



# Quelques nouveautés du domaine de la E-Santé

**Jean Charlet**

AP-HP et INSERM U1142

En collaboration avec J. Boudy et S.J. Darmoni



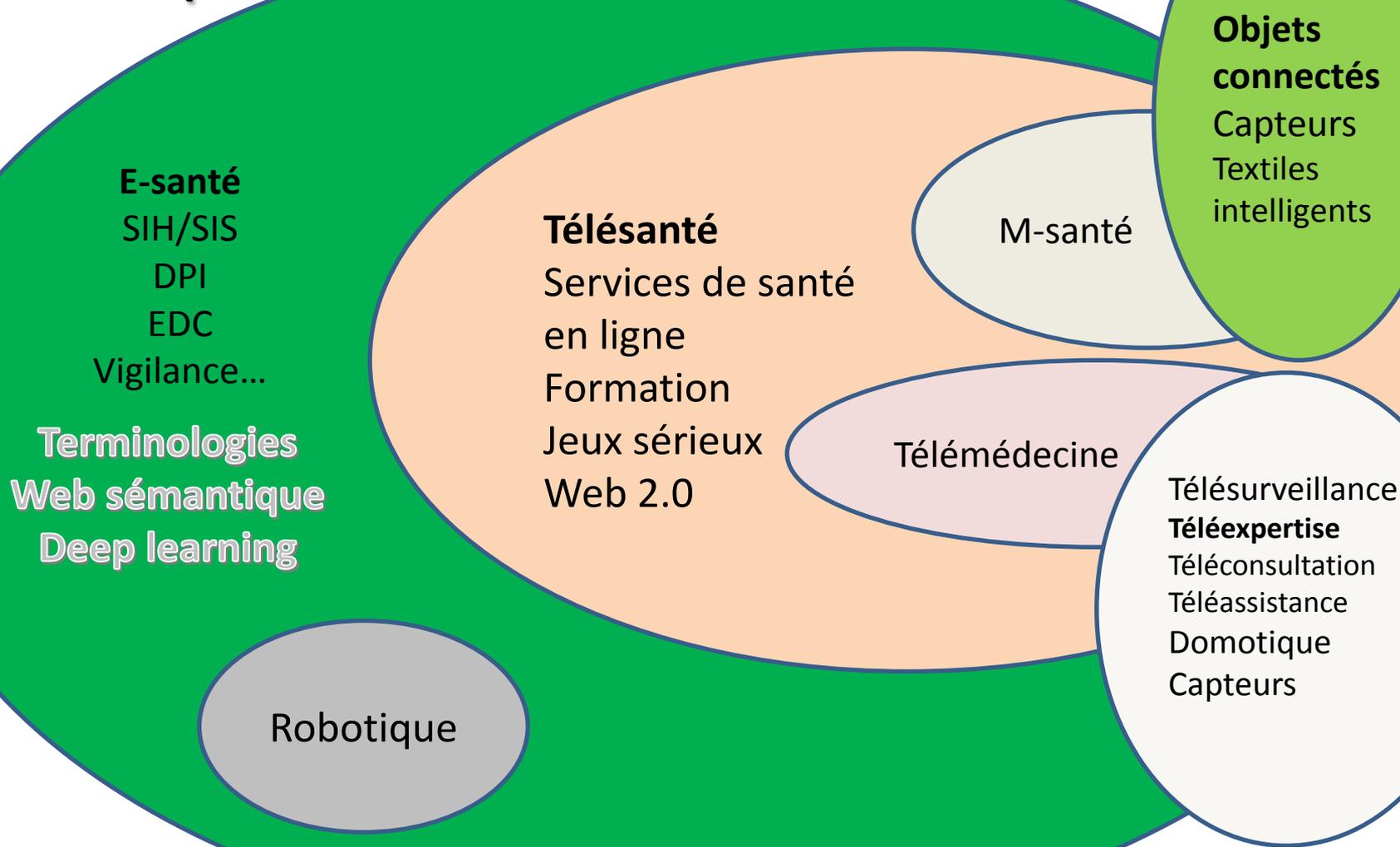
# Plan

- Rappels
- Sur le Web
  - Hype Cycle curve
  - **Watson**
- Un rapide **cours d'IA**
- Une analyse des performances de Watson et des propositions de travail
- Et pendant ce temps





# eSanté (CNOM)





# Définition et périmètre de la E-Santé

- E-santé (e-health en anglais) désigne
  - tous les aspects numériques touchant de près ou de loin la santé
  - correspond à du contenu numérique lié à la santé, appelé également la santé électronique ou télésanté
  - Recouvrement (total ?) avec le bien-être
- Nombreux domaines couverts par la E-Santé:
  - la télémédecine
  - le maintien à domicile, la prévention
  - le suivi d'une maladie chronique à distance (diabète, hypertension, insuffisance cardiaque ...)
  - les dossiers médicaux électroniques
  - Les Systèmes d'Information de santé et les entrepôt de données cliniques (EDC)
  - la domotique
  - la création de textiles intelligents



# E-Santé : définition et moyens techniques

- E-santé (e-health) désigne
  - tous les aspects numériques touchant de près ou de loin la santé
  - correspond à du contenu numérique lié à la santé, appelé également la santé électronique ou télesanté
  - Recouvrement (total ?) avec le bien-être
- La e-Santé concerne
  - L'acquisition (y compris à distance) de données hétérogènes
  - L'alimentation de bases de données interopérables
  - La représentation des connaissances (dictionnaires, terminologies, taxonomies, ontologies)
  - Le développement de SIAD
  - La fouille de données
  - L'intégration de données de toutes modalités
  - La fouille de texte...
  - **Beaucoup de domaines liés à l'Intelligence Artificielle**

