HAL
open science

VISUALISER LES DONNÉES DU CATALOGUE

Raphaëlle Lapôtre

▶ **To cite this version:**

Raphaëlle Lapôtre. VISUALISER LES DONNÉES DU CATALOGUE. Emmanuelle Bermès. Vers de nouveaux catalogues, Éditions du Cercle de la Librairie, pp.37-47, 2016, 9782765415138. 10.3917/elec.berme.2016.01.0037 . hal-02873679

HAL Id: hal-02873679

https://hal.science/hal-02873679v1

Submitted on 10 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Visualiser les données du catalogue

Raphaëlle Lapôtre

Introduction

Pourquoi poser aujourd'hui la question de la visualisation des données des bibliothèques, et en particulier celles des catalogues des bibliothèques ? Une réponse que nous pouvons tenter d'apporter à cette question est d'ordre juridique et technologique. En premier lieu, il est devenu plus aisé d'obtenir des données publiques aujourd'hui, notamment par le biais de plateformes de téléchargement comme data.gouv.fr, d'interfaces de programmation applicatives (API) ou encore, dans le contexte du web sémantique, via un SPARQL endpoint. Les métadonnées, ou données descriptives des catalogues des bibliothèques, n'échappent pas à ce processus, en témoigne notamment le projet data.bnf.fr, qui vise à diffuser en ligne et sous la licence libre Etalab les données des différents catalogues de la Bibliothèque nationale de France.
Les disciplines se sont également multipliées, pour essayer de tirer la substantifique moelle de ces données, parmi lesquelles le traitement de données, l'apprentissage automatique, les modèles de probabilités, ou encore la visualisation de données. En ce qui concerne cette dernière, elle devient de plus en plus accessible grâce à des logiciels grand public qui permettent de les générer : Excel et Google Sheet, par exemple, permettent de produire facilement de simples histogrammes, courbes et autres classiques diagrammes circulaires. Tableau Public, quant à lui, est un logiciel gratuit de type glisser-déplacer, qui reprend ces fonctionnalités et en propose d'autres plus approfondies et interactives. De même Gephy, logiciel libre qui s'est spécialisé dans la visualisation de réseaux ou graph de données, et qui a notamment été employé dans la visualisation de la New York Public Library (NYPL) sur laquelle s'appuiera une partie de notre réflexion.
Par ailleurs, de plus en plus de bibliothèques de programmation permettent de construire des représentations visuelles plus complexes et sur mesure. Citons à titre d'exemple jquery, d3.js, processing.js ou encore three.js pour les visualisations tridimensionnelles de données. C'est donc à première vue cette facilitation de l'utilisation et de la réutilisation des données qui nous conduit aujourd'hui à nous intéresser aux expérimentations visuelles conduites autour de l'information bibliographique.
Mais la question que souhaite poser cet article est surtout celle de la signification d'une telle démarche. Nous souhaiterions démontrer en effet que la visualisation des métadonnées, loin d'être un nouveau gadget contemporain, doit plutôt être vu comme la dernière évolution d'un attribut qui a toujours fait partie inhérente du projet moderne de description bibliographique.
Le catalogue, par sa dimension sémiologique est en effet un artéfact où la métaphore graphique est omniprésente, y compris sur le web où elle s'inscrit dans une logique représentationnelle qui relève du même projet classique d'ordonnancement de l'information et de l'attention. Le catalogue est en effet tableau, carte ou dessin, là où la visualisation est elle-même essentiellement cartographie, transposant le réel en signes et les disposant significativement à l'échelle d'un écran d'ordinateur. Cette dernière ouvre donc des perspectives heuristiques nouvelles sur le catalogue et les collections qu'il décrit, mais elle est aussi, par son aspect symbolique et réflexif, un nouveau "méta" à ajouter au "méta" de métadonnées : en tant que pensée sur la pensée documentaire et informationnelle, elle confère une visibilité nouvelle au catalogue et à son langage bibliographique qui participe pleinement au mouvement d'ouverture actuel de l'information publique.

