

AN OPEN-SOURCE FRAMEWORK FOR INTEGRATING MULTI-SOURCE LAYOUT AND TEXT

RECOGNITION TOOLS INTO SCALABLE OCR WORKFLOWS

KAY-MICHAEL WÜRZNER

KONSTANTIN BAIERER

OVERVIEW

1. Why OCR-D
2. The OCR-D initiative
3. Architecture
4. State of OCR-D tools
5. Scalability
6. Open source

USERS

WANT

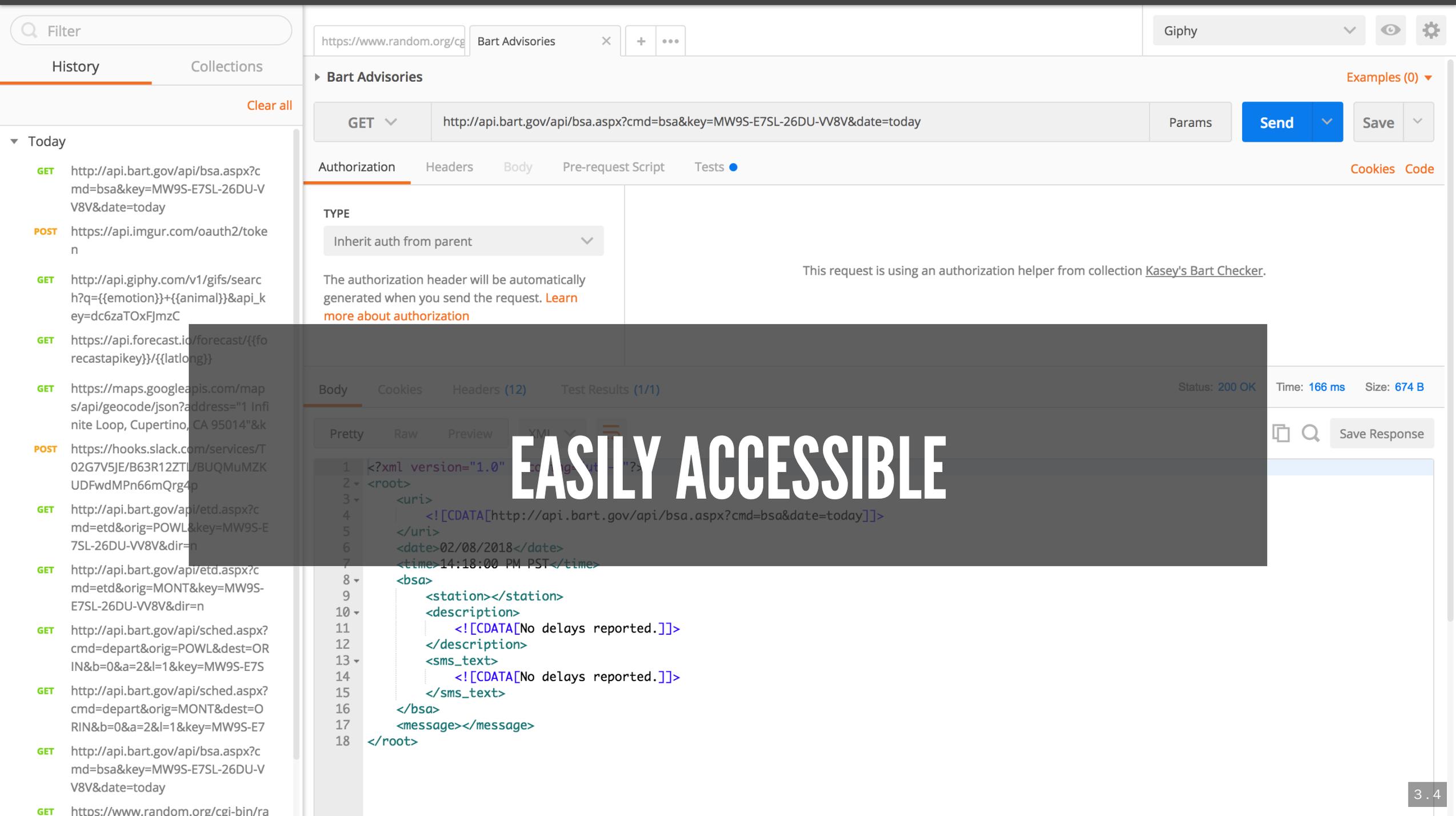
TEXT DATA



MASSIVE AMOUNTS

```
</Shape>
<TextLine ID="tl_5" BASELINE="860" HEIGHT="55" WIDTH="564" VPOS="805" HPOS="233" STYLEREFs="Arial_12.0_-----">
  <String ID="w_w1aab1b3b2b3b1ab1" HEIGHT="51" WIDTH="306" VPOS="807" HPOS="233" STYLEREFs="Arial_12.0_-----" CONTENT="Beantwortung"/>
  <String ID="w_w1aab1b3b2b3b1ac27" HEIGHT="38" WIDTH="61" VPOS="810" HPOS="569" STYLEREFs="Arial_12.0_-----" CONTENT="der"/>
  <String ID="word_1478541289590_808" HEIGHT="53" WIDTH="124" VPOS="806" HPOS="654" STYLEREFs="Arial_12.0_-----" CONTENT="Frage"/>
  <String ID="word_1478541289588_807" HEIGHT="53" WIDTH="19" VPOS="806" HPOS="778" STYLEREFs="Arial_12.0_-----" CONTENT=":"/>
</TextLine>
<TextLine ID="tl_6" BASELINE="941" HEIGHT="55" WIDTH="683" VPOS="886" HPOS="177" STYLEREFs="-----">
  <String ID="w_w1aab1b3b2b3b3ab1" HEIGHT="46" WIDTH="139" VPOS="888" HPOS="177" STYLEREFs="Arial_12.0_-----" CONTENT="Was"/>
  <String ID="w_w1aab1b3b2b3b3ab9" HEIGHT="51" WIDTH="56" VPOS="890" HPOS="362" STYLEREFs="Arial_12.0_-----" CONTENT="ift"/>
  <String ID="word_1478541293583_810" HEIGHT="52" WIDTH="367" VPOS="887" HPOS="465" STYLEREFs="Arial_12.0_-----" CONTENT="Aufklärung"/>
  <String ID="word_1478541293581_809" HEIGHT="52" WIDTH="28" VPOS="887" HPOS="832" STYLEREFs="Arial_12.0_-----" CONTENT="?"/>
</TextLine>
</TextBlock>
<TextBlock ID="r_2_3" HEIGHT="42" WIDTH="474" VPOS="977" HPOS="277">
  <Shape>
    <Polygon POINTS="277,977 751,977 751,1019 277,1019"/>
  </Shape>
</TextBlock>
a/kant_aufklaerung_1784/alto/kant_aufklaerung_1784_0017.xml xml 22%
</mets:mdWrap>
mets:dmdSec>
mets:fileSec>
<mets:fileGrp USE="OCR-D-IMG">
  <mets:file ID="INPUT_0020" MIMETYPE="image/tiff">
    <mets:Flocat LOCTYPE="URL" xlink:href="https://github.com/OCR-D/assets/raw/master/data/kant_aufklaerung_1784/kant_aufklaerung_1784_0017.tif" />
