





e-ginseng.org

# La solution GINSENG

Sébastien Cipièrre (Phd student, LPC)

**GINSENG SENTINELLE GRID NETWORK**  
Le projet est financé par l'ANR  
ANR 2014-2015

**GINSENG: Global Initiative for Sentinel E-health Network on Grid**

www.ginseng.org

### Contraintes du projet

- Bases de données distribuées
  - Pas d'export, pas de concentration de données
- Chaînage correct des données
- Niveau de sécurité suffisant
  - Chiffrement des informations
  - Contrôle des utilisateurs
  - Définition des droits (CPS)
- Contraintes légales
  - Dépôt d'une demande d'autorisation CNIL
  - Suivi des recommandations ASIP-Santé

### Les tâches du projet

Diagram illustrating the tasks of the project, centered around **GINSENG WPI management**. The tasks include:

- Recensement des besoins
- Installation - mise en œuvre
- Site web
- Communications
- Publications
- Mises à jour
- Mises à jour des données
- Mises à jour des utilisateurs
- Mises à jour des droits
- Mises à jour des paramètres
- Mises à jour des indicateurs
- Mises à jour des performances
- Mises à jour des coûts
- Mises à jour des risques
- Mises à jour des impacts
- Mises à jour des bénéfices
- Mises à jour des impacts sociaux
- Mises à jour des impacts environnementaux
- Mises à jour des impacts économiques
- Mises à jour des impacts culturels
- Mises à jour des impacts éducatifs
- Mises à jour des impacts sanitaires
- Mises à jour des impacts sociaux
- Mises à jour des impacts environnementaux
- Mises à jour des impacts économiques
- Mises à jour des impacts culturels
- Mises à jour des impacts éducatifs
- Mises à jour des impacts sanitaires

### Objectifs à atteindre

- Créer une infrastructure décentralisée d'échange de données de santé
- Besoins de la communauté médicale:
  - Information fiable
    - Validité → Chaînage performant des informations
  - Accessibilité de l'information
    - Disponibilité → Mettre à disposition les données tout en gardant leur contrôle
  - Information complète
    - Exhaustivité → Compléter en temps réel les informations

QuikProtect  
Copie... C:\Users\scipiere\Documents\Capt\_120928\_H1522\_54\_006.jpg



Ce projet est partiellement financé par :  
This project is partly funded by:



## GINSENG: Global Initiative for Sentinel E-health Network on Grid

Lydia Maigne (MCF, LPC),  
Vincent Breton (DR, LPC),  
Sébastien Cipièrre (IR LPC),  
David Sarramia (MCF, LPC),  
Quentin Bisson (Stagiaire, LPC),  
André Lautier (Association RSCA),  
Stéphanie Léger (MCF, CHU Clermont-Fd, LM),  
Audrey Lelong (Doctorante, CHU Clermont-Fd),  
Laurent Gerbaud (PU-PH, CHU Clermont-Fd),  
Lemlih Ouchchane (MCU-PH, ISIT-UA),  
Aline Guttmann (AHU, ISIT-UA),  
Xinran Li (Doctorant, ISIT-UA),  
Sébastien Gaspard (IR, maatG/gnúbila),  
Boris Vacher (IR, maatG/gnúbila),  
Nicolas Mugnier (IR, maatG/gnúbila),  
Jérôme Revillard (CTO, maatG/gnúbila),  
David Manset (CEO, maatG/gnúbila)

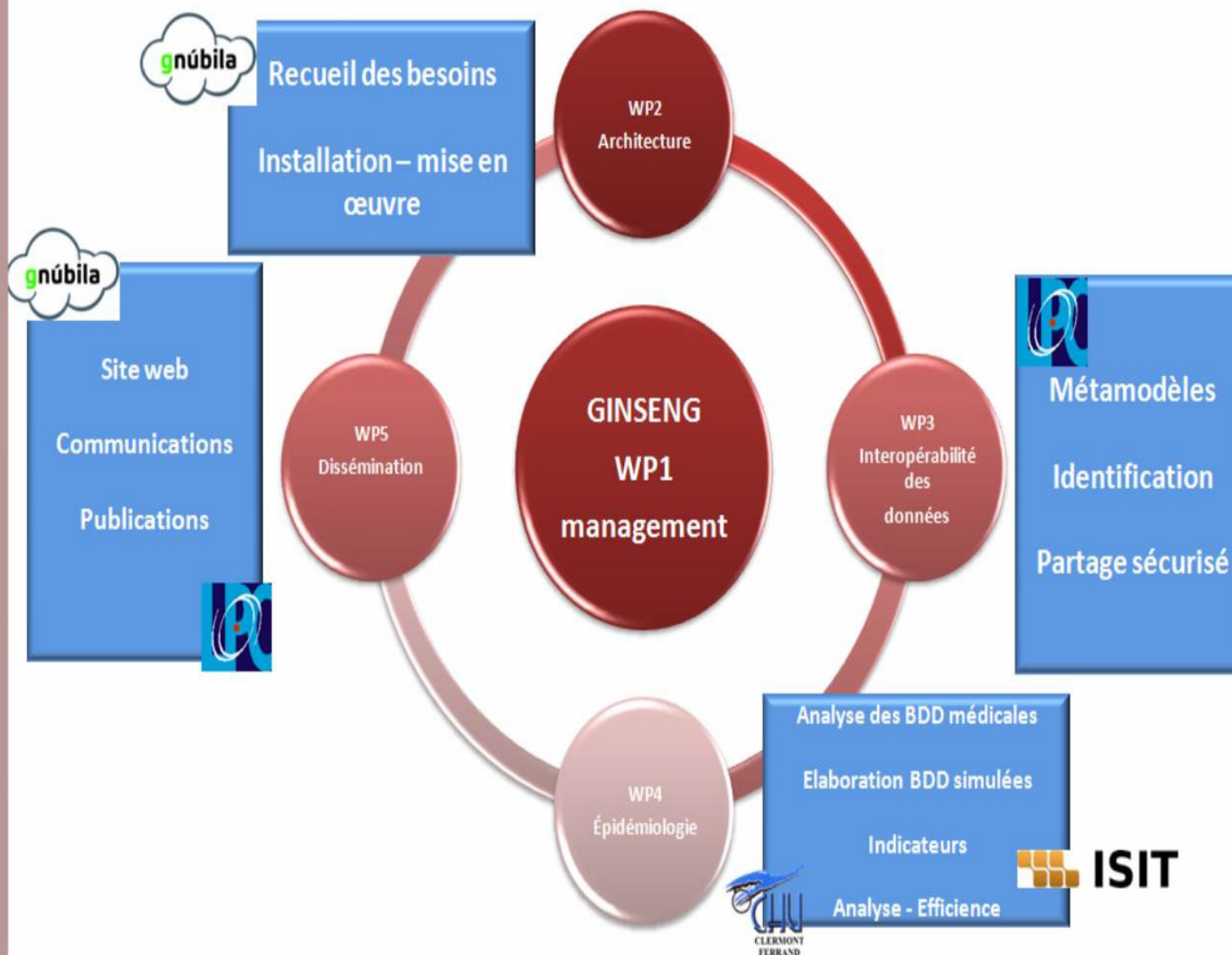
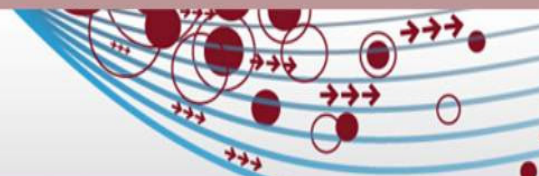
[www.e-ginseng.org](http://www.e-ginseng.org)

gnúbila



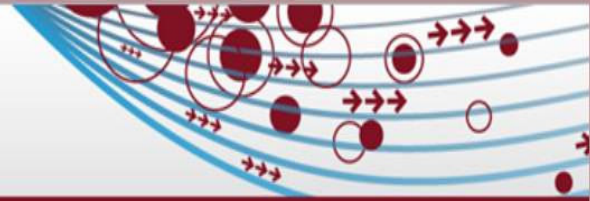


# Les tâches du projet



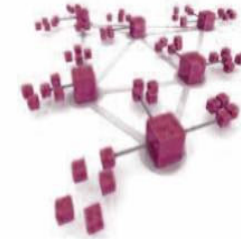


## Objectifs à atteindre



- Créer une infrastructure décentralisée d'échange de données de santé

- Besoins de la communauté médicale:



- Information fiable

**Validité**

→ Chaînage performant des informations

- Accessibilité de l'information

**Disponibilité**

→ Mettre à disposition les données tout en gardant leur contrôle

- Information complète

**Exhaustivité**

→ Compléter en temps réel les informations médicales du patient



# Contraintes du projet

- Bases de données distribuées
  - Pas d'export, pas de concentration de données
- Chaînage correct des données
- Niveau de sécurité suffisant
  - Chiffrement des informations
  - Contrôle des utilisateurs
  - Définition des droits (CPS)
- Contraintes légales
  - Dépôt d'une demande d'autorisation CNIL
  - Suivi des recommandations ASIP-Santé



# La solution GINSENG

Sébastien Cipièrre (Phd student, LPC)

**GINSENG SENTINELLE GRID NETWORK**

ANR 05A

**GINSENG: Global Initiative for Sentinel E-health Network on Grid**

www.e-ginseng.org

### Contraintes du projet

- Bases de données distribuées
  - Pas d'export, pas de concentration de données
- Chaînage correct des données
- Niveau de sécurité suffisant
  - Chiffrement des informations
  - Contrôle des utilisateurs
  - Définition des droits (CPS)
- Contraintes légales
  - Dépôt d'une demande d'autorisation CNIL
  - Suivi des recommandations ASIP-Santé

