



STATISTIQUES OPTIMISATION DATA

DOAAT – SoDATA – Une entité au carrefour de la donnée EDF
Intervention au Mastère Data Sciences – module IOT - S.TERNOT

07/01/2019



SOMMAIRE

1.

Contexte d'un
Opérateur de
Données

2.

Des données qui
augmentent en
volume et diversité

3.

Un Processus
générique de
production

4.

Un « dictionnaire »
pour partager

5.

Un patrimoine à
valoriser par la
diffusion et la
réutilisation



1.

1. Contexte d'un Opérateur de Données

- Quelques données d'EDF SA
- Viser l'équilibre production + Consommation
- La DOAAT : l'optimiseur d'EDF « Responsable d'équilibre »
- SoDATA : un Opérateur de Données
- Représentant EDF à l'externe
- Valoriser le patrimoine de données de l'entreprise

2. Des données qui augmentent en volume et diversité

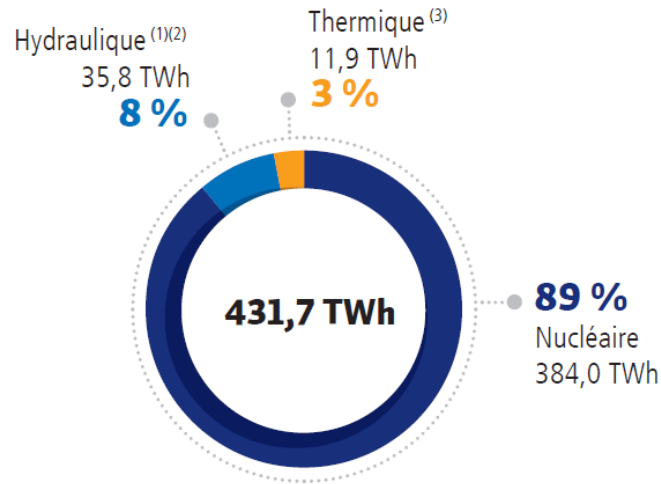
3. Un Processus générique de production

4. Un « dictionnaire » pour partager

5. Un patrimoine à valoriser par la diffusion et la réutilisation

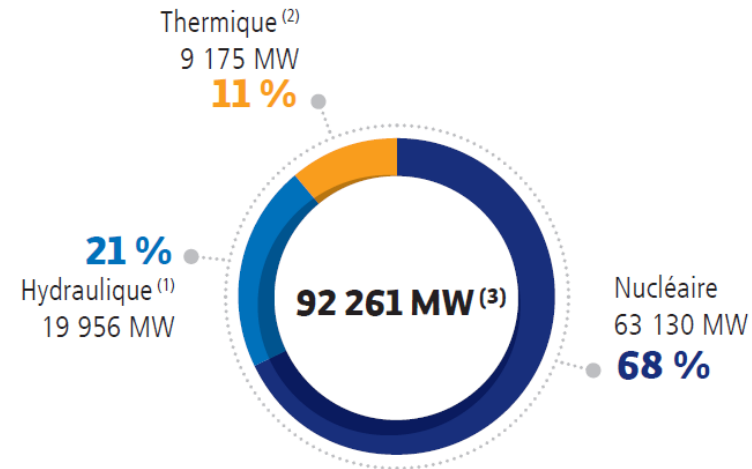
QUELQUES DONNÉES D'EDF SA (2016)

Production 2016 France continentale



(1) Hors Corse et outre-mer, soit 1,3 TWh en 2016.
(2) Production nette du pompage : la consommation d'électricité nécessaire au fonctionnement des STEP s'élève à 6,6 TWh en 2016, ce qui conduit à une production hydraulique non réduite de la consommation liée au pompage de 42,4 TWh, et comprenant la production marémotrice de la Rance, soit 0,5 TWh.
(3) Hors Corse et outre-mer, soit 4,5 TWh en 2016.

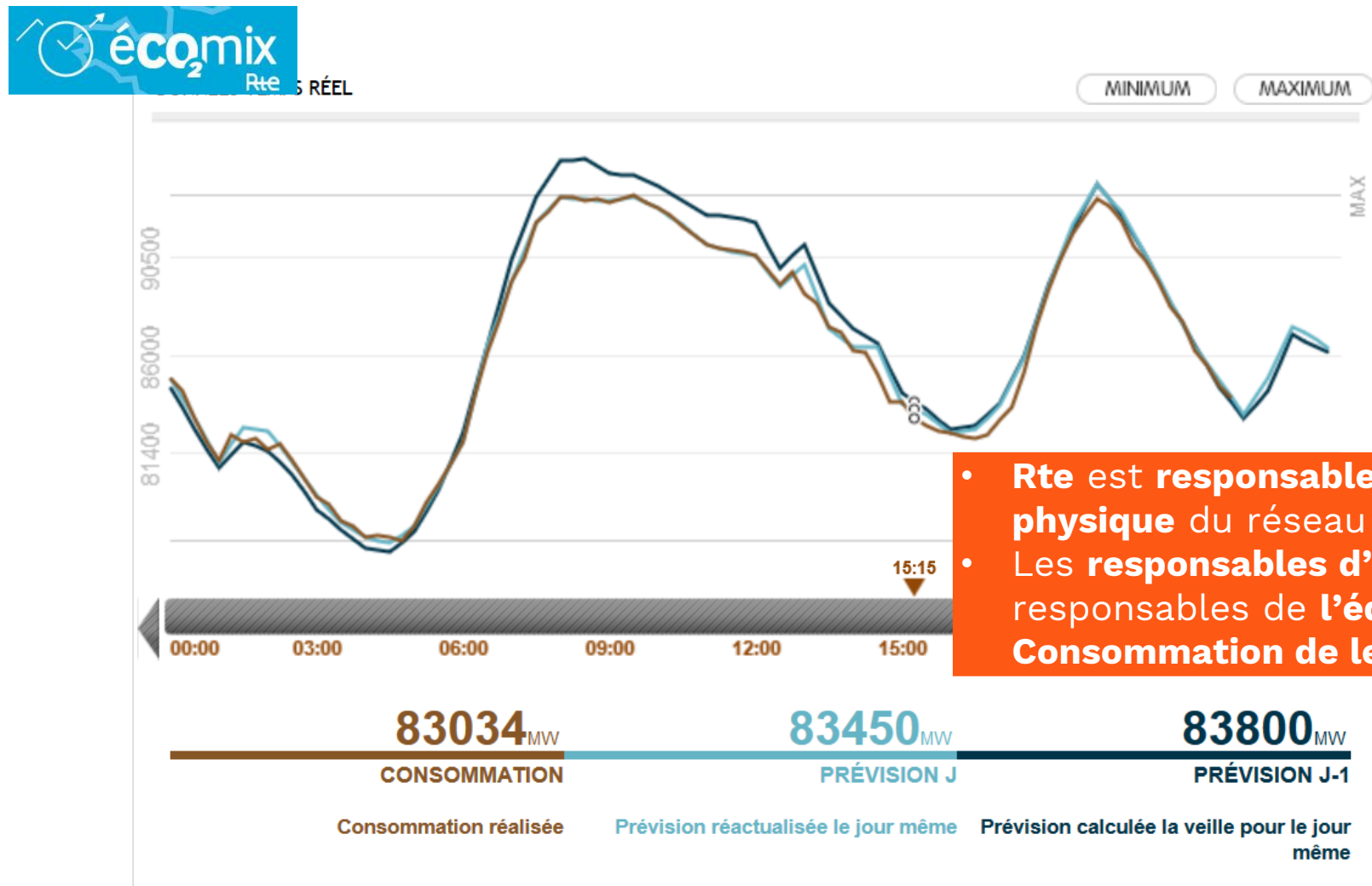
Capacité installée 2016 France continentale



NB : exprimé en mégawatts de puissance maximale couplée au réseau.
(1) Hors Corse et outre-mer, soit 440 MW en 2016.
(2) Hors Corse et outre-mer, soit 1623 MW en 2016.
(3) Hors capacités éoliennes de 12 MW et y compris capacités marémotrices de 240 MW.

