

DataViz : bien choisir son outil.

Grille d'analyse et panorama



DataViz : bien choisir son outil

Thomas Porquet

Département Services et Prospective, Couperin

Rachid Aliouat

Bureau informatique documentaire, Université de Lille

Hélène PIPET

Service informatique documentaire, UPEC et Commission SDSI

Bien choisir son outil en 3 temps



Protection et sécurité
des données



Contexte institutionnel



Les critères à
analyser

Dataviz ou Business intelligence, quelles différences ?

La **dataviz** (ou visualisation de données) est un ensemble de méthodes permettant de représenter des **données statistiques** sous **forme graphique** pour en extraire du **sens** ou mettre en évidence des **liens entre des données**.

Business intelligence (BI ou informatique décisionnelle) désigne les moyens, outils et méthodes qui permettent de **collecter, consolider, modéliser et restituer des données**. Il s'agit d'**outils d'aide à la décision**.



Protection et sécurité des données

Dataviz et protection des données

Marie Nikichine
Thomas Porquet

Les enjeux

- Le **RGPD** bien sûr
- La **sécurité** de toutes les données aussi
- Le contrôle de la **ré-exploitation** :
obtenir des garanties



Des informations parfois difficiles à exploiter

	TABLEAU	PANORAMA	METABASE	LODEX
Pays de l'éditeur	Salesforce/ États-Unis	Ebsco /États-Unis	Metabase/États-Unis	Inist-CNRS /France
Hébergement	Cloud ou local	Cloud	Cloud ou local	Cloud CNRS
Politique de confidentialité	https://www.salesforce.com/company/disclosure/	https://more.ebsco.com/privacy-policy https://www.ebsco.com/products/panorama/data-security-privacy	https://www.metabase.com/privacy-policies https://www.metabase.com/hosting/privacy_policy	https://www.istex.fr/politique-de-confidentialite/
Mesures de sécurité, anonymisation	https://trust.salesforce.com/ https://compliance.salesforce.com/ https://developer.salesforce.com/docs/atlas.en-us.security/implGuide/meta/security/implGuide/salesforce_security_guide.htm	Panorama offers a comprehensive strategy to secure library analytics. Your library's data is stored in a secured AWS / Snowflake cloud environment. These networks are architected to protect information, devices and identities. In addition, our team can work with you to understand your institution's privacy regulations to ensure they are upheld.	https://www.metabase.com/security SOC2 Type II GDPR	
Partenaires, transferts de données	FAQ sur les mécanismes de transfert de données en anglais	Library Analytics — Proving Value While Maintaining Privacy	« We may share your personal information with the following categories of third parties: » « All information processed by us may be transferred, processed, and stored anywhere in the world, including but not limited to, the United States or other countries, which may have data protection laws that are different from the laws where you live. »	Les données ne sont pas cédées à des tiers ni utilisées à d'autres fins.

Un affichage qui se veut rassurant

mais qui cache des risques réels

- **Failles de sécurité** / fuites de données :
Salesforces/Tableau
 - (responsabilité interne/responsabilité hébergeur)
- **Des livres blancs** vraiment très blancs: Ebsco/Panorama
- Le **contrôle de la ré-exploitation des données** :
 - Le phénomène de consolidation verticale (Ebsco/Panorama)
 - Des mécanismes difficiles à investiguer (Metabase)

Quels interlocuteurs sur ces questions ?

- RSSI
- GL ESR
- Groupe pistage Couperin et pôle logiciels (pistage@couperin.org)

**Contexte
institutionnel,
local et
national**

Etat des lieux :
les outils de datavisualisation
au SCD de Lille

Rachid Aliouat

Les critères à analyser

Présentation de la grille d'analyse

Hélène Pipet

4 grands groupes de critères :

- Cadre institutionnel, souveraineté et modalités d'implémentation
- Restitution des données et visualisations
- Interopérabilité et partage des données
- Ergonomie

10 outils analysés :

