

# Weisung 202112016 vom 14.12.2021 – E-AKTE: Bereitstellung von OCR-erkanntem Text bei Dokumenten aus der externen Digitalisierung von BA- Schriftgut

**Laufende Nummer:** 202112016

**Geschäftszeichen:** IT-AFM 12 - 1460 / 1461 / 1462 / 1463 / 3404 / 3313 / 3317 / II-5217.5 / 7010 / 8526.1

**Gültig ab:** 01.01.2022

**Gültig bis:** unbegrenzt

**SGB II:** Weisung - Relevanz §50 Abs. 3 SGB II

**SGB III:** Weisung

**Familienkasse:** Weisung

## **Bezug:**

- Weisung 201809020 vom 20.09.2018 – Abschließende Regelungen nach Einführung der E-AKTE im Rechtskreis SGB II auf Basis § 50 Abs. 3 SGB II

## **Aufhebung von Regelungen:**

- entfällt

---

Bei den aus dem Scanprozess beim externen Scandienstleister hervorgehenden digitalisierten Dokumenten wird zukünftig ein durch OCR-Erkennung gewonnener Text mitgeliefert. Mit diesem wird das Durchsuchen von PDF-Dokumenten sowie das Selektieren und Kopieren von Worten und Textpassagen in den PDF-Dokumenten und das Vorlesen der Dokumente mit entsprechender Arbeitsplatzausstattung ermöglicht. Bei einer Weiterverwendung ist der OCR-erkannte Text auf Richtigkeit und Vollständigkeit zu prüfen.

## 1. Ausgangssituation

Dokumente, die durch den externen Scandienstleister digitalisiert werden, werden als Bilddateien im PDF-Format in der E-AKTE gespeichert. Der in den Dokumenten enthaltene Text kann dadurch nicht für die weitere Bearbeitung genutzt werden.

## 2. Auftrag und Ziel

Im Rahmen der Digitalisierung wird durch den Scandienstleister zukünftig eine OCR-Volltexterkennung durchgeführt. Der erkannte Text wird als zusätzliches Metadatum an die E-AKTE übergeben und kann dann genutzt werden. Das PDF-Dokument wird wie bisher als Bilddatei in der E-AKTE angezeigt. Der OCR-Text wird im PDF-Dokument für Anwenderinnen und Anwender nicht sichtbar über das Dokument gelegt.

Durch den OCR-Text wird das Durchsuchen sowie das Selektieren und Kopieren von Worten und Textpassagen in den PDF-Dokumenten ermöglicht. Auch das Vorlesen der Dokumente wird mit dem OCR-Text bei entsprechender Arbeitsplatzausstattung möglich.

Die flächenweite Bereitstellung des OCR-Volltextes erfolgt in 2 Schritten:

- Mandant SGB III ab dem 03.01.2022
- Mandant SGB II ab dem 01.02.2022
- Mandant FamKa ab dem 01.02.2022

## 3. Einzelaufträge

Die Anwenderinnen und Anwender aller Rechtskreise prüfen bei einer Weiterverwendung (z. B. beim Kopieren und Einfügen) den OCR-erkannten Text auf Richtigkeit und Vollständigkeit.

Die gemeinsamen Einrichtungen berücksichtigen die zusätzlichen Kosten bei der Haushaltsplanung.

## 4. Info

OCR steht für Optical Character Recognition. Mit dieser Technik wird eine maschinelle Texterkennung durchgeführt. Eine manuelle Nachbearbeitung des erkannten Textes erfolgt nicht. Die Datenverarbeitung erfolgt wie bisher unter Beachtung einer angemessenen Sicherheit für die personenbezogenen Daten (Integrität, Vertraulichkeit, Richtigkeit, Vollständigkeit). Die originären Daten in der E-AKTE, also die PDF-Dokumente und die Metadaten, werden durch die OCR-Erkennung nicht verändert.

Ausschlaggebend für eine hohe Texterkennungsrate ist die Qualität des zu digitalisierenden

Schriftgutes. Aufgrund der sehr heterogenen Qualität des Schriftgutes in der BA ist auch die Qualität der Texterkennung nicht konstant, es kann keine vollständig richtige Erkennung gewährleistet werden. Die Erkennung von handschriftlichen Texten ist nur bedingt möglich und nach aktuellem Stand in der Massenverarbeitung bei der externen Digitalisierung des BA-

Schriftgutes durch den Scandienstleister nicht optimierungsfähig, so dass keine Verbesserungen in Aussicht gestellt werden können.

Dies hat zur Folge, dass die Anwenderinnen und Anwender sich nicht ohne weitere Prüfung auf die Richtigkeit der OCR-Erkennung verlassen können. Der für ein Kopieren und Einfügen („Copy&Paste“) genutzte Text ist bei der weiteren Verwendung durch die Anwenderinnen und Anwender stets auf Richtigkeit und Vollständigkeit (Übereinstimmung mit dem sichtbaren PDF-Dokument) zu prüfen. Bei einer Suche in Dokumenten sollte bei einem negativen Suchergebnis im Zweifelsfall noch eine zusätzliche visuelle Prüfung erfolgen. Trotz dieser Einschränkungen wurde in Erprobungsphasen für alle Mandanten ein relevanter Nutzen erkannt, z. B. die Reduzierung von Übertragungsfehlern und eine Reduzierung des Zeitaufwandes bei einzelnen Bearbeitungsschritten. Denkbare Anwendungsfälle sind z. B. das Kopieren von Fachschlüsseln/Kundennummern, Versicherungsnummern, IBAN, Kundennamen oder Betriebsbezeichnungen zur Verwendung in anderen Fachverfahren oder das Kopieren von Textteilen aus Stellungnahmen zur Verwendung in BK. Auch eine Suche nach Namen in umfangreichen Listen wird vereinfacht.

Bei dem IT-Verfahren E-AKTE Mandant SGB II handelt es sich um ein Verfahren nach § 50 Abs. 3 SGB II. Das BMAS hat die Weisung freigegeben.

## 5. Haushalt

Für die Einführung der OCR-Texterkennung entstehen im Rechtskreis SGB II Initialkosten von insgesamt 117.780,32 € netto = 140.158,58 € brutto. Die Aufteilung auf die einzelnen gemeinsamen Einrichtungen im Verwaltungskostennachweis (VKN) erfolgt entsprechend dem Anteil der gemeinsamen Einrichtung an der Gesamtdigitalisierungsmenge unter Berücksichtigung der durchschnittlichen Vorjahresmenge.

Daneben entstehen durch die OCR-Volltexterkennung zusätzliche laufende Kosten pro digitalisiertem Blatt in Höhe von 0,00045 € netto = 0,00054 € brutto. Diese werden der BA durch den Scandienstleister über einen geänderten Blattpreis in Rechnung gestellt. Im Mandanten SGB II führt dies dazu, dass den einzelnen gemeinsamen Einrichtungen im Rahmen der Abrechnung der Aufwendungen im VKN höhere Kosten entstehen. In der Kalkulationshilfe E-AKTE SGB II zur Haushaltsplanung ist ab der Version für das Jahr 2021 eine entsprechende Position enthalten.

## 6. Beteiligung

Der Hauptpersonalrat und die Hauptschwerbehindertenvertretung wurden beteiligt.

gez.

Unterschrift