Federated Trusted Third Party as an Approach for Privacy Preserving Record Linkage in a Large Network of University Medicines in Pandemic Research

Christopher Hampf (christopher.hampf@uni-greifswald.de)
Universitatsmedizin Greifswald https://orcid.org/0000-0002-4557-4783

Martin Bialke
Universitätsmedizin Greifswald: Universitatsmedizin Greifswald

Hauke Hund
Heilbronn University: Hochschule Heilbronn

Christian Fegeler
Heilbronn University: Hochschule Heilbronn

Stefan Lang
Gefyra Gmbh

Peter Penndorf
Universitätsmedizin Greifswald: Universitatsmedizin Greifswald

Nico Wöller
Universitätsmedizin Greifswald: Universitatsmedizin Greifswald

Thomas Bahls
Universitätsmedizin Greifswald: Universitatsmedizin Greifswald

Frank-Michael Moser
Universitätsmedizin Greifswald: Universitatsmedizin Greifswald

Lars Geidel
Universitätsmedizin Greifswald: Universitatsmedizin Greifswald

Arne Blumentritt
Universitätsmedizin Greifswald: Universitatsmedizin Greifswald

Ronny Schuldt
Universitätsmedizin Greifswald: Universitatsmedizin Greifswald

Florian Seidel
Charité Universitätsmedizin Berlin: Charite Universitatsmedizin Berlin

Peter Brunecker
Charité Universitätsmedizin Berlin: Charite Universitatsmedizin Berlin

Reto Wettstein
University Hospital Heidelberg: UniversitatsKlinikum Heidelberg

Lukas Arnecke
University Hospital Heidelberg: UniversitatsKlinikum Heidelberg
Wolfgang Hoffmann
Universitätsmedizin Greifswald: Universitatsmedizin Greifswald

Research Article

Keywords: federated Trusted Third Party, Privacy Preserving Record Linkage, Identity Management, Duplicate detection, Data quality, data sharing, data transfer, Pseudonymisation

Posted Date: December 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1053445/v1

Abstract

Background

The Federal Ministry of Research and Education funded the Network of University Medicine for establishing an infrastructure for pandemic research. This includes the development of a COVID-19 Data Exchange Platform (CODEX) that provides standardised and harmonised data sets for COVID-19 research. Nearly all university hospitals in Germany are part of the project and transmit medical data from the local data integration centres to the CODEX platform. The medical data on a person that has been collected at several sites is to be made available on the CODEX platform in a merged form. To enable this, a federated trusted third party (fTTP) will be established, which will allow the pseudonymised merging of the medical data. The fTTP implements privacy preserving record linkage based on Bloom filters and assigns pseudonyms to enable re-pseudonymisation during data transfer to the CODEX platform.

Results

The fTTP was implemented conceptually and technically. For this purpose, the processes that are necessary for data delivery were modelled. The resulting communication relationships were identified and corresponding interfaces were specified. These were developed according to the specifications in FHIR and validated with the help of external partners. Existing tools such as the identity management system E-PIX® were further developed accordingly so that sites can generate Bloom filters based on person identifying information. An extension for the comparison of Bloom filters was implemented for the federated trust third party. The correct implementation was shown in the form of a demonstrator and the connection of two data integration centres.

Conclusions

This article describes how the fTTP was modelled and implemented. In a first expansion stage, the fTTP was exemplarily connected through two sites and its functionality was demonstrated. Further expansion stages, which are already planned, have been technically specified and will be implemented in the future in order to also handle cases in which the privacy preserving record linkage achieves ambiguous results. The first expansion stage of the fTTP is available in the University Medicine network and will be connected by all participating sites in the ongoing test phase.

Background

The Network of University Medicines and the Medical Informatics Initiative promote medical research in Germany.

Motivated by the worldwide Corona pandemic, the Network of University Medicine (NUM), funded by the Federal Ministry of Research and Education (BMBF), is establishing infrastructures to bundle research at all university medicines across Germany. NUM aims to improve national pandemic research and the care of COVID-19 patients in the long term. To this end, COVID-19 related health data (for example, GECCO83 data set (1)) are to be used for medical research. In this way, 13 inter-clinical research projects (2) should be able to gain rapid and robust findings and to answer urgent research questions (3).

The aim of the CODEX (COVID-19 Data Exchange Platform) project is to set up a central COVID research platform to make data sets such as clinical data, image data and data on biospecimens available for research in a multicentre, patient-related and pseudonymised manner (4).

Essential for the central merging of medical patient data from more than 30 NUM locations in Germany is a fault-tolerant and secure record linkage process conforming to applicable legal regulations (EU-GDPR). This includes the data protection-compliant matching of person identifying information (PII) (without sharing of demographic information such as given name, family name, gender, etc.) and the correct linkage of COVID-19 and pandemic data distributed across several NUM-sites. This Privacy Preserving Record Linkage (PPRL) is a central component of CODEX and is to be implemented by establishing a federated Trusted Third Party (fTTP) within the framework of NUM-CODEX.

The Medical Informatics Initiative (MII) is funded by the BMBF as well. The focus of the MII is on improving patient care and medical research by means of innovative software solutions (5). A central point is the establishment of a data integration centre (DIC) at each site (5). This includes a local Trusted Third-Party (TTP) at each site, which: (a) matches, merges and manages person identifying information (record linkage); (b) generates and manages pseudonyms; (c) manages the patients' consents and withdrawals; and (d) coordinates local withdrawal processes, if requested by the data subject. Common TTP-components are an identity management, a pseudonym management and a consent management. Each MII site is free to choose necessary measures (and software tools) to implement these TTP-components.

The challenge of federated record linkage

In both NUM and MII, the challenge is to identify (potentially) similar persons across multiple sites in order to merge and scientifically use related medical data. Typically, a patient is identified at one site by PII. In order to merge data from several sites, a defined set of PII is matched with each other using a record linkage procedure. This record linkage process is an essential part of the identity management (6) and enables researchers to merge large data sets while, at the same time, ensuring the correct reference to the individual persons. This reference is pseudonymised in the later process.

In the research context, more than one pseudonym is typically generated for each patient and site. Thus, pseudonyms can be generated specifically for individual research projects and purposes and linked to the very same patient. With each pseudonymisation step, a hierarchy of pseudonyms emerges, consisting of

automatically generated multi-level-pseudonyms. These complex pseudonym structures prevent unauthenticated persons from findung out the real person concerned. If necessary, however, the pseudonym can be dissolved by the TTP. This may be necessary, for example, in the case of an incidental finding, which a patient shall be informed about or if a patient shall be contacted for a recruitment for new studies (if consented).

