

---

# Mettre en place le système d'information

## 3.1 Introduction

L'objectif de chapitre est de mettre en place un **système d'information local** au niveau de chaque système qui nous permet de collecter, stocker, traiter et distribuer les connaissances liées aux prédictions à un niveau de pilotage supérieur (Figure 30).

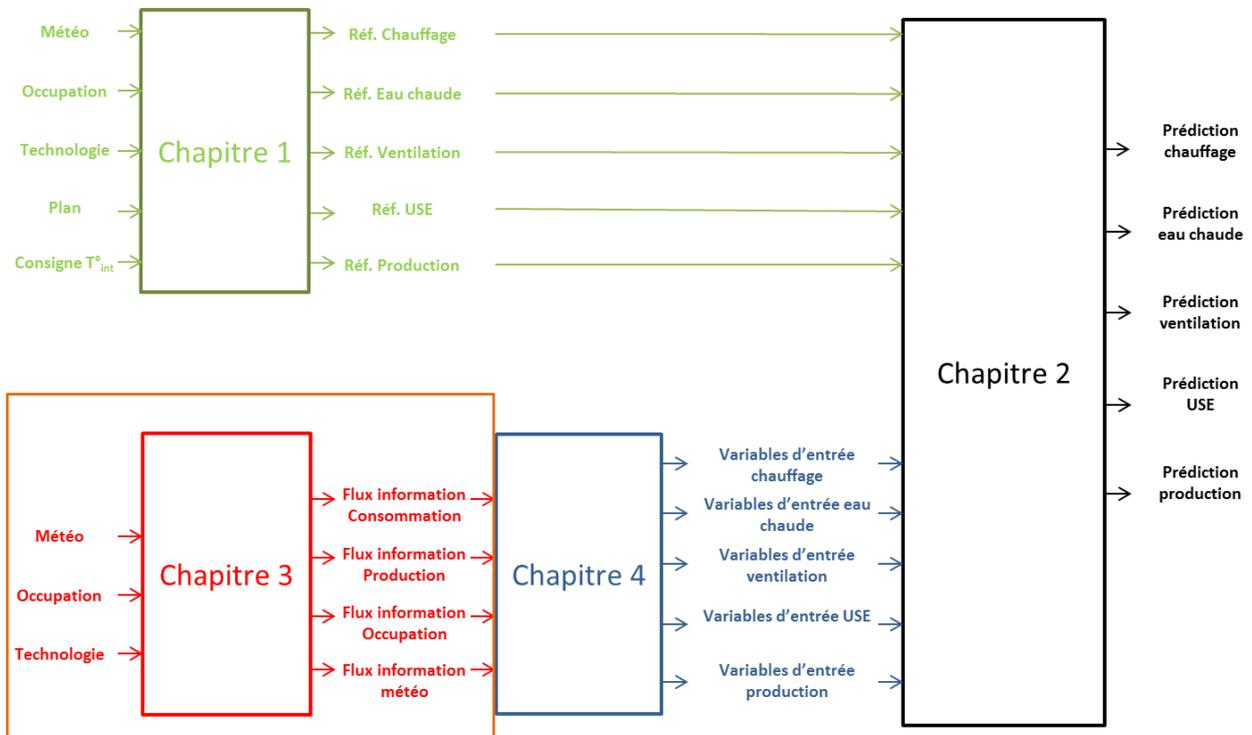


Figure 30 : Aide pour la lecture du document associée aux différents chapitres et au processus générale d'aide à la décision

L'étape de définition des modèles de référence permet de réaliser l'évaluation énergétique du système, en particulier au niveau du chauffage, du refroidissement et de la production solaire où des outils de simulations peuvent fournir des prévisions horaires. La partie « traiter » comprend la prédiction des différents flux énergétiques du système utilisant les modèles mathématiques décrits dans le chapitre 2 (Figure 31).

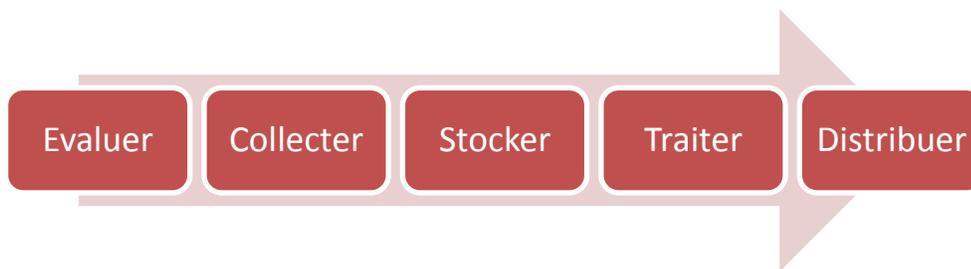


Figure 31 : Processus de mise en place du système d'information pour le pilotage énergétique pour nos systèmes d'étude

Les compteurs intelligents déployés aujourd'hui sont capables de recevoir et transmettre des informations entre le consommateur et le fournisseur (alertes sur la puissance soutirée, prix en temps réel, conseils automatiques concernant le report de consommation). Une couche de communication bidirectionnelle vient alors se superposer au réseau électrique, d'où la notion de réseau intelligent décrit dans le préambule.

## 3.2 Evaluer

### 3.2.1 Définition des acteurs

L'outil que nous créons correspond à une boîte noire qui doit fournir des services à son environnement. Par environnement, on entend les **acteurs** qui ont besoin de ce logiciel. Nous distinguons deux types d'acteurs : les acteurs principaux qui vont directement agir sur le système, dans notre cas le centre de pilotage et les acteurs secondaires qui n'ont pas d'accès direct à l'outil mais peuvent recevoir et échanger des informations avec celui-ci.

De plus, dans le langage UML (Unified Modeling Language) [BEN-2015], on appelle ce qu'on doit analyser, concevoir et réaliser : le système. Ici, le **système** est l'outil de prédiction des flux énergétiques qui permet de faire l'interface entre le centre de pilotage et les différentes charges du système.

Dans cette étude, notre outil fait le lien entre le système où il va tirer le maximum de connaissances et le centre de pilotage. La réactivité dans le pilotage de charges peut rendre des services aux systèmes de fournitures d'électricité. Les clients qui possèdent ou non une production décentralisée peuvent ainsi prendre part aux marchés de l'énergie et devenir de nouveaux acteurs côté demande par pilotage de leurs charges par un agrégateur [SUR-2004]. Des clients ou groupes de clients peuvent souscrire à des programmes orientés marché de l'énergie, dans lesquelles ils répondent à des appels tarifaires, en diminuant leur consommation.

De plus, ces programmes offrent la possibilité de participer activement aux échanges commerciaux d'électricité en proposant aux consommateurs de modifier leurs habitudes de consommation en retour de compensations financières. Ces dernières peuvent être un paiement direct de l'électricité non consommée, ou un paiement en fonction d'une garantie de disponibilité pour participer à un effacement de consommation à un moment opportun, faisant l'objet d'un accord [EU-2004]. Les clients concernés par ces programmes peuvent être des consommateurs individuels (gros clients industriels), ou des agrégations de petits consommateurs regroupés de manière à constituer une charge non négligeable au niveau du réseau (groupement de clients résidentiels).

Le centre de pilotage peut représenter un fournisseur et/ou producteur d'énergie ou un Virtual power plant [NIK-2012] qui aura agrégé des centaines de bâtiments/systèmes avec une production décentralisée photovoltaïque (Figure 32). Il peut être représenté par un logiciel ou par un être humain.



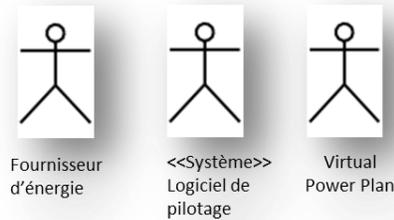


Figure 32 : Représentation des acteurs principaux dans notre outil

Les différents systèmes connectés sont définis par leurs secteurs d'activité. A l'intérieur de ces activités, nous proposons une différenciation des acteurs selon leurs contrats électriques qui les lient aux producteurs et fournisseurs d'énergie qui dépendent de leurs consommations énergétique. Nous présentons un exemple du système de facturation étudié en Suisse à Sierre dont les principaux paramètres tels que le raccordement ou le paiement du pic de puissance sont équivalents. Seul le prix va varier en fonction des pays Européens selon entre autre les différentes sources de production de l'électricité.