### Les tâches du projet

Recueil des besoins

Installation - mise en œuvre

Site web

Communications

Publications

GINSENG WPI management

Mécanismes

Identification

Partage sécurisé

Analyses des 100 modules

Élaboration 300 modules

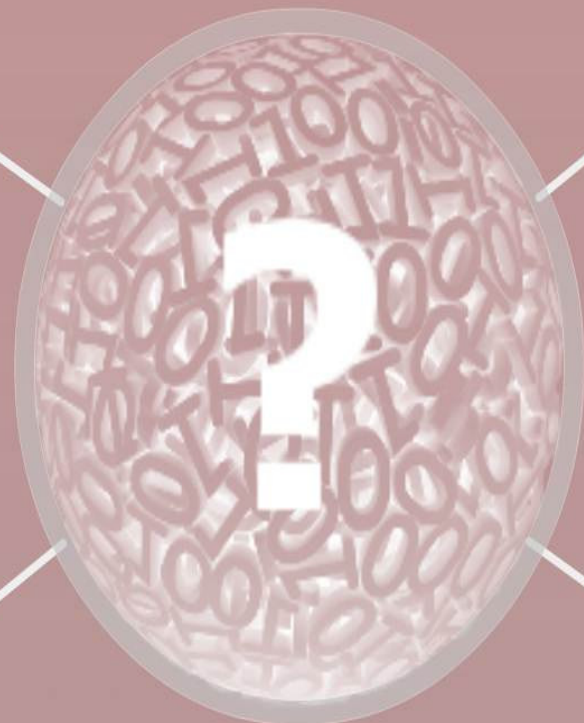
Indicateurs

Analyses d'efficacité

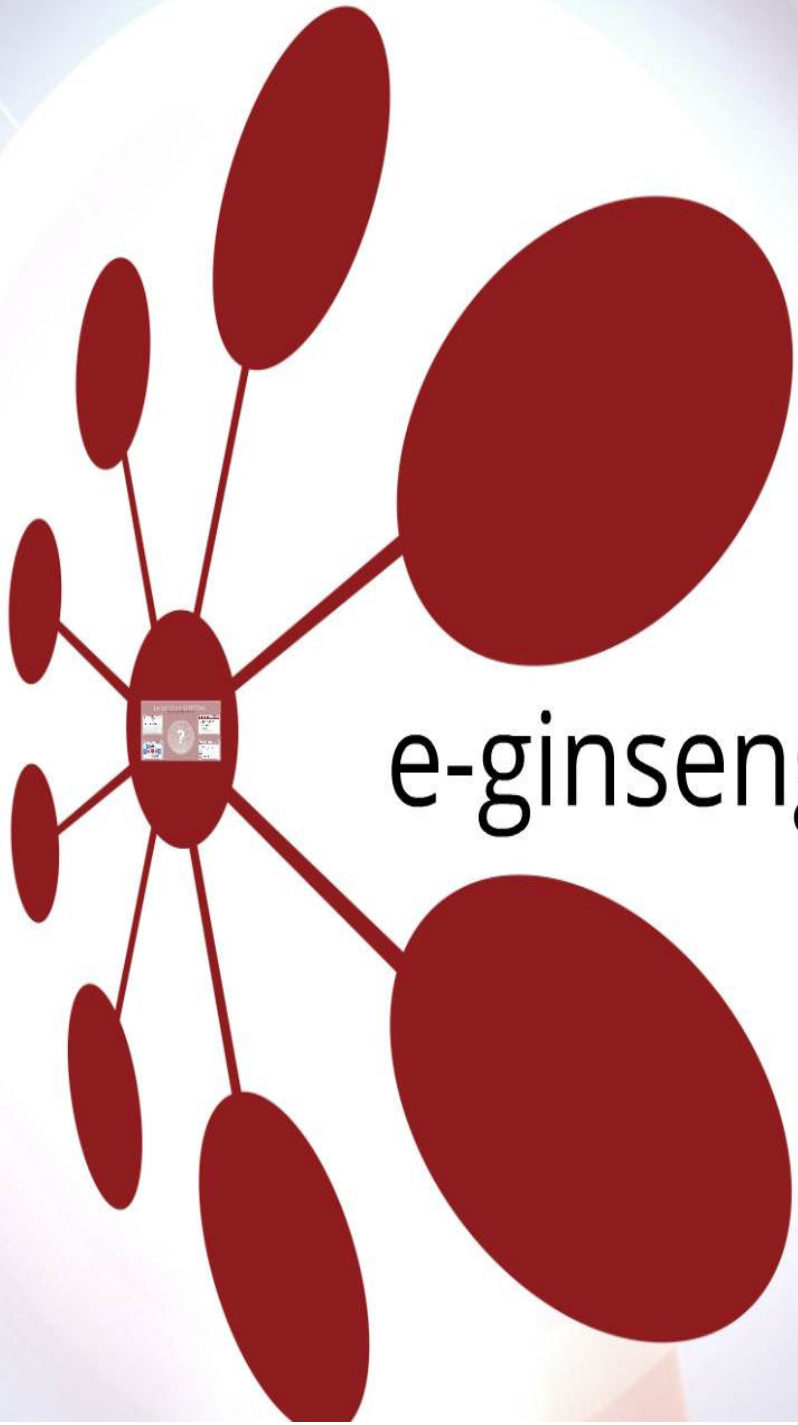
ISIT

### Objectifs à atteindre

- Créer une infrastructure décentralisée d'échange de données de santé
  - Besoins de la communauté médicale:
    - Information fiable
      - Validité → Chaînage performant des informations
    - Accessibilité de l'information
      - Disponibilité → Mettre à disposition les données tout en gardant leur contrôle
    - Information complète
      - Exhaustivité → Compléter en temps réel les informations médicales du patient

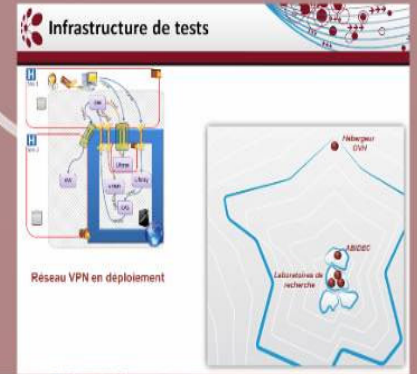
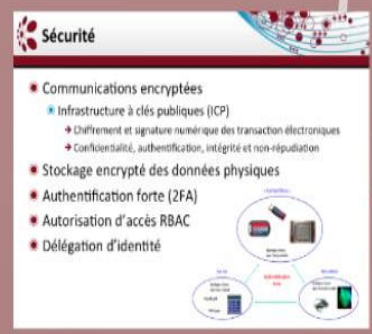
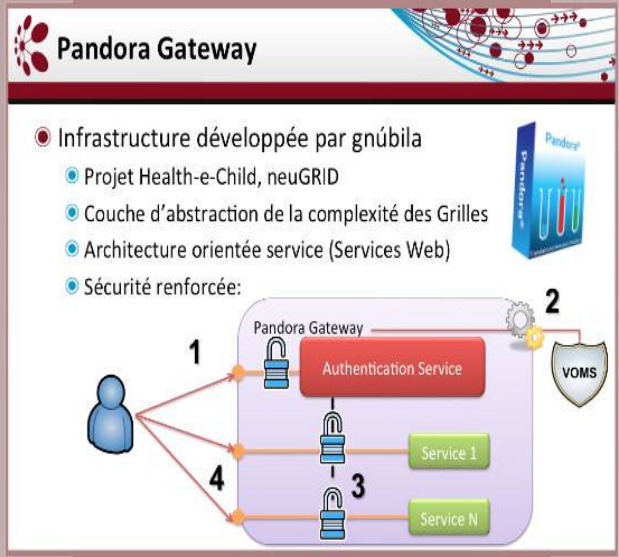
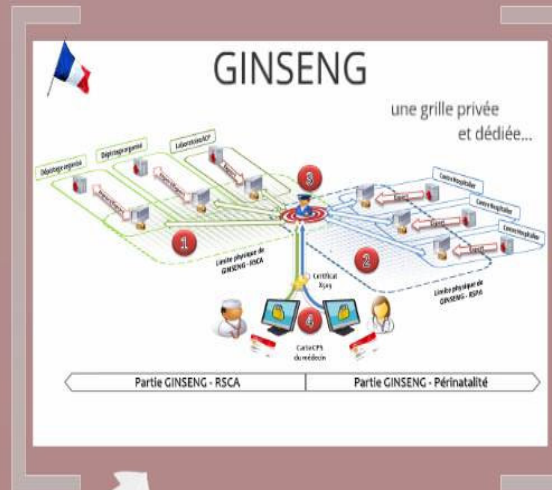






e-ginseng.org

# Infrastructure Distribuée

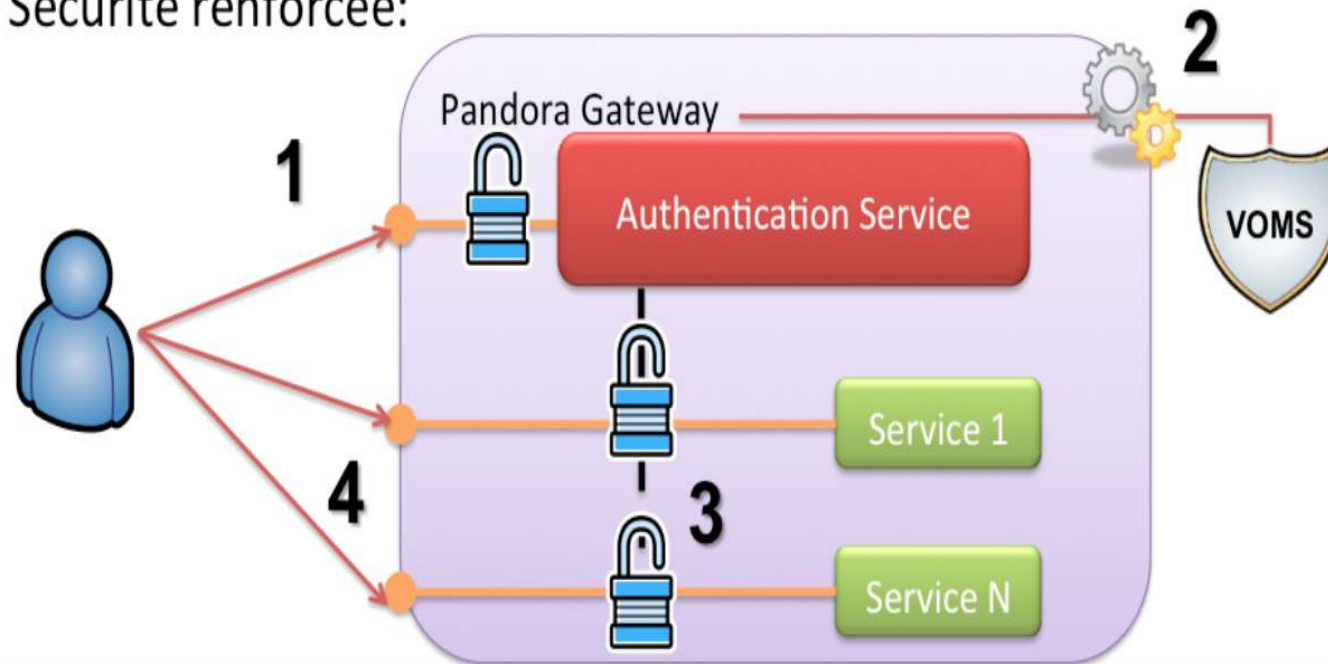


© 2010 ProjeCo. Tous droits réservés. Ce document est distribué sous licence Creative Commons Attribution-NonCommercial-ShareAlike. Toute réimpression ou utilisation non autorisée sans la permission écrite de la ProjeCo est formellement interdite. Toute réimpression ou utilisation non autorisée sans la permission écrite de la ProjeCo est formellement interdite.



