

CENTRALISER ET PARTAGER
LA QUALITÉ DES DONNÉES
AVEC UN
DATA CATALOGUE



L'ÉCOSYSTÈME DATA A L'ÈRE DU TAYLORISME

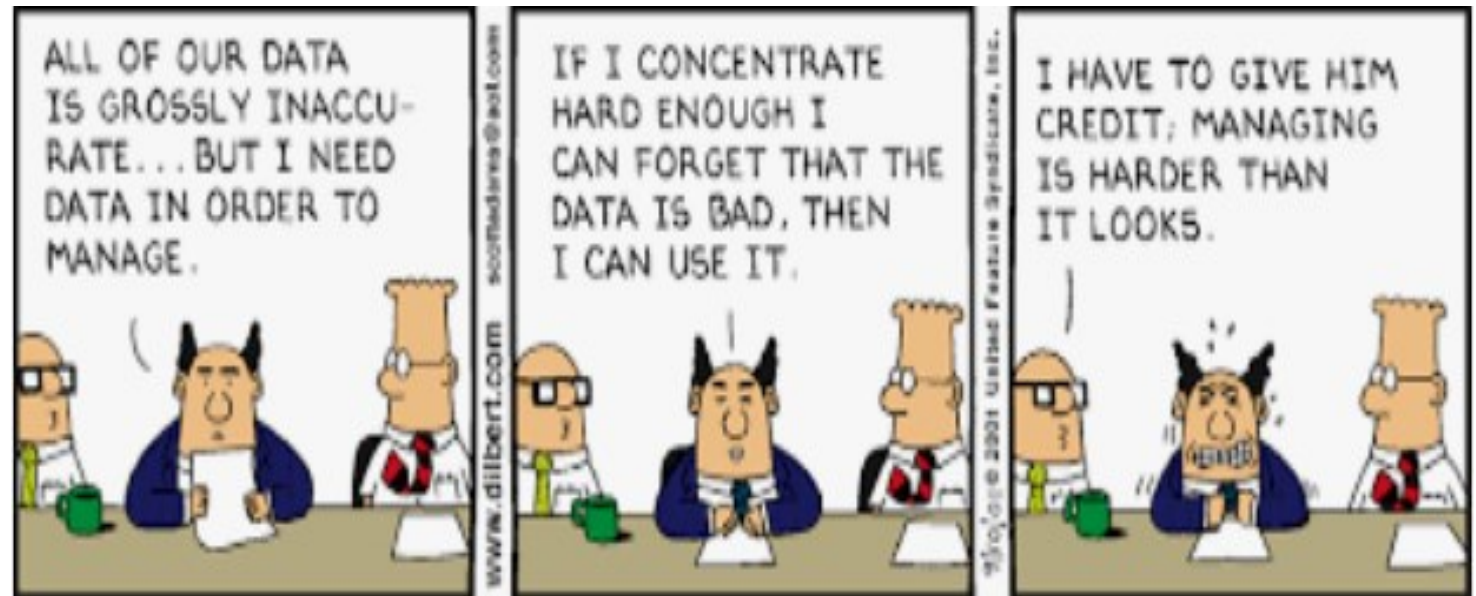
Octobre 2021

LA NON-QUALITÉ DES DONNÉES : UN OBSTACLE A L'USAGE DE LA DONNÉE

82%

*des responsables des données ...
ont déclaré que les problèmes
de qualité des données
constituaient un obstacle aux
projets d'intégration des
données*

Data Integrity Trends: Chief Data Officer Perspectives in
2021 - precisely



Prise de décision
Connaissance client
Productivité
Conformité

Non qualité
des
données

Organisation silotée
Opportunités manquées
Perte de revenus
Atteinte à la réputation

LA NON-QUALITÉ DES DONNÉES : DES CONSTATS BIEN IDENTIFIÉS

A la question « Où votre organisation a-t-elle le plus de difficultés avec la qualité des données ? », les experts data mettent en avant la rationalisation du dispositif dans l'entreprise

39%

Standardiser un ensemble commun de règles de qualité des données dans tous les systèmes

24%

Identifier les erreurs de données dans vos systèmes sources

19%

Définir des règles de qualité des données pour résoudre les problèmes de données découverts

18%

Piloter les erreurs de données et les exceptions



Notre enjeu : bâtir un socle data permettant de rationaliser le procédé d'évaluation de la qualité des données et ainsi fédérer les différentes parties prenantes autour d'un langage et des normes de fabrication en commun.