*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



- Rte est responsable de l'équilibre physique du réseau
- Les responsables d'équilibre sont responsables de l'équilibre Production / Consommation de leurs actifs au pas 30'

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

A-5
A-1

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

A-1
M-1

M-1
S-1

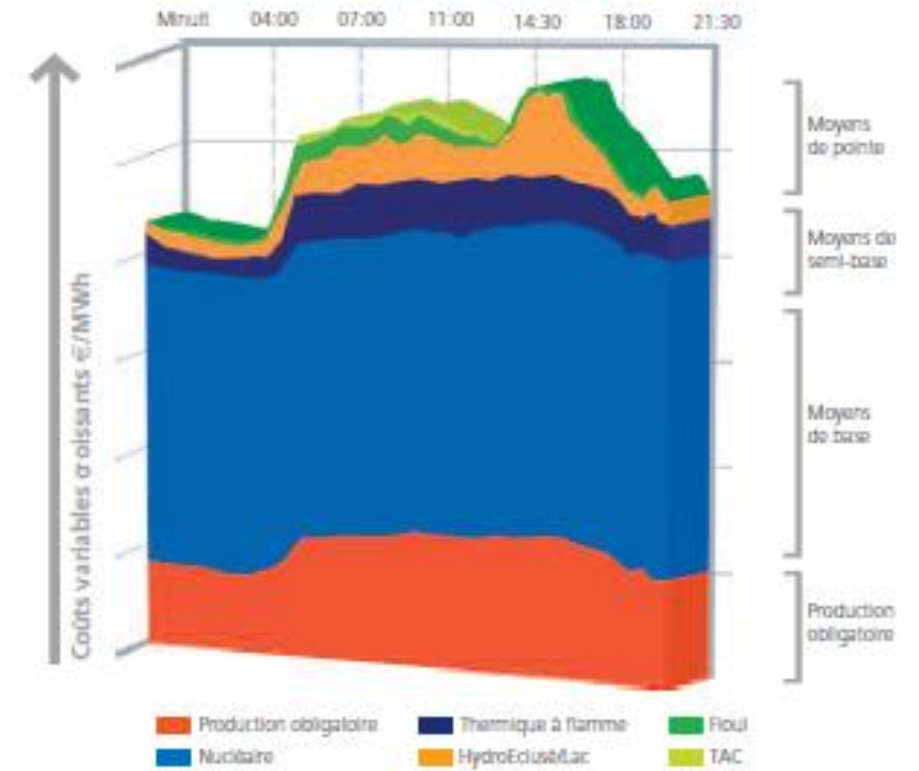
EDF achète sur le marché si le prix de gros est inférieur au coût variable de son dernier moyen de production démarré pour satisfaire la demande de ses clients jusqu'à égaler les deux prix. EDF vend sur les marchés dans le cas contraire

J-1

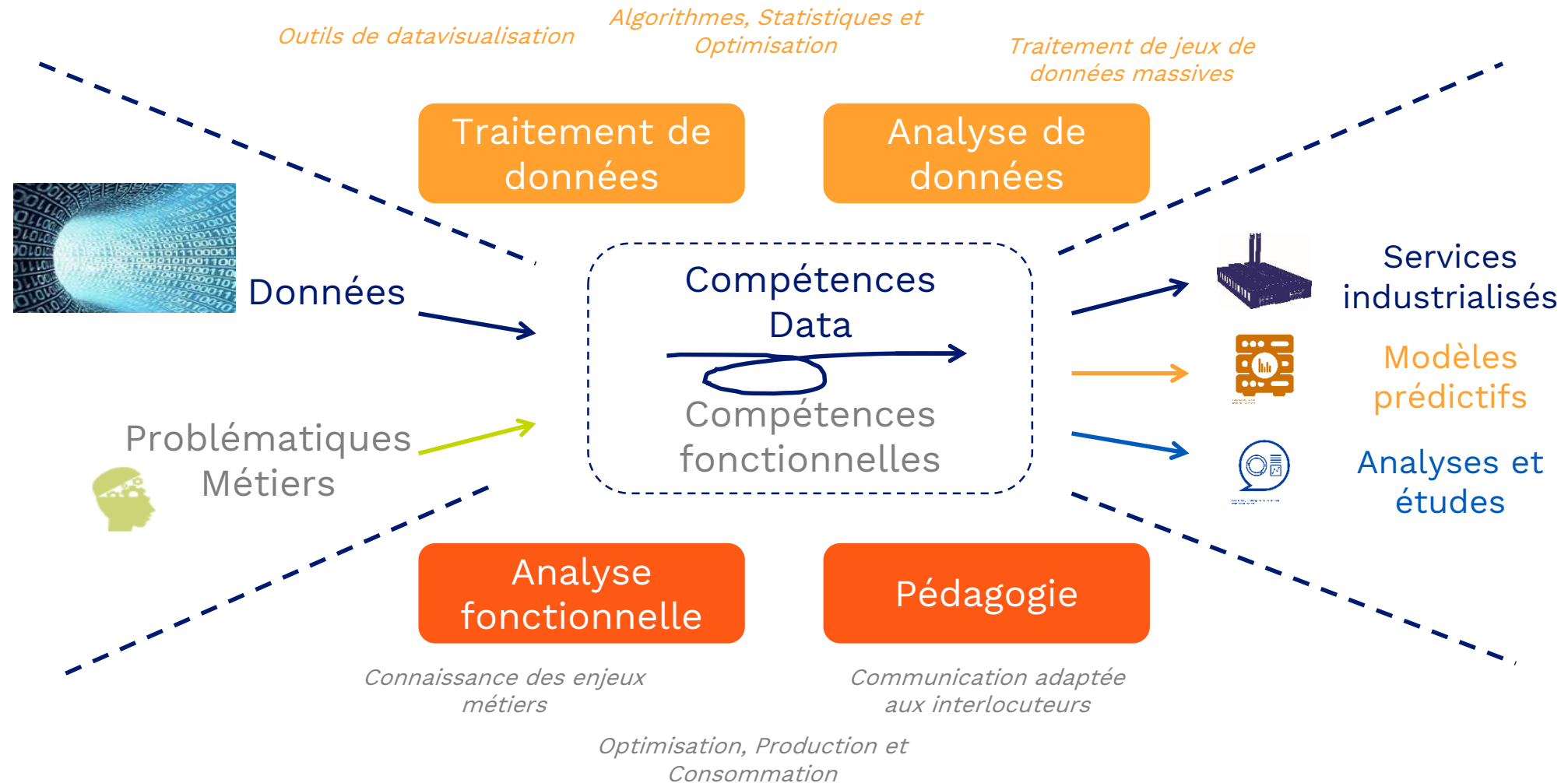
IJ

> J

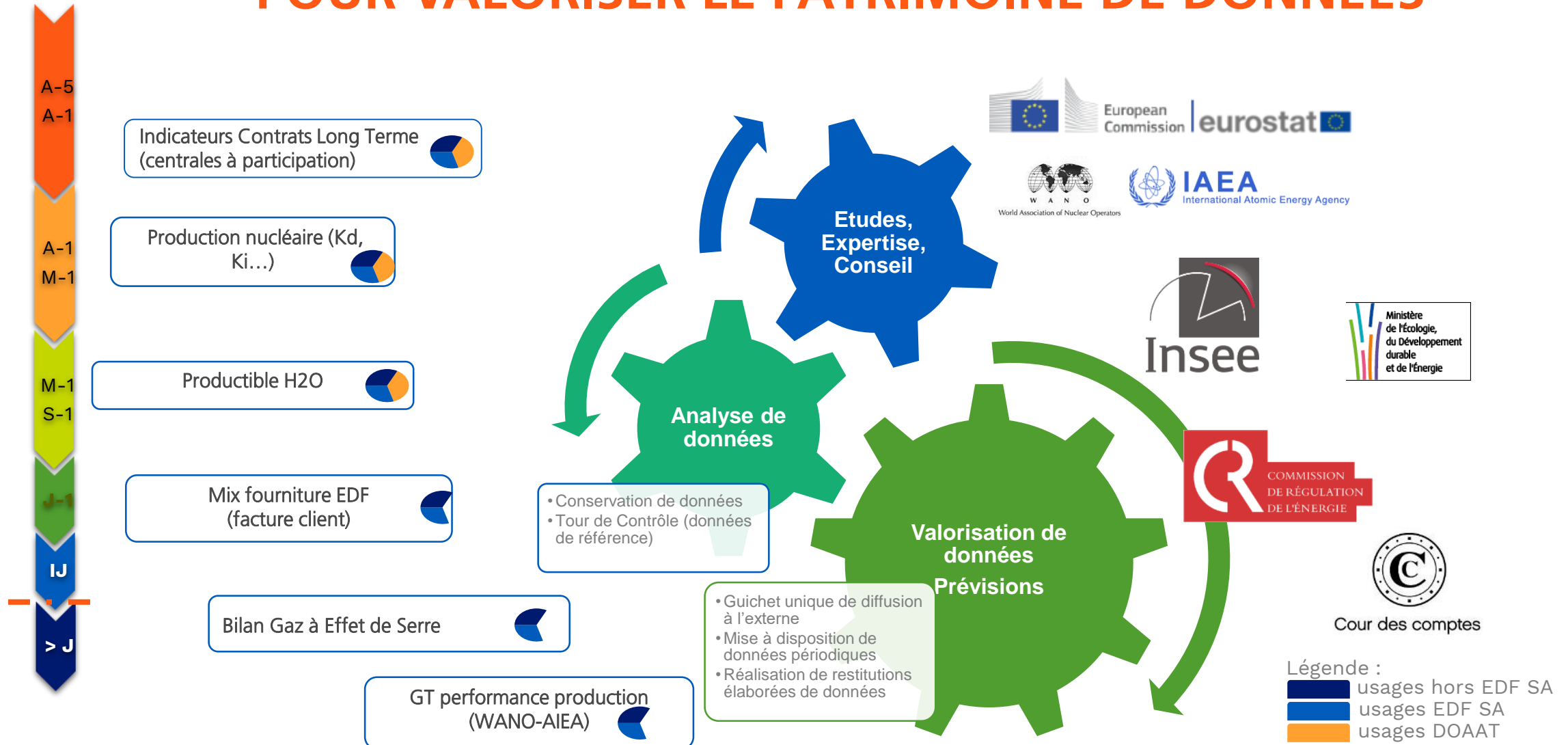
Empilement des moyens de production



SODATA : UN OPÉRATEUR DE DONNÉES



REPRÉSENTANT EDF À L'EXTERNE POUR VALORISER LE PATRIMOINE DE DONNÉES



UNE ENTITÉ AU CŒUR DE LA DONNÉE

Amont : Prod. nuc, THF, Hydraulique, EnR



Expertise autour des données de :

