

Le consultant indépendant en gestion des informations Ronald Damhof est l'auteur du modèle de quadrant de données

« Faire de la **gestion des données** une question d'actualité à l'échelle de l'entreprise »

**Ronald
Damhof**

CONSULTANT
INDÉPENDANT EN
GESTION DE
L'INFORMATION

Le domaine de la gestion des données regorge de toutes sortes de jargons. La plupart des dirigeants d'entreprise n'ont pas la moindre idée de ce que signifient tous ces mots, sans parler de les utiliser pour évoquer la valeur précise de données particulières, et la manière de les traiter. Pour permettre aux entreprises d'ouvrir une discussion sur les données, Ronald Damhof a développé le modèle de quadrant de données.

Ronald Damhof est consultant indépendant en gestion de l'information auprès d'établissements tels que Ahold, la banque centrale des Pays-Bas, l'administration fiscale néerlandaise, Alliander et des entreprises des secteurs financier et de la santé. Ces structures, qui font un usage intensif de données, prennent chaque jour conscience que la qualité de leur travail est de plus en plus liée à celle de leurs données. Mais comment transformer cette prise de conscience en stratégie efficace ? Une stratégie compréhensible par chaque membre de l'entreprise, du directeur à l'ingénieur informatique ? Ronald Damhof a développé un modèle de quadrant pour faire de la gestion des données une question d'actualité.

Pousser ou tirer ?

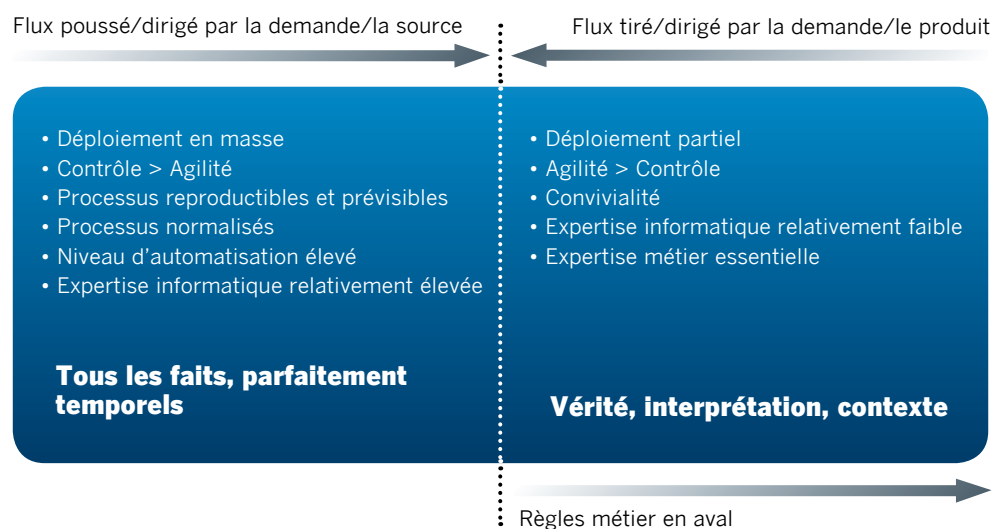
Ronald Damhof commence par expliquer un concept dont chacun d'entre nous a déjà entendu parler : le « pousser-tirer », qui concerne l'ampleur des effets de la demande sur le processus de production. Il prend l'exemple de la construction d'un yacht de luxe, dont le processus ne démarre que lorsque le client passe commande. Le point de découplage se situe alors au début du processus de production. Autre exemple, totalement opposé : la fabrication d'allumettes. Si un client veut des allumettes, il se rend au supermarché et en achète (sauf s'il souhaite des allumettes noires, plus difficiles à trouver). Ici, le point de découplage se situe complètement à la fin du processus de production. La construction d'une voiture, en revanche, implique des pièces standard et des éléments sur mesure. Les clients peuvent en effet demander une couleur particulière, une sellerie en cuir ou des jantes spécifiques. Dans ce cas, le point de découplage se situe quelque part au milieu du processus de production. « De même, dans la production d'un rapport, d'un tableau de bord ou d'un environnement analytique, le point de découplage se situe quelque part dans cette zone du milieu », explique Ronald Damhof.

