

Strukturierte Informationserfassung und Auswertungspotentiale im Projekt Buber-Korrespondenzen Digital

Julian Jarosch, M.A.
Dr. Lea Müller-Dannhausen
Akademie der Wissenschaften und der Literatur | Mainz

TOPs

- Zum Projekt
- CMIFerator
- BKD-Ontologie
- Fazit



Zum Projekt

BKD Korpus & Umfang

- Korrespondenzpartner:innen: ca. 7.000
- Korrespondenzen: ca. 41.400 Einheiten
- Zeitspanne: 1895 – 1965
- 16 Sprachen im weltweiten Raum

thematische Module

1. Jüdische Renaissance – Kulturzionismus – Palästina/Staat Israel
2. Religion(en) und Religionswissenschaft
3. Dialogisches Denken – Religionsphilosophie
4. Sprachphilosophie – Übersetzung – Bibelkommentar
5. Herausforderungen an das Christentum und kritischer Dialog mit der christlichen Theologie
6. Literatur – Kunst – Theater
7. Politische Philosophie und Sozialphilosophie
8. Philosophische Anthropologie – Pädagogik – Psychologie/Psychotherapie



CMIFerator

konfigurierbare flexible CMIF-Erstellung



Grundlagen

<https://github.com/digicademy/cmiferator>

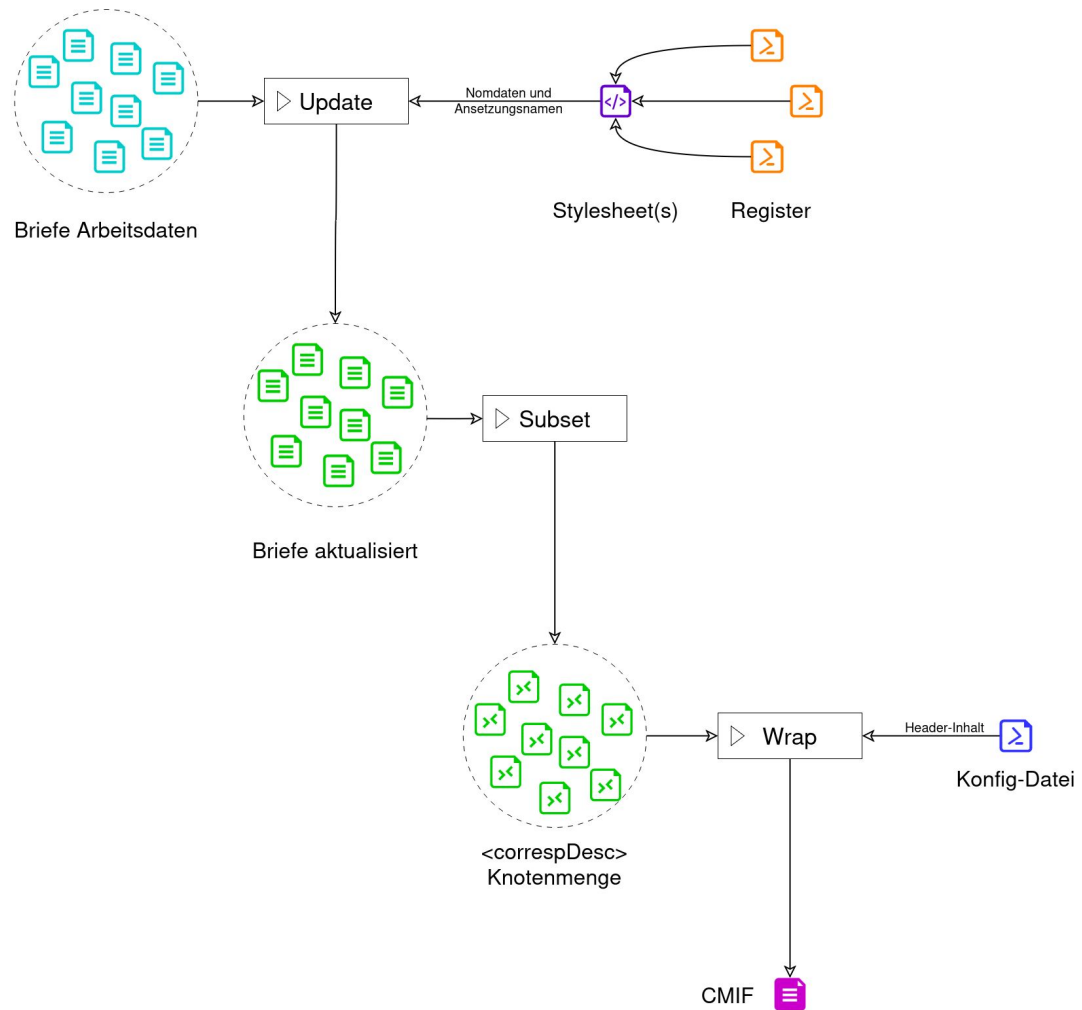
Der CMIFerator ist eine **Funktionsbibliothek**, die typische Transformationsschritte vom **Arbeits-XML** zur **CMIF-Datei** anbietet

