



# KI, Sprachtechnologie und Digital Humanities: Ein (unvollständiger) Überblick

Georg Rehm

georg.rehm@dfki.de  
DFKI GmbH, Berlin

DH-Rundgang – 05. Februar 2018



Deutsches  
Forschungszentrum  
für Künstliche  
Intelligenz GmbH



# Überblick

- Einführung – Überblick über Forschungsaktivitäten
- Künstliche Intelligenz und Digital Humanities
- Forschungsdaten und Metadaten
- Kuratierungstechnologien
- Beobachtungen und Schlussfolgerungen

[SECTIONS](#) [SEARCH](#)

**Los Angeles Times**

TUESDAY JUN. 20, 2017 LOCAL SPORTS POLITICS ENTERTAINMENT OPINION MOST POPULAR PLACE AN AD

SUBSCRIBE SALE: 12 wks / 99¢ [LOG IN](#)

64°

# Toyota invests \$1 billion in artificial intelligence in U.S.



Akio Toyoda, president of Toyota Motor Corp., announces Nov. 6, 2015, during a news conference in Tokyo that Toyota will establish a new artificial intelligence research and development company. (Kimimasa Mayama / EPA)

By [Associated Press](#)

**In Case You Missed It**

 They dismissed her as a lightweight. But California's top judge has found her voice — and uses it to call out Trump policies

**cnn tech** BUSINESS CULTURE GADGETS FUTURE STARTUPS

[f](#) [t](#) [i](#) [o](#)

# Elon Musk and tech heavies invest \$1 billion in artificial intelligence

by Heather Kelly [@heatherkelly](#) [Dec 12, 2015: 4:08 PM ET](#)



Elon Musk is sinking large amounts of money into something that scares him: artificial intelligence.

**Social Surge - What's Trending**

-  Elizabeth Warren wants the Wells Fargo board wiped out
-  Jeff Bezos is \$5 billion away from being the world's richest person
-  Off-camera, no audio broadcast: White House keeps undermining press

The New York Times Magazine

Home Share 448

## The Great A.I. Awakening

How Google used artificial intelligence to transform Google Translate, one of its more popular services — and how machine learning is poised to reinvent computing itself.

BY GIDEON LEWIS-KRAUS DEC 14, 2016



The Economist  
JANUARY 7TH-13TH 2017  
**Now we're talking**  
Turkey torn apart  
Theresa Maybe, Britain's indecisive PM  
Nestlé goes on a health kick  
Meet China's Shakespeare  
Voice computing comes of age

**THE VERGE** TECH • SCIENCE • CULTURE • CARS • REVIEWS • LONGFORM • VIDEO • MORE • [f](#) [t](#) [i](#) [o](#)

POLICY & LAW US & WORLD TECH

## Scientists advising the US military say fears of an AI existential threat are 'uninformed'

There are plenty of regular threats though by James Vincent | [@jvincent](#) | Jan 20, 2017, 9:53am EST

f SHARE t TWEET in LINKEDIN



Public perceptions of military AI can lean toward the apocalyptic — understandable when prominent figures like Elon Musk warn us about the possible rise of [Terminator-style artificial intelligence](#). But when it comes to the judgments the military actually relies on, things are a little more sober. As a recent report into the use of military AI funded by the Department of Defense states: "most computer scientists, the claimed 'existential threats' posed by AI seem at best uninformed."

**NOW TRENDING**

-  You won't be able to use your Surface Book if you want to take the bar in Tennessee
-  Making the OnePlus 5
- 

# Daten Intelligenz



**Aktuelle Durchbrüche mit maschinellen Lernverfahren (Deep Learning)  
Ebenfalls noch immer im Einsatz: Symbolische, regelbasierte Methoden**

# Sprachtechnologie

- **Sprachtechnologie** setzt theoretische Ergebnisse der sprachorientierten Forschung in praxisgerechte und technologisch verwertbare Anwendungen um.
- Nutzt Ergebnisse u.a. aus:
  - **Künstliche Intelligenz + Informatik**
  - **Computerlinguistik**
    - **Natural Language Processing**
    - **Natural Language Understanding**
  - **Psychologie, Psycholinguistik**
  - **Kognitionswissenschaft**
- **Sprache: *The next big thing for AI!***

## Beispielanwendungen

- Rechtschreibprüfung
- Diktiersysteme
- Übersetzungssysteme
- Suchmaschinen
- Reportgenerierung
- Expertensysteme
- Dialogsysteme
- Textzusammenfassen

# Evolution und Kontext

- Historisch gesehen sehr enge Verbindung zwischen Digital Humanities und Computerlinguistik
- Journal *Computers and the Humanities* (erstmalig: 1966)
- Association for Literary and Linguistic Computing (1973)
- Association for Computers and the Humanities (1978)
- Text Encoding Initiative Guidelines (TEI, 1987, P1: 1994)
- Aus TEI entstand der W3C-Standard XML (1998)

# Forschungsdaten und Metadaten

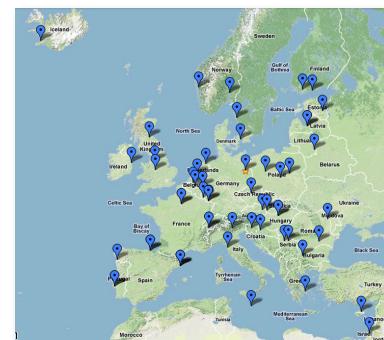


**60** research centres in **34** countries.

Chair of Executive Board: Jan Hajic (CUNI)

Dep.: J. van Genabith (DFKI), A. Vasiljevs (Tilde)

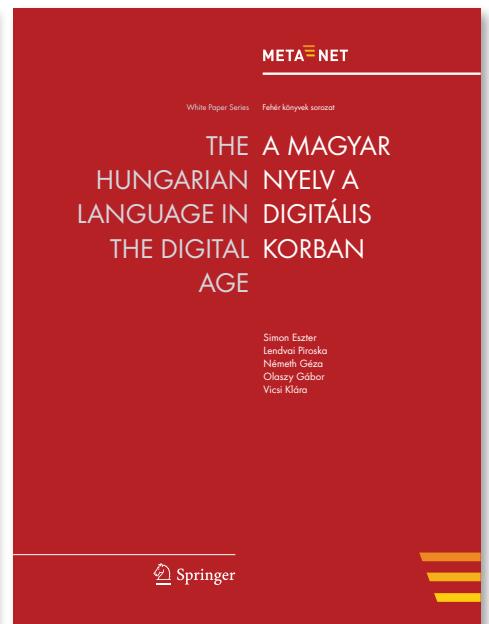
General Secretary: Georg Rehm (DFKI)



(published in 2013)



(31 volumes; published in 2012)



Multilingual Europe  
Technology Alliance.

**826** members in  
**67** countries



T4ME (META-NET)

CESAR

META-NORD

METANET4U

# META-SHARE: Was ist das?



- ❑ META-SHARE adressiert Probleme und Technologie- bzw. Infrastrukturbedarfe in Bezug auf die
  - ❑ Sichtbarkeit
  - ❑ Dokumentation
  - ❑ Identifizierung
  - ❑ Verfügbarkeit
  - ❑ Langzeitspeicherung
  - ❑ Interoperabilität
- ❑ von Sprachdaten und -ressourcen sowie Werkzeugen
- ❑ Implementiert als Netzwerk dezentraler Repositorys
  - ❑ Lokale Repositorys
  - ❑ Zentrale Repositorys
- ❑ <http://www.meta-share.org>

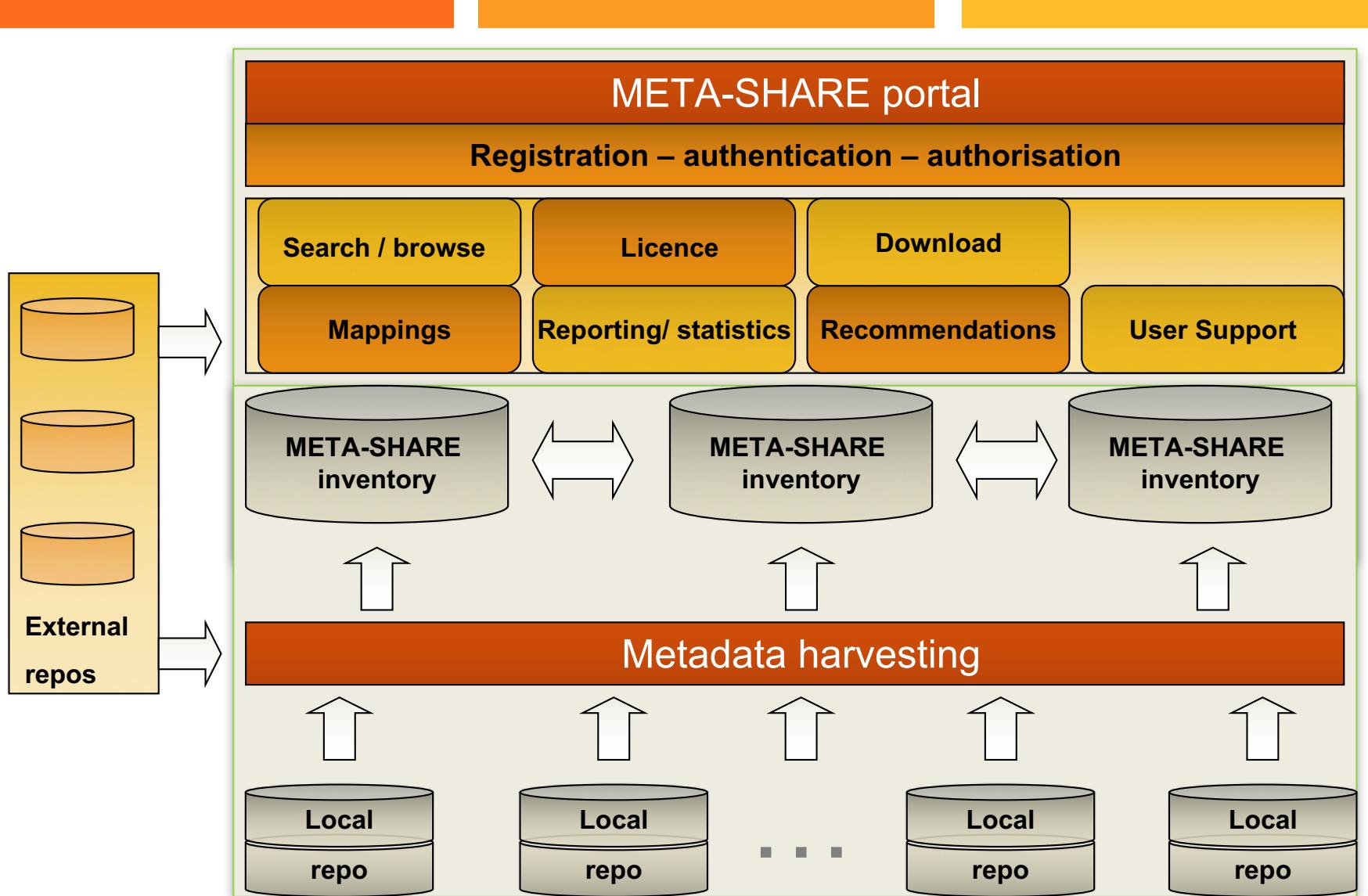
# META-SHARE: Was ist das?



- META-SHARE ist **offen** und **wächst**, **interoperabel** und sicher, es legt besonderes Augenmerk auf **Rechte** und **Workflows**.
- Zentrale Merkmale:
  - Kostenfrei benutzbar
  - Die zugrundeliegende Software ist frei verfügbar und Open Source
  - Unterstützung durch Helpdesk und Nutzerforum
- META-SHARE favorisiert offene Daten und Open Source
- Kostenpflichtige Ressourcen sind nicht ausgeschlossen
- Die kommerzielle Nutzung von Ressourcen soll stimuliert werden

# Architektur

METAFNET



# Die Seite des Providers



- Ein META-SHARE-Repository kann leicht installiert und angekoppelt werden:
  - Open Source Repository Software
  - Funktionen für das Dokumentieren, Aktualisieren, Beschreiben und Verknüpfen von Ressourcen
  - Hilfestellung (Helpdesk, Forum, Knowledge Base)
  - Jedes Repository umfasst ein Verzeichnis aller Metadaten und exportiert diese
  - Eingesammelte Metadaten werden in synchronisierten zentralen Servern gespeichert

The screenshot shows the 'META-SHARE backend for resource providers' interface. The top navigation bar includes 'Upload', 'Share/Create', 'Update', 'My Resources', and 'Community'. The user is logged in as 'dmavroelidis'. The main page displays a 'Greek Dependency Treebank' entry. The 'Required administration information' section contains fields for 'Resource name' (Greek Dependency Treebank), 'Language of this entry (IFC 3066 code, list)' (en), 'Description' (a detailed text about the resource), 'Resource short name' (GDT), 'Language of this entry (IFC 3066 code, list)' (en), 'Url' (http://gdt.ilsp.gr/), and 'Identifier' (a placeholder). The 'Distribution' section lists 'Contact person' (Despina Elina elina@ilsp.athena-innovation.gr) and 'Metadata creator' (Despina Elina elina@ilsp.athena-innovation.gr). The bottom right corner of the screenshot contains a small watermark: 'Digitale Kulturerbe - Eine Einführung'.

# Die Seite des Providers

- Ein META-SHARE-Repository kann leicht installiert und angekoppelt werden:
  - Open Source Repository Software
  - Funktionen für das Dokumentieren, Aktualisieren, Beschreiben und Verknüpfen von Ressourcen
  - Hilfestellung (Helpdesk, Forum, Knowledge Base)
  - Jedes Repository umfasst ein Verzeichnis aller Metadaten und exportiert diese
  - Eingesammelte Metadaten werden in synchronisierten zentralen Servern gespeichert

Add Distribution

Availability: available-unrestrictedUse

Specifies the availability status of the resource; restrictionsOfUse can be further used to indicate specific restrictions.

Licences (Hide)

Licence: #1

Licence:

AGPL  
ApacheLicence\_V2.0  
BSD-style  
CC

The licence of use for the resource

Restrictions of use:

academic-nonCommercialUse  
attribution  
commercialUse  
evaluationUse

Specifies the restrictions imposed by the type of the licence

Distribution access medium:

accessibleThroughInterface  
bluRay  
CD-ROM  
downloadable

Specifies the medium (channel) used for delivery or providing access to the resource

Download location:

Any url where the resource can be downloaded from

[+ Add Another Field](#)

# Die Seite des Konsumenten

META=NET

- Nutzer können
  - Ressourcen suchen
  - Durch den Katalog blättern

The screenshot shows the METAShare interface. At the top, there is a navigation bar with links for Home, Help, META=SHARE info, Editor, Profile, and Logout. Below the navigation bar is a search bar with a placeholder 'Search' and a magnifying glass icon. To the left of the main content area is a sidebar with various filter options: Language, Resource Type, Media Type, Availability, Licence, Restrictions of Use, Validated, Foreseen Use, Use is NLP Specific, Linguistic Type, Modality Type, Modality Type, MIME Type, Best Practices, Domain, Geographic Coverage, Time Coverage, and Subject. Below the filters are sections for Resource Type (Corpus, Lexical/Conceptual, Tool/Service, Language Description) and Media Type (Text, Audio, Image, Video). The main content area displays a list of 22 Language Resources (Page 1 of 2). Each resource entry includes a thumbnail, the resource name, and its availability status (Modern Greek (1653)). The resources listed are: ACCURAT Bilingual Comparable Corpora for under-resourced languages, COSIMOROE Travel Series Corpus, Cultural Parallels: Study and promotion of the cultural inheritance of the neighbouring areas in Greece and Bulgaria through internet technology, Cultural Thesaurus of the Greek Language, Database of Bulgarian sentences, Database of greek speech recordings, Educational Greek Corpus, EuroParl-NE Corpus, Greek Dependency Treebank, Greek Database from actual users, Greek PAROLE corpus, Greek Sign Language Corpus, Greek Textual Entailment Corpus, Hellenic National Corpus, INTERA Corpus, INTERBEG Corpus, Morphosyntactically annotated Greek Language Corpus, and MUSA Multilingual Multimodal Corpus.

# Die Seite des Konsumenten

META=NET

- Nutzer können
  - Ressourcen suchen
  - Durch den Katalog blättern
  - Ressourcen zugreifen, indem sie das Repository auswählen, die Lizenzen nutzen und die Ressourcen runterladen

The screenshot shows the METAShare interface. At the top, there is a navigation bar with links for Home, Help, META=SHARE info, Editor, Profile, and Logout. Below the navigation bar is a search bar with a 'Search' button. The main content area displays a list of '22 Language Resources (Page 1 of 2)'. Each resource entry includes a thumbnail icon, the resource name, and a small description. On the left side of the main content area, there is a sidebar with filtering options under 'Filter by:' and sections for 'Resource Type:' and 'Media Type:'.

Resource Name	Description
ACCURAT Bilingual Comparable Corpora for under-resourced languages	Corpus, English, German, French, Spanish, Modern Greek (1453), Renaissance, Byzantine
COSMOPOLIS Travel Series Corpus	English, Modern Greek (1453)
Cultural Parallels: Study and promotion of the cultural inheritance of the neighbouring areas in Greece and Bulgaria through internet technology	Bulgarian, Modern Greek (1453)
Cultural Thesaurus of the Greek Language	Modern Greek (1453)
Database of Bulgarian sentences	Bulgarian
Database of greek speech recordings	Modern Greek (1453)
Educational Greek Corpus	Modern Greek (1453)
EuroParl-NE Corpus	Modern Greek (1453)
Greek Dependency Treebank	Modern Greek (1453)
Greek Database from actual users	Modern Greek (1453)
Greek PAROLE corpus	Modern Greek (1453)
Greek Sign Language Corpus	Modern Greek (1453)
Greek Textual Entailment Corpus	Modern Greek (1453)
Hellenic National Corpus	Modern Greek (1453)
INTERA Corpus	Bulgarian, English, Modern Greek (1453), Serbian, Bosnian
INTERBEG Corpus	Bulgarian, English, Modern Greek (1453)
Morphosyntactically annotated Greek Language Corpus	Modern Greek (1453)
MUSA Multilingual Multimodal Corpus	English, French, Modern Greek (1453)

# Die Seite des Konsumenten



- Nutzer können
  - Ressourcen suchen
  - Durch den Katalog blättern
  - Ressourcen zugreifen, indem sie das Repository auswählen, die Lizenzen nutzen und die Ressourcen runterladen

A screenshot of a web page from the METASHARE beta platform. At the top, there's a navigation bar with the METASHARE logo and a 'beta' indicator. Below the header, the title 'Hunglish Corpus' is displayed. A large central area contains a 'Licence Agreement' section. It includes a detailed explanation of the Creative Commons Public License (CCPL) terms, a 'License' section, and a '1. Definitions' section with a detailed legal definition of 'Adaptation'. At the bottom of this section is a checkbox labeled 'I agree to these licence terms and would like to download the resource.' followed by a yellow 'Download' button. At the very bottom of the page, there's a footer note about funding from the European Commission through various projects.

# Die Seite des Konsumenten



- Nutzer können
  - Ressourcen suchen
  - Durch den Katalog blättern
  - Ressourcen zugreifen, indem sie das Repository auswählen, die Lizenzen nutzen und die Ressourcen runterladen

The screenshot shows a detailed view of a resource entry on the METASHARE platform. The top navigation bar includes links for Administration, Community, Documentation, Statistics, and Your Profile. The main title is "COSMOROE Caricature Corpus". Below the title, a brief description states: "Corpus of 50 caricatures by Yannis Ioannou covering the domains of politics and environment. They pertain to various types of visual-linguistic information. They have been annotated in different annotation levels in order to be used for the detection of the semantic relations between image and language." Three buttons are available: "Back", "Download", and "Edit Resource". The page is divided into several sections: "Distribution" (Availability: Available - Unrestricted Use, Licence: Proprietary, Contact person: Surname: Pastera, Email: kpastera@ilsp.gr, Communication: Organization name: Institute for Language and Speech Processing / Athena R.C., Email: www@ilsp.athena-innovation.gr), "Text" (Corpus text, Media: text, Linguality: Monolingual, Language: Language id: el, Language name: Modern Greek (1453-)), "Size" (Size: 50, Size unit: Images), "Annotation" (Annotation type: Speech Annotation, Modality Annotation - Body Movements, Annotation type: Modality Annotation - Hand Arm Gestures), and "Metadata" (Metadata creation date: 2012-02-02, Metadata creator: Surname: Labropoulou, Communication: Email: penny@ilsp.athena-innovation.gr, Organization name: Institute for Language and Speech Processing / Athena R.C., Communication: Email: www@ilsp.athena-innovation.gr). A note at the bottom indicates that people who looked at this resource also viewed the following: "Vocabulary/Word List for Greek Sign Language Synthesis" and "Hellenic National Corpus".

