

RESTITUTION

Journée de Veille DATA.IA

3 juillet 2023



Une journée de veille et d'idéation collective autour les enjeux data du Groupe La Poste



CONTEXTE DE LA JOURNÉE

Pour rappel, cette Journée de Veille et d'Idéation s'est inscrite dans le cadre de la Cellule de Veille Plateforme Data qui, portée par les CDOs des différentes branches et entités du Groupe, a pour but de stimuler l'innovation sur les problématiques de Technologies et Plateformes.

20
participants

4
ateliers

8
experts

12
books
remplis

RAPPEL AGENDA DE LA JOURNÉE



onepoint.
beyond the obvious

● 09:00 30min	ACCUEIL ET CAFÉ	Se retrouver	
● 09:30 15min	MOT DE PIERRE-ÉTIENNE	La Poste et la Cellule de Veille	Pierre-Etienne Bardin
● 09:45 10min	PRÉSENTATION PAR GONTRAN	Introduction de la journée	Gontran Peubez
● 09:55 15min	AGENDA	Présentation de l'agenda de la journée	Grégory Lecointe
● 10:10 1h	INTERVENTION 1	Data Trends & Issues 2023	Aude Perdriel-Vaissière, Nora Diep
● 11:10		PAUSE	
● 11:30 1 heure	INTERVENTION 2	Souveraineté Numérique	Matthieu Bourgeois
● 12:30		DEJEUNER	
● 13:30 10min	EXPLICATIONS ATELIERS	Présentation des ateliers	Grégory Lecointe
● 13h40 5min	ICEBREAKER	Mise en mouvement	Charlaine Crutz
● 13:45 1h40	2 ATELIERS		
● 15:25		PAUSE	
● 15:35 5min	ENERGIZER	Remise en mouvement	Charlaine Crutz
● 15:40 1h40	2 ATELIERS		
● 17:20 1h	CONCLUSION	Restitution des ateliers	Experts onepoint
● 18:30 2h30		COCKTAIL	
● 21:00		DÉPART DES LIEUX	

PARTICIPANTS GROUPE LA POSTE



Groupe	BGPN	BSCC	LBP	CNP	OpenValue	DIWO
Pierre-Etienne Bardin	Jeremy Tillet	Arnaud Lahoule	Matthieu Olivier	Benoit Lecussan	David Gayou	Eric Gabillet
Olivier Moraglia	Thibault Ledan	Julie Virly	Régine Buyss	Nicolas Schaaf		
Nestor Yepmeni	Thierry Leclere	Setra Rakotomavo	Muriel Combes			
Frederic Buissez	Thomas Schwender		Pierre Dehaene			
Omar Hamour			Frederic Laigre			
Maxime Canneva						

INTERVENANTS ET ORGANISATEURS



Organisation	Data Sharing	Data Engineering	Data Gouvernance	IA/ML/ LLM Ops	Data Trends	Introduction
Grégory Lecointe	Florent Pouget	Manuel Tang	Olivier Parrel	Stéphane Rion	Nora Diep	Gontran Peubez
Victor de Chantérac			Arnaud de Chambourcy Metayer		Aude Perdriel-Vaissiere	
Charlaine Crutz						
Naomi Darrigol						

klein • wenner





Présentation Étude Data Trends & Issues 2023

Nora Diep et Aude Perdriel-Vaissière

Onepoint

Plénière du matin

RAISON D'ÊTRE DE L'ÉTUDE



Cette seconde édition de nos Data Issues, Data Trends s'inscrit naturellement dans la continuité de ce qui fut le déclencheur de l'étude originelle :

« *les tendances que nous observons en matière de data et d'intelligence artificielle apportent-elles des réponses concrètes, opérationnelles et opérables aux problématiques rencontrées par nos clients ?* »

Bien sûr l'étude se garde de fournir une réponse tranchée qui n'aurait que peu de sens en dehors de contextes très spécifiques.

ETUDE DATA TREND 2023

82 CHIEF DATA OFFICERS



Grande
entreprise



Entreprise de taille
intermédiaire



Petite moyenne
entreprise

4 PROBLEMATIQUES



PEOPLE



GESTION DES DONNEES



INDUSTRIALISATION



« DATA FASTING »



8 TENDANCES 2023

Data Ecosystem
Frugality

Beyond 360° Data

Data Protection in
SecNumCloud

Data Productization

Data Marketplace

Low Code Data Platform

Beyond Data Governance

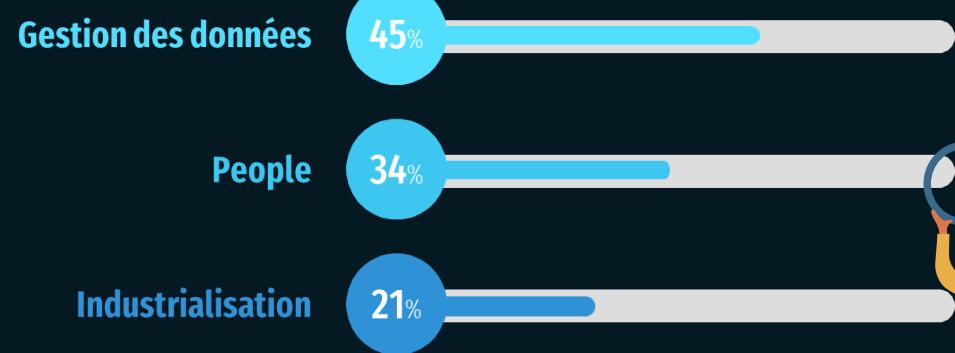
Data RESET

DATA ISSUES

EVOLUTION DES PROBLEMATIQUES
Identifiées dans la précédente étude



PRIORITES 2023



NOUVELLE PROBLEMATIQUE « DATA FASTING »

- Fin des invest de set up, mode run
- Paradoxe : résultats CT vs transfo

CONSTAT

- Recherche de ROI
- Complexité Gestion temps courts / temps longs

RAISONS

- Use cases métier en mode 'product'
- Frugalité + Ténacité stratégique

CONVICTIONS

ÉVOLUTION DES DATA TRENDS 2022 ET DATA TRENDS 2023

- Tendances qui font leur apparition en 2023
- Tendances qui se sont hybridées entre elles pour n'en former qu'une seule
- Tendances qui perdurent, parfois en s'amplifiant ou en mutant à la marge