Record linkage and the assignment of pseudonyms are basic procedures that can also be applied across sites. This can be implemented, for example, with a central TTP. This approach is commonly applied (7, 8), but requires the sharing of PII from the site per se, which is not intended within MII and NUM by default.

In the context of MII and NUM, merging of person identifying information has to comply with the requirements of PPRL. By design, within the PPRL-process, the identity of a person remains hidden from any central instance or other sites. For this purpose, the MII Taskforce for Data Protection has developed and systematised implementational concepts for cross-site record linkage and the establishment of an fTTP in the context of PPRL (9). These concepts are implemented accordingly based on existing identity management software components.

Bloom Filters as an established basis for PPRL

Within the framework of NUM, the concept developed by the MII was concretised. Two concepts utilise so-called Bloom Filters (BF) (10-12) as implementational basis. The record linkage process is performed based on BF, instead of the (clear text) PII. An essential part of the concept is the local Trusted Third-Party, which is established at each NUM site and allows a local record linkage. Moreover, the local TTP prepares the cross-site record linkage and generates for each patient (and available PII) corresponding BFs. While the responsibility of the PII remains at the respective site, the required cross-site record linkage is ultimately carried out by the fTTP for NUM.

A BF is a bit vector of length n and is initially occupied by n zeros (11). The BF is a data structure in which information is hashed so that the positions in the BF are set to ones. BFs can be implemented in such a way that only one element is encoded at a time (e.g. the first name) (13). However, this allows cryptographic attacks (12, 14, 15), which is why several or all attributes are encoded in one BF to reduce these attacks. This is called a Cryptographic Long-term Key (CLK) (8). To encode attributes in a BF, the individual attributes are split into n-grams (e.g. bigrams with the authors' name "Hampf": _H, HA, AM, MP, PF, F_) and then each individual n-gram is encoded into several bit positions, depending on the procedure, which are then set to 1 (12). To a certain extent, desired collisions occur within the BF, so that no conclusions can be drawn about the original input data. To check whether an element is contained in a BF, it is encoded with the corresponding procedure. If the bit positions match, the element is probably contained. This is not guaranteed, however, since these bit positions may also have been set by other elements. On the other hand, the presence of an element can be excluded if the bit positions differ at least

in one place. To detect contained differences, two BFs are checked for similarity. The more bit positions of two BFs are similar, the higher the probability that the input data was similar or the same.

As with a common record linkage process with PII, the identity management solution that performs PPRL-process classifies the two matched BFs into Match, Possible Match or No Match (6). In the case of a Possible Match, in the common record linkage process a manual resolution is performed by comparing the matching data and correcting it, if necessary. By design, this is not possible with a procedure using only BFs. Thus, the identity of a person is protected, but in cases where no clear assignment can be made, data cannot be merged correctly. Thus, in the case of a possible match with a BF, a following record linkage using PII may be necessary for selected cases (9). However, this execeptional case requires transmission of the PII to the fTTP.

Objectives

This publication describes the detailed conception and implementation of an fTTP for the NUM-CODEX project based on already consolidated concepts of the MII and the existing record linkage solution E-PIX®. The fTTP shall support both a cross-site record linkage with encoded PII (BFs) and, if necessary, a selective matching of PII. Finally, a publicly conducted demonstrator-event for all participating NUM sites shall provide proof for the successful implementation of the fTTP (in a first stage).

Methods

Delimitation of probability and clearing

One of the concepts followed by the MII provides an fTTP for record linkage. However, this concept only shows the required components, but not their specific implementation. For this reason, these components were specified and implemented both technically and organisationally as part of the NUM-CODEX project. In detail, the fTTP is based on two technically separate components, fTTP (probability) and fTTP (clearing).

(1) The fTTP (probability) performs record linkage based on BFs. Project and site-specific pseudonyms are generated and managed. In addition, pseudonyms can be re-pseudonymised.

(2) The fTTP (clearing) performs record linkage on temporarily cached PII. This operation is triggered only when a record linkage based on BFs within the fTTP (probability) reach a possible match. For projects with large cohorts, this component can be omitted if needed. Projects with a small number of participants (e.g., in the context of rare diseases) may require exact matching, so fTTP (clearing) can be added for quality improvement.

The fTTP (probability) represents the first stage, which is sufficient in NUM-CODEX for initial data transmissions. After successfull roll-out of the first stage, the fTTP (clearing) is establisehd as the second

stage at a later time.

Identification and coordination of use cases and processes

NUM-CODEX consists of central components, which include the fTTP, the GTH and the central platform (CODEX). These central components are available for the decentralised structures consisting of 34 DICs. The fTTP of the first stage basically requires two interfaces (to the sites and to the GTH). The necessary interfaces are shown in Figure 1.

The local TTP and the fTTP each provide data trustee capabilities, but operate in different modes. The local TTP may manage a person's PII for the site and performs record linkage based on PII.The fTTP per design does not store PII, but only manages BFs. BF can be categorised as person-relatable data, however, do not belong to PII. The local TTP generates a BF from the PII and transmits it to the fTTP (IF-1). The basis for the transmission of the BF is a corresponding informed consent from the patient. In case of an ambigious record linkage result in the fTTP, the PII can be requested from the local TTP for an additional classic record linkage procedure (9).

Data transfers from a DIC to the central platform require an additional pseudonymisation step (IF-2). The required pseudonyms are to be generated and managed within the fTTP.

In sum, two use cases were identified and modeled that need to be considered for fTTP implementation.

(UC1) Registration of persons and assignment of pseudonyms

Use case 1 comprises the registration of persons triggered by a site. In this case, a BF is transmitted to the fTTP. The fTTP performs a BF-based record linkage and assigns corresponding pseudonyms. During initial registration, two pseudonyms are generated:

1. The *DIC PSN* is a site-specific pseudonym that is only disclosed to the site sending the Bloomfilter. The site uses this pseudonym to reference the corresponding person.

2. The *CODEX PSN* is a cross-site pseudonym. This is only disclosed to the central platform (CODEX) and uniquely references a person in NUM-CODEX across all site boundaries.

Within the fTTP, all DIC PSNs for a person are assigned to a CODEX PSN. The sites cannot use the DIC PSN to determine which sites a person has visited.

