

Proposition d'un cadre de diagnostic

3.1 Introduction du besoin

Ce chapitre présente notre proposition de cadre méthodologique (cf. figure 3.1) pour le diagnostic organisationnel et l'évaluation des centres d'appels d'urgence.

Cette méthodologie a fait l'objet d'un article de journal publié en 2019 (Petitdemange et al., 2019) dans la revue *IEEE Transactions on Engineering Management* et ce chapitre emprunte une partie de ses résultats à cette publication. Ainsi, une première partie du chapitre est dédiée à la justification du besoin d'une telle méthodologie et une seconde à la justification des outils employés ainsi qu'à la présentation détaillée des étapes de la méthodologie proposée.

3.1.1 La notion de diagnostic

Le concept de processus de diagnostic ne se trouve pas uniquement en médecine où il est défini comme le « temps de l'acte médical permettant d'identifier la nature et la cause de l'affection dont un patient est atteint. »¹. En effet, il peut être transposé à d'autres domaines. En gestion par exemple, cela correspond à l'évaluation des points forts et des points faibles d'une entreprise. En industrie automobile, c'est le contrôle au moyen de dispositifs embarqués ou non permettant de déterminer l'état du véhicule lors de révisions ou lors de pannes ou d'incidents pour en déterminer les causes. (Française, 2020). On retrouve aussi la notion de diagnostic en automatique qui peut se définir comme « l'ensemble de mesures, de contrôles faits pour déterminer ou vérifier les caractéristiques techniques d'un système à des fins de maintenance ou d'amélioration. » On trouve en automatique toujours l'usage de la boucle commande-contrôle qui à l'aide de boucle de rétroaction permet de donner de l'information sur le système afin d'identifier rapidement des problèmes (Le Gallais, 1997).

1. Larousse - 2020

Le diagnostic médical

Lorsqu'un patient ressent une gêne, qu'il ne se sent pas bien, il va identifier ses symptômes, les choses qui le dérangent, qui ne sont pas normales. Il va partager ses symptômes avec son médecin qui va l'ausculter, afin d'avoir des informations supplémentaires (prise de tension, prise de sang, température, etc.). Sur la base des symptômes fournis par le patient (manifestation d'une maladie perçue par le malade) et des signes (manifestation d'une maladie constatable objectivement) mesurés par le médecin, un diagnostic va être fait. De ce fait, le diagnostic est déjà en cours, car il peut être considéré comme un processus, un enchaînement d'activités dans le but d'identifier le mal dont souffre le patient afin de le traiter (Jutel, 2009). Jutel dans son étude sociologique sur le diagnostic distingue deux définitions : Le diagnostic peut à la fois être un processus et une catégorisation. L'auteur considère que c'est une catégorisation car il correspond à l'agrégation de symptômes et de signes qui deviennent clairs, objectifs et justifiés dès lors qu'on peut les relier à une étiquette : le diagnostic. Le diagnostic comme processus correspond à la méthode d'évaluation et de traitement des maux du patient. C'est cette définition de processus que nous utiliserons pour la suite. En médecine, le processus de diagnostic a été défini par (National Academies of Sciences et al., 2015) comme un enchaînement complexe d'activités de collaboration, centrées sur le patient, qui impliquent la collecte d'information et le raisonnement clinique, dans le but de déterminer le problème de santé d'un patient.

Le diagnostic d'un patient est un processus en plusieurs étapes : Le patient constate qu'il a un problème de santé (1), il prend alors contact avec le système de santé (2). S'en suit un cycle itératif de récolte de l'information (3), d'analyse et d'interprétation de l'information (4), et de formulation du diagnostic en tant que tel (5). Si la quantité d'information recueillie n'est pas suffisante pour identifier le problème de santé du patient, le cycle reprend jusqu'à réussir à valider un diagnostic (ici, cela comprend la définition de catégorisation de la maladie). L'étape suivante est la communication et l'explication du problème de santé au patient (6). Enfin, un traitement s'appuyant sur le diagnostic est planifié et mis en place (7). Ce traitement fait l'objet d'un suivi via le cycle de récupération d'information afin de s'assurer de son effet (8). (National Academies of Sciences et al., 2015).

Le diagnostic organisationnel

Si on considère le SAMU comme un patient et que l'on file la métaphore, en se référant aux éléments discutés dans les chapitres précédents, on peut considérer que le SAMU présente des symptômes de dysfonctionnement, il est malade. Cette hypothèse se renforce si on considère les articles de presse généraliste relatifs aux SAMU qui se multiplient ces dernières années tels que : « Il faut prendre des mesures pour éviter que le drame du SAMU de Strasbourg ne se reproduise »², « Appels aux urgences : les SAMU accusés de ne pas toujours décrocher »³, « Classement

2. Le Monde, 15 mai 2018

3. Le Monde, 24 août 2018

des SAMU : quand le 15 ne répond pas »⁴. On constate dès lors un problème qu'il semble nécessaire d'analyser, de comprendre, et autant que faire ce peut, de résoudre

3.1.2 Contexte et justification de la question de recherche : Vers une méthode outillée de diagnostic organisationnel des SAMU

En un mot, les centres d'appels sont confrontés à de nombreux défis liés à leur performance, et plus particulièrement à leur réactivité. C'est extrêmement préoccupant, car une mauvaise performance dans ce secteur peut avoir un impact énorme sur la vie des gens. Pour améliorer leur performance, les centres d'appels doivent remettre en question le service de prise d'appels, c'est-à-dire la façon dont ils traitent et répondent aux appels. L'hétérogénéité de ces organisations et les spécificités du domaine des services d'urgence permettent d'expliquer les niveaux de performance variables dans des conditions de fonctionnement « normales » ainsi que les difficultés à anticiper le comportement de l'organisation en situation « extrême ». La contribution de ces travaux de recherche consiste en partie à relever ce défi complexe

Concernant les modèles formels, nous avons décidé de discuter ces différences selon les dimensions présentées au chapitre 1.1 : **la stratégie pré-hospitalière, l'accès au centre d'appels** et le modèle de décroché que l'on décline en quatre dimensions : **le type de plate-forme, la spécialité du centre d'appels, la gestion des files d'attente et les compétences en matière de ressources.**

La stratégie pré-hospitalière

La stratégie pré-hospitalière définit la façon dont le Centre de Réception et de Régulation des Appels (CRRA) répond aux besoins de secours. Deux approches principales ont été identifiées : *Scoop And Run* et *Stay And Play*. En France, le modèle est principalement du SAP. Malgré la grande quantité d'études existant sur le sujet, (Hannay et al., 2014 ; Levy et al., 2007 ; McCoy et al., 2013 ; Smith et al., 2009) nous ne sommes toujours pas capable de savoir quel modèle est le mieux pour la survie du patient. Ainsi, lors de l'étude d'un centre d'appels d'urgence, il est intéressant d'identifier le type de stratégie pré-hospitalière dans laquelle nous nous trouvons.

