Evo-KnoMo: How to Evolve with the Current Times

Oscar Corcho[1], Ana Iglesias-Molina[1], Edna Ruckhaus[1], Jhon Toledo[1], ChatOEG[1] and David Chaves-Fraga[1]

[1]*Universidad Politécnica de Madrid, Campus de Montegancedo, Boadilla del Monte, Spain*

Abstract

Our society is currently divided into two knowledge ecosystems. Know4All, as a public decentralized initiative, shares knowledge between peers under demand, while the private ecosystem GoKnow centralizes knowledge management and data services. The community-driven efforts in developing resources for Know4All have lead to the construction of KnoMo, a system composed of an aggregate of resources that enables peer-to-peer data sharing. These modules have been developed for each knowledge domain, and are well-established and maintained by the community. This maintenance and curation process, however, is still carried out manually. We propose Evo-KnoMo, a framework that automates the evolution of the KnoMo Modules. It is based on usage metrics, meta-neuro-genetic algorithms, and explainability algorithms in order to automate the deprecation, search and implementation of new resources; while producing a readable report of the modifications with their explanation. We test this framework in a user study with the Healthcare and Transport domains with users and curators. Results show that Evo-KnoMo can evolve modules with a high accuracy rate and a considerable level satisfaction by the participants. We believe that this framework can help in the maintenance of KnoMo and thus, enhance the user experience for all users within Know4All.

Keywords

Knowledge Graphs, Knowledge Management, Evolution

***Disclaimer**: This paper is a work of fiction, written in 2023 and describing research that may be carried out in 2043. For this reason, it includes citations to papers produced in the period 2024-2043, which have not been published (yet); all citations prior to 2024 refer instead to papers already in the literature. Any reference or resemblance to actual events or people or businesses, past present or future, is entirely coincidental and the product of the author's imagination.*

1. Introduction

Around three decades ago, chips that store personal data were created mainly to facilitate commercial transactions. In the following ten years, the technology of personal chips evolved and by 2028 the UniChip [1], a device that stores and organizes all our personal data, had already been widely adopted. This technology has radically changed society's lifestyle, enabling people to access, manage and share their personal information in a secure manner.

Together with the emergence of UniChip, two approaches for performing data transactions were developed: (1) GoKnow [2], a proprietary ecosystem in which Google controls data and

ESWC 2043 - The next 20 years track, ESWC 2023, May 31st 2023, Hersonissos, Greece

oscar.corcho@upm.es (O. Corcho); ana.iglesiasm@upm.es (A. Iglesias-Molina); edna.ruckhaus@upm.es (E. Ruckhaus); ja.toledo@upm.es (J. Toledo); david.chaves@upm.es (D. Chaves-Fraga)

0000-0002-9260-0753 (O. Corcho); 0000-0001-5375-8024 (A. Iglesias-Molina); 0000-0002-7735-2739 (E. Ruckhaus); 0000-0002-2924-7272 (J. Toledo); 0000-0003-3236-2789 (D. Chaves-Fraga)

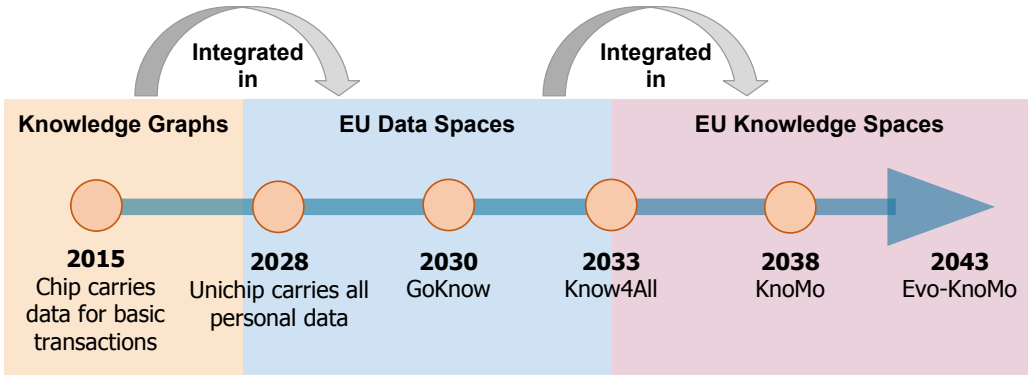


Figure 1: Timeline of approaches for personal data storage, management, and transference.

knowledge management, and (2) Know4All [3], an open-source and community-driven data management and modeling approach. The complete timeline of the development of approaches related to the personal data storage and transference can be seen in Figure 1.

