Unabhängige Treuhandstelle

UNIVERSITÄTSMEDIZIN GREIFSWALD

Community Dialog #11

Martin Bialke | Universitätsmedizin Greifswald | 13. Juni 2024

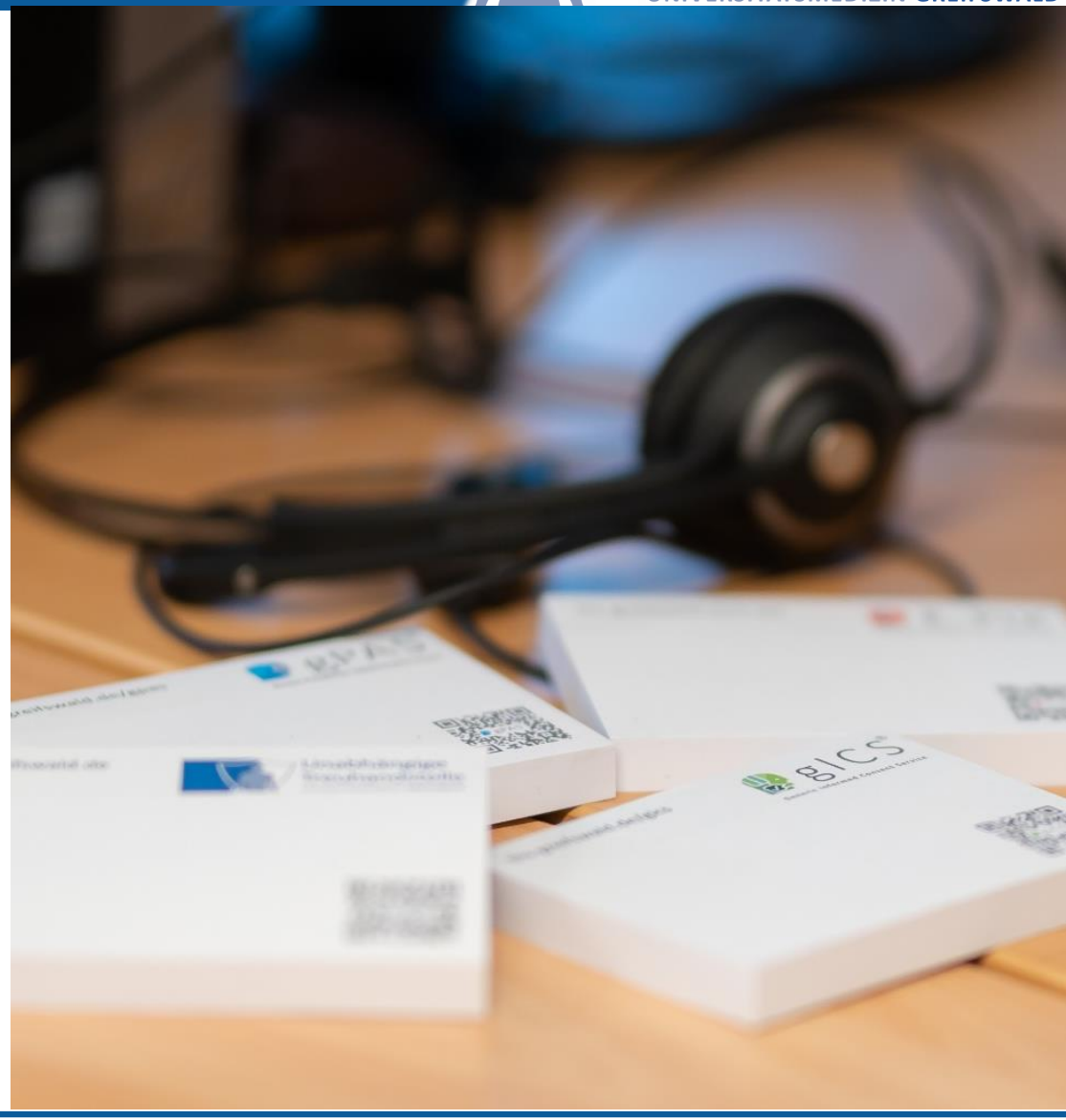
Community-Dialog: direkter Austausch mit den Anwendern

Unabhängige Treuhandstelle
UNIVERSITÄTSMEDIZIN GREIFSWALD

Der **THS Community Dialog**:

- bietet eine zentrale Anlaufstelle **für Anwender-Fragen, die über den Inhalt der Handbücher hinausgehen**
- hilft Hintergründe besser zu verstehen und gemeinsam mit den Entwicklern erforderliche Antworten zu finden.
- schafft die Basis für eine gemeinsam aufgebaute FAQ.
- ist Möglichkeit mit anderen eigene Ideen zu diskutieren und zu teilen

ths-greifswald.de/community



Agenda

Top 1 Frischer Wind: v2024.1.0 von E-PIX, gPAS und gICS

Top 2 Neue gICS-Funktionalität: Opt-Out Unterstützung

Top 3 Neue E-PIX-Funktionalität: Validatoren

Top 4 Q&A: Fokus E-PIX Matching

Top 5 Die Community stellt sich vor

Nico Schmid (RBK): Integration von E-PIX und der Statistiklösung R

TOP 1 - NEUE RELEASES V2024.1.0

TOP 1 Neue Releases (Juni 2024)

