



**École d'été interdisciplinaire en numérique de la santé
du 10 au 14 juillet 2023**

Des données aux modèles, de la logique aux relations

NDC02LL

EINS///EINS2023-NDC_Lavoie_Luc

version 0.1.1a, en date du 2023-07-10

(document de travail, ne pas citer)

Table des matières

1. Le problème
2. Une démarche
3. Un cas
 - 3.1. Caractérisation du problème
 - 3.2. Caractérisation des solutions
 - 3.3. Description d'une première solution
 - 3.4. Vérification de la solution
 - 3.5. Une première expérimentation
 - 3.6. Les premières conclusions
4. Trois réflexions (en guise de conclusion)

1. Le problème

- Comment accéder aux données de santé ?
 - Ces données sont-elles suffisantes pour faire la prévention, le soin, et la recherche en santé ?
- Comment accéder aux données *requis*es pour la prévention, le soin et la recherche en santé ?
 - La signification d'une donnée est-elle toujours univoque ?
 - Est-ce que la donnée est suffisante en elle-même ?
- Comment accéder aux données requises pour la prévention, le soin et la recherche en santé *et les interpréter (correctement)* ?

- Ce problème est-il susceptible d'avoir une solution consensuelle, indépendamment des cultures, des sociétés, des états, des ordres professionnels et des intérêts privés ?
- Ce problème est-il unique ou ne faudrait-il pas considérer une famille de problèmes, une famille de solution ?



- Il y aura plusieurs solutions partielles aux problèmes d'accès aux données requises pour la prévention, le soin et la recherche en santé.
- La couverture de ces solutions demeurera partielle et, pour cette raison, il serait souhaitable qu'elle soit interopérable.
- L'interprétation de ces données demeurera plurielle et, pour cette raison, il serait souhaitable qu'elle se fasse sur la base de processus et de modèles documentés, ouverts et traçables.



- Puisqu'il y aura plusieurs modèles, il importe de se doter d'une démarche pour les élaborer.
- Puisqu'il y aura plusieurs modèles, il faut déterminer un méta-modèle facilitant
 - le raisonnement
 - l'interopérabilité
 - la documentation
 - l'ouverture
 - la traçabilité

2. Une démarche

- Caractériser le problème
- Caractériser les solutions
- Concevoir, décrire et vérifier les solutions
- Choisir certaines solutions
- Expérimenter et valider celles-ci
- Tirer les conclusions



Explicitons maintenant notre démarche tout en l'appliquant à notre premier problème : trouver LE méta-modèle !

3. Un cas

3.1. Caractérisation du problème

- Cerner un problème afin d'y trouver une solution, passe par la modélisation et l'analyse dudit problème et donc de sa caractérisation.
- Comparer un problème avec d'autres est riche en enseignement, voire en solutions existantes ou adaptables.
- Choisir une solution, parmi d'autres, passe par la comparaison de celles-ci aux autres solutions.



Commençons par deux formes de caractérisation : les dimensions et les problèmes !

Caractérisation primaire (et partielle) : les dimensions

1. Volume (quantité de données devant être reçues, traitées, stockées, produites)
2. Variété (diversité et complexité des types utilisés par le modèle)
3. Vitesse (débit de données entrantes et sortantes)
4. Variabilité (relativité [ou provenance] : agent, spatio-temporalité, perspective, marge d'erreur et incertitude)
5. Valeur (richesse analytique des données... entre autres)
6. Virtualité (synthétisation [agrégation] des sources de données distribuées)
7. Vertu (gouvernance et sécurité des données)

- La mesure de ces dimensions n'est pas aussi triviale qu'on le voudrait.
- Les procédés et les techniques ne font pas (encore) consensus.

Le cas du méta-modèle.

Qu'en est-il des données de santé ?

1. Volume : pas considérable hors génomique et imagerie
2. Variété : considérable
3. Vitesse : pas considérable
4. Variabilité : considérable
5. Valeur : perception très variable
6. Virtualité : considérable
7. Vertu : complexe, sensible, débattue



Caractérisation secondaire (et partielle) : les besoins

besoin

chose considérée nécessaire à la définition d'un procédé ou au déroulement d'un processus (d'une activité, d'une tâche, d'une action).

Ne pas confondre avec désir attente exigence

procédé

...

processus

...

activité

ensemble de...

tâche

action portant sur une chose

chose

objet matériel ou information



Le cas du méta-modèle.

- Nous avons déjà identifié les besoins suivants
 - soutenir au moins une démarche (systématique) de modélisation
 - soutenir le raisonnement
 - soutenir l'interopérabilité
 - soutenir la documentation
 - soutenir l'ouverture
 - soutenir la traçabilité, voire l'explicabilité

- Voir également article Khnaïsser et coll. pour un approfondissement.