# Die Seite des Konsumenten



- Nutzer können
  - Ressourcen suchen
  - Durch den Katalog blättern
  - Ressourcen zugreifen, indem sie das Repository auswählen, die Lizenzen nutzen und die Ressourcen runterladen
  - Hilfestellung erhalten (Online-Forum und Helpdesk)
  - Eine Wissensbasis nutzen

The screenshot shows the META-SHARE Support Community forum. At the top, there are navigation links for 'Forum', 'Knowledge base', 'Search the forum', 'Administrators', 'dina@vanda', and 'log out'. Below the header, there's a search bar and a 'New question' button. The main area displays a list of questions. The first question is titled 'Schema location missed?' and was posted by Marta Villegas on 15-05-2012 at 08:46. It has 18 views and 1 answer. The second question is 'Payed resources' by Jérôme Caenro on 10-05-2012 at 14:33. The third is 'Some issues: META-SHARE editor' by Catarina Carvalho on 09-05-2012 at 14:16. Other visible questions include 'Using Google Analytics with META-SHARE', 'open source GNU GLP resources signed by the right holders', 'UIB comments regarding META-SHARE editor (beta version, 2.0)', 'Multiple editors for a given resource', 'Definition of NoRedistribution? (vs. No Derivatives)', 'UIB comments on the META-SHARE beta editor (II)', and 'Schema location for metashare v2?'. On the right side of the interface, there are sections for 'Sort questions' (Newest, Most active, Most discussed) and 'Editor actions' (Unpublished questions, Trashed questions).

# Metadaten für LRs



- Metadaten unterstützen META-SHARE Nutzer bei allen Services
  - Beschreibung von Ressourcen (Erzeugung, Speicherung etc.)
  - Browsing, Suche, Retrieval
  - Upload und Download
  - Regelmäßiges Metadaten-Harvesting
  - Monitoring von Ressourcen und verwandten Objekten
- Nutzer sind sowohl Bereitsteller von Daten als auch Nutzer von Daten

Heterogeneous situation outside the life cycle dictates and provides external factors and forces:

- LR/LT ecosystem and landscape with available resources and technologies for several languages.
- Gaps in terms of resources, technologies, languages.
- Regional, national, international data creation strategies, initiatives and goals to fill gaps and to foster research, innovation and development.

### Start of the Life Cycle

- *Linguistic research question* (incl. language documentation and preservation, creation of a resource in a new language based on a popular resource in some other language) or
- *Technology development goal*
- Project goal dictates type of LR to be created (lexicon, corpus, ontology etc.)
- Project goal also dictates some of the subsequent phases of the life cycle.

### Life Cycle continues ...

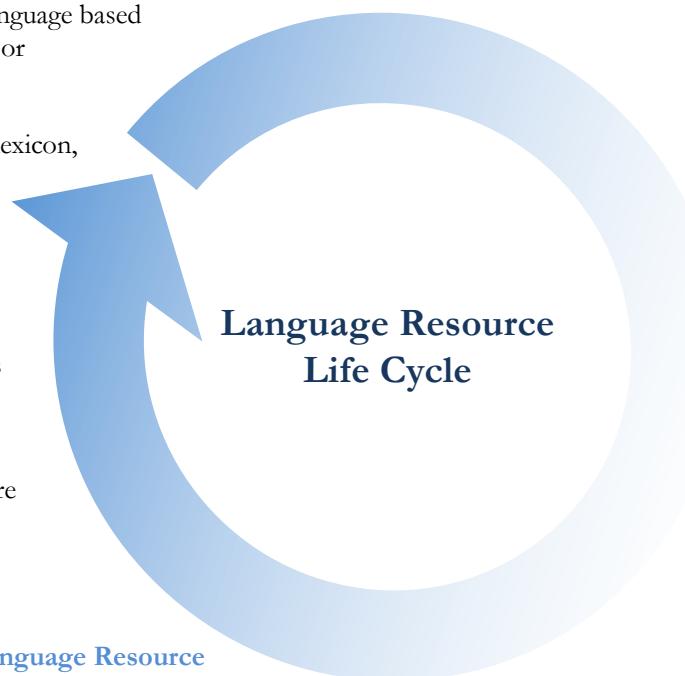
- Maintenance checks or user feedback: any errors in the LR? Corrections or data format upgrades needed? Release patch or update.
- New version or new research: acquisition of more data and subsequent annotation and/or annotation of additional linguistic analysis levels (sharing and re-use of language resources).

### Phase 7: Distribution and Publication of the Language Resource

- IPR and legal issues: license agreement between user/OEM and distributor
- Distribution through web interface
  - Querying: metadata, data annotations, both
  - Visualisation
  - Optional: distribution as stand-alone package
- Distribution through APIs (Linked data, REST etc.)
- Distribution as stand-alone package
  - LR provider distributes through own mechanism/website
  - Through a centralised or decentralised distribution infrastructure; these have their own life cycles: acquisition of LR, management of LR, maintenance, distribution, support

### Phase 1: Data Acquisition and Data Collection

- Data source:
  - *New data* (harvested from the web, acquired using specialised tools, outsourced to service provider etc.)
  - *Existing (legacy) data* in arbitrary research-oriented formats (alternative **start** or **continuation** of cycle)
- IPR and legal issues: who is the rightful owner or copyright holder of the primary data?



### Phase 2: Data Curation and Data Annotation

- Selection of linguistic analysis levels
- Formats, tag sets, schemas
  - Standards (SGML, XML, TEI, CES, LAF etc.)
  - Syntax and semantics registered in recognised repository
- Tools (GUIs etc.; software has its own life cycle)
- Annotation: manual, semi-automatic, fully automatic
- Definition/identification of objects through PIDs
- Data management (XML databases, editors, revision control systems etc.)

### Phase 3: Linguistic Analysis and Research

- Querying
  - Metadata
  - Data and their annotations
- Analysis: manual, semi-automatic, fully automatic
- Visualisation
- Combining data and annotations from various different sources
  - Interoperability standards

### Phase 4: Evaluation and Quality Control

- LR evaluation, quality evaluation
- Validation (schema compliance etc.)
- Reviews
- Independent assessments
- Publications
- Experiments

### Phase 5: Description

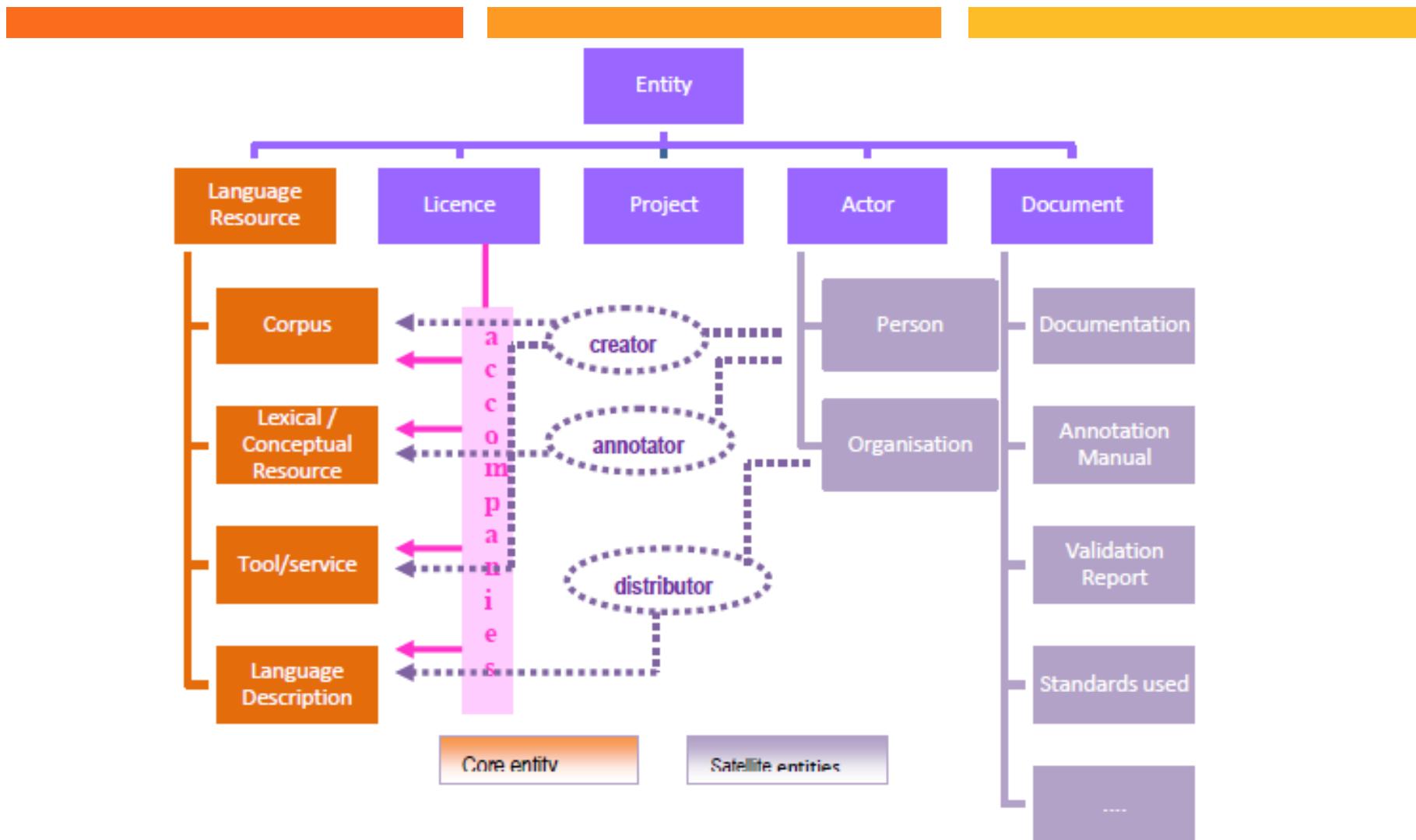
- Metadata records
  - Standards (OLAC, IMDI etc.)
- IPR and legal issues: who owns the annotations?
- Documentation, manuals, scientific publications
- Software used or implemented to create the LR

### Phase 6: Packaging

- Any constraints on the data and annotations based on IPR and legal aspects?
- Example: sentence shuffling

- Entitäten
  - **core entity** – die Sprachressource
  - **satellite entities** – verwandte Objekte, z.B.
    - Involvierte *actor: persons* und *organisations*, z.B. Erzeuger der Ressource, Drittmittelgeber, Distributoren etc.
    - Per *document*: werden Dokumente referenziert, z.B. Artikel, Papiere, Reports, Tagset-Beschreibungen etc.
    - Ferner können via *project*: Projekte beschrieben werden, die an der Schaffung einer Ressource beteiligt waren oder die sie nutzen
    - *licence*: zur Distribution einer Ressource

# Ontologie (Auszug)



# Typen von Ressourcen



- Zwei Klassifikationsachsen:
  - *resourceType*  
und
  - *mediaType*
- Jede Ressource besitzt einen Wert für *resourceType*, kann aber mehrere *mediaType* Werte umfassen, d.h. Ressourcen können aus Teilen bestehen, die zu unterschiedlichen Typen von Medien gehören

- **Das minimale Schema:**

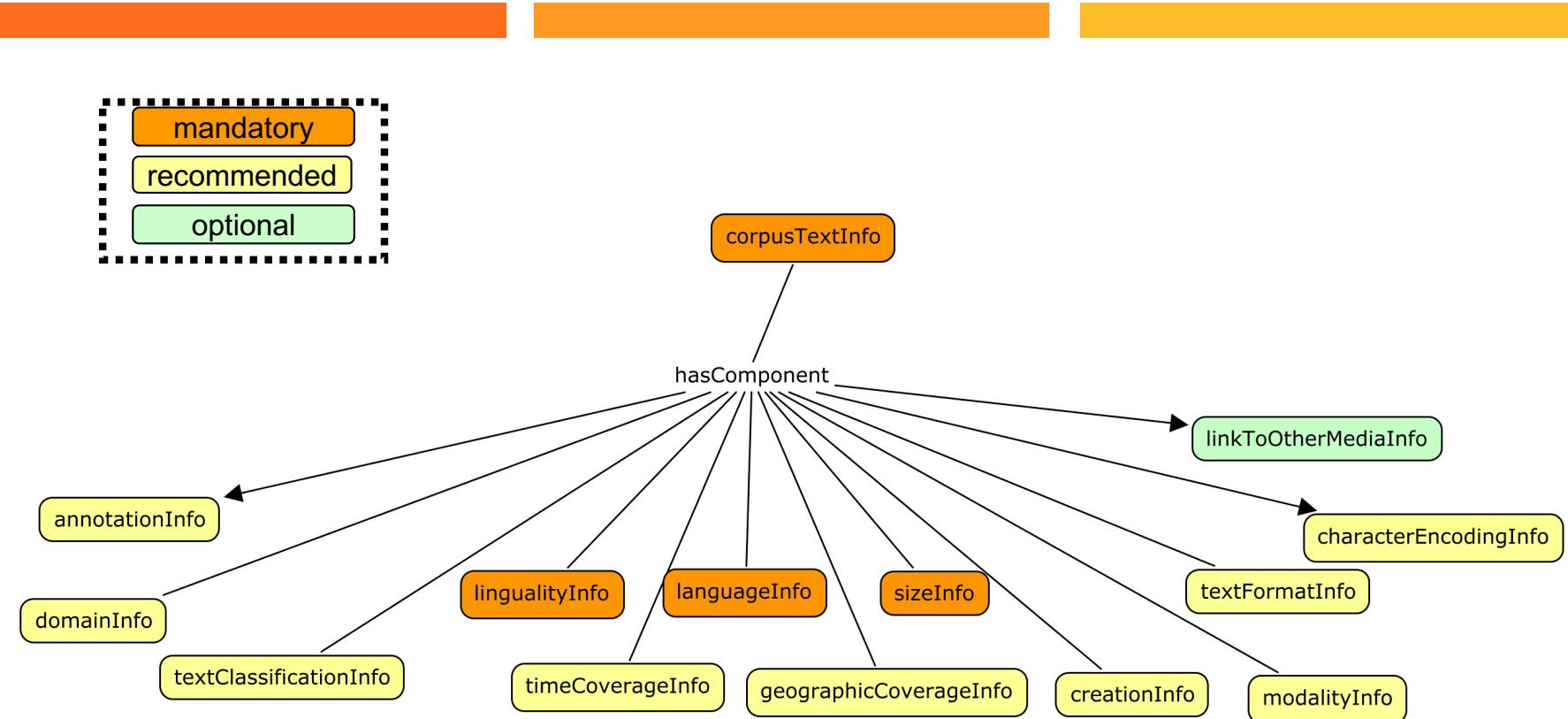
- Minimales Inventar von obligatorischen Elementen und Relationen, die für die effektive Suche, Identifizierung und Retrieval benötigt werden
- Beispiele: Identifizierung (Title, ID), Kontaktdetails, technische Informationen (Sprache, Inhalte, Format, Größe etc.)

- **Das maximale Schema:**

- Alle Elemente und Relationen, die für die Beschreibung einer Ressource benötigt werden, d.h. ein zusätzliches Set von empfohlenen und optionalen Elementen und Relationen für den vollen LR-Lebenszyklus (Produktion, Nutzung)
- Beispiele: Herkunft, Ursprung, Erzeugung, Validierung, Evaluation, intendierte Nutzung etc.

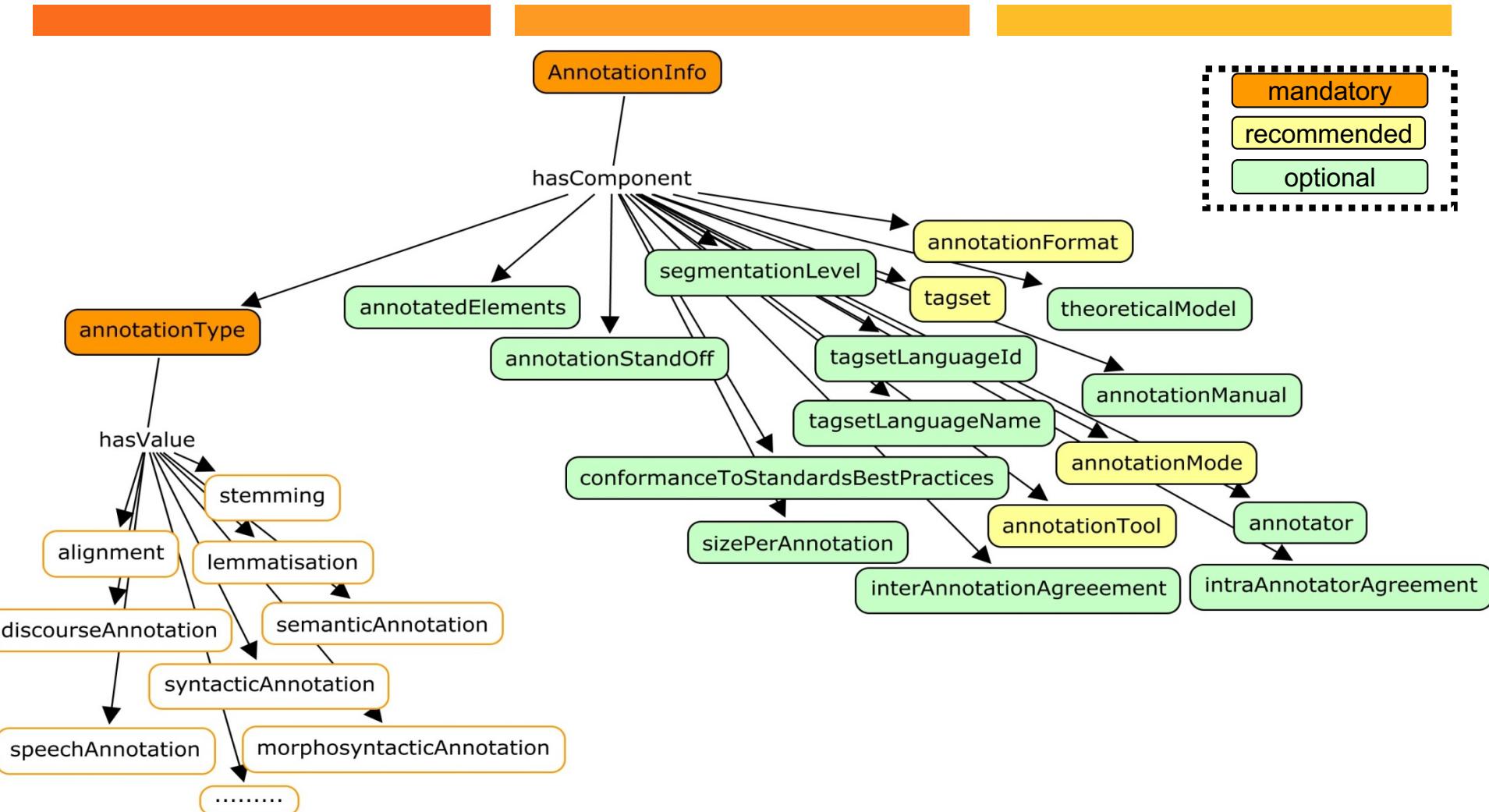
# corpusTextInfo

META=NET



# Modul für Annotationen

META=NET



- **Creative Commons** Lizenzen, z.B. Creative Commons Zero (CC-0) und alle möglichen CC-Kombinationen
- **META-SHARE Commons** Lizenzen, auf CC-basierende Lizenzen, die es META-SHARE-Mitgliedern erlauben, ihre Ressourcen nur innerhalb des META-SHARE-Netzwerkes zu teilen
- **META-SHARE “No Redistribution”** Lizenzen – erlauben die Nutzung einer Ressource während der Urheber weiterhin die volle Kontrolle bzgl. Distribution hat.
- Tools und Web Services werden zur Verfügung gestellt durch **Standard-Open-Source-** oder **kommerzielle Lizenzen**.

