

# EMILiA

Wie aus einem Problem viele wurden.  
Entwicklung einer teilautomatisierten Software für die  
Archivierung von E-Mails

Nico Beyer | Freie Universität Berlin / Archiv der Max-Planck-Gesellschaft

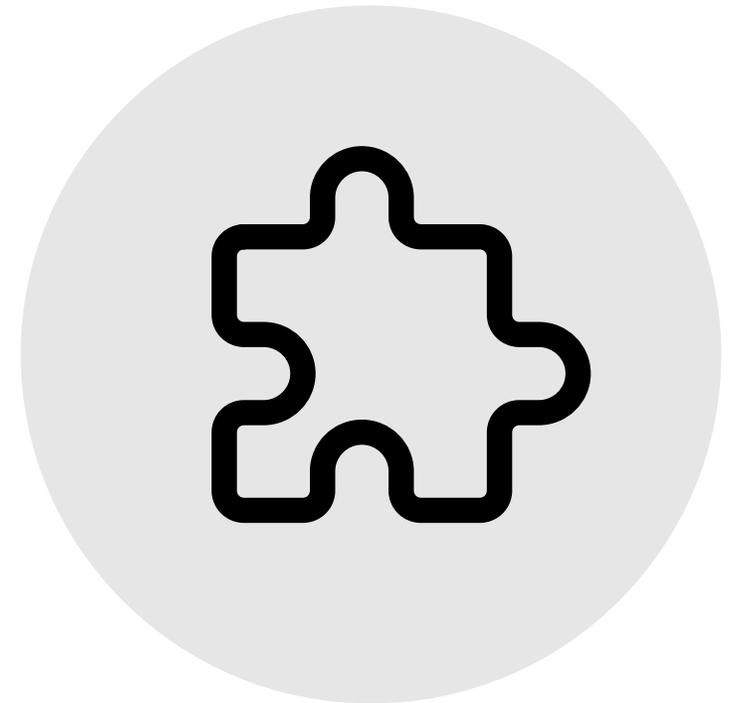




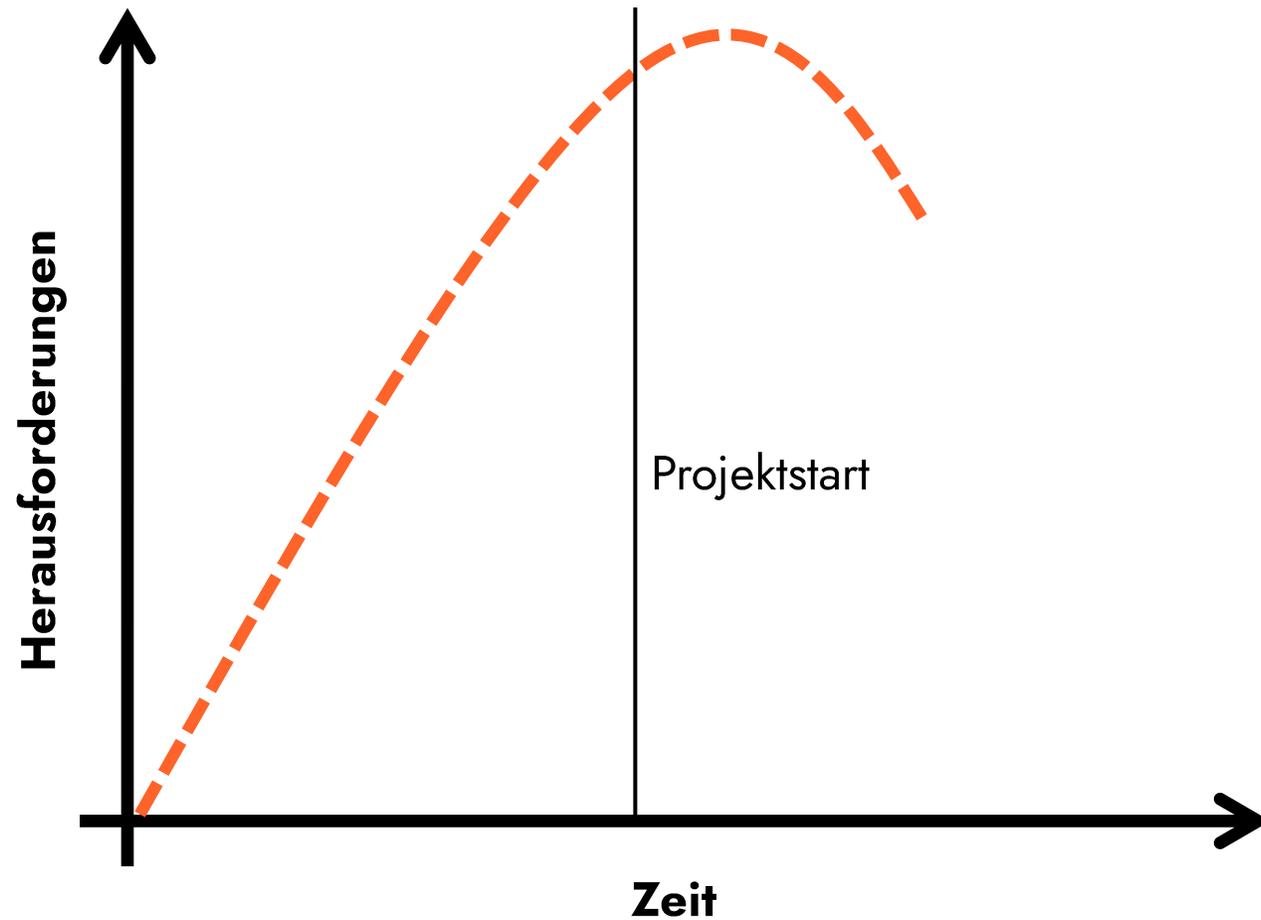
# Projektrahmen

# Projektrahmen

- 2015: Beginn der Konzeption im Archiv der MPG in Kooperation mit dem Fachbereich Informatik der Freien Universität Berlin
- 2017: Einstellung einer studentischen Hilfskraft und Beginn der Entwicklung eines Prototyps
- 2024: Förderprogramm „[ProValid](#)“ der Investitionsbank Berlin
- Aktuelles Kernteam: 2 Informatiker und 1 Archivar



# Die Spitze des Eisbergs



Herausforderungen

# Interdisziplinäre Kommunikation

- Unterschiedliche Herangehensweisen und Fachterminologien im Archivwesen und der Informatik können zu Missverständnissen führen
- Einige archivfachliche Konzepte erschließen sich anderen Fachbereichen nicht immer sofort, weshalb sie teilweise genauer erläutert werden müssen
