

IDS Repository

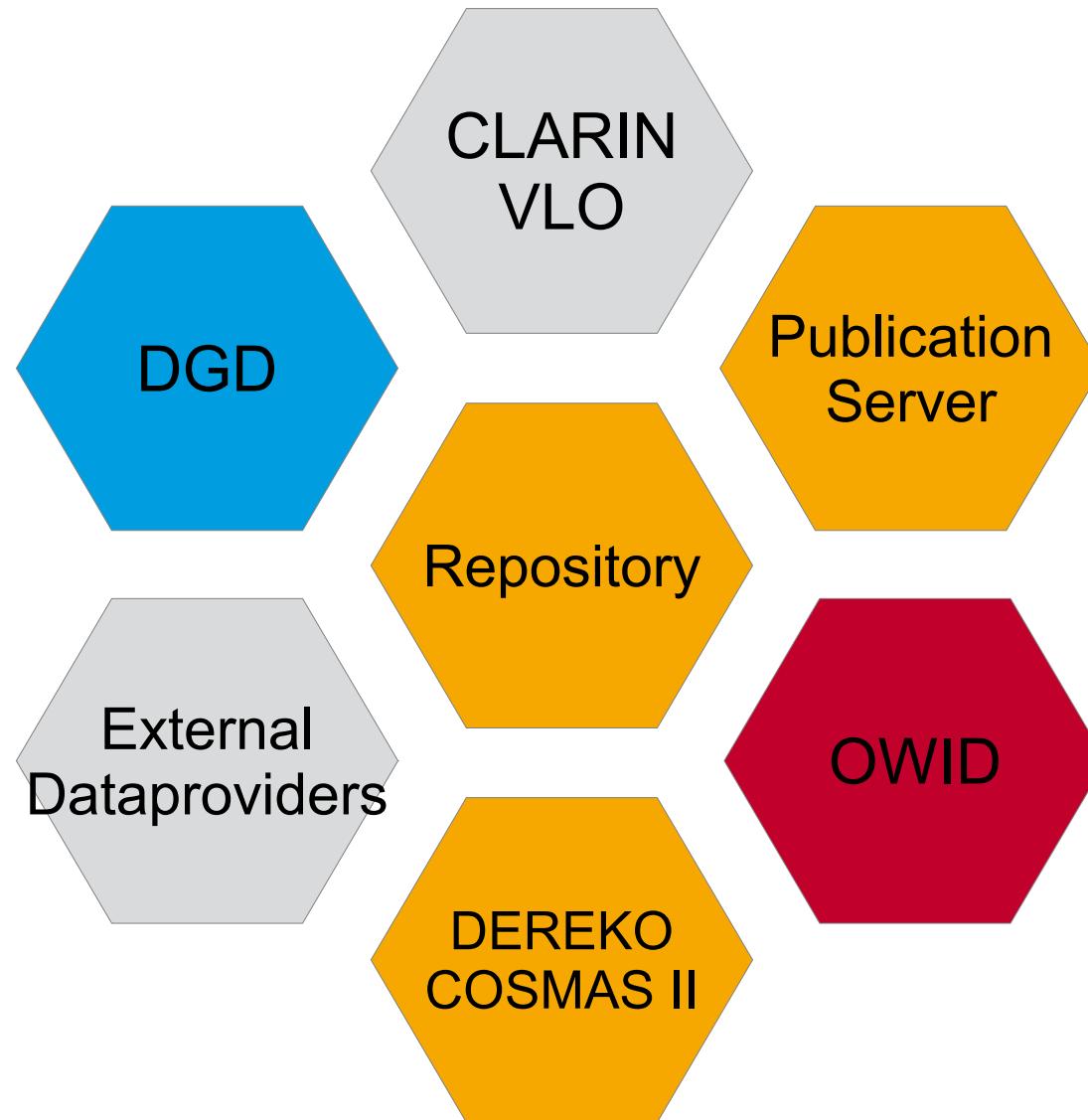
Peter Fankhauser

Mission Statement

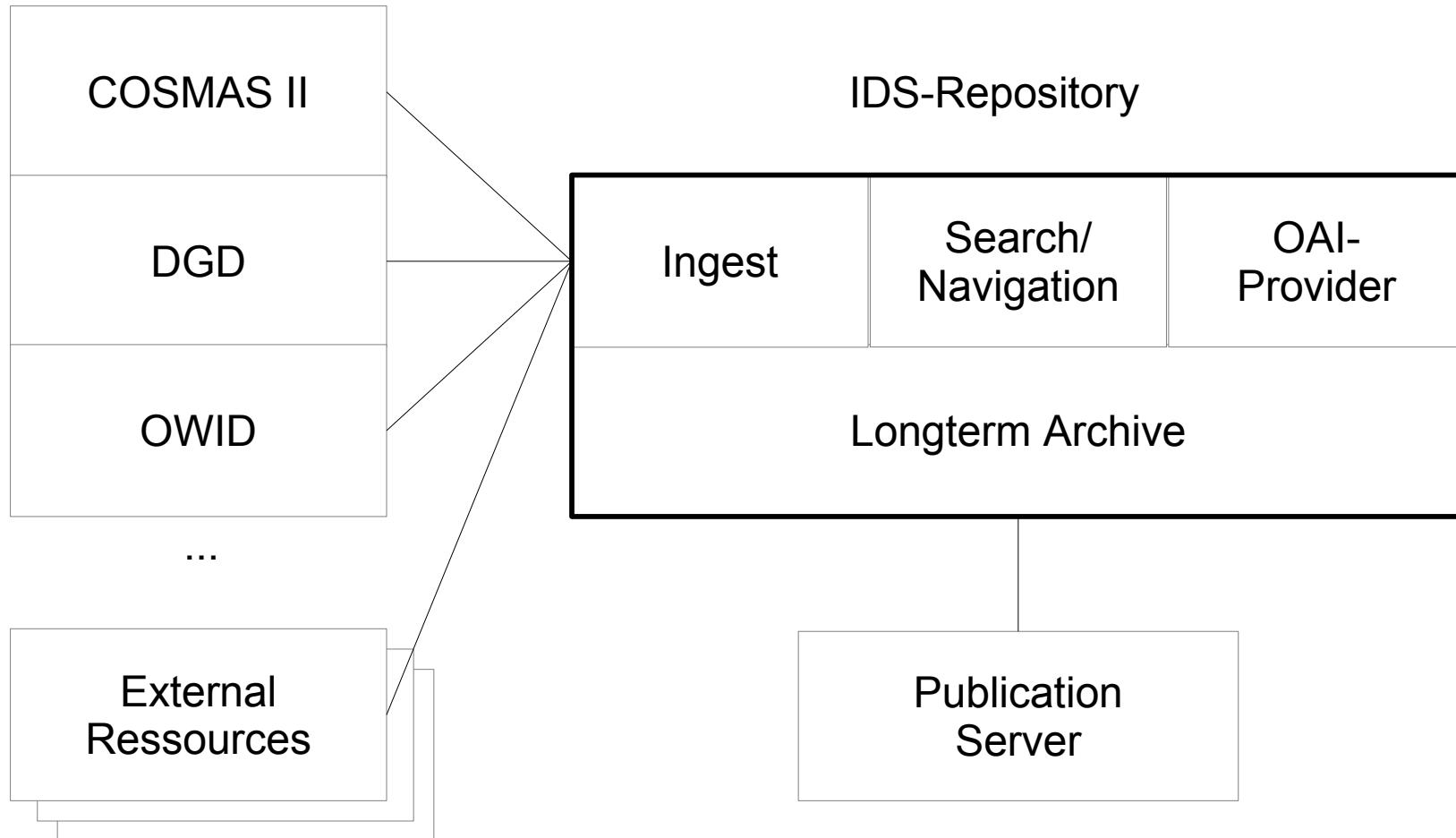
"The foundation pursues the purpose of scientifically researching and documenting the German language in its contemporary use and its more recent history.

