

- Introduction aux systèmes de communication, traitement et protection des données



# Introduction

- Avec le déploiement des technologies de l'information et de la communication (TIC) sur les réseaux publics d'électricité, le nombre et la variété des données disponibles croît de manière exponentielle.
- De nombreuses technologies déployées sur les réseaux en aval et en amont du compteur produisent des données : capteurs sur les réseaux, compteurs évolués installés chez les clients finals pour mesurer la consommation et la production et également la qualité sur les réseaux, objets connectés chez les consommateurs (électroménager, véhicules électriques, Energybox, GTB, ...).
- Le déploiement de l'ensemble de ces technologies générera plusieurs Pétaoctets (soit plusieurs millions de milliards) de données. Il faudra, donc, prévoir une architecture de réseau modulaire, évolutive et extensible.



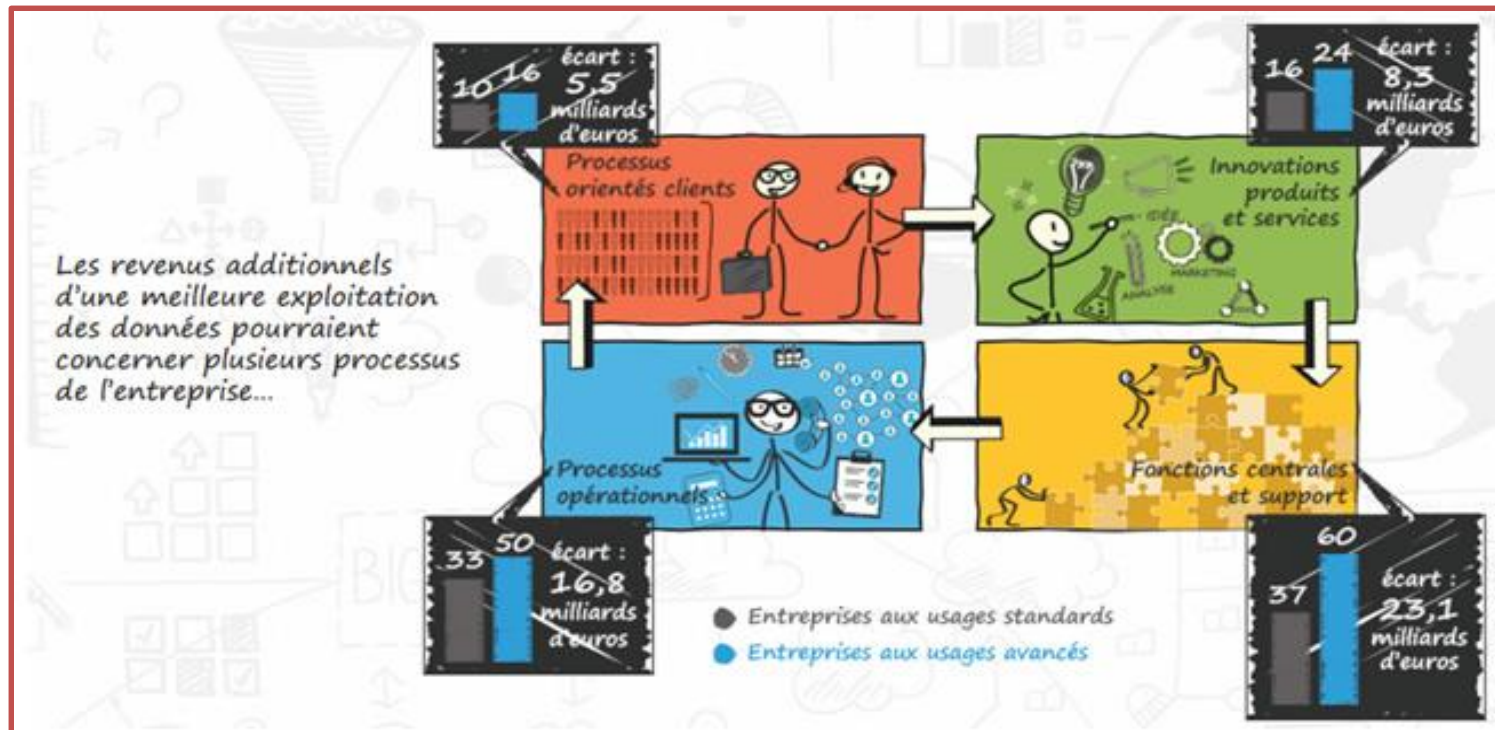
# Gestion des données des Smart grids

- La « gestion des données » fait référence à la transformation de données en connaissance utilisable, afin d'en extraire une valeur économique.
- La gestion des données comprend tous les aspects de collecte, de filtrage, d'analyse, de stockage et de mise à disposition des données à d'autres acteurs, d'autres applications et d'autres utilisateurs, **incluant** les problématiques d'identification des données, de validation, d'exactitude, de mise à jour, de synchronisation des équipements, d'homogénéité des bases de données, de sécurité et de confidentialité, etc.
- La gestion des données apporte des réponses sur les différentes questions sur les équipements de réseau, leur état et leur besoin de mise à niveau, sur la façon d'exploiter le réseau et d'améliorer cette exploitation, sur les besoins et les préférences des clients.