- Paket für eXist-DB
- Teilschritte optional auch einzeln nutzbar/kombinierbar
 - oder als Gesamtworkflow
- für dynamische Schnittstelle oder statische Generierung
- derzeit für CMIF 1.0 ausgelegt

gemeinsame Bedarfe der Mainzer Briefprojekte

- Aktualisierung der `<correspAction>`-Inhalte anhand Register
 - Hinzufügung von Normdaten in `@ref`
 - Aktualisierung der Ansetzungsnamen
- Filterung der umfangreicheren TEI-All-Möglichkeiten von `<correspDesc>` auf CMIF-Subset
 - z.B. Auslassung von `<correspContext>`
 - auch: Einfügung von obligatorischem `<persName>` wo vorher keins vorhanden
- Einfüllen in die vorgegebene CMIF-Dateistruktur
 - plus Metadaten

vollständiger Workflow



Einrichtung

Zur Benutzung des CMIFerator muss man:

- CMIFerator mit **Paket**-Manager in eXist-DB **installieren**
- **Konfigurationsdatei** anlegen
- **Register**-Transformation
 - Transformationen für ediarum-Register aus Git-Repo übernehmen
 - falls eigenes, abweichendes Register-Schema: selbst eine Transformation schreiben
- **XQuery**-Skript, das die CMIFerator-Konfig und -Funktionen aufruft anhand Beispiel aus Git-Repo anpassen

Beispiel Sozinianische Briefwechsel

```
<correspDesc>
  <correspAction type="sent">
    <persName key="Curtz">Curtz, Albert</persName>
    <placeName key="ucx_lgx_12b">Neuburg</placeName>
    <date when="1665-08-21"/>
  </correspAction>
  <correspAction type="received">
    <persName key="Lubieniecki_Stanisław">Lubieniecki, Stanisław</persName>
    <placeName key="Hamburg">Hamburg</placeName>
  </correspAction>
</correspDesc>
```

Aktualisierung des Ansetzungsnamens zu
Albert Curtz SJ und Hinzufügung von Normdaten

```
<correspDesc source="#b3b22a15-9906-406b-aae1-7d7fa2292e71" ref="https://sozinianer.de/id/
  <correspAction type="sent">
    <persName ref="https://d-nb.info/gnd/124643221">Albert Curtz SJ</persName>
    <placeName>Neuburg</placeName>
    <date when="1665-08-21"/>
  </correspAction>
  <correspAction type="received">
    <persName ref="https://d-nb.info/gnd/119357100">Stanisław Lubieniecki</persName>
    <placeName ref="https://www.geonames.org/2911298">Hamburg</placeName>
  </correspAction>
</correspDesc>
```

Beispiel STURM Edition

```
<correspDesc key="Bl.398" ref="https://sturm-edition.de/id/Q.01.19210921.JVH.01">
  <correspAction type="sent">
    <persName key="P.0000004" ref="http://d-nb.info/gnd/11854764X">Jacoba van Heemskerck</persName>
    <placeName key="O.0000004" ref="http://sws.geonames.org/2747373">Domburg</placeName>
    <date when="1921-09-21">21.9.21</date>
  </correspAction>
  <correspAction type="received">
    <persName key="P.0000001" ref="http://d-nb.info/gnd/118770950">Herwarth Walden</persName>
  </correspAction>
</correspDesc>
```

Normdaten sind bereits enthalten, Aktualisierung entfällt

Der Textinhalt von `<date>` wird entfernt (CMIF v 1.0!)

```
<correspDesc source="#STURM" ref="https://sturm-edition.de/id/Q.01.19210921.JVH.01">
  <correspAction type="sent">
    <persName ref="http://d-nb.info/gnd/11854764X">Jacoba van Heemskerck</persName>
    <placeName ref="http://sws.geonames.org/2747373">Domburg</placeName>
    <date when="1921-09-21"/>
  </correspAction>
  <correspAction type="received">
    <persName ref="http://d-nb.info/gnd/118770950">Herwarth Walden</persName>
  </correspAction>
</correspDesc>
```