# Pandora Gateway

- Infrastructure développée par gnúbila
  - Projet Health-e-Child, neuGRID
  - Couche d'abstraction de la complexité des Grilles
  - Architecture orientée service (Services Web)
  - Sécurité renforcée:

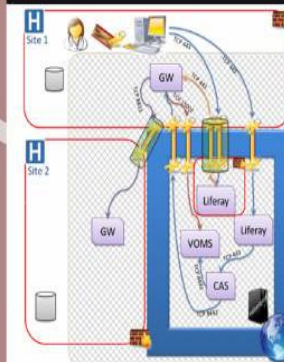


## Sécurité

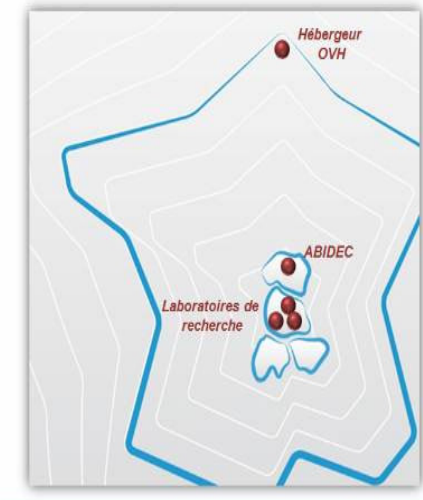
- Communications encryptées
  - Infrastructure à clés publiques (ICP)
    - Chiffrement et signature numérique des transaction électroniques
    - Confidentialité, authentification, intégrité et non-répudiation
- Stockage encrypté des données physiques
- Authentification forte (2FA)
- Autorisation d'accès RBAC
- Délégation d'identité



## Infrastructure de tests



Réseau VPN en déploiement



## Éléments de sécurité: CPS

- Mise en place d'un système d'authentification des utilisateurs par Carte de Professionnel de Santé [1]

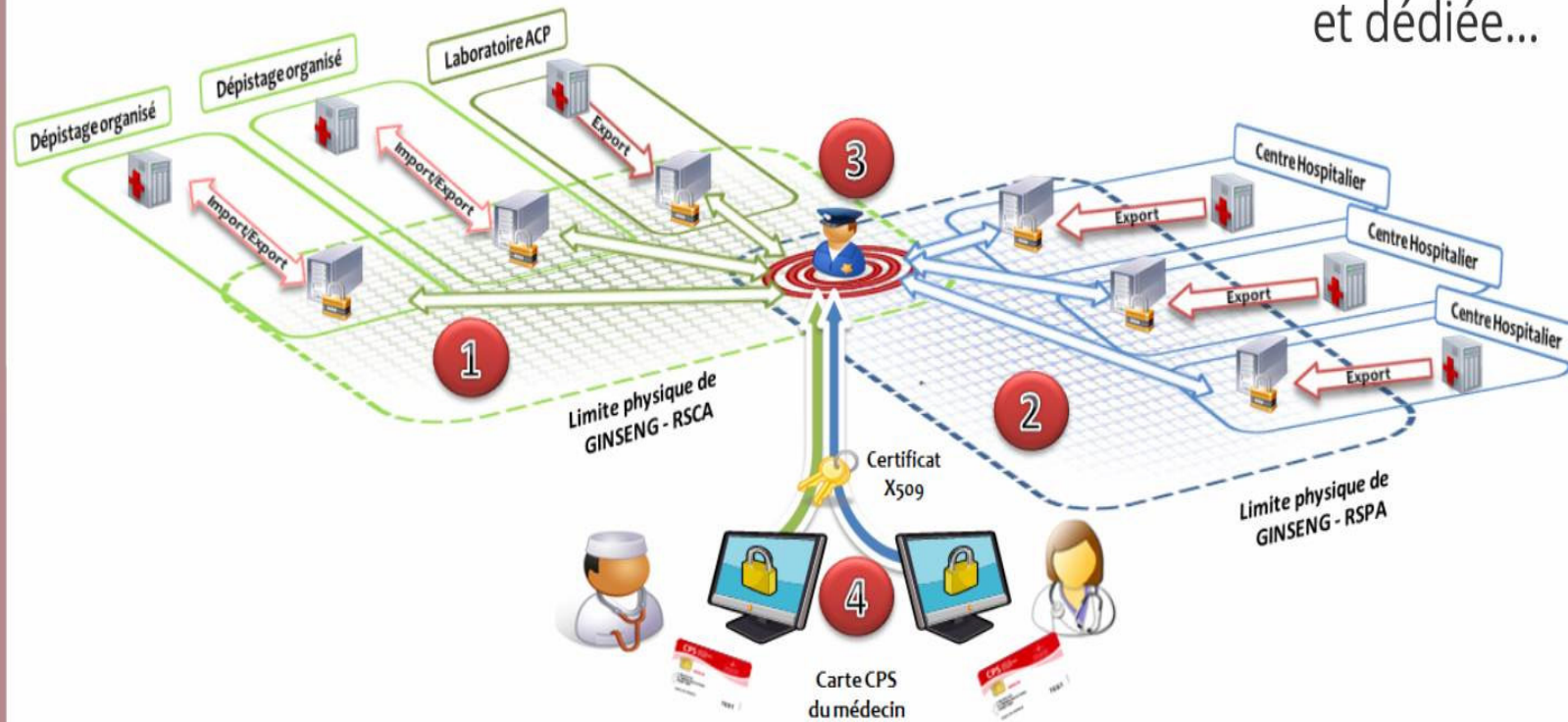


- Contenu d'une CPS:
  - Informations sur le porteur de la carte (personne, métier, ...)
  - Deux certificats X509: un pour l'authentification et un pour la signature



# GINSENG

une grille privée  
et dédiée...



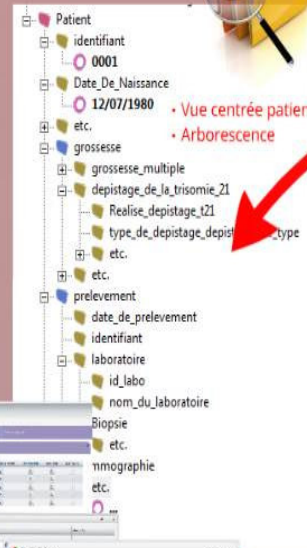
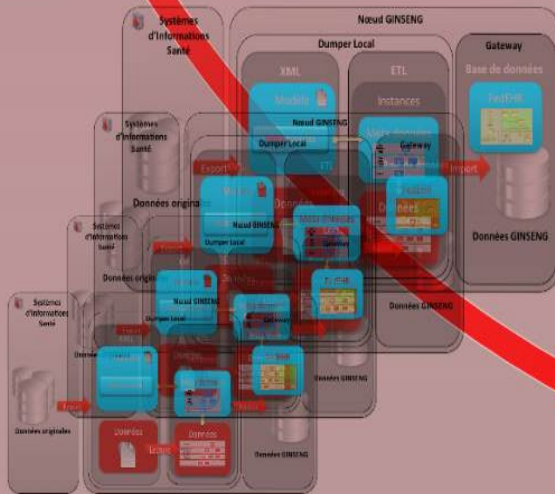
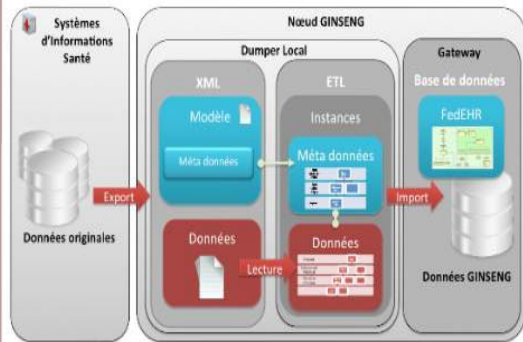
Partie GINSENG - RSCA

Partie GINSENG - Périnatalité

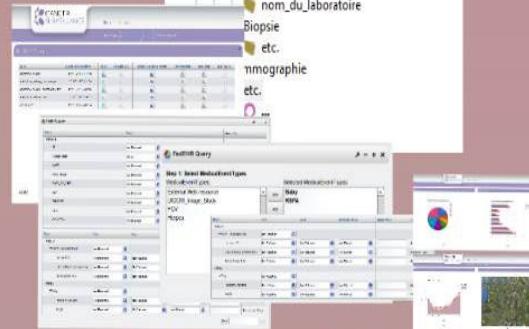
# Intégration des données

Sébastien Gaspard (IR, maatG/gnúbila),  
Boris Vacher (IR, maatG/gnúbila),  
Sébastien Cipièrre (IR, LPC), Lydia Maigne (MCF, LPC)  
David Sarramia (MCF, LPC)

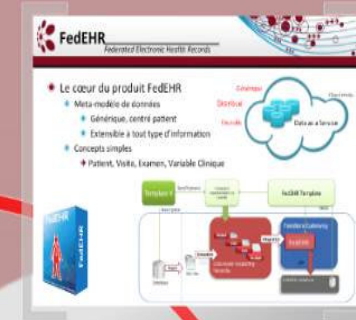
## Noeud GINSENG



• Vue centrée patient  
• Arborescence



• Interface requête générique  
• Couplage avec outils visualisation



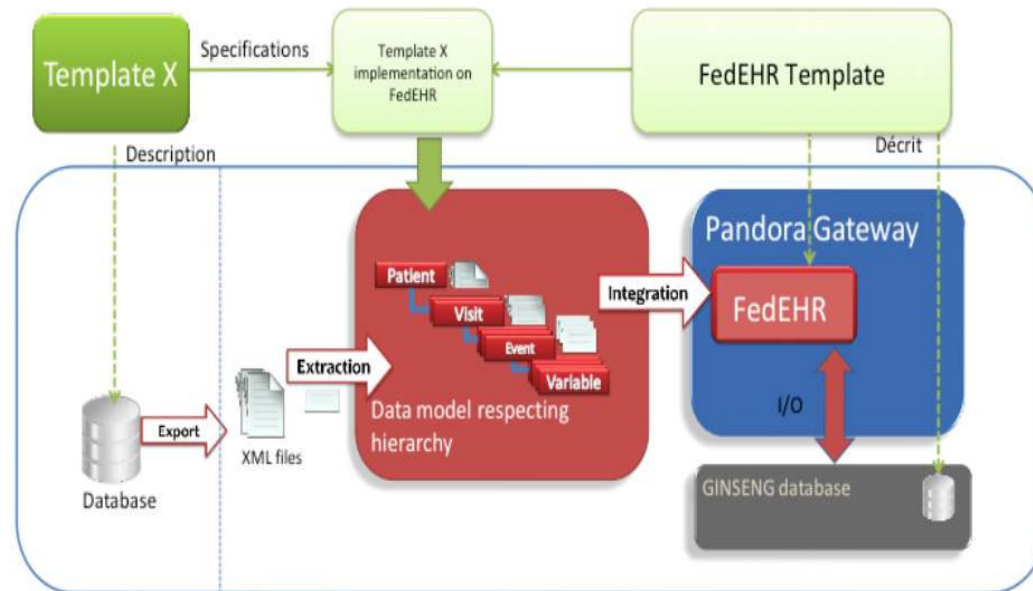


# FedEHR

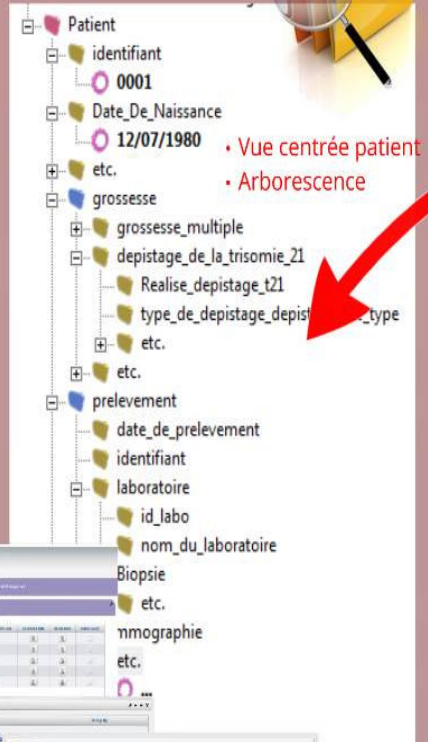
Federated Electronic Health Records

## Le cœur du produit FedEHR

- Meta-modèle de données
  - Générique, centré patient
  - Extensible à tout type d'information
- Concepts simples
  - Patient, Visite, Examen, Variable Clinique



