

Inhalt

Allgemein.....	1
1. Lokale Einrichtung des E-PIX	2
2. Erprobung der lokalen Domänen-Konfiguration.....	2
3. Zugang zur fTTP beantragen.....	5
3.1 Certificate Signing Request erzeugen.....	5
3.2 Zugang für die fTTP beantragen	6
3.3 Warten auf Bearbeitung des Antrags.....	7
3.4 Erhalt des Client-Zertifikats und API-Keys	7
4. Zugang zur fTTP einrichten und testen der Anbindung.....	7
4.1 Testen des Zertifikats	7
4.2 Testen der Anbindung	8
4.3 Abnahmeprotokoll	8
5. Durchführung der Test-Iterationen	9
5.1 Erhalt der Domänen-Konfiguration für die Iteration	9
5.2 Abstimmung der Seeds/Passwörter	10
5.3 Bloomfilter erzeugen.....	11
5.4 Bloomfilter bei der fTTP registrieren.....	12
5.5 Pseudonyme erhalten und ggf. Datensätze abgleichen.....	12
6. Testpaket	13
6.1 Beschreibung	13
6.2 Installation.....	13
6.3 Konfiguration.....	14
6.4 Durchführen der Test-Iterationen.....	15
Glossar	18
Abkürzungsverzeichnis	18

Allgemein

Dieses Dokument liefert einen Überblick über alle Schritte, die bei der Anbindung an die *federated Trusted Third Party* (fTTP) für CancerLink notwendig sind. Dabei wird auf die benötigten Ressourcen (Webseiten, Dokumente, etc.) verwiesen und ggf. auf weiterführende Dokumentationen aufmerksam gemacht. Im ersten Teil werden die einzelnen Schritte aufgelistet, welche danach im Detail beschrieben werden. Alle Begriffe, die *kursiv geschrieben* sind, werden im Glossar am Ende näher beschrieben und können bei Bedarf nachgeschlagen werden.

Die Anbindung an die fTTP umfasst folgende Schritte bzw. Voraussetzungen:

1. Lokale Einrichtung des Identitätsmanagements E-PIX
2. Erprobung der lokalen *Domänen-Konfiguration* anhand von Testdaten. Dies umfasst das Prüfen von erzeugten *Bloomfiltern* anhand von vorgegebenen Testdaten.
3. Zugang zur fTTP beantragen
4. Zugang zur fTTP einrichten und testen der Anbindung
5. Durchführung der Test-Iterationen

Die einzelnen Punkte werden im Folgenden näher beleuchtet.

1. Lokale Einrichtung des E-PIX

Der E-PIX ist ein Identitätsmanagement System und ermöglicht ein Record Linkage. Dabei werden Personendaten mittels sogenannter Identitäten verwaltet, welche wiederum die identifizierenden Daten (z.B. Vorname, Nachname, Geburtsdatum, Geschlecht, Adressen, etc.). Der E-PIX kann zu einer Person neben einer Hauptidentität, auch mehrere Nebenidentitäten verwaltet. Jede Identität enthält dabei andere Ausprägungen (z.B. verschiedene Schreibweisen des Vornamens „Annalena“, „Anna-Lena“). Der E-PIX erkennt diese Unterschiede und führt diese Identitäten in Abhängigkeit der hinterlegten Konfiguration zusammen oder erzeugt *Potential Matches*, welche bei Bedarf manuell zusammengeführt werden. Im Ergebnis wird jeder Person ein eindeutiger Identifikator (Master Patient Index, MPI) zugeordnet. Neben dem Abgleich identifizierender Daten, ermöglicht der E-PIX auch den Abgleich codierter Daten, in Form von *Bloomfiltern*. Außerdem kann der E-PIX solche *Bloomfilter* selbst erzeugen. Hierbei unterstützt der E-PIX mehrere verfahren, die variable konfiguriert werden können und mit mehreren Härtungsverfahren zusätzlich abgesichert werden können.

Die Bereitstellung des E-PIX erfolgt als Docker-Container. Der E-PIX kann auch als Deployment im Application Server WildFly betrieben werden, jedoch wird die Verwendung von Docker empfohlen. Eine Schritt-für-Schritt Anleitung zur Installation ist im E-PIX Handbuch (<http://www.ths-greifswald.de/e-pix/handbuch>) im Kapitel „Installation“ zu finden. Der Download zum E-PIX ist auf der THS-Webseite unter <https://www.ths-greifswald.de/forscher/e-pix/> zu finden.

Das heruntergeladene ZIP-Archiv kann am gewünschten Ort entpackt werden. Der E-PIX ist im Auslieferungszustand bereits fertig konfiguriert und kann direkt gestartet werden (`docker compose up -d`). Es können aber kleinere Anpassungen vorgenommen werden, um den E-PIX an das vorliegende System anzupassen. Eine Anpassung besteht darin, die Menge des zugelassenen Arbeitsspeichers einzustellen. Hierzu wird in der Datei `/envs/wf_commons.env` die Einstellung „JAVA_OPTS“ angepasst. Standardmäßig werden zumindest 1 GB und maximal 6 GB Arbeitsspeicher für den E-PIX reserviert. Bei Bedarf, z.B. großen Datenmengen (10.000 und mehr Personendatensätzen), kann der Arbeitsspeicher erhöht werden. Dies kann zusätzlich die Performance steigern.

Das erstmalige Hochfahren des E-PIX kann etwas Zeit in Anspruch nehmen (30-60 Sekunden) und steht dann über die Weboberfläche bereit.

2. Erprobung der lokalen Domänen-Konfiguration

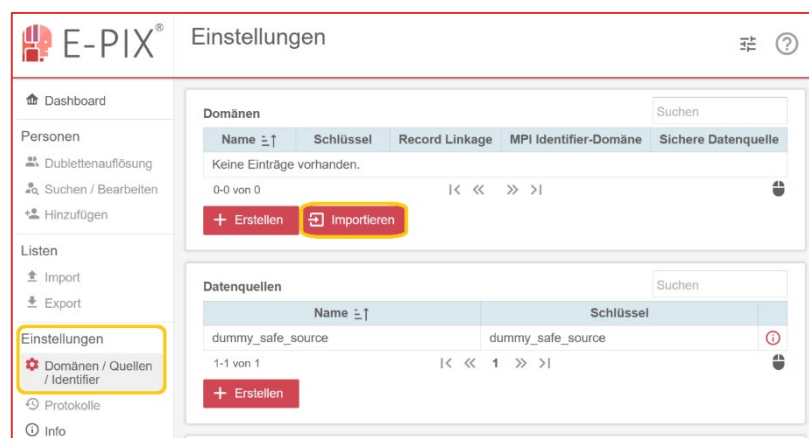
Im Folgenden wird die Einrichtung einer Domäne gezeigt und wie Daten in diese gespielt werden. Dies passiert auf Basis der beiliegenden Testdaten. Eine Domäne beschreibt dabei einen abgeschlossenen Kontext. Eingetragene Personendaten befinden sich immer in einer konkreten Domäne und werden nur innerhalb dieser miteinander abgeglichen. Es können mehrere Domänen nebeneinander existieren. In der späteren Projektphase werden mehrerer solcher Domänen angelegt. Jede Domäne repräsentiert dabei einer Iteration mit einer speziellen *Domänen-Konfiguration*.

Nun wird gezeigt, wie eine lokale *Domänen-Konfiguration* wird im E-PIX hinterlegt wird. Diese beinhaltet eine exemplarische Konfiguration, welche zur Erzeugung der beiliegenden Testdaten verwendet wurde. Diese beinhaltet unter anderem, die Art und Weise, wie identifizierende Daten (IDAT) in einen *Bloomfilter* überführt werden. In ähnlicher Form werden später *Domänen-Konfigurationen* bereitgestellt, welche dann die jeweilige *Bloomfilter*-Konfiguration repräsentieren.

Die beiliegenden Testdaten, stellen sicher, dass die Konfiguration vom E-PIX erfolgreich war und ermöglichen einen ersten Einstieg im Umgang mit dem E-PIX. Die beiliegende Datei `cancerlink_testdaten.csv` beinhaltet die Testdaten. Diese Datei beinhaltet 80 fiktive Datensätze, bestehend aus einer ID, Vorname, Nachname, Geburtsdatum, Geschlecht, Postleitzahl, Straße, Ort und *Bloomfilter*. Letzteres ist das Ergebnis, wenn die Daten mit der beiliegenden *Domänen-Konfiguration* (`cancerlink_testdata_epix_config.xml`) zu einem *Bloomfilter* umgewandelt werden. Die ID ist ein numerischer Wert, welcher die Personen-Zugehörigkeit der Datensätze darstellt. Jede fiktive Version weist zwei Datensätze in unterschiedlichen Ausprägungen auf. Die resultierenden *Bloomfilter* unterscheiden sich daher. Im Späteren gilt es, trotz der Unterschiede, entsprechende Zugehörigkeiten mittels der fTTP zu erkennen.

