



DFG-Projekt MOSAIC

Toolbox for Research

Martin Bialke

Institut für Community Medicine

Abt. Versorgungsepidemiologie und Community Health

Universitätsmedizin Greifswald

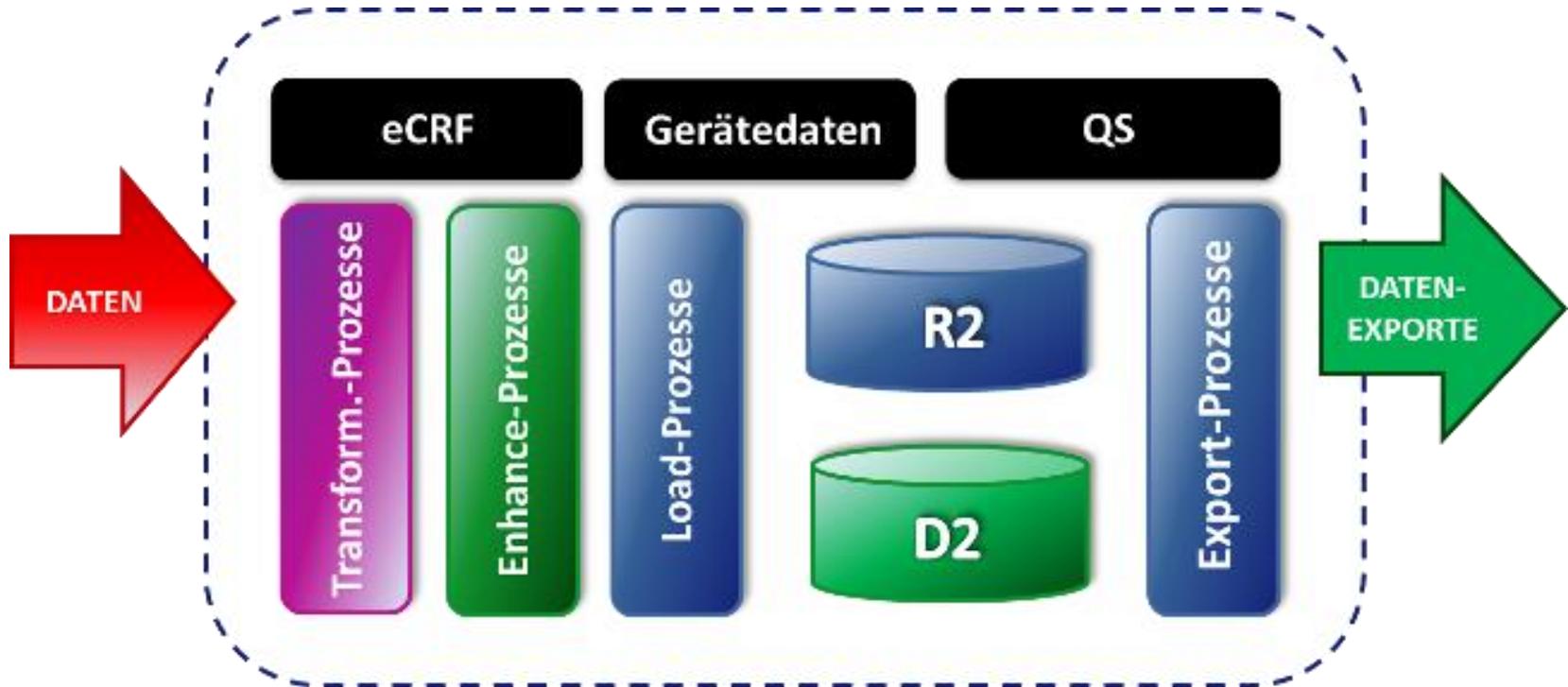
Stand: 05. August 2016



- Hintergrund:
 - vor allem kleinere Forschungsprojekte mit geringen IT-Ressourcen (Expertise, Personal, Infrastrukturen) waren bisher oftmals auf Excel und Co. angewiesen
 - manuelle Pseudonymisierung der erhobenen medizinische Daten erforderlich
 - Jährliche händische Zusammenführung der Daten zur Auswertung
- Ziel:
 - eine einfache Möglichkeit für web-basierten und standortübergreifende Datenerfassung bieten

