

Digitalisierung im Rahmen des E-Archiving-Projektes

Workshop in Bern, 18. Mai 2006

Dr. Matthias Töwe
Konsortium der Schweizer Hochschulbibliotheken

Gliederung

- Digitalisierung als Teil des Gesamtprojekts
- Digitalisierung: Pilot „SwissDML“
 - Ziele
 - Partner
 - Vorgehen
 - Exkurs: Automatische Metadaten-Extraktion
 - Standards
- Ergebnis
- Fortsetzung

Strategische Überlegungen und Ziele¹

- Zunehmende Entkopplung von Informationsversorgung und Bestandaufbau als Risiko gewertet
 - Nachhaltigkeit der Investitionen in Lizenzen sichern
- Zugänglichkeit zu vorhandenen Informationen verbessern durch gezielte Digitalisierung und Förderung von Dokumentenservern unter Open Access
 - Für neue Projekte im digitalen Raum von vorneherein dauerhafte Zugänglichkeit sichern
- Durch Koordination Aufwand verringern und Positionierung der Bibliotheken insgesamt verbessern

¹ Vgl. „Konzeptstudie E-Archiving“:
http://lib.consortium.ch/external_files/Konzeptstudie_D_V1_2.pdf

Zugänglichkeit vs. Langzeitarchivierung

- Für Hochschulbibliotheken stehen Zugänglichkeit und Nutzbarkeit der Information im Vordergrund.
→ Bedarf und Entscheidungen lokal unterschiedlich
- Archivierung als notwendige, aber nicht hinreichende Hintergrundfunktion, die weit über die Speicherung der Originaldaten hinausgehen muss.
→ Technisch-organisatorische Fragen, die zentral bearbeitet werden können
- Gefahr, dass technische Fragen dominieren, obwohl organisatorische Herausforderungen und langfristiges Engagement der Institutionen ausschlaggebend sind

Teilprojekte / Themenbereiche



Ein Treppelichter — bekanntlich ist diese
nach in den Kuppeln der großen Väter-
schmuck ausgeblendet. Kommt eine der schweizer-
schweizeren, die sich in der Schweiz verhalten,
— hat, nicht nur aus dem Streben nach Ver-
vollkommenheit, sondern auch aus dem Streben nach
wunderbarer Gesundheit sein Leben davon abge-
geben, das er schenkte in jedem Augenblick ge-
schrieben, Tag und Nacht auf dem Treppenlicht.
Aber nicht, wie man sich vorstellen könnte,
wurde durch einander abtunende Männer aus-
gesprochen, welche wissen und leben, was
Schweizerland für die Umwelt ergab sich aus
dieser Lebensweise nicht, nur während der
eigenen Programmstunden war es ein wenig
anders, daß er, wie sich unter anderem
oben gehalten war und das, was er sich
in seinen Tagen nicht ruhig verhalten, hat und
da ein Blick aus dem Publikum zu ihm abwärts.
Doch wieder ist dies die Dialektik, weil
er ein selbständiger, verantwortlicher Künstler
war. Auch ist nicht zufällig ein, daß er nicht
aus Mitleiden so leidet und eigentlich nur so
sich in diesem Oben schaute, nur so seine
Küsse in diese Vollkommenheit bezaubert können.
Doch war es oben auch seine Grenze und
wenn in der warmen Jahreszeit in der ganzen
Runde der Wirkung die Schönezeit aufsteigen
wird und mit der höchsten Luft die Sonne
schwebt in der oberen Kuppel, dann
dann war es dort sogar schön. Freilich, sein
menschenlicher Verstand war eingeschränkt, ein
menschliches Moment auf der Treppelichter ein Treppel-
lichter zu sein, denn, wenn er in der
dem Treppel, ist es nicht und nicht aus dem
Hintergrund und phantastisch, aber es verlor
Bewusstsein die Dicht- und wachsende einige Worte
mit ihm durch ein offenes Fenster, oder es über-
griff die Feuerherde, die Nachdenklichkeit
auf der oberen Gasse und der Treppel-
Respektvoll, aber wenig Verantwortlich zu. Seine
bleibt es um ihn still, nachdenklich ist er

E-Journals :
Zugänglichkeit
hat Priorität.
Zusätzlich
Einbezug der
gedruckten
Ausgaben
notwendig.

