

Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

**Sachbericht zur BMBF-Förderrichtlinie**

„Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle:

**Josua4DTM**

**Projekt: Josua für sichere FAIR-Datenenklaiven**

An die  
VDI/VDE Innovation + Technik GmbH  
Steinplatz 1  
10623 Berlin

Beauftragt durch das  
Bundesministerium für Bildung und Forschung (BMBF)  
Kapelle-Ufer 1  
10117 Berlin

**Anbieter**

Forschungsinstitut zur Zukunft der Arbeit GmbH (IZA)  
Schaumburg-Lippe-Straße 7-9  
53113 Bonn

**Ansprechperson**

Dr. Nikos Askitas  
Forschungsinstitut zur Zukunft der Arbeit GmbH (IZA)  
Schaumburg-Lippe-Straße 7-9  
53113 Bonn

Telefon: 0049 228 3894-525

Fax: 0049 228 3894-510

E-Mail: [askitas@iza.org](mailto:askitas@iza.org)

Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

## **Teil I: Kurzbericht zum Projekt "Josua für sichere FAIR-Datenenklaiven (Josua4DTM)"**

### **Ursprüngliche Aufgabenstellung sowie wissenschaftlicher und technischer Stand**

Das Ziel des Projekts war die Integration und Erweiterung der bestehenden Josua-Module für Remote Access und Remote Execution, die bereits in mehreren beim RatSWD akkreditierten Forschungsdatenzentren (FDZ) im Einsatz sind. Diese IT-Lösungen, Josua Web und Josua Direct, ermöglichen Forschenden den sicheren und datenschutzkonformen Fernzugriff auf sensible Forschungsdaten.

Josua Web automatisiert die Fernverarbeitung von Analyseskripten, während Josua Direct eine geschützte virtuelle Desktop-Umgebung bietet.

Die Trennung dieser Systeme führte bislang zu einem erhöhten Verwaltungsaufwand sowie redundanter Nutzer- und Projektadministration. Der aktuelle Stand der Technik und die Bedürfnisse der Datentreuhand erforderten daher eine vereinheitlichte Lösung, die Datensicherheit, Effizienz und Nutzerfreundlichkeit verbessert.

### **Ablauf des Vorhabens**

Das Projekt umfasste vier zentrale Arbeitspakete (AP):

#### 1. Integration von Josua Web und Josua Direct:

- Implementierung der Multidomainfähigkeit zur gemeinsamen Nutzung mehrerer Josua-Instanzen.
- Entwicklung eines zentralen Benutzeradministrationssystems mit API-basierter Nutzerverwaltung zur Reduzierung manueller Eingriffe.

#### 2. Export-Schnittstellen:

- Einrichtung einer Funktion zur automatisierten Erstellung und Verwaltung periodischer Nutzungsstatistiken in tabellenkompatiblen Formaten.
- Erweiterung der Operator-Oberfläche zur Konfiguration und Verwaltung dieser Exporte.

#### 3. Migration und Dokumentation

- Einrichtung einer spezialisierten Dataverse-Instanz für das IDSC, die die förderierte Nutzung von Daten ermöglicht.
- Vollständige Migration bestehender Datensätze und Metadaten auf die neue Plattform.

#### 4. Josua-Dokumentation:

- Erstellung einer umfassenden Systemdokumentation in Markdown zur Verwaltung und Nutzung aller Josua-Komponenten.
- Bereitstellung der Dokumentation im Git-Repository zur kontinuierlichen Pflege und Erweiterung.

### **Wesentliche Ergebnisse**

Das Projekt führte zu mehreren wichtigen Projektergebnissen

#### 1. Technische Verbesserungen:

- Die Implementierung der Multidomainfähigkeit erfolgte durch umfassende Modernisierungen, darunter die Migration des Systems auf die Plattform .NET 8 sowie die Integration des Kestrel-Webservers.
- Ein neuer Webservice zur API-basierten Nutzerverwaltung wurde auf Basis von ASP.NET Core entwickelt.
- Eine Benutzeroberfläche (Single-Page-Application) zur Nutzerverwaltung wurde mit Vue.js erstellt und ermöglicht eine effiziente Verwaltung von Benutzerkonten und Active Directory-Integrationen.

## Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

### 2. Neue Funktionen:

- Multidomainfähigkeit ermöglicht die gleichzeitige Nutzung mehrerer Josua-Instanzen auf leistungsstarken Servern.
- Implementierung einer API-basierten Nutzerverwaltung für Active Directory und Josua-Projekte.
- Eine Benutzeroberfläche (Single-Page-Application) zur Nutzerverwaltung wurde mit Vue.js erstellt und ermöglicht eine effiziente Verwaltung von Benutzerkonten und Active Directory-Integrationen.
- Eine Funktion zur periodischen Erstellung und Verwaltung von Nutzungsstatistiken wurde implementiert.
- Eine neue Vue-Komponente bietet eine verbesserte Visualisierung der Exportübersicht, inklusive responsiver Darstellung für unterschiedliche Bildschirmgrößen.

### 3. Erweiterte Infrastruktur:

- Die Migration auf Dataverse ermöglichte eine verbesserte Datenstruktur, erhöhte Interoperabilität und langfristige Archivierung nach FAIR-Prinzipien.
- Fünf neue Dataverse-Sammlungen für Forschungsdaten wurden eingerichtet, darunter das zentrale G<sup>2</sup>LM|LIC-Dataverse, das Wageindicator-Dataverse und das ECONtribute-Dataverse.
- Dataverse und die Eigenentwicklung einer Python-Exportschnittstelle dienen als Lösung zur Vernetzung und föderierten Nutzung getrennter Datensilos.

### Zusammenarbeit mit anderen Forschungseinrichtungen

Im Rahmen des Projekts fand eine enge Zusammenarbeit mit mehreren Partnern statt. Dazu zählen:

- **ECONtribute Exzellenzcluster** (Universität Bonn und Universität zu Köln): Aufbau eines gemeinsamen Datenrepositoriums für interdisziplinäre Forschungsdaten.
- **G<sup>2</sup>LM|LIC (Global Labour Markets)**: Bereitstellung umfangreicher internationaler Arbeitsmarktdaten in Kooperation mit dem britischen Foreign, Commonwealth & Development Office.
- **WageIndicator Foundation**: Entwicklung und Betreuung eines Dataverse-Repositories zur Bereitstellung von Lohn- und Arbeitsmarktdaten für die wissenschaftliche Forschung.
- Generierung eines föderierten Dataverse-Katalogs für Arbeitsmarktdaten

Diese Kooperationen tragen zur nachhaltigen Nutzung der Projektergebnisse und zur weiteren Verbreitung der Josua-Infrastruktur bei.

Mit diesen Entwicklungen wurde die Grundlage für eine sichere, effiziente und nachhaltige Nutzung sensibler Forschungsdaten gelegt, die sowohl für Forschungseinrichtungen als auch für die Datentreuhändergemeinschaft von hoher Relevanz ist.

Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

## **Teil II: Eingehende Darstellung zum Projekt "Josua für sichere FAIR-Datenenklaiven (Josua4DTM)"**

### **Arbeitspaket 1: Integration Josua Web und Josua Direct**

#### **1.1 Implementierung der Multidomainfähigkeit**

Die Implementierung der Multidomainfähigkeit des Systems wurde durch umfangreiche Anpassungen und Modernisierungen realisiert. Ein zentraler Bestandteil dieser Arbeit war die Überarbeitung des Node-Services. Dabei wurde die Entwicklungsumgebung auf Visual Studio 2022 aktualisiert und das Framework von .NET Framework 4.8 auf die moderne Plattform .NET 8 migriert. Zusätzlich wurde der Kestrel-Webserver integriert, um eine leistungsfähige und zuverlässige Webserver-Infrastruktur bereitzustellen, die alle Funktionen über HTTP bereitstellt. Die internen Methoden des Node-Services wurden überarbeitet und optimiert, während die interne Datenhaltung durch Anpassungen der Identifier und die Einführung einer Zuordnung der Jobs nach Endpunkten verbessert wurde. Darüber hinaus wurden die Routen zur Bereitstellung der Funktionalitäten über HTTP implementiert. Der Installationsassistent, der bisher auf der InstallShield Limited Edition basierte und für Visual Studio 2022 nicht mehr verfügbar ist, wurde durch Inno Setup ersetzt. Außerdem wurde die Benutzeroberfläche des Konfigurationstools des Node-Services entsprechend angepasst.