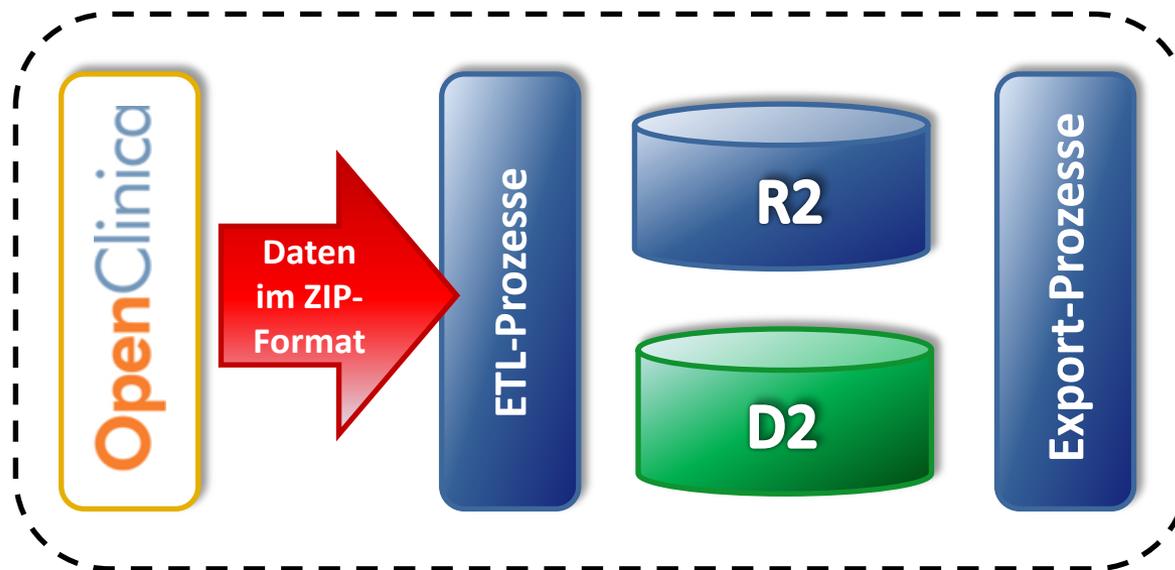


Toolbox for Research





- Generierung von web-basierten Fragebögen und Metadatenbeschreibungen
- Verwaltung von Standorten und Nutzern
- Vergabe von Rollen und Rechten



- automatisierter Export von Meta- und Forschungsdaten im ZIP-Format
- innerhalb der Research Box werden nur „neue“ Daten verarbeitet

