



Neues vom EMILiA-Projekt

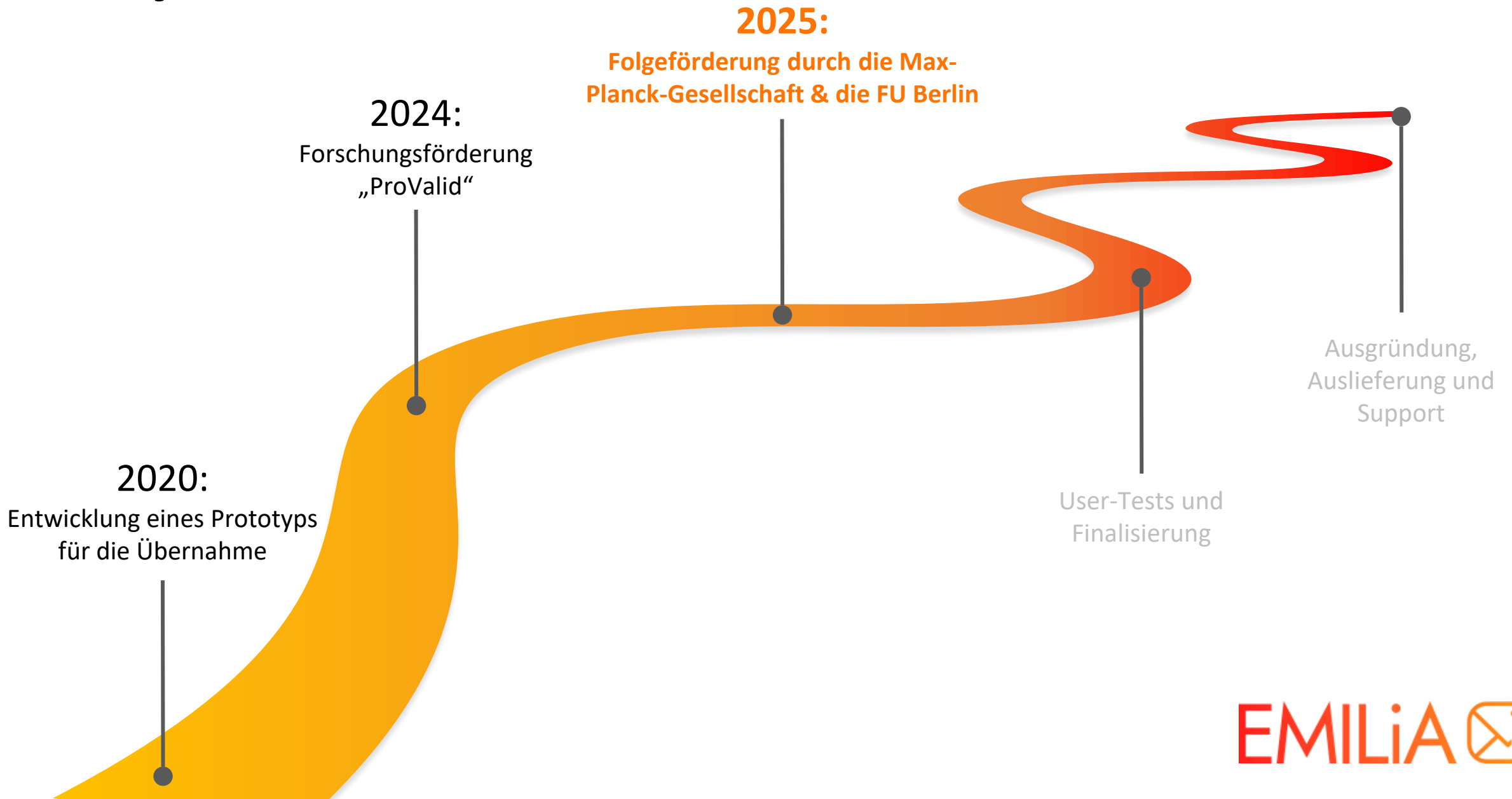
Aktueller Stand und Zukunftsperspektiven

Nico Beyer & Felix Gericke

Programm



Projektrahmen



Ausgangssituation

E-Mails sind für die meisten Menschen zu einem festen Bestandteil des beruflichen und privaten Alltags geworden.

