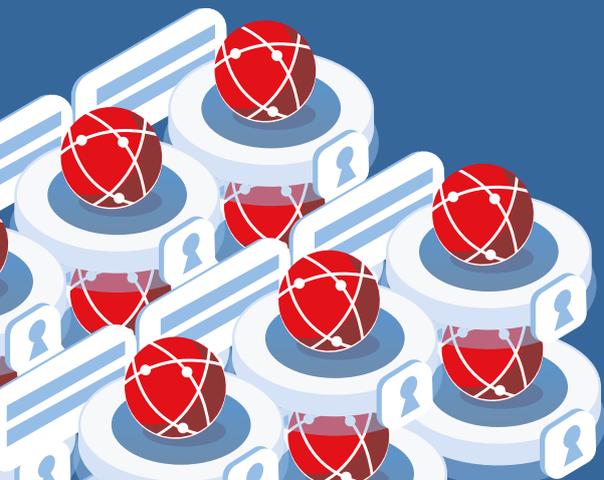


Synot^{is}

A SMILE GROUP COMPANY

— **Data Mesh : une réponse fédérative au silotage des données**



Sommaire

Sommaire	2
Qui sommes-nous?	4
Introduction	6
1. Les enjeux	9
1. Des architectures techniques rigides	10
2. Un manque d'agilité	11
3. Un manque de confiance	12
4. Un besoin d'implication des métiers	13
5. Un besoin de traçabilité de l'information	14
2. Comprendre le Data Mesh	17
1. Qu'est-ce que le Data Mesh?	18
2. Les principes fondamentaux du Data Mesh	19
3. Notre Vision	27
1. Data Mesh Factory	28
4. Data Mesher's Journey	53
1. Georgio a besoin de nouvelles données	54
2. Patrick, demande la création d'un nouveau data product	56
Vous avez encore des questions ?	58

Voici une **DATA**.



Découvrez le parcours d'une donnée selon synotis à travers toutes nos illustrations.

Qui sommes-nous?

Fondée en 2014, Synotis CH est une entreprise spécialisée dans la réalisation de projets Data, visant à accompagner les organisations dans leur transformation digitale. Notre objectif est de vous guider vers une démarche « Data Driven », en mettant en lumière la valeur de vos données. Pour ce faire, nous avons développé une offre de services basée sur deux Business Lines étroitement liées.

Notre première Business Line, « **Conseil** », intervient en amont de vos projets pour vous offrir un soutien stratégique et tactique, essentiel à la réussite de votre transformation. Nos experts travaillent main dans la main avec vous pour élaborer une vision claire et vous aider à concrétiser vos objectifs.

En complément, notre seconde Business Line, « **Data Management** », intervient dans la mise en œuvre concrète de vos projets en apportant une expertise technique pointue. De l'accompagnement au choix d'architecture à la réalisation, en passant par la conception, nos équipes se mobilisent pour garantir des résultats optimaux et une intégration harmonieuse de solutions data au sein de votre organisation.

Nos deux Business Lines collaborent étroitement, favorisant ainsi une culture du pragmatisme au cœur de nos actions, afin d'assurer un bon alignement entre les objectifs recherchés et la solution qui sera mise en œuvre. Par ailleurs, chez Synotis, nous mettons un point d'honneur à valoriser nos collaborateurs en adoptant les principes du flat management. Le bien-être de nos équipes est au cœur de nos préoccupations, ce qui se traduit par une stabilité remarquable de notre personnel.

Notre équipe, composée de plus de 40 consultants, incarne le sens du service par le biais de compétences variées et d'une expertise solide, tout en respectant scrupuleusement nos engagements envers nos clients.

Fort de plus de 10 ans d'expérience en Suisse romande, nous avons tissé des partenariats solides et durables avec nos clients dans de nombreux domaines d'activité, formant ainsi la base d'un développement progressif et constant.

Chez Synotis, nous nous engageons à fournir des solutions Data performantes et innovantes, contribuant ainsi à l'évolution et au succès de votre entreprise dans l'ère numérique en perpétuelle évolution.

L'équipe Synotis



Introduction

Si vous évoluez de près ou de loin dans le domaine de la donnée, il est probable que vous ayez entendu parler du **Data Mesh** et c'est peut-être même ce qui vous a amené à lire ces premières lignes. Avant que le concept ne soit récupéré par les équipes marketing des éditeurs de solutions, le concept de Data Mesh s'appliquait déjà à des enjeux bien réels que nous identifions depuis longtemps dans nos projets.

Le principe est simple : en augmentant le partage et l'accès à la donnée, en favorisant l'autonomie des équipes, nous accélérons le développement et l'émergence de nouvelles initiatives et usages autour de la donnée. Mais alors, pourquoi après le datawarehouse, DataLake et DataHub en sommes-nous arrivés au Data Mesh ? Pourquoi de nouvelles problématiques de partage de données donnent-elles naissance à des concepts aussi régulièrement et finalement ces concepts sont-ils si nouveaux que cela ?

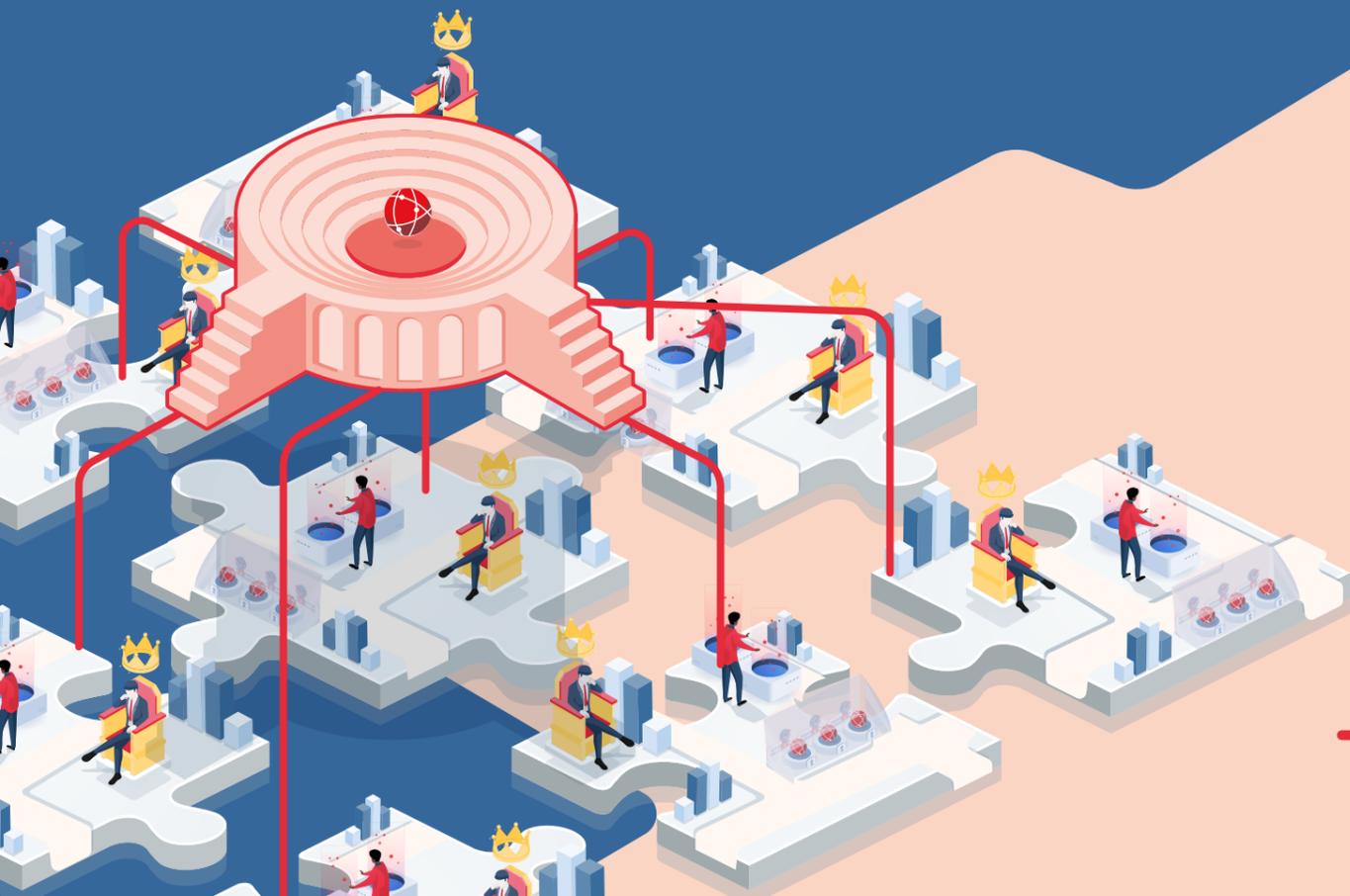
Car oui, le Data Mesh est avant tout un concept avec des principes et des approches mais il ne s'agit pas d'une solution technique. Autrement dit, il ne sera pas possible de simplement vous procurer un outil d'un éditeur pour faire du Data Mesh au sein de votre organisation, mais bien d'embrasser le sujet dans sa globalité en intégrant et sensibilisant vos utilisateurs dans un mouvement d'entreprise plus profond et global qu'un unique projet informatique.

Dans ce livre blanc, nous explorerons les **fondements du Data Mesh** et ses **avantages**, mais aussi les défis et **limites**. Nous analyserons également les meilleures pratiques pour implémenter une approche de Data Mesh efficace et durable. Notre objectif est de **vous donner une perspective critique et équilibrée sur le Data Mesh**, afin que vous puissiez décider si cette approche convient à votre organisation et comment vous pourriez la mettre en œuvre avec succès.



1.

Les enjeux



Synot^{is}

En tant qu'experts, nous ne pouvons l'ignorer, le data mesh répond à des enjeux que nous identifions souvent dans nos projets. Avant de vous dévoiler les fondements du Data Mesh, nous pensons qu'il est important de faire le point sur les enjeux qui ont amené à la naissance du concept.

Des architectures techniques rigides

Avec l'explosion des données et la multiplication de leurs sources, les systèmes d'informations sont devenus de plus en plus complexes. Les départements informatiques ont du mal à gérer efficacement toutes ces données, entraînant des retards dans les projets, des erreurs de traitement et une qualité de données souvent perfectible. En effet, au fil des années, les entreprises ont accumulé un nombre important de systèmes, d'applications, de bases de données et de technologies nécessaires pour gérer leurs données.

Ces systèmes ont souvent été développés par les équipes informatiques à partir d'initiatives indépendantes des métiers, et portent chacun une dette technique plus ou moins importante car il n'est pas toujours évident de faire évoluer ces systèmes de concert. En outre, la communication entre les différentes applications n'est pas toujours considérée comme une priorité lors du développement, ce qui complique la gestion des données globale et peut entraîner des coûts élevés pour leur adaptation.

Ainsi les entreprises contemporaines héritent d'un silotage tant au niveau de la conception des différentes applications que du stockage des données ce qui implique souvent des **niveaux de qualité variables**.

De plus, la **complexité** de ces systèmes d'information entraîne des problèmes de gouvernance, de qualité et de sécurité des données. Il est difficile pour les entreprises de savoir où se trouvent toutes leurs données, qui en est responsable, comment elles sont utilisées et qui y a accès. Ce manque de maîtrise amène des problèmes de conformité réglementaire (nLPD, GDPR...) et de risques liés à la sécurité des données.

Le concept de **Data Mesh** vise à **résoudre** cette **complexité** en créant une architecture distribuée où les données sont décentralisées et gérées par des équipes de domaines métier plutôt que par des équipes informatiques centralisées. Les équipes de domaines métier sont les plus à même de comprendre leurs données, et peuvent donc gérer les données de manière plus efficace et plus agile. Cela permet de réduire la complexité des systèmes d'information en éliminant les silos de données et en améliorant la collaboration entre les différents domaines métier.

Enfin, le Data Mesh a pour objectif de **réduire les coûts de gestion des données**, en évitant les redondances et en limitant les besoins en ressources humaines. Les domaines de données autonomes peuvent gérer leurs propres données de manière efficace, sans avoir besoin d'une équipe centrale dédiée à la gestion de celle-ci. Cela permet également de réduire le risque de perte de données ou de violation de la vie privée, car **chaque domaine de données est responsable de la sécurité et de la conformité de ses propres données**.

Un manque d'agilité

Nos clients sont confrontés à une concurrence accrue et doivent être en mesure de s'adapter rapidement aux changements du marché. Les départements informatiques traditionnels ne sont pas toujours en mesure de fournir des données pertinentes en temps voulu, ce qui entrave la prise de décision et la réactivité. Le Data Mesh permet une approche plus agile en donnant aux équipes métier la responsabilité de la gestion des données pertinentes pour leurs besoins spécifiques.

L'agilité dans la gestion des données est un enjeu majeur. Les approches traditionnelles de gestion des données sont souvent trop rigides et peu adaptées à l'environnement en constante évolution des entreprises. Les processus de gestion de données traditionnels sont souvent longs et complexes, ce qui ralentit la prise de décision et la réactivité.

Le concept de Data Mesh aborde ce problème en proposant **une approche plus agile de la gestion des données**.

D'un point de vue humain : au lieu d'avoir une équipe centrale de gestion des données, le Data Mesh repose sur des «domaines de données» autonomes, qui ont la responsabilité de gérer leurs propres données en fonction des besoins de leur métier. Les équipes de développement internes à ces domaines peuvent ainsi accéder plus rapidement aux données dont elles ont besoin pour créer de nouveaux produits ou services à mettre à disposition de l'ensemble des consommateurs de l'entreprise. Cela permet d'innover plus rapidement et d'accélérer les prises de décisions.

D'un point de vue technique : en proposant une plateforme unifiée qui permet à tous les domaines de diffuser leurs données à travers une seule et même plateforme permet **d'éviter les architectures spaghetti**. La gestion de cette plateforme centrale revient à l'équipe IT, qui a la responsabilité de fournir l'infrastructure et les outils qui soutiennent la mise en œuvre du Data Mesh dans l'ensemble de l'organisation.

Le Data Mesh permet également une collaboration plus étroite entre les équipes de développement et les experts métier, ce qui améliore la qualité des données et la pertinence des analyses. En effet, les experts métier ont une connaissance approfondie des données qu'ils utilisent au quotidien, et peuvent apporter des éclairages précieux aux développeurs qui cherchent à créer de nouveaux produits ou services.

