

# YANDOX Taxi et Big Data : un cas d'efficacité d'usage

*Ce cas est un extrait de la communication de E. Grigoryan et K. Lobre-Lebraty au Congrès AIM 2020.  
La version complète est en ligne sur le site de l'AIM*

## L'efficacité d'usage d'un système d'information

Pour Burton-Jones et Grange (2013), l'efficacité d'usage d'un SI (Système d'Information) est un usage du SI qui contribue à l'amélioration de l'atteinte des objectifs de l'organisation. Ces auteurs mettent alors en avant une théorie de l'efficacité d'usage (TEU) dans le SI. Les 3 éléments vecteurs de l'efficacité d'usage sont : les compétences et les motivations des usagers (1) ; l'intentionnalité du système (2) ; les caractéristiques des tâches (3). On peut constater ainsi que la Théorie de l'Efficacité d'usage (TEU) est une évolution naturelle des modèles TAM et UTAUT liée au passage nécessaire « de l'étude de l'usage » à « l'étude de l'efficacité d'usage ». Différentes recherches reliant théorie de l'efficacité d'usage et Big Data ont permis d'alimenter le modèle proposé par Surbakti et al. (2019) donnée en annexe. Ce modèle prend en compte les caractéristiques spécifiques du Big Data afin de contextualiser l'efficacité d'usage. Le modèle comprend les 41 facteurs, regroupés en 7 thèmes qui sont eux-mêmes classifiés en 3 catégories.

Le présent article fait état d'une recherche en cours dans le cadre d'un projet de thèse plus large sur l'efficacité d'usage du Big Data. L'objectif est ici de mettre en évidence la présence des facteurs contenus dans les 7 thèmes du modèle de Surbakti et al. (2019). Le directeur régional de Yandex Arménie a été retenu comme unique interviewé dans un premier temps, car la position de l'interlocuteur au sein de cette structure nous permet de nous situer à un niveau plus stratégique qu'opérationnel d'une part, et pour acquérir une véritable compréhension du fonctionnement de Yandex, en particulier au regard de l'utilisation du Big Data et des attentes qui y sont attachées, d'autre part. Au total, 4 entretiens non-directifs ont été réalisés d'une durée d'environ 1h15 chacun.

## Description du cas

Yandex est une organisation transnationale russe, un moteur de recherche, qui s'est installée en Arménie en 2018. Plus précisément, Yandex est un multi-portail qui propose différents services : Yandex Recherche, Yandex Maps, Yandex Market, Yandex Money, Yandex News, etc. C'est un moteur de recherche d'envergure mondiale. La mission de cette entreprise est d'aider les consommateurs (particuliers et entreprises) à mieux naviguer dans les mondes en ligne ou hors-ligne.

Cette entreprise technologique crée des produits et services intelligents utilisant le Machine Learning (ML), c'est-à-dire une technologie d'intelligence artificielle permettant aux ordinateurs d'apprendre sans avoir été programmés explicitement à cet effet. Cet apprentissage des ordinateurs repose sur les gros volumes de données disponibles à analyser. Ainsi Big Data et ML entretiennent des liens étroits, puisque le Big Data est la « matière première » du Machine Learning, qui constitue la technologie permettant d'exploiter pleinement le Big Data. Yandex dispose donc d'un vrai savoir-faire et d'une forte expertise dans le domaine de la gestion du Big Data. Remarquons que celle-ci est déjà ancienne, puisque dès 2007 elle ouvrait une école d'analyse des données, et qu'elle propose actuellement des formations en gestion des données massives. C'est en particulier cette expertise qui fait de Yandex un cas particulièrement intéressant pour tester empiriquement le modèle de l'efficacité d'usage du Big Data de Surbakti et al. (2019).

Comme terrain de notre recherche, nous avons choisi l'une des entités de Yandex Group, celle du service de Yandex Taxi Arménie. Yandex Taxi est un service de covoiturage lancé en 2011 par Yandex. En février 2018, pour le partage du marché, Yandex Taxi a fusionné avec un autre service de taxi en Russie, Arménie, Azerbaïdjan, Belarus, Géorgie et Kazakhstan, en créant une nouvelle entreprise privée immatriculée aux Pays-Bas. Cette entreprise fusionnée est possédée à 59.3% par Yandex, 36.9% par Uber son principal concurrent et partenaire, et 3.8% par les salariés du Groupe.