  </mets:file>
</mets:fileGrp>
<mets:fileGrp USE="OCR-D-GT-PAGE">
  <mets:file ID="PAGE_0020_PAGE" MIMETYPE="application/vnd.prima.page+xml">
    <mets:Flocat LOCTYPE="URL" xlink:href="https://github.com/OCR-D/assets/raw/master/data/kant_aufklaerung_1784/page/kant_aufklaerung_1784_0020.xml" />
  </mets:file>
</mets:fileGrp>
<mets:fileGrp USE="OCR-D-GT-ALTO">
  <mets:file ID="ALTO_0020_ALTO" MIMETYPE="application/vnd.prima.alto+xml">
    <mets:Flocat LOCTYPE="URL" xlink:href="https://github.com/OCR-D/assets/raw/master/data/kant_aufklaerung_1784/alto/kant_aufklaerung_1784_0020.xml" />
  </mets:file>
</mets:fileGrp>
assets/data/kant_aufklaerung_1784/mets.xml 3/3 71%
```

STRUCTURED



EASILY ACCESSIBLE

LIBRARIES

PROVIDE

TEXT DATA

92749 Suchergebnisse mit folgenden Kriterien:

Kategorien Historische Drucke

- reffer in 48 Kategorien
- Historische Drucke (92749)
- D18 digital (26341)
- heologie (22615)
- rechtswissenschaft (17651)
- prachen / Literaturen (13902)
- eschichte / Ethnographie / Geographie (9260)
- rieg 1914-1918 (7587)
- stasiatica (7302)
- llgemeines / Wissenschaftskunde / Literaturgeschichte (5580)
- usik (4997)
- inblattdrucke (3094)



Gesetzsammlung für die Königlichen Preußischen Staaten

1865. 1865 | Zeitschriftenband
Gr 3602-1865 | In: Gesetzsammlung für die Königlichen Preußischen Staaten



OCR

Guillié, Sébastien

LARGE AMOUNTS

Bibliothèque Ophtalmologique, Ou Recueil D'Observations Sur Les Maladies Des Yeux, Faites A La Clinique De L'Institution Royale Des Jeunes Aveugles
T. 1. 1820 | Band
Km 305-1,1/4 | In: Bibliothèque Ophtalmologique, Ou Recueil D'Observations Sur Les Maladies Des Yeux, Faites A La Clinique De L'Institution Royale Des Jeunes Aveugles



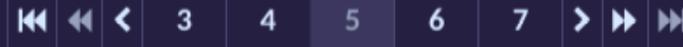
Der Herold

1839. 1839 | Zeitschriftenband
Gq 14350;Beil.3-1839 | In: Der Herold





[5] -- ▾



3

4

5

6

7



Gesetz-Sammlung
für die
Königlichen Preussischen Staaten.
1865.

Enthält die Gesetze, Verordnungen re. vom 2. Januar bis zum 28. Dezember 1865.,
nebst einigen Verordnungen re. aus den Jahren 1862. und 1864.