Omniscope, Tableau, Power BI, Cognos, Business Object, Metabase, Kibana, Lodex, Plotly, Seaborn



Cadre institutionnel, souveraineté et modalités d'implémentation

Les éléments à prendre en compte :

- Environnement institutionnel
- Type de logiciel : open Source ou propriétaire ?
- Type d'hébergement et souveraineté des données
- Modèle économique
- La gestion des accès



FOCUS : Hébergement des données



3 grandes options :

- Hébergement local sur un poste individuel
- Hébergement local sur un serveur institutionnel (on premise)
- Hébergement dans le cloud sur un serveur externe

NB : possibilité d'installations hybrides client/serveur (on premise/cloud prestataire)

FOCUS Modèle économique



Logiciel
Open
Source

Gratuité*

Support
payant
si hébergement
cloud

Logiciel
proprié
taire

Modèle
freemium

Licence par
utilisateur

Licence
serveur

Restitution et visualisation des données

Les éléments à prendre en compte :

- Possibilités de personnalisation
- Types de visualisation disponibles
- Interactivité
- Calculs et vectorisation, IA



FOCUS visualisations et interactivités



Quelques fonctionnalités incontournables :

- Graphiques :
 - **répartition** : camemberts, donuts, barres, courbes, nuages de points, treemaps ou cartes de carrés, bulles
 - **évolution dans le temps** : histogrammes, courbes et lignes
 - **relations entre les données** : diagramme circulaire, heatmaps ou cartes de chaleur...
 - **emplacements** : tous types de cartes
- Interactivité : filtres interactifs, drill-down (ou zoom dynamique)...



<https://opendatafrance.gitbook.io/editorialisation-des-donnees-publiques/mettre-en-oeuvre-son-projet-dataviz/etre-guide-dans-le-choix-du-type-dataviz/dataviz-project>

FOCUS IA et dataviz



L'IA comme fonctionnalité principale de la dataviz ?

- IA intégrée au coeur d'outils traditionnels de dataviz (copilot pour Power BI, Tableau +)
- Arrivée de nouveaux outils intégrant nativement l'IA (Veezoo par exemple).

Possibilité de poser des questions en langage naturel et de recevoir des visualisations et des tableaux de bord en réponse

Interopérabilité et partage des données

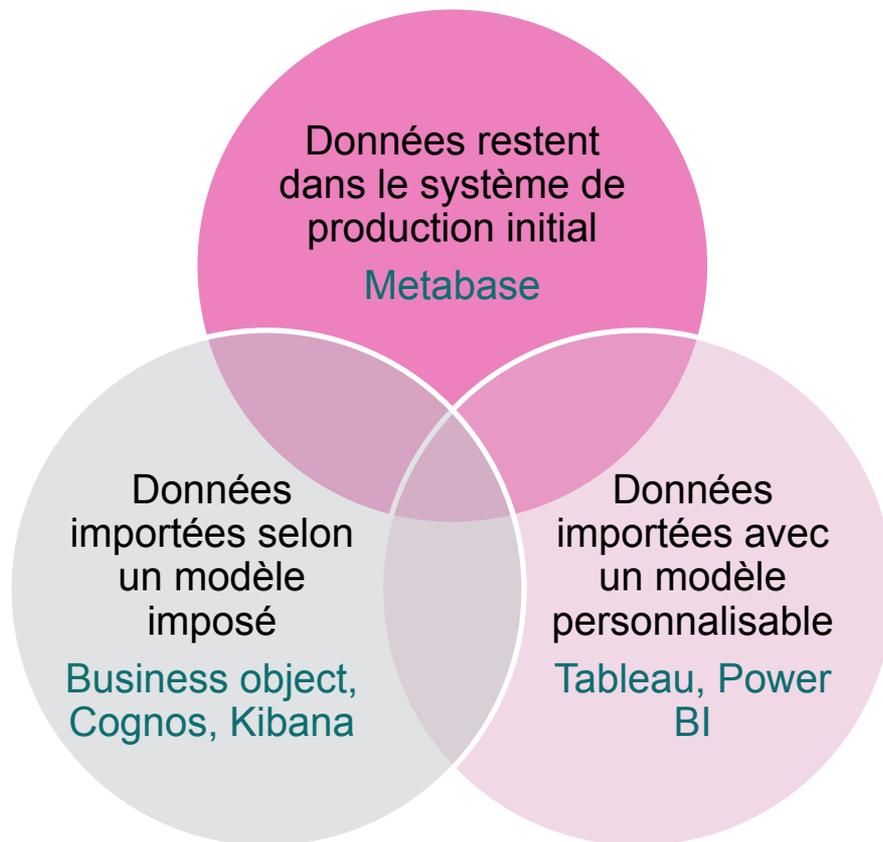
Les éléments à prendre en compte :