#### **Anpassungen am Operator Interface**

Auch am Operator-Interface wurden umfangreiche Änderungen vorgenommen. Die Node-Konfiguration wurde überarbeitet, um den neuen Anforderungen der Multidomainfähigkeit gerecht zu werden. Dazu wurden neue Methoden entwickelt sowie Templates angepasst bzw. neu erstellt. Das Datenbankmodell wurde aktualisiert und notwendige Migrationen erstellt, um die neuen Datenstrukturen zu unterstützen. Parallel dazu wurden die Model-Klassen aktualisiert, neue Methoden zur Kommunikation mit dem aktualisierten Node-Service erstellt und die Methoden zur Jobverteilung angepasst.

Zusätzlich wurden für den Node-Service neue API-Endpunkte erstellt und die Authentifizierung der Nodes angepasst. Da die Kommunikation zwischen dem Operator-Interface und dem Node-Service nun über HTTP stateless erfolgt und einige Aufgaben in bestimmten Intervallen durchgeführt werden müssen, wurde ein Scheduler integriert und in das Container-Build aufgenommen. Neue Konsolenkommandos wurden ebenfalls entwickelt, um die aktualisierte Funktionsweise zu unterstützen. Durch diese Maßnahmen wurde die Multidomainfähigkeit erfolgreich implementiert, wodurch das System nicht nur leistungsfähiger und flexibler, sondern auch sicherer und zukunftsfähig gestaltet wurde.

#### **1.2: Implementierung der zentralen Josua Benutzeradministration**

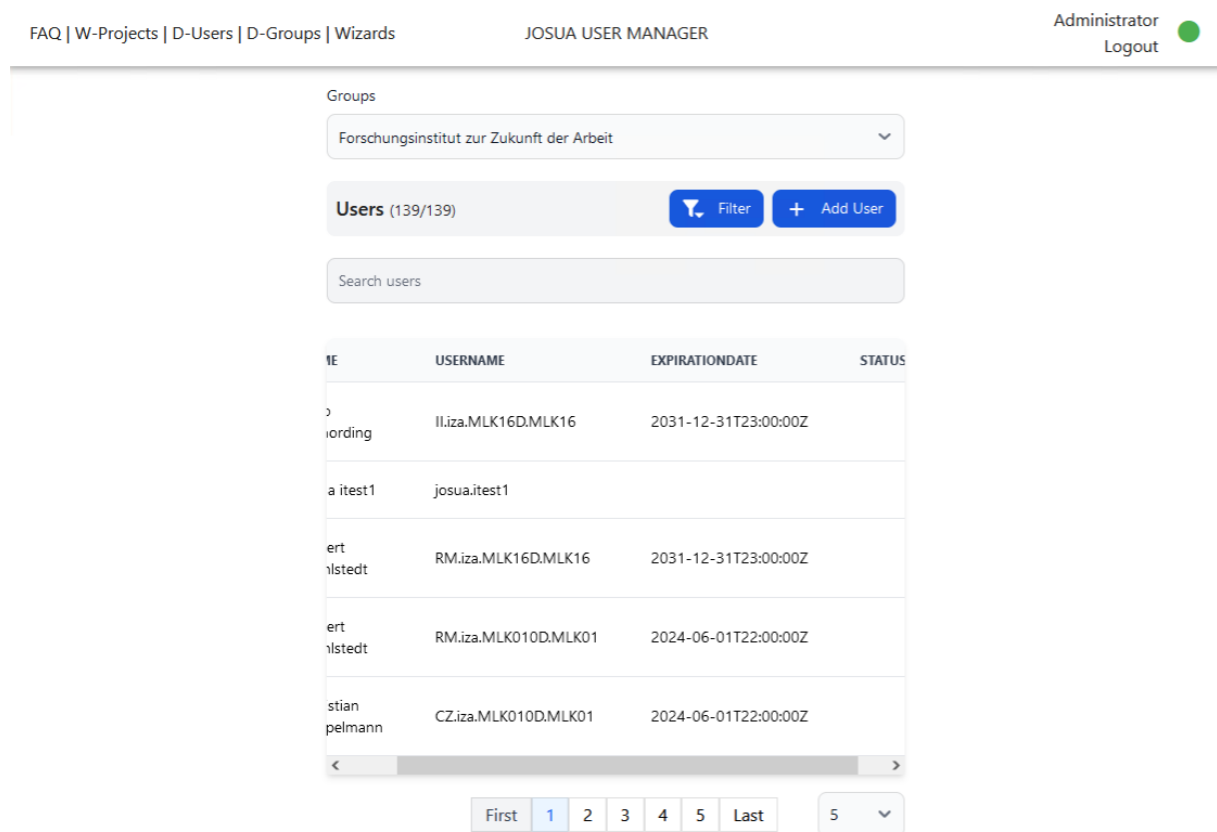
Die Optimierung der Benutzeradministration wurde in drei Schritten erreicht: (1) Entwicklung eines neuen Webservices zur Bereitstellung einer API für die Nutzerverwaltung, (2) Erstellung einer Benutzeroberfläche (UI) zur Nutzerverwaltung und (3) Erweiterung des Operator-Interfaces um eine API für die Nutzerverwaltung.

#### **Entwicklung des Webservices für API-basierte Nutzerverwaltung:**

Der neue Webservice wurde in C# auf Basis von ASP.NET Core entwickelt. Er kombiniert die Funktionen des Active Directory mit den bereits vorhandenen Nutzerverwaltungsfunktionen des Operator-Interfaces und stellt diese als Methoden über eine API bereit. Diese Methoden ermöglichen die Erstellung und Verwaltung von Active Directory-Benutzern und -Gruppen sowie deren Verknüpfung mit den entsprechenden Josua Web-Nutzern und -Projekten. AD-spezifische Konfigurationen können über eine dateibasierte Konfiguration gesetzt und mit den Josua-Operator-Endpunkten verknüpft werden.

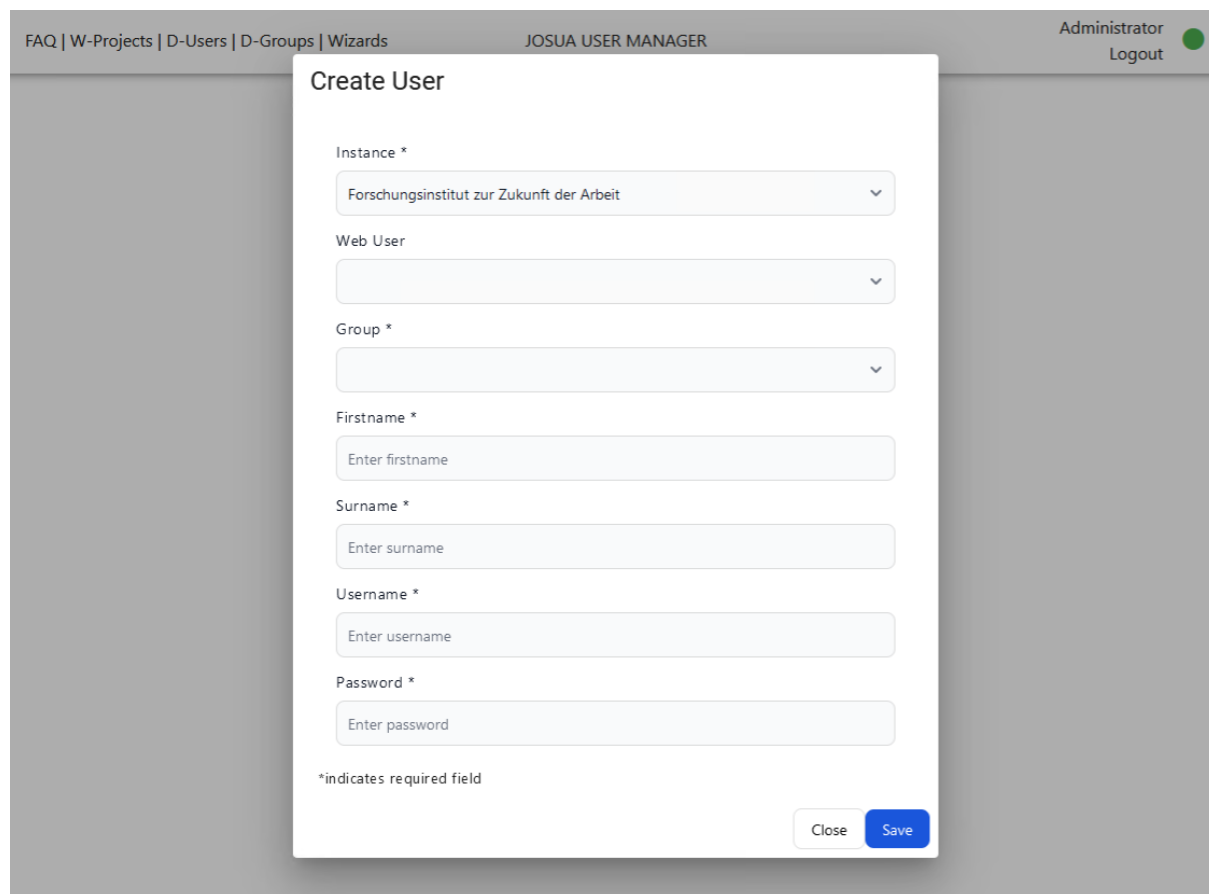
## Erstellung der Benutzeroberfläche zur Nutzerverwaltung