Le web, un théâtre de signes

Assurément, la visualisation des données du catalogue n'est pas un phénomène nouveau, plus exactement, elle ne date pas de cette "ère de la donnée" qui caractérise notre temps présent, celui des données massives, des sciences des données ou encore, des données

ouvertes et liées. Si en effet l'on essaye de théoriser la visualisation de données, force est de constater qu'il s'agit d'un objet d'étude qui se dérobe à notre regard : suffirait-il, pour la définir, de parler d'"affichage visuel d'informations quantitatives", à la manière d'Edward Tufte, de "visualisation de l'information" selon les mots de Lev Manovich, ou encore de "projections", "métaphores visuelles, transposant différents types de données quantitatives selon des formes spatiales et graphiques" ? Lorsque l'on sait que le mot "visualisation" peut désigner tout aussi bien le dispositif technologique permettant de lire des documents en ligne dans une bibliothèque numérique que l'interface graphique d'une application visant à diffuser de l'information bibliographique sur les collections d'une bibliothèque, on s'aperçoit que l'objet dont nous prétendons parler ici est loin d'être univoque, stable dans le temps et bien limité conceptuellement : que nous parlions d'un diagramme en bâtons, d'une liseuse numérique ou d'un Online Public Access Catalog (OPAC), nous parlons toujours de visualisation de données, la donnée étant elle-même un objet protéiforme, tantôt pixel, texte ou encore description brève et formalisée d'un objet réel. Mieux vaudrait peut-être parler, à l'instar de Johanna Drucker, de "graphesis", à savoir d'"expressions visuelles qui sont des arrangements de marques ou de formes visuelles organisées en vue d'être lues sur et en tant que surfaces planes (en d'autres termes, dans leur forme littérale et visible plutôt qu'en tant qu'illusions picturales)" : en bref, "les artéfacts graphiques présentent la connaissance à travers la combinaison de codes symboliques et les relations structurées de ces éléments sur un champ plat." Encore cette définition omet-elle peut-être un élément majeur et ubiquitaire aujourd'hui, à savoir le web et sa logique économique. En effet, l'OPAC est par définition une base de données bibliographiques accessible publiquement en ligne : cela signifie que l'interface du catalogue est non seulement un ensemble de signes organisés sur une surface qui se donne à lire au public, mais également un dispositif graphique qui tend à s'adapter de plus en plus aux conventions du web, qui cherche à se rendre visible aux usagers, à tirer parti de ce bien rare qu'est l'attention des internautes. Si l'on devait donc définir rapidement les catalogues sur le web, nous parlerions d'organisation graphique de signes désignant des documents selon un système de valeur concurrentiel fondé sur l'*apparaître* sur un écran. Pour tenter de comprendre tous les aspects de cette définition, peut-être serait-il opportun de faire appel à la notion de "représentation", telle que la développait Michel Foucault dans son ouvrage *Les Mots et les Choses*. Le catalogue est en effet une représentation, en ce qu'il est un système de signes visant à décrire un monde, en l'occurrence bibliographique et documentaire.

> "La conservation de plus en plus complète de l'écrit, l'instauration d'archives, leur classement, la réorganisation des bibliothèques, l'établissement de catalogues, de répertoires, d'inventaires représentent, à la fin de l'âge classique, plus qu'une sensibilité nouvelle au temps, à son passé, à l'épaisseur de l'histoire, une manière d'introduire dans le langage déjà déposé et dans les traces qu'il a laissées un ordre qui est du même type que celui qu'on établit entre les vivants."

Autrement dit, les catalogues sont en quelque sorte une déclinaison des systèmes de classement propres à l'Histoire Naturelle, discipline apparue au cours de ce que Foucault appelle l'âge classique et qu'il définit comme ayant présidé à la représentation taxinomique, à savoir l'établissement d'un système de signes fondé sur une analyse des "natures complexes" s'appuyant sur un processus de comparaison. De fait, les métadonnées, données descriptives dont les catalogues sont les recueils, relèvent bien de la sémiologie instaurée par l'ordre classique : à la manière des biologistes, le bibliothécaire instaure un nombre réduit de variables, universellement observables, qui permettront tout à la fois de désigner et de distinguer les documents les uns des autres. Ces variables sont les nom et prénom de l'auteur, le titre, les lieu et date de publication, l'éditeur, le format, le nombre de pages. Seront exclues les variables, pourtant essentielles à l'histoire du livre mais non univoques que sont la couleur de la couverture ou encore l'odeur et le grain du papier. Le catalogue est donc une représentation, en tant que filtre d'observation permettant la transformation du réel en langage, mais aussi parce que le catalogue est fondamentalement une table, un tableau, "étalement des connaissances dans un système contemporain de lui-même", selon un ordre établi par analyse et découpant les objets dans la nappe continue des êtres de papier.