- **performance / disponibilité des centrales,**
- **benchmark performances international**

En lien avec l'amont (DPNT, EDF Hydro, EDF SEI)

Optimisation – EOD – Environnement



Expertise autour des données de :

- **l'optimisation**
- **bilan électrique et construction du PMT**
- **impact environnemental**

En lien avec la DOAAT, la DDD (dév. durable)

Aval : Commercialisation, Météo, Marché



Expertise autour des données et modèles :

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

En lien avec l'aval (Commerce, Sowe, RTE, Enedis, Météo France)

SoDATA, une entité d'EDF au carrefour de la donnée, avec des missions de **collecte**, **nettoyage**, **qualification** de données variées, **d'intégration** de données et de **publication** de **données élaborées** à l'interne/externe.



2.

1. Contexte d'un Opérateur de Données

2. 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
- Plus de capteurs... quels correctifs à prendre en compte ?

3. Un Processus générique de production

4. Un « dictionnaire » pour partager

5. Un patrimoine à valoriser par la diffusion et la réutilisation

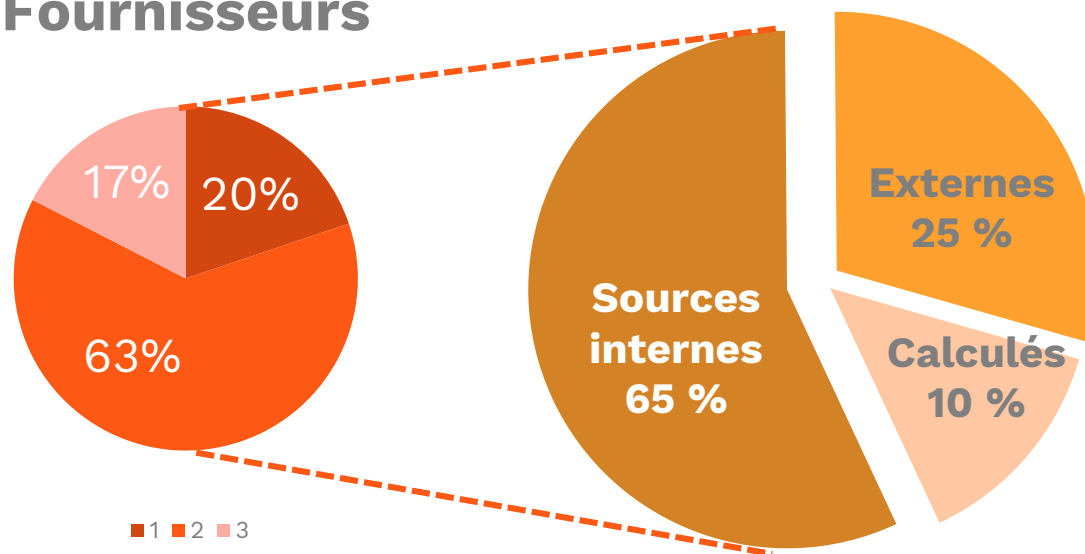
DES DONNÉES ADMINISTRÉES, UN PATRIMOINE À VALEUR AJOUTÉE

~15 millions d'enregistrements unitaires

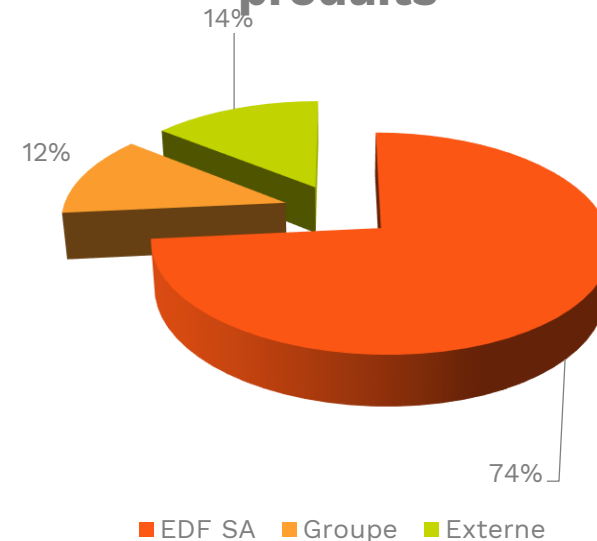
220 jeux de données administrées⁺

(depuis 1948 pour les plus anciens)