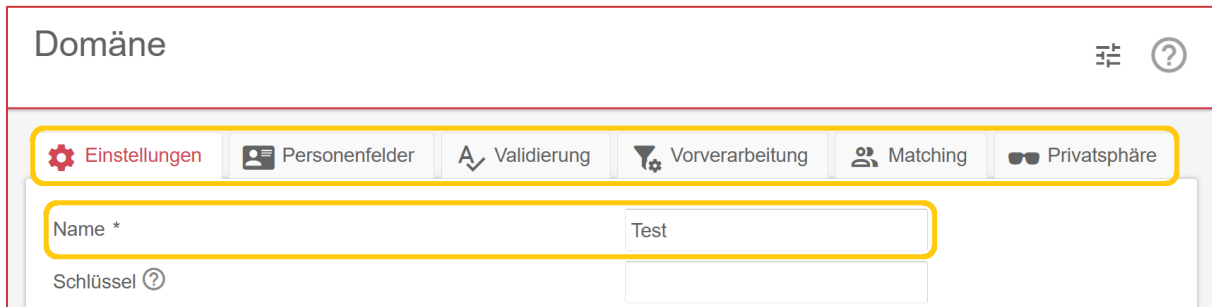
Im Folgenden wird gezeigt, wie diese *Domänen-Konfiguration* im E-PIX eingerichtet wird und wie die entsprechenden Testdaten eingespielt werden. Voraussetzung hierfür ist, dass der E-PIX erfolgreich installiert wurde und nun betriebsbereit ist. Grundsätzlich gilt, dass hier nur ein Überblick gegeben werden kann. Eine detailliertere Beschreibung zu den einzelnen Funktionalitäten des E-PIX, sind im Handbuch zu finden (<http://www.ths-greifswald.de/e-pix/handbuch>).

Zunächst muss die Weboberfläche vom E-PIX geöffnet werden. Standardmäßig erfolgt dies über den Browser über die URL <http://<IP-Adresse-des-Servers>:8080/epix-web>. Die IP-Adresse und ggf. der Port (sofern abweichend von 8080) müssen entsprechend angepasst werden. Die lokale IT, welche ggf. beteiligt war, kann nähere Informationen liefern. Beim erstmaligen Aufruf sind die meisten Menüpunkte auf der linken Seite ausgegraut. Zunächst sollte unter „Einstellungen“ der Menüeintrag „Domänen / Quellen / Identifier“ ausgewählt werden. Der E-PIX hat bereits standardmäßig eine Datenquelle und eine Identifier-Domäne hinterlegt (weitere Informationen hierzu im Handbuch im Kapitel 3 „Allgemein“). Im Folgenden und allen späteren Iterationen, können diese genutzt werden, ohne diese jedes Mal neu anzulegen. Um eine Domäne anzulegen, muss der Button „Importieren“ ausgewählt werden. Zwar lassen sich *Domänen-Konfigurationen* auch Schritt-für-Schritt einrichten, jedoch werden in CancerLink stets vorbereitete XML-Dateien bereitgestellt, sodass es nicht erforderlich ist, jeden Punkt einzeln auszufüllen.



Nun kann die beiliegende `cancerlink_testdata_epix_config.xml` ausgewählt und importiert werden. Die Einstellungen sind damit übernommen und verbergen sich hinter den jeweiligen Reitern (Personenfelder, Vorverarbeitung, Matching, Privatsphäre). Unter dem Reiter Einstellungen, muss noch ein Name vergeben werden. Hier kann ein beliebiger Name (z.B. „Test“)

gewählt werden. Weitere Einstellungen sind hier nicht erforderlich. An dieser Stelle soll jedoch noch ein kurzer Überblick über die jeweiligen Reiter gegeben werden.



Personenfelder: Hier können Pflichtfelder und Zusatzfelder definiert werden. Pflichtfelder müssen in späteren Tests angegeben werden. Ist z.B. der Vorname ein Pflichtfeld, so lehnt der E-PIX spätere Eintragungen ab, in denen dieser nicht ausgefüllt wurde. Zusatzfelder sind frei definierbare Felder („Feld 1“ – „Feld 10“). Hierunter fallen z.B. der *Bloomfilter*, welcher im E-PIX über ein Zusatzfeld als solcher benannt werden kann (in der Konfiguration „Feld 8“). Auch die Postleitzahl wird als Zusatzfeld definiert (in der Konfiguration „Feld 3“). Adressdaten können als solche nicht zum Abgleich verwendet werden. Daher können diese als Zusatzfeld aufgenommen werden und so auch miteinander abgeglichen werden.

Vorverarbeitung: Die Vorverarbeitung wird auf jeden Datensatz angewandt und bereitet diesen auf den Abgleich vor. So werden üblicherweise Vor- und Nachnamen so transformiert, dass diese nur noch aus Großbuchstaben bestehen (Anna → ANNA). Ebenso werden Diakritika entfernt (RENÈ → RENE). Darüber hinaus werden häufige Zusatzangaben wie „Dr.“ im Vornamen entfernt. Diese Maßnahmen dienen dazu, die Datensätze soweit aufzubereiten, dass Datenfelder auch dann als gleich erkannt werden können, wenn diese bei der Eintragung leichte Unterschiede aufwiesen.

Matching: Das Matching definiert, wann Daten zusammengeführt werden. Der E-PIX unterstützt mehrere Verfahren, wie die Ähnlichkeit eines jeden Attributes bestimmt werden kann und ab welcher Ähnlichkeit, diese als übereinstimmend gewertet werden sollen. Matching-Felder müssen verpflichtend Pflichtfelder sein.

Privatsphäre: Hier wird die *Bloomfilter*-Konfiguration vorgenommen. Diese Einstellungen werden bei späteren Iterationen stets verändert sein. Neben variablen *Bloomfilter*-Längen, können auch verschiedene Härten angewandt werden. Diese erhöhen die Sicherheit eines *Bloomfilters*, beeinflussen jedoch oft die Qualität des *Bloomfilters*. Jedes Eingabefeld (z.B. Vorname, Nachname, usw.) wird dabei anders codiert. Dies wird durch sogenannte *Seeds* (im späteren auch Passwörter genannt) ermöglicht, welche später unter den Krebsregistern abgestimmt werden. Nur wenn alle *Seeds* zu den jeweiligen Datenfeldern bei den Krebsregistern innerhalb einer Iteration identisch sind, können vergleichbare *Bloomfilter* generiert werden.

Info: Im Abschnitt „Durchführung der Test-Iterationen“ wird das Vorgehen exemplarisch gezeigt.

Mit einem Klick auf den Button „Hinzufügen“, wird die Domäne angelegt. Standardmäßig sollte nun die angelegte Domäne als „Aktive Domäne“ (siehe unten links im Menü) ausgewählt sein. Ansonsten kann dies nun mittels des Drop-Down-Feldes erledigt werden. Nun sind die Menüpunkte nicht mehr ausgegraut. Einzelne Datensätze können nun über den Menüpunkt „Hinzufügen“ eingetragen werden. Komplette Listen, so wie es im späteren Projektverlauf empfohlen wird, können unter dem Menüpunkt „Import“ eingetragen werden. Dies wird hier nun anhand der beiliegenden Datei (`cancerlink_testdaten.csv`) gezeigt.

Hierzu muss die Datei in den E-PIX geladen werden. In der beiliegenden Datei wird „;“ als Trennzeichen für die einzelnen Datenfelder verwendet. Diese Angabe kann also beibehalten werden. Die beiliegende Datei beinhaltet eine Kopfzeile. Das entsprechende Kontrollkästchen muss also angekreuzt werden. Danach kann die Datei ausgewählt werden. Anhand des ersten Datensatzes (die Kopfzeile wird übersprungen), können nun die Datenfelder zugeordnet werden. Hierzu werden die benötigten Felder angewählt (Vorname, Nachname, Geburtsdatum, Geschlecht und Postleitzahl). Nun wird der Typ ausgewählt. Dieser sollte dem jeweiligen Datenfeld

2. Felder auswählen und Typ festlegen ⓘ

Spalte	Typ	Beispieldatensatz 1	
☐ Spalte 1	Unbekannt	person	1
☒ Spalte 2	Vorname	firstname	Lukas
☒ Spalte 3	Nachname	lastname	Müller
☒ Spalte 4	Geburtsdatum	date_of_birth	1992-03-14
☒ Spalte 5	Geschlecht	gender	M
☒ Spalte 6	Postleitzahl	postalcode	10115
☐ Spalte 7	Unbekannt	street	Bahnhofstraße 12
☐ Spalte 8	Unbekannt	city	Berlin
☐ Spalte 9	Unbekannt	bloomfilter	Sn6HefKrSNIOLbslfpSLwpi4OLxKCDG

3. Optionen

Datenquelle * dummy_safe_source

entsprechen. Die Postleitzahl kann ebenso ausgewählt werden, weil das „Feld 3“ in der *Domänen-Konfiguration* so benannt wurde. Der *Bloomfilter* muss an dieser Stelle nicht importiert werden, da der E-PIX diesen gleich erzeugen wird. Unter „Optionen“ muss noch die Datenquelle angewählt werden. Hier genügt es, die standardmäßig vorhandene Datenquelle „dummy_safe_source“ anzugeben.

