"Who am I? And if so, how many?"
The E-PIX as innovative system to manage person identities

Bialke M, Langner D, Geidel L, Bahls T, Piegsa J, Havemann C, Hoffmann W

Institute for Community Medicine, Section Epidemiology of Health Care and Community Health, University Medicine Greifswald
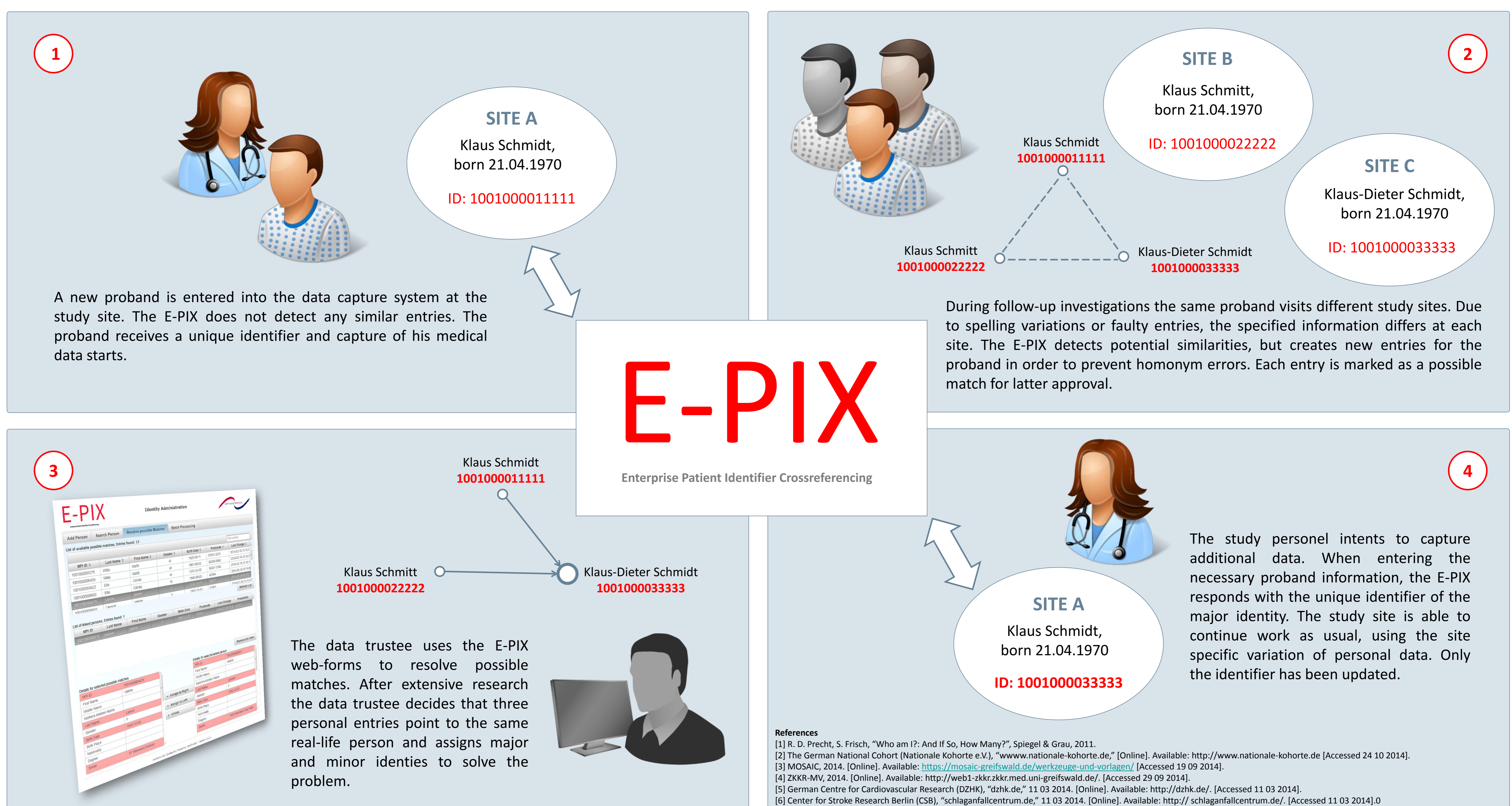
Background

Precht raised the question *"Who am I? And if so, How Many?"[1]*. Although it originally relates to multiple identities in a psychological context, this question invokes a central issue in the identity management in medical research. Rapidly growing studies and registers require an efficient data management. This comprises the electronic management of probands and patients. In order to support automated data processing and to avoid duplicate entries during the consolidation of research data from several study sites, each participant has to be faultlessly identifiable using a unique identifier (ID). Particularly, in larger projects (e.g. the German National Cohort [1] with about 200.000 probands) data management has to be fault-tolerant and reliable. Especially long-term prospective projects with varying data sources often have to deal with datasets with participant identities differing slightly, but still describing the same real-life person. This may occur as a result of changes in addresses (e.g. moving) or names (e.g. marriage/divorce). This necessitates both flexible and precise management of person identities. Yet, existing tools for proband management are able to store a single dataset per person only.