Eine Single-Page-Application (SPA) zur Nutzerverwaltung wurde mit dem JavaScript-Framework Vue entwickelt, wobei Flowbite als Komponentenbibliothek ausgewählt wurde. In den Abbildungen 1 und 2 werden exemplarisch einige Teile der Benutzeroberfläche dargestellt.



### Abb. 1: Benutzerliste

In der Benutzerliste kann die Gruppenzugehörigkeit der Nutzer über ein Dropdown-Menü ausgewählt werden. Innerhalb der ausgewählten Gruppe ermöglichen vordefinierte Filter sowie ein Textsuchfeld eine gezielte Suche. Die Filter lassen sich per Mausklick aktivieren. Im unteren Bereich der Benutzerliste stehen Funktionen zur Seitennavigation zur Verfügung. Hier können die Nutzer zwischen den Ergebnisseiten blättern und die maximale Anzahl der angezeigten Suchergebnisse pro Seite anpassen.



**Abb. 2: Dialog zur Erstellung eines neuen Josua Direct Benutzers.**

Im Dialog zur Benutzererstellung wird zunächst die Gruppe bzw. Instanz festgelegt, der der neue Nutzer angehören soll. Neben der Eingabe von Name, Benutzername und Passwort bietet ein Dropdown-Menü die Möglichkeit, den Nutzer mit einem bestehenden Josua Web-Account zu verknüpfen. Zusätzlich kann der Nutzer direkt einer bereits bestehenden Active Directory-Projektgruppe hinzugefügt werden.

## **Arbeitspaket 2 – Nutzungsstatistiken**

Um periodische Exporte von Nutzungsdaten in einem für Statistikprogramme kompatiblen Format zu ermöglichen, wurden zwei zentrale Maßnahmen umgesetzt: (1) die Erstellung einer Konfiguration für periodische Exporte und (2) die Anpassung der Übersicht und Darstellung von Exporten.

### **Erstellung der Konfiguration für periodische Exporte**

Zur Erleichterung der Verwaltung periodischer Exporte wurde eine Übersicht über alle periodischen Exporte (Abb. 3) entwickelt und ein intuitives Menü erstellt, das das Hinzufügen und Bearbeiten von Exporten ermöglicht (Abb. 4). Diese Vorgehensweise stellt sicher, dass Administratoren Exportaufgaben effizient planen und ändern können, wodurch die Flexibilität des Systems erhöht wird.

Dafür wurden mehrere technische Anpassungen vorgenommen, darunter:

- **Anpassung des Datenbankmodells:** Erweiterungen zur Speicherung periodischer Exportkonfigurationen.
- **Erstellung von Migrationen, Modellen und Templates:** Sicherstellung einer konsistenten Implementierung.

## Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

- **Anlegen von Routen und Anpassung der Controller:** Ermöglicht die Steuerung und Verarbeitung der Exportvorgänge.
- **Integration neuer Schedules im Kernel:** Automatisierung der periodischen Exporte zu festgelegten Intervallen.

Diese Schritte waren notwendig, um eine skalierbare und benutzerfreundliche Konfiguration periodischer Exporte zu gewährleisten.

Die Übersicht und Darstellung der Exporte wurde durch die Entwicklung einer neuen Vue-Komponente erheblich verbessert. Diese bietet eine bessere Visualisierung der Exporte, einschließlich der generierten Dateien, und ist für unterschiedliche Bildschirmgrößen responsiv gestaltet (Abb. 5). Diese Neuerungen tragen dazu bei, die Benutzerfreundlichkeit und Zugänglichkeit des Systems zu steigern.

Um diese Verbesserungen zu unterstützen, wurden folgende technische Änderungen vorgenommen:

- **Umstellung von Laravel Mix auf Vite:** Dies ermöglicht eine schnellere und effizientere Verarbeitung von Frontend-Ressourcen.
- **Upgrade von Node 12 auf Node 18:** Modernisierung der Entwicklungsumgebung, um die Kompatibilität mit aktuellen Technologien und Abhängigkeiten sicherzustellen.

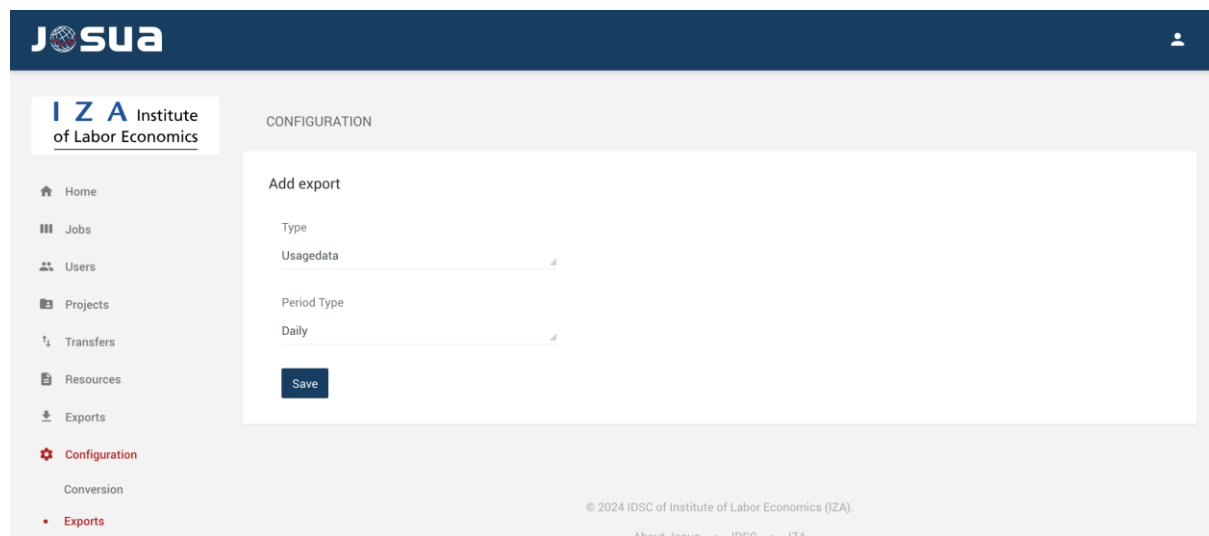
Diese Anpassungen waren notwendig, um die Performance und Wartbarkeit des Systems langfristig zu sichern. Die beschriebenen Maßnahmen basieren auf der Zielsetzung, eine nutzerfreundliche und zukunftssichere Plattform für die Verwaltung von Nutzungsdaten zu schaffen. Die Konfiguration periodischer Exporte ermöglicht eine automatisierte Bereitstellung von Daten für statistische Analysen, was die Effizienz und Genauigkeit der Datenauswertung erhöht.

