

# Absichtserklärung (LOI) für das DFG-Projekt „Implementierung der OCR-D-Software zur Volltextdigitalisierung historischer Drucke“

Bezugnehmend auf:

[https://www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung\\_ocr\\_implementierung.pdf](https://www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung_ocr_implementierung.pdf)

## Ziel

Das nachfolgend vorgestellte Konsortium möchte die bisher für Linux verfügbaren Bibliotheken von OCR-D in das vollständige Dokumenten-Management-System „bitfarm Archiv“ für Windows des Unternehmens „bitfarm Informationssysteme GmbH“ (kurz: bitfarm) integrieren. Durch die Integration in ein bereits verbreitetes und kostenlos, inkl. Source-Code, unter der GPLv2-Lizenz verfügbares Dokumenten-Management-System (kurz: DMS) wird OCR-D nicht nur in ein bereits am Markt etabliertes Produkt integriert, sondern profitiert gleichzeitig von einer nutzerevaluierten Oberfläche und zahlreichen Funktionen zur weiteren Verarbeitung der Dokumente.

Nach der Integration von OCR-D in „bitfarm Archiv“ wird der Prototyp für die Nutzung in der Praxis von verschiedenen Pilotkunden getestet. Hierzu verfügt bitfarm bereits über einen großen Kundenstamm, aus deren Reihen geeignete Pilotkunden mit Fokus auf die Be- und Verarbeitung frühzeitlicher Dokumente akquiriert werden sollen. Im Zuge dieses Piloteinsatzes wird gleichzeitig versucht, nicht nur Drucke zu evaluieren, sondern ebenfalls historische handschriftliche Dokumente. Dies bringt einen zusätzlichen Gewinn für OCR-D.

Der praktische Einsatz des Prototyps wird während des Projekts durch die „Juniorprofessur Wirtschaftsinformatik, insb. IT für die alternde Gesellschaft“ der Universität Siegen mittels sozio-informatischer Perspektive evaluiert. Hierbei wird der Fokus auf die Nutzbarkeit des Prototyps und der neuen Funktionen sowie die sich daraus verändernde Arbeitspraxis gelegt. Durch die Begleitung soll gewährleistet werden, dass die Implementierung den Anforderungen an die Praxis genügt, das System von den Mitarbeitenden bedient werden kann und somit ein nachhaltiger Einsatz des Prototypen ermöglicht wird.

Ein weiterer Aspekt in diesem Konsortium ist der Blick auf Barrierefreiheit, mit dem bitfarm und die Universität Siegen bereits über ein anderes Projekt in Kontakt stehen. Um Arbeitsplätze für Blinde und Menschen mit Sehbehinderung zu erhalten und zu schaffen soll die Integration der frühzeitlichen Schriftenerkennung in das DMS in einer für Blinde und Menschen mit Sehbehinderung geeigneten Art und Weise durchgeführt werden.

Die anvisierten Ziele des Projekts werden damit erfüllt und um einige Aspekte erweitert:

- Integration in ein vollständiges und frei verfügbares DMS (Open-Source)
- Erschließung der Windows-Welt
- Praxiseinsatz und -test durch Pilotkunden
- Wissenschaftliche Evaluation des Prototypen, der Arbeitspraktiken und des nachhaltigen Einsatzes von OCR-D
- Zusätzliche Evaluation von Handschriften
- Zusätzlicher Fokus auf Arbeitsplätze für Blinde und Menschen mit Sehbehinderung

## Konsortium

Das Konsortium besteht aus der „bitfarm Informationssysteme GmbH“, welche das DMS für die Implementierung bereitstellt, die Implementierung durchführt und den Zugang zu entsprechenden Pilotkunden hat.

Die „Juniorprofessur Wirtschaftsinformatik , insb. IT für die alternde Gesellschaft“ der Universität Siegen übernimmt in seiner Rolle als Experte für nutzerzentriertes, praxisorientiertes Mediendesign und partizipativer Aktionsforschung sowie der sozio-informatischen Perspektive auf die IT-Gestaltung die wissenschaftliche Begleitung des Prototyps, der Einführung sowie der nachhaltigen Nutzung von OCR-D.

Für die zielorientierte Aufgabenplanung und -festlegung steht das Konsortium bereits mit dem Koordinierungsprojekt in Kontakt und kontinuierlichem Austausch.

## Für die Pilotierung geplante Arbeiten

Zu Sicherstellung der Verfügbarmachung der OCR-D-Integration auch nach dem Projekt müssen noch Voraussetzungen geprüft werden. So ist aktuell noch keine „out of the box“-Lösung verfügbar, um die Linux-Komponenten von OCR-D auf einem Windows-System zu nutzen. Der derzeitige Prototyp verwendet das „Subsystem für Linux“ (WSL, ab Windows 10), wozu es im Rahmen einer kostenlos bereitgestellten Version des DMS noch keine Lösung gibt, die dieses eingerichtete Subsystem mit „ausliefern“ kann.

Weiterhin ist zusammen mit dem Koordinationsprojekt noch die Auswahl der zu integrierenden Module (Prozessoren, Workflows) abzustimmen.

## Absichtserklärung

Die hier genannten Parteien bekunden hiermit ihr Interesse und die Absicht, das DFG-Projekt „Implementierung der OCR-D-Software zur Volltextdigitalisierung historischer Drucke“ in der hier beschriebenen Art und Weise durchführen zu wollen. Die Absicht wird vorbehaltlich der späteren Projektförderung und der in der Pilotierung noch zu klärenden Rahmenbedingungen und Voraussetzungen erklärt.

bitfarm Informationssysteme GmbH  
Spandauer Str. 18  
57072 Siegen

---

Universität Siegen, Fakultät III  
Juniorprofessur Wirtschaftsinformatik/IT für die alternde Gesellschaft  
Kohlbettstraße 15  
57072 Siegen