L'accès

L'accès au centre d'appels définit la manière dont le centre d'appels peut être contacté. On considère plusieurs moyens d'accès aux centres d'appels d'urgences : un accès universel si le numéro est unique, et spécialisé s'il y a plusieurs numéros pour chaque type d'urgences différentes (Polices, Pompiers, Urgences Médicales). Les différents types de modèles que l'on retrouve majoritairement ont été discutés dans le chapitre 1.1. La différence majeure réside dans le fait que le patient ait besoin de choisir le

4. Le Point, 22 août 2018

bon numéro en fonction du type de secours requis ou si cette décision revient aux professionnels de l'urgence. De plus, dans le cas où le patient a besoin d'effectuer lui-même la détermination de la gravité de la situation, on observe en Europe 10% d'erreur dans les organisations qui opposent les services d'urgence et les services non urgents. Il y a ici un enjeu de simplification et de lisibilité sur l'utilisation des services d'urgences pour les citoyens.

La plate-forme

Le type de plate-forme peut être partagé ou spécifique. Elle est partagée si tous les appels sont traités au même endroit indépendamment qu'ils soient destinés aux urgences médicales, civiles, ou de sécurité incendie. Elles sont spécifiques si chaque spécialité dispose de son propre site pour la répartition et l'envoi des interventions. Par exemple, le SAMU de Carcassonne regroupe à la fois la plate-forme d'appels des pompiers et celle du SAMU, cette plate-forme est donc partagée. En revanche, les SAMU d'Albi, Rodez, Nantes ou Toulouse disposent de leur propre plateforme téléphonique indépendante de celle des pompiers. Au sein, d'un même pays les pratiques peuvent différer. Toutefois, en France les plate-formes sont majoritairement spécifiques, la configuration de plateforme commune ne concerne que 17 départements sur 101. Il n'y a actuellement pas d'études qui permettent de connaître l'impact de ces modèles sur les performances des centres d'appels. Il pourrait être intéressant d'étudier l'effet qu'une plate-forme commune peut avoir sur le temps de prise en charge des appels.

Le gestion des files d'attente

La gestion des files d'attente fait référence aux règles de gestion des appels. Fondamentalement, il peut être avec ou sans triage. Sans triage implique que les appels sont gérés selon le principe du premier arrivé, premier servi pendant tout le processus de l'appel, alors qu'avec le triage signifie que les appels sont classés par ordre de priorité après avoir été pris en charge en fonction de leur gravité. Ainsi, les appels les plus graves sont traités en priorité.

La spécialité

La spécialité renvoie aux types d'appels couverts par le centre d'appels : qu'il s'agisse d'appels urgents ou d'appels de médecine générale. Nous ajoutons cette catégorie, car cela apporte une nuance pour la priorisation des appels. Car dans le cas où le même centre d'appels gère à la fois de la médecine d'urgence et de la médecine générale, nous sommes face à deux types d'appels qui n'ont pas les mêmes besoins. In fine, ils n'ont pas non plus la même chaîne de valeur. En effet, les appels de médecine générale sont moins critiques en matière de temps en comparaison à ceux de médecine d'urgence. Ainsi, cela aura un effet sur la gestion des priorités dans le centre d'appels.

La gestion des ressources

Les compétences en matière de ressources font référence au fait que les ressources au sein d'un CRRA peuvent être spécialisées (pour une tâche spécifique, un flux d'appels, etc.) ou polyvalentes. (M. v. Buuren et al., 2017)

Étant donné qu'il y a deux possibilités pour chacune des dimensions organisationnelles que nous venons d'évoquer, nous avons une combinatoire de 64 (2^6). Cette combinatoire conduit mécaniquement à une grande hétérogénéité de systèmes SAMU, et à une grande complexité. Ainsi, il y a un enjeu majeur à disposer d'une méthode outillée complète permettant de poser un diagnostic étiologique^a organisationnel pertinent et complet.

a. ("Le diagnostic étiologique, consiste à identifier la cause de l'affection" - Larousse 2020)

Afin de confirmer la pertinence de notre objectif, nous avons effectué une analyse documentaire systématique à l'aide de ces mots clés sur le Web of Science :

(Methodology OR Method OR Method OR Framework OR Tool)
AND (« Emergency Call Center » OR « Emergency Contact Center »
OR « Emergency Dispatch Center » OR « Public Safety Answering
Point ») AND (Evaluation OR Diagnosis * OR Assess *) AND
(Responsiveness OR Performance OR Efficiency).

Cette requête a généré 1240 résultats. Parmi ces résultats, Gardett et al. (Gardett et al., 2016) passent en revue la littérature systématique sur la « recherche sur les dépêches d'urgence » depuis les années 1970 jusqu'en 2013. Dans leur revue, ils soulèvent un manque d'études comparatives sur le terrain, un cadre d'évaluation universel et une description précise du protocole utilisé dans les études individuelles. Nous avons donc concentré nos investigations de recherche à des articles publiés après 2013. Après un tri sur les titres et les résumés, nous avons sélectionné 10 articles qui traitent de l'organisation ou de la méthodologie des centres d'appels d'urgence. Toutefois, aucun d'entre eux ne présente de solution détaillée permettant d'adresser notre problématique. C'est pourquoi nous pensons que la proposition d'un cadre outillé est pertinente et vient compléter utilement la littérature tant pour les chercheurs qui veulent étudier les centres d'appels que pour les praticiens du domaine afin d'évaluer leur structure organisationnelle.

3.2 Présentation de la méthodologie

Notre proposition est donc une méthodologie exploitant différents outils. Elle peut être utilisée pour diagnostiquer et améliorer les centres d'appels tout en tenant compte de leurs spécificités. Cela permet d'effectuer une analyse plus précise en vue d'améliorer notamment la réactivité des centres d'appels. À partir du système réel, l'objectif est de pouvoir construire un modèle de connaissance. Ce modèle permet

de comprendre et de formaliser le système réel afin de créer son double numérique : la réplique numérique du centre d'appels. Le système réel et le double numérique couplé aux six dimensions déjà présentées dans la section 3.1.2 constituent le cadre proposé comme indiqué dans la Figure 3.1. Tout le principe est de partir du système réel, d'en extraire les données utiles afin de les transformer en connaissance. Tout ce travail dans le système réel nous permettra, dans un second temps, de construire et paramétrer le double numérique du système étudié. Ce double numérique va agir comme une zone expérimentale virtuelle, sur laquelle nous pourrions nous appuyer pour améliorer l'organisation du système réel.

La méthodologie est divisée en quatre phases que nous allons développer dans la suite :

1. Collecter et préparer les données, ce qui permet de comprendre le processus opérationnel du système réel et d'en extraire des données significatives.
2. Interpréter et intégrer les informations afin de transformer les données en connaissances.
3. Diagnostiquer, ce qui permet d'obtenir tous les paramètres d'entrée du modèle de simulation ainsi que d'analyser la performance actuelle du centre d'appels.
4. Améliorer, ce qui permet de tester de nouvelles configurations organisationnelles et d'évaluer l'impact sur la performance du système.