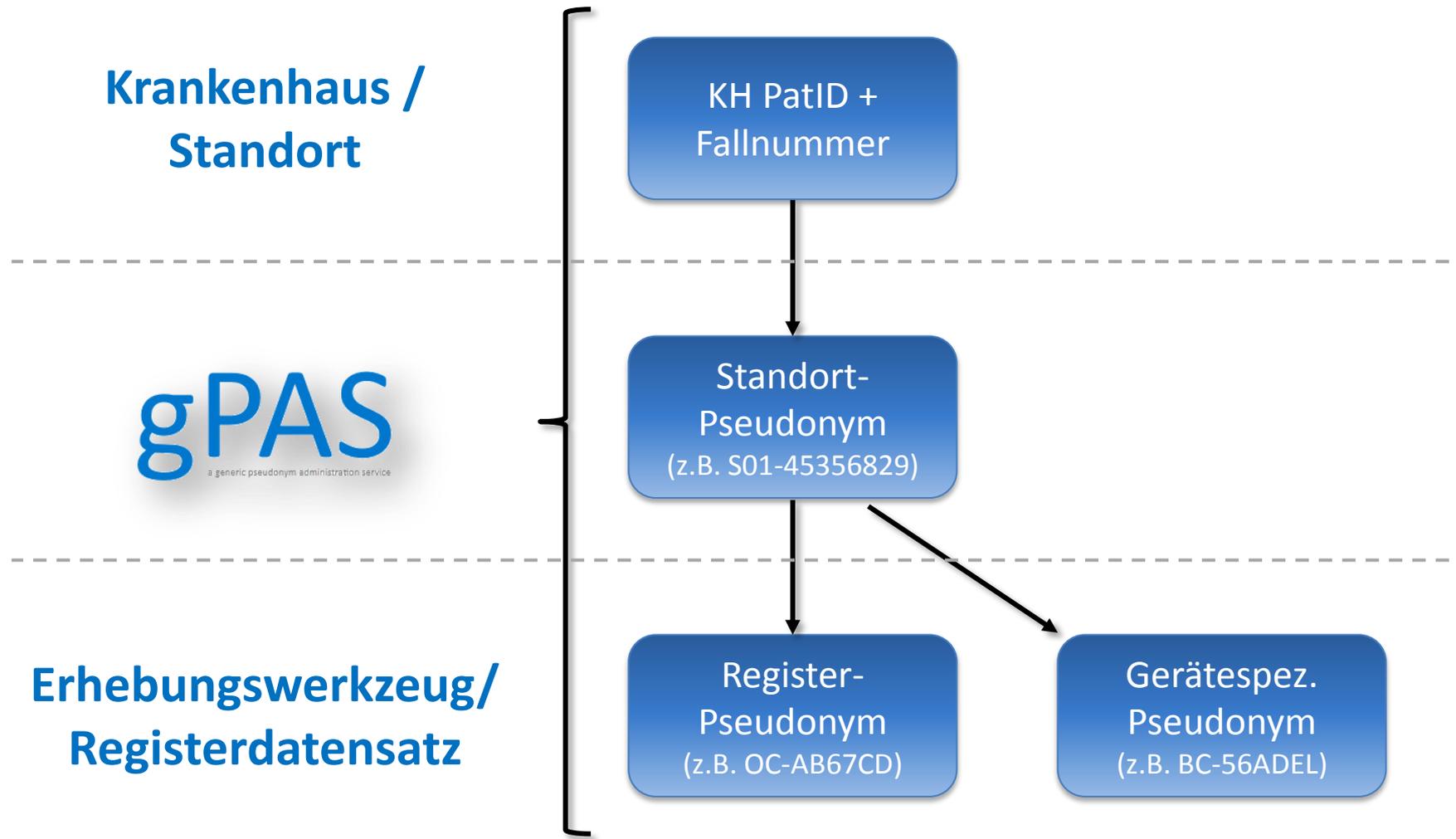
Technische Umsetzung - Pseudonymisierung mit gPAS