(Von Nr. 5991. bis Nr. 6235.)

Nr. 1. bis incl. 57.

Berlin,

zu haben im vereinigten Gesetz-Sammlungs-Debits- und Zeitungs-Komtoir.

ZBIÓR PRAW
DLA
PAŃSTW KRÓLESTWA PRUSKIEGO.
1865.

Zawiera ustawy, rozporządzenia i t. d. od 2. Stycznia aż do 28. Grudnia 1865.,
wraz z niektórymi rozporządzeniami i t. d. z roku 1862. i 1864.

(Od No. 5991. aż do No. 6235.)

No. 1. aż włącznie do 57.

W BERLINIE,

do nabycia w połączoném biurze sprzedaży Zbioru praw i gazet.

Gesetz-Sammlung

für die , ,

Königlichen Preussischen Staaten.

1865.

Enthält die Gesetze, Verordnungen re. vom 2 Januar bis ^ 28-Dezäm 186»,.,

„ebst einigen Verordnungen k. ans den Jahren 1862, und «**

(Bon Nr. 5991. bis Nr. 6235.)

Nr. 1. bis incl. 57.

zu

haben im vereinigten Gesetz -Sammlungs- Debits- und Zeitungs - Komtoir.

ZBIÓR PRAW

DLA

PANSTW

UNSTRUCTURED

link="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="http://content.staatsbibliothek-berlin.de/dms/PPN746523211/150/0/00002083.png"/>
le><mets:file ID="FILE_2084_THUMBS" MIMETYPE="image/png"><mets:Flocat xmlns:xlink="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="h
.staatsbibliothek-berlin.de/dms/PPN746523211/150/0/00002084.png"/></mets:file><mets:file ID="FILE_2085_THUMBS" MIMETYPE="image/png"><mets:FL
link="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="http://content.staatsbibliothek-berlin.de/dms/PPN746523211/150/0/00002085.png"
le><mets:file ID="FILE_2086_THUMBS" MIMETYPE="image/png"><mets:Flocat xmlns:xlink="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="h
.staatsbibliothek-berlin.de/dms/PPN746523211/150/0/00002086.png"/></mets:file><mets:file ID="FILE_2087_THUMBS" MIMETYPE="image/png"><mets:FL
link="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="http://content.staatsbibliothek-berlin.de/dms/PPN746523211/150/0/00002087.png"
le><mets:file ID="FILE_2088_THUMBS" MIMETYPE="image/png"><mets:Flocat xmlns:xlink="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="h
.staatsbibliothek-berlin.de/dms/PPN746523211/150/0/00002088.png"/></mets:file><mets:file ID="FILE_2089_THUMBS" MIMETYPE="image/png"><mets:FL
link="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="http://content.staatsbibliothek-berlin.de/dms/PPN746523211/150/0/00002089.png"
le><mets:file ID="FILE_2090_THUMBS" MIMETYPE="image/png"><mets:Flocat xmlns:xlink="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="h
.staatsbibliothek-berlin.de/dms/PPN746523211/150/0/00002090.png"/></mets:file><mets:file ID="FILE_2091_THUMBS" MIMETYPE="image/png"><mets:FL
link="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="http://content.staatsbibliothek-berlin.de/dms/PPN746523211/150/0/00002091.png"
le><mets:file ID="FILE_2092_THUMBS" MIMETYPE="image/png"><mets:Flocat xmlns:xlink="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="h
.staatsbibliothek-berlin.de/dms/PPN746523211/150/0/00002092.png"/></mets:file><mets:file ID="FILE_2093_THUMBS" MIMETYPE="image/png"><mets:FL
link="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="http://content.staatsbibliothek-berlin.de/dms/PPN746523211/150/0/00002093.png"
le><mets:file ID="FILE_2094_THUMBS" MIMETYPE="image/png"><mets:Flocat xmlns:xlink="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="h
.staatsbibliothek-berlin.de/dms/PPN746523211/150/0/00002094.png"/></mets:file><mets:file ID="FILE_2095_THUMBS" MIMETYPE="image/png"><mets:FL
link="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="http://content.staatsbibliothek-berlin.de/dms/PPN746523211/150/0/00002095.png"
le></mets:fileGrp><mets:fileGrp USE="FULLTEXT"><mets:file ID="FILE_0001_FULLTEXT" MIMETYPE="text/xml"><mets:Flocat xmlns:xlink="http://www.w
link" LOCTYPE="URL" xlink:href="http://digital.staatsbibliothek-berlin.de/ocrresolver/PPN746523211/00000001.xml"/></mets:file><mets:file
E_0002_FULLTEXT" MIMETYPE="text/xml"><mets:Flocat xmlns:xlink="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="http://digital.
bibliothek-berlin.de/ocrresolver/PPN746523211/00000002.xml"/></mets:file><mets:file ID="FILE_0003_FULLTEXT" MIMETYPE="text/xml"><mets:Flocat
http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="http://digital.staatsbibliothek-berlin.de/ocrresolver/PPN746523211/00000003.xml"/></
mets:file ID="FILE_0004_FULLTEXT" MIMETYPE="text/xml"><mets:Flocat xmlns:xlink="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="http:
.staatsbibliothek-berlin.de/ocrresolver/PPN746523211/00000004.xml"/></mets:file><mets:file ID="FILE_0005_FULLTEXT" MIMETYPE="text/xml"><mets
xmlns:xlink="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="http://digital.staatsbibliothek-berlin.de/ocrresolver/PPN746523211/0000
/mets:file><mets:file ID="FILE_0006_FULLTEXT" MIMETYPE="text/xml"><mets:Flocat xmlns:xlink="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlin
http://digital.staatsbibliothek-berlin.de/ocrresolver/PPN746523211/00000006.xml"/></mets:file><mets:file ID="FILE_0007_FULLTEXT" MIMETYPE="te
mets:Flocat xmlns:xlink="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="http://digital.staatsbibliothek-berlin.de/ocrresolver/PPN746
7.xml"/></mets:file><mets:file ID="FILE_0008_FULLTEXT" MIMETYPE="text/xml"><mets:Flocat xmlns:xlink="http://www.w3.org/1999/xlink" LOCTYPE="
ref="http://digital.staatsbibliothek-berlin.de/ocrresolver/PPN746523211/00000008.xml"/></mets:file><mets:file ID="FILE_0009_FULLTEXT"
E="text/xml"><mets:Flocat xmlns:xlink="http://www.w3.org/1999/xlink" LOCTYPE="URL" xlink:href="http://digital.staatsbibliothek-berlin.de/
lver/PPN746523211/00000009.xml"/></mets:file><mets:file ID="FILE_0010_FULLTEXT" MIMETYPE="text/xml"><mets:Flocat xmlns:xlink="http://www.w3.
~/Downloads/PPN746523211.mets.xml