Version 2024.1.0

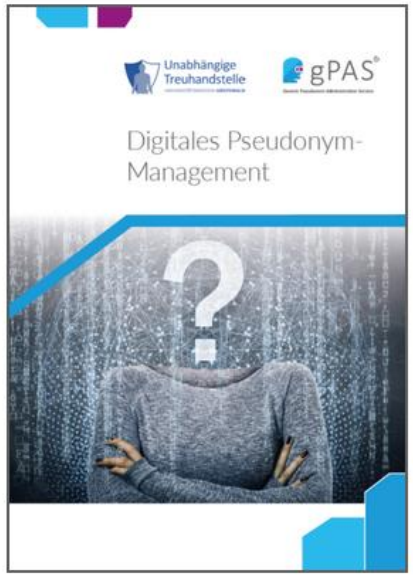
» Release Notes

» Handbuch

» API

» Update-Hinweise

HERUNTERLADEN



Version 2024.1.0

» Release Notes

» Handbuch

» API

» Update-Hinweise

HERUNTERLADEN

Version 2024.1.0

» Release Notes

» Handbuch

» API

» Update-Hinweise

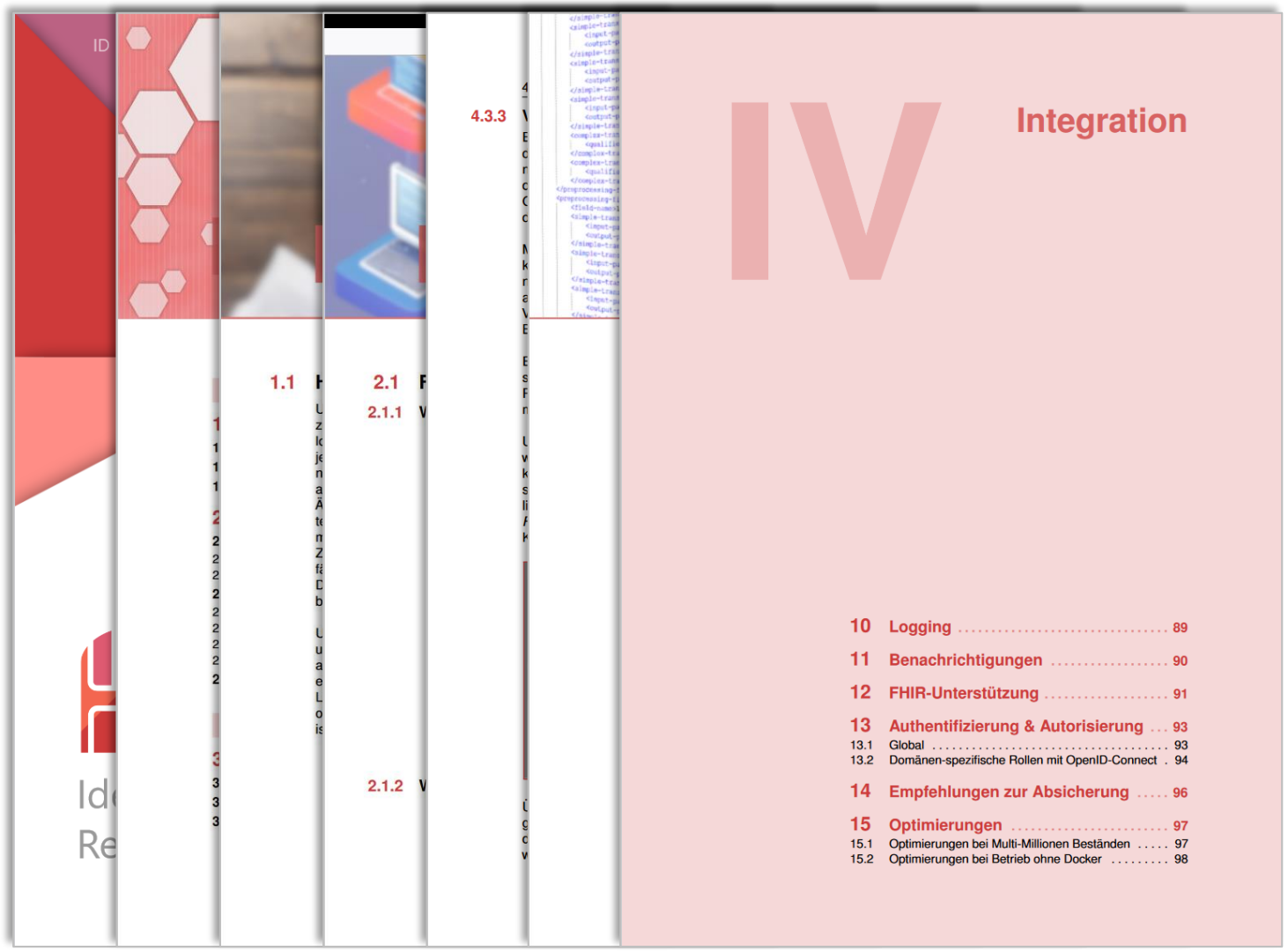
HERUNTERLADEN

Umfassendes Re-Factoring: Beginn Umstellung der Handbücher

Unabhängige Treuhandstelle
UNIVERSITÄTSMEDIZIN GREIFSWALD

Neues E-PIX Handbuch

- Neue technische Basis (Latex)
- Neues Layout und einheitliches Farbkonzept
- Neue (einheitliche) Struktur durch inhaltliche Abschnitte
- Grundlage für Umstellung gPAS, gICS und weitere Internationalisierung



Allgemeine Anpassungen – alle Tools

- **Einheitliche Umstellung auf JakartaEE**
- Anpassbarer Context Root der Weboberfläche und SOAP-Schnittstelle
- Allgemeine Verbesserung im Frontend
- Dashboard Statistiken für Summe aller Domänen
- Wählbarer Start und Endzeitpunkt in Statistik-Diagrammen
- Umbenennung von Konfig-Variablen

Version 2024.1.0

- Details unter ths-greifswald.de/e-pix/releasenotes

Neuerungen und Verbesserungen (Auswahl, inkl. 2023.2.0)

- Dashboard: Anzeige zurückgestellter Dubletten
- Mehr Hinweise bei Konfiguration einer Domäne in der Weboberfläche
- Export und Import der Felder Vitalstatus und Sterbedatum
- Verbessertes Feedback in der Oberfläche, wenn IDAT bei Änderung stark abweichen
- **Umfangreich konfigurierbare Validierung von Personen- und Zusatzfeldern**

E-PIX®

Version 2024.1.0

- Details unter ths-greifswald.de/gpas/releasenotes

Neuerungen und Verbesserungen (Auswahl, inkl. 2023.2.x)

- Import von Daten (Originalwert-Pseudonym Paare) über die Weboberfläche
- **Konfiguration zur Validierung von Originalwerten mittels Eltern-Domäne in der Weboberfläche**
- **Vorschau der Anzahl möglicher Pseudonyme bei gewählter Konfiguration**
- Möglichkeit zur nachträglichen Änderung der Eltern-Domänen

Und wo sind die MultiPsn-Values? Kommt im nächsten Release

gPAS®

Version 2024.1.0

- Details unter ths-greifswald.de/gics/releasenotes

Neuerungen und Verbesserungen (Auswahl, inkl. 2023.2.x)

- Export einer Vorlage und Ausfüllen einer Einwilligung aus Vorlagenliste heraus
- Umfangreiche Qualitätsprüfung in der Weboberfläche
- **Konfiguration der Qualitätsstatusoptionen in der Weboberfläche**
- Überarbeitete Domänenkonfiguration und Bearbeitung in der Weboberfläche
- **Speichern einer Einwilligung, ohne dass eine der angebotenen Moduloptionen gewählt wurde, sofern eine versteckte Vorauswahl definiert wurde**
- Export von Vorlagen als PDF
- **Bereitstellung der Vorlagen und Policynutzung als CSV Download**
- **Opt-Out-Unterstützung: neue Vorlagen-Typen, ConsentStatus-Angaben und validFrom-Properties**

gICS®

Wo finde ich was bei ths-greifswald.de?

Unabhängige Treuhandstelle
UNIVERSITÄTSMEDIZIN GREIFSWALD

E-PIX®
- /e-pix/releasenotes
- /e-pix/handbuch
- /e-pix/update

gPAS®
- /gpas/releasenotes
- /gpas/handbuch
- /gpas/update

gICS®
- /gics/releasenotes
- /gics/handbuch
- /gics/update
- /gics/consentstate-and-legalconsentdate

Dispatcher
- /dispatcher/releasenotes
- /dispatcher

HL7 FHIR
- /ttpfhirgw/releasenotes
- /gics/fhir
- /gpas/fhir
- /e-pix/fhir

KEYCLOAK
- /ttp-tools/keycloak
- /ttp-tools/domain-auth

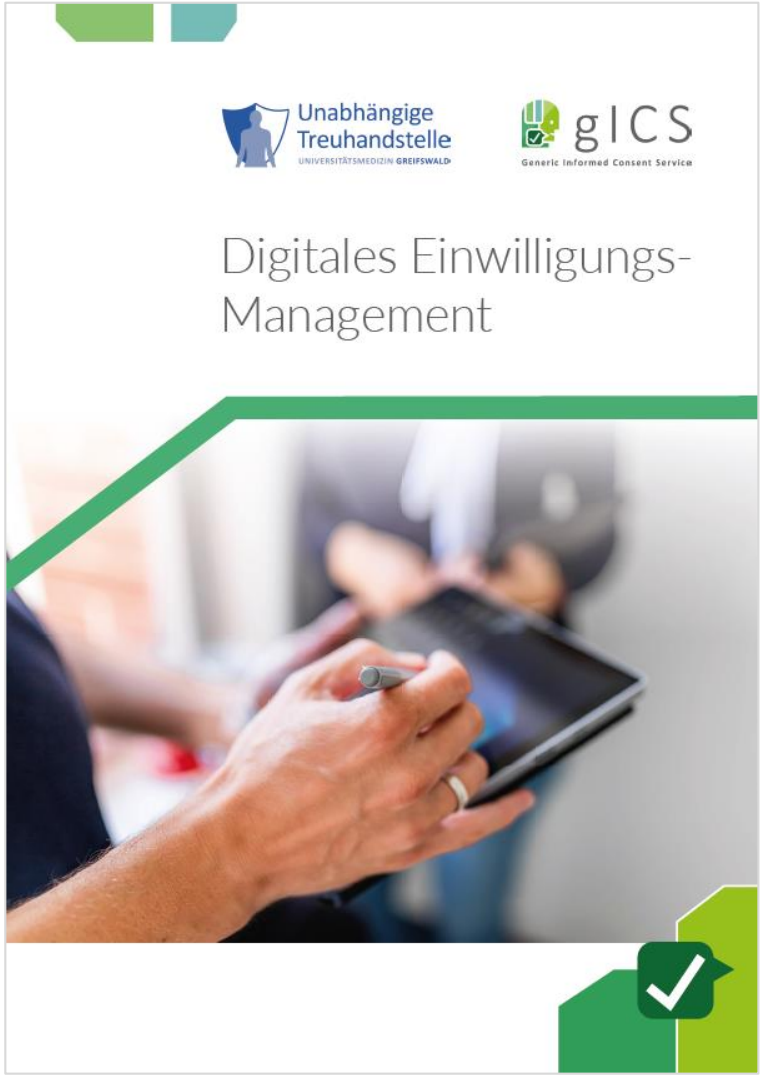
Notifications
- /ttp-tools/notifications