3.2. Caractérisation des solutions

- En présence d'une solution, il faut s'assurer qu'elle en est bien une.

- En présence de plusieurs solutions, il faut pouvoir les comparer.



Commençons par une forme de caractérisation, les exigences, et une façon de les utiliser, l'adéquation.

Exigences

Les exigences ont pour but de déterminer

- les conditions nécessaires et suffisantes pour qu'une solution soit acceptable ;
- les caractéristiques mesurables pertinentes permettant de comparer deux solutions.

Plusieurs procédés ont été proposés afin de permettre la détermination des exigences. Ces procédés se distinguent par l'organisation des activités suivantes et les méthodes préconisées pour chacune.

- exploration
- analyse
- spécification
- vérification
- validation

Le cas du méta-modèle.

Exploration

Exigences pratiques :

- Capacité de soutenir un processus de diagnostic puis de choix de traitement (fondée sur des données probantes en regard de pratiques médicales reconnues) ;
 - corolaire 1 : capacité de décrire l'état du patient,
 - corolaire 2 : capacité de formuler un diagnostic,
 - corolaire 3 : capacité de décrire un traitement,
- ...

Exigences induites :

- Capacité de transposer les théories et les modèles scientifiques dans un cadre unifié (appelé modèle)
 - conséquence : le méta-modèle doit fournir un formalisme permettant de décrire les théories et les modèles scientifiques.



- Capacité de formuler des hypothèses et de les vérifier (au sein du modèle)
 - conséquence 1 : le méta-modèle doit permettre le raisonnement et l'utilisation de données probantes en regard de théories et de modèles scientifiques décrits à l'aide du formalisme.
 - conséquence 2 : il serait souhaitable que le méta-modèle permette la vérification automatisée de prédicats en regard de données.
- Capacité de comparer deux modèles
- Capacité d'évaluer l'adéquation d'un modèle

Remarquons que les quatre capacités induites sont applicables tant au méta-modèle qu'aux modèles qu'il permet de décrire (modéliser).

Le processus d'exploration n'est pas terminé :

- une clarification est certainement possible ;
- d'autres exigences découlent des besoins exposés ;

Un processus de spécification devrait s'ensuivre afin

- de s'assurer que chaque exigence soit
 - claire (c'est-à-dire lisible et compréhensible),
 - exacte (c'est-à-dire précise et sans erreurs),
 - complète (c'est-à-dire comprenant tous les éléments requis et tous les éléments nécessaires) et
 - concise (c'est-à-dire sans éléments superflus) ;
- d'associer à chaque exigence un critère objectif permettant de vérifier qu'elle est satisfaite.

Finalement, la vérification et la validation devraient s'ensuivre afin de minimiser les risques de devoir recommencer la modélisation qui en découlera.

Adéquation

L'adéquation détermine à la fois la façon d'utiliser les exigences, de les concilier, mais aussi de les compléter au moment d'une évaluation globale d'une solution.

- Critères absolus
 1. Validité
 - conformité au modèle

- 2. Efficacité
 - conformité aux exigences
- 3. Cohérence
 - non contradiction interne
- Critères relatifs
 - 1. Complétude
 - couverture « suffisante » du problème
 - 2. Efficience
 - consommation « acceptable » de ressources
 - 3. Évolutivité
 - adaptation « aisée » aux changements
- Méta-critères
 - 1. Réfutabilité
 - comporte les mécanismes permettant d'exprimer et d'évaluer les « falsificateurs potentiels ».
 - 2. Acceptabilité (éthique)
 - comporte les mécanismes permettant d'exprimer et d'évaluer les « critères éthiques ».

Les critères relatifs sont souvent décomposés en sous-critères auxquels sont associés un protocole de mesure et un poids de façon à permettre une évaluation globale pondérée. D'autres méthodes d'évaluation globale sont possibles.

La réfutabilité est considérée en regard d'une épistémologie donnée ; en conséquence, une solution doit préciser son cadre épistémologique.

L'acceptabilité est considérée en regard d'une éthique donnée ; en conséquence, une solution doit préciser cadre éthique.



Le cas du méta-modèle.

Aux fins de l'exercice, posons

- pour l'épistémologie, celle de Karl Popper avec la réfutabilité discrète (de préférence binaire) ;
- pour l'éthique, celle de Karl Popper avec le principe de la nécessaire protection de la liberté par l'État... protection qui passe par sa limitation nécessaire et suffisante !

