

# OSGE

---

# L'OBSERVATOIRE STATISTIQUE

Une entité au carrefour de la donnée EDF

Intervention au Mastère Big Data  
S.TERNOT – 23/01/2017



# SOMMAIRE

- ▶ CONTEXTE D'UN OBSERVATOIRE STATISTIQUE
- ▶ DES DONNÉES QUI AUGMENTENT EN VOLUME ET DIVERSITÉ
- ▶ UN PROCESSUS GÉNÉRIQUE DE PRODUCTION
- ▶ UN « DICTIONNAIRE » POUR PARTAGER
- ▶ UN PATRIMOINE À VALORISER PAR LA DIFFUSION ET LA RÉUTILISATION

# CONTEXTE DE L'OBSERVATOIRE STATISTIQUES D'EDF SA

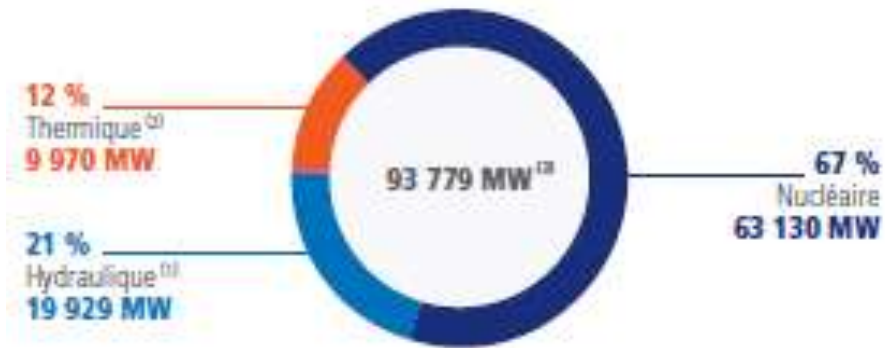
- QUELQUES DONNÉES D'EDF SA
- VISER L'ÉQUILIBRE PRODUCTION + CONSOMMATION
- LA DOAAT : L'OPTIMISEUR D'EDF « RESPONSABLE D'ÉQUILIBRE »
- OSGE : UN OPÉRATEUR DE DONNÉES
- REPRÉSENTANT EDF À L'EXTERNE
- VALORISER LE PATRIMOINE DE DONNÉES DE L'ENTREPRISE

## SOMMAIRE

- ▶ CONTEXTE D'UN OBSERVATOIRE STATISTIQUE
- ▶ DES DONNÉES QUI AUGMENTENT EN VOLUME ET DIVERSITÉ
- ▶ UN PROCESSUS GÉNÉRIQUE DE PRODUCTION
- ▶ UN « DICTIONNAIRE » POUR PARTAGER
- ▶ UN PATRIMOINE À VALORISER PAR LA DIFFUSION ET LA RÉUTILISATION

# QUELQUES DONNÉES D'EDF SA (2015)

Capacité installée 2015 France continentale



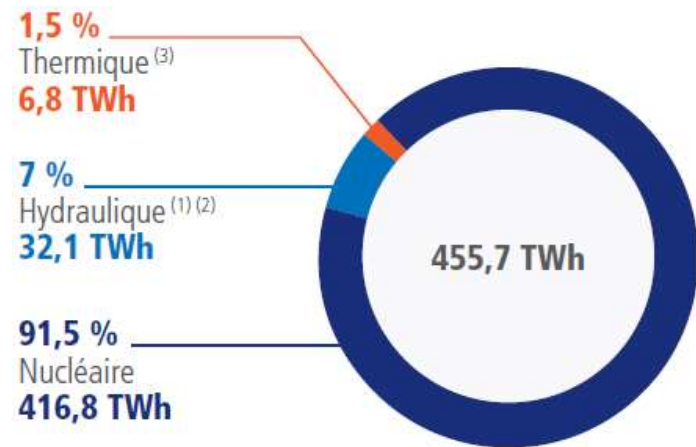
NB : exprimé en mégawatts de puissance maximale couplée au réseau.

(1) Hors Corse et outre-mer, soit 440 MW en 2015.

(2) Hors Corse et outre-mer, soit 674 MW en 2015.

(3) Hors capacités éoliennes de 12 MW et y compris capacités marémotrices de 240 MW.

Production 2015 France continentale



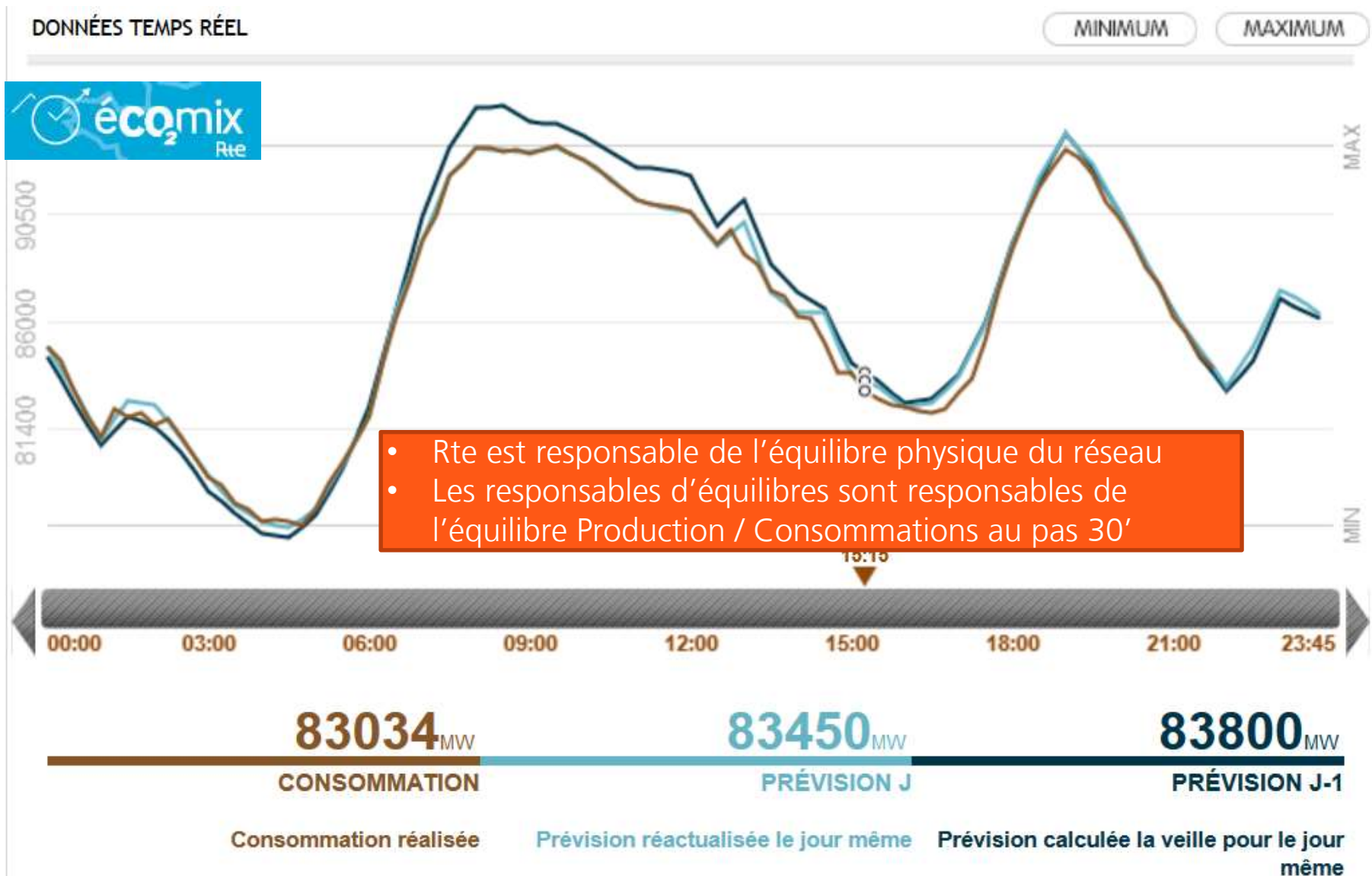
(1) Hors Corse et outre-mer, soit 1,3 TWh en 2015.

