EMILiA

Entwicklung einer lizenzbasierten Software für die Archivierung von E-Mails

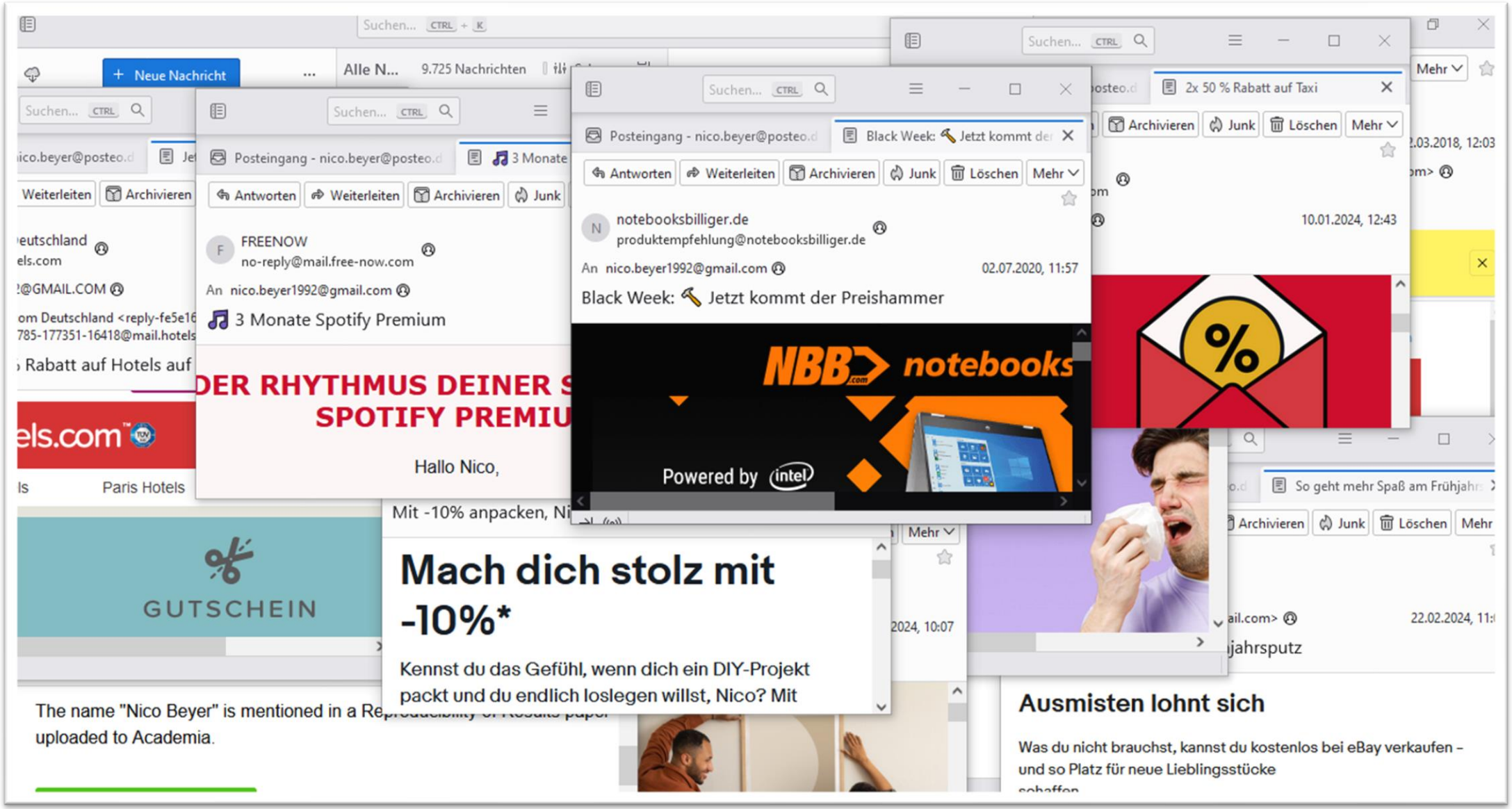
Nico Beyer (Freie Universität Berlin)

EMILiA

IBB Business Team

Freie Universität Berlin

ARCHIV DER MAX-PLANCK-GESELLSCHAFT


Neue Nachricht
Alle N...
9.725 Nachrichten
Posteingang - nico.beyer@posteo.d
Black Week: Jetzt kommt der
Antworten
Weiterleiten
Archivieren
Junk
Löschen
Mehr
FREENOW
no-reply@mail.free-now.com
An nico.beyer1992@gmail.com
3 Monate Spotify Premium
DER RHYTHMUS DEINER S
SPOTIFY PREMIU
Hallo Nico,
notebooksbilliger.de
produktempfehlung@notebooksbilliger.de
An nico.beyer1992@gmail.com
02.07.2020, 11:57
Black Week: Jetzt kommt der Preishammer
NBB notebooks
Powered by intel
2x 50 % Rabatt auf Taxi
10.01.2024, 12:43
Paris Hotels
GUTSCHEIN
Mit -10% anpacken, Ni
Mach dich stolz mit
-10%*
Kennst du das Gefühl, wenn dich ein DIY-Projekt
packt und du endlich loslegen willst, Nico? Mit
So geht mehr Spaß am Frühjahr
22.02.2024, 11:
jahrsputz
Ausmisten lohnt sich
Was du nicht brauchst, kannst du kostenlos bei eBay verkaufen –
und so Platz für neue Lieblingsstücke
The name "Nico Beyer" is mentioned in a Re
uploaded to Academia.