Mit einem Klick auf den Button „Importieren“, werden die Daten in die entsprechende Domäne geladen. Dabei wird auch der *Bloomfilter* erzeugt. In der Ergebnisliste werden die jeweiligen Datensätze nun aufgeführt. Die ersten 40 Datensätze wurden dabei als NO_MATCH klassifiziert. Danach wurden die Datensätze als MATCH oder POSSIBLE_MATCH erkannt. Dies sind jene Datensätze, welche leichte Abweichungen aufweisen. Possible Matches können unter „Dublettenauflösung“ eingesehen werden. Im produktiven Betrieb lassen sich so ähnliche Datensätze zusammenführen oder trennen. Außerdem wird in der Ergebnisliste bereits der erzeugte *Bloomfilter* angezeigt. Dieser wird später bei der fTTP registriert.

Das hier gezeigte Vorgehen ist später Grundlage für alle Iterationen. Es weichen nur die Parameter in der *Domänen-Konfiguration* ab. Die später genutzten Daten, sollten einmalig vorbereitet werden, damit diese stets als CSV-Datei eingelesen werden kann. Dies reduziert die Aufwände auf ein Minimum.

3. Zugang zur fTTP beantragen

Die Zugangsbeantragung umfasst die folgenden Schritte:

1. *Certificate Signing Request* (CSR) erzeugen
2. Zugang für die fTTP mit Informationen zu Ansprechpartnern und Serverinformationen, samt den CSR beantragen
3. Warten auf Bearbeitung des Antrags
4. Erhalt des Client-Zertifikats und API-Keys

3.1 Certificate Signing Request erzeugen

Ein CSR ist ein digitaler Antrag zur Erstellung eines Zertifikats. In diesem Fall wird ein Client-Zertifikat angefordert, welches später dem jeweiligen Krebsregister erlaubt, mit der fTTP eine Verbindung aufzubauen. Nur wer im Besitz des Zertifikats ist, kann am Ende eine Verbindung aufbauen. Dies eine

von mehreren Sicherheitsmaßnahmen, bezüglich der Kommunikation mit der Treuhandstelle Greifswald.

Zunächst muss eine Key-Datei für die Erzeugung des CSR erzeugt werden. Diese muss unbedingt aufbewahrt werden, da diese Datei später auch für die finale Erzeugung des Zertifikats benötigt wird.

Die notwendigen Schritte werden mit OpenSSL gezeigt.

Mit dem folgenden Befehl im Terminal, wird eine Key-Datei erzeugt. Hierzu sollte ein sprechender Dateiname vergeben werden (z.B. cancerlink.key). Dabei muss ein (sicheres) Passwort vergeben werden, welches unbedingt für später aufbewahrt werden muss.

```
openssl genrsa -des3 -out [KEYNAME].key 4096
```

Die entsprechende Key-Datei sollte sich nun im entsprechenden Verzeichnis befinden. Danach wird die CSR-Datei erzeugt. Für diesen Schritt wird die eben erzeugte Key-Datei benötigt.

Dabei werden während des Anlegens, mehrere Angaben gefordert, sie entsprechend korrekt angegeben werden sollten. Dabei ist im Besonderen zu beachten, dass die Angabe Organizational Unit Name (OU) mit fTTP CancerLink (Groß- und Kleinschreibung beachten!) ausgefüllt wird. Andernfalls, werden spätere Verbindungen nicht akzeptiert und die Erzeugung eines neuen Zertifikats ist erforderlich. Die CSR-Datei kann z.B. cancerlink.csr benannt werden.

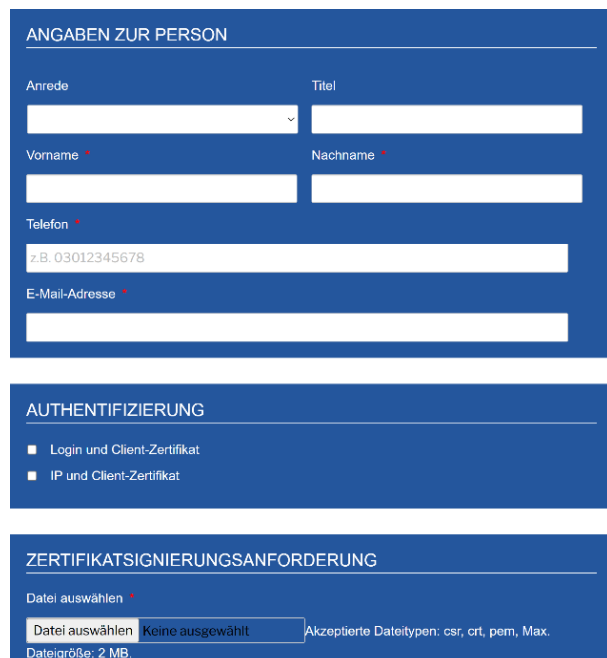
```
openssl req -new -key [KEYNAME].key -out [CSRNAME].csr
```

Die CSR-Datei wird für die Zugangsbeantragung zur fTTP benötigt.

3.2 Zugang für die fTTP beantragen

Auf der Webseite der Unabhängigen Treuhandstelle Greifswald befindet sich ein projektspezifisches Antragsformular (<https://www.ths-greifswald.de/forscher/ftp/cancerlink/register>).

Dieses muss mit Angaben zur Ansprechperson, der gewünschten Authentifizierungsmethode und der eben erzeugten CSR-Datei ausgefüllt werden. Bei der Authentifizierungsmethode kann zwischen einem Login (Name und Passwort) und einem IP-basierten Zugang gewählt werden. Die meisten Institutionen haben eine statische IP-Adresse oder -Bereich. Die Authentifizierung erfolgt dann über diese freigegebenen IP-Adressen. Sollte nur ein dynamischer Bereich vorliegen, oder Datenübermittlungen außerhalb dieses Bereichs stattfinden, so kann ein klassischer Login verwendet werden.



The screenshot shows a web form titled 'ANGABEN ZUR PERSON' and 'AUTHENTIFIZIERUNG'. The 'ANGABEN ZUR PERSON' section includes fields for 'Anrede' (dropdown), 'Titel' (text), 'Vorname' (text), 'Nachname' (text), 'Telefon' (text with example 'z.B. 03012345678'), and 'E-Mail-Adresse' (text). The 'AUTHENTIFIZIERUNG' section has two radio buttons: 'Login und Client-Zertifikat' (selected) and 'IP und Client-Zertifikat'. Below this is a section titled 'ZERTIFIKATSIGNIERUNGSANFORDERUNG' with a 'Datei auswählen' button, a 'keine ausgewählt' button, and a text 'Akzeptierte Dateitypen: csr, crt, pem, Max.' and 'Dateigröße: 2 MB'.

3.3 Warten auf Bearbeitung des Antrags

Der Antrag wird digital übermittelt und von der Unabhängigen Treuhandstelle meistens innerhalb von drei Werktagen bearbeitet. Bei Fragen kann sich jederzeit an die Kontaktadresse der Unabhängigen Treuhandstelle Greifswald gewandt werden (kontakt@med.uni-greifswald.de).

3.4 Erhalt des Client-Zertifikats und API-Keys

An die angegebene E-Mail-Adresse wird eine Antwort übermittelt. Diese beinhaltet eine Zertifikatsdatei (CER-Datei) und ein API-Key. Sowohl das Zertifikat, als auch der API-Key sind standortspezifisch und dürfen daher nicht an andere Standorte gegeben werden.

Die in der E-Mail enthaltene CER-Datei muss nun abschließend zu einem Client-Zertifikat umgewandelt werden. Hierfür wird die in Punkt 1 erzeugte Key-Datei benötigt.

```
openssl pkcs12 -export -in fTTP.cer -inkey [KEYNAME].key -out  
[Zertifikat].p12
```

Die resultierende P12-Datei kann in dieser Form bereits verwendet werden. Bei Bedarf kann die Dateiendung auf pfx umgestellt werden. Dabei genügt es, die Dateiendung entsprechend zu bearbeiten.

4. Zugang zur fTTP einrichten und testen der Anbindung

Das Testen der Verbindung zur fTTP schließt an die Beantragung des Zugangs an. Hierzu müssen der API-Key und das Zertifikat erhalten worden sein und das Zertifikat muss umgewandelt worden sein.