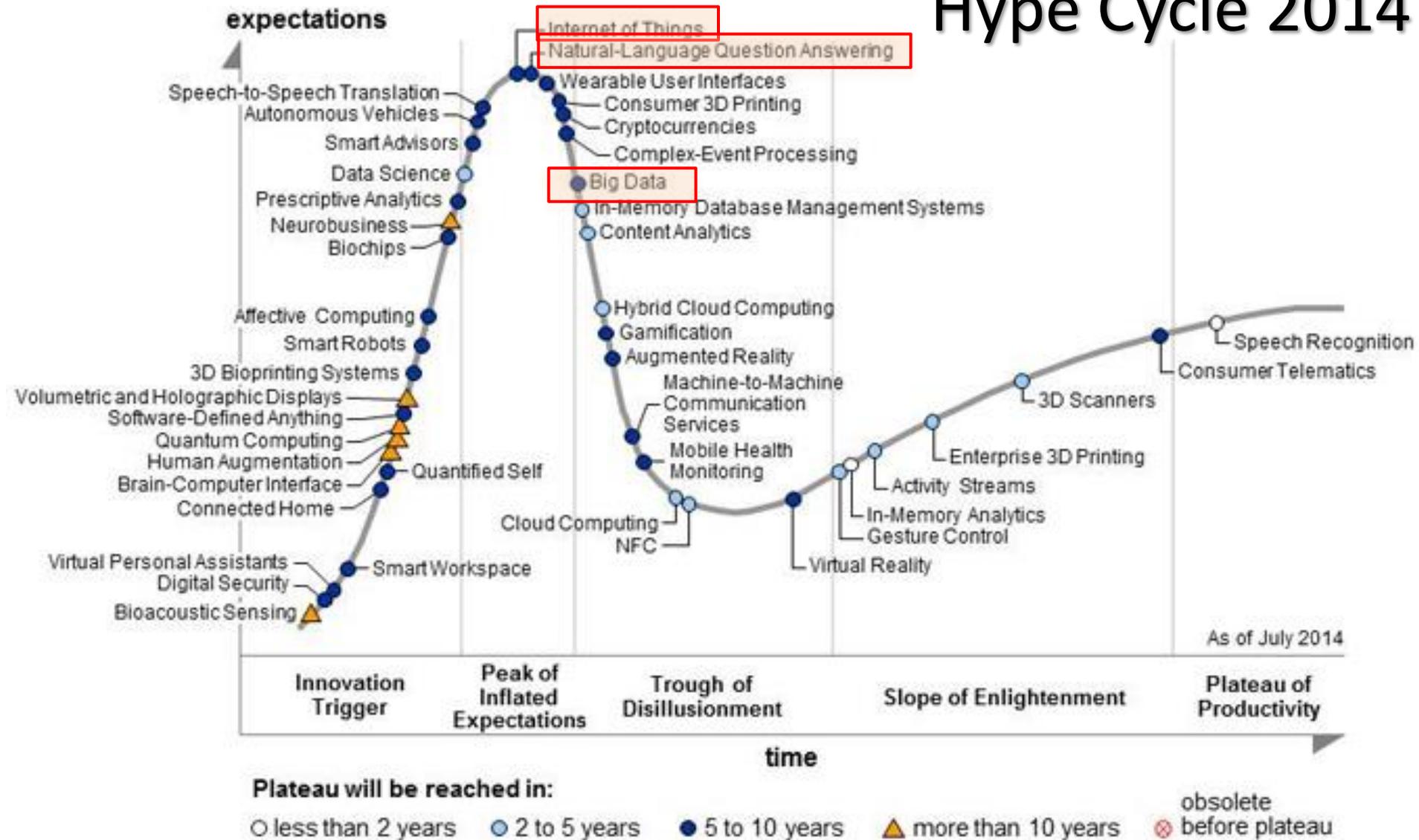


# Watson et le cycle *de la Hype*

- [https://fr.wikipedia.org/wiki/Cycle\\_du\\_hype](https://fr.wikipedia.org/wiki/Cycle_du_hype)
- Le Big Data ?
- Deep Learning = IA ?
- Et IBM/Watson ...