Le point de découplage divise le processus de production en deux parties : un côté tiré et un côté poussé, également appelés parties dirigée par l'offre et dirigée par la demande. Les systèmes à flux poussé visent à réaliser des économies d'échelle en cas d'augmentation du volume et de la demande, tout en garantissant la qualité du produit et des données associées. À l'inverse, les systèmes à flux tiré sont dirigés par la demande. Chaque type d'utilisateur souhaite manipuler les données de façon à produire « son » produit, sa réalité, sur la base de sa propre expertise et de son contexte.

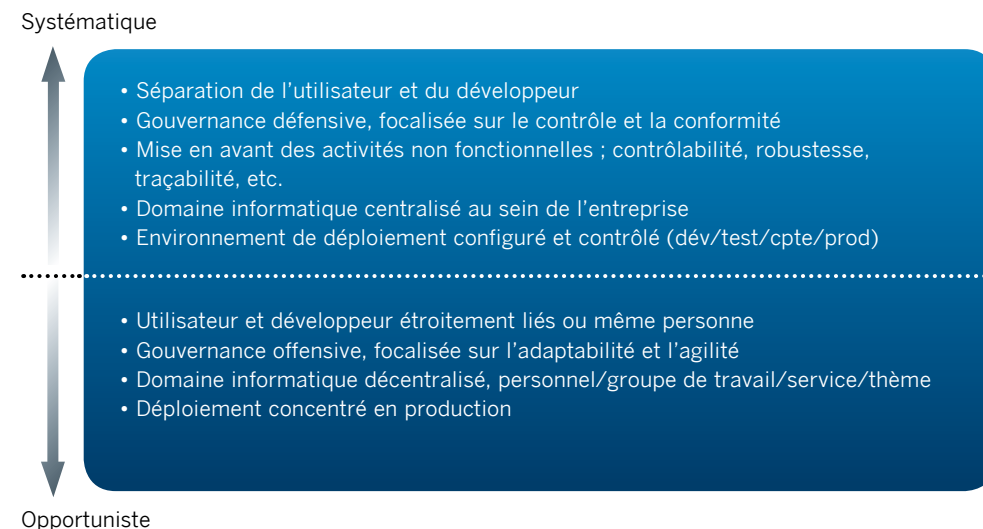
Développement opportuniste ou systématique ?

Sur l'axe Y, Ronald Damhof projette la dimension du style de développement. « Ce qui signifie : comment développez-vous un produit d'informations ? Il est ici possible d'agir de manière systématique, auquel cas l'utilisateur et le développeur sont deux personnes distinctes et vous appliquez une gouvernance défensive, axée sur le contrôle et la conformité. Cela met en pratique tout ce que les ingénieurs ont appris pour créer des logiciels sur une base solide. Ce schéma est souvent visible dans les données d'entreprise centralisées, comme les données financières et celles transmises aux autorités de contrôle ». Il est également possible d'adopter un style de développement opportuniste. « Dans ce cas, le développeur et l'utilisateur ne sont souvent qu'une seule et même personne. Prenons l'exemple du statisticien en quête de données innovantes pour produire et expérimenter des modèles analytiques, ou différentes situations où la rapidité de livraison est essentielle. Dans ces deux cas, la gouvernance est souvent offensive, et ce sont la flexibilité et l'adaptabilité qui priment ».