Publikationen der Schweizer Hochschulen

Fortschreibung der
traditionellen
Hochschulschriften
in elektronischer
Form:
Dokumentenserver

Grosses Interesse
für neue
Betätigungsfelder im
Zeichen von *Open*
Access

Noch nicht
digitalisierte
Zeitschriften:
Potential
und Bedarf
vorhanden.

Langzeitarchivierung

Themenbereiche

1. Zugangssicherung zu Online-Zeitschriften.
2. Dokumentenserver und Open Access.
3. Langzeitarchivierung.
4. Retrodigitalisierung von Zeitschriften.
5. Verteilte Archivierung gedruckter Zeitschriften.

Zusammenhang der Teilprojekte

<http://www.seals.ch>

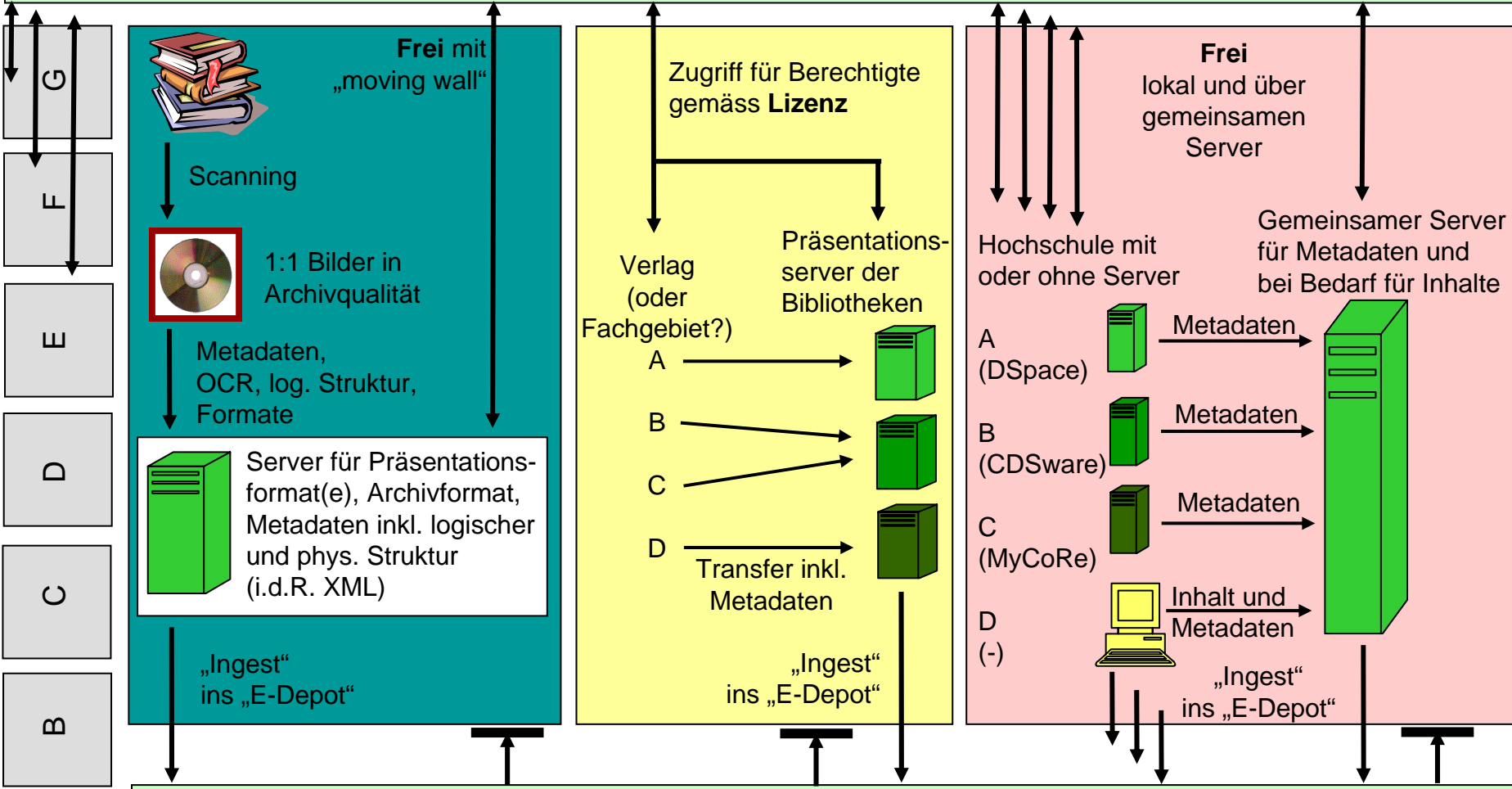
Gedruckte Titel: OPAC-Zugriff und SZP

Digitalisate: Zugriff aus externen bibliographischen DB, aus OPAC (Titalebene) und mit eigener Volltext-, Feld- und Indexsuche (<http://retro.seals.ch>)

Lizenzierte Inhalte: Zugriff aus OPAC (Titalebene), aus lokaler E-Journal-Liste, aus externen bibliographischen DB und über separate Oberfläche

Graue Literatur: Zugriff aus OPAC und über separate Oberfläche (Indices nach Institutionen, Fächern, Dokumenttyp) mit Volltext-, Feld- und Indexsuche

Archivbestände gedruckter Zeitschriften



„E-Depot“ aus verteilten koordinierten Archivservern. Beispielkonzept: Projekt KOPAL der DDB, wo DIAS (IBM, KB Den Haag) für verteilten Einsatz angepasst werden soll. Aufteilung inhaltlich, nach Verlagen oder anders? Es ist denkbar, dass mehrere der grün markierten Serverfunktionen und Archivierungsaufgaben von *einem* System wahrgenommen werden.

Hintergrund der Zeitschriftendigitalisierung

- Wissenschaftliche CH-Zeitschriften
- Sie erscheinen
 - als Eigenprodukte kommerzieller Verlage
 - Digitalisierung durch Verlag schon erfolgt oder wahrscheinlich, Sichtbarkeit generell hoch