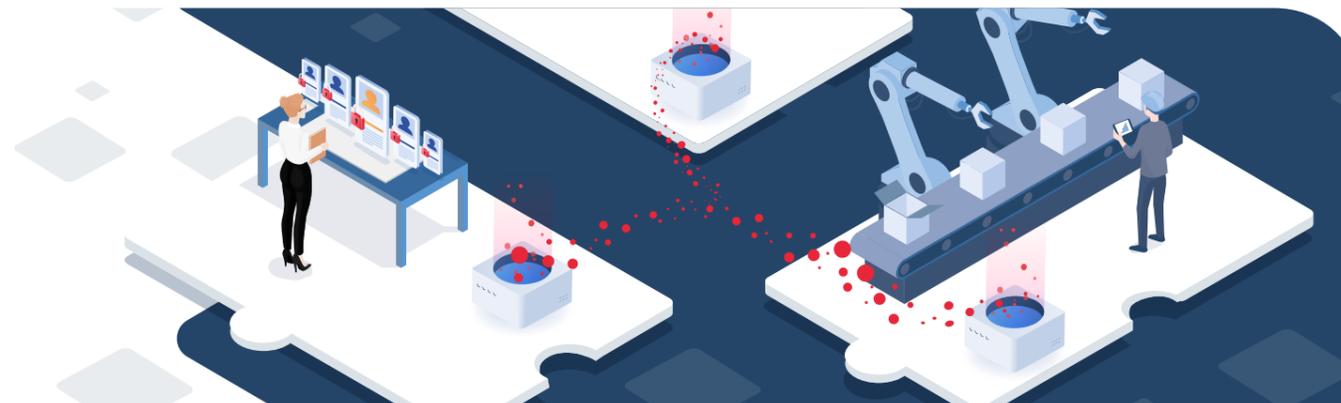


Un manque de confiance

La confiance dans les données est un enjeu majeur pour tous les consommateurs. En effet, la qualité des données qui leur sont fournies est essentielle pour prendre des décisions efficaces. Cependant, la confiance est souvent mise à mal par les écosystèmes en place, en particulier en raison d'architectures techniques et organisationnelles qui ne facilitent pas la gouvernance des données.

Historiquement, la **qualité** des données a toujours été **une préoccupation majeure des directions informatiques**, qui ont mis en place des leviers d'amélioration souvent techniques. Toutefois, cette tendance évolue aujourd'hui, avec la recrudescence de projets de gouvernance visant à **responsabiliser** les acteurs face à la qualité des données de leur périmètre. Malgré cela, ces évolutions ne sont pas suffisantes, et il est nécessaire de repenser les architectures techniques et organisationnelles des entreprises pour permettre une prise en compte plus efficace de la qualité des données.

C'est dans ce contexte que le Data Mesh émerge comme une réponse aux besoins croissants de consommation de données. Toutefois, pour assurer une consommation efficace et sécurisée des données, il est essentiel de garantir leur qualité et leur intégrité, afin d'éviter des risques de diffusion erronée vers l'extérieur et des décisions erronées en interne. En somme, **la confiance en les données est une condition sine qua non pour une institution digitalisée, performante et responsable.**



Un besoin d'implication des métiers

En qualité d'expert en données, nous savons que les départements informatiques ont traditionnellement été les principaux gestionnaires des données en raison de leur vision globale du système d'information de l'organisation. Les DSI ont une solide expérience en matière de gestion des interactions, de compréhension des spécificités des flux et d'acquisition progressive des connaissances métier. Toutefois, ils ne possèdent qu'une partie des éléments nécessaires pour maîtriser le capital «données» de l'organisation.

Les utilisateurs métier, quant à eux, doivent faire face quotidiennement aux inconvénients résultant des problèmes de partage de données, de qualité ou de l'absence de processus organisationnels clairs. Ils sont sans doute les plus touchés par ces insuffisances. Toutefois, ils sont également les mieux placés pour appréhender la réalité métier des données de manière précise.

De plus, de nombreuses autres parties prenantes s'orientent vers l'avenir et se préparent à des évolutions potentiellement importantes dans le traitement et la mise à disposition de l'information.

Or, nous avons constaté que les métiers ne sont pas assez impliqués dans les projets de gestion de données qui ne sont pas uniquement des projets informatiques mais bien des projets d'entreprises.

Le Data Mesh vise à rectifier cette lacune en confiant la propriété et la responsabilité des données aux divisions métier elles-mêmes, appelées «Data Domain» ou «domaine de données».

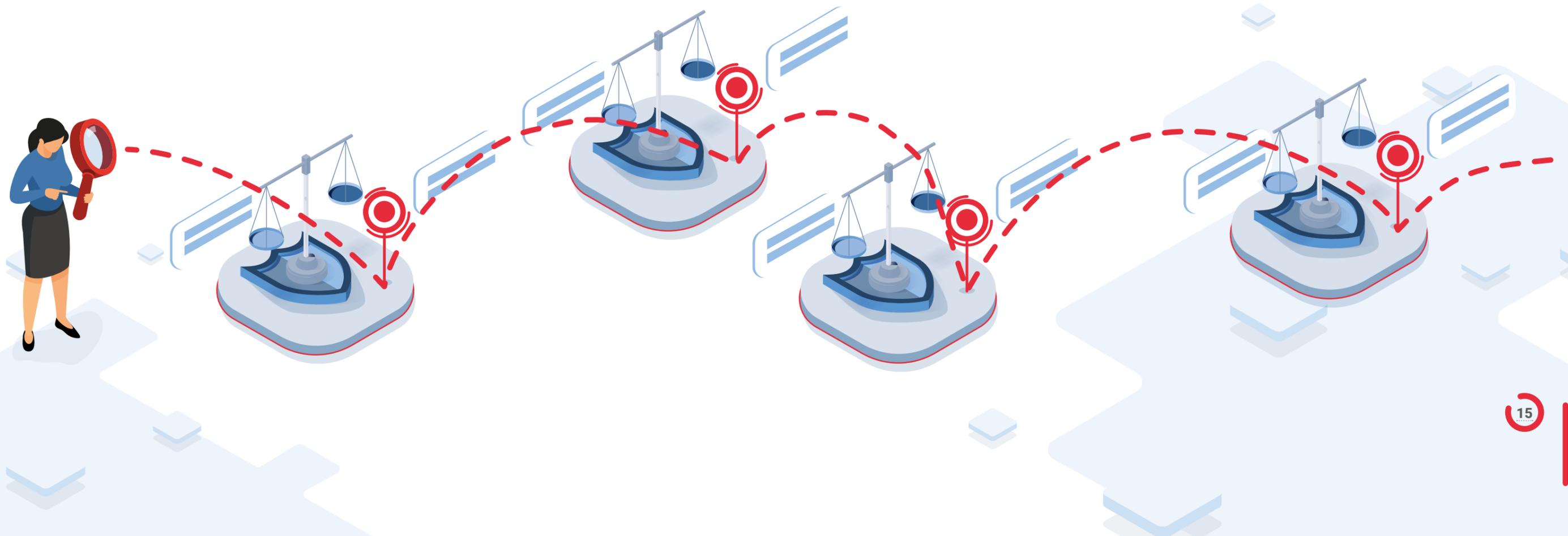
Un besoin de traçabilité de l'information

Les entreprises ont depuis longtemps le besoin de tracer les données et leurs accès. Mais de nos jours, la maîtrise du système d'information et la traçabilité sont devenues un impératif critique pour nos clients. Ceci est dû à l'introduction de **nouvelles réglementations** sur les données, qui **exigent des capacités d'audit** permettant de démontrer la conformité en toute transparence. Afin de se conformer à ces réglementations en toute confiance, les entreprises doivent être en mesure de suivre les données, leurs utilisations et leur cycle de vie complet.

Le besoin de traçabilité des informations et des accès ne doit cependant pas être guidé uniquement par le risque. Dans un système d'information pleinement intégré, les différents

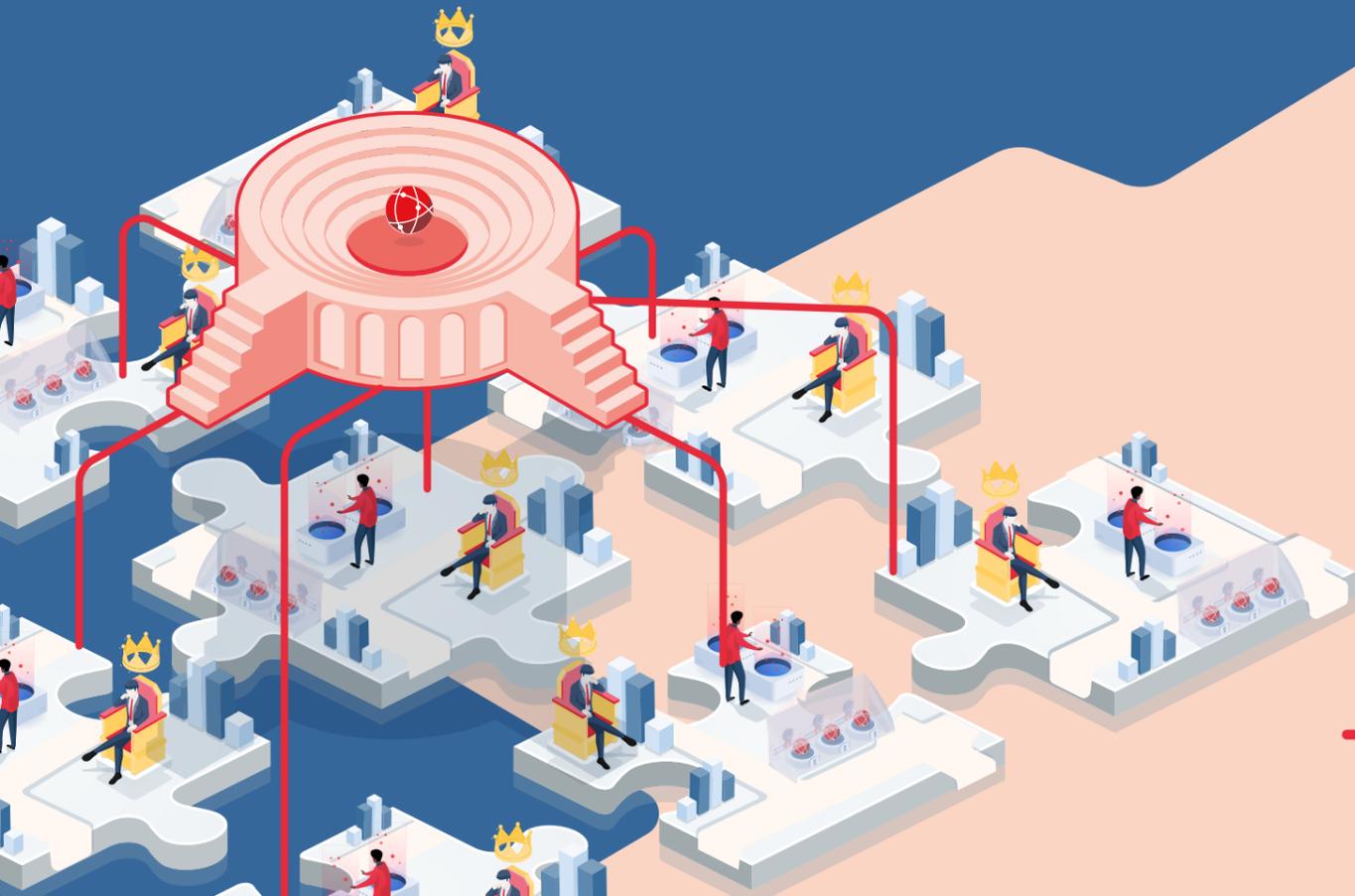
consommateurs, humains ou applicatifs, doivent être en mesure de vérifier l'origine et la pertinence des informations qui leur sont fournies. Pour ce faire, toujours dans cette optique de confiance évoquée précédemment, ils doivent être informés au fil de l'eau de toute modification apportée aux données.

Les architectures actuelles, telles que les architectures centralisées ou les architectures en **silos**, **ne sont pas toujours adaptées** pour répondre aux besoins de **traçabilité** des données et des accès. Ces architectures peuvent rendre la tâche très complexe, car il est souvent difficile de suivre les données lorsqu'elles sont stockées dans des systèmes hétérogènes. Le Data Mesh propose, à travers sa plateforme centrale et une gouvernance décentralisée de répondre aux enjeux de la traçabilité.



2.

Comprendre le Data Mesh



Synot^{is}

Qu'est-ce que le Data Mesh?

Le Data Mesh est une approche relativement récente et encore en évolution qui a été initialement proposée en 2018 par **Zhamak Dehghani**, architecte de données chez ThoughtWorks.

C'est une typologie d'architecture émergente qui permet aux entreprises de **gérer leurs données de manière distribuée, évolutive et résiliente**. L'objectif principal est de favoriser le partage des données entre les différentes équipes tout en leur permettant d'être autonomes dans la gestion de leurs données.

Comme décrit dans la précédente partie, le Data Mesh aborde les problèmes liés à la centralisation des données et à l'approche traditionnelle de leur gestion, où une équipe centrale est responsable de la collecte, du stockage et de l'analyse des données pour toute l'organisation. Cette approche implique une attention particulière aux performances, à la qualité des données et à la collaboration, ainsi qu'aux retards.

Ainsi, en permettant à chaque équipe de gérer ses propres données, le Data Mesh vise à **améliorer l'agilité de l'organisation, la qualité des données et à encourager l'innovation**. Le Data Mesh repose sur des composantes de Gouvernance fortes en prônant la **responsabilisation, la mise en place de normes de données communes, la documentation et le partage des connaissances**, afin de favoriser la collaboration entre les équipes et à participer à la rupture des silos organisationnels et techniques.

Le Data Mesh propose **quatre grands piliers** qui sont en adéquation avec la vision Synotis de la Gouvernance et du Data Management. Ils abordent des notions sous-jacentes d'agilité, de responsabilisation et de conception orientée service autour de la donnée.



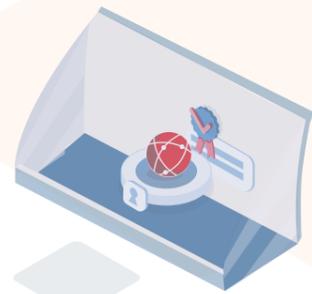
Les principes fondamentaux du Data Mesh



1. Domain Ownership

La notion de «Domain Ownership» est un pilier essentiel de l'architecture du Data Mesh. Elle précise que chaque domaine de données doit être sous la responsabilité d'un propriétaire clairement identifié. Ce propriétaire est ainsi chargé de la gouvernance des données qui relèvent de son domaine.

Un domaine de données est une agrégation logique de données qui peut être structurée autour de différents critères. Ces derniers peuvent être relatifs à un domaine organisationnel spécifique (finance, RH, etc.) ou encore à un concept précis (client, adresse, vente, etc.).