WHY THE DISCREPANCY?

OCR

Die OCR-Technik hat in jüngerer Zeit erheblich Fortschritte gemacht, und wesentliche Verbesserungen konnten auch bei Frakturschriften sowohl der Handpressen- als auch Maschinenpressenzeit erzielt werden. Gleichwohl können die Praxisregeln angesichts der dynamischen Entwicklung und mit Blick auf die zu erwartenden Ergebnisse des OCR-D Projektes zur Frage der OCR-Software und ihrer Anwendbarkeit derzeit noch keine abschließenden Empfehlungen geben.

Der OCR-Prozess besteht aus mehreren Schritten, die in der Reihenfolge von links nach rechts unterscheiden kann man das Preprocessing, in dem ein Image vorbearbeitet wird (cropping, despeckle, deskewing, binarisation), die Optical Layout Recognition (Segmentation bzw. Identifikation von Bild- und Textregionen), die Optical Character Recognition (eigentliche Texterkennung) und Postprocessing (Textkorrektur). Insofern hat die Qualität des Ausgangsbilds ganz entscheidenden Einfluss auf den Prozess der Binarisierung und dieser wiederum auf das Erkennungsergebnis. Daher sollten nur solche digitalen Vorlagen einer OCR-Bearbeitung unterzogen werden, bei denen sichergestellt ist, dass die Bildqualität ausreicht. Problematisch sind in diesem Zusammenhang auch intrinsische Phänomene wie Verschmutzungen, Widerdrucksschatten, manuelle Unterstreichungen oder Annotationen u.ä., die sich

UNDERSPECIFIED REQUIREMENTS ON OCR BY FUNDRERS AND USERS

- **The Reality**

- Masses of books/document are available & already scanned
- It is unclear if Antiqua and/or Gothic/Fraktur fonts are used in the documents
- Pre-Sorting is impossible, it would be too time/cost expensive

- **ABBYy OCR for Historical Documents**
Reduced the pricing for mixed "Old" + "Modern" font OCR projects

The pricing is now ready for "mass processing"

- **Examples Recognition Server 5.0 with "Gothic" enabled**

- 10.000 pages – 299 Euro – available online
- 500.000 pages* – 5.000 Euro = 1 Euro per page = ca 2.000 books a 250 pages
- Over 3 Mio pages* - ca 0,52 Euro cent per page = 12.000 books a 1,25 € (250 pages)
- Over 10 Mio pages* - ca. 40.000 books = ca. 0,5 € per book

OCR OF HISTORICAL DATA OF LITTLE ECONOMIC

INTEREST

LITTLE COMPETITION



ACADEMICAL SOLUTIONS NON-SUSTAINABLE

"Black Box"