# Das neue META-SHARE

META<sup>≡</sup>NET

The screenshot shows the homepage of the new META-SHARE platform. At the top, there is a navigation bar with four colored horizontal bars (orange, yellow, orange, yellow) followed by the META-SHARE logo and a menu with links: LEARN ▾, DISCOVER ▾, PARTICIPATE ▾, CONNECT ▾, and a LOGIN link. Below the header, the main content area features a dark background with various icons related to language resources and technology. On the left, there is a large call-to-action section with the heading "Search & exchange language resources" and a subtext explaining that META-SHARE is an open and secure network of repositories for sharing and exchanging language data, tools and related web services. It includes a search bar with placeholder text "Search the META-SHARE inventory" and a magnifying glass icon, along with a "OR LEARN MORE" link. On the right, there is a "Share your own resources!" section with a "JOIN OUR NETWORK NOW" button and a "Log in" link for existing members. At the bottom, there are four large, stylized orange and yellow text blocks with accompanying diagonal lines: "1,075 users", "2,733 language resources", "30 % text corpora", and "6,641 number of downloads".

LEARN ▾ DISCOVER ▾ PARTICIPATE ▾ CONNECT ▾ LOGIN

Search & exchange language resources

META-SHARE is an open and secure network of repositories for sharing and exchanging language data, tools and related web services

Search the META-SHARE inventory

OR LEARN MORE

Share your own resources!

JOIN OUR NETWORK NOW

Already a member? [Log in](#)

1,075 users

2,733 language resources

30 % text corpora

6,641 number of downloads

# In fünf Klicks zur Ressource

META<sup>E</sup>NET

The screenshot shows the homepage of the META-SHARE platform. At the top, there is a navigation bar with three horizontal bars of increasing length from left to right, colored orange, yellow, and light yellow. The navigation menu includes links for LEARN, DISCOVER, PARTICIPATE, CONNECT, and LOGIN. The main header features the META-SHARE logo. Below the header, there is a dark banner with various icons related to language resources and tools. On the left side of the banner, the text "Search & exchange language resources" is displayed, followed by a description of META-SHARE as an open and secure network of repositories. A search bar with the placeholder "Search the META-SHARE inventory" and a magnifying glass icon is located below the banner. To the right, there is a call-to-action button labeled "JOIN OUR NETWORK NOW" with a user icon, and a link for "Already a member? Log in". At the bottom of the page, there are four large, stylized orange and yellow downward-pointing arrows, each accompanied by a statistic: "1,075 users", "2,733 language resources", "30 % text corpora", and "6,641 number of downloads".

LEARN ▾ DISCOVER ▾ PARTICIPATE ▾ CONNECT ▾ LOGIN

**META<sup>E</sup>SHARE**

Search & exchange language resources

META-SHARE is an open and secure network of repositories for sharing and exchanging language data, tools and related web services

Search the META-SHARE inventory

OR LEARN MORE >

Share your own resources!

JOIN OUR NETWORK NOW

Already a member? [Log in](#)

1,075 users

2,733 language resources

30 % text corpora

6,641 number of downloads

# 1) Suche

META=NET

The screenshot shows the META-SHARE search interface. At the top, there is a navigation bar with links for 'Browse Resources', 'Manage Resources', 'Community', 'Statistics', 'Help', 'About', and 'Your Profile, Stelios'. On the far right of the bar is a 'Logout' link. Below the navigation bar, the main title 'META=SHARE' is displayed in large, bold letters. Underneath the title, it says '2,327 language resources at your disposal'. A search bar contains the text 'greek environment web text corpus' and a yellow 'Search' button. To the left of the search bar is a graphic of overlapping speech bubbles in various colors (yellow, green, blue, orange). Below this graphic, the word 'META=SHARE' is written again. To the right of the search bar, under the heading 'What is it? - About the project', there is a detailed description of the project's purpose and implementation. At the bottom of the page, there is a 'Go to top' button with a triangle icon.

greek environment web text corpus

Search

What is it? - About the project

META-SHARE, the open language resource exchange facility, is devoted to the sustainable sharing and dissemination of language resources (LRs) and aims at increasing access to such resources in a global scale.

META-SHARE is an open, integrated, secure and interoperable sharing and exchange facility for LRs (datasets and tools) for the Human Language Technologies domain and other applicative domains where language plays a critical role.

META-SHARE is implemented in the framework of the **META-NET Network of Excellence**. It is designed as a network of distributed repositories of LRs, including language data and basic language processing tools (e.g., morphological analysers, PoS taggers, speech recognisers, etc.).

Go to top ▲

The META-SHARE Infrastructure

## 2) Auswahl

The screenshot shows the METASHARE search interface. At the top, there is a navigation bar with links: Browse Resources, Manage Resources, Community, Statistics, Help, About, and Your Profile, Stelios. On the far right of the bar is a Logout button. Below the navigation bar is a search bar containing the text "greek environment web text corpus". To the right of the search bar is a yellow "Search" button. The main content area displays a search result titled "5 Language Resources". The results are ordered by "Resource Name A-Z". Each result item includes a thumbnail icon, the resource title, its language, and download statistics (0 downloads, 14 views). The results listed are:

- PANACEA Annotated Dependency Greek Environment Corpus Version 2 (Greek, Modern (1453-))
- PANACEA Annotated Dependency Italian Environment Corpus Version 2 (Italian)
- PANACEA English-French and English-Greek parallel corpus acquired for Environment domain (English, French, Greek, Modern (1453-))
- PANACEA Environment Corpus n-grams EL (Greek) (Greek, Modern (1453-))
- PANACEA Environment Greek monolingual corpus (Greek, Modern (1453-))

On the left side of the main content area, there are two sections: "Filter by:" and "Resource Type:". The "Filter by:" section contains links for Language, Resource Type, Media Type, Availability, Licence, Restrictions of Use, Linguistic Type, Multilinguality Type, MIME Type, and Domain. The "Resource Type:" section lists Corpus, Lexical/Conceptual, Tool/Service, and Language Description, each with a corresponding icon.

### 3) Beschreibung

The screenshot shows the METASHARE platform interface. At the top, there is a navigation bar with links: Browse Resources, Manage Resources, Community, Statistics, Help, About, Your Profile, Stelios, and Logout. Below the navigation bar, the title "PANACEA Environment Corpus n-grams EL (Greek)" is displayed, along with its URL: [http://nlp.ilsp.gr/panacea/D4.3/data/201209/gms/env\\_el/README.txt](http://nlp.ilsp.gr/panacea/D4.3/data/201209/gms/env_el/README.txt). A small status indicator shows 372 items, 3 checked, and 26 pending.

The main content area is divided into several sections:

- Distribution**:
  - Availability**: Available - Restricted Use
  - Licence**: CC - BY - SA
  - Restrictions**: Attribution, Share Alike
  - Download location**: hidden
  - Distribution Access/Medium**: Downloadable
  - Attribution Details**: This LR has been created by Athena R.C./ILSP ([www.ilsp.gr](http://www.ilsp.gr)) and is licensed under a CC-BY-SA licence
- Contact Persons**:
  - Prokopis Prokopidis ([View profile](#))
  - Vassilis Papavassiliou ([View profile](#))
- textngram**:
  - Monolingual textngram corpus**:
    - Languages**: Greek, Modern (1453-)
    - Linguality**: Linguality type: Monolingual
    - Text Format**: text/plain
    - Size**: 14,954,020 5 - Grams, 13,683,940 4 - Grams, 3,860,716 Bigrams, 9,767,383 Trigrams, 435,189 Unigrams
    - Domains**: environment
    - NGram**: Order: 5, Base item: Word
    - Creation**:
      - Creation mode details**: automatic web crawling, automatic language detection, data preprocessing (boilerpipe filtering, lemmatization & tagging)
      - Creation Tools**: ILSP-Lemmatizer, <http://code.google.c...>
- Resource Creation**:
  - Creation lasted**: 01/06/2011 - 31/08/2011
  - Funding Project**: Platform for Automatic, Normalized Annotation and Cost-Effective Acquisition of Language Resources for Human Language (PANACEA - ICT-248064)
    - URL**: <http://www.panacea-i...>
    - Funding Type**: Eu Funds
    - Funder**: ICT
    - Project duration**: 01/01/2010 - 31/12/2012
  - Metadata**:
    - Created**: 22/10/2012
    - Last Updated**: 22/10/2012
    - Metadata Language**: English (en)
    - Metadata Creator**: Penny Labropoulou ([View profile](#)), Prokopis Prokopidis ([View profile](#))
    - Version**: v1.0
    - Documentation**:
      - Samples Location**: <http://nlp.ilsp.gr/p...>, <http://nlp.ilsp.gr/p...>

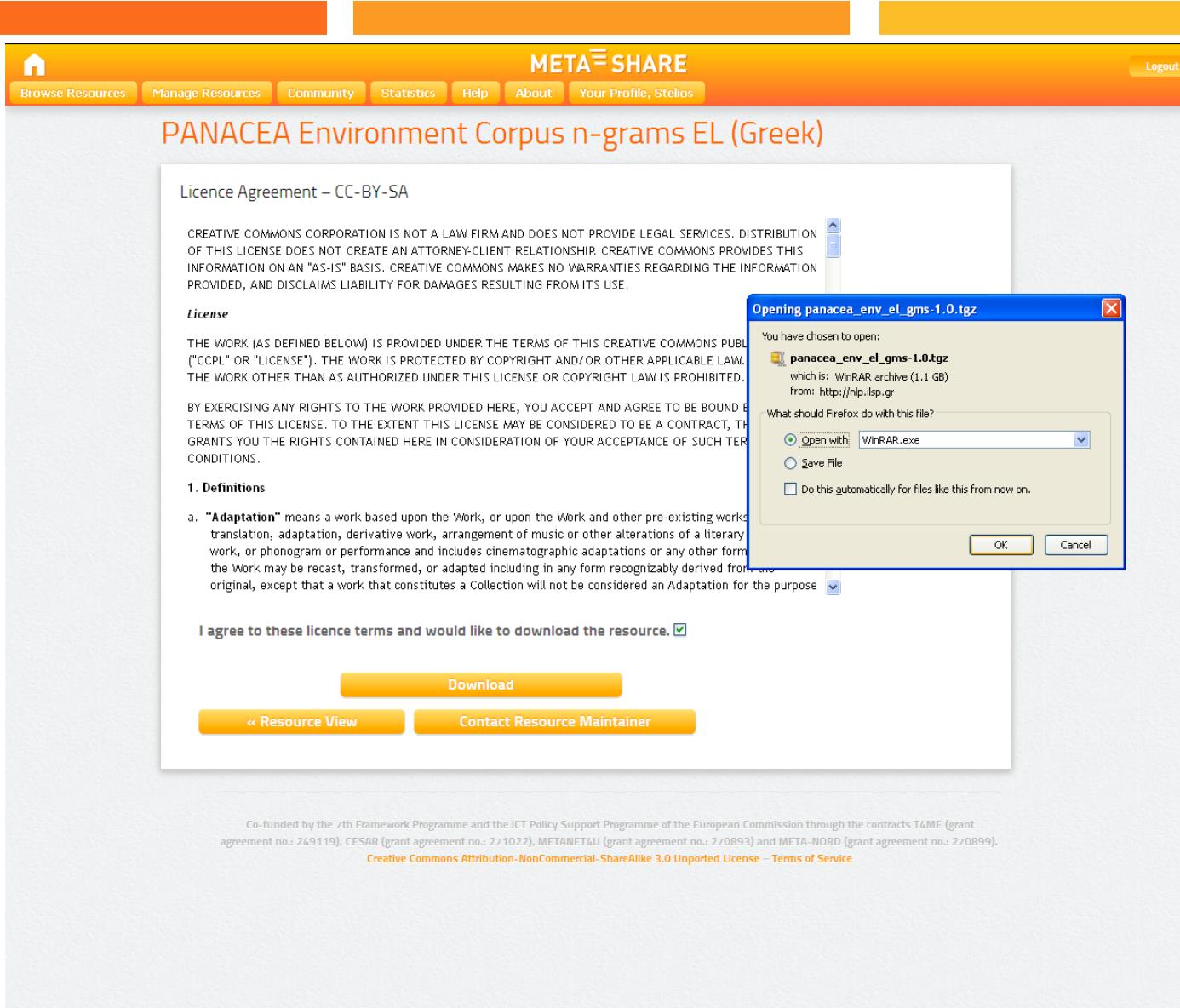
## 4) Lizenz



The screenshot shows a web interface for METASHARE. At the top, there is a navigation bar with links for 'Browse Resources', 'Manage Resources', 'Community', 'Statistics', 'Help', 'About', and 'Your Profile, Stelios'. On the far right of the bar is a 'Logout' button. Below the navigation bar, the title 'PANACEA Environment Corpus n-grams EL (Greek)' is displayed. The main content area is titled 'Licence Agreement – CC-BY-SA'. It contains a disclaimer from Creative Commons about the nature of the license. Below the disclaimer, there is a section titled 'License' which describes the terms of the CC-BY-SA license. Another section follows, detailing the terms of the license. A heading '1. Definitions' is present, followed by a detailed definition of 'Adaptation'. At the bottom of the main content area, there is a checkbox labeled 'I agree to these licence terms and would like to download the resource.' with a checked box. Below the checkbox are three buttons: 'Download', '« Resource View', and 'Contact Resource Maintainer'. At the very bottom of the page, there is a footer note about co-funding by the European Commission and links to Creative Commons license terms.

Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME (grant agreement no.: 269119), CESAR (grant agreement no.: 271022), METANET4U (grant agreement no.: 270893) and META-NORD (grant agreement no.: 270899).  
Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License – Terms of Service

# 5) Download



The screenshot shows the METASHARE interface. At the top, there is a navigation bar with links for 'Browse Resources', 'Manage Resources', 'Community', 'Statistics', 'Help', 'About', and 'Your Profile, Stelios'. On the right side of the header is a 'Logout' button. Below the header, the title 'PANACEA Environment Corpus n-grams EL (Greek)' is displayed. A modal window titled 'Licence Agreement – CC-BY-SA' is open, containing the Creative Commons Attribution-ShareAlike license text. To the right of the modal, a file download dialog box titled 'Opening panacea\_env\_el\_gms-1.0.tgz' is shown, asking what Firefox should do with the file. The options are 'Open with WinRAR.exe' (selected), 'Save File', and 'Do this automatically for files like this from now on'. At the bottom of the page, there is a footer with text about funding and a link to the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License – Terms of Service.

Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME (grant agreement no.: 249119), CESAR (grant agreement no.: 271022), METANET4U (grant agreement no.: 270893) and META-NORD (grant agreement no.: 270899).  
Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License – Terms of Service

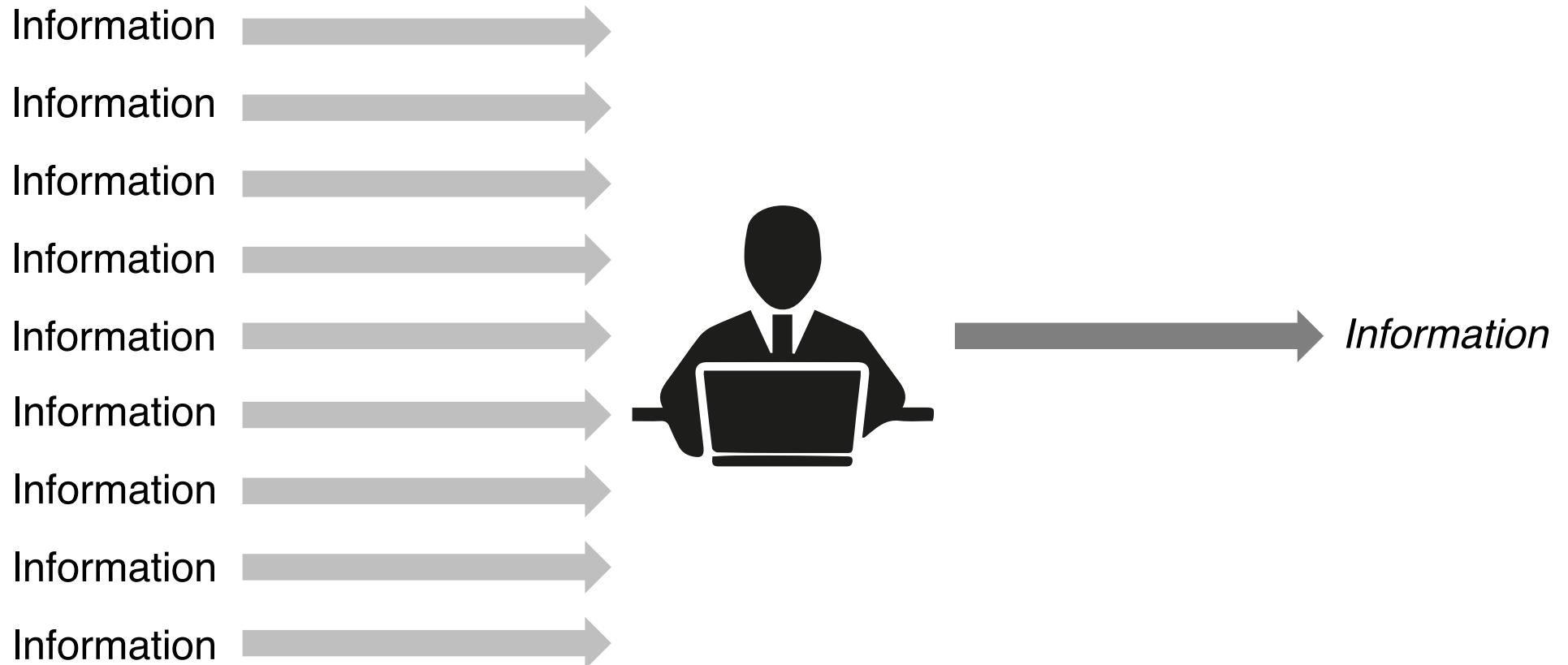
# META-SHARE heute



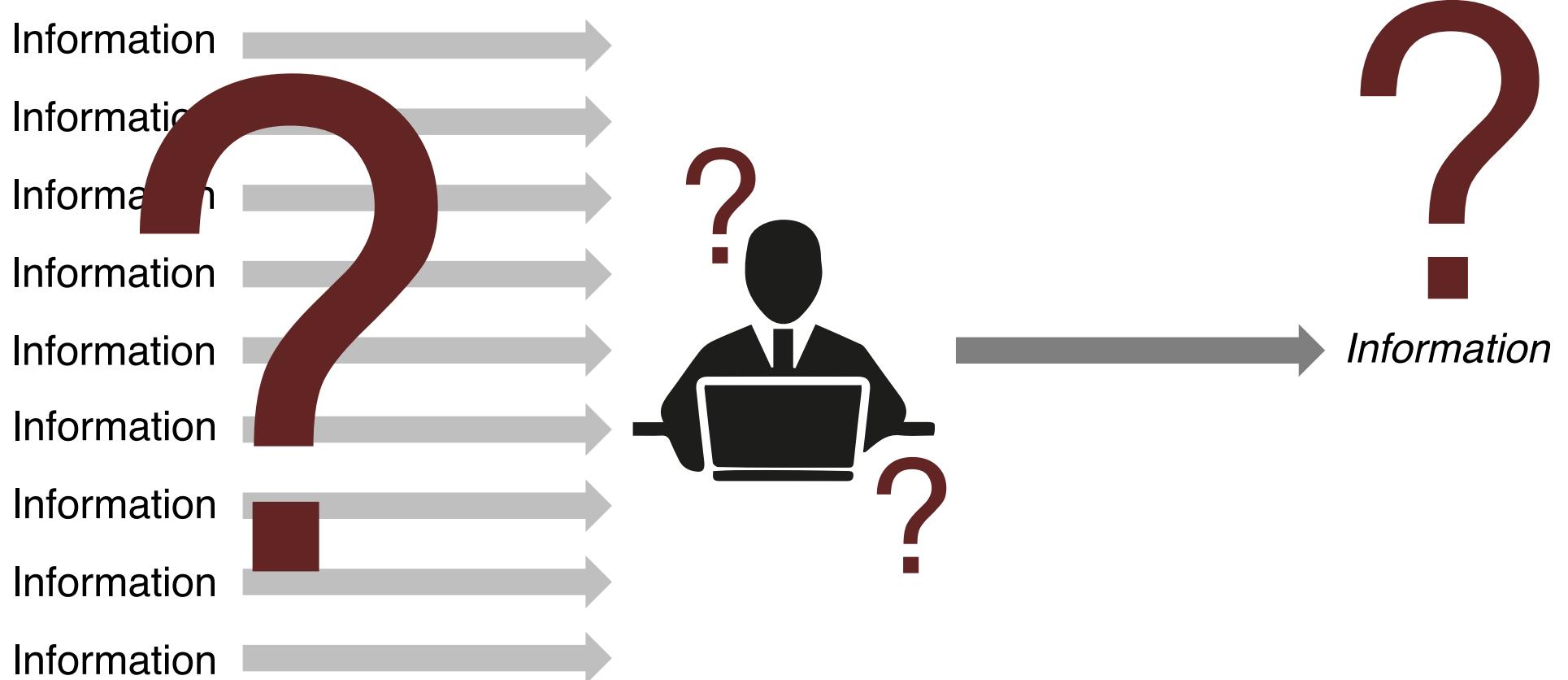
- ❑ 35 META-SHARE-Mitglieder/Organisationen in 25 Ländern
- ❑ 28 Repositorys
- ❑ 3.000+ Sprachressourcen
- ❑ 500.000+ Views
- ❑ 22.000+ Downloads von CC-lizenzierten Ressourcen
- ❑ Nachhaltigkeit durch das gesamte Netzwerk – die meisten Mitglieder haben sich zu Betrieb und Pflege verpflichtet
- ❑ META-SHARE wird benutzt als Basis des ELRC-SHARE Repositorys für Public-Sector-Language-Data –  
<https://www.elrc-share.eu>
- ❑ META-SHARE inspiriert bzw. wird benutzt für diverse nationale CLARIN-Infrastrukturen (GR, EE)