# Gestion des données des Smart grids

- Selon l'Electric Power Research Institute (EPRI, institut qui réalise des recherches pour l'industrie de la production d'électricité des États-Unis), les **données et leur analyse** sont les **nouveaux facilitateurs** du Smart grid.
- En effet, ces différentes données apporteront de nouvelles opportunités aux différents acteurs du monde de l'énergie dans les domaines de la gestion de l'énergie, de l'exploitation des réseaux ou de l'évolution de la relation avec le client, dans le cas où ces acteurs pourront analyser ces données et en tirer parti.



# Gestion des données des Smart grids

- **L'analytique** est la mise en œuvre de méthodologies statistiques et mathématiques dans les processus de l'entreprise (qu'ils soient opérationnels ou support), voire dans sa stratégie ou dans son modèle d'affaires. Dans la mesure où la matière première de l'analytique est la **donnée numérique**, cette mise en œuvre est consubstantielle (jointe) de la transformation digitale de nos sociétés, de notre économie et donc des entreprises.
- Business Intelligence, datamining, business analytics, big data et maintenant smart data : ont devenu depuis quelques années des concepts de l'analyse de la donnée en termes d'efficacité, de performance et d'opportunités de développement de nouvelles offres et services ciblés.

## Typologie d'acteurs autour de la data

- ▶ **Les producteurs de données** : ils constitueraient de véritables bases de données fiables qui, grâce à leur capital confiance, pourraient être revendues à des tiers.
- ▶ **Les agrégateurs de données** : il agrégeraient, croiseraient et contextualiseraient ces données à des fins d'analyse.
- ▶ **Les utilisateurs de données** : ils développeraient, à partir des analyses obtenues, des nouveaux services et nouvelles applications. Ces utilisateurs peuvent être des fournisseurs de services et des *start-ups* développant des applications spécifiques.
- ▶ **Les opérateurs d'infrastructures du big data** : nouveaux acteurs du numérique, opérateurs de télécommunications...
- ▶ **Les gestionnaires d'infrastructures** : ils pourront louer, en plus de leurs capacités de stockage, des capacités de calcul à des sociétés ayant des besoins ponctuels en matière de traitement de big data.
- ▶ **Des acteurs de confiance** s'assurant du maintien de l'intégrité / qualité des données, depuis leur *sourcing* jusqu'à leur utilisation, rassurant ainsi utilisateur et consommateur finaux.

- **Business Intelligence (BI)**, également connue sous le nom d'**informatique décisionnelle**, est l'ensemble des outils et méthodes visant à transmettre les informations pertinentes aux managers d'entreprise. Son but est de les aider à comprendre leur environnement et de **les accompagner dans leurs prises de décisions stratégiques**. Les besoins de l'utilisateur sont donc au cœur d'une Business Intelligence efficace.

- **Business analytics (BA)** est une gamme de produits informatiques destinés à l'aide à la prise de décision. L'accent est mis sur l'utilisation de la statistique descriptive et de l'analyse prédictive au service du développement commercial.

- **Smart data** : La notion de smart data qui peut être traduite comme une **donnée intelligente, ou plutôt pertinente**, est utilisée pour mettre en avant la nécessité de se concentrer sur les datas pertinentes dans la démarche de collecte et de traitement des données.