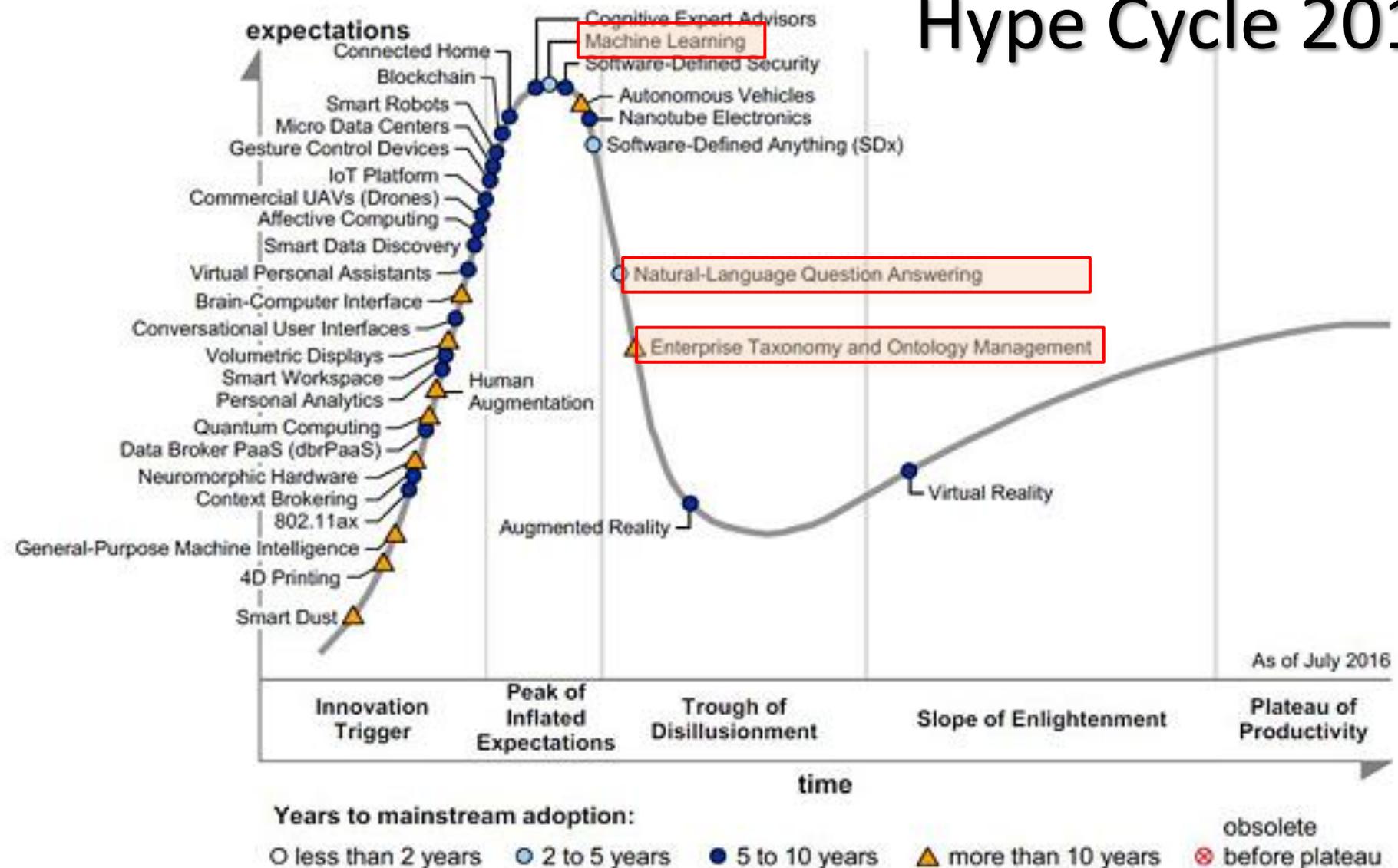


# Hype Cycle 2014





# Hype Cycle 2016



Source: Gartner (July 2016)



ibm watson



Tous Actualités Vidéos Images Shopping Plus Paramètres Outils

Environ 18 100 000 résultats (0,66 secondes)

### IBM Watson - Vous À La Puissance IBM - ibm.com

[Annonce](#) [www.ibm.com/](http://www.ibm.com/)

Osez L'Innovation Avec L'Intelligence Artificielle Pour Les Entreprises  
IBM Cloud · IBM Systems · Big Data & Analytics · Les Enjeux de la Fraude

### IBM Watson

<https://www.ibm.com/watson/> Traduire cette page

Watson is the AI platform for business.

Watson Conversation · Watson Discovery · Watson blog · Watson Media

### IBM Watson Talent - France

<https://www.ibm.com/watson/fr-fr/talent/>

Watson Talent étend l'expertise humaine, ce qui permet aux RH d'améliorer et d'accélérer l'impact des personnes sur l'entreprise.

### Watson (intelligence artificielle) — Wikipédia

[https://fr.wikipedia.org/wiki/Watson\\_\(intelligence\\_artificielle\)](https://fr.wikipedia.org/wiki/Watson_(intelligence_artificielle))

Watson est un programme informatique d'intelligence artificielle conçu par IBM dans le but de répondre à des questions formulées en langage naturel.

Participation à Jeopardy! · Applications actuelles et ... · Notes et références

### Watson : comment marche l'IA d'IBM dans la santé, la banque... - JDN

[www.journaldunet.com](http://www.journaldunet.com) Web & Tech Entreprises

26 oct. 2017 - Watson, l'intelligence artificielle d'IBM, est capable de répondre aux questions en langage naturel que lui posent ses utilisateurs dans de ...

### A la rencontre de Watson, l'intelligence artificielle star d'IBM - Le Monde

[www.lemonde.fr/.../a-la-rencontre-de-watson-l-intelligence-artificielle-star-d-ibm\\_50933...](http://www.lemonde.fr/.../a-la-rencontre-de-watson-l-intelligence-artificielle-star-d-ibm_50933...)

12 mars 2017 - Watson peut composer de la musique, conseiller des médecins ou même tenter d'évaluer l'humeur de son interlocuteur. Zoom sur cette ...

### Comment, avec son programme Watson, IBM veut changer le monde

[www.lemonde.fr/economie/article/.../watson-le-savant-calcul-d-ibm\\_5194573\\_3234.htm...](http://www.lemonde.fr/economie/article/.../watson-le-savant-calcul-d-ibm_5194573_3234.htm...)

1 oct. 2017 - Transport, santé, relation client... L'intelligence artificielle promet de changer le monde. Watson, le programme phare de Big Blue, peine à ...

**IBM Watson**  
Intelligence artificielle

Watson est un programme informatique d'intelligence artificielle conçu par IBM dans le but de répondre à des questions formulées en langage naturel. Il s'intègre dans un programme de développement plus vaste, le DeepQA research project. [Wikipédia](#)