# Kuratierungstechnologien

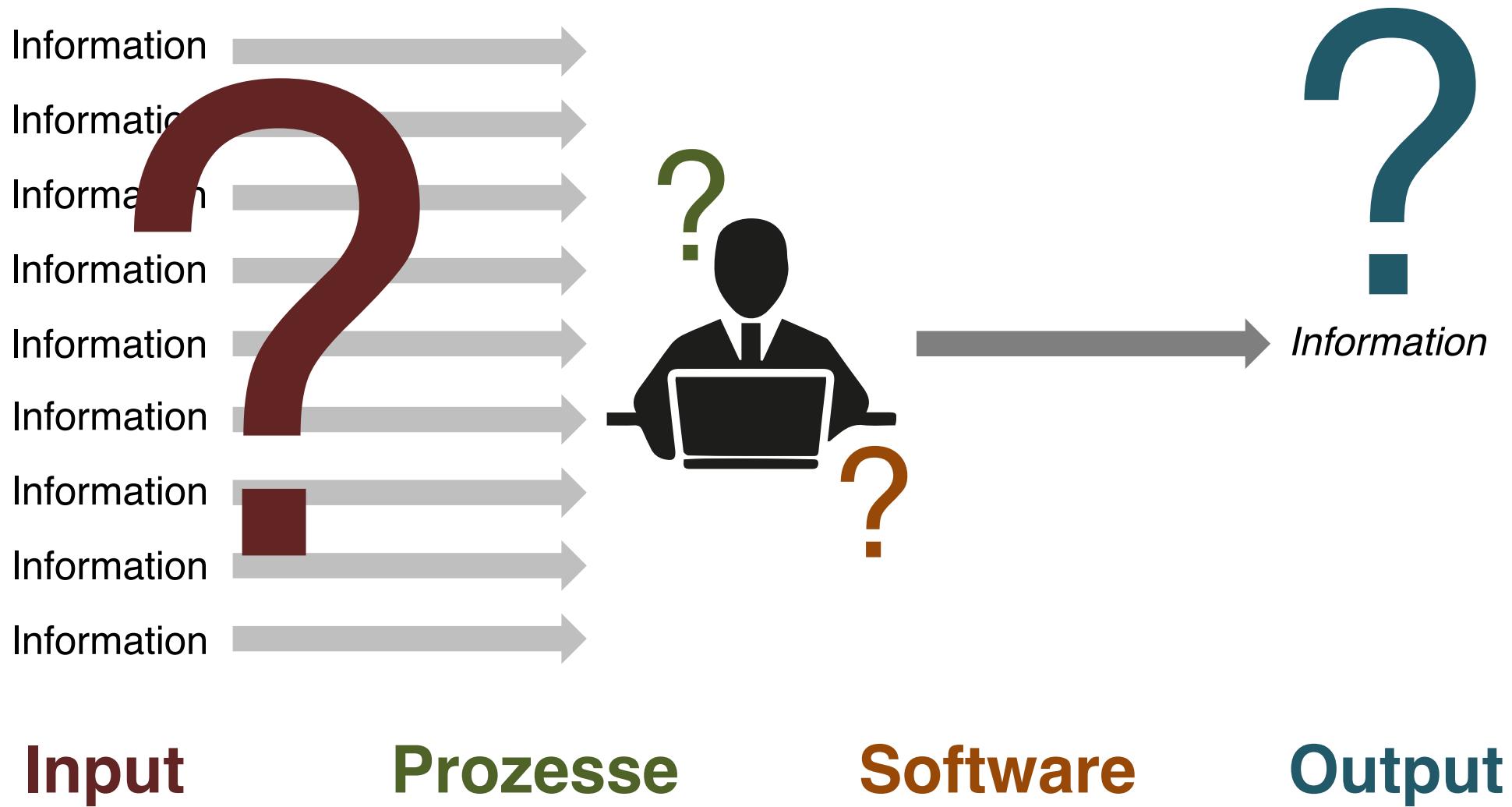
# Was ist digitale Kuratierung?



# Was ist digitale Kuratierung?



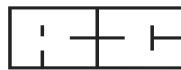
# Was ist digitale Kuratierung?



Branchen			
Input	Prozesse	Software	Output
Tweet Zeitungsa rtikel Agenturmeldung Facebook-Meldung Suchergebnis Email SMS Konzept Textdateien Video Karte Stockfotos In-house Datenbank Kalendereintrag Spreadsheets Archiv etc.	Analysieren Auswählen Fokussieren Überarbeiten Einlesen Schreiben Gestalten <b>Recherchieren</b> <b>Bewerten</b> <b>Evaluieren</b> <b>Ordnen</b> <b>Sortieren</b> <b>Strukturieren</b> <b>Zusammenfassen</b> Kürzen Übersetzen Informieren Kombinieren Abstrahieren Einordnen Visualisieren Generieren Annotieren Referenzieren etc.	Textverarbeitung Präsentationen Tabellenkalkulation Email Browser Groupware Branchenapplikationen CMS ECMS CRM Unternehmens-Software Grafik-/Layout-Software Telefonie etc.	Zeitungsa rtikel Multimedia-Website TV-Beitrag Ausstellungskatalog Mobile Applikation Mashup (z.B. Karte) Textbeitrag Konzept Zeitstrahl Fachartikel Studie Präsentation Faktensammlung Exponatsartikel Analysen etc.

**Beobachtungen**

- **Content:** textzentriert, mehrsprachig, multimedial
- **Kuratierung:** zeit- und wissensintensiv, interdisziplinär, evtl. durchgeführt in verteilten Teams
- **Branche:** domänen-/branchenspezifische Anforderungen
- **Einschränkung:** Traditionelle CMS-Systeme bieten kaum Unterstützung für Kuratierungsprozesse!
- **Sprachtechnologie** kann helfen – Kombinierung von Komponenten zu branchenspezifischen Workflows



ART+COM



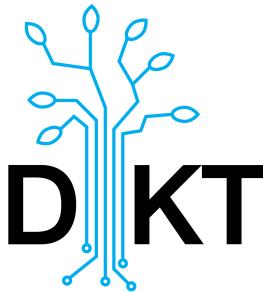
Deutsches  
Forschungszentrum  
für Künstliche  
Intelligenz GmbH

**3pc**

Neue Kommunikation

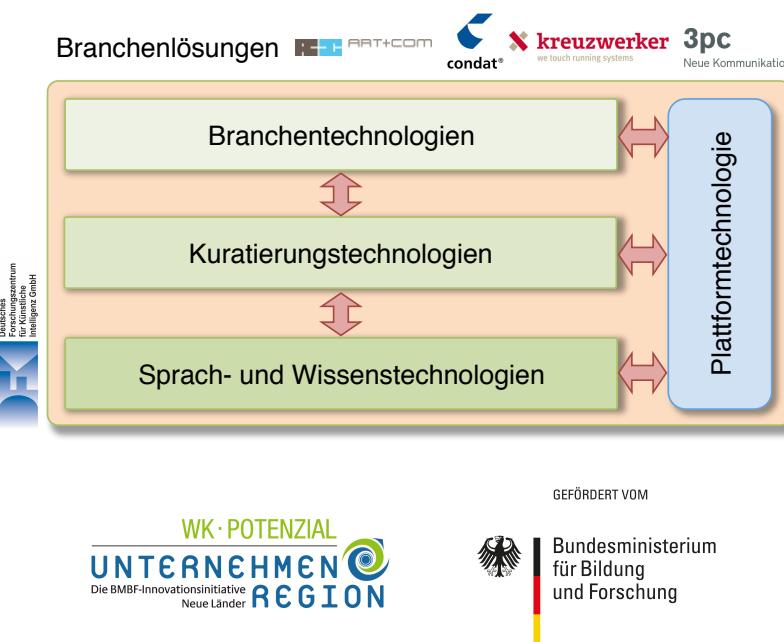


**kreuzwerker**  
we touch running systems



Digitale  
Kuratierungs-  
technologien

- Unterstützung und Optimierung digitaler Kuratierung durch Sprach- und Wissenstechnologien.
- Entwicklung innovativer Prototypen bei den KMU-Partnern.
- Weiterentwicklung der DFKI-Technologien und Transfer mittels **Plattform für digitale Kuratierungstechnologien**.



DKT Kick-off-Veranstaltung – 25. September 2015



ART+COM

**Branche:**

Museen, Showrooms,  
Ausstellungen



**condat®**

**Branche:**

TV, Radio, Web-TV, Medien



Deutsches  
Forschungszentrum  
für Künstliche  
Intelligenz GmbH

**3pc**

Neue Kommunikation

**Branche:**

Öffentliche Archive



**kreuzwerker**  
we touch running systems

**Branche:**

Journalismus

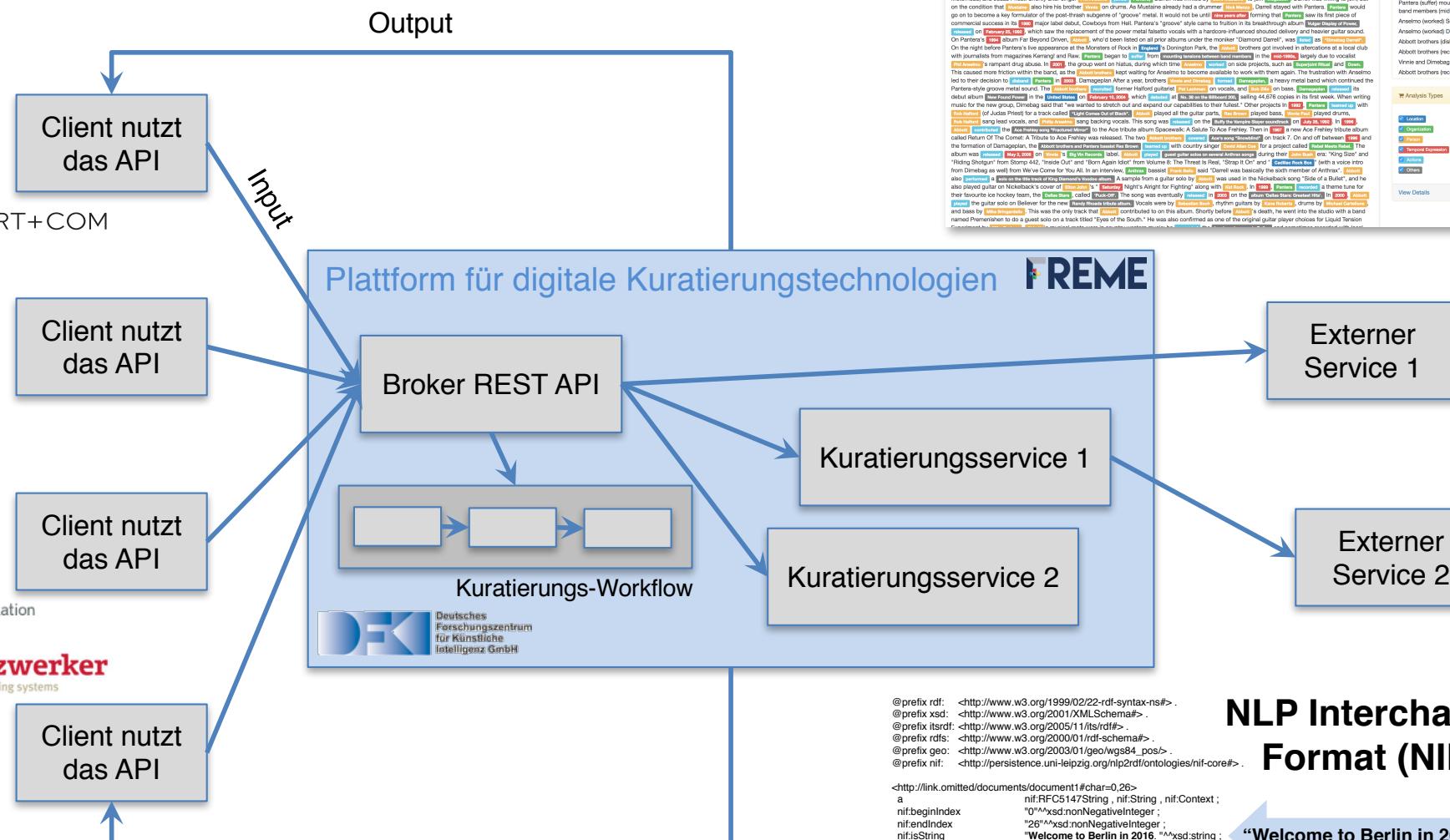


Deutsches  
Forschungszentrum  
für Künstliche  
Intelligenz GmbH

# **Prototypisch implementierte Plattform und Services**



Digitale  
Kuratierungs  
technologien



- Durch (Semi-)Automatisierung der Kuratierungsprozesse zeitliche und finanzielle Aufwände reduzieren
  - Flexible, robuste, skalierbare Services
  - Gemeinsamer Bau von Prototypen (proofs-of-concept)
  - Modulare Inhalte in neuen Produktionen aggregieren
  - Interoperabilität durch generische APIs
  - Human in the loop

 Events
<hr/>
Abbot Formed Parties (1981)
Pal Awadhi Served Parties ()
Parties (and) Vulu Display of Power (February 25, 1992)
Abott (tele) "Dirnebag Damel" (1994)
Parties (buffer) mounting tensions between central and state governments (1994)
Anand (worked) Supersonic Ritual (2001)
Babu (worked) Dook! (2001)
Abbott brothers (diseased) Parties (2003)
Abbott brothers (retired) Pat Leichman (2003)
Vinne and Dirnebag (formed) Maggepan, ()
Abbott brothers (retired) Pat Leichman ()
<hr/>
 Analysis Types
<hr/>
 <a href="#">Catalysis</a>
 <a href="#">Aggregation</a>
 <a href="#">Purity</a>
 <a href="#">Aerobic Respiration</a>
 <a href="#">Adolescence</a>
 <a href="#">Others</a>

# NLP Interchange Format (NIF)

"Welcome to Berlin in 2016 "

NLP Inter Format	
“Welcome To Be	
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .	
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .	
@prefix itsrdf: <http://www.w3.org/2005/11/its/rdf#> .	
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .	
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .	
@prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#> .	
<http://link.omitted/documents/documents#document1#char>_0_26	
a nif:RFC5147String , nif:String , nif:Context ;	
nif:beginIndex "0"^^xsd:nonNegativeInteger ;	
nif:endIndex "26"^^xsd:nonNegativeInteger ;	
nifisString "Welcome to Berlin in 2016."^^xsd:string ;	
dif:kind#averageLatitude "52.51666666666666"^^xsd:double ;	
dif:kind#averageLongitude "13.38333333333333"^^xsd:double ;	
dif:kind#stdDevLatitude "0.0"^^xsd:double ;	
dif:kind#stdDevLongitude "0.0"^^xsd:double ;	
nif:meanDateRange "201601010000_20170101010000"^^xsd:string .	
<http://link.omitted/documents/documents#document1#char>_21_25	
a nif:RFC5147String , nif:String ;	
itsrdf:labelRef <http://link.omitted/ontologies/nif#date>_20160101000000_20170101000000	
nif:anchorOf "2016"^^xsd:string ;	
nif:beginIndex "21"^^xsd:nonNegativeInteger ;	
nif:endIndex "25"^^xsd:nonNegativeInteger ;	
nif:entity <http://link.omitted/ontologies/nif#date> .	
<http://link.omitted/documents/documents#char>_11_17	
a nif:RFC5147String , nif:String ;	
nif:anchorOf "Berlin"^^xsd:string ;	
nif:beginIndex "11"^^xsd:nonNegativeInteger ;	
nif:endIndex "17"^^xsd:nonNegativeInteger ;	
itsrdf:labelClassRef <http://dbpedia.org/ontology/Location> ;	
nif:referenceContext <http://link.omitted/documents#char>_0_26 ;	
geo:lat "52.51666666666666"^^xsd:double ;	
geo:lon "13.38333333333333"^^xsd:double ;	
itsrdf:labelRef <http://dbpedia.org/resource/Berlin> .	

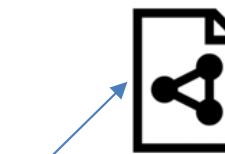
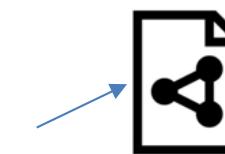
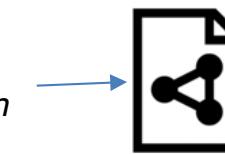
# Aktueller Stand

- **Plattform:** Services und Service-Workflows
- **Implementierte Kuratierungsservices:**
  - **Named Entity Recognition** – e-entityrecognition e-service
  - **Geolocation** – e-entityrecognition, Visualisierung
  - **Temporal Analyser** – e-entityrecognition, Visualisierung
  - **Classification** – e-classification e-service
  - **Clustering** – e-clustering e-service
  - **Textzusammenfassen** – e-summarisation e-service
  - **Maschinelle Übersetzung** – e-translation e-service
  - **Sentiment Analysis** – work in progress
  - **Event Extraction** – work in progress
  - **Semantic Storytelling** – work in progress
- **Kuratierungs-Dashboard:** Erster Prototyp

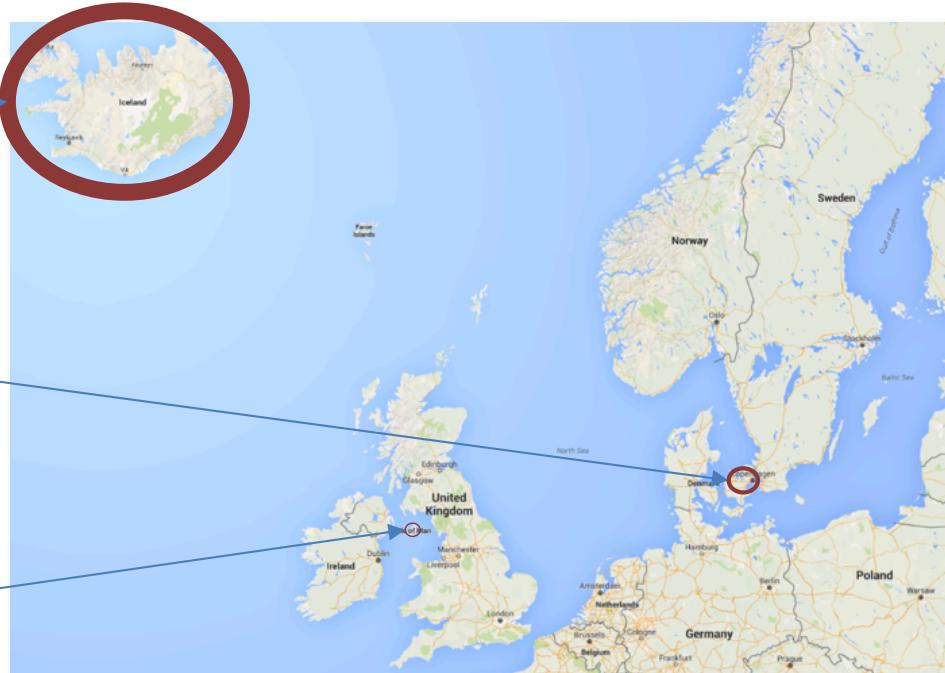
# NER, Linking, Geolokalisierung

- **Modus 1:** Modell-basiert (für Domänen, für die annotierte Trainingsdaten verfügbar sind)
- **Modus 2:** Wörterbuch-basiert (für Domänen, für die lediglich Namenslisten verfügbar sind)
- Basiert auf OpenNLP (mit NIF-Integration)

...  
*In the Viking colony of Iceland, an extraordinary vernacular literature blossomed in the 12th through 14th centuries*



- Entity-Linking durch SPARQL-Querys auf DBpedia.
- Für Lokationen werden GPS-Koordinaten bezogen.
- Es werden Durchschnittsangaben berechnet auf Dokumentebene (über alle Lokationen), um diese auf einer Karte visualisieren zu können.