Le point de découplage des données



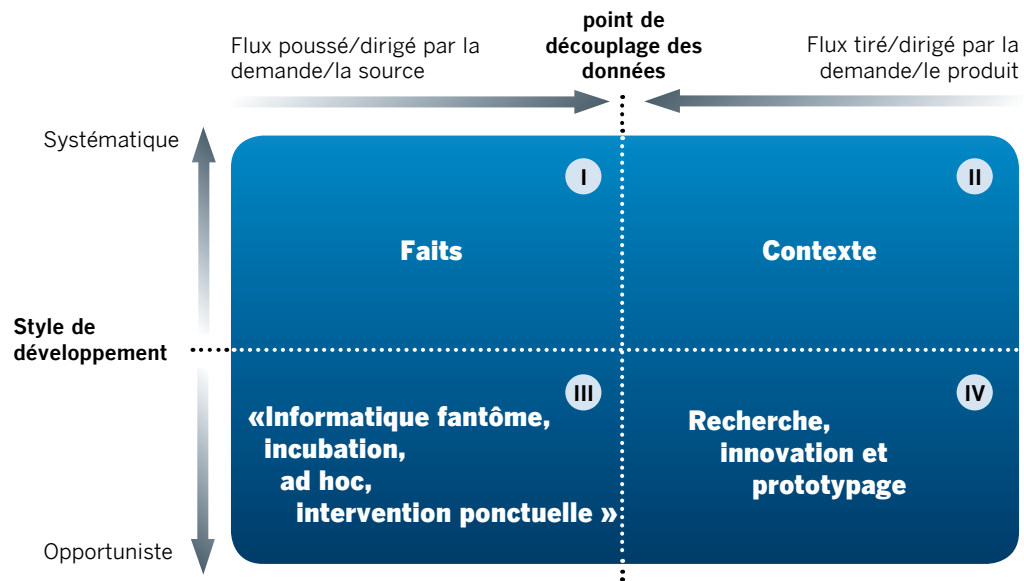
Style de développement



Modèle de quadrant de données

La réunion de ces deux dimensions produit le tableau suivant.

Quadrant de déploiement de données



« Les faits réels se trouvent dans le quadrant I », explique Ronald Damhof. « Ces données peuvent être intégralement fournies de manière intelligible aux quadrants II et IV. Les données du quadrant I étant produites par des systèmes et processus hautement normalisés, elles sont entièrement prévisibles et reproductibles ».

Les données du quadrant IV, diagonalement opposées, se caractérisent par l'innovation et le prototypage. « C'est le quadrant dans lequel travaillent les spécialistes des données, qui n'ont en réalité que trois exigences : données, puissance informatique et convivialité du logiciel ». Différentes équipes sont progressivement mises en place à mesure que les laboratoires qui alimentent ces « scientifiques » en données leur donnent carte blanche pour les utiliser à des fins d'expérimentation et d'analyse, dans une optique d'innovation. « Ce type de gestion des données est nécessaire pour avoir des idées et les expérimenter. Lorsqu'un concept fonctionne, il doit remonter du quatrième au deuxième quadrant, car seules la génération et l'analyse systématiques de données permettent des économies d'échelle dans ce domaine. Le concept peut ensuite être généralisé à toute l'entreprise ».

« Je discute souvent avec des spécialistes des données qui tirent de très précieux enseignements d'un environnement de type sandbox », poursuit Ronald Damhof. « Mais ils oublient ou sont incapables de monétiser ces connaissances dans un système de production. Ils ne parviennent pas à les faire passer du quadrant IV au quadrant II. C'est là qu'intervient la gouvernance ». Mais c'est également à ce stade que réside la principale difficulté pour bien des entreprises, phénomène bien



« Les entreprises génèrent des volumes de données toujours plus importants et ne peuvent plus se permettre d'être aussi négligentes que par le passé. Elles doivent à présent s'assurer que la gestion des données et la gouvernance associée sont correctement définies. C'est précisément le rôle du modèle de quadrant de données »

Ronald Damhof

connu de Ronald Damhof. « Si vous expliquez ce modèle aux responsables et les interrogez sur leurs priorités, ils répondent invariablement qu'ils doivent commencer par mettre en place des bases solides – le premier quadrant. Mais si vous leur demandez dans quoi ils investissent actuellement, dans quels domaines ils innovent, il s'agit souvent du quatrième quadrant. C'est très bien de s'engager ainsi dans une forme de gestion des données plus expérimentale et exploratoire, mais ce n'est possible que si les bases sont solides. Sinon, cela revient à être équipé de toilettes ultra-modernes qui ne sont pas raccordées au réseau d'assainissement, et c'est le chaos ». Demandez au statisticien moyen ce qui lui prend le plus de temps et il vous répondra : récupérer les données au niveau de qualité adapté – ce qui est l'objectif du quadrant 1. « Seul un spécialiste des données disposant d'un solide logiciel analytique, d'une grande puissance informatique et de données de qualité fera véritablement la différence ».