 - im Auftrag von nicht-gewinnorientierten wissenschaftlichen Fachgesellschaften bei kommerziellen Verlagen
 - Von Digitalisierungsprogrammen der Verlage z. T. ausgenommen, Sichtbarkeit aber generell hoch
 - in eigener Regie der Gesellschaften, z. T. im Hochschulumfeld
 - Retrodigitalisierung aus eigener Kraft fraglich, teilweise auch keine aktuelle Online-Produktion vorgesehen: Eingeschränkte Sichtbarkeit und Zugänglichkeit

Kriterien für die Auswahl der Zeitschriften

- Wissenschaftlicher Charakter
 - peer review nicht zwingend, aber mit hohem Gewicht
- Relevanz älterer Inhalte ist fachabhängig
 - Interesse bei Nutzerschaft, Bibliotheken, Herausgeberschaft
- „Internet-Affinität“ der Leserschaft ist fachabhängig
- Erscheinen über längeren Zeitraum hinweg (> 25 Jahre; bei Vorgänger- und Nachfolgetiteln unübersichtlich)
- Bezug zur Schweiz
 - herausgebende Institution, Wirkungsort der Autoren, Sprachen
- Nicht-gewinnorientierter Charakter
 - nur scheinbar klar: Fachgesellschaften subventionieren z. T. ihre Publikationen, z. T. fließen ihnen aber auch Mittel daraus zu

Pilotprojekt SwissDML

- SwissDML
 - Anlehnung an DML / WDML: (World) Digital Mathematics Library
- Anstoss durch Schweizerische Mathematische Gesellschaft (SMG)
- Nutzung auch älterer Inhalte
- Mathematik als Disziplin mit langer „Online-Tradition“
- Nennenswerter Umfang, aber überschaubar (ca. 90'000 Seiten)
- Erscheinungsbeginn der drei Zeitschriften: 1899, 1929, 1946
- Bezug zur Schweiz
 - Herausgeber und Verlag in der Schweiz, mehrsprachig
- Nicht-gewinnorientierter Charakter
 - Trägerschaft und wirtschaftliches Risiko bei SMG bzw. Fondation L'Enseignement Mathématique

