

Datenblatt zu Neu- und Fortsetzungsanträgen im Programm „Digitalisierung und Erschließung“

Projekttitel: VD-Volltext: Zentrale OCR-D-basierte Erschließung der Volltexte der digitalisierten Drucke aus dem deutschen Sprachraum des 16. (VD 16), 17. (VD 17) und 18. Jahrhunderts (VD 18)

A. Allgemeine Angaben		Kommentare
<i>Bitte für jede antragstellende Einrichtung separat ausfüllen.</i>		
Name der antragstellenden Einrichtung	Herzog August Bibliothek Wolfenbüttel (HAB)	
Träger	Land Niedersachsen	
Name der antragstellenden Einrichtung	Staatsbibliothek zu Berlin (SBB)	
Träger	Stiftung Preußischer Kulturbesitz	
Name der antragstellenden Einrichtung	Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB)	
Träger	Georg-August-Universität Göttingen - Stiftung Öffentlichen Rechts	
Name der antragstellenden Einrichtung	Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)	
Träger	Gesellschafter: Max-Planck-Gesellschaft; Georg-August-Universität Göttingen - Stiftung Öffentlichen Rechts	

B. Projektgesamtkosten (aufgeteilt nach beantragten Mitteln und Eigenleistungen) Kommentare

Bitte die Mittel getrennt nach antragstellender Einrichtung auflisten.

Einrichtung	Arbeitspaket	Kostenart*	Bei der DFG beantragte Mittel		Eigenleistungen		Mittel in Euro	Mittel in Euro		
			ggf. PM / Stellenanteil / Eingruppierung	Mittel in Euro	Kostenart**	ggf. PM / Stellenanteil / Eingruppierung				
HAB	Arbeitspaket 1	Personal	18 PAM, E13	132.300 €	Personal	A16, E14, WHK	67.880 €			
SUB	Arbeitspaket 1	Personal	18 PAM	132.300 €	Personal	E13, E13, E13	64.340 €			
HAB	Arbeitspaket 2	Personal	18 PAM, E13	132.300 €	Personal	A16, E14, WHK	67.880 €			
SUB	Arbeitspaket 2	Personal	18 PAM	132.300 €	Personal	E8	73.032 €			
SUB	Arbeitspaket 3	Personal	18 PAM	132.300 €	Personal	E13	68.124 €			
SBB	Arbeitspaket 3	Personal	6 PAM, E13	44.100 €	Personal	E13	22.710 €			
GWDG	Arbeitspaket 3	Personal	18 PAM	132.300 €	Personal	E13	35.779 €			
SUB	Arbeitspaket 4	Personal	18 PAM	132.300 €	Personal	E13	68.124 €			
GWDG	Arbeitspaket 4	Personal	18 PAM	132.300 €	Personal	Diverse	103.000 €			
SBB	Arbeitspaket 5	Personal	18 PAM, E13	132.300 €	Personal	E13	68.130 €			
HAB		Reisen		900 €						Kick-off / Workshop 1
SUB		Reisen		900 €						Kick-off / Workshop 1
GWDG		Reisen		900 €						Kick-off / Workshop 1
GWDG		Reisen		900 €						Workshop 2
SBB		Reisen		900 €						Workshop 2
SUB		Reisen		900 €						Workshop 2
HAB		Reisen		900 €						Workshop 3
SBB		Reisen		900 €						Workshop 3
HAB	Arbeitspaket 1	Workshop		2.400 €						Kick-off / Workshop 1
HAB	Arbeitspaket 1	Workshop		2.400 €						Workshop 2
HAB	Arbeitspaket 1	Workshop		2.400 €						Workshop 3
GWDG		HPC-Rechenleistung		126.975 €	Infrastruktur	/.	62.000 €			
Gesamtsumme der Mittel:				1.376.175 €	Gesamtsumme der Mittel:		700.999 €	Alle Personalkosten (Eigenanteil) inkl. 3% Personalkostensteigerung		

* Bspw. Personal, Reisen, Publikationen, Projektspezifische Workshops, Geräte, Sonstiges. Bitte orientieren Sie sich für die bei der DFG beantragten Mittel an den Personalmittelsätzen der DFG (DFG-Vordruck 60.12).