2022

Data Product

Data science as BI

Monétisation

Data Sharing

Low code Data pipeline

Data gouvernance augmentée

Data RESET

Data academy

Data Xops

Data Office as Profit center



2023

Data Ecosystem Frugality

Beyond 360° Data

Data Protection in SecNumCloud

Data Productization

Data Marketplace

Low Code Data Platform

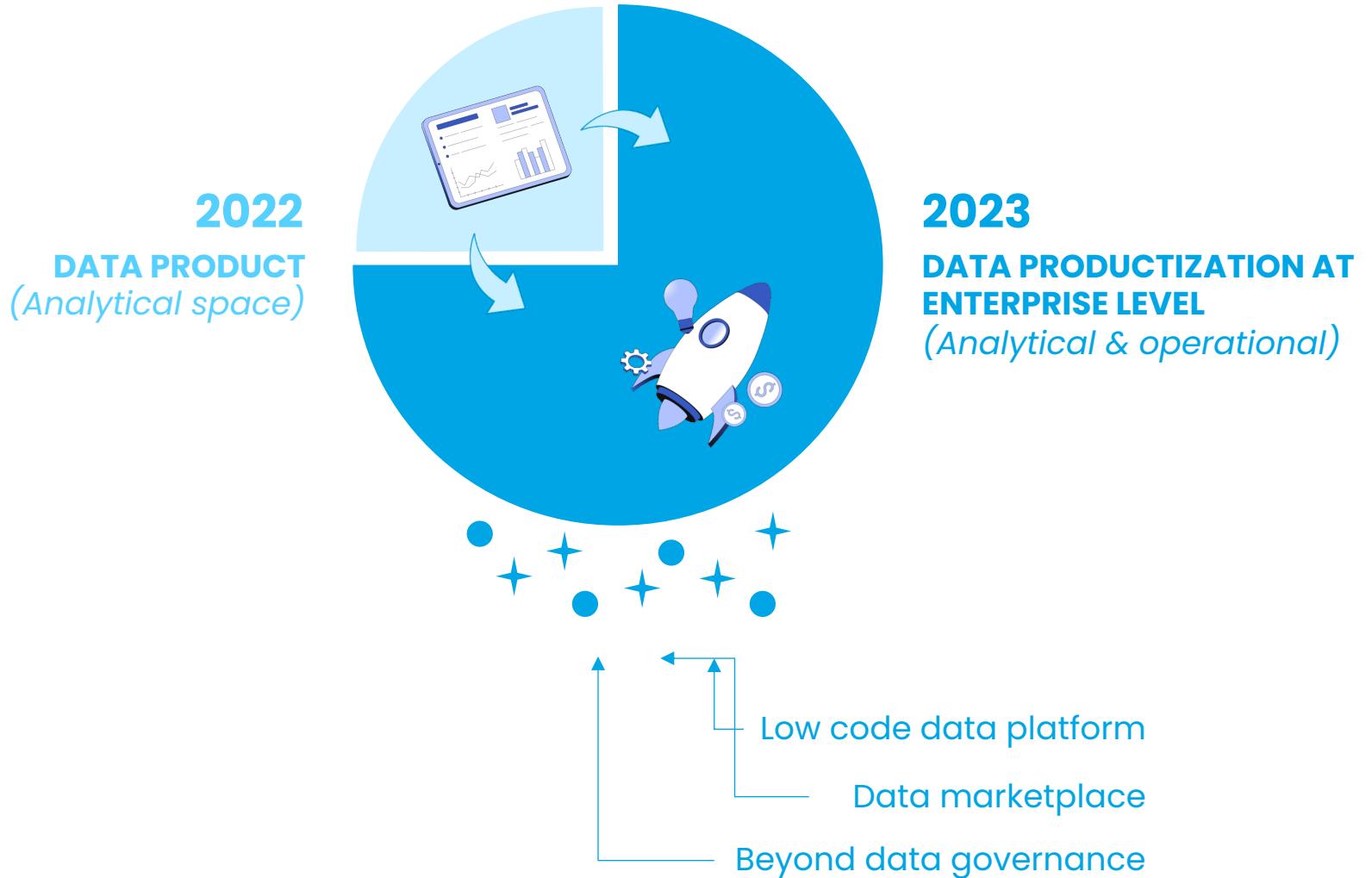
Beyond Data Governance

Data RESET

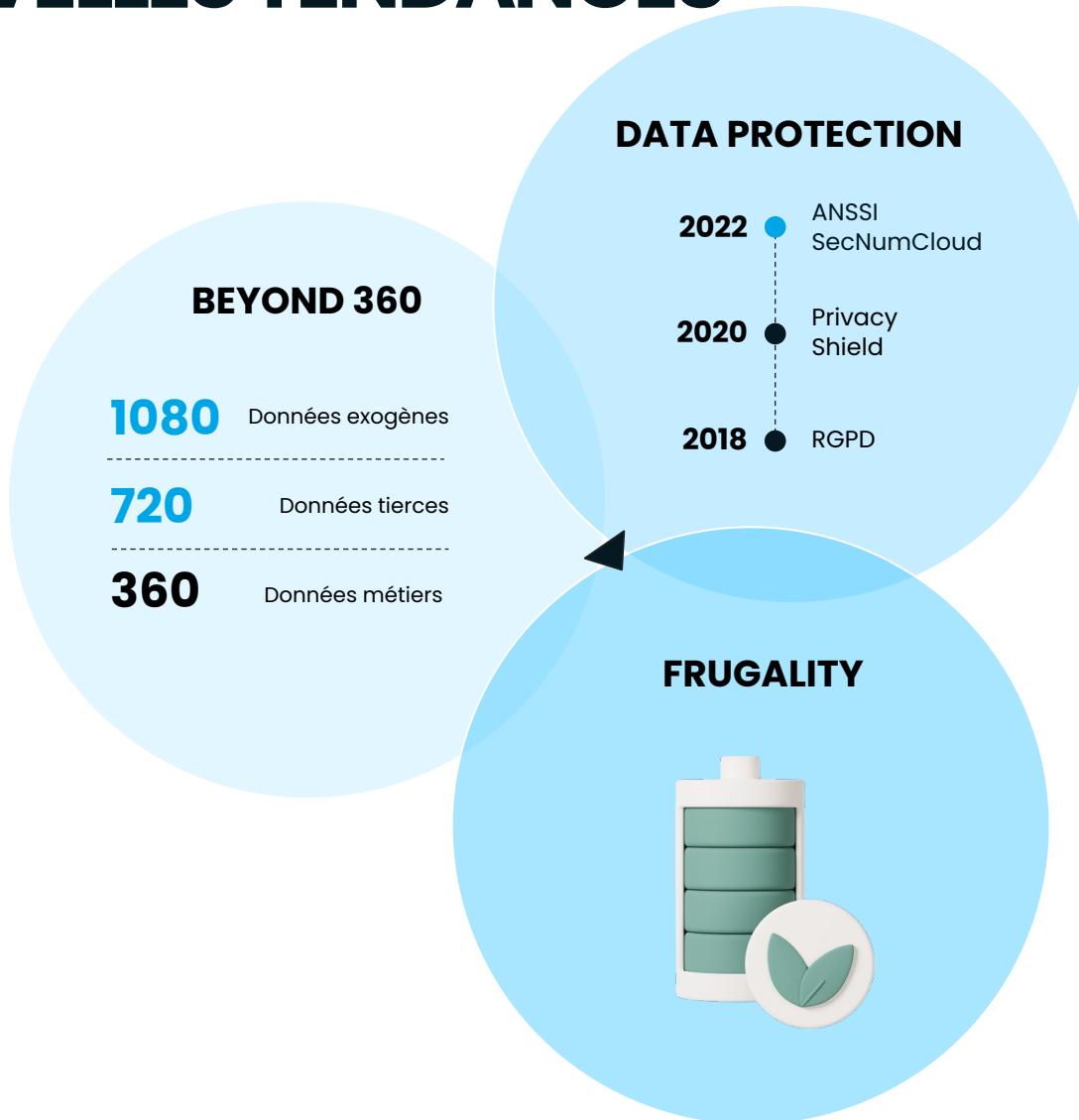
Intégrées dans le business as usual

Tendance ajournée

DATA PRODUCTIZATION

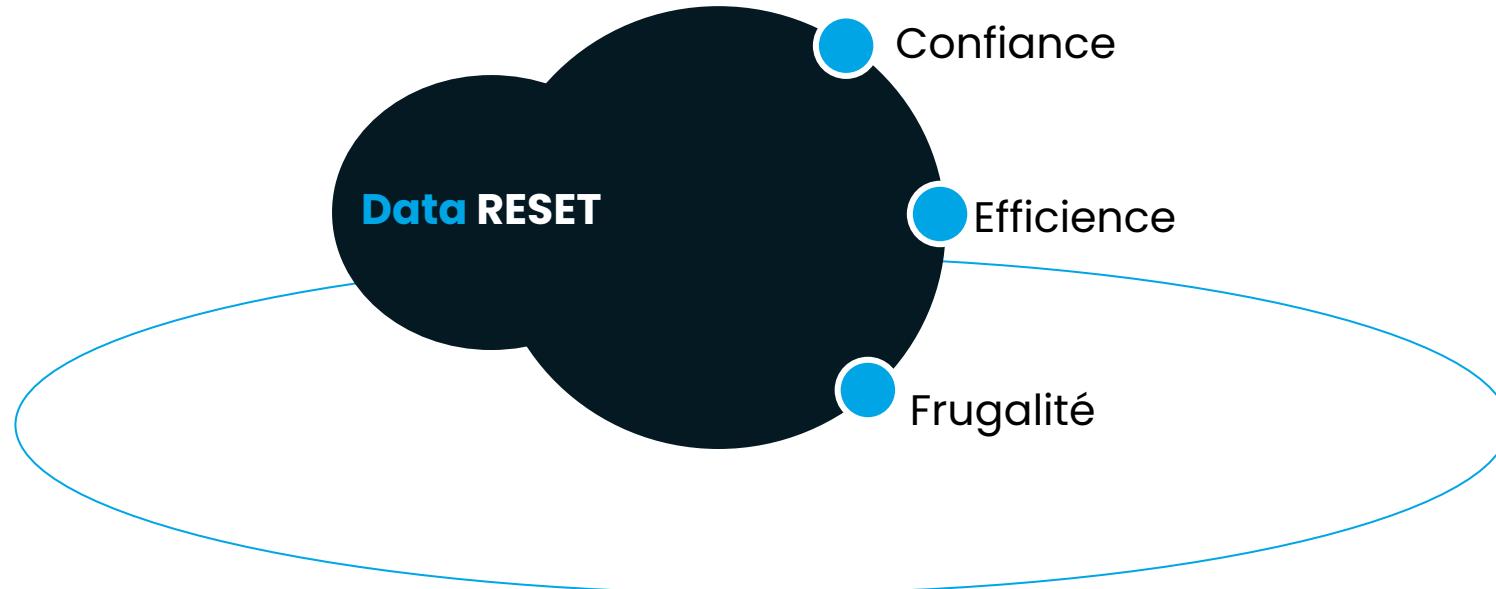


NOUVELLES TENDANCES



RESET

Responsabilité Environnementale, Sociétale, Économique et Technologique





Souveraineté Numérique

Matthieu Bourgeois

Klein Wenner

Plénière du matin



Souveraineté numérique : Essai pour une reconquête

3 juillet 2023 – OnePoint

Public



Une étude issue des travaux conjoints du cercle de la donnée et l'Agora 41

- Un groupe de travail formé il y a 2 ans
- 20 auteurs et contributeurs du Cercle de la donnée et de l'Agora 41
- Préface de Guillaume Poupard, DG de l'ANSSI