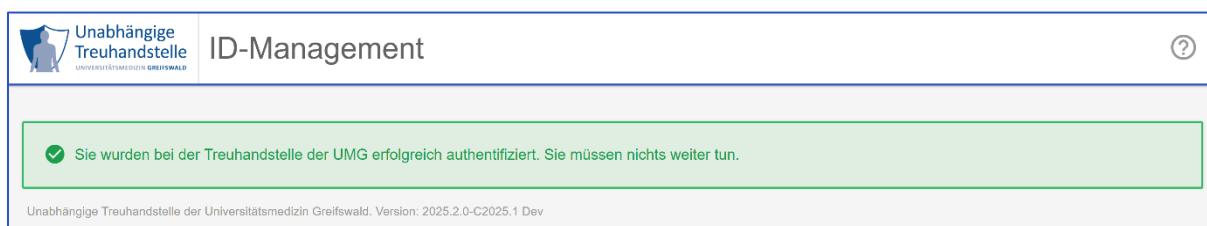
Folgende drei Schritte sollten durchgeführt werden, um eine Erfolgreiche Anbindung an die fTTP vorzunehmen:

1. (optional) Testen des Zertifikats
2. Testen der Anbindung und exemplarischer Erhalt von Test-Pseudonymen
3. Ausfüllen des Abnahmeprotokolls

4.1 Testen des Zertifikats

Als optionalen Schritt kann das Client-Zertifikat getestet werden. Dies schließt Fehler durch ein fehlerhaftes Zertifikat aus. Hierzu wird zunächst das entsprechende Zertifikat im Browser installiert. Hierzu kann der bereitgestellten Anleitung gefolgt werden (<https://www.ths-greifswald.de/forscher/ftp/cancerlink/installcert>).

Zur Prüfung der Authentifizierung mittels Client-Zertifikat, kann im Browser die URL <https://cancerlink.med.uni-greifswald.de/check> aufgerufen werden. Erscheint die nachfolgende Meldung, war die Authentifizierung erfolgreich. Wurde bei der Beantragung des Zugangs die Authentifizierungsmethode „Login“ gewählt, so muss zusätzlich der Name und das Passwort im Dialog eingetragen werden.



Sollte statt der gezeigten Seite ein Fehler auftreten, so sollte zunächst geprüft werden, ob das korrekte Zertifikat gewählt wurde.

Falsches Client-Zertifikat ausgewählt



Ihr Client-Zertifikat erlaubt keinen Zugang zu dieser Seite. Bitte folgen Sie den unten stehenden Lösungsschritten, um die Fehlermeldung zu beheben.

Lösungsschritte

- 1 Prüfen Sie, ob Sie das korrekte Client-Zertifikat ausgewählt haben.
- 2 Prüfen Sie, ob Sie mit einem Test-Client-Zertifikat das Produktiv-System versuchen zu erreichen.
- 3 Kontaktieren Sie uns via Telefon oder E-Mail, sofern der Fehler weiterhin auftritt.

Sollte der Fehler trotzdem bestehen, kann die Unabhängige Treuhandstelle per E-Mail kontaktiert werden. Ggf. muss ein fehlerhaftes Zertifikat neu erzeugt werden, was das beschriebene Vorgehen aus Abschnitt (1. Certificate Signing Request erzeugen) voraussetzt.

4.2 Testen der Anbindung

Das schnelle Testen der Anbindung an die fTTP kann mit dem bereitgestellten Testpaket erreicht werden. Dies umfasst einen *jmeter*-Test, welcher eine Anfrage an die fTTP durchführt. Im Erfolgsfall liefert die fTTP ein Pseudonym. Ist dies der Fall, so ist sichergestellt, dass die Anbindung an die fTTP korrekt durchgeführt wurde. Dies beinhaltet sowohl die korrekte Authentifizierung, als auch die korrekte Übermittlung eines Bloomfilters. Außerdem ist so sichergestellt, dass etwaige Portfreigaben am Standort korrekt erfolgt sind oder lokale Proxies die entsprechenden Anfragen korrekt durchleiten. Aus Erfahrung, sind dies potentielle Fehlerquellen, die eine Anbindung verzögern.

Das Testpaket (<https://www.ths-greifswald.de/forscher/fttp/cancerlink/testpaket>) beinhaltet zusätzlich eine Anleitung (HowTo-Client-Zertifikat_mit_jmeter.md), welche die erforderlichen Schritte beschreibt, um das Client-Zertifikat für *jmeter* zu hinterlegen (die Datei kann mit einem üblichen Texteditor geöffnet werden). Außerdem sollten die beigelegte ReadMe.md beachtet werden.

Der *jmeter*-Test muss entsprechend angepasst werden (erhaltenen API-Key eintragen und ggf. Login-Daten eintragen). Danach kann der *jmeter*-Test ausgeführt werden. Sofern die Kommunikation mit der fTTP erfolgreich geklappt hat, wird eine Textdatei (results.txt) angelegt, welche die abgerufenen Pseudonyme beinhaltet. Die Datei sollte kurz überprüft werden, ob sich die abgerufenen Pseudonyme darin befinden.

Es ist auch möglich, die Anbindung anderweitig zu realisieren. Die abgerufenen Pseudonyme (oder die results.txt) sollten im nächsten Schritt an die fTTP übermittelt werden.

4.3 Abnahmeprotokoll

Das Abnahmeprotokoll enthält alle notwendigen Schritte, um die Anbindung abzuschließen. Die Übermittlung des Abnahmeprotokolls ist verpflichtend und stellt sicher, dass die Anbindung korrekt stattgefunden hat. Das ausgefüllte Abnahmeprotokoll muss an kontakt@med.uni-greifswald.de

übermittelt werden. Das entsprechende Protokoll ist unter <https://www.ths-greifswald.de/forscher/ftp/cancerlink/abnahme> zu finden.

5. Durchführung der Test-Iterationen

⚠ Hinweis: Es steht ein Testpaket zur Verfügung, welches das Anlegen der E-PIX Domäne, das Registrieren im lokalen E-PIX (zur Erzeugung der Bloomfilter) und der Registrierung der Bloomfilter bei der fTTP übernimmt. Das Vorgehen wird im nachfolgenden Kapitel beschrieben, sodass die hier gezeigten manuellen Schritte teilweise entfallen.

Die eigentlichen Testdurchläufe werden realisiert, indem voneinander unabhängige Iterationen mit unterschiedlichen Konfigurationen durchgeführt werden. Die Erzeugung der *Bloomfilter* basiert auf den Echtdaten der Krebsregister und wird im jeweiligen Krebsregister durchgeführt. Hierzu müssen folgende Schritte für jede Iteration durchlaufen werden:

1. Erhalt der *Domänen-Konfiguration* für die Iteration
2. Abstimmung der *Seeds*/Passwörter
3. *Bloomfilter* erzeugen
4. *Bloomfilter* bei der fTTP registrieren
5. Pseudonyme erhalten und ggf. Datensätze abgleichen

5.1 Erhalt der Domänen-Konfiguration für die Iteration

Die Treuhandstelle Greifswald erzeugt pro Iteration eine Vorlage für eine E-PIX *Domänen-Konfiguration* im XML-Format, welche spezifische Einstellungen für die Erzeugung der *Bloomfilter* enthält. Darüber hinaus werden pro Iteration an der fTTP Einstellungen vorgenommen, die sich hauptsächlich auf Schwellwerte bezieht (ab welcher Ähnlichkeit, werden zwei *Bloomfilter* derselben Person zugeordnet).