Fournisseurs



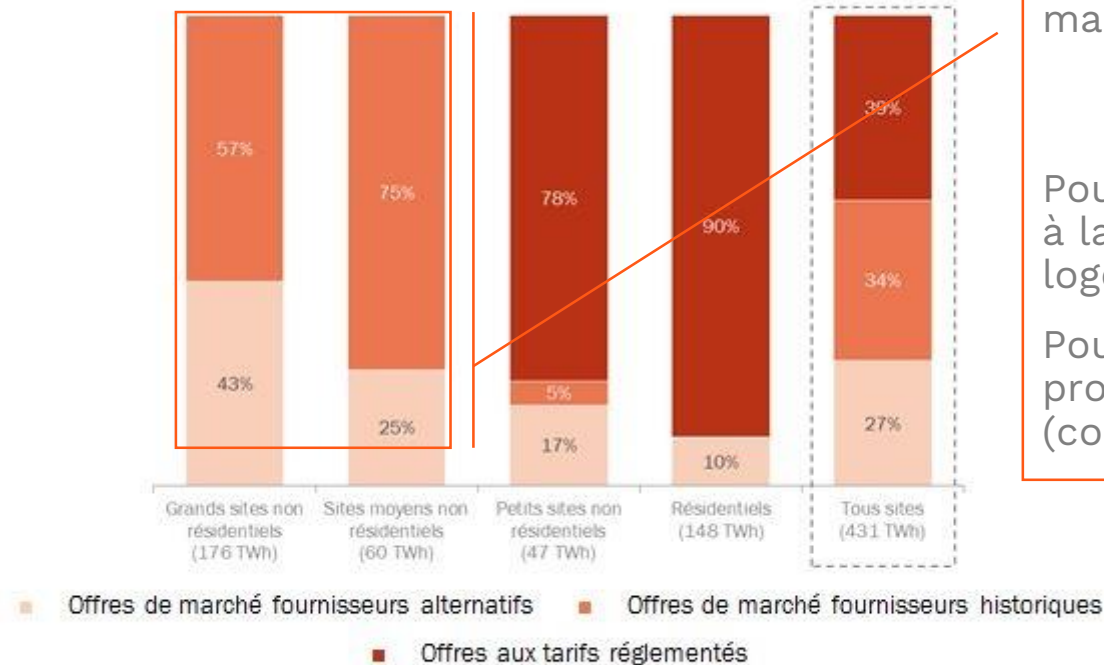
Clients des données et produits



+ 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



~170 TWh (bas de portefeuille) en relève manuelle

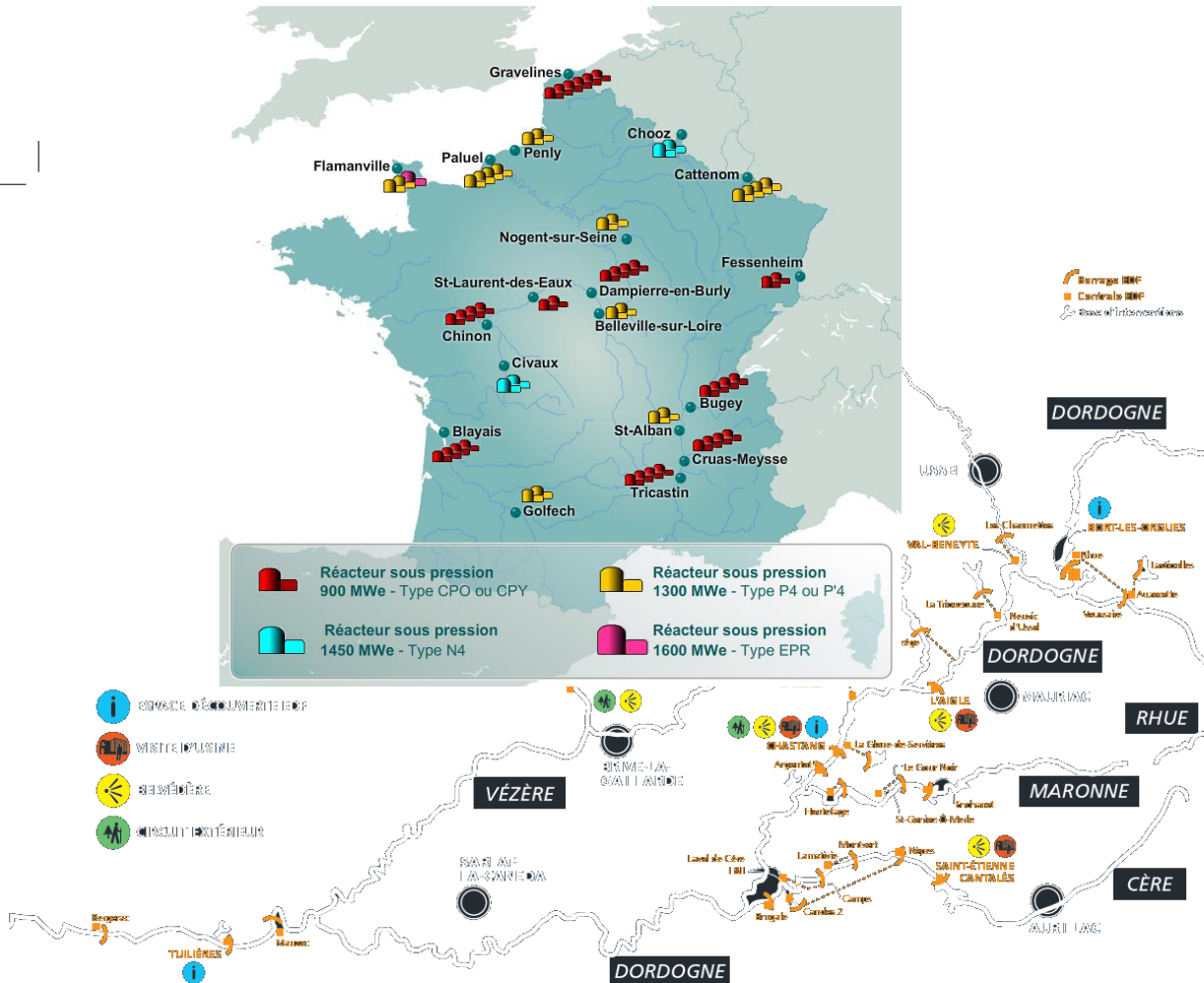
- Préviction de consommation réalisée par prévision statistique

Pour les particuliers, forte dépendance à la météo et « caractéristiques » du logement

Pour les « pros », dépendance au processus métier et écosystème (conjoncture économique, etc.)

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

UN PORTEFEUILLE DIVERSIFIÉ SUR L'AMONT (PRODUCTION)

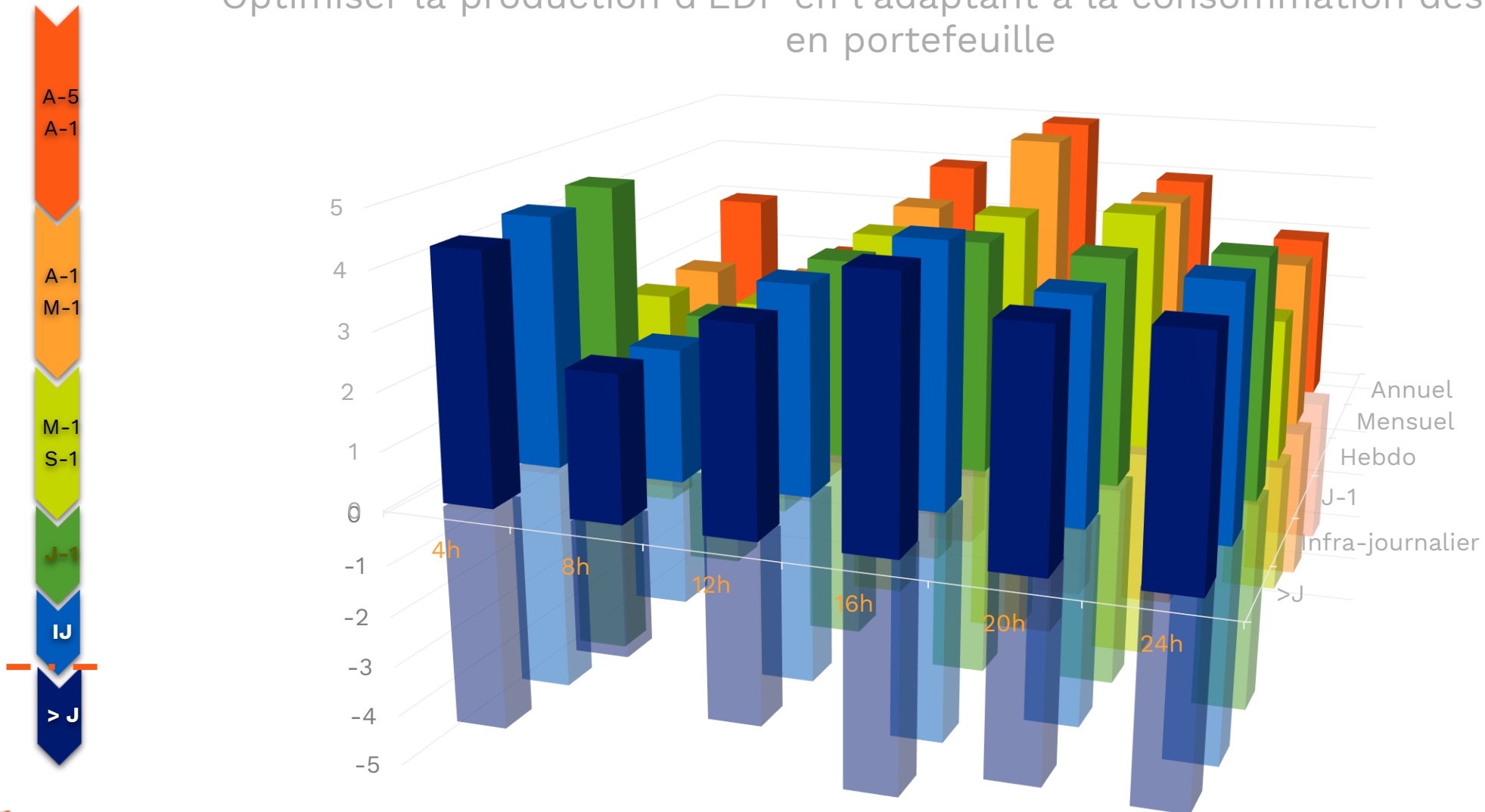


Le parc en France continentale se composait au 31 décembre 2015 de :

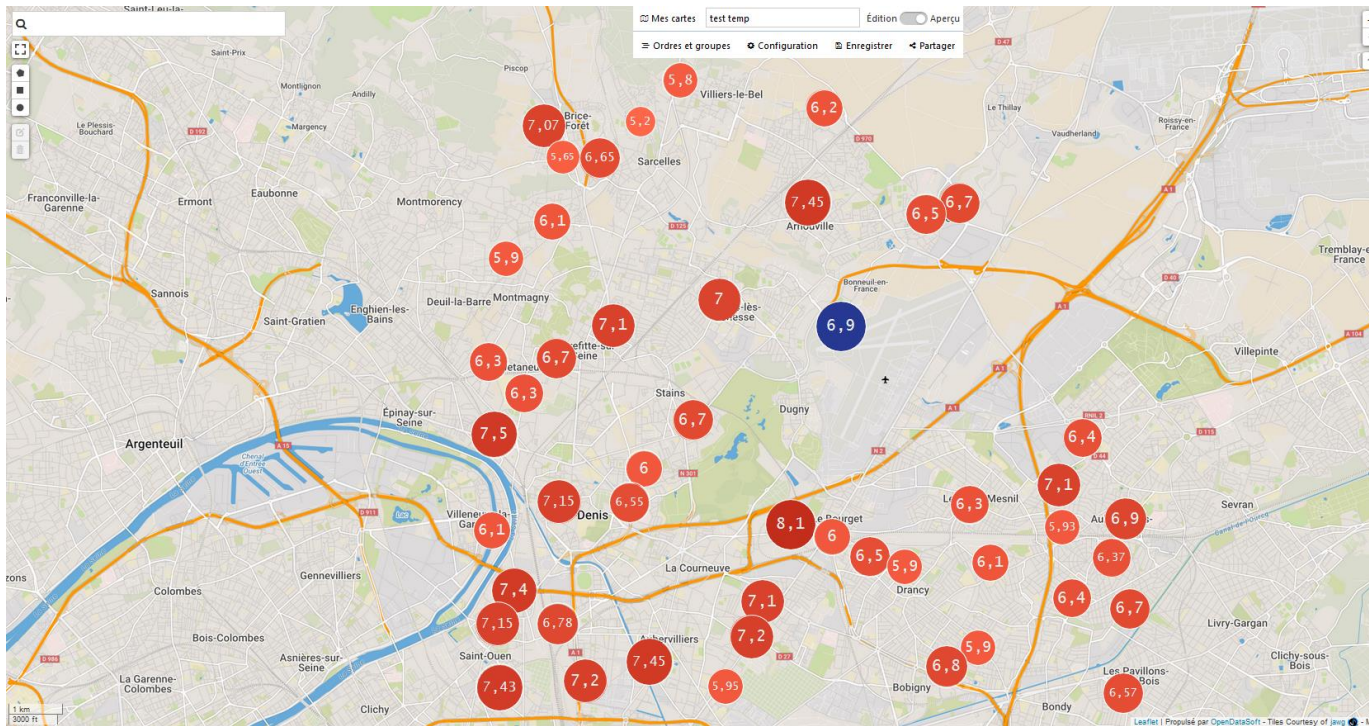
- 58 tranches nucléaires, de puissances électriques variables allant de 900 MW à 1 450 MW, réparties sur 19 sites
- 27 tranches thermiques en fonctionnement,
- 433 centrales hydrauliques dont:
 - environ 11 % des centrales ont une puissance unitaire supérieure à 100 MW ; elles représentent environ 60 % 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



PLUS DE CAPTEURS... QUELS CORRECTIFS À PRENDRE EN COMPTE ?



