

# MathArc:

Aufbau eines  
verteilten digitalen Langzeitarchivs

Olaf Brandt – Markus Enders



## Projektziel:

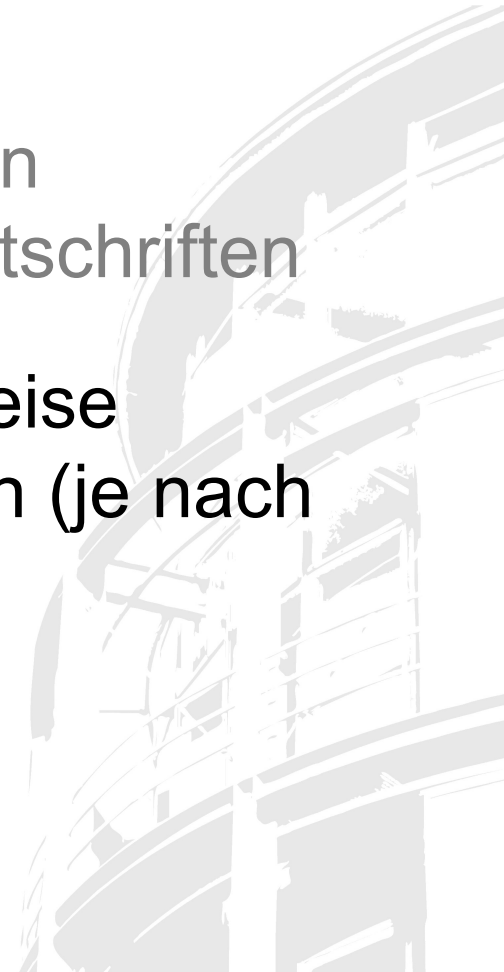
Aufbau einer Infrastruktur zur verteilten  
Archivierung von mathematischen Zeitschriften

elektronische Zeitschriften  
retrodigitalisierte Zeitschriften

## Projektziel:

Aufbau einer Infrastruktur zur verteilten  
Archivierung von mathematischen Zeitschriften

Verteilt : Daten sollen ganz oder teilweise  
redundant im System gehalten werden (je nach  
Policy)



## Projektziel:

Aufbau einer Infrastruktur zur verteilten  
Archivierung von mathematischen Zeitschriften

### Ziel:

- durch Redundanz die Vorteile unterschiedlicher Preservation-Systeme nutzen
- Gefahr des Verlustes / nicht Zugreifbarkeit auch über längeren Zeitraum hinweg verringern.

## Projektziel:

Aufbau einer Infrastruktur zur verteilten  
Archivierung von mathematischen Zeitschriften

## Finanziert über:

- NSF – für Cornell University Library
- DFG – für SUB Göttingen

derzeit "nur" zwei Partner für Entwicklung und  
prototypische Implementierung

EMANI:

electronic mathematical archive network  
initiative

Cornell University Library, USA

MathDoc, Frankreich

SUB Göttingen, Deutschland

Tsinghua University Library, China



EMANI:

electronic mathematical archive network  
initiative

Ziel:

"...to store the digital content in mathematics from the content providers at the reference libraries..."

"...to retrodigitise printed publications in mathematics from the content providers at the reference libraries..."

"...to care about the long-term preservation of this content in readable form..."

## Aufbau des MathArc-Systems:

### Asset als funktionale Einheit in der Archivierung

muss bestimmte Strukturbeschreibung und Metadata enthalten

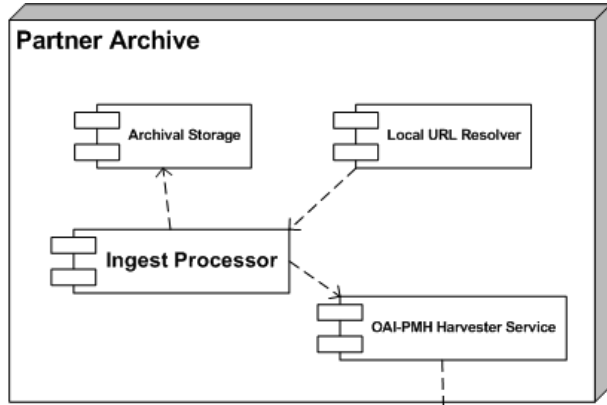
bspw. Zeitschriftenband, Artikel, Webpage etc...