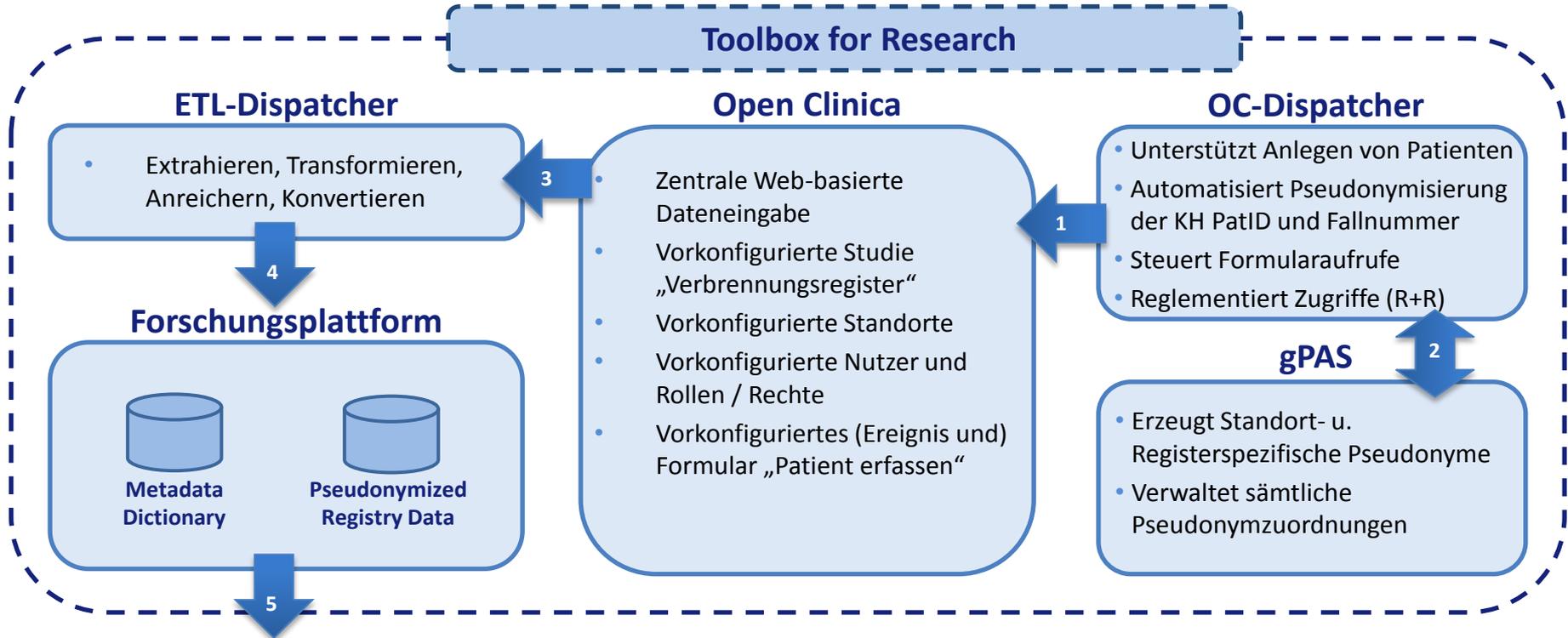
- Modularer Webservice
- Verarbeitung beliebiger Zeichenfolgen für Fallnummern, Kennungen, etc.
- Mehrfache Pseudonymisierung durch Pseudonym-Domänen möglich
- Depseudonymisierung
- Anonymisierung durch Löschung von Zuordnungen
- Flexible Konfiguration: Alphabete, Länge, Prüfzifferalgorithmus
- Validierung von Pseudonymen
- Stapelverarbeitung (CSV)
- Integration von Altbeständen

<https://mosaic-greifswald.de/werkzeuge-und-vorlagen/pseudonymverwaltung-gpas.html>

August 2016



Technische Umsetzung - Workflow



1 OC-Dispatcher vereinfacht und automatisiert das Anlegen von Patienten in Open Clinica mit Geburtsdatum, Geschlecht und KH Pat-ID+Fallnummer

2 Erzeugung einer einheitlichen Register-ID (Sekundärpseudonym)

August 2016



3 Pseudonymisierte Forschungsdaten werden transformiert und für Speicherung in Forschungsplattform vorbereitet (OC, BurnCase3D)

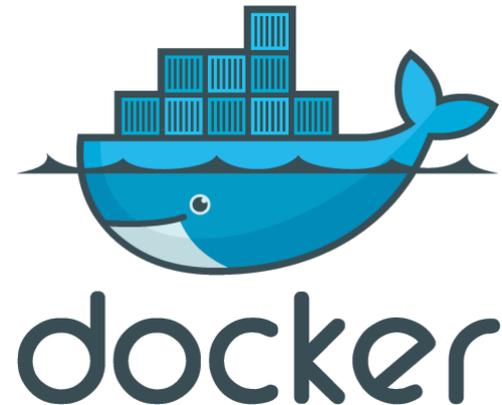
4 Zentrale Speicherung der Registerdaten

5 Forschungsplattform stellt Daten aus allen angeschlossenen Quellen für Exporte und Analysen zur Verfügung, z.B. im SPSS-Format