#### Détail du système de facturation dans le secteur tertiaire à Sierre, Suisse

Il existe deux grands types de facturation pour cette localité : la facturation de l'électricité chez les particuliers et les « petits consommateurs d'électricité », alimentés en moins de 100A, et celle pour les « grands consommateurs d'énergie », alimentés en plus de 100A.

**Pour les compteurs de plus de 100A**, les entreprises soumises à ce type de facturation bénéficient de tarifs préférentiels sur le kWh. En revanche, la pointe de puissance appelée mensuelle leur est facturée particulièrement cher (7.56CHF/kW). De plus, l'énergie réactive consommée peut leur être facturée si la quantité consommée dépasse 50% de l'énergie active consommée car cela déséquilibre le réseau électrique. L'énergie réactive correspond à l'énergie consommée par les appareils pour se mettre sur la même phase que l'énergie fournie par le réseau (Figure 33).

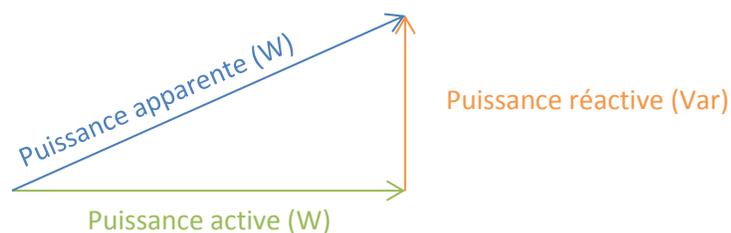


Figure 33 : Triangle des puissances indiquant la relation entre la puissance active, réactive et apparente

Selon le type de facturation, une prédiction précise du pic de puissance devient nécessaire à la fois pour la stabilisation du réseau mais également dans la diminution du prix dans la facture d'électricité

du consommateur pour les compteurs de plus de 100A. Par exemple, le pic de puissance représente 31% de la facture globale d'électricité sur l'année pour un restaurant d'entreprise (Figure 34).

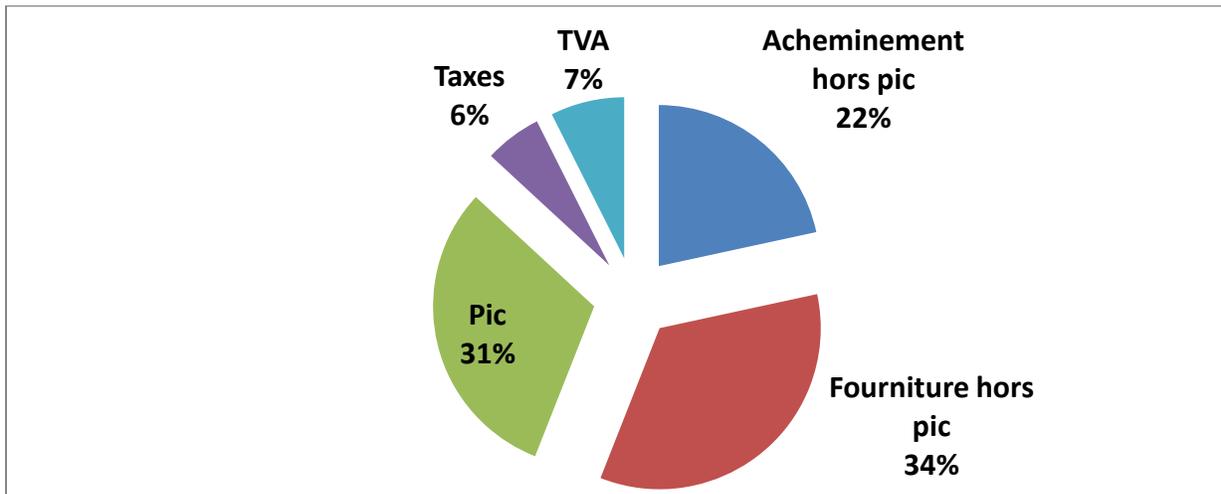


Figure 34 : Séparation de la facture d'électricité selon la fourniture et l'acheminement hors pic de puissance de l'électricité, la TVA, les différentes taxes et le pic de puissance qui comprend la part active et réactive pour le restaurant d'entreprise

A titre de comparaison, nous avons sélectionné trois entreprises avec trois secteurs d'activité différents (Studio de télévision, bureaux, restauration). Nous avons l'entreprise Canal 9 (Studio de télévision) qui consomme le plus parmi les trois entreprises, pourtant la pointe moyenne de consommation atteinte pendant l'année 2013 s'élève seulement à 38kW. En revanche, le Mikado (Restaurant d'entreprise ouvert seulement le midi), qui consomme l'équivalent de 36% de la consommation de Canal 9, a un pic moyen s'élevant à 52kW. Quant à l'entreprise Eticolle (Fabrication d'étiquettes), elle consomme à peu près la même quantité d'énergie que Canal 9 pourtant sa pointe de consommation s'élève à 99kW en moyenne sur l'année soit 2.6 fois plus que celle de Canal 9 (Figure 35). Ces trois entreprises, alimentées en plus de 100A, sont soumises à une facturation de la pointe de consommation mensuelle.

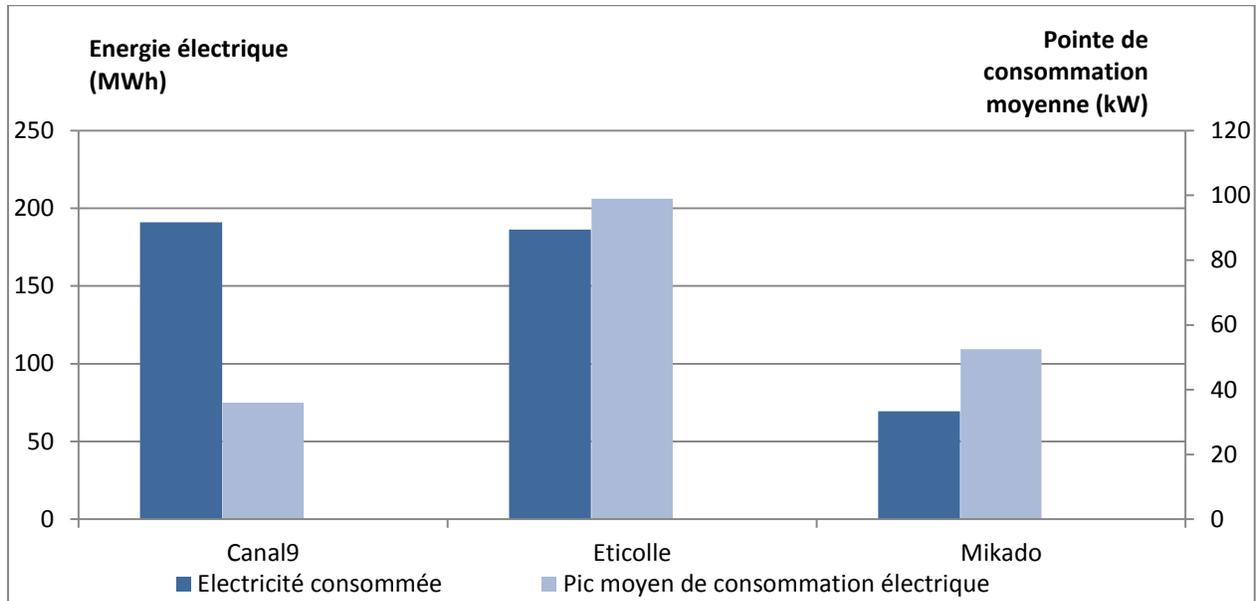


Figure 35 : Quantités d'énergies électriques consommées moyennes sur l'année 2013 et leurs pics de puissance associés pour trois consommateurs d'un des sites d'étude, Canal 9 (Studio de télévision), Eticolle (Fabrication d'étiquettes) et le Mikado (Restaurant d'entreprise)

Ces histogrammes sont expliqués par les courbes de charge journalières des entreprises. Le restaurant d'entreprise a une faible bande de base de 3kW, qui correspond à la consommation des groupes de froid. Cette entreprise ne consomme pas beaucoup d'électricité en soi, mais appelle beaucoup de puissance à certaines heures de la journée qui correspondent aux heures de préparation et de service des repas (Figure 36) avec la consommation des fours, des plaques électriques et des micro-ondes.