**Geolokalisierung als visuelles Zusammenfassen!**

## Plain Text

<http://api.digitale-kuratierung.de/api/e-nlp/namedEntityRecognition?analysis=ner>

## NIF-Anreicherung

<http://http://dev.digitale-kuratierung.de/admini/pages/geolocalization.php>

## Visualisierung

# NE Recognition und Linking

A

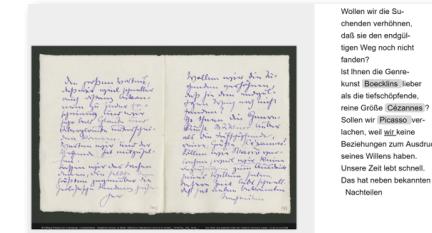
- <http://api.digitale-kuratierung.de/api/e-nlp/trainModel?analysis=dict>
- Falls lediglich Listen von Namen oder Termen und deren URLs in einer Ontologie zur Verfügung stehen.

person	PND	pseudonym
Mendelsohn, Erich	<a href="http://d-nb.info/gnd/11858071X">http://d-nb.info/gnd/11858071X</a>	Erich
Mendelsohn, Luise	<a href="http://d-nb.info/gnd/128830751">http://d-nb.info/gnd/128830751</a>	Luise
Beethoven, Ludwig van	<a href="http://d-nb.info/gnd/118508288">http://d-nb.info/gnd/118508288</a>	Beethoven
Löwe, Ferdinand	<a href="http://d-nb.info/gnd/119273853">http://d-nb.info/gnd/119273853</a>	Ferdinand Loewe
Mottl, Felix	<a href="http://d-nb.info/gnd/188311017">http://d-nb.info/gnd/188311017</a>	Mottl
Lessing, Gotthold Ephraim	<a href="http://d-nb.info/gnd/118572121">http://d-nb.info/gnd/118572121</a>	Lessing
Dehmel, Richard	<a href="http://d-nb.info/gnd/118679236">http://d-nb.info/gnd/118679236</a>	Dehmel
Haydn, Joseph	<a href="http://d-nb.info/gnd/118547356">http://d-nb.info/gnd/118547356</a>	Haydn

Datenbank-Dump der Mendelsohn-Briefe



- **Linking** per Extraktion der DBpedia-URI
- NE-Typspezifische **SPARQL-Querys** für **Personen** (Geburtsdatum), **Lokationen** (Koordinaten), **Organisationen** (Typ)
- Wörterbuch kann URLs enthalten



Wollen wir die Su  
chenden verhören,  
daß sie den endgül  
tigen Weg noch nicht  
fanden?  
Ist ihnen die Gege  
bung Beethovens lieber  
als die telefondende,  
reine Größe Cézanne's?  
Sollen wir Picasso ver  
lachen, weil wir keine  
Beziehungen zum Ausdruck  
seines Willens haben.  
Unsere Zeit lebt schnell.  
Das hat neben bekannten  
Nachteilen

statistisches  
NER-Modell

Mittlere Qualität  
Benötigt weniger annotierte Daten

auf neuem Input nutzbar  
(auch gemeinsam)

Hohe Qualität  
Benötigt annotierte Daten

Mittlere Qualität  
Menschliche Intervention notwendig  
Benötigt keine annotierten Daten

C

- Falls – z.B. bei Spezialdomänen – weder das eine (A) noch das andere (B) vorliegt, können potenzielle Entitäten in Kollektionen berechnet werden.
- Diese Liste kann vom Wissensarbeiter überprüft und anschließend als Wörterbuch (A) eingesetzt werden.

# Zeitausdrücke

...  
The ships were scuttled there  
in the **11th century**, to block a  
navigation channel and thus  
protect Roskilde, then  
Copenhagen from seaborne  
assault  
...

...  
Viking Age inscriptions have  
also been discovered on the  
Manx runestones on the  
Isle of Man.

...  
In the Viking colony of Iceland,  
an extraordinary vernacular  
literature blossomed in the **12th**  
through **14th centuries**

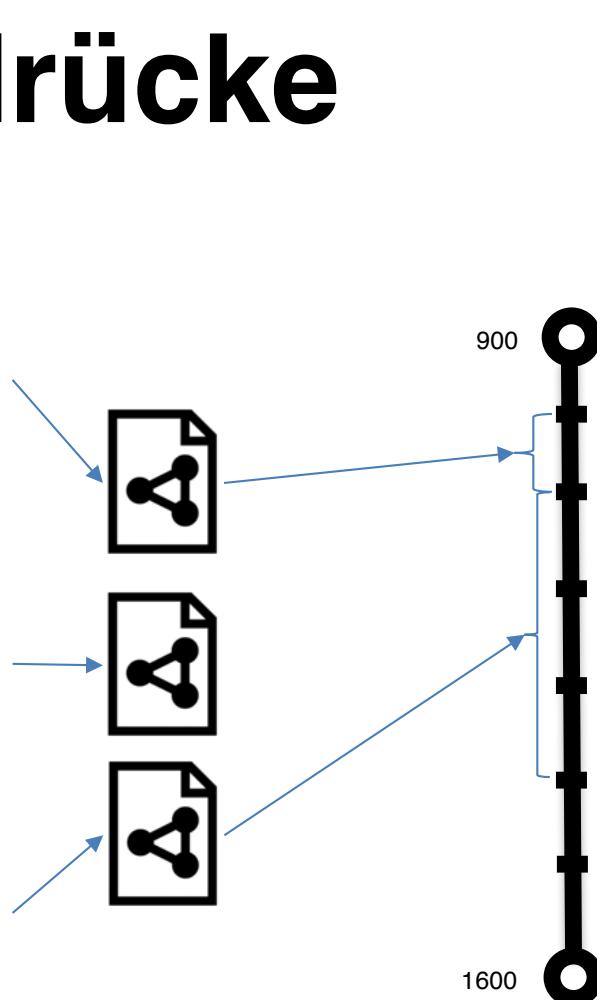
## Plain-Text

<http://api.digitale-kuratierung.de/api/e-nlp/namedEntityRecognition?analysis=temp>

## NIF-Anreicherung

<http://dev.digitale-kuratierung.de/admin/pages/timelining.php>

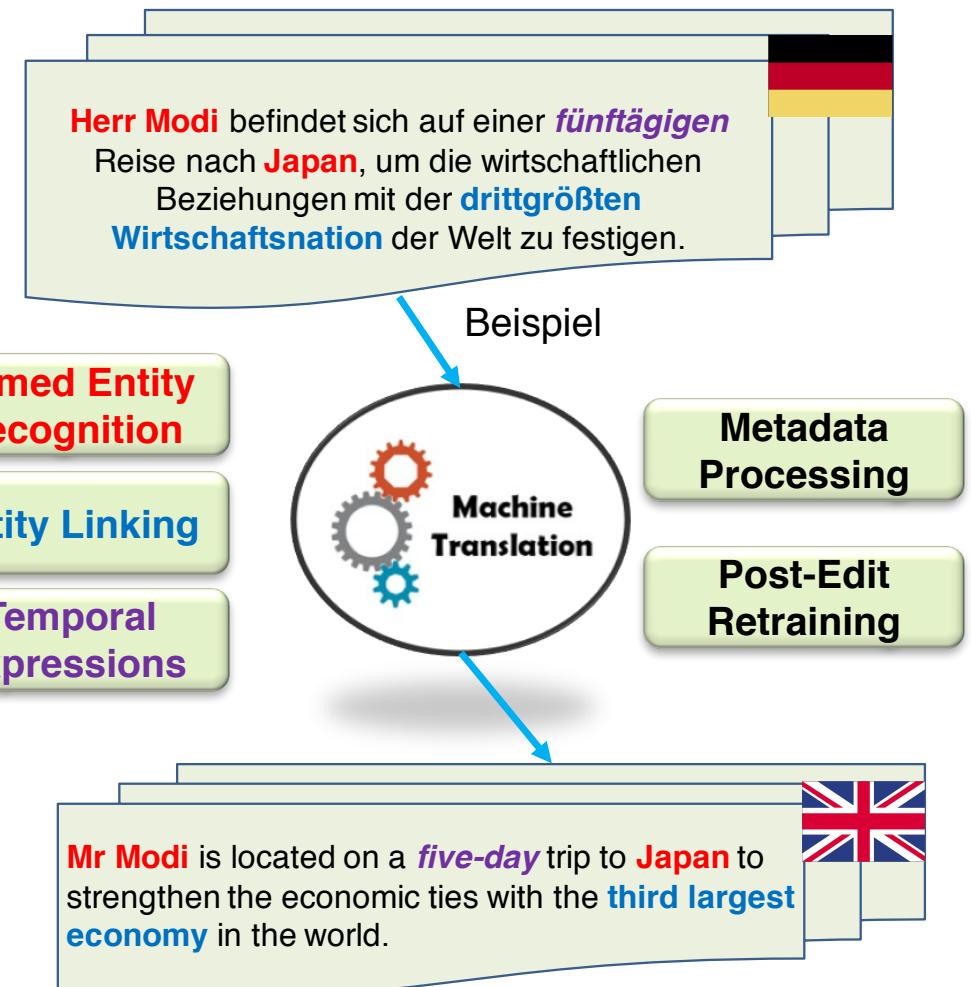
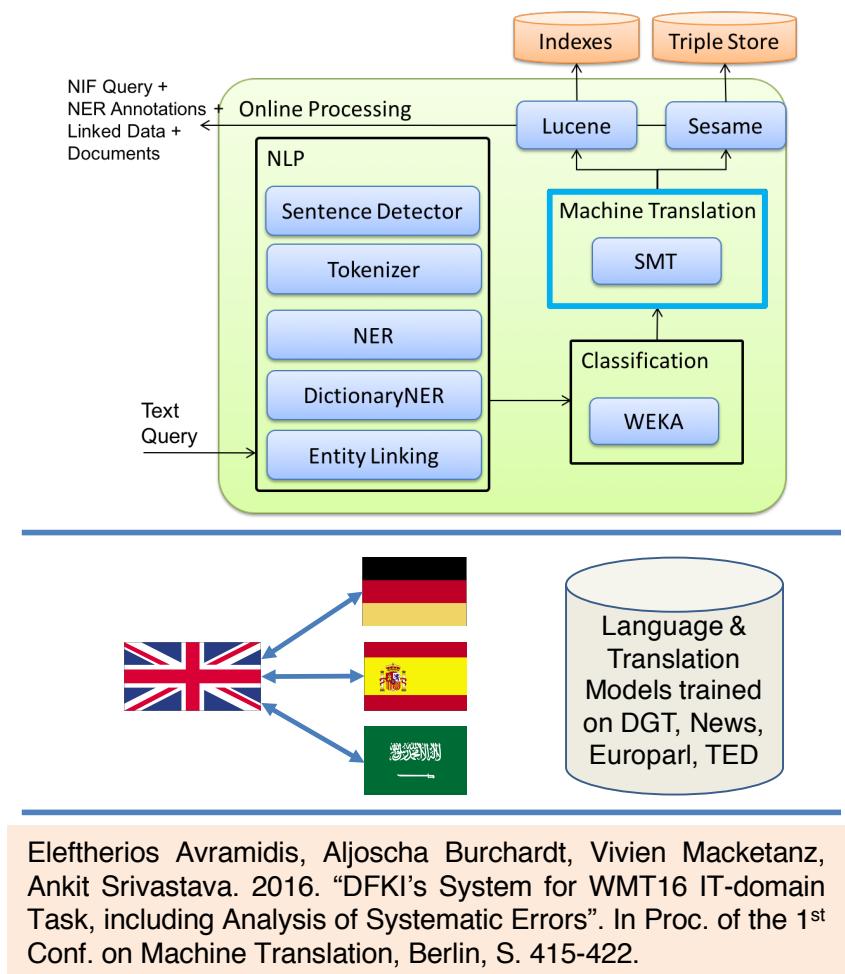
## Visualisierung



- Sortiert Dokumente auf einer chronologischen Skala.
- Regelbasiertes System, um unsere Zielsprachen bestmöglich bedienen zu können (EN, DE).
- Analyse von Zeitausdrücken in einem Dokument.
- Berechnet Durchschnittswerte und Intervalle.
- Plan: Mechanismus für nutzerbasierte Regeln.
- Verwandte Arbeiten: SUTime, HeidelTime, Tango, Tarsgi.

# Maschinelle Übersetzung

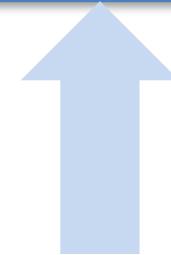
- Robuste, adaptierbare MT-Modelle (nutzen Moses, Cdec, Giza++, SRILM etc.)
- Parallele und monolinguale Korpora: Europarl, DGT-TM, TED, UN, Newscrawl u.a.
- Kombination mit anderen DKT-Services (Summariser, NER, Temporal Analyser); ITS 2.0, NIF
- Diverse Linked-Data-Datenquellen unterstützen MT (z.B. Dbpedia, BabelNet, WordNet)



# Textzusammenfassen

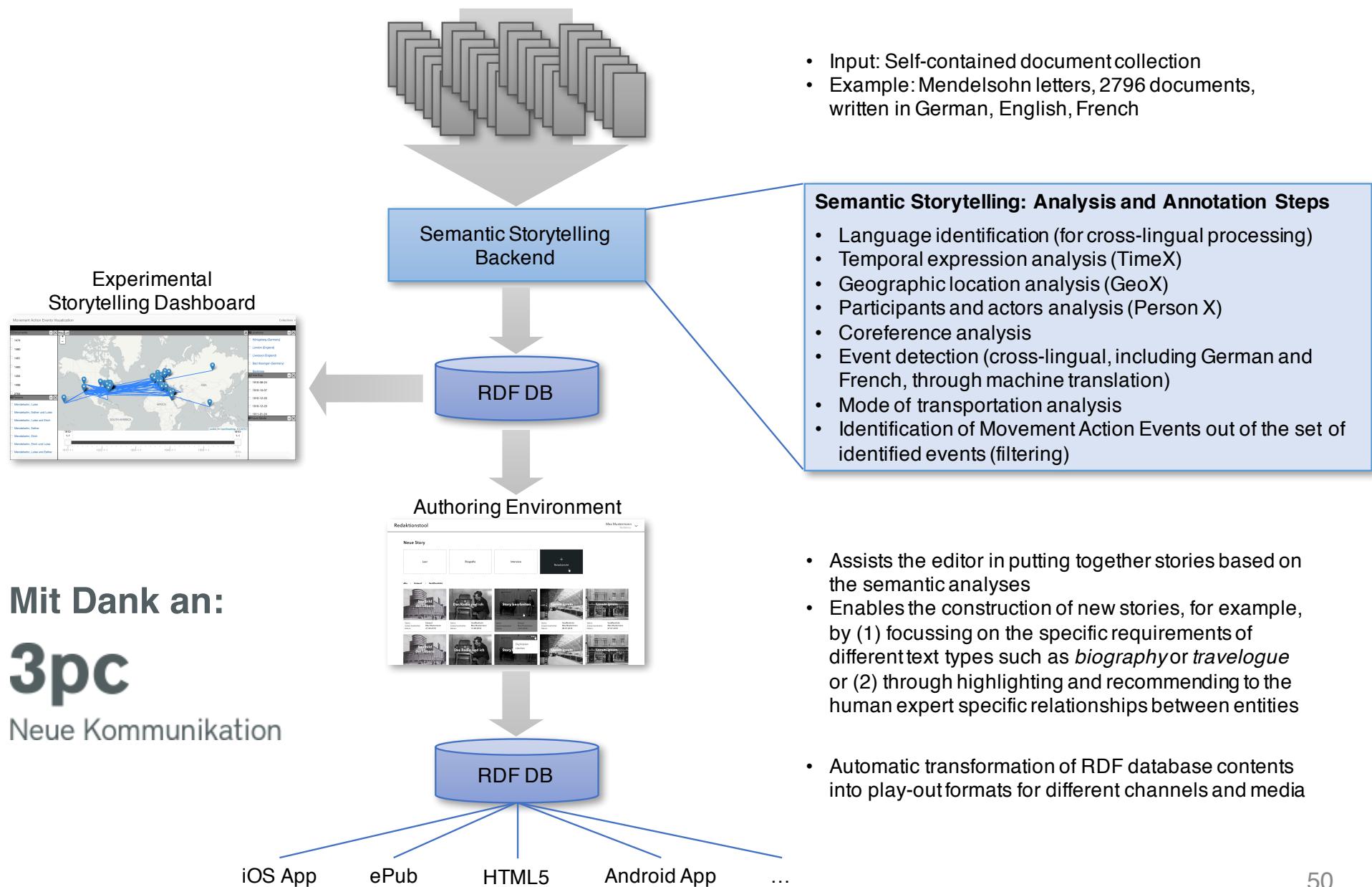
- Kuratierungsservice rankt Sätze – basierend auf div. Features – hinsichtlich ihrer Wichtigkeit.
- Beispiel: Artikel über den fallenden Aktienkurs von RWE (Daten stammen von Condat).
- Integration der Analyseergebnisse anderer DKT-Services in den Algorithmus.

Im letzten Monat und den letzten 3 Monaten verlor die RWE-Aktie 3,79% bzw. 18,95% und in den letzten 3 Tagen 3,55%.