To complete the person registration process, the fTTP sends a BF-specific response containing the DIC PSN to the requesting site. If a person was already known because he or she was already registered by another site, no new CODEX PSN is generated. In this case only a new DIC PSN that is assigned to the

CODEX PSN is generated. If a local site accidentally sends an unmodified BF of the same person to the fTTP multiple times, the fTTP will always respond with the same DIC PSN. No additional pseudonyms will be generated. The process is summarised in Figure 2.

(UC2) Re-pseudonymisation of a site-specific pseudonym to a cross-site pseudonym

If medical data is transmitted from a site to the central platform, use case 2 describes the re-pseudonymisation of the data with the support of the fTTP. This means that the CODEX PSN cannot be used in the CODEX platform to determine which site has supplied data on this person. This re-pseudonymisation is initiated by the GECCO Transfer Hub (GTH). For this purpose, the GTH first sends the corresponding DIC PSN to the fTTP. The fTTP validates the incoming pseudonym and returns the correspondingly assigned CODEX PSN to the GTH. The GTH is responsible to replace the existing DIC PSN in the medical data with the CODEX PSN received (re-pseudonymisation) and then sends the re-pseudonymised MDAT to the central platform. The process is summarised in Figure 3.

NUM-wide coordination of required interfaces and responsibilities

Based on the modeled use cases for person registration and re-pseudonymisation, the necessary interfaces were identified. These were coordinated between the partners involved in the implementation or who uses these interfaces later.

To implement the two use cases of the fTTP (probability), initially only two fTTP-interfaces are required:

(1) fTTP-IF1: is used for person registration and is only called by the DICs. The fTTP receives a BF and returns a DIC pseudonym.

(2) fTTP-IF2: is used for re-pseudonymisation and expects a DIC pseudonym and returns the corresponding CODEX pseudonym. This interface is only used by the GTH.

The interfaces were implemented as generically as possible so that they can also be applied in other projects.

Specification and implementation of the interfaces

The technical interfaces were specified in FHIR format, as required in MII and NUM. To ensure FHIR conformity, the company Gefyra was commissioned to support the specification-process, to validate corresponding implementations and to propose corrections, if applicable. The interfaces were

continuously documented in the public Simplifier project of the Independent Trusted Third-Party of the University Medicine Greifswald (16).

The technical implementation of the interfaces required for UC1 and UC2 expanded the existing Trusted Third Party FHIR Gateway (TTP-FHIR Gateway). This FHIR-specific, supplementary module of the TTP tools coordinates and validates incoming FHIR requests and forwards them to assigned ttp-components (for example dispatchers, E-PIX®) as required. For this purpose, the TTP-FHIR Gateway is based on HAPI FHIR (17) (version 5.4.0).

Extension of the E-PIX®

The E-PIX® is used at some sites within the local TTP as identity management for the management of PII and local record linkage. Likewise, this software solution will be used within the fTTP for the planned PPRL processes.

Therefore, E-PIX® has been extended with functionalities for the generation of BFs and matching of BFs were implemented.

Setup of the fTTP infrastructure

A data protection concept and a data protection impact assessment were prepared in advance and coordinated across NUM. Both basically describe the security concepts, tools used and the procedures of the fTTP.

The fTTP infrastructure, of course, relies on the concepts of the Independent Trusted Third-Party Greifswald. This comprises a set of secured network-zones separated by firewalls and with restrictive communication-capabilities consisting of a demilitarised network zone, a transfer network zone and a data trustee network zone. The virtual machines with the tools of the fTTP are operated in the network data trustee zone. Access to the interfaces is only granted if a site has been registered beforehand and enabled by the respective IP range of the site or a site-specific login.

The tools used in the fTTP are configured accordingly and provided in the fTTP. The E-PIX® is provided as identity management for the sites, if required, in an already pre-configured instance.

Demonstrator

First, the required test infrastructure of the fTTP was set up and selected sites were authorised to use the provided fTTP functionalities and enabled for corresponding tests.

In the form of a NUM-wide demonstrator event for interested sites, the correct implementation of UC1 on the part of the fTTP for patient registration was first shown with the help of the Charité - University

Medicine Berlin and the University Hospital Heidelberg. In addition, the correct support of the fTTP for the implementation of UC2 for re-pseudonymisation during a data transfer was then demonstrated by using the GTH, which is operated at Heilbronn University.

Results

Overview of the realised fTTP architecture

The fTTP (probability) set up for NUM-CODEX focuses in first stage on the cross-site matching of personal data based on BFs. The internal architecture of the fTTP comprises several software components:

- TTP-FHIR Gateway: Receives FHIR requests and calls the corresponding workflows in the TTP dispatcher.

- TTP Dispatcher: Workflows can be modeled and executed using the Dispatcher (18). This enables data-holding systems such as identity management and pseudonym management to be linked within the processes.

- E-PIX®: Is the identity management and manages the identities of persons within the fTTP. Identities are mapped in the fTTP (probability) exclusively via BFs. To each person several identities, and/or characteristics can be assigned (6).

- gPAS®: Is the pseudonym management and enables the management of pseudonym hierarchies and the generation of pseudonyms in separate domains.

In the specific case of a person registration, BFs are transmitted in FHIR format to the TTP-FHIR gateway. The gateway validates and converts the FHIR message and transfers the contents to the dispatcher, which triggers the fTTP-internal person registration in the form of a workflow.

This initially provides a record linkage, which is why the BF is transferred to the E-PIX®. The E-PIX® performs a record linkage by comparing all previously registered BFs with the transferred BF. The E-PIX® classifies the determined matching result and assigns the BF to an existing person or creates a new person accordingly. The matching result is transmitted back to the dispatcher. Afterwards, pseudonyms are created if necessary and a DIC PSN is delivered for the respective person depending on the matching result:

- person already exists and is already registered by the site: existing DIC PSN is returned

- person already exists but is registered from a different site: new DIC PSN is generated and returned

- person not known yet: new DIC PSN and new CODEX PSN are generated and the DIC PSN is returned

The corresponding fTTP internal communication is shown in simplified form in Figure 4.

Interface specification of the fTTP freely available

Based on the FHIR operations specified in cooperation with Gefyra (16), the TTP-FHIR gateway was extended by the fTTP-specific functionalities listed in Table 1. These form the basis for the FHIR-compliant implementation of the identified use cases on the part of the fTTP.