#	Type	Period	Last schedule at	Last run at/status	Created by	Created at	
1	Usagedata	daily	2024-10-08 08:00:01	2024-10-08 08:00:02 (ok)	Schöpf, David	2024-09-09 08:26:05	Edit

Abb. 3: Übersicht der periodischen Exporte.

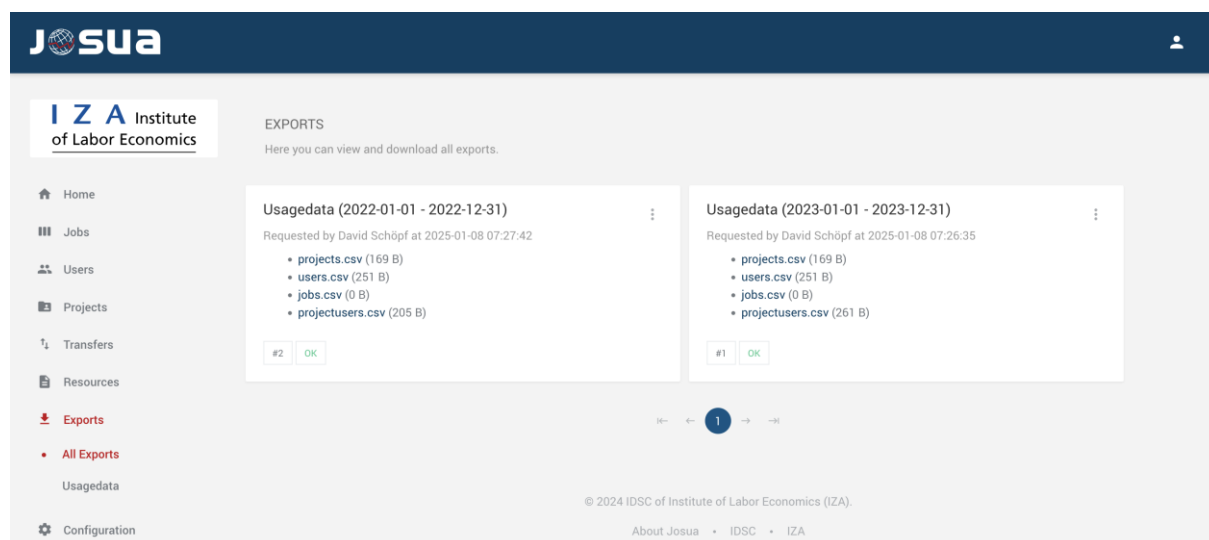
Die tabellarische Darstellung zeigt eine Übersicht aller periodischen Exporte. Dabei werden folgende Informationen aufgeführt: Export-ID, Export-Typ, Intervall, der zuletzt geplante Ausführungszeitpunkt, der Zeitpunkt der letzten erfolgreichen Ausführung sowie der aktuelle Ausführungsstatus. Zusätzlich werden der Name und der Zeitpunkt der Konfiguration des Exports angezeigt. Eine Schaltfläche zur Änderung der Exporteinstellungen ermöglicht eine direkte Anpassung der Konfiguration.

# Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle



**Abb. 4: Periodische Exporte: Menü zur Erstellung und Anpassung**

Das Menü ermöglicht die Konfiguration periodischer Exporte. Über Dropdown-Menüs können sowohl der Export-Typ als auch das gewünschte Intervall für die automatische Erstellung ausgewählt werden. Diese Optionen bieten eine flexible Anpassung an die spezifischen Anforderungen der Nutzer.



**Abb. 5: Darstellung der Exporte: Übersichtskacheln mit Detailinformationen.**

Jeder Export wird in einer eigenen Kachel dargestellt. Im Header sind der Export-Typ, der abgedeckte Zeitraum, der Name des Erstellers sowie das Erstellungsdatum angegeben. Im mittleren Bereich werden alle erfassten Dateien aufgelistet, einschließlich ihrer jeweiligen Dateigröße. Der untere Bereich der Kachel zeigt Tags für die Export-ID sowie den aktuellen Status des Exports an. Diese strukturierte Darstellung erleichtert die schnelle Orientierung und Verwaltung der Exporte.

## Arbeitspaket 3: Vollständige Migration

Im Rahmen von Milestone 3 wurde eine spezialisierte Dataverse-Instanz für das IDSC eingerichtet. Die Installation erfolgte auf einer virtuellen Maschine (Ubuntu 22.04) unter Verwendung von DockerCE zur Containerisierung. Nach Abschluss der technischen Implementierung wurde das Metadatenmodell der bestehenden Datenbank erfolgreich in die Dataverse-Struktur migriert, wodurch die Interoperabilität und Nutzerfreundlichkeit der Forschungsdatenplattform signifikant verbessert werden konnten.

Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

## **Dataverse-Installation und -Konfiguration**

Die Installation und Konfiguration von Dataverse bildeten die technische Grundlage für die Umsetzung dieses Meilensteins. Sie legten die Basis für eine effiziente Bereitstellung, Verwaltung und nachhaltige Nutzung der Forschungsdaten. Dabei kamen aktuelle Technologien und gezielte Anpassungen zum Einsatz, um eine leistungsfähige, flexible und zukunftsfähige Infrastruktur zu gewährleisten. Im Folgenden werden die wesentlichen Schritte der technischen Umsetzung detailliert beschrieben.

### **Technische Implementierung der Dataverse-Instanz**

Im Rahmen der technischen Umsetzung wurde eine spezialisierte Dataverse-Instanz für das IDSC eingerichtet. Die Installation erfolgte auf einer virtuellen Maschine mit Ubuntu 22.04, wobei DockerCE zur Containerisierung eingesetzt wurde. Zur flexiblen Verwaltung des Festplattenspeichers wurde ein Logical Volume Manager (LVM) implementiert, der eine Erweiterung der Speicherressourcen während des laufenden Betriebs ermöglicht. Die zentrale **/data-Partition** dient als Speicherort für alle wesentlichen Systembestandteile, einschließlich Docker- und Dataverse-Daten, Logs und Skripte. Die Ressourcenausstattung der virtuellen Maschine mit 4 CPU-Kernen und 12 GB RAM wurde bewusst gewählt, um den RAM-Anforderungen der Java-basierten Dataverse-Plattform gerecht zu werden und eine skalierbare Leistungsfähigkeit zu gewährleisten. Darüber hinaus ist die VM in das systematische Rechenzentrum-Backup integriert, wodurch eine Wiederherstellung im Bedarfsfall sichergestellt ist.

### **Systembereitstellung und Sicherheitsarchitektur**

Die Wahl der Docker-Version erfolgte unter Berücksichtigung der Kompatibilität mit dem Basis-VM-Template des Rechenzentrums. Zusätzlich wurde die Bereitstellung des Servers über einen HA-Proxy-Server realisiert, der neben der Lastverteilung auch für die SSL-Verschlüsselung und somit für die sichere Datenübertragung zuständig ist. Nach dem Klonen des Dataverse Docker-Projekts aus dem offiziellen Git-Repository wurden in der **.env-Datei** systemspezifische Anpassungen vorgenommen, darunter die Konfiguration des Domainnamens, die Einrichtung des E-Mail-Servers und die Anpassung von Standardpasswörtern zur Erhöhung der Systemsicherheit. Die Instanz wurde mittels Docker-Compose gestartet und umfasst vier Container: den Hauptcontainer für Dataverse, einen PostgreSQL-Datenbankcontainer, einen Solr-Cachecontainer sowie einen Minio-Container für Object Storage.

#### **3.1 Systemkonfiguration und Anpassung der Benutzeroberfläche**

Um eine effiziente Systemverwaltung und gezielte Anpassungen der Plattform zu ermöglichen, wurde ein Bash-Skript entwickelt, das sowohl für die interne Konfiguration der Docker-Container als auch für die Steuerung des Payara Java Backends über asadmin-CLI genutzt wird. Darüber hinaus wurden Layout und Funktionalität der Dataverse-Website über dieses Skript modifiziert, etwa durch individuelle Anpassungen an Header, Footer, die About-Seite sowie Lizenzbedingungen.