Composants du socle data combinés :



Datalake



Plan de contrôles métier



Data Catalogue

Modernizing Data Quality & Governance - white paper 2016 Informatica

QUELQUES CONCEPTS CLÉS



Data catalogue



Inventaire organisé des actifs de données dans l'organisation.

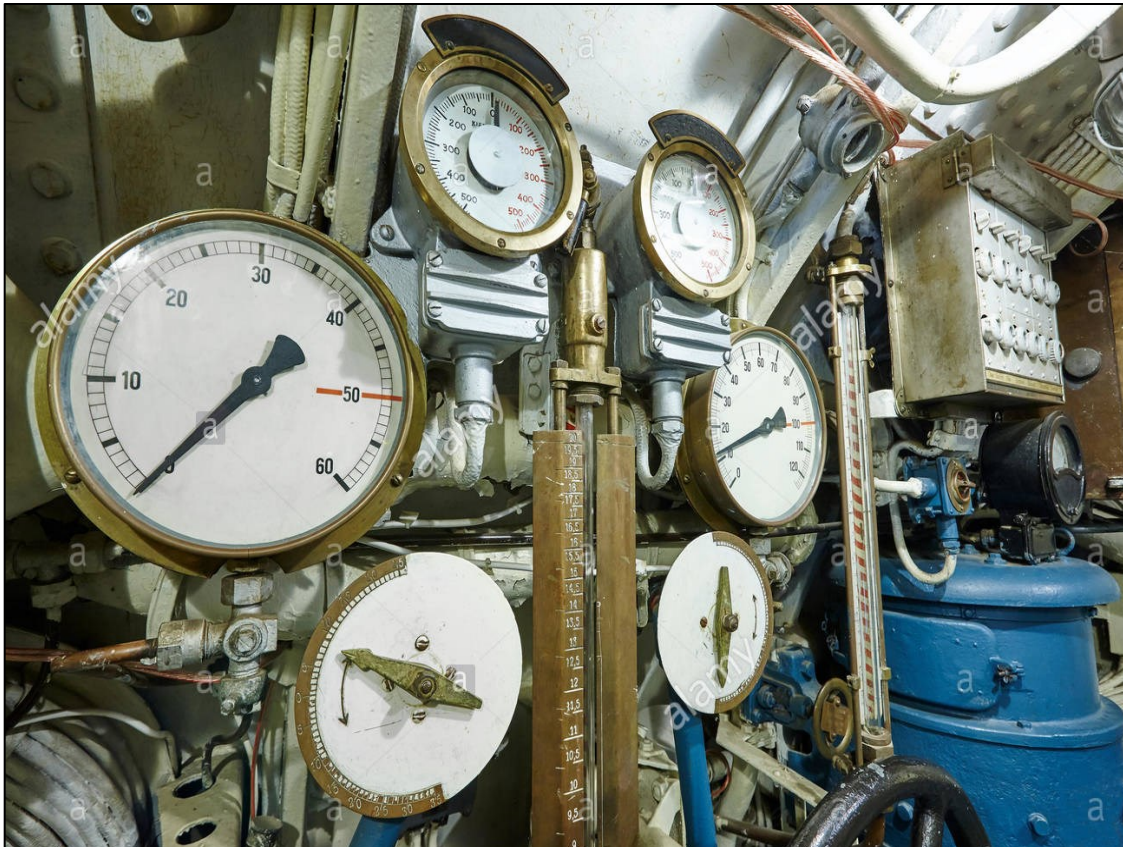
Ses caractéristiques :

- Documente les dimensions métiers et techniques de la donnée
- Développe un langage d'entreprise avec une taxonomie commune
- Facilite la recherche des données
- Accroît la collaboration entre acteurs autour d'un langage commun
- Publie les exigences de qualité des données et les résultats de contrôle
- S'interface avec le système d'information pour automatiser la collecte des métadonnées (et éventuellement les données elle-même)

QUELQUES CONCEPTS CLÉS



Plan de contrôles métier



Stratégie de contrôles apportée aux données des processus de l'entreprise.

Ses caractéristiques :

- Sécurise la relation client/fournisseur en garantissant la fiabilité des données consommées pour l'usage souhaité
- Fabriqué au cas pas cas, en fonction des silos de l'entreprise, de la maturité des acteurs, de la nature des processus, des réglementations, du mille-feuille IT
- Mutualisation limitée des moteurs de contrôles
- Publication locale des résultats de contrôles

QUELQUES CONCEPTS CLÉS



Datalake

Référentiel centralisé qui permet de stocker des données structurées et non structurées à n'importe quelle échelle.