EMILiA

Herausforderungen

Spam, Werbung, E-Mails ohne Informationswert

Händische Bewertung und Erschließung ist aufgrund von riesigen Datenmengen nicht zu bewerkstelligen

Personenbezogene Daten

Viele E-Mails müssen anonymisiert werden

Schadhafte Anhänge

E-Mails können Viren enthalten

EMILiA

Automatisierung als Chance

- Fortschritte im Bereich der künstlichen Intelligenz können dabei helfen, Prozesse zu vereinfachen und zu automatisieren
- Wichtige Entscheidungen werden nach wie vor von Archivfachkräften getroffen
- KI ist in erster Linie als Hilfsmittel für die Bewältigung der riesigen Datenmengen zu verstehen

EMILiA

Themen

1 Rückblick

2 Aktueller Stand

3 Herausforderungen

4 Vision

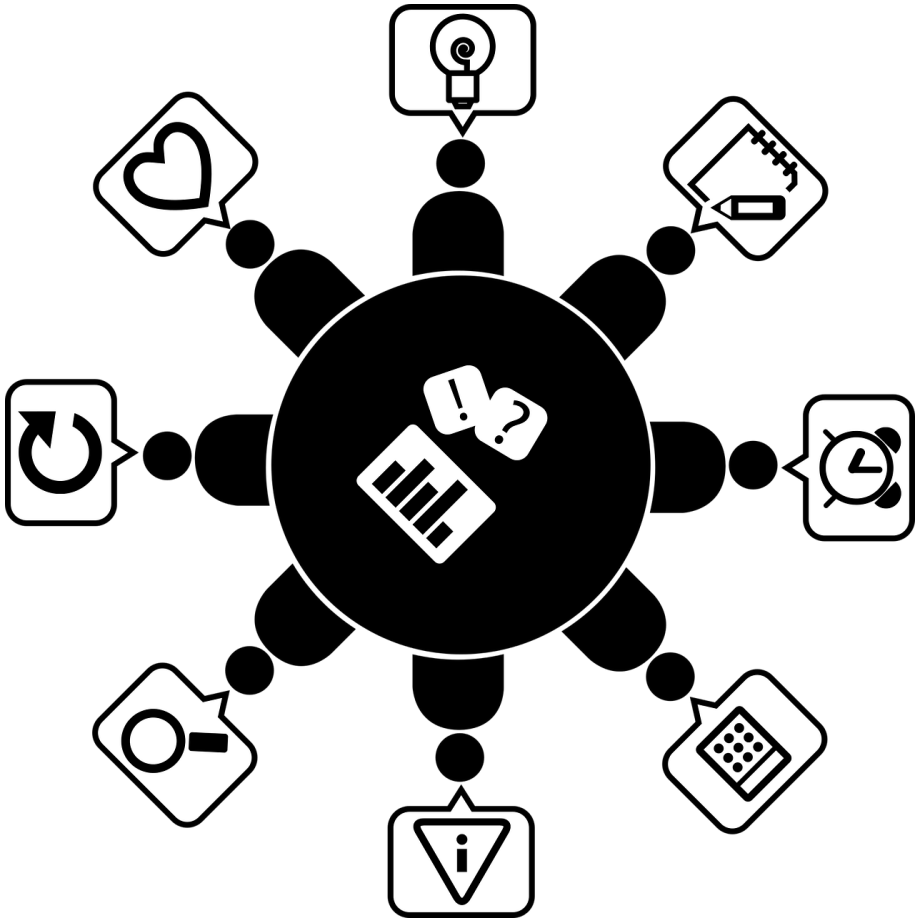
5 Zeit für Fragen

EMILiA

1

Rückblick

Rückblick



- 2015: Beginn der Konzeption im Archiv der MPG in Kooperation mit dem Fachbereich Informatik der FU Berlin
- Einstellung einer studentischen Hilfskraft im Archiv der Max-Planck-Gesellschaft
- Spezifikation und Entwicklung eines Prototyps
- Tests mit Echtdaten des Archivs der Max-Planck-Gesellschaft
- 2024: Förderprogramm „ProValid“ der Investitionsbank Berlin

EMILiA

2

Aktueller Stand

Grundkonzept

Erfassung, Management, Indizierung, Limitierung, intelligente Analyse

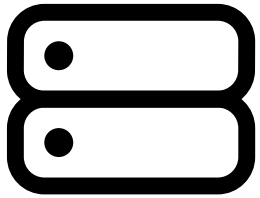
Abgabe an das Archiv

Übernahme

- Import
- Virenprüfung
- Formaterkennung
- Authentizität
- Integrität
- Erstellung von SIPs