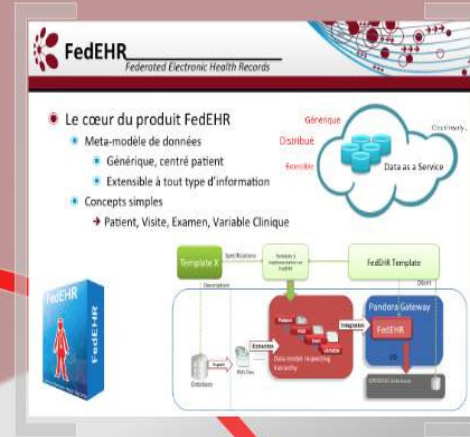
Sébastien Gaspard (IR, maatG/gnúbila),  
Boris Vacher (IR, maatG/gnúbila),  
Sébastien Cipièrre (IR, LPC), Lydia Maigne (MCF, LPC)  
David Sarramia (MCF, LPC)



• Vue centrée patient  
• Arborescence

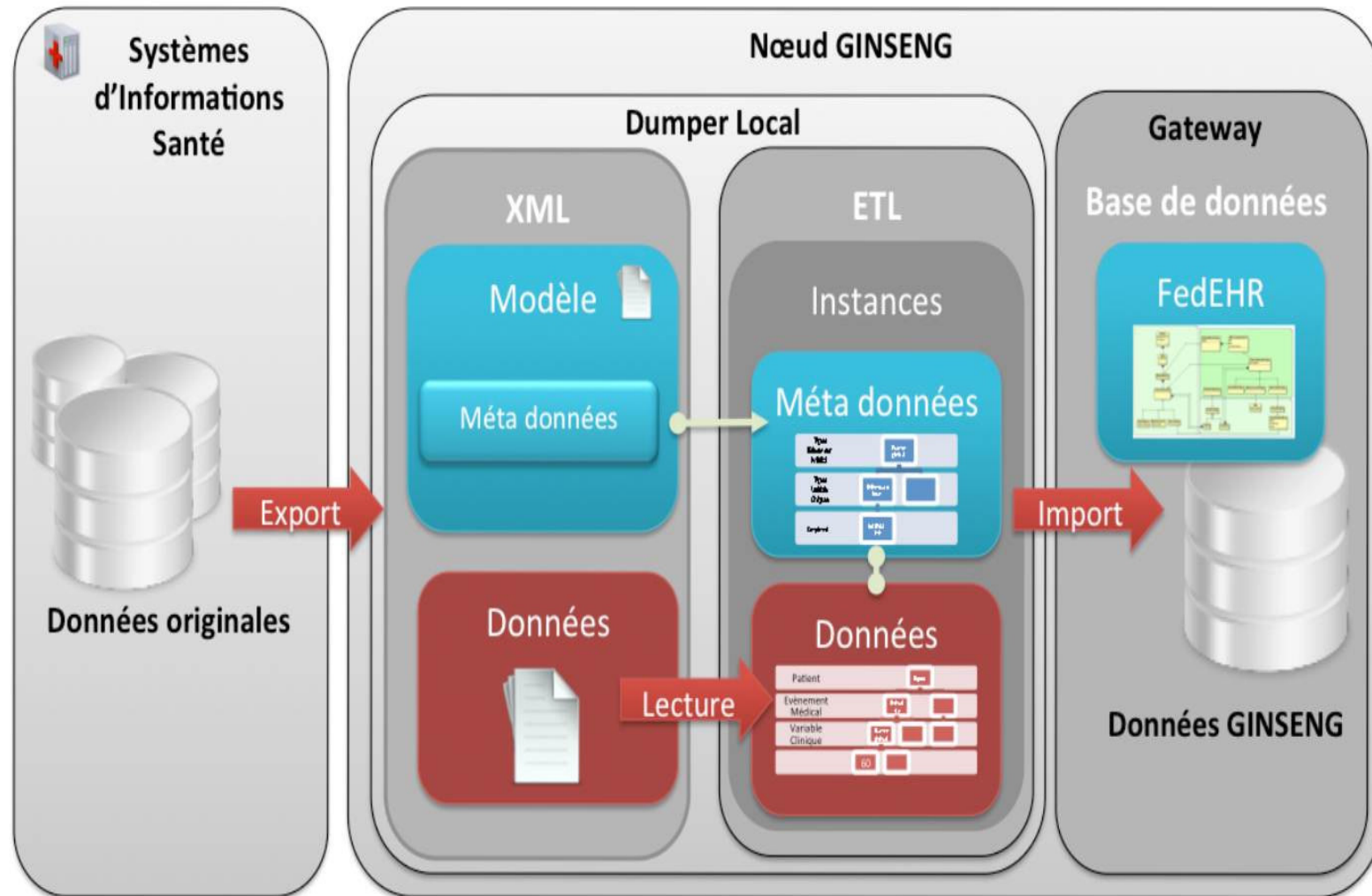
A collection of screenshots from the FedEHR interface. One screenshot shows a 'CANCER SURVEILLANCE' table with columns for patient ID, date, and status. Another screenshot shows a 'FedEHR Query' window with a 'Step 1: Select MedicalEventTypes' dialog box. A third screenshot shows a 'Biopsie' window with a 'Select MedicalEventTypes' dialog box. A fourth screenshot shows a dashboard with a pie chart and a map.

• Interface requêtage générique  
• Couplage avec outils visualisation





# Noeud GINSENG



Sébastien  
Boris  
Sébastien  
David

Query	Last Data Calculation	Table	Calculate Data	Generate Google Earth Data	Download Data	Backup Data	Delete Query
abdominal_delivery	2012-03-27 04:13:38						
abdominal_delivery_pourcentage	2012-03-27 04:21:06						
abdominal_delivery_pourcentage_BfA	2012-04-25 04:37:46						
abdominal_delivery_ratio	2012-03-27 04:23:53						
cancer_rsc	2012-03-27 04:38:54						

nom\_du\_laboratoire  
 Biopsie  
 etc.  
 nmographie  
 etc.  
 D ...

**FedEHR Query**

Name	Select	Group By
▼ Patient		
ID	Not Returned	
PATIENTREF	Value	
NAME	Not Returned	
FIRSTNAME	Not Returned	
DATE_OF_BIRTH	Not Returned	
SEX	Not Returned	
ZIPOCODE	Not Returned	
CITY	Not Returned	
COUNTRY	Not Returned	

Name	Date	Value
▼ RSPA		
▼ RSPA - Accouchement	Not Returned	
Dernier FCV	Not Returned	Not Returned
Date de Debut de Grossesse	Not Returned	Not Returned
Annee dernier FCV	Not Returned	Not Returned
▼ Baby		
▼ Baby	Not Returned	
Pregnancy duration	Not Returned	Not Returned
Weight	Not Returned	Not Returned

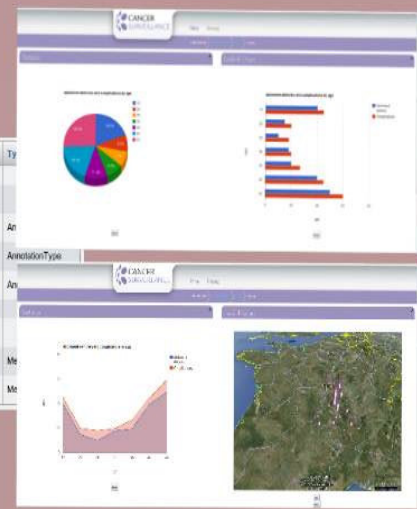
**FedEHR Query**

Step 1: Select MedicalEventTypes

MedicalEventTypes: External Web resource, DICOM\_Image\_Study, FCV, Herpes

Selected MedicalEventTypes: Baby, RSPA

Name	Date	Value	Where Condition	Where Value	Type
▼ RSPA					
▼ RSPA - Accouchement	Not Returned				
Dernier FCV	Not Returned	Not Returned	Not Filtered		An
Date de Debut de Grossesse	Not Returned	Not Returned	Not Filtered		An
Annee dernier FCV	Not Returned	Not Returned	Not Filtered		An
▼ Baby					
Pregnancy duration	Not Returned	Not Returned	Not Filtered		Me
Weight	Not Returned	Not Returned	Not Filtered		Me



\*intègre Google Earth, Maps et Visualisation

- Interface requêtage générique
- Couplage avec outils visualisation

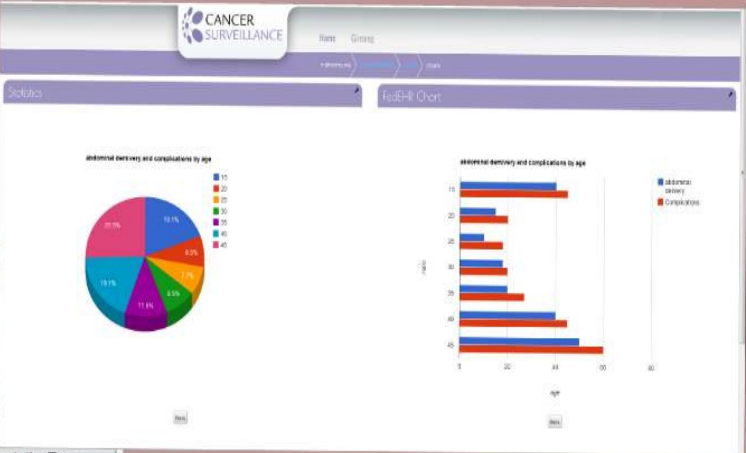
Selected MedicalEventTypes:

**Baby RSPA**

Date	Value	Where Condition	Where Value	Type
Not Returned				
Not Returned	Not Returned	Not Filtered		An
Not Returned	Not Returned	Not Filtered		An
Not Returned	Not Returned	Not Filtered		An
Not Returned				
Not Returned	Not Returned	Not Filtered		Me
Not Returned	Not Returned	Not Filtered		Me

ntType

Next



\*intègre Google Earth, Maps et Visualisation

requêtage générique  
avec outils visualisation



# Confidentialité

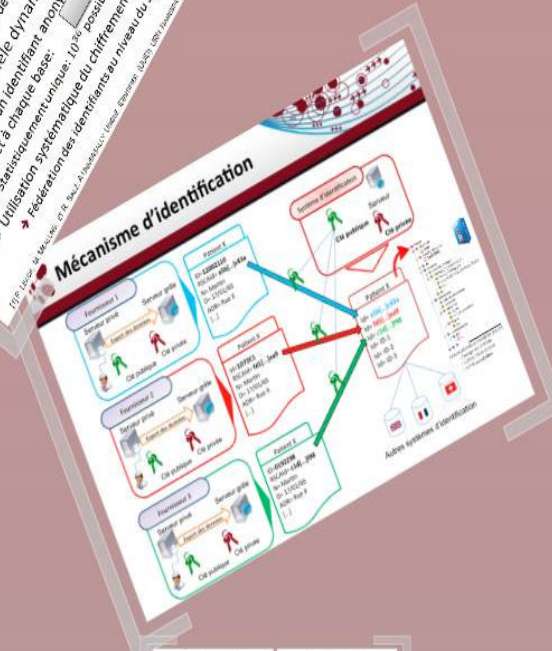
Sébastien Cipièrre (IR, LPC),  
Quentin Bisson (Stagiaire, LPC),  
Lydia Maigne (MCF, LPC)