Recherches associées

- Deep Blue
- Ordinateur central
- Raspberry Pi
- IBM PC
- ThinkPad

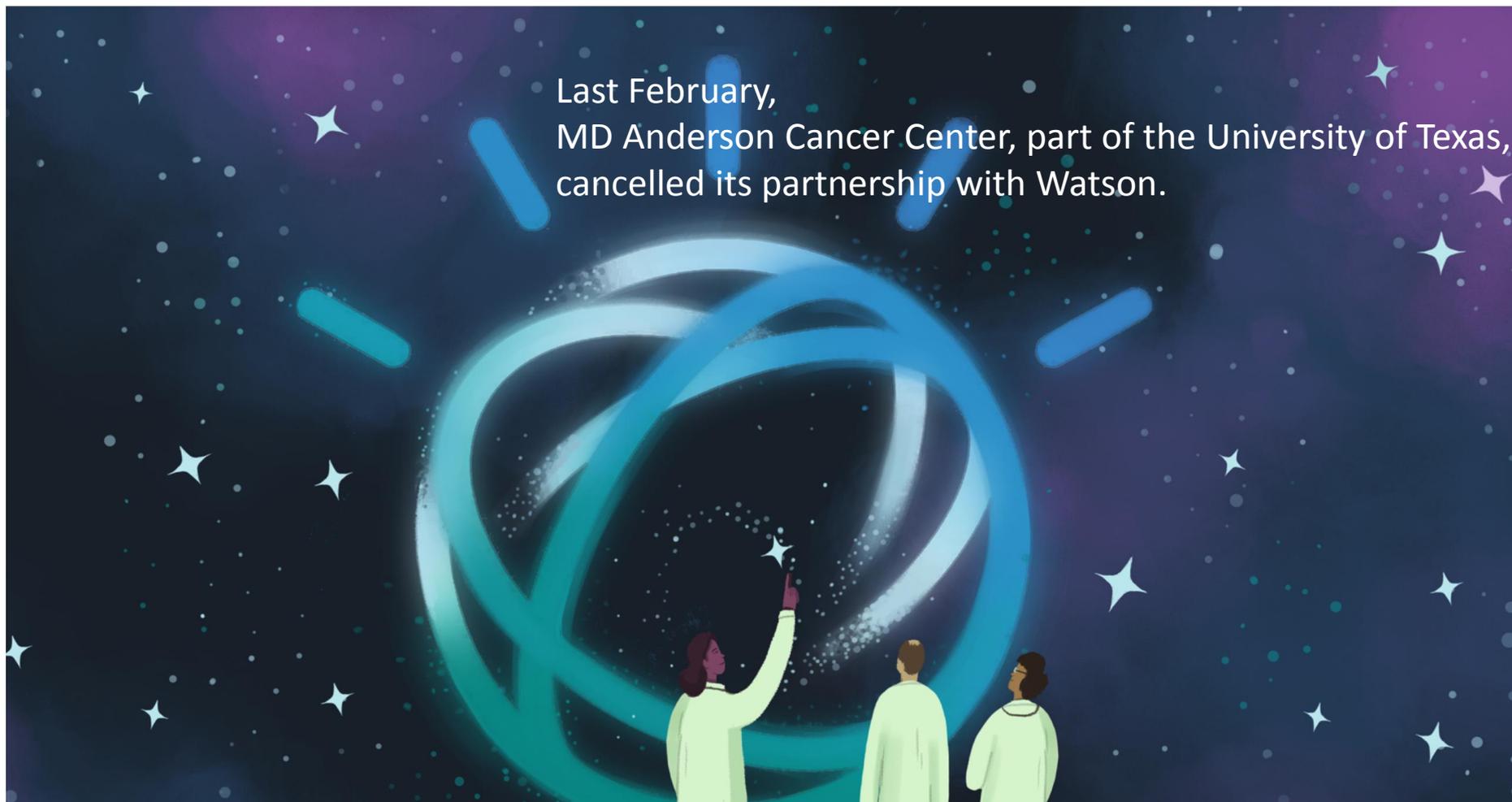
Commentaires



A STAT INVESTIGATION

## IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close

By CASEY ROSS @caseymross and IKE SWETLITZ @ikaswetlitz / SEPTEMBER 5, 2017



<https://www.statnews.com/2017/09/05/watson-ibm-cancer/>



# M Économie

ÉCONOMIE

Les données du « Monde »

Économie mondiale

Économie française

Entreprises

Emploi

A

ÉDITION  
ABONNÉS



ARTICLE SÉLECTIONNÉ DANS LA MATINALE DU  
01/10/2017

> Découvrir l'application

## Comment, avec son programme Watson, IBM veut changer le monde

Transport, santé, relation client... L'intelligence artificielle promet de changer le monde. Watson, le programme phare de Big Blue, peine à relever le défi de l'industrialisation.

LE MONDE ÉCONOMIE | 02.10.2017 à 06h39 • Mis à jour le 02.10.2017 à 15h11 |

Par Chloé Hecketsweiler et Sandrine Cassini

Abonnez vous à partir de 1 €

Réagir

★ Ajouter



f Partager

Twitter



Kendall Square, c'est un peu la Silicon Valley de la Côte est américaine. Ce quartier de Cambridge – un confetti de 2 kilomètres carrés – est le royaume du Massachusetts Institute of Technology (MIT), l'une des Mecque de la recherche aux États-Unis.

[http://www.lemonde.fr/economie/article/2017/10/02/watson-le-savant-calcul-d-ibm\\_5194573\\_3234.html?xtmc=watson&xtcr=13](http://www.lemonde.fr/economie/article/2017/10/02/watson-le-savant-calcul-d-ibm_5194573_3234.html?xtmc=watson&xtcr=13)



# L'“IA” Watson, le pour et le contre

## *Pour*

- Trouve les articles pertinents
- Propose des protocoles de soins connus/prévus
- Affiche toutes les options pour discuter

## *Contre*

- Incapable de proposer un bon protocole en cas d'échec du premier
- En accord avec les professionnels dans 33% des cas (Danemark)
- Ne comprend pas bien les dossiers médicaux
- A du mal à contextualiser les propositions (par pays)
- Impossible, à ce jour, d'écrire un article médical démontrant l'avantage d'utiliser Watson