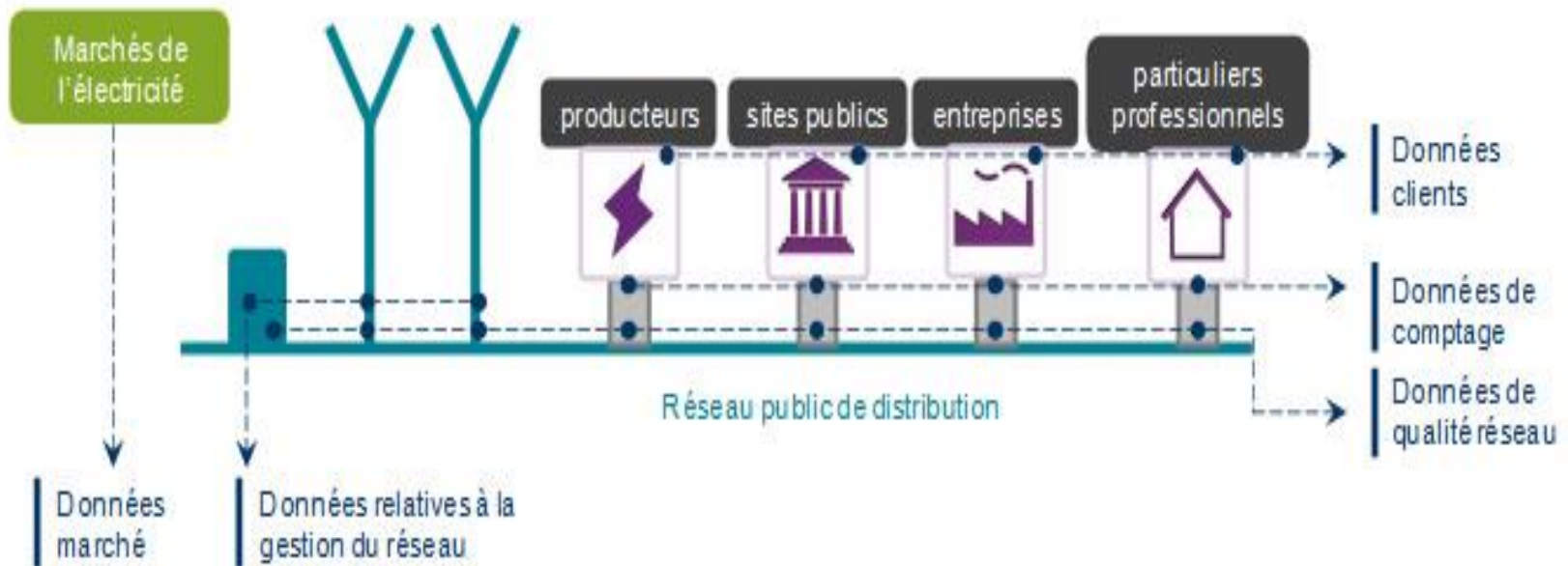
Le terme de smart data est apparu en réaction à l'explosion des volumes de données collectées dans le cadre du phénomène big data afin de ne pas entraîner un gaspillage des ressources techniques et des coûts de traitement additionnels.

- **Big Data**, signifie **mégadonnées**, grosses données ou encore **données massives**. Ils désignent un ensemble très volumineux de données qu'aucun outil classique de gestion de base de données ou de gestion de l'information ne peut vraiment travailler.

- **Data Mining**, désigne **l'analyse de données depuis différentes perspectives et le fait de transformer ces données en informations utiles, en établissant des relations entre les données** ou en repérant des patterns (modèles).

# Types de données des Smart grids

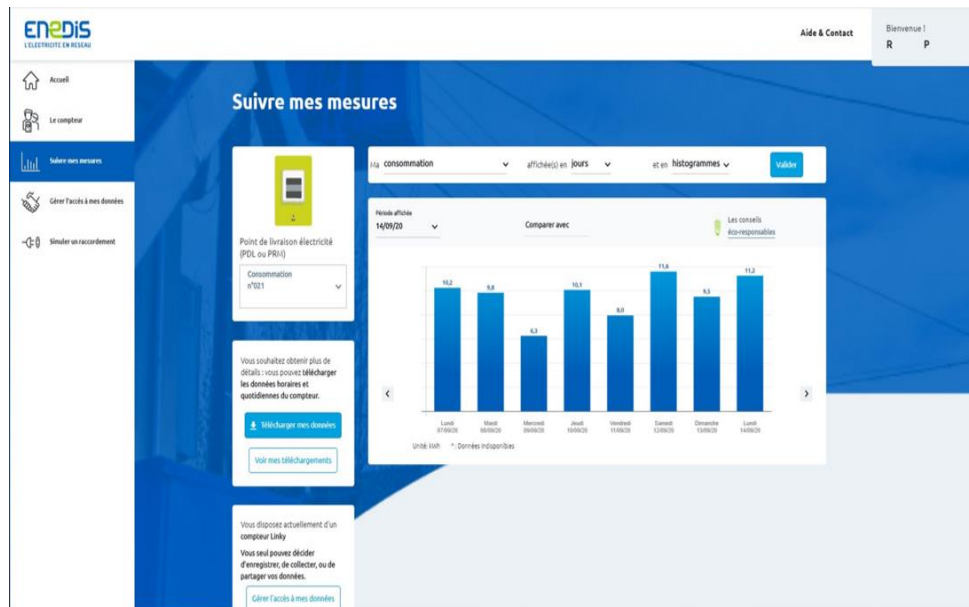
- ERDF est gestionnaire d'un nombre croissant de données : données de comptage, données techniques relatives au réseau, mais également données client et données relatives au fonctionnement des marchés de l'électricité.
- La collecte d'un nombre considérable et varié de données, qu'ERDF doit gérer, se fait tout en préservant la confidentialité des informations commercialement sensibles et des données à caractère personnel.