Dans l'essence même du projet de catalogue moderne il y a donc la visualisation, l'organisation du réel documentaire au travers de signes disposés sur une surface plane,

qu'elle soit métaphorique ou physique : "Il est caractéristique que l'exemple premier d'un signe que donne la Logique de Port-Royal, ce ne soit ni le mot, ou le cri, ni le symbole, mais la représentation spatiale et graphique, - le dessin : carte ou tableau." Il est caractéristique que parmi les rédacteurs de ce que l'on peut considérer comme l'un des premiers catalogues modernes, à savoir le catalogue général de la Bibliothèque nationale, figure Nicolas Clément, également inventeur de la cotation du même nom : la cote, petit code qui tout en désignant un exemplaire, le localise spatialement mais aussi et surtout, le situe dans l'ordre des connaissances et du savoir d'une époque. "De cet espace en tableau, il faut entreprendre maintenant l'analyse, là où il apparaît sous sa forme la plus claire, c'est-à-dire dans la théorie du langage, de la classification et de la monnaie".
Nous l'avons dit plus haut, parler des catalogues aujourd'hui implique nécessairement de décrire le fonctionnement de ces derniers dans l'environnement du Web. Or, force est de constater que par sa logique économique, le web est, lui aussi, affaire de représentation. Ainsi, analysant l'algorithme kNN utilisé par Google mais également par la plupart des sites web commerciaux dont le commerce repose sur le principe de la recommandation, Thomas Neal situe la philosophie économique de Google du côté de l'utilitarisme :

> "les interfaces web modernes permettent désormais la définition et la satisfaction de "situations problématiques" socialement contextualisées, de manière à pouvoir agir en tant que médiatrices d'une information transformatrice. Le point focal en est la production perpétuelle et collective de sens, et le cadre théorique est une correspondance utilitaro-économique entre le sujet et l'objet. Plus simplement, elles sont basées sur la théorie du choix rationnel."