TOP 2 - **OPT-OUT UNTERSTÜTZUNG**

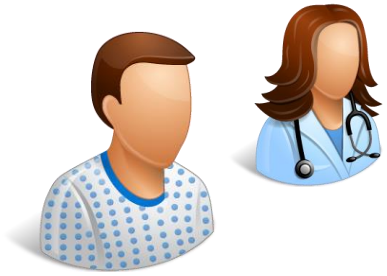
Einwilligungsmanagement gICS®

Unabhängige Treuhandstelle
UNIVERSITÄTSMEDIZIN GREIFSWALD

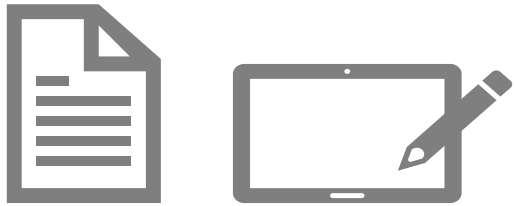
- Erstellung und Verwaltung **Einwilligungen, Widerrufe und Widersprüchen**
- Ermittlung des Einwilligungsstatus von Teilnehmern
- Speicherung digitaler Unterschriften und Scans
- Im- und Export von Konfigurationen und Vorlagen
- Unterstützung für Qualitätssicherung
- Umfangreiche Suchfunktionen
- Automatisierbarer Versand von Benachrichtigungen
- Technische Schnittstellen (SOAP, FHIR) zur System-Integration
- OIDC-Unterstützung per KeyCloak



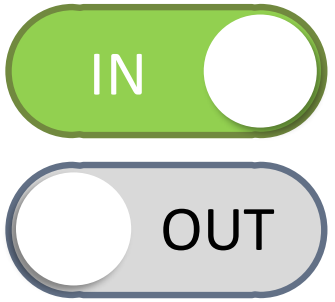
Unterstützung für verschiedene Einsatz-Szenarien



Medizinische Forschung,
z.B. Studien und Register



Papierbasiert & rein digital



Opt-In & Opt-Out

Opt-*?

- Datenverarbeitung (für die Forschung) auf Basis der expliziten Einwilligung einer Person (vgl. Art. 6, Abs.1 lit.a).
- Umkehrbar mittels **Widerruf**
- Mögliche Anwendungsszenarien sind z.B. Register und/oder Forschungsvorhaben: MII Broad Consent der Medizininformatik-Initiative, NAKO Gesundheitsstudie, Krankheitsspezifische Register

- *„Die Bereitstellung für die Forschung erfolgt automatisch, es sei denn es erfolgt ein* ***Widerspruch****.“* [1]
- Mögliche Anwendungsszenarien sind z.B. Register und/oder Forschungsvorhaben **mit gesetzlichem Auftrag:** Krebsregister (Krebsfrüherkennungs- und -registergesetz KFRG), Gesundheitsdatennutzungsgesetz (GDNG), Elektronische Patientenakte (ePA)

Quellen:
[1] https://www.bundesgesundheitsministerium.de/ministerium/gesetze-und-verordnungen/guv-20-lp/gesundheitsdatennutzungsgesetz/faq-gesundheitsdatennutzungsgesetz.html, Zugriff am 2024-06-10

Anforderungen an eine Opt-Out-Lösung

Unabhängige Treuhandstelle
UNIVERSITÄTSMEDIZIN GREIFSWALD

- Unterstützu … ruch“ -> Konformität … ent
- Kompatibilit … licies) und Funktionen
- Berücksichti … tsverzögerung von Opt-Out-Ein
- Automatisie … haltezeiten

Ideal

- Abwärtskom
- Ressourcen-

Vorlagentyp	Coding	Beispielformulierung	Mögliche Antworten code (display)	Resultierender provisionCodeType
Einwilligung (Typ: Opt-In)	System: http://loinc.org Code: 59284-0	*Ich willige ein, dass …*	`Gültig` (ja)	`Permit`
			`Ungültig` (nein)	`Deny`
			`Unbekannt` (unbekannt)	`Deny`
Widerruf	System: http://fhir.de/ConsentManagement/CodeSystem/TemplateType Code: WITHDRAWAL	*Hiermit widerrufe ich meine bestehende Einwilligung …*	`Gültig` (ja)	`Deny`
			`Ungültig` (nein)	keine Änderung des bislang gültigen Wertes
			`Unbekannt` (nicht gefragt/ nicht gewählt)	keine Änderung des bislang gültigen Wertes
Ablehnung	System: http://fhir.de/ConsentManagement/CodeSystem/TemplateType Code: REFUSAL	*Die betroffene Person lehnt … ab.*	`Gültig` (ja)	`Deny`
			`Ungültig` (nein)	keine Änderung des bislang gültigen Wertes
			`Unbekannt` (nicht gefragt/ nicht gewählt)	keine Änderung des bislang gültigen Wertes
Widerspruch	System: http://fhir.de/ConsentManagement/CodeSystem/TemplateType Code: OBJECTION	*Die betroffene Person hat einen Widerspruch … erteilt.*	`Gültig` (ja)	`Deny`
			`Ungültig` (nein)	keine Änderung des bislang gültigen Wertes
			`Unbekannt` (nicht gefragt/ nicht gewählt)	keine Änderung des bislang gültigen Wertes
Einwilligung (Typ: Opt-Out)	System: http://fhir.de/ConsentManagement/CodeSystem/TemplateType Code: CONSENT-OPT-OUT	*Die betroffene Person hat auf gesetzlicher Grundlage (ePA) eingewilligt.*	`Gültig` (ja)	`Permit`
			`Ungültig` (nein)	`Deny`
			`Unbekannt` (unbekannt)	`Deny`

Umsetzung Opt-Out im gICS

Paradigmen

1. Nur registrierte Personen(-kennungen) werden durch gICS berücksichtigt
   Vorteil: verhindert falsch-positive Aussagen zum Einwilligungsstatus
2. Nur dokumentierte Einwilligungsaspekte (Policies) sollen bezogen auf ein Forschungsvorhaben abfragbar sein
   Vorteil: verhindert „Generalvollmacht zur beliebigen Datenverarbeitung“
3. Dokumentation in gICS muss ohne Unterschrift und komplexe Modulstatuskonfigurationen auskommen
   Vorteil: ermöglicht niederschwellige Umsetzung und reduziert intern Datenmengen
4. Dokumentation von Widersprüche und Teilwidersprüchen sollen umsetzbar sein
   Vorteil: Konsistenz zu Widerrufsvorgehen und Konformität zu HL7 AG EWM

Möglicher Ablauf zum Einsatz der Opt-Out-Lösung

1

Konfiguration

- Vorlage Opt-Out-Consent
- Vorlage Widerspruch

gICS®

2

Dokumentation und Datenpflege

- Registrierung von Teilnehmern (z.B. automatisiert im Rahmen der Patientenanlage)
- Dokumentation von Widersprüchen