BKD

```
<correspDesc ref="https://bkd.adwmainz.net/correspondences/BKD00022/BKD00022.19020112.01.S">
  <correspAction type="sent">
    <persName key="P.0000001">Buber, Martin</persName>
    <placeName key="0.0000007">Wien</placeName>
    <date when="1902-01-12" evidence="conjecture" cert="high"/>
  </correspAction>
  <correspAction type="received">
    <persName key="P.0000018">Achad Ha'am</persName>
    <placeName key="0.0000565" evidence="conjecture" cert="high">Odessa</placeName>
  </correspAction>
</correspDesc>
```

BKD ist ein ›klassischer‹ Anwendungsfall
für den vollständigen Workflow

Allerdings erweitert um eine
projektspezifische Zusammenstellung
von Informationen nach CMIF Version 2 in `//correspDesc/note`

Das XSLT für CMIFv2 ist hier dem CMIFerator-Workflow
einfach vorangestellt


```

<correspDesc source="#ae451a0a-2186-450c-ae5d-72024981937c" ref="https://lod.academy/bkd/id/BKD00022.19020112.01.S">
  <correspAction type="sent">
    <persName ref="https://d-nb.info/gnd/118516477">Buber, Martin</persName>
    <placeName ref="https://www.geonames.org/2761369">Wien</placeName>
    <date when="1902-01-12"/>
  </correspAction>
  <correspAction type="received">
    <persName ref="https://d-nb.info/gnd/118717510">Achad Ha'am</persName>
    <placeName ref="https://www.geonames.org/698740">Odessa</placeName>
  </correspAction>
  <note>
    <ref type="https://lod.academy/cmif/vocab/terms#hasTextBase" target="https://lod.academy/cmif/vocab/terms#Manuscript"/>
    <ref type="https://lod.academy/cmif/vocab/terms#isAvailableAsTEIfile" target="https://lod.academy/bkd/id/BKD00022.19020112.01.S"/>
    <ref type="https://lod.academy/cmif/vocab/terms#hasLanguage" target="de"/>
    <ref type="https://lod.academy/cmif/vocab/terms#isPublishedWith" target="https://lod.academy/cmif/vocab/terms#Abstract"/>
    <ref type="https://lod.academy/cmif/vocab/terms#isPublishedWith" target="https://lod.academy/cmif/vocab/terms/#Transcription"/>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/124434290">Ben-Ami, Mordechai</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/118972472">Berdyczewski, Micha Josef</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/118663453">Birnbaum, Nathan</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/116383747">Ehrenpreis, Marcus</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/119226723">Gaster, Moses</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/116126264">Kellner, Leon</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/119010046">Peretz, Isaac Leib</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/1029911517">Schiller, Salomon</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/127056637">Thon, Osias</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/118629387">Wassermann, Jakob</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/118808281">Zangwill, Israel</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/11679724X">Friedlaender, Israel</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/130053007">Nossig, Alfred</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/1057673374">Weisengrün, Paul</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsPerson" target="https://d-nb.info/gnd/143744232">Rawnitzki, Yehoshua Hana</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsOrg" target="https://www.wikidata.org/entity/Q1186113">Demokratische Fraktion</ref>
    <ref type="https://lod.academy/cmif/vocab/terms#mentionsOrg" target="https://www.wikidata.org/entity/Q1637904">Jüdischer Verlag</ref>
  </note>
</correspDesc>