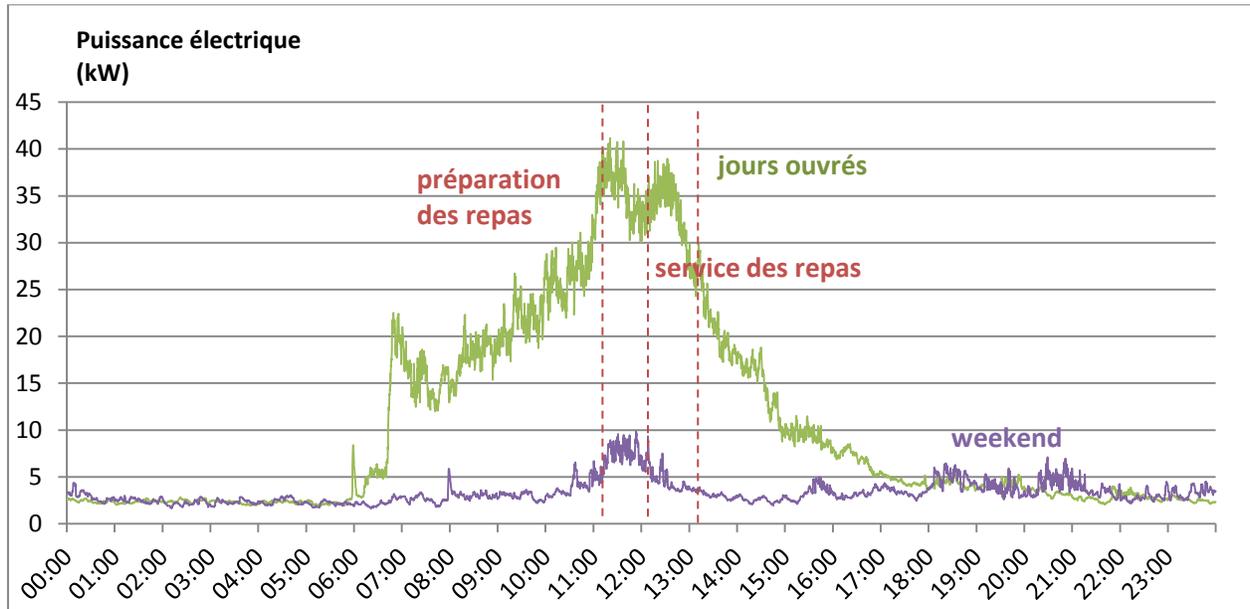


Figure 36 : Courbes de charges journalières moyennes en semaine et le weekend du restaurant d'entreprise moyennées sur le mois de Mars 2013

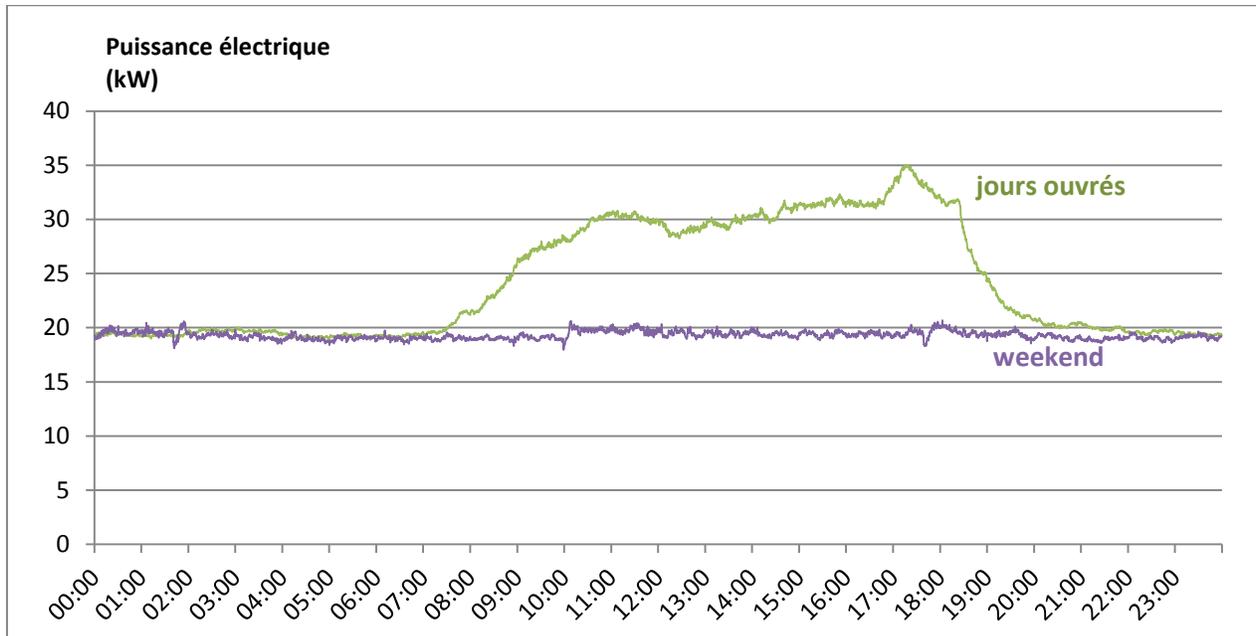


Figure 37 : Courbes de charges journalières moyennes en semaine et le weekend de l'entreprise de télévision Canal 9 moyennées sur le mois de Mars 2013

L'entreprise Canal9, qui diffuse des programmes de télévision, a une tendance de consommation très différente du restaurant d'entreprises, due à son secteur d'activité. En effet, les programmes et les supports pour leur diffusion devant être disponibles à toute heure du jour et de la nuit, 81% de la consommation de l'entreprise se situe en bande de base la journée (Figure 37).

**Pour les compteurs de moins de 100A**, les factures sont générées une fois par an. Si le prix du kWh est plus élevé, il n'y a pas de pointe de consommation ni d'énergie réactive facturée.

Ainsi, nous différencions les systèmes par leurs secteurs d'activité et leurs systèmes de facturation décrits dans la figure 38.