It cooperates with other national and international institutions with a similar goal, and also provides scientific services."

IDS Repository in Context (1)



IDS Repository in Context (2)



Corpora/Resources

- Written Language
 - German Reference Corpus (Dereko)
 - Mannheimer Corpus of Historical Newspapers
 - Discourse in the Weimar Republic
- Spoken Language
 - DGD (Datenbank Gesprochenes Deutsch)
 - Gewiss (Uni Leipzig)
- Lexical Resources
 - OWID

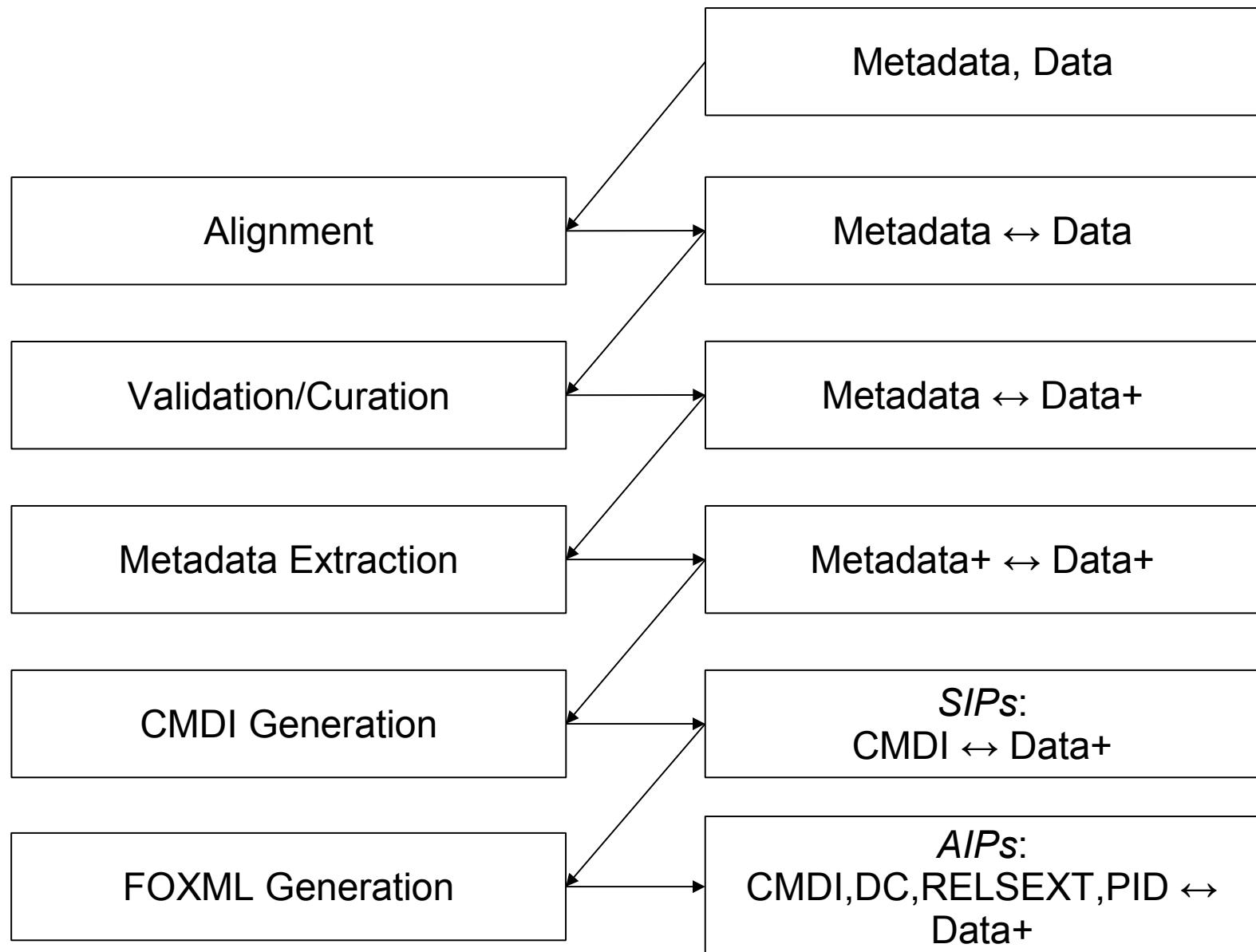
Challenges

- Resources consist of many objects
 - Ingest = Migration
- Heterogeneous Data Formats
 - TUSTEP
 - .doc/.rtf
- Heterogeneous, incomplete Metadata
 - Comma separated
 - Informal documentation
 - Various XML Dialects