Bewertung & Erschließung

- Erkennung von Spam und Dubletten
- Erkennung von Themen, Personen und Orten
- Identifikation personenbezogener Daten



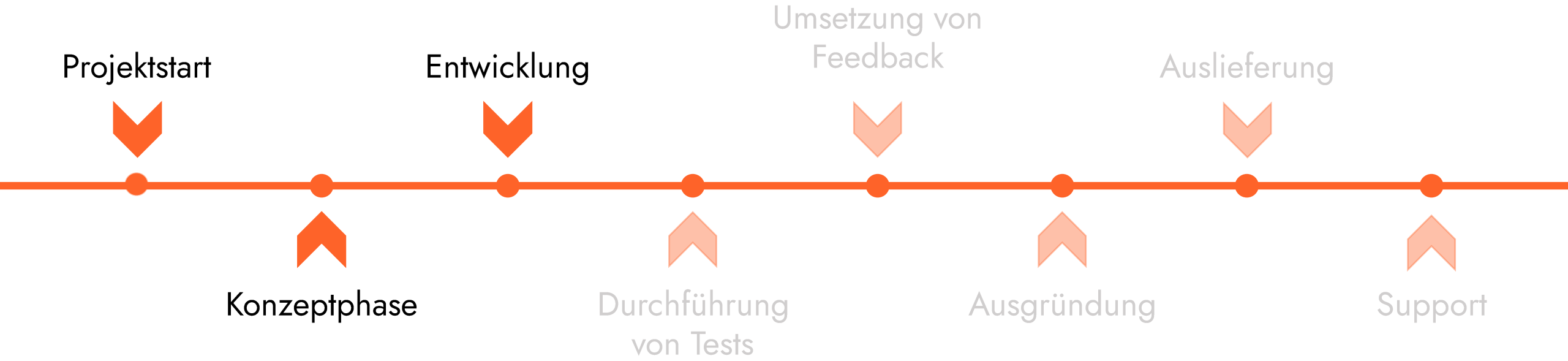
Übergabe an digitales Langzeitarchiv

Nutzung

- Recherchedatenbank
- Anonymisierung
- Erstellung von DIPs
- Darstellung

EMILiA

Meilensteine



Entwicklungsstand

Übernahme

- Übernahme und Erstellung von SIPs funktioniert bereits

Bewertung & Erschließung

- Erkennung von Klarnamen, Sprachen, Themen und Threads wurde getestet aber noch nicht implementiert
- Dubletten- und Spamerkennung müssen noch evaluiert werden

Nutzung

- Recherchedatenbank und Viewer müssen noch entwickelt werden

EMILiA

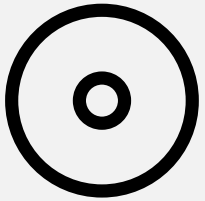
Optionen für die Übernahme

1

Abgabe über Mailserver

Deponent*In meldet sich auf Mailserver an und importiert den Account

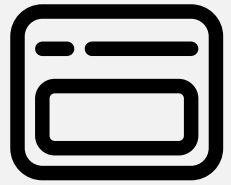
2



Physische Abgabe

Deponent*In überträgt E-Mails mithilfe eines Abgabetools auf physisches Speichermedium