(2) Production nette du pompage : la consommation d'électricité nécessaire au fonctionnement des STEP s'élève à 6,8 TWh en 2015, ce qui conduit à une production hydraulique non réduite de la consommation liée au pompage de 38,9 TWh, et comprenant la production marémotrice de la Rance (530 GWh).

(3) Hors Corse et outre-mer, soit 1,7 TWh en 2015.

*une baisse de la température de **1 °C** en hiver entraîne une hausse de la consommation d'électricité en France de l'ordre de **2 400 MW** et l'amplitude de production hydraulique entre **deux** années extrêmes peut atteindre une **vingtaine** de **térawattheures***

# VISER L'ÉQUILIBRE PRODUCTION + CONSOMMATION



# LA DOAAT : L'OPTIMISEUR D'EDF RESPONSABLE D'ÉQUILIBRE

A-5  
A-1

A-1  
M-1

M-1  
S-1

J-1

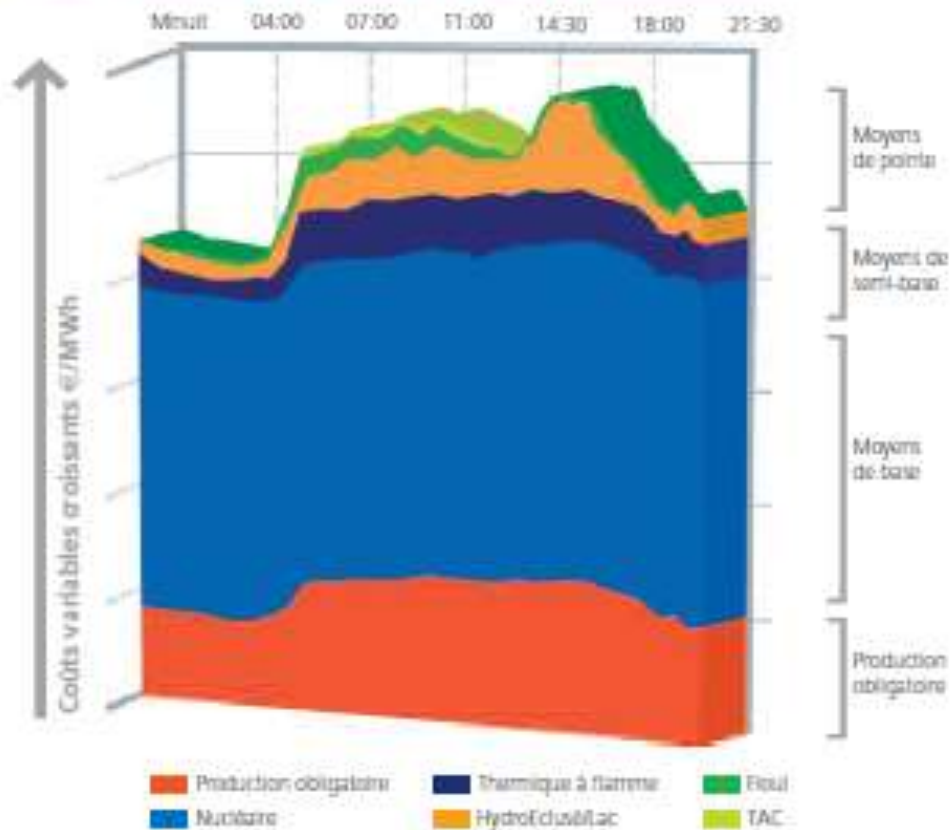
J

> J

EDF programme le fonctionnement des outils de production en les mobilisant par coûts variables croissants jusqu'à satisfaire la demande prévisionnelle de ses clients.

EDF achète sur les marchés si le prix de gros est inférieur au coût variable de son dernier moyen de production démarré pour satisfaire la demande de ses clients jusqu'à égaler les deux prix. EDF vend sur les marchés dans le cas contraire.

## Empilement des moyens de production



# OSGE : UN OPÉRATEUR DE DONNÉES

## ■ CERTIFIÉ ISO 9001 ET ORGANISÉ EN MIROIR DES MÉTIERS D'EDF

A-5  
A-1  
  
A-1  
M-1  
  
M-1  
S-1  
  
J-1  
  
IJ  
  
> J

### PRODUCTION – ENVIRONNEMENT – INTERNATIONAL

En lien avec l'amont (DPI) et la DDD (développement durable).



EXPERTISE AUTOUR DES DONNÉES DE :

- performance/disponibilité des centrales,
- impact environnemental,
- benchmark performances international

### COMMERCIALISATION – MÉTÉO & MARCHÉ – RÉFÉRENTIEL

En lien avec l'aval (DCO, RTE, ERDF, Météo France)



EXPERTISE AUTOUR DES DONNÉES :

- commerciales,
- météorologiques,
- de responsable d'équilibre

### BILAN – SYNTHÈSE – HISTORISATION

En lien avec la DOAAT.