Fiabilité et flexibilité

« Les dirigeants insistent sur la fiabilité et la flexibilité des systèmes, mais ces qualités sont inversement proportionnelles. Un système particulièrement fiable et robuste est moins souple, tandis qu'un système extrêmement flexible implique de revoir ses exigences de fiabilité à la baisse », décrypte Ronald Damhof. « C'est parfaitement clair dans le modèle de quadrant de données. Dans le quadrant I, la fiabilité prévaut sur la flexibilité, et c'est l'inverse dans les quadrants II et IV ». Une utilisation optimale des données exige par conséquent une expertise et des compétences différentes.

Expertise et compétences

Les entreprises exigent souvent qu'une seule personne offre l'expertise et les compétences couvrant l'intégralité du quadrant. Cette personne n'existe pas. Les employés du quadrant I ont un profil d'ingénieur. Ce sont des spécialistes de l'informatique et des données, ayant une formation en architecture et modélisation. « Vous remarquerez qu'il ne s'agit pas du profil informatique type. Ces ingénieurs sont capables d'effectuer un développement orienté modèle et comprennent parfaitement les besoins en modélisation conceptuelle et logique ». Un tel niveau d'expertise est très rare. À l'opposé, les quadrants II et IV exigent des spécialistes dans leur domaine respectif, ayant en outre des compétences en informatique décisionnelle et/ou en analyse.

« Dans le monde de l'entreprise, il est souvent question d' »une seule version de la vérité », mais cela n'existe pas. Il existe « une seule version des faits » et plusieurs « vérités ». Après tout, la manière dont chacun interprète les faits varie selon le type d'entreprise, les perspectives, les connaissances et l'expérience ».

Ronald Damhof

Les faits et la vérité

Ronald Damhof surnomme aussi le quadrant I « la seule version des faits ». Ces derniers sont mis à la disposition des employés au niveau des quadrants II et IV, et leur permettent d'établir leur propre vérité. Dans la mesure où les mêmes faits servent à créer plusieurs vérités dans la moitié droite du modèle – selon le contexte et les connaissances de l'utilisateur des données – Ronald Damhof appelle cette moitié « les multiples versions de la vérité ». Vous devez garder à l'esprit que la « vérité » change souvent au fil du temps. « Dans les entreprises, il est souvent question d' »une seule version de la vérité », mais cela n'existe pas. Après tout, la manière dont chacun interprète des faits particuliers varie selon le contexte, les perspectives, les connaissances et l'expérience ».