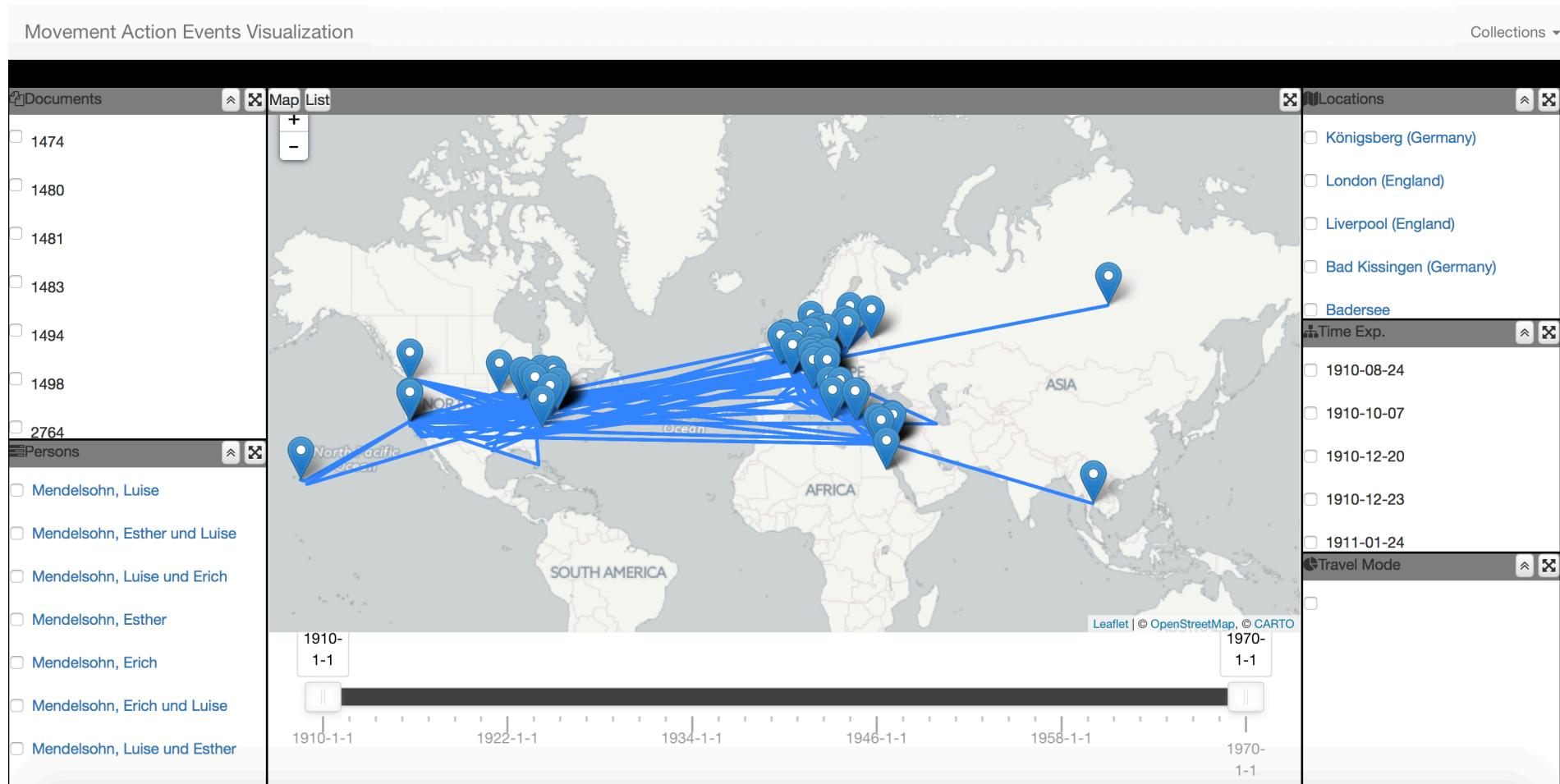
Die Aktie der RWE AG fiel am Donnerstag um 0,21% auf 19,16 EUR und schwankte am Handelstag zwischen 19,08 und 19,32 EUR. Das Handelsvolumen der Aktie lag bei 1,79 Millionen Aktien und so unter dem 52-Wochen und 150-Tagesvolumen von 3,40 Millionen bzw. 3,96 Millionen Aktien. Im letzten Monat und den letzten 3 Monaten verlor die RWE-Aktie 3,79% bzw. 18,95% und in den letzten 3 Tagen 3,55%. Das PE und PB-Verhältnis der Unternehmensaktie liegt aktuell bei 11,44 bzw. 1,29, während die historischen PE und PB-Werte jeweils bei 11,77 bzw. 2,13 liegen.

# Beispiel: Die Mendelsohn-Briefe

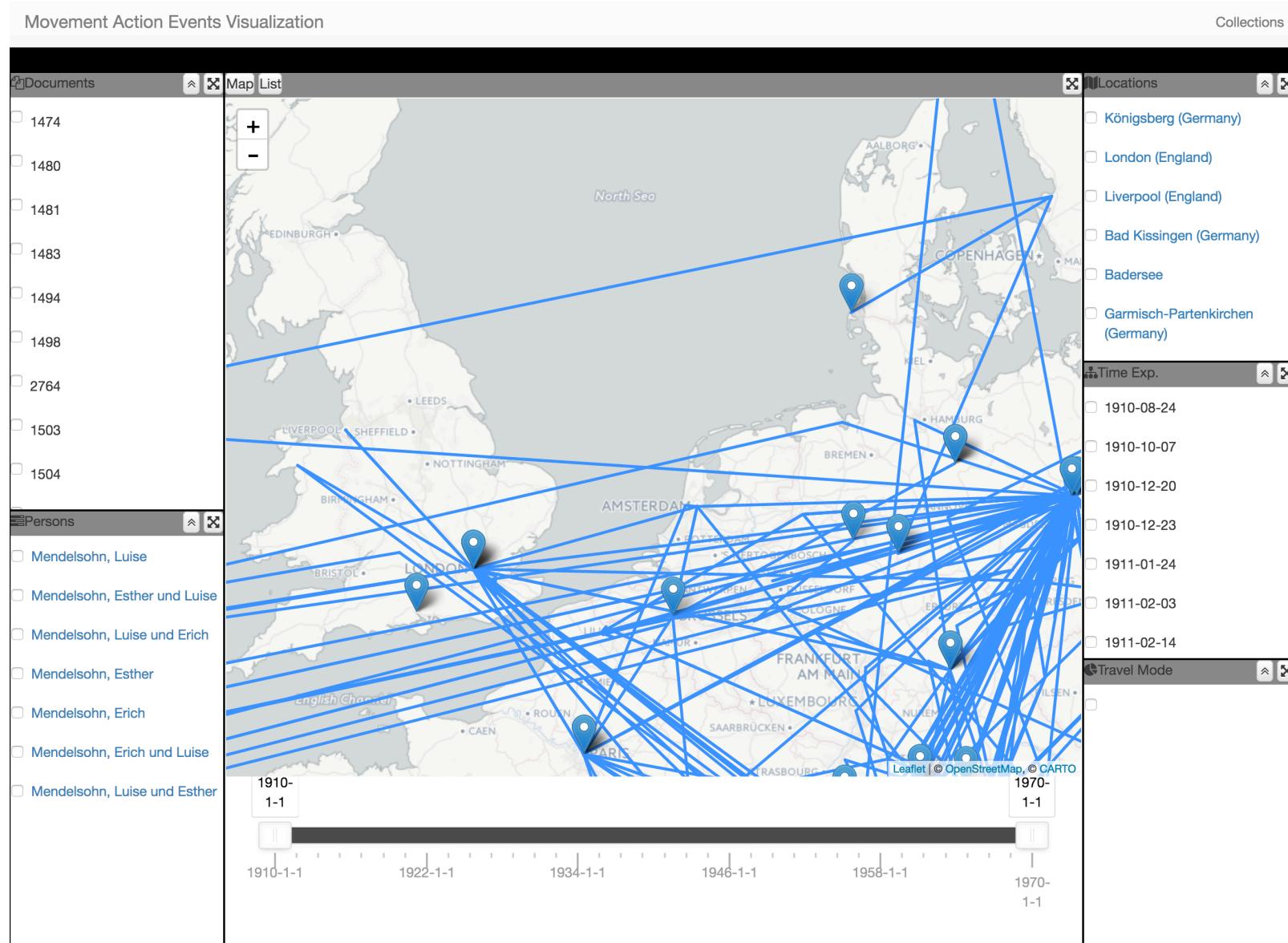


Mit Dank an:  
**3pc**  
Neue Kommunikation

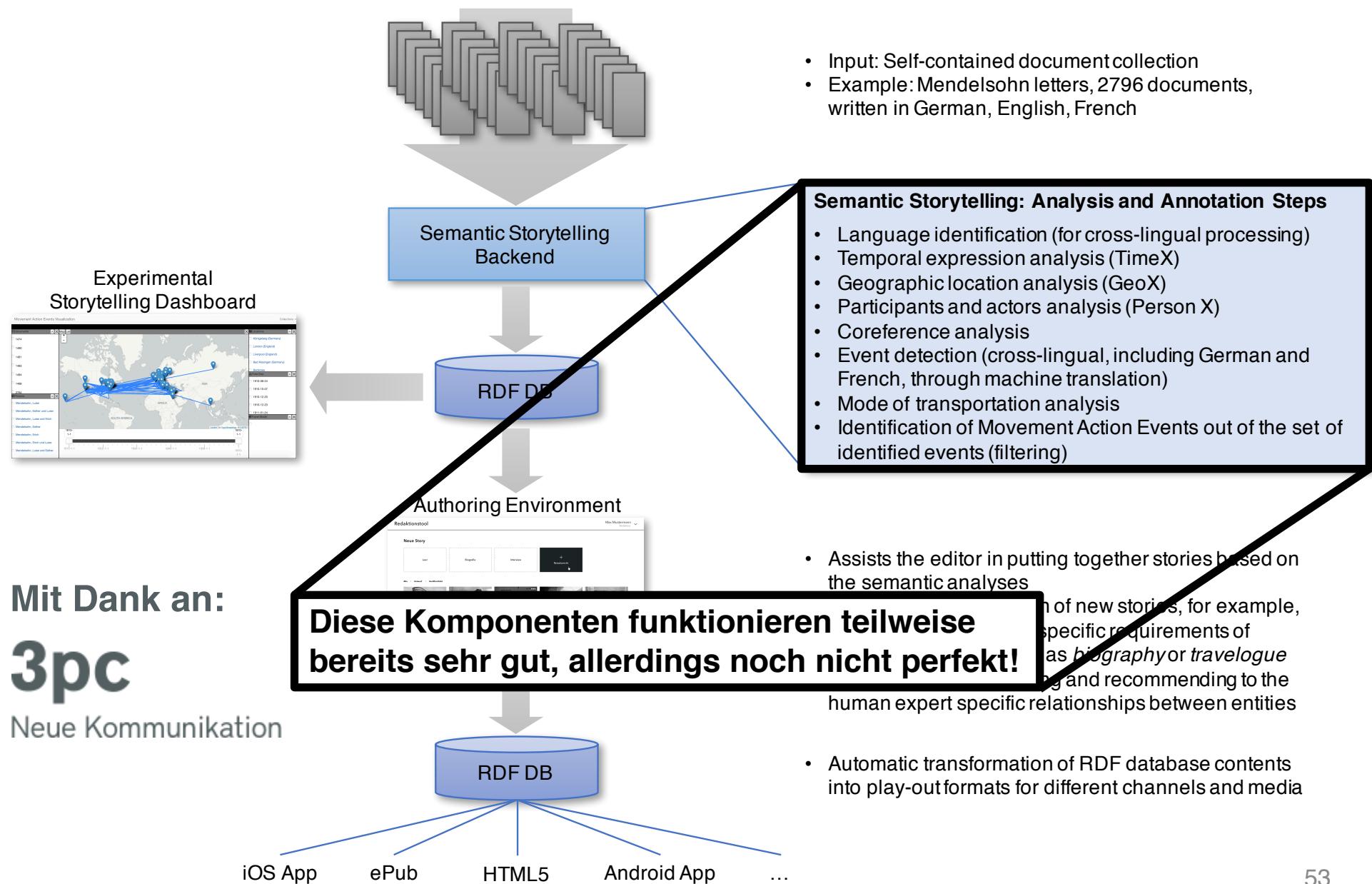
# Beispiel: Die Mendelsohn-Briefe



# Beispiel: Die Mendelsohn-Briefe



# Beispiel: Die Mendelsohn-Briefe



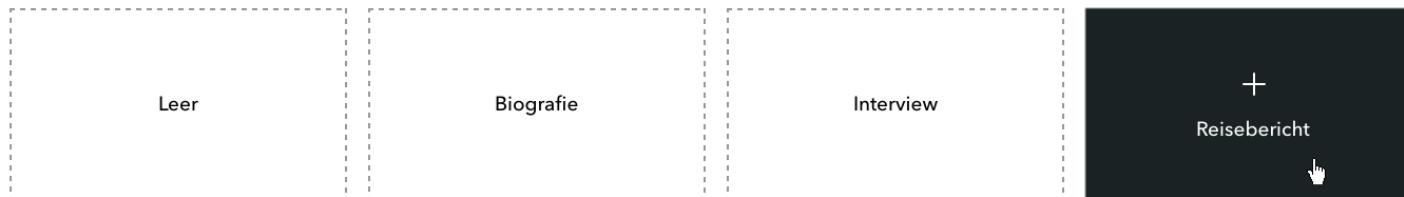
# Beispiel: Die Mendelsohn-Briefe

Redaktionstool

Max Mustermann  
Redakteur



## Neue Story



Alle / Entwurf / Veröffentlicht



Status: Entwurf  
Zuletzt bearbeitet: Max Mustermann  
Datum: 27.08.2018



Status: Veröffentlicht  
Zuletzt bearbeitet: Max Mustermann  
Datum: 12.08.2018



Status: Entwurf  
Zuletzt bearbeitet: Max Mustermann  
Datum: 14.07.2018



Status: Veröffentlicht  
Zuletzt bearbeitet: Max Mustermann  
Datum: 08.07.2018



Status: Veröffentlicht  
Zuletzt bearbeitet: Max Mustermann  
Datum: 07.07.2018



# Beispiel: Die Mendelsohn-Briefe

Redaktionstool

Max Mustermann  
Redakteur

## Neue Story

Leer

Alle / Entwurf / Veröffentlicht

Status: Entwurf  
Zuletzt bearbeitet: Max Mustermann  
Datum: 27.08.2018

Story anlegen

**Neue Story anlegen**

**Titel**  
Das Leben des Erich Mendelsohn

**Beschreibung**  
Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery

**Tags**  
Erich Mendelsohn | X

Story anlegen

# Beispiel: Die Mendelsohn-Briefe

< Das Leben des Erich Mendelsohn  
Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery

Letzte Änderung vor 2 Minuten

Vorschläge Stories Merkliste

Bilder Videos Audio Dokumente

Brief No. 1 an Luise Medelsohn Dokument

Portrait Erich Medelsohn Bild

Portrait Erich und Luise Medelsohn Bild

Brief No. 24 an Luise Medelsohn Dokument

Beginnen Sie eine neue Geschichte.

DFK Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Filter

The screenshot shows a dark-themed digital archive interface. At the top left, there's a navigation bar with a back arrow, the title 'Das Leben des Erich Mendelsohn', a subtitle 'Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery', and a timestamp 'Letzte Änderung vor 2 Minuten'. On the right side of the header are three buttons: 'Vorschläge' (with a star icon), 'Stories' (with a grid icon), and 'Merkliste' (with a bookmark icon). Below the header, there are four categories: 'Bilder', 'Videos', 'Audio', and 'Dokumente'. Under 'Dokumente', there are two items: 'Brief No. 1 an Luise Medelsohn' (labeled 'Dokument') and 'Portrait Erich Medelsohn' (labeled 'Bild'). Under 'Bilder', there are two items: 'Portrait Erich und Luise Medelsohn' (labeled 'Bild') and 'Brief No. 24 an Luise Medelsohn' (labeled 'Dokument'). A large, semi-transparent text box in the center says 'Beginnen Sie eine neue Geschichte.' with a dashed outline and a plus sign in the center. On the far left, there's a vertical toolbar with icons for 'A' (text), a minus sign, a camera, a minus sign, and a play button. At the bottom right, there are '+' and '-' buttons and a 'Filter' bar. The DFK logo and name are at the bottom left.

# Beispiel: Die Mendelsohn-Briefe

< Das Leben des Erich Mendelsohn  
Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery

Letzte Änderung vor 2 Minuten

Vorschläge Stories Merkliste

Bilder Videos Audio Dokumente

Portrait Erich Medelsohn

Brief No. 1 an Luise Medelsohn Dokument

Portrait Erich und Luise Medelsohn Bild

Brief No. 24 an Luise Medelsohn Dokument

Drag & Drop!

Filter

The screenshot shows a dark-themed digital archive interface. At the top left is a navigation bar with the title 'Das Leben des Erich Mendelsohn' and a subtitle 'Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery'. A timestamp 'Letzte Änderung vor 2 Minuten' is also present. On the right side, there are buttons for 'Vorschläge', 'Stories', and 'Merkliste'. Below these are tabs for 'Bilder', 'Videos', 'Audio', and 'Dokumente'. The main content area displays several items: a portrait of a man (Erich Medelsohn) writing at a desk, a letter document titled 'Brief No. 1 an Luise Medelsohn', another portrait of the couple (Erich and Luise Medelsohn), and another letter document titled 'Brief No. 24 an Luise Medelsohn'. A large, semi-transparent rectangular overlay in the center contains a smaller image of the couple and a dashed blue rectangle with a plus sign, indicating a 'Drag & Drop' feature. On the far left, a vertical sidebar has a letter 'A' at the top and icons for image, video, and audio below it. At the bottom right, there are zoom controls (+/-) and a 'Filter' button.

# Beispiel: Die Mendelsohn-Briefe

# Beispiel: Die Mendelsohn-Briefe

< Das Leben des Erich Mendelsohn  
Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery

Letzte Änderung vor 2 Minuten

Vorschläge Stories Merkliste

Bilder Videos Audio Dokumente

Bild

Dokument

Bild

Dokument

Filter

A

—

—

—

+

+

-

The screenshot shows a dark-themed digital archive interface. At the top left, there's a back arrow, the title 'Das Leben des Erich Mendelsohn' with a subtitle 'Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery', and a search bar with a magnifying glass icon labeled 'Archiv durchsuchen'. On the right side, there are navigation buttons for 'Vorschläge', 'Stories', and 'Merkliste'. Below these are tabs for 'Bilder', 'Videos', 'Audio', and 'Dokumente'. The main content area displays two columns of items. The left column contains a thumbnail of a couple sitting, a dashed box for cropping, and a document thumbnail. The right column contains a portrait of a man writing, a document thumbnail, a portrait of the couple, and another document thumbnail. Each item has a 'Bild' or 'Dokument' label below it. A vertical sidebar on the left features icons for a magnifying glass, a camera, and a play button, with a plus sign at the bottom. A horizontal zoom slider with '+' and '-' buttons is at the bottom right.

# Beispiel: Die Mendelsohn-Briefe

Zum Löschen hier ablegen

Vorschläge Stories Merkliste

Archiv durchsuchen

A

Bild Dokument

Filter

A screenshot of a digital archive interface. At the top right are navigation icons for 'Vorschläge' (star), 'Stories' (grid), and 'Merkliste' (bookmark). Below is a search bar with the placeholder 'Archiv durchsuchen'. On the left, there's a vertical toolbar with icons for 'A' (archive), 'Bild' (image), and 'Dokument' (document). The main area displays a grid of items. In the center, a thumbnail of a couple is shown within a dashed rectangular frame, indicating it can be enlarged. To the right, there are two columns of items. The first column contains a handwritten letter titled 'Brief No. 1 an Luise Medelsohn' and a portrait of Erich Medelsohn, both labeled 'Dokument' and 'Bild'. The second column contains another handwritten letter titled 'Brief No. 24 an Luise Medelsohn' and a portrait of a man, also labeled 'Dokument' and 'Bild'. At the bottom right is a 'Filter' button.

Portrait Erich Medelsohn

Brief No. 1 an  
Luise Medelsohn  
Dokument

Bild

Brief No. 24 an  
Luise Medelsohn  
Dokument

# Beispiel: Die Mendelsohn-Briefe

Zum Löschen hier ablegen

Archiv durchsuchen

A

—

—

—

+

-

Filter

The screenshot shows a digital archive interface with a dark theme. At the top left is a red bar with a trash icon and the text "Zum Löschen hier ablegen". Below it is a search bar with a magnifying glass icon and the placeholder "Archiv durchsuchen". On the far left is a vertical sidebar with a letter "A" at the top, followed by three horizontal line icons, then a photo icon, and finally a circular arrow icon. In the center, there is a grid of five small thumbnail images of historical figures, likely members of the Mendelsohn family. To the right of these thumbnails is a larger image of a man with glasses, identified as Erich Mendelsohn. Further right, there are two large boxes containing historical documents. The top box is labeled "Brief No. 1 an Luise Medelsohn Dokument" and features a portrait of a woman. The bottom box is labeled "Brief No. 24 an Luise Medelsohn Dokument" and also features a portrait of a woman. At the very bottom right of the interface is a "Filter" button.

Vorschläge Stories Merkliste

Bilder Videos Audio Dokumente

Brief No. 1 an Luise Medelsohn  
Dokument

Portrait Erich Medelsohn  
Bild

Brief No. 24 an Luise Medelsohn  
Dokument

# Beispiel: Die Mendelssohn-Briefe

# Beispiel: Die Mendelsohn-Briefe

< Das Leben des Erich Mendelsohn ⚙ Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery

Letzte Änderung vor 2 Minuten ▶

Text X

Archiv durchsuchen

Schriftgröße: H1 H2 H3 H4 p

Schriftfarbe:

Ausrichtung:

Textfeld: Text eingeben

+

-

The screenshot shows a dark-themed digital archive interface. At the top left is the title "Das Leben des Erich Mendelsohn" with a pencil icon, followed by the subtitle "Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery". A timestamp "Letzte Änderung vor 2 Minuten" is at the top right. On the far left is a vertical toolbar with icons for text (A), line (—), image (img), and video (play). The main area displays three items: a black and white photograph of a man and a woman, a handwritten letter on aged paper with the text "Baldwaring den 31. 1. 19", and a white text input field with the placeholder "Text eingeben". To the right of these items is a sidebar titled "Text" containing styling options: "Schriftgröße" with buttons for H1 through p, "Schriftfarbe" with color swatches, "Ausrichtung" with alignment icons, and a large "Textfeld" input box. At the bottom right are zoom controls (+, -).