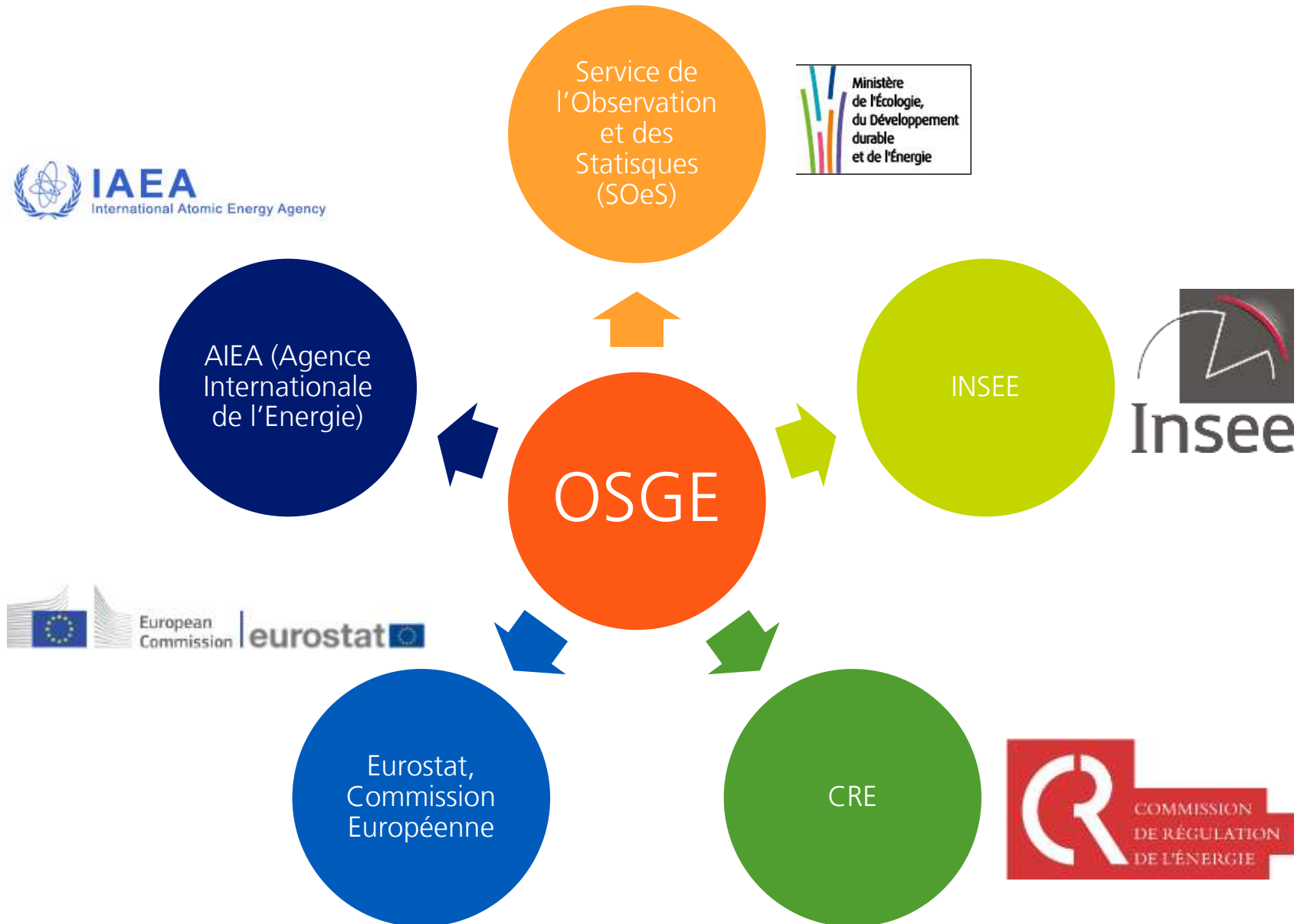
EXPERTISE AUTOUR DES DONNÉES DE :

- l'optimiseur,
- bilan électrique et construction du PMT

L'OSGE, une entité d'EDF au carrefour de la donnée, avec des problèmes de récupération/nettoyage/qualité de données variées, d'intégration de données et de publication de données élaborées à l'interne/externe.



# REPRÉSENTANT EDF À L'EXTERNE





# VALORISER LE PATRIMOINE DE DONNÉES DE L'ENTREPRISE

ILLUSTRÉS AU TRAVERS DE QUELQUES PRODUITS...

Analyse de l'écart RE

Indicateurs CLT  
(centrales à participation)

Mix fourniture EDF  
(facture client)

Kd, Ki  
(parc nucléaire)

Correction climatique

Productible H2O

Données de marché  
(Platts, Argus, ...)

Indicateurs éco  
(parc ThF)




Gestion des référentiels (RE,  
RP, AA,...)

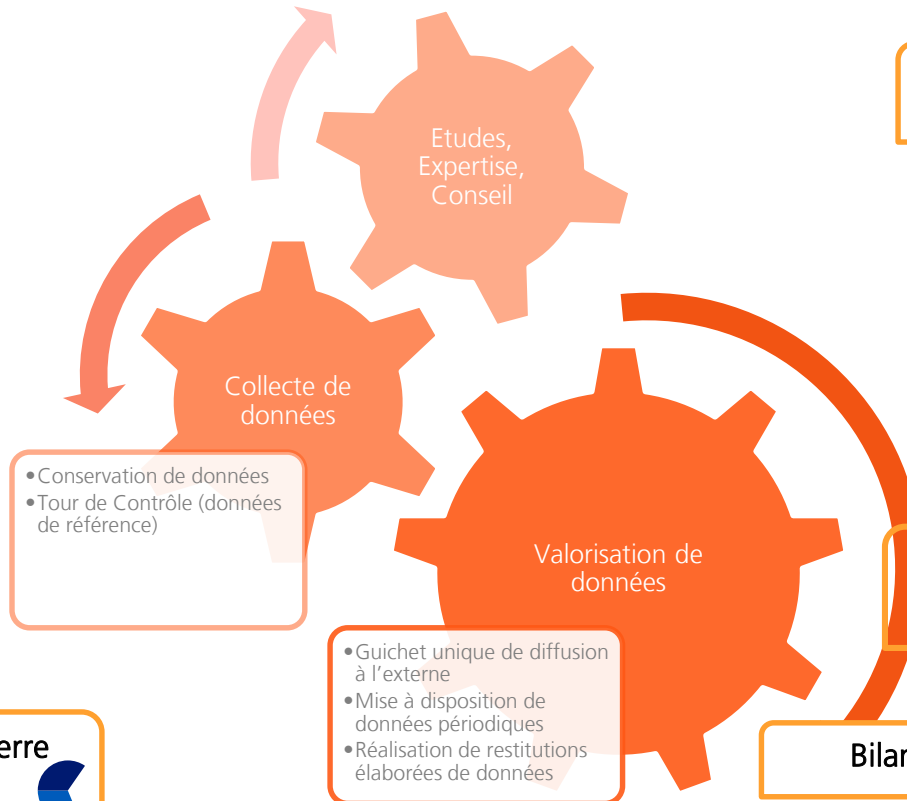
Bilan Gaz à Effet de Serre

Bilan Electrique C+P

GT performance production  
(WANO-AIEA)

Légende :

-  usages hors EDF SA
-  usages EDF SA
-  usages DOAAT



A-5  
A-1

A-1  
M-1

M-1  
S-1

J-1

IJ

> J

# DES DONNÉES QUI AUGMENTENT EN VOLUME ET DIVERSITÉ

- DES DONNÉES ADMINISTRÉES, UN PATRIMOINE À VALEUR AJOUTÉE
- UN PORTEFEUILLE DYNAMIQUE SUR L'AVAL (CONSOMMATION)
- UN PORTEFEUILLE DIVERSIFIÉ SUR L'AMONT (PRODUCTION)
- 48 DEMI-HEURE PAR JOUR... 17520 DONNÉES EMPILÉES PAR AN

## SOMMAIRE

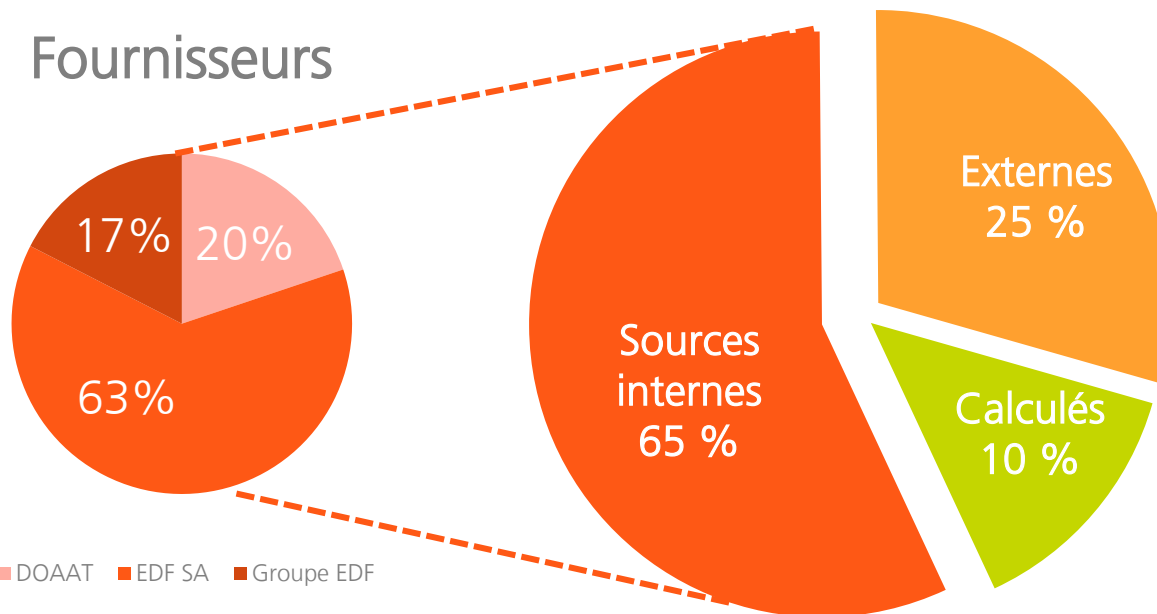
- ▶ CONTEXTE D'UN OBSERVATOIRE STATISTIQUE
- ▶ DES DONNÉES QUI AUGMENTENT EN VOLUME ET DIVERSITÉ
- ▶ UN PROCESSUS GÉNÉRIQUE DE PRODUCTION
- ▶ UN « DICTIONNAIRE » POUR PARTAGER
- ▶ UN PATRIMOINE À VALORISER PAR LA DIFFUSION ET LA RÉUTILISATION

