

ATELIERS DATAVIZZ

Liste des projets

Etablissement porteur du projet	Descriptif rapide	Données disponibles (typologie / format)	Résultat attendu	Objectif final	Historique
DBU Sorbonne Nouvelle - Paris 3	Utiliser la dataviz pour mieux exploiter les données d'usages de notre portail documentaire : http://www.dbu.univ-paris3.fr	Données issues de Google Analytics Données issues du serveur web (logs)	Produit web avec mise à jour régulière (quotidienne/hebdomadaire)	Indicateur de pilotage et aide à la décision lors des refontes (nouvelles fonctionnalités, toilettage graphique) Données d'utilisation essentielles pour le prochain projet web (2019)	Aucune tentative précédente Peu d'utilisation des données pour le moment
ENS Paris-Saclay	Exploitation des données lecteurs / prêts issues de notre SIGB (Koha)	BD SQL (Koha)	Alimentation automatique (périodicité à définir) au fur et à mesure des inscriptions et interface web de visualisation avec possibilité d'export tableur	Reporting et présentation aux instances avec objectif final d'automatiser tout le rapport d'activité	Rapports SQL dans Koha mais pas automatisés et qui n'alimentent rien
Université Le Havre-Normandie	Exploiter les données sur les collections afin d'avoir un outil visuel et partagé de pilotage de la politique documentaire de l'établissement.	Données issues du SIGB Horizon Base SQL adossée (BiblioGQL) Fichiers excel récapitulatifs sur les collections de livres papier (budget, acquisitions, désherbage etc) et leurs usages (prêts, âge, rotation, disponibilité, etc.)	Application web interne partagée entre les acquéreurs Mise à jour mensuelle	Objectif de pilotage pour la gestion des acquisitions et des collections par les acquéreurs et par le responsable de la politique documentaire	Aucune tentative précédente

Université Jean Moulin Lyon 3	Exploiter les données prêt / niveau d'étude du SIGB Koha pour valoriser les documents qui sortent (incitation pour les autres étudiants de même niveau et discipline) et pour favoriser les interactions avec les acquéreurs	Fichiers CSV/excel (données extraites du SIGB Koha)	Produit visant à enrichir : - le site web (pages collections disciplinaires) - le catalogue Koha (et/ou l'outil de découverte) - l'intranet des étudiants (informations poussées en fonction des disciplines)	Objectif principal : favoriser la réussite des étudiants de licence 1) Renforcer le travail des acquéreurs en favorisant les interactions avec les étudiants 2) Valoriser les fonds pour les étudiants de licence en améliorant leur visibilité et leur adéquation avec les besoins des usagers	Aucune tentative précédente
Université Jean Monnet - Saint-Etienne	Mise en valeur des données d'activité du SCD (dans un cadre autre ou pour compléter l'annuel rapport d'activité)	Toutes les données relatives au SIGB sont dans une base dédiée que l'on requête via MySQL (phpMyAdmin) Données autres : sous fichiers Excel	Application web externe Mise à jour annuelle Visualisation de l'activité du SCD sur plusieurs années	Utiliser la datavisualisation à des fins de communication institutionnelle, pour "glamouriser", rendre plus attractives et compréhensibles les données d'activité du SCD, auprès des politiques et des lecteurs	Aucune tentative précédente pour ce qui est de la datavisualisation, mais une nouvelle approche de la représentation des données via la rédaction d'un " Digest " original du rapport d'activité
Bibliothèque SciencesPo Paris	Mise en valeur des données d'interrogation de l'outil de découverte (environ 600 000 mots annuels)	Extrait des requêtes sur l'outil de découverte	Présentation chronologique et /ou thématique des sujets et des ouvrages et des auteurs recherchés	Mieux comprendre les intérêts des étudiants ainsi que la façon dont ils interrogent le catalogue	Aucune exploitation de ces données dans le passé
Université de Bourgogne	Mettre en valeur les données d'activité du SCD et les principaux indicateurs en matière de politique documentaire	Données extraites du SIGB AbsysNet interrogé via BO Données rassemblées sous forme de fichiers Excel	Application web avec mises à jour annuelles	Mieux communiquer sur l'activité du SCD en direction des publics. Valoriser l'action du service auprès de la gouvernance de l'université	Pas d'expérience antérieure