3

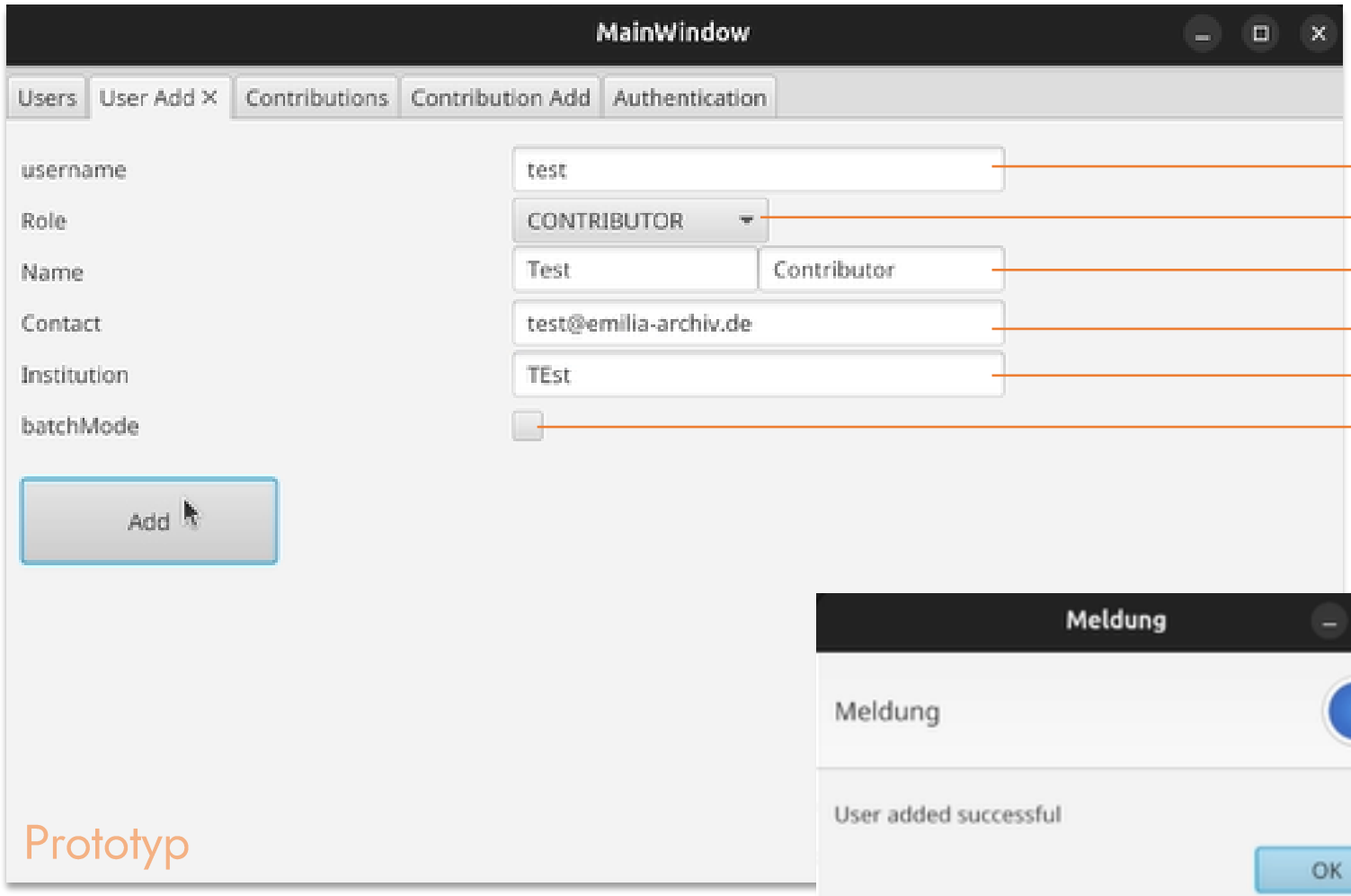


Abgabetool

E-Mails werden mithilfe eines Abgabetools übertragen

EMILiA

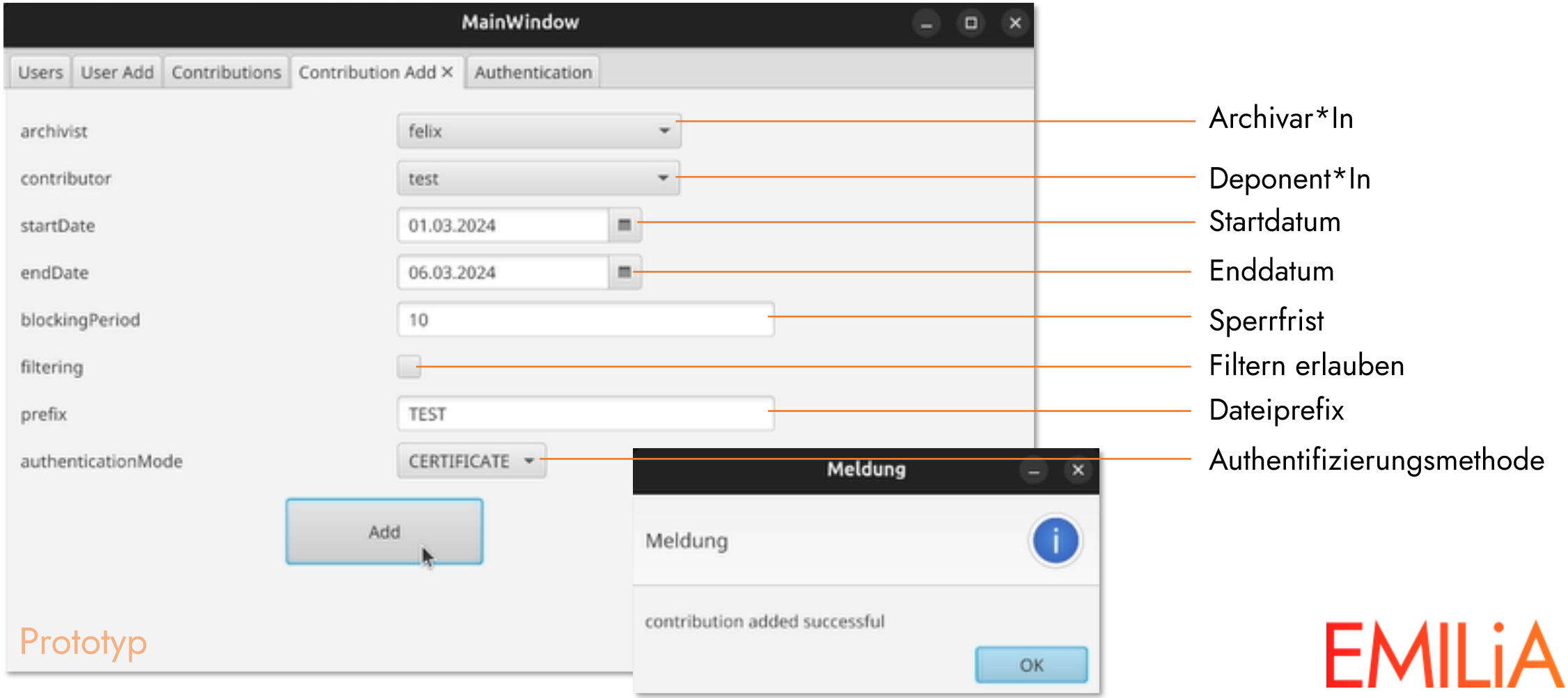
Archiv: Deponent*In anlegen



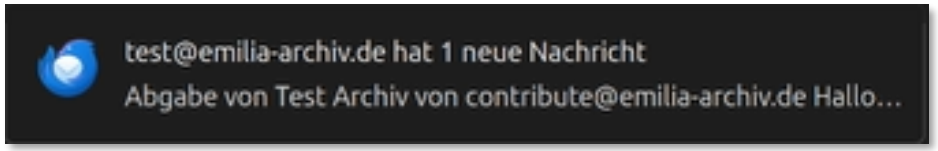
- Name
- Rolle
- Vor- und Nachname
- E-Mail-Adresse
- Institution
- Stapelverarbeitung