However, the possibility to assign different identities to a single person is crucial for cohort studies and registries, where keeping variations in spelling (e.g. for names, addresses, etc.) is inevitable.

Methods

The Institute of Community Medicine has developed a tool that efficiently facilitates the management of multiple person identities. The E-PIX (Enterprise Patient Identifier Cross Referencing) allows the fault-tolerant consolidation of identifying data. In detail, for each new identifying dataset, a new identity will be created internally. These identities are grouped and consolidated to a single person. Thus, each variation in spelling can be captured permanently. For each person a major identity and an arbitrary number of assigned minor identities are defined. The major identity represents the actual status of a person. The assignment of minor identities to a single major identity is adjustable at any time, but is restricted to authorized personnel after extensive interactive reviewing of primary data.

1 SITE A — Klaus Schmidt, born 21.04.1970 — ID: 1001000011111

A new proband is entered into the data capture system at the study site. The E-PIX does not detect any similar entries. The proband receives a unique identifier and capture of his medical data starts.

2 SITE B — Klaus Schmitt, born 21.04.1970 — ID: 1001000022222; SITE C — Klaus-Dieter Schmidt, born 21.04.1970 — ID: 1001000033333; Klaus Schmidt 1001000011111; Klaus Schmitt 1001000022222; Klaus-Dieter Schmidt 1001000033333

During follow-up investigations the same proband visits different study sites. Due to spelling variations or faulty entries, the specified information differs at each site. The E-PIX detects potential similarities, but creates new entries for the proband in order to prevent homonym errors. Each entry is marked as a possible match for latter approval.

E-PIX — Enterprise Patient Identifier Crossreferencing

3 Klaus Schmidt 1001000011111; Klaus Schmitt 1001000022222; Klaus-Dieter Schmidt 1001000033333

The data trustee uses the E-PIX web-forms to resolve possible matches. After extensive research the data trustee decides that three personal entries point to the same real-life person and assigns major and minor identies to solve the problem.

4 SITE A — Klaus Schmidt, born 21.04.1970 — ID: 1001000033333

The study personel intents to capture additional data. When entering the necessary proband information, the E-PIX responds with the unique identifier of the major identity. The study site is able to continue work as usual, using the site specific variation of personal data. Only the identifier has been updated.

References
[1] R. D. Precht, S. Frisch, "Who am I: And If So, How Many?", Spiegel & Grau, 2011.
[2] The German National Cohort (Nationale Kohorte e.V.), "wwww.nationale-kohorte.de," [Online]. Available: http://www.nationale-kohorte.de [Accessed 24 10 2014].
[3] MOSAIC, 2014. [Online]. Available: https://mosaic-greifswald.de/werkzeuge-und-vorlagen/ [Accessed 19 09 2014].
[4] ZKKR-MV, 2014. [Online]. Available: http://web1-zkkr.zkkr.med.uni-greifswald.de/. [Accessed 29 09 2014].
[5] German Centre for Cardiovascular Research (DZHK), "dzhk.de," 11 03 2014. [Online]. Available: http://dzhk.de/. [Accessed 11 03 2014].
[6] Center for Stroke Research Berlin (CSB), "schlaganfallcentrum.de," 11 03 2014. [Online]. Available: http:// schlaganfallcentrum.de/. [Accessed 11 03 2014].0

Results and Conclusion

The E-PIX enables precise person management following the principles of a master patient index and allows the utilisation of several assigned identities. The assignment process itself is designed to be both flexible and revertible. In addition, the E-PIX provides automated historization mechanisms in order to document essential user decisions, which is of special importance for analyses, assurance of data quality and improvement of process transparency.

The E-PIX has a service-oriented architecture and a central service-interface to provide all necessary functionality. The functionality can be deployed comfortably through web-based user-interfaces (e.g. to resolve possible matches). The E-PIX is open source software [3] and is provided by the DFG-funded MOSAIC-project for free for non-commercial and commercial use. It has already successfully implemented in several projects including the German National Cohort, the Central Clinical Cancer Registry in Mecklenburg-Western Pomerania [4], the German Centre for Cardiovascular Research [5] and is presently implemented at the Centre for Stroke Research Berlin at the Charité (University Medicine Berlin) [6].

Contact

Martin Bialke, M. Sc.
Institute for Community Medicine
Section Epidemiology of Health Care and Community Health
University Medicine Greifswald
Ellernholzstrasse 1-2, 17487 Greifswald, Germany
Tel.: +49 3834 / 86-7580, Fax: +49 3834 / 86-7752
E-Mail: martin.bialke@uni-greifswald.de

Download this poster from
mosaic-greifswald.de