SCD Aix-Marseille Université	Exploitation des données lecteurs : inscriptions, prêts A titre indicatif : - Inscrits 2016 : 40 580 - Prêts 2016 : 396 492	Données issues SIGB Koha	Affichage web pour la visualisation des données avec : > mise à jour automatique (périodicité à définir) > conservation de l'historique des affichages antérieurs	Pilotage et communication externe : suivre et rendre visible l'activité de prêts/inscriptions dans son évolution	Présentation des rapports d'activité SCD 2014 et 2015 en utilisant Piktochart pour la datavisualisation (optique communication). Mais aucune tentative pour l'automatisation des données.
SCD de l'Université Paris Descartes (Paris V)	Avoir un outil moderne qui exploite pleinement les indexations des notices bibliographiques	Notre corpus de notices bibliographiques	Une navigation (web) graphique dans les thésaurus, rameau par exemple (mais ça pourrait être fmesh) en fonction des données de notre catalogue : nœuds liés par de nombreuses occurrences communes, zoom avant/arrière pour atteindre les nœuds parents/enfants, affichage des notices correspondant à la descendance du nœud sélectionné (et non pas que le nœud comme sur gallica), auteurs phares pour cette descendance etc...	Ce serait pour faire une interface publique pour naviguer plus intuitivement dans le catalogue	Pas d'historique
SCD de l'Université Rennes 2	Avec la mise en œuvre des projets EZparse/EZmesure et AGIMUS des progrès importants ont été consentis pour tendre vers une meilleure analyse et compréhension des usages des ressources et des services numériques. Cependant, nous continuons (et c'est sûrement le cas d'autres établissements) à rencontrer des difficultés pour professionnaliser la collecte d'indicateurs plus traditionnels de l'activité de nos bibliothèques. Notre volonté serait donc de capitaliser sur l'expérience des projets sus-cités (et éventuellement de se rapprocher	Données provenant du SIGB open source KOHA. Plus précisément, il s'agira d'extractions SQL issues de plusieurs tables (statistics, borrowers, biblio, biblioitems, items). Il est possible de fournir les données dans plusieurs formats si nécessaire : sql, csv, json Ce corpus de données permettra d'obtenir : - des données caractérisant les documents/exemplaires empruntés : identifiant notice, titre, auteur, éditeur, date de publication, collection, identifiants (ISBN,	La solution pourrait être constituée de plusieurs parties : - un module de traitement des données, servant à la collecte, à la transformation et à l'enrichissement des données de circulation. Dans l'idéal, la collecte pourrait se faire quotidiennement/hebdomadairement de manière (semi)-automatique ; - un moteur d'indexation (elasticsearch, influxDB, autre ?) servant au stockage des données enrichies ; - une application de rendu visuel (kibana, etc.) qui permettra un affichage dynamique des indicateurs par requête temps réel ou encore la création de tableaux de bord ; - des services web qui permettront de réutiliser, publier facilement les visualisations/extractions générées dans différents contextes, sur différentes plateformes, dans différents formats.	En interne (pour le SCD) : - aider à l'évaluation et au pilotage des acquisitions - connaître et comprendre l'usage qui est fait de la documentation imprimée - tendre vers une meilleure adéquation de l'offre et des besoins - prévisualiser l'activité de la bibliothèque et rendre compte des rythmes universitaires - connaître les pratiques informationnelles, comprendre les stratégies documentaires de recherche des publics, les typologies et les communautés de lecteurs - disposer d'indicateurs d'usage pour la pédagogie : impact des formations documentaires, etc. - identifier les mutations et les tendances en œuvre dans les usages Pour la communauté universitaire :	Fonctionnement actuel basé sur le module "Rapports" de Koha. Plusieurs requêtes SQL pré-paramétrées. Quelques tableaux de bord réalisés sous des logiciels de type tableur (excel, libreoffice calc).

