

CancerLink – Use of PPRL approaches for data linkage in three selected cancer registries

Hampf C¹, Krauß A¹, Stein A¹, Schuldts R², Wöller N², Theen S³, Dettmer D⁴, Mitnacht L⁵, Bialke M¹, Leddig T¹, Weitmann K¹, Stahl D², Hoffmann W¹

¹ Institute for Community Medicine, Section Epidemiology of Health Care and Community Health, University Medicine Greifswald, Greifswald

² Trusted Third Party of the University Medicine Greifswald, Greifswald, Germany

³ Cancer Registry Mecklenburg-Western Pomerania, Institute for Community Medicine, University Medicine Greifswald, Greifswald

⁴ Clinical Cancer Registry of Lower Saxony, Hannover

⁵ Cancer Registry Rhineland-Palatinate, Mainz, Germany

Introduction

Data from cancer registries are an important source of information for cancer research. Optimal use of the information collected by the registries requires the option to request data from various cancer registries simultaneously. Currently, there is no systematic method to detect duplicates for record linkage. As a result, the data sets may contain unrecognized homonym errors or false-positives (entries are incorrectly assigned to the same patient, although they actually refer to different patients) and synonym errors or false-negatives (entries are not merged, although they actually belong to the same patient). Those errors can lead to problems in the analysis, and thus significantly impair the quality and informative value of the data set. For this project, data on selected cancer entities (ICD-10: C18, C19, C20, C34, C43, C53, C54, C55, C61) with a diagnosis in 2018-2024 were requested from three cancer registries: Mecklenburg-Western Pomerania, Rhineland-Palatinate, and Lower Saxony.

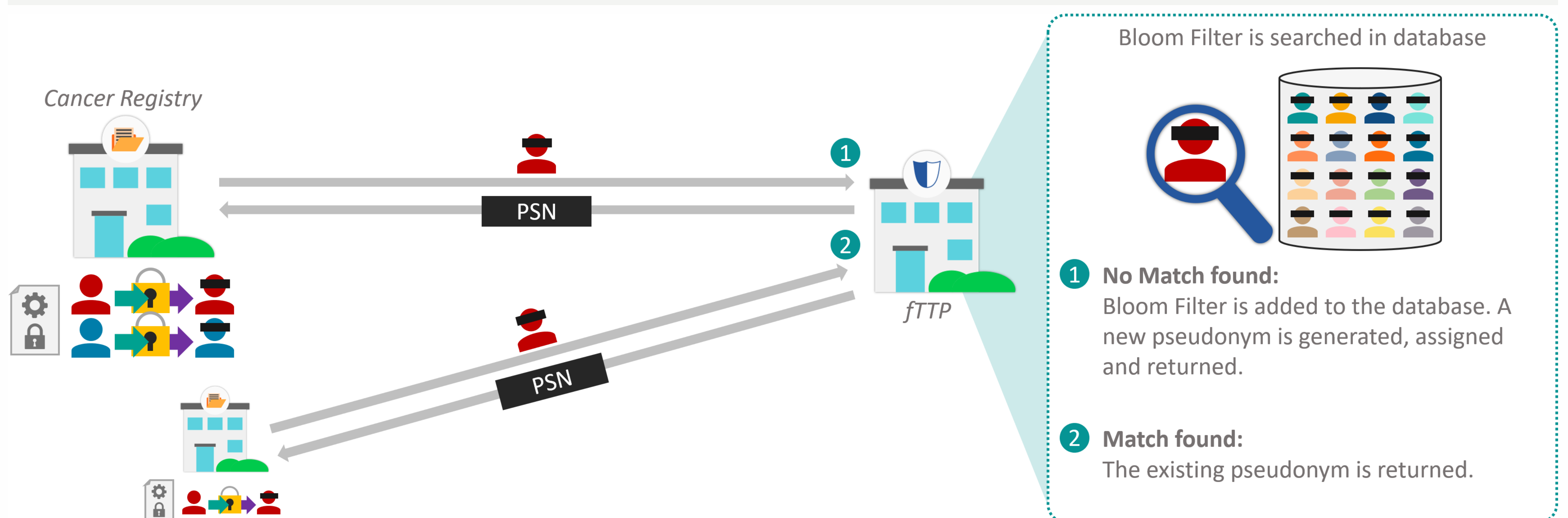


Figure 1: Cancer registries transform person identifying information (PII) to project specific bloom filters. Bloom filters are transferred to the fTTP. The fTTP compares the bloom filters and generate and assign pseudonyms. The pseudonyms are used to verify potential matches between the cancer registries. The verification is based on PII.

Methods

A privacy-preserving record linkage was implemented with help of a federated Trusted Third Party (fTTP)^{1,2}. The cancer registries generated a bloom filter for each person based on personally identifiable information (PII). Four test cases with different bloom filter configurations were conducted in which up to 250,000 data sets were transferred to the fTTP. The fTTP compared the bloom filters, detected potential matches, assigned pseudonyms and returned the pseudonyms to the cancer registries. The cancer registries used the PII to check whether a potential match found is a match and report the result back. Cancer registries exchange patient data at least twice a year between the patient's place-of-residence registry and the patient's place-of-treatment registry via the cross-registry data exchange (RÜD). Patients involved in the exchange are therefore listed in both registries. The RÜD information was subsequently used to determine whether a confirmed match was previously known or unknown.

Discussion

This project implemented a random hashing² for generating bloom filters and generating balanced bloom filters² for additional security (hardening). Each configuration used different attributes (first name, surname, birthdate, postal code), which were equally included in the hash. In future projects, configurable proportions are desirable, so that specific attributes can be evaluated individually.

Results

The fTTP was able to find matches that are not covered by the RÜD information including previously unknown matches within the individual cancer registries. A very low threshold for potential matches was chosen for the test cases. This resulted in a high number of potential matches that are actually no matches. For example in one test case 536 potential matches resulted in only 13 matches. For future projects, a stricter threshold can prevent this high number of potential matches. In conclusion, the fTTP is able to find matches based on bloom filters, which allows personal data to be compared in a secure manner.

References

1. Hampf C, Bialke M, Hund H, Fegeler C, Lang S, Penndorf P, et al. Privacy-preserving record linkage by a federated trusted third party (fTTP) – unlocking medical research potential in Germany. GMS Medizinische Informatik, Biometrie und Epidemiologie [Internet]. German Medical Science GMS Publishing House; 2025;21. <https://doi.org/10.3205/MIBE000277>
2. Kästner A, Hampf C, Naumann P, Fiedler-Lacombe L, Kron A, Spier A, et al. Concept and feasibility of privacy-preserving record linkage of cancer registry data and claims data in Germany: results from the DigiNet study on stage IV non-small cell lung cancer. J Cancer Res Clin Oncol. 2025;152:6. <https://doi.org/10.1007/s00432-025-06384-7>
3. Schnell R, Borgs C. Randomized Response and Balanced Bloom Filters for Privacy Preserving Record Linkage. 2016. p. 218–24. <https://doi.org/10.1109/ICDMW.2016.0038>

Funded by