E-Mails sind ein zentrales Kommunikationsmedium

Ein großer Teil der meisten Postfächer besteht aus **Spam, Werbung** oder nur **kurzfristig relevanten Informationen**

Viele wertlose Informationen

Es gibt jedoch auch **historisch oder rechtlich relevante E-Mails**, die langfristig bewahrt werden sollten

Es muss selektiert werden

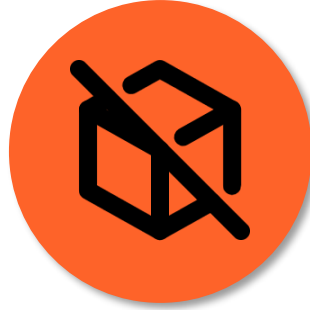
Eine **fachgerechte Auswahl, Archivierung und Auswertung** ist nur mithilfe automatisierter Prozesse möglich

Automatisierung als Chance

Herausforderungen



E-Mail-Standard macht nur wenige klare Vorgaben



Gängige E-Mail-Container sind nicht gut für die Archivierung geeignet



Anhänge in den unterschiedlichsten Formaten



Signierte und verschlüsselte E-Mails

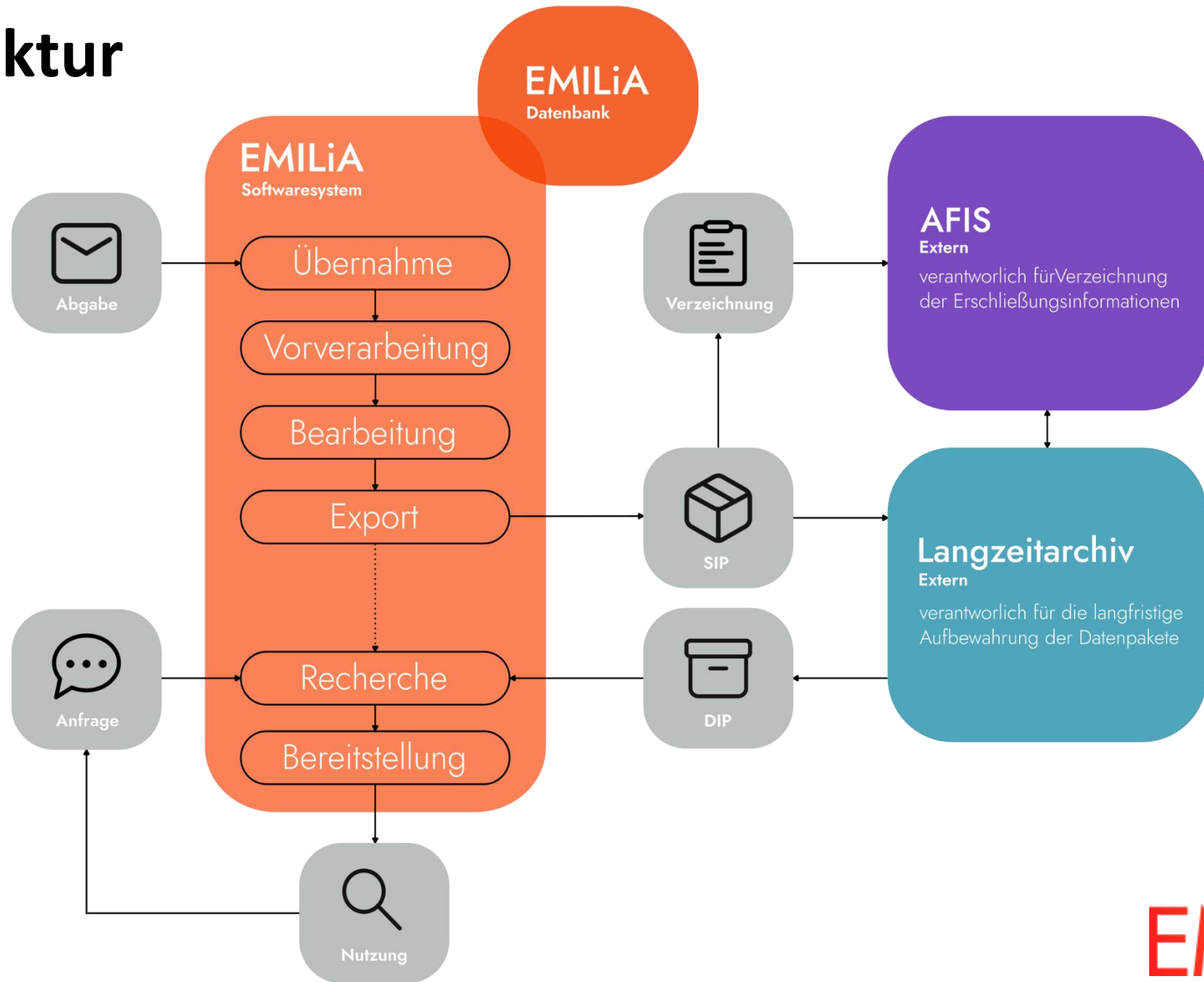


Personenbezogene Daten und urheberrechtlich relevante Dokumente



Archivwürdiger Kern ist nur schwer ausfindig zu machen

Architektur



Entwicklungsstand

➤ *Mit der aktuellen Version von EMILiA kann ein **vollständiger Archivierungsprozess** durchlaufen werden.*

Übernahme: Sicherer Transfer, Konvertierung, Virenprüfung, Formaterkennung, Integritätssicherung

Bewertung & Erschließung: Spracherkennung, Spam- und Dublettenerkennung, Themenerkennung, Identifikation personenbezogener Daten, Kassation, Merkliste