**Types de données collectées par ERDF**

# Les données de comptage

- Afin de s'adapter aux besoins des clients et à leurs installations, ERDF a développé toute une gamme de compteurs permettent de relever:
  - les puissances de consommation et production relevées à intervalles réguliers ;
  - les puissances maximales atteintes quotidiennement ;
  - d'autres mesures comme l'énergie réactive ou la tension moyenne.
- Fig : tableaux de bord de la consommation d'un ménage via une interface Web



- La solution « Mon suivi Electricité » a mis à disposition des particuliers un comparatif de consommation ce qui a conduit à une diminution de la consommation des foyers sans avoir recourt à des équipements particuliers.



# Les données clients

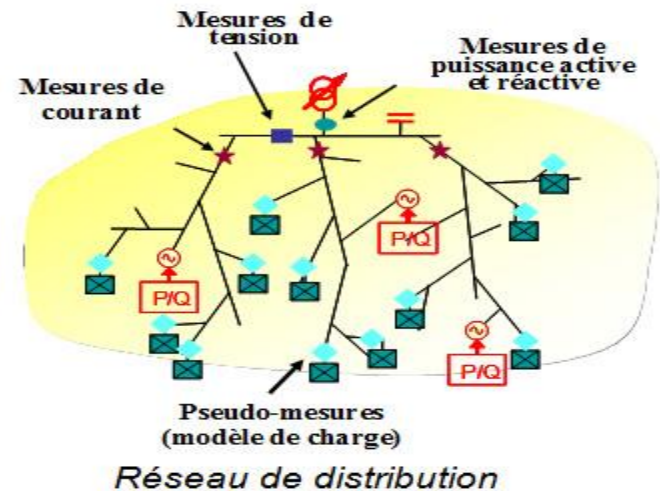
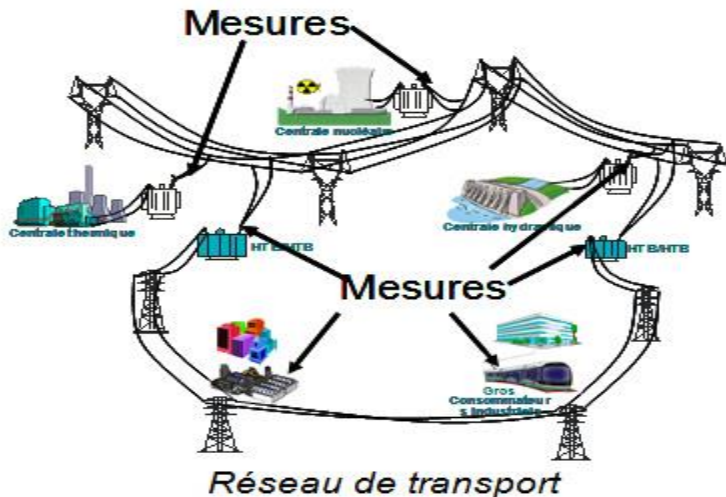
- ERDF est également dépositaire des données techniques relatives à l'ensemble des clients finaux et des producteurs d'électricité comme le type de compteur, la puissance installée, le domaine de tension, l'existence de dispositifs particuliers de limitation des perturbations, etc. Le distributeur est chargé de tenir ces données à jour et de les mettre à la disposition du client, du fournisseur d'électricité titulaire du point et d'autres acteurs légitimes et autorisés, tels que les agrégateurs ou les opérateurs d'effacement.
- De manière équivalente, dans le cadre des contrats uniques, les fournisseurs d'énergie partagent avec les distributeurs les données client nécessaires à l'exercice de leurs missions. Ils communiquent un certain nombre de données contractuelles avec ERDF, telles que l'identité du client, l'offre de fourniture d'électricité ou le responsable d'équilibre en charge du point de livraison concerné, ainsi que d'autres informations utiles à ERDF (adresse de contact, type d'entreprise le cas échéant, etc.).

