



HAL
open science

Interface d'interrogation graphique de parcours de soins à base de chroniques

Pegdwendé N Sawadogo, Thomas Guyet, Etienne Audureau

► To cite this version:

Pegdwendé N Sawadogo, Thomas Guyet, Etienne Audureau. Interface d'interrogation graphique de parcours de soins à base de chroniques. Conférence Extraction et gestion de connaissances, Jan 2023, Lyon, France. hal-04181253

HAL Id: hal-04181253

<https://hal.science/hal-04181253>

Submitted on 15 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Interface d’interrogation graphique de parcours de soins à base de chroniques

Pegdwendé N. Sawadogo*, Thomas Guyet**
Etienne Audureau***

*Fondation de l’AP-HP, Paris, France
nicolas.sawadogo-ext@aphp.fr

**INRIA, AISTroSight, Lyon, France
thomas.guyet@inria.fr

***AP-HP, Hôpital Henri Mondor, Université Paris Est Créteil, France
etienne.audureau@aphp.fr

Résumé. Pour mener des études épidémiologiques, les épidémiologistes sont amenés à cibler des patients d’intérêt sur la base de leurs parcours de soins. Cette tâche est particulièrement complexe car elle nécessite de pouvoir exprimer des contraintes et relations temporelles entre les événements d’une trajectoire de soins. Les approches existantes pour y parvenir se montrent limitées soit en terme d’expressivité, ou à cause d’une grande technicité d’utilisation inadaptée au profil des épidémiologistes. Pour y remédier, nous proposons un outil permettant de définir graphiquement une trajectoire de soins et de retrouver les patients correspondants.

1 Introduction

Au cours des dernières décennies, la prise en charge des patients dans les hôpitaux a été révolutionnée par l’intégration progressive d’outils numériques. Des bases de données médico-administratives ont alors été mises en place pour supporter la digitalisation des services médicaux à travers le stockage des données relatives à la prise en charge des patients. Sont ainsi stockées des informations biologiques sur les patients (sexe, date de naissance, groupe sanguin, etc.), des comptes rendus de consultations, des résultats d’examens médicaux, des diagnostics, des prescriptions de médicaments, ou encore un historique de mesures réalisées sur les patients (poids, taille, tension artérielle, etc.) (Sharma et al., 2019).

Ces bases de données supportent non seulement la prise en charge des patients, mais également, de façon indirecte, la recherche en épidémiologie (De Moor et al., 2014). En effet, l’exploitation à grande échelle des données de patients offre de nouvelles opportunités de recherche à travers, par exemple, de l’analyse de l’efficacité des traitements, de la surveillance d’effets secondaires et même de l’anticipation d’évènements probables pour un patient (De Moor et al., 2014; Bakalara et al., 2021). Cependant, la réalisation de ces études nécessite une étape primordiale d’identification des patients d’intérêt (par exemple, des patients atteints d’une maladie). Il s’agit alors de retrouver des patients correspondant à un ensemble de critères portant sur les données disponibles : des critères biologiques (poids, taille, etc.) mais

Interface d'interrogation graphique de parcours de soins

aussi des critères sur leur parcours de soin, c'est-à-dire, un ensemble d'évènements médicaux survenus dans la vie du patient (prescriptions de médicaments, consultations, diagnostics, etc.). L'ensemble de ces critères forme une *trajectoire de soin*.

L'identification de patients sur la base de leur parcours de soins est particulièrement complexe (Bakalara et al., 2021). En effet, cela nécessite de pouvoir exprimer des contraintes temporelles et des relations de délais, d'antériorité ou de postériorité entre plusieurs évènements apparaissant dans un parcours de soins. Dans ce contexte, les outils classiques d'interrogation de bases de données comme le langage SQL se sont montrés limités. En effet, ce type d'outils devient très technique lorsqu'il s'agit de rechercher des trajectoires de soins complexes (Bache et al., 2015). Pourtant, les études épidémiologiques sont généralement conduites par des médecins ayant des compétences techniques limitées. Certes, des approches alternatives ont été proposées dans la littérature pour y remédier, mais celles-ci restent limitées soit, en terme d'expressivité (Chapman et al., 2021; Uciteli et al., 2020), ou à cause de leur complexité d'utilisation (Bacry et al., 2020; Bakalara et al., 2021).