EMILiA

Archiv: Übernahme vorbereiten

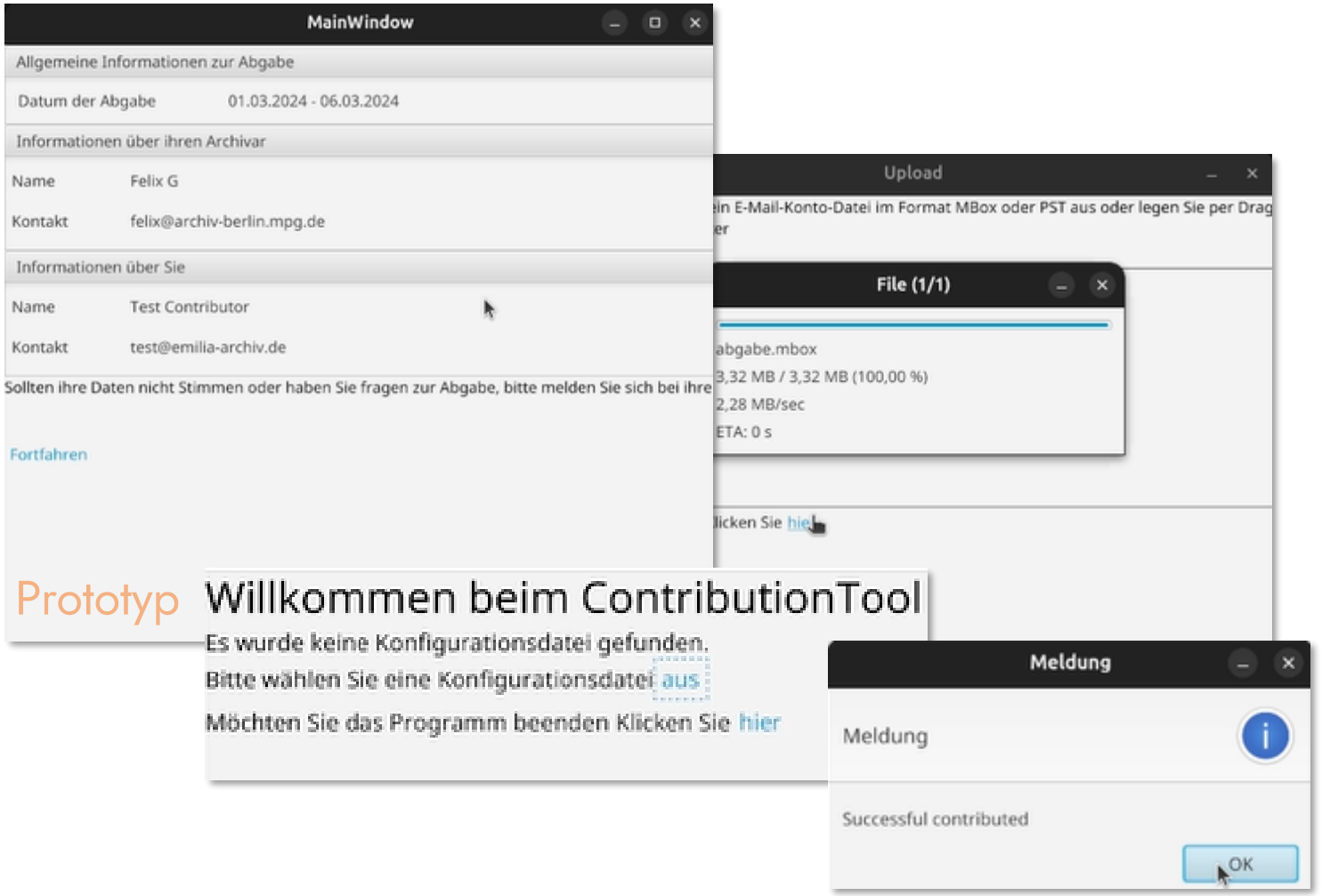


Deponent*In: Benachrichtigung



Konfigurationsdatei mit Spezifikationen für Abgabe wird automatisch versendet

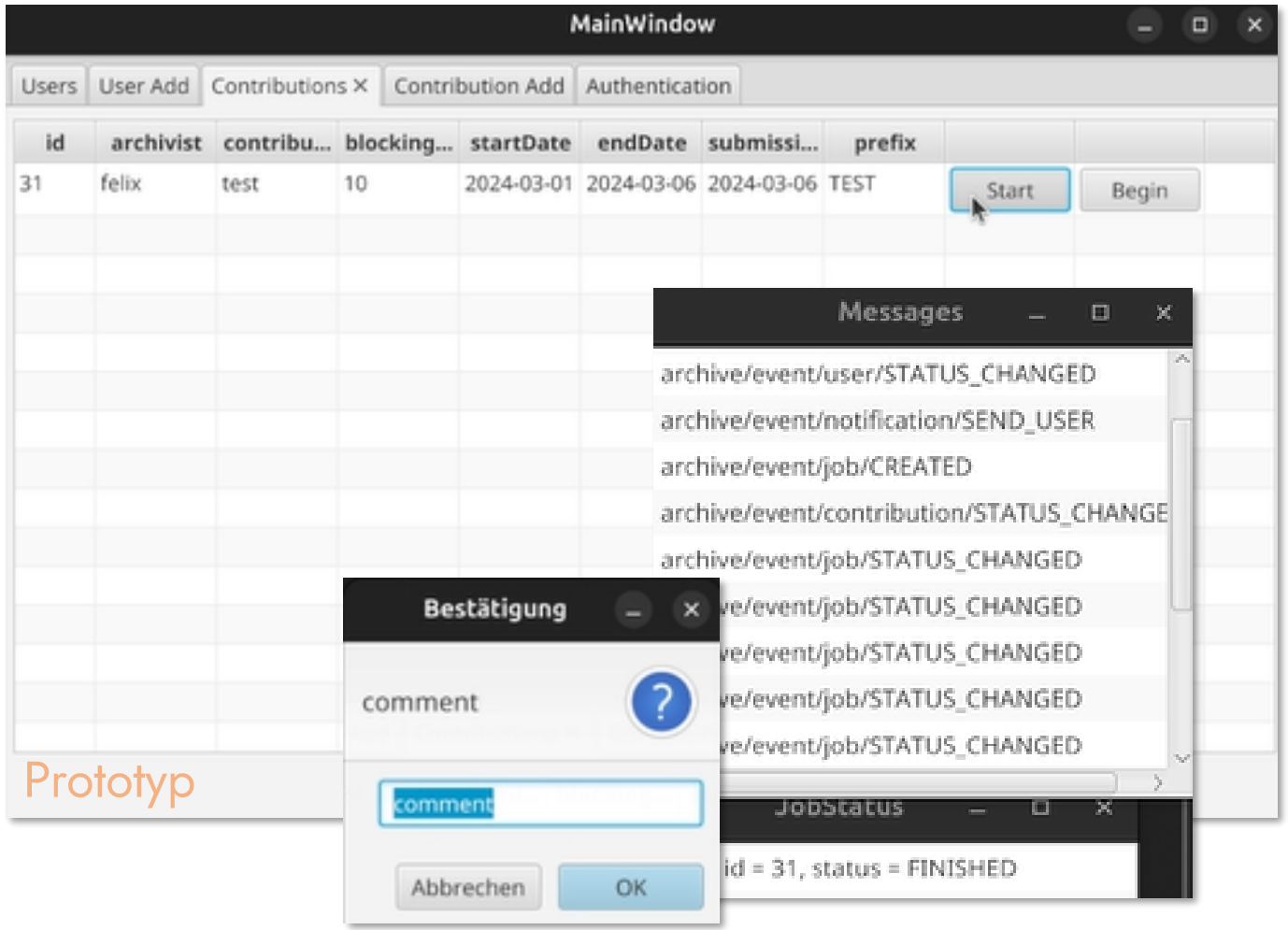
Deponent*In: Abgabe



1. Abgabetool öffnen
2. Konfigurationsdatei auswählen
3. Daten bestätigen
4. MBOX oder PST auswählen
5. Kommentar abgeben (optional)