3.2.1 Étape 1 : Collecter et préparer les données

1.a Collecter et nettoyer les données

La complexité d'obtenir le journal des événements des données réside dans la variété de la structure des données, qui peut être différente d'une organisation à l'autre. Afin de faciliter ce processus il est préférable que les données soient préparées selon un format standardisé. Van Dongen et van der Aalst (Dongen, van B.F. et al., 2005) ont développé un métamodèle pour le traitement par fouille de données (*Data Mining*). Chaque événement du journal doit avoir un nom, un horodatage et un identifiant (ID) de trace. Dans notre cas spécifique, nous ajoutons la ressource par qui l'activité a été effectuée et l'origine de la trace. La structure de nos données (que l'on appellera *event log*) comprendra ainsi les éléments suivants :

- **Événement** : le nom de l'événement qui se produit ;
- **Horodatage** : l'heure de l'événement - année, mois, jour, heure, minute, seconde ;
- **ID** : une code unique pour identifier tous les événements liés à un appel ;
- **Ressource** : le numéro d'identification de la ressource exécutant l'événement ;
- **Origine** : le numéro composé par l'appelant pour accéder au centre d'appels

La figure 3.2 reprend les grands principes de cette première étape. En entrée, nous avons besoin des données de téléphonie brutes, sur lesquelles nous allons appliquer des algorithmes d'associations (Hand et al., 2001) pour rassembler les événements correspondant au même appel. Prenons un exemple, un appel à lui seul va générer

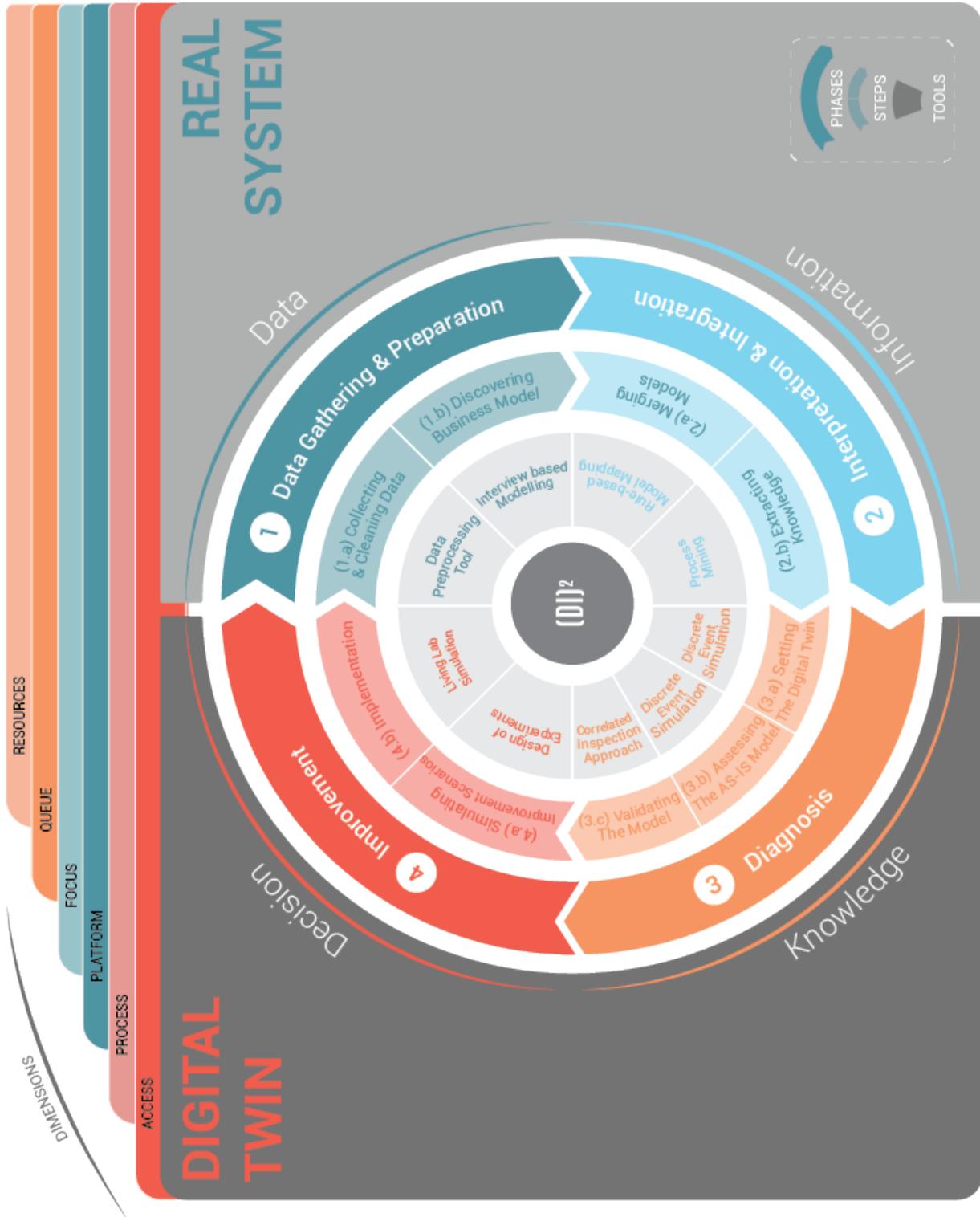


Figure 3.1 – La méthodologie outillée proposée (Petitdemange et al., 2019)

plusieurs événements. Afin de pouvoir retracer le cycle de vie de cet appel, il faut que chaque événement en lien avec ce dernier ait un identifiant commun qui permette de le relier à cet appel précisément. Cette suite d'événements horodatés en rapport avec un et un seul appel se nomme une trace. Cela va nous permettre de faire un tri sur les traces incomplètes en utilisant de l'analyse de séquence (Van der Aalst et al., 2004) et de les mettre de côté. Les traces considérées comme incomplètes sont celles qui ne vont pas avoir d'événement de début, ou qui n'auront pas d'événement de fin. Ainsi nous obtenons des données qui sont nettoyées et complètes pour la suite.

Le but de cette première étape est d'obtenir des données qui puissent être utilisées en entrée des prochaines étapes où l'on va appliquer du *Process Mining*.

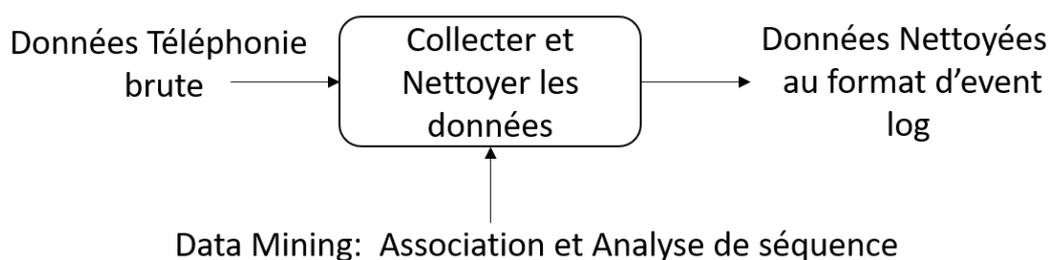


Figure 3.2 – Collecter et nettoyer : entrées, sorties et outils

1.b Découvrir le processus métier

L'étape suivante consiste à comprendre le processus opérationnel et à obtenir plus d'information sur les opérations du centre d'appels. Comme nous l'avons déjà mentionné, les centres d'appels peuvent être organisés de différentes façons d'une région à une autre (Penverne et al., 2019). L'une des difficultés est de saisir et de comprendre ces différences. Dans ce cas, nous avons besoin d'information sur les ressources :

- Combien de types de ressources sont utilisés ?
- Quelles sont les différences entre elles ?
- Quelles sont les activités réalisées par chaque type ?