- Parution le 24 janvier 2022



PREMIÈRE PARTIE DÉFINITIONS	
Chapitre 1 : LA SOUVERAINETÉ	7
Chapitre 2 : L'ESPACE NUMÉRIQUE	29
DEUXIÈME PARTIE RÉFLEXION	
Chapitre 1 : LA SOUVERAINETÉ APPLIQUÉE À L'ESPACE NUMÉRIQUE	47
Chapitre 2 : LES DÉFIS À RELEVER.....	67
TROISIÈME PARTIE PROPOSITIONS	
Proposition n° 1 : DÉFINIR ET PILOTER UNE STRATÉGIE DE CYBERSOUVERAINETÉ.....	95
Proposition n° 2 : INVESTIR DANS LA RECHERCHE NUMÉRIQUE FONDAMENTALE FRANÇAISE ET EUROPÉENNE.....	97
Proposition n° 3 : STIMULER LES INVESTISSEMENTS PRIVÉS DANS LE NUMÉRIQUE.....	99
Proposition n° 4 : IMPOSER UNE SOUVERAINETÉ JURIDIQUE EUROPÉENNE POUR LES DONNÉES LES PLUS SENSIBLES.....	103
Proposition n° 5 : ARMER LE DROIT DE LA CONCURRENCE FACE À L'ÉCONOMIE DE LA DONNÉE.....	107
Proposition n° 6 : CRÉER DES PROFESSIONS RÉGLEMENTÉES POUR LE NUMÉRIQUE.....	111
Proposition n° 7 : CAMPAGNE DE SENSIBILISATION CITOYENNE AVEC L'ÉDITION D'UN LIVRET D'INSTRUCTION CIVIQUE NUMÉRIQUE.....	115

De la souveraineté, à la souveraineté numérique, entamons la reconquête!

