



KA³

Kölner Zentrum Archivierung und Analyse von AV-Daten

BMBF-Zentrum KA³

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

- Kölner Zentrum **Analyse** und **Archivierung** von **AV-Daten**
- Verbundprojekt
 - Universität zu Köln
 - Regionales Rechenzentrum
 - Data Center for the Humanities
 - Institut für Linguistik
 - Externe Partner
 - Fraunhofer-Institut für Intelligente Analyse und Informationssysteme (IAIS)
 - FernUniversität Hagen, Archiv Deutsches Gedächtnis
 - Max-Planck-Institut für Psycholinguistik

BMBF-Zentrum KA³

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Fachwissenschaftliche Partner

- Kölner Zentrum **Analyse** und **Archivierung** von **AV-Daten**
- Verbundprojekt
 - Universität zu Köln
 - Regionales Rechenzentrum
 - Data Center for the Humanities
 - **Institut für Linguistik**
 - Externe Partner
 - Fraunhofer-Institut für Intelligente Analyse und Informationssysteme IAIs
 - **FernUniversität Hagen, Archiv Deutsches Gedächtnis**
 - Max-Planck-Institut für Psycholinguistik



BMBF-Zentrum KA³

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Technische Partner

- Kölner Zentrum **Analyse** und **Archivierung** von **AV-Daten**
- Verbundprojekt
 - Universität zu Köln
 - **Regionales Rechenzentrum (Archivierung)**
 - Data Center for the Humanities
 - Institut für Linguistik
 - Externe Partner
 - **Fraunhofer-Institut für Intelligente Analyse und Informationssysteme (Analyse)**
 - FernUniversität Hagen, Archiv Deutsches Gedächtnis
 - **Max-Planck-Institut für Psycholinguistik (Archivierung)**