Partner

- Anstoss durch Schweizerische Mathematische Gesellschaft (SMG)
- Fondation L'Enseignement Mathématique
- European Mathematical Society Publishing House als neuer Verlag für zwei der Zeitschriften
- Niedersächsische Staats- und Universitätsbibliothek, Göttingen
- EMANI-Projekt (Electronic mathematical archiving network initiative)
- Kommerzielle Dienstleister für Scanning und Metadatenextraktion
- Softwarelieferant

Vorgehen

- Schriftliches Memorandum of Understanding aller Partner (SMG bzw. Fondation, aktueller Verlag, Konsortium...)
- Analyse des gedruckten Bestandes:
 - Charakter, Schriften, Grafiken, Abbildungen, Formate, Sonderfälle
 - Vollständigkeit des Bestandes
 - Aufschneiden der Bände möglich für Durchlaufscanner?
- Diskussion relevanter Standards mit Anbieterfirmen
- Tests von Scans externer Firmen und Auftragsvergabe
 - Grosse Bandbreite bei den Kosten und bei der Qualität
 - Entscheidung gegen Hybridlösung mit Mikrofilmerstellung
- Anforderungen an Erschliessung, Suche, Präsentation
 - Qualifizierte Metadaten, Suche im Volltext, einfacher PDF-Download
- Auswahl Software und Metadatenerfassung

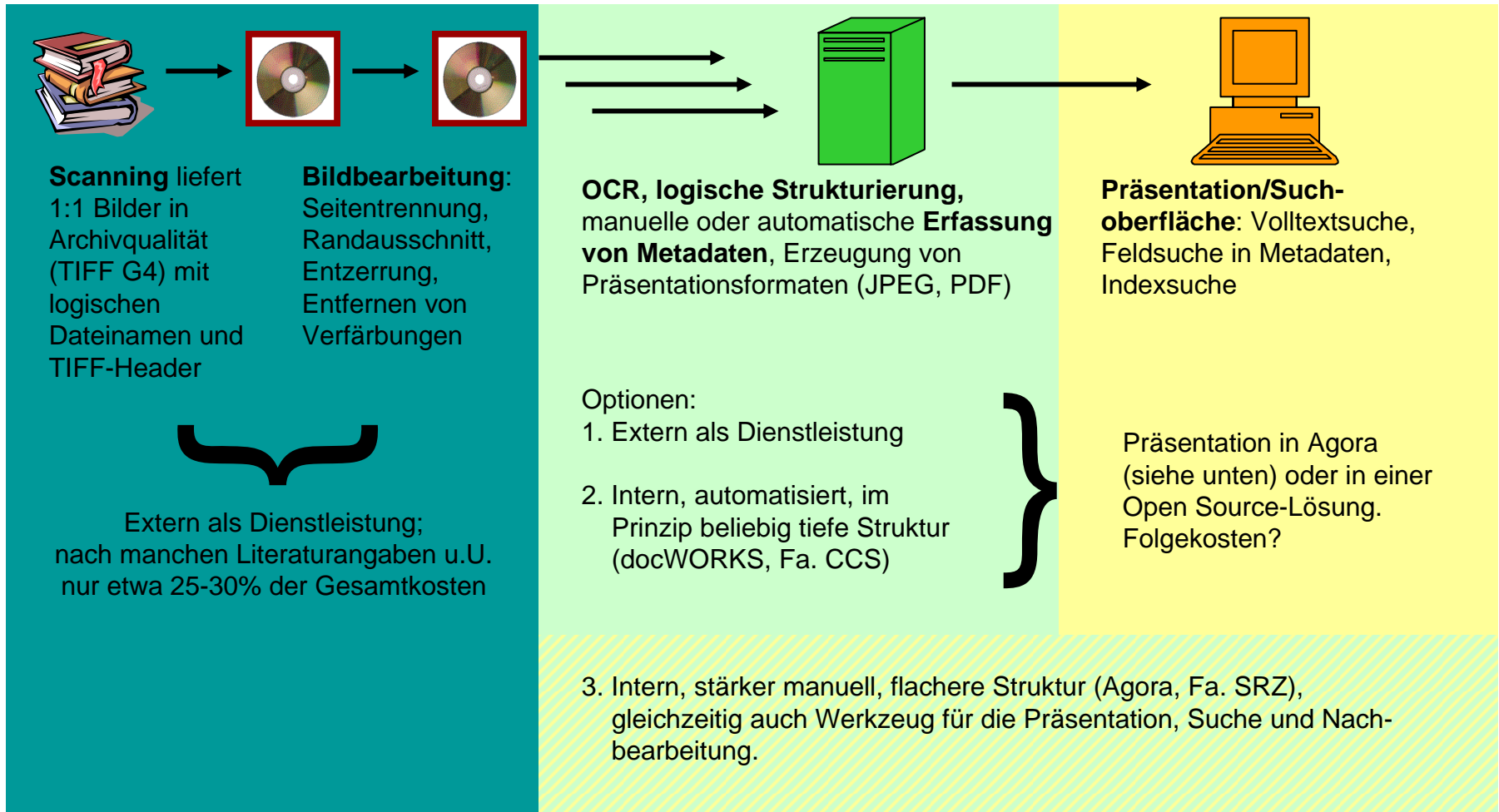
Exkurs: Automatische Metadatenextraktion