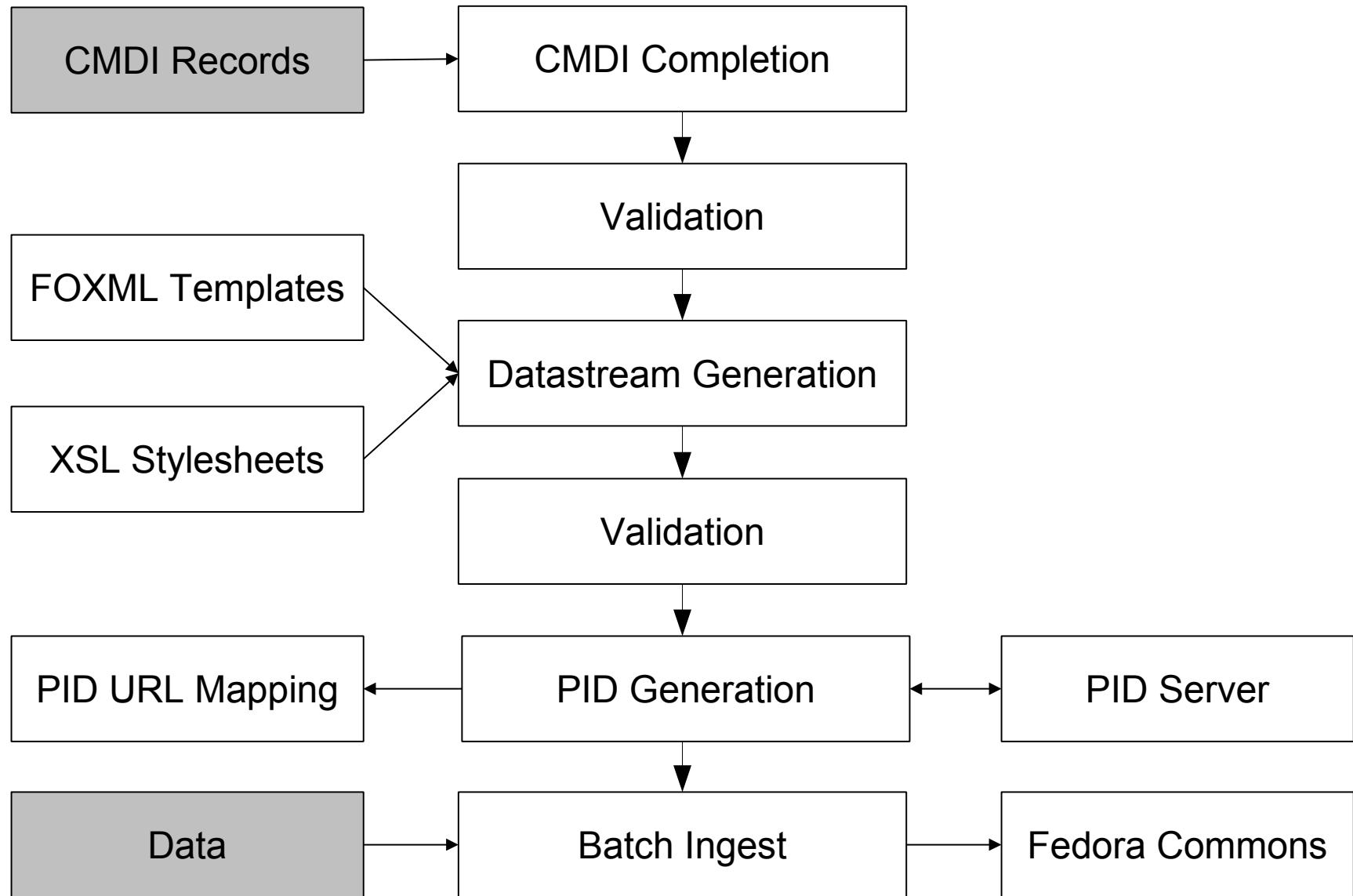
Modelling Principles

- Data Formats
 - Maintain original formats: support current use
 - Generate sustainable formats (e.g. TEI)
- Metadata
 - Maintain original metadata (if XML)
 - Formalize metadata by way of CMDI + ISOcat
- Persistent Identifiers
 - PIDs for Metadata
 - PIDs for Data