# DES DONNÉES ADMINISTRÉES, UN PATRIMOINE À VALEUR AJOUTÉE

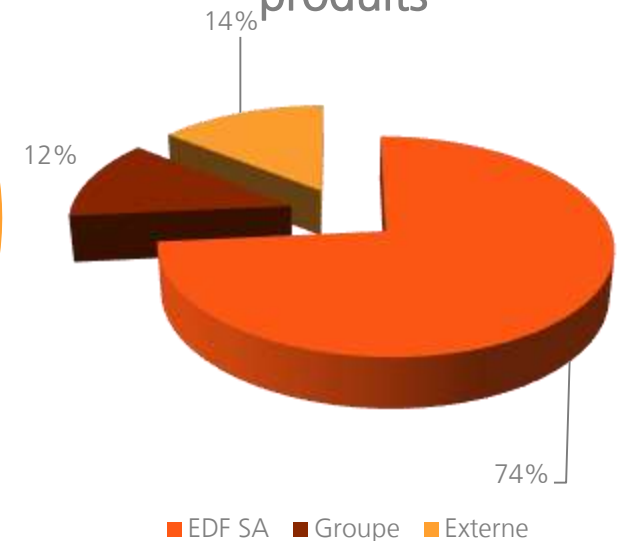
■ ~15 millions d'enregistrements unitaires

220 jeux de données administrées+  
(depuis 1948 pour les plus anciens)

## Fournisseurs



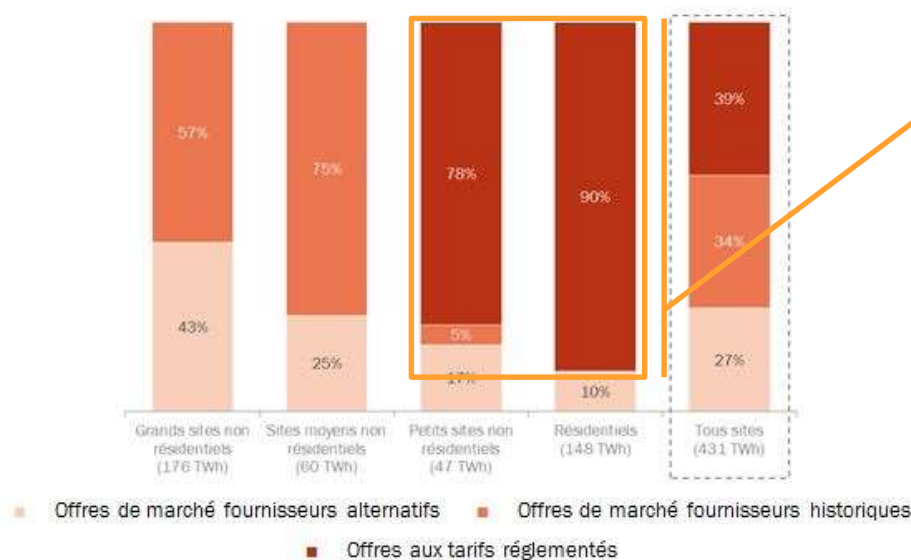
## Clients des données et produits



+ 288 jeux de données historisés pour audit/litige  
+ environ 180 produits construits

# UN PORTEFEUILLE DYNAMIQUE SUR L'AVAL (CONSOMMATION)

Répartition des consommations par type d'offre au 30 juin 2016



Sources : GRD, RTE, Fournisseurs historiques – Analyse : CRE

- ~170 TWh (bas de portefeuille) en relève manuelle
  - Prévion de consommation réalisée par prévision statistique
- Pour les particuliers, forte dépendance à la météo et « caractéristiques » du logement
- Pour les « pros », dépendance au processus métier et écosystème (conjoncture économique, etc.)

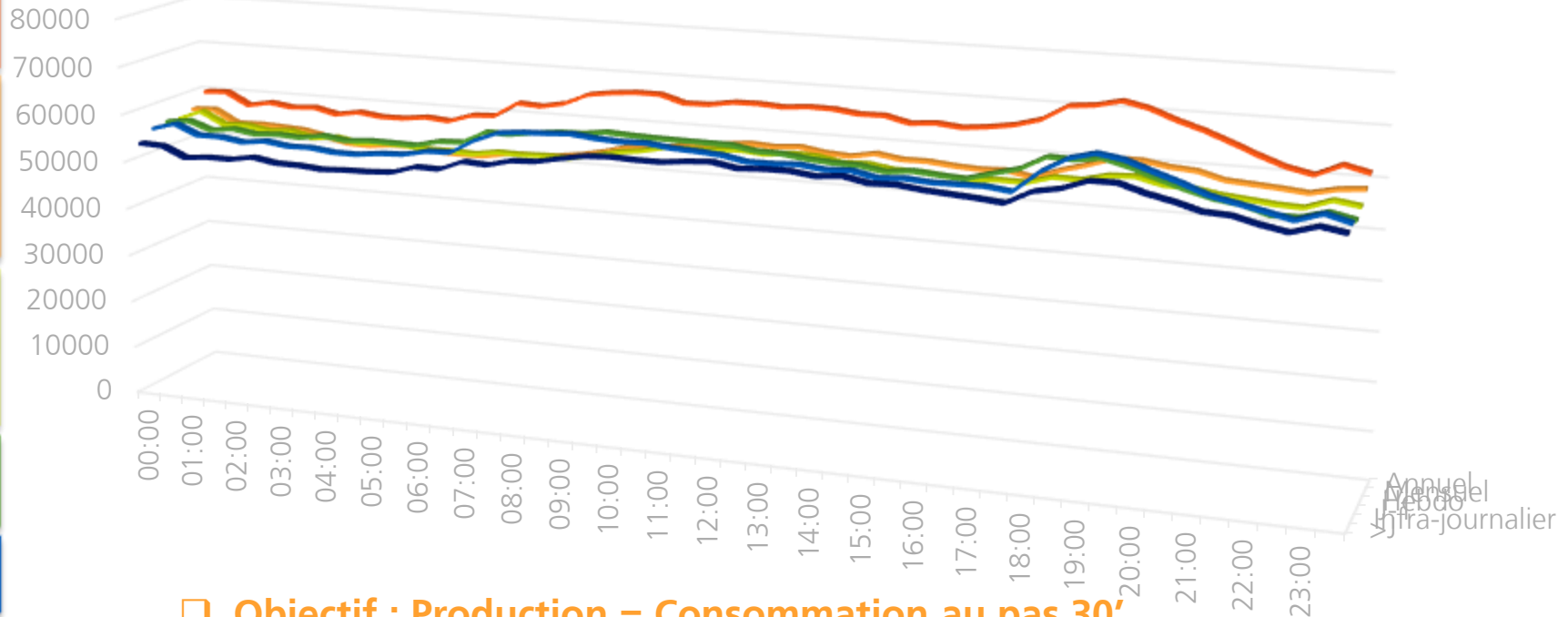
# UN PORTEFEUILLE DIVERSIFIÉ SUR L'AMONT (PRODUCTION)