Nous avons également besoin d'information sur les flux entrants et la gravité des appels. Ensuite, nous devons comprendre les données afin de pouvoir relier le numéro d'identification de l'agent à sa fonction, et le numéro d'identification du flux entrant à sa catégorie. Cela implique d'avoir un glossaire des numéros d'identification ainsi qu'un accès au journal du centre d'appels.

Une partie de ces informations pourrait être récupérée à travers les journaux (logs) de données. Toutefois, il est nécessaire de les compléter avec de l'information provenant directement du système réel afin d'avoir une meilleure compréhension globale du système. Cette connaissance peut s'obtenir en analysant les processus, donc lors de démarches d'observation. Pour collecter ces données, on retrouve deux types de méthodes : la recherche quantitative et la recherche qualitative.

La recherche qualitative est une méthode d'exploration qui permet de décrire, de comprendre et d'expliquer un phénomène particulier en répondant aux questions « quoi ? », « qui ? » et « comment ? », en opposition à la recherche quantitative qui

répond à des questions du type « combien ? » et « à quelle fréquence ? » (Jones, 1999). La recherche qualitative se caractérise par l'utilisation de méthodes particulières comme les interviews semi-structurées afin de récupérer des données ou d'observer des événements.

Elle est surtout utilisée pour étudier les processus, pour permettre de comprendre comment des actions sont effectuées, par quelles étapes on doit passer pour réaliser l'action (Ma, 2000). Gill et Stewart nous montrent qu'il s'agit d'une méthode de collecte de donnée très utilisée dans le domaine de la santé (Gill et al., 2008). C'est pourquoi nous avons effectué des interviews avec des experts métiers. Lors de ces interviews (voir Figure 3.3), il s'agit de récupérer des informations sur le processus métier, sur la manière dont s'enchaînent les différentes activités et ainsi obtenir un premier processus formalisé des activités. On doit aussi utiliser ces moments de rencontre privilégiés afin de donner un sens métier aux données de téléphonies. Par exemple, en récupérant la liste des identifiants des agents on pourra par la suite identifier quel agent est un assistant de régulation médicale (ARM) ou un médecin régulateur (MR).

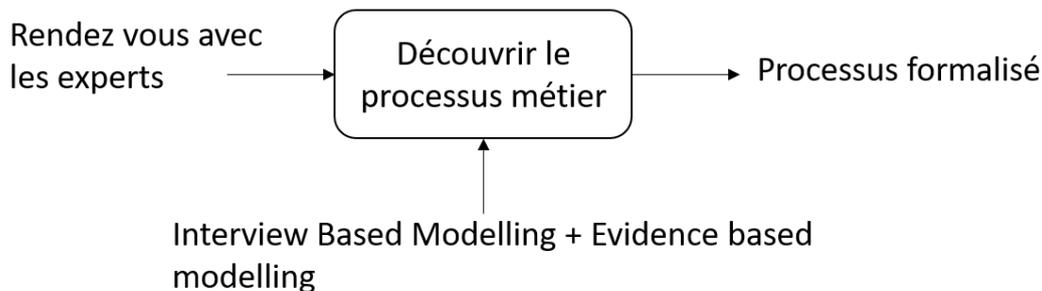


Figure 3.3 – Découvrir le processus : entrées, sorties et outils

3.2.2 Étape 2 : Interpréter et intégrer

2.a Fusionner les modèles

Cette étape crée le lien entre l'événement et l'activité, ce qui permet de déterminer la durée des activités, grâce à l'horodatage des événements. Les données du log sont composées de traces, qui elles-mêmes sont composées d'événements. Il s'en suit qu'une trace (une séquence d'événements) correspond à un appel dans le centre d'appels. Tous les événements sont instantanés. En outre, selon la ressource liée à l'événement, il peut s'agir d'une activité différente dans le processus de gestion. Ici, l'objectif est de déterminer la durée de chaque activité pour chaque appel à l'aide du log des événements. Nous avons donc élaboré des règles métiers pour relier les événements et les activités.

Dans un appel on peut enregistrer plusieurs durées, ce qui permet de décrire le type d'appel en termes d'état. Un appel a deux états principaux :

- En attente : l'appel est dans une file d'attente, en attente d'une ressource pour le traiter.

- En ligne : l'appel est traité par une ressource.

Un processus d'appel est une séquence d'états - transitions, de type "en attente" - "en ligne". L'état d'attente est facultatif facultatif (ou équivalent potentiellement à une durée nulle) et deux états en ligne peuvent être consécutifs. Un appel peut être géré par une ou plusieurs ressources. Par exemple, s'il y a deux ressources qui gèrent l'appel, il y aura deux durées correspondant à un état en ligne, une pour chaque ressource concernée. La durée entre les événements peut alors être facilement calculée.

Dans notre cas, lorsque plusieurs actions sont liées par une seule ressource, nous considérons cette séquence d'action comme une seule activité. Ceci est possible parce que nous considérons que les actions ne peuvent pas être interrompues. Le constat effectué lors des observations terrains est que lorsqu'un appel commence à être traité, il n'est que très rarement interrompu avant la fin de son traitement. Il s'ensuit qu'un temps de communication dans le log serait lié à une activité (ou à un ensemble d'action) effectuée par une ressource. Pour lier les processus métier et le log des événements il faut identifier au moins quatre types d'événements :

- *E1-Ringing* : Le premier événement pour un appel, lorsqu'il arrive au centre d'appels et qu'il attend d'être pris en charge par une ressource.
- *E2-Established* : L'appel est décroché et la ressource est en ligne.
- *E3-Released* : L'appel est libéré.
- *E4-Held* : L'appel est suspendu.

À titre d'exemple, nous considérons deux modèles d'appels du point de vue de l'événement et du processus métier décrit à la figure 3.4. Dans cette figure, il y a deux types de formalisme, d'une part les enchaînements d'événements qui proviennent des données de téléphonie, d'autre part les enchaînements d'activités qui proviennent du processus métier. L'objectif est de réussir à faire le lien entre les deux modèles. Dans ce processus de gestion, deux séquences d'activités sont possibles : i) A0-A2-A3-A3-A4, avec toutes les actions exécutées par la ressource R1, ou ii) A0-A2-A3-A5, les trois premières actions étant effectuées par R1 et la dernière par R2. Puisque nous supposons que les activités ne peuvent pas être interrompues, la séquence A2-A3-A4 est équivalente à l'activité A1 et la séquence A2-A3 à l'activité A1'. Ces activités peuvent être décrites comme suit :

- A0 correspond au temps d'attente, c'est la présence d'un appel dans la file d'attente.
- A1 représente la réponse d'un appel par un opérateur ; la gravité de l'appel est évaluée et s'il n'a pas besoin d'être diagnostiqué par un médecin, l'opérateur donnera quelques conseils à l'appelant.
- A1' représente la réponse d'un opérateur à un appel avant qu'il ne soit transféré à un médecin.
- A5 représente le diagnostic d'un appel d'un médecin durant lequel il décide ou non d'envoyer des moyens au patient.