C'est pourquoi nous proposons dans cet article un outil graphique d'interrogation des parcours de soins dans une base de données médico-administratives. Nous exploitons ainsi la richesse expressive du concept de chronique sémantique (Bakalara et al., 2021) pour assister les épidémiologistes dans l'interrogation des bases de données médico-administratives. À travers de cette interface graphique, les épidémiologistes peuvent représenter une trajectoire de soins composée de plusieurs évènements médicaux (visites, délivrance de médicaments, diagnostics et décès), en définissant des contraintes temporelles sur lesdits évènements. Cette trajectoire de soin est alors exécutée comme une requête pour identifier les patients correspondants.

La suite de cet article est constituée comme suit : la Section 2 présente un bref état de l'art des outils d'interrogation des parcours patients. Dans la Section 3, nous détaillons notre approche d'interrogation des parcours de soins. Dans la Section 4, nous présentons des scénarios d'utilisation de l'outil que nous proposons. Enfin, nous concluons l'article dans la Section 5.

2 État de l'art

Dans cette section, nous présentons quelques outils proposés dans la littérature pour l'interrogation des parcours de soins. Nous distinguons d'une part, les systèmes basés sur des langages informatiques et, d'autre part, les systèmes intégrant une interface graphique.

2.1 Systèmes basés sur des langages informatiques

Pour aider à la constitution de cohortes à partir de données patient, Bache et al. (2015) ont proposé le langage de requêtage *ECLECTIC* (*Eligibility Criteria Language for Clinical Trial Investigation and Construction*). *ECLECTIC* possède une grande richesse expressive, supportant à la fois des combinaisons logiques et des contraintes temporelles lui permettant d'interroger le parcours de soins.

Le système *SCALPEL3* (*SCALable Pipeline for hEaLth data*) propose des fonctionnalités analogues à *ECLECTIC* en se basant sur un système d'API¹ en langage Python (Bacry et al., 2020). Cependant, la présentation sous la forme d'une API python rend ce système difficilement exploitable par les épidémiologistes.

1. Application Programming Interface (Interface de programmation d'application)

Bakalara et al. (2021) adoptent quant-à-eux une présentation plus intuitive pour supporter l'interrogation de parcours de soins. Chaque trajectoire y est conçue à travers une chronique, c'est-à-dire un modèle graphique constitué d'un ensemble d'évènements et de contraintes temporelles. Cependant, la représentation par chroniques se limite au niveau conceptuel dans le système de Bakalara et al. (2021). En pratique, la chronique est concrétisée sous la forme de requêtes SPARQL.

2.2 Systèmes d'interrogation graphiques

Le système *Informatics for integrating biology and the bedside (i2b2)* est un des pionniers en matière d'interrogation graphique des parcours de soins ou des données démographiques des patients (Murphy et al., 2010). Il intègre une interface graphique à travers laquelle les utilisateurs peuvent définir une trajectoire constituée d'une succession d'évènements et de contraintes temporelles.

En guise d'alternative à *i2b2*, l'outil *Cohort360*² est en cours de conception par l'Assistance Publique - Hôpitaux de Paris (AP-HP) pour remédier à certaines limites d'*i2b2*. Ainsi, l'interface de *Cohort360* est axée sur le support des combinaisons logiques (conjonction, disjonction, négation) entre les évènements de la trajectoire. *Cohort360* permet également de visualiser les données patients, une fois la cohorte constituée.

Finalement, l'outil *Phenoflow* (Chapman et al., 2021) propose un système de *workflow* pour interroger les données de patients. Les critères de définition de la cohorte ciblée sont ainsi agencés et présentés visuellement sous la forme d'un ensemble d'étapes successives. Toutefois, *Phenoflow* est, à l'image de *Cohort360*, centré sur le support de combinaisons de critères et supporte peu la définition de contraintes temporelles.