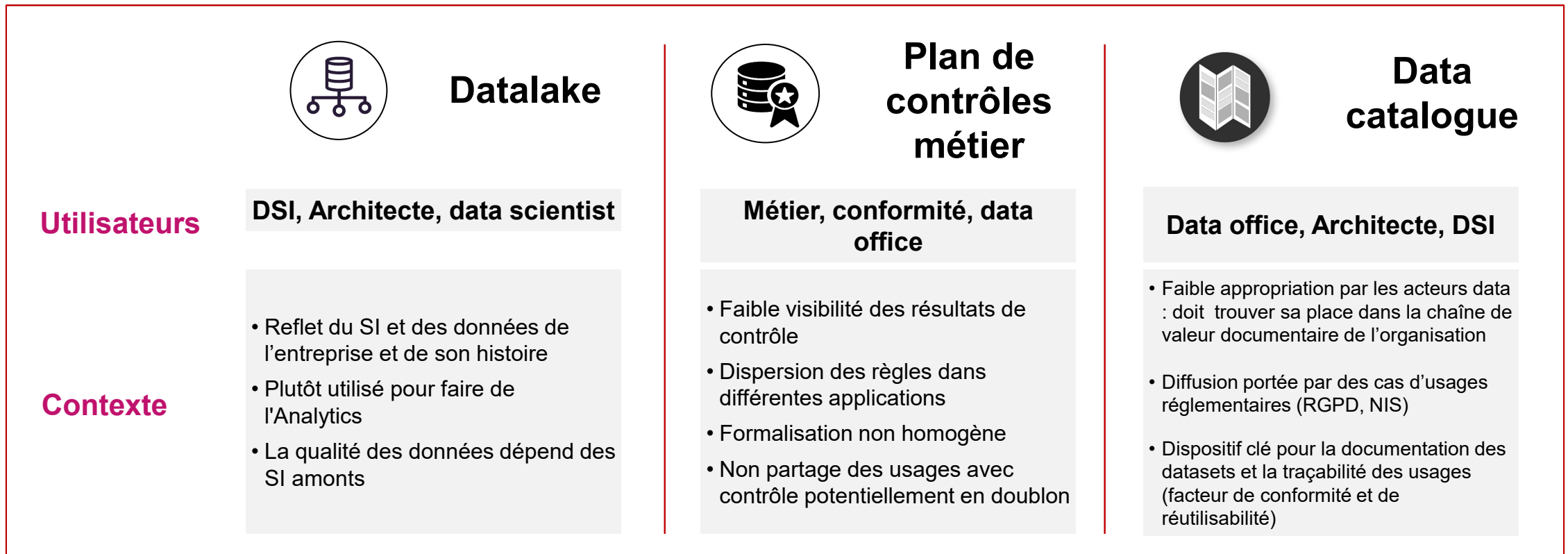
Ses caractéristiques :

- Alimenté par les systèmes d'information clés de l'entreprise
- Porte de plus en plus les cas d'usage stratégiques de l'entreprise
- Grandes variétés de données dans un environnement technologique unique et une gouvernance centralisée propice à une taylorisation du processus de fabrication des projets data (division des tâches \normalisation)



EVOLUER D'UN CONTEXTE SEGMENTÉ ...