**La panne française et
européenne**
(1980 – 2022)

7 propositions pour la reconquête



Volet Economique

Stimuler les investissements privés dans le numérique

Armer le droit de la concurrence face à l'économie de la donnée

Créer des professions réglementées pour le numérique



Volet Education & Recherche

Investir dans la recherche numérique fondamentale française et européenne

Campagne de sensibilisation citoyenne avec l'édition d'un livret d'instruction civique numérique



Volet Juridique & Politique

Imposer une souveraineté juridique européenne pour les données les plus sensibles

Définir et piloter une stratégie de cybersouveraineté

L'impact de cette étude

- Lien vers l'étude :
<https://www.lecercledeladonnee.org/souverainete-numerique-essai-pour-une-reconquete/>
- Présentation auprès de députés, à l'**Assemblée Nationale** le 24 janvier 2022
- Présentation à l'**Institut Diderot**, le même jour
- Nombreuses reprises dans la presse, dont France Info, et Smart Tech (Bsmart)
- Envoi d'exemplaires papiers à chacun des **partis politiques crédités > 5% vote** dans les sondages au printemps 2022



Qui sommes-nous ? Intervenants

SOUVERAINETÉ NUMÉRIQUE / DATA
RECONQUÉRIR LA SOUVERAINETÉ NUMÉRIQUE
Publié en janvier 2022

MATTHIEU BOURGEOIS et BERNARD DE COURREGES D'ISTOU

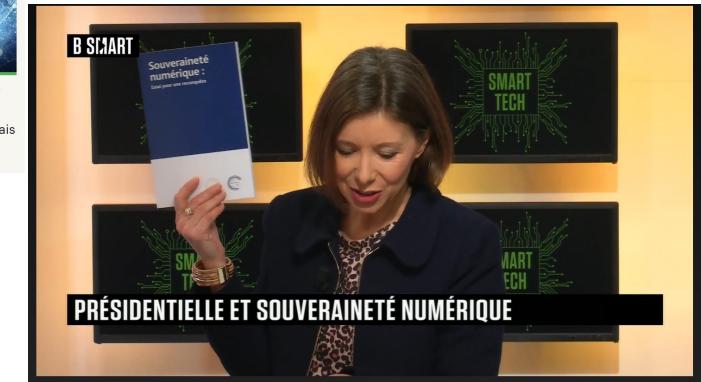


l'Eclaireur FNAC

FNAC LIVE • CULTURE • POP CULTURE • TECH • SOCIÉTÉ

Le numérique, un domaine impliquant l'ensemble de la société. Isidorecor / Shutterstock

Dans une étude, deux groupes de réflexion français offrent des suggestions sur le sujet qu'ils considèrent comme crucial.



franceinfo:

vidéos | radio | jt | magazines | |

examens Q&A politique vrai ou fake société faits-divers santé éco/conso monde europe culture sport en

Cet article date de plus d'un an.

Présidentielle 2022 : le numérique s'invite dans la campagne

écouter (3min)

La suite de cette étude

- Un groupe qui s'est remis au travail fin 2022
- Le choix du sujet : « *Créer des professions réglementées pour le numérique* »
- Un lien évident avec la souveraineté : garantir la qualité/probité, pour retrouver la confiance

[*Ebauche de plan*]

SOUVERAINETE NUMERIQUE ET PROFESSIONS REGLEMENTEES

I – DEFINITIONS

- A. Les professions réglementées
- B. Les métiers du numérique

II – REFLEXION

- A. Les insuffisances du cadre actuel
- B. Opportunité / Cadre souhaité

III – PROPOSITIONS

- A. Au niveau des entreprises
- B. Au niveau de l'Etat
- C. Dimension internationale



Notes des ateliers collectifs

Séquence de l'après-midi

RAPPEL DU FONCTIONNEMENT DES ATELIERS

Déroulé

- **4 groupes** de 4 à 6 personnes mixants branches et entités
- Chaque groupe couvre **4 sujets**
- 30 minutes de **réflexion commune** autour du sujet
- 10 minutes **d'appropriation par branche**

Objectif

1. Découvrir de **nouvelles manières** de résoudre des problématiques
2. **Partager** entre branches
3. **Échanger avec les experts one point** pour s'inspirer de méthodes et solutions externes au Groupe

Stéphane
RION



IA/ML/LLM
Ops

Florent
POUGET



Data Sharing



Data
Engineering

Manuel
TANG



Arnaud DE
CHAMBOURCY

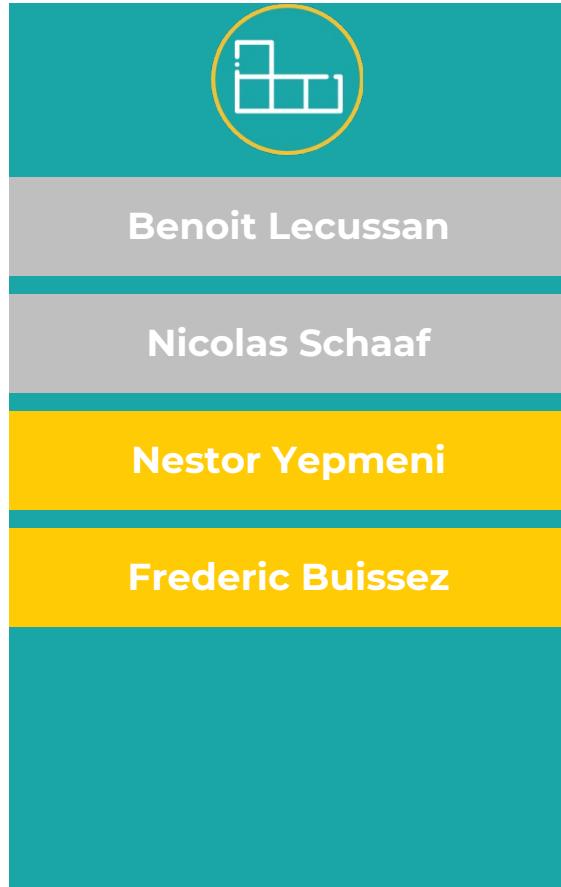


Data
Gouvernance

Olivier
PARREL



RAPPEL DES GROUPES



Légende :

Groupe	BSCC	BGPN	LBP
CNP	OpenValue	DIWO	

NOTE D'INTRODUCTION

Les planches suivantes présentent la synthèse par thème abordé des discussions qui se sont tenues avec les quatre groupes.

Les participants étaient incités à réfléchir collectivement à sur un problème donné et partager comment chacun envisageait la ou les solutions dans le cadre de sa compétence et de son activité.

Les experts Onepoint animant les ateliers étaient également incités à présenter leurs propres solutions émanant de leurs expériences en dehors du Groupe.

Les idées capturées sont regroupées par sous-thèmes ou par solutions selon l'animation spécifique de chaque atelier.

DATA SHARING



Une vision long terme qui s'appuie sur :

De la confiance

Un respect des **documentations**, qui renforce la **confiance** et qui peut passer par des accréditations, labels, certifications, notes de qualité, ainsi qu'une documentation de langage commun entre les métiers.

Des règles

Un **respect des règlementations**, de l'anonymisation, avec un **contrôle des accès, et du lineage** sur l'utilisation pour prévenir des comportements dangereux ou néfaste au partage des données (ex. duplication).