- Herausforderungen, die zunächst komplex wirken, sind technisch zum Teil sehr einfach zu lösen und andersherum

Bei der Entwicklung sollten alle archivfachlichen Anforderungen so genau wie möglich im Team diskutiert werden, um Missverständnisse vorzubeugen. Um eine enge und dynamische Zusammenarbeit zu gewährleisten, haben sich Herangehensweisen aus der agilen Softwareentwicklung bewährt.



# Technische Aspekte

- Die gängigen E-Mail-Formate MBOX und PST sind für die digitale Langzeitarchivierung ungeeignet
- Der offene E-Mail-Standard, das Alter der Daten und unterschiedliche Zeichenkodierungen erschweren die Verarbeitung der Daten
- Umgang mit signierten und verschlüsselten Mails ist problematisch



Bei der Archivierung von E-Mail-Postfächern müssen zahlreiche technische und organisatorische Barrieren überwunden werden.



# Rechtliche Aspekte

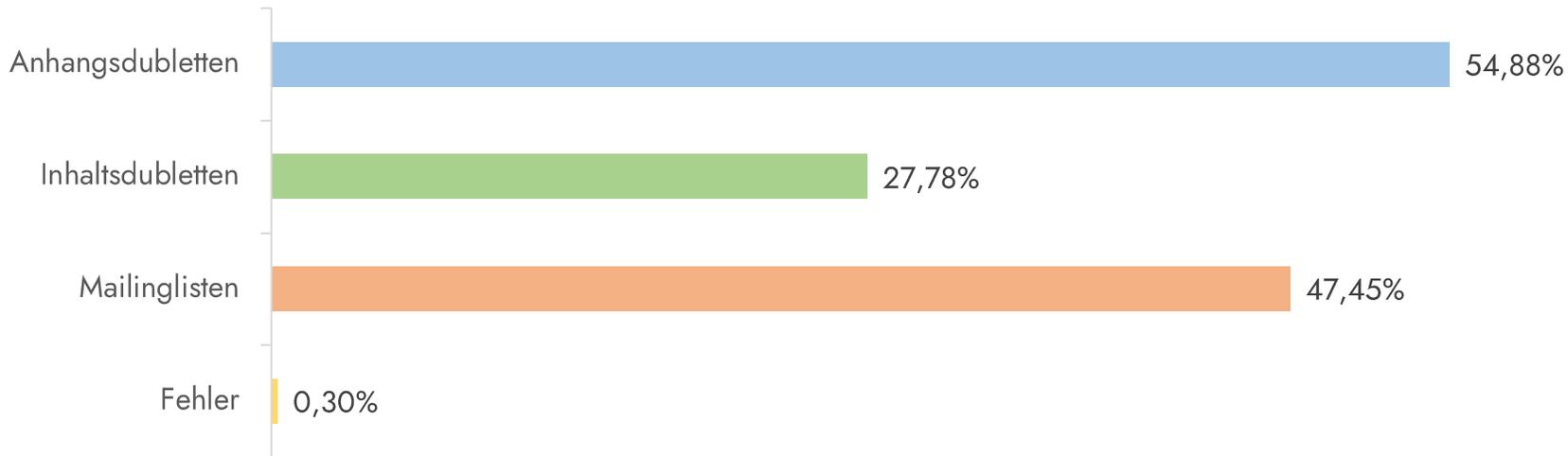
- E-Mails enthalten große Mengen personenbezogener Daten
- E-Mails und E-Mail-Anhänge können urheberrechtlich relevant sein
- Außerhalb der Anbietungspflicht können Vorbehalte gegen die Archivierung von E-Mail-Postfächern bestehen