## Nature du service

Yandex Taxi est un service de taxi en ligne fondée sur la maraude et une application de géolocalisation. Lors de son apparition, ce type de service a été considéré comme « une innovation de rupture » par les chercheurs qui s'intéressaient surtout alors à son impact sur l'économie, l'environnement et la société. Il se trouve que ce service novateur de taxi a totalement révolutionné l'industrie du service traditionnel des taxis. Ce service proposé par Yandex fonctionne actuellement dans 15 pays à travers le monde. La Russie reste leur marché

essentiel, mais Yandex le propose également aujourd'hui dans les pays suivants : Arménie, Bélarusse, Géorgie, Kazakhstan, Israël, Côte d'Ivoire, Kirghizistan, Lettonie, Lituanie, Moldavie, Serbie, Ouzbékistan, Finlande et Estonie.

### **Fonctionnement du service et Big Data**

A l'instar d'Uber, Yandex Taxi ne possède pas de voiture, ni de chauffeurs salariés. Pour servir ses clients, l'entreprise met en relation des chauffeurs-partenaires et des passagers via une application mobile. A l'aide de cette application, le passager peut commander une course dans les villes où le service est disponible. La commande passée est reçue par les chauffeurs à proximité. Une fois acceptée par l'un d'entre eux, l'application indique au client le délai d'arrivée du chauffeur sur le lieu de prise en charge, puis elle lui signale la présence du chauffeur sur la place. L'application transmet également au client la marque et le modèle du véhicule, le numéro de taxi ainsi que le moyen de contacter le chauffeur. Le client peut quant à lui indiquer au chauffeur s'il le souhaite, l'itinéraire à emprunter. L'application lui permet également d'obtenir une estimation du prix de la course qui se calcule automatiquement. Une fois arrivé à destination, le prix de la course peut, soit être débité sur la carte du client, soit être payé en espèces. Enfin pour clore la transaction, l'application sollicite le client pour noter son chauffeur et laisser un commentaire, le chauffeur peut également laisser son commentaire concernant son client.

Cette brève description du processus de fonctionnement de Yandex Taxi permet d'illustrer les empreintes numériques générées par le service. Ces empreintes sont la source du Big Data de Yandex. Les données y sont stockées et traitées en temps réel. Par exemple, le suivi du trajet de la course en temps réel permet de détecter le comportement potentiellement irrégulier d'un chauffeur, qui sans raison ne suivrait pas l'itinéraire recommandé par l'application pour une course donnée, itinéraire qui est généré par l'algorithme « On A Trade ». Autre exemple, l'algorithme « hausse de prix » permet de connaître l'évolution de la demande du service au temps réel, d'estimer l'élasticité de la demande par rapport au prix, et de gérer ainsi le surplus de clients sur une zones donnée, à un moment donné.

Les algorithmes sophistiqués de traitement du Big Data de Yandex permettent ainsi à l'entreprise d'accomplir trois activités majeures : déterminer la demande, octroyer les ressources nécessaires pour satisfaire cette demande et fixer les tarifs. Les capteurs GPS qui équipent les véhicules du service de taxi laissent de nombreuses empreintes numériques des activités et des mouvements quotidiens de leurs clients. Elles constituent ainsi d'importantes et riches sources de données (Big Data), qui combinées à d'autres données permettent de mieux comprendre et d'analyser en profondeur les dynamiques de comportement d'un individu ou d'un groupe social.

Pour conclure cette présentation soulignons que Yandex utilise ses propres technologies de cartographie (mapping), de routage (routing) et de navigation. Une partie du Big Data créée dans le cadre de l'activité de taxi est composée de données structurées (celles issues des capteurs GPS par exemple), de données semi-structurées et données non-structurées (notes et commentaires des clients et des chauffeurs-partenaires par exemple), ce qui suppose notamment des capacités de traitement adaptées pour une utilisation efficace.

### **Résultats**

L'analyse thématique nous a permis d'identifier la présence des 7 thèmes du modèle de Surbakti et al. (2019) dans l'exploitation du Big Data de Yandex Taxi. En plus, pour certains thèmes, nous avons pu identifier clairement certains facteurs inclus dans les thèmes (ces facteurs sont marqués en rouge dans la Figure 1), ce qui permet d'enrichir notre analyse. Les paragraphes suivants présentent brièvement les résultats sur chaque thème du modèle, en insérant les verbatims les plus illustratifs.