### Identification du patient

- Problème « ouvert » dans le système de soin français
- Utilisation du numéro de sécurité sociale prohibé par la CNIL
- Adoption de l'Identifiant National de Santé Français  
1 79 09 69 877 12 56  
13 196 952 729 686 106 357
- Besoin de définir un nouveau modèle patient
- Respect des contraintes de l'ASip
- Création d'un modèle dynamique et distribué du patient
- Définition d'un modèle dynamique et distribué du patient et à chaque base:
  - Statiquement unique: 10<sup>10</sup> possibilités
  - Possibilité de l'attribuer à chaque patient
- Utilisation systématique du chiffrement de l'identifiant
- Réalisation des identifiants au niveau de serveur central

17/11/2013 - Ministère de l'Enseignement Supérieur, de la Recherche et de l'Éducation - Unité d'Enseignement Supérieur - 2008 - 070 - 002



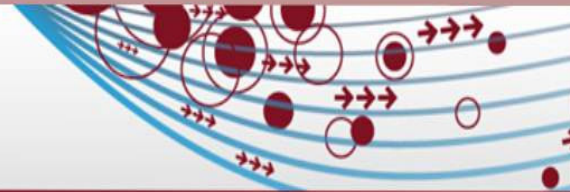
### Dispositions légales

- Demande d'autorisation déposée pour RSCA sans réponse depuis 2008
- Nouvelle demande d'autorisation déposée début mai 2013 via le CL du CNRS
  - Une demande d'autorisation / application
  - RSCA application serveur
  - RSCA application patient
- Deroulement en cours

### Rapport de patient



# Identification du patient



- Problème « ouvert » dans le système de soin français

- Utilisation du numéro de sécurité sociale prohibé par la CNIL

**1 79 03 63 877 122 56**

- Adoption de l'Identifiant National de Santé marginal

**13 196 952 729 666 105 357**

- Besoin de définir un nouveau modèle

- Respect des contraintes de l'ASIP

- Création d'un modèle dynamique et distribué du patient

- Définition d'un identifiant anonyme de type 'uuid' propre à chaque patient et à chaque base:

**162bb7e6-8fe4-4c69-854d-48c184e29825**

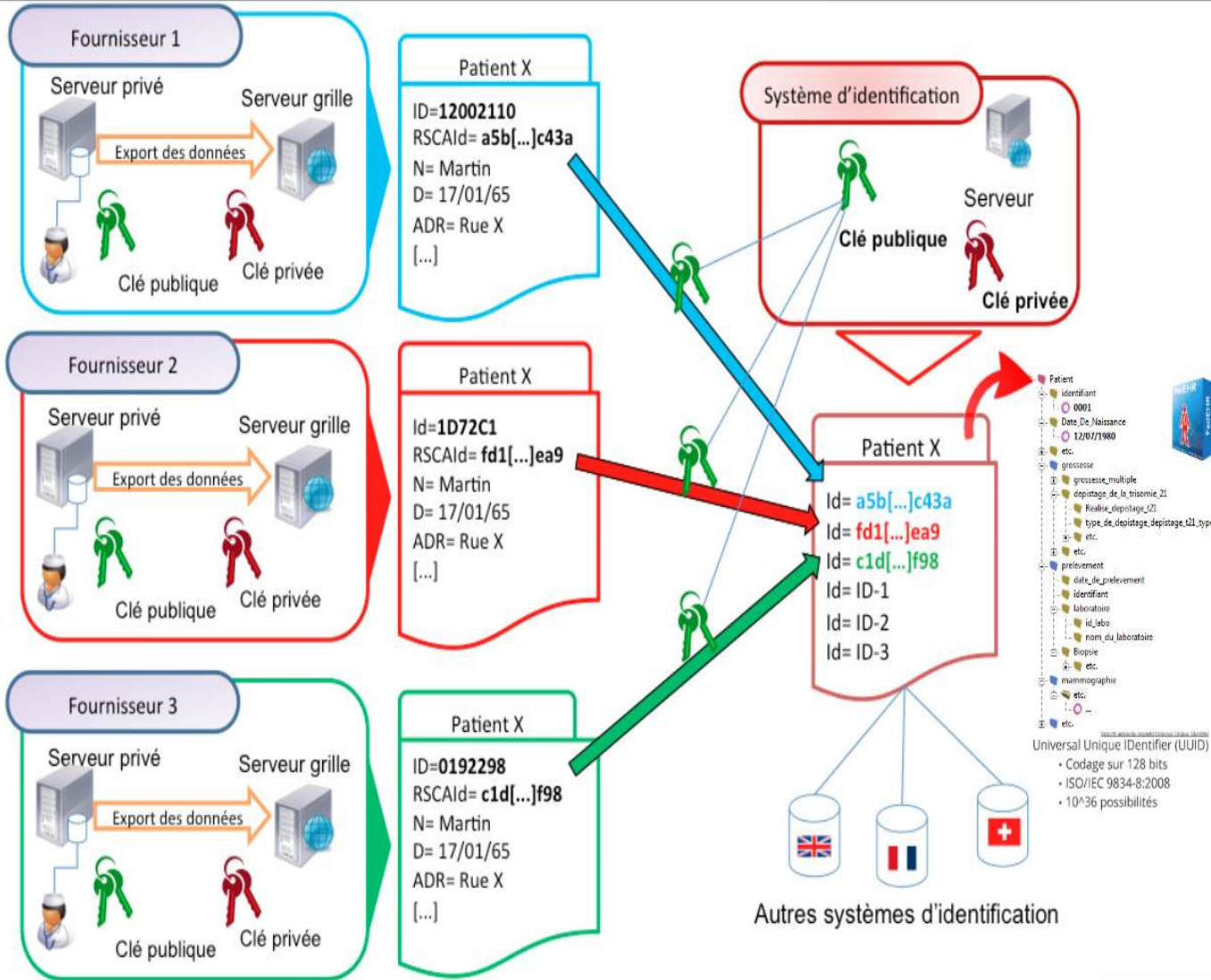
→ Statistiquement unique:  $10^{36}$  possibilités

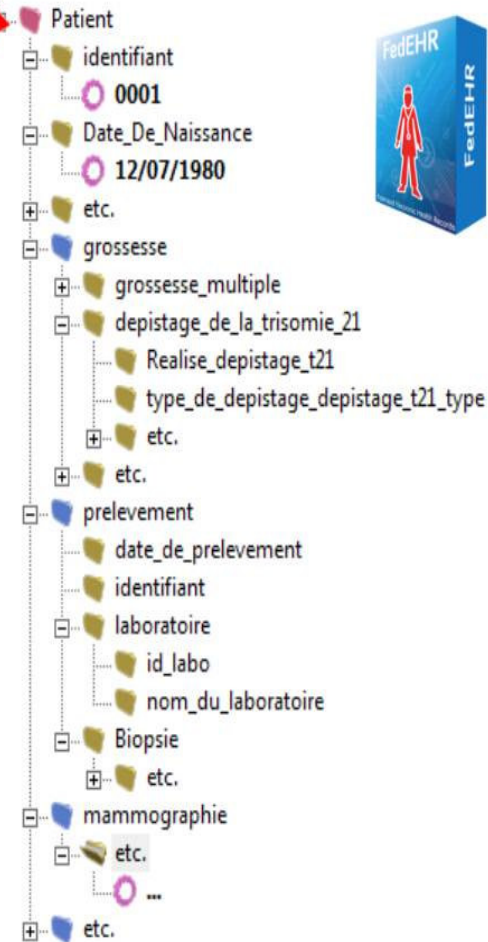
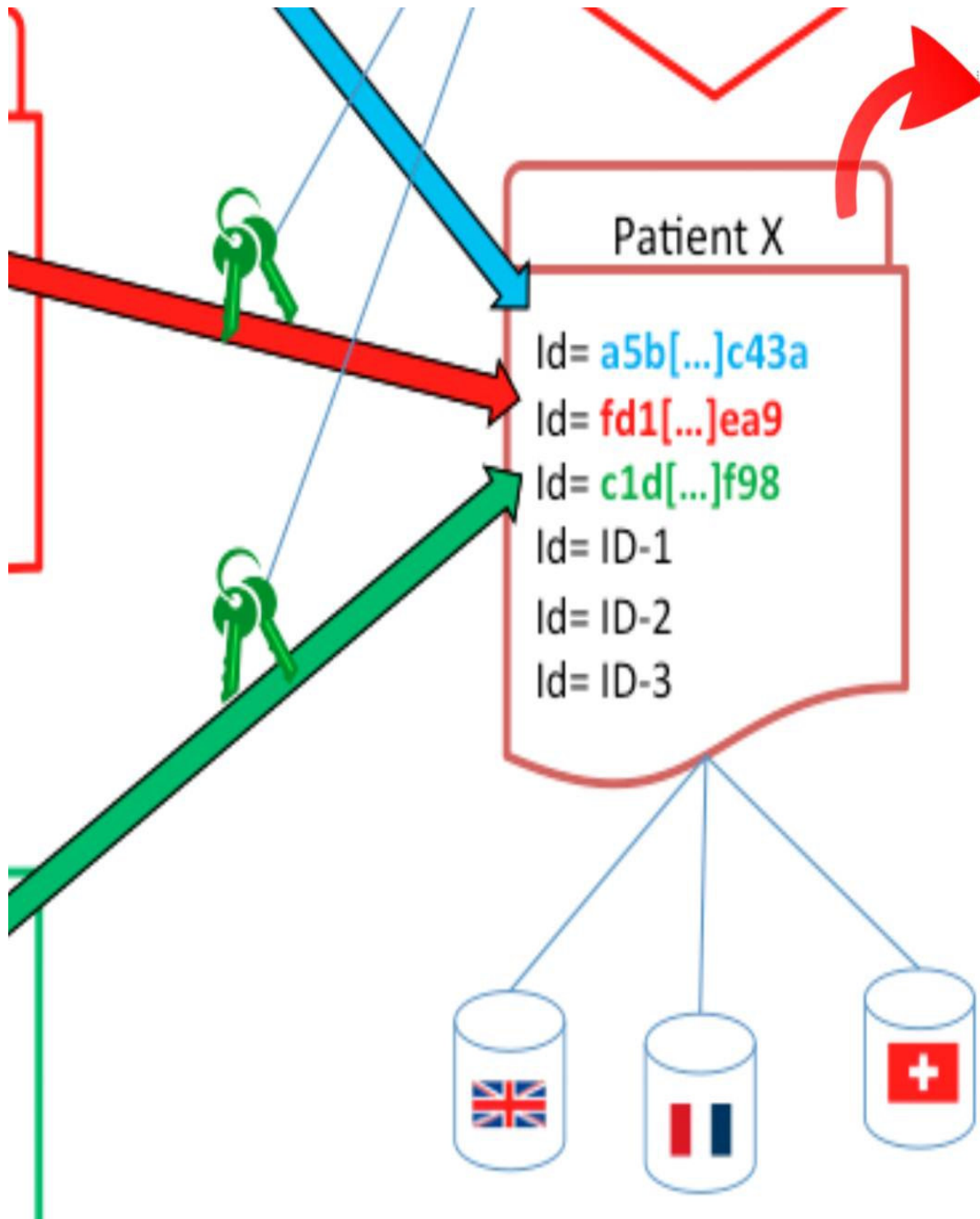
- Utilisation systématique du chiffrement de l'identifiant