Un changement de référentiel qui peut induire des **biais** dans les modèles de calculs par rapport à des **modèles historiques** du fait d'un changement de « centre de gravité »



3.

1. Contexte d'un Opérateur de Données

2. Des données qui augmentent en volume et diversité

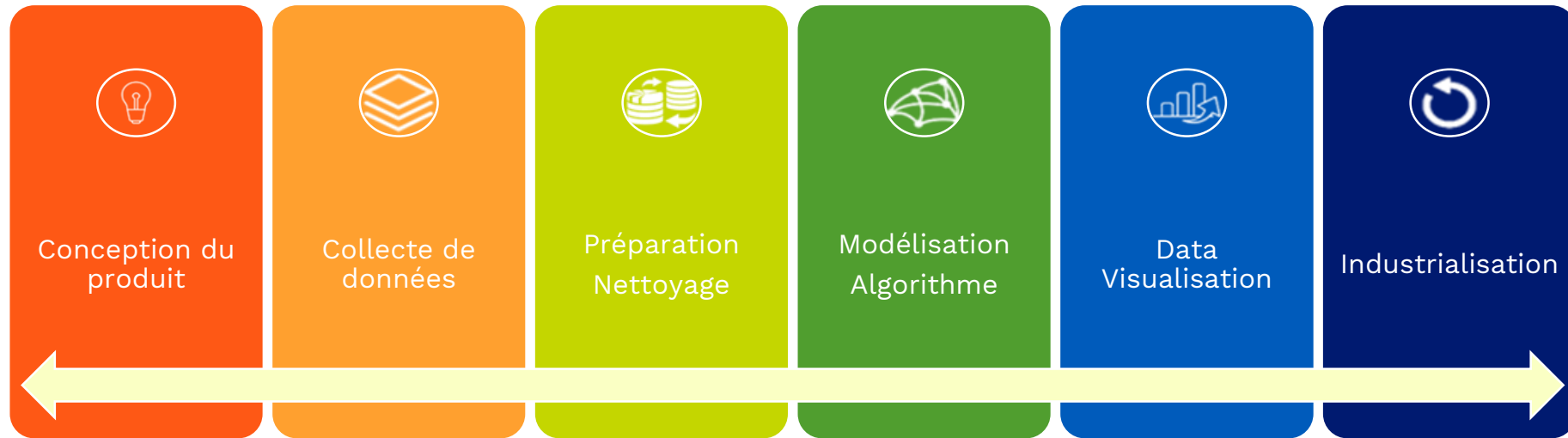
3. Un Processus générique de production

- Travailler les données au plus tôt du processus
- Confidentialité des livrables
- Axe de données manipulées

4. Un « dictionnaire » pour partager

5. Un patrimoine à valoriser par la diffusion et la réutilisation

TRAVAILLER 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

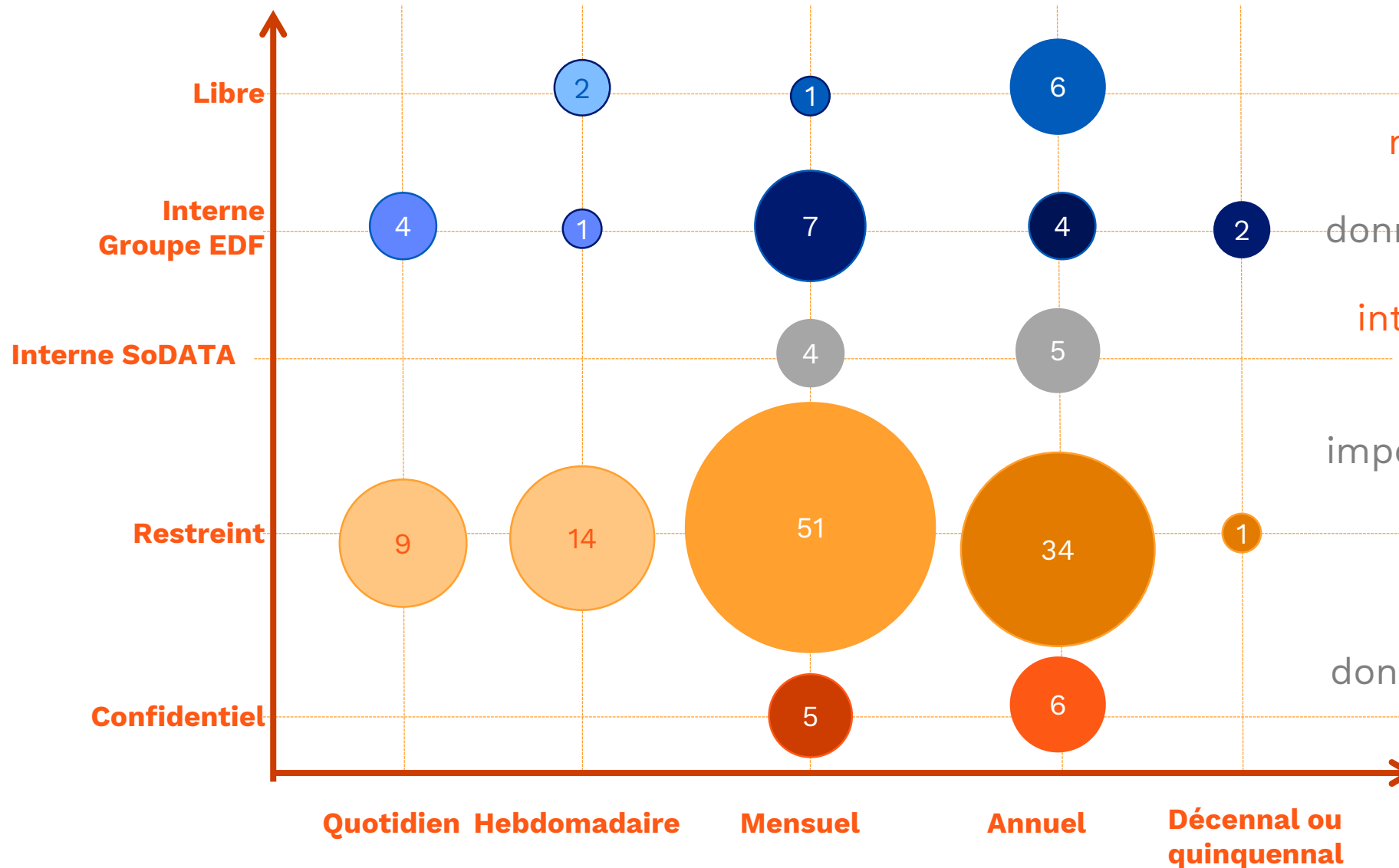
Préparation, Nettoyage, Transformation

Modélisation et algorithme pour optimiser la production d'un produit récurrent ou d'une étude