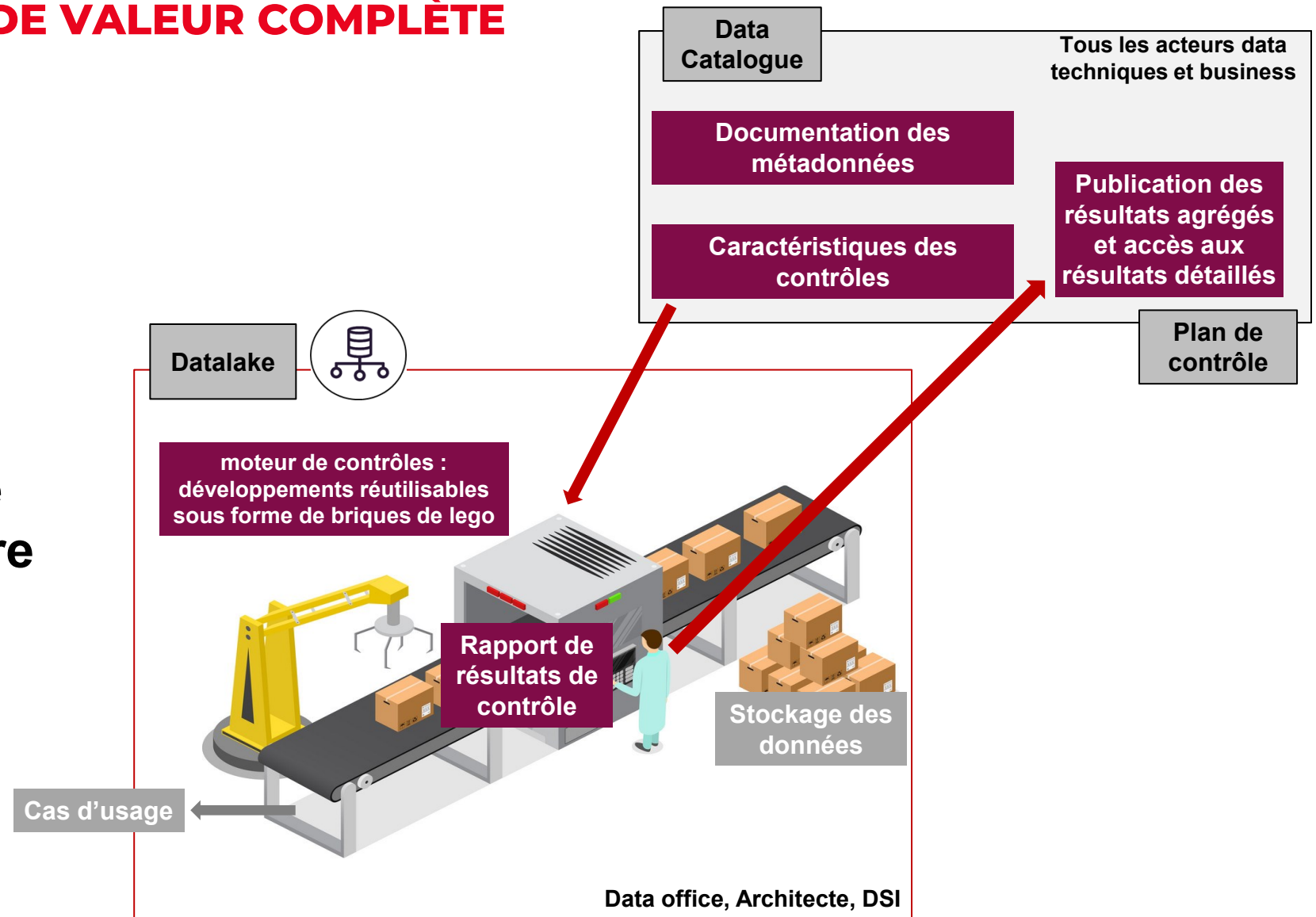
Dans une approche traditionnelle, orientée processus, chaque métier gère son propre patrimoine de données, en étant souvent réticent à l'ouvrir aux autres métiers de l'entreprise.



L'utilisation dispersée de ces dispositifs limite l'efficacité de la stratégie data et augmente le risque de défaut de couverture

... VERS UNE CHAÎNE DE VALEUR COMPLÈTE

Faire converger ces différentes sphères d'expertises, pour augmenter la valeur des chantiers de l'entreprise en définissant un cadre homogène de mesure de la qualité des données



Une solution d'architecture centralisée sur le datalake qui démocratise l'accès à la qualité des données

Bénéfices

	Solution	Gouvernance
Procédés de fabrication standardisés (modèle lego)	✓	
Coûts de fabrication réduit (capitalisation sur la plate-forme)	✓	
Time-to-market réduit	✓	
Gouvernance des données by design (prise en compte dans les projets)	✓	✓
Expérience utilisateur sans couture intégrant les enjeux de restitution	✓	✓
Appropriation par le top management (pilotage centralisé)		✓
Collaboration entre acteurs autour d'un langage commun (sensibilisation par capillarité)		✓
Optimisation du plan de contrôle (mutualisation \réduction des contrôles en double)		✓

Facteurs limitatifs

Mesure de la qualité des données sur des golden copies (remédiation par redressement en bout de chaîne)

Exhaustivité et fréquence de rafraîchissement des données

Resistance au changement des processus disposant déjà d'une offre de service (legacy)

**CENTRALISER ET PARTAGER
LA QUALITÉ DES DONNÉES
AVEC UN
DATA CATALOGUE**



L'ÉCOSYSTÈME DATA RENTRE DANS L'ÈRE DU TAYLORISME

Octobre 2021

Q&A