- Le parc en France continentale se compose au 31 décembre 2015 de :
  - 58 tranches nucléaires, de puissances électriques variables allant de 900 MW à 1 500 MW, sont réparties sur 19 sites
  - 27 tranches thermiques en fonctionnement,
  - 433 centrales hydrauliques :
    - ⇒ environ 11 % des centrales ont une puissance unitaire supérieure à 100 MW ; elles représentent environ 60 % de la production totale
    - ⇒ environ 51 % des centrales ont une puissance unitaire inférieure à 12 MW ; elles représentent environ 6 % de la production totale.

# 48 DEMI-HEURE PAR JOUR... 17520 DONNÉES EMPILÉES PAR AN

Optimiser la production d'EDF en l'adaptant à la consommation des clients en portefeuille



- ❑ Objectif : Production = Consommation au pas 30'
- ❑ Empilement des prévision de consommation par segment de client
- ❑ Empilement des programmations de productions
  - Prise en compte des planning d'indisponibilités
  - Aléas climatiques
  - Aléas prix de marché

# UN PROCESSUS GÉNÉRIQUE DE PRODUCTION

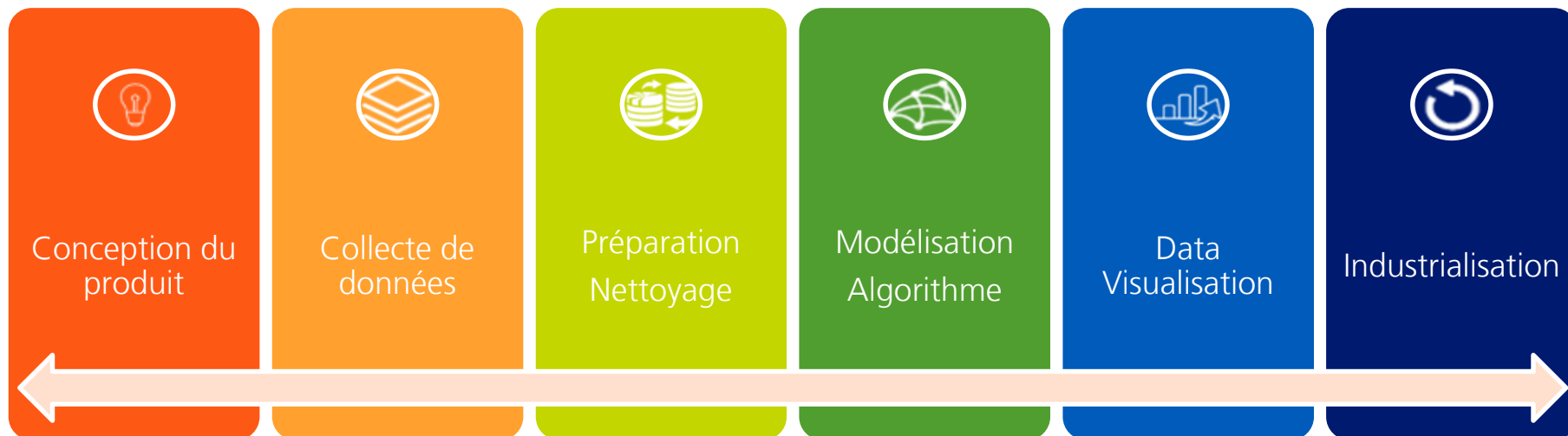
- MANIPULER LES DONNÉES AU PLUS TÔT DU PROCESSUS
- CONFIDENTIALITÉ DES LIVRABLES
- AXE DE DONNÉES MANIPULÉES

## SOMMAIRE

- ▶ CONTEXTE D'UN OBSERVATOIRE STATISTIQUE
- ▶ DES DONNÉES QUI AUGMENTENT EN VOLUME ET DIVERSITÉ
- ▶ **UN PROCESSUS GÉNÉRIQUE DE PRODUCTION**
- ▶ UN « DICTIONNAIRE » POUR PARTAGER
- ▶ UN PATRIMOINE À VALORISER PAR LA DIFFUSION ET LA RÉUTILISATION