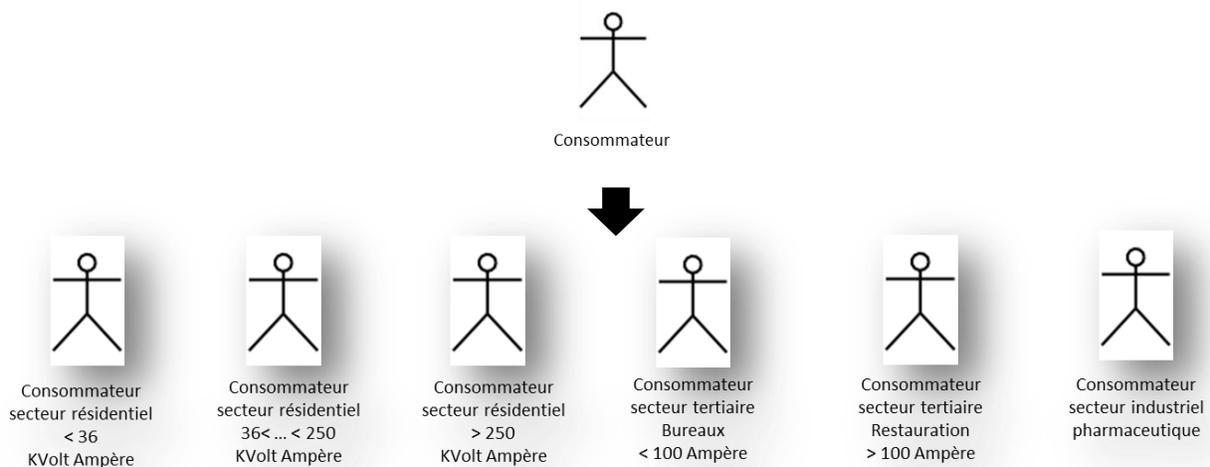


Figure 38 : Acteurs secondaires du système étudié qui est utilisé pour différencier les systèmes

### 3.2.2 Définition des fonctionnalités

Notre objectif est de réaliser un **bilan énergétique prédictif** pour l'heure suivante. Cela permettra d'avoir des possibilités de **scénarios de surproductions ou surconsommations** sur un quartier donné à **court terme au niveau de chaque flux énergétiques (t + 1 heure)**. Cela implique une prédiction fine des différents flux énergétiques d'un système, en particulier le chauffage et l'eau chaude, qui représentent le plus gros potentiel de flexibilité dans les bâtiments et permettra ainsi de **compléter l'offre de stockage**.

Il nous faut ainsi définir une « **Capacité d'effacement** » (Figure 39). Le potentiel d'effacement se situe avant tout au niveau des charges thermiques liées au besoin en eau chaude et en chauffage. Si le système possède une production solaire décentralisée, il devient utile de réaliser une prédiction fine de la consommation et de la production pour charger ou décharger la batterie suivant ces prédictions et des consignes de pilotage liés à l'autonomie du système ou à la vente d'énergie.

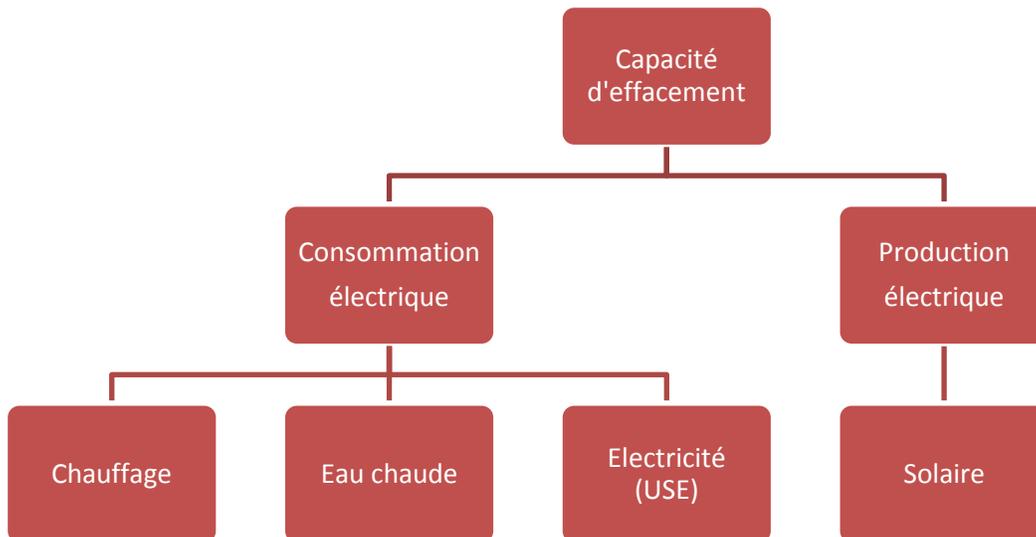


Figure 39 : Caractérisation de la capacité d'effacement qui nécessite la prédiction du chauffage, de l'eau chaude sanitaire, de la consommation électrique globale et de la production solaire

Une partie des données analysées peut-être envoyée aux consommateurs finaux qui sont les **acteurs secondaires**. L'**identification des appareils** regroupe les fonctionnalités qui permettent à partir du compteur global électrique d'identifier les différents usages. L'identification des appareils permet ainsi de fournir au consommateur final propriétaire ou non des connaissances sur la consommation globale du site et par usage qui diffèrent selon les secteurs d'activité. Nous nous attacherons à définir les appareils qui représentent la plus grande part dans la consommation énergétique du système. Dans le secteur résidentiel, le chauffage et l'eau chaude, en particulier sur les bâtiments les plus récents, représentent les plus gros consommateurs d'énergie.

L'usage de l'énergie électrique dans le secteur tertiaire est moins dominé par le chauffage que dans le résidentiel. A titre d'exemple, les usages d'eau chaude sanitaire et de cuisson sont très

développés dans la branche café-hôtel-restaurant alors que les usages spécifiques de l'électricité sont prédominants dans les commerces et les bureaux.

Le système avec seulement un appareil de mesure au niveau du compteur global, fournit **un outil de suivi de la consommation par usage** pour le consommateur final. Il permet également **d'augmenter la précision de la consommation électrique** en détectant les habitudes de consommation (par exemple, allumage de l'électroménager et du chauffage la nuit). De plus, si le système d'étude permet un effacement du chauffage et de l'eau chaude, cela permet de **limiter le nombre de données en transit** et stockées. Seules des informations de pilotage de type ON-OFF seraient envoyées.

En tant que	Je veux	Pour
Centre de pilotage	Prédire la consommation électrique globale	Adapter la consommation, la production et la capacité d'effacement
	Prédire la production globale	Adapter la consommation et la production
	Prédire le pic de puissance	Adapter la consommation et la production
	Définir une capacité d'effacement liée à l'eau chaude	Sécuriser le réseau
	Définir une capacité d'effacement liée au chauffage	Sécuriser le réseau
	Définir une capacité d'effacement liée au stockage électrique	Sécuriser le réseau
	Définir les données à collecter	Maintenir un haut niveau de prédiction
	Définir les données minimum à collecter	Limiter le nombre de capteurs à installer
	Définir les données minimum à collecter	Maintenir un haut niveau de prédiction
	Définir les modèles et méthodes de prédictions	Maintenir un haut niveau de prédiction
	Définir la fréquence de collecte	Limiter la capacité de stockage et de besoin en ressources informatiques
Consommateur	Connaître ma consommation totale et les principaux usages	Diminuer ma facture énergétique
	Un système non intrusif	Préserver ma vie privée
	Un système autonome de pilotage énergétique du chauffage, eau chaude sanitaire et du stockage	Me donner du temps
	Maximiser l'autoconsommation de mon système	Répondre aux enjeux climatiques à mon échelle
	Un système rentable	Gagner de l'argent

Tableau 4 : Tableau des fonctionnalités de l'outil

## 3.3 Collecter les données

### 3.3.1 Collecter les données au niveau des compteurs

Pour la collecte des données de consommation et de production dans le secteur résidentiel et tertiaire, nous nous sommes basés sur le projet i-BATS **[IBA-2013]** développé au sein de l'institut Informatique de Gestion à Sierre. Dans une optique de recherche, plusieurs niveaux de capteurs de la consommation ont été testés. Dans les locaux de la HES-SO Valais-Wallis, la température intérieure, le débit d'eau entrant et sortant du système de régulation de la température, l'intensité lumineuse, la détection de présence, la gestion des stores et la puissance active des appareils sont collectés. Dans ce cas, ces capteurs utilisent les protocoles de communication KNX et Zigbee décrits dans le projet européen IoT6 **[ANT-2014]**.

Un protocole de communication repose sur des règles qui définissent la transmission d'un message à travers un réseau. Un protocole de communication est donc nécessaire pour que deux équipements puissent échanger des données. Il existe des protocoles de communication standard comme le protocole **TCP/IP**. Des protocoles de communication sont spécialement adaptés au comptage électrique. Des protocoles de communication standardisés internationaux ont été définis. Les différents protocoles et caractéristiques liés au comptage électrique sont décrits dans **[COM-2010]**.