Le **propriétaire du domaine**, aussi appelé «**Data Owner**», généralement **entouré d'une équipe ou d'une personne disposant d'une expertise technique et fonctionnelle du domaine de données concerné**, est le garant de la qualité et de l'intégrité des données de ce domaine. Au-delà de cette responsabilité, le propriétaire se charge aussi de la documentation, de la mise en place des politiques de sécurité et de confidentialité, et de la gestion de l'accès aux données et de leur qualité.

Ce principe de propriété par domaine favorise l'autonomie et la responsabilité des équipes en charge des différents domaines de données. Il leur confère la possibilité de gérer leurs propres données de manière indépendante, facilitant ainsi la collaboration entre équipes et diminuant les conflits potentiels liés à la gestion des données.

De plus, cette notion de propriété par domaine encourage la **décentralisation** et la **prise de décision locale**, favorisant ainsi une **prise de décision plus rapide, plus efficace et plus agile** face aux changements du contexte d'entreprise. Le principe de «**Domain Ownership**» est intrinsèquement lié à celui de «**Federated Computational Governance**», qui sera décrit plus loin.

2. Data as Product

Le concept «Data as Product», ou «**Données en tant que Produit**», constitue un pilier fondamental du Data Mesh. Il marque une révolution dans la perception et le traitement des données, les considérant non plus simplement comme une ressource brute ou un **sous-produit** des applications informatiques, mais plutôt comme **des produits à part entière**, créés, développés, entretenus et mis à disposition pour répondre aux exigences des utilisateurs.

Le Data Mesh suggère que **chaque ensemble de données soit envisagé comme un produit distinct**. Chaque **domaine** se voit alors confier la responsabilité de développer, maintenir et valoriser ces produits, de la même manière qu'une entreprise le ferait pour un produit ou un service vendu à ses clients. Il incombe donc à chaque équipe associée à un domaine de collaborer étroitement avec les utilisateurs finaux afin de garantir que le data product soit de qualité, réponde à leurs besoins et soit accessible de manière fiable et cohérente.

De plus, le concept de «Data as Product» précise que les données doivent être traitées de façon à être facilement consommables par d'autres parties prenantes. Cela implique que les données soient découvrables, disponibles de manière transparente, compréhensibles grâce à une documentation claire, et conformes à des normes de qualité partagées par tous les domaines. Les utilisateurs doivent également avoir la certitude que ces données sont fiables, sécurisées et en accord avec les réglementations en vigueur.

En résumé, le principe de «Data as Product» incarne un changement de paradigme dans la perception et la gestion des données, en les traitant comme des produits en soi.

LE DÉCRYPTAGE DE SYNOTIS



- **Ajoute une autonomie** vis-à-vis de la gestion et du partage des données
- **Favorise l'agilité**. Les domaines de données sont à même de prioriser et réaliser leur «data product» en toute autonomie.
- Favorise une **meilleure lisibilité de la responsabilité** des données vis-à-vis des utilisateurs
- **Soutien à la construction d'un programme de gouvernance des données.**

- **Difficile de fixer les «limites»** des domaines de données au démarrage.
- **Identifier et sélectionner les bons responsables** peut être difficile.
- **Redistribuer certaines responsabilités** actuellement attribuées à la DSI peut être mal perçu.
- Nécessite des **compétences techniques** dans chaque domaine
- **Aligner les points de vue** différents entre les équipes qui produisent des données similaires ou qui ont des vues différentes sur la façon dont les données doivent être utilisées **afin d'éviter les «conflits»**.



LE DÉCRYPTAGE DE SYNOTIS



- **Favorise la perception de la data comme un actif**
- **Favorise l'augmentation de valeur** autour des données
- **Place la documentation et la description comme un besoin utilisateur**

- Si les Data Owners ne communiquent pas assez, ou si les domaines de données sont mal définis alors cela peut amener une **multiplication des data products similaires**.



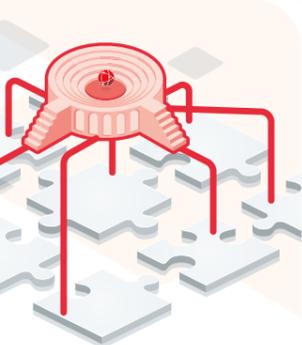
3. Federated Computational Governance

Le terme «Federated Computational Governance» se réfère à une approche de la gouvernance des données et de la gouvernance de l'informatique qui repose sur la collaboration et la coordination entre plusieurs **parties prenantes**.

Une fédération se définit par l'existence d'un État fédéral se superposant à des entités fédérées, selon **une organisation «à double étage»**. **Les fédérés restent souverains de leur domaine**, mais **délèguent certaines de leurs compétences à une instance commune**. Il en résulte donc **une indépendance des états fédérés, avec un objectif de cohésion sur certains aspects**.

Le principe de Federated Computational Governance est une application des principes de la fédération aux architectures de données. Dans cette approche, **la gouvernance est distribuée et décentralisée**, ce qui signifie que les différentes parties prenantes collaborent pour définir et mettre en œuvre les politiques, les processus et les normes qui régissent l'utilisation et la gestion des données. Mais les différentes parties prenantes conservent une certaine autonomie et une certaine responsabilité pour la définition de la gouvernance des données au sein de leur propre domaine d'activité, tout en travaillant ensemble pour garantir la conformité, la cohérence et la qualité de l'ensemble du système.

Le principe de Federated Computational Governance est une approche de la gouvernance qui permet aux parties prenantes de travailler ensemble de manière coordonnée pour gérer les données et les systèmes informatiques, tout en préservant leur autonomie et leur responsabilité dans leur propre domaine. Cette approche favorise la collaboration et la transparence, tout en garantissant une gouvernance efficace et adaptée aux besoins de l'organisation.

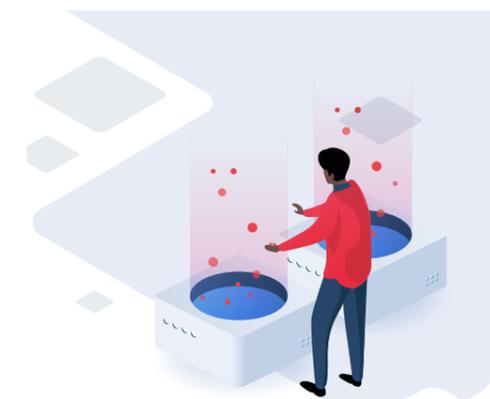


4. Self-service platform

Dans le contexte du Data Mesh, une «self-service platform» (plateforme en libre-service) est une plateforme de données qui permet aux équipes de domaine de gérer de manière autonome leurs propres produits de données, sans avoir besoin de dépendre d'une équipe centrale de gestion des données.

L'objectif principal est de permettre aux équipes d'être autonomes pour qu'elles puissent gérer leurs propres produits de données, de la conception à la production, en réduisant leur dépendance à l'égard des équipes centrales de gestion des données et en leur permettant de se concentrer sur les besoins spécifiques de leurs utilisateurs. Cela signifie que les équipes de domaines disposent d'un accès facile et convivial à une infrastructure et des outils nécessaires pour gérer leurs propres produits de données, tels que le stockage de données, les pipelines d'ingestion, les services de traitement et d'analyse des données, les outils de gouvernance des données, etc. Cette plateforme leur permet de créer, déployer et gérer leurs produits de données de manière autonome et tout en répondant aux besoins spécifiques de leur domaine.

La plateforme en libre-service est donc un pilier de l'approche du Data Mesh, nous décrivons plus en détail les composantes de cette plateforme.



LE DÉCRYPTAGE DE SYNOTIS



- Favorise la responsabilisation des équipes.
- Favorise la rupture des silos.
- La gouvernance fédérée favorise les déploiements flexibles et progressifs.

- Peut générer des **incohérences dans les pratiques d'entreprises**.
- Ce type de gouvernance peut être **inadapté dans les organisations très silotées**. En effet elle nécessite une communication et un partage fort.
- La gouvernance fédérée avec une orientation data domaine peut amener à des **difficultés à déterminer la propriété des objets**.



LE DÉCRYPTAGE DE SYNOTIS



- **Innovation**: Les échanges de métier à métier peuvent amener de nouvelles perspectives qui n'auraient pas été identifiées dans l'architecture actuelle.
- **Time to market clairement amélioré** car la latence liée à la disponibilité de l'équipe informatique est évitée.
- **Améliore la maîtrise des coûts** si les processus sont suffisamment autonomes et bien définis.
- **Facilite le travail des profils de type Data Analyst / Data Scientist**

- **Dépendance forte aux notions de responsabilité, sécurité et qualité de données**. Ces 3 aspects doivent être matures afin d'avoir un terrain fertile pour une approche self-service.
- **Danger sur l'homogénéité des technologies** utilisées par les data domain pour pousser dans la self-service platform.
- **S'assurer de la complétude de la documentation de la plateforme pour s'éviter une charge importante de support**.
- **Changement de paradigme pour l'équipe informatique en place**.



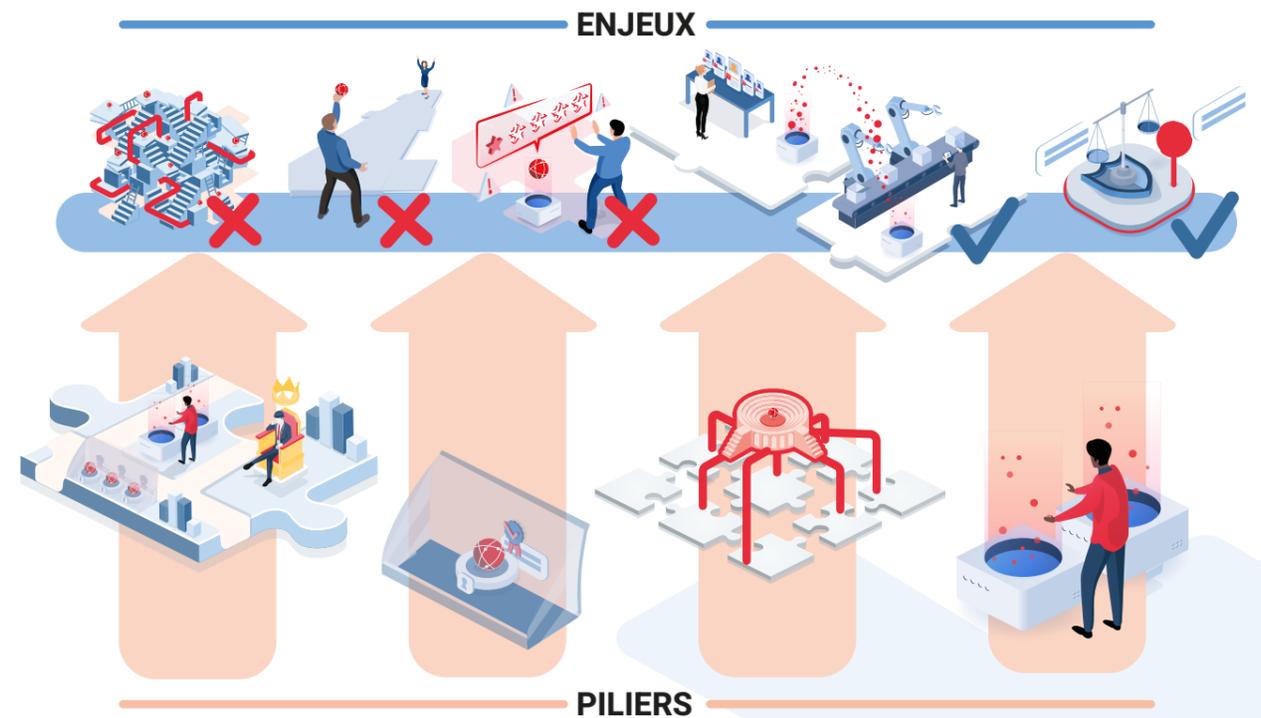
Notre Avis

Le Data Mesh représente une approche de gestion des données qui met l'accent sur **l'autonomie des équipes et la responsabilisation de chaque domaine métier** dans la gestion de leurs propres données. Le Data Mesh est **autant une architecture socio-organisationnelle que technique**, par conséquent elle nécessite un changement de paradigme important pour les entreprises qui choisissent de l'adopter. Nous notons notamment le **repositionnement du rôle de la DSI** pour favoriser la collaboration entre les équipes métier et techniques.

Le Data Mesh peut offrir des avantages considérables en termes de flexibilité, de réactivité et d'efficacité, mais il exige également un niveau de maturité élevé dans la gestion des données, ainsi qu'une mise en place progressive et gouvernée pour éviter des erreurs coûteuses. Il est donc important **d'avancer par petits pas** et de mettre en place les processus et les outils appropriés pour soutenir une architecture de données distribuée et coordonnée.

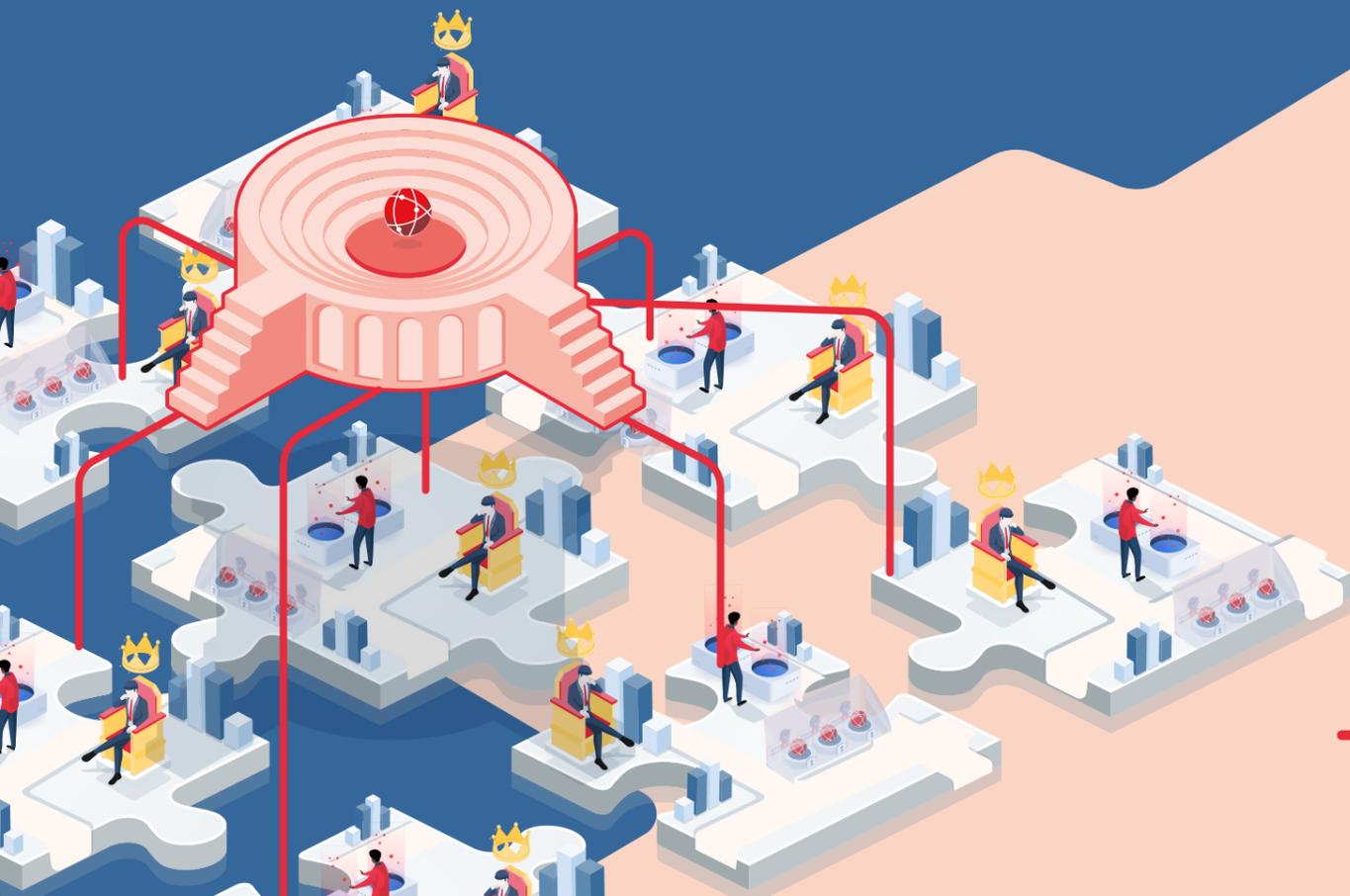
En fin de compte, **les avantages du Data Mesh peuvent largement l'emporter sur les inconvénients potentiels**, à condition que les entreprises abordent cette approche avec une attitude ouverte et un **engagement ferme envers la transformation de leur culture et de leur fonctionnement**.