Since the inception of Know4All in 2033, based on the previous Solid project [4], a large part of society has actively contributed to improving the system in terms of peer-to-peer communication and user experience in such a way that currently, Know4All is competitive with GoKnow [5]. One of the most important challenges that we have faced these last years is achieving the interoperability of personal data with the different models that organizations have created. Take, for example, the healthcare domain where users need to carry out health-related transactions, e.g., go to a doctor's appointment. In Know4All, the hospital's data on each person is not persistent, so when users arrive at the doctor's office, the personal data is transferred to the hospital's system according to a domain model common to all hospitals. This transference of personal data is carried out through automatic alignments with this common model. Once the doctor writes a diagnosis and prescribes drugs, data including these additions is transferred back to the person's Unichip and deleted from the hospital's repository.

We can then say that Know4All maintains a community-driven system –KnowMo [6], that is comprised of several components: a set of interconnected conceptual models –ontologies that correspond to different domains, e.g. healthcare, transport, education; validation constraints; and actionable mappings for automatic data transfer and integration. The KnoMo system is able to recognize and align the user's data with the Healthcare module, and validate the transferred data with SHACLv3[1] shapes available for this module. Consequently, the data is available for the doctor's appointment in an automatic way, and once the patient leaves the hospital, his chip's personal data has been updated and enriched with additional data.

Currently, KnoMo is actively maintained and curated by stakeholders with the support of some digital tools (e.g. Knoops! [7], MappingPedia [8]). Thus, hospitals and other health-related facilities contribute to the Healthcare KnoMo; transport agencies contribute to the Transport KnoMo, among others. As time passes, the workload for curators has increased exponentially, creating bottlenecks for proper maintenance of the different KnowMo modules and resources.

[1]https://www.w3.org/TR/shacl-3/

The main problem lies in monitoring the user's evolving transfer requests (and alignments) where new or updated concepts and relationships may be required, or on the contrary, not used anymore and need to be deprecated.

In this paper, we propose Evo-KnoMo, an approach that enables KnoMo to evolve dynamically and automatically according to the users' interaction with the system. Evo-KnoMo takes into consideration which parts of the model are not being used during data transfer, and which concepts and relationships are requested but not found, in order to apply changes to KnowMo that are aligned with the users needs. In order to facilitate the management of this evolution, Evo-KnoMo also provides a notification system to inform about the changes performed, so that curators and maintainers are updated on the latest changes and are able to roll back these changes if necessary. We evaluate Evo-KnoMo with two KnoMo modules, the Healthcare and the Transport modules, during an eight-week trial. The goal of this experimental study is to assess the accuracy of the framework and the curator's satisfaction regarding its usability.

The remainder of this paper is structured as follows: Section 2 introduces the related work; Section 3 presents Evo-KnoMo; Section 4 describes the experimental evaluation that was conducted and discusses the results obtained; and Section 5 draws conclusions and future steps.

2. Related Work

Data integration has been an actual problem for almost 80 years [9]. For many years, the construction of knowledge graphs [10] helped to integrate the myriad of data uploaded to the web, but most of the work was done manually by knowledge engineers. With the advent of Large Language Models (e.g., GPT-4 [11]), Solid [4] and Data Spaces [12], data wrappers started to be automatically constructed, impacting knowledge management in two ways: i) the manual work done by experts was minimized; but, ii) the understandability of the data integration pipelines was completely lost. Therefore, users allow access to their data through these models because of the benefits of the provided services developed around them [5].