#### **F1 : Intérêt organisationnel perçu**

Ce thème peut être considéré comme la priorité initiale de Yandex pour construire un avantage comparatif. « *Yandex est née technologique, sa mission c'est aider ses consommateurs à mieux naviguer dans les mondes en ligne ou hors-ligne, en suivant et utilisant les avancements technologiques dans les services et produits proposés. Sa stratégie c'est de proposer un meilleur service grâce à la technologie* ».

Nous avons repéré les facteurs suivants :

• Valeur perçue du Big Data (f1.1) et son efficacité perçue (f1.3) (voir Figure 1) : « *...chez nous, la valeur des services et celle de toute l'organisation repose sur nos capacités technologiques d'exploitation du Big Data ; ce type de service de taxi n'a de sens sans l'exploitation du Big Data ...* » ; « *...pour un projet de service de taxi, on ne voit pas de grande valeur dans un modèle traditionnel aujourd'hui* ».

• Facilité d'utilisation perçue (f1.4) : « *l'exploitation du Big Data dans ce service innovateur de taxi rend le service plus rapide, plus sécuritaire, plus confortable et donc plus agréable...* ».

☐ observabilité perçue (f1.6) et analyse coût-avantage (f1.7) : « ...par rapport au service traditionnel du taxi, ce service innovateur renvoie aux résultats plus visibles et plus avantageux dans les divers contextes ... ».

On observe ainsi, que cette catégorie « motivationnel » est éminemment stratégique dans le cas d'une entreprise technologique puisque le Big Data et son usage sont intimement liés aux objectifs stratégiques de l'entreprise. Autrement dit l'intérêt organisationnel perçu du Big Data dans une organisation née technologique apparaît nécessairement fort et par ricochet son impact sur l'efficacité d'usage. Pour autant les perceptions des opérationnels devront compléter ces premiers résultats, celles des chauffeurs dans le cas de Yandex Taxi, mais aussi celles des data scientists et autres opérationnels dont le Big Data est au coeur de l'activité quotidienne.

## **F2 : Management des processus**

Dans la catégorie « opérationnel », le premier thème est celui du management des processus. Il s'agit de la capacité d'une organisation à structurer ou restructurer ses processus pour tirer parti des opportunités du Big Data. Le succès de Yandex Taxi souligne un rapport qualité-prix compétitif du service proposé par l'entreprise. Plus précisément Yandex Taxi a su prendre des parts de marché au service de taxi traditionnel, en proposant un service de qualité supérieure, à un prix supérieur au service traditionnel sans toutefois être proportionnel au niveau d'augmentation de la qualité.

Nous avons pu repérer tous les facteurs de ce thème opérationnel :

☐ orientation vers processus (f2.1) : « ...c'est sûrement la puissance opérationnelle qui est la plus critique pour l'exploitation du Big Data, la réussite pour un projet axé sur big data... » ; « ...notre entreprise bénéficie de la forte compétence et expérience développée par le Groupe dans ses diverses activités, notamment dans la gestion de ses données et dans l'alignement de ses processus métiers avec ceux de Big Data ».

☐ intégration du business processus SI (f2.2) et management des données (f2.3) : «...sûrement, pour exploiter le big data, l'entreprise doit avoir une maturité en management de ses processus..., le management de ses données : les opérations avec les bases de données pour qu'elles soient propres, la sécurité des données, leur stockage, etc... ».

☐

## **F3 : Confidentialité, sécurité et gouvernance des données**

Le deuxième thème de la catégorie « opérationnel » concerne la confidentialité, sécurité et gouvernance des données. La gouvernance des données s'applique à l'échelle d'une organisation. Elle se définit comme l'ensemble des procédures mises en place au sein de cette organisation afin d'encadrer la collecte de données et leur utilisation. Il s'agit à la fois de respecter les obligations légales imposées, et d'instaurer un cadre permettant d'optimiser l'utilisation des données. Ce thème joue un rôle majeur dans la confiance que les clients accordent à une entreprise, et donc dans leur propension à lui confier des données personnelles. Par ailleurs, sans données pas d'exploitation du Big Data et pour une entreprise comme Yandex Taxi pas de service de taxi !

