à les publier sous les mêmes conditions.

4.2 Attribution

Bien que permettant de renoncer à la plupart des droits, certaines licences ouvertes demandent aux réutilisateurs de l'objet d'en mentionner sa paternité (clause d'attribution).

Exemples de licences :

- Licence *Open Data Commons Attribution License* (ODC-BY)[3]
- Licence *Creative Commons Attribution 4.0* (CC-BY)[4]

Avantages :

- Garantie d'une meilleure visibilité du producteur des données sur le web.

4.2.1 L'exemple OCLC

OCLC (9) a publié quatre jeux de données sous licence ODC-BY : WorldCat.org Linked Data, WorldCat Works data, VIAF et FAST. Les identifiants OCLC (exemples : OCLC Numbers, VIAF Ids, URIs, WorldCat Work URIs), indépendamment de l'objet qu'ils décrivent, ont été déclarés comme relevants du domaine public.

OCLC a publié des directives sur la manière de mentionner la paternité des données lors d'une réutilisation. Divers cas de figures avec solutions sont détaillés, par exemple pour un document, un site ou une page web, un jeu de données, etc. En cas de problème d'application de cette clause d'attribution, l'utilisation de l'identifiant OCLC uniquement est considéré comme suffisant.

4.3 Partage dans les mêmes conditions

Certaines licences permettant aux réutilisateurs de copier et modifier l'objet peuvent comporter, en plus de la clause d'attribution, une clause demandant, en cas de redistribution, à utiliser la même licence que celle avec laquelle l'objet original a été diffusé.

---

1 https://creativecommons.org/publicdomain/zero/1.0/
2 http://opendatacommons.org/licenses/pddl/1.0/
3 http://opendatacommons.org/licenses/by/1-0/
4 https://creativecommons.org/licenses/by/4.0/

Exemples de licences :

- Licence *Open Database License* (ODbL)[5]
- Licence *Creative Commons Attribution-ShareAlike 4.0* (CC-BY-SA)[6]

Avantages :

- Garantie d'une meilleure visibilité du producteur des données sur le web (attribution).
- Promotion délibérée de l'Open Data en garantissant que tout service se basant sur les données du fournisseur soit également un service Open Data.

4.4 Licences pas totalement ouvertes

Les autres clauses supplémentaires restreignant par exemple la modification ou la réutilisation à des fins commerciales sont considérées par la fondation *Open Knowledge* comme non conformes à la définition de Open Data (10). Ces licences ne sont pas conséquent par recommandées dans ce contexte.

Exemples de licences :

- Licence *Creative Commons Attribution-NoDerivatives 4.0* (CC-BY-ND)[7]
- Licence *Creative Commons Attribution-NonCommercial 4.0* (CC-BY-NC)[8]

5 Enjeux dans le contexte des Linked Open Data

Les clauses d'attribution ou de partage dans les mêmes conditions peuvent représenter des obstacles dans le contexte des Linked Open Data. En effet, un service LOD réutilise des données de provenances diverses, en les fusionnant, les reliant et en effectuant diverses autres modifications. Le respect de telles restrictions juridiques peut rendre la création d'un service à valeur ajoutée basé sur ces données très difficile, voire impossible. En effet, il s'agirait alors de mentionner la provenance des données à un niveau de granularité très précis, par exemple au niveau du triplet RDF, ce qui est techniquement complexe. Cette situation de blocage est parfois appelée « attribution stacking » (7).

Ainsi, la décision prise par OCLC d'exiger l'attribution lors de toute réutilisation de ses données (licence ODC-BY) soulève une certaine polémique au sein des bibliothèques. D'une part, cette base de données n'est pas forcément protégée selon le copyright américain, et d'autre part les données ne proviennent – à l'origine – pas toutes d'OCLC mais devraient être publiées avec cette mention (11). Par ailleurs, OCLC s'attribue la paternité des données de milliers de bibliothèques dans le monde.

Sur cette base, le groupe allemand DINI-AG KIM, compétent et actif dans la gestion des métadonnées et du web sémantique, encourage les fournisseurs à mettre à disposition leurs données en accès ouvert (gratuitement), selon des standards ouverts (formats et protocoles) et sous licence ouverte. Pour cette dernière, la publication des données de bibliothèque dans le domaine public (licence CC0 ou PDDL) est recommandée (12).

Pour cette raison, plusieurs institutions ont fait le choix de renoncer à tous droits (domaine public), avec *recommandation* par exemple de l'attribution de la paternité ou du partage dans les mêmes conditions.

---

5 http://opendatacommons.org/licenses/odbl/1-0/
6 http://creativecommons.org/licenses/by-sa/4.0/
7 http://creativecommons.org/licenses/by-nd/4.0/
8 http://creativecommons.org/licenses/by-nc/4.0/

6 Conclusion

La question des licences nécessite la réalisation préalable d'un diagnostic juridique au sein de son institution. Cette analyse mènera au choix d'une licence particulière, une décision politique de première importance car en principe irréversible. De plus, cette décision aura un impact sur les tiers souhaitant réutiliser les données et promouvra (ou non) le mouvement de l'Open Data.

Les deux licences ouvertes les plus utilisées sont sans doute la licence *Creative Commons Zero* et la licence *Open Data Commons Attribution*. La première permet de renoncer à tous ses droits en libérant les données dans le domaine public, tandis que la seconde exige la mention de la paternité des données.