L'application de l'algorithme 1 implique que la séquence (i) est liée au modèle 1 (Pattern 1 - voir figure) et que la durée d0 d'un appel dans la file d'attente est la différence entre l'horodatage de l'événement *E2-Established* et le premier événement *E1-Ringing*. L'heure de communication est la différence entre l'horodatage du dernier

événement *E3-Released* et l'événement *E2-Established*. L'algorithme 1 est également appliqué au modèle 2 pour obtenir la durée qui nous permet de lier la durée des activités du processus métier au journal des événements.

Algorithme 1 : How to link events to duration of communication

Résultat : Obtain the duration of each communication time

R_j : $j = 1..n$, $n \in (N)$ the number of resources

d_j : the communication time of the resource j

$E_{i,j}$: $i = 1..4$ the timestamp of the event i realized by the resource j ;

si $n = 1$ **alors**

$d_0 = E_{2,1} - E_{1,1}$

$d_n = E_{4,1} - E_{2,1}$

sinon

$d_0 = E_{2,1} - E_{1,1}$

$d_j = E_{3,j} - E_{2,j}$

$d_n = E_{4,n} - E_{2,n}$

fin

Ainsi, cette étape, comme résumée dans la figure 3.5 prend en entrée, les données sous forme de log issue de la première étape et les données de processus formalisées issues de l'étape 2. Le fait de croiser les deux résultats et de fusionner les deux modèles nous permet d'obtenir un jeu de données contenant plus d'information et notamment les temps d'attente et de communication des différents appels.

2.b Extraire des connaissances

En utilisant les données précédemment générées on peut extraire de la connaissance de ces dernières en faisant usage du *Process Mining* et du *Data Mining* (voir Figure 3.6). L'intérêt ici est de récupérer les données qui vont nous permettre de paramétrer le modèle de simulation.

Identifier la gravité

Les appels peuvent être classés par ordre de priorité et il est possible d'établir un niveau de gravité pour chaque appel en fonction de l'ensemble des ressources qui le traite. Nous pouvons classer par ordre de priorité le type de ressources : les ressources essentielles traitent les appels les plus urgents, tandis que les ressources non essentielles traitent les appels les moins urgents.

Par exemple, dans l'algorithme 2 avec $p=3$, Rt1 est un MR et la ressource la plus critique, Rt2 est un médecin généraliste et une ressource intermédiaire, et Rt3 est ARM et une ressource non critique. Rt3 reçoit tous les appels. Il/elle traite les appels non critiques et transfère le reste soit vers Rt2, soit directement vers Rt1. Rt2 transfère à Rt1 les appels qu'il considère comme très critiques. Essentiellement, tous les appels transférés à Rt1 par Rt2 et Rt3 sont considérés comme les appels les plus

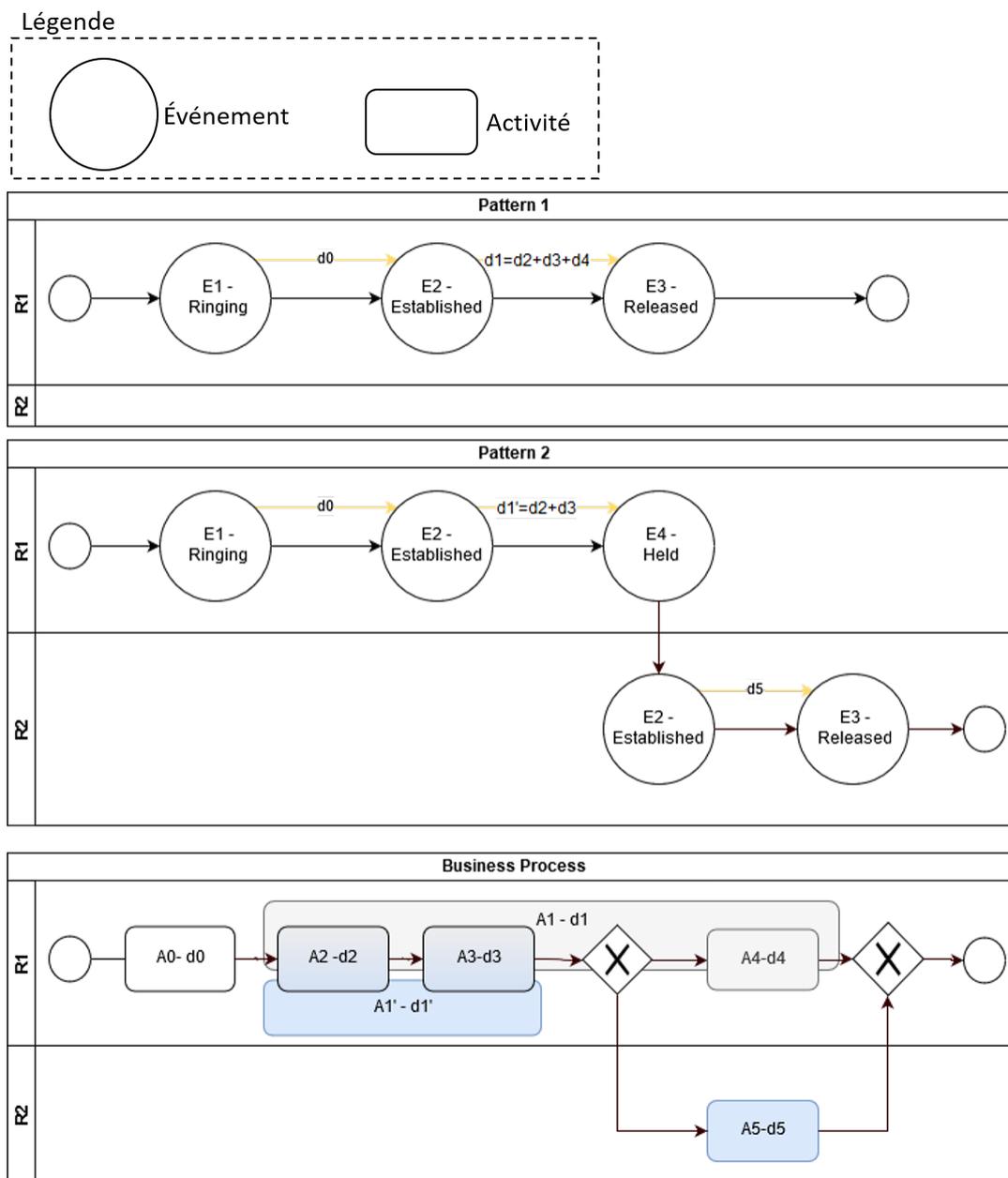


Figure 3.4 – Un processus métier et des exemples de patrons d'appels

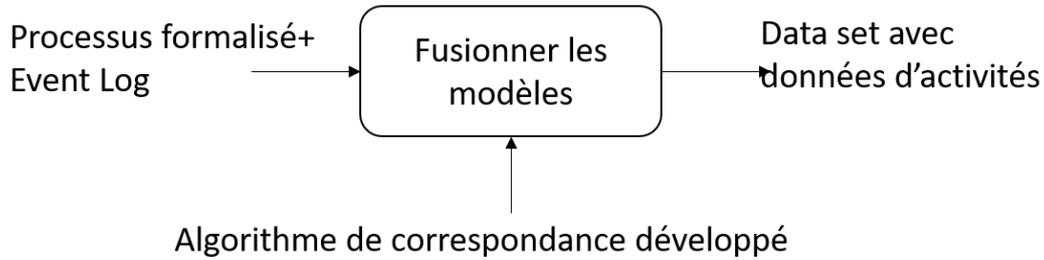


Figure 3.5 – Fusionner les modèles : entrées, sorties et outils

urgents (graves), tandis que ceux qui sont traités par Rt2 sont considérés comme étant d'un niveau de gravité intermédiaire.