```



BKD-Ontologie

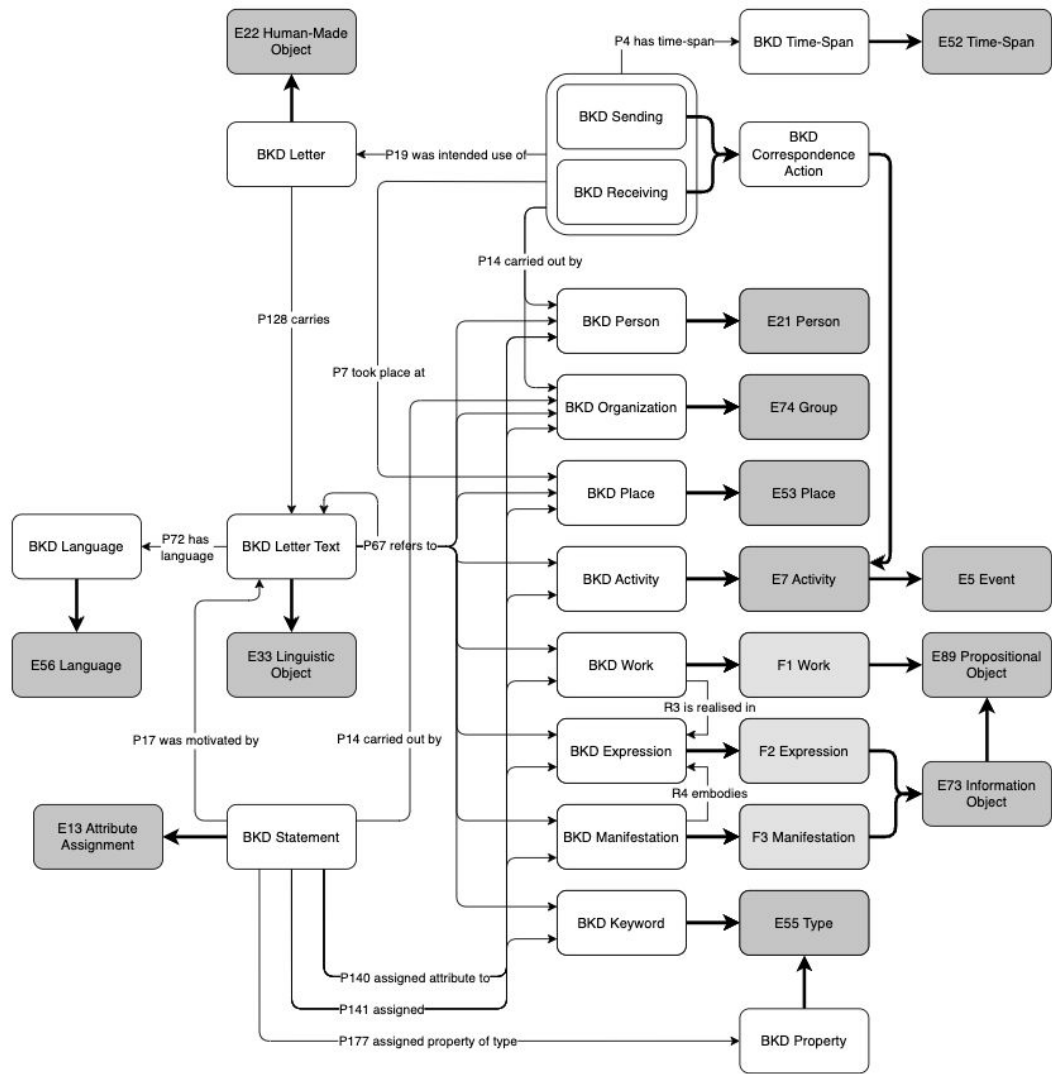
Ontologie

- ermöglicht formale Wissensrepräsentation in Form eines **Knowledge Graphen**
- als Beschreibung von Klassen (Arten von Dingen) und ihren Eigenschaften (Beziehungen)
- auf Basis von Technologien des **Semantic Web**:
 - Uniform Resource Identifier (URI)
 - Resource Description Framework (RDF) -> Syntax (Tripel)
 - RDF Schema (RDFS) -> Klassen und ihre Beziehungen