Des process

- Une **gouvernance et un cadre** définis qui définissent les règles et usages, ainsi que les politiques de réutilisation, de consommation et de responsabilité.
- Une **vision centralisée des achats et acquisitions** pour permettre une optimisation des dépenses et des travaux de mise en qualité.
- Un **processus clair et industrialisable**, réplicable et compréhensible par toutes parties prenantes.

Des solutions évoquées

- Intégrer un **partage par API relié au Data Catalog** qui peut permettre un contrôle et la mise en place d'autorisations.
- Une **diffusion de data products**, validés par les métiers et liés à un contrat d'usage, une source de données, des utilisateurs, etc.
- Un **point d'entrée et de sortie unique** (ex. Data catalog, marketplace) bien alimenté, en adéquation avec les systèmes actuels, user-friendly, qui peut permettre de trouver les produits data existants, d'évaluer les nouveaux besoins data, et de partager sur un référentiel commun.
- Une recommandation de **découpage par data domain** pour faciliter le déploiement et l'usage d'une marketplace interne.

DATA GOUVERNANCE (1/2)

Une approche « by design », data avant, pendant et après les projets



Collecte

- En amont : **Expression d'un besoin data**, identification des objets métiers et des données nécessaires à mon projet.
- Bien **inclure les DPO** pour toutes créations de nouveaux flux.
- Mise en place de **contrats d'usages, cadres de références** pour contrôler les usages, qui suivent et évoluent avec le besoin du métier.
- Au sein du catalogage, impératif d'indiquer la **propriété des données et les sources maîtresses** afin d'assurer le pilotage qualité.
- En fonction du profil métier de l'utilisateur, **définir ses accès** par jeux de données.
- Adopter d'un **glossaire métier** commun.
- **Engager voire responsabiliser les métiers** sur les données à utiliser par cas d'usages.
- Imposer la mise à disposition des données, data lineage, cataloguing comme **systématique** et non par opportunisme.
- **Automatiser** de la documentation via l'ingestion des métadonnées.

DATA GOUVERNANCE (2/2)

Une approche « by design », data avant, pendant et après les projets



Construction

- Favoriser une **co-construction** avec la DSI/DIWO pour garantir les bonnes pratiques et la sécurité.
- Définition et partage plus global des **rôles et responsabilités** des acteurs data.
- **Communiquer** sur l'utilisation de l'univers BO pour le passage à l'échelle.
- **Synchronisation des équipes** sur les développements data pour réduire le time to market
- Un cas d'usage, s'il doit être industrialisé doit passer par une **Data Factory** (ex. BO).
- Permettre l'utilisation **d'environnements de dev avec des données** métiers.
- Mettre en place / définir des **contrats de pattern** réutilisables

Exposition

- Mettre en place une **checklist** pour le passage en prod, et intégrer les updates pour de nouvelles librairies.
- Intégrer dans le contrat d'usage une partie **sur l'exposition des données**, ainsi que leur **industrialisation** pour responsabiliser les métiers.
- Assurer la **mise en qualité** des données avant leur exposition.
- Définition et mise à disposition d'un **catalogue de service** au travers d'espaces collaboratifs tels que SharePoint.
- **CI/CD complète** sur l'ensemble de l'environnement

DATA ENGINEERING (1/2)



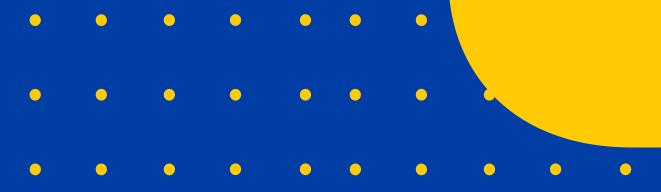
Solution 1

- Utiliser des **outils existants** plutôt que des solutions personnalisées pour réduire les coûts de développement et profiter des fonctionnalités avancées offertes.
- **Valoriser** l'utilisation de l'IA générative.
- Utiliser des outils tels que **Collibra** et **Qualitadd**, qui offrent des fonctionnalités étendues pour la gestion et l'analyse de données, la gestion des flux, la détection d'anomalies ou encore l'optimisation des processus de qualité des données grâce à leurs fonctionnalités avancées d'IA.

Solution 2

- Créer un **outil personnalisé** basé sur CFT (cross/continuous File Transfer) et utiliser un datalake pour traiter les données sous forme de topics (Kafka)
- Mettre en place une équipe "**Digital Factory**" dédiée à la gestion des données.
- Développer un système basé sur **Java ou Scala** pour le traitement des données en batch ou utiliser de frameworks tels que **Apache Spark** pour le traitement des données à grande échelle.
- Utiliser des instances de **Confluence** (Jira) afin de faciliter le partage de code et la collaboration
- Mettre en place **un jeu de tests de production** pour tester les outils et utiliser des outils comme Datastage pour l'ingestion des données.

DATA ENGINEERING (2/2)

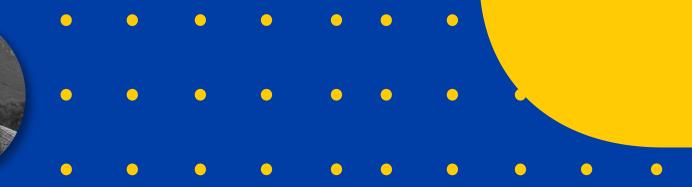


Solution 3

- Adopter une architecture **Lambda** pour les socles batch ou streaming, et envisager une architecture **Kappa** pour une approche hybride batch/streaming.
- Utiliser de la **virtualisation des données** à partir de sources diverses (Denodo, Staburst) afin de créer un hub centralisé pour源源er les données.
- Tenir compte des défis tels que la **réPLICATION DES DONNÉES** et la gestion de la **qualité** des données.
- Utiliser des outils de **data profiling** pour évaluer la qualité des données et mettre en œuvre des stratégies appropriées pour l'historisation des données.

Solution 4

- **Documenter** tous les aspects techniques du projet, y compris l'architecture, les flux de données, les schémas, les transformations, etc.
- **Développer des mécanismes flexibles** pour gérer les flux de données en fonction des besoins spécifiques du projet, tels que l'utilisation d'outils ETL (Talend, Informatica, SSIS)
- Mettre en place des **équipes dédiées** à l'observabilité, au monitoring et à l'orchestration des ingestions de données
- Concevoir un **schéma de base de données** dès le début du projet en tenant compte des besoins spécifiques.
- Exploiter les **services cloud** pour leur scalabilité, disponibilité et facilité de déploiement, ainsi que les APIs pour permettre une intégration facile avec d'autres systèmes.
- Réduire les efforts de déploiement en automatisant les processus avec des outils de **CI/CD**
- Utiliser des bibliothèques comme DataFaker pour générer des **jeux de données fictives réalistes**.