# Beispiel: Die Mendelsohn-Briefe

## Das Leben des Erich Mendelsohn

Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery

Letzte Änderung  
vor 2 Minuten



Vorschläge

Stories

Merkliste

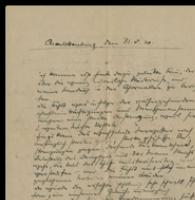
Bilder Videos Audio Dokumente



Portraits von  
Erich Mendelsohn

Bild

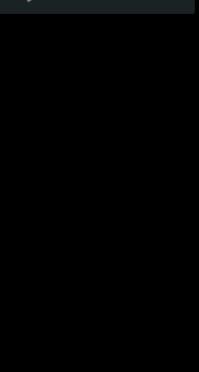
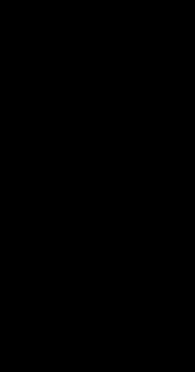
Dokument



Brief No. 24 an  
Luise Mendelsohn

Sinnbild des Lebens

Dokument



Story

Ergebnisse für **Erich Mendelsohn** | X Reise | X Lor|em ipsum

► Lorem ipsum  
Lord lorem  
Lorocius Pangasius  
Lorio

**Brief an Luise Mendelsohn über Amerikareise**

Ich komme erst heute dazu, geliebte Frau, Dir über die wenig erfreuliche Rückreise und meine Landung in der Ahornallee zu berichten.

Die Fahrt war infolge der vorhergehenden Schlafbeeinträchtigungen und Menschenansammlungen eine starke Anstrengung, weil sie 2 weitere Nächte

Medium: Brief  
Verfasser: Erich Mendelsohn  
Quelle: EMA - Erich Mendelsohn Archiv  
Datum: 31.05.1920  
Entities: Charlottenburg, Hans Heyden...

**Erich Mendelsohns Rede am 27.06.1920**

Ich komme erst heute dazu, geliebte Frau, Dir über die wenig erfreuliche Rückreise und meine Landung in der Ahornallee zu berichten.

Die Fahrt war infolge der vorhergehenden Schlafbeeinträchtigungen und Menschenansammlungen eine starke Anstrengung, weil sie 2 weitere Nächte

Medium: Audio  
Verfasser: Erich Mendelsohn  
Quelle: EMA - Erich Mendelsohn Archiv  
Datum: 27.06.1920  
Entities: Charlottenburg, Hans Heyden...

**Sinnbild des Lebens**

Ich komme erst heute dazu, geliebte Frau, Dir über die wenig erfreuliche Rückreise und meine Landung in der Ahornallee zu berichten.

Die Fahrt war infolge der vorhergehenden Schlafbeeinträchtigungen und Menschenansammlungen eine starke Anstrengung, weil sie 2 weitere Nächte

Medium: Story  
Verfasser: Erich Mendelsohn  
Quelle: EMA - Erich Mendelsohn Archiv  
Datum: 31.05.1920  
Entities: Charlottenburg, Hans Heyden...

**Bild**

Quelle: EMA - Erich Mendelsohn Archiv  
Ort: Amerika  
Datum: 31.05.1920  
Entities: Charlottenburg, Hans Heyden...

**Brief an Luise Mendelsohn über Amerikareise**

Ich komme erst heute dazu, geliebte Frau, Dir über die wenig erfreuliche Rückreise und meine Landung in der Ahornallee zu berichten.

Die Fahrt war infolge der vorhergehenden Schlafbeeinträchtigungen und Menschenansammlungen eine starke Anstrengung, weil sie 2 weitere Nächte kostete.

Medium: Brief  
Verfasser: Erich Mendelsohn  
Quelle: EMA - Erich Mendelsohn Archiv  
Datum: 31.05.1920  
Entities: Charlottenburg, Hans Heyden...

**Erich Mendelsohns Rede am 27.06.1920**

Ich komme erst heute dazu, geliebte Frau, Dir über die wenig erfreuliche Rückreise und meine Landung in der Ahornallee zu berichten.

Die Fahrt war infolge der vorhergehenden Schlafbeeinträchtigungen und Menschenansammlungen eine starke Anstrengung, weil sie 2 weitere Nächte

Medium: Audio  
Verfasser: Erich Mendelsohn  
Quelle: EMA - Erich Mendelsohn Archiv  
Datum: 27.06.1920  
Entities: Charlottenburg, Hans Heyden...

**Sinnbild des Lebens**

Ich komme erst heute dazu, geliebte Frau, Dir über die wenig erfreuliche Rückreise und meine Landung in der Ahornallee zu berichten.

Die Fahrt war infolge der vorhergehenden Schlafbeeinträchtigungen und Menschenansammlungen eine starke Anstrengung, weil sie 2 weitere Nächte

Medium: Story  
Verfasser: Erich Mendelsohn  
Quelle: EMA - Erich Mendelsohn Archiv  
Datum: 31.05.1920  
Entities: Charlottenburg, Hans Heyden...

**Bild**

Quelle: EMA - Erich Mendelsohn Archiv  
Ort: Amerika  
Datum: 31.05.1920  
Entities: Charlottenburg, Hans Heyden...

# Beispiel: Die Mendelsohn-Briefe

## Das Leben des Erich Mendelsohn

Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery

Letzte Änderung  
vor 2 Minuten



Vorschläge

Stories

Merkliste

Videos Audio Dokumente Stories

Erich Mendelsohn | X Reise | X



Bilder Videos Audio Dokumente

325 Ergebnisse



Ich komme erst heute dazu, geliebte Frau,  
Dir über die wenig erfreuliche Rückreise  
und meine Landung in der Ahornallee zu  
berichten.

Die Fahrt war infolge der vorhergehenden  
Schlafbeeinträchtigungen und  
Menschenansammlungen eine starke  
Anstrengung, weil sie 2 weitere Nächte

Medium: Brief  
Verfasser: Erich Mendelsohn  
Quelle: EMA - Erich Mendelsohn Archiv  
Datum: 31.05.1920  
Entities: Charlottenburg, Hans Heyden...

Ich komme erst heute dazu, geliebte Frau,  
Dir über die wenig erfreuliche Rückreise  
und meine Landung in der Ahornallee zu  
berichten.

Die Fahrt war infolge der vorhergehenden  
Schlafbeeinträchtigungen und  
Menschenansammlungen eine starke  
Anstrengung, weil sie 2 weitere Nächte

Medium: Audio  
Verfasser: Erich Mendelsohn  
Quelle: EMA - Erich Mendelsohn Archiv  
Datum: 27.06.1920  
Entities: Charlottenburg, Hans Heyden...

Medium: Story  
Verfasser: Erich Mendelsohn  
Quelle: EMA - Erich Mendelsohn Archiv  
Datum: 31.05.1920  
Entities: Charlottenburg, Hans Heyden...

Medium: Bild  
Quelle: EMA - Erich Mendelsohn Archiv  
Ort: Amerika  
Datum: 31.05.1920  
Entities: Charlottenburg, Hans Heyden...



Ich komme erst heute dazu, geliebte Frau,  
Dir über die wenig erfreuliche Rückreise  
und meine Landung in der Ahornallee zu  
berichten.

Die Fahrt war infolge der vorhergehenden  
Schlafbeeinträchtigungen und  
Menschenansammlungen eine starke  
Anstrengung, weil sie 2 weitere Nächte  
kostete.

Medium: Brief  
Verfasser: Erich Mendelsohn  
Quelle: EMA - Erich Mendelsohn Archiv  
Datum: 31.05.1920  
Entities: Charlottenburg, Hans Heyden...

Ich komme erst heute dazu, geliebte Frau,  
Dir über die wenig erfreuliche Rückreise  
und meine Landung in der Ahornallee zu  
berichten.

Die Fahrt war infolge der vorhergehenden  
Schlafbeeinträchtigungen und  
Menschenansammlungen eine starke  
Anstrengung, weil sie 2 weitere Nächte

Medium: Audio  
Verfasser: Erich Mendelsohn  
Quelle: EMA - Erich Mendelsohn Archiv  
Datum: 27.06.1920  
Entities: Charlottenburg, Hans Heyden...

Medium: Story  
Verfasser: Erich Mendelsohn  
Quelle: EMA - Erich Mendelsohn Archiv  
Datum: 31.05.1920  
Entities: Charlottenburg, Hans Heyden...

Medium: Bild  
Quelle: EMA - Erich Mendelsohn Archiv  
Ort: Amerika  
Datum: 31.05.1920  
Entities: Charlottenburg, Hans Heyden...

Zur Merkliste hinzufügen

Brief No. 24 an  
Luise Medelsohn

Dokument

Story

Filter

# Beispiel: Die Mendelsohn-Briefe

< Das Leben des Erich Mendelsohn  
Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery

Letzte Änderung vor 2 Minuten >

🔍 Erich Mendelsohn | X Reise | X Luise Mendelsohn | X

↑ ↗

+ -

## Luise Mendelsohn

Bilder Videos Audio Dokumente

Portraits von Erich Mendelsohn  
Bilder

Brief No. 1 an Luise Medelsohn  
Dokument

Brief No. 24 an Luise Medelsohn  
Dokument

Sinnbild des Lebens  
Story

# Beispiel: Die Mendelsohn-Briefe



# Beispiel: Die Mendelsohn-Briefe

< Das Leben des Erich Mendelsohn  
Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery

Letzte Änderung vor 2 Minuten >

🔍 Erich Mendelsohn | X Reise | X 1911 - 1951 | X

World map showing travel routes. A green location pin is placed on San Francisco, USA. Numerous small green pins are scattered across Europe and the Middle East.

Vorschläge Stories Merkliste

Bilder Videos Audio Dokumente

  
Portraits von  
Erich Mendelsohn  
Bilder

  
Brief über Amerikareise  
an Luise Medelsohn  
Dokument

  
Sinnbild des Lebens  
Story

  
Brief No. 24 an  
Luise Medelsohn  
Dokument

# Beispiel: Die Mendelsohn-Briefe

< Das Leben des Erich Mendelsohn  
Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery

Letzte Änderung vor 2 Minuten

Vorschläge Stories Merkliste

Archiv durchsuchen

Bilder Videos Audio Dokumente

A

—

—

—

+

-

Filter

The screenshot displays a dark-themed digital archive interface. At the top left, there's a navigation bar with the title 'Das Leben des Erich Mendelsohn' and a subtitle 'Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery'. On the right side of the top bar are icons for 'Vorschläge', 'Stories', and 'Merkliste'. Below the top bar is a search bar labeled 'Archiv durchsuchen' with a magnifying glass icon. To the left of the main content area is a vertical sidebar with icons for 'A' (for alphabet), three horizontal lines (for filters), and a play button (for media). The main content area is divided into several sections. At the top right, there are tabs for 'Bilder', 'Videos', 'Audio', and 'Dokumente'. Below these tabs are two columns of items. The first column contains a thumbnail of a handwritten letter titled 'Brief No. 1 an Luise Medelsohn Dokument' and a thumbnail of a portrait of Erich Mendelsohn titled 'Portrait Erich Medelsohn Bild'. The second column contains a thumbnail of a photograph of Erich and Luise Mendelsohn titled 'Portrait Erich und Luise Medelsohn Bild' and a thumbnail of another handwritten letter titled 'Brief No. 24 an Luise Medelsohn Dokument'. At the bottom right of the main content area are plus and minus signs, likely for zooming, and a 'Filter' button.

Das Leben des Erich Mendelsohn

Archiv durchsuchen

Bilder Videos Audio Dokumente

—

—

—

+

-

Filter

Brief No. 1 an Luise Medelsohn  
Dokument

Portrait Erich Medelsohn  
Bild

Portrait Erich und Luise Medelsohn  
Bild

Brief No. 24 an Luise Medelsohn  
Dokument

The image displays a series of screenshots from a digital archive system, likely for historical research or journalism. The interface is designed for managing and presenting a collection of documents, images, and other media related to the life of architect Erich Mendelsohn.

- Top Left:** A dashboard titled "Redaktionstool" showing a grid of items labeled "Leer", "Biografie", "Interview", and "Reisebericht". Below the grid are several thumbnail images of historical documents and photographs.
- Top Middle:** A "Neue Story" creation dialog box. It asks for a "Titel" (Title) "Das Leben des Erich Mendelsohn", a "Beschreibung" (Description) "Eine Biografie geschrieben von Jeremy Larson und Kathleen Montgomery", and a "Tags" (Tags) "Erich Mendelsohn". A "Story anlegen" (Create Story) button is at the bottom.
- Top Right:** A dark-themed interface for viewing a biography. It shows a large thumbnail of a portrait of Mendelsohn, a sidebar with navigation links like "Bilder", "Videos", "Stories", and "Merkliste", and a main content area with text and smaller images.
- Middle Left:** A "Branche: Öffentliche Archive" section featuring a large "3pc Neue Kommunikation" logo. Below it is a grid of historical images and documents.
- Middle Center:** A detailed view of a document titled "Brief am Luise Mendelsohn, über Amerikareise". It includes text, images, and a small map.
- Middle Right:** A semantic network diagram centered on "Erich Mendelsohn". Nodes include "Luise Mendelsohn", "Peter Bourgonje", "Ankit Srivastava", "Jan Nehring", "Armin Berger", "Luca König", "Sören Räuchle", and "Jens Gerth". Relationships are shown by lines between nodes.
- Bottom Left:** A timeline visualization showing the frequency of documents over time from 1887 to 1993, with a prominent peak around 1920.
- Bottom Center:** A world map where document locations are plotted as points.
- Bottom Right:** Another view of the biography interface, showing a grid of images and documents.

Georg Rehm, Julian Moreno Schneider, Peter Bourgonje, Ankit Srivastava, Jan Nehring, Armin Berger, Luca König, Sören Räuchle, and Jens Gerth. Event Detection and Semantic Storytelling: Generating a Travelogue from a large Collection of Personal Letters. In Tommaso Caselli, Ben Miller, Marieke van Erp, Piek Vossen, Martha Palmer, Eduard Hovy, and Teruko Mitamura, editors, *Proceedings of the Events and Stories in the News Workshop*, Vancouver, Canada, August 2017. Association for Computational Linguistics. Co-located with ACL 2017.

**kreuzwerker**  
we touch running systems

## Branche: Journalismus

Search: Advanced Saved

PARAMETERS FILTERS

ENTITY TYPES: person (45786), organization, location, event.

SENTIMENT: negative.

TOPICS: Politics (889), Economy (412), Health (255), Science and Tech... (117), Sociology (89).

KEYWORDS: None.

GENRES: None.

SEARCH

Total: 1641 Sort: Updated

News items:

- 24 The Bag Attached To Airplane Oxygen Masks Does Serve A Purpose
- 16 Wearing Pre-Distressed Sneakers Is Not 'Subversive'
- 111 Oklahoma City Thunder Agrees To Contract Extensions With Victor Oladipo And Steven Adams
- 27 The Normalization Of Pete Rose
- 36 Why So Many Millennials Experience Imposter Syndrome
- 23 There's Still Room For Disruption In The Cloud

**INTERNATIONAL**

Monitoring

ALL Updated

BBC SEARCH: 2758

HEADLINE Required: Sinosphere: Making Online Literature Pay Big in China (53/42 characters)

ABSTRACT: A scene from Douluo Dalu, an online game based on Zhang Wei's digital novel of the same name.

BYLINE: Add content here.

DATELINE: Mosul battle.

BODY Required: New York Times

Zhang Wei, China's top-earning online novelist, is not modest about his success. "I love writing, and I'm gifted," Mr. Zhang, 35, said in a recent interview.

The numbers back him up. In 2015, Mr. Zhang, better known by his pen name, Tang Jia San Shao, earned 110 million renminbi, about \$16.8 million at the time, according to China Daily. Much of his fortune was made from selling his so-called IPs, a buzzword in China referring to intellectual properties or original content that is often adapted into movies, television shows and games. It's a strategy that in recent years has become a major source of revenue for China's online literature websites and writers. Mr. Zhang's earnings would put him on par with best-selling authors like Stephen King and George R. R. Martin.

Semantic Analysis

ENTITIES: ADD NEW ENTITY

Person: Zhang Wei (4), Tang Jia San Shao (1), Stephen King (1), George R. R. Martin (1).

Location: China (4).

TOPICS: None.

TEMPORAL INFORMATION: None.

**INTERNATIONAL**

Monitoring

ALL Updated

BBC SEARCH: 2758

HEADLINE Required: Sinosphere: Making Online Literature Pay Big in China (53/42 characters)

ABSTRACT: A scene from Douluo Dalu, an online game based on Zhang Wei's digital novel of the same name.

BYLINE: Add content here.

DATELINE: Mosul battle.

BODY Required: New York Times

Zhang Wei, China's top-earning online novelist, is not modest about his success. "I love writing, and I'm gifted," Mr. Zhang, 35, said in a recent interview.

The numbers back him up. In 2015, Mr. Zhang, better known by his pen name, Tang Jia San Shao, earned 110 million renminbi, about \$16.8 million at the time, according to China Daily. Much of his fortune was made from selling his so-called IPs, a buzzword in China referring to intellectual properties or original content that is often adapted into movies, television shows and games. It's a strategy that in recent years has become a major source of revenue for China's online literature websites and writers. Mr. Zhang's earnings would put him on par with best-selling authors like Stephen King and George R. R. Martin.

**INTERNATIONAL**

Monitoring

ALL Updated

BBC SEARCH: 2758

HEADLINE Required: Sinosphere: Making Online Literature Pay Big in China (53/42 characters)

ABSTRACT: A scene from Douluo Dalu, an online game based on Zhang Wei's digital novel of the same name.

BYLINE: Add content here.

DATELINE: Mosul battle.