Data Visualisation pour favoriser la diffusion

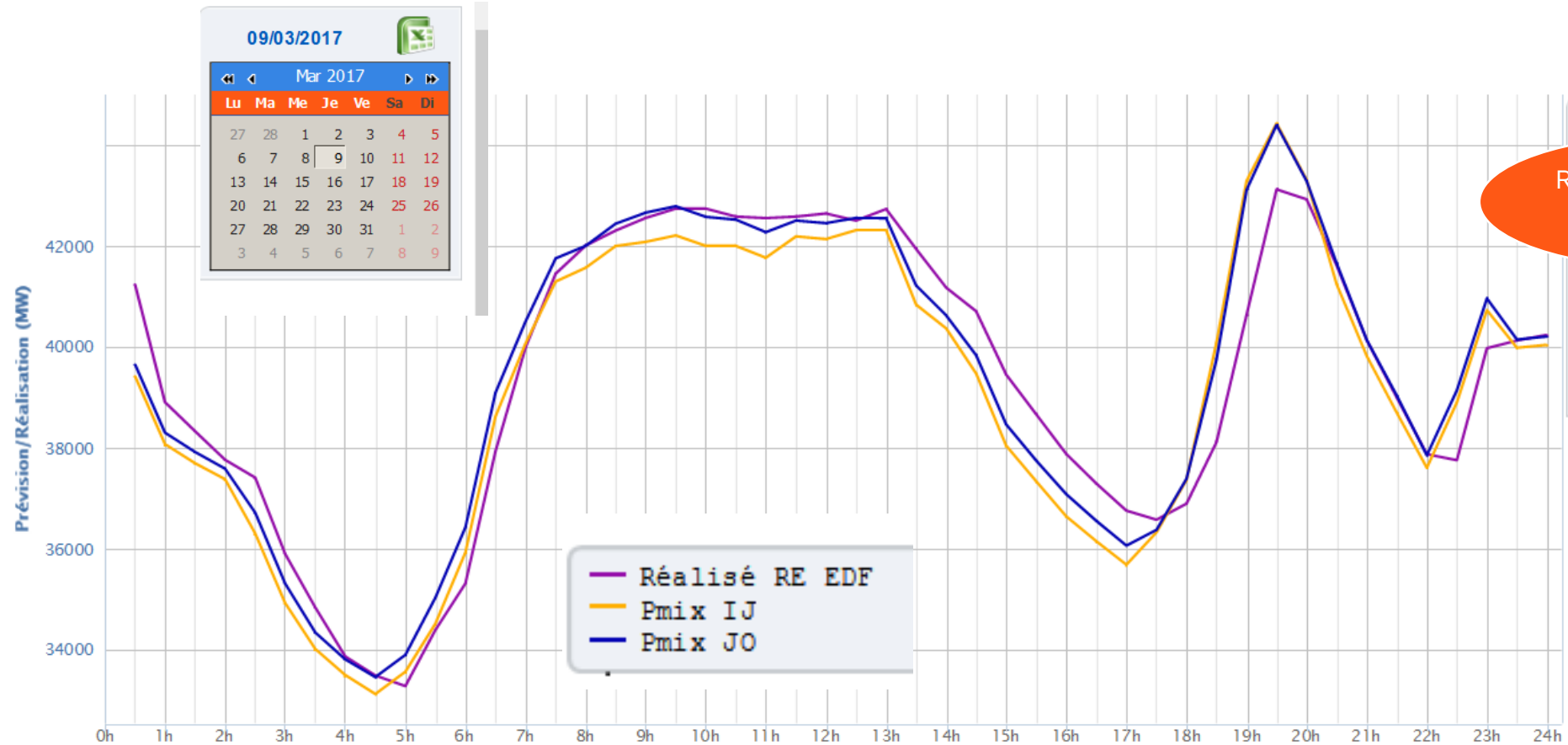
Industrialisation dans les cas où les utilisateurs sont demandeur d'une alimentation récurrente

CONFIDENTIALITÉ DES INFORMATIONS

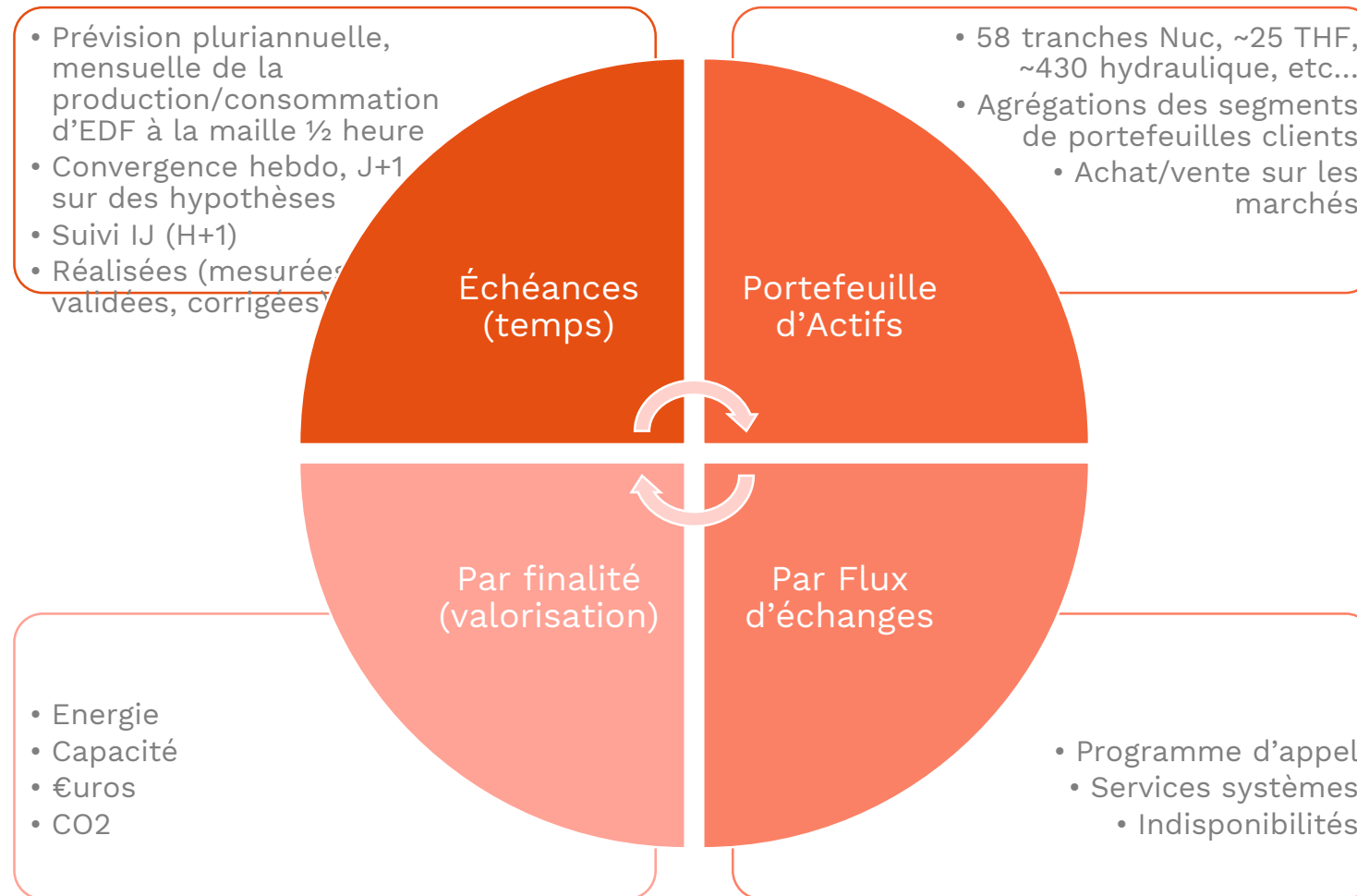


La généralisation progressive de **nouveaux systèmes d'entreposage** de données et documents (**local**, datacenters **internes**, datacenters externes, **cloud** public/privée, ...) impose de questionner le **périmètre d'accessibilité et conservation / destruction** des données dans la **durée**

CONSO > MODÈLES PRÉDICTIFS > MÉTHODE



AXE DE DONNÉES MANIPULÉES





4.

1. Contexte d'un Opérateur de Données

2. Des données qui augmentent en volume et diversité

3. Un Processus générique de production

4. Un « dictionnaire » pour partager

- Quelques définitions
- Modalités d'acquisition versus stockage et représentation
- Importance des métadonnées

5. 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 SoDATA reçoit, analyse 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

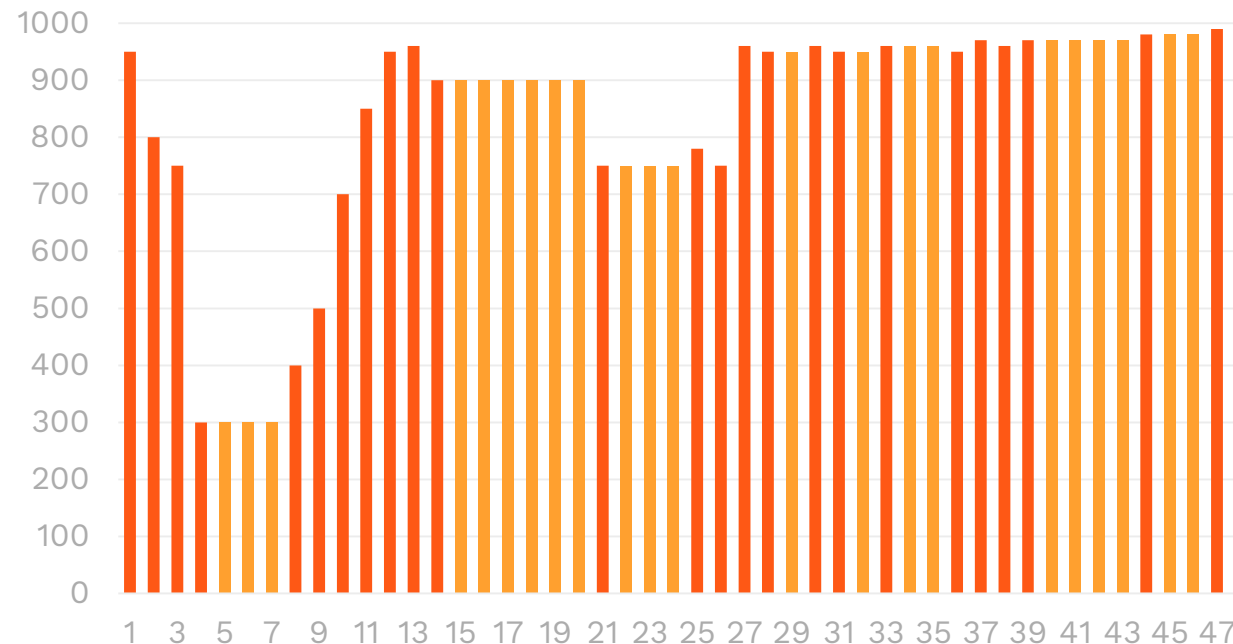
MODALITÉ D'ACQUISITION VERSUS STOCKAGE / REPRÉSENTATION