Recherche: Viewer, Optionale Anonymisierung, Such- und Filterfunktionen, Statistiken und Netzwerke, Zusammenfassung von Threads mittels LLM

Export: MBOX oder strukturierter BagIt-Container mit Metadaten

Roadmap

0.8.0



Verbesserte und individualisierbare Personendatensätze und Verbesserung der Volltextsuche

0.8.1



Anzeige von HTML und Verbesserung der Performance bei großen Datensätzen

0.9.0



Rechtemanagement, Sperrfunktion und Anhangsviewer

**In den nachfolgenden Versionen wird sich das Projektteam in erster Linie auf das Recherche- und Bereitstellungstool fokussieren*

User tests

1

Mitarbeiter*Innen des Archivs der Max-Planck-Gesellschaft
→ Seit 12/24

2

Erste externe User tests mit ausgewählten Institutionen
→ Ab 05/25

3

Offene Testrunde, an der sich alle Interessierten beteiligen dürfen
→ Abhängig vom Umfang des Feedbacks der vorherigen Runden

Live-Demonstration

Übernahme, Vorverarbeitung, Statistik-Dashboard, Viewer, Kassation, Export

Vielen Dank für Ihre Aufmerksamkeit!

E-Mail: info@emilia-archiv.de

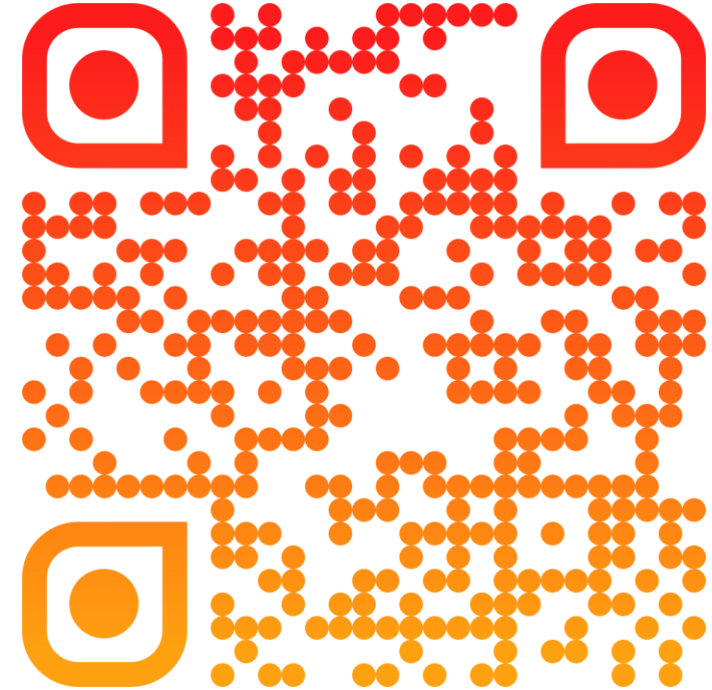
Telefon: +49 30 841 337 15

Archiv der Max-Planck-Gesellschaft

EMILiA-Projekt

Boltzmannstraße 14

14195 Berlin-Dahlem



www.emilia-archiv.de