Dans les propos de son directeur régional, nous avons pu repérer les deux facteurs contenus dans ce 2ème thème opérationnel :

☐ confidentialité et sécurité (f3.1) et gouvernance des données (f3.2) : « ...nous stockons d'énormes masses de données personnelles, avant de les traiter, nous sécurisons leur confidentialité au niveau opérationnel... Une fois l'application téléchargée, le client doit accepter les modalités et les conditions d'emploi qui définissent particulièrement que l'utilisateur donne son accord à Yandex Taxi de traiter son information personnelle, y compris, ses données personnelles. En plus il donne son consentement aussi pour que cette information soit transmise aux partenaires de l'entreprises pour de futurs traitements... En plus le passager accepte que toutes ses conversations avec des personnes de l'entreprise puissent être enregistrées... ».

Il apparaît clairement que chez Yandex Taxi la réglementation concernant la captation et l'utilisation des données personnelles est respectée. Par ailleurs, des mécanismes techniques et organisationnels sont déployés afin de sécuriser et protéger les données contre la destruction non autorisée, la perte, l'abus et des formes illégales de traitement.

## **F4 : Qualité des données**

Le dernier thème de la catégorie « opérationnel » concerne la qualité des données. Pour une entreprise comme Yandex Taxi, dont les données sont la principale ressource clé du modèle d'affaire, ce thème est fondamental. Dans ce thème-là, nous avons repéré les facteurs suivants :

☐ qualité des données et qualité de l'information (f4.1) : « ...le point le plus délicat avec l'exploitation du Big Data chez nous, c'est qu'on n'a pas de contrôle sur la qualité des données dans leur ensemble... ».

☐ complétude des données (f4.2) et cohérence des données (f4.7) : « ...une autre partie vient des sources comme les commentaires et les enregistrements des conversations, ou les déclarations... leur qualité est moins facile à vérifier, elles sont de la nature semi-structurée et non-structurée, et la cohérence n'est pas toujours possible à constater... » ; « ...l'idéal c'est quand par exemple la note du client peut se compléter par un commentaire qui est prouvé par les empreintes structurées du GPS, mais ce n'est pas toujours le cas... »

☐ pertinence des données (f4.5) : « ...une autre partie des données qui vient des empreintes numériques des capteurs GPS, constitue la majeure partie pour l'exécution de l'activité principale de la firme, ce sont des données faciles car elles sont pour la plupart de nature structurée... ».

☐ exactitude des données (f4.6) : « ...une grosse partie des données sont saisies dans la base par les clients et les partenaires eux-mêmes, mais ils disent ce qu'ils veulent... ».

En réalité, tous les facteurs de ce thème sont présents dans les propos du directeur régional de Yandex. Il précise ensuite que c'est la capacité de l'entreprise, en termes de mécanismes de supports technologiques, qui va permettre d'appliquer des processus de nettoyage des données, pour assurer la qualité nécessaire à une utilisation efficace du Big Data.

### **F5 : Aspect humain**

L'aspect humain est le premier thème de la catégorie « mécanismes de support ». Il représente l'un des moteurs essentiels de l'efficacité d'usage du Big Data. C'est à travers le facteur humain que tous les autres thèmes sont maîtrisés. Dans ce thème, nous avons repéré les facteurs suivants :

• savoirs et compétences des individus (f5.1) et champions (f5.3) : « ...c'est vraiment très difficile de trouver des spécialistes du domaine, des data-scientistes, des techniciens, des chefs de projet avec une expertise assez forte en Big Data... et l'on rencontre ce problème partout, en Arménie, en Russie, en Europe... Parfois on peut rechercher un spécialiste pendant des mois...mais on recrute vraiment des professionnels de leur métier... ».

☐ confiance (f5.2) et engagement des salariés (f5.4) : «...évidemment, il y a une relation symbiotique entre l'entreprise et ces spécialistes : ils sont confiants que dans ce domaine, Yandex possède une forte expertise et la compétence, qui assure que tout souci lié aux tâches reçoit l'attention nécessaire et soit réglé, ce qui assure aussi leur fort engagement... ».

☐ participation des utilisateurs (f5.5) : « ...la plupart des individus recrutés sont chez nous des hommes dans la tranche d'âge 26-35. Même pour les chauffeurs-partenaires ... On constate que les personnes plus âgées ont plus de mal avec les technologies novatrices, et ont de l'anxiété à utiliser les nouveautés. Il est plus difficile de les inclure dans les processus. Dans un pays africain par exemple, nous avons des cours préparatoires pour les chauffeurs pendant plus d'un mois afin de les former à utiliser l'application...et c'est vraiment compliqué... ».