# Données relatives aux marchés de l'électricité

- L'électricité se stockant difficilement, un équilibre est continuellement nécessaire entre la production et la consommation. Un dispositif spécifique confie à des acteurs appelés **responsables d'équilibre (RE)** la charge financière des déséquilibres **entre injection et soutirage** à l'échelle de leur périmètre.
- Lorsqu'il y a un déséquilibre entre la production et la consommation, **le mécanisme d'ajustement** permet de procéder à des adaptations des comportements de producteurs et de consommateurs. Ce mécanisme peut alors faire intervenir des capacités de production ou d'effacement raccordées aux réseaux publics de distribution.
- Le mécanisme NEBEF (notifications d'échanges de blocs d'effacement) permet aux **opérateurs d'effacement** de vendre de l'énergie effacée sur les marchés au même titre qu'un moyen de production.
- Enfin, le **marché de capacités** permettra de sécuriser l'approvisionnement en électricité et contribue à garantir qu'en cas de pic de consommation, les capacités de production et d'effacement installées sont suffisantes pour couvrir les besoins de consommation.
- **Le fonctionnement de ces marchés nécessite de nombreux échanges d'informations entre les différents acteurs concernés : responsables d'équilibre, acteurs de l'ajustement, exploitants de capacités, gestionnaires de réseaux, opérateurs d'effacement, etc.**

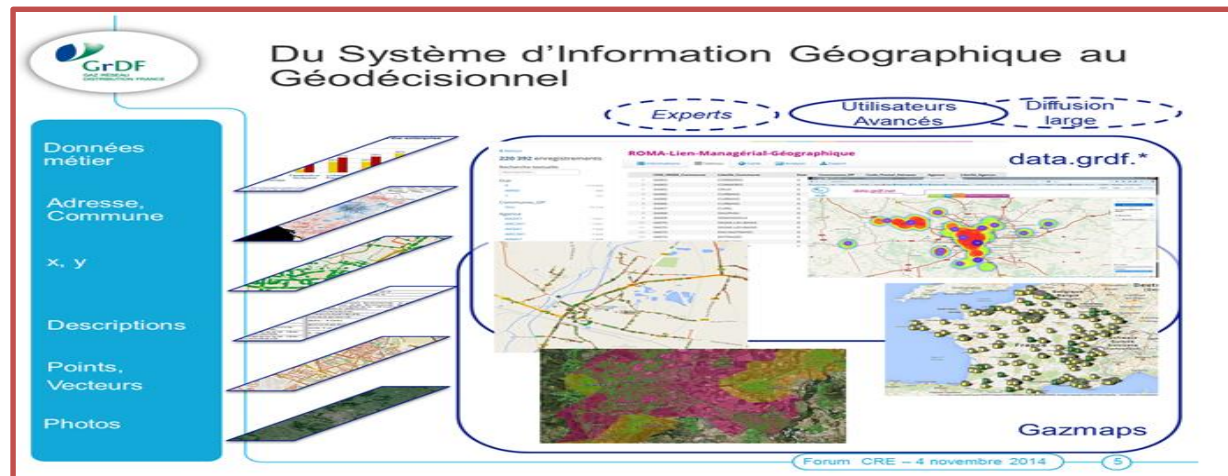
# Les données de qualité réseau

- La qualité de l'électricité distribuée aux consommateurs se définit par :
  - la continuité de l'alimentation électrique ;
  - la qualité de l'onde électrique, en termes de stabilité de la tension et de la fréquence notamment.
- Plusieurs dispositifs de comptage installés sur le réseau permettent de réaliser des mesures de qualité, par exemple d'enregistrer et de dater les coupures ou les variations de la tension délivrée.
- En particulier, les compteurs communicants Linky permettent de relever ces indicateurs. Avec la généralisation de ce type de compteurs, ERDF disposera d'une vue très fine et quotidienne de la qualité d'alimentation sur le réseau public de distribution.



# Données liées à la gestion du réseau

- Pour assurer les missions d'exploitation et de développement du réseau public de distribution, ERDF dispose d'un système d'information géographique (SIG). Cet outil lui permet de construire une description fine du réseau public de distribution d'électricité, des postes de transformation jusqu'aux dispositifs de comptage.
- ERDF déploie notamment des capteurs permettant de mesurer les flux d'énergie au niveau des équipements du réseau. Ainsi, des postes sources instrumentés permettent à ERDF de mesurer et collecter les quantités d'énergie y transitant à des intervalles très fins.
- Grâce à ces données, ERDF peut, notamment, adresser aux demandeurs de raccordement les propositions techniques et financières adaptées à leur demande et aider les clients et les collectivités à optimiser les effets liés à l'évolution du réseau sur leurs projets (raccordement, aménagement, etc.).



# Traitement des données

- **Les étapes de traitement des données**

Après l'étape de collecte grâce aux différents capteurs, ERDF opère les traitements génériques suivants :



- **Contrôle** : cette étape permet de s'assurer que les données brutes collectées sont valides.
- **Correction** : sur certains types de données, des mécanismes d'estimation ou de correction peuvent être mis en œuvre. Ces dispositifs sont concertés (arrangés) avec l'ensemble des parties prenantes.