EMILiA

Archiv: Übernahme initiieren



1. Abgabetool öffnen
2. Kommentar abgeben (optional)
3. Übernahme starten
4. Übertragung abwarten

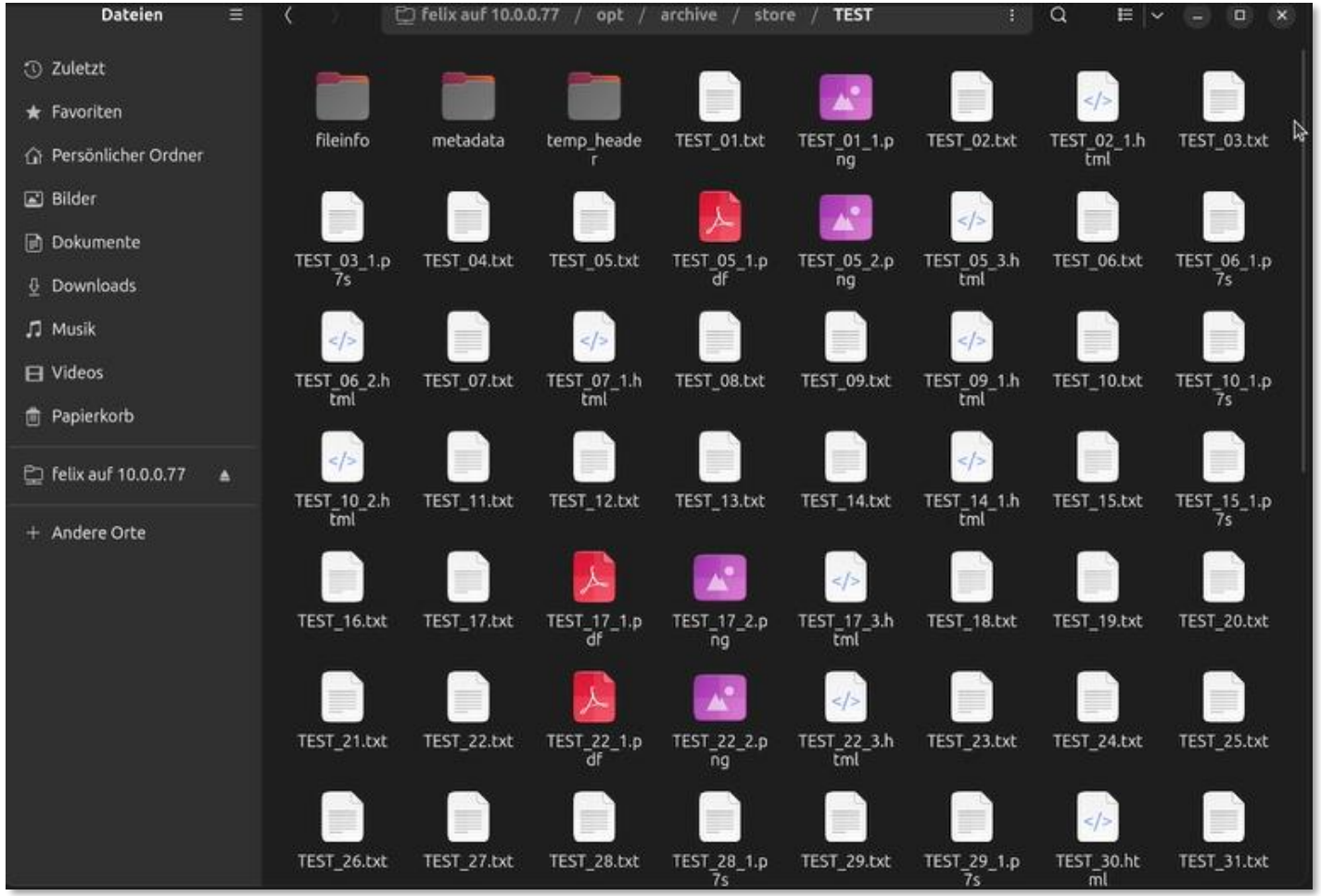
Dauer der Übernahme hängt von der Leitung der abgebenden Person und des Archivs ab

EMILiA

Übernahme-Workflow

EMILiA

Ergebnis der Übernahme

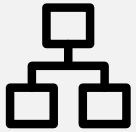


- Container mit Strukturdatei, Prüfsummen, PREMIS-Metadaten, E-Mails und Anhängen

EMILiA

Inhalt der SIPs

Angaben zum SIP (.xml)



Strukturdatei (.xml)

PREMIS-Metadaten (.xml)

E-Mails (.txt) und Anhänge (Original)

Prüfsummen (.txt)