BMBF-Zentrum KA³

GEFÖRDERT VOM

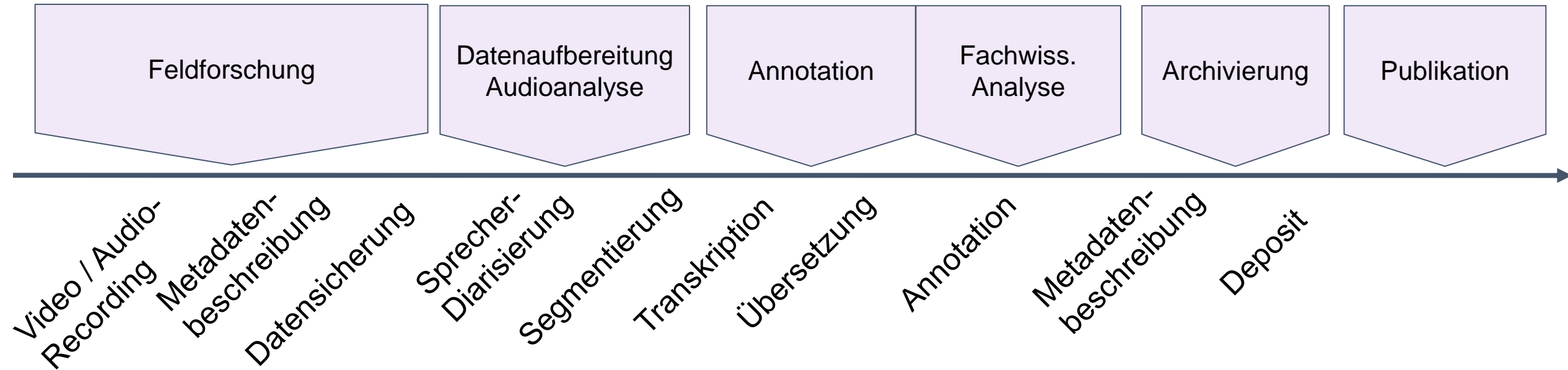


Bundesministerium
für Bildung
und Forschung

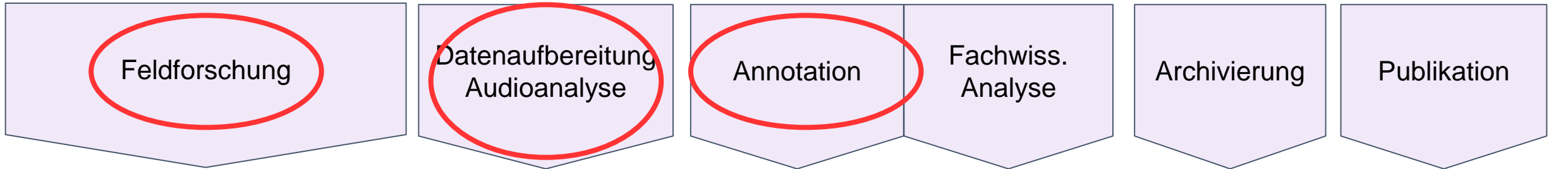
Fachwissenschaftlich-technischer Partner

- Kölner Zentrum **Analyse** und **Archivierung** von **AV-Daten**
- Verbundprojekt
 - Universität zu Köln
 - Regionales Rechenzentrum
 - **Data Center for the Humanities**
 - Institut für Linguistik
 - Externe Partner
 - Fraunhofer-Institut für Intelligente Analyse und Informationssysteme IAIs
 - FernUniversität Hagen, Archiv Deutsches Gedächtnis
 - Max-Planck-Institut für Psycholinguistik

Forschungsprozess



Forschungsprozess



Video / Audio-Recording
Metadatenbeschreibung
Datensicherung
Sprecher-Diarisierung
Segmentierung
Transkription
Übersetzung
Annotation
Metadatenbeschreibung
Deposit

AV

Werkzeuge & Dienste

Werkzeuge & Dien

Feldforschung

Video / Audio-
Recording
Metadaten-
beschreibung
Datensicherung
Sprecher-
Diarisierung
Segmentierung
Transkription



Sciebo
Dropbox
Drive
> 20GB!



offline

Werkzeuge & Dienstleistungen

Datenaufbereitung
Audioanalyse

Video / Audio-Recording
Metadaten-beschreibung
Datensicherung
Sprecher-Diarisierung
Segmentierung
Transkription



iFinder
AudioMining

AudioMining powered by Fraunhofer IAIS

Suche: Titel, Transkript, Analysedatum

Überblick | Suchtreffer | Transkript | Empfehlungen

Wladimir, Susdal und Kideksha, Russland, Folge 176

Schatze der Welt | SWR Fernsehen | 10.12.2000 | 00:14:41 min.

Kirche, Klöster, Jahrhundert, Tataren, Mongolen, Wladimir, Fürst, Gotteshaus, Star, Feuersbrunst, Schutz, Ruß, Costas, Maria, Glocken

Traum und Wirklichkeit: Russland und die Modernisierung - 27.08.2011

Breitengrad - Bayern 2 | Bayern 2 | 27.8.2011 | 00:24:13 min.

Modernisierung, Innovation, Medwedew, Moskau, Rußland, Guder, Jahr, Petrov, Golkar, Politologe, Goll, Zug, Popow, Nikolai, Projekt

Russland will neuen Hilfskonvoi in die Ukraine schicken

Keywords: Rußland, Kiew, Konvoi, Hilfsaktionen, Parade, Putin, Osten, Moskau, Schenkung, Lohans, Lihl, Outkast, Regionenkreisen, Geisterfahrt

Werkzeuge & Dier

AudioMining
powered by Fraunhofer IIS

Suchen
Leeren

Überblick | Suchtreffer | Transkript | Empfehlungen

Hintergrundgeräusche | Militäreinheiten

Wladimir, Susdal und Kideksha, Russland, Folge 176
Schätze der Welt | SWR Fernsehen
10.12.2000 | 00:14:41 min.
Kirche | Klöster | Jahrhundert | Tataren
Mongolen | Wladimir | Fürst | Gotteshaus | Star
Feuersbrunst | Schutz | Ruß | Costas | Maria
Glocken