BODY Required: New York Times

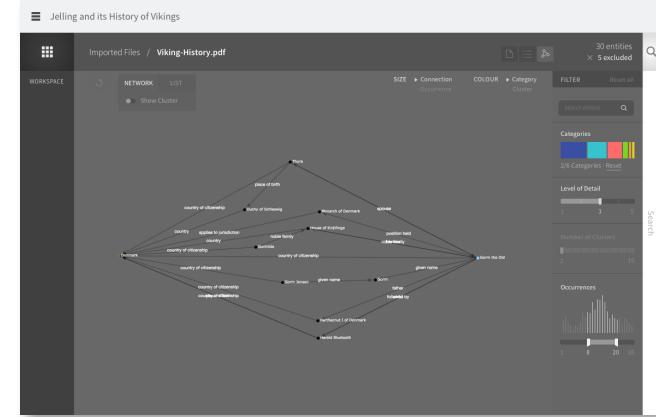
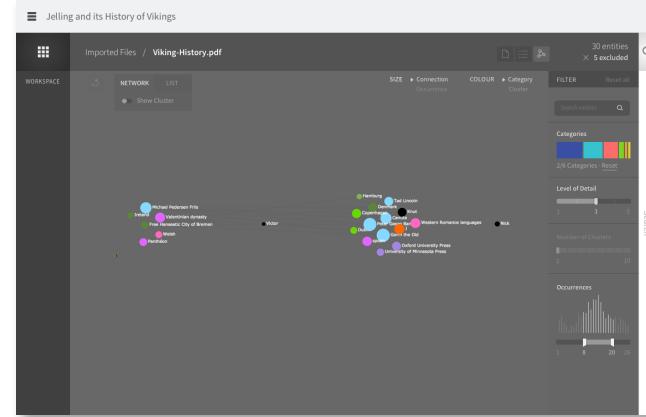
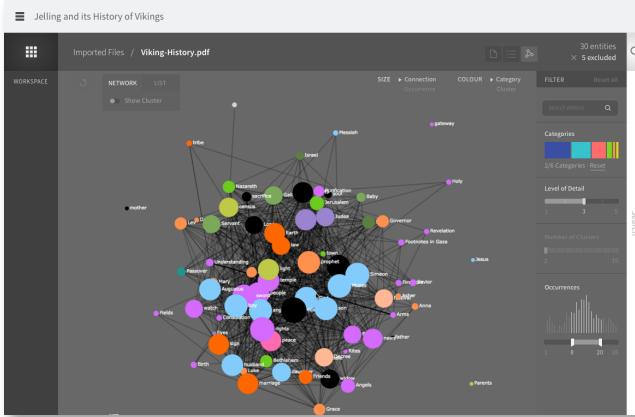
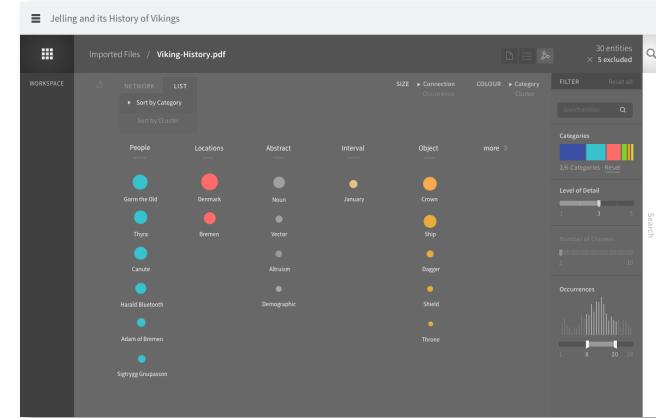
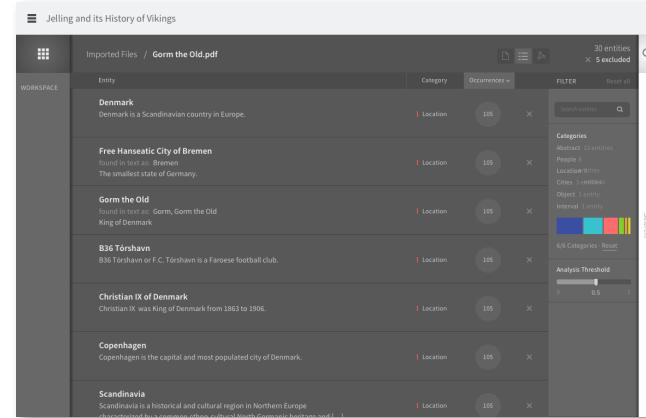
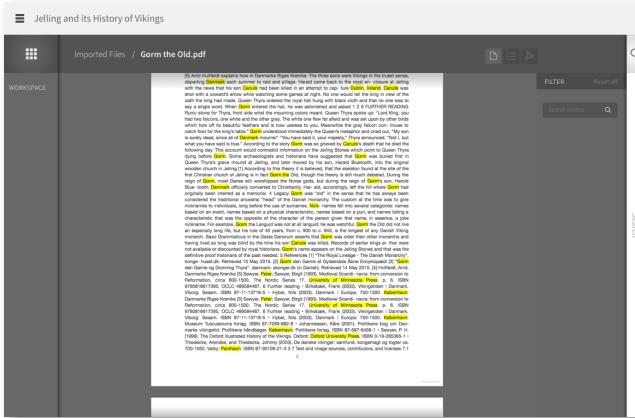
Zhang Wei, China's top-earning online novelist, is not modest about his success. "I love writing, and I'm gifted," Mr. Zhang, 35, said in a recent interview.

The numbers back him up. In 2015, Mr. Zhang, better known by his pen name, Tang Jia San Shao, earned 110 million renminbi, about \$16.8 million at the time, according to China Daily. Much of his fortune was made from selling his so-called IPs, a buzzword in China referring to intellectual properties or original content that is often adapted into movies, television shows and games. It's a strategy that in recent years has become a major source of revenue for China's online literature websites and writers. Mr. Zhang's earnings would put him on par with best-selling authors like Stephen King and George R. R. Martin.

Related Content

- Shane Warne Facebook video earns cricketers
- Nine Qualities Of A Good Boss And How...
- Three Leadership Lessons From The Cubs...
- Sinosphere: Making Online Literature P...
- Ubisoft Says It's Learned DLC Lessons From...

Julian Moreno-Schneider, Ankit Srivastava, Peter Bourgonje, David Wabnitz, and Georg Rehm. Semantic Storytelling, Cross-lingual Event Detection and other Semantic Services for a Newsroom Content Curation Dashboard. In Octavian Popescu and Carlo Strapparava, editors, Proceedings of Natural Language Processing meets Journalism - EMNLP 2017 Workshop (NLPMJ 2017), Copenhagen, Denmark, September 2017. 7. September.



**Branche: Museen,  
Showrooms, Ausstellungen**

Georg Rehm, Jing He, Julian Moreno Schneider, Jan Nehring, and Joachim Quantz. Designing User Interfaces for Curation Technologies. In Sakae Yamamoto, editor, *Human Interface and the Management of Information: Information, Knowledge and Interaction Design, 19th International Conference, HCI International 2017*, number 10273 in Lecture Notes in Computer Science (LNCS), pages 388-406, Vancouver, Canada, July 2017. Springer.



Thema

Suche

Beiträge Perlenfänger Serendipities Summarization Timelining

Regiona report Antenne Brandenburg 05:30

Regiona report Büro Pferleberg 05:30

Kurznachrichten Antenne Brandenburg 05:30

Regional report Studio Frankfurt 05:30

Verkehrsexport INFOradio

(Hintergrund) Neuartige Antriebe bei Omnibussen

072915/NR/PZ/5.30 Regio Prenzlau

072915/NR/PZ/5.30 Regio Prenzlau

MÄRKTE 3-Chinesische Börsen drehen leicht ins Minus

0530 FritzNachrichten 29-07-15

r1wv05:45-29.07.2015

Geheimdienste/Spionage /Israel/USA/ (Zusammenfassung 0530 - neu mit Weissem Haus, Details zu Pollard, Hintergrund) USA lassen Ex-Spion Pollard nach 30



Thema

Suche

Beiträge Perlenfänger Serendipities Summarization Timelining

Die Stadt Luckau übernimmt heute symbolisch den Schlabendorfer See. Brandenburgs Finanzminister Görke übergibt ihn am Vormittag dem Bürgermeister der Stadt, Lehmann. Gestern hatte bereits die Stadt Brandenburg an der Havel den Fruchtwetzel erhalten.

Das Land Brandenburg hatte insgesamt 128 Seen vom Bund gekauft, um den Bürgern weiterhin Zugang zu gewähren. Die Gewässer sollten ursprünglich privatisiert werden, später aber als ein Naturpark ausgewiesen werden.

Spanier lernen Lausitzer Handwerksbetriebe unterschiedlichen Handwerkskammer Lukas Rotnicki. Das Nadelohr/Kokuvit geht in die nächste Berliner Straße wird der Straßenbelag Umleitung eingerichtet. Betroffen sind dauernd.

Wetter Aktuell werden in Berlin und Brandenburg es anfangs oft länger sonnig Schauer, vereinzelt Gewitter. Maximal Verkehr Eine Meldung zur A9 Berlin Richtung Zwischen Klein Marzehns und Koselitz. Weitere Meldungen aus der Verkehrs

Ein 14-jähriger Abiturient wird am 1. C aufnehmen Conrad Lempert geboren soeben erfahren, dass seine Bewerber gestiftete und finanzierte Institut an der am HPI-Schülerkollegprogramm teilgenommen Informatik, in diesem Jahr den zweiten



## Branche: TV, Web-TV, Medien

Zugleich geht Ankara seit der vorigen Woche gegen die Terrormiliz Islamischer Staat in Syrien vor. Die NATO hatte der Türkei gestern ihre Solidarität im Kampf gegen den Terror versichert.

Die türkische Luftwaffe hatte gestern ihre Angriffe auf Stellungen der kurdischen PKK ausgeweitet und erstmals auch Einrichtungen auf türkischem Boden bombardiert. Das türkische Parlament will heute mittag in einer Sondersitzung über die Sicherheitslage beraten. Zuvor hatte Ankara sich darauf beschränkt, kurdische Ziele im Nordirak zu attackieren.

USA verteidigen Erdogan

Die US-Regierung hat die türkischen Luftangriffe auf die kurdische PKK als Akt der Selbstverteidigung eingestuft. Die PKK habe Anschläge auf türkische Polizisten verübt und sei der Aggressor, hieß es von ranghohen Regierungsbeamten in Washington.

Auch die NATO hatte gestern auf Kritik an dem türkischen Vorgehen gegen kurdische Rebellen verzichtet. Allgemein erklärte sich der NATO-Rat aber solidarisch mit dem türkischen Vorgehen gegen den Terrorismus. Die Türkei bombardiert seit voriger Woche erstmals Stellungen des Islamischen Staates in Syrien.

Die US-Regierung stuft die türkischen Luftangriffe auf die kurdische PKK als einen eindeutigen Akt der Selbstverteidigung ein. Das sagten ranghohe Regierungsvertreter in Washington. Die PKK habe Anschläge auf türkische Polizisten verübt und sei der Aggressor.

Washington hatte in den vergangenen Tagen mehrfach betont, man betrachte die PKK noch wie vor als Terror-Organisation. Die NATO hatte Ankara gestern ihre - wie es heißt - "starke Solidarität" im Kampf gegen den Terrorismus zugesichert. Türkische Kampfjets griffen danach erstmals auch wieder kurdische Stellungen im Südosten der Türkei an. Zudem hatte Staatspräsident Erdogan den Friedensprozess mit der PKKfrage gestellt. Das türkische Parlament wird heute über die angespannte Sicherheitslage im Land beraten.

6 - Der türkische Staatspräsident Erdogan hat den Friedensprozess mit den Kurden offiziell für beendet erklärt. Er sagte laut Medienberichten, Politiker mit Verbindungen zu terroristischen Gruppen müssten juristisch belangt werden. Erdogan forderte die Justiz auf, gegen HDP Abgeordnete zu ermitteln, da diese mit der verbotenen PKK zusammenarbeiteten.



Deutsches  
Forschungszentrum  
für Künstliche  
Intelligenz GmbH

# **Kuratierungstechnologien für unterschiedliche Anwendungsfälle**

- **BMBF-Projekt Digitale Kuratierungstechnologien:**

- Museen, Showrooms, Ausstellungen
- TV, Web-TV, Medien
- Öffentliche Archive
- Journalismus



**3pc**  
Neue Kommunikation



Digitale  
Kuratierungs-  
technologien

- **BMBF-Projekt Digitale Kuratierungstechnologien:**

- Museen, Showrooms, Ausstellungen
- TV, Web-TV, Medien
- Öffentliche Archive
- Journalismus

ART+COM



3pc  
Neue Kommunikation

kreuzwerker  
we touch running systems



Digitale  
Kuratierungs-  
technologien

- **Außerdem Kuratierungstechnologien konzipiert bzw. in der Entwicklung befindlich für:**

- Juristische und regulatorische Inhalte
- Bibliothekswissenschaft und Digital Libraries
- Customer-Relationship-Management
- Medizinische Inhalte
- Film- und Kinobranche
- Digital Humanities



Georg Rehm. KI für die Kundenkommunikation: Der Markt der Zukunft. Rethink! Connected Customer 360°. Hamburg, June 22/23, 2017.

Georg Rehm. Artificial Intelligence for the Film Industry. FilmTech Meetup Berlin, July 25, 2017.

Clemens Neudecker und Georg Rehm. „Digitale Kuratierungstechnologien für Bibliotheken“. *Zeitschrift für Bibliothekskultur* 027.7, Open Access. Nov. 2016.

Georg Rehm. Der Mensch bleibt im Mittelpunkt – Smarte Technologien für alle Branchen. *Vitako Aktuell. Zeitschrift der Bundes-Arbeitsgemeinschaft der Kommunalen IT-Dienstleister e.V.*, 2-2016:26-27, 2016.

# Kuratierungstechnologien

- **Kuratierungstechnologien:** Branchenspezifische Anwendungen, die auf KI-Technologien basieren
- **KI-Technologien:** Symbolische Verfahren, statistische Verfahren, maschinelles Lernen, Deep Learning
- Anwendung auf typischerweise hochgradig spezielle **Digital Humanities**-Datensammlungen ist ambitioniert

Kuratierungs-technologien für	Anforderungen	Erste Priorität
Diverse Branchen	Eher flache Analyse	Hoher Abdeckungsgrad
Humanities	Eher tiefe Analyse	Hohe Präzision

# **Beobachtungen und Schlussfolgerungen**

# Schlussfolgerungen

- Kuratierungstechnologien unterstützen Wissensarbeiter – inkl. DH-Anwender – beim Verarbeiten digitaler Inhalte.
- Kuratierungstechnologien werden eingesetzt, um digitalen Content semantisch zu erschließen und für den jeweiligen Anwendungsfall nutzbar zu machen.
- Auch im Bereich Digital Humanities kann durchaus von „Kuratierungstechnologien“ gesprochen werden.
- Ziele: Bessere Analysen und Nutzbarkeit der Daten; Findbarkeit; Kontextualisierung und Visualisierung (Karten, Zeitstrahl, Verknüpfung, LOD etc.).

# KI – Reality Check

- **Künstliche Intelligenz**
  - Beeindruckende Durchbrüche in den vergangenen Jahren
  - Basieren u.a. auf extrem großen Datenmengen
  - Entwicklung disruptiver, revolutionärer KI-Tools für die Arbeit mit DH-Daten ist eher unwahrscheinlich
  - Aber: Standardwerkzeuge wie NER etc. werden kontinuierlich verbessert
- **Arbeit mit DH-Daten**
  - Hochgradig spezifische Datensammlungen und Anwendungsfälle
  - Anforderungen: Hohe Präzision der Annotation sowie der Metadaten
  - Eher kleine und sehr spezielle Datenmengen
  - Sehr tiefe und spezifische Anforderungen bzw. Annotationen, aufwendig in der Automatisierung

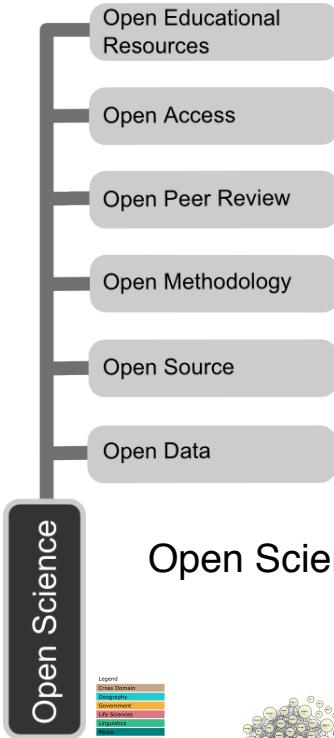
# Prognosen

- Immer präzisere Basistechnologien (NER etc.)
- Entwicklung generischer, adaptiver Daten-Annotations- und Analyse-Workbenches für DH-Forschung/-Anwender
- Im Hintergrund: „AI-as-a-Service for Digital Humanities“
- KI-Technologien (maschinelle Lernverfahren) in der Basisinfrastruktur, auch von Nicht-Experten mit generischen Annotations-Werkzeugen nutzbar
- Domänenspezifische Spezialwerkzeuge, Semantiken, Metadaten etc. – inhärent vernetzt mit LOD, ORKG etc.

# Beobachtungen

- In der Vergangenheit zahlreiche fragmentierte Digital-Humanities-Projekte (Deutschland und darüber hinaus)
- Unterschiede bzgl. Software, Plattformen, Datenformate
- Angestrebt werden sollte eine Zusammenführung von:
  - Forschung
  - Forschungsergebnissen (Dokumentation ggf. standardisiert)
  - Forschungsdaten (Annotation ggf. standardisiert)
  - Metadaten und Datenformaten (standardisiert)
  - Infrastrukturen/Plattformen (Interoperabilität, standardisiert)
  - Repositorien von Forschungsdaten und Services
- „Digital Humanities“ allein ist zu kurz gedacht
- Eigentlich sollten wir über „Digital Science“ sprechen

# Wichtige Initiativen

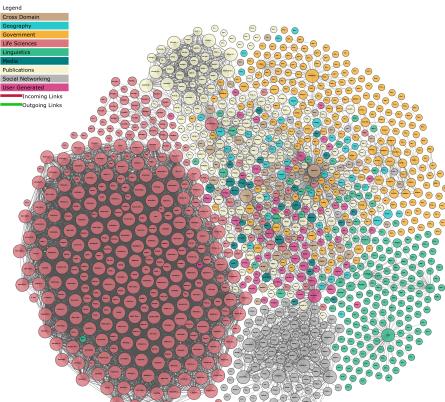


Open Science



RI	
Link zu diesem Datensatz	<a href="http://dbpedia.org/resource/Person_(ontology)">http://dbpedia.org/resource/Person_(ontology)</a>
Typ	Person (pz)
Person	Appelbaum, Anne
Geschlecht	weiblich
Zeit	Lebensdaten: 1908-1998
Land	Deutschland (XA-DE); USA (XD-US)
Geografischer Bezug	Wirkungsort: New York, NY
Beruf(e)	Psychologin
Beziehungen zu Personen	Cassier, Ernst (Vater)
Beteiligt an	1. Nachgelassene Manuskripte und Texte / Bd. 1. Zur Metaphysik der symbolischen Formen 1995

Gemeinsame Normdatei



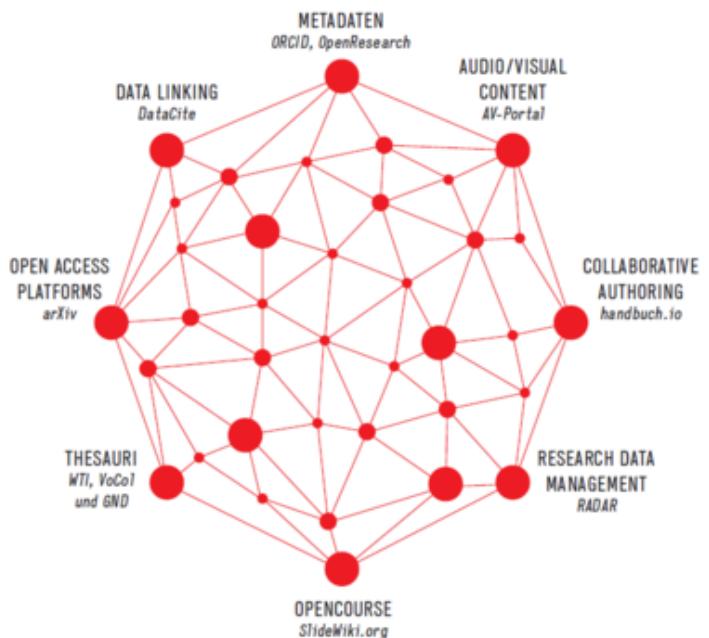
Linked Open Data Cloud

Schaffung einer nationalen Forschungsdateninfrastruktur



EOSC pilot  
The European Open Science Cloud for Research Pilot Project

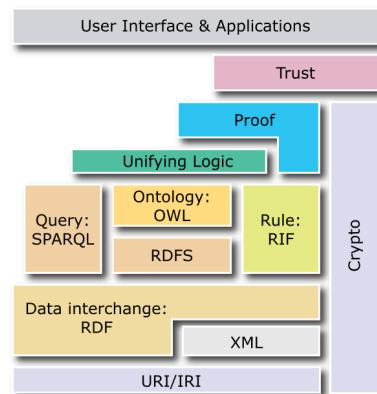
European  
Open Science  
Cloud



Open Research Knowledge Graph



Standards



Semantic Web

# DFKI: Schwerpunkte bislang

- **Infrastrukturen für Forschungsdaten:** META-SHARE
- **Generische Sprach- und Wissenstechnologien für branchenspezifische Kuratierungstechnologien:** DKT
- **Nutzung, Annotation, Dashboards:** DKT
- **Metadaten und Standards:** W3C, Web Annotations, Semantic Web, XML, XSLT, SPARQL etc.
- **Metadaten, Dokumentation von Forschungsdaten und Prozesse:** Language Resource Life Cycle

# DH am DFKI Berlin: Zukunft

- Am DFKI in Berlin werden wir uns in Zukunft vermehrt auch um das Thema Digital Humanities kümmern
- Projekte mit Bezügen zu Digital Humanities:
  - **DKT** (BMBF) – abgeschlossen 2015-2017
  - **Open-Science-Projekt** (TUB) Start Q2/2018
  - **KT/DH-Projekt** (Berliner Hochschule) Start Q2/2018
  - **DKT2** (BMBF) – Förderung empfohlen Start Q4/2018
  - **DFG-Verbundprojekt** – in Vorbereitung Start Q4/2018
- Das DFKI freut sich auf eine intensivere Kooperation mit ifIDHib und der DH-Forschung in Berlin!

# Vielen Dank!