Algorithme 2 : How to set the severity of a call

Résultat : Obtain the Severity of each Call

Rt_j : $j = 1..p$ the type of resource

S_i : $i = 1..m$, $m \in (N)$ The severity of the call i ;

$$A_{i,j} = \begin{cases} 1 & \text{if the Resource } Rt_j \text{ handled the call } i \\ 0 & \text{if not} \end{cases}$$

$\forall i, S_i = j$ pour $A_{ij} \neq 0$

Modéliser le temps de service pour les activités

Avec la durée des activités, on peut dès lors trouver la distribution de probabilité de chaque activité, ou du moins la distribution qui correspond le mieux aux données observées sur le terrain. Cela nous permet de calibrer le modèle de simulation avec précision. Dans la littérature existante, on suppose généralement que les temps de service suivent une distribution log-normale (Zuzakova et al., 2012)(Bhulai, 2009). Cette répartition peut convenir dans des cas particuliers où les ressources et les appels sont considérés comme homogènes et identiques.

Dans le présent document, nous faisons l'hypothèse que ce n'est pas nécessairement vrai puisque les ressources et les appels sont hétérogènes. Ainsi afin de modéliser les temps de service il y a deux possibilités : trouver une distribution théorique qui correspondent aux données réelles ou travailler directement avec les données réelles et la distribution empirique des données.

Dans le premier cas, une manière de faire est la suivante : en analysant les temps de service collectés à partir de données réelles, il est possible d'appliquer un algorithme de correspondance afin d'identifier les distributions théoriques les plus adaptées à notre cas. Cela signifie que pour chaque distribution théorique on va chercher le jeu de paramètres qui permet de minimiser l'écart entre les données réelles et la distribution théorique. Pour mesurer cet écart, on va utiliser la somme résiduelle des carrés (Dwars et al., 2013), également connue sous le nom de Somme des erreurs au carré (SSE). Une faible valeur de SSE indique un ajustement serré entre les données réelles et le modèle ce qui permet de choisir la meilleure distribution

Dans le deuxième cas, il s'agit de travailler avec les données réelles et d'utiliser la distribution empirique des données directement pour modéliser les temps d'activités.

Obtenir le profil des appels entrants

Le profil des appels peut être défini en fonction des données réelles, qui sont représentées sous la forme d'un histogramme avec un intervalle de classe de 30 minutes. En se basant sur les données réelles, l'objectif est de réussir à obtenir une distribution théorique de l'arrivée des appels. La plus grande difficulté sur cette partie est de disposer de suffisamment de données pour être capable d'identifier une distribution théorique.

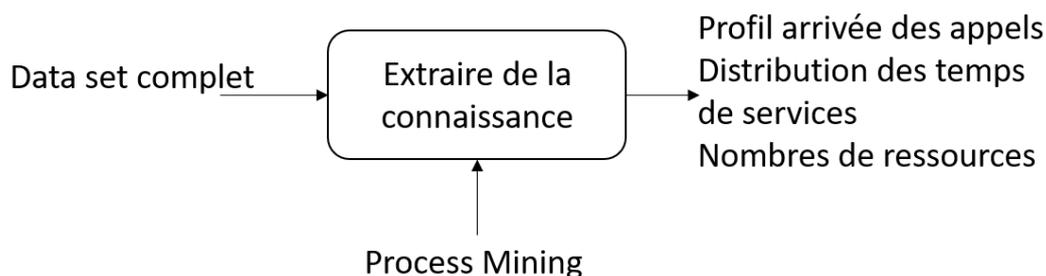


Figure 3.6 – Extraire les connaissances : entrées, sorties et outils

3.2.3 Étape 3 : Diagnostiquer

La première étape du diagnostic consiste à choisir la manière dont on va modéliser le réel. On considère deux grands types de modèles : les modèles analytiques et les modèles de simulation. Les modèles analytiques sont construits en définissant le modèle par des équations mathématiques. On lui donne en entrée des paramètres, le modèle effectue des calculs exacts et on obtient les résultats en sortie. Le modèle de simulation quant à lui va fonctionner avec un système de règles pour passer d'un état au suivant.

Lorsqu'il s'agit de décrire des systèmes dynamiques et hétérogènes, les modèles analytiques s'avèrent limités. Si on considère l'exemple de la théorie des files d'attente, on constate que si on complexifie la situation, assez rapidement nous ne disposons plus de modèle analytique permettant de résoudre le problème. (Borshchev, 2013b).

C'est pourquoi nous avons choisi d'utiliser un modèle de simulation dans notre étude. La simulation présente plusieurs avantages (Borshchev, 2013b) :

- Cela permet de tester et de trouver des solutions à des problèmes où d'autres méthodes n'ont pas fonctionné (comme la programmation linéaire ou les méthodes analytiques).
- Une fois le niveau d'abstraction établi, il est relativement facile de créer le modèle et de le faire évoluer de manière incrémentale et modulable. Cela apporte plus de flexibilité qu'un modèle analytique.
- Il est possible d'utiliser des éléments visuels de manière à ce que le modèle de simulation ressemble à la réalité. Cela rend le modèle plus accessible et il est plus aisé de communiquer avec les experts métiers.

- Il est possible de mesurer n'importe quelle valeur au sein du processus, il est facile de rajouter des calculs pour obtenir des statistiques sur le fonctionnement du modèle à n'importe quel moment dans le temps.
- On peut faire tourner le modèle en temps réel et voir son évolution au fil du temps, ce qui permet de voir l'état du système à chaque étape. Cette fonctionnalité est non seulement utile à des fins de communication mais aussi pour la construction et la vérification du modèle.
- Enfin l'aspect visuel est moins austère que peut l'être un modèle analytique avec uniquement des formules. Ce qui en fait un outil didactique très pertinent pour expliquer les résultats aux experts métiers.

3.a Paramétrer le double numérique

Cette partie consiste ainsi à construire le modèle de simulation (voir Figure 3.7). Il faut choisir le type de simulation que l'on souhaite utiliser puis le logiciel ou l'outil de programmation avec lequel on souhaite réaliser la modélisation. Le modèle est ensuite paramétré avec les données et distributions obtenues lors de l'étape précédente.