Dans la section suivante, nous présenterons notre vision d'architecture organisationnelle et technique pour mettre en place un Data Mesh efficace au sein de votre entreprise. Cette approche est basée sur notre expérience en matière de gestion et de gouvernance des données auprès de nos clients, ainsi que sur une analyse des défis et des opportunités liés à cette approche.



3.

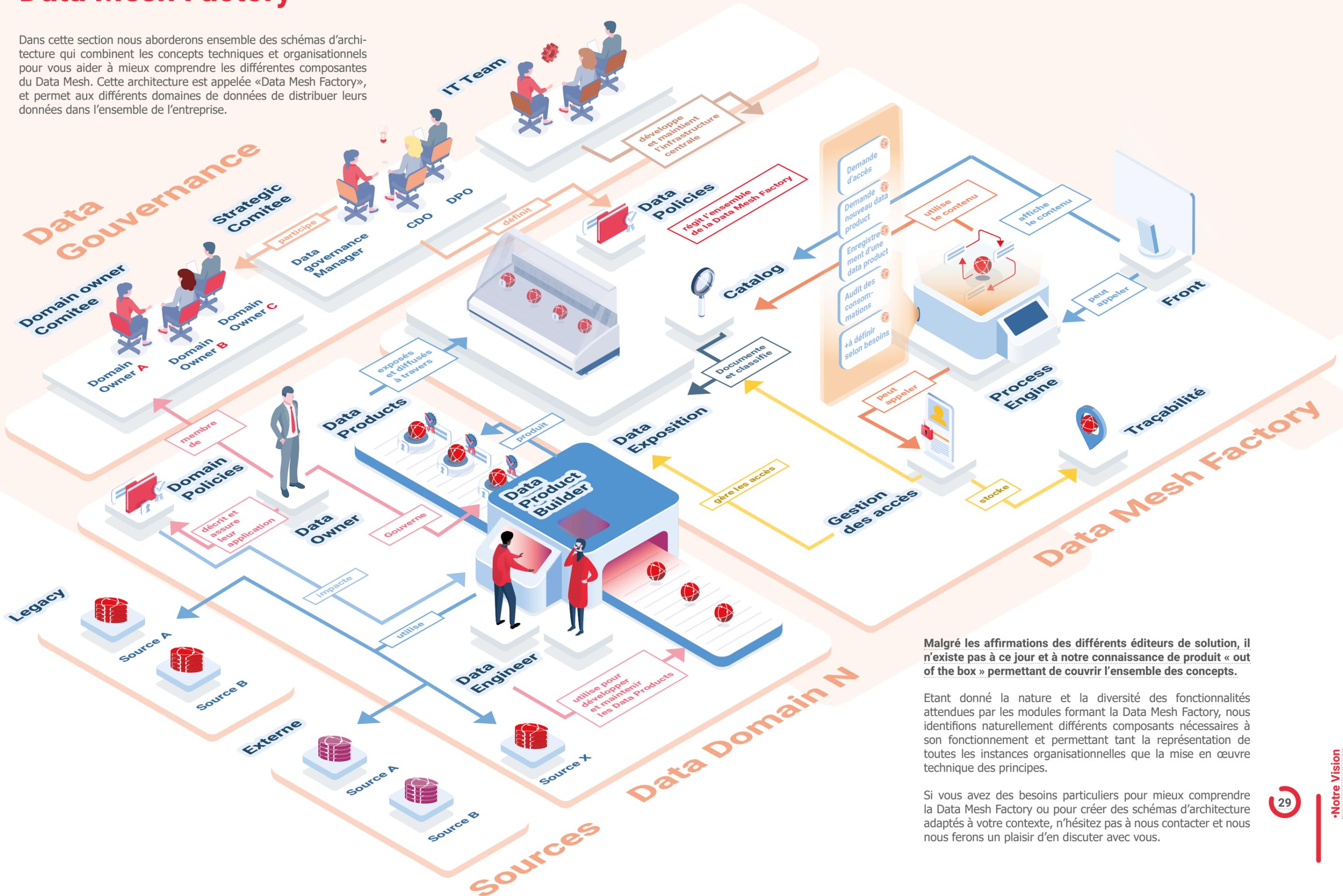
Notre Vision



Synotis

Data Mesh Factory

Dans cette section nous aborderons ensemble des schémas d'architecture qui combinent les concepts techniques et organisationnels pour vous aider à mieux comprendre les différentes composantes du Data Mesh. Cette architecture est appelée «Data Mesh Factory», et permet aux différents domaines de données de distribuer leurs données dans l'ensemble de l'entreprise.



Malgré les affirmations des différents éditeurs de solution, il n'existe pas à ce jour et à notre connaissance de produit « out of the box » permettant de couvrir l'ensemble des concepts.

Etant donné la nature et la diversité des fonctionnalités attendues par les modules formant la Data Mesh Factory, nous identifions naturellement différents composants nécessaires à son fonctionnement et permettant tant la représentation de toutes les instances organisationnelles que la mise en œuvre technique des principes.

Si vous avez des besoins particuliers pour mieux comprendre la Data Mesh Factory ou pour créer des schémas d'architecture adaptés à votre contexte, n'hésitez pas à nous contacter et nous nous ferons un plaisir d'en discuter avec vous.

Data Mesh Factory

Focus Data Domain

Data Owner

Le Data Owner est une personne qui porte la responsabilité des données relatives à son périmètre. Il possède une expertise métier approfondie. Il a notamment la charge de définir et gérer les data products, piloter leur qualité, assurer le respect des politiques et réglementations de l'entreprise.

Data Domain

Le principe de «Federated Computational Governance» repose sur l'idée que les données sont mieux gérées lorsque la responsabilité est répartie entre plusieurs entités indépendantes, mais coordonnées appelées domaines de données.

En garantissant la souveraineté de son domaine de données, cette organisation peut assurer les développements, la qualité, la sécurité et la fiabilité de ses données, tout en contribuant aux objectifs globaux de l'entreprise. Cela permet de créer un écosystème dans lequel chaque entité peut bénéficier des avantages de la collaboration tout en préservant son autonomie.

Data Product

Un data product est un ensemble de données organisées, structurées, nettoyées et transformées en un format qui peut être consommé par d'autres applications ou utilisateurs.

Il va au-delà de la simple accumulation de données brutes. Il comprend généralement des données qui ont été transformées, nettoyées et structurées d'une manière qui les rend utiles et faciles à comprendre pour les utilisateurs finaux.

Data Product Builder

Le but principal de ce composant est de générer des produits de données adaptés aux exigences de consommation de votre entreprise. Il offre la possibilité de consolider des sources de données internes et externes à l'entreprise, lesquelles peuvent être de différentes natures (structurées ou non structurées). Ce processus peut nécessiter une variété de démarches et de techniques pour garantir la qualité et la pertinence des produits de données.

La panoplie de technologies employées dans ce composant peut être vaste. En effet, la production des data products relève de la responsabilité des domaines de données. Ces derniers peuvent donc choisir les technologies qu'ils souhaitent.

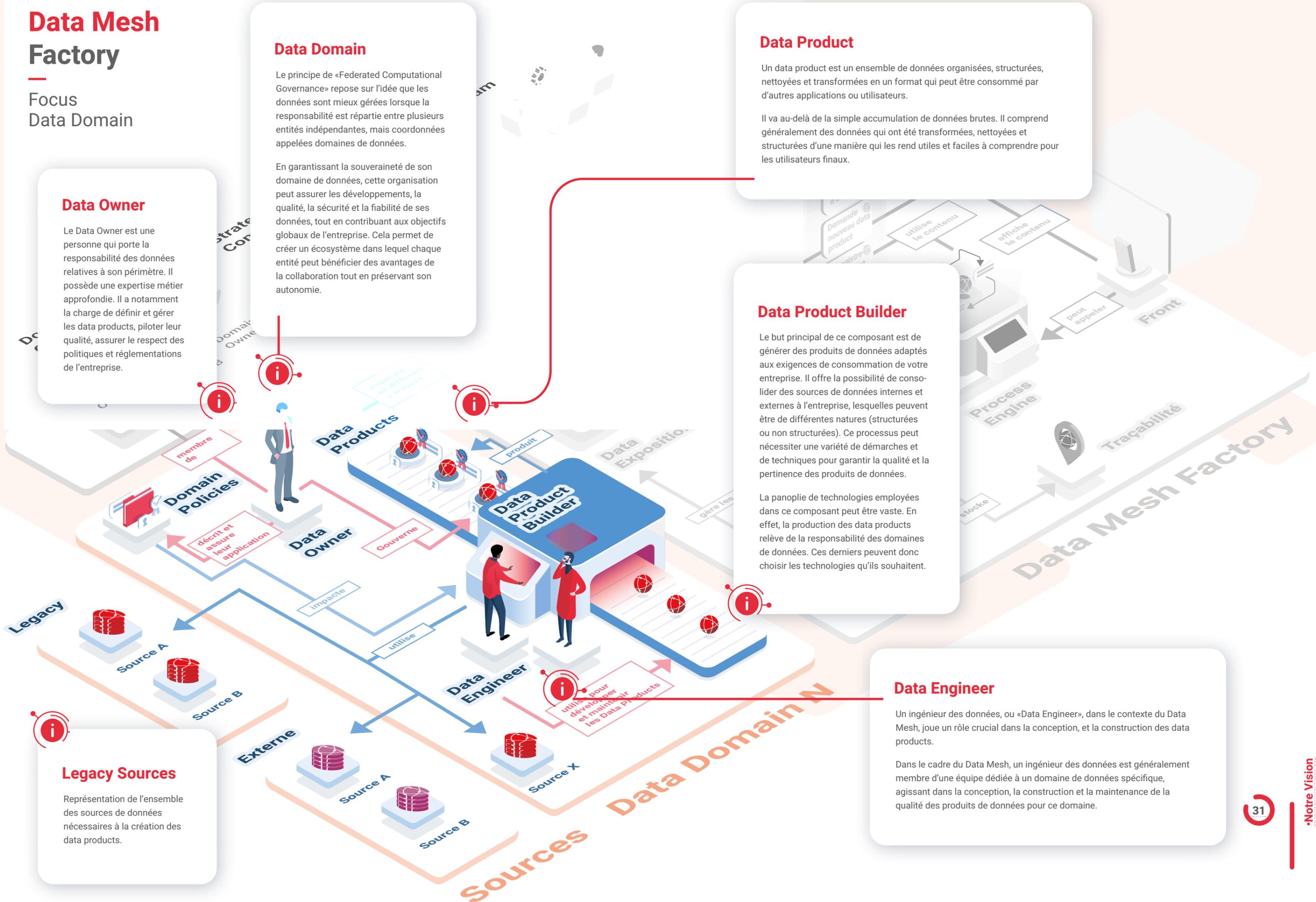
Data Engineer

Un ingénieur des données, ou «Data Engineer», dans le contexte du Data Mesh, joue un rôle crucial dans la conception, et la construction des data products.

Dans le cadre du Data Mesh, un ingénieur des données est généralement membre d'une équipe dédiée à un domaine de données spécifique, agissant dans la conception, la construction et la maintenance de la qualité des produits de données pour ce domaine.