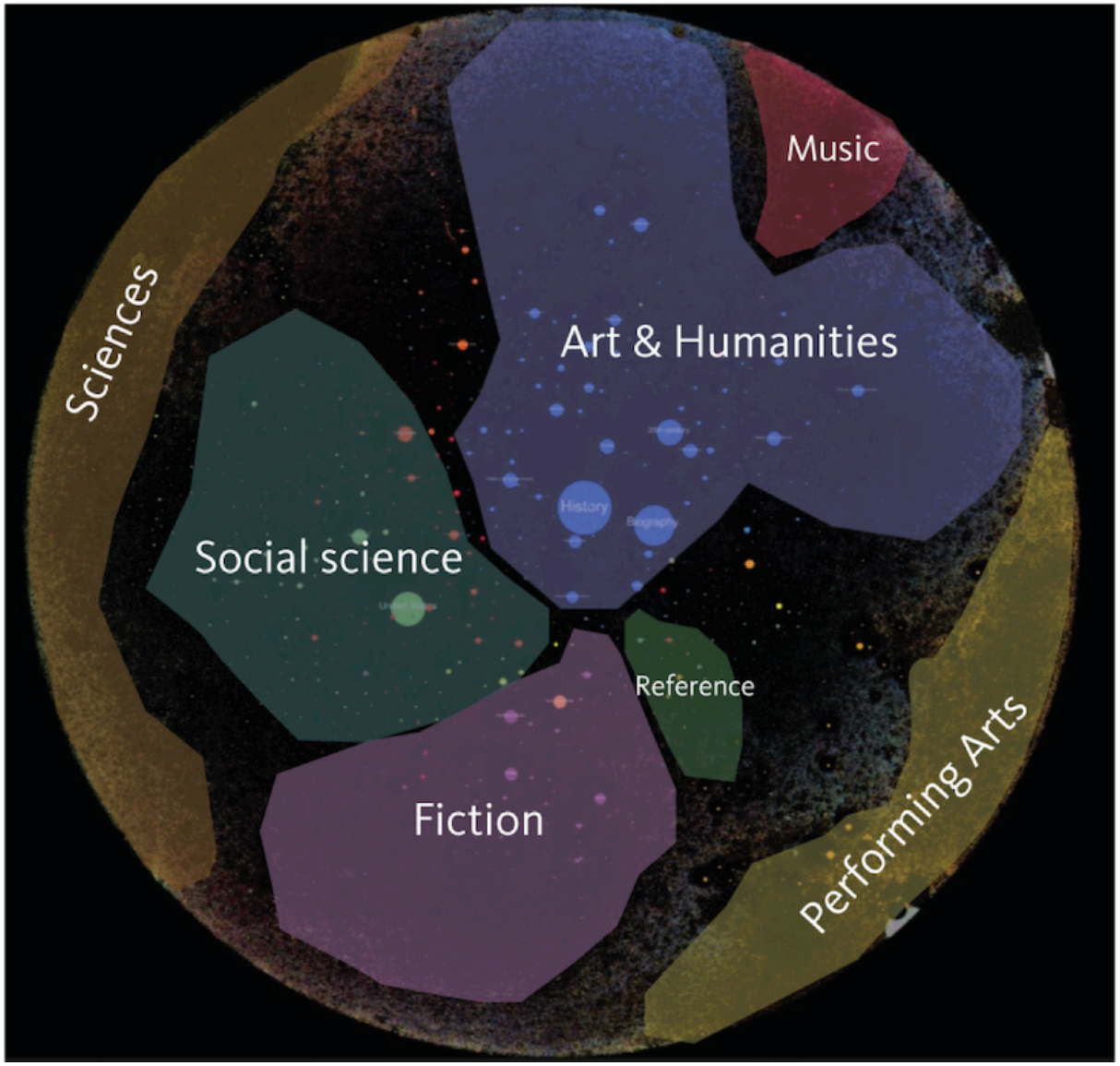
Or la logique utilitariste relève elle aussi du domaine de la représentation, et en particulier de ce domaine de la représentation où se "rencontrent la théorie de la monnaie et de la valeur, - science des signes qui autorisent l'échange et permettent d'établir des équivalences entre les besoins ou les désirs des hommes." Dans ce contexte, l'utilitarisme est ce qui, pour Michel Foucault, place la valeur d'un bien du côté de sa réception, c'est-à-dire de la mesure du besoin dont en ont les hommes lorsqu'ils sont placés en situation d'échange. Dans cet ordre d'idées, plus la nécessité d'un bien est grande, plus le bien a de la valeur par rapport à d'autres biens. C'est ainsi que l'algorithme KNN ne se préoccupe pas de classer les biens eux-mêmes mais de classer les usagers qui en ont parcouru la représentation sur le web. De là naît la recommandation, qui mesure la valeur du signe à sa place dans le parcours de consultation des usagers : ceux qui ont consulté *x* ont également consulté *y*.
"Les "utilitaristes" fondent sur l'articulation des échanges l'attribution aux choses d'une certaine valeur ; les Physiocrates expliquent par l'existence des richesses le découpage progressif des valeurs." Si l'on pose le principe fondamental de l'économie de l'attention comme étant le paradoxe d'un regard humain par définition limité, porté sur une information très abondante, alors on s'aperçoit que nous avons d'une part des acteurs, parmi lesquels Google, qui s'appuient davantage sur une analyse de l'attention et de la circulation de cette dernière d'un bien informationnel à son voisin, d'autre part, des sites importants tels Amazon et E-bay qui poussent leur analyse du côté de la profusion de ces mêmes biens informationnels et se proposent de gérer la quantité massive d'information en la synthétisant au travers d'un classement hiérarchique, c'est-à-dire, en définitive, par une présentation de cette dernière sous forme de facettes classificatoires. Or, c'est précisément de ce côté là que se situent, historiquement et dans leur principe fondamental, les catalogues des bibliothèques sur le web. C'est ainsi que la visualisation métaphorique que représentait le catalogue dans son principe de tableau des identités et des différences bibliographiques rejoint la visualisation graphique qui est au fondement de la logique représentationnelle du web : le catalogue sur le web est en réalité une représentation qui se représente, représentation d'une représentation. Assurément, donc, la visualisation du catalogue n'est pas chose nouvelle, elle est même au coeur de son projet descriptif initial.