Die *Domänen-Konfiguration* enthält zudem Platzhalter, in denen pro verwendeten Personenattribut (z.B. Vorname, Nachname, etc.) später noch *Seeds* eingetragen werden müssen. Dies kann auch für das verwendete Verfahren zutreffen (z.B. benötigt das Verfahren „Balanced Bloomfilter“ zusätzlich einen *Seed*). Diese *Seeds* müssen bei allen Standorten identisch sein. Hierbei enthält die Vorlage entsprechende Verweise, z.B. `HIER_SEED_A_EINTRAGEN` oder `HIER_SEED_B_EINTRAGEN`.

```

31 <privacy>
32   <bloomfilter-config>
33     <algorithm>org.emau.icmvc.ttp.deduplication.impl.bloomfilter.RandomHashingStrategy</algorithm>
34     <field>value8</field>
35     <length>1000</length>
36     <ngrams>2</ngrams>
37     <bits-per-ngram>25</bits-per-ngram>
38     <fold>0</fold>
39     <alphabet>ABCDEFGHIJKLMNOPQRSTUVWXYZ.-0123456789mfoux</alphabet>
40     <balanced>
41       <seed>HIER_SEED_A_EINTRAGEN</seed>
42     </balanced>
43     <source-field>
44       <name>firstName</name>
45       <seed>HIER_SEED_B_EINTRAGEN</seed>
46     </source-field>
47     <source-field>
48       <name>lastName</name>
49       <seed>HIER_SEED_C_EINTRAGEN</seed>
50     </source-field>
51     <source-field>
52       <name>birthDate</name>
53       <seed>HIER_SEED_D_EINTRAGEN</seed>
54     </source-field>
55     <source-field>
56       <name>value3</name>
57       <seed>HIER_SEED_E_EINTRAGEN</seed>
58     </source-field>
59   </bloomfilter-config>
60 </privacy>

```

5.2 Abstimmung der Seeds/Passwörter

Nach Erhalt der *Domänen-Konfiguration* müssen einheitlich *Seeds* vergeben werden. Diese *Seeds* dienen dazu, dass die Codierung überall auf dieselbe Weise passiert. Unterschiedliche *Seeds* würden bei der Codierung identischer Daten, zu unterschiedlichen *Bloomfiltern* führen. Diese wären dann nicht mehr vergleichbar. Die *Seeds* können daher auch als ein Geheimnis oder Passwort gesehen werden. Ohne diesen *Seed*, kann der *Bloomfilter* nicht auf dieselbe Weise erzeugt werden. Die *Seeds* sollten dabei nur zwischen den Krebsregistern bekannt sein. Obwohl das alleinige Wissen über einen *Seed* nicht ausreicht, um einen *Bloomfilter* wieder zurück in die ursprünglichen Klardaten überführen zu können, würde dies das Ausprobieren häufiger Namenszusammensetzungen ermöglichen und damit das gezielte Erzeugen von *Bloomfiltern*. Daher sollen diese *Seeds* der fTTP nicht bekannt gemacht werden.

Die gemeinsame Abstimmung der *Seeds* sollte für jede Iteration erfolgen und jedes Mal neu vergebene *Seeds* enthalten. Über alle Krebsregister hinweg, sollten diese dann bekannt gemacht werden und in der *Domänen-Konfiguration* Vorlage ausgetauscht werden. Ein *Seed* besteht dabei immer aus einer Ganzzahl.

Beispiel:

Iteration 1:

Vorname (Seed A): 4623829476

Nachname (Seed B): 456542343

Iteration 2:

Vorname (Seed A): 374027465

Nachname (Seed B): 8364293

Im gleich gezeigten Beispiel, wurden für die Iteration folgende *Seeds* festgelegt:

Seed A (Balanced Bloomfilter / balanced): 1111111111

Seed B (Vorname / firstName): 2222222222

Seed C (Nachname / lastName): 3333333333

Seed E (Geburtsdatum / birthdate): 4444444444

Seed F (Postleitzahl / value3): 5555555555

```
31 <privacy>
32   <bloomfilter-config>
33     <algorithm>org.emaui.icmvc.ttp.deduplication.impl.bloomfilter.RandomHashingStrategy</algorithm>
34     <field>value8</field>
35     <length>1000</length>
36     <ngrams>2</ngrams>
37     <bits-per-ngram>25</bits-per-ngram>
38     <fold>0</fold>
39     <alphabet>ABCDEFGHIJKLMNOPQRSTUVWXYZ.-0123456789mfoux</alphabet>
40     <balanced>
41       <seed>1111111111</seed>
42     </balanced>
43     <source-field>
44       <name>firstName</name>
45       <seed>2222222222</seed>
46     </source-field>
47     <source-field>
48       <name>lastName</name>
49       <seed>3333333333</seed>
50     </source-field>
51     <source-field>
52       <name>birthDate</name>
53       <seed>4444444444</seed>
54     </source-field>
55     <source-field>
56       <name>value3</name>
57       <seed>5555555555</seed>
58     </source-field>
59   </bloomfilter-config>
60 </privacy>
```

Die einzelnen Krebsregister haben die *Seeds* übernommen. Alternativ kann das Krebsregister, welches die *Seeds* festgelegt hat, die angepasste Konfiguration einfach verteilen (z.B. per Mail).

5.3 Bloomfilter erzeugen

Zunächst muss die nun angepasste *Domänen-Konfiguration* in den E-PIX geladen werden. Über die Weboberfläche kann dies unter dem Menüpunkt „Domänen / Quellen / Identifier“ über die Schaltfläche „Importieren“ erfolgen. Die Einstellungen wurden übernommen. Es muss nur noch ein Name gewählt werden und die Domäne über den Button „Hinzufügen“ angelegt werden.

Bloomfilter

Algorithmus

RandomHashingStrategy

Alphabet *

ABCDEFGHIJKLMNOPQRSTUVWXYZ-0123456789mfoux

Länge *

1.000

Bits

Länge der N-Gramme * ⓘ

2

Bits pro N-Gramm * ⓘ

25

Bits

Anzahl der XOR-Faltungen * ⓘ

0

Balanced Bloomfilter ⓘ

☒

Seed *

1.111.111.111

Speicherfeld *

Feld 8

Quell-Felder

Feld	Seed
Vorname	2222222222
Nachname	3333333333
Geburtsdatum	4444444444
Feld 3	5555555555

+ Quell-Feld hinzufügen

Speichern

Abbrechen

Über die Import Funktion können wie bereits in Abschnitt „2. Erprobung der lokalen Domänen-Konfiguration“ gezeigt, können Listen von Personendatensätzen eingespielt werden. Die *Bloomfilter* werden dabei direkt erzeugt.

E-PIX®

Import

Dashboard

Personen

Dublettenauflösung

Suchen / Bearbeiten

Hinzufügen

Listen

Import

Export

Einstellungen

Domänen / Quellen / Identifier

Protokolle

Info

Aktive Domäne

Iteration 1

Erfolgreiche Importe: 80

Matchstatus	MPI	Vorname	Nachname	Geschlecht	Geburtsdatum	Postleitzahl	
NO_MATCH	1001000000011	Lukas	Müller	Männlich	14.03.1992	10115	Q1V9m+Tu/Y3vSd3rsJgofBkCOLRZJi2QcnloxFpF2ruDhTAc
NO_MATCH	1001000000028	Anna Sophie	Schneider	Weiblich	22.07.1988	50667	E0ohDCVoHF7/883TJnzeOnT4XVzHoBdyBDHLAOIXONIRi
NO_MATCH	1001000000035	Maximilian	Weber	Männlich	05.11.2001	80331	0x/CzJ5a5S35z83wYqSiTTKAIDL9Z/0YBWL83tLyzi3entPm
NO_MATCH	1001000000042	Marie	Lange	Weiblich	08.01.1997	1067	pEOXfmas/UaRiWOSrbbvIQtCKi06LnWx/TMINBxMajydKaN
NO_MATCH	1001000000059	Felix	Becker	Männlich	30.06.1990	28195	yZc3AHT0uef5XnvMejAOXDQdnK51dLWVeC3oWEfd6B3bI
NO_MATCH	1001000000066	Laura Marie	Hofmann	Weiblich	17.09.1995	40210	wsq5H0quWEQw/UnQKj6fPzx4M+wgxmeli2e4J+B6UKj43at
NO_MATCH	1001000000073	Paul	Richter	Männlich	12.04.1985	99084	BZNRqGVknHLpaN8AZLsXID2pg+5DeUvgLTggHkDaJTR8
NO_MATCH	1001000000080	Sophie Charlotte	Wagner	Weiblich	25.12.2003	4109	spsgHkXmvmxnt10X62iyNkHzy1hp5iIRSVTuAEG5Fer81z7Qi
NO_MATCH	1001000000097	Leon	Schäfer	Männlich	08.02.1993	20095	XzJlwa9svO+bz01QepENXhZDqRZdBuJGSot9tZPah/5mUi
NO_MATCH	1001000000103	Emilia	Neumann	Weiblich	19.10.2000	90402	0x2ITFBsCU2xu87wQHaxPz.Jps34MYO2AzoolqIgaRp77ahI

1-10 von 80

CSV herunterladen

Neuen Import beginnen

Zwar könnte bereits an dieser Stelle ein Export der *Bloomfilter* vorgenommen werden, jedoch würde die resultierende CSV-Datei alle gezeigten Angaben (Matchstatus, MPI, Vorname, Nachname, ...) beinhalten. Um nur die *Bloomfilter* zu extrahieren, sollte die Export Funktion verwendet werden.

Unter dem Menüpunkt „Export“ können über den Button „Alle Personen extrahieren“ und anschließend „Suche“ exportiert werden. Mit einem Klick auf „Leere Spalten ausschließen“, werden alle leeren Spalten vom Export ausgeschlossen. Die letzten Spalten (Vorname, Nachname, etc.) müssen

händisch mit einem Klick auf das „X“ abgewählt werden. Nur „Feld 8“ (bzw. der *Bloomfilter*) soll angewählt bleiben.