Le facteur relatif aux savoirs et compétences des individus est considéré comme crucial dans les processus d'exploitation du Big Data. Il apparaît également problématique car il y a pénurie de spécialistes. Dans ce contexte, l'attractivité de l'entreprise, puis l'implication et l'engagement de ces spécialistes, lorsqu'ils ont pu être captés, reposent en partie sur leur confiance dans les compétences de l'entreprise en Big Data.

### **F6 : Aspect organisationnel**

Dans ce 2ème thème de la catégorie de « mécanismes de support », nous avons repéré les facteurs suivants :

☐ compétence organisationnelle (f6.1) et accent sur l'innovation (f6.13) : « Yandex, née technologique a une expertise très forte dans le domaine qui assure une culture organisationnelle favorable à l'exploitation de Big Data. Sa mission c'est aider ses consommateurs à mieux naviguer dans les mondes en ligne ou hors-ligne, en suivant et utilisant les avancements technologiques dans les services et produits proposés. Sa stratégie c'est de proposer un meilleur service grâce à la technologie ».

☐ gestion des talents (f6.2), collaboration interdépartementale (f6.8), support de haute direction (f6.10), taille de l'entreprise (f6.7) : « ...le Groupe joue un rôle majeur sur l'activité Yandex Taxi en termes de partage des connaissances, d'expérience, des compétences, des bonnes pratiques, des perspectives de développement... Il s'agit bien d'un partage de ressources en particulier humaines et technologiques... ».

☐ alignement stratégique (f6.4), management du projet (f6.5), management de la performance (f6.6) : « ...le

*Big Data est dans le coeur de l'activité de Yandex...l'on développe et poursuit le "best case practice" dans le Groupe et évidemment, on applique souvent les méthodes prédéfinies et valides qui assure le succès... ».*

■ effet environnemental (f6.11) : « ... soulignons, que c'est au niveau du Groupe que l'accent est mis sur l'innovation et que la veille environnementale est assurée... ».

### **F7 : Systèmes, outils et technologies**

Le dernier thème de la catégorie « mécanismes de support » dans l'efficacité d'usage du Big Data regroupe les facteurs « qualité du système », « infrastructure TI » et « support des fournisseurs ». Ces facteurs assurent le soutien technologique nécessaire aux initiatives Big Data. Conformément au modèle de Surbakti et al. (2019), pour permettre une utilisation efficace du Big Data, les systèmes doivent être fiables, flexibles face aux nouvelles exigences ou conditions, permettre l'intégration des données provenant de différents ensembles, facilement accessibles aux utilisateurs, et doivent fournir des données en temps opportun tout en protégeant les informations. Par ailleurs, les initiatives Big Data ont besoin d'outils et technologies appropriés pour aider les organisations à exploiter à leur profit les données volumineuses. Elles ont besoin d'une infrastructure informatique flexible pour s'adapter facilement aux exigences d'utilisation diverses des projets Big Data. Elles ont besoin de plates-formes robustes pour traiter de multiples sources de données. Nous avons repéré tous les 3 facteurs composants du thème :

• qualité du système (f7.1) et infrastructure TI (f7.2) : « ...notre entreprise a développé ses propres technologies de cartographie, de routage, de navigation qui assurent la qualité du système pour le traitement du Big Data et l'adéquation de son infrastructure SI aux besoins du service... ».

■ support des fournisseurs (f7.3) : « ...en utilisant nos propres technologies de routage, cartographie, navigation, nous enlevons presque le risque de dépendance de bon fonctionnement des fournisseurs clés... ».

Les sept thèmes contribuant à l'efficacité d'usage du Big Data sont bien présents chez Yandex Taxi, terrain choisi pour tester empiriquement le modèle de l'efficacité d'usage du Big Data de Surbakti et al. (2019). Autrement dit, dans ce cas, chaque thème semble bien contribuer à ce que le Big Data contribue aux objectifs de Yandex Taxi.