Installation mittels Docker



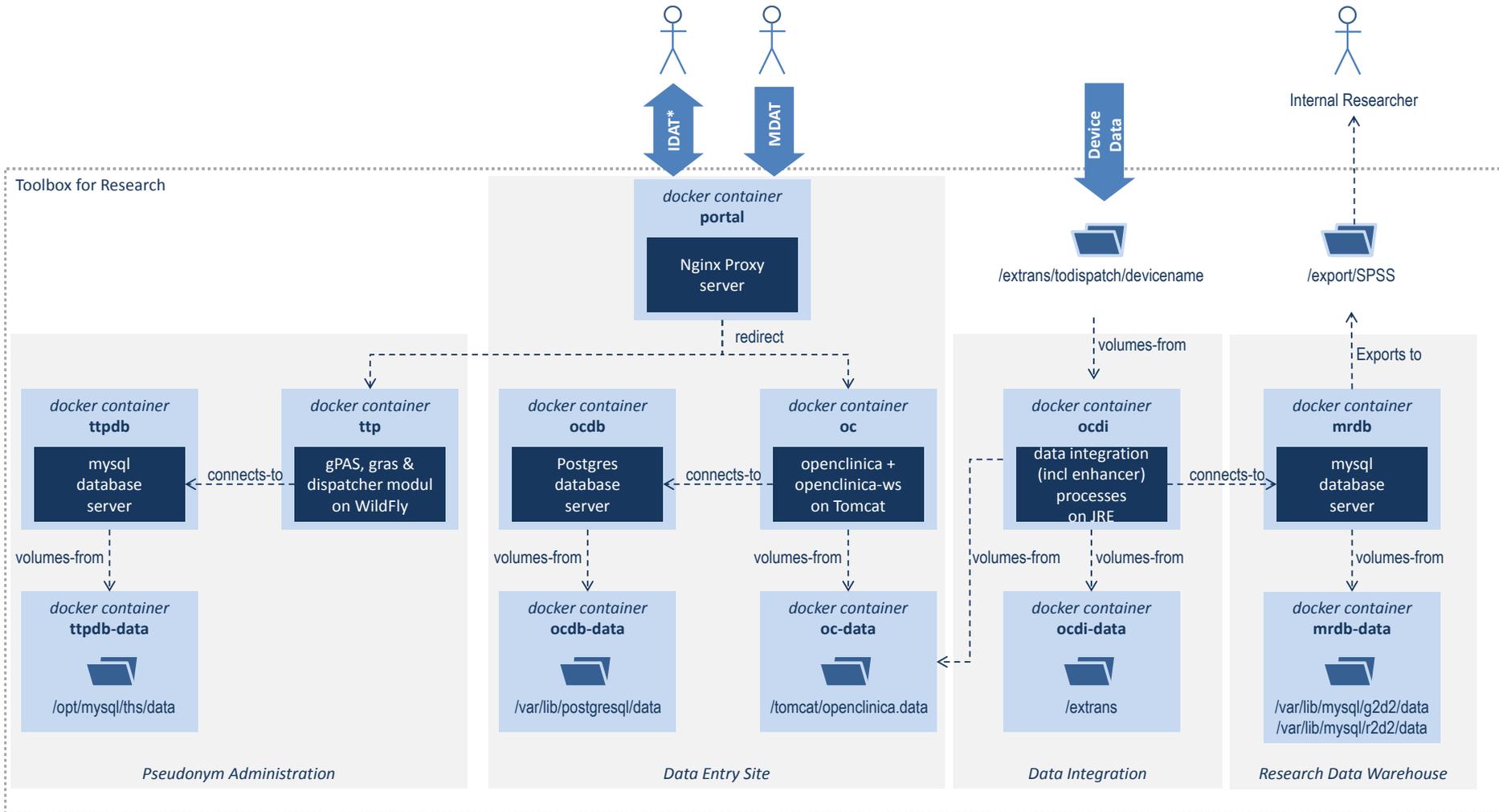
- automatisierte Installation von Containern per Skript
- Jeder Container enthält die gewünschte Software und notwendige Abhängigkeiten
 - Ressourcensparend
 - Kurze Startzeiten
 - Erleichterte Wartung, Update, Recovery
 - Es muss „nur ein Skript“ ausgeführt werden
- Free Open Source Software
- Verfügbar für Windows, Mac und Linux



Literatur : *heise.de*, *Anwendungen mit Docker transportabel machen*
<http://www.heise.de/developer/artikel/Anwendungen-mit-Docker-transportabel-machen-2127220.html?artikelseite=1>,
Zugriff am 29. Mai 2015

August 2016

Docker-Architektur



Was leistet die Toolbox?