Abhängig von der Konfiguration: **Benachrichtigung externer Systeme**

3

Nutzung

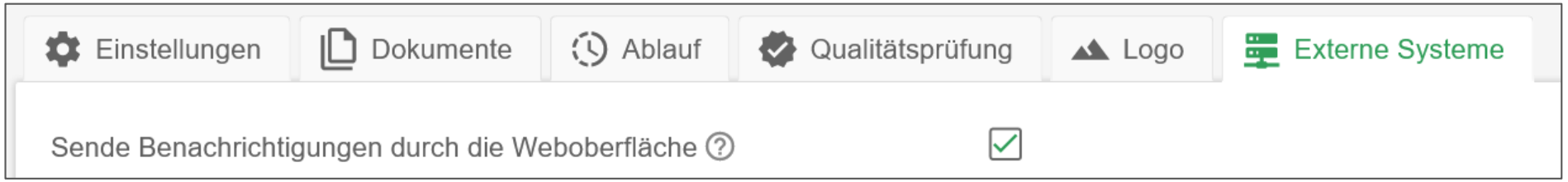
- Abfrage und Analyse von Einwilligungsständen

Unabhängige Treuhandstelle
UNIVERSITÄTSMEDIZIN GREIFSWALD

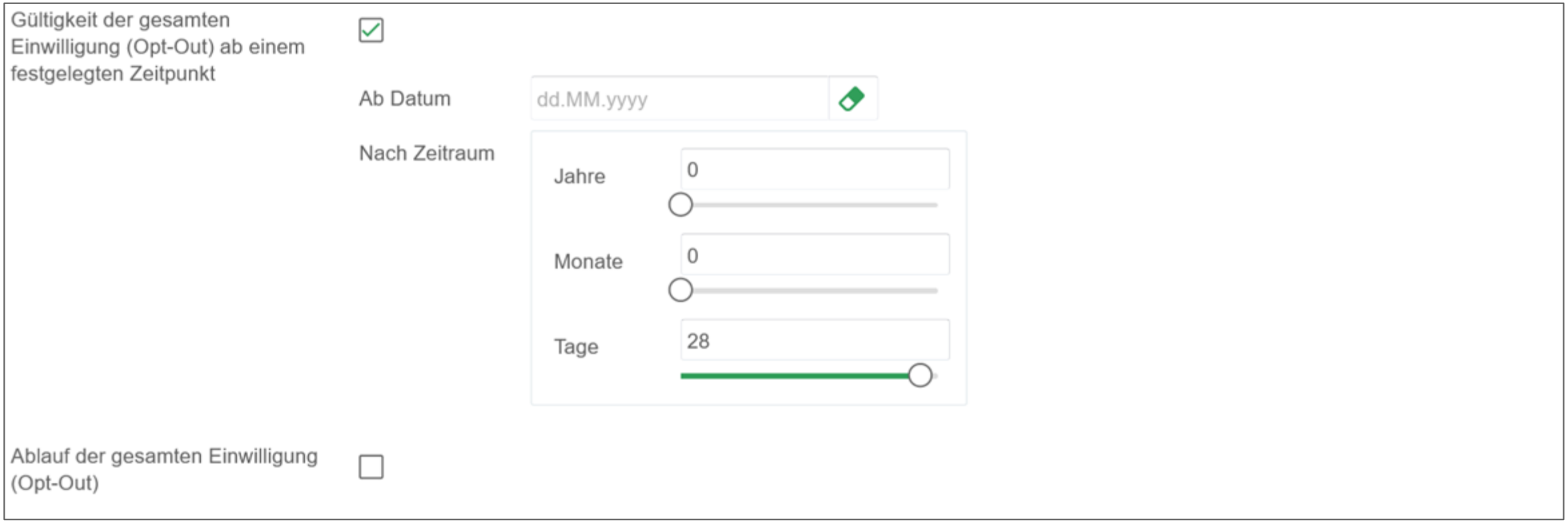
Einrichtung der Domain

- Auswahl Dokumenttypen „Einwilligung (Opt-Out)“ und „Widerspruch“

- QS-Vorgaben (z.B. per Default „geprüft & gültig“)
- ggf. Aktivierung Versand von Benachrichtigungen

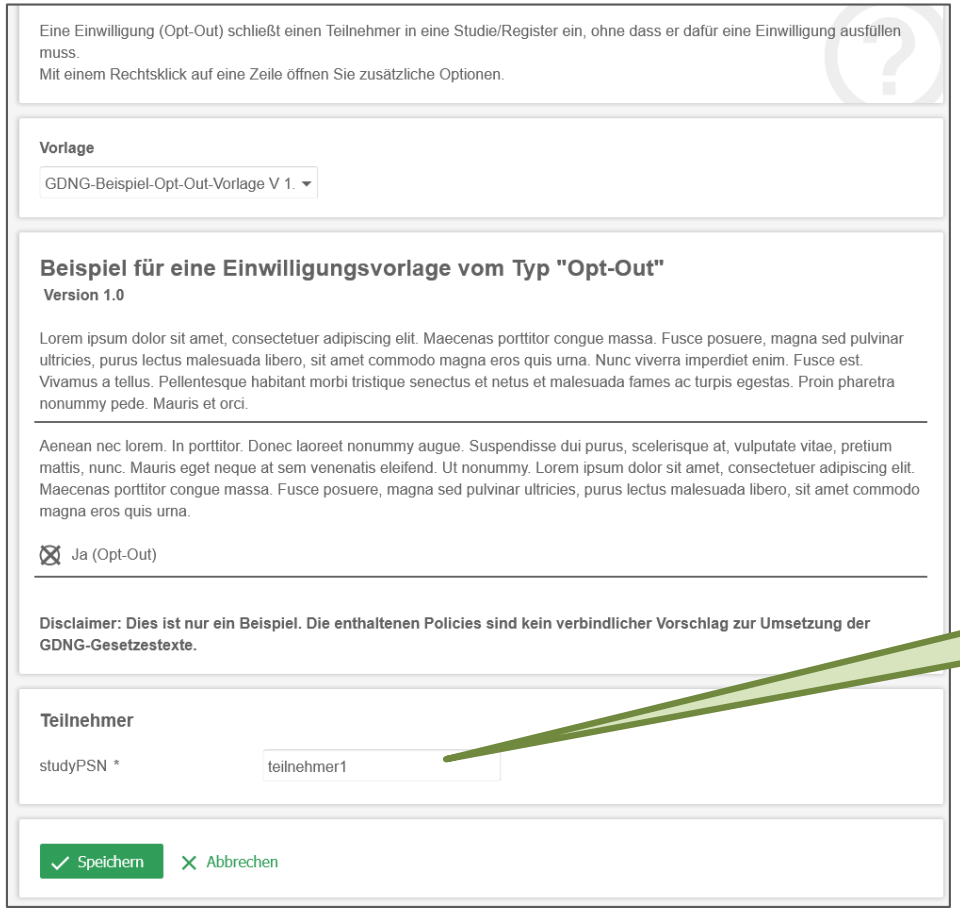


Einrichtung der Vorlagen



Erfassung von Einwilligung (Typ: Opt-Out) zur **Registrierung der Teilnehmer** über

Web-UI



Schnittstelle (SOAP)