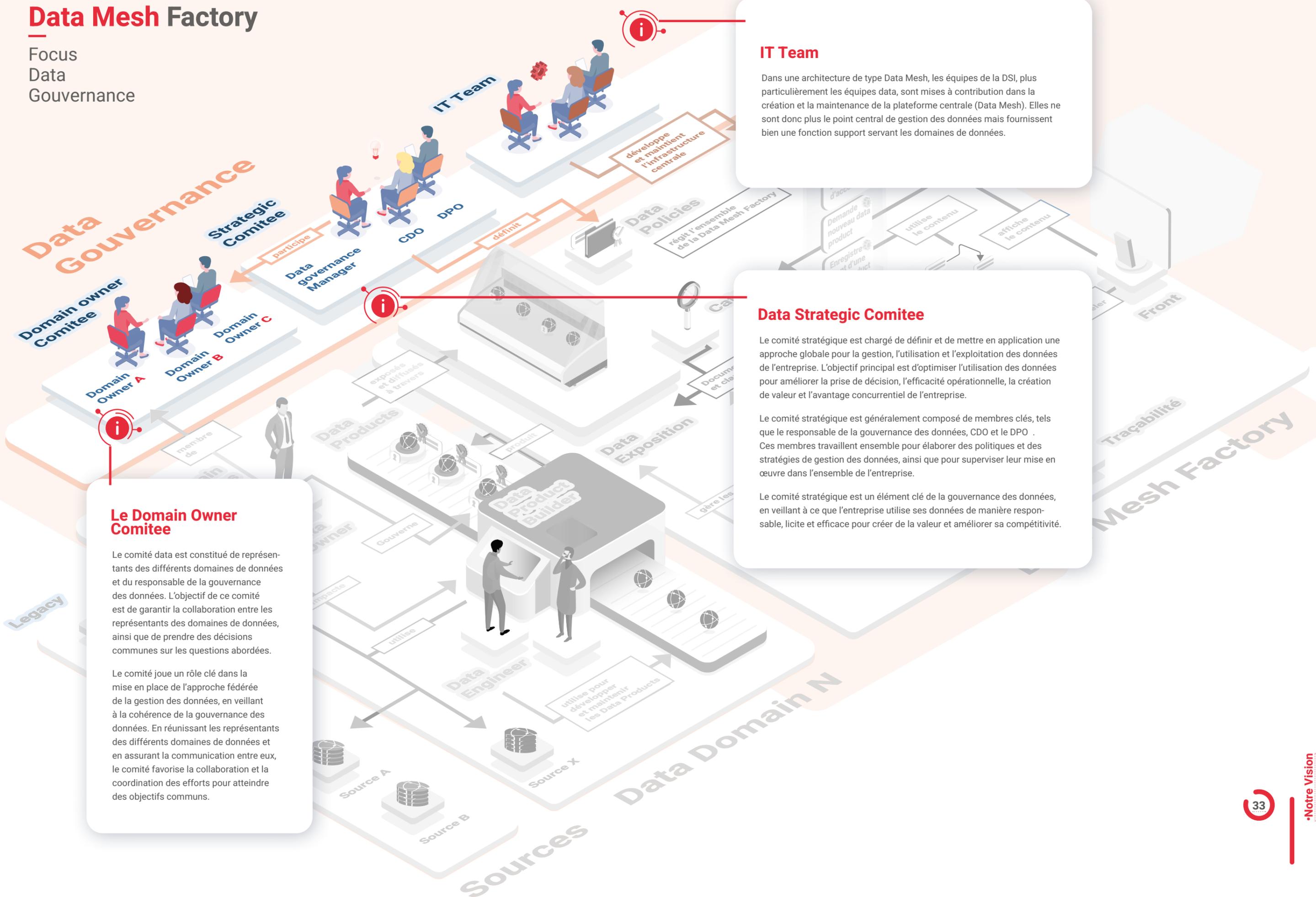
Legacy Sources

Représentation de l'ensemble des sources de données nécessaires à la création des data products.



Data Mesh Factory

Focus
Data
Gouvernance



IT Team

Dans une architecture de type Data Mesh, les équipes de la DSI, plus particulièrement les équipes data, sont mises à contribution dans la création et la maintenance de la plateforme centrale (Data Mesh). Elles ne sont donc plus le point central de gestion des données mais fournissent bien une fonction support servant les domaines de données.

Data Strategic Comitee

Le comité stratégique est chargé de définir et de mettre en application une approche globale pour la gestion, l'utilisation et l'exploitation des données de l'entreprise. L'objectif principal est d'optimiser l'utilisation des données pour améliorer la prise de décision, l'efficacité opérationnelle, la création de valeur et l'avantage concurrentiel de l'entreprise.

Le comité stratégique est généralement composé de membres clés, tels que le responsable de la gouvernance des données, CDO et le DPO . Ces membres travaillent ensemble pour élaborer des politiques et des stratégies de gestion des données, ainsi que pour superviser leur mise en œuvre dans l'ensemble de l'entreprise.

Le comité stratégique est un élément clé de la gouvernance des données, en veillant à ce que l'entreprise utilise ses données de manière responsable, licite et efficace pour créer de la valeur et améliorer sa compétitivité.

Le Domain Owner Comitee

Le comité data est constitué de représentants des différents domaines de données et du responsable de la gouvernance des données. L'objectif de ce comité est de garantir la collaboration entre les représentants des domaines de données, ainsi que de prendre des décisions communes sur les questions abordées.

Le comité joue un rôle clé dans la mise en place de l'approche fédérée de la gestion des données, en veillant à la cohérence de la gouvernance des données. En réunissant les représentants des différents domaines de données et en assurant la communication entre eux, le comité favorise la collaboration et la coordination des efforts pour atteindre des objectifs communs.

Après avoir exploré en profondeur notre perspective d'une architecture conforme à la vision du datamesh dans la section précédente, il est temps de se plonger dans le cœur pratique de sa mise en œuvre. Un Data Mesh réussi n'est pas uniquement basé sur la conception de l'architecture; il s'appuie également sur une série de principes, de pratiques et d'outils qui facilitent sa gestion et son évolution.

Découvrons ensemble comment aborder les modules clés de la Data Mesh Factory et les choix technologiques possibles.

1. Structurer ses data domaines

Dans le contexte de la gestion de données, un «domaine de données» fait généralement référence à une segmentation de l'entreprise en sous-ensemble cohérent afin d'y associer des rôles et des responsabilités.

Cela signifie qu'au lieu de simplement organiser les données en fonction de leur type, de leur structure, ou de leur source, les domaines de données sont organisés en fonction de la manière dont les données sont utilisées et valorisées dans l'entreprise. Cela aide à rendre la gestion des données plus efficace et utile pour les personnes qui travaillent dans ce domaine. Par exemple, un domaine de données pourrait être «les ventes», «les clients», «les produits», etc.

Comment structurer ses data domaines ?

Selon **Jill Dyché et Analise Polsky**, deux spécialistes de la donnée ayant travaillé sur la description des modèles de stewardship, il n'y a pas de réponse unique à cette question. La vérité est que la façon dont vous structurez votre organisation dépend de la situation, du contexte et de vos objectifs. **Il faut prendre le soin de l'analyser**, car c'est une des étapes déterminantes de la constitution d'une organisation de Gouvernance et nécessaires à la mise en œuvre du Data Mesh.

Afin de vous aider, et en nous inspirant de certains modèles de Dyché et Polsky, **ci-dessous certains axes de développement** que peuvent prendre une organisation de gouvernance ainsi que leurs avantages et inconvénients.

Orienté activités

Dans un contexte d'organisation d'entreprise structurée par activités, on observe souvent une similitude marquée entre cette structuration et l'organisation des data owners.

En effet, chaque propriétaire de données a sous sa responsabilité un domaine spécifique : par exemple, les propriétaires de données liées aux ventes ne sont pas les mêmes que les responsables de données du marketing ou des ressources humaines, etc.

L'atout majeur de cette structuration par activité réside dans la facilité de sa mise en œuvre. Les limites de chaque domaine sont clairement définies, ce qui simplifie l'attribution des responsabilités. Toutefois, ce modèle présente aussi des défis, notamment quand il s'agit d'attribuer la propriété des données utilisées par plusieurs entités.



Prenons le cas du «client», une notion transverse à plusieurs domaines. Il peut être complexe de déterminer à qui incombe la responsabilité des données client, étant donné que ces informations sont utilisées à la fois par le domaine des ventes et par le domaine du marketing. La mise en place d'un système efficace de propriété des données nécessite donc une attention particulière pour les éléments de données partagés à travers plusieurs domaines.

Orienté processus

Dans une organisation où les domaines de données sont structurés autour de processus, un propriétaire de données est associé à un ou plusieurs processus ou sous-processus spécifiques. Par exemple, un propriétaire de données chargé du processus de facturation serait responsable de diverses données impliquées dans ce processus métier spécifique.

Cette approche permet une évaluation plus simple du succès, avec la qualité et la disponibilité des données pouvant être mesurées par le suivi de l'utilisation de données de haute qualité dans le processus métier concerné. Cependant, cette structure présente un défi en termes d'attribution de la propriété des données, du fait des nombreuses interconnexions dues à l'usage des données. Un processus peut en effet impliquer plusieurs applications et silos organisationnels. Il est donc crucial de bien déterminer et vérifier la structure des processus avant de recommander ce modèle.

En outre, ce modèle nécessite une cartographie claire des processus existants au sein de l'entreprise. Il peut donc être particulièrement adapté aux organisations ayant déjà mené des travaux alignés sur la norme ISO 9001.

Nous estimons que ce type de modèle est approprié pour les entreprises dont la structure des processus est en adéquation avec les structures organisationnelles et applicatives.

Orienté Système

Dans le type de modèle orienté système, nous avons une vision de la construction des domaines de données très techniques, c'est-à-dire que la construction va se baser sur la connaissance et la compétence techniques des données. Ce modèle attribue le stewardship aux personnes qui gèrent les systèmes, ce qui ressemble beaucoup à la vision que nous avons donc des «Technical Steward».

Dans ce modèle, par exemple, un responsable des données pourrait être affecté au système de facturation.

L'idée est qu'en traitant les problèmes de données en amont des métiers, une entreprise peut propager des données de meilleure qualité à d'autres systèmes et utilisateurs de manière durable. Cela permet d'économiser du travail et du temps à mesure que les systèmes en aval acquièrent des données opérationnelles robustes et précises, mais cela offre une vision IT à la gestion des données. Cette approche n'est pas la plus adaptée dans le cadre de la mise en place d'un Data Mesh.

Pour résumer, lors de l'initiation d'un tel projet au sein de votre organisation, il n'y a généralement pas de chemin prédéfini à suivre. Il vous faudra naviguer à travers les spécificités de votre entreprise et les objectifs de la mise en œuvre pour déterminer le bon équilibre en termes de répartition et de structure des responsabilités.

2. Les politiques

Au cœur de la mise en œuvre d'un Data Mesh, les «politiques», incarnent **un ensemble de principes directeurs, de règles et de normes**. Ces dernières dictent comment les données doivent être traitées, sécurisées, conservées et exploitées dans le contexte de cet écosystème.

Ces politiques ont un spectre d'action large et touchent à divers aspects de la gestion des données. **Elles disposent notamment des standards en termes de qualité des données, exigeant une précision, une exhaustivité et une cohérence de chaque «data product»**. Elles encadrent également la documentation, en définissant comment les produits de données doivent être décrits et classifiés.

Elles jouent aussi un rôle en termes d'accessibilité, en précisant le degré de sensibilité des données et en déterminant qui peut en connaître l'existence. La sécurité des données n'est pas en reste, ces politiques déterminant les mesures de protection indispensables pour prévenir toute violation de données. Elles instaurent aussi des règles relatives à la confidentialité et garantissent la conformité réglementaire, en établissant la manière dont les données personnelles doivent être traitées pour respecter la législation en matière de protection de la vie privée et autres réglementations applicables.

Ces politiques définissent en outre qui peut accéder à quelles données et dans quelles situations, tout en traçant des lignes directrices concernant l'utilisation des données. Cela comprend la définition des utilisations permises des données et celles qui sont interdites.

A quel niveau faut-il décrire les politiques ?

Les «politiques» au sein du Data Mesh sont essentielles pour établir un cadre de travail rigoureux et clair qui régit la manipulation des données à travers l'organisation. L'intégration des politiques dans un écosystème Data Mesh requiert d'une part une définition claire au niveau central pour l'ensemble des data products publiés et partagés, mais chaque data domaine doit aussi décrire des politiques qui le régissent.

Les politiques au niveau central englobent les activités transversales indispensables à la cohérence des actions de gouvernance des données dans leur globalité.

Les politiques au niveau du domaine de données comprennent un ensemble de règles fixées par chacun des domaines de données. Ces politiques permettent au domaine de données de maîtriser leur périmètre en se développant de manière gouvernée.

Cette structuration en deux niveaux présente plusieurs avantages, notamment une forte responsabilisation et une indépendance des domaines de données, ainsi qu'une agilité face aux besoins d'adaptation spécifiques à chaque domaine. Notre expérience démontre que cette responsabilisation, associée à des étapes de contrôle plus globales, facilite les échanges et l'application des principes de gouvernance des données centrale, améliorant ainsi la fluidité et la qualité.

Il est important de comprendre que les politiques sont le prélude aux décisions prises par les domaines de données. Elles servent à la fois de contraintes et de guides, permettant aux domaines de données de respecter les diverses attentes du système de gouvernance des données. Les moyens d'atteindre ces attentes restent cependant à la discrétion des domaines de données.

Quels sont les types de politiques ?

Nous distinguons plusieurs sujets essentiels qui doivent être abordés par les politiques afin d'assurer un fonctionnement fluide du Data Mesh.

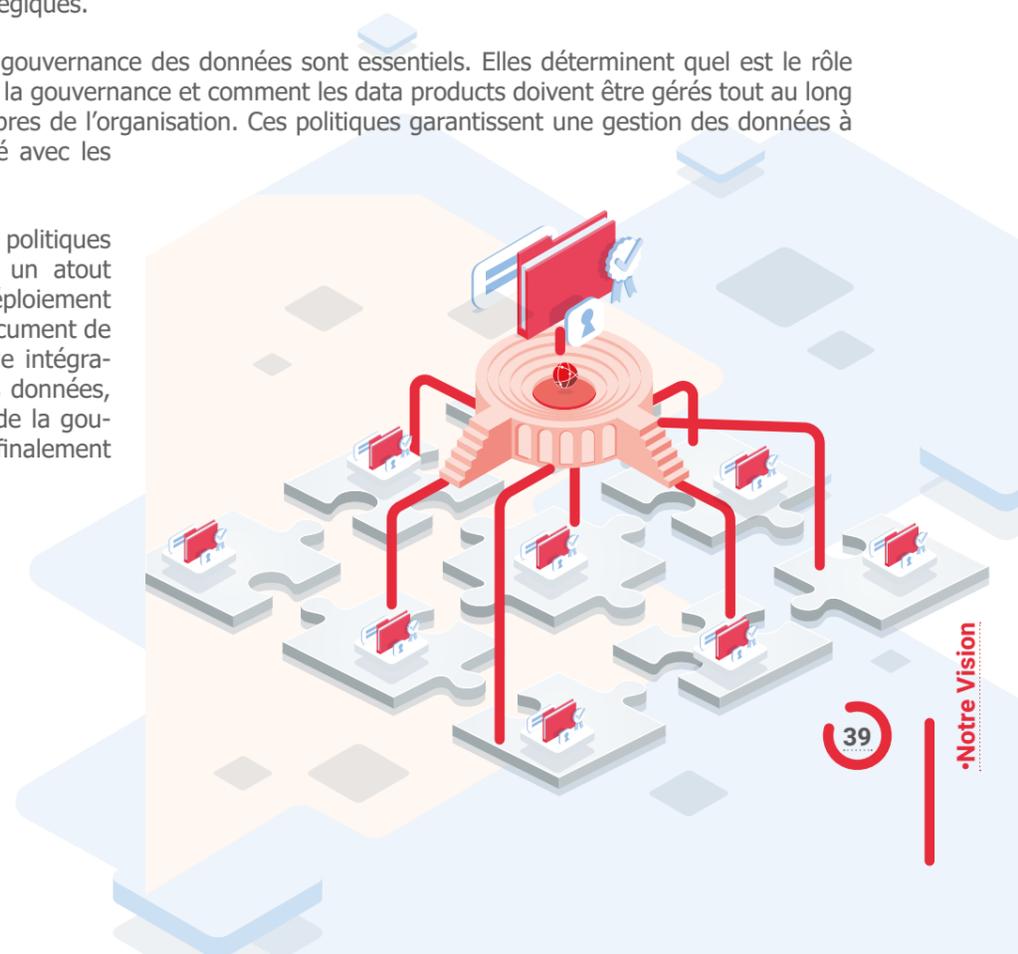
Une des priorités est la nécessité de décrire de manière exhaustive chaque data product. Une description précise, qui couvre l'identité du produit, sa fonction, le type de données qu'il intègre, et ses modalités d'utilisation, est impérative pour garantir une compréhension cohérente et transparente de chaque produit de données par les consommateurs.