- zentrale Vergabe standort- und studienspezifischer Pseudonyme dank [gPAS](#)-Integration
- Suche von Pseudonymen/Depseudonymisierungsfunktion für autorisiertes Personal
- vereinfachtes Anlegen und registrieren von Studienteilnehmern/Patienten dank Dispatcher-Integration
- Formulargenerierung mittels Leitfäden des MOSAIC Projektes und OpenClinica-Mechanismen
- Web-basierte und standort-übergreifende Datenerfassung durch OpenClinica-Integration
- Getrennte Speicherung von Metadaten und Forschungsdaten
- Möglichkeit zur Gerätedatenintegration (z.B. im CSV-Format)
- Datenexport im SPSS-Format



Was leistet die Toolbox NICHT?

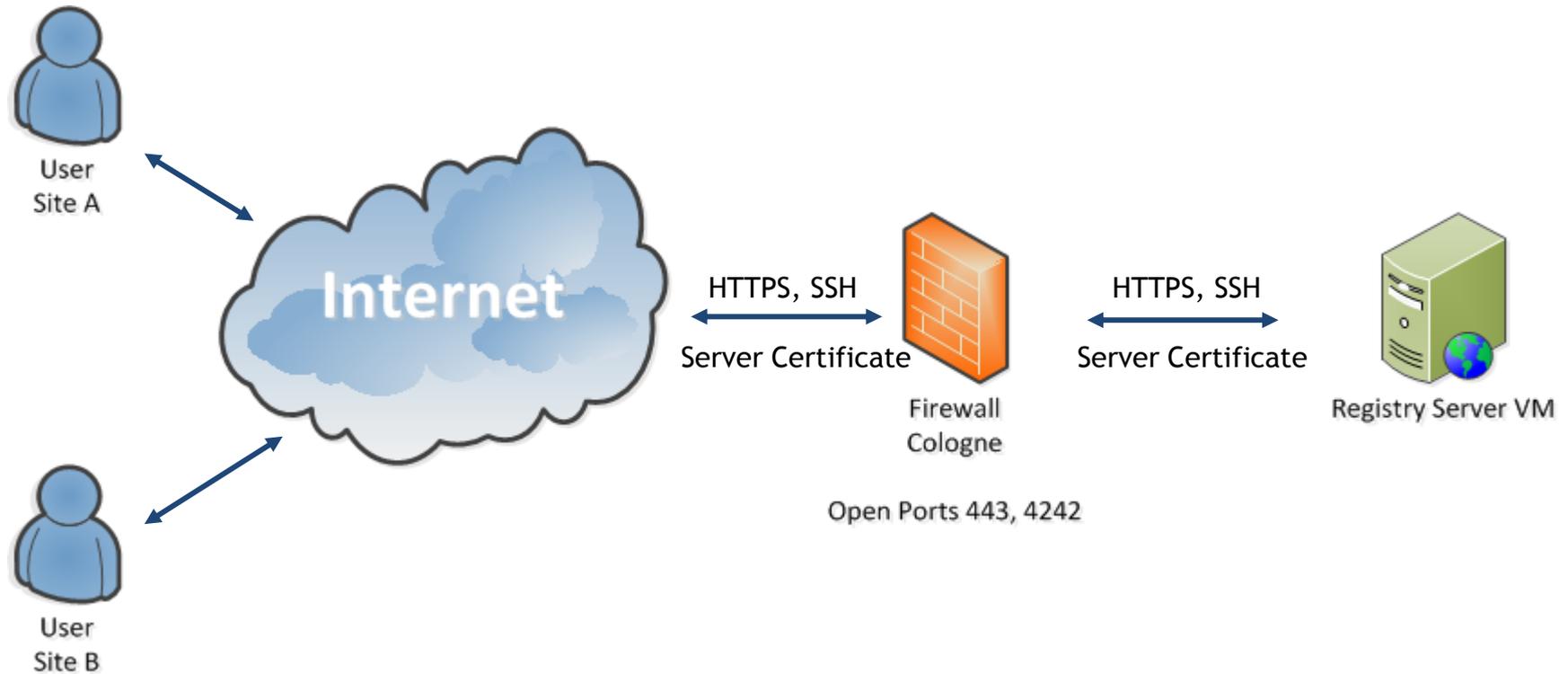
- Keine Definition des Data Dictionary
- Keine Definition der Fragebögen
- Keine Dublettenerkennung, denn die Toolbox speichert KEINE personenidentifizierenden Daten (IDAT) außer Geschlecht und Geburtsdatum (Beides ist zum Anlegen von Teilnehmern in OpenClinica erforderlich.)
- Keine Treuhandstellen-Funktion
- Keine automatische Datenauswertung
- Keine Kontrolle der wissenschaftlichen Qualität des Data Dictionary, der Fragebögen, der Daten und Datenauswertung
- Keine Erstellung des notwendigen Datenschutzkonzeptes
- Limitierungen bei der Formularerstellung durch OC in Bezug auf Sprünge, die Validierung von Werten und Eingaben, Variablenübergreifende Plausibilitätsprüfungen, u.a.