- Software docWORKS (aus dem EU-Projekt METAe, jetzt als Produkt von Firma CCS)
- Auf Basis des Volltextes (Optical Character Recognition – OCR) weitgehend automatische Erfassung und Wiedergabe der logischen Dokumentstruktur sowie aller relevanten Metadaten in XML-Dateien
- Software-Installation vs. Dienstleistung durch Anbieter
- Kosten erheblich und mit geringem Skaleneffekt
- Resultat grundsätzlich überzeugend, aber nicht voll transparent
- Weiterverarbeitung auch für Experten nicht einfach, weil eigenwilliges Format – aber als XML grundsätzlich analysierbar und verwendbar
- Lohnt sich für tiefgehende Strukturierung und Erschliessung

Teilprojekt Digitalisierung: Ablauf und Optionen

Ziel: Gedruckte Bestände in digitaler Form für eine besser Zugänglichkeit online bereitzustellen. Die Digitalisierung ist aber hier selbst keine Erhaltungsmassnahme für die *gedruckten* Bestände.

03.04.2006, M. Töwe



Ablauf

L'ENSEIGNEMENT MATHÉMATIQUE

REVUE INTERNATIONALE

Organe officiel
de la Commission internationale de l'Enseignement Mathématique

Fondée en 1899

par

H. FEHR et C.-A. LAISANT

II^e SÉRIE

TOME V

Scanned Image

GENÈVE
IMPRIMERIE KUNDIG
1960

OCR



L'ENSEIGNEMENT

MATHÉMATIQUE

REVUE INTERNATIONALE

Organe officiel

de la Commission internationale de
l'Enseignement Mathématique

Fondée en 1899

par H. FEHR et C.-A. LAISANT

II^e SÉRIE

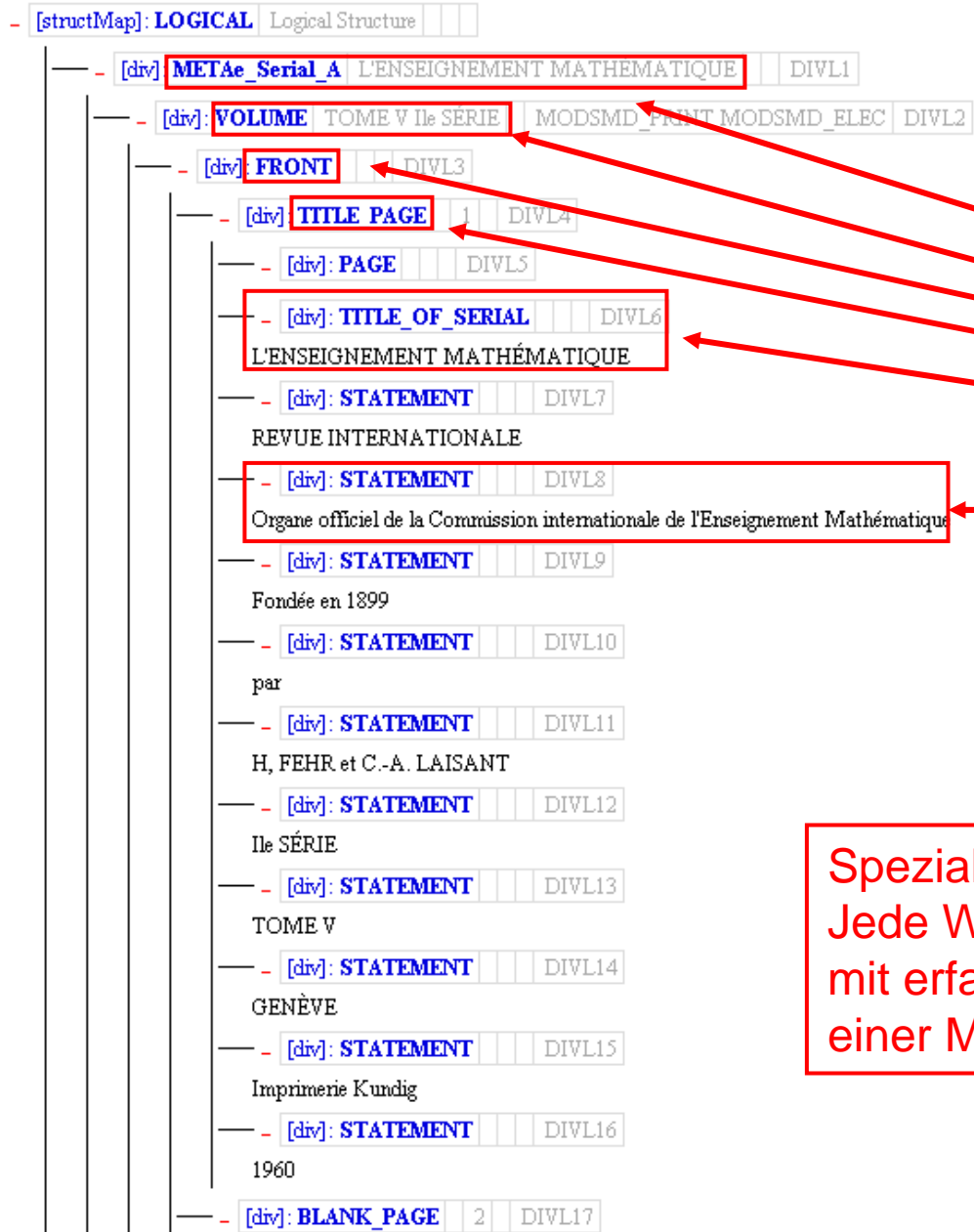
TOME V

Erkannter Text,
Maschinenlesbar.

GENÈVE

IMPRIMERIE KUNDIG

1960

Logische Struktur: Bandebene

•Logisch-hierarchische Struktur

- Zeitschrift
- Band
- Frontmatter
- Titelblatt mit erkannten Bestandteilen („Title of Serial“) und Teilen, die ohne Anpassung nicht als Strukturelement eines bestimmten Typs erkannt werden können („Statement“).