POST {{baseUrlAndPort}}/gics/gicsService Send

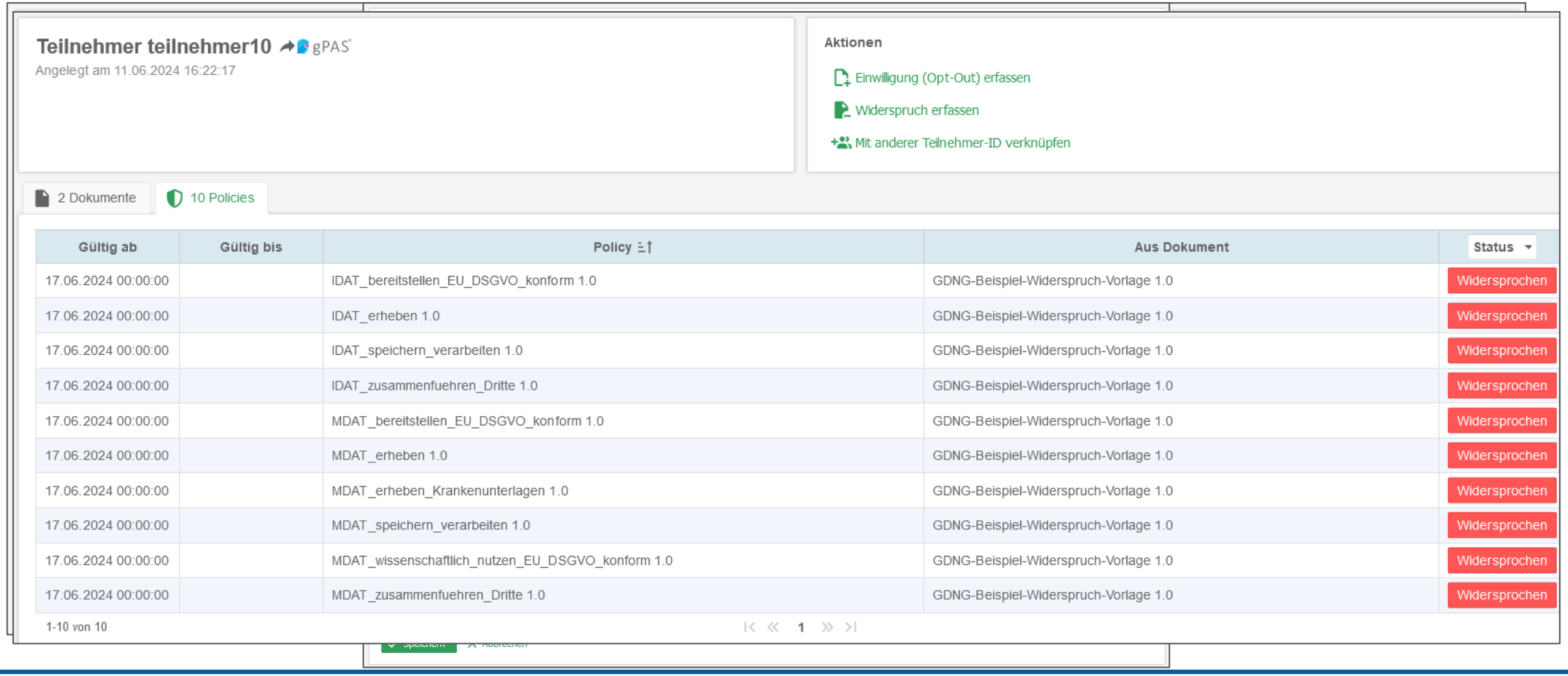
Params Auth Headers (9) Body Pre-req. Tests Settings Cookies

raw XML Beautify

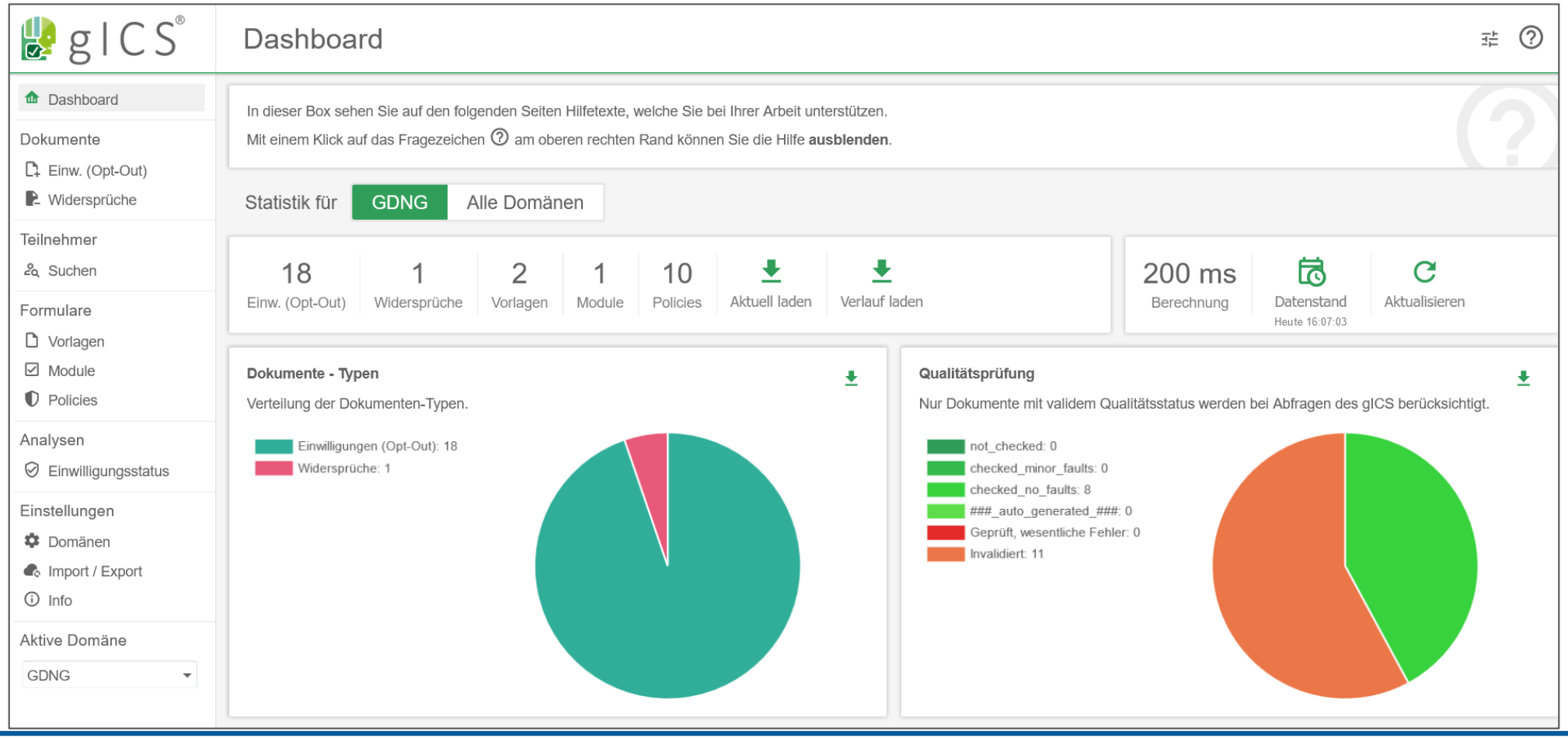
```xml
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/"
    xmlns:cm2="http://cm2.ttp.ganimed.icmvc.emau.org/">
    <soapenv:Header/>
    <soapenv:Body>
        <cm2:addConsentOptOut>
            <consentTemplateKey>
                <domainName>GDNG</domainName>
                <name>GDNG-Beispiel-Opt-Out-Vorlage</name>
                <version>1.0</version>
            </consentTemplateKey>
            <signerIds>
                <id>teilnehmer1</id>
                <idType>studyPSN</idType>
            </signerIds>
        </cm2:addConsentOptOut>
    </soapenv:Body>
</soapenv:Envelope>
```

Reduziert aufs Wesentliche: **Wer & Was**. Das Wann passiert automatisch. **Regelfall: Automatisierter System-Prozess**

Erfassung von Widersprüchen und Teilwidersprüchen analog zum bisherigen Vorgehen mit Widerrufen über die Oberfläche und die Schnittstelle (SOAP)



Die **Abfrage und Analyse von Einwilligungsständen** erfolgt **wie gewohnt** über vielfältige Funktionalitäten in Web-UI und Schnittstelle (SOAP)



Nächste Schritte

Ausbau der HL7 FHIR Unterstützung für Einwilligung (Typ: Opt-Out) und Widerspruch nach Vorgaben der **HL7 AG Einwilligungsmanagement**

- zur Registrierung von OptOut-Teilnehmern
- zur Ausleitung von entsprechenden Consent-Ressourcen