	d'autres comme prevu.fr en mettant cette fois-ci, le focus sur la question de l'usage des ressources imprimées en bibliothèque universitaire. Partir de l'exploitation des données d'usage issues des SIGB locaux, en s'appuyant en particulier sur les données de circulation (enrichies de données bibliographiques ou encore de données caractérisant les utilisateurs) pour construire une solution permettant une remontée automatique des indicateurs de l'utilisation de la documentation imprimée.	ISSN), PPN, type de document, identifiant exemplaire, code-barres, cote, type de prêt, site, localisation, statut, date de création, niveau cible - des données caractérisant les emprunteurs : catégorie de lecteur, année inscription, statut, établissement, niveau, code étape, diplôme, discipline, affectation, campus - des données caractérisant les transactions réalisées (données d'exemplaire) : site, date et heure, type de transaction Volumétrie : 2 400 000 lignes Rétrospectif : 2013 à aujourd'hui		- faire connaître l'activité de la bibliothèque par la valorisation de ses collections imprimées = meilleure (re)connaissance du SCD par le public (usager ou non) ainsi que par la tutelle - imaginer des actions de valorisation/sensibilisation physiques (matérialisées dans l'espace) en utilisant la data-visualisation : visualisation des données sous la forme d'expositions par exemple (en lien avec certains enseignements réalisés à l'université (partenariats à construire avec le Master Arts - Spécialité Création et Management Multimédia ou encore le Master ergonomie) - constituer d'un socle favorable à la conception/construction de nouveaux services à valeur ajoutée : * valoriser, recommander les collections par l'usage qui en est fait * dispositif de médiation numérique et de recherche dans le catalogue et/ou le site web de la bibliothèque * favoriser de nouvelles pratiques de recherche documentaire et de nouveaux usages scientifiques par le biais de visualisations des données Si le projet est porté au niveau de plusieurs établissements : L'ambition d'un tel projet pourrait être de ne pas forcément se limiter à des expériences locales ou parfois isolées. En effet, agréger l'ensemble de ces données dans un entrepôt unifié facilement exploitable (comme le propose Ezmesure pour la documentation électronique) permettrait plusieurs choses : - disposer d'un outil libre générique qui collecte à la fois localement et nationalement les données d'usage et les analyse sans dépendre de la technique ou des modules/outils fournies par les prestataires desdits SIGB (= sortir les SCD de l'artisanat dans la collecte des indicateurs) - favoriser la mutualisation et le partage d'expérience entre établissements : réflexion sur la définition des données d'usage attendues, des indicateurs, des scénarios de visualisation, sur les questions d'informatique et libertés - proposer une interface en ligne de visualisation dynamique, consolidée et comparative des données entre les établissements ce qui pourrait offrir une alternative potentielle à ASIBU (par exemple).	
--	--	--	--	--	--

SciencesPo Paris	Donner aux étudiants de Sciences Po (et par extension à tout étudiant dans les mêmes disciplines) selon leurs cursus et leurs niveaux les documents qui les concernent selon les critères documentaires et d'audience par exemple. Visualisation de l'environnement documentaire par cursus : - Etape 1 : la visualisation de tous les ouvrages (ou d'une partie seulement, si la volumétrie est trop importante) et les e-plateformes consultées dans l'année précédente par profil d'étudiants (ex. Les masters Affaires Publiques ; les étudiants en 1A) ... en incluant la langue de prédilection de l'étudiant. - Etape 2 : La possibilité de représenter sur une carte la géolocalisation de ces ouvrages (tel livre traite du Venezuela, tel autre de la Tunisie) ... et de faire varier la sélection en sélectionnant les dates traitées dans les ouvrages en faisant varier un curseur sur une timeline.	Transactions de prêts sur 6 mois incluant ID étudiant (et données descriptives programme, école, niveau, langue, pays d'origine de l'étudiant, plateforme accédées) , et données bibliographiques : (titre, auteur, langue, période et zone géographique couvertes, mot-clés) (csv ou xls) Consultation de ressources numériques (en ligne / téléchargement) sur 6 mois incluant ID étudiant (idem ci-dessus) et spécification des ressources selon descriptif ezpaarse. (csv ou xls)	Visualisation web externe à un instant T : Visualisation des ressources proposées avec filtre sur le cursus, le niveau, le type de ressources (numérique ou papier), la période couverte, la langue de l'ouvrage et le pays d'origine de l'étudiant. Concrètement, un étudiant s'identifie, en fonction de son cursus, on lui propose des ressources imprimées et électroniques qui correspondent à ce que les étudiants du même cursus ont consulté/emprunté dans les 6 derniers mois, soit sous forme de liste, soit sous forme d'une carte (domaine géographique traité) + time line (période historique traitée). Cette dernière proposition n'est possible que pour les imprimés indexés et pas pour les e-ressources.	Aider les étudiants à mieux cibler leurs lectures, les inciter à lire davantage, découvrir les livres et auteurs de référence, approfondir leurs recherches. Valoriser les ressources électroniques Favoriser la serendipité	Pas d'exploitation de ces données jusqu'à présent, sauf pour les données ezpaarse (exploitation en interne ne donnant pas lieu à des propositions sur le site de la bibliothèque)
--------------------------------	---	---	---	---	--