- Formats de données et compatibilité des sources
- Possibilités d'intégration technique
- Types de partage possibles
- Collaborations possibles



3 grandes logiques de manipulation des données

FOCUS Traitement et import de données



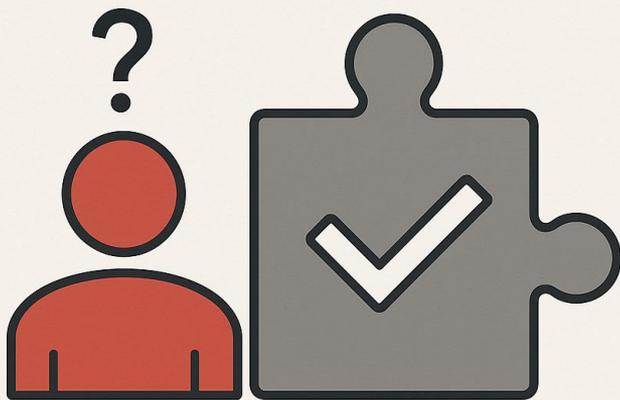
Ergonomie

Les éléments à prendre en compte :

- Facilité d'apprentissage (quelles sont les compétences disponibles ?)
- Ergonomie de l'interface



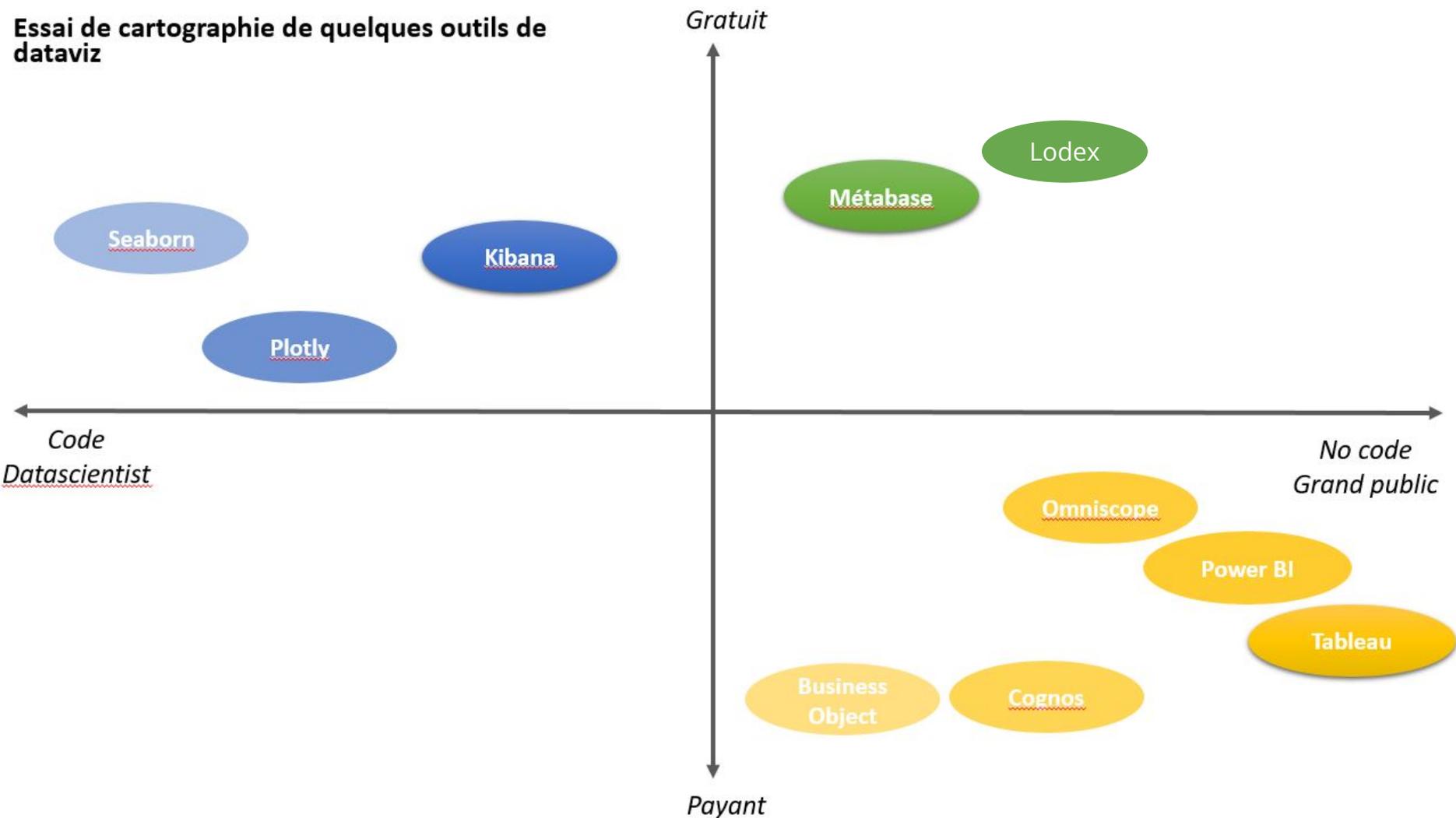
En conclusion



Le bon outil, c'est
avant tout celui qui
nous correspond



Essai de cartographie de quelques outils de dataviz



Extrait : tableau comparatif des outils de Dataviz

Grille d'analyse comparative d'outils de datavisualisation									
	Critère	Omniscope	Tableau	Power BI	Cognos	Business Object	Metabase	Kibana	Lodex
Souveraineté des données et modalités d'implantation	Open-source	Non	Non	Non	Non	Non	Oui	Oui	Oui
	Editeur	Visokio	Salesforce	Microsoft	IBM	SAP	Metabase	Elastic	INIST-CNRS
	Hébergement	Hébergement on-premise, cloud privé ou public	Tableau Server (on-premise), Tableau Cloud, Tableau Public (freemium)	Power BI Desktop (on-premise), Power BI Service (cloud Microsoft)	Déploiement sur serveurs IBM cloud data center ou cloud privé	Hébergement on-premise ou via SAP BI Platform	Hébergement on-premise ou cloud (Metabase Cloud via AWS)	Hébergement on-premise ou via Elastic Cloud	Hébergement on-premise (code accessible librement) ou cloud (infrastructure ISTEK)
	Emplacement des serveurs si hébergement cloud	Pas de data center imposé. cloud public (comme AWS, Azure, Google Cloud), ou chez des prestataires français (par exemple OVHcloud, Enova, OKTEO)	Hébergement sur Amazon Web Services (en Europe mais pas nécessairement en France)	Variable en fonction de la demande des clients. Data centers français : OVHcloud ou DecIvision OU cloud SAP (pas d'hébergement systématique en France).	Data center IBM à Clichy-La-Garenne	Cloud privé ou public (Microsoft Azure, OVHcloud, ou d'autres fournisseurs d'hébergement certifiés SAP)	Francfort sur Amazon Web Services	Hébergement possible via Amazon Web Service, Microsoft Azure et Google Cloud Platform. Lors de la création d'un cluster, il est possible de choisir la région d'hébergement, notamment « eu-west-3 » (Paris) sur AWS.	Hébergement en France via infrastructure ISTEK