Danach kann der Export über den Button „CSV herunterladen“ gestartet werden. Die resultierende Datei enthält nun nur noch die *Bloomfilter*. Die ersten zwei Zeilen der CSV-Datei können bei Bedarf entfernt werden.

```
1 sep=;
2 value8
3 Q1V9m+Tu/Y3vSd3rsJg4fhIKOLRZB12QcnIowFJV2ruDhTAc+oOnKqevGVPu41wVvSXR460W87J94qHtq8RWIYBovMc5yDsQu7nflFTF9AoiMLok
19eEQi/6SN9ouz9nfNvSr3teKzirkfS4I/ODDj0ZS4Qg/KWgATB/8Dpxj1S+AZnY23Z5nWL8GWNN7h3e7V6xiG0iicC6APucA5nlakio9YmmIa94
dwiSFYM/g7p/gK6oz+x++9H3BRUE+Hyre6KGE2c4/g10qKHoyP2WoxkcbMm6UoMLkw1EE71DhqCwVedQFFPumiVoqsBvhkJVr4ColxFBZzPnZg==
4 E0ohDCVgHF7/883Sjnze0HT5XFzFoBDyBDGKA0iXenFrRvKbnZnM9VN3DHOvWBUKQXz58Sgfc4UuiKy14MxsfZJf+kTLKRAERiNaJrj9cRe06Yh
D08X0EL7Jv580pkXaMFOPp1t2xFZCOEen/tFADx0/uMpBVuIhhEu6ctAXhhXoeFcQSTWjmRbCf/UYw1WabTQYCLD00YPpb1UB0bJ7NepRy96sv74
w611b0N4xn6/QmnrW0KPvJk5/W3fXOtIjplYm4ouuj87LicT4oM2DmfB8+sFpdHkeUvJztiFg8ZoXE/EO+M3B48cJfORKdof8Y9X2sz3LVVw==
5 0x/Czj5s535z83wYqSittKAIDL9Z/0YBWLb3tgLyzi3entPmsFIN1ivVpJib76N41Avy5TCb0+IFyebdZ0bEnUW6Y0qfJ8NtDNApNvxspImmLi
JDDuXyWPSUnpb+3GdkY8ydhniZBjCxlTYE1FheZrHzyibmyj1UrD6g13G1Dmga1h6ShXh8LqbYCYZ/4FXOyck9CZPxsMSonW1HTHHbuHmRTuYryo
Ww0HuMp8uyINKikwbLUQJNe+rp5bzDpjiXqH2WkjtdBGLd87eM94qBi7qnT/z5XTtgWEV3AIwaMeS1TPzgtacX2f9Cu1j2MR/Ko3mG9N1GNQ==
6 JE9Xfmas/UaTiWOSrbbv1QtkK1w6LnWx/TMINBwMaj6dKaPom6Fq+RZWPf114NUuNWX5oyz0zw1+/KJ3CB6G0+71trkTmKDSY7xDbhMsvYpP4uWZ
5YRUZzS/0c3wjL+dGayj9ZNYWxS5AAN0NvelhpRtUKduNB0q3Mk2QPFsU1DWSAfZS2QPDSPqIwJS4zTUD1tReYPTCUCS0RNQ1Yf1SEbjpm4EgIwp
+qRdn2vGoBm8qvsWcmYpADDKbjqfBc47roLo/w+hoXMSWG6Lr8K3u8Ts/OuhJ00eZMBj5NBDG+1GXB3F0y0MtAuVTPmgT9o5c6Yj8jy+03Zw==
7 yzc3AHT0Pef7X3vI8jAOXHSdnK51ZLWEWC3oUMbd6BzbdRiR78JcVwdpJVPtUdsIq+1ArHQeFdNHs8GUb5bQwxS1cM4qxd8B5SZfcCMZWoiaaW5
30WAZyZyCX1slv1mevOamIRrA5tqS/zTVh0YxS9kn6S+ZWcjmdiyxgl+Pkj0ZiLLSiR/pkDmWxCw4wVEqSLlyN2GUCfCPP21E5gLNpJjKaDmxs48
1y0nMpgcUFY3aq/p3W8KAUM0z4vF3izVzKJom9cusfKb1a8j7FCjsrJypKtns4peISonLxn1rI+CeVTFM0jCjBmWfNL8sm9Hb5QW33E3ZH0Pw==
8 wsq5m0qmGAQw/U3YaJ6XPzz4M+wgxmEiyu4N+Ba00h43SDGyzLsD3IYZ0quizfVsnTQxv2yXi/s6GKQPTKQKgm1yw5EJ9KXqbcCMv3Ks1skwVEK6d
0H7ULDvYk8tkpDXQPZVL71U+n8Tmkf7iNKzGppF805Qjxd081Nj8l1j0AL2wetUVTHko0e0UnOcX4VdHsJmueuDW0KZcCjJdf3hMRnj9p4Dx7c/
foHDR9oZw1s00KBk012D1U7An/V4ExS5YHJJFIM34x08rG21XckBkFwe1hfe1nt8g0yZ7aFsm+w+0BFAzAy9m01RPzXYtS61oYP0LZys39Q==
9 BZNRqGvKnHlpn8AZLsv1D2pg+5DeUvGLTggHkDaJTR8ChKnvEio3TV96HEZqZcmTX+06hWENEJxoeblyCzttxoa4GQ5bfpbRbmb70RuIfMoq8
wsfPne4/bvvh+pf3ejergmEtuFKGie3oi/4NM729Xbmjdsps19+E6MnGyJVET1RSBW7u0Hbc12ZTu2G51lx2AaH2Mg01+hxAwKreE3i6Q5KbvrP
0mxm31K1U0ccUyguAcG38htCGfMb/DajPYKEZR0ojw1bVLHGDZCf6ReQPjbyhr1naRdTr1CGGkcyg0XfK6Cfjce1JzuQfx146SiNc0be/ptw==
```

5.4 Bloomfilter bei der fTTP registrieren

Die soeben erzeugten *Bloomfilter* müssen nun an die fTTP übermittelt werden. Dies kann auf verschiedene Weisen passieren. Sofern keine eigene Lösung zur Verfügung steht, kann die Registrierung über einen *jmeter*-Test durchgeführt werden. Die gerade erzeugte CSV-Datei wird dabei automatisch eingelesen und die *Bloomfilter* einzeln bei der fTTP registriert. Alternativ kann ein fertiger Docker-Container bereitgestellt werden, der die *Bloomfilter* aus dem E-PIX extrahiert und an die fTTP übermittelt. Beide Lösungen können bei Bedarf zur Verfügung gestellt werden.

5.5 Pseudonyme erhalten und ggf. Datensätze abgleichen

Die fTTP gleicht die *Bloomfilter* ab und erzeugt kontextfreie, zufallsgenerierte Zeichenketten als Pseudonyme. Diese beinhalten keinerlei Informationen über den *Bloomfilter*. Identische oder sehr ähnliche *Bloomfilter* werden einer Person zugeordnet und erhalten dasselbe Pseudonym. In

unsicheren Fällen sollen die Krebsregister untereinander prüfen, ob zwei Datensätze tatsächlich zu einer Person zugehörig sind, oder ob es sich um verschiedene Personen handelt. Hierzu führen die Krebsregister untereinander einen Abgleich durch. Das genaue Vorgehen befindet sich noch in Abstimmung und wird hier entsprechend ergänzt.

Nach Abschluss einer Iteration, löscht die fTTP alle *Bloomfilter*, die während der Iteration übermittelt wurden. Danach erfolgen die Bereitstellung einer neuen *Domänen-Konfiguration* und eine neue Iteration beginnt.

6. Testpaket

Das Testpaket automatisiert einige oben genannte Schritte. Hierzu wird intern das Programm *jmeter* verwendet. Es handelt sich hierbei um ein separates Testpaket (Abweichend vom Testpaket zum Testen der Anbindung an die fTTP (Abschnitt 4.2 Testen der Anbindung)). Dieses beinhaltet drei Docker-Container: einen E-PIX, eine MySQL-Datenbank und *jmeter* um mit der fTTP zu kommunizieren. Das Docker-Paket kann unter <https://www.ths-greifswald.de/forscher/ftp/cancerlink/docker-testpaket> heruntergeladen werden.