The society changed profoundly with the incorporation of UniChip [1], moving from data management to actual knowledge management. Google took this chance and created GoKnow [2], a unique, centralized and integrated knowledge ecosystem, which provides high-quality services to its users, but increases its benefits exponentially by exploiting personal data. From the public perspective, the joint effort between the US and the EU, transforms all data spaces (e.g., Transport [13]) into Know4All, materializing KnoMo [6], a public knowledge ecosystem that provides support for all public services and it is able to share data with the personal UniChip.

The tools that currently support the modification of KnoMo resources, such as Knoops [7] or MappingPedia 2.0 [8], are mainly supported by domain-related teams that incorporate or deprecate properties based on experiences and user feedback. Some implementations were proposed, such as AutoKnoMo [14], that try to reduce the manual work of experts, ensuring a proper evolution of these resources. However, their results reflect low efficiency and accuracy in capturing user knowledge; hence, they cannot be integrated into the tool suite of KnoMo. We propose the novel approach Evo-KnoMo, in order to completely automatize the evolution of the KnoMo system.

3. Evo-KnoMo

Evo-KnoMo is a framework that enables KnoMo to evolve automatically and dynamically by exploiting the input that is gathered when a person transfers his data through a service that addresses a certain area's functionality, e.g. healthcare doctor's appointment. Evo-KnoMo is able to detect which concepts and relationships of the module are not being used to deprecate them; and add new ones when users request them more frequently (Figure 2). These changes are applied to all resources in the KnoMo module: (1) the ontologies that provide the agreed shared conceptualization of the domain, (2) the mappings that enable the transformation of personal data according to the KnoMo model and back, and (3) the validation constraints that check the correctness of the transformed data. The ultimate objective of Evo-KnoMo is to improve the accuracy of KnowMo with respect to evolving user requirements, facilitate the management of changes to the domain models, and thus reduce the workload of curators and maintainers and the bottlenecks in the maintenance process, while keeping the door open for manual changes.

3.0.1. Deprecation.

Evo-KnoMo adds a usage measure layer on top of the model. In this way, the framework measures how much the current model entities are being used. This information is stored as weights – numeric values assigned to concepts and relationships. These weights change daily, taking into account the amount of usage of the concept or relationship in a certain lapse of time, and the time elapsed since the last use. Concepts and relationships taken into account in this usage measure are those to which the personal data is aligned. This measured is calculated as

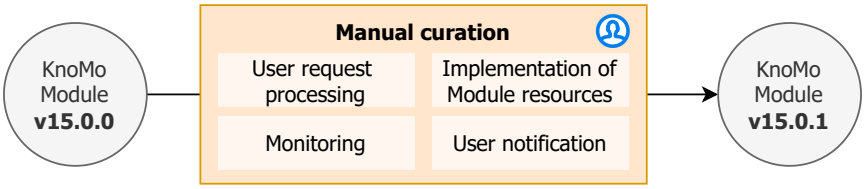
$$\mathscr{W} = \frac{U}{d + U} \tag{1}$$

where $\mathscr{W}$ is the weight measure, d is the days that have passed since the last use, and U is the total number of uses in the last 30 days.

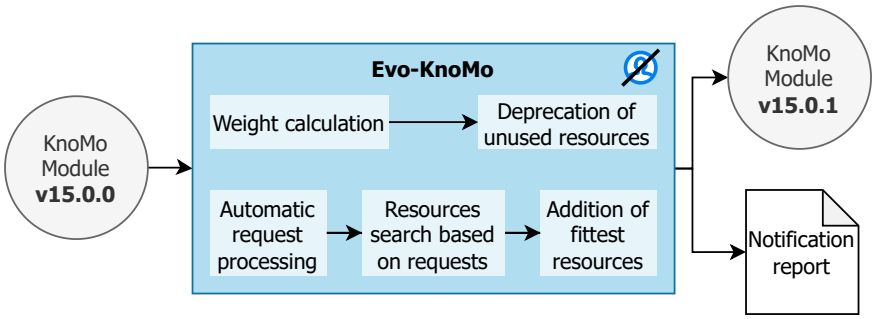
However, not all resources are subject to deprecation. There are relationships and entities that serve as a bridge between the different submodules within a KnoMo Module; and between different Modules, i.e. other modules have "dependencies" on these concepts/relationships. Thus, these bridges must be kept active and not deprecated. Evo-KnoMo outputs a notification to warn that the connection between the different modules may be improved to promote interoperability.