Table 1
Extract FHIR Operations of the fTTP.

Operation	Description	Specification and examples
requestPsnWorkflow	Query or create pseudonyms based on a preconfigured pseudonymisation workflow for a given scope (study and site). Return the generated site- and study-specific pseudonyms as parameters.	https://simplifier.net/guide/ttp-fhir-gateway-ig/requestPsnWorkflow
requestPsnFromBfWorkflow	Creation and matching of patients purely based on BFs (PPRL) for a given scope (study and site). Return of generated pseudonyms (e.g. DIC PSN(s)) as parameters.	https://simplifier.net/guide/ttp-fhir-gateway-ig/requestPsnFromBfWorkflow

Based on the specifications and detailed examples provided, all relevant partners and NUM sites were able to start planning and implementing site-related processes at an early stage.

Creation of consistent Bloom Filters is essential

The BFs must be generated using the same procedure, otherwise the BFs cannot be compared with each other. Accordingly, the BF creation is performed at all sites with the same configuration and is therefore defined accordingly. The documentation was made known to all sites (19).

This includes the methods used, the assumed hardening, length of BFs, the attributes used that are hashed, the alphabet used for random hashing, initial values for random generators, the steps required in the normalisation of the respective attributes, value ranges and formatting. A ready configuration for the E-PIX® was provided to all sites so that all users of the E-PIX® already had an out-of-the-box application. An instance of E-PIX® for testing is freely available (20).

A random hashing method (12) is used to generate the BFs, which is additionally hardened using Balanced BFs (12). The attributes first name, last name, date of birth and gender were selected as input values. After a test phase, the adjustment of the configuration is conceivable, but is sufficient for the planned demonstration at the moment.

Initial infrastructure of fTTP is ready for first tests

The setup of the technical infrastructure of the fTTP follows the BSI's recommended measures for segmenting the network (M 5.117 Integration of a database server into a security gateway from the BSI catalog of measures (21)). This includes a DMZ (Demilitarised Zone), which implements the authentication of a site by means of a client certificate. In addition, access is granted only for approved IPs or via a site-specific login. Connections are routed through an internal transfer zone. Access control databases are stored there. Identity management, pseudonym management, workflow engine and TTP-FHIR Gateway are operated in a separate trustee zone. Each site receives a site-specific API key that only allows access to the respective enabled functions. For example, the GTH may only perform re-pseudonymisation, and a DIC may only register persons within its site-specific domain of the fTTP.

Demonstration of first stage successful

Before demonstrating the technical components, organisational requirements first had to be met. The DICs of Charité - University Medicine Berlin, University Hospital Heidelberg and GTH had to register for access to the fTTP. These were set up with the specific access authorisations. This process must be carried out by all sites in NUM.

For the demonstration at the end of March 2021, the two data integration centers participated and set up an exemplary TTP, which either uses the E-PIX® as identity management and can generate BFs from identity data or ready-made BFs. For the demonstration, BFs were created for this purpose where a merge was expected in the fTTP. The DIC PSN was transmitted according to the site. The assignment of the PSNs is shown in Figure 5. After receiving the DIC PSN, a data transfer to the CODEX platform was simulated. For this purpose, the DICs sent the received DIC PSN with test data to the GTH. This performed a re-pseudonymisation by addressing the fTTP with the DIC PSN. The fTTP returned the corresponding CODEX PSN for the DIC PSN. The previous merging of BFs results here in the return of the same CODEX PSN, so that the MDAT were automatically assigned to one person, even though they come from different sites.

Discussion

Step-wise expansion of fTTP functionalities

The establishment of the fTTP infrastructure for NUM-CODEX and the functionalities follow an iterative approach. Thus, the successfully performed demonstration at the end of March 2021 focused exclusively on cross-location processes based on the first stage (PPRL based on probabilities). In June 2021, this test infrastructure was expanded to include the second stage (selective clearing process to achieve uniqueness of patients). An additional NUM-wide demonstration event is planned for Q4 2021. The test infrastructure is already available for all locations.

3 ways to provide Bloom Filters for patient registration

For successful patient registration from the sites, consolidated specifications for the generation of BFs ensure their technical comparability. The site Greifswald supports this standardised BF generation within each local TTP in three different ways. (a) Sites that use the E-PIX® for record linkage are already able to generate the appropriate BFs. No additional measures are required. (b) Sites with alternative record linkage solutions are able to utilise E-PIX® solely for BF generation (appropriate configuration is provided. (c) Sites intending to implement custom solutions were provided with a comprehensive specification of necessary BF parameters (including methods, hardenings, hashed attributes, alphabets, seed values for random number generators, length of BF, value ranges, normalisation, formates). Additional refinement of this specification based on the experience gathered in the current test stage is conceivable.

Supporting orchestration of consent and withdrawal processes

The valid and informed consent of the patient (based on the MII broad consent) represents the legal basis for processing and scientific use of health data in NUM. At present, the patient's consent is documented and processed locally at the sites. In order to be able to implement the patients' data protection rights in conformity with the GDPR, the CODEX platform may only provide research data for scientific analyses if the validity of the consent and thus the legal admissibility of the data provision has been unequivocally ensured by CODEX. A federated consent management would be an expedient extension of the fTTP-concept for orchestrating cross-site consent and withdrawal processes in order to support the correct and timely implementation of data subjects' rights at the NUM sites and in NUM-CODEX in a consistent and uniform manner.

Conclusion

The federated Trusted Third-Party for NUM-CODEX was successfully designed and implemented in a first stage. The fTTP implements the MII concepts for a cross-site privacy-preserving record linkage based on E-PIX® using BFs. In the process, original MII concepts were concretised together with the partners involved (local Trusted Third-Parties, GECCO Transfer Hub), consolidated, specified by means of FHIR (16) and practically implemented using the FHIR HAPI (17).

The established test infrastructure of the fTTP is operated by the University Medicine Greifswald and can be used by appropriately activated and authenticated NUM sites since March 2021. The correct operation of the fTTP and the complex interaction between connected sites and GTH was successfully demonstrated in a NUM-wide demonstrator-event on 31 March 2021. Two local data integration centers and Trusted Third-Parties (Berlin, Heidelberg) were exemplary connected to the fTTP. The processes for PPRL, central pseudonym allocation and transfer of pseudonymised medical data to the GTH were demonstrated live to a broad audience with representatives from all 34 sites (22).