Dans le contexte du web sémantique et d'un fort mouvement de réutilisation des données, les licences comportant des clauses d'attribution ou de partage dans les mêmes conditions peuvent mener à des difficultés techniques lors du développement de services web basés sur plusieurs sources de données aux situations juridiques diverses.

L'une des solutions couramment employée en bibliothèque est l'adoption d'une licence de type « domaine public », tout en recommandant aux tiers la mention de la source en cas de réutilisation.

7 Index des jeux de données bibliographiques et licences

Bibliothèque ou jeu de données	Licence	Source
Biblioteca nacional de España	CC0	http://www.bne.es/es/Inicio/Perfiles/Bibliotecarios/DatosEnlazados/
Bibliothèque nationale de France (data.bnf.fr)	Licence ouverte française (type CC-BY)	http://data.bnf.fr/about#Ancre10
Bibliothèque nationale suisse	CC0	http://www.nb.admin.ch/dienstleistungen/04161/index.html
British Library	CC0	http://www.bl.uk/bibliographic/datafree.html
CERN	PDDL	http://library.web.cern.ch/library_projects/bookdata
Deutsche Nationalbibliothek	CC0	http://www.dnb.de/DE/Service/DigitaleDienste/Datendienst/datendienst_node.html
Digital Public Library of America	CC0	http://dp.la/info/wp-content/uploads/2013/04/DPLADataUseBestPractices.pdf
Europeana	CC0	http://www.europeana.eu/portal/rights/terms-of-use.html
Harvard Library (USA)	CC0	http://openmetadata.lib.harvard.edu/bibdata/useterms
Hochschulbibliothekszentrum Nordrhein-Westfalen (Allemagne)	CC0	http://lobid.org/about
Library of Congress : LCSH, LOCNA, MARC Countries, etc.	[domaine public]	http://id.loc.gov/about/#aboutTOS
OCLC : dewey.info	CC-BY-NC-ND	http://dewey.info/
OCLC : WorldCat.org Linked Data, WorldCat Works data, VIAF, FAST	ODC-BY	http://www.oclc.org/data/attribution.en.html
RERO (Suisse)	CC0	http://data.rero.ch/
SUDOC	Licence ouverte française (type CC-BY)	http://www.abes.fr/Liens-utiles/Mentions-legales-et-CGU

8 Bibliographie

1. BERMÈS, Emmanuelle. IFLA 2012 (suite): quelques questions existentielles sur le Linked Data. *Figoblog* [en ligne]. 21 août 2012. [Consulté le 16 février 2015]. Disponible à l'adresse : http://figoblog.org/2012/08/21/2010/
2. BERNERS-LEE, Tim. Linked Data. *World Wide Web Consortium* [en ligne]. 2010. [Consulté le 3 mars 2015]. Disponible à l'adresse : http://www.w3.org/DesignIssues/LinkedData.html
3. JISC. Transfer and use of bibliographic records: guidance on legal issues. *JISC legal information* [en ligne]. 2011. [Consulté le 20 février 2015]. Disponible à l'adresse : http://www.jisclegal.ac.uk/Projects/TransferandUseofBibliographicRecords.aspx
4. INSTITUT FÉDÉRAL DE LA PROPRIÉTÉ INTELLECTUELLE (SUISSE). Questions fréquentes: droit d'auteur, domaine public. *Institut fédéral de la propriété intellectuelle* [en ligne]. 22 décembre 2014. [Consulté le 2 janvier 2015]. Disponible à l'adresse : https://www.ige.ch/fr/service/questions-frequentes/droit-dauteur/public-domain.html
5. KREUTZER, Till. *Open Data: Freigabe von Daten aus Bibliothekskatalogen* [en ligne]. Köln : Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen, 2011. [Consulté le 17 février 2015]. Disponible à l'adresse : http://www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/open-data-leitfaden.pdf
6. SCHMACHTENBERG, Max, BIZER, Christian et PAULHEIM, Heiko. State of the LOD cloud 2014. *University of Mannheim, Data and Web Science Group* [en ligne]. 30 août 2014. [Consulté le 3 mars 2015]. Disponible à l'adresse : http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/
7. KORN, Naomi et OPPENHEIM, Charles. *Licensing Open Data: a practical guide* [en ligne]. juin 2011. JISC Discovery. [Consulté le 20 février 2015]. Disponible à l'adresse : http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf
8. DPLA. Policies. *Digital Public Library of America* [en ligne]. 17 septembre 2014. [Consulté le 23 février 2015]. Disponible à l'adresse : http://dp.la/info/about/policies/
9. OCLC. Data licenses and attribution. [en ligne]. 14 avril 2014. [Consulté le 19 février 2015]. Disponible à l'adresse : http://www.oclc.org/data/attribution.en.htmlData licenses & attribution
10. OPEN KNOWLEDGE. Non-conformant licenses. *Open definition* [en ligne]. 16 février 2015. [Consulté le 16 février 2015]. Disponible à l'adresse : http://opendefinition.org/licenses/nonconformant/
11. VOLLMER, Timothy. Library catalog metadata: open licensing or public domain? *Creative Commons* [en ligne]. 14 août 2012. [Consulté le 23 février 2015]. Disponible à l'adresse : http://creativecommons.org/weblog/entry/33768
12. DINI-AG KIM. Empfehlungen zur Öffnung bibliothekarischer Daten. *Kompetenzzentrum interoperable Metadaten* [en ligne]. 31 octobre 2011. [Consulté le 23 février 2015]. Disponible à l'adresse : https://wiki.dnb.de/pages/viewpage.action?pageId=45419980