3.0.2. Addition.

In order to add new concepts and relationships that may be useful for each KnoMo Module, Evo-KnoMo registers every request that involves concepts or relationships that are not directly mapped in KnoMo. When similar requests are made, Evo-KnoMo starts a search process in the ontologies and graphs available on the web (e.g. Wikidata). Implementing meta-neuro-genetic algorithms [15], the platform tries different options and combinations of potential solutions until it finds one that best suits the user's requests. Once found, Evo-KnoMo implements the change into the module and outputs the notification.



(a) Current manual evolution workflow.



(b) Evo-KnoMo automated evolution workflow.

Figure 2: Evolution workflows in (a) current manual scenario and (b) automated with Evo-KnoMo.

To keep users in the loop, Evo-KnoMo notifies every change in the KnoMo module with a brief report explaining the criteria applied to make the change. This report is generated using the ExplaPy library[2] for explainability in combination with ChatOEG [16] for writing the report. This allows curators and maintainers to keep track of the current status of the module, and to modify the module updates when they are not accurate or correct.

4. Experimental Evaluation

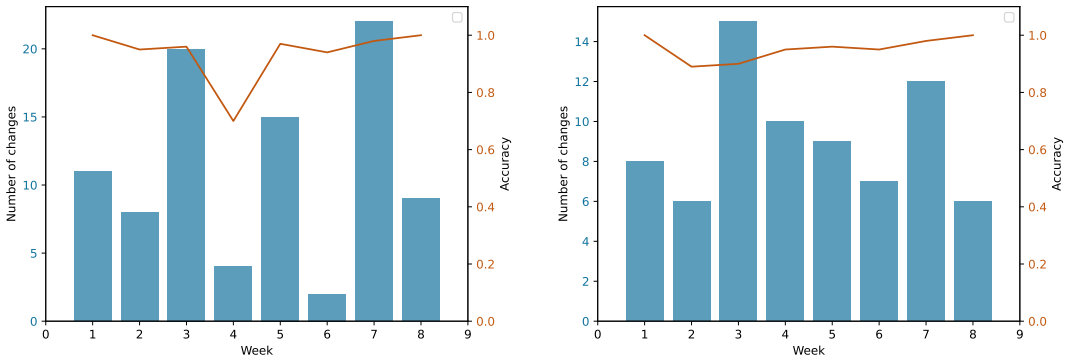
4.1. Modules

We use two KnoMo Modules for the evaluation: Healthcare and Transport. The Healthcare Module is used within hospitals and medical clinics to address clinical knowledge: personal health records, diagnosis, tests, drugs, treatments, etc. The Transport Module represents the public transport means for mobility within and among countries, the different lines of trains, subway, buses, time of arrival, current status, warnings, and payments.

4.2. Experimental Setup

We evaluate Evo-KnoMo in a user study from two perspectives. First, we assess the accuracy of the framework by comparing the results of the evolution of the two modules with Evo-KnoMo

[2]https://pypi.org/project/explapy/



(a) Healthcare KnoMo Module. **(b)** Transport KnoMo Module.

Figure 3: Accuracy results and number of changes per week in the (a) Healthcare and (b) Transport KnoMo Modules.