→ Fédération des identifiants au niveau du serveur central



# Mécanisme d'identification



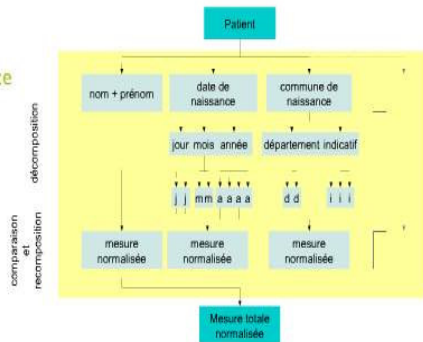


[http://fr.wikipedia.org/wiki/Universal\\_Unique\\_Identifier](http://fr.wikipedia.org/wiki/Universal_Unique_Identifier)  
 Universal Unique Identifier (UUID)

- Codage sur 128 bits
- ISO/IEC 9834-8:2008
- $10^{36}$  possibilités

## Rapprochement de patients

- Données patient
  - Nom et prénom
  - Date de naissance
  - Commune de naissance
  - Sexe
- Structure extensible
- Traitement indépendant de chaque champ
  - Point différent sur chaque information
  - Phonétique non-compatible avec les nombres
- Regroupement des résultats en une mesure



## Rapprochement de patients

- Erreur (Humaines et techniques)
  - Inversion nom et prénom
  - Prénoms composées ( Jean-Pierre <=> Jean Pierre ? )
  - Accentuation (é,ñ,ö, etc.)
  - Fautes courantes sur 1 ou 2 lettres
    - Rajout ( Paul => Paull )
    - Suppression ( Paul => Pul )
    - Inversion ( Paul => Pual )
    - Remplacement ( Paul => Pauk, clavier  
Paul => Pol, phonétique )

## Rapprochement de patients

- Mesure de similitude (exemple : Bisson Jean <=> Bisosn Jeann)
  - Analyse des caractères
    - Jaro-Winkler

□ Compte les caractères communs et les transpositions

Nom 1	B	I	S	S	O	N	Nom 1	B	I	S	S	O	N
Caractères communs	1	2	3	4	5	6	Transpositions	1	2				
Nom 2	B	I	S	O	S	N	Nom 2	B	I	S	O	S	N

□ 11 communs, 2 transpositions => 91% similaire

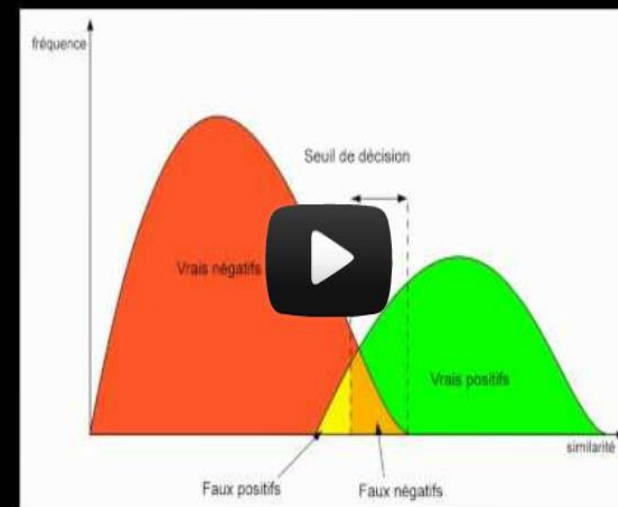
- Analyse phonétique

- Phonex

- Transforme en code phonétique et compare chiffre par chiffre

Code phonétique	an, am, en, em	oua, oi, oy	ou	in, ein, ain, eim, aim	ch, sch, ... sh
Équivalent numérique	0	1	2	3	4 ...

- Bisson Jean => Fison GIn, Bisosn Jeann => Fisonn GIn
- => 75% similaire

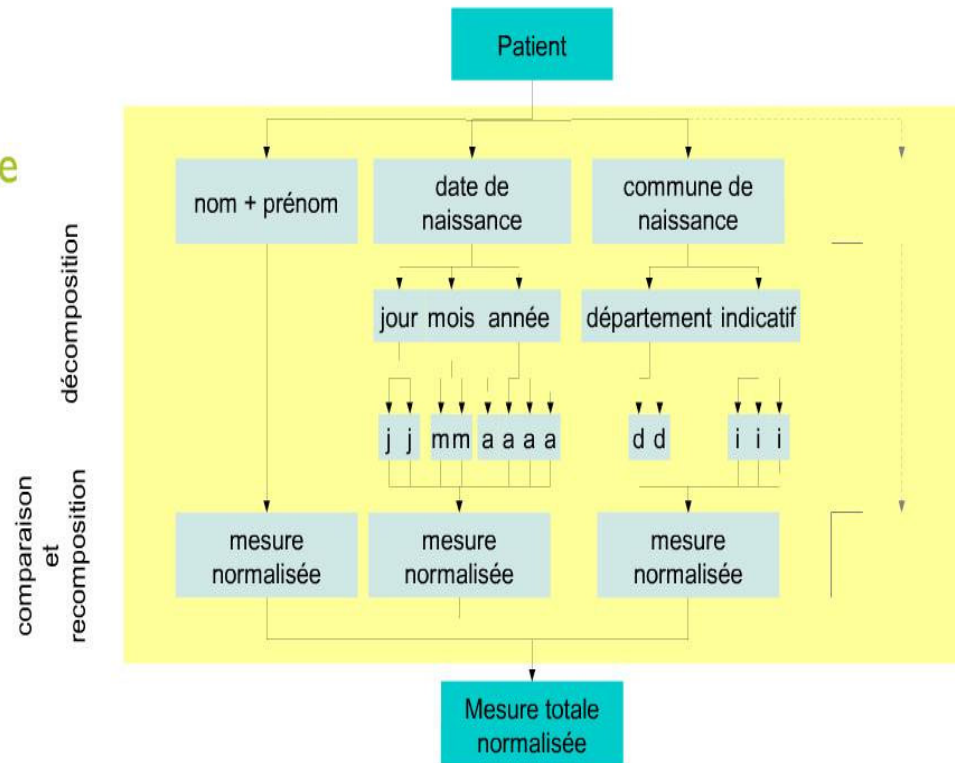


YouTube



# Rapprochement de patients

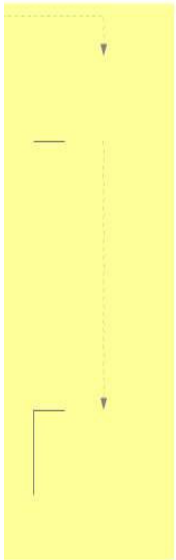
- Données patient
  - Nom et prénom
  - Date de naissance
  - Commune de naissance
  - Sexe
- Structure extensible
- Traitement indépendant de chaque champ
  - Point différent sur chaque information
  - Phonétique non-compatible avec les nombres
- Regroupement des résultats en une mesure



- Erreurs
- Ir
- P
- A
- F

# Rapprochement de patients

- Erreur (Humaines et techniques)
  - Inversion nom et prénom
  - Prénoms composées ( Jean-Pierre <=> Jean Pierre ? )
  - Accentuation (é,ñ,ö, etc.)
  - Fautes courantes sur 1 ou 2 lettres
    - Rajout ( Paul => Paull )
    - Suppression ( Paul => Pul )
    - Inversion ( Paul => Pual )
    - Remplacement ( Paul => Pauk, clavier  
Paul => Pol, phonétique )



# Rapprochement de patients

- Mesure de similitude (exemple : Bisson Jean <=> Bisosn Jeann)

- Analyse des caractères

- Jaro-Winkler

- Compte les caractères communs et les transpositions

Nom 1	B	I	S	S	O	N	Nom 1	B	I	S	S	O	N
Caractères communs	1	2	3	4	5	6	Transpositions			1	2		
Nom 2	B	I	S	O	S	N	Nom 2	B	I	S	O	S	N