Since April 2021, additional sites have been successively connected to the fTTP and the GTH. As of 2021-09-08: 30 of 34 sites incl. GTH are able to access the provided FHIR-based fTTP-functionalities. In addition, the second stage of the fTTP (clearing) was implemented in Q2 2021 and also put into operation for all sites for test purposes in June 2021. As of September 2021, 17 locations will use the E-PIX® as a local identity management tool.

The successful execution of the tests is documented by the NUM sites through a corresponding “fTTP connection test protocol”. As soon as the required local consent processes are established at the NUM sites and the CODEX platform launches live operation to process “real medical data”, these sites can be activated for the productive environment of the fTTP. Thus, the federated Trusted Third-Party can support a data protection-compliant sharing of medical data for COVID-19 and pandemic research for the Network of University Medicine in Germany.

Abbreviations

BF	Bloom Filter
BMBF	Federal Ministry of Education and Research
CODEX	COVID 19 Data Exchange Platform
DIC	Data Integration Center
fTTP	federated Trusted Third Party
GTH	GECCO Transfer Hub
IC	Informed Consent
IP	Internet Protocol
MDAT	Medical Data
MII	Medical Informatics Initiative
NUM	Network of University Medicine
PII	Person Identifying Information
TTP	Trusted Third Party

Declarations

Ethics approval and consent to participate:

Not applicable.

Consent for publication:

Not applicable.

Availability of data and material:

Not applicable.

Competing interests:

All authors declare that they have no conflict of interest regarding this work.

Funding:

The NUM-project is funded by the Federal Ministry of Education and Research (BMBF) (Grant Number 01KX2021) and MIRACUM (German Federal Ministry of Education and Research, grant number 01ZZ1801M).

Authors' contributions:

Drafting of the manuscript: CHampf, MBialke. Expansion of E-PIX®: CHampf, FMoser, LGeidel. Harmonisation of interfaces: CHampf, HHund, MBialke. Concept and implementation of the TTP-FHIR Gateway: MBialke, PPenndorf. Implementation of fTTP Infrastructure: NWöller, CHampf. Participation for Demonstration: CFegler, HHund, PBrunecker, FSeidel, RWettstein, LArnecke. Dispatcher-Customisation: PPenndorf. FHIR Profiling and implementational consulting: SLang. Revision of the manuscript: CHampf, MBialke, HHund, CFegeler, Slang, PPenndorf, NWöller, TBahls, FMoser, LGeidel, ABlumentritt, RSchuldt, FSeidel, PBrunecker, RWettstein, LArnecke, WHoffmann. All the authors approved the final version of the manuscript.

Acknowledgements:

Not applicable.

References

1. Sass J, Bartschke A, Lehne M, Essenwanger A, Rinaldi E, Rudolph S, et al. The German Corona Consensus Dataset (GECCO): a standardized dataset for COVID-19 research in university medicine and beyond. BMC Med Inform Decis Mak. 2020;20(1):341.

2. Nationales Forschungsnetzwerk der Universitätsmedizin zu Covid-19. Forschungsarbeiten für die bestmögliche Patientenversorgung [2021-08-15]. Available from: https://www.netzwerk-universitaetsmedizin.de/projekte.
3. Bundesministerium für Forschung und Bildung. FAQ zum nationalen Forschungsnetzwerk der Universitaetsmedizin [2021-08-15]. Available from: https://www.bmbf.de/de/faq-zum-nationalen-forschungsnetzwerk-der-universitaetsmedizin-11570.html.
4. Nationales Forschungsnetzwerk der Universitätsmedizin zu Covid-19. CODEX | COVID-19 Data Exchange Platform [2021-08-15]. Available from: https://www.netzwerk-universitaetsmedizin.de/projekte/codex.
5. Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods of information in medicine. 2018;57(S 01):e50-e6.
6. Hampf C, Geidel L, Zerbe N, Bialke M, Stahl D, Blumentritt A, et al. Assessment of scalability and performance of the record linkage tool E-PIX((R)) in managing multi-million patients in research projects at a large university hospital in Germany. J Transl Med. 2020;18(1):86.
7. Schwaneberg T, Weitmann K, Dosch A, Seyler C, Bahls T, Geidel L, et al. Data privacy management and data quality monitoring in the German Centre for Cardiovascular Research's multicentre TranslatiOnal Registry for CardiomyopatHies (DZHK-TORCH). ESC Heart Fail. 2017;4(4):440–7.
8. German National Cohort C. The German National Cohort: aims, study design and organization. Eur J Epidemiol. 2014;29(5):371–82.
9. Hampf C, Bahls T, Hund H, Drepper J, Lablans M, Speer R. Record Linkage: Optionen für standortübergreifende Datenzusammenführungen. medizin://dokumentation/informatik/informationsmanagement/. 2019;21(4):117-21.
10. Bloom BH. Space/time trade-offs in hash coding with allowable errors. Communications of the ACM. 1970;13(7):422–6.
11. Schnell R, Bachteler T, Reiher J. Privacy-preserving record linkage using Bloom filters. BMC Med Inform Decis Mak. 2009;9:41.
12. Schnell R, Borgs C, editors. Randomized Response and Balanced Bloom Filters for Privacy Preserving Record Linkage. 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW); 2016 12-15 Dec. 2016.
13. Schnell R, editor Privacy-preserving record linkage. Methodological Developments in Data Linkage; 2015.
14. Domingo-Ferrer J, Muralidhar K. New directions in anonymization: permutation paradigm, verifiability by subjects and intruders, transparency to users. Information Sciences. 2016;337:11–24.
15. Kuzu M, Kantarcioglu M, Durham E, Malin B, editors. A Constraint Satisfaction Cryptanalysis of Bloom Filters in Private Record Linkage2011; Berlin, Heidelberg: Springer Berlin Heidelberg.
16. Unabhängige Treuhandstelle der Universitätsmedizin Greifswald. Workflow-basierte Verwaltung 2021 [2021-09-09]. Available from: https://simplifier.net/guide/ttp-fhir-gateway-ig/Workflow-basierteVerwaltung.