1. Gestion et acquisition de la donnée

- Extraction de la donnée (Source, disponibilité, qualité, volume)
- Définition du besoin en donnée labélisée (récurrent ou pas, etc)
- Envisager séparation apprentissage/test/validation
- Symétrie code et environnement expérimentale et opérationnelle

2. Expérimentation/modélisation

- Choix du type du ou des modèles (plusieurs composants possible)
- Préparation de l'infra IT. (approche distribuée ou non, CPU/GPUs, etc.)
- Recherche des features/caractéristiques pertinentes
- Définition des metrics de performance (technique/business)
- Metadata tracking, stockage artifacts

3. Déploiement

- Choix de l'approche « serving » : Batch, API (Microservices) ou embarquée (mobile)
- Gestion du code : Stockage, modularisation
- Type de déploiement : Juste le modèle ou le training pipeline aussi ?
- Fréquence de déploiement

4. Monitoring (performance)

- Contrôle écart du schéma de donnée, écart de valeurs de données
- Retraining (automatisée via CI/CD), fréquence régulière suivant l'acquisition de nouvelles données labélisés ou concept drift ou adhoc.
- Transparence et interprétabilité

IA/ML/LLM OPS (2/2)



Concept vu pendant l'atelier :

Trois niveaux de maturité sur le MLOps:

Niveau 0: Modèle construit par les DS mais déployé par les Data engineers.

Niveau 1: Construction d'un pipeline de génération de modèle (training pipeline avec déploiement automatisé)

Niveau 2: Automatisation du modèle généré en production (CICD) avec monitoring de performance

Outils vus pendant l'atelier :

Outil pour le **tracage de la donnée** (Data version control):

- Pachyderm ([Home Page | Pachyderm](#))
- Dvc ([Data Version Control · DVC](#))

Outil pour **l'expérimentation**

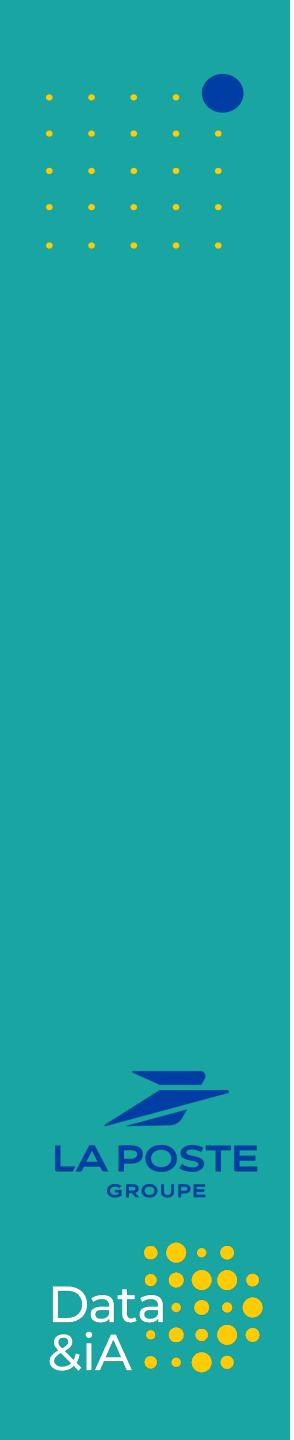
- Mlflow ([MLflow - A platform for the machine learning lifecycle | Mlflow](#))

Outil pour le **déploiement automatisé**:

- Seldon core ([Seldon Core — seldon-core documentation](#))

Retranscription des notes par branches et par ateliers

Séquence de l'après-midi



A l'issue de chaque séance de réflexion collective nous avons demandé à chaque participants de réfléchir à l'implication des échanges sur leur propre branche et entité.

Vous trouverez dans les planches ci-dessous la synthèse de ces notes groupées par branche et par atelier.

Appropriation Groupe

<p>Équivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i></p> <p>Projet de pilotage commercial PAG</p>	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p> <p>PO, métiers, architectes, DM Data stewart + Data Owner -> assure unicité ce qui est à la source</p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p> <p>Accélérer le partage des données</p> <ul style="list-style-type: none"> - Data Catalog à rendre userfriendly (LLM?) - Comité de validation des échanges : DPO, juriste, direction data groupe (data office) 	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <ul style="list-style-type: none"> - Mettre en place une data marketplace interne - Passer du data catalog au data marketplace <ul style="list-style-type: none"> - produit par produit, en prenant ce qui est mature - vue sur les 1ers produits pr démontrer ce qui marche - Métriques de consommation - Expés chiffrement/anonymisation/habilitation - Datascore, certification : producteur, bien décrite, homologuée, RGPD - Organisation centralisée au niveau du groupe pour l'acquisition externe <ul style="list-style-type: none"> - 1 gouvernance à mettre en place (pilotage, achats, licences, etc.) - Domaine transverses : opendata pour tous avec des contrôles de qualité et finalité
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p>	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p> <p>Avancer dans la mise en place de data products</p>

<p>Équivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i></p> <p>Banque d'ingestion du datalake</p>	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p> <p>Management Groupe (Pole Data, DIWO, PD360)</p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p> <ul style="list-style-type: none"> - Question du make or buy (ex : solution open source ingestion) - Bande passante nécessaire pour toutes les solutions 	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <ul style="list-style-type: none"> - Contrôles de type / format - Pilotage/Monitoring + observabilité / orchestration plus reprise - Moteur d'ingestion - Environnement coding hors "dataiku" - Documentation - ETL - Utilisation de l'IA pour la génération du code à injecter <p>Dev-ops -> Tests auto : Comment la "REP" devient un non évènement</p>
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p> <p>Stratégie groupe difficile à prendre (Groupe, BSCP, BGPN)</p>	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p> <p>Mise en place d'un POC</p>

<p>Equivalence dans ma branche</p> <p><i>Quels sont les cas applicables à ma branche ?</i></p> <p>Equipe Probayes, Openvalue Tous les sujets ML Mise en place d'un niveau 0 du Mlops - Choses déjà faisables avec Dataiku</p>	<p>Acteurs</p> <p><i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p> <p>Equipe Probayes, Openvalue DE/DS/DEVOPS/Archi Infra</p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche</p> <p><i>Qu'est-ce qui me permet d'avancer ?</i></p> <ul style="list-style-type: none"> - Acculturation aux étapes de construction de modèles - Contraintes machines / Bonnes pratiques 	<p>Nouvelles idées issues de l'atelier</p> <p><i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <ul style="list-style-type: none"> - Standardisation des modèles - Maturité du déploiement 3 niveaux de Google (0, 1, 2) <ul style="list-style-type: none"> - Niveau 0 : DS -> DE - Niveau 1 : standardisation + reproduction selon les normes - Niveau 2 : cycle du devis du modèle - Briques de supervision des données ou de modèles - Déploiement -> pas que du batch, aussi API Rest embarqué - Plus globalement mise en place d'un framework MLOps
<p>Freins à la résolution du cas dans ma branche</p> <p><i>Qu'est ce qui m'empêche d'avancer ?</i></p> <ul style="list-style-type: none"> - Mise à dispo des données - Implication des data ops dans le projet IA - Data scientist isolé - Mettre à disposition les environnements nécessaires aux équipes DS 	<p>Actions nécessaires à la mise en place de ces nouvelles idées</p> <p><i>Quelles sont les grandes étapes ?</i></p> <ul style="list-style-type: none"> - Diffuser la culture MLOps auprès des Data Scientists

