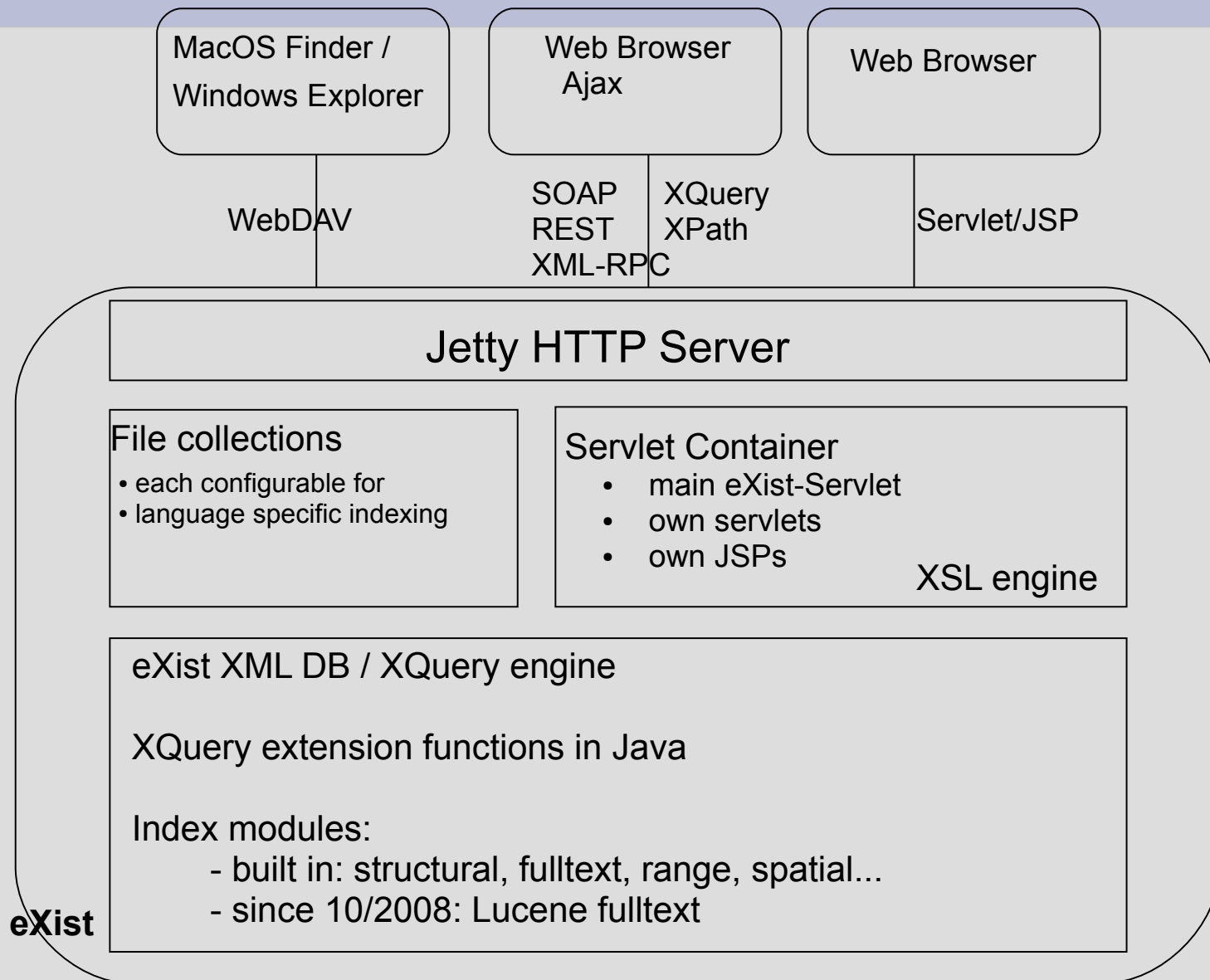


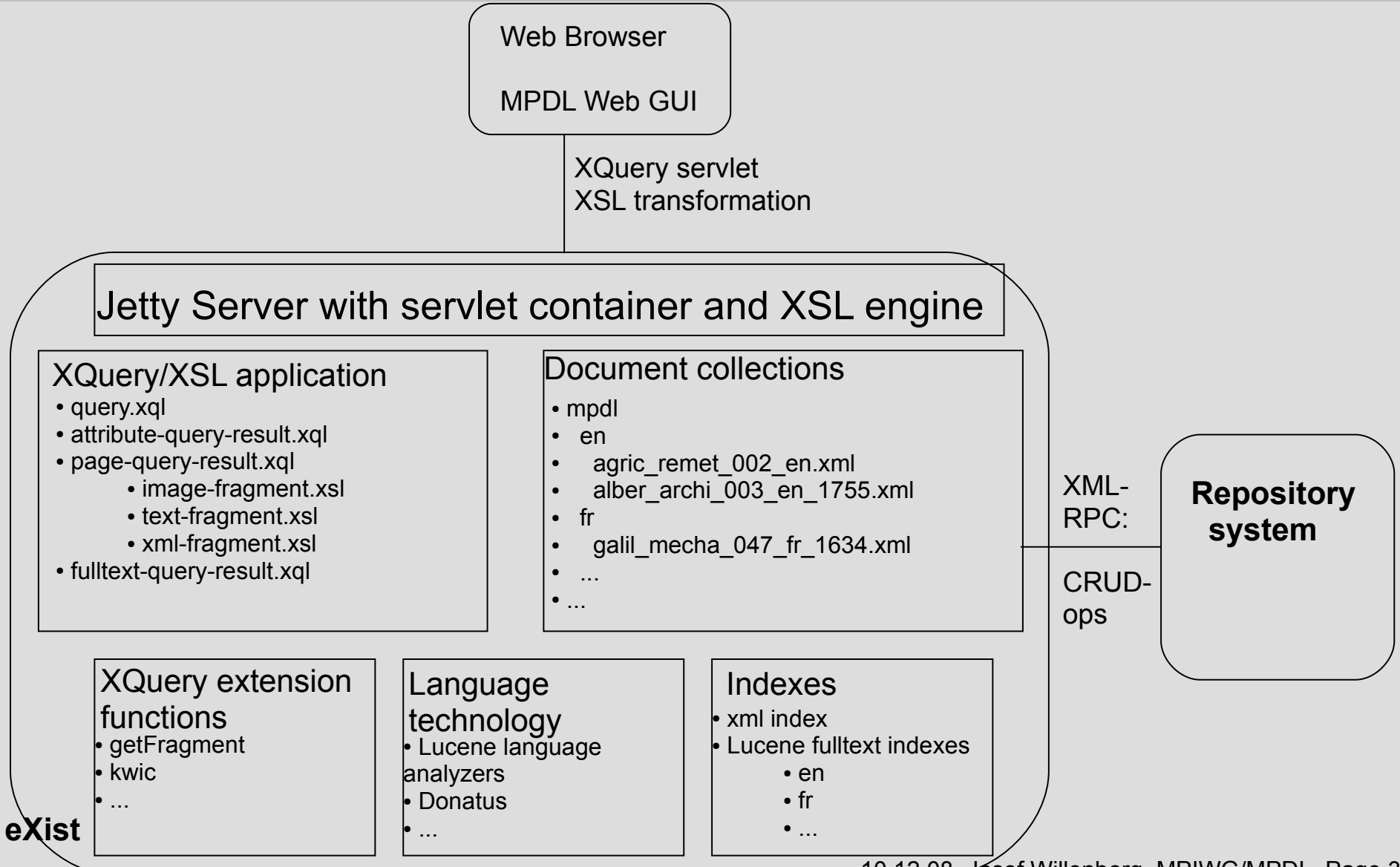
# eXist with Lucene

First developments and experiments

# eXist with Lucene architecture



# eXist in MPDL



# eXist

## heavy bugs

- after mistakenly trying an own developed trigger in a test collection
  - indexes (especially Lucene indexes) were corrupt
  - data was lost
  - huge log files are generated
- consequences / demands
  - be even more cautious with open software especially between „stable“ releases (done)
  - backup all downloaded and used software (done)
  - store all own applications in version management system (to be done)
  - backup all data (done)
  - learn and use the restore process (done)
  - separate development and test server (done)

# eXist development with Eclipse project definition

- Import eXist with Lucene as project (existDev) with sources (own .classpath and .project, fetch missing libraries, ...)
- Build project
- Import Lucene and Lucene contrib sources into project
- Configure Jetty server, conf.xml, ...
- Develop own eXist extensions
  - e.g.: mpdl/extensions/modules/src
- Start eXist in project (debug mode)

# eXist with Lucene

## first experiments

- Lucene language analyzers in archimedes test collection
  - language specific subcollections of documents (with its own configuration)
  - used: en, de, fr, el, zh, nl
  - not yet used: Thai, CJK, Brazilian, cz, ru
- Java sources are available
- only simple analyzers
- de analyzer
  - query with \* and special characters (such as ö, ü, ä, ...): incorrect results
  - only simple implementation
- ...