Par exemple, en cas de dysfonctionnement d'un dispositif de comptage d'un client particulier, l'énergie consommée est estimée à l'aide de méthodes reconnues par les organismes compétents (cf. article 20 du cahier des charges fonctionnel sur le comptage électrique joint à la Communication de la CRE du 29 janvier 2004).

# Traitement des données

- **Stockage** : ERDF procède ensuite à un stockage sécurisé des données validées et corrigées dans ses systèmes informatiques. référentiel de sécurité rédigé par ERDF, évalué et certifié par l'Agence nationale de la sécurité des systèmes d'information (ANSSI).
- **Traitements adaptés** : des traitements adaptés peuvent être effectués sur les données. Par exemple, il est possible de retraiter les données de consommation afin d'en déduire des données de facturation à destination des fournisseurs titulaires des contrats de fourniture, d'effectuer des agrégations de données de consommation et de production sur le périmètre d'une concession, etc.
- **Mise à disposition** : les données sont ensuite utilisées par ERDF pour l'exploitation, la maintenance et le développement du réseau public de distribution. ERDF transmet également aux acteurs externes les données nécessaires pour l'exercice de leurs activités ou missions (clients finaux, fournisseurs, collectivités, tiers, régulateur, etc.) dans le respect du cadre juridique, notamment concernant la protection des don

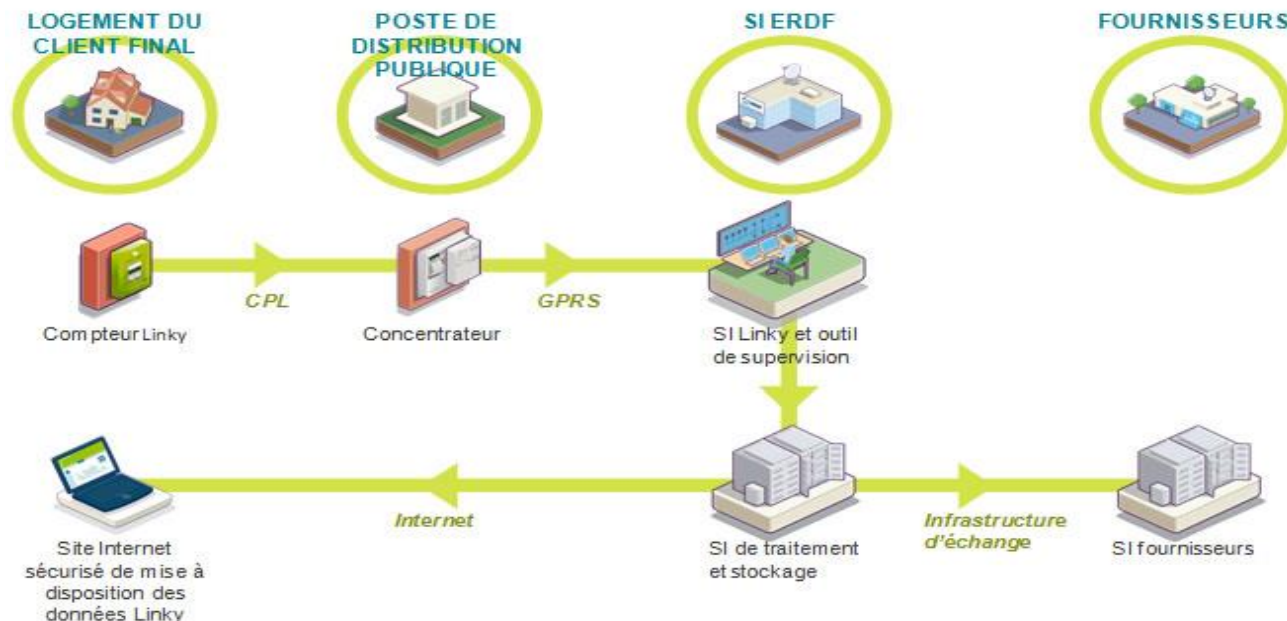


# Traitement des données

## Un exemple, la chaîne de traitement des données Linky

- Une fois mesurées par les compteurs communicants Linky, les données brutes sont transmises cryptées (codée) par courant porteur en ligne (CPL) aux concentrateurs situés dans les postes de distribution publique, puis par GPRS à un système d'information national. ERDF valide les données avant de les stocker sur des serveurs sécurisés, en conformité avec les recommandations légales en la matière. Ces données sont ensuite mises à disposition des acteurs autorisés. Par exemple, dans le cas de données individuelles de consommation, elles sont affichées dans l'espace client du consommateur concerné et transmises à son fournisseur titulaire pour l'exécution du contrat de fourniture.