<p>Équivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i></p> <p>Data catalogue DD Offre self-service Phases « Lab » Datahub de la CNP</p>	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p> <ul style="list-style-type: none"> - Process usage = Emergence -> qualif -> instruction-> réalisation - Qui ? Le métier, le data office, les experts self-services, les architectes si non standard
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p> <ul style="list-style-type: none"> - Le catalogue de données avec la vraie interaction entre le catalogue d'objet / attributs métiers et les métadonnées des tables du DLG - Mise en place "best practices / transfert de connaissances DSI -> Métiers (normes, outil...) 	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <ul style="list-style-type: none"> - Engagement des métiers sur : <ul style="list-style-type: none"> - Le cas d'usage - Les données, les sources de données, les KPIs - Relecture de code par la DSI - Data lineage, data catalog pas que par opportunisme - Documentation partagée (DSI...) - Objectifs dans les fiches de poste - Acculturation sur le IN/OUT usages / règles du DLG - La mise en application du contrat d'usage jusqu'au dernier KM = rédaction et signature du contrat, venue périodique, checks réguliers conformité
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p> <p>Des process assez lourds qui font intervenir de nombreux acteurs</p>	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p> <ul style="list-style-type: none"> - Réduire le nombre d'acteurs mais les sanctionner "...« - Scripter / automatiser le plus possible le process usage

Appropriation

LBP

<p>Equivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i></p> <p>ESG</p>	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p> <p>Comité de validation des échanges</p>	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <p>Ex. fichier COFACE</p> <ul style="list-style-type: none">- Vision d'ensemble acquisition externe / Organisation centralisée<ul style="list-style-type: none">- achat x fois & licensing- Mettre du LLM sur Collibra ou Confluence (Poc côté BSCC)- Data Score pour scorer le Confluence / data catalogue
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p> <p>Modèle Collibra</p>	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p>

<p>Équivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i></p> <p>Supervision, monitoring Moteur d'ingestion</p>	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p> <p>OBS, métiers</p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p> <ul style="list-style-type: none">- Quels OKR, besoin autour de l'ingestion- Data virtualisation sur les SIO	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <ul style="list-style-type: none">- Socle spécifique pour chaque type d'ingestion : batch- bibliothèque de code
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p> <ul style="list-style-type: none">- Supervision- Ne veulent pas une connexion directe aux SIO de prod	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p>

<p>Équivalence dans ma branche</p> <p><i>Quels sont les cas applicables à ma branche ?</i></p> <ul style="list-style-type: none"> - référentiel des cas d'usages - Déploiement data gouvernance - Data catalog - Marketing (consolidation des données clients de toutes les filiales) 	<p>Acteurs</p> <p><i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p> <ul style="list-style-type: none"> Comex Data Office, Métiers, DSI, Directoire/Comex
<p>Leviers d'accélération pour la résolution du cas dans ma branche</p> <p><i>Qu'est-ce qui me permet d'avancer ?</i></p> <ul style="list-style-type: none"> - Déployer l'ensemble de la data gouv - Support Comex - Datalabs - Nettoyage des usages - Normer les expositions (patterns) - Optimisation des environnements amonts - Data quality - Mise à jour itérative des data domains 	<p>Nouvelles idées issues de l'atelier</p> <p><i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <ul style="list-style-type: none"> - CI/CD sur périmètre - Mise en place d'un "contrat d'usage" des données d'un entrepôt -> cf Olivier Moraglia <ul style="list-style-type: none"> - données utilisées, utilisateurs, valorisation ou référent d'usage - Enregistré dans Harmada -> Quid de la MAJ du "contrat" d'usage ? Durée de la vie du contrat ? - Acculturer les métiers et les data scientist aux contraintes de la DSI - Checklist de la DSI à destination des DS Squad industrialisation
<p>Freins à la résolution du cas dans ma branche</p> <p><i>Qu'est ce qui m'empêche d'avancer ?</i></p> <ul style="list-style-type: none"> - Avoir des data stewards à temps plein dans chaque entités métiers / sans data domain - Data lineage - Déploiement partiel de la data gouvernance - Connaissance partielle de l'outil de "cataloging" - Offre de service potentiellement définis & partiellement déployés 	<p>Actions nécessaires à la mise en place de ces nouvelles idées</p> <p><i>Quelles sont les grandes étapes ?</i></p> <ul style="list-style-type: none"> - Accélérer la qualité du meta données - CI/CD à mettre en place - Demander à la DSI d'établir une checklist <ul style="list-style-type: none"> - Immersion des DS avec la DSI (lj) - Définition d'un contrat d'usage co-construit avec la DSI et les métiers - Intégration de la méthodologie projet du groupe - Revue et contrôles des contrats d'usage

Equivalence dans ma branche

Quels sont les cas applicables à ma branche ?

Ensemble des UC LPB

Acteurs

Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?

- Data scientists
- Validation des modèles/contrôle 2nd niveau
- DSIBA : accès aux données + infras pour faire tourner les modèles en prod)
- (Probayes ?)

Leviers d'accélération pour la résolution du cas dans ma branche

Qu'est-ce qui me permet d'avancer ?

- Data product -> feature store
- Environnement d'apprentissage miroir de l'environnement ops

Nouvelles idées issues de l'atelier

Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?

- Monitoring
- S'assurer que les modèles construits en "dev" soient reproductibles en production
- Différents niveaux de maturité des MLOps
 - N0 DS fournit le modèle et le pousse à la DSI pour déploiement
 - N1 Pipeline de données pour construire les modèles avec data engineer
 - N2 Automatisation de l'apprentissage

Freins à la résolution du cas dans ma branche

Qu'est ce qui m'empêche d'avancer ?

- Hétérogénéité des SIO
- Dispose-t-on des infrastructures suffisantes pour faire tourner les modèles ?

Actions nécessaires à la mise en place de ces nouvelles idées

Quelles sont les grandes étapes ?