Sur la base du processus de gestion et des données, la simulation à événements discrets (SED) est utilisée pour créer un modèle du centre d'appels avec les paramètres extraits lors de l'étape précédente. La plupart du temps, les problèmes des centres d'appels sont modélisés en utilisant la théorie des files d'attente (Koole et al., 2002), qui est plus adaptée à un seul type d'appel et d'agent. Mais, les centres d'appels ont évolué et utilisent maintenant des ressources hétérogènes et reçoivent plusieurs types d'appels. C'est ce qu'on appelle des centres d'appels multi-sources avec des agents multicompetents, le SED semble donc être plus appropriée comme le démontrent Bhulai (Bhulai, 2009) et Roubos et al (Roubos et al., 2013). La simulation n'est pas seulement utilisée comme un outil d'aide à la décision, mais aussi comme un outil d'aide à la gestion (Mason, 2013).

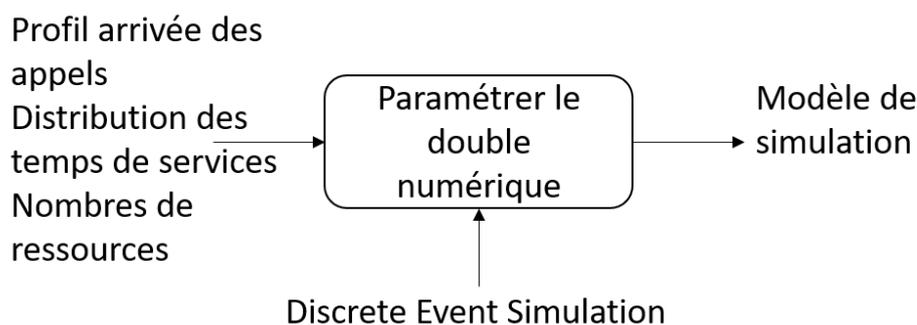


Figure 3.7 – Paramétrer : entrées, sorties et outils

3.b Evaluer l'AS-IS

Dans cette étape, il s'agit de définir les indicateurs de performance pour évaluer le centre d'appels. À l'aide des données, le temps d'attente pour accéder au centre d'appels est déterminé et utilisé pour évaluer la qualité de service (QoS). Le QoS est

une mesure de la réactivité d'un centre d'appels (voir chapitre 1.3) (Avramidis et al., 2009). Il s'agit habituellement de la fraction des appels desservis en deçà d'un seuil donné divisée par le nombre total d'appels entrants. Cet indicateur est généralement utilisé pour l'évaluation des centres d'appels et spécialement pour les centres d'appels d'urgence (Passmore et al., 2013)(Penverne et al., 2017). Dans notre étude, nous avons également tenu compte du temps d'attente maximal.

Par exemple, si on souhaite mettre l'accent sur la réactivité du centre d'appels. Il est naturel de se concentrer sur l'évaluation du temps d'attente avant qu'un appel ne soit traité. En générant l'histogramme des temps d'attente, nous pouvons rapidement évaluer le QoS et le temps d'attente maximum du centre d'appels.

En fonction du seuil défini par les experts du système réel, il est aisé de vérifier si les exigences de qualité sont respectées par le centre d'appels. Si ce n'est pas le cas, il faut alors chercher à améliorer la situation actuelle. On sera donc ensuite capable de tester de nouvelles configurations organisationnelles en vue de répondre aux exigences.

Le principe de cette étape (voir figure 3.8) est d'utiliser la SED afin de reproduire la réalité pour mieux comprendre où le système a des difficultés de fonctionnement, en identifiant les configurations dans lesquelles les temps d'attentes sont les plus longs.

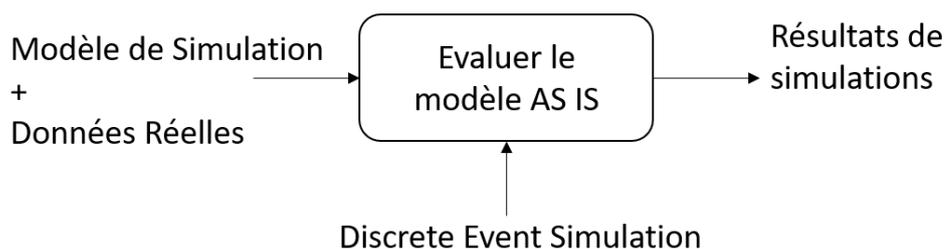


Figure 3.8 – Évaluer le AS-IS : entrées, sorties et outils

3.c Validation du modèle

Plus de 15 techniques de validation ont été identifiées dans la littérature par Robert G. Sargent (Sargent, 2013). Cet auteur indique que le choix d'une ou de plusieurs de ces techniques est fondé sur de multiples critères objectifs ou subjectifs. Nous avons donc examiné de nombreuses méthodes statistiques et opté pour l'approche de l'inspection corrélée (*correlated inspection approach*) proposée par (Law et al., 2000). Ici, nous souhaitons avoir une distribution qui permet de reproduire les temps de traitement des appels arrivant de manière totalement aléatoire suivant un profil d'arrivée prédéfini (défini par un nombre d'appels par tranche horaire).

La performance du centre d'appels réel doit être comparée à celle du centre d'appels simulé par rapport à l'indicateur de performance (KPI) le plus précis. Plus la performance du centre d'appels réel est proche du modèle de simulation, plus le niveau de confiance du modèle est élevé. Tukey (Tukey, 1977) a mis au point plusieurs méthodes qui sont utilisées pour comparer des séries univariées de données,

comme c'est le cas pour notre étude. Nous préconisons d'utiliser trois tests - deux graphiques et un analytique - pour évaluer la validité du modèle. Les tests graphiques doivent permettre de comparer la distribution des données à l'aide de leurs boîtes à moustaches (boxplots) et histogrammes. Cette méthode nous fournit des informations sur la distribution des données, la valeur moyenne et les valeurs extrêmes. Pour que le modèle soit valide, il doit avoir une forme similaire au modèle réel. Enfin, nous proposons d'utiliser le test analytique de Pearson pour évaluer la corrélation entre les deux séries. Si la valeur d'essai est supérieure ou égale à 95%, le modèle est considéré comme valide.

L'important dans cette étape est d'être capable d'attester que le modèle correspond à la réalité selon les hypothèses qui ont été formulées. (voir Figure 3.9). Cette étape est primordiale car si le modèle est trop éloigné de la réalité, toutes les préconisations qui pourront être faites en utilisant ce modèle seront caduques.