General Ingest Pipeline



From CMDI to FOXML



Example Configuration

```
<?xml version="1.0"?>
- <config xmlns="http://repos.ids-mannheim.de">
  - <datastreams xmlns="http://repos.ids-mannheim.de">
    <datastream template="../templates/cmddiDatastream.xml" validate="yes" filepattern=".cmddi"/>
    <datastream template="../templates/dcDatastream.xml" stylesheet="../stylesheets/dc.xslt"/>
    <datastream template="../templates/relextDatastream.xml" stylesheet="../stylesheets/relext.xslt"/>
    <datastream template="../templates/xmlDatastream.xml" filepattern=".xml"/>
  </datastreams>
</config>

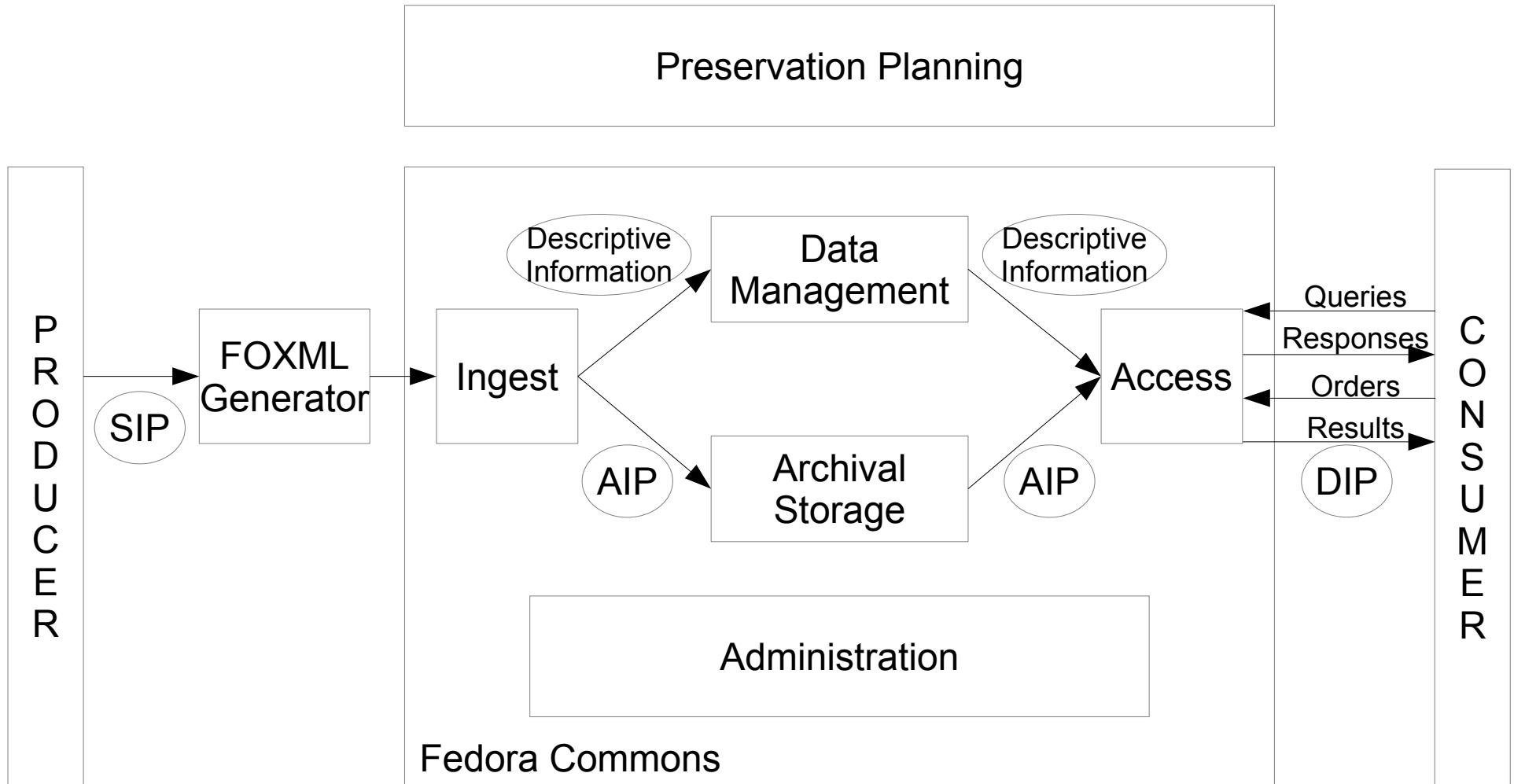
<?xml version="1.0"?>
- <foxml:datastream CONTROL_GROUP="X" STATE="A" ID="DC" xmlns:foxml="info:fedora/fedora-
  system:def/foxml#">
  - <foxml:datastreamVersion ID="DC.0" LABEL="Dublin Core Record for this object"
    MIMETYPE="text/xml" FORMAT_URI="http://www.openarchives.org/OAI/2.0/oai_dc/">
    <foxml:xmlContent/>
  </foxml:datastreamVersion>
</foxml:datastream>

<?xml version="1.0" encoding="UTF-8"?>
- <xsl:stylesheet xmlns:dc="http://purl.org/dc/elements/1.1/" version="1.0"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <xsl:output version="1.0" indent="yes" encoding="UTF-8" method="xml"/>
  <xsl:param name="id"/>
  - <xsl:template match="/">
    - <oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:cmdi="http://www.clarin.eu/cmd/"
      xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/">
      - <xsl:element name="dc:identifier">
        <xsl:value-of select="//cmdi:MdSelfLink"/>
      </xsl:element>
      - <xsl:if test="//cmdi:IDSAGD_Corpus/cmdi:Name">
```

Foxml Ingest Object

```
<?xml version="1.0" encoding="UTF-8"?>
<foxml:digitalObject xsi:schemaLocation="info:fedora/fedora-system:def/foxml# http://www.fedora.info/definitions/1/0/foxml1-1.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" PID="clarind-ids:mkhz.000002" VERSION="1.1" xmlns:foxml="info:fedora/fedora-system:def/foxml#">
  - <foxml:objectProperties>
    - <foxml:property VALUE="A" NAME="info:fedora/fedora-system:def/model#state"/>
    - <foxml:property VALUE="Weltgeschichte0001, Seite 25" NAME="info:fedora/fedora-system:def/model#label"/>
  </foxml:objectProperties>
  - <foxml:datasream CONTROL_GROUP="X" STATE="A" ID="CMDI">
    - <foxml:datasreamVersion ID="CMD.0" LABEL="CMDI Record for this object" FORMAT_URI="http://www.clarin.eu/cmd/
      http://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/profiles/clarin.eu:cr1:p_1288172614026/xsd" MIMETYPE="application/xml">
      - <foxml:xmlContent>
        - <CMD xsi:schemaLocation="http://www.clarin.eu/cmd/ http://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/profiles/clarin.eu:cr1:p_1288172614026/xsd
          xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" CMDVersion="1.1" xmlns="http://www.clarin.eu/cmd/">
          - <Header>
            <MdCreator>generateFoxml.pl</MdCreator>
            <MdCreationDate>2013-02-04</MdCreationDate>
            <MdSelfLink>http://hdl.handle.net/10932/00-017B-E0F5-76B7-4F01-D</MdSelfLink>
            <MdProfile>clarin.eu:cr1:p_1288172614026</MdProfile>
            <MdCollectionDisplayName>Institut für Deutsche Sprache, CLARIN-D Zentrum, Mannheim</MdCollectionDisplayName>
          </Header>
          - <Resources>
            - <ResourceProxyList>
              - <ResourceProxy id="TEXT">
                  <ResourceType mimetype="text/plain">Resource</ResourceType>
                  <ResourceRef>http://hdl.handle.net/10932/00-017B-E0F5-8B27-5001-0</ResourceRef>
                </ResourceProxy>
              - <ResourceProxy id="JPEG">
                  <ResourceType mimetype="image/jpeg">Resource</ResourceType>
                  <ResourceRef>http://hdl.handle.net/10932/00-017B-E0F5-9F37-5101-6</ResourceRef>
                </ResourceProxy>
            </ResourceProxyList>
            <JournalFileProxyList/>
            <ResourceRelationList/>
            - <IsPartOfList>
              <IsPartOf>http://hdl.handle.net/10932/00-017B-E0F5-6287-4E01-8</IsPartOf>
            </IsPartOfList>
          </Resources>
          - <Components>
            - <OLAC-DcmiTerms>
              <description xml:lang="de">Der allerneuesten Europäischen Welt= und Staats=Geschichte II. Theil No. IV. 2. Woche. Freytags den 10. Jenner, MDCCXLIV.
                Schluß derer Gedancken von Neuen=Jahrs= Zetteln.</description>
              <format>text/plain</format>
              <format>image/jpeg</format>
              <hasPart>http://hdl.handle.net/10932/00-017B-E0F5-8B27-5001-0</hasPart>
              <hasPart>http://hdl.handle.net/10932/00-017B-E0F5-9F37-5101-6</hasPart>
              <Identifies>http://hdl.handle.net/10932/00-017B-E0F5-6287-4E01-8</Identifies>
            </OLAC-DcmiTerms>
          </Components>
        </CMD>
      </foxml:xmlContent>
    </foxml:datasreamVersion>
  </foxml:datasream>
</foxml:digitalObject>
```

IDS Repository Functional Model



JPEG

TUSTEP

Mainzer Journal.