INFLEXIBLE WORKFLOWS

Output, I

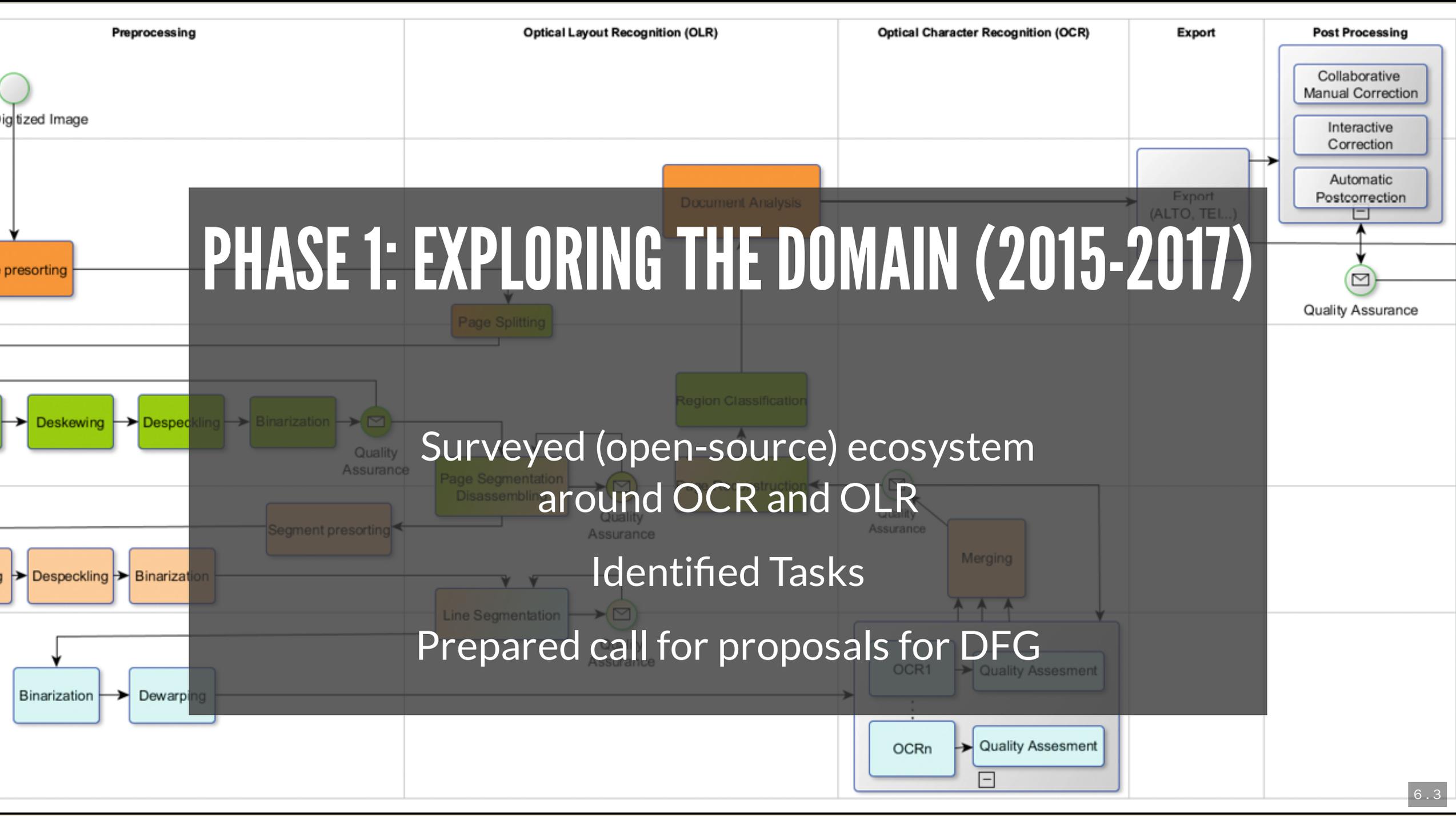
THE OCR-D INITIATIVE

DFG-organized expert Workshop (2014)
Verfahren zur Verbesserung von OCR-Ergebnissen

Result: A concerted effort for improving OCR is seen as required.

OCR-D COORDINATION PROJECT





PHASE 1: EXPLORING THE DOMAIN (2015-2017)

