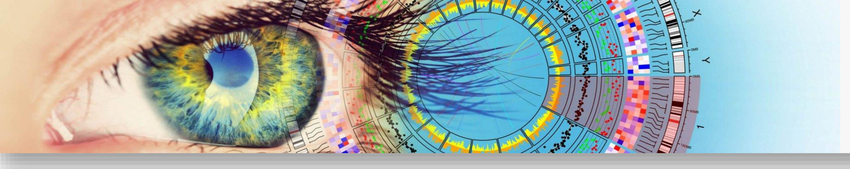
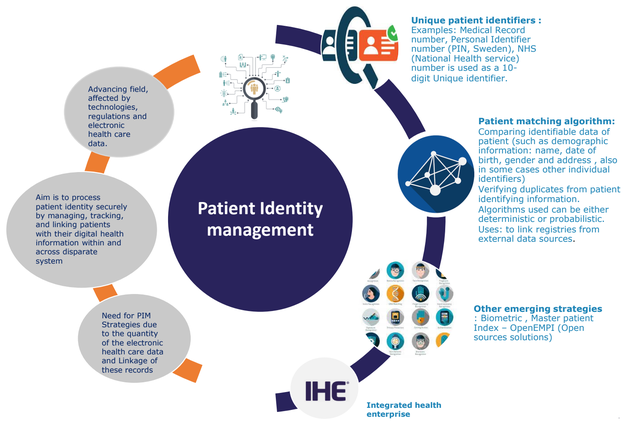


What are the approaches to mitigate patient identity-related challenges in an attempt to achieve interoperability between different healthcare entities?

Anju Philip, Jonathan Okereke, Prof. Dr. med. Georgi Chaltikyan, European Campus Rottal-In, Technische Hochschule Deggendorf, Uniklink Köln.

Introduction

Background



Problem/Challenges

Local (MeDIC) Problems:

- Semantic interoperability or ability to share digitalized healthcare information across disparate system.
- Availability of clinical data (used during routine care) made ready for secondary use (Research and Big data analytics)

MeDIC : Medical Data Integration center in Uniklinik Köln

Federated (HiGHmed) Problems:

- De-identification of patient demographic information to enable federated sharing.
- Elimination of possible duplicates of patient records in before a Virtual Molecular Tumour board.
- Cross-Institutional identification of patients/subjects participating in a project (clinical trial)
- Authentication and Authorization of Users

HiGHmed: One of the Consortiums of Medical informatic initiatives.

Goals and Objectives

Goals

- Aim of this research work is to identify the patient identity-related challenges within a healthcare infrastructure.
- Mitigating those challenges by proposing solutions, based on understanding the gaps and requirements in a healthcare infrastructure.

Objectives

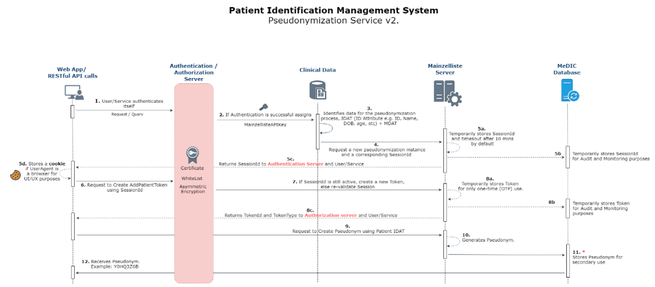
- Understanding the current infrastructure of the hospital.
- Planning, designing and implementing the patient identity solution based on the requirements.
- Proposing a solution for data security and privacy of the patient identity.
- Comparing other industry standard solutions.

Reference:

Gliklich, R. E. (2014, April 17). Managing Patient Identity Across Data Sources.https://www.ncbi.nlm.nih.gov/books/NBK208618/

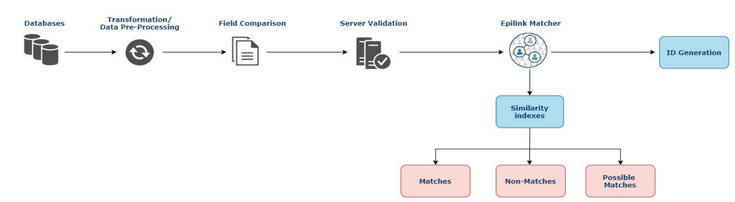
Methodology

Implementation of pseudonymisation service



Note. ©Copyright 2019 MeDIC Infrastructure Uniklinik Koeln

Process flow of de-duplication service (Record-Linkage)



Process flow of Record-Linkage service

Datasets

Primary Datasets:

Name Data:

- Prefix
- Suffix
- Firstname
- Lastname
- Birthname

Demographic data:

- Date of Birth
- Gender
- Street Address
- Postcode
- Place
- Insurance number

Altered Dataset

The altered dataset consists of datasets modified from the *primary datasets* based on difference scenarios which are generally encountered during routine care, especially, during patient on-boarding. Scenarios such as :

Scenario 1: Interchanging firstname, lastname and birthname which exists in the database for selected patients and keeping other attributes (i.e. date of birth, postcode and city) of the same patients.
For example, if the Patient details are firstname: Anju Lastname: Philip, Birthname: Anju, date of birth: 05.07.1970, Postcode: 96403 and City: Coburg, then firstname, lastname and birthname is completely interchanged with another patient which is existing in the database.