Das Mainzer Journal erscheint täglich (mit Ausnahme der höchsten Feiertage) und zwar so, daß das Exemplar mit den Mittwochen verhandelt wird. Bezahlungen nehmen alle Postämter an; für Mainz und die nächste Umgebung die Handels-Beilage am Vormittage des betreffenden Tages gleich ausgetragen wird. Bezahlungen nehmen alle Postämter an; für Mainz und die nächste Umgebung die Buchhandlung von Kirchheim, Scott und Eitzenmann am Nachmittag. Der Preis des Blattes ist wie in Mainz jährlich zu 8. Pfennig, alle Art weiter aufzurunden und die dreispaltige Petitzile über deren Raum mit 8. Kr. berechnet.

Nr. 8. Donnerstag, den 22. Juni. 1848.

Wegen des hohen Frohleichtagsfestes erscheint morgen keine Zeitung. Die zu der heutigen Nummer gehörnde Beilage wird Freitag Vormittag um 11 Uhr ausgegeben.

Der erste Artikel der Grundrechte des deutschen Volkes nach dem Entwurf der Verfassungskommission.

II Vom Rheine 20. Juni. „Dem deutschen Volke ist gewidmete Freiheit des Betriebs und Wegertheit, welche der Verhandlung dieser Freiheit bedienten, sowie der Verhandlung dieser Freiheit, welche der Betrieb und Wegertheit aller freien und öffentlichen Dienste und der Betrieb des Staates und der Städte befandt es nicht. Für die Betreuer der aller Religionen Gleichheit vor dem Gesetz. Es ist ausdrücklich die Bildung neuer Religionsgesellschaften gestattet. (Das Verhältniß von Kirche und Staat bestreitet.) Niemand kann zu einer kirchlichen Handlung oder Geistlichkeit gezwungen werden. Die Kirche ist ausdrücklich aufzuheben.“ Also lautet dieser Artikel.

Mit Freuden haben alle wahren und einflussreichen Freunde des Vaterlandes und der Religion wahre Freiheit begrüßt, weil nur auf dem Boden der Freiheit wahre Religionsfreiheit blühen kann. Sie glauben, daß jetzt endlich die unwürdige Bevormundung der Religion durch die Staatspolitik aufhorten, sie haben mit Zuversicht erwartet, daß die deutsche Nationalversammlung den großen und herzlichen Wunsch der religiösen Freiheit und der Unabhängigkeit der Kirche vom Staat, wie in Belgien und Nordamerika bestellt, für alle Konfessionen klar, bestimmt und rechtlich anprechbar werde. Sie haben sich darüber sicherlich gemacht, als nun durch die sohn der ehrlichen Durchführung des Grundbegriffs der religiösen Freiheit und der Unabhängigkeit der Kirche vom Staat eine aufrichtige und innige politische Einheit Deutschlands möglich ist. All den Erscheinungen entwegen hat sie der angeführte Artikel des Commissionsentwurfs bitter gefaßt.

Statt jener klaren und großen Säße der religiösen Freiheit, finden wir nichts als geschriften, bedeutungslose Bestimmungen und hinterhältige Gauleins. Das jeder seine Religion frei betreuen darf, versteht sich von selbst. Allein was wir fordern und verlangen müssen, ist unendlich mehr. Wir müssen auch fordern, daß jede religiöse Gesellschaft, jede Konfession, jedes Kirche als als Corporation in ihrer Verfassung, Verwaltung, in ihrer Lehre, in ihrem Gottesdienst, in der Wahr und Anstellung ihrer Diener, in dem Betrag, dem Erwerb und der Verhandlung ihres Eigentums frei und unabhängig sei und ohne jede Behinderungen unterliege, als denen, welche das allgemeine Volk, allen Staatsbürgern, in privatrechtlicher und staatsrechtlicher Bevormundung leidet, anfangs das bißchen der Staat sich in alle kirchlichen Sündhaftigkeiten einzumischen, bis zur Haftung von Spenden, die Gläubiger Kirchlicher Abstellungen oder Schreinen, so sogar die Gültigkeit aller kirchlichen Urteile von seiner Genehmigung abhängt und damit die Wahl und Ablösung der Geistlichen vielfach bis sehr weit verzögert, die Bermischungsverwaltung der Kirche seinen Besitzern ununterbrochen, so sogar die Freiheit des Gottesdienstes betreffende Bestimmungen trifft. Allein die französische Verfassungskommission sah nicht daran zu haben, daß dieser unrechte Zustand der Kirche auch unter der Herrschaft der neuen Freiheit fortbesteure. Oder meint man vielleicht,

I Siehe Beilage 3. zum Mainzer Journal.

@@@<filename:MainzerJournal(1)0079.jpg/>

\$0 @@1@<SE79>

\$0079.01 @z <Z><W+1>Mainzer Journal</W+1>.</Z>

\$0079.02 #(IMAGE)

\$0079.03 <P>_<W-1>Das Mainzer Journal erscheint täglich (mit Ausnahme der höchsten Feiertage) „blättern“ schon am Vorabende, die ständige Beilage am Vormittage des betreffenden Tages für Mainz und die nächste Umgebung die Buchhandlung von #S+Kirchheim#S- und #S+Luziaß schriftliche Buchhandlung von 2 fl.; in dem gesammten Gebiete des Fürstentums Mainz werden aufgenommen und die dreispaltige Petitzile oder deren Raum in Mainz jährlich zu 8. Pfennig, alle Art weiter aufzurunden und die dreispaltige Petitzile über deren Raum mit 8. Kr. berechnet.

\$0079.08 #(IMAGE)

\$0079.09 <P>{Numero} 43. <Z><W+1>Freitag, den 28. Juli</W+1>.</Z> <R>1848.</R></P>

\$0079.10 #(IMAGE)

\$0079.11 @z <duolan 2><Z>Der zwölften Juli und der sechste August.