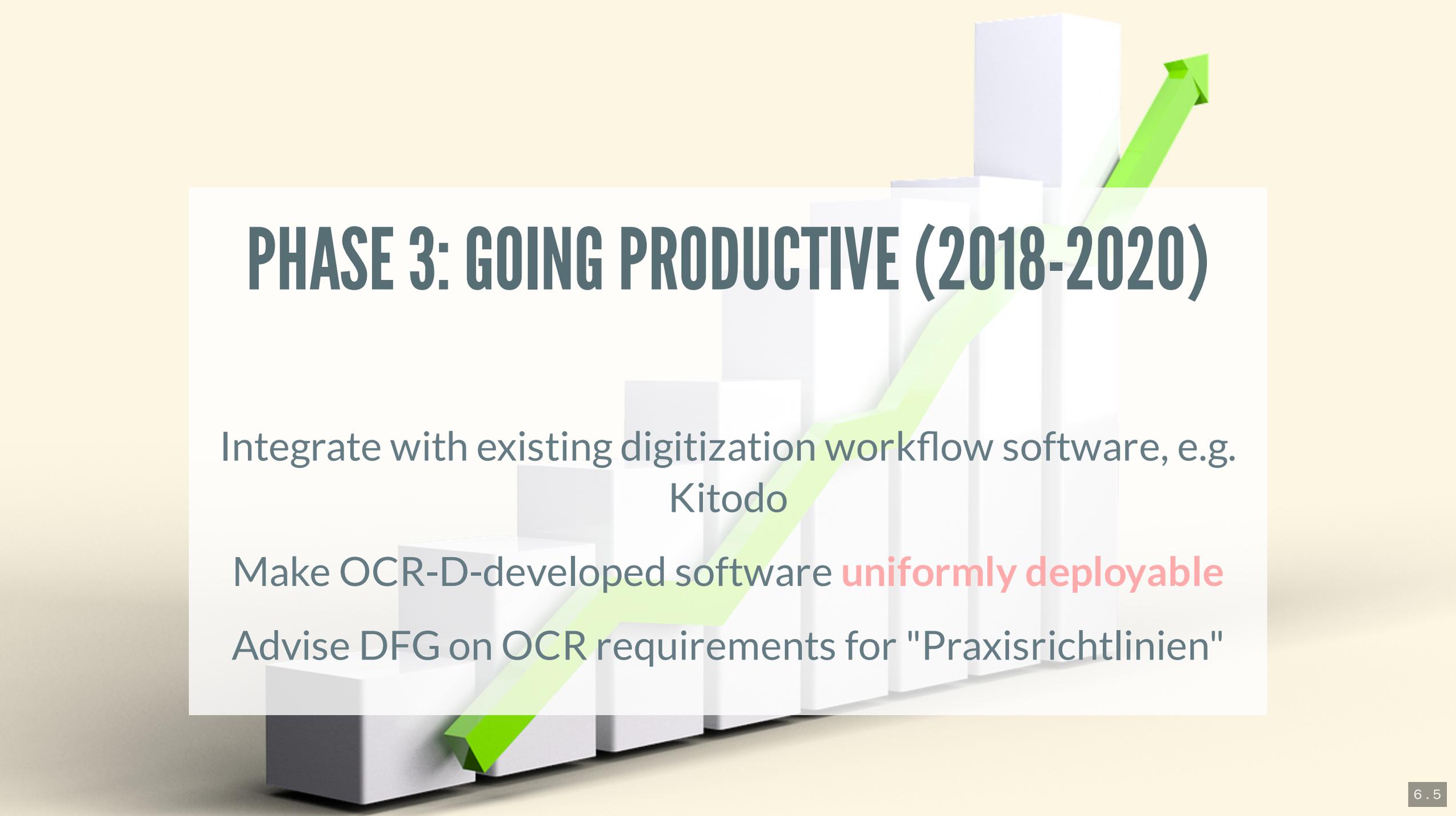
Surveyed (open-source) ecosystem
around OCR and OLR

Identified Tasks

Prepared call for proposals for DFG

PHASE 2: MODULE PROJECT STAGE (2018-2019)





PHASE 3: GOING PRODUCTIVE (2018-2020)

Integrate with existing digitization workflow software, e.g.
Kitodo

Make OCR-D-developed software **uniformly deployable**

Advise DFG on OCR requirements for "Praxisrichtlinien"

NOT IN THIS TALK

(BUT IN OCR-D)

GROUND TRUTH FOR OCR

OCR ENGINE TRAINING

OCR RESEARCH DATA REPOSITORY

WORKFLOW COMPOSITION AND PROVENANCE

Konventionen für diese
Richtlinien

In den Konventionen werden einzelne
Begriffe, Formatierungen, bestimmte
Verwendungsweisen definiert und
erklärt, die im Text der Richtlinien

Layout und Struktur

Richtlinien zur Erfassung des Layouts
und der Struktur von gedruckten
Texten für die Nutzung als Ground
Truth

Dokumentation zum
OCR-D Struktur-Ground-
Truth

Der OCR-D Struktur-Ground-Truth
beinhaltet Publikationen bei denen
ausschließlich die Strukturen oder

Page XML
Erweiterungen

ARCHITECTURE

"MULTI SOURCE"

existing tools by OCR-D partners
(tesseract, PoCoTo, LAREX...)

new developments within OCR-D
(font identification, post-correction...)

existing tools **outside OCR-D**
(ocropus, kraken, ScanTailor, OLENA...)

□ **MODULAR** □

□ **MONOLITHIC** □

SPECIFICATION

+

IMPLEMENTATION



Intro

Overview of OCR-D technical document structure and command line tools provided by MP

OCR-D / SPEC

METS

OCR-D METS conventions

METS + PAGE-XML (+ ALTO)

OCR-D-ZIP

METS packages realized as ZIP

ecrd-tool.json

OCR-D tool description

PAGE

PAGE conventions

STRUCTURED TOOL DESCRIPTIONS

Dockerfile

OCR-D Dockerfile conventions

COMMAND LINE INTERFACE

Base Dockerfile for OCR-D services: `workshop/2018/01/01/018`

HTTP INTERFACE



```
1 {
2   "version": "0.1.2",
3   "git_url": "https://github.com/OCR-D/ocrd_tesseract",
4   "dockerhub": "ocrd/tesseract",
5   "tools": {
6     "ocrd-tesseract-recognize": {
7       "executable": "ocrd-tesseract-recognize",
8       "categories": ["Text recognition and optimization"],
9       "description": "Recognize text in lines with tesseract",
10      "steps": ["recognition/text-recognition"],
11      "parameters": {
12        "textequiv_level": {
13          "type": "string",
14          "enum": ["region", "line", "word", "glyph"],
15          "default": "line",
16          "description": "PAGE XML hierarchy level to add the TextEquiv results to (requires existing layout annotation up to c
17        },
18        "model": {
19          "type": "string",
20          "description": "tessdata model to apply (an ISO 639-3 language specification or some other basename, e.g. deu-frak or
21        }
22      }
23    },
24    "ocrd-tesseract-segment-region": {
25      "executable": "ocrd-tesseract-segment-region",
26      "categories": ["Layout analysis"],
27      "description": "Segment regions into lines with tesseract",
28      "steps": ["layout/segmentation/region"],
29      "parameters": {}
30    },
31    "ocrd-tesseract-segment-line": {
32      "executable": "ocrd-tesseract-segment-line",
33      "categories": ["Layout analysis"],
```

