Centre de Calcul de l'Institut National de Physique Nucléaire et de Physique des Particules

Nothing is lost, nothing is created, everything is <xml:transformed>

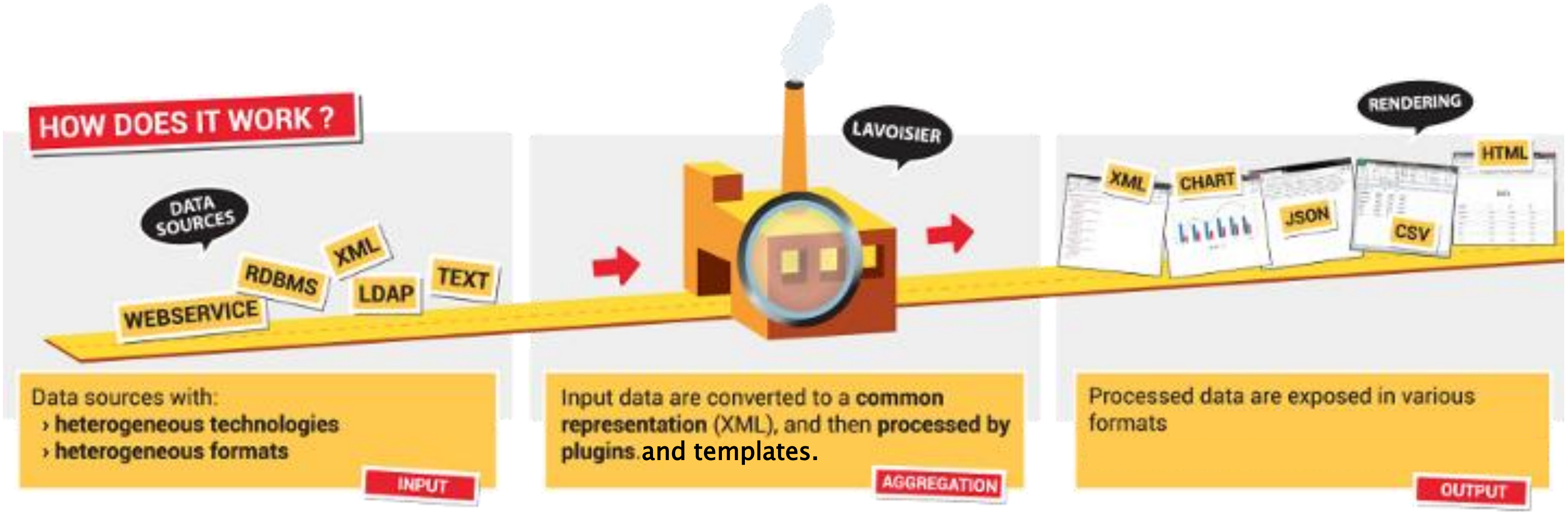
LAVOISIER

DATA AGGREGATION FRAMEWORK

CCIN2P3

What is Lavoisier ?

An open source framework for developing data aggregation applications



Data sources with:
- heterogeneous technologies
- heterogeneous formats

Input data are converted to a **common representation** (XML), and then **processed by plugins**.and templates.

Processed data are exposed in various formats

CCIN2P3

What are the main issues when developing a data aggregation application ?

Heterogeneity
- Protocols
- Data formats

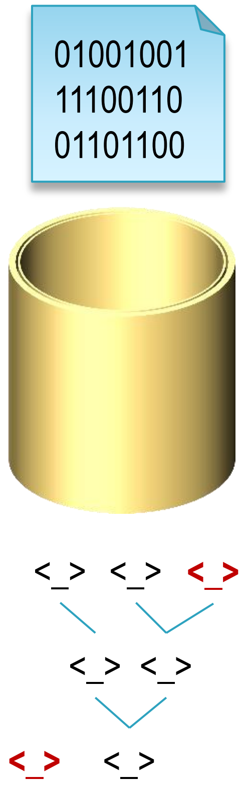
Performance
- Too big for memory
- Data source latency

Quality
- Reliability
- Robustness
- Monitorability

Security
- Authentication
- Authorization

How does Lavoisier help ?

- Common data format : XML
- Common query language : XPath
- Bytes/XML processed in **streaming**
  - detect the smallest data structure to build



CCIN2P3

What are the main issues when developing a data aggregation application ?