\$0079.12 @Z -----</Z>

\$0079.13 <P>_{sj} Wunderbar ist das Walten der Wiedervergeltung in der

\$0079.14 Geschichte und es scheint ihr zu fallen, oft selbst in kleinen,

\$0079.15 scheinbar zufälligen Umständen eine Spur und ein Merkzeichen

\$0079.16 ihres Waltens auszudrücken.</P>

\$0079.17 <P>_Am 12. Juli 1848 hat der Reichsverweser, Erzherzog Joseph von Oesterreich sein Amt feierlich angetreten -- und am

\$0079.18 künftigen 6. August wird er vollständig von seiner Gewalt Besitz

\$0079.19 ergreifen, indem an diesem Tage die ganze bewaffnete Macht

\$0079.21 Deutschlands ihm huldigen und dadurch alle bisherigen Landes-

\$0079.22 und Bundesstruppen in Reichstruppen verwandelt werden sollen.</P>

\$0079.23 <P>_Am 12. Juli des Jahres 1806 aber war es, wo der schmach-

\$0079.24 vollste, und am 6. August desselben Jahres, wo der traurigste

\$0079.25 Akt der ganzen deutschen Geschichte stattfand. Am 12. Juli 1806

\$0079.26 nämlich schlossen die Könige von Bayern und Württemberg, die

\$0079.27 souveränen Fürsten von Regensburg, Baden, Berg, Hessen-

\$0079.28 Darmstadt, Nassau und andere größere Fürsten des südlichen und

\$0079.29 westlichen Deutschlands #S+in Paris den Rheinbund unter

\$0079.30 dem Protektorat Napoleons, #S- und sagten sich dadurch

\$0079.31 los vom deutschen Reich, und kündeten dem Kaiser, in Übereinstimmung mit dem französischen Gesandten, ihre bisherige

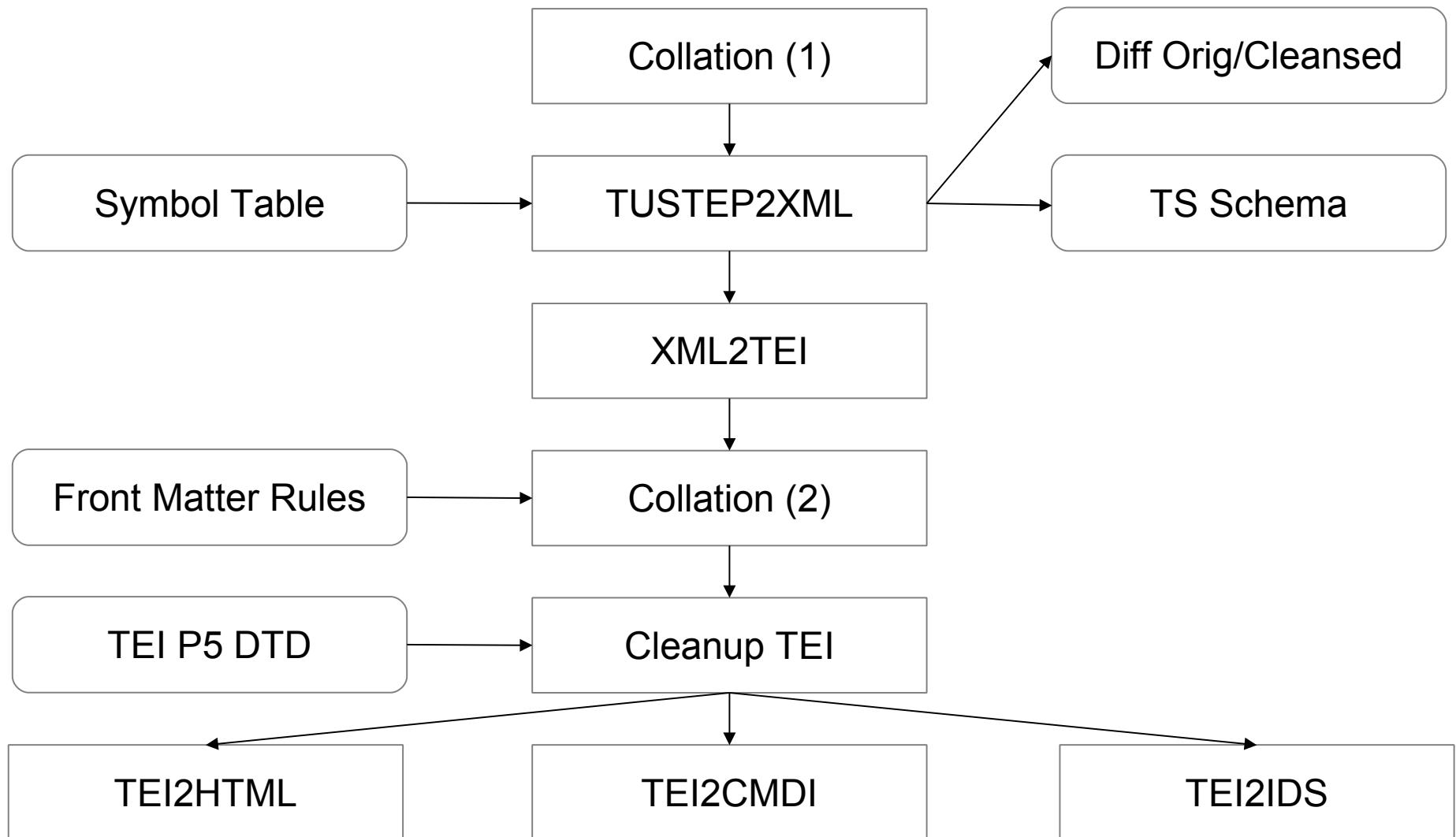
\$0079.32 Abhängigkeit auf. Um den eitlen Flitter einer scheinbaren Sou-