Université de Limoges	Notre projet est d'utiliser les métriques fournies par l'outil d'analyse PIWIK sur notre catalogue KOHA des bibliothèques (catalogue.unilim.fr) pour créer des outils de reporting stables tant pour les instances universitaires (rapport d'activité, conseils de documentation par composantes) que pour les étudiants à des fins d'animation de séances de formation à la recherche documentaire. L'idée est de partir de ce que cherchent les étudiants et de comment ils le cherchent afin de pouvoir établir à ce propos des séquences pédagogiques quelque peu orientées par la « data visualisation ». L'objectif est d'intégrer des iframe sur un ou plusieurs sites web ou sur une plateforme moodle ou bien de prévoir la génération de rapports graphiques comme réponse à une requête de recherche documentaire sur le catalogue.	Application PIWIK. Toutes les données mentionnées ci-dessous sont accessibles aux formats : CSV, TSV, XML, JSON, Php. Toutes les données mentionnées sont disponibles depuis le 1/01/2016 (le but étant de construire pour la fin de cette année des visualisations comparées de ces deux années pleines : 2016 et 2017). * Visites * Durée des visites * Actions par visites * Pages vues * Nombre de recherches * Nombre de recherches uniques * Périphériques d'entrée (analyse des matériels, systèmes d'exploitation et navigateurs) * Liens sortants (vers applications tierces de l'université) * Informations sur les logs : horaires, provenances géographiques, sites de provenances (sites tiers de l'université, partenaires institutionnels comme l'ABES).	Côté institutionnel. La DSI de l'université travaille au développement de l'application KIBANA qui pourrait fonctionner rapidement pour certains outils. Notre objectif en déposant pour l'AAP bibliovizz pourrait être de pouvoir bénéficier d'une expertise précise afin d'implémenter un outil de ce genre et donc de générer des mises à jour en temps réel. Pour l'instant, des intégrations de widgets issus de l'application Piwik sont en cours mais offrent des possibilités de visualisation limitées. Les intégrations et/ ou visualisations professionnelles concerneraient : * Un site internet public (wordpress) * Un intranet professionnel (wordpress également) Côté public * Les visualisations recherchées pourraient être « fixes » (rétrospectif sur des années pleines) et viendraient alimenter une plaquette pédagogique commune au SCD. * Une version en ligne de ces données pourrait être hébergée sur une plateforme moodle, déjà existante, et permettrait de générer des séquences pédagogiques (questions/ réponses sur les logiques de recherche, d'exploitation du catalogue ...). Grosso modo, prévoir des résultats graphiques types en fonction de certaines requêtes de recherche simple sur le catalogue (quel contenu par site, par type de documents ...etc.).	L'objectif final est d'avoir une assise statistique et graphique permettant de dire « comment les étudiants et enseignants cheminent dans leur parcours de recherche documentaire ? ». Il s'agit donc plus d'une visée de communication (pédagogique peut être) plus que de pilotage administratif. Un outil de datavisualisation adaptée permettrait de pouvoir faire de la pédagogie inversée et / ou par l'exemple de façon assez efficace.	Les données ne sont exploitées que : * Au format image brute à partir des graphiques générées par l'outil PIWIK lui-même (pour Pwpt internes, rapports d'activité ... etc.) * En utilisant des widgets assez sommaires proposés par l'application PIWIK.
-------------------------------------	--	--	---	---	--

DDE de l'Université de Lorraine	J'aimerais établir une cartographie des documents empruntés par tranches de cotes ou grands domaines et par filières ou types de lecteurs	2 fichiers Excel : > 1 avec la liste des documents empruntés et leurs cotes avec les 5 dernières dates d'emprunt, les 5 dernières dates de retour et les 5 codes barres des lecteurs ayant empruntés > 1 avec la liste des codes barres lecteurs et leurs filières ou leurs catégories de lecteurs	2 tableaux annuels sur web interne avec 1 tableau des filières de lecteurs et les tranches de cotes empruntés avec leurs périodes et 1 ou plusieurs tableaux par bibliothèques avec les tranches de cotes et les filières emprunteuses avec leurs périodes	Cartographie de ce qui est emprunté pour les acquéreurs et les responsables de pôles et plus largement pour le reste du personnel	Pour l'instant à l'état de réflexion (Macro sur Excel ?)
Université Paris-Est Marne-la-Vallée (UPEM)	Représenter les co-publications présentes dans HAL entre les structures de la recherche du (nouveau) I-SITE Future — rassemblant l'EIVP, l'ENSAVT, l'ENSG, l'ESIEE Paris, l'Ifsttar et enfin l'UPEM.	Données bibliographiques/affiliations des publications scientifiques (articles, actes, ouvrages...) référencées dans HAL, sur les 5 dernières années. Format au choix JSON, XML, CSV	Un rendu interactif web externe serait souhaitable, et idéalement un affichage en temps réel (via l'API de HAL).	Rendre compte des connexions déjà présentes entre les nombreuses structures de la recherche de cette nouvelle entité qu'est l'I-SITE Future. Placer HAL, et par la même l'Open Access, comme outil central pour référencer et diffuser la production scientifique écrite.	Deux rapports bibliométriques réalisés (avec HAL, Scopus et le Wos) à l'échelle de l'UPEM