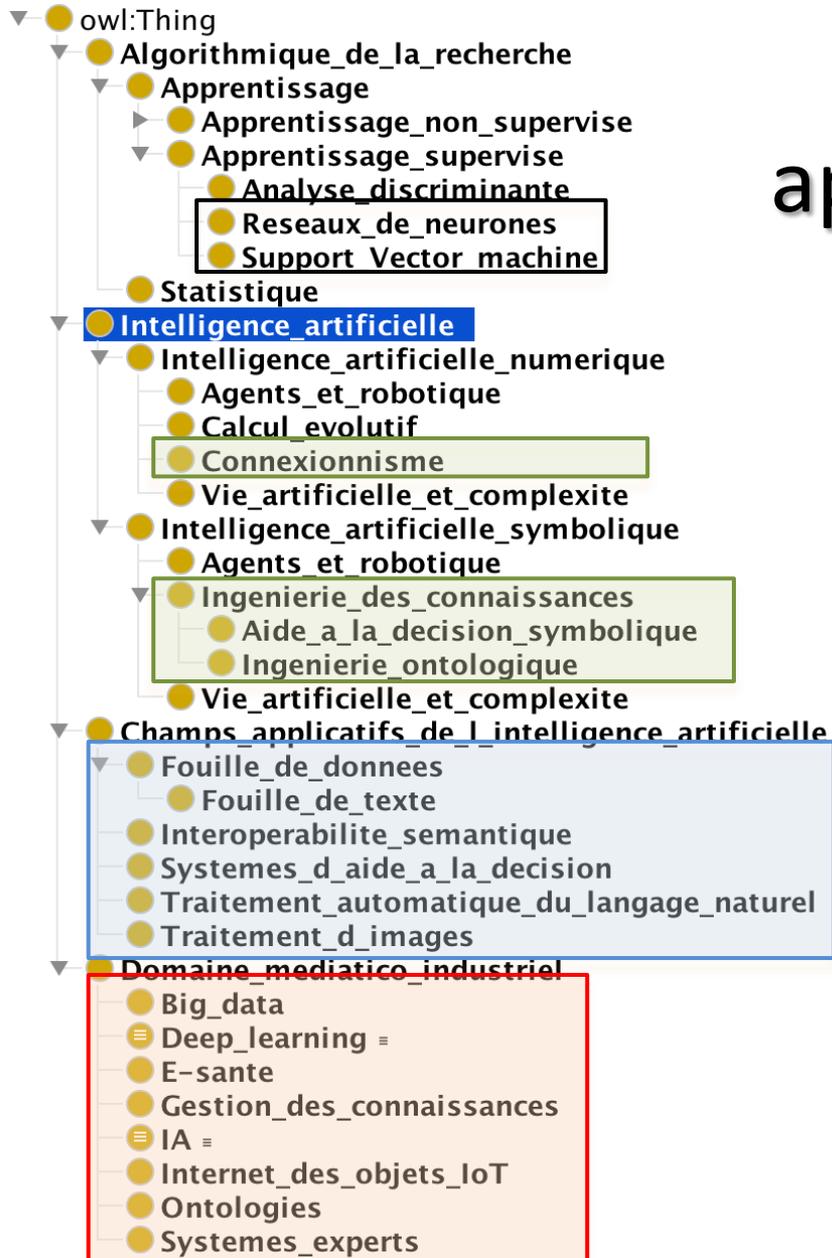


# Watson et le cycle *de la Hype*

- [https://fr.wikipedia.org/wiki/Cycle\\_du\\_hype](https://fr.wikipedia.org/wiki/Cycle_du_hype)
- Où est parti le Big Data ?  
Dans le paquet des buzz inutiles ?
- Deep Learning = IA ?
- Et IBM/Watson fait peut-être des choses bien ?
  
- Il y a des modes médiatiques... et pendant ce temps on travaille



# Un petit cours d'I.A. appliqué à la médecine (1)



- Dans les journaux, le Web
- Les domaines qu'aborde l'IA
- Les sous-disciplines de l'IA
- Les algorithmes



# Un petit cours d'I.A. (2)

		ALL		
	Team	P	R	F
Official runs	SIBM-run1	<b>.857</b>	<b>.689</b>	<b>.764</b>
	LITL-run2	.666	.414	.510
	LIRMM-run1	.541	.480	.509
	LIRMM-run2	.540	.480	.508
	LITL-run1	.651	.404	.499
	TUC-MI-run2	.044	.026	.033
	TUC-MI-run1	.025	.015	.019
	<b>average</b>	.475	.358	.406
	<b>median</b>	.541	.414	.508
Non-official	LIMSI-run2	.872	.784	.825
	LIMSI-run1	.883	.760	.817
	TUC-MI-run1-corrected	.883	.539	.669
	TUC-MI-run2-corrected	.882	.536	.667
	UNIPD-run1	.629	.468	.537
	UNIPD-run2	.518	.384	.441
	Mondeca-run1	.375	.131	.194
	Frequency baseline	.339	.237	.279

- Retrouver des codes de maladies CIM-10 dans les textes des certificats de décès du CépiDc (et aussi sur un échantillon anglais)
- Un cas théoriquement *très* facile
- VP, FP, VN, FN, Précision (VP/(VP+FP)), rappel (VP/(VP+FN)), F-mesure (moyenne harmonique(Rappel, Précision))
- Sur cet exercice, les équipes françaises sont très bonnes dans les 2 langues (fr, en) mais spécialement bonnes en français (cette figure)
- Les meilleures équipes utilisent des techniques mixtes d'annotation sémantique et d'apprentissage (svm)

CLEF eHealth 2017 Multilingual Information Extraction task overview: ICD10 coding of death certificates in English and French  
<https://pdfs.semanticscholar.org/7726/a6eff024adee59c1bf21d88f5a0f75239f29.pdf>



# Watson, qu'est-ce que c'est ? (1/2)

- Un programme utilisant des techniques de TALN et de fouille de données
- Un programme entraîné par des être humains donnant les « gold standards » des traitements (heureusement ?)

*It's taken nearly six years of painstaking work by data engineers and doctors to train Watson in just seven types of cancer, and keep the system updated with the latest knowledge*

- Fouillant les données de l'hôpital sur les patients



# Watson, qu'est-ce que c'est ? (2/2)

- Un programme fouillant aussi (tous) les textes de l'hôpital pour *apprendre* des connaissances et pour *retrouver* d'autres données sur les patients
- Utilisant des technologies principalement numériques (et des technologies symboliques ?) pour le TALN
- Utilisant des technologies numériques pour la fouille de données (connexionnisme)
- Dans un contexte médiatique fort : le *Digital Canary* de l'IA en médecine



# Watson, un échec ?

- Quelles que soient les méthodes, pour une annotation complexe, une F-mesure de 0,7 est
- Si dans les textes médicaux, il y a des imprécisions ou des sources de confusion, il est impossible de les corriger
- **Aujourd'hui**, une analyse correcte ( $> 0.9$ ) des éléments textuels d'un dossier standard est impossible sans intervention humaine
- Et IBM/Watson fait des choses **surement très bien... mais pas tout à fait ce qu'ils ont (sur)vendu**

Episode Dépressif Caractérisé  
Every Day Carry  
Entrepôt de Données Cliniques  
Electronic Data Capture



# Des propositions

- **Continuer les recherches**
  - Approches mixtes, ontologies, apprentissage, etc.
- Faire intervenir les **professionnels de santé** dans l'annotation sémantique des textes (sur de petits échantillons – EDC)
- **Communiquer** sur la modélisation, les processus de TALN et d'apprentissage auprès de ceux qui écrivent les textes

## *Pour*

- Une meilleure structuration des données en général
- Une amélioration du format et de l'expression et de la véracité des textes

® Meilleurs reconnaissance et résultat



# Et pendant ce temps...



# KRM (Yearbook of MI 2017)

- Une « review » des articles pubmed en représentation et gestion des connaissances<sup>1</sup>
- L'**intégration** des données avec l'annotation des textes (~80% de l'inf. méd.) fondée sur des terminologies/ontologies commence à donner des résultats
- Des articles montrent des efforts de modélisation<sup>2</sup>
- Mais dès qu'il faut traiter des textes, l'intervention humaine est nécessaire<sup>3</sup>