\$0079.34 verainetät, und den Vorheil eigener Gebietsvergrößerungen,

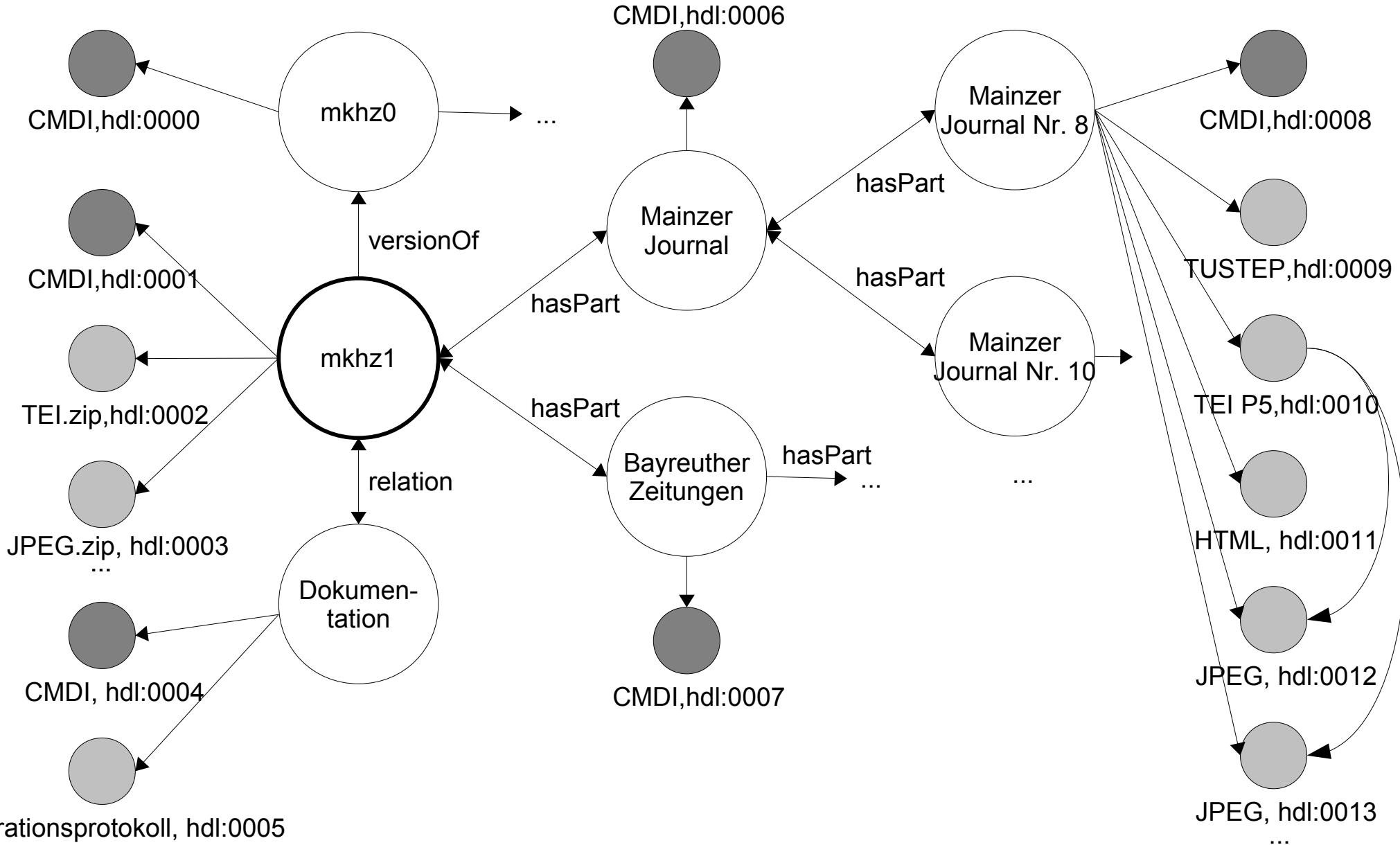
MKHZ: Problems and Solutions

- Encoding Issues
 - Filenames (diacritics, whitespace)
 - Normalization of filenames
- Granularity
 - Pages → (n:1) → Transcriptions → (n:1) → Journal
 - Collate into individual Issues
- TUSTEP Markup
 - Rulebased Cleansing and Extraction
- Missing formal meta data
 - Manual Extraction from Documentation
 - Automatic Extraction of dc:title, dc:date, ... from filenames and transcriptions

TUSTEP to TEI



MKHZ: Object Model



Navigation based on CMDI

Mannheimer Korpus Historischer Zeitungen und Zeitschriften

[de] Das Mannheimer Korpus Historischer Zeitungen und Zeitschriften besteht aus 21 deutschsprachigen Zeitungen und Zeitschriften des 18. und 19. Jahrhunderts und umfasst etwa 652 einzelne Ausgaben mit insgesamt über 4.1 Mio laufenden Wortformen auf 4678 Druckseiten. Dieses Korpus wurde zwischen 2009 und 2011 zusammengestellt und digitalisiert, und 2013 von TUSTEP nach TEI P5 konvertiert.

[en] The Mannheim Corpus of Historical Newspapers and Magazines consists of 21 German newspapers and magazines from the 18th and 19th century. It comprises about 652 individual volumes with over 4.1 Mio word tokens on 4678 pages overall. This corpus has been assembled and digitized from 2009 to 2011, and been converted to TEI P5 in 2013.

Type: collection
PID: <http://hdl.handle.net/10932/00-01B8-AE41-41A4-DC01-5>
Publisher: Institut für Deutsche Sprache

hasPart : Der allermeisten Europäischen Welt= und Staats=Geschichte II. Theil (1744)
hasPart : Allgemeine Auswanderungs=Zeitung (1846 - 1848)
hasPart : Allgemeine Zeitung (1871)
hasPart : Arbeitgeber. Archiv für die gesamte Volkswirtschaft, Central-Anzeiger für Stellen- und Arbeitsergesuche (1870 - 1877)
hasPart : Die Bayerische Presse (1850)
hasPart : Bayreuther Zeitungen (1752)
hasPart : Berlinische Nachrichten von Staats- und gelehrten Sachen (1740 - 1741)
hasPart : Berlinische Privilegierte Zeitung (1737)
hasPart : Conversations=Blatt zur Unterhaltung und Belehrung für alle Stande (1836 - 1838)
hasPart : Das Heller=Blatt oder Magazin zur Verbreitung gemeinnütziger Kenntnisse (1834)
hasPart : Das Pfennig=Magazin der Gesellschaft zur Verbreitung gemeinnütziger Kenntnisse (1836 - 1855)
hasPart : Das wohlfeilste Panorama des Universums zur ertheilenden Belehrung für Jedermann und alle Lander (1834 - 1836)
hasPart : Deutsche Auswanderer=Zeitung (1852)
hasPart : EUROPA Wochenschrift für Kultur und Politik (1905)