Bulac	Projet 1 : Visualisation des statistiques de la fréquentation des salles de lecture (étude fine) Projet 2 : Pilotage interne et communication au sujet des acquisitions d'imprimés (suivi mensuel).	Projet 1: Tableur csv issu du système de contrôle d'accès sur une semaine (Heure d'entrée / Catégorie de lecteur / Niveau d'étude / Etablissement / Spécialité étudiée) Projet 2: Tableur csv issus d'une requête sur la base de données SQL du SIGB Koha (Exemplaires Commandés/ Exemplaires Reçus / Coût/ Ligne Budgétaire / Fournisseur / Type de doc/ Catégorie de prêt / Indice Thématique BULAC / Indice Géographique BULAC / Domaine BULAC/ Langue / Mois / Localisation)	Projet 1 : Suivi journalier du volume et de la composition du public fréquentant (évolution horaire) Projet 2: Tableau de bord de suivi des acquisitions et de la politique documentaire avec possibilité de croiser l'ensemble des indicateurs collectés	Projet 1 : Communication à destination des lecteurs en vue d'améliorer l'information sur la saturation des espaces / communication institutionnelle sur la composition du public fréquentant la bibliothèque Projet 2 : Pilotage interne (pôle Flux et données / pôle Développement des collections) / communication (lecteurs/institutions/gis...) sur la composition et le volume des acquisitions imprimées et audiovisuelles)	Projet 1: Rapport csv exploité à la main et a minima pour le rapport d'activité, pas de communication à destination des lecteurs Projet 2 : Tableur csv traité manuellement pour le rapport d'activité (suivi annuel uniquement et pas de croisement fin des données pour le travail courant)
Bibliothèque nationale de France	Depuis 2011 existe l'observatoire du dépôt légal, panorama sur l'année n-1 de l'édition de tous les documents rentrés par dépôt légal à la BnF. Fondé sur les statistiques d'entrées par typologie de documents et sur les notices bibliographiques, il montre les évolutions sur la forme (et un peu sur les contenus) de tout ce qui est diffusé au public en France. Il est destiné au grand public et est repris par différents médias (professionnels principalement).	Tableaux excel décrivant les 80 000 livres, 260 000 fascicules, 2 000 vinyles, 2 milliards d'url de sites internet, etc. rentrant par DL (avec une colonne par métadonnée). Ces tableaux existent pour les données de 2010 à 2016 inclus.	Un changement de forme est attendu, a minima une visualisation graphique/ en images (et non pas une synthèse textuelle avec quelques tableaux), qui peut prendre différentes formes : application web / intégration à des outils en ligne (pas de possibilité d'intégrer au SI maison de la BnF), etc.	L'objectif final souhaité est de renouveler la communication autour de ces données, aussi bien en termes de forme que de fond : 1/ sortir du traditionnel PDF qui ressemble à un rapport annuel pour mettre en valeur et prioriser les statistiques et résultats ; 2/ permettre de faire émerger d'autres éléments plus qualitatifs (notamment sur les contenus des documents) si c'est possible pour de futures exploitations de ces données.	Le dernier observatoire du DL est ici : http://www.bnf.fr/fr/professionnels/depot_legal_definition/s.depot_legal_observatoire.html?first_Art=non

Université des Antilles	Exploiter les données sur les collections : composition, évolution et usage des collections imprimées. Usage des collections électroniques.	Données sur les collections et leur usage extraites du SIGB (Horizon - format Excel) Données de consultation des ressources électroniques - format Excel	Application web interne permettant de visualiser l'usage des collection et l'évolution pluriannuelle d'indicateurs.	Outil de pilotage de la politique documentaire pour les responsables de collecitons et le responsable Poldoc. Outil de suivi transversal en vue de produire l'ESGBU et les rapports d'activité. Outil de valorisation de la docelec.	Aucune exploitation des données précédentes, en dehors des analyses produites pour les rapports d'activité.
--------------------------------	---	---	---	--	---