### Chaîne de traitement des données Linky



# Le big data

ERDF s'est dotée d'un outil nommé Système de traitement des mesures (STM), fondé sur des nouvelles technologies telles que Linky et les capteurs installés partout sur le réseau.

STM est un véritable entrepôt (magasin) des données de la distribution d'électricité, il stocke et gère la totalité des données de mesure d'électricité produites par ERDF (réseaux et comptage), les informations descriptives des points associés à ces mesures (clients, contrats, responsables d'équilibre, fournisseurs, localisation, etc.), les données météorologiques, les profils de consommation utilisés pour la reconstitution des flux, etc.

L'outil STM s'appuie sur une technologie big data évolutive en termes de capacités de stockage et de traitement des données.





# Le big data

La technologie big data utilisée dans l'outil STM se caractérise par les « 4 V » :

**volumétrie** : le déploiement des compteurs communicants entraîne un télé relevé quotidien. Les volumes concernés se comptent en téraoctets, voire en pétaoctets.

**vélocité** : les temps de traitement peuvent être réduits d'un facteur compris entre 10 et 20 par rapport aux systèmes traditionnels de gestion de données.

De plus, le système STM permet d'effectuer simultanément deux natures de traitements différents : des traitements analytiques fondés sur des grandes quantités de données et des traitements transactionnels rapides sur des mesures unitaires ;

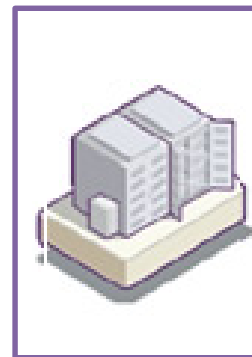
**variété** : STM prend en charge une forte diversité de types de données (mesures, données météorologiques, description du réseau, etc.). Une multitude de croisements devient possible et engendre de nouvelles connaissances issues de l'analyse collective de ces données ;

**véracité** : la qualité des analyses menées repose sur la validité des données collectées. La multiplication des sources de données et l'accroissement des volumes mis en jeu apportent des moyens supplémentaires de contrôler la véracité des données et de détecter les erreurs éventuelles.

<https://youtu.be/ByahUk76yal>

<https://youtu.be/7k8JdXqTHZQ>

<https://youtu.be/uMLamohaHGo>



## Les 4 "V" du Big Data

Volumétrie

Véracité

Variété

Vélocité

# Outils d'exploitation et d'analyse des données

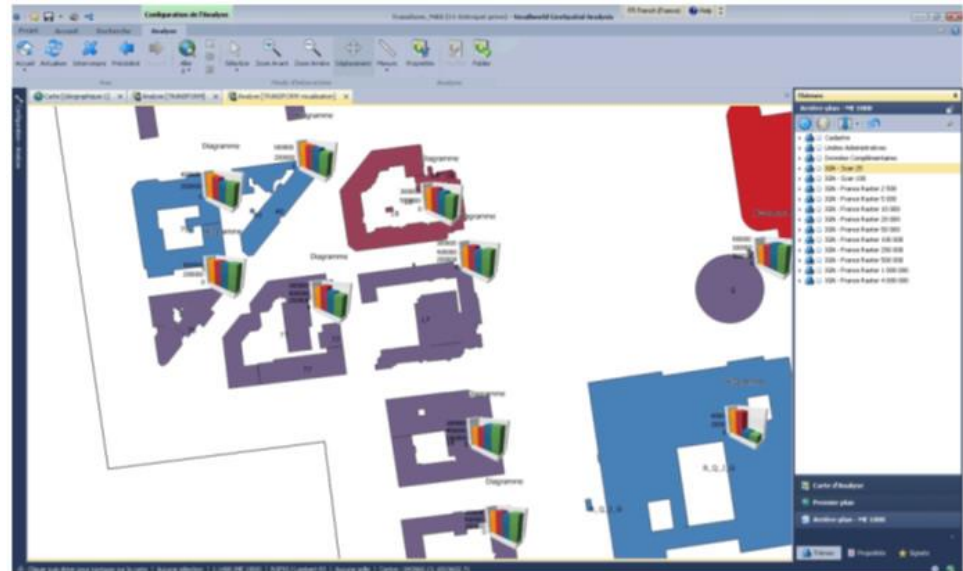
Les données collectées irriguent un grand nombre d'outils, qui permettent de les exploiter, les croiser, les analyser et les mettre à disposition des parties prenantes.

Citons par exemple **ÉRABLE** (**É**tudes pour un **R**éseau dur**ABLE**) est l'outil de développement du réseau public de distribution de ERDF. Il intègre l'ensemble des données réseau (production décentralisée, contraintes, etc.) pour réaliser les études réseau ERDF et anticiper le dimensionnement du futur réseau public de distribution.