EMILiA

3

Herausforderungen

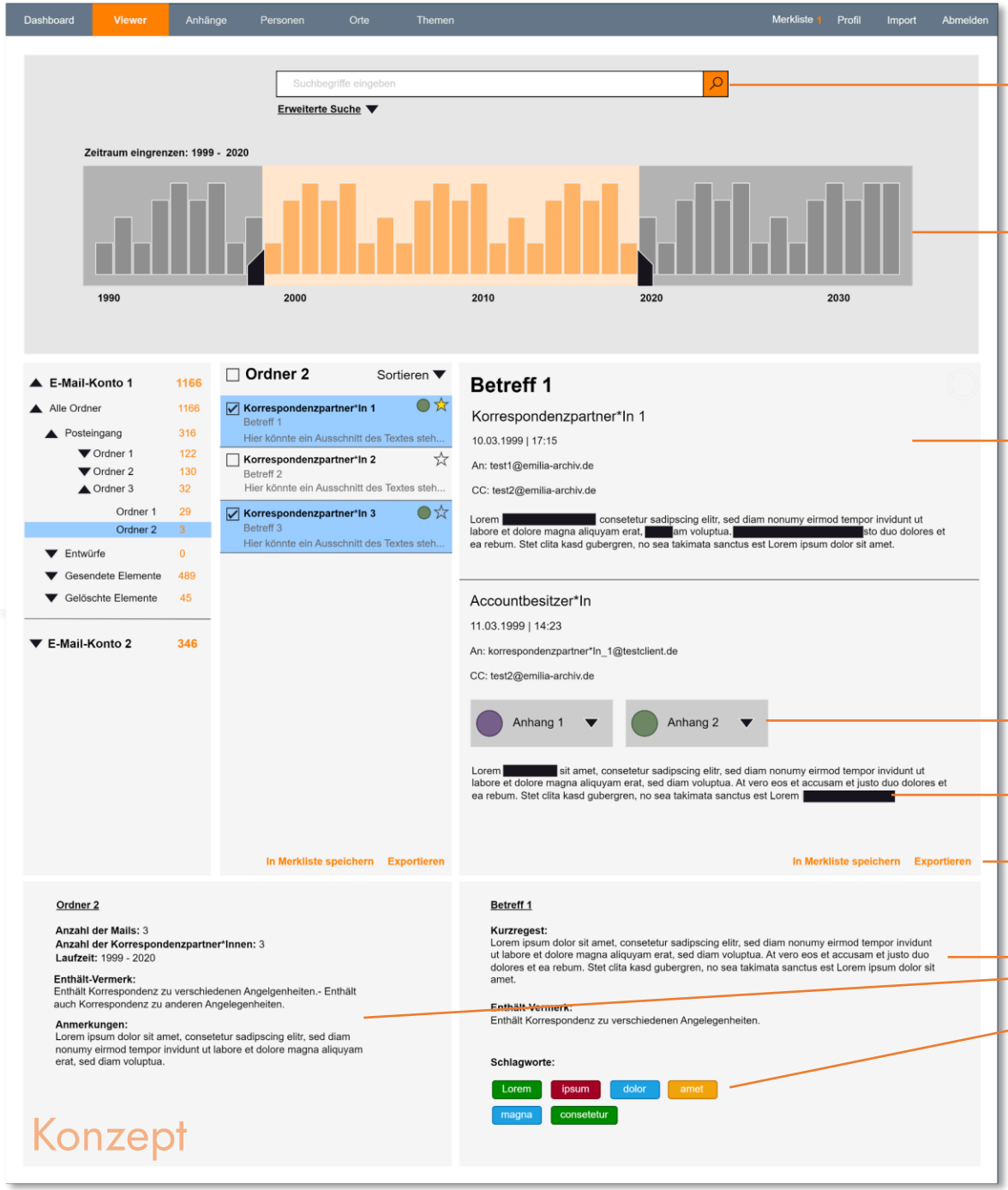
Herausforderungen

1. E-Mail Standard ist sehr offen, weshalb die eingehenden Metadaten sehr unterschiedlich aussehen können
2. Unterschiedliche E-Mail-Clients, proprietäre Dateiformate und Alter der Daten können zu Defekten und Anomalien führen
3. Umgang mit verschlüsselten E-Mails ist noch unklar
4. Unterschiedliche Zeichenkodierungen können dazu führen, dass einige E-Mails nicht korrekt dargestellt werden
5. Entscheidung für ein archivtaugliches Containerformat steht noch aus
6. Test an Echtdaten zeigt aktuell, dass die Informationsdichte sehr gering ist

EMILiA

4

Vision



Suchschlitz

Histogramm

An E-Mail-Clients angelehnter Viewer

Anhänge

Schwärzung

Merkliste und Exportfunktion

Erschließungsdaten und von der KI erkannte Entitäten

EMILiA

Zeit für Fragen

5

Vielen Dank für Ihre Aufmerksamkeit!

Fragen und Vorschläge sind jederzeit willkommen.

Nico Beyer
Freie Universität Berlin
AG Zuverlässige Systeme
EMILiA-Projekt

E-Mail: nico.beyer@emilia-archiv.de
Telefon: +49 30 841 337 15

Boltzmannstraße 14
14195 Berlin-Dahlem

www.emilia-archiv.de

EMILiA