Spezialität von docWORKS:
 Jede Wortposition ist im Hintergrund mit erfasst und kann später zur Anzeige einer Markierung genutzt werden.

```

- [div]: CONTRIBUTION SUR L'EXEMPLE, DONNÉ PAR M. DE RHAM, D'UNE FONCTION CONTI
- [div]: FRONT | | DIVL1545
- [div]: TITLE_OF_CONTRIBUTION | | DIVL1546
  SUR L'EXEMPLE, DONNÉ PAR M. DE RHAM, D'UNE FONCTION CONTINUE SANS DÉRIVÉE
- [div]: AUTHOR_OF_CONTRIBUTION | | DIVL1547
  Jean-Pierre Kahane (Montpellier)
- [div]: MOTTO | | DIVL1548
- [div]: TEXT | | DIVL1549
  ? PAR (Reçu le 16 juin 1958)
- [div]: MAIN | | DIVL1550
- [div]: BODY | | DIVL1551
- [div]: BODY_CONTENT | | DIVL1552
- [div]: PARAGRAPH | 1 | DIVL1553
  - [div]: TEXT | | DIVL1554
    Un exemple très simple de fonction continue sans dérivée a été donné par M. G. de Rham.
- [div]: FORMULA | 2 | DIVL1555
- [div]: IMAGE | | DIVL1556
- [div]: PARAGRAPH | 3 | DIVL1557
  - [div]: TEXT | | DIVL1558
    (i)
- [div]: PARAGRAPH | 4 | DIVL1559
  - [div]: TEXT | | DIVL1560
    où
- [div]: FORMULA | 5 | DIVL1561
- [div]: IMAGE | | DIVL1562
- [div]: PARAGRAPH | 6 | DIVL1563
  - [div]: TEXT | | DIVL1564
    Cet exemple se prête à des remarques intéressantes. \En premier lieu, nous mettrons en é
    simple exercice, pouvant illustrer un cours d'analyse. En second lieu, nous verrons que l'ex
    fonction, stricte plus faible que la condition de Lipschütz, n'assure la déri fut-ce en un seul
    continuité en chaque point est «aussi mau qu'on veut ».
- [div]: FOOTNOTE | | DIVL1565
  - [div]: TEXT | | DIVL1566
    i) C'est ainsi que M. de Vito a montré, dans le vol. IV, fasc. 3, que / appartient à la classe Lip a
    rapportée par M. de Vito, qu'on peut construire des fonctions continues sans dérivées, n'appart
  
```

•Logisch-hierarchische Struktur. Hier: Ebene des einzelnen Artikels:

- Artikeltitlel
- Autor
- Absatz
- Formel
- Fussnote
-

•Welche Bestandteile erfasst werden sollen und mit welcher Benennung, ist zu Beginn zu definieren.

•Formeln können von gängigen OCR-Programmen nicht erkannt werden. Sie werden als Bilder behandelt.

Daraus erstelltes Inhaltsverzeichnis

TOC for Contributions		
TOME V Ite SÉRIE (METAe_Serial_A)		
orderNo	contribution title	pageNo
-	[[ISSUE] -	
	FONCTIONS ENTIÈRES D'ORDRE FINI ET FONCTIONS MÉROMORPHES	1
	SUR LES COURBES LIMITES DE POLYGONES OBTENUS PAR TRISECTION	29
	ON THE PROJECTION OF A FOUR-POINT SYSTEM OF CONICS INTO A FAMILY OF CIRCLES	44
	SUR L'EXEMPLE, DONNÉ PAR M. DE RHAM, D'UNE FONCTION CONTINUE SANS DÉRIVÉE	53
	CALCUL D'UNE INTÉGRALE DOUBLE PAR DEUX INTÉGRATIONS SIMPLES SUCCESSIVES (INTÉGRALE DE RIEMANN)	58
	THE MATHEMATICS EDUCATION OF YOUTH A COMPARATIVE STUDY	61
-	[[ISSUE] -	
	SUR LES POLYGONES ET LES POLYÈDRES RÉGULIERS ENTIERS	81
	DIVERGENCE DE LA SÉRIE HARMONIQUE D'APRÈS MENGOLI (1650)	86
	SUR L'EXISTENCE D'UNE SPHÈRE PASSANT PAR UN NOMBRE DONNÉ DE POINTS AUX COORDONNÉES ENTIÈRES	89
	THE POSSIBLE TRANSFORMATIONS OF A REAL CURVE INTO A CURVE WITH REAL EQUATION AND PASSING THROUGH THE ISOTROPIC POINTS	91
	ON THE SOLUTION OF SIMULTANEOUS IMPLICIT EQUATIONS	107
	PROGRAMME DE MATHÉMATIQUES DE LA COMMISSION DES MATHÉMATIQUES DU CONSEIL DES EXAMENS D'ADMISSION DANS LES COLLÈGES (COLLEGE ENTRANCE EXAMINATION BOARD) 1)	115
	A COMPARATIVE STUDY OF METHODS OF INITIATION INTO GEOMETRY	119
	ANNEXE AU RAPPORT SUR L'ÉTUDE COMPARÉE DES MÉTHODES D'INITIATION A LA GÉOMÉTRIE	140
-	[[ISSUE] -	
	POINTS RATIONNELS SUR CERTAINES COURBES ET SURFACES CUBIQUES	
	A SPECIAL CASE OF KUMMER'S CONGRUENCE (mod 2e)	171
	DIE WISSENSCHAFTLICHEN GRUNDLAGEN DER SCHULMATHEMATIK	176
	DER MATHEMATISCHE UNTERRICHT DER 6- BIS 15-JÄHRIGEN JUGEND IN DER BUNDESREPUBLIK DEUTSCHLAND	183
	SCIENTIFIC FOUNDATIONS OF SCHOOL MATHEMATICS	196
	DES PRINCIPES DE L'ENSEIGNEMENT MATHÉMATIQUE	203
	[[ISSUE] -	
javascript:parent.frames["topFrame"].showPageTIF(this.frames["name"], '7');		