zentrales Registry enthält bestimmte Policy-Regeln zum Steuern der Replizierung

Austausch der Assets erfolgt mittels des OAI-PMH

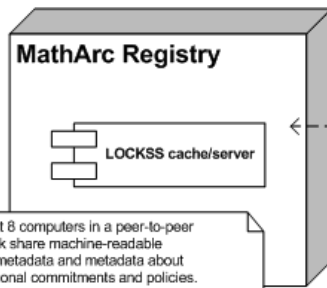
dank des XML-Formats können darüber neben Metadaten auch Contentdaten (METS-basiert) ausgetauscht werden.

# Aufbau des MathArc-Systems

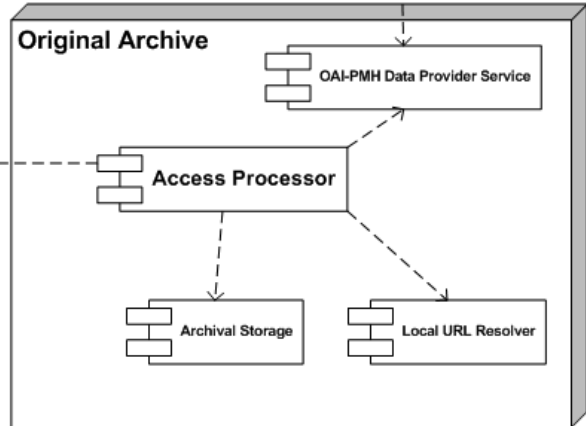


The Partner Archive's (PA) Ingest Processor uses its OAI-PMH harvester to communicate with the Original Archive's (OA) OAI-PMH Service Provider. The Ingest Processor requests asset-description metadata from the OA. The OA checks rights and commitment metadata in the MathArc Registry, and returns a list of appropriate assets. The PA requests individual files from the OA. After performing quality assurance checks, the Ingest Processor stores the Asset in the PA storage, mapping local addresses to persistent identifiers in the Local URL Resolver.

When an Original Archive requests its own Assets from a Partner Archive, the roles are reversed. (Each archive can perform all request and reply functions and can create Dissemination Information Packages and Submission Information Packages from its stored Archival Information Packages.)



At least 8 computers in a peer-to-peer network share machine-readable rights metadata and metadata about institutional commitments and policies. The LOCKSS software provides a self-repairing mechanism, insuring against data corruption. Because the network is distributed among many machines, each replicating the data on the others, the network is owned by all participants.



## Aufbau des MathArc-Systems:

mögl. Befehle zum Abgleich des Archivs:

ingest

update im Fall einer Migration

löschen

Administrative Informationen  
Ergebnis von Konsistenzprüfungen etc.



## Standards:

für internationale Zusammenarbeit ist die  
Nutzung internationaler Standards wichtig:

**METS – als Containerformat für Assets**

<http://www.loc.gov/mets/>

**DublinCore für deskriptive Metadaten**

<http://www.dublincore.org>

**PREMIS für preservation Metadata**

<http://www.loc.gov/premis/>

Standards:

Premis:

Metadatenchema für Preservation Metadata  
Erarbeitet von internationaler Community  
aus unterschiedlichen institutionellen  
Bereichen

Premis-Event-Metadaten veranlassen  
Aktionen bei den Partnerarchiven und  
steuern so das verteilte Archiv

# Integration in lokale Preservation-Systeme:

SUB Göttingen:

DIAS-System als Preservation System

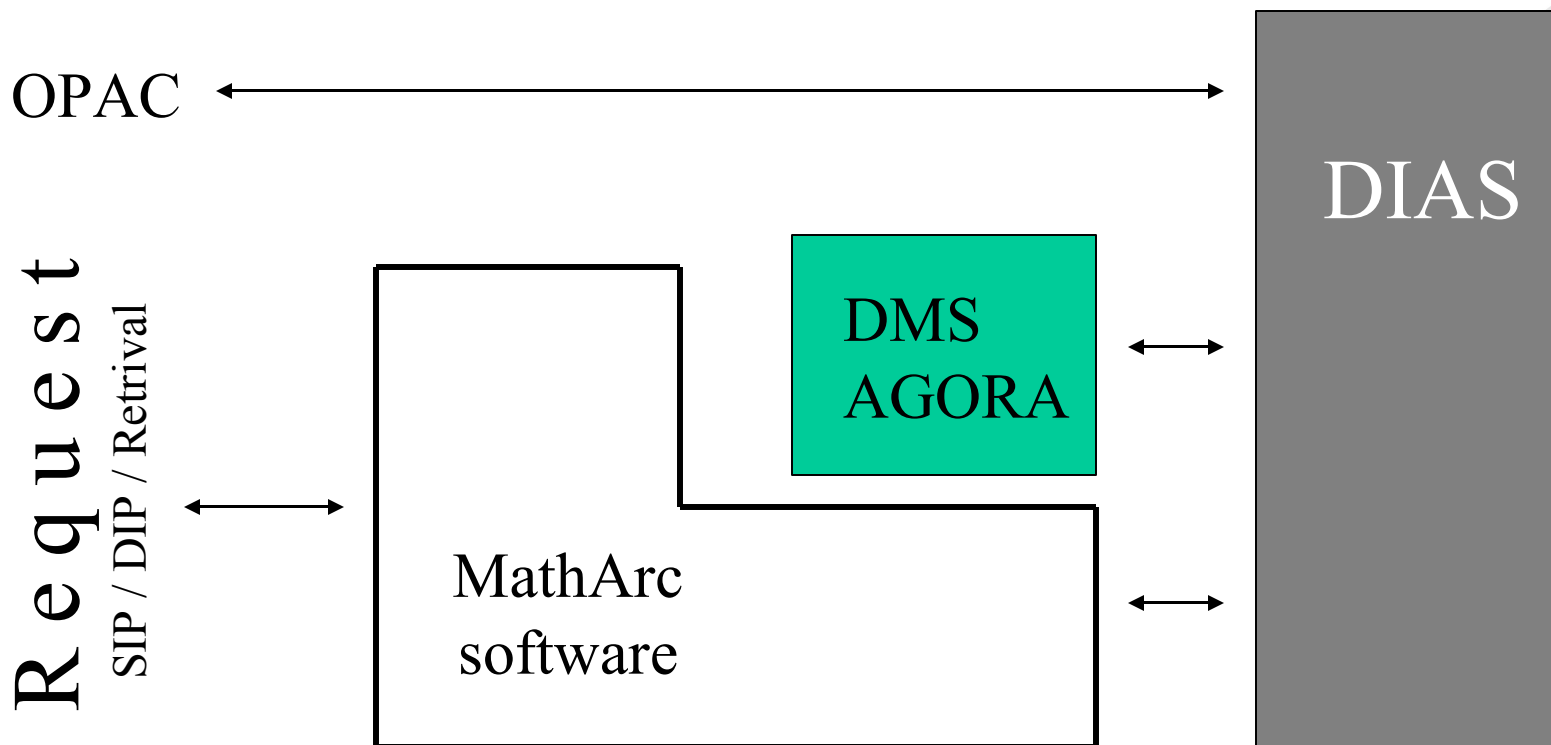
Entwicklung geeigneter Schnittstellen

(neben SIP/DIP auch administrative Schnittstellen notwendig)

und Konvertierungstools

Preservationmetadaten: LMER <- -> Premis

## Integration in lokale Preservation-Systeme:



Stand:

Metadaten-Schema fertiggestellt

Beispieldateien fertig gestellt

Szenario-Analyse beendet;

Mapping LMER <- -> Premis fertiggestellt

to do:

## Implementierung des OAI-Protokolls

inkl. Auswertung der Premis-Events  
und Verbindung zum Preservation-System

Mapping LMER <- -> Premis  
implementieren (mittels XSLT)

Query-Mediator zum zentralen Zugriff auf  
verteilten Archiv implementieren

kein Endnutzer-Zugang geplant; MathArc-System ist als reines  
Back-Office System konzipiert.

Projekt-Homepage:

<http://www.library.cornell.edu/dlit/MathArc/web/index.html>

Vielen Dank!