Cartographier les collections de la bibliothèque

Nous avons défini la visualisation du point de vue plutôt large de la représentation Foucaldienne, une définition qui s'attache à considérer comme graphisme jusqu'à l'élaboration même du catalogue et de ses métadonnées. Comment situer, dans ce contexte, ce que l'on appelle couramment aujourd'hui "visualisation de données", diagrammes et graphiques que l'on peut générer de plus en plus facilement à partir d'outils de traitements de données tel Excel ou encore via des logiciels libres tels Gephy ? En d'autres termes, si l'interface graphique du catalogue est autant une représentation visuelle qu'une carte ou un graphe arborescent, qu'est-ce qui pourtant fait que nous ne pouvons les confondre dans notre langage courant ?
Peut-être pouvons-nous considérer ceci : alors que les interfaces graphiques cherchent à se rendre transparentes et à faire oublier leur nature fondamentale de système sémiologique, le fonctionnement d'un graphique repose quant à lui sur un temps d'appropriation des conventions qu'il utilise, il s'accompagne la plupart du temps si ce n'est d'une légende, a minima d'un texte introducteur en décrivant le contexte et les objectifs.
Un exemple d'un tel fonctionnement peut être trouvé dans une expérience conduite par le laboratoire de recherche de la Bibliothèque Publique de New York (NYPL Labs) et intitulée "Le Catalogue en Réseau" (The Networked Catalog). Sur le blog du NYPL Labs, Matthew Miller, conduisant au NYPL Labs la recherche sur les données liées et leurs applications, a publié un article décrivant la méthode mise en oeuvre pour élaborer une projection graphique en réseau d'environ 430 000 vedettes-matières utilisées pour indexer les collections de la NYPL. L'article explique ainsi que l'intensité des connexions entre les sujets a été calculée en fonction du nombre de fois où les termes apparaissent ensemble dans l'indexation d'un ouvrage. Il en résulte une sorte de planisphère où sont rassemblées des pastilles de couleurs et de tailles variées figurant les termes et leur proximité sémantique. En guise de légende, l'article nous apprend que les couleurs ont été attribuées en fonction de regroupements automatiques des termes en familles, et que la taille des pastilles dépend du nombre de fois où la vedette matière qu'elles désignent a été employée pour indexer un ouvrage.
D'une certaine manière, ce que nous appelons couramment "visualisation de données" désigne donc une sorte de cartographie, c'est-à-dire une projection à caractère métaphorique qui "transpose différents types de données quantitatives sous formes spatiales et graphiques."

> "En tant que métaphores, les visualisations font correspondre des domaines sources (jeux de données) à des domaines cibles (structures graphiques)."

Ainsi, de même que le principe d'une carte est de transposer l'espace réel sous forme graphique en attribuant à des coordonnées géographiques des coordonnées géométriques fondées sur un repère orthonormé, de même, la visualisation proposée par la Bibliothèque Publique de New York se propose de retranscrire l'intensité d'une connexion entre deux sujets d'indexation en terme de distance graphique. À cet égard, il est significatif d'observer que le résultat final de la visualisation ressemble dans son ensemble à une mappemonde sur laquelle chacune des grandes familles thématiques identifiées automatiquement évoque des sortes de continents de constellations sémantiques.



Légende à compléter.

Dans l'esprit de ses concepteurs, cette visualisation du catalogue doit pouvoir jouer le rôle d'un bouton "voir tout", permettant de donner une vue d'ensemble non seulement des collections mais également de ce que connaît du monde une bibliothèque. Projection métaphorique, la visualisation est donc également une projection panoramique d'un paysage documentaire : "de la même manière que l'allure rapide de la locomotive de passagers offrait de nouvelles perspectives balayant de larges étendues géographiques, les visualisations (...) tentent de révéler des liens significatifs (...) à des échelles auparavant inconcevables".

Donnant à voir de vastes ensembles de données et révélant ainsi sous un angle nouveau des phénomènes jusque là invisibles, la visualisation serait donc un moyen heuristique qui participerait à la production de connaissances. Reste à savoir quel genre de connaissances, et selon quelles conditions ces dernières peuvent être générées.