Malgré l'essor ces dernières années pour l'internet des objets, les capteurs testés qui permettent un relevé de la consommation d'énergie sans fil ne sont pas stables dans le temps (durée de vie supérieure à un an à comparer à celle d'un bâtiment). De plus, l'installation de capteurs filaires de type KNX **[ANT-2015]** par exemple dans les systèmes pour collecter la température ou le contrôle d'une vanne de chauffage demande des travaux complexes en particulier sur des logements existants.

Ceci nous conforte dans notre approche de limiter l'étude au niveau des compteurs globaux d'électricité concernant la consommation et la production si le système possède des capteurs solaires. Dans le cadre d'un outil qui se veut robuste et généralisable, nous nous sommes ainsi concentrés sur les communications filaires entre le capteur et les bases de données et sur les mesures globales de la consommation électrique par entreprise et de production solaire.

Dans le cadre du projet i-BATS, le choix des compteurs a été fait de façon à regrouper les communs et les entreprises consommant le plus pour obtenir 85% de la consommation électrique totale du site pour 20 compteurs installés. Les données relevées sont les nombres d'impulsions au niveau de chaque compteur d'électricité.

Chaque seconde, un capteur créé par la société Alro relève le nombre d'impulsions émises par le compteur auquel il est affecté. Une impulsion correspond à une quantité d'énergie consommée spécifique à chaque compteur. Le nombre d'impulsions est mesuré par seconde et est remonté à travers un bechhoff **[ANT-2014]** sur les différents serveurs soit par fibre optique soit par des câbles type RJ45 (Figure 40).

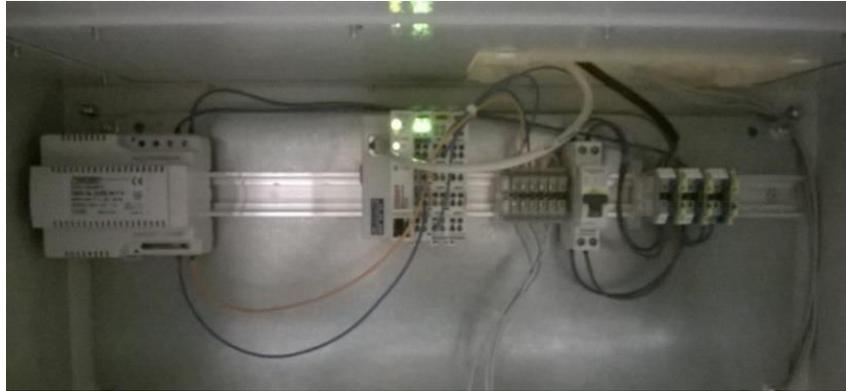


Figure 40 : Représentation d'un bekhoff qui fait le lien entre les compteurs d'électricité et les bases de données

Chaque compteur est identifié par une adresse IP qui lui permet de faire remonter les données acquises sur le serveur de l'entreprise. Il est affecté à ces données des coefficients (poids des impulsions et facteur compteur) qui sont stockées sous forme d'une puissance (Tableau 5). Ainsi la donnée en sortie du capteur est un nombre d'impulsions qui, après application d'un poids spécifique lié à cette impulsion, permet d'avoir une puissance pendant une durée déterminée, c'est-à-dire une énergie.

Cas d'étude	Données	Nom du compteur	Fréquence	Paramètres	Description Paramètres	Historique
Secteur tertiaire, entreprise connectée au système d'information	Consommation électrique globale	Beckhoff	1s	Sensorname	Identifiant unique par entreprise	18 mois
				gran	Granularité voulue	
				format	csv/json	
				fronts	timestamp	
				tots	timestamp	
				param	Puissance active (W)	

Tableau 5 : Données récoltées chaque seconde au niveau du transformateur basse tension pour mesurer la production photovoltaïque

Au niveau de la station transformatrice basse tension du quartier, un capteur Elco [ELC-2015] est placé au niveau de chaque entrée de bâtiment et au niveau de la production solaire. Ce capteur est branché sur chacune des phases du compteur et permet ainsi de renvoyer la puissance active et la puissance réactive par phase, à la fois pour la consommation en entrée des bâtiments, et pour la production à laquelle nous nous intéresserons dans cette étude (Tableau 6).

Cas d'étude	Données	Nom du compteur	Fréquence	Paramètres	Description Paramètres	Historique
Secteur tertiaire, panneaux photovoltaïques	Production électrique globale	ELCO	1s	Sensorname	Identifiant unique	18 mois
				gran	Granularité voulue	
				format	csv/json	
				fronts	timestamp	
				tots	timestamp	
				param	Puissance active (Watt) et puissance réactive (VAR) par phase	

Tableau 6 : Données récoltées chaque seconde au niveau de la transformatrice basse tension pour mesurer la production photovoltaïque

Au niveau du secteur résidentiel, un système permettant de mesurer la consommation électrique globale d'un foyer a été déployé sur 4 ménages tests. Ce système est installé à l'entrée du tableau électrique. Il comprend un appareil de mesures (Schneider Electric PM810) et un système de collecte de données (PC Engine) [ROD-2014]. Nous mesurons, à l'aide du PM810 de Schneider Electric les paramètres tels que la tension, le courant, la puissance active, la puissance réactive, chaque seconde directement sur chaque câble du compteur avec une précision de mesure de la puissance active de 1W (Tableau 7).

Cas d'étude	Données	Nom du compteur	Fréquence	Paramètres	Description Paramètres	Historique
Secteur résidentiel, maison	Consommation électrique globale	PM 810	1s	Sensorname	Identifiant unique par maison	18 mois
				gran	Granularité voulue	
				format	csv/json	
				fronts	timestamp	
				tots	timestamp	
				param	Puissance active (Watt) et puissance réactive (VAR) par phase	

Tableau 7 : Données récoltés chaque seconde au niveau du compteur global dans le secteur résidentiel

Pour la collecte des données dans le **secteur industriel**, nous avons travaillé à partir du projet Adaptive IES **[ADA-2015]**. Le matériel utilisé s'inscrit dans la stratégie globale de standardisation de l'automatisation du site de production de DEBIOPHARM RESEARCH & MANUFACTURING SA. En effet la CPU Siemens (S7-315 2PN/DP) choisie comporte 2 ports de communication Ethernet et suffisamment de ressources (Mémoire et Processeur) pour permettre de faire évoluer le système facilement. Les cartes d'entrées et de sorties sont tout à fait standards également. Le fait d'utiliser ce type de matériel permet une plus grande flexibilité et facilite la maintenance des installations.

### 3.3.2 Collecter les données météorologiques

Pour la majorité des installations photovoltaïques de faible puissance, aucune station météo n'est installée sur le site d'étude. Pour les autres installations, une station est installée et collecte les données de température et de radiation lumineuse. Il est possible que cette station collecte la température des cellules. C'est le cas pour notre station installée sur le site du technopôle. Cependant, nous avons voulu rester au plus proche de la réalité, c'est-à-dire dans le cas d'installations décentralisées de faible puissance avec aucune station météo présente sur le système.

Pour la collecte des données météorologiques, nous nous sommes basés sur le projet ELF4U **[ELF-2015]** développé au sein de l'institut Informatique de Gestion à Sierre. Nous nous sommes basés ainsi sur des services déjà existants et les solutions suivantes ont été analysées :

- ⇒ World weather online
- ⇒ Open Weather map
- ⇒ Meteoblue
- ⇒ Meteosuisse
- ⇒ Meteonews
- ⇒ Meteocentrale
- ⇒ Zygrib
- ⇒ Awekas

Les deux premières solutions proposent une version gratuite et fournissent les données sous forme API. Il s'agit de plateformes web automatiques qui fournissent des données mondiales. L'avantage de fournisseurs tels que Meteoblue, Meteosuisse, Meteonews et Meteocentrale est le contrôle des données par des météorologues qui les commentent et les ajustent en fonction de l'évolution du contexte météorologique **[ELF-2015]**. L'accès à ces services varie entre 300 et 2000 CHF/an.