Quadrant III

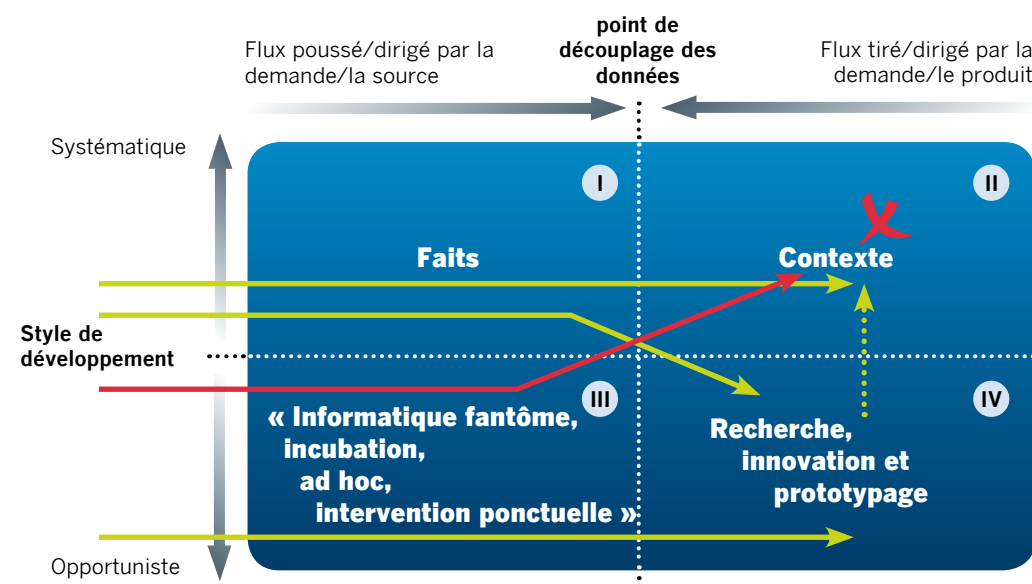
Le quadrant III, rarement évoqué jusqu'à présent, est d'une importance capitale. C'est le quadrant des sources de données qui ne font l'objet d'aucune gouvernance, comme un téléchargement ad hoc auprès d'un fournisseur de données ouvertes, une liste Excel dont vous souhaitez vous servir ou une série de données de vérification reçues sur un CD. « Vous pouvez même associer sans problème des données administrées du quadrant I à votre propre jeu de données dans le quadrant IV », affirme Ronald Damhof.

La traversée des quadrants

Pour valoriser vos données, vous pouvez effectuer différents déplacements dans le modèle. Vous pouvez ainsi passer d'une gestion des données factuelle à un modèle où le contexte est également important (C'est-à-dire des quadrants I à II). « C'est vraiment le parcours classique de « déblocage des données et d'élaboration d'un produit d'informations » », analyse Ronald Damhof. Il est pourtant souvent inefficace, car ce processus suppose des exigences et des attentes connues de l'utilisateur. « Or l'utilisateur ne dispose pas réellement de ces connaissances à l'avance ». La plupart des entreprises adoptent une démarche axée sur l'agilité, en passant par exemple des quadrants I à II via le IV. Demandez aux employés du quadrant IV d'élaborer un produit d'informations de manière itérative à partir des données du quadrant I ou III. Remontez ensuite le produit au quadrant II, uniquement s'il est important de le gérer.

Il est également possible de passer du quadrant III au quadrant IV. « Vous disposez de jeux de données propres et vous souhaitez faire des essais ? Parfait », note Ronald Damhof. Le seul mouvement totalement proscrit consiste à passer du quadrant III au quadrant II. « Car dans ce cas, vous utilisez des données dont vous n'êtes pas entièrement sûr, puisqu'elles n'ont pas fait l'objet d'une gouvernance éclairée, contrairement aux prescriptions. Un exemple à éviter à tout prix : une déclaration de conformité destinée à un organisme de contrôle, que vous souhaitez produire à partir de données non soumises à une gouvernance ».

How we produce, process variants



Faire de la gestion des données une question d'actualité

Dans le cadre de son travail quotidien, Ronald Damhof constate que son modèle aide les entreprises à parler de gestion des données. « Chez mon client actuel, la Banque des Pays-Bas, j'entends régulièrement des phrases telles que « Je veux faire passer ce produit de données du quadrant IV au quadrant II », « Nous devons tout d'abord placer les données dans le quadrant I, mais les données de ce quadrant sont sous la responsabilité de l'émetteur » ou « J'ai besoin d'espace pour stocker provisoirement les données dans le quadrant III ». Tout le monde comprend ce dont il s'agit, ce qui est nouveau. L'établissement n'avait jamais envisagé les données sous cet angle auparavant, et cela peut vraiment s'appliquer à toutes les sociétés qui font un usage intensif de données. Les entreprises parlent depuis longtemps des données comme d'un « actif » mais dans les faits, elles les gèrent de façon très déstructurée. Résultat : elles ne monétisent jamais cet actif. Les entreprises génèrent des volumes de données toujours plus importants et ne peuvent plus se permettre d'être aussi négligentes que par le passé. Elles doivent à présent s'assurer que la gestion des données est correctement définie. C'est précisément le rôle du modèle de quadrant de données ». ■