#### **3.2 Integration in Suchmaschinen und Sichtbarkeitsoptimierung**

Ein zusätzlich eingerichteter Cronjob generiert über Nacht eine Sitemap, die automatisch in die Google Search Console überführt wird, um eine verbesserte Indexierung durch Suchmaschinen zu ermöglichen. Dies trägt dazu bei, die Sichtbarkeit der veröffentlichten Dataverse-Sammlungen und Datensätze des IDSC im Internet zu erhöhen und den Zugang zu den Forschungsdaten zu erleichtern. Weitere Integrationen umfassen die Anbindung von Google Analytics zur Analyse der Plattformnutzung sowie spezifische Konfigurationen des E-Mail-Servers.

## **4. Optimierung der Suchfunktion und Datenverwaltung**

Das Skript automatisiert zudem die Neuindexierung der Solr-Cachedatenbank und stellt sicher, dass Log-Dateien an den richtigen Speicherorten abgelegt werden. Eine regelmäßige Neuindexierung des

Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

Solr-Caches war erforderlich, da während der Erprobung festgestellt wurde, dass bestimmte Objekte, insbesondere Datensätze in Dataverse, nicht korrekt angezeigt wurden. Nach einer manuellen Neuindexierung waren die fehlenden Objekte wieder sichtbar, was auf einen potenziellen Fehler in der verwendeten Dataverse-Version hinweist. Um diesem Problem proaktiv entgegenzuwirken und die Konsistenz der Suchergebnisse sicherzustellen, wurde daher eine regelmäßige Neuindexierung des Solr-Caches implementiert. Darüber hinaus wurden Anpassungen vorgenommen, um die Föderation mehrerer Dataverse-Instanzen zu ermöglichen. Ein speziell entwickeltes Kommando namens „dv“ erlaubt eine direkte Fehlerbehebung, indem es automatisch die Container-ID ermittelt und eine Bash-Shell innerhalb des Dataverse-Containers öffnet. Zusätzlich wird dieses Kommando als Cronjob verwendet, um nach einem Systemneustart API-Aufrufe zu initiieren und eine regelmäßige Neuindexierung des Solr-Caches durchzuführen, was sich während der Testphase als notwendig erwiesen hat.

## 5. DOI-Problematik und erprobter Workaround

Im Zuge der technischen Evaluierung wurde festgestellt, dass die von Dataverse unterstützten DOI-Provider, **DataCite und EZID**, nicht mit dem am IDSC des IZA genutzten DOI-Anbieter **CrossRef** kompatibel sind. Darüber hinaus erfolgt die DOI-Vergabe am IDSC nicht zufällig, sondern folgt einer **internen Systematik**, die spezifische **Regelungen und Strukturen** beinhaltet. Diese Systematik berücksichtigt unter anderem die **Versionsnummern von Publikationen und Datensätzen**, wodurch eine eindeutige und nachvollziehbare Identifikation gewährleistet wird.

Da Dataverse standardmäßig eine **automatische, zufällige DOI-Vergabe** durchführt, wurde ein alternativer Workflow entwickelt. Dabei wird zunächst ein neues Dataset mit vollständigen Metadaten erstellt, woraufhin Dataverse einen DOI generiert. Vor dem Upload der Daten und der endgültigen Veröffentlichung im System wird dieser DOI in der **PostgreSQL-Datenbank** modifiziert. Die Anpassung erfolgt über eine **SSH-Verbindung mit Tunnel auf den PostgreSQL-Port 5433 des Ubuntu-Systems** und wird über **pgAdmin** durchgeführt.

## Alternative DOI-Workflows und Optimierung des Zugriffs

Grundsätzlich sind alternative DOI-Workflows denkbar, die eine **manuelle Änderung der Datenbank überflüssig machen**. Eine Möglichkeit wäre die Implementierung eines **sicheren, dedizierten Interfaces zur Verwaltung von DOI-Daten**, das die spezifischen Anforderungen des IDSC direkt in die Vergabestruktur integriert. Allerdings wurde der **Entwicklungs- und Bereitstellungsaufwand** für eine solche Lösung im Verhältnis zur derzeit erforderlichen Arbeitszeit für manuelle Eingriffe als **zu hoch für das aktuelle Vorhaben** eingeschätzt. Eine zukünftige Entwicklung in diese Richtung bleibt jedoch eine Option.

Um den bestehenden Prozess dennoch zu optimieren, wurde ein **webbasiertes pgAdmin-Interface** (eine Datenbank-Administrationssoftware für PostgreSQL) auf dem Server installiert. Dieses ermöglicht einen **direkten, aber kontrollierten Zugriff auf die Dataverse-Datenbank**, wobei **Sicherheitsmaßnahmen wie eine firewallbasierte Zugriffsbeschränkung** implementiert wurden. Damit wird eine **gezielte Verwaltung der DOIs** im Einklang mit der hausinternen Systematik des IDSC erleichtert, ohne ein vollständig neues Interface entwickeln zu müssen.

## 6. Föderation und Skalierung der Dataverse-Instanz

Neben der produktiven Dataverse-Instanz wurden mehrere zusätzliche Instanzen eingerichtet. Zwei dieser Instanzen wurden gezielt für die Erprobung und Demonstration der Föderation zwischen verschiedenen Dataverse-Umgebungen genutzt. Der Lösungsansatz zur erfolgreichen Durchführung dieser Föderation zwischen zwei Dataverse-Instanzen wird im Folgenden detailliert dargestellt.

Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

### Datenmigration zwischen verschiedenen Dataverse-Instanzen

Harvesting ist ein Prozess des Austauschs von Metadaten mit anderen Repositorien. Als sog. Harvesting Client kann dabei eine Dataverse-Installation Metadatenätze aus entfernten Quellen sammeln oder wiederum als Harvesting Server anderen Harvesting Clients zugänglich machen. Solche Clients bzw. Server können andere Dataverse-Installationen oder andere Archive sein, die das OAI-PMH-Protokoll als Standard-Harvesting-Protokoll unterstützen. Die gesammelten Metadatenätze werden indiziert und durchsuchbar gemacht. Durch Klicken auf einen geharvesteten Datensatz in den Suchergebnissen gelangen Benutzer zum ursprünglichen Repository. Geharvestete Datensätze können in der eigenen Dataverse-Installation allerdings nicht (lokal) bearbeitet werden und stehen nicht zum Download zur Verfügung<sup>1</sup>.

Auch das IDSC Repository bietet seine Dataverses im Sinne des FAIR-Prinzips als Harvesting Server externen Harvesting Clients an, was eine effektive und standardisierte Integration verschiedener Datensätze des IZA in unterschiedlichen Umgebungen unterstützt. In der Dataverse-Community besteht allerdings auch der Bedarf, Metadaten, Dataverses und insbesondere Datensätze zwischen verschiedenen Dataverse-Instanzen und Repositorien automatisiert zu migrieren.<sup>2</sup>

Im Rahmen des Projekts wurde hierzu eine innovative Lösung zur Migration verschiedener Dataverse-Instanzen entwickelt. Zur Veranschaulichung der Funktionsweise und des Lösungsansatzes wurden eigens zwei Dataverse-Demonstrationsinstanzen im Vorhaben so aufgesetzt und konfiguriert, dass sie untereinander Metadaten, Dataverses und Datensätze austauschen können<sup>3</sup>.

Ein wesentliches Element für diese Datenföderation bildet die Erstellung von Python-Programmen, die die beiden Dataverse-Demonstrationsinstanzen automatisiert zurücksetzen, neu befüllen und für den gegenseitigen Datenaustausch konfigurieren. Diese Python-Module extrahieren die Metadaten, Dataverses und Datensätze aus einer Dataverse-Instanz, speichern sie lokal und ermöglichen den vollständigen Export zu einer anderen (externen) Instanz.

Die Python-Module ermöglichen im Detail:

- das Erstellen, Löschen und Veröffentlichen von Dataversen.
- den Up- und Download aller Metadaten und Daten in ein Dataverse.
- das Veröffentlichen und Löschen von Datensätzen.
- die Anpassung von persistenten Identifikatoren (DOIs etc.) eines Datensatzes an eigene Publikationsnomenklaturen und -standards.
- die Migration der Nutzungsrechte.

Damit steht eine leistungsfähige Lösung zur Verfügung, um Daten effizient zwischen verschiedenen Dataverse-Instanzen zu migrieren. Dies eröffnet auch neue Möglichkeiten zum Export in andere offene Repositorien (wie Zenodo, Figshare, re3data.org etc. oder weitere fachspezifischen Repositorien wie Pangea, etc.), die zum gegenwärtigen Projektstand, jedoch noch nicht geprüft werden konnten.