Nous y reviendrons après avoir décrit le méta-modèle, au moment de l'évaluer.

3.3. Description d'une première solution

- Raisonnement
 - la capacité de soutenir le raisonnement, voire de l'automatiser, est au coeur des capacités recherchées et des exigences à satisfaire.
- Logique
 - la logique décrit les lois du raisonnement gouvernant les prédicats.
- Relation
 - les relations sont une représentation privilégiée des prédicats qui en facilitent le calcul grâce à une algèbre appropriée [adéquate].
- Typage
 - le typage, fondé sur la théorie des ensembles, réduit les risques d'ambiguïté, de paradoxes et d'indécidabilité dans la formulation des prédicats et le calcul des prédicats.

Ce cheminement et la théorie relationnelle qui en découle sont le fruit de travaux que Edgard F. Codd entreprit au cours des années 1960 et mena jusqu'au début des années 1990. La publication séminale date de 1969.

Typage

- valeur, représentation, type
- type de base et sous-type
- type scalaire et type non scalaire
- type prédéfini et constructeur de type

Modèle relationnel

- concepts
 - valeur
 - variable
 - fonction
 - état
 - procédure
 - automatisme
- structure
 - attribut
 - tuple
 - relation

- base

Algèbre relationnelle

- union, intersection, différence
- renommage
- projection, restriction, jointure

3.4. Vérification de la solution

Satisfaction des exigences

La démonstration de la satisfaction des quatre exigences spécifiées plus tôt dépasse la portée et le temps imparti.

Les écrits de nombreux scientifiques (et parmi les plus grands) le démontrent. Écrits qui furent revus, publiés et rendus accessibles.

Adéquation

- Critères absolus
 1. Validité : oui
 2. Efficacité : oui
 3. Cohérence : oui
- Critères relatifs
 1. Complétude : suffisante (complétude au sens de Turing démontrée)
 2. Efficience : la plus efficiente connue, parmi les solutions générales
 3. Évolutivité : pas meilleure que les autres !
- Méta-critères
 1. Réfutabilité
 2. Acceptabilité (éthique)

À vous de voir, en fonction de l'épistémologie et de l'éthique choisie !

3.5. Une première expérimentation

Vive l'APP ! On compte sur vous !

3.6. Les premières conclusions

Acquis et défis

Pourquoi présenter acquis et défis conjointement ? Parce que de nombreux acquis théoriques tardent encore à être mis à disposition, sinon utilisés, en pratique !

- Ambigüité : une certaine amélioration grâce à l'algèbre relationnelle et aux raisonneurs, mais le problème demeure en théorie (lire notre ami Gödel) et en pratique (complexité algorithmique des algorithmes de raisonnement).
- Complétude : la finitude résoud théoriquement le problème de complétude... mais persiste toujours en pratique (voir complexité algorithmique et volume de données).
- Données manquantes : plusieurs solutions ont été proposées dont deux seulement sont solides (Codd:logique_4V et Date:décomposition) ; malheureusement, aucun langage utilisé couramment ne permet de les utiliser commodément.
- Agents : résolu en théorie — en pratique : complexité et expressivité.
- Axes spatiaux : résolu en théorie — en pratique : outils disponibles, mais la complexité demeure et l'expressivité ainsi que l'efficacité demande encore des avancées significatives.
- Axes temporels : plusieurs approches intéressantes, aucune exhaustive, peu d'outils en pratique... mais qu'est-ce que le temps ?
- Cohérence : Que se passe-t-il si on ajoute l'exigence suivante « Capacité de transposer les règles de pratiques médicales » dans le modèle ?

Pistes

- Utiliser les ontologies appliquées pour décrire les prédicats
- S'appuyer sur les connaissances.
- Réduire (encore plus) les sources d'ambigüité.

4. Trois réflexions (en guise de conclusion)

« Du problème à la solution, en passant par les besoins, les modèles, les exigences et l'adéquation. »

« Des données aux modèles en passant par la connaissance et les relations. »

« Du raisonnement aux relations en passant par la logique. »

Merci à Pythagore (vers -580 à -495), Socrate (vers -470 à -399), Platon (vers -428 à -347), Aristote (vers -384 à -322), Eulcide (vers -300)... Descartes (1596 à 1650), Pascal (1623 à 1662), Russel (1872 à 1970), Wittgenstein (1889 à 1951), Gödel (1906 à 1978), Floyd (1936-2001), Codd (1923 à 2003), Chomsky (1928...), Hoare (1934...) et Date (1941...).

document de travail, ne pas citer 0.1.1a

Last updated 2023-07-11 06:03:55 -0400