** Bspw. Personal, Reisen, Publikationen, Projektspezifische Workshops, Geräte, Sonstiges. Bitte orientieren Sie sich bei den in Eigenleistung eingebrachten Posten an Ihren lokalen Personalmittelsätzen.

C. Angaben zur Digitalisierung		Kommentare
<i>Bei mehreren Beständen bitte - wenn erforderlich - pro Bestand eine Tabelle ausfüllen. Bei der Berechnung der Kosten sollten sowohl die bei der DFG beantragten Mittel als auch die in Eigenleistung erbrachten Mittel mit einbezogen werden.</i>		
Befindet sich der Bestand in öffentlich-rechtlicher Trägerschaft?	Ja	
Wurden die rechtlichen Voraussetzungen für die zu digitalisierenden Objekte geklärt?	Ja	Gemeinfrei
Materialart	Drucke des 16.-18. Jahrhunderts	Bitte geben Sie an, um welche Materialart(en) es sich bei dem/den zu digitalisierenden Bestand/Beständen handelt. (bspw. Handschriften, Briefe, Tonkulpturen, Muscheln, Tonbänder, ...)
Gesamtzahl der zu digitalisierenden Objekte	Zu erstellende Volltexte für 368.000 Drucke	Bitte beziffern Sie die Gesamtzahl der zu digitalisierenden Objekte, ggf. differenziert nach Beständen.
Gesamtzahl der zu erstellenden Digitalisate	48 Mio Seiten Volltext	Bitte geben Sie an, wie viele Digitalisate erstellt werden. Hinweis: Für die Digitalisierung eines Objektes kann die Erstellung mehrerer Digitalisate notwendig sein. (bspw. mehrere Buchseiten pro Handschrift, mehrere Ansichten pro Skulptur, ...)
An welchen technischen Richtlinien (national/international) orientiert sich die Digitalisierung? Werden neue Digitalisierungsformate und -standards entwickelt?	DFG-Praxisregeln „Digitalisierung“	https://www.dfg.de/resource/blob/176108/12-151-v1216-de.pdf-Fassung_2022 https://zenodo.org/records/7435724
Kosten pro digitalisiertem Objekt (einschließlich Strukturdatenerfassung)	0,0026 € pro prozessierender Seite (auf vier Nachkommastellen gerundet). Als Basis für die Berechnung werden die 48.576.000 Seiten gemäß Tabelle 1 im Antrag gewählt.	Bitte geben Sie die gesamten Kosten pro Objekt an, d.h. beziehen Sie sowohl die beantragten als auch die in Eigenleistung erbrachten Mittel in die Berechnung mit ein.
Sind über die Image-Digitalisierung hinausgehende weitere Nachbearbeitungen und/oder Analysen vorgesehen?	Keine Image-Digitalisierung, ausschließlich Volltext-Erkennung	Beispielsweise Volltexterkennung, Bildererkennung o.Ä.