ACTIONABLE DOCUMENTATION

Code

Issues 9 Pull requests 1 Projects 2 Wiki Releases 37 More Settings

Collection of OCR-related python tools and scripts from @OCR-D

Edit

Manage topics

OCR-D / CORE

538 commits 3 branches 37 releases 5 contributors Apache-2.0

Branch: master

v0.8.5

Create new file Find file Clone or download

- kba changelog
- .circleci
- bashlib
- docs
- ocrd
- test
- .coveragerc
- .dockerignore
- .gitignore

VALIDATION AND HELPER FUNCTIONS

PYTHON LIBRARY

SHELL LIBRARY

bashlib: subtle bug which clobbered retval, fix #161 2 months ago

bring app... a month ago

changelog 6 days ago

Merge branch 'master' into drop-from-folder 12 days ago

Add coverage a month ago

dockerignore: relative paths for exceptions 6 months ago

replace xmllint callout with lxml solution, fix #72 2 months ago

WHY PYTHON?

Python widely used in computer vision and machine learning
(keras, pytorch...)

Wrapping existing tools with minimal friction
(ocropus, kraken ...)

Bindings for low-level APIs
(opencv, tesseract ...)

WHY SHELL?

Lowest common denominator

Wrap arbitrary command line tools

Process callout possible in every
framework/workflow engine/programming environment



STATE OF THE OCR-D TOOLSET

PREPROCESSING

Tool	Developer	Functionality	Wrapper
anyOCR	DFKI Kaiserslautern	binarization, cropping, deskewing, dewarping	(python)
OLENA	OCR-D	binarization	shell
tesseract	UB Mannheim, ASV Leipzig	binarization	python
OCRopus	OCR-D	binarization	python
kraken	OCR-D	binarization	python
ImageMagick	OCR-D	binarization, conversion	shell

LAYOUT RECOGNITION

Tool	Developer	Functionality	Wrapper
anyOCR	DFKI Kaiserslautern	block+line seg8n, block class7n, document analysis	(python)
LAREX	Uni Würzburg	block+line seg8n, block class7n	(shell)
OCRopus	OCR-D	line seg8n	python
kraken	OCR-D	line seg8n	python
tesseract	UB Mannheim, ASV Leipzig	block+line seg8n	python
dh_segment	OCR-D	block+line seg8n	(shell)

TEXT RECOGNITION

Tool	Developer	Functionality	Wrapper
OCROpus	OCR-D	text recognition	python
kraken	OCR-D	text recognition	python
tesseract	UB Mannheim, ASV Leipzig	text recognition	python
calamari	OCR-D	text recognition	(python)
ocrad	OCR-D	text recognition	(shell)

POSTPROCESSING

Tool	Developer	Functionality	Wrapper
corASV	ASV Leipzig	post correction	(python)
PoCoTo	CIS München	post correction	python
keraslm	ASV Leipzig	post correction	python
ocrevalUAtion	OCR-D	evaluation	(shell)

YOUR TOOL?





SCALABILITY

<IMPRESSIVE NUMBER HERE>



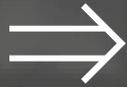
GEARED TOWARDS REAL

DIGITIZATION SCENARIOS

Cooperation with Kitodo and commercial providers

Frequent reality check with current practices
("Pilotbibliotheken")