Il est à cet égard significatif d'observer que dans le texte accompagnant la visualisation en réseau de la NYPL, l'auteur évoque le désir, pluri-centenaire selon lui, de cartographier les connaissances : visualiser les données d'indexation du catalogue ainsi que l'évolution dans le temps de leurs multiples interconnexions permettrait de dessiner sous nos yeux une histoire des idées au prisme de la politique documentaire de la Bibliothèque Publique de New York, à la manière d'un miroir limpide dans lequel viendrait se refléter en quelque sorte le réel. Mais cela serait peut-être oublier le caractère avant tout culturel des données : entre le réel et sa représentation se glissent les erreurs de catalogage, la personnalité des catalogueurs, les présupposés du modèle de données, les évolutions technologiques qui ont affecté supports et

environnements de production, les politiques institutionnelles des établissements producteurs et plus largement encore, le contexte historique qui a présidé à la création des données. Avant de décrire les objets auxquels elles sont associées, les métadonnées renvoient donc inévitablement à leur créateur, et le seul réel qu'elles peuvent valablement révéler est par conséquent celui qui a précédé et organisé leur apparition.
C'est ainsi que dans un essai en ligne intitulé *The Life and Death of Data*, Yanni Loukissas et Krystelle Denis ont produit une vue chronologique des données enregistrant les dates d'entrée des arbres et plantes conservés dans l'arboretum d'Harvard (Arnold Arboretum). Le résultat, loin d'être un tableau exact des collections de l'arboretum, ne prend son sens qu'au regard de l'essai interprétatif qui l'accompagne :

> "La visualisation peut être faite pour souligner la diversité des données enregistrées par le personnel de l'Arboretum au fil des années. Pour commencer, elle peut refléter des variations dans la manière avec laquelle l'Arboretum a créé les numéros d'accession, un aspect à première vue banal de la façon selon laquelle les plantes sont placées dans la collection. Et pourtant, on peut facilement observer comment la formule utilisée pour produire les numéros d'accession a changé plusieurs fois en réponse à la complexité de plus en plus grande de la collection ainsi qu'à des limites techniques importantes. Par exemple, les dates d'acquisition sont devenues un élément de systèmes de numérotation plus récents dans le but d'établir des distinctions parmi des spécimens similaires collectés tout au long de l'histoire de l'Arboretum. (...) Même des changements mineurs peuvent affecter de manière significative la manière dont les collections gèrent les données."

Il faudrait alors utiliser la visualisation des métadonnées non comme un reflet limpide du savoir d'une époque et des collections d'une bibliothèque mais comme trace de l'activité et de la vie d'une établissement culturel qui se donne à lire et interpréter.
À travers les deux exemples de visualisation proposés par le NYPL labs d'une part et à partir de l'Arnold Arboretum d'autre part, nous avons pu observer trois manières par lesquelles la visualisation, selon la chercheuse Johanna Drucker, permet de "connaître" la bibliothèque. Un premier moyen est "l'analogie visuelle ou ressemblance morphologique" : la métaphore cartographique utilisée pour représenter le savoir de la NYPL permet d'ouvrir les collections de la bibliothèque à leur exploration virtuelle. Deuxièmement, la visualisation est également le moyen de fournir "une image" d'un phénomène non perceptible directement : en l'occurrence, la frise chronologique développée par Yanni Loukissas et Krystelle Denis fournit un négatif fragmentaire de la vie de l'Arnold Arboretum depuis sa fondation.
En dernier lieu, la visualisation désigne également des conventions visuelles qui peuvent être utilisées pour structurer des opérations et procédures. C'est ainsi que la représentation cartographique des collections signalées par les données du catalogue, participe à la représentation mentale que se fait l'usager de sa bibliothèque et peut en devenir en quelque sorte le symbole, de même que les frontières hexagonales de la France en sont venues à représenter la nation française en tant que signe participant à la fois à l'objet qu'il désigne et pouvant se substituer à lui. En tant que représentation symbolique s'appuyant sur des signes conventionnels, la visualisation structure donc les interactions des usagers avec la bibliothèque.

Ouvrir le catalogue

Dans son essai sur les métadonnées de l'Arboretum d'Harvard, Yanni Loukissas s'interroge sur la pertinence du mouvement vers l'ouverture des données, notamment lorsque ce dernier concerne les métadonnées de collections culturelles :

> "Les chargés de collection se concertent pour savoir comment construire des interfaces de programmation applicatives, qui permettront à des publics extérieurs d'interagir directement avec leurs données. Des discussions accompagnent ce changement, se demandant dans quelle mesure les données peuvent être libres : libres d'erreurs, libres de droits de propriété intellectuelle, libres de frontières institutionnelles, libres de leurs origines disciplinaires, libres de valeurs et présupposés locaux, et même libres de redevances. Dans le mouvement d'ouverture des données, nous devons nous demander quel genre d'accès elles fournissent, pour qui et dans quel but."