Szenario	Bezug	TemplateType	ResultType	Consent enthält aggregierte Informationen?	Kardinalität Provenance.entity	Level1 Provision.Type	Level2 Provision.Type(s)
Opt-Out	Einwilligung (Typ Opt-Out)	System: http://fhir.de/ConsentManagement/CodeSystem/TemplateType Code: CONSENT-OPT-OUT	document	Nein	0..1	PERMIT	PERMIT
	Widerspruch (vollständig)	System: http://fhir.de/ConsentManagement/CodeSystem/TemplateType Code: OBJECTION	document	Nein	0..1	DENY	PERMIT, DENY
	Widerspruch (teilweise)	System: http://fhir.de/ConsentManagement/CodeSystem/TemplateType Code: OBJECTION	document	Nein	0..1	DENY	PERMIT, DENY
	Einwilligungsstand (Typ Opt-Out)	n.a.	consent-status	Ja	0..*	DENY	PERMIT, DENY

HL7 Deutschland

TOP 3 - **VALIDATOREN**

Validatoren - Allgemein

Der E-PIX 2024.1 ermöglicht die Validierung von Feldern mittels Validatoren. Der E-PIX lehnt dabei ungültige Feldwerte ab. Die Weboberfläche ermöglicht die Korrektur von Eingaben, bevor diese in den E-PIX aufgenommen werden.

Bisher implizite Überprüfung:

- Pflichtfelder → leere Pflichtfelder werden abgelehnt
- Geburtsdatum (z.B. 32.04.1999 wird abgelehnt)
- Geschlechterangabe (abweichend von Männlich, Weiblich, Divers, Unbekannt und Sonstiges wird abgelehnt)

Zusätzlich ab E-PIX 2024.1:

- Felder können zusätzlich auf Inhalt validiert werden (Validierung findet pro Feld statt)
- Anfragen mit *nicht validen* Feldern werden abgelehnt. Eingaben können korrigiert werden.

Konfiguration von Validatoren

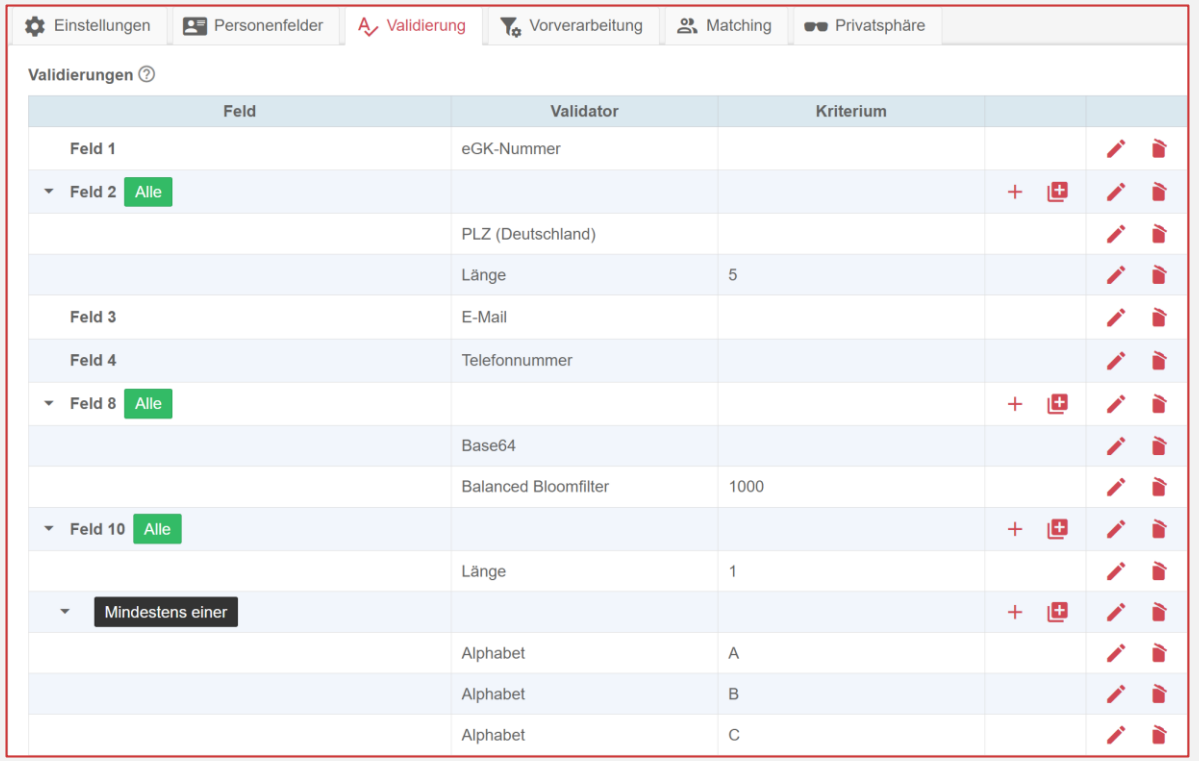
Unabhängige Treuhandstelle
UNIVERSITÄTSMEDIZIN GREIFSWALD

XML-Domänenkonfiguration

```
<validator-config>
    <field>value1</field>
    <validator>
        <qualified-class-name>org.emau.icmvc.ttp.deduplication.impl.validation.AlphabetValidator</qualified-class-name>
        <validation-criterion>ABCDEFGHIJKLMNOPQRSTUVWXYZ</validation-criterion>
    </validator>
</validator-config>
<validator-config>
    <field>value2</field>
    <validator>
        <qualified-class-name>org.emau.icmvc.ttp.deduplication.impl.validation.AlphabetValidator</qualified-class-name>
        <validation-criterion>A</validation-criterion>
    </validator>
</validator-config>
<validator-config>
    <field>value3</field>
    <validator>
        <qualified-class-name>org.emau.icmvc.ttp.deduplication.impl.validation.RegExValidator</qualified-class-name>
        <validation-criterion>B</validation-criterion>
    </validator>
</validator-config>
<validator-config>
    <field>value4</field>
    <validator>
        <qualified-class-name>org.emau.icmvc.ttp.deduplication.impl.validation.GermanZipCodeValidator</qualified-class-name>
    </validator>
</validator-config>
<validator-config>
    <field>value5</field>
    <validator>
        <qualified-class-name>org.emau.icmvc.ttp.deduplication.impl.validation.EGKValidator</qualified-class-name>
    </validator>
</validator-config>
<validator-config>
    <field>value6</field>
    <validator>
        <qualified-class-name>org.emau.icmvc.ttp.deduplication.impl.validation.PhoneNumberValidator</qualified-class-name>
        <validation-criterion>A</validation-criterion>
    </validator>
</validator-config>
<validator-config>
    <field>value7</field>
    <validator>
        <qualified-class-name>org.emau.icmvc.ttp.deduplication.impl.validation.EMailValidator</qualified-class-name>
    </validator>
</validator-config>
<validator-config>
    <field>value8</field>
    <validator>
        <qualified-class-name>org.emau.icmvc.ttp.deduplication.impl.validation.BalancedBloomFilterValidator</qualified-class-name>
        <validation-criterion>2000</validation-criterion>
    </validator>
</validator-config>
```

▶ **Handbuch Kapitel 6.9**

Domänenkonfiguration über Weboberfläche



▶ **Handbuch Kapitel 4.3.3**

Ausgelieferte Validatoren

Validator	Parameter	Beschreibung
Alphabet	Zeichen, die im Feld akzeptiert werden.	Prüft, ob ein Feld nur Zeichen des angegebenen Alphabets beinhaltet.
BalancedBloomFilter	*Keiner*	Prüft, ob ein Feld den Kriterien eines Balanced Bloomfilter entspricht. Dieser muss Base64 kodiert sein.
Base64	*Keiner*	Prüft, ob ein Feld in Base64 kodiert ist.
EGK	*Keiner*	Prüft, ob das Feld der zehnstelligen Krankenversichertennummer (KVNR) auf der Elektronische Gesundheitskarte (eGK) entspricht.
EMail	*Keiner*	Prüft, ob das Feld einer E-Mail Adresse entspricht.
EmptyField	true, wenn Leerzeichen ignoriert werden sollen, ansonsten false.	Prüft, ob ein Feld leer ist. Dabei können wahlweise ignoriert oder berücksichtigt werden.
GermanZipCode	*Keiner*	Prüft, ob das Feld eine deutsche Postleitzahl beinhaltet.
Length	Zahl, welche die erlaubte Länge des Feldes angibt.	Prüft, ob das Feld die angegebene Anzahl an Zeichen aufweist.
RegEx	Regex-Zeichenkette	Prüft, ob das Feld der Bedingung des angegebenen Regex entspricht.
PhoneNumber	*Keiner*	Prüft, ob das Feld eine Telefonnummer enthält. Dabei wird nur geprüft, ob die Zeichen im entsprechendem Alphabet beinhaltet sind. Es wird auf kein spezifisches Format geprüft.