August 2016



- **Personell**
 - Definition und Aktualisierung von Data Dictionary und Fragebogen
 - Eigenverantwortlicher Betrieb der Toolbox, d.h. Systembereitstellung, Installation, Wartung, Monitoring, Pflege, IT-Sicherheit, Datensicherheit, Benutzerverwaltung, u.a.
 - Weitere Infos siehe Check-Liste für Administratoren
- **Technisch**
 - 64-Bit Hostsystem Debian8 (Linuxkernel 3.10 oder höher)
 - Mindestens 6GB Arbeitsspeicher
 - 50GByte freier Festplattenspeicher (ohne Hostsystem)
 - Docker
 - Internetzugang
 - Offizielles und gültiges Server-Zertifikat für gesicherte Datenkommunikation
 - Domainregistrierung

Integrationsempfehlung



→ Bereitstellung einer Checkliste für die Vor-Ort-Integration

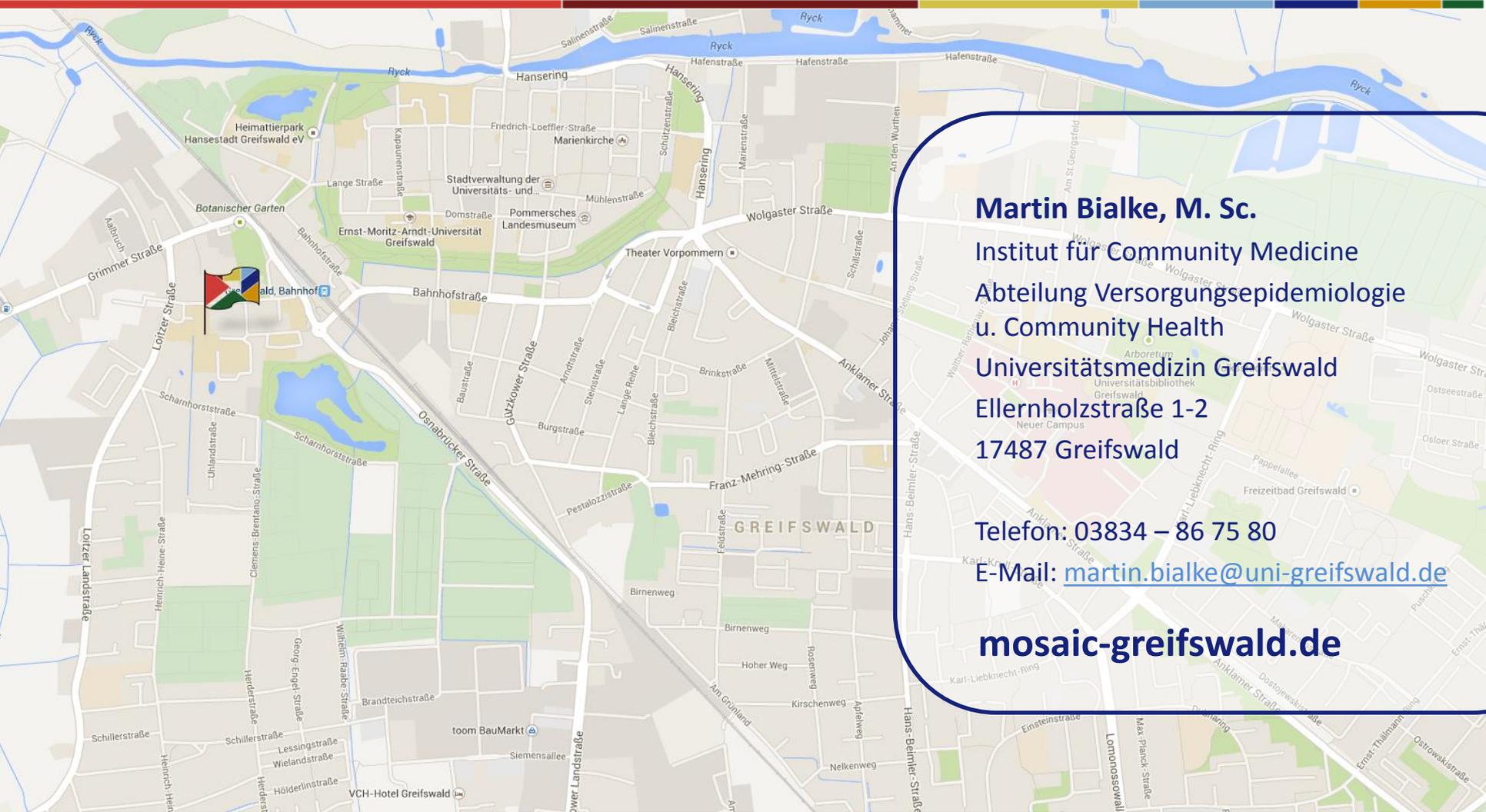
Bereitstellung seit 7. März 2016



-  [1 Automatische Installation.pdf](#)
Installationsanleitung (automatisch), v.1.0
-  [2 Einrichtung.pdf](#)