Peut-être pouvons-nous reformuler la remarque de l'auteur par cette question : dans quelle mesure les métadonnées peuvent-elles être considérées comme des collections à part entière que l'on mettrait à disposition du public au même titre que des archives, et notamment sans que ce mouvement d'ouverture n'affecte les pratiques actuelles de production des données, dans la mesure où "rendre de telles informations publiques changera inévitablement ce que l'on considère comme de "bonnes" données, la manière avec laquelle elles sont produites, et même, la façon dont le personnel les utilise"? En d'autres termes, le regard et l'attention nouvelle accordés à un objet issu d'une activité humaine, loin d'être simplement passif, peut le modifier en profondeur. De fait, dans cette dynamique entre le regard public et les données diffusées, se pose de nouveau la question de la visualisation.

Il nous semble en effet que, concernant l'ouverture des données, il paraît quelque peu rapide de juger comme séparées et étanches l'utilisation des métadonnées pour accéder à des collections, d'une part, et la réutilisation de ces mêmes données au sein de projets recherche ou de développement : le public des utilisateurs et celui des réutilisateurs est en effet lié par un même besoin d'exploration et de découverte de jeux de données, en dépit des différences d'objectif qui peuvent les caractériser. OpenStreetMap, par exemple, est une application de cartographie qui permet à la fois d'effectuer un repérage topographique et d'exporter les données décrivant le lieu recherché par une sélection qui peut être faite manuellement à partir de la carte en ligne. Si donc l'on considère que l'ouverture des données ne consiste pas simplement en la mise à disposition en ligne d'informations qu'il sera difficile d'obtenir sans un bagage technologique adéquat, mais également en une médiation de ces mêmes données impliquant leur obtention via une interface graphique, alors il semble que de nouveau la représentation imagée soit un élément incontournable de la mise à disposition des métadonnées culturelles auprès du public.

Reprenant la remarque initiale de Yanni Loukissas, nous pouvons donc nous demander comment, dans le contexte présent d'ouverture, la visualisation peut affecter la production même des données du catalogue.

En premier lieu, la présentation des collections sous forme graphique peut être utilisée à des fins de correction d'erreurs dans les données, ou plutôt de ce que Matthew Battles préfère appeler "artéfact de données", à savoir des "motifs structurels, conçus à dessein ou non, qui témoignent de l'histoire sociale et matérielle de leur accumulation", c'est-à-dire de la constitution progressive de collections de métadonnées. Ainsi l'Observatoire de la Bibliothèque (Library Observatory), une application permettant de visualiser hiérarchiquement les métadonnées de la Digital Public Library of America (DPLA), a-t-il mis à disposition de ses utilisateurs un tumblr sur lequel ils sont invités à publier des captures d'écran indiquant des anomalies qu'ils ont repérées visuellement dans les données. De fait, certains bibliothécaires ayant expérimenté l'application ont exprimé leur intérêt pour l'utilisation de la visualisation dans le but d'améliorer la qualité des données.

En second lieu, la représentation cartographique des métadonnées peut servir de support à une action direct du public sur les données du catalogues. La plateforme en ligne MetadataGames développée par le Tiltfactor Labs de l'université de Dartsmouth propose par exemple un dispositif d'indexation collaborative d'objets numérisés qui s'appuie sur un jeu sur navigateur Web. Le principe fondamental de ces "Jeux de Métadonnées" rejoint donc celui des "data games", jeux dont le contenu se fonde sur des données du monde réel, et qui ont en effet souvent pour caractéristique de proposer une forme d'exploration et d'apprentissage à partir de ces données. À la manière des jeux vidéos, ces data games peuvent être dotés d'une simulation virtuelle sur laquelle l'exploration s'appuie.

On peut trouver un tel exemple de MetadataGames à la British Library : "Ship Tag est un jeu en solitaire sur navigateur Web qui permet de "tagger" une importante collection d'images à thématique nautique", de même pour Book Tag, qui permet de tagger un grand nombre de livres numérisés. L'interface permet simplement de visualiser la première de couverture d'un exemplaire numérisé et d'en proposer une indexation, qui sera virtuellement récompensée par un décompte de points. Si en l'occurrence le jeu ne s'appuie pas sur une représentation graphique des métadonnées, rien n'interdit d'imaginer un jeu qui s'appuierait sur une exploration interactive d'une projection cartographique des données et récompenserait la