17. HAPI FHIR - The Open Source FHIR API for Java 2021 [2021-09-05]. Available from: https://hapifhir.io/.
18. Bialke M, Penndorf P, Wegner T, Bahls T, Havemann C, Piegsa J, et al. A workflow-driven approach to integrate generic software modules in a Trusted Third Party. J Transl Med. 2015;13.
19. Hampf C, Bialke M. Patientenregistrierung 2021 [2021-09-14]. Available from: https://confluence.adesso.de/display/NUM/Patientenregistrierung.
20. Trusted Third Party of the University Medicine Greifswald. E-PIX Demo: Trusted Third Party of the University Medicine Greifswald; 2021 [Available from: https://demo.ths-greifswald.de/epix-web/.
21. Bundesamt für Sicherheit in der Informationstechnik. BSI-Grundschutz Kataloge. 2016.
22. Unabhängige Treuhandstelle der Universitätsmedizin Greifswald. Erfolgreiche Live Demo der föderierten Treuhandstelle (fTTP) in NUM-CODEX 2021 [2021-09-08]. Available from: https://www.ths-greifswald.de/erfolgreiche-live-demo-der-foederierten-treuhandstelle-fttp-in-num-codex/.

Figures

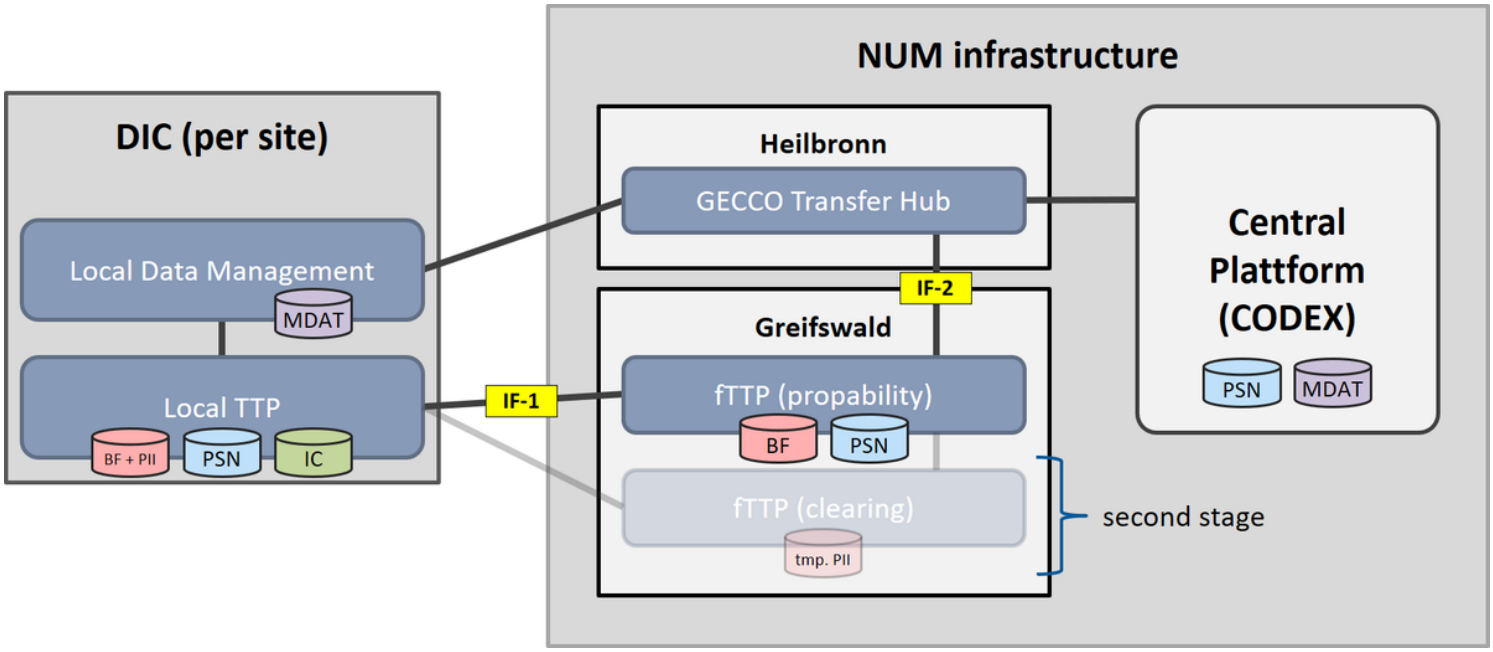

Figure 1

Interfaces between the sites and the central NUM infrastructure. Further communications are required between the central components.

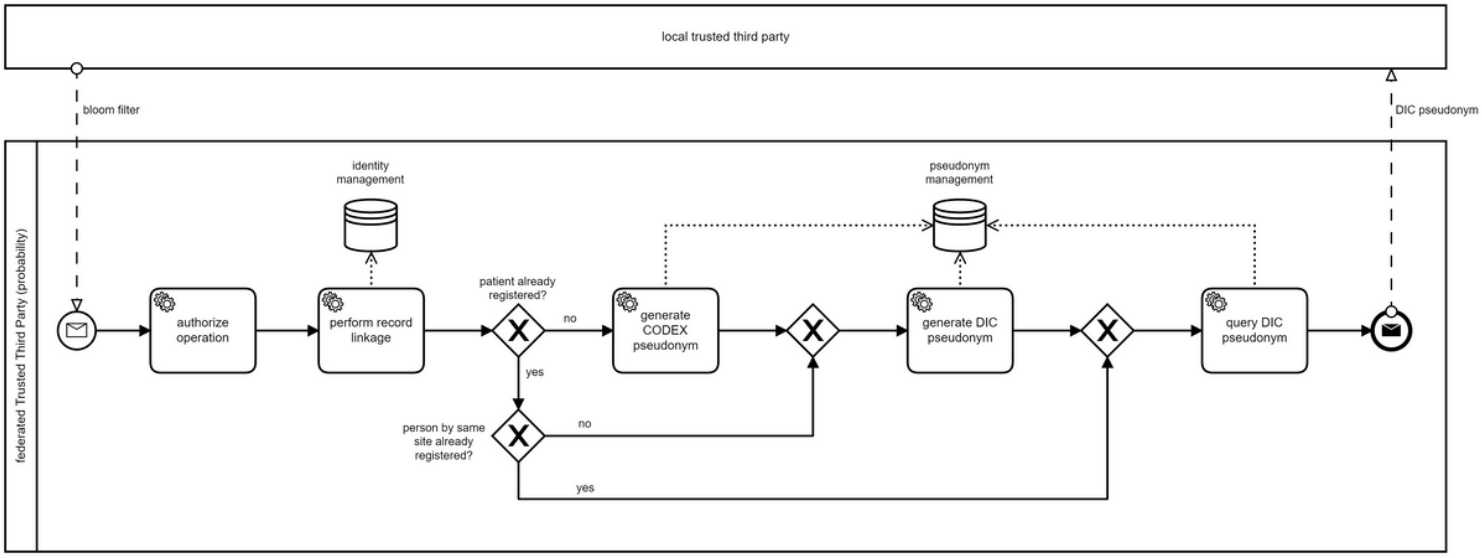

Figure 2

Process to register a person in federated Trusted Third Party (without clearing process).