Heterogeneity	• Protocols • Data formats
Performance	• Too big for memory • Data source latency
Quality	• Reliability • Robustness • Monitorability
Security	• Authentication • Authorization

How does Lavoisier help ?

- Common data format : XML
- Common query language : XPath

- Bytes/XML processed in **streaming**
  - detect the smallest data structure to build
- Tune cache according to constraints of

data:	size, time-to-live, dependencies
technology:	latency, throughput, availability
users:	patience, access frequency

- Data validation
- Fallback, cache
- Web console

- Authentication by chaining plugins
- Authorization with XPath expressions

What are the main issues when developing a data aggregation application ?

Heterogeneity
- Protocols
- Data formats

Performance
- Too big for memory
- Data source latency

Quality
- Reliability
- Robustness
- Monitorability

Security
- Authentication
- Authorization

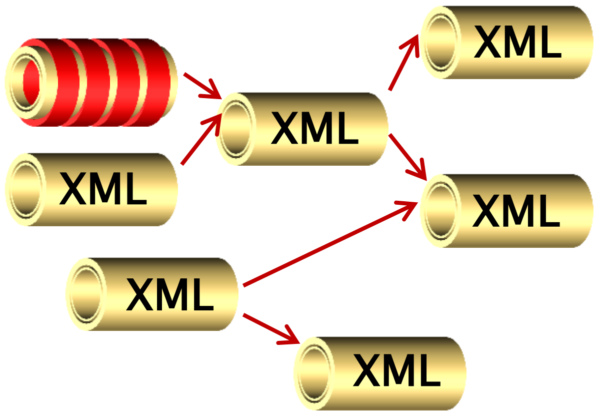
Lavoisier do all this for you…

…enabling you to focus on business code

CCIN2P3

Main benefits

- Faster to develop **quality** applications
  - performance, robustness, monitorability, testability, security…
- Easier to maintain: **declarative** programming language
  - a Lavoisier application is made of inter-dependent data views
  - each data view is a chain of plugins and templates



