



HAL
open science

Personal Knowledge Base Systems

Serge Abiteboul, David Montoya

► **To cite this version:**

Serge Abiteboul, David Montoya. Personal Knowledge Base Systems. PAP 2017, Personal analytics and privacy, Sep 2017, Skopje, Macedonia. hal-01592601

HAL Id: hal-01592601

<https://inria.hal.science/hal-01592601v1>

Submitted on 25 Sep 2017

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.

Personal Knowledge Base Systems

Serge Abiteboul

David Montoya

INRIA & ENS, Paris*
serge.abiteboul@inria.fr

Engie Ineo & ENS Cachan
david@montoya.one

September 24, 2017

The typical Internet user has personal data spread over several devices and across several online systems. Geeks already know how to control their personal data. It is now (or soon will be) possible for everyone to do the same, and there are many advantages to doing so. Everyone should now manage his/her personal information management system. We explain the main features of *personal information management systems* (sometimes called, self-data management systems). We consider advantages they provide from societal viewpoints. We argue that they are coming because of a number of reasons, political, economical, technical, and societal.

We believe that a main argument for their future success is that they enable new functionalities based on knowledge integration, and thus that the future systems are personal *knowledge* management systems. Such a system should integrate into a coherent whole data of very different nature, in particular, emails and other messages, calendars, contacts, GPS locations, web searches, bills, bank and medical records, etc. The integration of data that typically have to be exported from various systems should be performed by transforming all the information into a single knowledge base on a machine the user controls. The resulting system thus acts as a digital home for all the digital knowledge of the person.

This integration allows the user to query her personal information within and across different dimensions. It also enables performing analytics to learn from all this information. We believe that many users would be reluctant to give access to all their information to a single company, or if they did that, they would require strong guarantee of privacy. The fact that the integration, the query evaluation, and analysis happen on a system directly controlled by the user guarantees her privacy.

To illustrate the possibly very rich derivation of knowledge, suppose that a user, say Alice, met a friend, say Bob. The agenda indicates the time of a meeting with “Bob”. Some text messages may remove the ambiguity on which

*Département d’informatique de l’ENS, École normale supérieure, CNRS, PSL Research University, 75005 Paris, France

Bob she met (Alice may know several Bobs). The phone number he used has long been aligned with a specific Bob. The location of Alice, as provided by her phone's GPS, provides the name of the place where they met. Bob's timeline in his social network provides pictures of the event. Finally, from her web search history, we can even learn about topics they discussed. A meeting that had traces in different systems turned into a rich event in the KB.

Designing such a personal KB is not easy: Data of completely different nature has to be modeled in a uniform manner, pulled into the knowledge base, and integrated with other data. Different from related work in the area of information integration, we cannot rely on outside sources, we cannot expect redundancy, and we have a very limited and particular set of attributes.

We will mention such a system, the Thymeflow system that for instance aligns events based on time, e.g., a meeting in the user's calendar and a GPS position, or on space, e.g., the address of a contact and a GPS position.

Acknowledgments We will discuss works performed with colleagues that we would like to thank:

- Benjamin André, Cozy Cloud
- Daniel Kaplan, Fing
- Amélie Marian, Rutgers University
- Thomas Pellissier Tanon, ENS Lyon
- Pierre Senellart, ENS
- Fabian M. Suchanek, Telecom ParisTech

The bibliography provides references to these works. More references can of course be found in these papers.

References

- [1] Serge Abiteboul, Benjamin Andr, Daniel Kaplan: Managing your digital life. *Commun. ACM* 58(5): 32-35 (2015)
- [2] Serge Abiteboul, Amlie Marian: Personal Information Management Systems, Tutorial, EDBT/ICDT Conference, 2015.
<https://www.slideshare.net/ameliemarian>
- [3] David Montoya, Thomas Pellissier Tanon, Serge Abiteboul, Fabian M. Suchanek: Thymeflow, A Personal Knowledge Base with Spatio-temporal Data. *CIKM 2016*: 2477-2480, 2015
- [4] David Montoya, Serge Abiteboul, Pierre Senellart: Hup-me: inferring and reconciling a timeline of user activity from rich smartphone data. *SIGSPATIAL/GIS 2015*: 62:1-62:4