



OCR-D

Koordinierte Förderinitiative zur Weiterentwicklung
von Verfahren der Optical Character Recognition (OCR)

OCR-D in a Nutshell

12.02.2020

Elisabeth Engl



Auftrag

2

- Entwickeln von Lösungen für den Lückenschluss zwischen Forschung und Praxiseinsatz im Bereich der automatischen Texterkennung
- Anwendung von OCR auf historische Materialien von VD16, VD17, VD18
- Aus dem Auftrag laut DFG-Ausschreibung vom 28. Mai 2014:
 - Optimierung und Vereinheitlichung der Volltexterstellung auf Basis von Image-Digitalisaten
 - Standardisierbare, offene Prozesse etablieren
 - Erstellung von Referenzkorpora [Ground Truth] unterstützen
 - Konzept für die Interaktion der Module erstellen
 - DFG- Digitalisierungsrichtlinie weiter entwickeln



Aufgaben und Deliverables im Koordinationsprojekt

3

Koordination:

- der Modulprojekte, Pilotbibliotheken, Community und Wissenschaftl. Beirat
- untereinander (Wiki, Videokonferenzen)

Integration:

- Schnittstellen
- GitHub

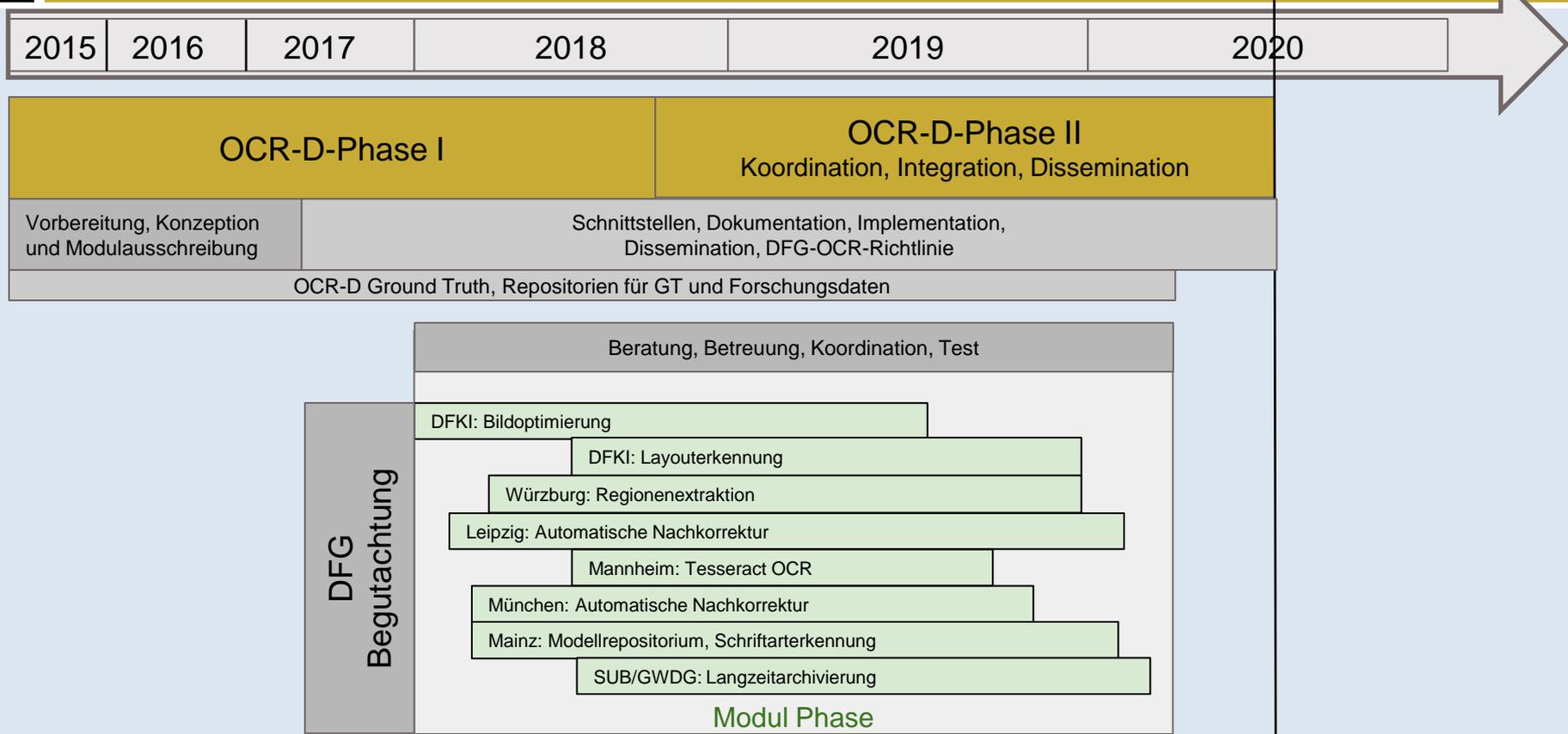
Dissemination:

- Webseite, Gitter (Chat)
- Publikationen
- Konferenzen (DATeCH, Bibliothekskongress, DHd-Jahrestagung)
- Workshops
- Pilotbibliotheken



Zeitstrahl

4





Etappen des Projekts

5

03/2018

Kick-Off-Workshop

06/2018

1. Entwicklerworkshop mit den MP
(Vorstellung der Schnittstellen)

02/2019

2. Entwicklerworkshop mit den MP
(Stand der MP, Vorstellung Ground-Truth-Repository)

03/2019

Workshop mit Anwendern auf der DHd 2019

03/2019

7. Bibliothekskongress in Leipzig: Offene Arbeitssitzung mit Dienstleistern und Anwendern sowie Vortrag zum Projekt in der Sektion "Texterkennung: aktuelle Entwicklungen"



Etappen des Projekts

6

- 05/2019 2 Vorträge auf der DATeCH2019 (**Best Paper Award** für OCR-D)
- 08/2019 3. Entwicklerworkshop mit den MP
- 09/2019 Vortrag bei der ICDAR 2019
- 12/2019 Teststellung der Modulprojektergebnisse
- 02/2020 Workshop zur Vorstellung der (vorläufigen) Ergebnisse



Meilensteine 2020

7

- 03/2020 Übernahme der Modulprojekte und Integration in OCR-D-Framework weitgehend abgeschlossen
- 03/2020 Dokumentation OCR-D-Gesamtsystem (Spezifikationen, Infrastrukturanforderungen, Installation, Nutzer guide) weitgehend abgeschlossen
- 05/2020 Konzept für massenhafte Volltextdigitalisierung der VD-Bestände
- 05/2020 Workshop mit Bibliotheken und anderen Anwendern auf dem 8. Bibliothekskongress in Hannover
- 06/2020 Vorschläge zur Weiterentwicklung der DFG-Praxisregeln "Digitalisierung"



Vielen Dank für Ihre Aufmerksamkeit!