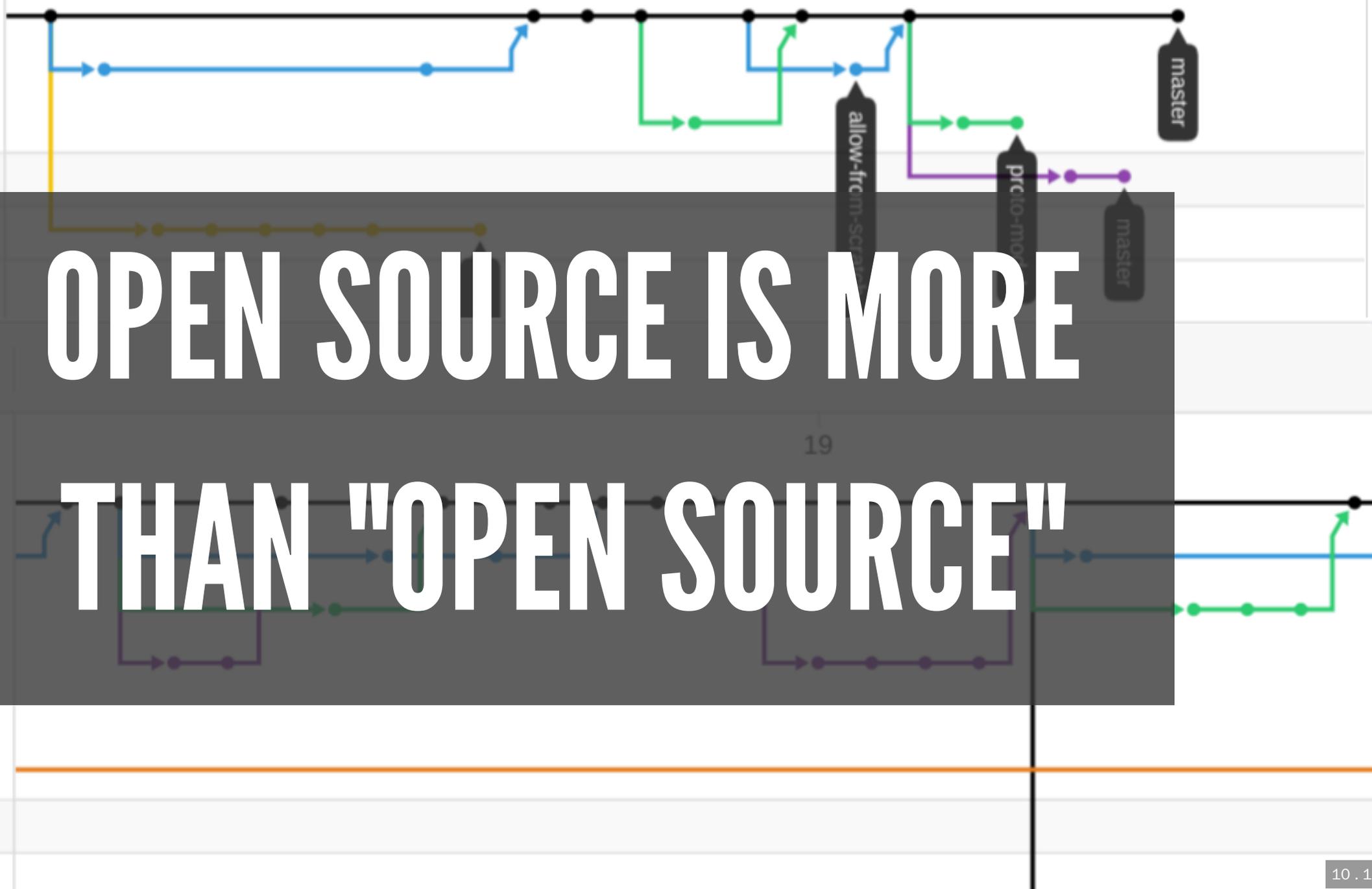
MODULARITY + UNIFORM INTERFACES



ADAPTIVE WORKFLOWS

(Instantiation and composition up to users)

OPEN SOURCE IS MORE THAN "OPEN SOURCE"





STEP 1: GET FUNDED!

```
* @author jeff
*/
public class Main {

    public static String AppName = "SQL Mail";
    public static String AppVersion = " 0.0.1 ";
    public static String AppAuthor = "Jeffrey Cobb";
    public static String AppDate = "August 8th, 2007";
    public static String AppPath = System.getProperty("user.dir");
    public static String AppDriver = "smallsql.database.SSDriver";
    public static String AppDBHeader = "jdbc:smallsql:";
    public static String AppDBPath = AppPath + "/sqlmail";
    public static String AppPreferences = AppPath + "/sqlmail_prefs";
    /** Creates a new instance of Main */
    public Main() {
    }

    /**
     * @param args the command line arguments
     */
    public static void main(String[] args) throws Exception {
        // TODO code application logic here

        boolean bDBConnect = false;
        int result = 0;
        frmMain SQLMailForm = new frmMain();
        System.out.println("\r\n" + AppName + "\r\nVersion" + AppVersion + "\r\nAuthor: " + AppAuthor + "
.. " + AppDate + "\r\n");
    }
}
```

STEP 2: DEVELOP!



STEP 3: PUBLISH CODE!



SUSTAINABILITY AND REUSE!

BEST PRACTICES

Transparency from day one

Unit tests

Unified **test assets**

Continuous Integration

Semantic versioning

Docker base image

Releases to GitHub, PyPI, DockerHub



COMMUNITY

Issues

Pull requests

Code review

Support [chat](#)



OCR-D/DOCS

DEVELOPER DOCUMENTATION

"COOKBOOK"

USER GUIDE

DOCUMENTATION DOCUMENTATION

Anatomy of an OCR-D module project (MP)

`ocrd-tool.json`

Mechanics of the `ocrd-tool.json`

Metadata about the module project

Metadata about the tools

Metadata about parameters

`Makefile`

`Makefile` for Python MP

`Makefile` for generic MP

`ocrd workspace` - Working with METS

From METS to Workspace

Lines in code examples

▣ THANK YOU ▣

ocr-d.de

ocr-d.github.io

ocr-d.github.io/docs

github.com/OCR-D

gitter.im/OCR-D/Lobby