2.3 Discussion

Les systèmes de constitution de cohortes basés sur des langages informatiques partagent l'inconvénient d'être inadaptés pour les épidémiologistes. Ils exigent des compétences techniques avancées, et sont *de facto* difficilement utilisables par ces utilisateurs (Sawadogo et al., 2022). Sur cet aspect, les systèmes d'interrogation graphiques sont donc préférables.

Cependant, les systèmes d'interrogation graphiques existant dans la littérature possèdent également des limites. Ainsi, *i2b2* qui est basé sur des requêtes SQL supporte difficilement les requêtes temporelles complexes (Sholle et al., 2020). De plus, l'interface graphique proposée dans *i2b2* paraît peu intuitive et limitée pour exprimer des relations d'antériorité complexes entre des évènements. D'ailleurs, *Cohort360* qui est encore en cours de développement, semble partager cette insuffisance. Enfin, le système de *workflow* de *Phenoflow* est inadapté à l'expression de contraintes temporelles complexes.

Pour dépasser les limites identifiées dans la littérature, nous proposons dans cet article un nouvel outil offrant aux épidémiologistes une interface à la fois intuitive et expressive pour l'interrogation des parcours de soins au travers d'une trajectoire exprimée sous la forme d'une chronique. Ainsi, l'outil que nous proposons intègre à la fois l'expressivité des chroniques et la facilité d'utilisation d'une interface graphique.

2. <https://docs.cohort360.org/>

3 Approche d'interrogation d'un parcours de soins

Dans cette partie, nous détaillons l'approche de modélisation sous-jacente au système d'interrogation que nous proposons. Pour ce faire, nous montrons comment un parcours de soins peut être représenté sous la forme d'un graphe, et interrogé grâce au concept de chroniques.

3.1 Représentation du parcours de soins par un graphe

Un parcours de soins représente un ensemble d'évènements médicaux prescrits à un patient. Ces évènements sont associés soit à un horodatage (naissance, examen médical, décès, etc.) ou à un intervalle de temps (exposition à des médicaments, maladie, etc.) (Sawadogo et al., 2022). Dans notre cas, nous concevons les parcours de soins comme des graphes. Ce choix est particulièrement motivé par la flexibilité et l'intuitivité offertes par cette représentation.

Pour constituer ce graphe, chaque évènement d'un parcours de soins est formellement représenté par un couple (p, v) , avec p le patient, v l'évènement. Cet évènement est constitué de trois composantes : le temps (l'horodatage ou l'intervalle de temps), le type d'évènement (p. ex. *diagnostic*) et, dans certains cas, une description détaillée (p. ex. le code CIM³ pour les diagnostics).

3.2 Interrogation des parcours de soins par des chroniques

Pour retrouver les patients à inclure dans une étude, les épidémiologistes définissent une trajectoire qui, en plus des évènements, intègre des contraintes temporelles de délai et d'ordre. Par exemple, les patients souffrant d'une thrombose veineuse profonde (TVP) peuvent être caractérisés par la trajectoire suivante (Bakalara et al., 2021) : « *Patients ayant fait l'objet d'un diagnostic de TVP ou d'embolie pulmonaire (EP) suivi, un à deux jours plus tard, de plusieurs prescriptions de médicaments anti-coagulants (AC); les prescriptions étant séparées entre elles de deux mois au plus* ».

Ce type de requêtes peut être représenté et exécuté efficacement à l'aide du concept de *chronique*. Les chroniques permettent en effet de définir ce type de trajectoire, et de vérifier son occurrence dans un ensemble de séquences (Dousson et Le Maigat, 2007). Les chroniques offrent le double avantage d'une grande expressivité en termes de contraintes temporelles, et d'une représentation intuitive sous la forme d'un graphe. Ainsi, la trajectoire de soins définie dans l'exemple précédent peut être représentée par la chronique de la Figure 1. Les nœuds y représentent des évènements (un diagnostic et deux délivrances de médicaments) tandis que les arcs traduisent des intervalles de délais acceptables.

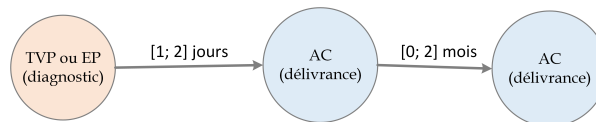


FIG. 1 – *Chronique correspondant à la trajectoire d'un patient atteint d'une thrombose veineuse profonde (Bakalara et al., 2021).*

3. <https://icd.who.int/en>. Nous utilisons la version 9 de la CIM.