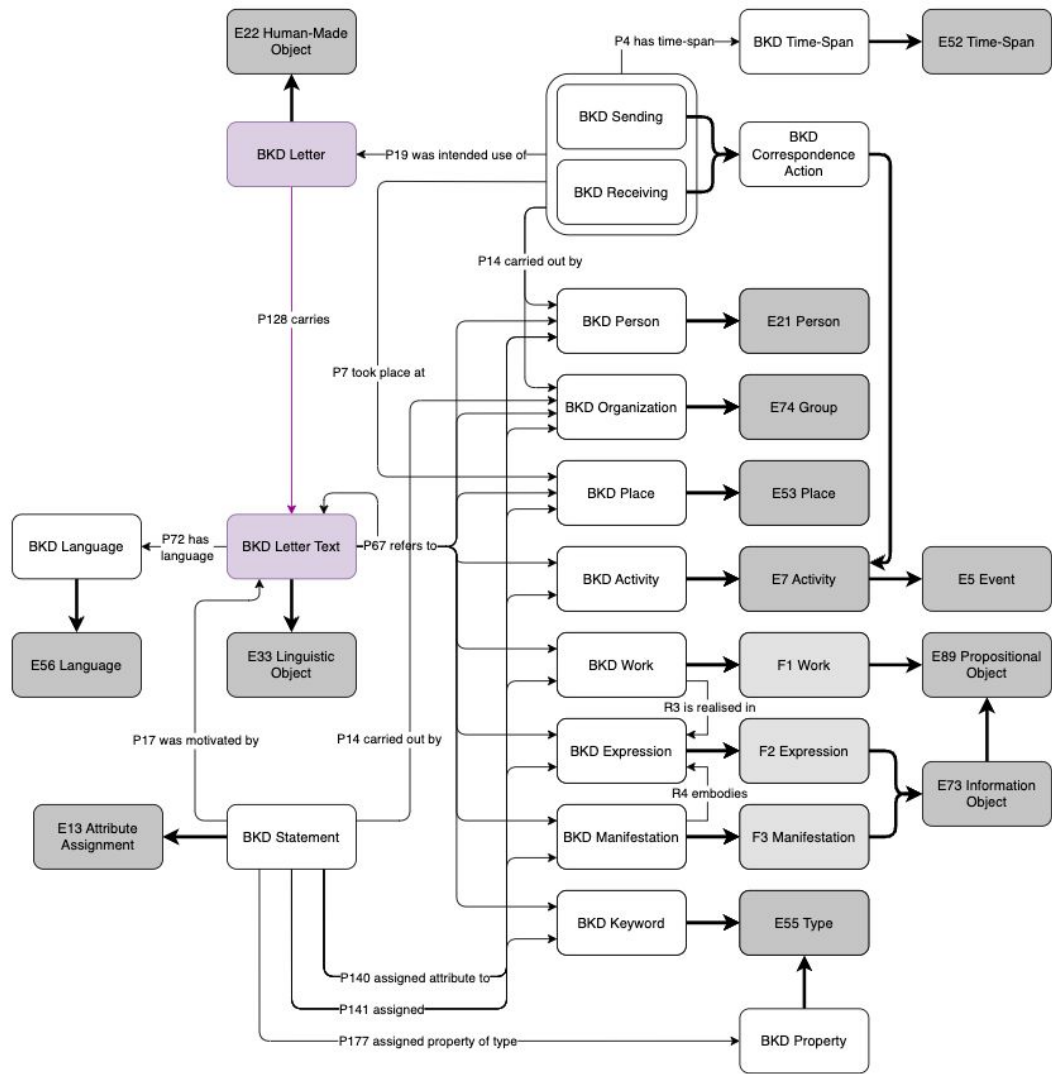
BKD- Ontologie

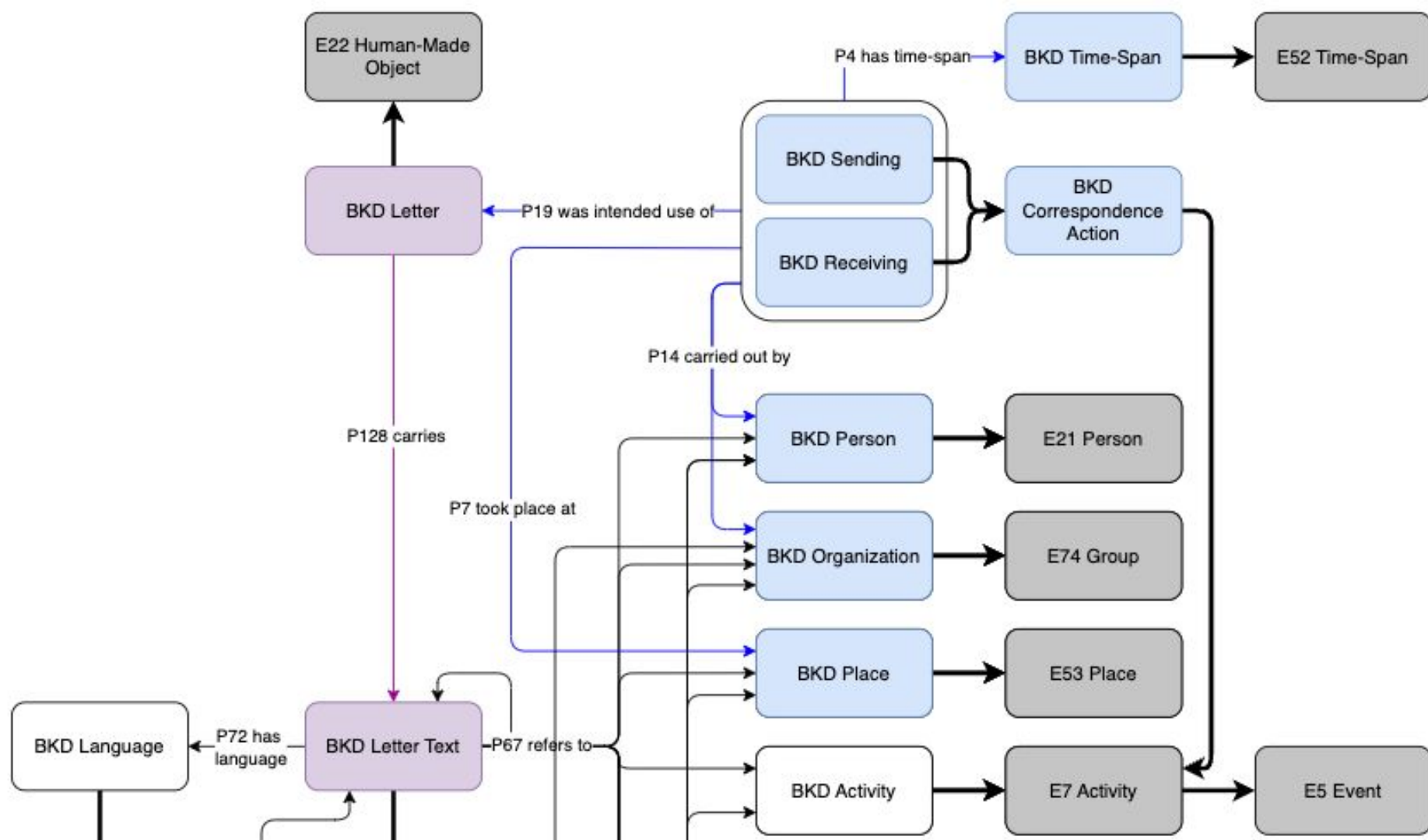
- Status: Entwurfsphase
- Konzentration auf “Rahmendaten” der Korrespondenz/Kommunikationsakte (<correspDesc>) und deren inhaltliche Erschließung (<abstract> und <listRelation>)
- Verwendung bestehender und bewährter (Upper) Ontologies, Vokabulare und Thesauri: CIDOC Conceptual Reference Model (CIDOC CRM), Library Reference Model (LRMoo/FRBROo), GND, Geonames etc.