® sur des échantillons à taille humaine

<sup>1</sup>Dhombres, Charlet, PMID: [29063556](https://pubmed.ncbi.nlm.nih.gov/29063556/)

<sup>2</sup>Hochheiser et al., PMID: [27629872](https://pubmed.ncbi.nlm.nih.gov/27629872/)

<sup>3</sup>Sarntivijai, PMID: [27011785](https://pubmed.ncbi.nlm.nih.gov/27011785/)



RESEARCH ARTICLE

Open Access

# An information model for computable cancer phenotypes



Harry Hochheiser<sup>1,2\*</sup> , Melissa Castine<sup>1</sup>, David Harris<sup>3</sup>, Guergana Savova<sup>3</sup> and Rebecca S. Jacobson<sup>1,2,4</sup>

## Abstract

**Background:** Standards, methods, and tools supporting the integration of clinical data and genomic information are an area of significant need and rapid growth in biomedical informatics. Integration of cancer clinical data and cancer genomic information poses unique challenges, because of the high volume and complexity of clinical data, as well as the heterogeneity and instability of cancer genome data when compared with germline data. Current information models of clinical and genomic data are not sufficiently expressive to represent individual observations and to aggregate those observations into longitudinal summaries over the course of cancer care. These models are acutely needed to support the development of systems and tools for generating the so called clinical “deep phenotype” of individual cancer patients, a process which remains almost entirely manual in cancer research and precision medicine.

**Methods:** Reviews of existing ontologies and interviews with cancer researchers were used to inform iterative development of a cancer phenotype information model. We translated a subset of the Fast Healthcare Interoperability Resources (FHIR) models into the OWL 2 Description Logic (DL) representation, and added extensions as needed for modeling cancer phenotypes with terms derived from the NCI Thesaurus. Models were validated with domain experts and evaluated against competency questions.

**Results:** The DeepPhe Information model represents cancer phenotype data at increasing levels of abstraction from mention level in clinical documents to summaries of key events and findings. We describe the model using breast cancer as an example, depicting methods to represent phenotypic features of cancers, tumors, treatment regimens, and specific biologic behaviors that span the entire course of a patient’s disease.

**Conclusions:** We present a multi-scale information model for representing individual document mentions, document level classifications, episodes along a disease course, and phenotype summarization, linking individual observations to high-level summaries in support of subsequent integration and analysis.

**Keywords:** Cancer, Deep phenotyping, Information extraction, Information model



RESEARCH

Open Access



# Linking rare and common disease: mapping clinical disease-phenotypes to ontologies in therapeutic target validation

Sirarat Sarntivijai<sup>1,2\*</sup> , Drashtti Vasant<sup>1,2</sup>, Simon Jupp<sup>1</sup>, Gary Saunders<sup>1,2</sup>, A. Patrícia Bento<sup>1,2</sup>, Daniel Gonzalez<sup>1,2</sup>, Joanna Betts<sup>2,3</sup>, Samiul Hasan<sup>2,3</sup>, Gautier Koscielny<sup>2,3</sup>, Ian Dunham<sup>1,2</sup>, Helen Parkinson<sup>1</sup> and James Malone<sup>1,2</sup>

## Abstract

**Background:** The Centre for Therapeutic Target Validation (CTTV - <https://www.targetvalidation.org/>) was established to generate therapeutic target evidence from genome-scale experiments and analyses. CTTV aims to support the validity of therapeutic targets by integrating existing and newly-generated data. Data integration has been achieved in some resources by mapping metadata such as disease and phenotypes to the Experimental Factor Ontology (EFO). Additionally, the relationship between ontology descriptions of rare and common diseases and their phenotypes can offer insights into shared biological mechanisms and potential drug targets. Ontologies are not ideal for representing the *sometimes associated* type relationship required. This work addresses two challenges; annotation of diverse big data, and representation of complex, *sometimes associated* relationships between concepts.

**Methods:** Semantic mapping uses a combination of custom scripting, our annotation tool 'Zooma', and expert curation. Disease-phenotype associations were generated using literature mining on Europe PubMed Central abstracts, which were manually verified by experts for validity. Representation of the disease-phenotype association was achieved by the Ontology of Biomedical Association (OBAN), a generic association representation model. OBAN represents associations between a subject and object *i.e.*, disease and its associated phenotypes and the source of evidence for that association. The indirect disease-to-disease associations are exposed through shared phenotypes. This was applied to the use case of linking rare to common diseases at the CTTV.

**Results:** EFO yields an average of over 80 % of mapping coverage in all data sources. A 42 % precision is obtained from the manual verification of the text-mined disease-phenotype associations. This results in 1452 and 2810 disease-phenotype pairs for IBD and autoimmune disease and contributes towards 11,338 rare diseases associations (merged with existing published work [Am J Hum Genet 97:111-24, 2015]). An OBAN result file is downloadable at <http://sourceforge.net/p/efo/code/HEAD/tree/trunk/src/efoassociations/>. Twenty common diseases are linked to 85 rare diseases by shared phenotypes. A generalizable OBAN model for association representation is presented in this study.

**Conclusions:** Here we present solutions to large-scale annotation-ontology mapping in the CTTV knowledge base, a process for disease-phenotype mining, and propose a generic association model, 'OBAN', as a means to integrate disease using shared phenotypes.

**Availability:** EFO is released monthly and available for download at <http://www.ebi.ac.uk/efo/>.

**Keywords:** Rare disease, Phenotype disease associations, OBAN, CTTV, EFO



# Les objets connectés en santé

## Le grand défi, se centrer sur l'utilisateur



- Apporter une valeur d'usage
  - Co-création de valeur au travers de l'interaction
- Intégrer l'expérience utilisateur dans l'élaboration du produit/service
- Comprendre l'appropriation des objets connectés
  - Processus avec facteurs/variables impactant cette appropriation. A accompagner et mesurer
- Mesurer l'expérience utilisateur
  - Impacts sur le bien-être du patient, l'amélioration de la qualité de vie et l'observance

<https://programmesante.wp.imt.fr/page-d-exemple/evenements-publics/journee-objets-connectes-en-sante-5-octobre-2017/>



CNIL.