En parallèle, l'importance des politiques de sécurité des données ne saurait être sous-estimée. Elles jouent un rôle essentiel pour protéger les informations sensibles et pour se conformer aux réglementations en vigueur. Ces politiques délimitent précisément qui est autorisé à accéder à quelles données, et dans quelles conditions. Elles contribuent à préserver l'intégrité des données, tout en sécurisant les informations sensibles.

En ce qui concerne la qualité des données, des politiques ad hoc sont requises. Elles définissent les normes de qualité que chaque domaine doit respecter, notamment en termes de précision, d'intégrité et de cohérence des données. De telles politiques assurent que les données sont dignes de confiance, et peuvent servir de base solide pour des prises de décisions stratégiques.

En dernier lieu, les politiques de gouvernance des données sont essentielles. Elles déterminent quel est le rôle exact de chacun des membres de la gouvernance et comment les data products doivent être gérés tout au long de leur cycle de vie par les membres de l'organisation. Ces politiques garantissent une gestion des données à la fois efficiente et en conformité avec les exigences réglementaires.

La mise en place d'un système de politiques solide et bien documenté offre un atout majeur : il sert de guide pour le déploiement du Data Mesh, agit comme un document de référence lors de chaque nouvelle intégration au sein de la fédération des données, facilite la diffusion de la culture de la gouvernance des données, et assure finalement la conformité réglementaire.



3. Interface utilisateur universelle

Le Data Mesh introduit des concepts fondamentaux d'accessibilité et de découvrabilité des données, qui doivent être accessibles à tous les types d'utilisateurs cibles. En fonction de nos observations et de nos expériences, nous pouvons classer les utilisateurs en trois catégories distinctes, ayant chacune des attentes et des besoins spécifiques.

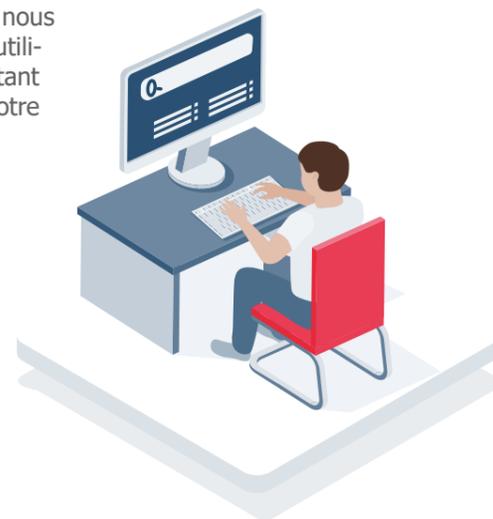
Premièrement, **les utilisateurs de type «grand public» recherchent une interface d'accès facile et intuitive**, rappelant le fonctionnement d'un moteur de recherche de données ou d'un Wikipédia. Leur objectif premier est d'accéder à des données existantes et de les comprendre facilement.

Deuxièmement, **les utilisateurs «avertis», qui englobent les analystes de données, les data scientists ou encore les membres d'autres domaines de données, se distinguent par leur compréhension technique**. Leurs objectifs sont souvent d'utiliser les données à des fins avancées, tels que la conception de nouveaux tableaux de bord ou la consommation automatisée d'un data product. Ces utilisateurs, capables d'une montée en compétence rapide, auront accès, avec le soutien de l'équipe centrale (DSI), à des fonctionnalités avancées et aux couches sous-jacentes de la Data Mesh Factory.

Enfin, **les utilisateurs membres de la fédération dits « fédérés », ont besoin d'une interface unique facilitant leurs activités**. Ils doivent être en mesure de gérer efficacement les data products qu'ils mettent à disposition. Leurs attentes résident principalement dans la gestion de la qualité des données, la gestion des accès, la documentation, etc., de manière centralisée et efficace.

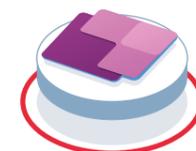
Ces trois catégories d'utilisateurs, bien que différentes, doivent pouvoir accéder au contenu de la Data Mesh Factory de manière simple, rapide et efficace et répondant aux attentes découlant de leurs rôles. À notre connaissance, il n'existe pas de produit éditeur couvrant de manière exhaustive l'ensemble des besoins. Dans le cadre de notre Data Mesh Factory, nous nous sommes plutôt orientés vers l'intégration d'une interface utilisateur personnalisée servant également de portail permettant l'accès aux interfaces des autres produits exploités dans notre solution.

De manière à proposer une expérience utilisateur capable d'offrir les capacités de personnalisation requises, nous avons naturellement sélectionné des solutions techniques proposant suffisamment de souplesse pour couvrir toutes les attentes.



Angular JS

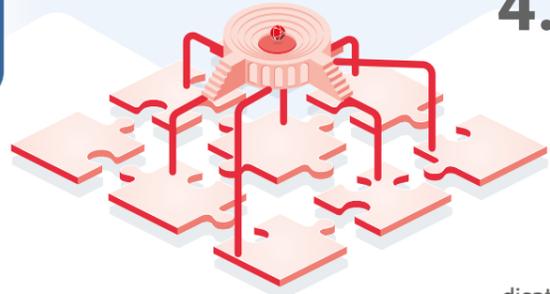
Les développements à façon via une technologie comme AngularJS reste une approche ayant fait ses preuves et pour lesquelles il existe de nombreuses ressources expertes sur le marché. Grâce à son architecture MVC, AngularJS permet une organisation claire du code et facilite la maintenance et l'évolutivité de l'application. Son système de binding de données bidirectionnelle garantit des mises à jour en temps réel de l'interface utilisateur, offrant une expérience interactive et réactive. De plus, la communauté conséquente et l'écosystème riche d'AngularJS fournissent un soutien solide. Des bibliothèques graphiques sont disponibles en abondance, ce qui en fait un choix pertinent pour développer des portails et interfaces web performants et qui seront en mesure de suivre l'évolution de votre Data Mesh Factory.



PowerApps

Les outils Low Code comme PowerApps peuvent également servir ce besoin. Ils proposent une approche simplifiée permettant d'obtenir un aperçu immédiat du rendu visuel tout en intégrant le développement du back end de manière transparente. Grâce à une interface visuelle conviviale et à une riche bibliothèque de composants graphiques, les développeurs peuvent concevoir et créer des portails web en s'affranchissant de l'écriture de code qui peut se révéler complexe. Ainsi, ces outils accélèrent considérablement le processus de développement, réduisent les erreurs et facilitent la collaboration entre les développeurs et les utilisateurs métier. Ces outils proposent une approche moderne, agile et rapide pour développer des interfaces utilisateur sur mesure s'inscrivant parfaitement dans ce que l'on peut attendre de l'interface de la Data Mesh Factory.

Ce type de technologie permet d'embarquer facilement vos utilisateurs en répondant avec une grande réactivité à leurs demandes.



4. Assurer la fédération des domaines de données

L'approche du Data Mesh offre une gouvernance décentralisée, favorisant ainsi l'autonomie des domaines de données tout en formalisant leur responsabilité en matière de diffusion de l'information. À l'instar des politiques qui soutiennent la standardisation des procédures dans l'écosystème du Data Mesh, la construction et le développement de domaine de données souverain ne sauraient s'opérer sans un certain nombre de facteurs communs.

Posséder des rôles communs

La présence de rôles communs dans tous les domaines de données stimule l'uniformité et la cohérence tout au long du processus de production et de maintien des data products. Malgré le caractère unique que peut revêtir chaque domaine de données, l'instauration de rôles communs apporte un cadre de normes et de pratiques partagées, orientant ainsi l'exploitation et l'usage des données au sein de l'organisation globale.

La définition de rôles communs est déterminante dans l'amélioration de la collaboration et de la communication entre les divers domaines de données. En conséquence, **une meilleure compréhension des responsabilités et des attentes mutuelles se dessine, favorisant une collaboration plus fluide et un partage d'informations optimisé entre les domaines de données.**

Dans un écosystème Data Mesh, la présence de rôles communs est essentiel pour garantir la cohérence, la collaboration, l'interopérabilité et la promotion d'une culture de la donnée. Dans les sections suivantes, nous détaillerons certains rôles et instances que nous considérons cruciaux dans ce contexte.

Il est important de souligner que lors de la mise en place d'une architecture Data Mesh, la **prise en compte de l'organisation actuelle**, son **enrichissement** et sa **rationalisation** (si nécessaire) sont des étapes fondamentales. Cette démarche est cruciale pour aligner la nouvelle architecture avec les structures et les pratiques existantes, tout en maximisant son efficacité et son impact.

Quel est le rôle du Data Owner ?

Dans le contexte d'une organisation, le «propriétaire de données» ou «data owner» est un **rôle clé qui consiste à superviser, contrôler et garantir l'utilisation correcte d'un ensemble spécifique de données**, aussi appelé domaine de données. L'ampleur de ses responsabilités est délimitée par les frontières de ce domaine de données.



Le propriétaire de données est responsable de plusieurs aspects importants en définissant les règles ou critères minimaux de qualité des données, sécurité, confidentialités, d'intégrités, d'exactitude et s'assure du respect des règlements en vigueur.

Il s'agit généralement de la figure de référence pour toutes les décisions concernant les données de son domaine. Il a le pouvoir de déterminer qui peut accéder à ces données, de quelle manière elles sont utilisées et la stratégie de protection à adopter. De plus, il est tenu de garantir que l'utilisation de ces données se fasse dans le respect des normes éthiques et légales.

Que devient le Data Engineer ?

Dans le contexte d'un écosystème Data Mesh, le rôle du Data Engineer se réinvente pour s'aligner avec la philosophie décentralisée de ce nouvel écosystème. Cette fonction se distingue nettement de l'approche traditionnelle, qui se concentre principalement sur la création et la maintenance d'une plateforme de données centralisée au sein d'une équipe IT.

Une des responsabilités primordiales du Data Engineer dans un cadre Data Mesh réside dans la conception, la maintenance et l'amélioration continue des produits de données (Data Products) associés à son domaine spécifique. En tant qu'ingénieur des données, il assure l'intégrité, l'exactitude, la sécurité et l'accessibilité de ces données en suivant les directives établies par le data owner, ce qui garantit leur utilisation optimale et leur conformité aux politiques de gouvernance des données établies par l'organisation.

Les Data Engineers coopèrent étroitement avec d'autres parties prenantes clés, notamment les consommateurs des produits de données, les analystes de données et les scientifiques des données. Leur objectif est de comprendre les besoins de ces acteurs, de leur fournir un support technique et de les aider à accéder aux données dont ils ont besoin.

Les Data Engineers jouent aussi un rôle pédagogique important. Ils sont amenés à former et sensibiliser d'autres membres de l'organisation sur l'utilisation et la manipulation des données. Ce rôle est essentiel pour permettre à l'ensemble de l'organisation d'exploiter au mieux ses données.

En résumé, au sein d'un écosystème Data Mesh, le Data Engineer a pour mission de veiller à ce que les données soient disponibles, accessibles, sécurisées et de haute qualité, tout en assurant leur utilisation efficace pour soutenir les objectifs de l'organisation. Il s'agit d'un rôle multifacette, essentiel à la bonne gestion des données dans le cadre d'un Data Mesh.

Comment mettre en relation les différents membres ?

De par sa nature, un écosystème Data Mesh est structuré de manière fédérée, et ses participants doivent disposer d'instances d'échange communes pour une co-construction harmonieuse. Si votre entreprise a déjà une organisation de gouvernance établie, il est probable que vous disposiez d'un ou de plusieurs comités dédiés à ces pratiques. Au sein d'un écosystème Data Mesh, la composition de tels comités est généralement diversifiée, rassemblant fréquemment les propriétaires des données (data owners) ainsi que certains membres de l'équipe centrale comme le data gouvernance manager.

On distingue généralement deux comités. Le premier, **le Data Mesh committee vise à définir les directives stratégiques** pour la gestion des données au sein de l'écosystème. Il précise les objectifs principaux et les principes fondamentaux qui orientent l'approche de l'organisation en matière de gestion des data products, tout en tenant compte des objectifs commerciaux globaux de l'organisation. Il favorise les évolutions de l'écosystème et la priorisation de l'évolution ou de la création de nouveaux data products nécessaires à plusieurs domaines de données.

Le second, généralement déjà inscrit dans les entreprises contemporaines, **le comité data, rassemble les membres de l'IT et des directions, il traite plutôt de la stratégie des usages de la donnée, de la mitigation des risques relatifs aux données et de la construction de la gouvernance des données dans son ensemble.**

En outre, les comités sont responsables de l'élaboration et de l'évolution des politiques et des standards en matière de gestion des données. Cela englobe la définition de normes de qualité, de règles de documentation, de standards techniques et d'autres directives qui doivent être respectées par les ingénieurs de données (Data Engineers) et les autres membres de l'organisation pour la publication et l'utilisation des data products disponibles.

5. Construire ses data products

Comme évoqué dans notre section principes (Data as Product), les data products se caractérisent par plusieurs attributs clés qui les distinguent des données brutes et en font des produits finis utiles et exploitables. Tout d'abord, un data product est organisé et structuré, ce qui signifie que les données sont agencées de manière logique et cohérente, facilitant ainsi leur accès et leur compréhension. De plus, un data product est transformé et enrichi, ce qui implique que les données soient traitées, nettoyées et améliorées pour garantir leur qualité, leur intégrité et leur pertinence. Les data products peuvent être également accompagnés de métadonnées descriptives, qui fournissent des informations supplémentaires sur les données, telles que leur provenance, leur signification et leur contexte d'utilisation.

En synthèse, les caractéristiques clés d'un data product sont son organisation, sa transformation, ses métadonnées, sa qualité et sa pertinence pour répondre aux besoins spécifiques des utilisateurs et des applications de votre entreprise.

Comment choisir les données appropriées à la construction d'un data product ?

Pour un Domain Owner, la définition d'un data product est une opération cruciale. Pour faire un choix pertinent, il doit identifier clairement les objectifs et les cas d'utilisation spécifiques du data product. L'analyse détaillée des besoins exprimés par les utilisateurs est primordiale de manière à élaborer un data product au plus proche des attentes. Les sources de données disponibles au sein de son domaine ainsi que leur qualité doivent également être évaluées afin de vérifier que les objectifs envisagés soient atteignables. Sans oublier d'assurer la conformité des contraintes légales et réglementaires liées à la confidentialité et à la sécurité des données.