Einrichtung der Toolbox (v.1.0)
-  [3 Betrieb.pdf](#)
Betrieb der Toolbox (v.1.0)
-  [4 Checkliste fuer Administratoren.pdf](#)
Checkliste für Administratoren (v.1.0)
-  [5 Anwenderhandbuch.pdf](#)
Anwenderhandbuch (v.1.0)
-  [6 Geraetedatenintegration.pdf](#)
Anleitung zur Integration von Gerätedaten (in Arbeit)
-  [7 Einspielen von BurnCaseDaten.pdf](#)
Anleitung zum Einspielen von Gerätedaten (am Beispiel von Burncase 3D) (v.1.0)
-  [burncase3d.zip](#)
Metadatenbeschreibung für Gerätedaten (am Beispiel von Burncase 3D, inkl. Datenbeispiel)
-  [toolbox-docker.zip](#)
Toolbox-Installationsdateien (v1.0.1, ZIP-Datei, Größe 113 MB)

August 2016

14



Martin Bialke, M. Sc.

Institut für Community Medicine
Abteilung Versorgungsepidemiologie
u. Community Health

Universitätsmedizin Greifswald
Ellerholzstraße 1-2
17487 Greifswald

Telefon: 03834 – 86 75 80

E-Mail: martin.bialke@uni-greifswald.de

mosaic-greifswald.de