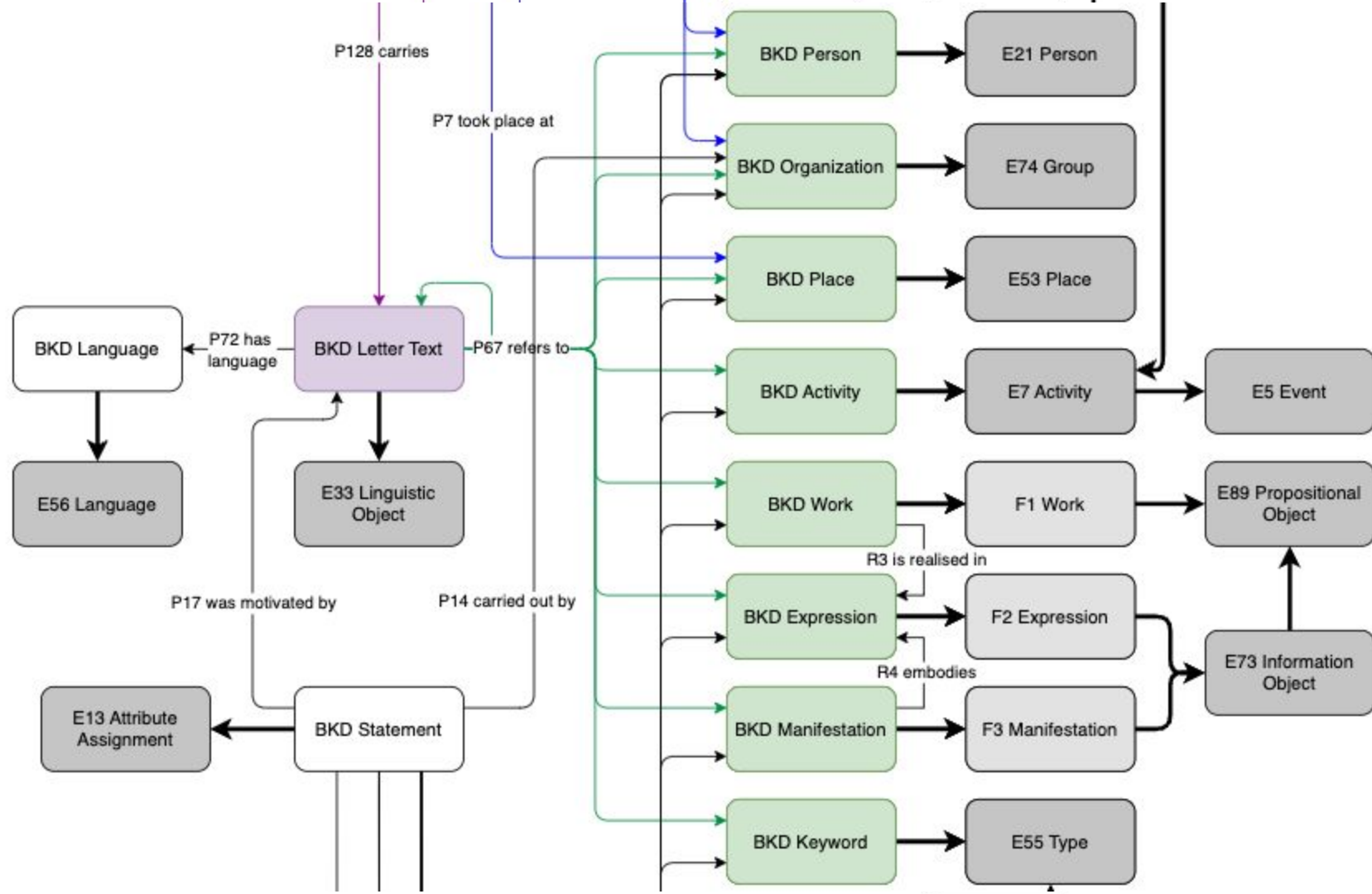
BKD- Ontologie



BKD- Ontologie

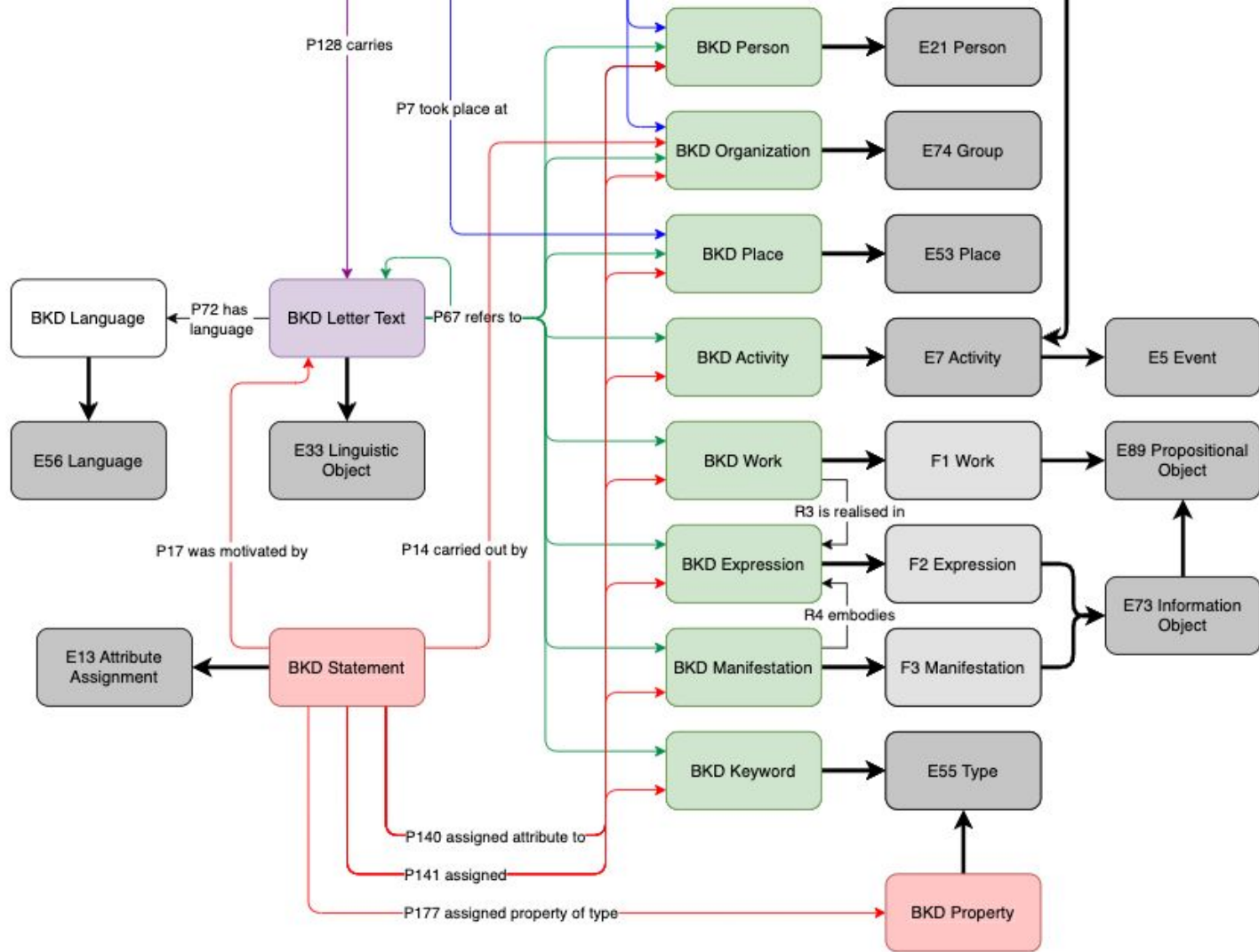






Erwähnungen

```
<abstract>
  <list type="content">
    <item>...</item>
  </list>
  <list type="persons">
    <item>
      <name key="P.0002131">Herzl, Theodor</name>
    </item>
  </list>
  <list type="organizations">
    <item>
      <name key="K.0000004">Büro des Zionistenkongresses</name>
    </item>
  </list>
  <list type="places">
    <item>
      <name key="O.0000177">Salzburg</name>
    </item>
  </list>
  <list type="works">
    <item>
      <name key="W.0000287.01.01">Herzl, Theodor: Eröffungsrede Dr. Herzls</name>
    </item>
  </list>
  <list type="events">
    <item>
      <name key="E.0000005">5. Zionistenkongress</name>
    </item>
  </list>
  <list type="keywords">
    <item>
      <name key="S.0000001">Zionistische Bewegung</name>
    </item>
  </list>
```



Relationen

```
<listRelation>
  <relation active="P.0002131" name="recommends" passive="W.0000284.01.01"/>
  <relation active="P.0002131" name="rejects" passive="W.0000286.01"/>
  <relation active="P.0000001" name="supporterOf" passive="P.0005878"/>
  <relation active="P.0006964" name="employeeOf" passive="K.0000007"/>
  <relation active="E.0000005" name="topicOf" passive="W.0000297.01"/>
</listRelation>
</abstract>
```


BKD- Relationen

activelyIn – adviserOf – affiliatedWith – allyOf – attacks –
basedIn – bornIn – characterIn – creatorOf – commissionedBy –