Scenario 2: Positional changes: Changing firstname with lastname and vice versa, positional changes within the field, if there are two or more names in one field. For example: Firstname is Josef Richard Siegfried and Lastname is: Otto, Positional changes within the field would be: Richard Josef Siegfried

Scenario 3: Alteration in Birthname, by replacing completely, changing letters, addition of letters or omitting of letters. For example, Birthname is Anju, it was altered in the following ways: Anj or Anu or Annju or Anjuu

Scenario 4: Modification in Date of Birth was made by following categories:
Modification in : Date, month, year, data and year, date and month, month and year.

Scenario 5: Inclusion of characters in the firstname, lastname and birthname such as letters, numbers. For example, Firstname: Anju Lastname: Philip , manipulated in the following ways: Anju1, Phillip , Annju Phllip, Anju Philip22.

Scenario 6: Addition of spaces in firstname, lastnames, birthname and city. For example, data will altered in such way: Firstname: Anju , Anj u Lastname: Philip, Phili p etc.

Scenario 7: Addition of Prefixes and suffixes in firstname and lastname. For example, Firstname: Dr. Anju Lastname: Philip B.E M.Sc Phd.

Scenario 8: Alteration in the German special charaters (umlaut) in a field such as ü,ö,ä and ß to english characters. For example: ü -> ue , ö-> oe , ä -> ae and ß -> ss

Scenario 9: Inclusion of Delimiters in the Identifying attributes such as dots, umlauts, comas, and dashes.
Scenario 10: Leap year as date of birth. For instance, if there is an entry 29 Feb as date and month, then the year is not a leap year. For example: 29 Feb 2015, 2015 is not leap year.
Scenario 11: Similarity in firstname, lastname and date of birth and other attributes (birthname, postcode and city) are different.

Results

- Pseudonym is generated as unique identifier from patient demographic information and Mainzelliste generates the corresponding weight.
- Record linkage experiment was conducted based on defined scenarios and corresponding weight generation.

Data Analysis

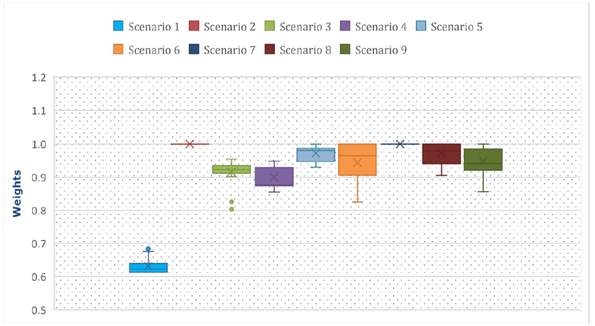
Variations were observed for some of scenarios excluding scenario 2 and 7. The below plot shows the mean, median, upper and lower limits and any outliers for each scenario
Margin of Error for CI 95% : ±2.93% (highest) - Mainzelliste service is able to detect duplication and other data entry errors not just in attribute level, but also in character level.

Possible outliers and their Rationale:

Scenario 1:
Outlier Value: 0.6835990573548800(that is higher than upper bound value) was received for the below altered data, this was due to similarity in birthdate of the patient as shown in the below primary dataset table.

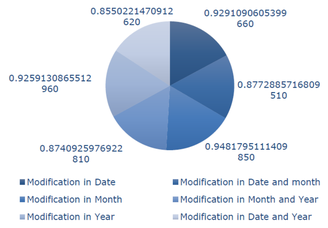
Scenario 3: Changes in Birthname
Outlier Value: 0.8032834783128977 (lower than the lower bound value) was received for the below altered data, since there was complete replacement of birthname, due to which the weight value was lower than other weight values in the scenario.

Summarized plot of all the scenarios with mean, median, upper and lower limits and outliers



Relationships and Trends

Correlations in date of Birth



Trends

Trend 1 : Mainzelliste always gives result "Found match with IDxx" when the weight value is 1,0. But there also cases where in Mainzelliste gives weighted value less than 1,0 and returns a outcome "Found match with IDXX"(Scenario 2,3,5,6,8,9).

Trend 2 : When completely or partially replacing firstname, lastname, birthname date of birth, city or including extra characters in these fields(Scenario 1,3,4,6,8,9) Mainzelliste is unable to find an exact match, but it provides a weight value, indicating there could be a patient with same demographic data, and to verify the data entered.

Trend 3 : If there are special characters (Scenario 7,9) or number (Scenario 5) or non leap year date and month entries (Scenario 10), then Mainzelliste is able to detect it as bad request or wrong format.

Trend 4: Similarities in firstname, lastname and date of birth but other attributes(birthname, postcode, city) are different (Scenario 11), then Mainzelliste creates a new pseudonym, pointing out that Mainzelliste is able to perform similarity index and finds it as unique patient.

Conclusion

This research work primarily started with reviewing various patient identity management strategies and standards used to share information across systems, such as IHE, MPI etc. The objective was to conceal patient demographics and to use their clinical data for research purposes, continuity of care and its importance is highlighted throughout this work. This research also attempts to validate the operations of the Mainzelliste (Software by TMF-BMBF) based on specific scenarios. The rationale for the defined scenarios are due to some common errors associated with patient identity. Extensive data analysis of conducted experiments, which showed the reliability and validity of the service. Some limitations with pseudonymization service were project-specific pseudonym, addition of identifiers to improve generation of pseudonyms for special scenarios, pre-defined data types in some of the fields, and scalability of record-linkage service which was not evaluated by this work. The methodology was implemented based on the current usecase and requirement of MeDIC (Uniklinik Köln) and HiGHmed.

Medical informatic initiative, HiGHmed is establishing data sharing framework across other consortium partners for continuity of care, research purposes and patient privacy. In an open community, this service can be used by any researchers and implement further enhancements based on their requirement.

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
MEDIZIN INFORMATIK INITIATIVE