distance parcouru dans l'espace virtuel de la collection ainsi que le marquage des objets informationnels rencontrés.
En dernier lieu, le public est amené à influer indirectement sur la production de données au travers des systèmes d'acquisition à la demande, par le biais desquels ils peuvent obtenir l'acquisition d'ouvrages, et avec ces derniers, l'acquisition de données descriptives qui pourront figurer dans le catalogue au côté des métadonnées des ouvrages déjà présents dans l'établissement. Si ce type de systèmes est appelé à se développer à l'avenir, ils seront donc une autre source d'alimentation des catalogues : le choix du public conditionnerait alors l'apparition de tel ouvrage, tandis que l'éditeur, qui pourra également être à l'origine de la base de données dans laquelle l'usager aura sélectionné sa ressource, sera responsable du format et de la qualité des données accompagnant l'item. Car les acquisitions à la demande ne peuvent être envisagées sans un outil de recherche préalable, doté d'une interface et d'un moteur de recherche : la subjectivité du professionnel qui était autrefois pleinement responsable des acquisitions de son établissement est alors déplacée dans les spécifications des algorithmes et de la présentation visuelle des données qui ont orienté le choix des lecteurs : c'est donc de nouveau par le biais d'une représentation de représentation bibliographique que le lecteur peut être amené à intervenir sur les données du catalogue. Autant dire qu'il serait préférable, de notre point de vue, que ces systèmes de découvertes de contenu ne soient pas entièrement laissés à des intérêts privés. Non pas que la subjectivité des professionnels des bibliothèques serait a priori meilleure, mais plutôt que l'élaboration du catalogue doit être elle-même ouverte à l'espace public et que le bibliothécaire doit pouvoir faire la synthèse entre ce qu'il perçoit des usages de recherche de son public et sa propre expertise en matière de recherche documentaire, avant d'exprimer le résultat de cette réflexion du point de vue des technologies représentationnelles d'aujourd'hui et de demain.

Conclusion

L'influence croissante du web sur les pratiques de recherche de l'information propulse les catalogues des bibliothèques sur le devant de la scène numérique : ils sont pour les internautes la première image de la bibliothèque, ils en viennent par leur architecture même à représenter symboliquement l'ordre du savoir imaginé au fil de l'histoire par les professionnels qui ont présidé à sa conception. Du point de vue du contexte épistémologique de la recherche documentaire, on ne saurait séparer les données du catalogue de leur mise en scène visuelle, que cette dernière s'exprime au travers de la mise en page de volumes physiques reliés ou bien sur l'écran d'un ordinateur : sur le web, les instruments de recherche bibliographiques sont avant toute chose représentation d'une représentation.
Le catalogue est donc un média et, en tant que tel, est nécessairement un objet politique qui, dans un contexte démocratique, doit pouvoir être questionné autant qu'utilisé : cela implique que de la représentation sémiologique, nous passions à la représentation symbolique, le symbole, seul, étant capable de pointer vers la nature politique de l'information bibliographique. Les métadonnées ne vont pas d'elles-mêmes : privilégier une illusoire transparence des données vers les contenus qu'elles signalent au prétexte de l'efficacité de la recherche serait de notre point de vue un contresens. Les données ne sont en effet ni atemporelles ni universelles : tout comme les archives, elles sont un produit de l'activité humaine, ancré dans un contexte local et social, politique et financier, toutes choses qu'une projection cartographique commentée des données peut donner à lire à qui veut et sait regarder.
Une autre erreur serait à nos yeux de séparer conceptuellement l'utilisation des données à des fins de recherche documentaire de leur réutilisation dans le cadre de projets de développement ou de recherche scientifique : autant par leurs usages que par leurs réutilisations, les données doivent être perçues comme un contenu documentaire à part entière qu'il faudrait pouvoir épaissir et matérialiser.
Paradoxalement, ouvrir le catalogue en ligne signifie donc rendre son langage moins immédiat et plus opaque, tout en sachant que la représentation virtuelle, métaphorique et interactive du catalogue est aussi l'élément qui permettra de le rendre ludique : lorsqu'il est bien conçu, le

jeu peut être le lieu où le joueur se contraint volontiers à l'apprentissage de règles et conventions, il peut être à la fois un instrument de séduction et de distance réflexive entre un sujet actif et cet objet lui-même issu d'une autre subjectivité humaine qu'est le catalogue.