Donnée

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

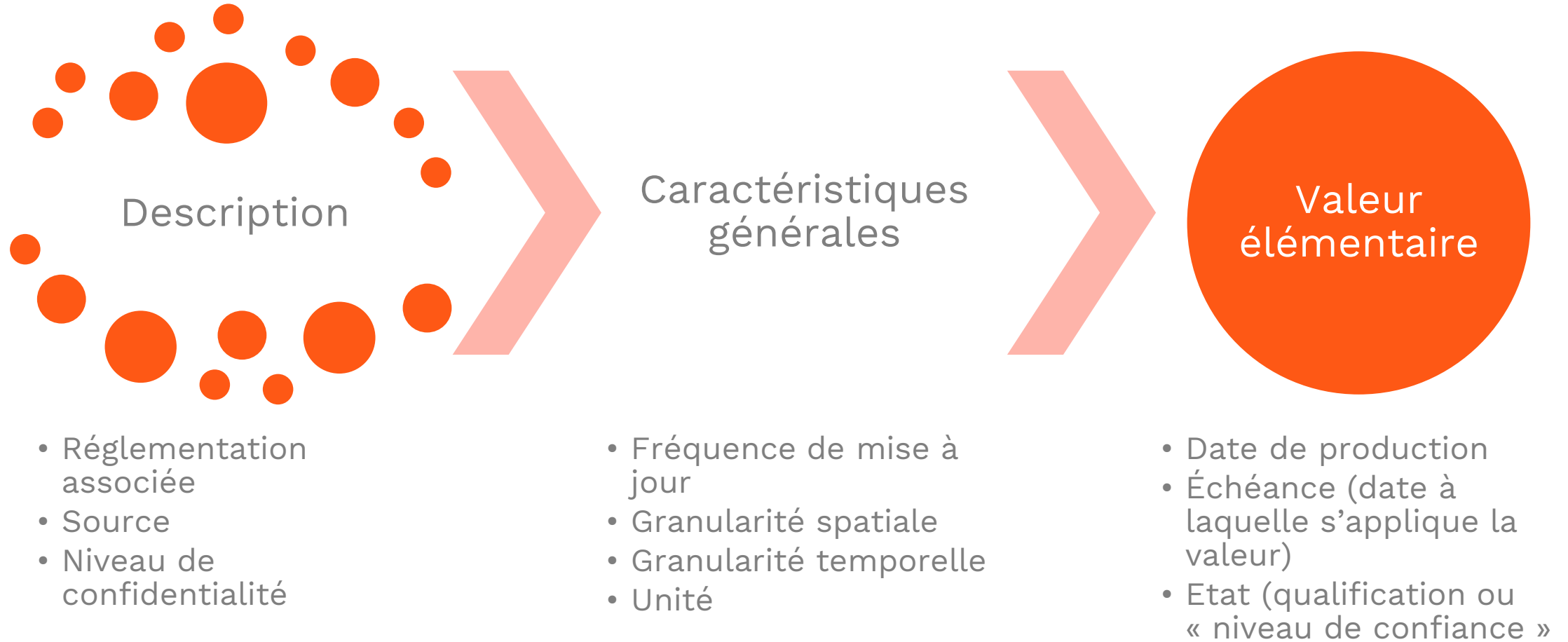
Valeur transmise selon

- 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





5.

1. Contexte d'un Opérateur de Données

2. Des données qui augmentent en volume et diversité

3. Un Processus générique de production

4. Un « dictionnaire » pour partager

5. 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
- Les données de référence en France
- Un potentiel à développer

CHOISIR UN « STANDARD »

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

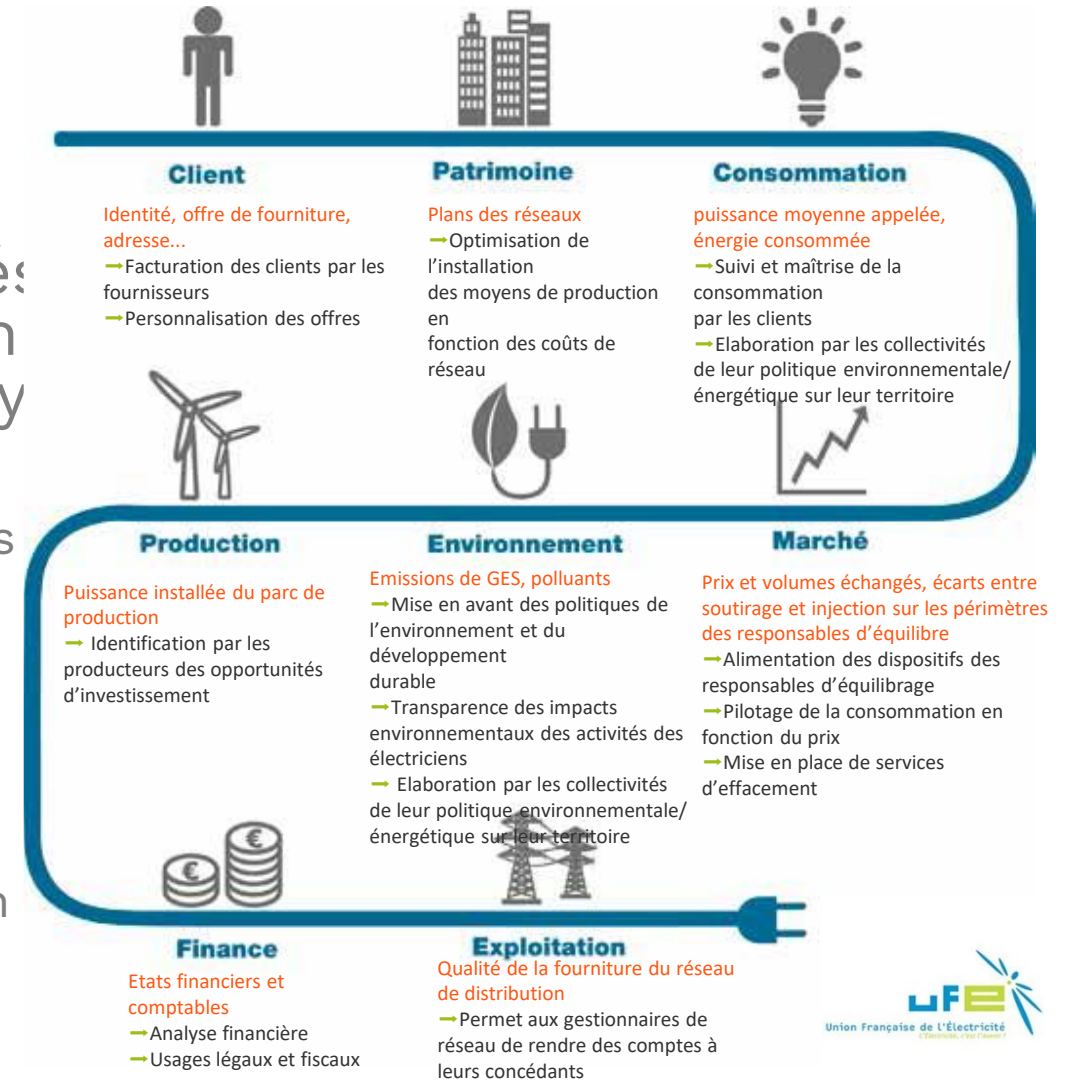
- Modèle Européen des échanges liés aux marchés électriques (Common Information Model (CIM) for Energy Markets – ENTSOE)