4 Présentation de l'outil

Dans cette partie, nous présentons plus en détails le fonctionnement de l'outil que nous proposons pour l'interrogation graphique de parcours de soins. Nous nous focalisons plus précisément sur les fonctionnalités offertes et l'architecture qui permet de les supporter.

4.1 Architecture du système

L'outil d'interrogation que nous proposons est structuré en trois composants principaux (cf. Figure 2) :

Le *graphe de propriétés* permet de stocker l'ensemble des parcours de soins. L'approche par graphe offre en effet une flexibilité permettant de supporter aisément de nouveaux types d'évènements, sans remettre en cause le modèle de données. Dans notre cas, le graphe de propriétés est supporté par la base de données Neo4J⁴.

L'*API REST* sert d'intermédiaire entre le graphe de connaissance représentant l'ensemble des parcours de soins, et l'interface graphique à partir d'où les parcours de soins sont interrogés. Elle permet ainsi de convertir les chroniques définies graphiquement par l'utilisateur en requêtes CYPHER (le langage de requêtage dédié à Neo4J) et de les exécuter. L'API REST est développée dans notre cas en exploitant la bibliothèque Python Flask⁵.

L'*interface graphique* permet aux utilisateurs de définir visuellement une trajectoire, et de recueillir la liste des patients correspondants. En guise de trajectoire, les utilisateurs dessinent une chronique composée d'évènements médicaux. L'interface graphique a été conçue à l'aide de la bibliothèque Javascript D3JS⁶.

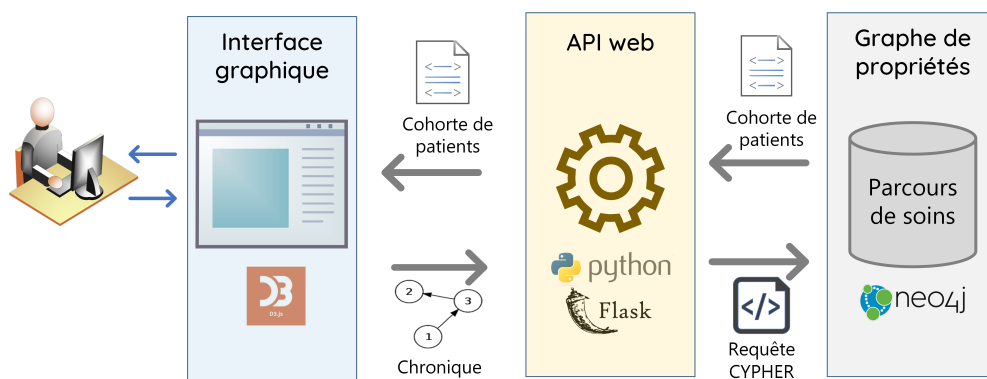


FIG. 2 – Architecture de l'outil d'interrogation.

4. <https://neo4j.com/>

5. <https://flask.palletsprojects.com/>

6. <https://d3js.org/>

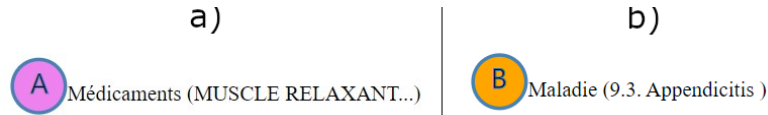


FIG. 3 – Exemples de représentations d'évènements uniques.

4.2 Fonctionnalités et exemples d'utilisation

Le système que nous proposons permet aux utilisateurs de définir une trajectoire en combinant quatre types d'évènements que sont la *visite*, la *délivrance* de médicaments, le *diagnostic* de maladie et le *décès*. Chacun des quatre types d'évènements est associé à une couleur et à un ensemble de valeurs possibles issues des valeurs observées (visites), ou de codages standards (ATC⁷ pour les délivrances et CIM pour les diagnostics). L'utilisateur peut alors interroger les parcours de soins à travers des requêtes visuelles plus ou moins sophistiquées :

Évènements uniques : Il s'agit du cas où la trajectoire est définie par l'occurrence d'un seul évènement. Ce type de trajectoire est traduit sur l'interface graphique par un nœud unique. Ce nœud peut être associé à un intervalle de temps précisant une durée minimale et une durée maximale attendue pour l'apparition des évènements.