En suivant ces recommandations, le domain owner sera en mesure de sélectionner les données appropriées pour construire un data product solide et aligné sur les besoins de votre entreprise

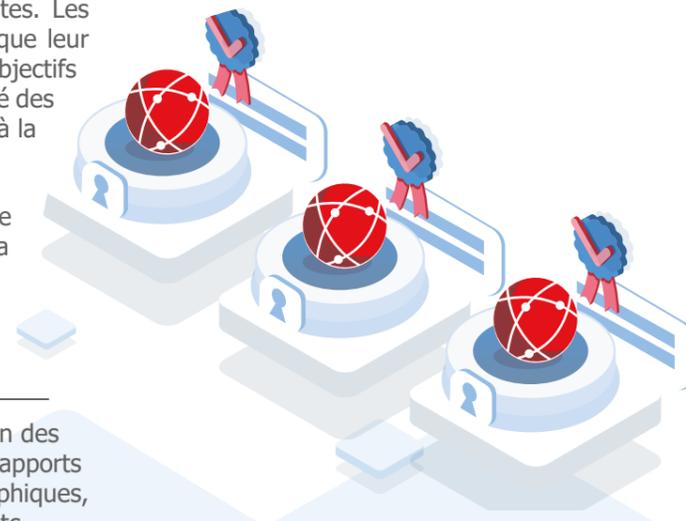
Quel est le format d'un data product ?

Les data products peuvent prendre différentes formes en fonction des besoins et des objectifs spécifiques. Ils peuvent inclure des rapports prédéfinis, des tableaux de bord interactifs, des visualisations graphiques, des fichiers CSV ou Excel, des API (REST / SOAP), des vues DB, etc...

Au final, la forme importe peu, à partir du moment qu'il respecte à la fois les politiques définies au sein du domaine ainsi que les politiques définies au niveau de la Data Mesh Factory.

Les domaines ont la souveraineté technologique !

Les principes du Data Mesh le définissent clairement : chaque domaine est libre de s'organiser et choisir les technologies qu'il souhaite du moment que les politiques générales sont respectées.



En théorie, dans ce que nous nommons Data Product Builder, chaque domaine est libre d'apporter sa propre implémentation en utilisant les technologies qu'il souhaite en fonction des compétences ou produits à disposition.

Dans la pratique, il sera généralement plus rationnel de **mutualiser une approche commune en termes de produit, technologie ou encore méthode d'implémentation**, mais cela va dépendre de la taille de votre entreprise. Mais, on le rappelle, il s'agit ici d'une interprétation pragmatique qui va à l'encontre du principe de souveraineté des domaines.

Parmi les innombrables possibilités d'implémentation du Data Product Builder, nous avons choisi des approches qui peuvent vous donner des pistes d'orientation technologique lors de la mise en œuvre d'un Data Product Builder :



Talend

Talend est une solution puissante et complète pour la création de data products. Talend propose une large gamme de connecteurs préconstruits pour se connecter à différentes sources de données, qu'elles soient internes ou externes, ce qui facilite l'intégration des données dans le processus de création du data product. Nous apprécions l'IDE qui permet de transformer / enrichir / nettoyer les données simplement et rapidement. Les fonctionnalités de qualité des données intégrées permettent de détecter et de corriger les erreurs, les doublons et les incohérences dans les données. Ces caractéristiques font de Talend une solution à envisager pour réaliser un Data Product Builder répondant aux attentes.



Azure

La plateforme Azure est une autre solution pertinente pour la création de data products. Cette solution orientée cloud offre une gamme complète d'outils et de services utiles à la mise en place d'un Data Product Builder. En plus d'une solution de stockage, Azure offre de nombreux services de traitement des données puissants, tels que Azure Databricks qui permet d'exécuter des tâches de transformation, de nettoyage et d'analyse de données à grande échelle. La fonctionnalité Azure Machine Learning offre même la possibilité d'intégrer de l'intelligence artificielle durant l'élaboration des data products. La sécurité et la conformité sont également au cœur de la plateforme Azure, avec des fonctionnalités robustes pour la protection des données sensibles et la conformité aux réglementations en vigueur. Azure est donc une solution à considérer pour créer des data products si le cloud fait partie de votre culture d'entreprise.

6. Rendez facilement accessibles vos data products

Un data product élaboré au sein d'un domaine peut sur demande être mis à disposition des utilisateurs ou applications intéressées. Cependant cette approche non formalisée nécessiterait des échanges, explications et autres opérations spécifiques de la part du domaine owner. Une autre alternative répliquant les data products dans une plateforme centrale nécessiterait de coûteux processus d'intégration. Toutes ces options vont à l'encontre du principe fondamental de self-service platform énoncé par le Data Mesh.

Notre Data Mesh Factory adresse cette exigence en automatisant toute la phase d'exploration en passant par la partie documentation jusqu'à la mise à disposition technique du data product aux utilisateurs intéressés.

En considérant l'ensemble de ces attentes, nous avons envisagé diverses possibilités d'implémentation s'appuyant sur différentes familles de produits. **La solution qui nous semble apporter le plus de valeur est sans conteste les outils de data virtualisation.**

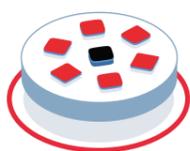
Pourquoi la data virtualisation est une solution appropriée ?

La data virtualisation est une approche qui offre de nombreux avantages dans la gestion, l'intégration et l'accès aux données, ce qui en fait une solution candidate pour les architectures s'inspirant des concepts du Data Mesh. L'un des principaux atouts de la data virtualisation est sa capacité à créer une vue unifiée et cohérente des data products dispersés au travers des différents systèmes de chaque domaine, sans dépendre de processus complexe d'intégration ou de duplication de données.

Une autre caractéristique importante de la data virtualisation est son agnosticisme, c'est-à-dire sa capacité à être totalement indépendante sur le plan technique. Cela signifie que chaque domaine peut choisir les technologies et les outils qui conviennent le mieux à ses besoins spécifiques, sans être limité par des contraintes techniques imposées par la Data Mesh Factory. Cette décorrélation technique permet une plus grande flexibilité et agilité, permettant aux équipes d'**intégrer rapidement leurs data products.**

De plus, la centralisation offerte par les solutions de data virtualisation facilite la gestion de la sécurité et de la gouvernance des données. En regroupant les data products au sein d'une plateforme virtuelle centralisée, il devient plus facile de mettre en place des politiques de sécurité, de contrôler l'accès aux données sensibles et de garantir la conformité avec les politiques définies au sein de la Data Mesh Factory. La data virtualisation facilite aussi le suivi des échanges de données, offrant ainsi une traçabilité complète et facilitant les processus d'audit.

Pour Synotis, l'utilisation de solutions de data virtualisation telles que Denodo ou Dremio se révèle être un véritable accélérateur lors de la mise en œuvre d'une architecture de type Data Mesh.



Denodo

La solution Denodo offre une connectivité étendue avec une large gamme de sources de données, y compris les bases de données relationnelles, les applications cloud, les services web et bien d'autres encore. Sa capacité à s'intégrer facilement avec diverses sources de données en fait une solution polyvalente et adaptable à différents environnements. De plus, de par ses fonctionnalités avancées de modélisation des données et de transformation de données, il sera en mesure d'assurer un rôle de Data Product Builder pour des domaines qui ne seraient pas en capacité de produire techniquement leurs data products. Avec sa performance optimisée et sa gestion intelligente du cache, Denodo garantit des temps de réponse rapides et une évolutivité élevée.



Dremio

La solution Dremio offre des performances exceptionnelles grâce à son moteur de requêtes distribué et optimisé offrant des mécanismes de cache. Les connecteurs proposés couvrent une grande partie des produits du marché. Dremio propose une interface agréable qui facilite l'exploration et la préparation des données, offrant aux utilisateurs la possibilité d'interagir facilement avec les données. Tout comme son concurrent, Dremio offre des fonctionnalités de modélisation et transformation de données suffisamment avancées pour assurer le rôle de Data Product Builder aux domaines qui en auraient besoin. Dremio propose une version standard sans licence permettant aux entreprises de tester ou mettre en œuvre avec limitation le produit.



Un outil de data virtualisation c'est quoi ?

Un outil de data virtualisation est une solution informatique qui permet d'accéder, de consolider et rendre accessible les données à partir de sources de données multiples et hétérogènes, sans nécessité de les déplacer ou de les stocker dans un espace centralisé.

7. Maîtriser son patrimoine

Dans le cadre d'un domaine de données spécifique, il est primordial de maîtriser son patrimoine de données avant même d'envisager la diffusion de produits de données. Cette maîtrise joue un rôle crucial dans l'atteinte de l'un des objectifs clés du Data Mesh, à savoir la minimisation des dépendances entre les domaines de données. En effet, chaque domaine doit avoir une compréhension approfondie et un contrôle de ses propres données pour fonctionner de manière autonome, ce qui facilite la flexibilité et réduit les goulots d'étranglement.

Il est essentiel de ne pas interpréter ces informations comme une invitation à entreprendre des campagnes de documentation massives du patrimoine de données de l'entreprise. Bien que chaque domaine doive maîtriser son patrimoine de données, il reste souverain et cette maîtrise n'est pas un prérequis au lancement du Data Mesh. Il s'agit plutôt d'un guide pour choisir les ressources appropriées pour sa mise en œuvre.

Toutefois, dans le contexte de la Data Mesh Factory, la maîtrise du patrimoine de données est un prérequis essentiel à son fonctionnement. Pour promouvoir la consommation des données et l'autonomie des utilisateurs, les data products doivent être décrits en respectant des normes spécifiques indiquées dans les politiques. Ces normes définissent notamment les métadonnées nécessaires à la publication des produits de données sur la plateforme centrale.

Ces métadonnées peuvent être regroupées en deux catégories principales : celles relatives au contenu et à la compréhension du data products, et celles liées à la classification du data product. Cette dernière peut concerner la sensibilité du contenu, la visibilité par les autres membres, ou simplement l'identification du domaine de données auquel il appartient.

En prenant l'architecture proposée, nous positionnons le catalogue de données comme une source d'information permettant la bonne exploitation des processus d'accès et d'audit ainsi qu'une interface de recherche se destinant aux utilisateurs finaux.

Le catalogue de données s'avère ainsi indispensable dans ce contexte, offrant des fonctionnalités telles que l'exploration et la recherche de données, la collaboration, ainsi que le suivi et la traçabilité. Le catalogue de données est donc un outil essentiel pour toute entreprise cherchant à optimiser l'utilisation de la Data Mesh Factory. **Il permet aux utilisateurs de découvrir et de comprendre les produits de données et permet aux propriétaires de domaines de documenter les data products pour faciliter leur utilisation, le tout de manière autonome et sans intervention de l'équipe centrale.**



DataHub

Le produit DataHub est un outil de gestion des métadonnées open source et hautement personnalisable. Contrairement aux produits de catalogage du marché, le produit DataHub offre une approche décentralisée qui s'aligne parfaitement avec les principes fondamentaux du Data Mesh. Il permet aux équipes des domaines de cataloguer et documenter leur data products de manière autonome et totalement automatisée s'ils le souhaitent. DataHub offre également une interface utilisateur riche, avec un moteur de recherche type Google efficace, de multiples plugins permettant de créer des interfaces sur mesure répondant aux attentes de chaque acteur de la Data Mesh Data Mesh Data Factory est donc un gage de souplesse et une garantie d'évolutivité pour votre entreprise.

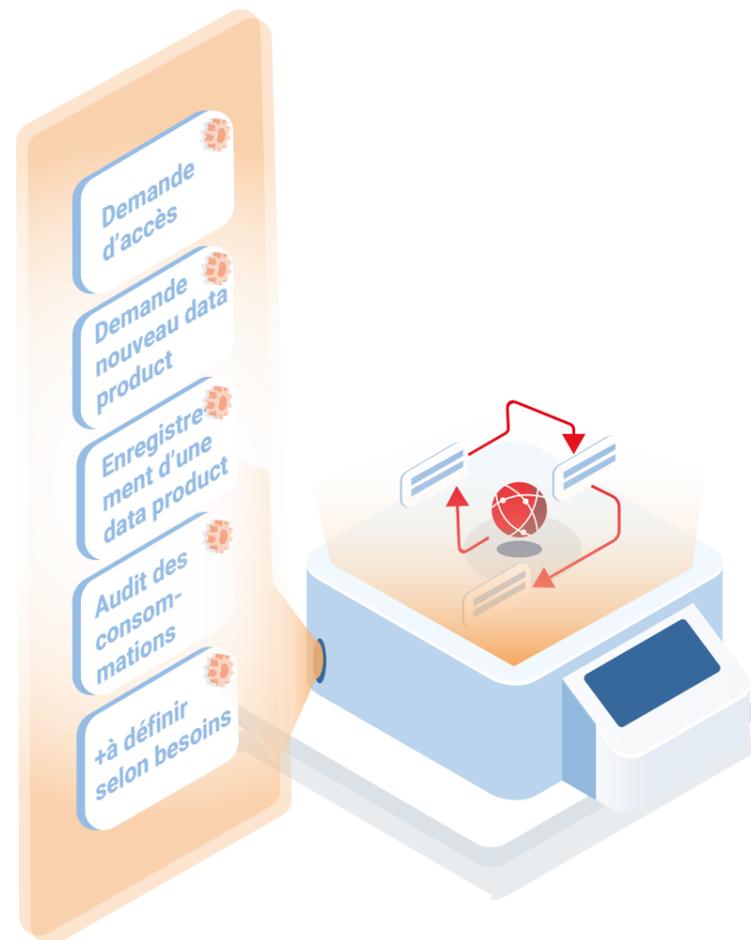


Azure data catalogue

Azure data catalogue est un catalogue de données disponible sur le cloud Azure permettant aux équipes de votre entreprise de fournir une vision centralisée et homogène du patrimoine données de votre entreprise. Il clarifie les définitions des données, les synonymes et les attributs commerciaux essentiels afin que tous les utilisateurs comprennent et puissent exploiter de manière autonome les data products. Grâce à ses fonctionnalités avancées de recherche et de métadonnées, Azure data catalogue permet aux utilisateurs d'explorer facilement les données de manière autonome et agile. Dans un environnement cloud, Azure data catalogue regroupe toutes les fonctionnalités requises pour assurer son rôle dans notre Data Mesh Factory.