Les meilleures solutions fournissant des données complètes et contrôlées sont Meteoblue, Meteonews, Meteosuisse. Nous avons utilisé les données de Météosuisse dans ce projet pour son meilleur rapport qualité / coût des données.

Notons que Meteonews ne fournit pas d'historique des prévisions météos et que l'historique des mesures météos provient des stations de Météosuisse. Les paramètres météorologiques récoltés sont : température, précipitations, irradiation, durée d'ensoleillement, hauteur de neige, vitesse et direction du vent, couverture nuageuse. Pour certaines des stations, les données réelles et prédites sont disponibles.

Sur la figure 41, nous avons un aperçu des stations météorologiques connectées à notre système d'information où nous distinguons 4 types de stations différentes :

- **Systèmes présentés dans ce rapport**
- **Stations avec des données météorologiques réelles et prédites**
- **Stations utilisées dans cette étude**
- **Stations avec uniquement des données hydrographiques.**

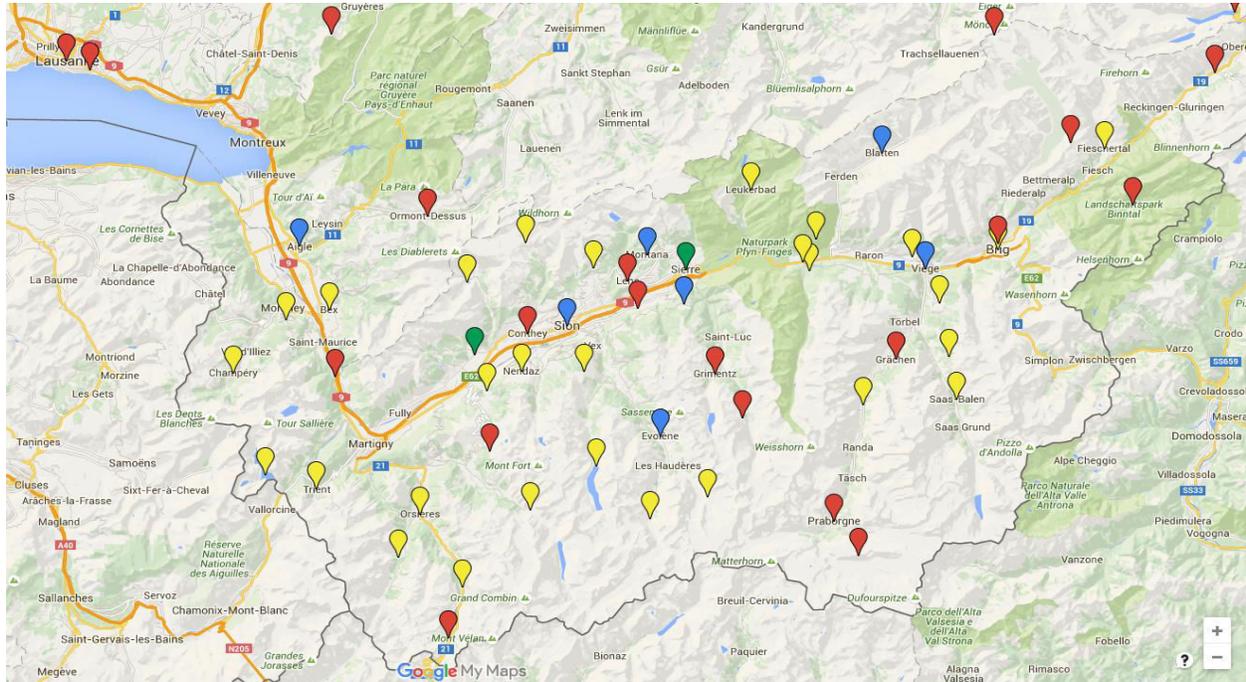


Figure 41 : Stations météorologiques connectées à notre système d'information

Pour caractériser à la fois la consommation énergétique et la production solaire décentralisée, nous nous sommes concentrés dans cette étude à récupérer chaque heure les données de température (°C) et de radiation lumineuse ( $W/m^2$ ) réelles et prédites (Tableau 8).

Capteur	Nom du compteur	Fréquence	Paramètres	Description Paramètres	Historique
Station météo	Météo Suisse	1h	Sensorname	Identifiant unique par station	5 ans
			gran	Granularité voulue	
			format	json	
			fronts	timestamp	
			tots	timestamp	
			param	Température réelle et prédite	

				(°C), radiation lumineuse réelle et prédite (W/m <sup>2</sup> )	
--	--	--	--	---	--

Tableau 8 : Données météorologiques utilisées dans notre outil de prédiction

### 3.3.3 Identifier la présence humaine

L'identification de la présence humaine est un des points fondamentaux dans la prédiction de la consommation électrique et dans la création de scénarios liés au pilotage de charges. Lorsque l'on veut surveiller l'intérieur d'une zone considérée, pour nous une pièce intérieure, on parle de détection intérieure [BEL-2010]. Plusieurs détecteurs existent : des détecteurs hyperfréquence, des détecteurs ultrason ou encore des détecteurs infrarouges passifs. Une étude comparative des différents systèmes est disponible [BEL-2010] et résumée dans le tableau 9. Une méthode d'identification consiste à calculer un taux d'occupation basé sur la mesure de la consommation au niveau des prises pour caractériser les usages spécifiques électriques utiles. Nous pouvons également citer les travaux qui consistent à placer un bracelet connecté sur les personnes pour connaître précisément le nombre de personnes [VIC-2013].

Type de détecteurs	Précision	Information				Coût	Intrusive
		Détection de mouvement	Nombre d'occupants	Localisation	Activité Physique		
Barrières lumineuses	Basse	+/-	+	-		-	-
Détecteur micro-ondes	Basse	+	-	-	-	+/-	-
Ultrasons	Basse	+	-	-	+/-	+/-	-
Détecteurs de chocs	Basse	+	-	-	+/-	+/-	-
Caméra infrarouge	Très haute	+	+	+	+	++	+
Pyrodétecteurs	Très haute	+	+	+	+	+/-	-
Transpondeurs	Basse	+	-	-	-	+/-	-
Caméra Vidéo	Très haute	+	+	+	+	+	+
Systemes biométriques	Basse	-	+	-	-	+	-

Tableau 9 : Techniques d'identification de la présence humaine

Il est toujours possible comme première approche et définition des modèles de référence d'utiliser les normes SIA 380/4 et SIA 2024 mais les valeurs restent des moyennes, qui ne répondent pas à un pilotage réactif.

En suivant nos contraintes pour la création d'un outil non intrusif, nous pouvons remarquer que les méthodes d'identification nécessitent la mise en place de capteurs à l'intérieur de notre système et ne sont pas adaptés à notre outil.

Nous préférons ajouter le planning de production comme information a priori sur nos différents systèmes pour à la fois améliorer nos données d'entraînement et identifier des zones de pilotage en fonction de la présence humaine (Tableau 10).

Cas d'étude	Données	Fréquence	Paramètres	Description Paramètres	Historique
Secteur tertiaire et industriel	Planning de production journalier	1 an	Id	Identifiant spécifique à un système	3 ans
			Heure_in	Heure d'arrivée des employés	
			Heure_out	Heure de sortie des employés	
	Planning de production par semaine		Id	Identifiant spécifique à un système	
			Jour_in	Premier jour de la semaine travaillé	
			Jour_out	Dernier jour de la semaine travaillé	

Tableau 10 : Données de présence utilisées dans notre outil de prédiction pour identifier les périodes de production

Dans le secteur tertiaire et industriel, les heures d'arrivée et de départ du personnel et les jours travaillés comme informations d'entrée permettront d'affiner la qualité du set d'entraînement en enlevant des données que l'on peut considérer comme du bruit (Figure 42). Dans le cas du secteur résidentiel, cela passe par la connaissance sur les consignes données aux systèmes de chauffage et d'eau chaude.