Erfolgt die Digitalisierung inhouse oder durch Dienstleister?	Inhouse-Volltexterkennung OCR-D	Bilddigitalisierung ist bereits erfolgt. Im hier beantragten Projekt steht die Erstellung von Volltexten im Mittelpunkt der Arbeiten
Inhouse-Digitalisierung	Inhouse-Volltexterkennung OCR-D	
Welche Scanner werden verwendet? Welche Digitalisierungsverfahren kommen zum Einsatz?	keine Scanner, nur Volltext-Erkennung	
Durchsatz pro Scanner und Stunde	keine Angabe, es wird nicht gescannt	
Mit welcher Workflowssoftware wird gearbeitet?	OCR-D	
D. Angaben zur Erschließung		Kommentare
Befindet sich der Bestand in öffentlich-rechtlicher Trägerschaft?	Ja	
Materialart	Drucke des 16.-18. Jahrhunderts	Bitte geben Sie an, um welche Materialart(en) es sich bei dem/den zu erschließenden Bestand/Beständen handelt. (bspw: Handschriften, Briefe, Tonskulpturen, Muscheln, Tonbänder, ...)
Gesamtzahl der zu erschließenden Objekte	48 Mio Seiten (Volltext)	Bitte beziffern Sie die Gesamtzahl der zu erschließenden Objekte, ggf. differenziert nach Beständen.
An welchen Erschließungsrichtlinien (national/international) orientiert sich die Erschließung? Werden neue Erschließungsstandards entwickelt?	Erschließung ist nicht Teil des Projekts, ggf. Resource Description and Access (RDA), ggf. VD 16-, VD 17- und VD 18-Erschließungsregeln	
Welche Normdaten/Thesauri werden verwendet?	keine Angabe	
Wie viele Aufnahmen pro Tag und Mitarbeiter/in werden erzielt?	Es werden keine Aufnahmen durch Mitarbeitende intellektuell erstellt.	Bitte nennen Sie die durchschnittliche Anzahl der pro Tag/Mitarbeiter*in vollständig aufgenommenen bzw. erschlossenen Objekte.
Kosten pro Aufnahme	nicht zutreffend	Erweiterung der METS-Dateien, keine Ersterschließung
E. Angaben zur Bereitstellung und Nachweis der Projektergebnisse		Kommentare
Bereitstellung in internationalen Portalen (bspw. Europeana)	Europeana	wird durch Schnittstellen unterstützt
Bereitstellung in nationalen Portalen (bspw. DDB)	VD-Portal (in Beantragung); VD 16; VD 17, VD 18; OLA-HD;	
Bereitstellung in material- und/oder fachspezifischen Portalen	grundsätzlich möglich	wird durch Schnittstellen unterstützt
Bereitstellung auf der Webseite (Webadresse)	https://ocr-d.de/	Bereitstellung durch VD-Portal und bereitstellende Einzeleinrichtungen
Welche Präsentationssoftware/welcher Viewer wird verwendet?	verschiedene Viewer (IIIF-Technologie)	Viewer des VD-Portals, diverse Viewer der bereitstellenden Einzeleinrichtungen
Gibt es eine normierte Sammlungs- bzw. Bestandsbeschreibung in einem standardisierten Format?	Beschreibungen zu den Einzel-VD liegen vor	VD 16: u.a. Gisela Moncke: Uas „Verzeichnis der im deutschen Sprachbereich erschienenen Urucke des 16. Jahrhunderts“ (VD 16) als Teil einer deutschen retrospektiven Nationalbibliographie. In: Zeitschrift für Bibliothekswesen und Bibliographie. Bd. 51, Heft 4, 2004, ISSN 0044-2380, S. 207–212. - VD 17 u.a. Gerhard Lauer, Saskia Limbach, Christoph Reske, Michaela Scheibe, Nikolaus Weichselbaumer: Zukunft der VD – Vision einer forschungs-adäquaten
In welchem Umfang können Datensätze aus rechtlichen Gründen bei Projektende noch nicht im Open Access bereitgestellt werden? *	keine (0%)	Bitte beziffern Sie den Umfang der Datensätze, die NICHT im Open Access bereitgestellt werden können, prozentual.
* Maximal 5% des zu digitalisierenden Bestandes darf noch Schutzfristen unterliegen. Dieser Teil kann nur in Eigenleistung bearbeitet werden, nicht mit DFG-Förderung.		
F. Angaben zum Datenmanagement/zur Langzeitarchivierung		Kommentare
In welchem Format erfolgt die Speicherung der digitalen Master?	Es werden keine neuen digitalen Master erstellt	Nutzung vorhandener digitaler (Bild-)Daten
In welchem Format werden die Volltexte bereitgestellt?	METS als Container Format und PAGE-XML für Layout Erkennung und Texterkennungsergebnisse	Andere OCR Formate wie ALTO, hOCR oder ABBYY FineReader XML werden auch unterstützt
In welchem Format werden die deskriptiven Metadaten bereitgestellt?	JSON	Manifest-Dateien
Welche Schnittstellen werden bedient?	OAI-PMH für Metadaten, IIIF für Bildzugriff, OAI-PMH-, IIIF- und METS-Schnittstellen	
Wie erfolgt die Vergabe der persistenten Identifier?	Über das ePIC Konsortium (https://www.pidconsortium.net/) werden Handle-basierte Persistent Identifier vergeben (kompatibel mit den in NFDI verwendeten PIDs gemäß Basisdienstprojekt PID4NFDI).	
Mit welchen standardisierten Rechten (Public Domain, CC-Lizenzen etc.) werden die einzelnen Projektergebnisse (Metadaten, Digitalisate, Volltexte, neue Digitalisierungs- und Erschließungsformate, Software) versehen?	Public Domain, CC0	