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

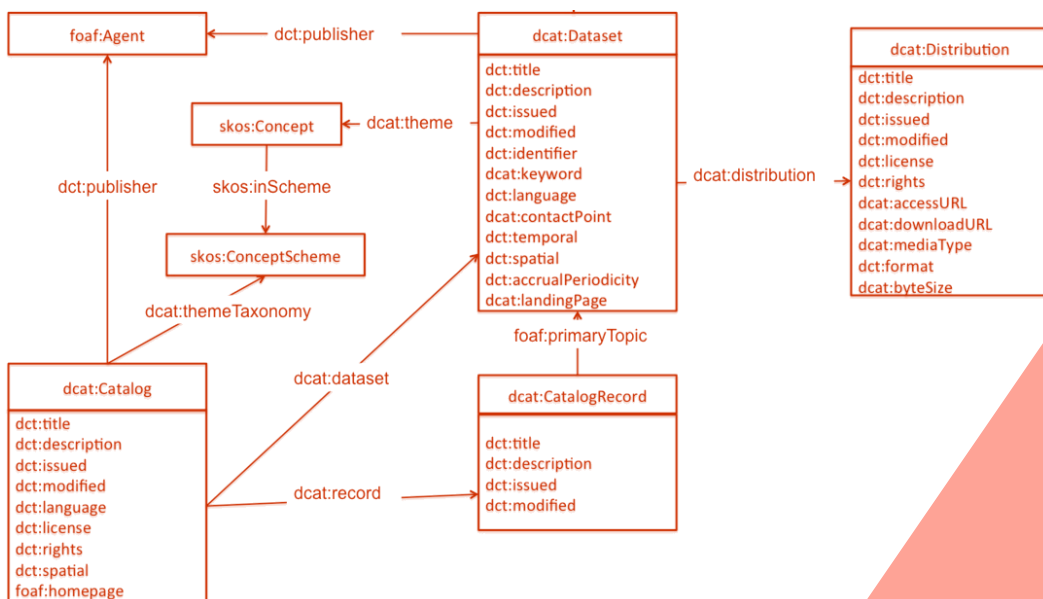


CLASSIFICATION DES JEUX DE DONNÉES AU SEIN D'UN CATALOGUE

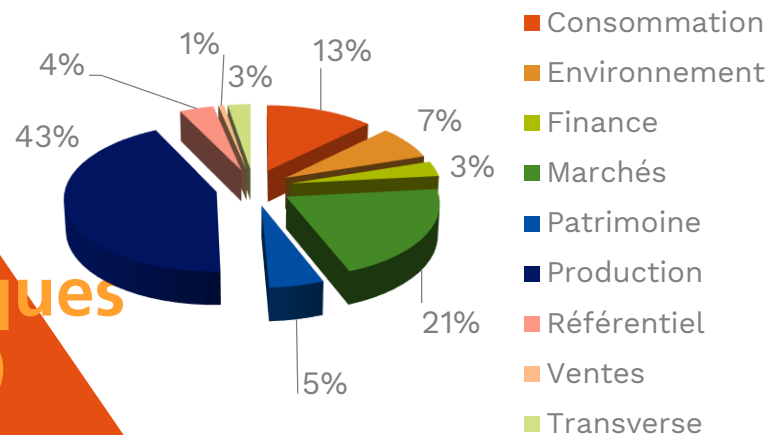
Préparer la bascule vers des outils « big data »

- Moins de modélisation « à priori »
- Plus d'index de recherche

Cataloguer avec des méta-données standards pour faciliter l'accès aux « nouveaux » utilisateurs et aux « machines »

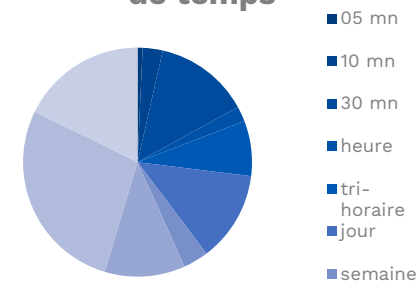


8 thématiques (UFE)



64 macros-données

Répartition par pas de temps



450 jeux de données

EXEMPLE DE MACRO-DONNÉES



Référentiels

- Contrat d'achat
- Contreparties
- Eléments de preuve
- Eléments 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



Exploitation

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



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



Consommation

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



Production

- Courbes de charges
- Production d'énergie



Patrimoine

- Bâtiments tertiaires
- Flotte de véhicules



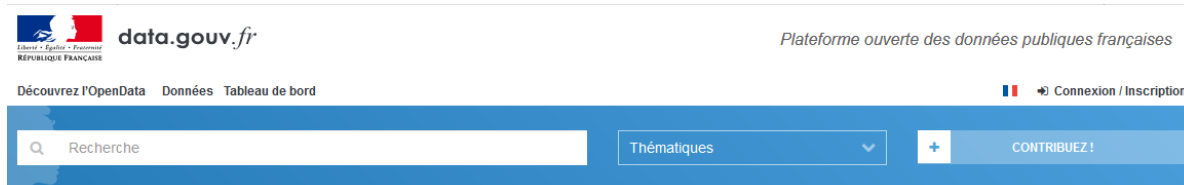
Finances

- Achat de biens et de services
- Immobilisations
- PMT



Clients

LES DONNÉES DE RÉFÉRENCE EN FRANCE







Service public de la donnée : des données sur lesquelles vous pouvez compter

Le service public de la donnée créé par l'Article 14 de la loi pour une République numérique vise à mettre à disposition, en vue de faciliter leur réutilisation, les jeux de données de référence qui présentent le plus fort impact économique et social. Il s'adresse principalement aux entreprises et aux administrations pour qui la disponibilité d'une donnée de qualité est critique. Les producteurs et les diffuseurs prennent des engagements auprès de ces utilisateurs. La mission Etalab est chargée de la mise en oeuvre et de la gouvernance de ce nouveau service public. Elle référence l'ensemble des données concernées sur cette page.

Les données de référence

À ce jour, neuf jeux de données, qui couvrent un large champ thématique ont été identifiés comme des données de référence.

 BAN - Base Adresse Nationale 26 ★ 30	 Base Sirene des entreprises et de leurs établissements (SIREN, SIRET) 35 ★ 45	 Code Officiel Géographique (COG) 5 ★ 15
 Plan Cadastral Informatisé (PCI) 1 ★ 44	 Registre parcellaire graphique (RPG) : contours des îlots cultureux et leur groupe de cultures majoritaire 1 ★ 6	 Référentiel de l'organisation administrative de l'Etat 4 ★ 13
 Référentiel à grande échelle (RGE) 2 ★ 8	 Répertoire National des Associations (RNA) 0 ★ 13	 Répertoire Opérationnel des Métiers et des Emplois (ROME) 1 ★ 7

Chacun de ces « référentiels » sont des « clés de jointure potentielle »

- Code communes INSEE
 - Données statistiques aux mailles administratives
- Données géographiques
 - Projections cartographiques sur fond de plan
 - Calculs de distances, ...
- Codes Siren
 - Données descriptives économiques

UN POTENTIEL À DÉVELOPPER

Favoriser l'exploration par le croisement possible de données

- Normalisation des métadonnées (standards [RDF](#))
- Partage de catalogues (standard [DCAT](#))

Tirer parti des **nouvelles technologies** de stockage et analyse

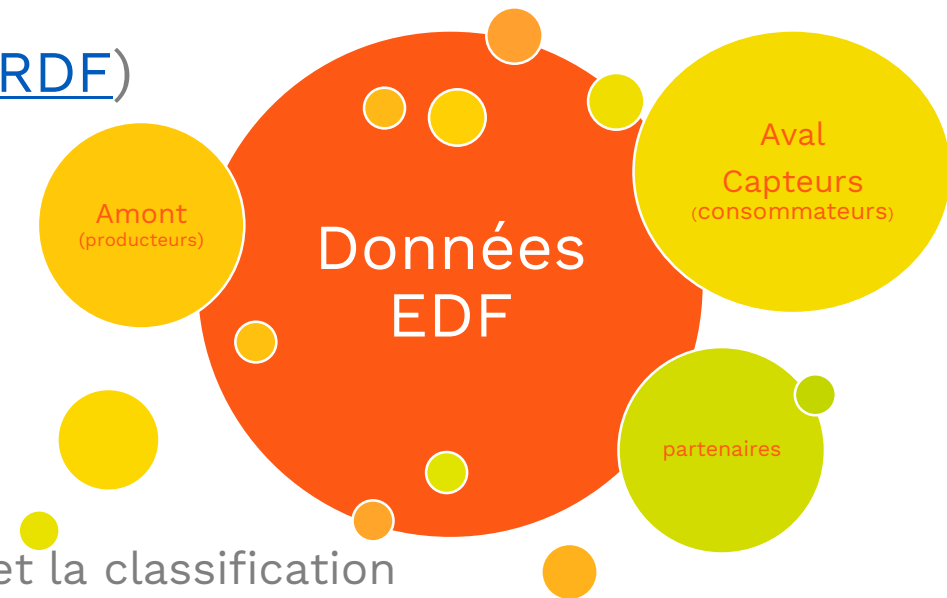
- Hadoop, spark
- Python, R
- API json
- ...

Utiliser des **automates** (**chatbots**) pour assister la qualification et la classification

- Analyse sémantique de textes issus des journaux de bord d'exploitation et comparaison avec des données structurées remontées par ailleurs

S'appuyer sur **des infrastructures partagées** avec nos partenaires pour éviter les duplications

Travailler en **open-innovation** maîtrisée



Merci
stephane.ternot@edf.fr



Sources :

www.edf.fr

www.rte-france.com

www.asn.fr

ufe-electricite.fr

www.entsoe.eu

www.cre.fr

data.gouv.fr