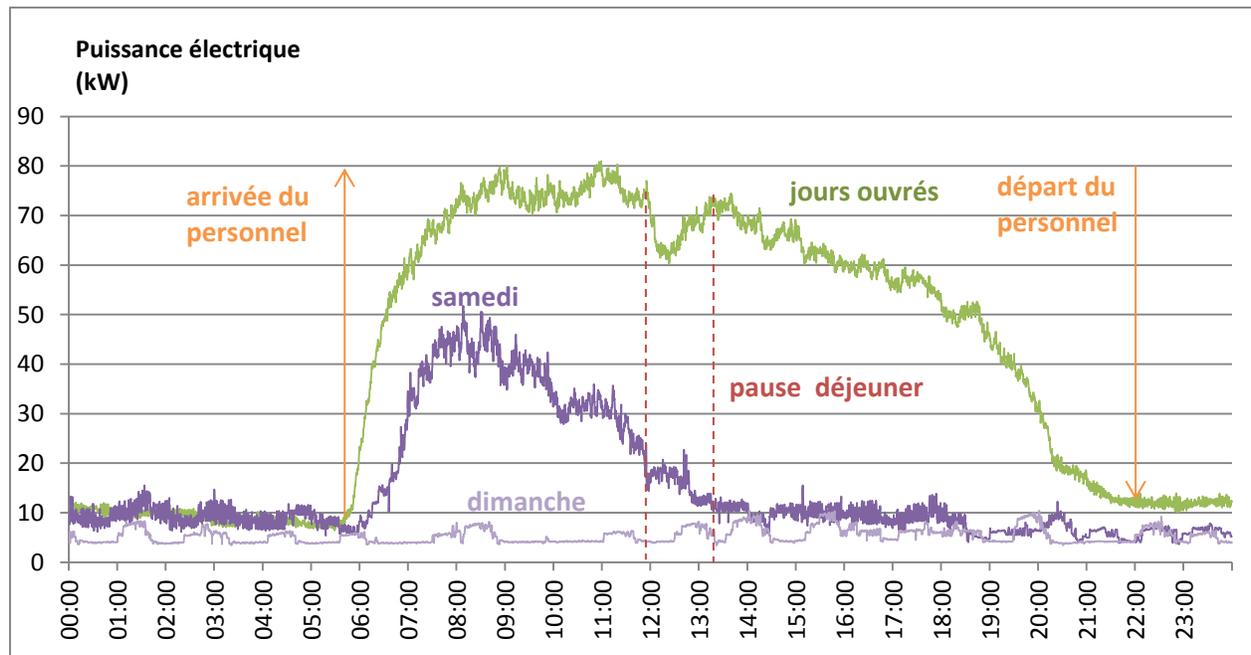


Figure 42 : Evolution des courbes de charges journalières moyennées sur le mois de Mars 2013 caractérisant les jours de la semaine, le samedi et le dimanche pour une industrie de fabrication d'étiquettes

### 3.4 Stocker les données

Selon l'objectif voulu et la taille du système étudié (maison, quartier) en question, le système de stockage mis en place diffère. Lors de nos premiers essais en 2012, lors du lancement du projet I-BATS, nous avons commencé le plus simplement possible avec des bases de données MySQL et nous avons très vite migré vers des bases de données NoSql [NOS-2015]. Cependant, notre objectif n'est pas seulement de stocker des données mais d'y faire appel sur des temps très courts (la seconde ou la minute par exemple pour le suivi de contraintes électriques) pour accéder aux informations des capteurs. La solution finale a été de créer plusieurs bases de données en parallèle distinctes MONGODB pour les données concernant la consommation et la production.

Pour les données météorologiques, nous nous sommes basés sur le projet ELF4U. Un centre de stockage et d'importations de données énergétiques a été créé. Après analyse, le choix de l'outil s'est porté sur une base de données 'Open Source' spécialisée dans le stockage et le traitement de séries temporelles : Influx DB [ELF-2015].

Cette base de données a été choisie pour ses fonctionnalités avancées en terme de traitement de séries temporelles notamment avec l'agrégation de données, l'interfaçage avec Grafana (un outil de visualisation) ainsi que pour son aspect 'Open source' et la communauté active qui l'entoure. Aujourd'hui le centre e-Energy center à travers le centre Icare nous met à disposition à travers une API spécifiques les données stockées (Figure 43).

A partir d'un logiciel d'analyse décrit dans le paragraphe suivant, nous pouvons accéder à tout moment aux données voulues. Pour communiquer avec les autres composants du système d'information, une solution à base de services web de type REST utilisant le format JSON a été utilisée. Le format JSON est un standard d'échange de données et de nombreuses bibliothèques permettant de le traiter sont présentes sur toutes les plateformes du marché. De tels services permettent la construction d'applications lisibles et facilement utilisables dans le cadre de développement d'applications clientes [CAT-2011] [BAK-2012].

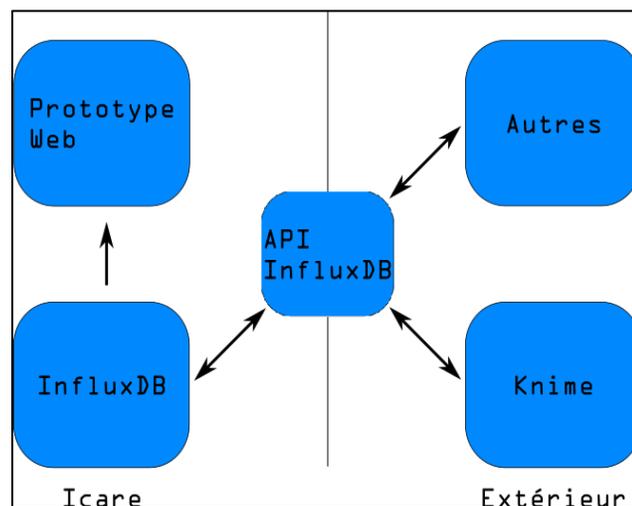


Figure 43 : Echanges d'information mises en place dans le cadre du projet ELF4U, [ELF-2015]

A ce stade, nous utilisons les bases de données MONGODB pour les données relatives à la consommation et la production de nos différents systèmes et les bases de données Influx pour les données météorologiques (Tableau 11). L'objectif final sera de basculer l'ensemble du système vers des bases de données Influx DB.

Cas d'étude	Données	Nom du compteur	Fréquence	Historique	Type base de données	Accès	Format
Secteur tertiaire	Consommation électrique globale	Bechkoff	1s	18 mois	MONGO DB	API REST	JSON
	Données météorologiques	Météo Suisse	3h	5 ans	Influx DB		

Tableau 11 : Accès aux données collectées par les capteurs dans le secteur tertiaire

### 3.5 Traiter et distribuer l'information

Par rapport à notre contrainte imposant de limiter au maximum nos coûts d'installation, nous sommes intéressés aux logiciels libres. Il existe des logiciels open source (gratuits), à l'instar de RapidMiner ou encore Weka, R, Knime résumé dans le tableau 12 et open source :

- ⇒ **KNIME** acronyme de Konstanz Information Miner, est un logiciel libre édité par un laboratoire de l'université de Constance dénommé Nycomed Chair for Bioinformatics and Information Mining. Il intègre tous les modules d'analyse de Weka et permet de créer des scripts en langage R.
- ⇒ **R** est un langage et un environnement permettant d'effectuer des calculs statistiques et de créer leurs graphiques. R fournit un large éventail de techniques statistiques et graphiques telles que la modélisation linéaire et non linéaire, les tests statistiques classiques, l'analyse des séries chronologiques, la classification. R est publié selon les termes de la licence GNU sous forme de code source.
- ⇒ **Orange** est un logiciel libre créé à l'université de Ljubljana en Slovénie. Ce logiciel est doté d'une interface homme-machine conviviale. Il est développé en C++ et en Python.
- ⇒ **Weka** est un logiciel libre de fouille de données créé par l'université de Waikato (Nouvelle-Zélande).