- Socle Datahub
- CI/CD
 - Resultat batch
 - Model API

Appropriation

BGPN

<p>Equivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i></p>	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p>	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <ul style="list-style-type: none">- Moteur d'ingestion- Pattern d'ingestion à partager avec les acteurs- Monitoring des flux d'ingestion
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p>	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p>

<p>Équivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i></p>	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p>	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <ul style="list-style-type: none">- Data Management en émergence puis identification des données- Créer des solis dans les projets pour découvrir les données- ROI des cas d'usages- Travailler sur l'exposition- Connaitre les liaisons entre les différents domaines de données au niveau du SI
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p>	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p>

<p>Équivalence dans ma branche</p> <p><i>Quels sont les cas applicables à ma branche ?</i></p> <ul style="list-style-type: none"> - Digitalisation des factures - Détection des moments de vie - Problème de la qualité de la donnée à l'entrée - Problème d'identification des données critiques qui viendraient de références 	<p>Acteurs</p> <p><i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p> <ul style="list-style-type: none"> - Métiers - IT pour la partie solution
<p>Leviers d'accélération pour la résolution du cas dans ma branche</p> <p><i>Qu'est-ce qui me permet d'avancer ?</i></p> <ul style="list-style-type: none"> - Modèle de machine learning - Définition du process d'interaction entre les acteurs 	<p>Nouvelles idées issues de l'atelier</p> <p><i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <ul style="list-style-type: none"> - Symétrie des environnements - Drop du modèle - Déploiement (batch vs API vs embarqué) - Monitoring - Comment au sein de la BGPN le DM intervient ? - Features engineering / features stores
<p>Freins à la résolution du cas dans ma branche</p> <p><i>Qu'est ce qui m'empêche d'avancer ?</i></p> <ul style="list-style-type: none"> - Organisation / process - Interaction entre les équipes - Outilage -> qualité / observabilité (lineage / qualité) 	<p>Actions nécessaires à la mise en place de ces nouvelles idées</p> <p><i>Quelles sont les grandes étapes ?</i></p>

Appropriation

CNP Assurance

<p>Équivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i></p> <p>Partages donnés avec LBP (banque & assurances) Données open (DIALOG -> Université)</p>	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p> <p>Data marketplace</p>	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <ul style="list-style-type: none">- Documenter le langage / Les sources au niveau "object métier"- Renfort Data (métier) vs Data manager (fait le catalogue) par projet ou certifié dans le data office
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p>	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p> <p>Enrichir le data catalogue</p>

<p>Équivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i></p> <ul style="list-style-type: none">- Ingestion des fichiers partenaires- Migration Data stage (le cas échéant)	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p> <ul style="list-style-type: none">- La DSI
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p>	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <ul style="list-style-type: none">- Framework d'ingestion clef en main, basé sur des template d'ingestion
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p> <ul style="list-style-type: none">- Délais d'acquisition d'une nouvelle solution	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p>

Équivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i>	Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i>
Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i>	Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i> Add ML-Flow
Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i>	Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i>

Appropriation

BSCC

<p>Equivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i></p> <p>Data score : niveau de doc sur les datasets</p>	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p> <p>Data Nova (open data groupe)</p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p> <ul style="list-style-type: none">- Gouvernance de l'achat/optimisation des données niveau groupe- Data catalogue (Collibra)- Data domaine traverse<ul style="list-style-type: none">- achat, briques de données externes, exposition = fiches api- Certification des données : Qui ? Description ? Homologation ? Contrôle altération, réglementaire	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p>
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p> <p>Visibilité des acquisitions de la donnée sur l'ensemble de la branche / du groupe</p>	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p>

<p>Equivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i></p>	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p> <p>DSI/DS/OPS/métier</p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p>	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <p>Pipeline de déploiement</p>
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p>	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p> <ul style="list-style-type: none">- Gérer le drift- AutoML

Appropriation

DIWO

<p>Équivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i></p> <ul style="list-style-type: none">- Datalake Groupe	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p> <ul style="list-style-type: none">- Qualité des données sources- CI/CD + VRAIE préprod + Test AVANT PROD validée	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p>
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p> <p>Silotage métier / Data office / DSI ROI : Data Catalogue</p>	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p> <p>Formation et sensibilisation du client</p>

<p>Équivalence dans ma branche</p> <p><i>Quels sont les cas applicables à ma branche ?</i></p> <p>Déploiement ML dans OneDLG</p> <ul style="list-style-type: none"> - Environnements ? Entraînements vs Prod vs Tech - Populations ? - Source de données ? - Symétrie expérimentale vs opérationnelle ? 	<p>Acteurs</p> <p><i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche</p> <p><i>Qu'est-ce qui me permet d'avancer ?</i></p> <ul style="list-style-type: none"> - MLOPS : <ul style="list-style-type: none"> - Niveau 0 : DS vers DVC - Niveau 1: pipeline d'apprentissage et iteration rapide - Niveau 2 : evolution du modèle + CI/CD complet sur la base de metrics - MLFlows : « DVC » fit pour la donnée 	<p>Nouvelles idées issues de l'atelier</p> <p><i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <ol style="list-style-type: none"> 1. Acquisition : volume, data pipeline 2. Expérimentale : choix du modèle 3. Déploiement : batch, ou API ou embarqué 4. Monitoring : profit, performance
<p>Freins à la résolution du cas dans ma branche</p> <p><i>Qu'est ce qui m'empêche d'avancer ?</i></p>	<p>Actions nécessaires à la mise en place de ces nouvelles idées</p> <p><i>Quelles sont les grandes étapes ?</i></p>

<p>Équivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i> DataHub, Datalake, ingestion de sources chaud à froid</p>	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p> <ul style="list-style-type: none">- Framework<ul style="list-style-type: none">- Code générique ou code généré (dataiku + lighthouse)	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p> <p>5 typologies de flux différent</p> <p>IDEE : Data virtualisation mais effort de modélisation par le OWNER</p> <ul style="list-style-type: none">- data quality- Historisation- consistency
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p>	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p>

Appropriation Openvalue

<p>Équivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i></p> <p>Présent chez plusieurs clients</p>	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p> <p>Consultants et clients</p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p> <p>Sensibilisation des consultants</p>	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p>
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p> <p>Propre aux clients</p>	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p> <p>Formation et sensibilisation du client</p>

<p>Équivalence dans ma branche <i>Quels sont les cas applicables à ma branche ?</i></p> <p>Plusieurs missions</p>	<p>Acteurs <i>Qui impliquer dans la mise en œuvre, les décisions, les rituels... ?</i></p>
<p>Leviers d'accélération pour la résolution du cas dans ma branche <i>Qu'est-ce qui me permet d'avancer ?</i></p> <ul style="list-style-type: none">- Plus de clients- Plus de consultants formés	<p>Nouvelles idées issues de l'atelier <i>Qu'est-ce qui m'a inspiré(e) que je voudrais essayer ?</i></p>
<p>Freins à la résolution du cas dans ma branche <i>Qu'est ce qui m'empêche d'avancer ?</i></p>	<p>Actions nécessaires à la mise en place de ces nouvelles idées <i>Quelles sont les grandes étapes ?</i></p>

**Vos retours sur la cellule de veille data
et ses activités sont essentiels !**

**Merci de nous donner votre avis si
vous ne l'avez pas déjà fait**



**Merci à tous pour votre
participation !**

Rendez-vous à la rentrée...