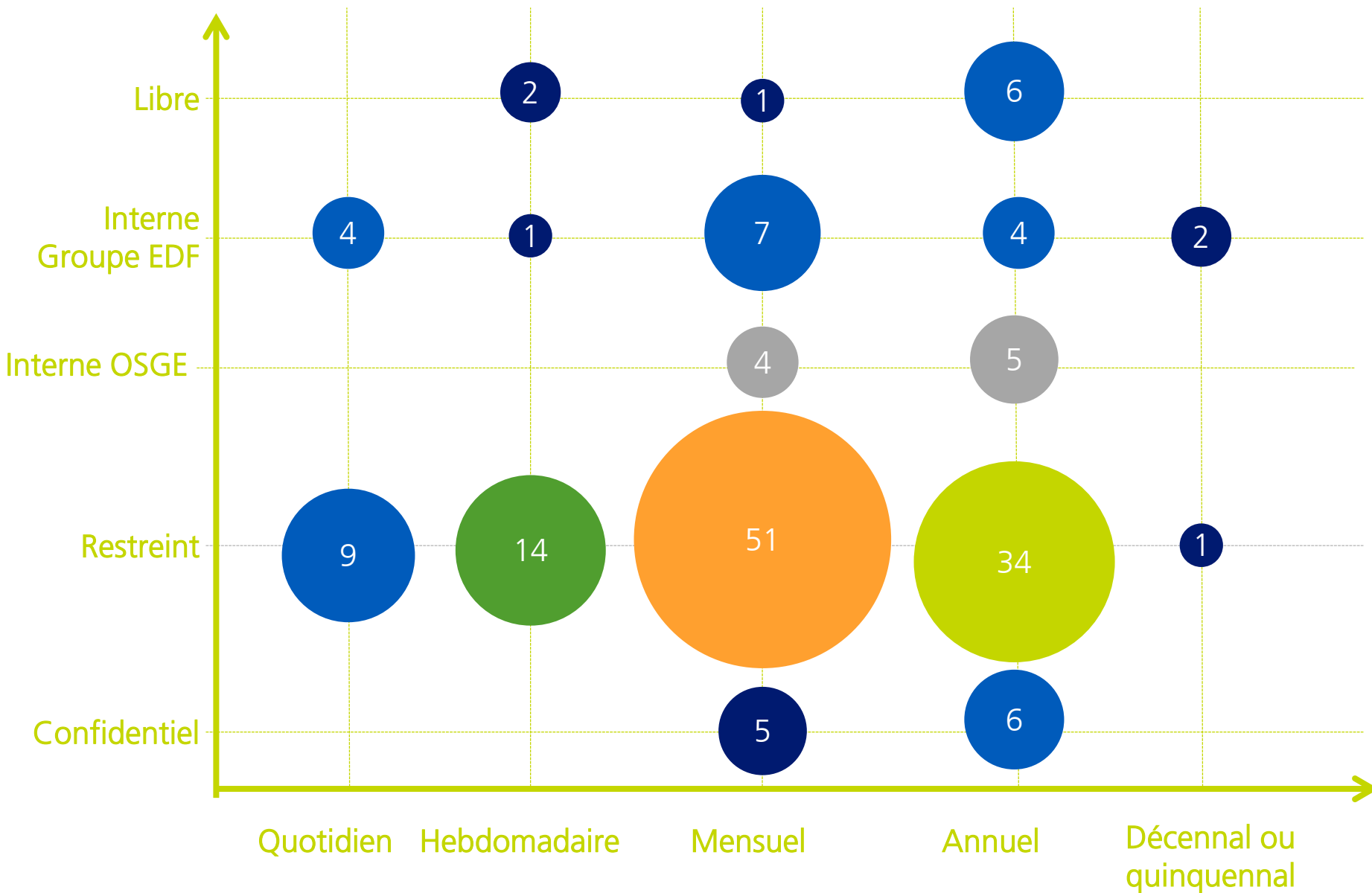


# MANIPULER LES DONNÉES AU PLUS TÔT DU PROCESSUS

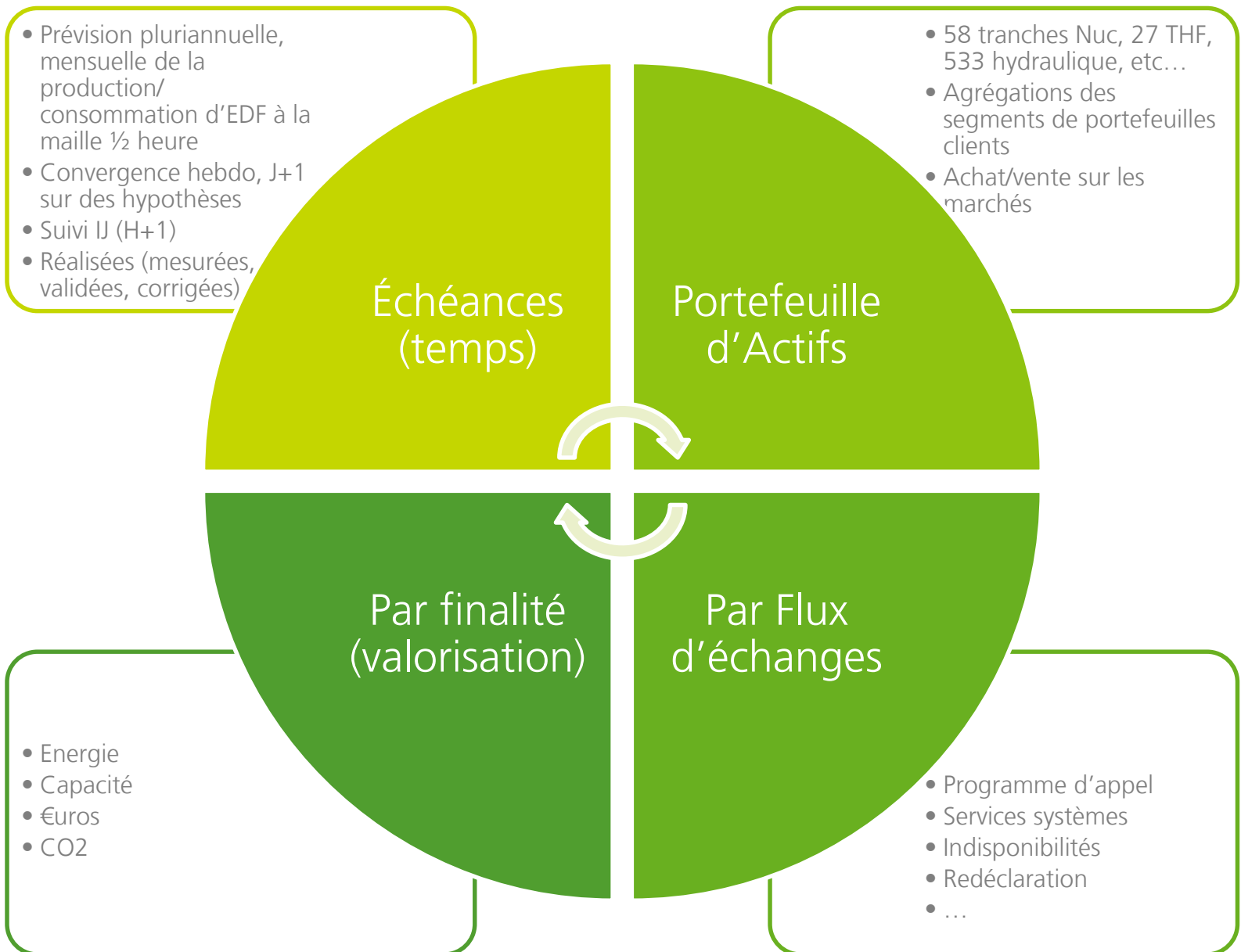


- Travaux avec le client sur son besoin
- Identification des données nécessaires, des sources, des « droits », des « protocoles » d'acquisition
- Préparation, Nettoyage, manipulation
- Modélisation et algorithme pour optimiser la production d'un produit récurrent ou d'une étude
- Data Visualisation pour favoriser la diffusion
- Industrialisation dans les cas où les utilisateurs sont demandeur d'une alimentation récurrente

# CONFIDENTIALITÉ DES LIVRABLES



# AXE DE DONNÉES MANIPULÉES



- QUELQUES DÉFINITIONS
- MODALITÉS D'ACQUISITION VERSUS STOCKAGE ET REPRÉSENTATION
- IMPORTANCE DES MÉTADONNÉES

## SOMMAIRE

- ▶ CONTEXTE D'UN OBSERVATOIRE STATISTIQUE
- ▶ DES DONNÉES QUI AUGMENTENT EN VOLUME ET DIVERSITÉ
- ▶ UN PROCESSUS GÉNÉRIQUE DE PRODUCTION
- ▶ **UN « DICTIONNAIRE » POUR PARTAGER**
- ▶ UN PATRIMOINE À VALORISER PAR LA DIFFUSION ET LA RÉUTILISATION

# QUELQUES DÉFINITIONS

---

## ■ Nécessité d'un dictionnaire pour

- Garantir une continuité des données dans la durée (à des fins statistiques)
- Lever les doutes avec un partenaire/client
- Tracer le « cycle de vie » de la donnée (dans le cadre d'un processus d'élaboration et/ou d'un processus de qualification)

## ■ Donnée

- « Dans les technologies de l'information (TI), une donnée est une description élémentaire, souvent codée, d'une chose, d'une transaction d'affaire, d'un événement. »
- Les données sont exclusives par leur nature et non leurs usages.
  - ⇒ 3 natures de donnée : opérationnelle, pilotage, référence

## ■ Catalogue

- « Inventaire des données exploitables, et leurs attributs, que l'OSGE reçoit, manipule et transforme. »

## ■ Gouvernance

- « La gouvernance de Données pose le cadre décisionnel pour tout ce qui touche aux données.
- C'est un ensemble de procédures, sous la responsabilité d'acteurs, s'appliquant aux données utiles à l'entreprise et s'appuyant sur un outillage technologique