8. Gouvernez ses processus

Les processus doivent être au cœur de l'animation de la Data Mesh Factory. L'action d'automatiser et de standardiser certaines tâches relatives aux data products est le meilleur moyen d'une part d'uniformiser les pratiques autour du partage de données, mais aussi d'impliquer les métiers en fournissant un moyen clair d'obtenir, de demander, ou de faire évoluer des data products existants. Cela permet de rendre le Data Mesh vraiment dynamique.



Les processus métier doivent être décrits de manière claire et structurée. Pour cela, l'utilisation d'un gestionnaire de processus nous paraît être la meilleure option. Il permet de définir des séquences de tâches, d'envoyer des notifications, de suivre l'état d'avancement des tâches et de gérer les exceptions de tous les processus déclarés dans notre moteur de processus.

Parmi la liste des processus à traiter nous identifions :

De manière à assurer une cohérence globale, il est important de pouvoir déclencher, exécuter et assurer le suivi depuis l'interface de la Data Mesh Factory qui doit demeurer l'unique point d'entrée pour les utilisateurs.

Au final, le gestionnaire de processus doit présenter les fonctionnalités suivantes :

- **Modélisation de processus** : Les traitements du process engine sont définis via un format visuel pour faciliter la compréhension, la maintenance et l'évolutivité des processus.
- **Automatisation de tâches** : Le fonctionnement du process engine se fait de manière automatique. Les sollicitations aux acteurs de la Data Mesh Factory n'interviennent que sur les étapes nécessaires prédéfinies par des règles lors de la modélisation des processus.
- **Définition des délais** : Des notifications en cas de non-réponse de la part des acteurs de la Data Mesh Factory permettent d'assurer une prise en compte rapide des demandes d'accès ou de mise à disposition des data products.
- **Suivi et reporting** : Par exemple, l'utilisateur sollicitant l'accès à un data product pourra en tout temps visualiser le statut de sa demande et connaître les personnes impliquées dans la validation de sa requête.
- **Gestion des exceptions** : Les cas d'exceptions et les erreurs qui surviennent pendant l'exécution du processus sont gérés en proposant des mécanismes de remédiation pouvant faire intervenir d'autres acteurs de la Data Mesh Factory.

Il existe de nombreuses solutions disponibles sur le marché. A titre d'exemple nous avons choisi de mettre en avant les produits suivants :



Activiti

Activiti est un moteur de workflow open-source qui offre des fonctionnalités avancées pour la modélisation, l'exécution et l'optimisation.



Power Automate

Power Automate est une plateforme de gestion de processus basée sur le cloud offrant des fonctionnalités de création, d'automatisation et de suivi des flux de travail. Power Automate permet de connecter différents services et applications hébergés dans le Cloud ou on premise pour créer des flux de travail automatisés et simplifier les processus métier. Il présente l'avantage d'exploiter le standard de modélisation BPMN facilitant ainsi une modélisation cohérente, efficace et compréhensive de toutes les parties prenantes.

4.

Data Mesher's Journey

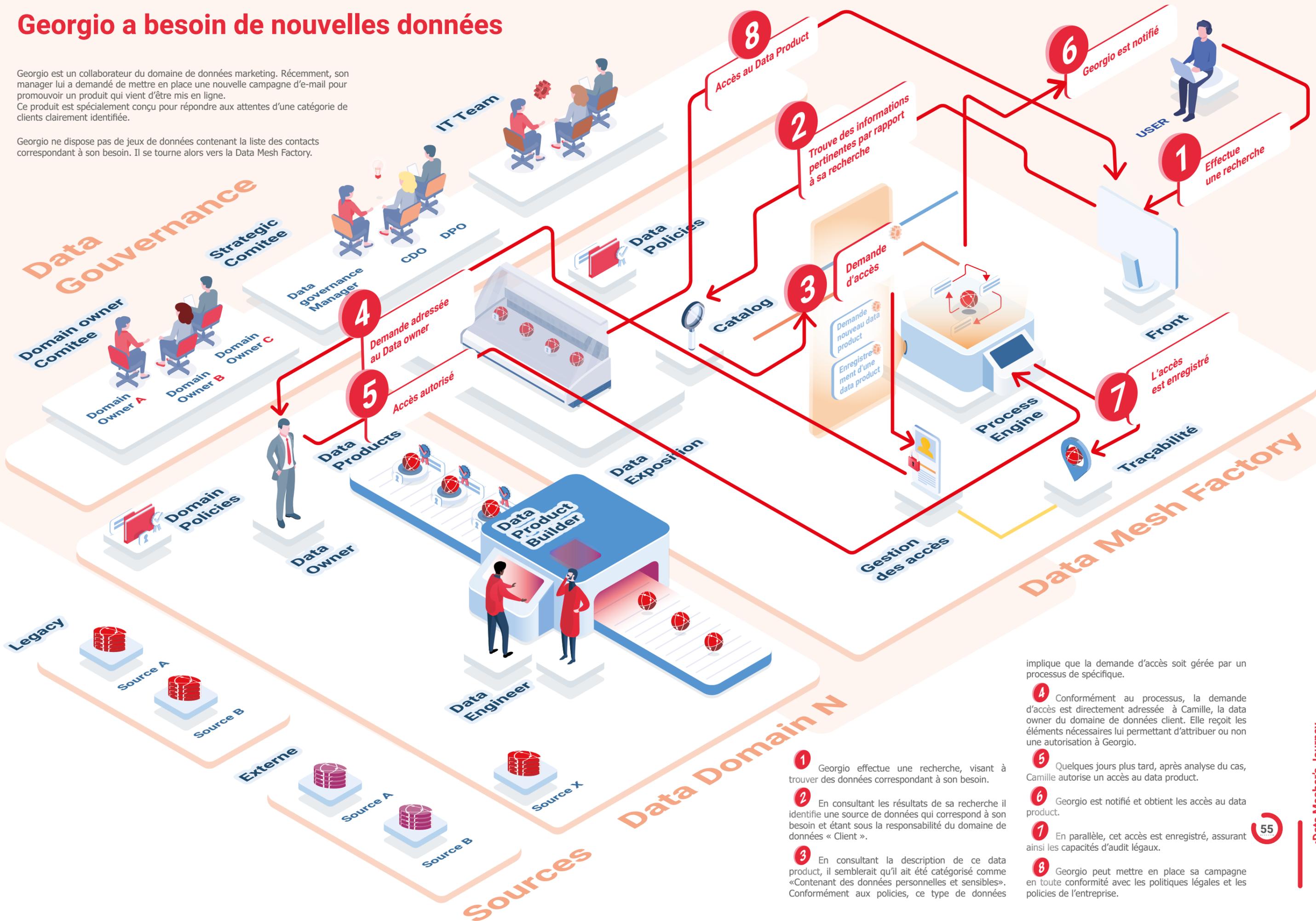


Synot^{is}

Georgio a besoin de nouvelles données

Georgio est un collaborateur du domaine de données marketing. Récemment, son manager lui a demandé de mettre en place une nouvelle campagne d'e-mail pour promouvoir un produit qui vient d'être mis en ligne. Ce produit est spécialement conçu pour répondre aux attentes d'une catégorie de clients clairement identifiée.

Georgio ne dispose pas de jeux de données contenant la liste des contacts correspondant à son besoin. Il se tourne alors vers la Data Mesh Factory.



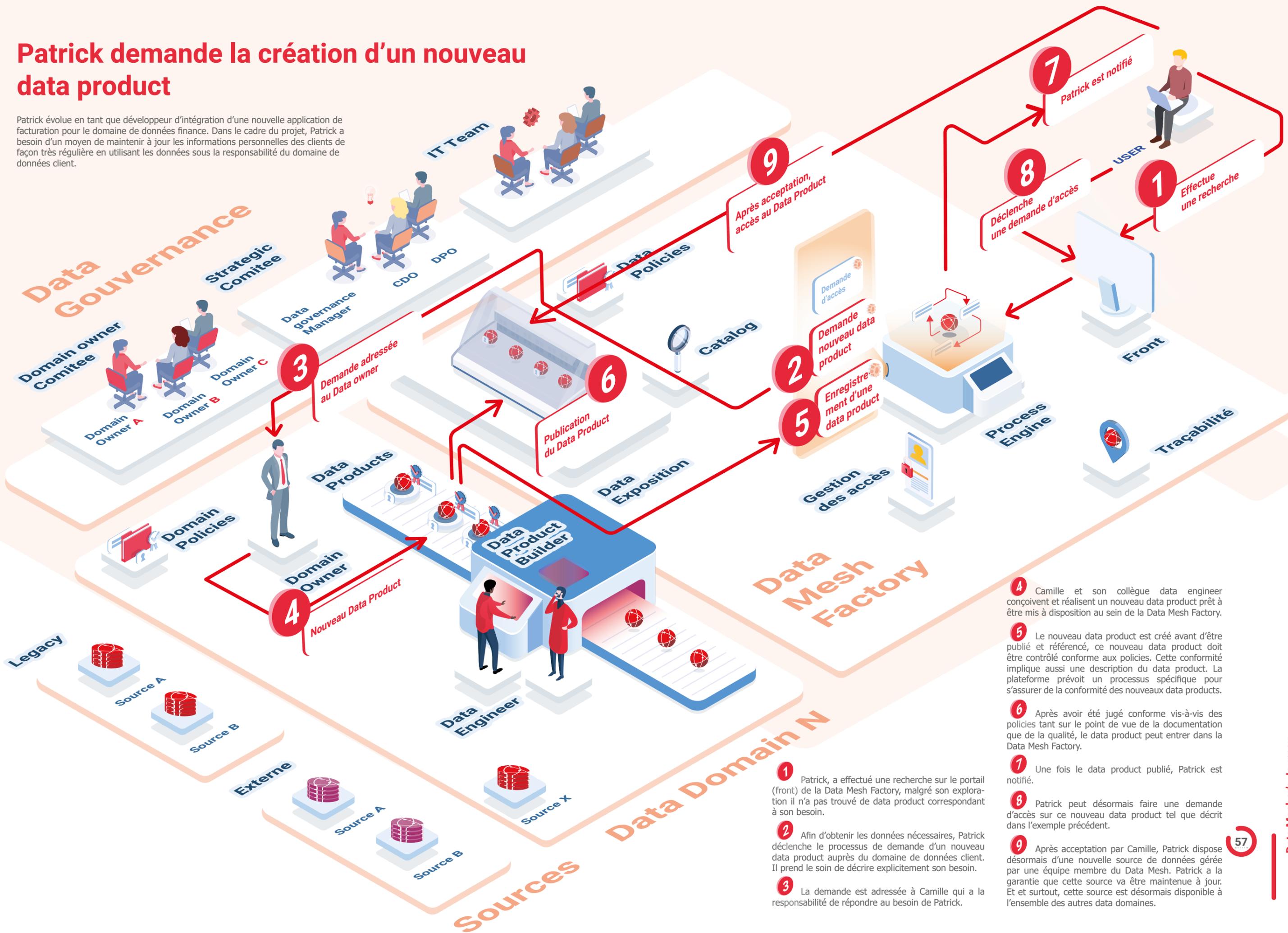
- 1 Georgio effectue une recherche, visant à trouver des données correspondant à son besoin.
- 2 En consultant les résultats de sa recherche il identifie une source de données qui correspond à son besoin et étant sous la responsabilité du domaine de données « Client ».
- 3 En consultant la description de ce data product, il semblerait qu'il ait été catégorisé comme «Contenant des données personnelles et sensibles». Conformément aux politiques, ce type de données

implique que la demande d'accès soit gérée par un processus de spécifique.

- 4 Conformément au processus, la demande d'accès est directement adressée à Camille, la data owner du domaine de données client. Elle reçoit les éléments nécessaires lui permettant d'attribuer ou non une autorisation à Georgio.
- 5 Quelques jours plus tard, après analyse du cas, Camille autorise un accès au data product.
- 6 Georgio est notifié et obtient les accès au data product.
- 7 En parallèle, cet accès est enregistré, assurant ainsi les capacités d'audit légaux.
- 8 Georgio peut mettre en place sa campagne en toute conformité avec les politiques légales et les politiques de l'entreprise.

Patrick demande la création d'un nouveau data product

Patrick évolue en tant que développeur d'intégration d'une nouvelle application de facturation pour le domaine de données finance. Dans le cadre du projet, Patrick a besoin d'un moyen de maintenir à jour les informations personnelles des clients de façon très régulière en utilisant les données sous la responsabilité du domaine de données client.



4 Camille et son collègue data engineer conçoivent et réalisent un nouveau data product prêt à être mis à disposition au sein de la Data Mesh Factory.

5 Le nouveau data product est créé avant d'être publié et référencé, ce nouveau data product doit être contrôlé conforme aux politiques. Cette conformité implique aussi une description du data product. La plateforme prévoit un processus spécifique pour s'assurer de la conformité des nouveaux data products.

6 Après avoir été jugé conforme vis-à-vis des politiques tant sur le point de vue de la documentation que de la qualité, le data product peut entrer dans la Data Mesh Factory.

7 Une fois le data product publié, Patrick est notifié.

8 Patrick peut désormais faire une demande d'accès sur ce nouveau data product tel que décrit dans l'exemple précédent.

9 Après acceptation par Camille, Patrick dispose désormais d'une nouvelle source de données gérée par une équipe membre du Data Mesh. Patrick a la garantie que cette source va être maintenue à jour. Et surtout, cette source est désormais disponible à l'ensemble des autres data domaines.

1 Patrick, a effectué une recherche sur le portail (front) de la Data Mesh Factory, malgré son exploration il n'a pas trouvé de data product correspondant à son besoin.

2 Afin d'obtenir les données nécessaires, Patrick déclenche le processus de demande d'un nouveau data product auprès du domaine de données client. Il prend le soin de décrire explicitement son besoin.

3 La demande est adressée à Camille qui a la responsabilité de répondre au besoin de Patrick.

Vous avez encore des questions ?

Nous espérons que vous avez pris autant de plaisir à le lire que nous à le rédiger. Comme vous avez pu le constater, il existe de nombreuses possibilités pour mettre en oeuvre les concepts du Data Mesh. Cela va dépendre de la maturité de votre entreprise quand à l'utilisation des données, la culture qu'elle a vis-à-vis des données et bien sur de votre domaine d'activité.

Si vous souhaitez aller plus loin dans la compréhension des concepts du Data Mesh et imaginer de quelle façon vous pourriez mettre cela en musique dans votre organisation, nos experts de la BL conseil et nos architectes de la BL Data Management peuvent venir à votre rencontre.

Au plaisir de vous rencontrer.

N'hésitez pas à nous contacter

- Téléphone : 021 519 06 71
- Mail : contact@synotis.ch



Robert Bounheng

**CoFondateur &
Directeur du développement**



Synotis

A SMILE GROUP COMPANY

Avenue de Rhodanie 64B, 1007 Lausanne

021 519 06 71

contact@synotis.ch