Par exemple, la Figure 3-a présente le cas où l'on recherche les patients ayant été exposés à des médicaments de type « relaxants musculaires », tandis que la Figure 3-b représente les patients à qui on a diagnostiqué une « appendicite ».

Évènements multiples : Il s'agit ici du cas d'une conjonction de plusieurs évènements. Cela se traduit sur l'interface graphique par la création de n nœuds (pour n évènements). Dans certains cas, un même nœud peut être dupliqué pour exprimer le fait que l'évènement apparaisse plusieurs fois dans le parcours de soins.

En guise d'illustration, la trajectoire définie dans la Figure 4 permet de retrouver les patients à qui on a diagnostiqué une blessure ou un empoisonnement, à qui on a délivré au moins une fois des médicaments antalgiques et qui sont décédés.

Évènements multiples avec des délais : La trajectoire est ici définie via une conjonction d'évènements multiples avec en plus des contraintes d'ordre et de délais entre les évènements. Ces contraintes sont ajoutées par l'utilisateur en reliant les nœuds (qui représentent des occurrences d'évènements) par des arcs. La direction des arcs précise l'ordre dans lequel les évènements apparaissent dans la trajectoire, tandis qu'un intervalle de temps associé à l'arc permet de spécifier un délai minimal et un délai maximal attendu entre deux évènements.

La Figure 5 illustre la trajectoire associée aux patients qui ont eu une blessure ou un empoisonnement, et qui ont eu une prescription d'antalgique moins d'un mois après, puis une deuxième prescription d'antalgique moins d'un mois après la première.

En plus de la trajectoire définie, des critères démographiques peuvent servir à affiner le résultat en filtrant suivant la date de naissance ou le sexe biologique des patients. On obtient alors une cohorte de patients, c'est-à-dire une liste de patients correspondants. Cette liste est observable sur l'interface graphique sous la forme d'un tableau, ou de graphiques permettant

7. https://www.whooc.no/atc_ddd_index/

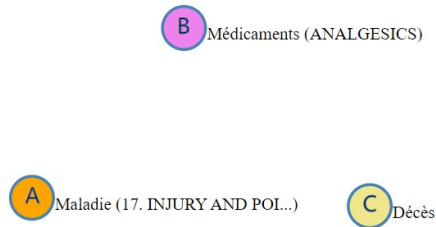


FIG. 4 – Chronique avec des événements multiples.

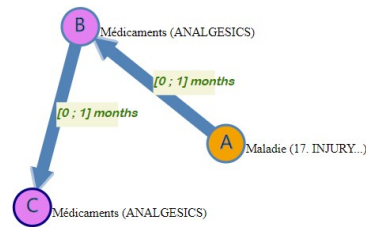


FIG. 5 – Chronique avec des événements multiples et des délais.

de visualiser, par exemple, la répartition des âges ou des sexes biologiques. Cette liste peut ensuite être exportée pour poursuivre les analyses.

5 Conclusion

Dans cet article, nous avons proposé un outil destiné à l'interrogation graphique des parcours de soins dans les bases de données médico-administratives. Pour ce faire, nous avons combiné la richesse expressive du concept de chroniques et l'intuitivité offerte par la représentation graphique. Nous avons ainsi conçu un outil plus complet que les systèmes existants dans la littérature pour l'interrogation de parcours de soins.

Dans cette première implémentation, nous nous sommes servis des données MIMIC, une base de données médico-administrative libre d'usage (Johnson et al., 2016). L'outil reste toutefois suffisamment flexible pour supporter de nouveaux types d'événements, ou même d'autres bases de données. D'ailleurs, cet outil peut servir au-delà du domaine médical à interroger toute base de données d'événements. Il pourrait par exemple servir à retrouver des combinaisons d'événements à partir de fichiers *logs*.