- ⇒ **RapidMiner** est un logiciel libre distribué par la société Rapid-I, basée à Dortmund en Allemagne. Il intègre le *Business Intelligence* dont les principales fonctionnalités sont l'ETL, l'OLAP, la production d'états et l'exploration de données et les techniques classiques comme les SVM, l'ACP, les arbres de décision et les réseaux neuronaux.
- ⇒ **Tanagra** est un logiciel libre d'exploration de données développé du laboratoire ERIC de l'Université Lumière Lyon 2.

Volume de données	Produit	Spécialité	Editeur
Faible (dizaines de milliers d'enregistrements)	NeuralWorks Predict	Réseau de neurones	NeuralWare
	NeuroOne	Réseau de neurones	Netral
	Wizwhy		Wizsoft
	WEKA		"open source" Université de Waikato, Nouvelle-Zélande)
	R		"open source"
	RapidMiner		Rapid-I
	DataLab	Prétraitement des données	Complex Systems
	Alice	Arbre de décision	Isoft
Moyen (centaines de milliers d'enregistrements)	KnowledgeSEEKER	Arbre de décision	Angoss
	KnowledgeSTUDIO		Angoss
	C 5.0	Arbre de décision	RuleQuest Research
	CART	Arbre de décision	Salford Systems
	MARS	Arbre de décision	Salford Systems
	TreeNET	Gradient boosting	Salford Systems
	Predictive Modeling		Salford Systems
	Polyanalyst	Data mining + Text mining	Megaputer
	TANAGRA		Université de Lyon
	JMP		SAS Institute
	SPAD		Coheris
	S-PLUS		TIBCO Software
	Elevé (millions d'enregistrements)	InfiniteInsight	
KNIME		Data mining + Text mining	Knime AG incorporated in Switzerland
Oracle Data Mining			Oracle
IBM SPSS Statistics			IBM
IBM SPSS Modeler		Data mining + Text mining	IBM
Statistica Data Miner		Data mining + Text mining	Statsoft
Insightful Data Miner			TIBCO Software
SAS/STAT			SAS Institute
Enterprise Miner		Data mining + Text mining	SAS Institute

Tableau 12 : Tableau récapitulatif des logiciels d'analyse de données [DAT-2015]

Avec nos différentes bases de données qui représentent des centaines de millions d'enregistrements, nous nous sommes concentrés sur l'utilisation du logiciel KNIME. Nous avons ainsi pu créer autour de cet outil un système d'information complet qui part des données fournies par les capteurs, les stocke, est connecté avec KNIME qui nous permet le traitement des données et d'envoyer les informations relatives aux différentes prédictions à un niveau supérieur de pilotage. Plusieurs modèles mathématiques peuvent être testés. Nous pouvons ainsi comparer plusieurs méthodes et algorithmes de prédictions décrites dans le chapitre 2 :

- ⇒ Méthode linéaire : Modèle ARIMA et une régression linéaire
- ⇒ Méthode non linéaire : MLP, PNN, SVM, Random Forest et le gradient Boosted Tree

Une fois les résultats obtenus, il est possible de les réécrire dans la base de données, d'envoyer l'information vers un web service ou directement à des automates programmables (Figure 39).

L'ensemble du processus de prédiction a été testé et industrialisé dans le cadre de deux projets. Le premier nommé ELF4U [ELF-2014] consistait à déployer un système de prédiction pour les installations solaires pour le fournisseur et producteur d'énergie local ESR (Energie Sion Région). Le deuxième nommé APEAS avait pour objectif de fournir un outil de désagrégation de la courbe de charges de 200 maisons [APE-2015]. Dans ces projets, nous avons utilisé le logiciel KNIME qui met à disposition des interfaces qui permettent la connexion directement avec les bases de données ou à travers d'API (Figure 44). Nous avons utilisé cette dernière à travers des requêtes de type GET qui peuvent être envoyées avec une fenêtre de temps choisie. Après analyse et prédiction du ou des flux considérés, les données sont reconverties en format JSON puis envoyées sur la base de données par des requêtes POST [KNI-2017].

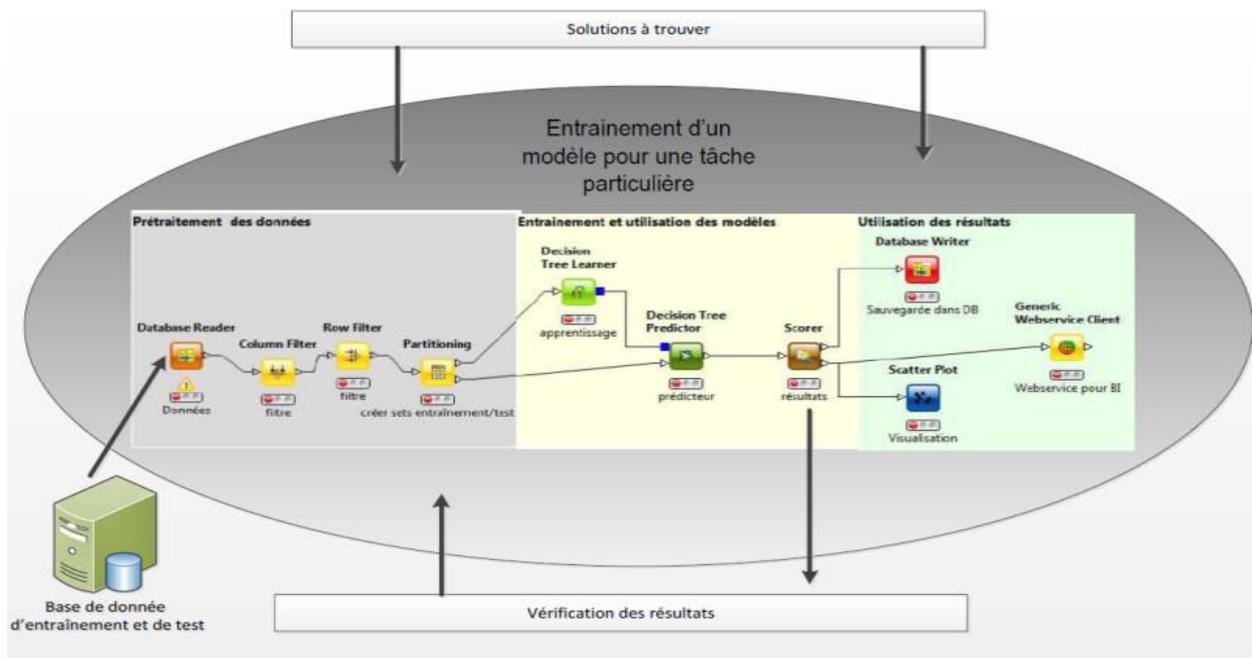


Figure 44 : Représentation du traitement des données à l'intérieur du logiciel d'analyse KNIME [HES-2012]

## 3.6 Conclusion

Nous avons décrit le processus qui nous a permis de mettre en place notre système d'information. Au vu de la bibliographie en particulier sur l'identification des appareils électriques, nous avons choisi une collecte des données de consommation à la seconde. Nous récoltons la puissance active et réactive au niveau du secteur résidentiel pour identifier les plus gros consommateurs et ceux ayant le plus gros potentiel de flexibilité : le chauffage et l'eau chaude. La production solaire est également récoltée sur chacun de nos sites.

Puisque nous voulons limiter le nombre de capteurs présents au sein de nos systèmes, nous apportons des informations a priori sur la technologie et l'énergie associées au chauffage et à l'eau chaude. Enfin le planning de production est également ajouté dans le secteur tertiaire et industriel pour limiter le bruit dans l'apprentissage de nos modèles mathématiques.