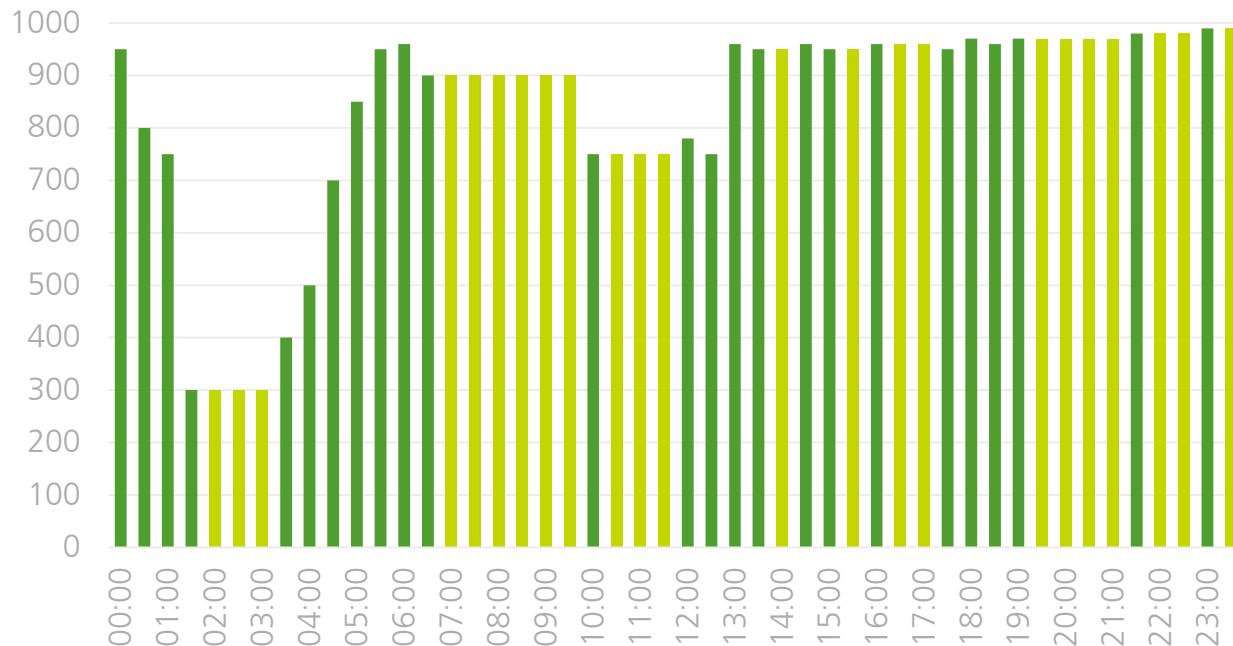
# MODALITÉ D'ACQUISITION VERSUS STOCKAGE / REPRÉSENTATION

## ■ Donnée

- instantanée représentant l'état (ou index) d'un « capteur »
- de « consigne » d'un équipement (hausse/baisse de puissance à HH:MM)

## ■ Valeur transmise selon

- Une fréquence régulière
  - ⇒ selon le cas, peut-il exister une valeur intermédiaire « différente » ?
- une fréquence irrégulière
  - ⇒ cas d'une relève manuelle
- Un « événement »
  - ⇒ par exemple : une modification (supposition que la valeur précédente était « stable » et connue)



Liste des 48  
valeurs par  
échéance  
ou  
Liste des 25  
événements

# IMPORTANCE DES MÉTADONNÉES



- Identifier les nomenclatures, normes, standard favorisant la description
  - ISO, Web Semantique (RDF, SKOS, OWL), ...
- Exemple : lang:fr, heure au standard ISO:8601, Unités SI ISO/IEC 80000, etc...



# UN PATRIMOINE À VALORISER PAR LA DIFFUSION ET LA RÉUTILISATION

- CHOISIR UN STANDARD
- CLASSIFICATION DES JEUX DE DONNÉES AU SEIN D'UN CATALOGUE
- EXEMPLE DE MACRO-DONNÉES
- UN POTENTIEL À DÉVELOPPER

## SOMMAIRE

- ▶ CONTEXTE D'UN OBSERVATOIRE STATISTIQUE
- ▶ DES DONNÉES QUI AUGMENTENT EN VOLUME ET DIVERSITÉ
- ▶ UN PROCESSUS GÉNÉRIQUE DE PRODUCTION
- ▶ UN « DICTIONNAIRE » POUR PARTAGER
- ▶ UN PATRIMOINE À VALORISER PAR LA DIFFUSION ET LA RÉUTILISATION

# CHOISIR UN « STANDARD »

- Existence d'une « sémantique » liée aux activités de responsable d'équilibre

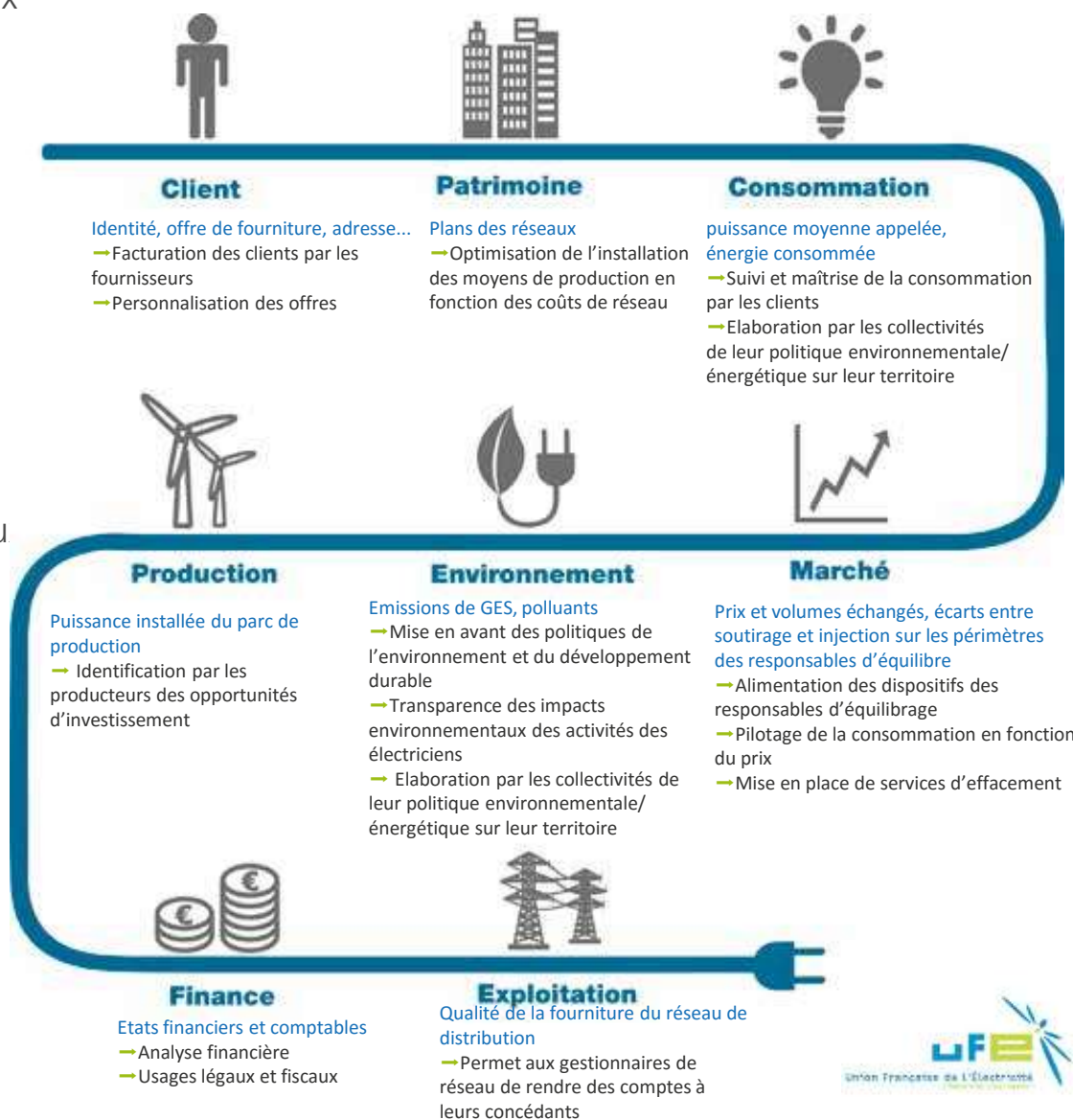
⇒ Modèle Européen des échanges liés aux marchés électriques (Common Information Model (CIM) for Energy Markets – ENTSOE)