	Modèle économique	Licence payante (par utilisateur ou serveur), essais gratuits possibles	Licence payante (par utilisateur ou serveur), version public gratuite limitée	Licence payante (sauf Power BI Desktop gratuit, licences Pro et Premium payantes)	Licence entreprise (sur devis), modèle par utilisateur ou serveur	Licence SAP avec coût élevé, souvent intégré à des offres groupées	Gratuit en open-source (auto-hébergé), version cloud payante avec fonctionnalités avancées (Metabase Cloud)	Gratuit (open-source) avec support payant via Elastic	Hébergement gratuit pour les unités affiliées au CNRS, ouverture possible pour les établissements adhérents ISTEK
	Gestion des accès	Authentification avec login/mot de passe interne à l'application, gestion des rôles	Gestion fine des accès, SSO, sécurité renforcée via Tableau Server	Intégration Active Directory	Authentification LDAP ou AD, gestion des droits utilisateur avancée	Gestion de sécurité intégrée à SAP	Authentification LDAP, gestion des rôles basique	Contrôle par rôles, sécurité d'accès par index Elasticsearch	Comptes locaux avec 2 niveaux d'accès (administrateur et utilisateur)
	Partage et réutilisation des données								
	Communauté et support	Communauté restreinte, support payant par Visokio, inclus dans la licence	Très grande communauté, supports officiels, nombreux tutoriels	Support Microsoft, communauté active, forum	Support IBM, communauté professionnelle	Communauté SAP + support entreprise	Grande communauté open-source, documentation abondante et claire	Large communauté Elastic, documentation technique	Liste de diffusion dédiée, documentation complète en ligne, support technique INIST