---

<sup>1</sup> Dies ist erkennbar an dem Symbol für externe Links, welches sich in der oberen rechten Ecke von geharvesteten Datensatzbeschreibungen befindet, sowie dem Hinweis in den Metadaten: “This Dataset is harvested from our partners. Clicking the link will take you directly to the archival source of the data.”

<sup>2</sup> [https://groups.google.com/g/dataverse-community/c/PfKIZFxZHE/m/RBC0j4q1HQAJ?utm\\_medium=email&utm\\_source=footer](https://groups.google.com/g/dataverse-community/c/PfKIZFxZHE/m/RBC0j4q1HQAJ?utm_medium=email&utm_source=footer)

<sup>3</sup> Die beiden Dataverse-Demonstrationsinstanzen sind öffentlich unter folgenden Adressen erreichbar: Demonstrationsinstanz 1: <https://demodv2.iza.org/> & Demonstrationsinstanz II: <https://demodv1.iza.org/>

## Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

Die Python-Module<sup>4</sup> stehen der interessierten Öffentlichkeit und Community als open source code in GitHub zur Verfügung<sup>5</sup> und in den Guidelines von Dataverse selbst<sup>6</sup>. Die Entwicklung der Python-Module zur Datenföderation sowie das Aufsetzen der Demonstrationsinstanzen erfolgte zusätzlich und freiwillig über das Arbeitsprogramm des Vorhabens hinaus, ohne die übrigen Projektmodule und Arbeitspakete zu beeinträchtigen. Dieser Ansatz erleichterte zugleich die Verarbeitung und das Verständnis der komplexen Datenstrukturen, was die effiziente Implementierung und Nutzung der föderierten Lösung förderte.

### **Vollständige Migration**

Grundlage der Migration bildet der identische Metadaten-Standard DDI 2.5, der sowohl in Dataverse wie auch im Vorgänger-Repositorium des IDSC einheitlich vorliegen<sup>7</sup>. Die Migration nach einheitlichem Metadaten-Standard in verschiedenen Instanzen erhöht die Interoperabilität, Qualität und Nachvollziehbarkeit der Daten und erleichtert deren effizienten Austausch und die Integration.

Zu Beginn der Migration konnte auf keine programmatische oder anderweitig softwaregestützte Migration sowohl der Daten wie auch der Metadaten zurückgegriffen werden. So besteht in der Dataverse-Community auch im Verlauf des Projekts weiterhin der Bedarf, Datensätze und Dataverses und deren Metadaten zwischen verschiedenen Dataverse-Instanzen und weiteren Repositorien zu migrieren<sup>8</sup>. Daher ist wichtig zu betonen, dass die Migration vom bisherigen in das neue Dataverse-Datenrepositorium zunächst manuell und nicht programmatisch erfolgte. Eine programmatische Umsetzung wurde zu Projektbeginn als zu zeitaufwendig und ineffektiv bewertet, da der Aufwand nur einmalig für das eigene Repositorium angefallen wäre, ohne Synergieeffekte für andere Repositorien zu erzielen. Um solche Effekte zu erreichen, hätten zahlreiche (Meta-)Datenparameter unterschiedlicher Art und Anbieter berücksichtigt und durch ein komplexes Mapping harmonisiert werden müssen.

Das manuelle Vorgehen bietet zusätzlich den Vorteil, ein tieferes Verständnis der Dataverse-Software und ihrer Anpassungsmöglichkeiten zu erlangen, was die langfristige eigenständige Nutzung und Weiterentwicklung der Plattform erleichtert.

Erst im Verlauf des Projekts wurde eine programmatische Migrationslösung als add-on zwischen verschiedenen Dataverse-Instanzen entwickelt und der Community bereitgestellt um eine technische Lücke der programmgestützten Migration zwischen einheitlichen Dataverse-Instanzen zu schließen.

Folgende Durchführungsschritte sind bei der Erreichung des Milestones „Vollständige Migration“ zu unterscheiden:

---

<sup>4</sup> <https://github.com/iza-institute-of-labor-economics/idsc.dataverse>

<sup>5</sup> <https://github.com/IQSS/dataverse/blob/develop/doc/sphinx-guides/source/api/client-libraries.rst>

<sup>6</sup> <https://guides.dataverse.org/en/latest/api/client-libraries.html#python>

<https://guides.dataverse.org/en/latest/api/apps.html#idsc-dataverse>

<sup>7</sup> Vgl. die Übersicht über die verschiedenen in Harvard Dataverse enthaltenen (Meta)datenstandards:

<https://guides.dataverse.org/en/latest/user/appendix.html>

Vgl. Crosswalk zu DDI 2.5 und weiteren Metadatenstandards: The Dataverse Project. (2020, February).

*Dataverse 4+ Metadata Crosswalk.* <https://docs.google.com/spreadsheets/d/10Luzti7svVTVKTA-px27oq3RxCUM-QbiTkm8iMd5C54>

<sup>8</sup> [https://groups.google.com/g/dataverse-community/c/PfKIZFxFzE/m/RBC0j4q1HQAQ?utm\\_medium=email&utm\\_source=footer](https://groups.google.com/g/dataverse-community/c/PfKIZFxFzE/m/RBC0j4q1HQAQ?utm_medium=email&utm_source=footer)

## Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

### 1. Aufbau der Harvard Dataverse Instanz:

- Konfiguration der Instanz auf einem virtuellen Ubuntu Linux Server.
- Vollständige Containerisierung der Instanz in Docker.
- Bereitstellung der Instanz über einen zentralen Web Proxy, zur Sicherstellung einer Sicherstellung hohen Verfügbarkeit durch Backups, etc.

### 2. Metadatenmigration:

- Umfassende Migration des Metadatenmodells der bestehenden Datenbank in die Harvard Dataverse Instanz.
- Sicherstellung der identischen Abbildung des DDI 2.5 Metadatenstandards in der neuen Instanz.

### 3. Datensatzmigration:

- Vollständige Übertragung sämtlicher Datensätze der Migrationsbasis in die neue Dataverse Instanz.
- Gewährleistung der vollständigen Verfügbarkeit und Downloadbarkeit aller Datensätze in der Dataverse Umgebung.
- Sicherstellung aller Datenschutz und -nutzungsrechte für die Daten sowie Aktualisierung und Weiterentwicklung des IDSC Lizenzmodells

### 4. Website-Aktivierung:

- Launch der neuen Website des Datenrepositoriums im Internet am 17.01.2024. Insgesamt wurden 42 Datensätze mit 111 Dateien (Daten und Dokumentation) migriert. Bis zum Erreichen des Milestones wurden weitere Studien in das Repositorium eingebunden. Seite und Service sind somit produktiv im Einsatz und wurden weiter ausgebaut.

## Föderierte Datennutzung und Dokumentation

Arbeitspaket 3 konzentriert sich auf die Öffnung zusätzlicher Datensilos sowie die Verbesserung der föderierten Datennutzung und -dokumentation durch die Einführung und Implementierung der Dataverse-Technologie. Mit zunehmender Fragmentierung der Angebote verschiedener Datentreuhänder wird die Auffindbarkeit und Dokumentation von Forschungsdaten erschwert. Die Dataverse-Infrastruktur bietet hierfür eine skalierbare Lösung, die den FAIR-Prinzipien (Findable, Accessible, Interoperable, Reusable) entspricht und eine zentrale Plattform für die konsistente Verwaltung von Metadaten und Datensätzen ermöglicht.

Zudem war es im alten Repositorium bislang nicht möglich, Container für Datensätze (Forschungsdaten, Code, Dokumentation und Metadaten) mit thematischen oder institutionellen Schwerpunkten zu erstellen und besonders hervorzuheben. Solche Hervorhebungen erlauben eine höhere Sichtbarkeit und direkte Nutzung von solchen Daten-Sammlungen.