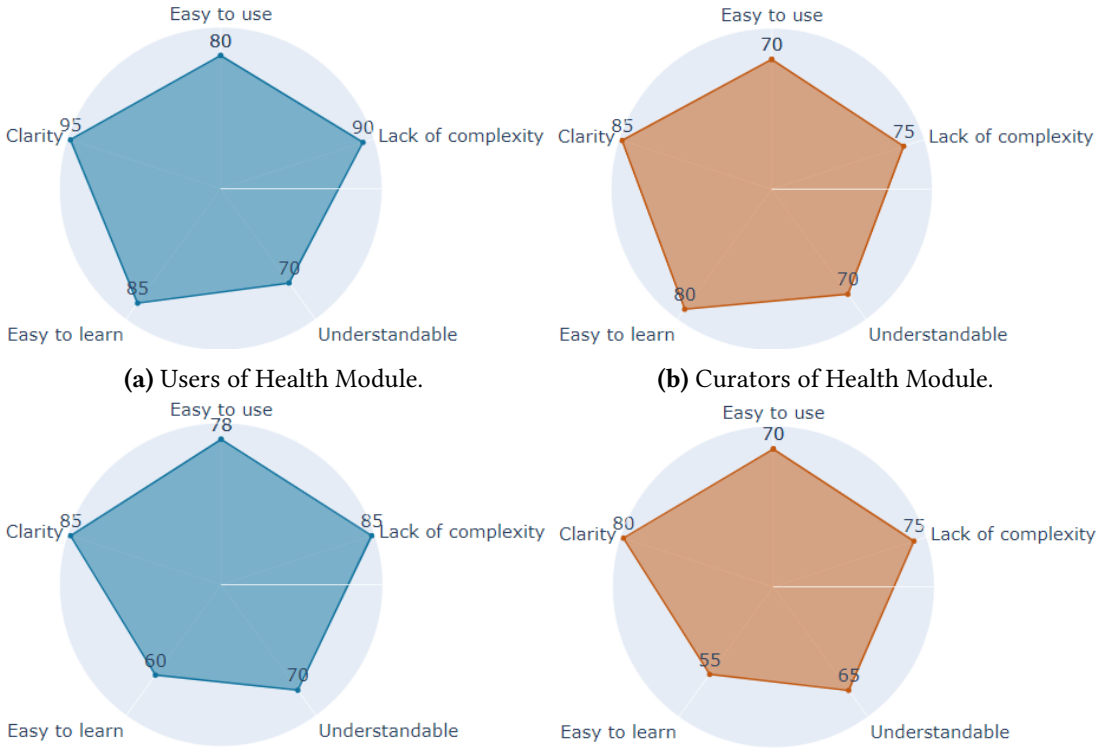
against the manual evolution performed by curators during an eight-week trial. We measure the number of changes per week, and the accuracy as the ratio between the correct evolution changes against the total of changes. Then, we ask curators and users to use the framework to later gather their impressions on the interaction with Evo-KnoMo. This is carried out by providing them with a usability questionnaire that considered 5 aspects: ease of use and learn, lack of complexity, understandability, and clarity. This two-fold evaluation took place separately for each module. For the Healthcare module, 5 institutions and a total of 63 curators participated; while 3 institutions and 55 curators participated for the Transport Module.

4.3. Results

Figure 3 presents the results obtained when evaluating the accuracy of the evolution changes performed by Evo-KnoMo. Overall, it can be observed that the accuracy rate for both modules is high, usually above 0.9. The framework performs exceptionally well for the Transport module (Figure 3b), in which its behavior is barely affected by the number of changes. For the Healthcare module (Figure 3a), in week 4 the accuracy drops to 0.7, but it is recovered in the following weeks. Regarding the usability for user interaction, Figure 4 shows that both users and curators are overall satisfied with Evo-KnoMo, but there is still room for improvement. We extract from these results that Evo-KnoMo is able to evolve accurately the KnoMo modules based on the user requirements, and that it can provide a good user experience.

5. Conclusions and Future Work

In this paper, we present Evo-KnoMo, an approach that empowers KnoMo with the ability to evolve dynamically to create and delete concepts or relationships. This evolution approach relies on the user-model interaction, so that the components that are not used for some time are deprecated; and new ones can be added with users requests that do not align with existing



(a) Users of Health Module.

(b) Curators of Health Module.