Mainzer Journal

Type: collection
PID: <http://hdl.handle.net/10932/00-01B8-B174-96E3-9C01-0>
Publisher: Institut für Deutsche Sprache

isPartOf : Mannheimer Korpus Historischer Zeitschriften und Zeitungen
hasPart : Mainzer Journal, Nr. 8, 22-06-1848
hasPart : Mainzer Journal, Nr. 10, 25-06-1848
hasPart : Mainzer Journal, Nr. 13, 28-06-1848
hasPart : Mainzer Journal, Nr. 15, 30-06-1848
hasPart : Mainzer Journal, Nr. 16, 01-07-1848
hasPart : Mainzer Journal, Nr. 19, 04-07-1848
hasPart : Mainzer Journal, Nr. 22, 07-07-1848
hasPart : Mainzer Journal, Nr. 26, 11-07-1848
hasPart : Mainzer Journal, Nr. 29, 14-07-1848
hasPart : Mainzer Journal, Nr. 30, 15-07-1848
hasPart : Mainzer Journal, Nr. 34, 19-07-1848
hasPart : Mainzer Journal, Nr. 37, 22-07-1848
hasPart : Mainzer Journal, Nr. 39, 24-07-1848
hasPart : Mainzer Journal, Nr. 42, 27-07-1848
hasPart : Mainzer Journal, Nr. 43, 28-07-1848
hasPart : Mainzer Journal, Nr. 45, 30-07-1848
hasPart : Mainzer Journal, Nr. 46, 31-07-1848
hasPart : Mainzer Journal, Nr. 47, 01-08-1848
hasPart : Mainzer Journal, Nr. 48, 02-08-1848
hasPart : Mainzer Journal, Nr. 49, 03-08-1848

Mainzer Journal, Nr. 10

Type: text
PID: <http://hdl.handle.net/10932/00-01B8-B176-54E3-A701-7>
Publisher: Institut für Deutsche Sprache

isPartOf : Mainzer Journal
source : HTML
source : TEI
source : TUSTEP
source : TUSTEP
source : MainzerJournal/MainzerJournal(1)0007.jpg
source : MainzerJournal/MainzerJournal(1)0008.jpg
source : MainzerJournal/MainzerJournal(1)0009.jpg
source : MainzerJournal/MainzerJournal(1)0010.jpg
source : MainzerJournal/MainzerJournal(1)0011.jpg
source : MainzerJournal/MainzerJournal(1)0012.jpg

Metadata Information

Creator: generateCMDI.pl
Createon date: 2013-06-23
Profile: clarin.eu:cr1:p_1366895758244
Collection: Institut für Deutsche Sprache, CLARIN-D Zentrum, Mannheim

Titel: Mainzer Journal.
Ort: Mainz
Datum: 25-06-1848
Nummer: 10
Seite: 7 [Facsimile](#) [TEI](#) [Tustep](#)

[1] Mainzer Journal.

[2] [Image](#)
[3] Das Mainzer Journal erscheint täglich (mit Ausnahme der höchsten Festtage) und zwar so, daß das Hauptblatt mit den „Rheinischen Unterhaltungs=blättern“ schon am Vorabende, die ständige Beilage am Vormittage des betreffenden Tages selbst ausgegeben wird. Bestellungen nehmen alle Postämter an; für Mainz und die nächste Umgebung die Buchhandlung von Kirchheim, Schott und Thielmann am Leichhofe. Der Preis des Blattes ist hier in Mainz jährlich 8 fl. in vierteljährlichen Vorausbezahlungen von 2 fl.; in dem gesamten Gebiete des Fürstlich Thurn- und Taxisschen Postbezirkes jährlich ebenfalls 8 fl. Inserate aller Art werden aufgenommen und die dreispaltige Petitzeile oder deren Raum mit 3 kr. berechnet.

[8] [Image](#)

[9] N° 10.

[10] [Image](#)

Sonntag, den 25. Juni.

1848

[11] Die absolute Demokratie und die Geschichte.
[12] J Mainz 23. Juni. Das, worauf eine durch Zahl ansehnliche und durch ihre Talente und Ihre Energie kraftige Partei lossteuert, ist die absolute Demokratie und eine einzige und untheilbare Republik für ganz Deutschland. Hierin allein, glauben jene Männer, könne die Freiheit vollständig verwirklicht und die Größe Deutschlands nach Außen begründet werden. Was das letztere betrifft, so würde in der That vielleicht Deutschland in der gewaltigen Auffregung, in welcher es dadurch sich versetzt fände, eine ähnliche Kraft nach Außen entwickeln, wie das revolutionäre Frankreich am Ende des vorigen und am Anfang dieses Jahrhunderts. Was aber das Ende der eingenommenen demokratischen deutschen Republik seyn würde, darüber mag die Geschichte uns belehren. Bei den Männern freilich, die man die Männer des „binden“ Fortschritts nennen könnte, weil sie gleichsam mit verbundenen Augen voranstürmen, ohne rück-

(...) alle Gewalt in der Weise in den Händen des Volkes liegt, daß dasselbe rein nach der Kopfzahl alles entscheidet und regiert oder durch eine Versammlung, die es durch eine ebenso nach der Kopfzahl berechnete Majorität gewahrt hat, beschließen und regieren läßt, so zwar, daß die von der Majorität gefassten Beschlüsse das einzige und höchste Gesetz sind. Dieses ist jener majestatische, unumschränkte, allmächtige Volkswille, der Wille der 40 Millionen, von denen Hecker in seinem Manifest an das Parlament mit einer Art abgöttischer Schwärmerie spricht. Eine solche absolute Demokratie aber, sagen wir, hat nie auf die Dauer bestanden. In allen antiken Republiken war in einem Senat, mochte er heißen wie immer, in allen adeligen Geschlechtern ein höchst zähes, tief in die Geschichte und das Leben verwachsenes conservatives und aristokratisches Element vorhanden, das dem leicht beweglichen und fluthenden Willen der Menge gegenüber die Wage hielt. Außerdem waren die Magistrate, die Consuln, Prätoren des alten Roms nichts weniger als bloße Beamten, son-

Some Open Issues with CMIDI

- Representation of Relationships
 - Beyond part-of (e.g. versioning)
 - Ranking in VLO
- Unnecessary Heterogeneity within IDS
 - Written vs. Spoken Resources
- Complex Framework for External Dataproviders
 - Suggest Profiles (DC-oriented)
 - Experiment with low tech SIPs (e.g. metadata as csv)

Ongoing Work

- Stabilization of Ingest Workflows & Repository
 - Documentation
 - Versioning
 - Integrity Checks
 - Reconsideration of Authentication Policies
- Homogenization of Metadata
 - VLO Taskforce (Clarin-D)
 - Revision of CMDI-Profiles
- Assistance for external Dataproviders
 - Documents, Guidelines
 - Recommended Profiles/Formats