Validator-Gruppen

- Sollen mehrere Eigenschaften überprüft werden, können mehrere Validatoren mittels Validator-Gruppe miteinander verknüpft werden.
- Komplexe Validierungen mit verschachtelten Validator-Gruppen möglich
- Verschiedene Verknüpfungen möglich

Validator-Gruppe			Gruppen-Verknüpfung			
Validator 1	Validator 2	Validator 3	Alle	Mindestens einer	Genau einer	Alle oder keiner
nicht valide	nicht valide	nicht valide	nicht valide	nicht valide	nicht valide	valide
valide	nicht valide	nicht valide	nicht valide	valide	valide	nicht valide
nicht valide	valide	nicht valide	nicht valide	valide	valide	nicht valide
valide	valide	nicht valide	nicht valide	valide	nicht valide	nicht valide
nicht valide	nicht valide	valide	nicht valide	valide	valide	nicht valide
valide	nicht valide	valide	nicht valide	valide	nicht valide	nicht valide
nicht valide	valide	valide	nicht valide	valide	nicht valide	nicht valide
valide	valide	valide	valide	valide	nicht valide	valide

Validatoren

KURZE LIVE-DEMO

TOP 5 - FRAGEN UND ANTWORTEN

Umgang mit der Fehlermeldung „Identität stimmt mit mehreren Personen überein“

Beim **Hinzufügen von Patienten** mit demselben Vornamen, Geburtsdatum und Geschlecht, aber mit einem leicht abweichenden Nachnamen, gibt E-PIX ein **Fehler-Pop-up** aus, **dass der Patient nicht hinzugefügt werden kann, da die Identität mit mehreren Personen übereinstimmt**. Was ist der Grund für die Anzeige dieses Fehler-Pop-ups? Wir gehen davon aus, dass E-PIX den Benutzer warnen will, dass diese Person bereits in E-PIX vorhanden ist, und die leicht abweichenden Daten zur gleichen Identität hinzufügt. **Ist das richtig, oder ist das wirklich ein Fehler**, den wir vermeiden sollten?

1. Der E-PIX ermittelt anhand von **Ähnlichkeiten**, ob die Person bereits bekannt ist oder erstmalig registriert wird.
2. Wird erkannt, dass der zu registrierende Datensatz hohe Ähnlichkeiten zu einem bestehenden Datensatz hat, wird in der Regel ein Possible Match ermittelt. Diese **mögliche Dublette** kann mittels Dublettenauflösung aufgelöst werden.
3. Ob und wann mögliche Dubletten erkannt werden sollen, kann mittels **Domänen-Konfiguration** beeinflusst werden.
4. Wenn der Datensatz mit mehreren Datensätzen mehrerer Personen entsteht ein **Multiple Match**. Da ein Datensatz (bzw. die Identität) nur einer Person zugeordnet werden kann, wird eine Fehlermeldung herausgegeben.
5. In einem solchen Fall **sollte geprüft werden, ob ein offensichtlicher Fehler korrigiert werden kann**, sodass eine Zuordnung zu einer Person erfolgen kann.

Nutzung von Adressinformationen zum Matching

Unabhängige Treuhandstelle
UNIVERSITÄTSMEDIZIN GREIFSWALD

Warum wird die **Adresse** nicht als Entität für den **Abgleichsprozess** verwendet, kann sie durch die Definition selbst definierter Werte hinzugefügt werden? Warum würden Sie uns empfehlen, die Adresse (nicht) als zusätzliche Entität hinzuzufügen?

Derzeit besitzt der E-PIX **feste Felder**, die für das Matching verwendet werden können. Dies soll in einer späteren Version flexibler werden. Jeder Person können mehrere Adressen zugeordnet werden (Hauptwohnsitz, etc.). Um nun Adressfelder auch in das Matching einzubeziehen, können die **Freitextfelder** verwendet werden. Dies wird von einigen Projekten auch genutzt (z.B. Postleitzahl des Hauptwohnsitzes). Dass Adressfelder standardmäßig nicht in dem Matching-Prozess verwendet werden können, ist daher keine Empfehlung, sondern der internen Datenverwaltung des E-PIX geschuldet, welche **künftig** aber **flexibler** gestaltet wird.

Weitere Informationen unter https://www.ths-greifswald.de/e-pix/handbuch Abschnitt 6.13

Wann wird die Verwendung von Bloomfiltern empfohlen?

In welchen Situationen oder Projekten würden Sie die Verwendung von Bloomfiltern empfehlen?

Das Record Linkage (bzw. **Privacy-Preserving Record Linkage) mittels Bloomfiltern dient für Standort-übergreifende Abgleiche.** Ein Standort hat meist die IDAT einer Person und kann ein lokales Record Linkage darauf basierend umsetzen. Bei übergreifenden Vorhaben **kann** die **Anforderung** bestehen, dass diese **IDAT nicht** regelhaft **an eine zentrale Stelle** geleitet werden. Hierbei kommen **stattdessen Bloomfilter** zum Einsatz, welche ohne Weiteres keine Rückschlüsse auf die IDAT geben. Der Bloomfilter wird dabei auf Basis der lokal vorliegenden IDAT erzeugt. Der zentralen bzw. föderierten Stelle werden dann nur die Bloomfilter mitgeteilt, welche dann übergreifende Abgleiche durchführen kann. Lokal sollten stets die IDAT abgeglichen werden, da diese auch eine Dublettenauflösung ermöglichen, was bei Bloomfiltern nicht ohne Weiteres erfolgen kann, da keine Rückschlüsse auf die IDAT möglich sind.

Weitere Informationen unter https://www.ths-greifswald.de/white-paper-record-linkage

fTTP Bloomfilter Konfig unter https://confluence.imi.med.fau.de/display/MIIC/Use+Case%3A+Patientenregistrierung

Parallele Matching-Anfragen

Können E-PIX und die Matching-Algorithmen Daten von mehreren Kliniken gleichzeitig verarbeiten?
Im E-PIX-Handbuch werden die Parameter `parallel-matching-after` und `number-of-threads-for-matching` erwähnt. Würden Sie uns empfehlen, die **Standardwerte** von **1000** und **4** zu verwenden?

Setting: Verschiedene Kliniken mit getrennter Verwaltung in 1 E-PIX
-> 1 E-PIX, 1 Domäne *je Klinik*, Matching innerhalb einer Klinik-Domain

Setting: Verschiedene Kliniken mit gemeinsamer Verwaltung/Zusammenführung in 1 E-PIX
-> 1 E-PIX, 1 Domäne *für alle Kliniken*, 1 Datenquelle je Klinik, Matching innerhalb der Gesamt-Domain

Parallelisierung: Ja, die Parameter `parallel-matching-after` und `number-of-threads-for-matching` können so übernommen werden bzw. werden so standardmäßig gesetzt

Weitere Informationen unter https://www.ths-greifswald.de/e-pix/handbuch Abschnitt 6.13

Verwendung der Standard-Konfiguration und Anpassungsempfehlungen

Unabhängige Treuhandstelle
UNIVERSITÄTSMEDIZIN GREIFSWALD

Warum sind die Standardparameter für das Matching so gewählt, was ist der Grund für diese Wahl? Unter welchen Umständen würden Sie vorschlagen, von den Standardeinstellungen abzuweichen? Wenn beispielsweise kein möglicher Abgleich erwünscht ist, wurde uns klar, dass der Schwellenwert für den automatischen Abgleich auf denselben Wert wie der Schwellenwert für den möglichen Abgleich gesetzt werden sollte (z. B. beide Schwellenwerte auf 2,99). Wie sieht es mit dem Schwellenwert und den Gewichtungen für die Abgleichsfelder aus? Warum wird z. B. der Vorname (8) höher gewichtet als der Nachname (6), warum wird das Geburtsdatum (9) höher gewichtet, usw.?