- 11 communs, 2 transpositions => 91% similaire

- Analyse phonétique

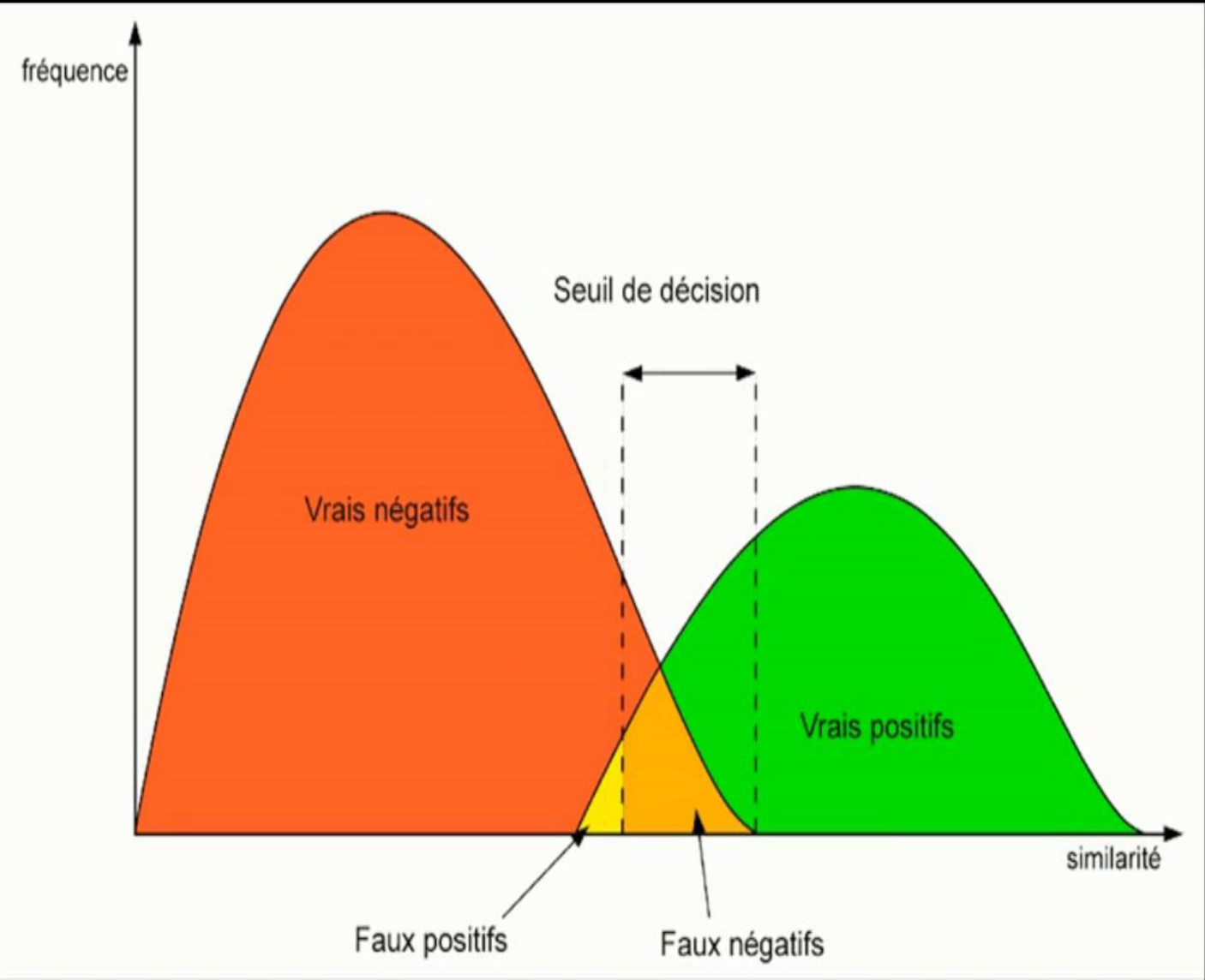
- Phonex

- Transforme en code phonétique et compare chiffre par chiffre

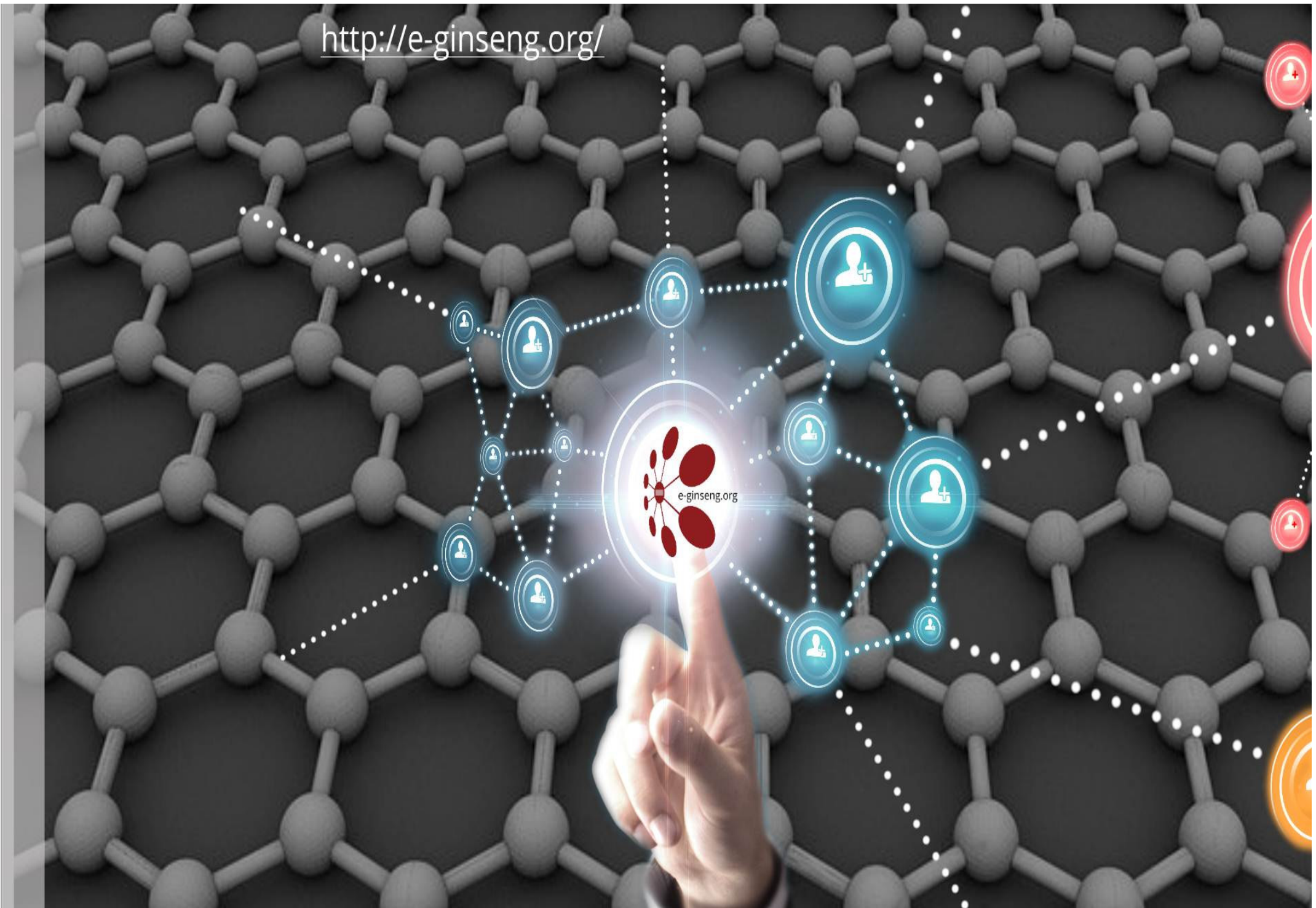
Code phonétique	an, am, en, em	oua, oi, oy	ou	in, ein, ain, eim, aim	ch, sch, sh	...
Équivalent numérique	0	1	2	3	4	...

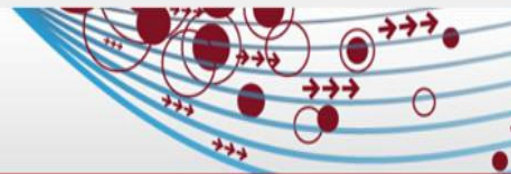
- Bisson Jean => Fison Gln, Bisosn Jeann => Fisosn Gln
        - => 75% similaire

S O N  
↔ 2 ↔  
D S N



<http://e-ginseng.org/>





## Veille Épidémiologique

- Recherche d'agrégats spatiaux/temporels
- Performance des recueils et qualité du signal

## Évaluation des pratiques

- Amélioration des performances du système dans son ensemble
- Tri et hiérarchisation des facteurs agissant sur une réalité complexe



# Biostatistiques, épidémiologie et évaluation

Stéphanie Léger (MCF, CHU Clermont-Fd, LM),  
Audrey Lelong (doctorante, CHU Clermont-Fd),  
Laurent Gerbaud (PU-PH, CHU Clermont-Fd)  
Françoise Vendittelli (PH, RS Audipog, CHU Clermont-Fd)



**Stratégie et Faisabilité**

- **Stratégie générale pour l'épidémiologie et l'évaluation**
  - Point de départ : Déterminer de données cliniques
  - Inventaire des besoins d'interopérabilité
  - Simulation d'un environnement de BDD distribuées
- **Faisabilité vis-à-vis d'aspects épidémiologiques fondamentaux**
  - **Interopérabilité par ajout d'observations**
    - Augmenter la puissance,
    - Améliorer la représentativité,
  - **Interopérabilité par ajout de variables**
    - Optimiser la validité de l'information,
    - Réduire la confusion

**Augmenter la puissance des analyses  
Améliorer la représentativité des données**

Techniquement : ajout d'observations

- **Fusion de données de structures identiques**
  - Regrouper les données de santé d'un même type d'activité
  - Exemple : maternités d'une même région...
- **L'augmentation du nombre d'observations...**
  - ...améliore la précision des estimations (IC95%)
  - ...améliore la puissance des analyses (réduction du risque  $\beta$ )
- **L'élargissement du nombre d'acteurs...**
  - ...améliore la représentativité typologique des échantillons
  - ...améliore la représentativité géographique des échantillons

**Optimiser la validité des informations  
Réduire le phénomène de confusion**

Techniquement : ajout de variables

- **Concaténation de données de structures différentes**
  - Regrouper les données de santé d'acteurs d'une même filière
  - Exemple : centre de dépistage et laboratoires de pathologie...
  - Compléter les données de niveau par d'autres données informatives (biologie de la naissance, compte rendu d'hospitalisation formalisé)
- **L'augmentation du nombre de champs...**
  - ...améliore la validité des informations par recoupements
  - ...enrichit l'analyse de nouveaux facteurs, y compris de confusion

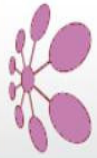
**Applications: Péri natalité et  
Dépistage des cancers**

- **Pratique de la césarienne**
  - Parmi les mères de la région Auvergne
  - Étude des facteurs de risque de césarienne
  - Facteurs au facteur « centre » et « filière »
  - Acquisition des données ICDS : Compte rendu d'hospitalisation
- **Dépistage de la trisomie 21**
  - ISVA + laboratoires d'analyses médicales + échographistes
  - Mixte à disposition mutuelle des données en temps réel
- **Couverture du dépistage du cancer du sein**
  - ARDCC - AIOCC + laboratoires de cytologie
  - Évaluer les déterminants de participation aux étapes
- **Pratique du frottis CV (FCV) et suivi de grossesse**
  - ISVA + laboratoires de cytologie
  - Distinuer le caractère de la cervix FCV
  - Passer d'une information descriptive à une information réelle

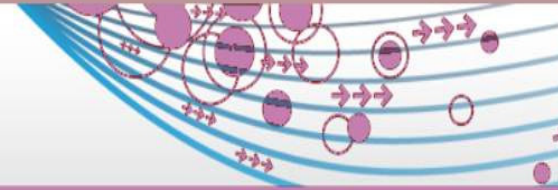


Lemlih Ouchchane (MCU-PH, ISIT-UA), Aline Guttmann (AHU, ISIT-UA),  
Xinran Li (doctorant, ISIT-UA), Héliène Laurichesse (PH, DPN),  
Philippe Chatron (DG, GenBio)





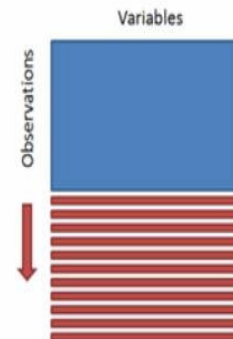
# Stratégie et Faisabilité



- **Stratégie générique pour l'épidémiologie et l'évaluation**
  - Point de départ : Détenteurs de données cliniques
  - Inventaire des besoins d'interopérabilité
  - Simulation d'un environnement de BDD distribuées
- **Faisabilité vis-à-vis d'aspects épidémiologiques fondamentaux**

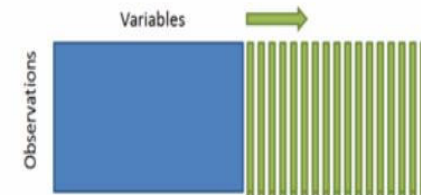
- **Interopérabilité par ajout d'observations**