(c) Users of Transport Module.

(d) Curators of Transport Module.

Figure 4: Results of the level of satisfaction when using Evo-KnoMo framework for (a) users and (b) curators of the Healthcare Module; and (c) users and (d) curators of the Transport Module.

components. In addition, Evo-KnoMo implements a notification system to report the modifications that are performed automatically, and facilitates the manual edition when necessary, facilitating the curation task for modules maintainers. We test our approach with two modules: the Health and Transport KnoMo Modules. Evaluation results show that Evo-KnoMo is capable of detecting and evolving the model with a high accuracy rate. We can also extract from the user study that the users are mostly satisfied when interacting with the system.

As future work we plan to implement the Evo-KnoMo framework with the remaining modules while improving the evolution and notification techniques. We believe that this effort is a contribution to enhance Know4All, moving towards the complete democratization of data.

Acknowledgments

We want to thank Raúl García-Castro, Andrea Cimmino and Daniel Garijo for their feedback and insights, which made possible the development of this paper.

References

[1] J. Rojas, R. Verborgh, P. Colpaert, Unichip: Semantics for personal data in small devices, Data And Knowledge Engineering (2028).

[2] T. Alrashed, D. Paparas, O. Benjelloun, Y. Sheng, N. Noy, Goknow: The final platform for your data, in: International Semantic Web Conference, 2030.

[3] D. Garijo, P. Groth, A. Polleres, O. Corcho, M.-E. Vidal, A. Dimou, Know4all: The peer2peer eu public knowledge space, in: International Conference on Information and Knowledge Management, 2033.

[4] A. V. Sambra, E. Mansour, S. Hawke, M. Zereba, N. Greco, A. Ghanem, D. Zagidulin, A. Aboulnaga, T. Berners-Lee, Solid: a platform for decentralized social applications based on linked data, MIT CSAIL & Qatar Computing Research Institute, Tech. Rep. (2016).

[5] D. Garijo, P. Groth, A. Polleres, O. Corcho, M.-E. Vidal, A. Dimou, Catching up: Successes, challenges and lessons learned of a decade of know4all, Transactions on Graph Data and Knowledge (2040).

[6] A. Iglesias-Molina, D. Chaves-Fraga, M.-E. Vidal, A. Dimou, Knomo: Community-driven knowledge management platform for know4all, in: ESWC, 2040.

[7] M. Poveda-Villalón, Knoops! an knowledge pitfall scanner for knomo, in: International Semantic Web Conference, ISWC, P&D, 2035.

[8] D. Chaves-Fraga, A. Iglesias-Molina, J. Toledo, A. Dimou, Mappingpedia 2.0 – upload, store, and share your mappings!, Semantic Web (2038).

[9] M. Lenzerini, Data integration: A theoretical perspective, in: Twenty-first ACM SIGMOD-SIGACT-SIGART symposium, 2002, pp. 233–246.

[10] O. Corcho, F. Priyatna, D. Chaves-Fraga, Towards a new generation of ontology based data access, Semantic Web 11 (2020) 153–160.

[11] OpenAI, Gpt-4 technical report, 2023. arXiv:2303.08774.

[12] S. Geisler, M.-E. Vidal, A. Dimou, S. Kirrane, R. Filgueira, Data spaces: The story so far, Transactions on Graph Data and Knowledge (2035).

[13] J. Toledo, J. Rojas, M. Aguado, O. Corcho, Using semantics for enhancing the mobility data space, in: International Semantic Web Conference, 2033.

[14] Á. Sicilia, G. Nemirovski, Autoknomo: Automated modification of the knomo model, in: Knowledge Engineering and Knowledge Management: EKAW, 2041.

[15] D. Adams, Finding the answer to life the universe with meta-neuro-genetic algorithms, International Journal for Numerical Methods in Engineering (2042).

[16] C. Badenes-Olmedo, P. Calleja, P. Martin-Chozas, E. Montiel-Ponsoda, Writing in our own style with chatoeg, Transactions on Graph Data and Knowledge (2025).