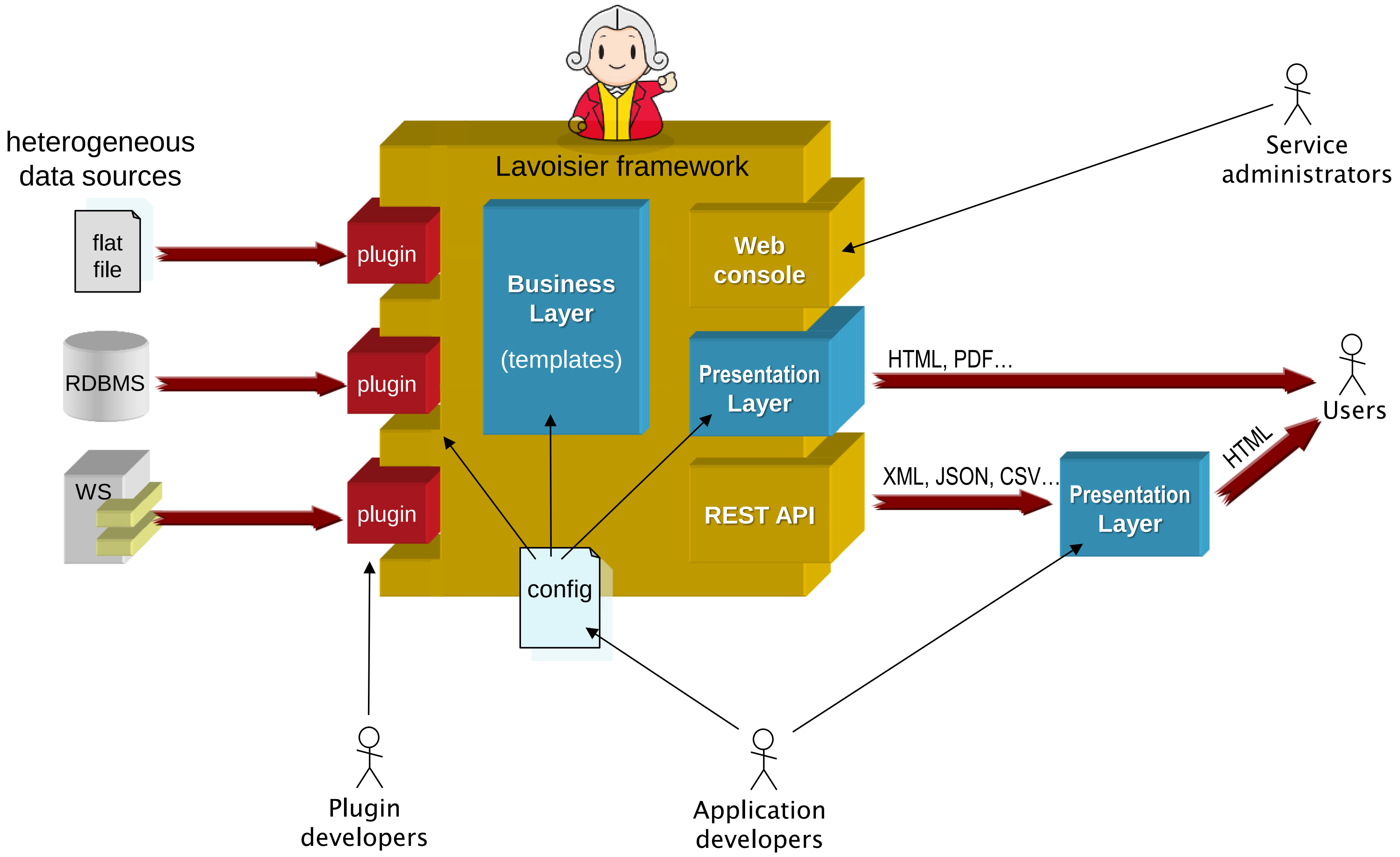
Main benefits

- Faster to develop **quality** applications
  - performance, robustness, monitorability, testability, security…
- Easier to maintain: **declarative** programming language
  - a Lavoisier application is made of inter-dependent data views
  - each data view is a chain of plugins and templates
- Factorize development efforts : **about 100 plugins** for

Technologies	HTTP, SQL, SSH, grid (JSAGA)…
Formats	JSON, YAML, CSV, LDIF, text…
Cache mechanism	on disk/in memory, indexed, BaseX…
Security	CAS, password, X509, IP, OAuth…

- Separate actors responsibilities *(see next slide)*

Main benefits : separate actors responsibilities



CCIN2P3

Projects using Lavoisier

EGI projects

- OPERATIONS PORTAL
- **ARGO**
  - availabilities, reliabilities
- **VAPOR**
  - toolkit for VO management