- Die Parameter sind auf Basis vorheriger und guter Ergebnisse gewählt worden.
- **Änderungen** erforderlich, wenn der eigene Bestand **Unterschiede in der Datenqualität** aufweist, sodass Felder z.B. anders gewichtet werden sollten.
- Um nur zwischen MATCH und NON-MATCH zu unterschieden (also keine möglichen Dubletten zu berücksichtigen), empfehlen wir den Schwellwert auf 1000 zu setzen. Der E-PIX führt dann nur noch PERFECT-MATCHES, also jene Datensätze zusammen, die identisch sind. *Dieses Verhalten kann gewünscht sein, wenn die Datenqualität sehr hoch ist, sodass unterschiedliche Datensätze stets unterschiedlichen Personen zugeordnet werden sollen.*
- Die **Wichtungen** der einzelnen Felder richten sich hauptsächlich nach der **Qualität und der Wahrscheinlichkeit von Änderungen**.
  Das **Geburtsdatum** wird z.B. höher gewichtet, da es (eigentlich) unveränderlich ist.
  Der **Nachname** kann sich z.B. durch Heirat ändern.
  Es kann daher sinnvoll sein, vor dem produktiven Betrieb die Einstellungen gemäß des eigenen Datensatzes anzupassen und zu prüfen, ob ggf. bessere Matching-Ergebnisse erzielt werden können.
- **Die von uns bereitgestellte Standardkonfiguration passt für viele Projekte** zunächst, muss aber z.B. dann angepasst werden, wenn andere projektspezifische Parameter beim Matching beachtet werden sollen (z.B. die Postleitzahl des Hauptwohnsitzes).

Weitere Informationen unter https://www.ths-greifswald.de/e-pix/handbuch Abschnitt 6.13

Zusammenspiel der Schwellwerte

Wie gestaltet sich das Zusammenspiel zwischen den Schwellenwerten, der möglichen Übereinstimmung und der automatischen Übereinstimmung sowie den Schwellenwerten für übereinstimmende Felder?

- Die **Ähnlichkeit** (bzw. Übereinstimmung) zwischen den Felder wird ermittelt (z.B. Vorname ANNA und ANNE).
- **Unterscheidung:** Die Felder werden dabei in **MATCH und NON-MATCH** unterschieden.
- Der Parameter „matching-threshold“ gibt dabei an, ab welcher Übereinstimmung die Felder als MATCH gewertet werden.
- **Berechnung:** Pro Datensatz werden dabei die matchenden und nicht-matchenden Felder gesammelt und jeweils mit der Wichtung multipliziert.
- **Entscheidung**: überwiegt das Produkt der matchenden Felder werden die Schwellwerte threshold-possible-match und threshold-automatic-match herangezogen und geprüft, ob diese überschritten werden und in Abhängigkeit dazu das Matching-Ergebnis ermittelt.

Bedeutung der Matching-Schwellwerte

Haben die Schwellenwerte für automatische (Standard: 14,5) und mögliche Übereinstimmungen (Standard: 2,99) eine interpretierbare Bedeutung? Wie werden diese Werte berechnet?

Diese Schwellwerte sind **abhängig vom verwendeten Algorithmus** (beim E-PIX Fellegi-Sunter). Der Algorithmus gibt entsprechend vor, wie diese Schwellwerte berechnet werden können.
Um diese Werte künftig besser interpretierbar zu machen, prüfen wir derzeit, ob und wie diese umgerechnet werden können, um so z.B. auf eine Skala von 0 bis 100 abgebildet werden zu können.

Weitere Informationen unter:
Fellegi IP, Sunter AB. A Theory for Record Linkage. J Am Stat Assoc. 1969 Dec 1;64(328):1183–210.
http://dx.doi.org/10.1080/01621459.1969.10501049

Verwendung von multiplen Vornamen (Multiple-Value-Felder)

Uns ist aufgefallen, dass es möglich ist, mehrere Vornamen einzubeziehen. Würden Sie empfehlen, diese Option zu nutzen? Gibt es irgendwelche Standardwerte, die Sie für dieses zusätzliche Abgleichsfeld empfehlen?

Ja, sofern nicht sichergestellt werden kann, dass z.B. nie mehr als ein Vorname angegeben wird, ist es zu empfehlen diese Option zu nutzen. Damit können **erheblich bessere Matching-Ergebnisse** erzielt werden. Die Option nennt sich **Multiple-Value-Felder**. Die empfohlenen Standardwerte hierfür sind in der **Standardkonfiguration** hinterlegt.

Dabei ist zu beachten, dass bis zu sechs Werte innerhalb eines Feldes enthalten sein können. Anhand des angegebenen Separators werden die Werte innerhalb des Feldes getrennt und einzeln miteinander verglichen. Dabei muss darauf geachtet werden, dass bei der Vorverarbeitung (pre-processing, konfiguriert innerhalb der Domänenkonfig), nicht das Separator-Zeichen entfernt wird (z.B. Leerzeichen). Bei der Berechnung in MATCH und NON-MATCH wird dann einbezogen, ob einzelne Teile miteinander matchen (z.B. „Klaus Dieter“, „Dieter Klaus“, „Dieter Klaus Peter“). Gute Schwellwerte für verschiedene Betrachtungen sind in der Standkonfiguration enthalten und können übernommen werden.

Weitere Informationen unter https://www.ths-greifswald.de/e-pix/handbuch , Abschnitt 6.13.7

Berechnungsdetails bei der Datensatzverknüpfung

Unabhängige Treuhandstelle
UNIVERSITÄTSMEDIZIN GREIFSWALD

Als Standardalgorithmen für die Datensatzverknüpfung werden Levenshtein und Fellegi-Sunter genannt. Ist es möglich, die genauen Berechnungen, die von E-PIX durchgeführt werden, einzusehen?

Der Algorithmus für das Probabilistische Matching ist **Fellegi-Sunter.** Für die einzelnen Feldabgleiche wird meist die Levenshtein-Distanz berechnet. Da hierbei eine **Vielzahl an Berechnungen** durchgeführt wird, sind die einzelnen **Zwischenergebnisse** der Felder **nicht einsehbar**. Zwischen den Datensätzen ist meist nur das höchst erzielte Ergebnis (also die höchste ermittelte Übereinstimmung) relevant. Dieser Wert wird abgespeichert und kann entweder über das **Protokoll** oder die **Datenbank** (identity_history-Tabelle) eingesehen werden. Im Fall eines Matches, ist dann auch die entsprechend matchende Identität verlinkt.

Weitere Informationen unter:
Fellegi IP, Sunter AB. A Theory for Record Linkage. J Am Stat Assoc. 1969 Dec 1;64(328):1183–210.
http://dx.doi.org/10.1080/01621459.1969.10501049

Die Community stellt sich vor

Referent: Nico Schmid
Robert Bosch Krankenhaus

Robert Bosch Krankenhaus

Integration von E-PIX und der Statistiklösung R

E-PIX®

Unser TEAM sagt Danke. Fragen?

Unabhängige Treuhandstelle
UNIVERSITÄTSMEDIZIN GREIFSWALD

Kontakt

Dr. Martin Bialke
Universitätsmedizin Greifswald
Institute für Community Medicine
Ellernholzstr. 1-2
17475 Greifswald

+49 (3834) 86-7580
martin.bialke@uni-greifswald.de
ths-greifswald.de