•Im Beispiel: Beiträge heftweise als Artikel erkannt

•Seitenzahlen physisch (d.h. inklusive Leerseiten etc.) und logisch korrekt zugeordnet:
Die Seite mit der Nummer 1 ist in Wirklichkeit das siebte Seitenbild.
Lücken in der Paginierung werden automatisch geschlossen, wenn dies widerspruchsfrei möglich ist

•Im Hintergrund logische Struktur als XML-Datei gemäss einem modifizierten METS-Standard

Verwendete Standards

Ziel: Attraktives Angebot und Vorbereitung der Langzeitarchivierung

- TIFF für gescannte Images
 - Text und Liniengrafiken bitonal (Tests nötig – Scan in Graustufen kann sich lohnen)
Fotos und Flächen-Grafiken in Graustufen, 8 bit
 - 600dpi horizontal und vertikal (für OCR mindestens 400dpi)
 - TIFF 6.0, single pages, Fax G4 Kompression (verlustfrei, aber man muss sich bewusst sein, dass der eigentliche „Master“ verloren geht)
- Online-Anzeige: GIF*
- Ausdruck: PDF 1.4 (Acrobat 5.x)*
- Metadaten (inkl. Volltext)
 - METS/XML
 - Temporär bei der Metadatenextraktion: METS-ALTO/XML

* *Nicht zur Archivierung bestimmt.*

Ergebnis

- Online-Angebot mit Volltextsuche unter <http://retro.seals.ch>.
→ Demonstration

Fortsetzung

- Letzten Titel des Pilotprojektes einbinden.
- Verknüpfung mit Einträgen in bibliographischen Fachdatenbanken (noch Teil des Pilotprojekts).
- Ergänzung von Abläufen, um Folgejahrgänge laufend zu integrieren
- Integration der digitalisierten Versionen von Schweizerische Bauzeitung / tec 21 und tracés des SIA.
Zusätzlich Hosting der laufenden Ausgaben.
- Prüfung und Auswahl weiterer vorgeschlagener Zeitschriftentitel, Rücksprache mit Bibliotheken.
- Vereinbarungen mit den Herausgebern.
→ Ziel: Betrieb langfristig sichern.
- Digitalisierung und Integration dieser Titel.

Vielen Dank!

Dr. Matthias Töwe
Konsortium der Schweizer Hochschulbibliotheken
c/o ETH-Bibliothek
Rämistrasse 101
CH-8092 Zürich
0041-(0)44 632 60 32
matthias.toewe@library.ethz.ch
<http://lib.consortium.ch>