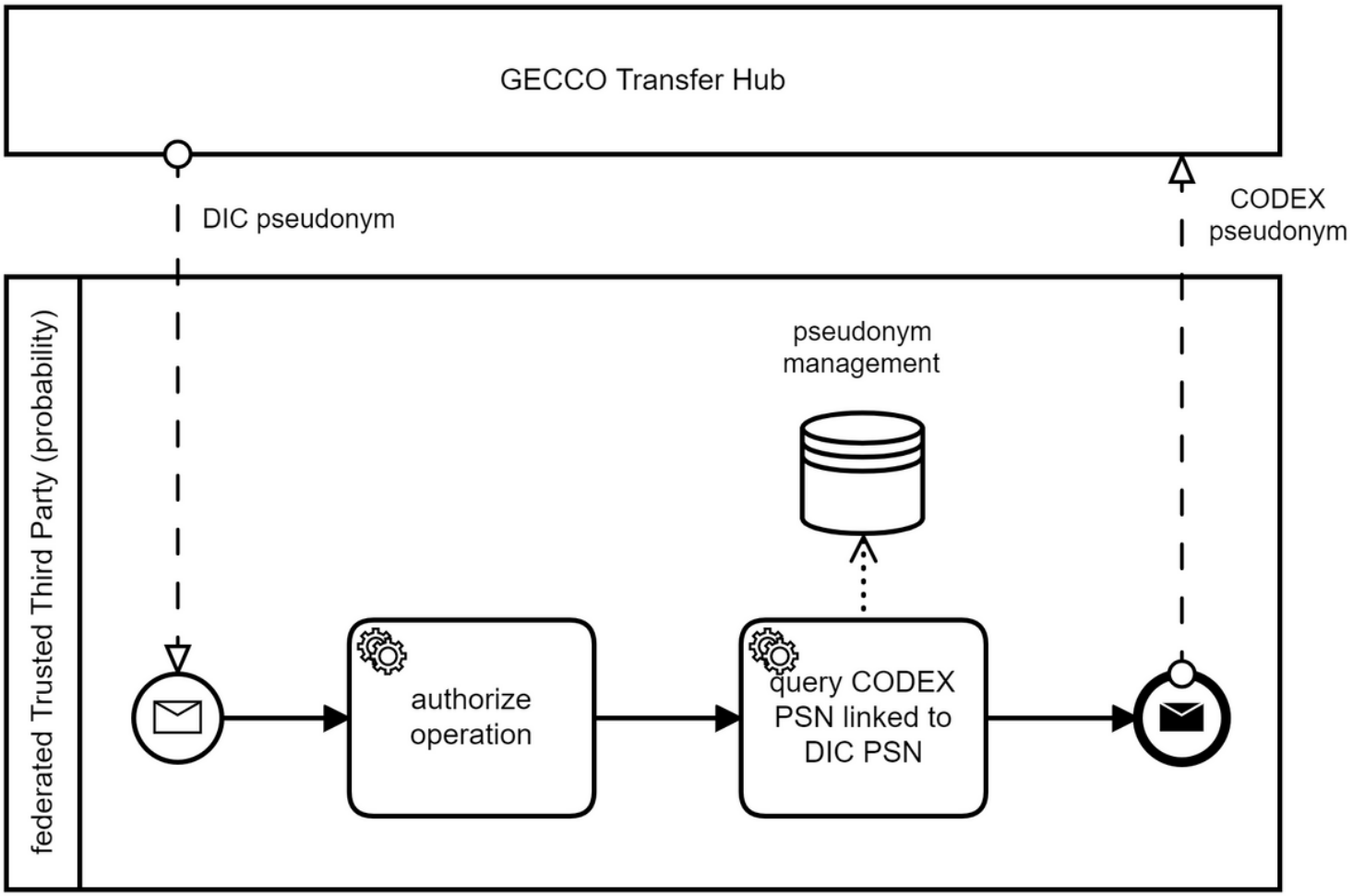

Figure 3

Process of re-pseudonymisation (dic pseudonym to codex pseudonym).