## Einrichtung von fünf Dataverse-Sammlungen

Im Rahmen des Vorhabens wurden fünf englischsprachige Dataverse-Repositories erfolgreich eingerichtet, die den Zugang zu Forschungsdaten im Bereich der Arbeitsmarkt- und Verhaltensökonomie nachhaltig verbessern. In Kooperation mit dem Harvard Dataverse wird unser Repository dort gespiegelt<sup>9</sup>:

### 1. [IZA Dataverse](#)

Das zentrale Repositorium für das Forschungsdatenzentrum des IZA (IDSC) bietet eine umfassende Plattform zur Auffindbarkeit, Speicherung und Bereitstellung arbeitsmarktrelevanter Forschungsdaten. Es ermöglicht die Forschung sowie die Replikation von Ergebnissen durch ein strukturiertes Metadatenverzeichnis, das kontinuierlich erweitert wird.

---

<sup>9</sup> [https://dataverse.harvard.edu/dataverse/idsc\\_harvested](https://dataverse.harvard.edu/dataverse/idsc_harvested)

# Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

## 2. [ECONtribute Dataverse](#)

Diese Sammlung wird in Kooperation mit der Universität Bonn und der Universität zu Köln betrieben und unterstützt die interdisziplinäre Forschung im Rahmen der Exzellenzinitiative ECONtribute. Ziel ist die Förderung von hochqualitativen Datensammlungen zur wirtschaftswissenschaftlichen Forschung, insbesondere in den Bereichen Märkte und öffentliche Politik.

## 3. [G<sup>2</sup>LM|LIC Dataverse](#)

Ein Kooperationsprojekt zwischen dem IZA und dem britischen [Foreign, Commonwealth & Development Office \(FCDO\)](#), das umfangreiche Forschungsdatensätze zur globalen Arbeitsmarktforschung bereitstellt. Dies umfasst auch Daten zu COVID-19-bezogenen Studien und unterstützt evidenzbasierte Entscheidungsfindung sowie den internationalen Kapazitätsaufbau.

## 4. [WageIndicator Dataverse](#)

Diese Sammlung basiert auf einer Kooperation mit der WageIndicator Foundation und stellt Lohn- und Arbeitsmarktdaten für die wissenschaftliche Nutzung bereit. Das IZA fungiert dabei als exklusiver akademischer Datenprovider für die Foundation.

## 5. [Research and Replication Data for Labor Economics](#)

Das Forschungsdatenzentrum des IZA (IDSC) bietet mit dieser Sammlung ein durchsuchbares Katalogsystem für Metadaten zu relevanten Datensätzen im Bereich der Arbeitsmarkt- und Verhaltensökonomie. Die Sammlung verweist auf externe Repositories und hat das Ziel, ein möglichst umfassendes Metadatenverzeichnis zu schaffen. Auf diese Weise werden auch schwer auffindbare Daten für Forschende leichter zugänglich gemacht.

### Nutzungsstatistiken

Die Nutzungsstatistiken zeigen die Aktivität und Akzeptanz des neuen Dataverse-Portals. Im Folgenden werden die jährlichen und monatlichen Entwicklungen der Nutzerzahlen analysiert, um die Auswirkungen der Systemeinführung auf die Datennutzung in der Forschung zu verdeutlichen.

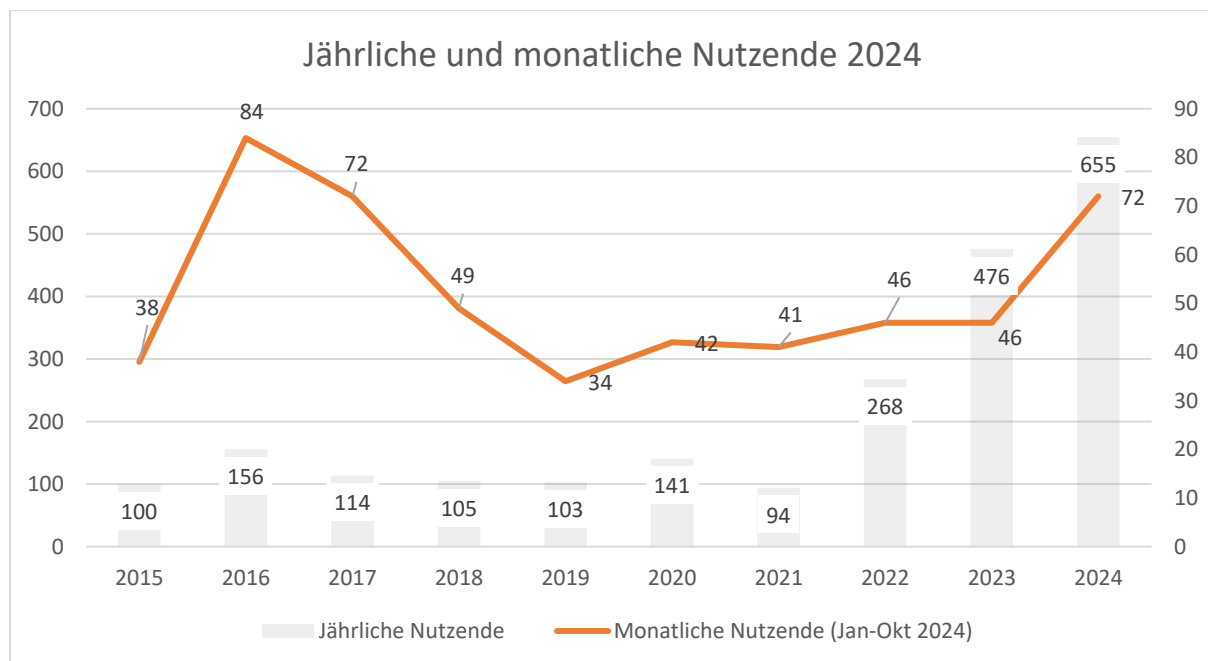


Abb. 6 Jährliche und monatliche Nutzer

Die Analyse der monatlichen Nutzungszahlen zeigt deutliche Schwankungen, insbesondere in den ersten drei Monaten des Jahres 2024. Im Januar erfolgte der Launch, und alle bestehenden Datensätze wurden integriert. Werbemaßnahmen wie Newsletter und Mailings folgten im Februar und erneut im

## Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

März. Der Januar weist mit 38 Einträgen einen vergleichsweise niedrigen Wert auf, während im Februar mit 84 Einträgen der Höchststand erreicht wurde, gefolgt von 72 Einträgen im März. Ab April stabilisieren sich die Zahlen, was auf eine reguläre Bearbeitung durch Forschende hindeutet. Im Oktober und November sind erneut Peaks zu erkennen, die mit der Veröffentlichung neuerer Studien nach der Sommerpause zusammenhängen. Insgesamt konnten im Projektverlauf in 2023 44 und in 2024 55 Datensätze veröffentlicht werden.

Die intensive Nutzung übertrifft zudem deutlich die Werte vergangener Jahre. Vor 2022 lag die maximale Jahresnutzung bei nur 156 Nutzern (2016), während die heutige monatliche Nutzung diesen Wert nahezu erreicht. Diese Zahlen verdeutlichen die gestiegene Akzeptanz und das Vertrauen in den Dienst.

Ein entscheidender Schritt für das Wirken des IDSC Repositoriums war die Einführung des Vorgänger-Portals **datasets.iza.org** im Jahr 2020. Mit dem neuen Dataverse-Portal und seiner verbesserten Datenstruktur konnte eine weitere signifikante Steigerung der Nutzung erreicht werden. Insgesamt zeigen die Analysen, dass die Nutzeraktivität seit der Einführung des neuen Systems erheblich zugenommen hat.

### **Mehrwert durch die Dataverse-Technologie**

Die Dataverse-Infrastruktur bietet eine skalierbare und föderierte Lösung für die Datenbereitstellung. Über standardisierte Schnittstellen (APIs) können Datensätze und Metadaten automatisiert importiert und integriert werden. Dies erleichtert die langfristige Archivierung, Replikation und gemeinsame Nutzung der Daten. Gleichzeitig fördert die Plattform die interinstitutionelle Zusammenarbeit, insbesondere für Datentreuhänder mit begrenzten Ressourcen zur eigenständigen Verwaltung und Dokumentation ihrer Datensätze. Durch die erfolgreiche Implementierung der Plattform und der fünf neuen Dataverse-Sammlungen ist das IDSC nun aktiv Teil der internationalen Dataverse-Entwicklungs- und Nutzergemeinschaft. Die IDSC Dataverse-Instanz ist auf den globalen Plattformübersichten sichtbar und das IDSC ist in der Lage, andere Partner durch den Austausch von technischem Know-how dabei zu unterstützen, ihre eigenen Dataverse-Repositories aufzubauen. Zusätzlich bietet das IDSC Hosting-Services für weitere Institutionen an und steht hierzu bereits in Gesprächen mit potenziellen Partnern.