### **Conclusion**

Dans cette recherche exploratoire chez Yandex Taxi les 7 thèmes du modèle semblent bien contribuer à l'efficacité d'exploitation du Big Data, autrement dit à la contribution du Big Data à l'atteinte de ses objectifs de performance. Cependant notre travail ne valide pas en totalité ce modèle. En effet, les interrelations entre les 3 catégories (motivationnel, opérationnel et mécanismes de supports) dans lesquelles sont regroupés les sept thèmes n'émergent que partiellement.

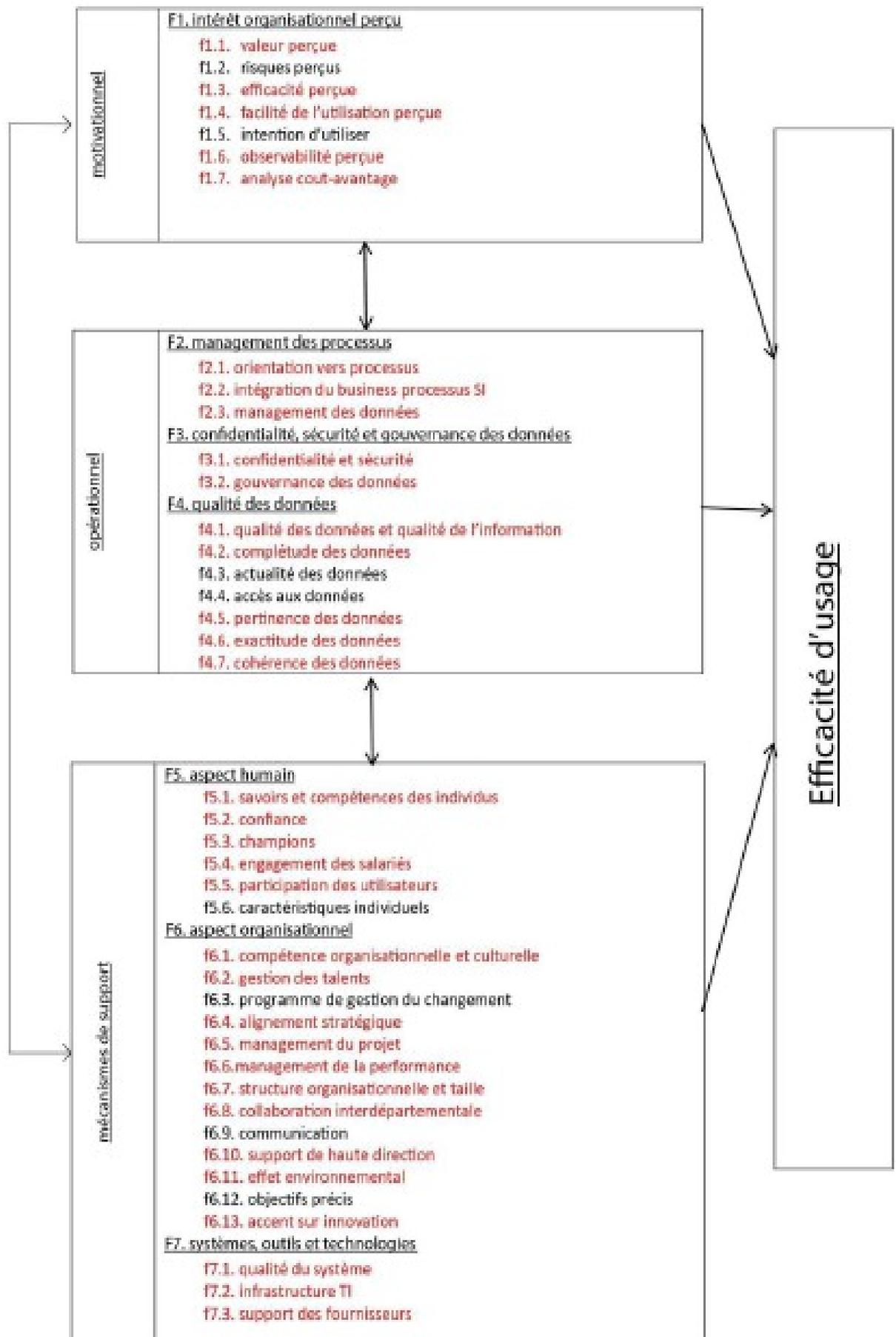


Figure 1 : Modèle de l'efficacité d'usage, traduit de Surbakti et al. (2019)<sup>1</sup>