Restitution des données et visualisations	Personnalisation	Graphisme très personnalisable (couleurs, interactions, calculs, formats) Transformation des données	Hautement personnalisable (couleurs, interactions, calculs, formats)	Personnalisable avec le langage DAX	Personnalisation possible mais interface plus rigide	Personnalisation via paramètres prédéfinis	Thèmes simples, personnalisation graphique limitée	Tableaux de bord personnalisables mais axés sur les séries temporelles	Thèmes graphiques personnalisables. Extensions avancées pour ajout de fonctionnalités
	Types de visualisations	Graphiques classiques, cartes, diagrammes de réseau, flux de données	Nombreuses visualisations : barres, courbes, cartes, nuages de points, treemaps, etc.	Nombreuses visualisations : barres, courbes, cartes, nuages de points, treemaps, etc. + intégration facile avec Excel	Mise en page classique, peu flexible. Graphiques, jauges, tableaux croisés, cartes.	Formats prédéfinis SAP. Graphiques classiques, tableaux croisés, diagrammes.	Tableaux et graphiques simples, peu de formats alternatifs	Tableaux de bord orientés logs et séries temporelles, histogrammes, cartes de chaleur, focus sur séries temporelles	Graphiques : barres, histogrammes, diagramme à bulles, radars, treemaps, heatmaps, flux, réseaux, cartes géographiques...
	Interactivité	Filtres, drill-down, interactions entre visualisations	drill-down, actions contextuelles, filtres dynamiques	Filtres, drill-down, synchronisation Excel	Drill-down, rapports interactifs	Rapports interactifs, drill-down	Filtres, drill-down basique	Filtres par séries temporelles, zoom, drill-down	Filtres dynamiques, info-bulles
	Intégration de référentiels géographiques	Cartographie avancée, prise en charge de shapefiles (format fichier SIG)	Cartes de chaleur, géolocalisation	Visualisations géographiques riches	Cartes basiques	Fonctions géo limitées	Cartes simples avec coordonnées	Intégration avec GeoIP, cartes logiques	Cartes de chaleur, cartes de flux, cartes géographiques

Les suites de la journée d'étude "De la gouvernance des données à la datavisualisation" :



17/06/2025

JOURNÉE DATAVIZ