PARTICULIER

JE SUIS UN  
PROFESSIONNEL

Protéger les données personnelles, accompagner l'innovation, préserver les libertés individuelles

MES DÉMARCHES | THÉMATIQUES | À LA UNE | RÈGLEMENT EUROPÉEN | LA CNIL |   

 > Recherches dans le domaine de la santé : le nouveau chapitre IX est applicable



## Recherches dans le domaine de la santé : le nouveau chapitre IX est applicable

30 juin 2017

*Le nouveau chapitre IX de la loi Informatique et Libertés est désormais applicable en matière de recherche, d'étude ou d'évaluation dans le domaine de la santé. Son entrée en vigueur effective se fera progressivement.*



Le chapitre IX de la loi distingue deux grandes catégories de recherches : d'une part, les recherches impliquant la personne humaine et, d'autre part, les recherches, études et évaluations n'impliquant pas la personne humaine.

### Les recherches impliquant la personne humaine

Elles sont définies par le code de la santé publique. Il s'agit des recherches organisées et pratiquées sur l'être humain en vue du développement des connaissances biologiques ou médicales.

Il en existe trois types, en fonction du niveau d'implication du patient :

1. des recherches interventionnelles qui comportent une intervention sur la personne non justifiée par sa prise en charge habituelle ;
2. des recherches interventionnelles qui ne comportent que des risques et des contraintes minimales dont la liste est fixée par arrêté du ministre chargé de la santé, après avis du directeur général de l'Agence nationale de sécurité du médicament et des produits de santé ;
3. des recherches non interventionnelles qui ne

### Les recherches, études ou évaluations n'impliquant pas la personne humaine

Il s'agit des recherches dans le domaine de la santé qui n'appartiennent pas aux recherches impliquant la personne humaine. Sont en particulier visées les recherches nécessitant exclusivement la réutilisation de données de santé à caractère personnel (par exemple celles issues de dossiers médicaux, de cohortes existantes ou du SNDS).

Les recherches qui relevaient anciennement du chapitre X de la loi appartiennent à cette catégorie.

Les traitements de données à caractère personnel ayant pour finalité ces recherches sont soumis à l'autorisation de la CNIL.

**Comment obtenir une autorisation de la CNIL ?**



## Le CIL et le futur délégué à la protection des données

*Avec une fonction située au coeur de la conformité Informatique et Libertés, le CIL veille à la sécurité juridique et informatique de son organisme. Le CIL a vocation à devenir le délégué à la protection des données dans le cadre de la nouvelle réglementation, applicable en 2018.*

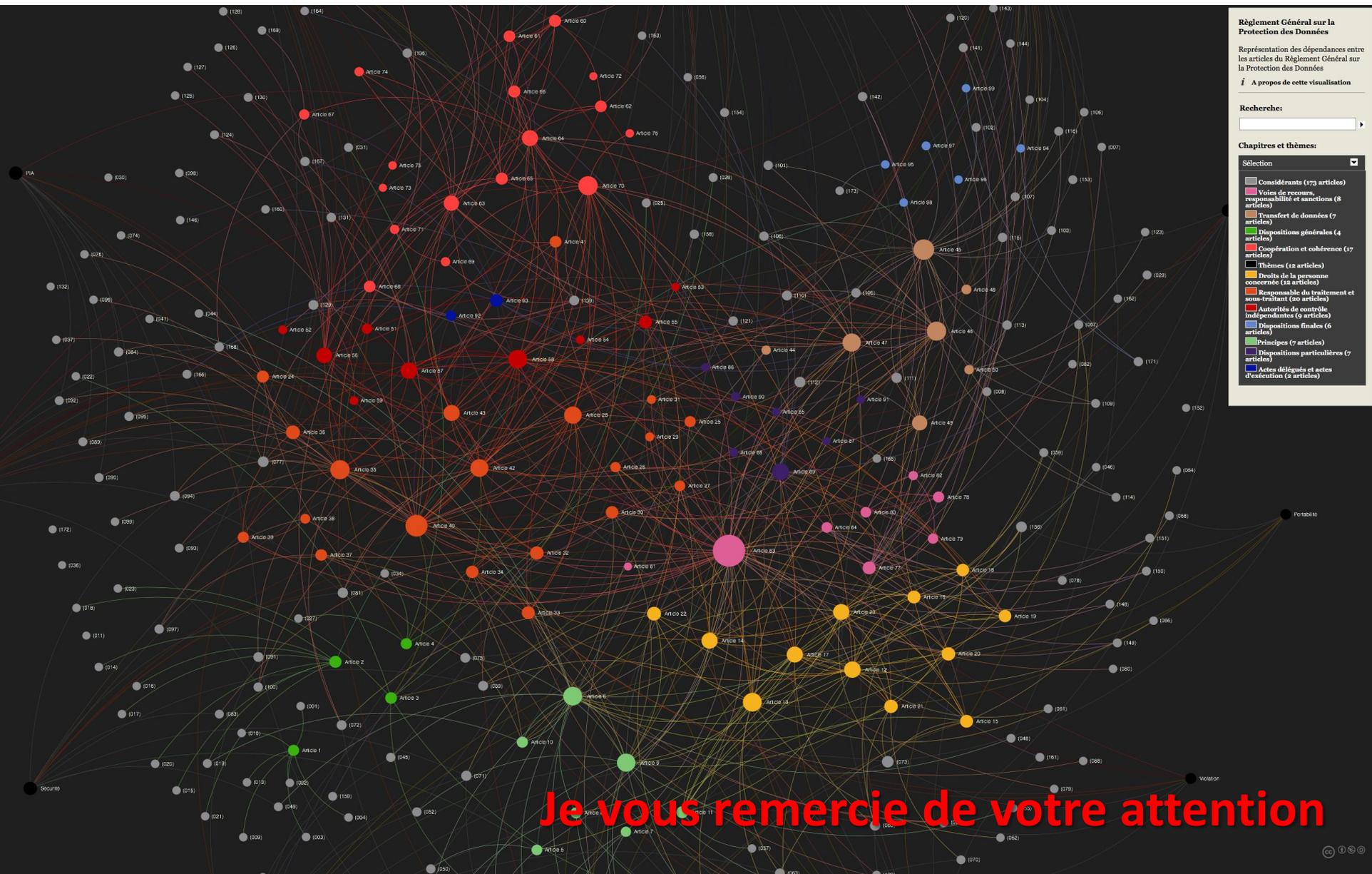


### Devenir délégué à la protection des données

Le délégué à la protection des données est au cœur du nouveau règlement européen, qui constituera le cadre juridique applicable en mai 2018 dans toute l'Europe.