6.1 Beschreibung

Das bereitgestellte Paket besteht aus einem angepassten Docker-Compose, welches neben dem E-PIX (und der MySQL-Datenbank) auch einen angepassten *jmeter*-Test beinhaltet. In den folgenden Abschnitten werden die Installation und Konfiguration beschrieben. Nach dem Hochfahren startet der hinterlegte *jmeter*-Test. Dieser legt die iterationsspezifische Domäne im E-PIX an. Die dafür benötigte Domänenkonfiguration muss pro Iteration entsprechend ausgetauscht werden. Danach werden die hinterlegten Daten (im CSV-Format) in den E-PIX eingespielt und ein Bloomfilter erzeugt. Dieser Bloomfilter wird an die fTTP übermittelt und ein Pseudonym zurückgeliefert. Dieses Pseudonym wird dem jeweiligen Datensatz im E-PIX angefügt. Danach erfolgt der Export über die E-PIX Oberfläche. Pro Iteration ergeben sich daraus zwei manuell durchzuführende Aktionen:

1. Hinterlegen der Iterationsspezifischen Domänenkonfiguration (im XML-Format)
2. Export der Datensätze mit dem erhaltenen fTTP-Pseudonym.

Die Test-Iterationen sind damit weitestgehend automatisiert und sollen die manuellen Aufwände auf ein Minimum reduzieren.

6.2 Installation

Die Installation des Testpakets erfolgt zunächst so, wie als wenn der E-PIX installiert werden würde (siehe Handbuch). Das vorliegende Paket ist jedoch etwas optimiert, sodass weitere Anpassungen nur nach Bedarf erfolgen müssen. Folgende Voraussetzungen sollten gegeben sein:

1. Docker ist installiert
2. Das verwendete System besitzt zumindest 8 GB Arbeitsspeicher
3. Das verwendete System kann die fTTP erreichen
4. Die Docker Images (E-PIX, MySQL und *jmeter*) sind vorhanden

Sofern der E-PIX bereits installiert wurde und die Anbindung an die fTTP bereits getestet wurde, sind fast alle Voraussetzungen schon erfüllt. Sollte das verwendete System freien Zugang zum Internet haben (von diesem Zustand wird nicht ausgegangen), so kann das vorliegende Paket einfach gestartet werden. Alle Abhängigkeiten werden dann aus dem Docker-Hub heruntergeladen. Da jedoch nicht davon ausgegangen wird, müssen die entsprechenden Abhängigkeiten anderweitig heruntergeladen werden. Dies wird hier gezeigt:

1. System wählen, welches einen Internet-Zugang besitzt (und Docker installiert hat)
2. Abhängigkeiten/Docker-Images herunterladen

```
docker pull mysql:8
docker pull mosaicgreifswald/wildfly:38
docker pull mosaicgreifswald/jmeter:5.6.3
```

3. Abhängigkeiten/Docker-Images verpacken

```
docker save mysql:8 -o mysql8.tar
docker save mosaicgreifswald/wildfly:38 -o wildfly38.tar
docker save mosaicgreifswald/jmeter:5.6.3 -o jmeter563.tar
```

4. Verpackte „tar“-Dateien auf das Zielsystem übertragen

5. „tar“-Dateien entpacken:

```
docker load -i mysql8.tar
docker load -i wildfly38.tar
docker load -i jmeter563.tar
```

Die benötigten Abhängigkeiten sind nun auf dem Zielsystem. Nun kann die einmalige Konfiguration durchgeführt werden.

6.3 Konfiguration

Die Konfiguration beinhaltet alle Schritte, die einmalig durchgeführt werden müssen. Etwaige Anpassungen für die Test-Iterationen, werden im nächsten Abschnitt erläutert. Im vorliegenden Testpaket, wird folgende Datenstruktur vorgefunden:

```
cancerlink-docker/
├── addins/
├── certificate/
├── deployments/
├── envs/
├── jboss/
├── jmeter-props/
├── jmeter-tests/
├── logs/
├── sqls/
└── docker-compose.yml
```

Folgende Schritte müssen durchgeführt werden:

1. Das Client-Zertifikat (.pfx-Datei) in `cancerlink.pfx` umbenennen.
2. Die eben benannte Datei in den Ordner `certificate/` kopieren.
3. Im Ordner `envs/` die Datei `secrets.env` anpassen:

Die Datei besteht aus den drei Werten: `USERNAME`, `PASSWORD` und `APIKEY`. Wurde bei der Zugangsbeantragung ein IP-Login gewählt, so bleiben die Werte für `USERNAME` und `PASSWORD` leer.

Bitte wählen Sie eine Zugangs-Variante aus

- ☐ Login und Client-Zertifikat
☐ IP und Client-Zertifikat

Wurde hingegen ein Zugang mit Login gewählt, so müssen bei `USERNAME` der gewählte Benutzername und bei `PASSWORD` das gewählte Passwort eingetragen werden. Danach muss in jedem Fall der API-Key eingetragen werden. Dieser wurde bei der Zugangsgenehmigung übermittelt und wird als Wert für `APIKEY` gesetzt.

1	<code>USERNAME=MeierM</code>	1	<code>USERNAME=</code>
2	<code>PASSWORD=qJ_H4T'Uvf+K>NB\$</code>	2	<code>PASSWORD=</code>
3	<code>APIKEY=dS8mG7mThnQai3YwicfzJIq6iAoBsgT5</code>	3	<code>APIKEY=dS8mG7mThnQai3YwicfzJIq6iAoBsgT5</code>
4		4	

4. Im Ordner `jmeter-tests/data/` werden die Daten, welche zur Analyse verwendet werden, in eine CSV-Datei `data.csv` geschrieben. Folgendes ist dabei zu beachten:
 - a. Die CSV-Datei ist UTF-8 kodiert.
 - b. Werte sind mit einem Semikolon separiert.

- c. In der ersten Zeile befindet sich die Kopfzeile, welche folgende Werte enthält:
firstname, lastname, birthdate, gender, postalcode
 - d. Innerhalb der Daten muss sichergestellt sein, dass keine Semikolons auftreten
 - e. Das Geburtsdatum ist im Format JJJJ-MM-TT
 - f. Das Geschlecht ist kodiert als:
 - i. M für Male bzw. männlich
 - ii. F für Female bzw. weiblich
 - iii. D für Divers bzw. divers
 - iv. U für Unknown bzw. unbekannt
 - g. Beispiel:
Lukas;Müller;1992-03-14;M;10115
 - h. Im Ordner liegt eine exemplarische Datei. Diese muss ersetzt werden.
5. In der Datei `docker-compose.yml` muss in Zeile 35 der Platzhalter `<PASSWORD>` mit dem Passwort ersetzt werden, welches bei der Erzeugung des *Certificate Signing Requests* gewählt wurde (siehe Abschnitt 3.1 Certificate Signing Request erzeugen).

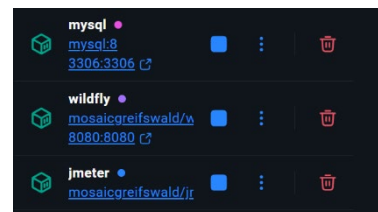
Damit ist die Konfiguration abgeschlossen. Alle weiteren Einstellungen erfolgen pro Test-Iteration und werden im nächsten Abschnitt erläutert.

6.4 Durchführen der Test-Iterationen

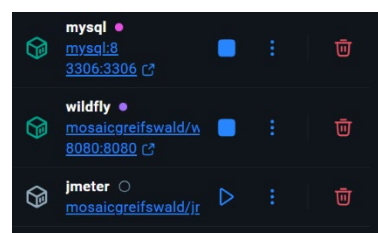
Jeder Durchlauf beginnt mit dem Austausch der E-PIX Domänenkonfiguration. Die Vorlage für jede Iteration wird von der Treuhandstelle Greifswald als XML-Datei bereitgestellt. Die darin enthaltenen Seeds müssen einheitlich ersetzt werden. Dies kann z.B. von einem Standort übernommen werden und den anderen Standorten bereitgestellt werden.

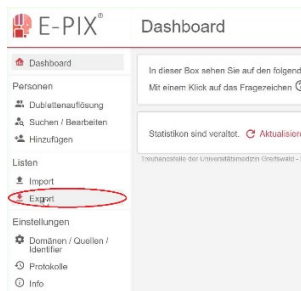
Die Treuhandstelle Greifswald teilt den Start einer Iteration mit. Die Standorte fügen die `domain.xml` mit den definierten Seeds im Ordner `jmeter-tests/domain_xml/` ein. In diesem Ordner liegt stets nur eine Datei, welche die entsprechende Konfiguration beinhaltet (diese muss stets `domain.xml` heißen).

Die Standorte können dann beliebig die Iteration starten. Dies geschieht, indem das Docker-Paket gestartet wird. Hierzu kann z.B. die Konsole geöffnet werden und in das Stammverzeichnis des Pakets (`cancerlink-docker/`) gewechselt werden. Mit dem Befehl `docker compose up -d` starten die Docker-Container. Im Bild links wird gezeigt, dass alle Container hochgefahren sind. Nachdem der E-PIX hochgefahren ist, startet automatisch der Durchlauf. Dabei werden alle Datensätze der hinterlegten `data.csv` im lokalen E-PIX registriert und der resultierende *Bloomfilter* wird an die fTTP übermittelt. Die fTTP liefert ein Pseudonym zurück, welches im E-PIX an den entsprechenden Datensatz angefügt wird.