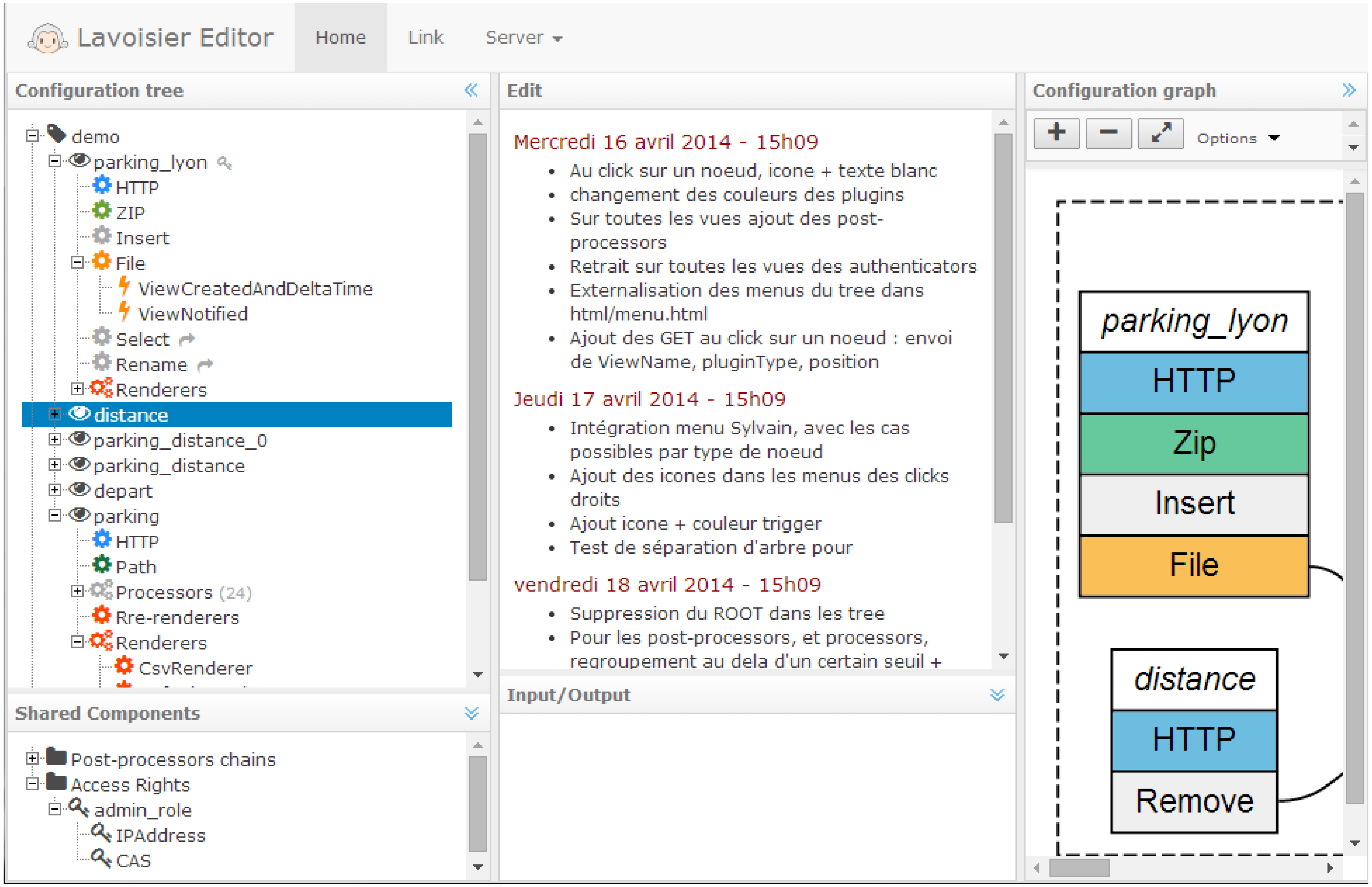
(almost) all these use-cases have heterogeneous data sources

CC-IN2P3 projects

- Data import for **CMDB**
  - (see Emmanouil's talk)
- **CC-Status**
  - A. Vedaee, O. Lequeux
- Cache for **Grid Engine**
  - J-R. Rouet
- **CostModel**
  - R. Vernet
- Automatic generation of CE static information for site **BDII**
  - C. Eloto
- Part of data import (iRods) in **decision-making tool**
  - C. Evesque

Perspectives : GUI under development (A. Zicaro, C. Flieller)

- GUI for assisting development of Lavoisier applications



Configuration tree:

- demo
  - parking_lyon
    - HTTP
    - ZIP
    - Insert
    - File
      - ViewCreatedAndDeltaTime
      - ViewNotified
    - Select
    - Rename
    - Renderers
  - distance
  - parking_distance_0
  - parking_distance
  - depart
  - parking
    - HTTP
    - Path
    - Processors (24)
    - Rre-renderers
    - Renderers
      - CsvRenderer

Shared Components:

- Post-processors chains
- Access Rights
  - admin_role
    - IPAddress
    - CAS

Edit:

Mercredi 16 avril 2014 - 15h09

- Au click sur un noeud, icone + texte blanc
- changement des couleurs des plugins
- Sur toutes les vues ajout des post-processors
- Retrait sur toutes les vues des authenticators
- Externalisation des menus du tree dans html/menu.html
- Ajout des GET au click sur un noeud : envoi de ViewName, pluginType, position

Jeudi 17 avril 2014 - 15h09

- Intégration menu Sylvain, avec les cas possibles par type de noeud
- Ajout des icones dans les menus des clicks droits
- Ajout icone + couleur trigger
- Test de séparation d'arbre pour

vendredi 18 avril 2014 - 15h09

- Suppression du ROOT dans les tree
- Pour les post-processors, et processors, regroupement au dela d'un certain seuil +