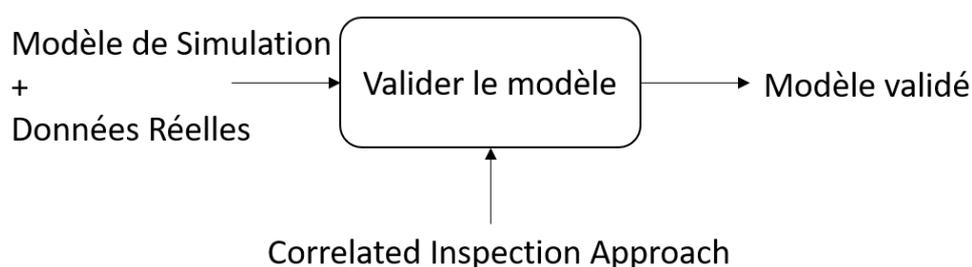


Figure 3.9 – Valider le modèle : entrées, sorties et outil

3.2.4 Étape 4 : Améliorer

4.a Simuler des scénarios d'amélioration

Avec la collecte des données et des informations extraites des journaux, une SED est mise en place pour tester plusieurs scénarios et évaluer l'impact sur la réactivité du centre d'appels. Le principe de cette étape est itératif (voir figure 3.10). Nous allons tester des scénarios d'amélioration en jouant sur différents paramètres et/ou variables illustrées par les onglets de la figure 3.1. Par exemple, il est possible de faire varier les règles des files d'attente : mettre une file d'attente par type d'appel ou une seule file d'attente pour tous les types d'appels ? Si tel est le cas est-ce que la file d'attente doit être gérée en FIFO ? (*First In, First Out*, le premier appel qui entre dans la file d'attente sera le premier appel décroché) Ou alors faut-il mettre une priorité en fonction de la provenance de l'appel (Urgence ou PDSA) ? Il est aussi possible de modifier la spécialité des ressources : tester l'effet qu'auraient des ARM spécialisés dans le décroché des appels d'urgence ou dans celui des appels de la PDSA. A partir de là, il est possible d'envisager des scénarios mixtes avec certaines ressources spécialisées et d'autres qui puissent répondre à tous les types d'appels. Un autre type de scénario envisageable est de modifier le processus de décroché des appels en mettant en place un niveau supplémentaire de décroché qui aurait pour but d'évaluer la gravité de l'appel le plus rapidement possible.

Une fois un scénario joué, il est analysé pour voir son impact sur les différents indicateurs de performance défini lors de l'étape 3b. De cette analyse des résultats obtenus, il est possible de définir un nouveau scénario afin d'affiner la préconisation. Plusieurs itérations peuvent être faites jusqu'à obtenir des résultats satisfaisants.

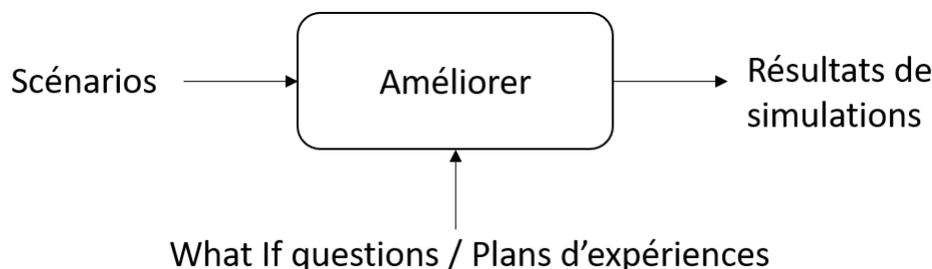


Figure 3.10 – Améliorer : entrées, sorties et outils

4.b Mettre en oeuvre

Si un ou plusieurs scénarios répondent aux objectifs des experts, ils pourraient être testés dans un laboratoire d'usages (*Living Lab*) ou par un jeu sérieux *Serious Game* dans une zone sécurisée pour évaluer le scénario en conditions réelles avec les ressources du métier. Cette étape permettra d'une part de présenter les scénarios retenus aux experts métiers (enjeux) et d'autre part d'évaluer la difficulté de mise en place de cette solution (accessibilité). Cette étape peut aussi s'avérer très intéressante pour obtenir les avis de l'équipe par rapport à des difficultés de mise en place qui pourraient être rencontrées et auxquelles nous n'aurions pas pensé (conduite du changement).

3.3 Synthèse

Face à la grande hétérogénéité des SAMU et notamment de leur composante aval, il semble pertinent de proposer un cadre méthodologique d'évaluation et d'amélioration de leurs organisations. A ce jour dans la littérature, il n'y a pas de travaux qui proposent une méthode complète, générique et outillée pour l'étude organisationnelle des centres d'appels d'urgence. La contribution proposée met donc en place un cadre méthodologique en deux temps qui s'appuie sur l'étude des données réelles pour soutenir la formulation d'un état des lieux objectif, puis sur la création d'un double numérique pour soutenir la conception et l'évaluation de propositions d'améliorations organisationnelles. Chaque partie est décomposée en étapes pour lesquelles nous proposons des outils opérationnels permettant d'atteindre les objectifs visés. Nous proposons aussi 6 dimensions organisationnelles (**la stratégie pré-hospitalière, l'accès au centre d'appels** et le modèle de décroché que l'on décline en quatre dimensions : **le type de plate-forme, la spécialité du centre d'appels, la gestion des files d'attente et les compétences en matière de ressources**). Ces six dimensions permettent de caractériser l'hétérogénéité des centres d'appels d'urgence. Elles sont autant de leviers sur lesquels il est possible d'agir afin de modifier les

organisations des centres d'appels d'urgence et mesurer l'effet de ceux ci sur les performances des centres d'appels.

Avec cette contribution nous souhaitons proposer un cadre d'étude structuré des centres d'appels associé à une boîte à outils. Les outils identifiés pour les différentes étapes n'ont pas la prétention d'être uniques et peuvent être modifiés ou adaptés. Nous ne prétendons pas non plus avoir choisi le meilleur outil pour chaque étape, mais simplement un outil qui permet de réaliser l'objectif de l'étape. Afin de faire progresser encore ce travail, il pourrait être utile de développer les transitions entre les étapes, de manière à ce que le format de sortie d'une étape corresponde exactement à celui d'entrée de l'étape suivante. L'idéal serait de pouvoir proposer une boîte à outils consolidée et unifiée qui utiliserait des données d'entrées harmonisées et réaliserait automatiquement toutes les étapes les unes à la suite des autres. Mais pour arriver à ce résultat, de nombreux verrous restent à lever. Notamment au niveau de la génération automatique du modèle de simulation à partir de la connaissance extraite des logs de données. Le tableau 3.1 propose une synthèse des contributions de ce chapitre et des principales limitations associées.

Dans la suite, nous proposons d'appliquer cette méthodologie pas à pas, afin d'évaluer dans le détail les bénéfices et les limites de la proposition faite dans un cadre d'application réelle. Le chapitre 4 correspond ainsi à l'étude et l'analyse des données réelles du SAMU associée aux étapes de collecte (section 3.2.1) et d'interprétation (section 3.2.2) des données. Le chapitre 5 présente quant à lui la contribution en lien avec les préconisations d'amélioration de l'organisation des SAMU. Ce chapitre 5 correspond à la partie gauche de la figure 3.1, donc celle qui traite du double numérique et fait référence aux étapes de diagnostic (section 3.2.3) et d'amélioration (section 3.2.4).

Tableau 3.1 – Contributions et Limites : Chapitre 3

Chapitre 3	
Contributions	Proposer des dimensions pour caractériser la diversité organisationnelle des SAMU
	Une méthodologie outillée pour le diagnostic et l'amélioration des centres d'appels d'urgence
Limitations	Une boîte à outils pouvant être adaptée et améliorée
	Les différentes étapes ne sont pas totalement consolidées.