Dans nos travaux futurs, nous souhaitons mener à terme une campagne d'évaluation de l'utilisabilité de l'outil, basée sur le protocole *System Usability Scale* (Brooke, 1996). Cette étude a actuellement été menée sur 5 utilisateurs et donne des premiers résultats prometteurs.

Remerciements Une partie des recherches présentées dans cette démo est subventionnée par la Fondation de l'AP-HP, dans le cadre de la Chaire AI-RACLES et a reçu l'accord du Comité scientifique et éthique du CDW de l'AP-HP (CSE-20-11-COVIPREDS).

References

- Bache, R., A. Taweel, S. Miles, et B. C. Delaney (2015). An eligibility criteria query language for heterogeneous data warehouses. *Methods of information in medicine* 54(01), 41–44.
- Bacry, E., S. Gaiffas, F. Leroy, M. Morel, D.-P. Nguyen, Y. Sebiat, et D. Sun (2020). SCALPEL3: a scalable open-source library for healthcare claims databases. *International Journal of Medical Informatics* 141, 104203.

- Bakalara, J., T. Guyet, O. Dameron, A. Happe, et E. Oger (2021). An extension of chronicles temporal model with taxonomies-application to epidemiological studies. In *Proceedings of the International Conference on Health Informatics (HEALTHINF)*, pp. 133–142.
- Brooke, J. (1996). Sus: a quick and dirty usability scale. *Usability evaluation in industry 189*, 4–7.
- Chapman, M., L. V. Rasmussen, J. A. Pacheco, et V. Curcin (2021). Phenoflow: A microservice architecture for portable workflow-based phenotype definitions. In *Proceedings of the AMIA Summits on Translational Science*, pp. 142.
- De Moor, G., M. Sundgren, D. Kalra, A. Schmidt, M. Dugas, B. Claerhout, T. Karakoyun, C. Ohmann, P.-Y. Lastic, N. Ammour, R. Kush, D. Dupont, M. Cuggia, C. Daniel, G. Thienpont, et P. Coorevits (2014). Using electronic health records for clinical research: The case of the EHR4CR project. *Journal of Biomedical Informatics 53*, 162–173.
- Dousson, C. et P. Le Maigat (2007). Chronicle recognition improvement using temporal focusing and hierarchization. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, Volume 7, pp. 324–329.
- Johnson, A. E., T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, et R. G. Mark (2016). MIMIC-III, a freely accessible critical care database. *Scientific data 3*(1), 1–9.
- Murphy, S. N., G. Weber, M. Mendis, V. Gainer, H. C. Chueh, S. Churchill, et I. Kohane (2010). Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *Journal of the American Medical Informatics Association 17*(2), 124–130.
- Sawadogo, P., T. Guyet, et E. Audureau (2022). Conceptions de phénotypes computationnels pour la recherche en santé publique. In *Actes de la journée Santé et IA*.
- Sharma, H., C. Mao, Y. Zhang, H. Vatani, L. Yao, Y. Zhong, L. Rasmussen, G. Jiang, J. Pathak, et Y. Luo (2019). Developing a portable natural language processing based phenotyping system. *BMC Medical Informatics and Decision Making 19*(3), 79–87.
- Sholle, E. T., M. Cusick, M. A. Davila, J. Kabariti, S. Flores, et T. R. Campion (2020). Characterizing basic and complex usage of i2b2 at an academic medical center. *Proceedings of the AMIA Summits on Translational Science 2020*, 589.
- Uciteli, A., C. Beger, T. Kirsten, F. A. Meineke, et H. Herre (2020). Ontological representation, classification and data-driven computing of phenotypes. *Journal of Biomedical Semantics 11*(15), 1–17.

Summary

For the purpose of epidemiological studies, epidemiologists have first to retrieve using health data patients corresponding to a defined care trajectory. This task is particularly complex because it requires the expression of temporal constraints and relationships between events. Existing approaches to achieve this are limited either in terms of expressiveness, or because of a high level of technicality unsuitable to epidemiologists. To remedy this, we propose an intuitive tool to define a care trajectory graphically and find the corresponding patients.