### **Arbeitspaket 4: Josua-Dokumentation**

Das Ziel des Arbeitspakets 4 ist die Erstellung einer umfassenden Dokumentation, die es Datentreuhändern ermöglicht, die Josua-Komponenten eigenständig zu implementieren und anzupassen. Diese Dokumentation umfasst sämtliche Elemente der Josua-Lösung und deren Zusammenspiel. Sie wurde in Markdown verfasst und wird über ein Git-Repository verwaltet, um kontinuierliche Aktualisierungen und Anpassungen zu ermöglichen<sup>10</sup>. Wesentliche Aspekte sind die Interfaces, die im Folgenden detailliert beschrieben werden:

#### **1. User Interface:**

Das User Interface dient den Forschenden als zentrale Arbeitsumgebung für die sichere Analyse sensibler Daten. Die Dokumentation beschreibt folgende Funktionen:

- Navigationsstruktur: Überblick über die Benutzerführung, Menüoptionen und Hauptfunktionen. Forschende erhalten eine klare Orientierung, um ihre Arbeitsschritte effizient durchführen zu können.

---

<sup>10</sup> Vgl. <https://github.com/iza-institute-of-labor-economics/idsc-josua-doku/>

Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

- Interaktionen: Anleitungen für den Zugriff auf Datensätze sowie die Ausführung von Analyse-Skripten. Auch sicherheitsrelevante Aspekte wie die Anzeige von Wasserzeichen bei Ergebnissen werden erläutert.

Dieses Interface stellt sicher, dass Forschende sich in einer benutzerfreundlichen und gut strukturierten Umgebung bewegen können, die gleichzeitig die hohen Datenschutzerfordernungen erfüllt.

## 2. **Operator Interface:**

Das Operator Interface ist für die Verwaltung und Überwachung der Josua-Instanz verantwortlich. Die Dokumentation erklärt alle relevanten administrativen Funktionen, darunter:

- Benutzer- und Rollenmanagement: Anleitung zur Erstellung, Verwaltung und Zuordnung von Benutzerrollen, um differenzierte Zugriffsrechte festzulegen.
- Systemkonfiguration: Beschreibungen zur Einrichtung der Josua-Komponenten, einschließlich der Integration von Datensätzen und Sicherheitsmechanismen.
- Überwachung und Protokollierung: Tools und Funktionen zur Überprüfung von Systemaktivitäten, Zugriffen und der ordnungsgemäßen Funktion der Plattform.

Dieses Interface stellt sicher, dass die Betreiber die Kontrolle über die Plattform behalten und individuelle Anforderungen flexibel umsetzen können. Die kombinierte Dokumentation der User- und Operator-Interfaces stellt somit eine wesentliche Grundlage für die nachhaltige Nutzung und Weiterentwicklung der Josua-Plattform dar. Sie ermöglicht sowohl Forschenden als auch Betreibern eine effiziente und sichere Nutzung der Dateninfrastruktur.

Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

## **Ergänzungen zur Teil II: Eingehende Darstellung**

### **Wesentliche Positionen des zahlenmäßigen Nachweises**

Überschneidungen mit anderen Positionen lagen nicht vor, da ausschließlich Personalausgaben angefallen sind. Die Einhaltung der Ausgabenplanung wurde vollständig sichergestellt; die Ausgaben entsprechen exakt den geplanten Beträgen.

### **Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten**

Die Projektarbeiten erwiesen sich als notwendig, um die Projektziele – insbesondere die Öffnung von Datensilos, die Verbesserung der Dateninfrastruktur und die Entwicklung förderierter Nutzungsmöglichkeiten – zu erreichen. Die durchgeführten Maßnahmen, einschließlich der Integration neuer Technologien wie Dataverse und der umfassenden Dokumentation der Josua-Komponenten, ermöglichten eine nachhaltige und zukunftssichere Nutzung.

### **Voraussichtlicher Nutzen und Verwertbarkeit des Ergebnisses**

Mit der Implementierung und Bereitstellung der Dataverse-Infrastruktur wurde eine Grundlage geschaffen, die auch langfristig für andere Institutionen und Partner nutzbar ist. Die Sichtbarkeit und Verfügbarkeit der Daten wurden signifikant verbessert. Die Josua-Plattform sowie die Dataverse-Module fördern eine erhöhte interinstitutionelle Zusammenarbeit und bieten Synergieeffekte für Datenbereitstellung, Dokumentation und Replikation. Hosting- und Beratungsservices sind bereits in Planung.

### **Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen**

Während des Projekts wurde bekannt, dass innerhalb der Dataverse-Community ähnliche Anforderungen an Datenmigration und -föderation bestehen. Diese Erkenntnisse flossen direkt in die Entwicklung spezifischer Lösungen, wie die Python-Module zur Migration zwischen Dataverse-Instanzen, ein.

### **Erfolgte und geplante Veröffentlichungen**

Im Rahmen des Projekts wurden Ergebnisse und Entwicklungen, insbesondere zur Dokumentation des Operator- und User-Interfaces für die Datentreuhand-Community veröffentlicht und geteilt.

Sachbericht im Rahmen der BMBF-Förderrichtlinie: „Förderung von Projekten zur Erforschung oder Entwicklung praxisrelevanter Lösungsaspekte („Bausteine“) für Datentreuhandmodelle

### **Teil III: Erfolgskontrollbericht**

#### **Wissenschaftlich-technische Ergebnisse und gesammelte Erfahrungen**

Das Projekt „Josua4DTM“ führte zur erfolgreichen Integration der Module **Josua Web** und **Josua Direct**, die bereits in akkreditierten Forschungsdatenzentren genutzt werden. Die wesentlichen technischen Fortschritte beinhalten:

- Multidomainfähigkeit: Mehrere Josua-Instanzen können nun gleichzeitig leistungsstarke Serverressourcen nutzen.
- Automatisierte API-basierte Nutzerverwaltung: Entwicklung eines zentralen Services für Active Directory und Projektkonten.
- Erweiterte Exportmöglichkeiten: Automatisierte Erstellung von Nutzungsstatistiken.

Die Migration bestehender Datensätze auf die Dataverse-Plattform ermöglichte die förderierte Nutzung nach FAIR-Prinzipien. Erfahrungen in der Zusammenarbeit mit Kooperationspartnern haben zur Weiterentwicklung der Plattform beigetragen.

#### **Fortschreibung des Verwertungsplans**

##### **Schutzrechte und Innovationen:**

- Es wurden keine neuen Schutzrechte beantragt.
- Die Nutzung bestehender Open-Source-Software (z. B. Dataverse) wurde optimiert.

##### **Wirtschaftliche Erfolgsaussichten:**

- Die neue Infrastruktur bietet erhebliche Vorteile durch verbesserte Effizienz und Kostensenkung bei der Datenverarbeitung.
  - Zielgruppen sind nationale und internationale Forschungseinrichtungen sowie Datentreuhänder.
  - Erste Anfragen für Hosting- und Beratungsservices zur Dataverse-Plattform bestehen bereits.

##### **Wissenschaftliche Erfolgsaussichten:**

- Ausbau der Kooperationen mit dem ECONtribute Exzellenzcluster, G<sup>2</sup>LM|LIC, und der WageIndicator Foundation.
- Entwicklung neuer technischer Lösungen wie Python-Module zur Datenmigration und -föderation.

#### **Angaben zu erfolglosen Arbeiten**

- Die technische Herausforderung der automatisierten DOI-Vergabe wurde als manuell zu lösendes Problem definiert.

#### **Einhaltung der Ausgaben- und Zeitplanung**

Die Arbeiten wurden im geplanten Zeitraum abgeschlossen und die Budgetvorgaben vollständig eingehalten. Die technische Implementierung und Dokumentation wurden rechtzeitig fertiggestellt.