# MPDL development

## first steps (1)

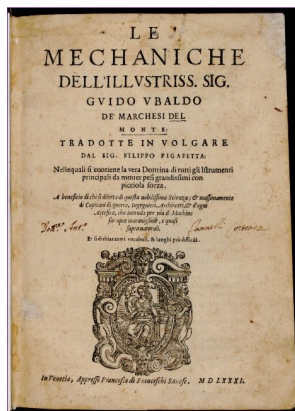
- first web GUI for archimedes test (mpdl-test:8080/archimedes/query.xql)
  - attribute query (done)
  - retrieve fragment between page breaks (done)
  - page presentations: image, text, xml (done)
  - fulltext queries: with and without language analyzers (done)
  - query collection specific (done)
    - e.g.: collection('/db/mpdl/de', '/db/mpdl/nl')
  - first integration into digilib (done)
- XQuery modules: own Java extensions
  - find words in sentences (first steps done)
  - get fragment between page breaks in big documents (to be done)
- Lucene: own language analyzers
  - integration into eXist (done)
  - improvement of the predefined analyzers (to be done: later)
  - design and implementation of analyzers for project important languages such as latin (to be done: later)

# Page presentations image, text, xml

Monte, Guidobaldo del. *Le Mechaniche*. Venezia, 1581

[Text] [Image] [XML]

◀ 1 / 270 ▶ Page: 1



Elapsed time: 10 ms. Back to [query page](#), see the [XQuery source](#) and [XSL source](#) of this page

Monte, Guidobaldo del. *Le Mechaniche*. Venezia, 1581

[Text] [Image] [XML]

◀ 1 / 270 ▶ Page: 1

LE  
MECHANICHE  
DELL'ILVSTRISS SIG.  
GUIDO VBALDO  
DE' MARCHESI DEL  
MONTE:  
TRADOTTE IN VOLGARE  
DAL SIG. FILIPPO PIGAFETTA:  
Nellequali fi contiene la vera Dottrina di tutti gli Ifrumenti  
principali da mouer peſi grandisſimi con  
picciola forza.  
A beneficio di chi fi diletta di queſta nobiliſſima ſcienza; & maffimamente  
di Capitani di guerra, Ingegneri, Architetti, & d'ogni  
Artefice, che intenda per via di Machine  
far opre marauigliofe, e quaſi  
fopra naturali.  
Et fi dichiarano i vocaboli, & luoghi più difficili.



Figure: 1

In Venetia, Appreſſo Franceſco di Franceſchi Sanefe. MD LXXXI.

Elapsed time: 294 ms. Back to [query page](#), see the [XQuery source](#) and [XSL source](#) of this page

Monte, Guidobaldo del. *Le Mechaniche*. Venezia, 1581

[Text] [Image] [XML]

◀ 1 / 270 ▶ Page: 1

```
<text id=id.0.0.0.3> <pb id=p.0001 xlink:href=037/01.001.jpg> </pb>
<front id=id.1.0.0.0> <section> <p id=id.2.1.2.0.0 type=head> <s id=id.2.1.2.1.0>LE </s> </p>
MECHANICHE </p> </p>
DELL'ILVSTRISS SIG. </p> </p>
GUIDO VBALDO </p> </p>
DE MARCHESI DEL </p> </p>
MONTE: </p> </p>
</s>
<s id=id.2.1.3.1.0>TRADOTTE IN VOLGARE </s> </p>
DAL SIG. FILIPPO PIGAFETTA: </p> </p>
Nellequali fi contiene la vera Dottrina di tutti gli Ifrumenti </p> </p>
principali da mouer peſi grandisſimi con </p> </p>
picciola forza. </p> </p>
</s>
<s id=id.2.1.4.1.0> <emph type=italics> </emph>
A beneficio di chi fi diletta di queſta nobiliſſima ſcienza; & maffimamente </p> </p>
di Capitani di guerra, Ingegneri, Architetti, & d'ogni </p> </p>
Artefice, che intenda per via di Machine </p> </p>
far opre marauigliofe, e quaſi </p> </p>
fopra naturali. <emph.end type=italics> </emph.end>
</p> </p>
</s>
<s id=id.2.1.5.1.0>Et fi dichiarano i vocaboli, & luoghi più difficili. </p> </p>
</s>
<figure id=id.037.01.001.1.jpg xlink:href=037/01.001/1.jpg> </figure>
<s id=id.2.1.7.1.0> <emph type=italics> </emph>
In Venetia, Appreſſo Franceſco di Franceſchi Sanefe. </s>
<s id=id.2.1.7.2.0>MD LXXXI. <emph.end type=italics> </emph.end>
</p>
</p>
</section>
</front>
</text>
```

Elapsed time: 292 ms. Back to [query page](#), see the [XQuery source](#) and [XSL source](#) of this page



# MPDL development first steps (2)

- Interface to repository
  - as an open web interface
  - create document (started ...)
  - provide further operations (to be done early)
- technical specification: first draft (to be done)
- Separation of eXist instances (done)
  - „MPDL development“ and „MPDL test“ on one computer
  - ant scripts for deploying application from development to test server
- Move eXist instance „MPDL test“ to server: xserve08 (to be done)
- ...

# Development environment