congratulates – consentWith – contributorTo – createdIn –
criticizes – dedicatedTo – defends – dependentOn – diedIn –
discourages – dissentWith – editorOf – employeeOf –
encourages – enemyOf – foundedIn – founderOf – friendOf –
fundedBy – greets – headOf – informs – inspiredBy – invites –
invitedTo – keptBy – lauds – livesIn – meets – memberOf –
negotiatesWith – offendedBy – opponentOf – organizedBy –
participantOf – partOf – partsWith – perceives – presentedIn –
publishedIn – publisherOf – recommends – reconciledWith –
refersTo – rejects – relativeOf – reviewerOf – revises –
sendsWork – staysIn – storedIn – studentOf – sues –
supporterOf – takesPlace – topicOf

Pläne und Potentiale

- vielfältige **Analysemöglichkeiten** der Forschungsdaten durch Maschinenlesbarkeit und Maschinen-Auswertbarkeit
- **Verfügbarmachung** der Forschungsdaten in Form von Linked Open Data (LOD.ACADEMY)
- **Verknüpfung** mit Daten anderer digitaler Editionen
- **Übertragbarkeit** -> Verwendung der Ontologie für ähnliche Korpora

Fazit

Wir tun viel dafür, unsere Projektergebnisse **nachnutzbar**, **anschlussfähig** und **findbar** zu machen.

- **nachnutzbar** durch projektübergreifende **Generalisierung** der CMIF-Erstellung
- **anschlussfähig** durch systematisches und strukturiertes Erschließen mit Mitteln, die die Daten **interoperabel** machen
- **findbar**, indem wir ausgiebig **correspSearch** bedienen

Danke für's Zuhören!

eine leere
Inhaltsfolie

das ist die Vorlage für normale Inhaltsfolien