Traum und Wirklichkeit: Russland und die Modernisierung - 27.08.2011
Breitengrad - Bayern 2 | Bayern 2
27.8.2011 | 00:24:13 min.
Modernisierung | Innovation | Medwedew
Moskau | Rußland | Guder | Jahr | Petrov
Golkar | Politologe | Goll | Zug | Popow | Nikolai
Projekt

Russland will neuen Hilfskonvoi in die Ukraine schicken
Mittwoch, 1. Dezember 2011

IFinder / AudioMining

Spracherkennung (Deutsch)

- ~95% bei hochqualitativen Aufnahmen
- ~40% bei Alltagsaufnahmen

Herausforderungen:

- Hintergrundgeräusche
- Schnelle Sprecherwechsel
- Überlappungen beim Sprecherwechsel
- ...

Weiterentwicklung (KA³ 2. Phase):

- Trainierbare Sprachmodelle

Video / Audio-
Recording
Metadaten-
beschreibung
Datensicherheit

Werkzeuge & Dier

The screenshot shows the AudioMining interface, powered by Fraunhofer IIS. It features a search bar with fields for 'Titel', 'Transkript', and 'Analysedatum'. Below the search bar, there are navigation tabs for 'Überblick', 'Suchtreffer', 'Transkript', and 'Empfehlungen'. The main content area displays search results for 'Wladimir, Susdal und Kideksha, Russland, Folge 176', including a video player, a description, and a list of related terms like 'Kirche', 'Klöster', 'Jahrhundert', 'Tataren', etc.

DH Analyse Infrastruktur?

- Sehr unterschiedliche Hardwareanforderungen
- Unterschiedliche Parallelisierungstechniken und SDKs
- HPC vs. HTC, Tasks pro Sekunde vs. Tasks pro Tag / Woche / Monat

Wer muss sich nach wem richten? Infrastruktur nach Entwicklern von Analysesystemen oder umgekehrt?

Video / Audio-
Recording
Metadaten-
beschreibung
Datensicher

Werkzeuge & Dienste

iFinder
AudioMining
Export

Annotation



Annotation: MR.anvil

Track (primary): trl
Time: 00:17:48 - 00:20:00 (63 frames)

Attributes
lexeme: Progressive
handedness: 2H
path: curved
hand-height-1: chest
hand-body-dist-1: normal
hand-radius-orient-1: front
elbow-inclination-1: normal
2H-distance-start (25x127 00:18:00) (92x126 00:18:00)
lex affix: **gerpakt**