➔ Vor der Übernahme von E-Mails muss mittels rechtlicher Vorkehrungen eine solide Vertrauensbasis geschaffen werden. Eine zeitnahe und rechtskonforme Nutzbarmachung von E-Mails ist nur mithilfe einer Anonymisierung denkbar.



# Inhaltliche Aspekte

## Statistische Auswertung eines Postfachs mit 133.194 E-Mails und 258.180 Anhängen



Bei der E-Mail-Archivierung haben wir es mit riesigen Datenmengen zu tun. Um eine aussagekräftige Überlieferung bilden zu können, ist eine umfangreiche Bewertung und Datenreduktion erforderlich.

Aktueller Stand

# Funktionsumfang

Erfassung, **M**anagement, **I**ndizierung, **L**imitierung, **i**ntelligente **A**nalyse



Abgabe an das  
Archiv



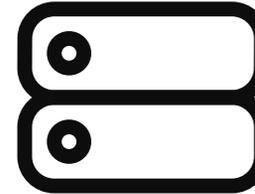
Übernahme

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



Bewertung &  
Erschließung

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



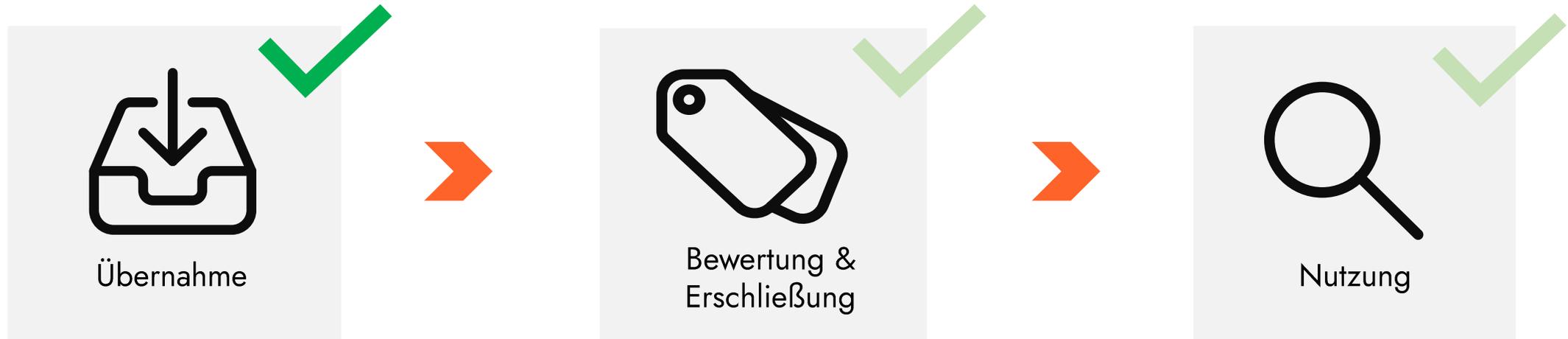
Übergabe an  
digitales  
Langzeitarchiv



Nutzung

- Recherchedatenbank
- Anonymisierung
- Erstellung von DIPs
- Darstellung

# Entwicklungsstand



Wie geht es weiter?

# Bleiben Sie auf dem Laufenden

- Wir arbeiten aktiv an der Implementierung weiterer Funktionen. Um auf dem Laufenden zu bleiben, können Sie unseren Newsletter abonnieren und unsere Webinare besuchen.



# Vielen Dank für Ihre Aufmerksamkeit!

Fragen und Vorschläge sind jederzeit willkommen.

**E-Mail:** [nico.beyer@emilia-archiv.de](mailto:nico.beyer@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](http://www.emilia-archiv.de)

EMILiA The logo icon for EMILiA, featuring a stylized orange envelope with a red dot and a series of horizontal lines to the right.

# Lizenz

- Diese Präsentation kann gemäß der Creative Commons Lizenz [CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/) verwendet werden.
- Die in dieser Präsentation enthaltenen Logos sind von dieser Lizenz ausgenommen und dürfen ohne ausdrückliche Genehmigung des Rechteinhabers nicht weiterverwendet oder verändert werden.