Für die Digitalen Geisteswissenschaften (Digital Humanities) ist die OCR (Optical Character Recognition) [[1]](#footnote-0) von großer Relevanz, da sie es erlaubt, eine der wichtigsten Ressourcen der Digitalen Geisteswissenschaften – Text, das schließt Kunsthistoriker aus, auch in Lit.-wissenschaften sind Bilder und Noten in manchen Kontexten (Minnesang) relevant – effizient zu erschließen und in maschinenverarbeitbarer Form zu präsentieren. Diese Bedeutung zeigte sich auch auf der DHd Jahrestagung 2019, auf der sowohl die beiden angebotenen OCR-Workshops, als auch die Session „Schrift und Zeichen“ großen Zulauf erhielten. Aufgrund der Relevanz für das Feld der DH und dem offenkundigen Interesse innerhalb der deutschsprachigen DH Community, wird hiermit die Einrichtung einer DHd Arbeitsgruppe zum Thema OCR beantragt.

Die Idee für eine AG entstand auf einem OCR-Workshop am 14.06.2019 an der Universitäts- und Landesbibliothek Sachsen-Anhalt in Halle/Saale. Die dortigen Teilnehmer spiegeln die erhoffte heterogene Zusammensetzung der zu gründenden AG hervorragend wieder:

* Nutzer, die bewährte, einfach zu bedienende aber dennoch mächtige Tools und Technologien für verschiedenste Zwecke einsetzen wollen.
* Vermittler (u. a. die Leiter beider OCR Workshops der letzten DHd Jahrestagung), deren Ziel es ist diese Tools und Technologien den Nutzern in Workshops und anderen Lehrveranstaltungen näher zu bringen.
* Entwickler und Forscher (u. a. Vertreter des DFG geförderten Verbundprojekts OCR-D und des Würzburger Zentrums für Philologie und Digitalität „Kallimachos“), die sowohl mit der Optimierung bestehender Software und Technologien wie auch vollkommenen Neuentwicklungen beschäftigt sind.

Ausgehend von dieser breit aufgestellten Basis an Mitgliedern setzt sich die AG folgende Ziele:

* Vernetzung von Nutzern, Vermittlern, Entwicklern und Forschern durch mindestens jährliche Treffen sowie Workshops und gemeinsame Projekte mit dem Fokus auf freier und quelloffener Software beziehungsweise Publikationen zum Thema.
* Erarbeitung, Sicherung und Verbreitung von Best Practices zu praktischen Anwendungsfragen der OCR, z.B. Workflow, Formate, Parameter
* Technologietransfer aus der Entwicklung in die Praxis.
* Klare Kommunikation spezifisch geisteswissenschaftlicher Anforderungen den Entwicklern von OCR-Software, damit diese sich noch stärker als bisher auf die Programmierung von Software für den Produktivbetrieb statt wie bisher vielfach auf die Erstellung von Protoytpen konzentrieren.
* Identifikation und kollaborative Bearbeitung offener Forschungsfragen auch im engen Austausch mit bestehenden bereits bestehenden AGs der DHd wie z. B. der „AG Zeitungen & Zeitschriften“.

1. Auch wenn die Verwendung des Ausdruckes OCR (= ›Optical Character Recognition‹) nicht mehr korrekt ist und von ›optischer Abschnittserkennung‹ gesprochen werden muß, wird der Einfachheit und besseren Allgemeinverständlichkeit halber nachfolgend gleichwohl der Ausdruck OCR für aktuelle Verfahren der Digitalisierung gedruckter und geschriebener Texte benutzt. [↑](#footnote-ref-0)