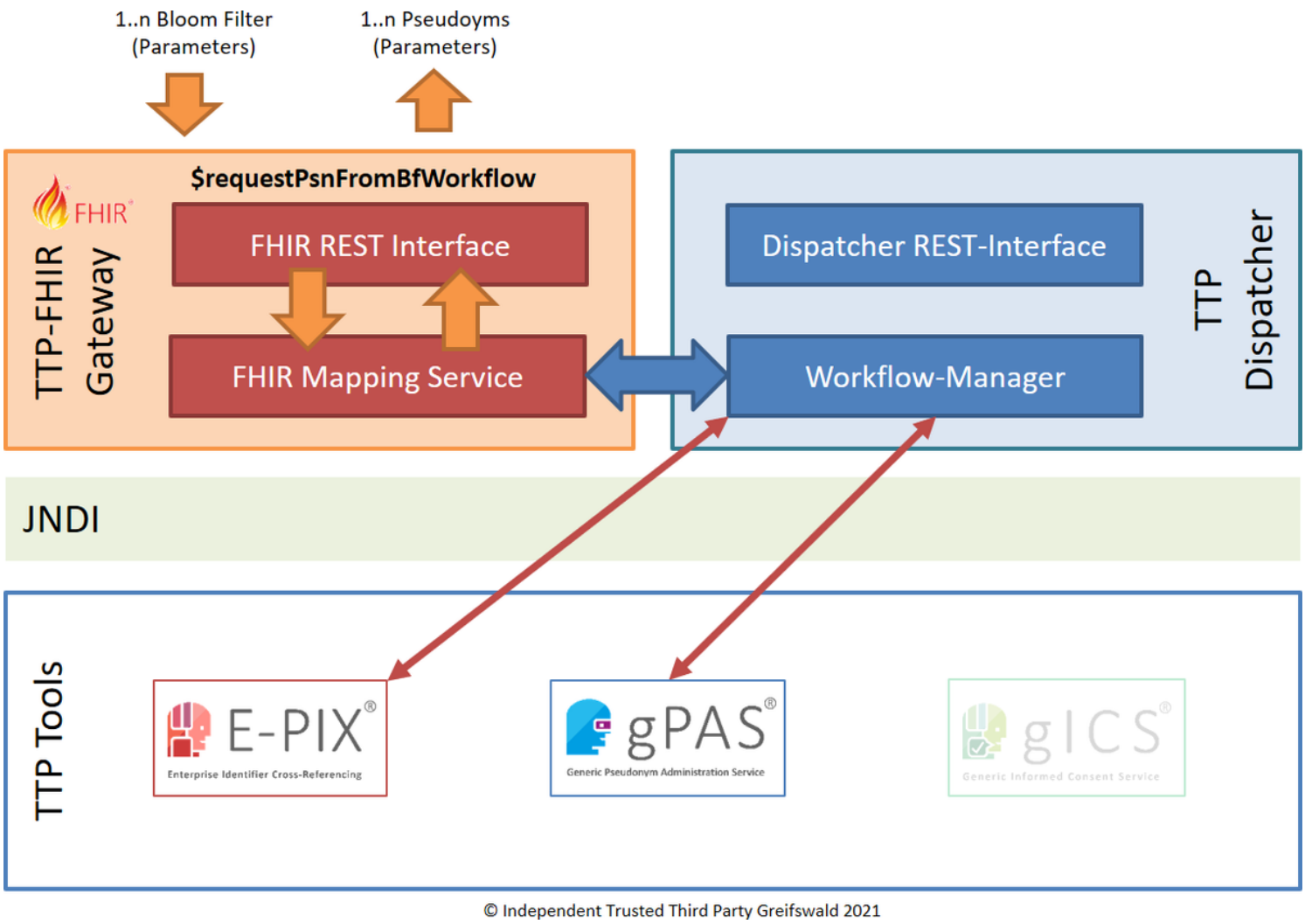

Figure 4

Required components and internal communication in the fTTP (probability) using the FHIR operation "requestPsnFromBfWorkflow" as an example.

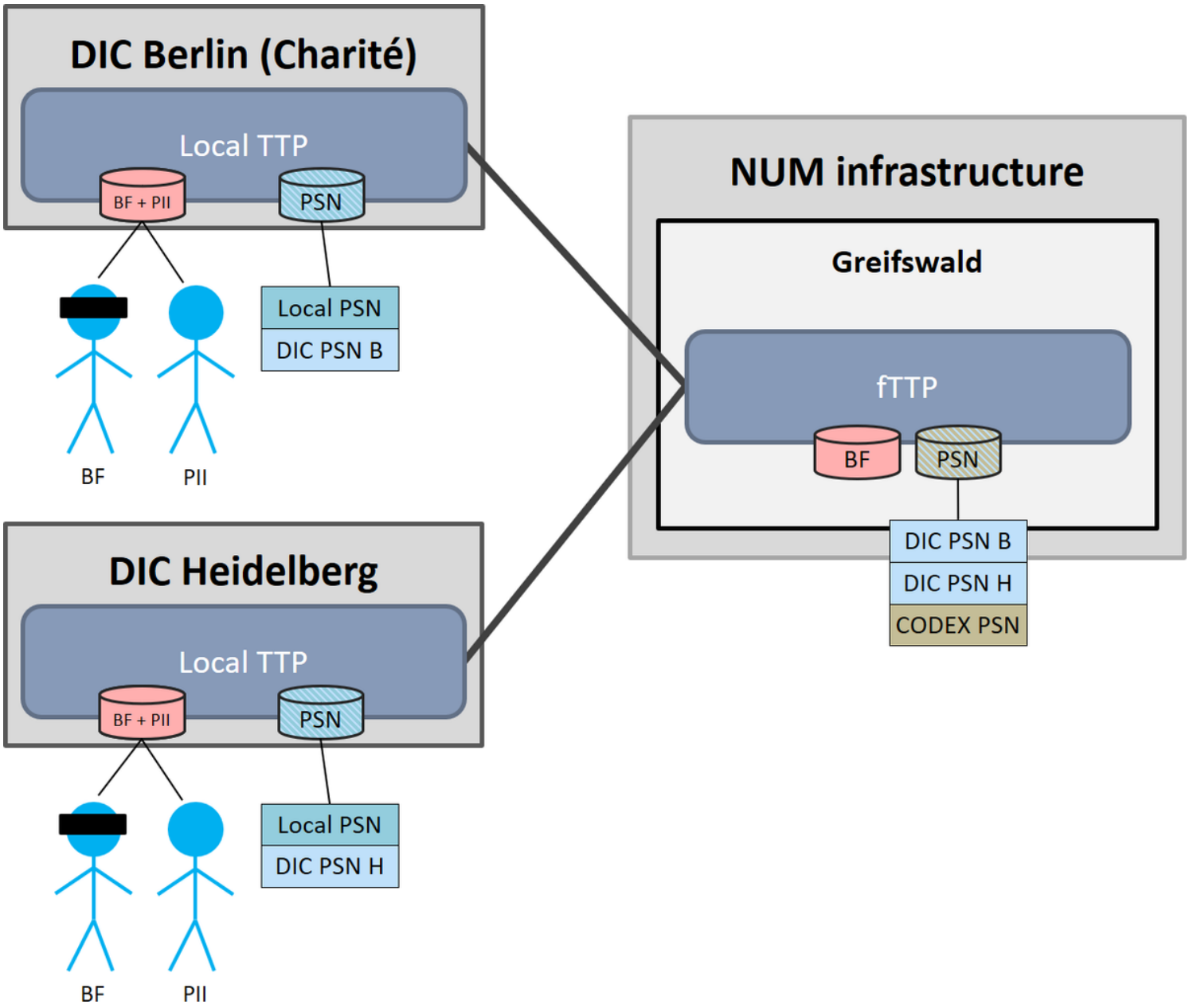

Figure 5

Pseudonym assignment at the sites. Berlin and Heidelberg have received different DIC PSNs for the same person. In the fTTP, these PSNs are assigned to a CODEX PSN so that MDAT available at several sites for this person can be merged during subsequent data transmission.