

# Pour des données de Santé ouvertes au bénéfice de la science

Arslan, Mazouni, DataScientiste



# L'explosion des données digitales et de la puissance de leur traitement

1. Notre capacité de production des données double chaque 40 mois depuis 1980. Aujourd'hui elle est estimée à 10 exabytes( $2.5 \times 10^{18}$ ) par jour
2. Les Ordinateurs sont aujourd'hui 4 milliards de fois plus puissants
3. qu'en 1960 et ils le seront 8 milliards de fois plus dans 2 ans
4. En 1956 le 1<sup>er</sup> disque dur pesait une tonne et coutait 50 000 \$ en ayant une capacité de 5Mo( une chanson Mp3)

La Majorité écrasante des données produites n'est pas structurées(texte, photos, vidéos) Elles relèvent moins du régime « réponse a questionnaire » que du régime purement observationnel



# Les Données Médicales

- Historiquement les principaux producteurs des données médicales étaient les Médecins dont les sujets sont les patients
- Ce qui est nouveau c'est que les Patients se mettent a produire des données, à travers notamment les récits de leur Maladie, sans que nécessairement les destinataires de leurs récits soient leurs médecins( réseau sociaux par exemple)



# Les Récits Patients

- Les récits des Patients n'obéissent pas au protocole de l'anamnèse médicale, mais sous forme d'expression libre, ils ont souvent pour but le partage avec d'autres patients d'une expérience purement empirique.
- Qu'en est-il donc de ce savoir expérientiel?

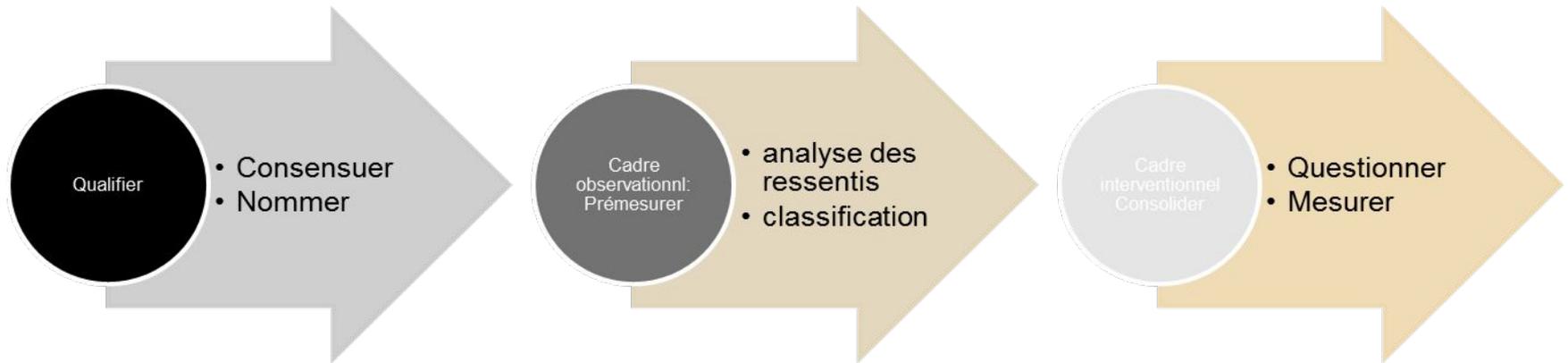


# Illustration de Cas à travers un forum de patients

- Le forum de L'AFA: Présentation
- <https://www.afa.asso.fr/forum/forum.html>



# Méthodologie Faiblement Supervisée: Métrilogos



# 25 Préoccupations centrales des Patients atteints de MICI ou de RCH

**Biothérapies**



**Chirurgie**



**Coloscopie**



**Connaissance et Recherche**



**Douleurs**



**Droits Sociaux**



**Encouragement/Soutien Moral**



**Entourage**



**Espoir**



**États d'Âme**



**Fatigues et Impacts Physiques**



**Fonctionnement du Forum**



**Grossesse**



**Maladie de Crohn**



**Orientation Spécialiste/Avis Spécialisé**



**Parcours de Soins**



**Poussées**



**Problèmes Extra-Digestif**



**Rechute**



**Régime Alimentaire**



**S'interroger**



**Témoignages**



**Temporalité\*\***



**Thérapies Complémentaires**



**Traitements Hors Biothérapies**



# La plate forme Metrilogos

- <https://arslanmazouni.shinyapps.io/metrilogosafa/>



# La question du parcours de soin des enfants atteints de DYS

- Le constat: La faible conformité du réel au cadre du parcours de soin préconisé par la HAS
- Or :
- qu'en est-il réellement sur le terrain?
- Qu'en est-il des parcours de soins effectifs que l'on constate sur les terrains opérationnels relativement au niveau 2 et niveau 3 ?
- Sont-ils d'ores et déjà conformes aux recommandations des rapports de la CNNSE et de l'HAS?
- Sinon, comment mesurer leurs écarts? Et quels efforts mettre en place pour assurer une meilleure convergence vers la conformité préconisée?



# Quel Cadre Méthodologique pour cette étude?

- **Par une méthode rétrospectivement observationnelle et sans a priori, que l'on fera remonter au plus loin que nous le permettront les archives médicales, et notamment les Anamnèses, la reconstitution de tous les parcours de soin dans leurs ordres chronologiques, ainsi que dans leurs différents adressages, devra nous permettre non seulement de répondre à ces questions, mais aussi de mesurer le chemin qui reste à parcourir pour ce faire et partant l'effort à mettre en place.**



# Méthode: Algorithmique, rétrospective et observationnelle

A partir du corpus, non-structuré, et avec des techniques de NLP (Natural language processing) constitué par les observations – Type rédigées à partir du recueil des témoignages des parents des enfants, lors de leur première réception, et accumulées depuis plusieurs années:

1. Comme préalable d'abord, il s'agira de définir un cadre et un protocole pour la Collecte de ce corpus, non Structuré, sa centralisation, son Anonymisation par des identifiants uniques de parcours, sa Classification temporelle afin de:
  1. Reconstituer toutes les Balises pointant un point de contact du patient-anonymisé avec le système de soin (dans le sens le plus général possible dans un premier temps, quitte à déborder vers les parcours de vie-parents, maître d'école...)
  2. De mettre en relation chronologique ces balises ainsi identifiées par des chemins, unidirectionnel, ou bidirectionnels, en en dégagant à chaque fois l'autorité contrôleuse ou prescriptrice ordonnatrice de ce parcours
  3. De projeter ces chemins ainsi constitués dans leurs horizons temporels respectifs, en en dégagant particulièrement les discontinuités qui correspondraient aux zones où le patient-anonymisé serait hors tout parcours de soin.
  4. Enfin de confronter chacun des parcours de soins effectifs aux parcours tant hypothétiquement conformes que ceux prescrits.
2. Tous les parcours de soins ainsi collectés, ordonnés et structurés au niveau de chaque patient-anonymisé dans notre base nous pourrions ainsi:
  1. Caractériser par agrégation et Classification les grandes familles ou type de parcours de soin, depuis les plus erratiques aux plus conformes
  2. De Mesurer les poids respectifs à chaque famille de parcours de soins ainsi identifiée
  3. Enfin, de modéliser, si les données à notre disposition nous le permettent, les mécanismes sous-jacents de chaque type de parcours de soins.



# Fin

- Merci