Comment

Start Edit Cut Extend

ELAN Annotation Tool

ung

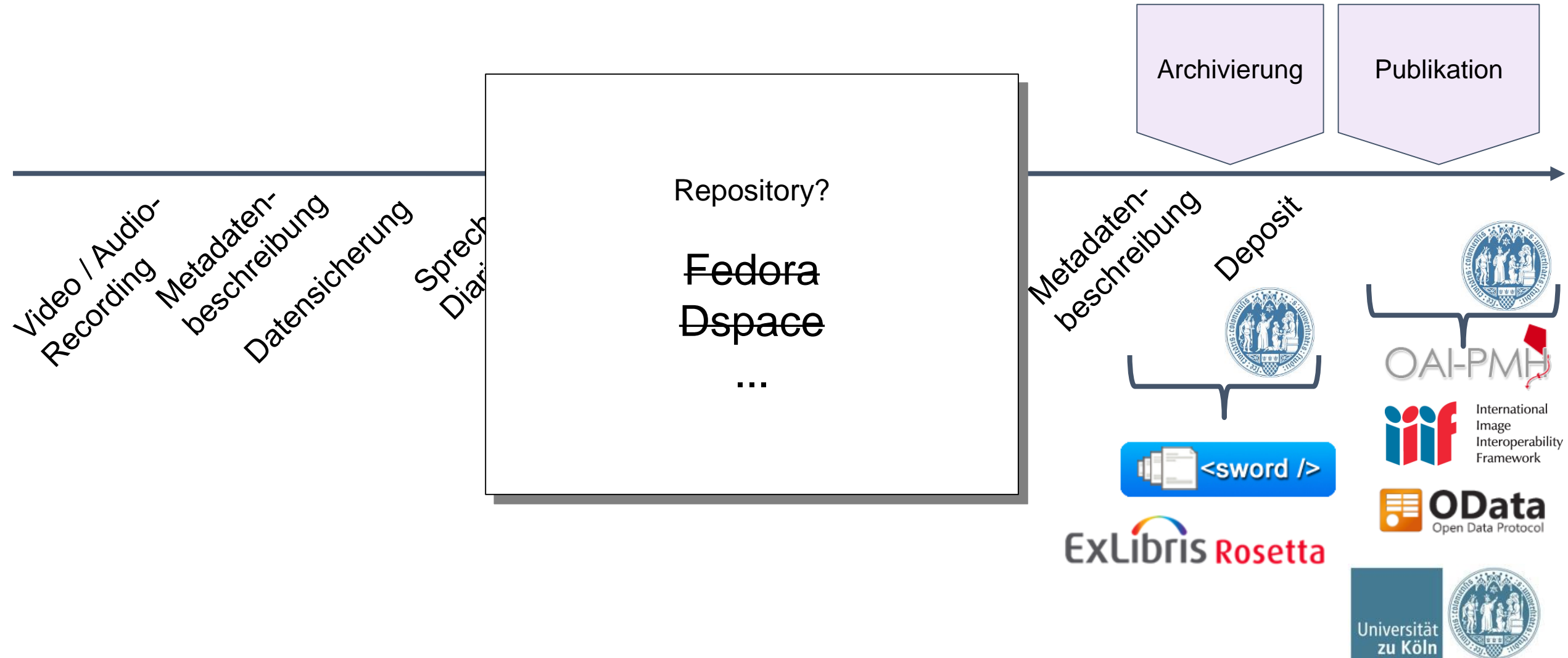
Annotation

Metadaten-
beschreibung

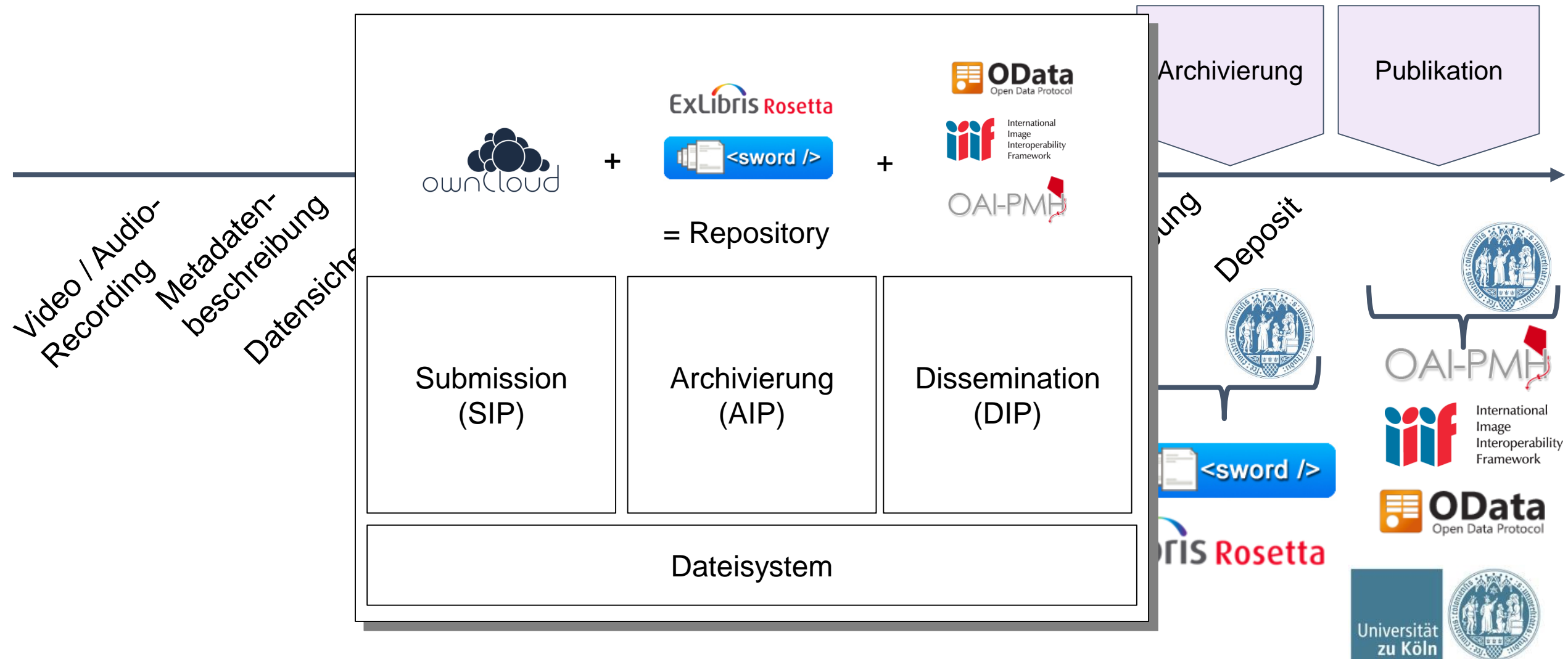
Deposit



Werkzeuge & Dienste



Werkzeuge & Dienste



Werkzeuge & Dienste

ExLibris Rosetta

Pilotprojekt zum Thema Langzeitverfügbarkeit

- HBZ – Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen
- RWTH Aachen – Rheinisch-Westfälische Technische Hochschule Aachen
- USB Köln – Universitäts- und Stadtbibliothek Köln
- RRZK – Regionales Rechenzentrum, Universität zu Köln