### Le rôle du CIL et ses bénéfices

En gérant les données personnelles qui leur sont confiées dans le respect des règles, les collectivités territoriales, les entreprises publiques ou privées ainsi que les associations, réduisent leur exposition aux risques et optimisent leurs ...

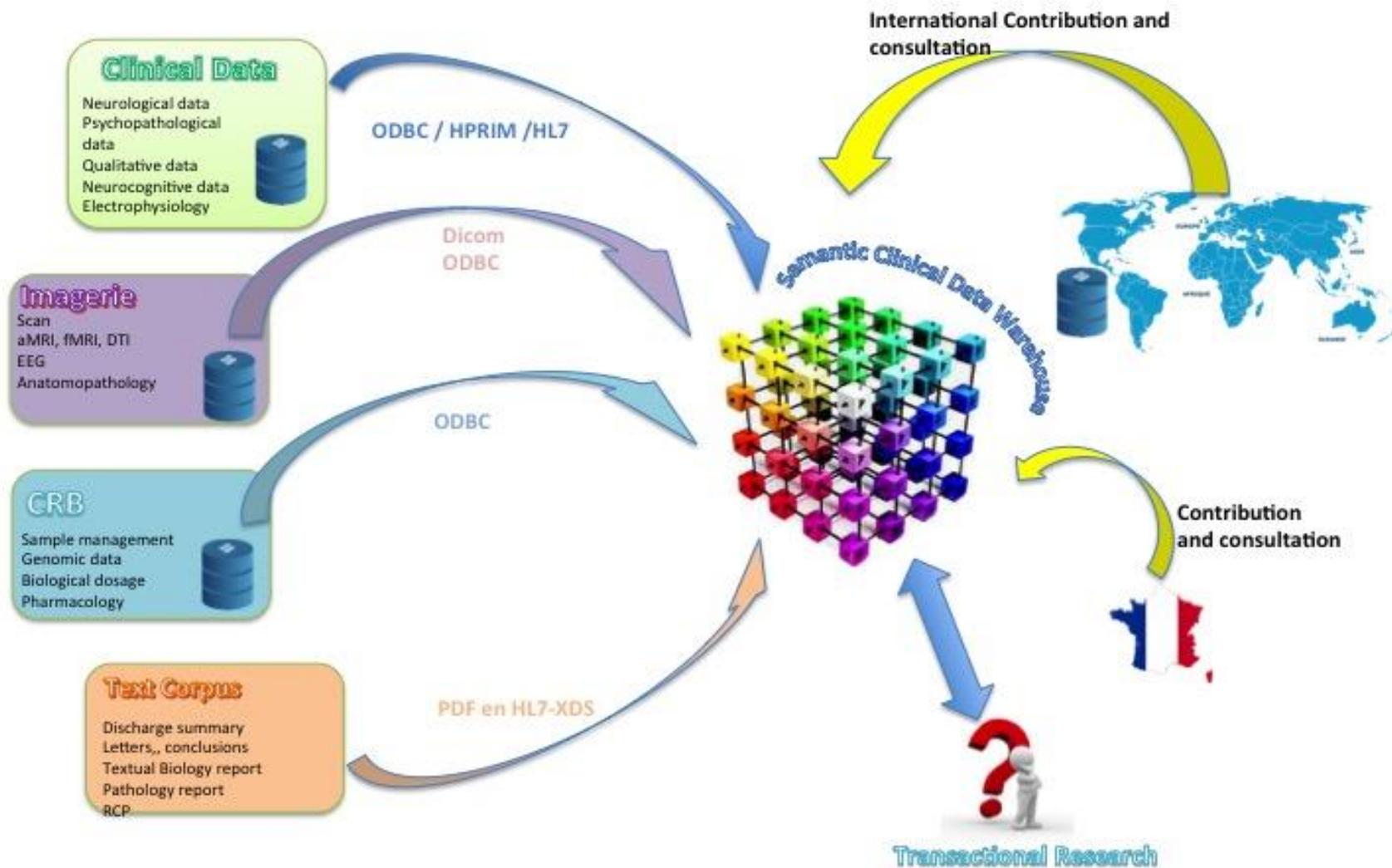






# Intégration de données

- **Que ce soit** dans le contexte du cancer où le séquençage du génome ouvre des voies de compréhension et, encore plus vite, de thérapeutiques...
- **Que ce soit** pour les MR pour 80% Mendéliennes mais dont on n'a souvent que le phénotype...
- **Que ce soit** en psychiatrie où les troubles sont principalement décrits par des syndromes cliniques et où l'on a besoin de mieux caractériser le phénomène (*e.g. EMA, oculomotricité*) et le croiser avec des variants génomiques...
- **Que ce soit**...





# Intégration de données

- Mise à disposition de bases génomiques-phénotypes<sup>1</sup>
- Constitution de sa propre base/étude intégrant les modalités nécessaires
- Modèles termino-ontologiques pour la clinique (et autre ?) et annotations (*cf. supra*)
- Formats pour les données et méta-données (SMART on FHIR Genomics ?) et en fonction du niveau d'interprétation<sup>2,3</sup>
- Modèles d'échanges, de partage de tâches, économique (mission NGS)

<sup>1</sup>AJ Brookes, PN Robinson, Human genotype–phenotype databases: aims, challenges and opportunities, Nature Reviews Genetics, 10 November 2015

<sup>2</sup>G Alterovitz et al., SMART on FHIR Genomics: Facilitating standardized clinico-genomic apps, JAMIA 2015

<sup>3</sup>C Cabot, LF Soualmia, SJ Darmoni, Intégration de données cliniques et omiques pour la recherche d'information dans le Dossier Patient Informatisé, IC2015