- Dans le contexte « opendata » et des évolutions des marchés énergétiques en cours, l'UFE (Union Française de l'Electricité) a publié

□ Une **cartographie** des **usages** liés aux données énergétiques en France

- Le raisonnement usage favorise une compréhension « client » (des données)

□ Adoption de ces thématiques



# CLASSIFICATION DES JEUX DE DONNÉES AU SEIN D'UN CATALOGUE

■ Cataloguer pour faciliter l'accès aux « nouveaux » utilisateurs

■ Préparer la bascule vers des outils « hadoop »

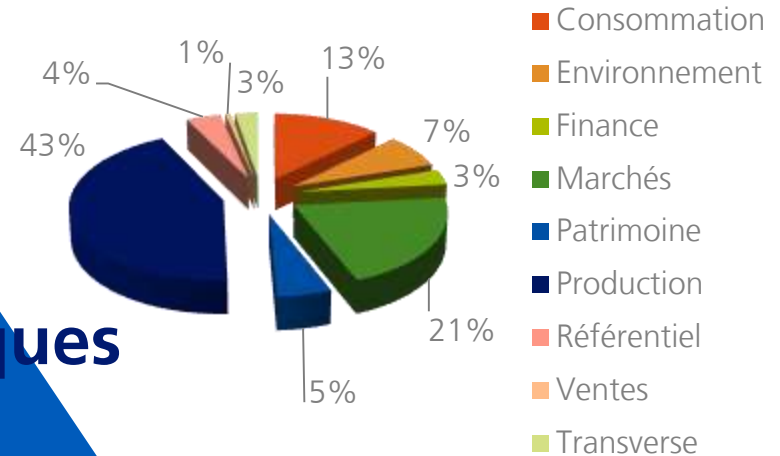
□ Moins de modélisation « globale »

□ Plus d'index de recherche

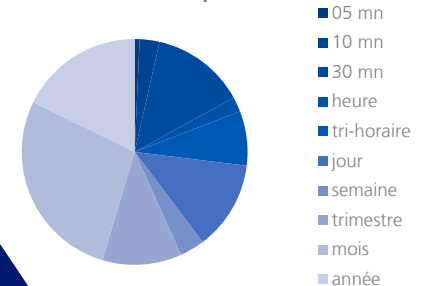
**8  
thématiques  
(UFE)**

**64 macros-  
données**

**450 données**



Répartition par pas de temps



# EXEMPLE DE MACRO-DONNÉES



## Référentiels

- Contrat d'achat
- Contreparties
- Éléments de preuve
- Éléments déclaratifs
- Entités de production
- GRD
- Périmètres
- Points de raccordement
- Stations météo



## Marchés

- Achat / vente d'énergie
- Ajustements
- Données publiques du mécanisme d'ajustement
- Modélisation DTG
- Prix de marché
- Services système
- Synthèses
- Veille marché Energie



## Production

- Courbes de charges
- Production d'énergie



## Exploitation

- Correctifs
- Exploitation centrales
- Exploitation réservoirs
- Hypothèses compensation
- Indisponibilité
- NEB équilibre
- Programme d'appel



## Patrimoine

- Bâtiments tertiaires
- Flotte de véhicules



## Environnement

- Autres rejets dans l'eau
- Biodiversité
- Consommables
- Consommation Bâtiments tertiaires
- Consommation Flotte de véhicules
- Déchets conventionnels
- Déchets radioactifs
- Eau
- Emissions atmosphériques
- Emissions liés aux déplacements
- Facteur d'émission
- Rejets radio-actifs
- Sous-produits issus de la combustion



## Finances

- Achat de biens et de services
- Immobilisations
- PMT



## Consommation

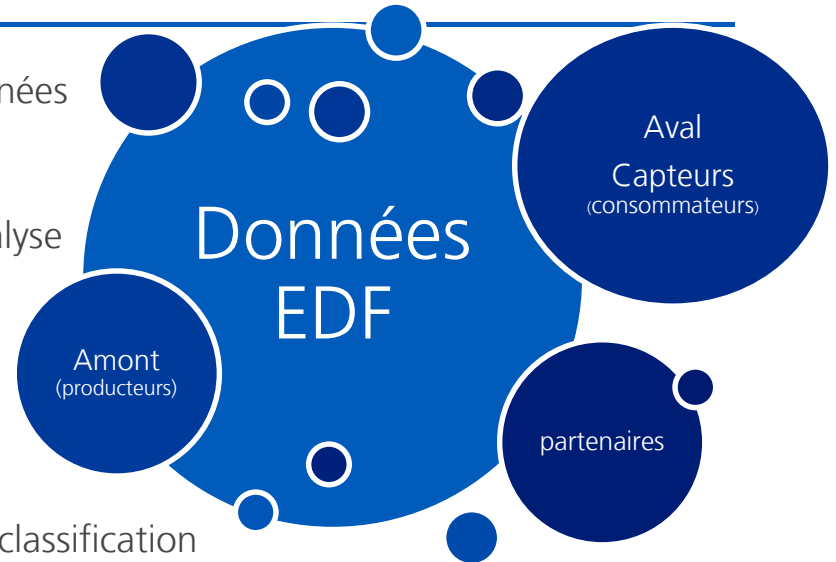
- Comptage RPD - BGC
- Comptage RPT
- Consommation réalisée
- Correction climatique
- Nébulosité
- Prédiction de consommation
- Température
- Vent



## Clients

# UN POTENTIEL À DÉVELOPPER

- Favoriser l'exploration par le croisement possible de données
  - Normalisation des métadonnées (standards RDF)
- Tirer parti des **nouvelles technologies** de stockage et analyse
  - Hadoop, spark
  - Python, R
  - API json
  - ...
- Utiliser des **automates** pour assister la qualification et la classification
  - Analyse sémantique de textes issu des journal de bord d'exploitation et comparaison avec des données structurées remontées par ailleurs
- S'appuyer sur **des infrastructures partagées** avec nos partenaires pour faciliter les **croisements**
- Travailler en **open-innovation** maîtrisée



# MERCI !



Contact

[stephane.ternot@edf.fr](mailto:stephane.ternot@edf.fr)

Sources :

[www.edf.fr](http://www.edf.fr)

[www.rte-france.com](http://www.rte-france.com)

[www.asn.fr](http://www.asn.fr)

[ufe-electricite.fr](http://ufe-electricite.fr)

[www.entsoe.eu](http://www.entsoe.eu)

[www.cre.fr](http://www.cre.fr)