Video / Audio-Recording
Metadatenbeschreibung
Datensicherung

Archivierung

Publikation

Metadaten-
beschreibung

Deposit



ExLibris Rosetta



OAI-PMH

iif International Image Interoperability Framework

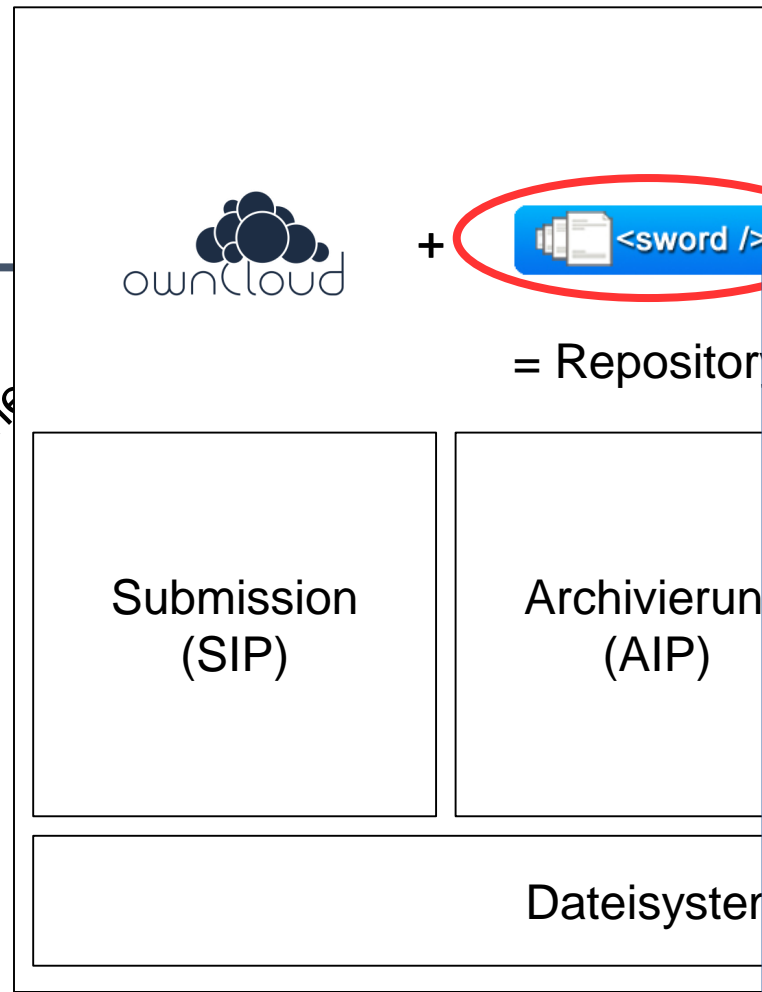
OData Open Data Protocol

Universität zu Köln



Werkzeuge & Dienste

Video / Audio-Recording
Metadaten-beschreibung
Datensicherung



Deposit
Sword Version 3, Mai 2018


- Cloud Storage
- Große Dateien, z.B. AV

Publikation

The screenshot shows the SWORDv3 website. The navigation bar includes links for HOME, ABOUT SWORD, SWORD V1, SWORD V2, SWORDV3 (highlighted), THE SWORD COURSE, and CONTACT. The main content area is titled 'SWORDv3' and contains the following text: 'SWORDv3 is a proposed revision of the specification to break the dependence on Atom and the Atom Publishing Protocol, and to engineer the protocol around JSON and a pure RESTful approach. You can read the working documents here. The documents themselves:'. A list of links follows: 'The proposed spec (v4 – 2018-05-08)', 'The Protocol Behaviours – a companion to the spec which expands the requirements out in full, per operation', 'Our usage patterns and requirements analysis', and 'SWORD community discussion document'. Below this, it states 'The research, statements and use cases upon which the new proposal is based:' followed by a list of links: 'An annotated version of the current SWORDv2 spec', 'Jisc's statement of requirements, working document', 'Data Deposit Scenarios blog post', 'SWORD: Facilitating Deposit Scenarios in DLib', and 'SWORD community consultation document from Jisc'. On the right side, there are two sections: 'Categories' with links for Case studies, Clients, News, Repositories, SWORD v1, SWORD v2, SWORD v3, and Uncategorized; and 'Archives' with links for May 2018, January 2018, November 2017, October 2017, and December 2013.

Workflows & Dienste

KA³ v. 0.58 | Grails v. 3.2.4 Home Login



Search

Explore: Browse the archive by category (showing top 5)

LANGUAGES	COUNTRIES	KEYWORDS
English 353	Papua New Guinea 239	Unspecified 445
Tok Pisin 221	Germany 202	Discourse 328
Garzei 220	India 64	Interactive 208
German, Standard 121	Iran 29	Family 145
Zaghawa 88	Vancouver 30	Conversation 243
more...	more...	more...

KA³

KA³ Home Sign In Registration	About Project Information Term of use Privacy policy	Support FAQ Help desk Contact us
--	--	--

© 2017 Copyright KA³

OData
Open Data Protocol

iif International Image Interoperability Framework

OAI-PMH

Dissemination (DIP)

Archivierung

Publikation

Archivierung

Deposit



OAI-PMH

iif International Image Interoperability Framework

OData Open Data Protocol

ORIS Rosetta


Universität zu Köln



Dateisystem

W

← → ↻ iif.io/community/groups/av/



IIF A/V Technical Specification Group

About

While IIF has focused primarily on Image based content, there has always been interest around the co...
 The IIF A/V Technical Specification Group aims to extend to A/V the benefits of interoperability and the British Library to gather [use cases for extending IIF to audio and video](#), and discussion at the 2016 IIF group welcomes contributions to their collection of [audiovisual user stories](#) and mockups and prototype...
 The [IIF A/V charter](#) provides an outline and timeline for anticipated work.

Organization

Chairs:

- Tom Crane (Digirati)
- Jon Dunn (Indiana University)

Communication Channels:

- Calls every other week on Tuesdays at 12:00pm Eastern (opposite the general IIF Community)
- Virtual meetings announced on the [IIF-Discuss](#) email list
- General discussion on the [# av IIF Slack channel](#) (Join Slack)

Call Notes and Group Documents:

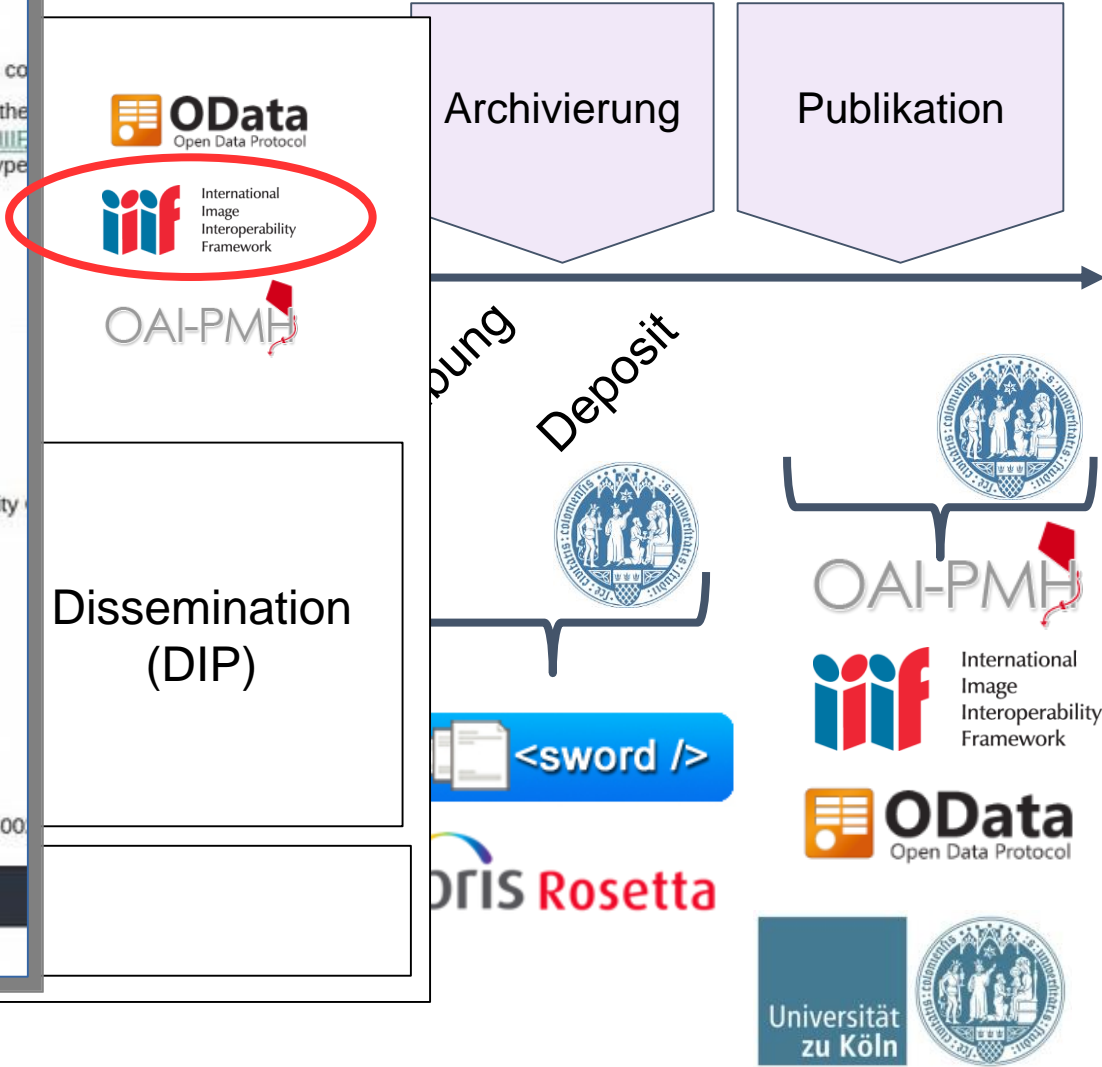
- [IIF A/V Tech Spec Group folder](#)

Call Connection Information:

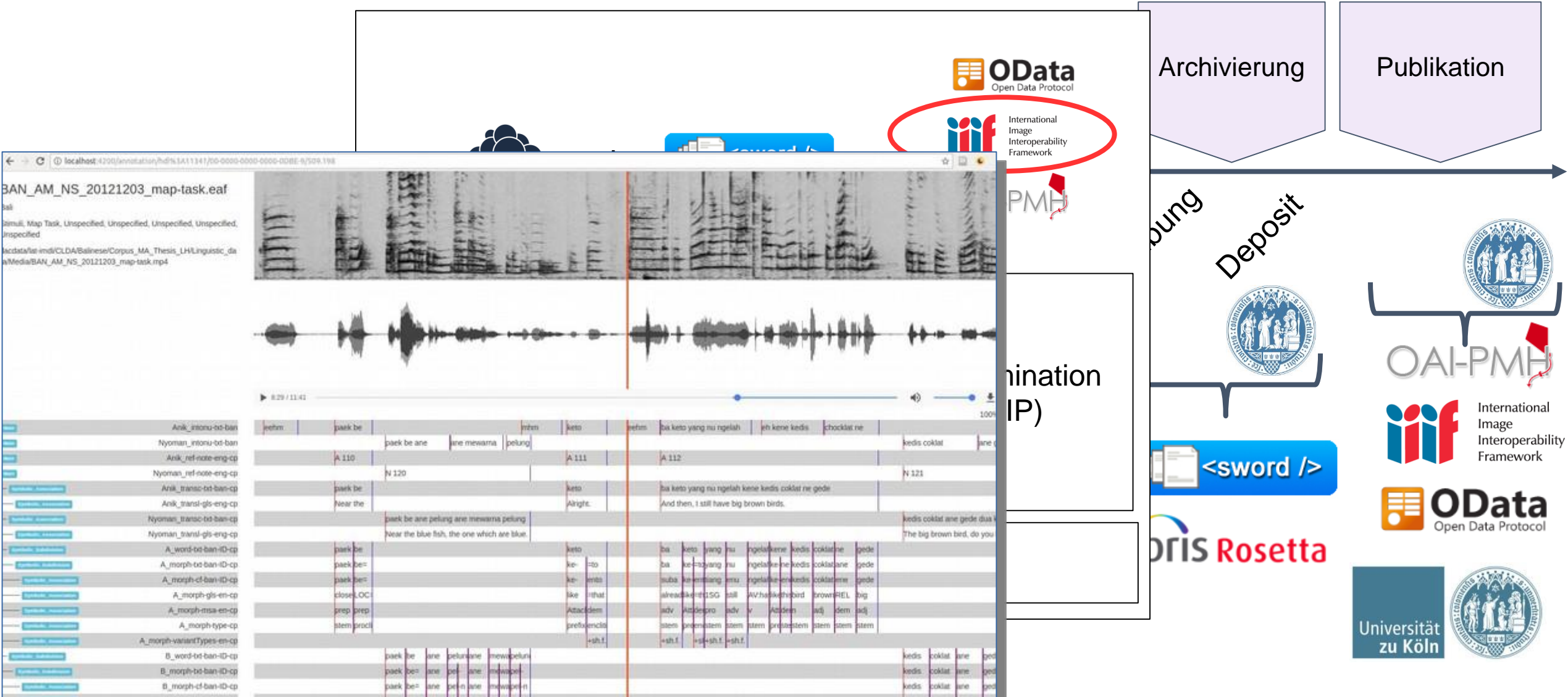
- Online: <https://bluejeans.com/222002812>
- Phone: +1.888.240.2560 (US Toll Free) or see [international numbers](#) - Enter Meeting ID: 22200

Feedback: iif-discuss@googlegroups.com Get involved: [Join IIF](#)

Video / Audio-Recording Meeting



Werkzeuge & Dienste



KA³ Infrastruktur

Anbieter

Regionales Rechenzentrum (RRZK)

- Standardisierte **APIs** für Deposit und Dissemination von AV-Daten (Sword, IIF, OData, OAI-PMH)
- **Drittanbieter Software** (OwnCloud, Rosetta, iFinder, AudioMining)
- **Deployment- und Entwicklungsumgebung** (Docker, Jenkins, Redmine)

Nutzer / Anbieter

Data Center for the Humanities (DCH)

- Fachwissenschaftliche **Portale** und Tools
- **Metadatenkuration**

Nutzer

Institut für Linguistik (IfL)

Danke!
jochen.graf@uni-koeln.de