1. Augmenter la puissance,
2. Améliorer la représentativité,

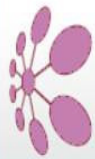


- **Interopérabilité par ajout de variables**

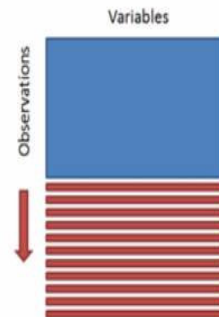
3. Optimiser la validité de l'information,
4. Réduire la confusion







# Augmenter la puissance des analyses Améliorer la représentativité des données



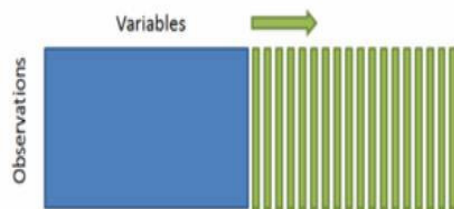
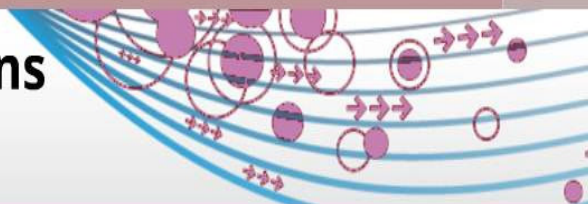
Techniquement : ajout  
d'observations

- **Fusion de données de structures identiques**
  - Regrouper les données de santé d'un même type d'activité
  - Exemple : maternités d'une même région,...
- **L'augmentation du nombre d'observations...**
  - ...améliore la précision des estimations (IC95%)
  - ...améliore la puissance des analyses (réduction du risque  $\beta$ )
- **L'élargissement du nombre d'acteurs...**
  - ...améliore la représentativité typologique des échantillons
  - ...améliore la représentativité géographique des échantillons



# Optimiser la validité des informations

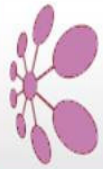
## Réduire le phénomène de confusion



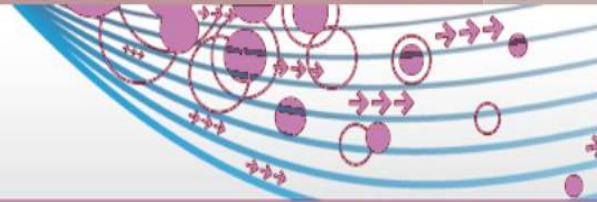
Techniquement : ajout de variables

- **Concaténation de données de structures différentes**
  - Regrouper les données de santé d'acteur d'une même filière
  - Exemple : centre de dépistage et laboratoires de pathologie,...
  - Compléter les données de réseau par d'autres données informatisées (biologie de la naissance, compte rendu d'hospitalisation formalisé)
- **L'augmentation du nombre de champs...**
  - ...améliore la validité des informations par recoupement
  - ...enrichit l'analyse de nouveaux facteurs, y compris de confusion

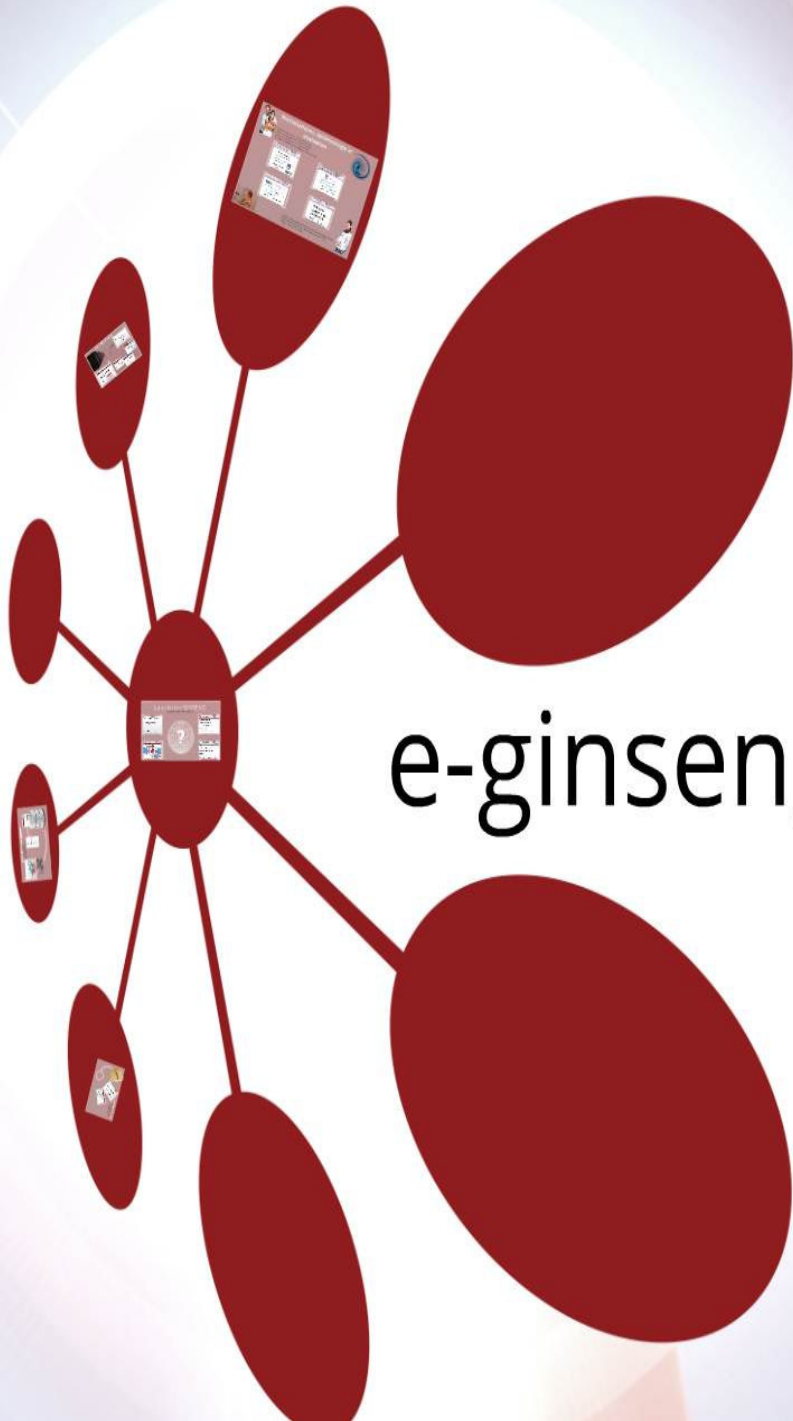




## Applications: Péri natalité et Dépistage des cancers

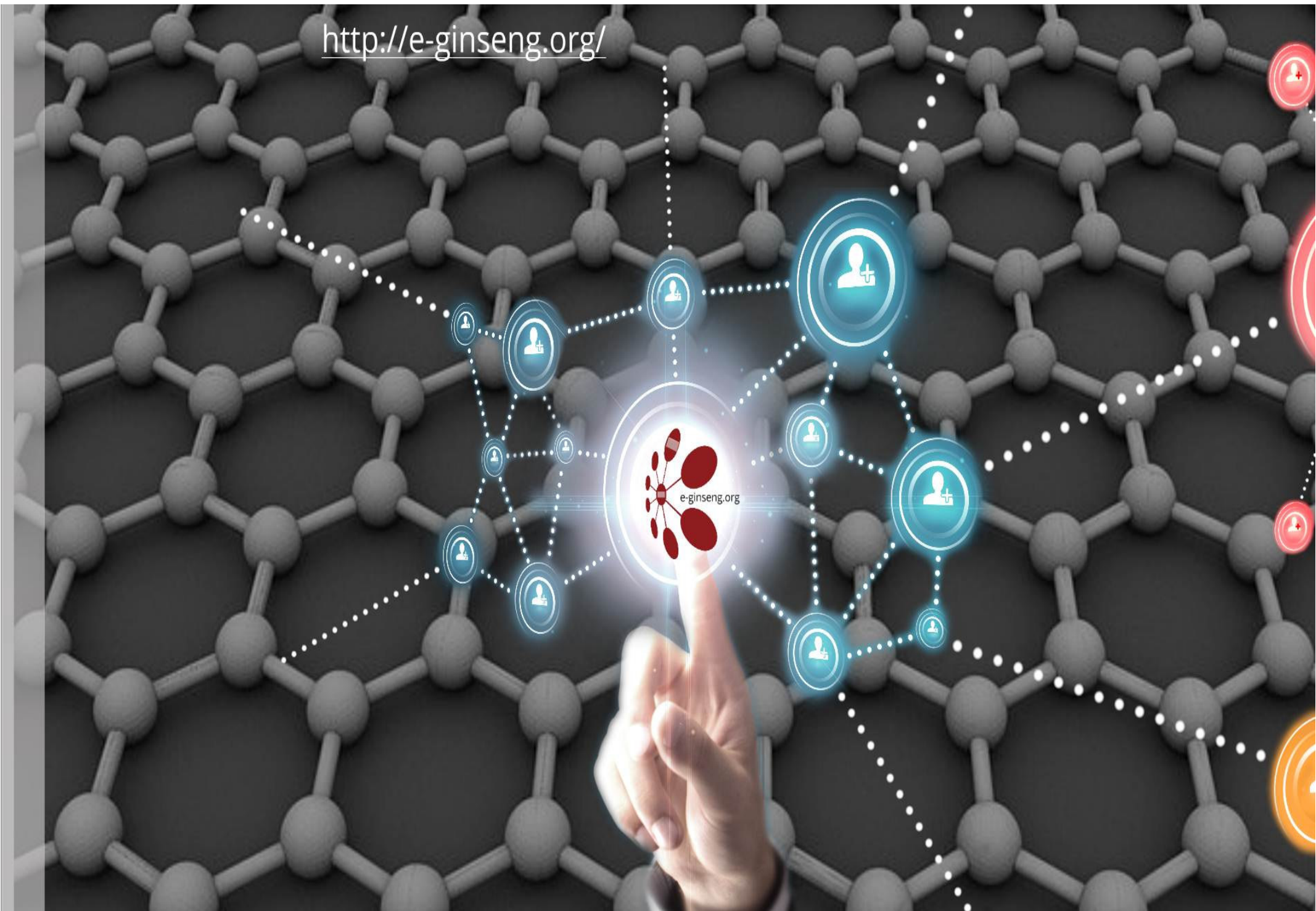


- **Pratique de la césarienne**
  - Parmi les membres du RSPA (maternités de la région Auvergne)
  - Étude des facteurs de risque de césarienne
  - Extension au facteur « centre » et « filière »
  - Acquisition des données ICOS – Compte rendu d'hospitalisation
- **Dépistage de la trisomie 21**
  - RSPA + laboratoires d'analyses médicales + échographistes
  - Mise à disposition mutuelle des données en temps réel
- **Couverture du dépistage du cancer du sein**
  - ARDOC + ABIDEC + laboratoires de cyto-pathologie
  - Évaluer les déterminants de participation aux dépistages
- **Pratique du frottis CV (FCV) et suivi de grossesse**
  - RSPA + laboratoires de cyto-pathologie
  - Optimiser la validité de la date du FCV
  - Passer d'une information déclarative à une information réelle



e-ginseng.org

<http://e-ginseng.org/>





<http://e-ginseng.org/>