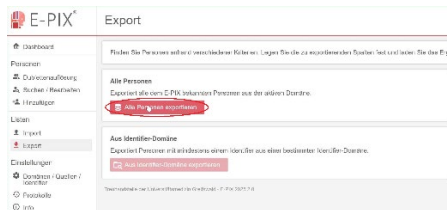


Wurden alle *Bloomfilter* an die fTTP übermittelt, beendet sich der *jmeter*-Container automatisch. Das Bild links zeigt, dass der *jmeter*-Container beendet wurde. Im Log kann dies über die Zeile `[run.sh:149] successfully finished action /opt/mosaic/jmeter/run_jmeter.sh` entnommen werden. Es kann nun die Weboberfläche vom E-PIX aufgerufen werden, um die erhaltenen Pseudonyme zu exportieren: <http://localhost:8080/epix-web>.

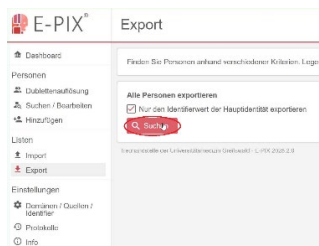




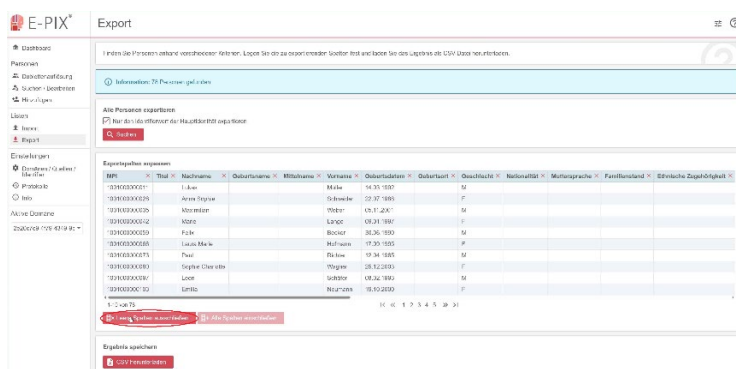
1. Zunächst wird der Reiter Export ausgewählt.



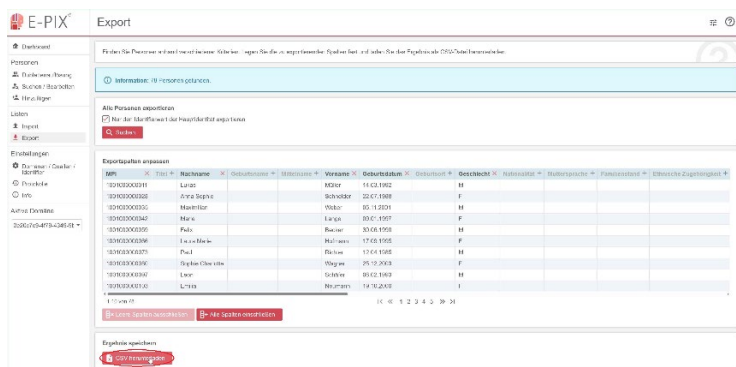
2. Als nächstes wird die Schaltfläche *Alle Personen exportieren* angewählt.



3. Danach können alle Personen exportiert werden, indem auf die Schaltfläche *Suchen* geklickt wird.



4. Es werden alle Datensätze aufgeführt. Bei einem Export werden alle Spalten exportiert. Um die Anzahl der Spalten auf jene zu beschränken, die überhaupt belegt sind, können mit Klick auf die Schaltfläche *Leere Spalten ausschließen* alle leeren Spalten ausgeblendet werden. Bei Bedarf kann auch die Spalte „Bloomfilter“ ausgeblendet werden. Diese wird für eine Auswertung nicht benötigt.



5. Als letztes kann mit Klick auf die Schaltfläche *CSV herunterladen* der Export durchgeführt werden. Die Datei sollte so umbenannt werden, dass diese der Iteration zugeordnet werden kann. Beispiel: `result_iteration_1.csv`

Nach dem Export ist die Iteration abgeschlossen und die Daten können gelöscht werden. Hierzu genügt es, dass Docker-Paket per `docker compose down` herunterzufahren. Um alle Daten der Iteration

restlos zu entfernen, kann mit dem Befehl `docker volume prune` das soeben erzeugte Volume restlos entfernt werden. Sobald alle Standorte diese Iteration vollendet haben, werden auch die Bloomfilter in der fTTP restlos entfernt. Danach beginnt die Abstimmung zur nächsten Iteration.

Hier sind alle Schritte nochmals im kurzgefasst dargestellt:

1. XML-Datei mit Domänenkonfiguration mit jeweiligen Seeds anpassen und im Ordner `jmeter-tests/domain_xml/` als `domain.xml` ablegen. Ggf. vorbereitete XML-Datei verwenden.
2. Mit `docker compose up -d` das Testpaket starten und warten, bis der Test abgeschlossen ist (jmeter-Container ist beendet).
3. Datenbestand über die E-PIX Oberfläche exportieren und die Datei umbenennen, sodass diese später der iteration zugeordnet werden kann.
4. Mit `docker compose down` das Testpaket herunterfahren. Das Testpaket ist damit bereinigt und beim nächsten Hochfahren wird die nächste Iteration ausgeführt.

Glossar

Begriff	Beschreibung
Domänen-Konfiguration	Dies umfasst alle Einstellungen einer E-PIX Domäne, für eine spezifische Iteration. Dies umfasst z.B. die Bloomfilter-Konfiguration (Länge, Anzahl verwendeter Bits, Härtingsverfahren), Normalisierungen etc. Diese Domäne wird im XML-Format beschrieben und kann vom E-PIX eingelesen werden.
Bloomfilter	Ein Bloomfilter ist ein Bit-Vektor, der initial mit 0 besetzt ist. Daten die in den Bloomfilter codiert werden, werden derart umgewandelt, dass diese Daten mehrere Positionen im Bloomfilter auf 1 setzen. Verschiedene Daten können dabei dieselbe Position auf 1 setzen. Zusätzliche Härtingsverfahren, sichern den Bloomfilter zusätzlich ab, sodass keine Rückschlüsse auf die eigentlichen Daten gezogen werden können. Es handelt sich dabei um eine Einweg-Verschlüsselung, wie auch von anderen Hash-Verfahren bekannt ist. Der Unterschied besteht darin, dass ähnliche Eingangsdaten auch zu ähnlichen Bloomfiltern führen und somit auf Ähnlichkeit verglichen werden können. Andere Hash-verfahren führen auch bei sehr ähnlichen Eingangsdaten zu völlig unterschiedlichen Hashwerten.
Seed	Ein Seed ist ein Startwert für einen Pseudozufallsgenerator. Ein Pseudozufallsgenerator erzeugt zufällig aussehende Zahlenserien. Beim gleichen Seed werden also dieselben Zahlenserien generiert. Der Seed dient in diesem Kontext als Passwort, da nur mit dem Wissen des Seeds die spezifische Zahlenserie erzeugt werden kann und damit eine Codierung auf dieselbe Weise stattfinden kann.
Certificate Signing Request	Ein Certificate Signing Request ist ein digitaler Antrag zur Erstellung eines Zertifikats. Dieses enthält diverse Angaben zum Antragssteller und finden sich so später im fertigen Zertifikat wieder. Das CSR wird dabei auf Basis einer Key-Datei erzeugt, sodass nur der Antragssteller das finale Zertifikat verwenden kann.
federated Trusted Third Party	Die federated Trusted Third Party ist eine Stelle, welche Datensätze miteinander abgleicht. Im Unterschied zu konventionellen Treuhandstellen, welche identifizierende Daten zentral oder im Klartext abgleichen, ist eine fTTP speziell für den übergreifenden Abgleich codierter Daten konzipiert. Die identifizierenden Daten bleiben dabei bei den datenhaltenden Standorten und müssen nicht zentral bekannt gemacht werden. Die fTTP ermöglicht es über codierte Daten dennoch Daten übergreifend zusammen zu führen.
jmeter	jmeter ist eine Software zur Ausführung von Lasttests zwischen Client und Server. Es ermöglicht Anfragen zu modellieren und diese an einen Server zu übermitteln und die entsprechenden Antworten entgegen zu nehmen.

Abkürzungsverzeichnis

BF	Bloomfilter
CSR	Certificate Signing Request
E-PIX	Enterprise Identifier Cross-Referencing
fTTP	federated Trusted Third Party
gPAS	generic Pseudonym Administration Service
IDAT	Identifizierende Daten
IP	Internet Protokoll
MPI	Master Patient Index