De même, le Système d'information géographique d'ERDF (SIG), référentiel des données descriptives du réseau, permet d'alimenter des outils d'analyses thématiques des données. Par exemple, en croisant les données géographiques des référentiels réseau et clientèle, il permet d'illustrer géographiquement la consommation par quartier et type de client dans le cadre d'études réseau et de projets d'aménagements.

**Exemple d'analyse d'évolution des consommations sur un système d'information géographique**

[https://youtu.be/sAkbabX\\_dec](https://youtu.be/sAkbabX_dec)

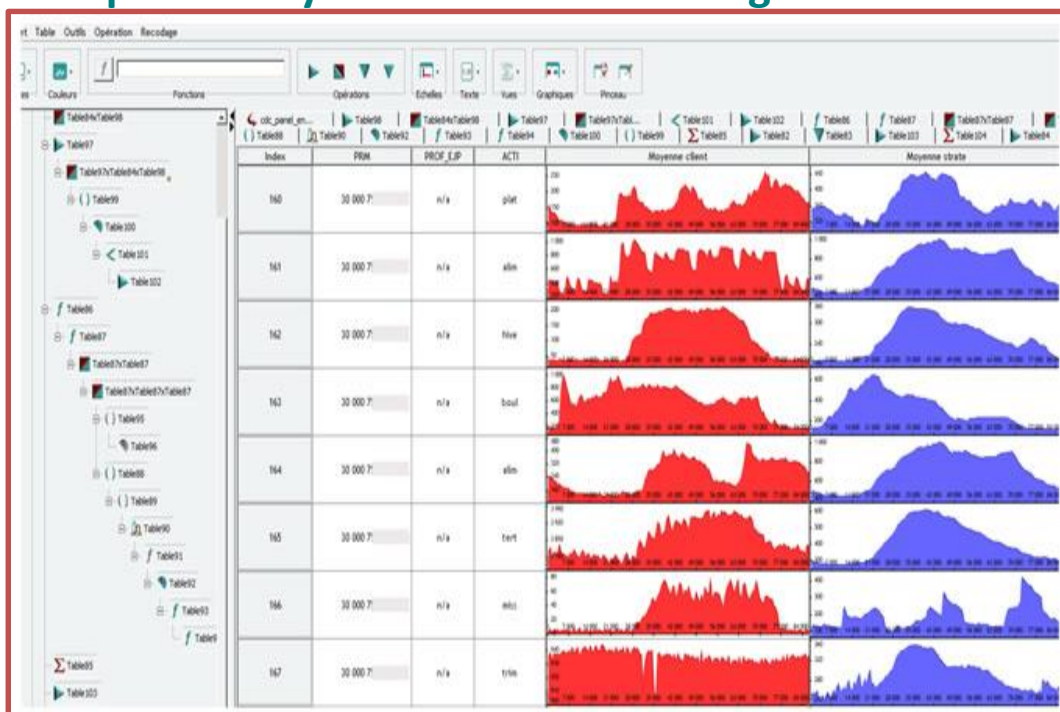


# Outils d'exploitation et d'analyse des données

Les données collectées sont également utilisées par différents outils d'analyse et de traitement statistique en vue de la reconstitution des flux pour les marchés de l'électricité. Grâce à des outils spécifiques dits de « **data mining** », elles permettent par exemple de **construire des profils type**, d'**identifier les pertes non techniques** potentielles ou d'**élaborer des méthodes d'évaluation de l'effacement** via des panels de consommateurs. [https://youtu.be/OZ\\_LueQDbm0](https://youtu.be/OZ_LueQDbm0)

## Exemple d'analyse de courbes de charge de consommation

## Data Mining



# Protection de la vie privée des usagers

La mise en place des compteurs communicants peut comporter des risques importants en matière de protection de la vie privée des usagers.

La CNIL a notamment insisté sur les risques de traçage des usagers compte tenu du volume de données collectées et des informations qu'elles fournissent sur le quotidien des usagers. **En** effet, les informations de consommation d'énergie transmises par les compteurs peuvent être très détaillées (par exemple, enregistrement de la consommation au pas demi-horaire) et permettent de présager (supposer) certaines habitudes de vie des habitants d'un foyer (temps de présence, d'absence, horaire de coucher et de réveil, etc.).

Pour cette raison, la sécurité et la confidentialité des données collectées par ces compteurs doivent notamment être garanties par la mise en œuvre de mesures adéquates (Délibération de la CNIL n° 2012-404 du 15 novembre 2012 portant recommandation, relative aux traitements des données de consommation détaillées collectées par les compteurs communicants).