

## Situation

E-Mails sind als Kommunikationsform nicht mehr wegzudenken. Obwohl sie seit über 40 Jahren existieren, ist erst in den letzten Jahren die Notwendigkeit einer Erhaltung dieses flüchtigen Kommunikationsmediums in das Bewusstsein der Menschen gerückt.

Staatliche und private Archive erhalten daher von Unternehmen und Institutionen vermehrt E-Mails als Teil der übergebenen Akten. Diese befinden sich in proprietären Formaten, die für die langfristige Aufbewahrung in quelloffene, archivtaugliche Formate überführt werden müssen. Die Archive haben neben der gesetzlichen Aufbewahrungsfrist für steuerlich relevante Aufzeichnungen auch die Aufgabe, die Daten „für die Ewigkeit“ aufzubewahren.

Das Archiv der sozialen Demokratie (AdsD) der Friedrich-Ebert-Stiftung hat als erstes deutsches Archiv zwei Pilotprojekte zu diesem Thema durchgeführt. Zurzeit plant es ein Projekt zur Übernahme schwach strukturierter Daten. Hierbei sollen insbesondere auch E-Mails berücksichtigt werden.

Zu diesem Zweck ist es sinnvoll, bereits vorhandene E-Mail-Archivierungslösungen zu evaluieren und auf ihre mögliche Einsetzbarkeit in das zu entwickelnde Ingest-Modul zu testen. Da das neue Archivsystem XML-basiert sein soll, ist es ratsam, die E-Mails in dieses Format zu überführen.



Abb. 1: Weg ins Archiv

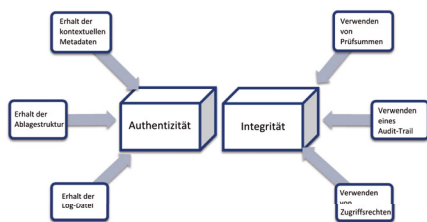
## Methode



## Durchführung

### Anforderungen an die E-Mail-Archivierung

- Wesentliche Anforderungen an die E-Mail-Archivierung sind der Erhalt der Authentizität und der Integrität



- Authentizität beschreibt die Eigenschaft eines Objektes genau das zu sein, was es vorgibt zu sein. Dies wird im analogen Bereich beispielsweise durch Unterschriften, Stempel oder Siegel erreicht
- Integrität beschreibt die Eigenschaft eines Objektes unverändert und vollständig zu sein.

Abb. 2: Anforderungen an die E-Mail-Archivierung

- Verschiedene Schritte beweisen die Authentizität und Integrität von E-Mails (s. Abb. 2).

### Testen und Analysieren vorhandener Software-Tools in Hinblick auf die Erfüllung der Anforderungen

- Getestet wurden 8 verschiedene Software-Tools für die Umwandlung von E-Mails in ein archivtaugliches XML-Format.
- Auf Grund der starken Dominanz von Microsoft Outlook unter den Desktop-Clients (s. Grafik) wurde der Fokus der Versuche auf das von Outlook verwendete PST-Format gelegt.
- Klare Sieger der Testreihe waren zum Einen der **CERP Email Parser** des Collaborative Electronic Records Projektes und zum Anderen der **PeDALS Email Extractor** des Persistent Digital Archives and Library System Projektes.
- Der CERP Email Parser bietet durch die Verwendung des MBOX-Formates als Ausgangsformat eine größere Flexibilität.



Abb. 3: Verteilung von E-Mails-Clients (<http://www.campaign-monitor.com/resources/will-it-work/email-clients>)

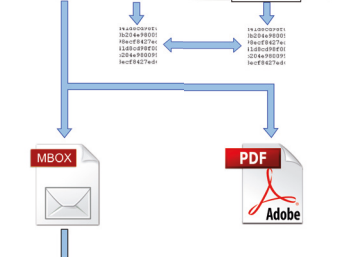
## Ergebnis

### Modell eines Ingestprozesses mit Hilfe des CERP Email Parser

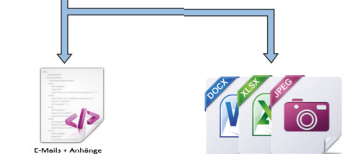
Ausgangslage: Abgabe von E-Mails in unterschiedlichen Formaten  
Erstellen einer exakten Kopie und überprüfen dieser über Prüfsummen



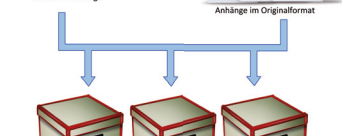
Umwandeln der Ausgangsformate in MBOX-Dateien (eine MBOX-Datei für jeden Ordner des E-Mails-Postfaches) beispielsweise mit dem Tool Gegebenenfalls Erstellen eines PDF/A-Abbildes der E-Mails als „Visual Aid“



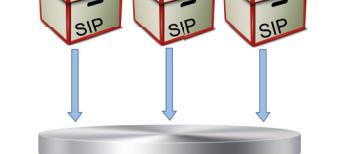
Umwandeln der MBOX-Dateien in archivtaugliche XML-Dateien mit Hilfe des CERP Email Parsers  
Extraktion der Anhänge in ihrem Originalformat zum leichteren Zugriff



Verpacken der Objekte in einen XML-Wrap mit den nötigen Metadaten



Übergabe an das Archivsystem



## Fazit

- Für den Erhalt von E-Mails sind die Bewahrung von Authentizität und Integrität unabdingbar
- XML hat sich als Speicherformat von E-Mails als am Vorteilhaftesten erwiesen
- Bei einer Versuchsreihe erwies sich der **CERP Email Parser** zusammen mit dem **PeDALS Email Extractor** als am Besten für die Umwandlung von proprietären Formaten in XML als am Zufriedenstellendsten

## Literaturangaben

- Prom, Christopher J. (2011): Preserving Email - DPC Technology Watch Report 11-01. Online verfügbar unter [http://www.dpconline.org/component/docman/doc\\_download/739-dpctw11-01.pdf](http://www.dpconline.org/component/docman/doc_download/739-dpctw11-01.pdf), zuletzt geprüft am 03.03.2012.
- Zuchet, Mike (2011): Pilotprojekte zur Langzeitarchivierung digitaler Email-Korrespondenz im Archiv der sozialen Demokratie (AdsD). In: Archiv-Nachrichten. Internet-Newsletter aus dem Archiv der sozialen Demokratie. Online verfügbar unter [http://www.fes.de/archiv/adsd\\_neu/newsletter/01\\_2011/zuchet.html](http://www.fes.de/archiv/adsd_neu/newsletter/01_2011/zuchet.html), zuletzt geprüft am 13.08.2012.

## Kontakt

Sina Haase  
E-Mail: [sinahaase@aol.com](mailto:sinahaase@aol.com)

Studiengang Konservierung Neuer Medien und Digitaler Information  
<http://www.mediaconservation.abk-stuttgart.de>