

# METADATOVÝ EDITOR PRO DIGITÁLNÍ KNIHOVNY

*Martin Šárfy, Ústav výpočetní techniky, MU Brno*

## **1. Úvod**

Metadatový Editor (<http://sf.net/projects/dme>, dále jen ME) je nástroj původně vyvíjený na Masarykově univerzitě v Brně pro potřeby projektu „Česká digitální matematická knihovna DML-CZ“ programu „Informační společnost“ Akademie věd ČR (<http://project.dml.cz>, Projekt 1ET2001 90513) [1], na jehož část navazuje Moravská zemská knihovna v Brně v rámci grantového projektu „Metadatový editor pro systém Kramerius“ programu Ministerstva kultury ČR. Vývoj probíhá formou Open Source Software a nástroj je (v podobě „as it is“) k dispozici zdarma.

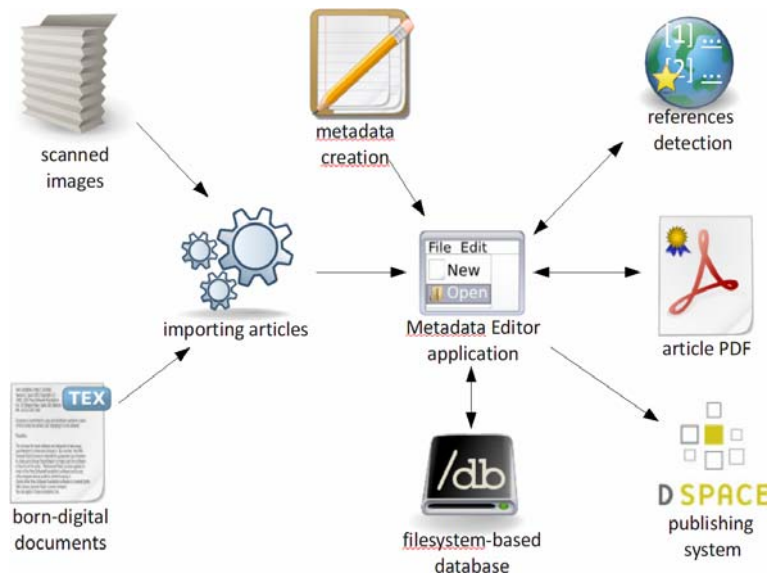
## **2. Klíčové vlastnosti ME**

Cílem Metadatového Editoru je umožnit kompletní digitalizaci tištěných materiálů, od skenování předloh až po podobu vhodnou pro převzetí publikačním systémem. Důraz je kladen na flexibilitu pracovního postupu, jakož i na volnost úrovně a formátu metadatového popisu. ME podporuje následující pracovní postup tvorby digitální knihovny:

1. Vkládání skenovaných stránek (časopisů, sborníků či monografií).
2. Manipulace se stránkami (korekce pořadí, duplicit, vakátů apod.).
3. Vytvoření textové reprezentace stránek pomocí OCR nástrojů.
4. Vytvoření hierarchické struktury (článek, kapitola, číslo, ročník, ...).
5. Detekce článkových citací, provázání s citačními databázemi.
6. Popis metadat – článků, periodik, sborníkových řad, monografií apod.
7. Vytvoření popisu z MARC záznamu staženého přes Z39.50 protokol.
8. Správa autoritní databáze – jmenné formy, transliterace apod.
9. Generování dvouvrstvých článkových PDF souborů včetně metadat.
10. Export dat do publikačního systému (např. DSpace [6] nebo Kramerius [4]).

Není nutno provádět všechny kroky uvedeného postupu. Alternativně lze na vstupu místo skenovaných stránek vkládat rovnou hotová vysázená PDF (a využít ME pouze k vytvoření bibliografického popisu) nebo načítat data uložená v systému Kramerius. Rovněž není nutno budovat autoritní bázi či vytvářet až článkovou úroveň hierarchie. Systém nechává uživateli veš-

kerou volnost při vytváření obsahu digitální knihovny, naprostá flexibilita je však vyvážena nutností netriviální konfigurace (a místy až zásahy do zdrojových kódů aplikace). Ve svém základním nastavení však tvoří ucelenou sbírku všech potřebných nástrojů pro tvorbu digitální knihovny [3].



Obr. 1: Celkový přehled vytváření digitální knihovny Metadatovým Editorem.

Velký důraz je kladen na schopnost jednoduché adaptace na zvolený bibliografický formát. Při editaci popisných metadat je využíván systém *SForms* (viz dále) založený na XML Schema šablonách, pomocí kterých lze definovat seznam polí, jejich vzájemné zanoření či integritní omezení vyplňovaných hodnot. Tím lze bibliografický popis vytvářet v téměř libovolném XML formátu, od standardně připraveného formátu pro systém Kramerius po formát vytvořený na míru instituci, která ME používá.

### 3. Postup zpracování dat

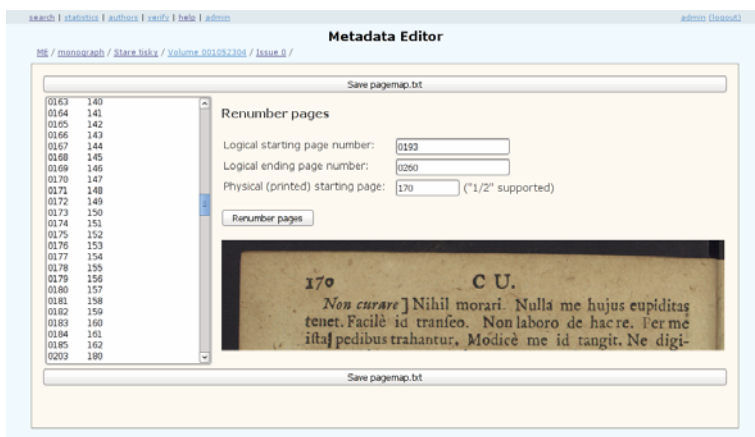
Datový obsah digitální knihovny je vytvářen vložením skenovaných stránek, jejich hierarchickým začleněním a obohacením o bibliografický popis jednotlivých částí.

### 3.1. Vytvoření základních struktur, vložení obrazových souborů

V ME lze zakládat nové publikace (periodika, sborníky či monografie) a ty dělit do nižších hierarchických úrovní (periodika na ročníky a čísla, sborníkové řady na jednotlivé sborníky, monografické série na jednotlivé monografie). Na úrovni čísla, sborníku nebo jednotlivé monografie pak lze vkládat skeny jednotlivých stran. Vkládání jednotlivých obrazových souborů stran je možné několika způsoby:

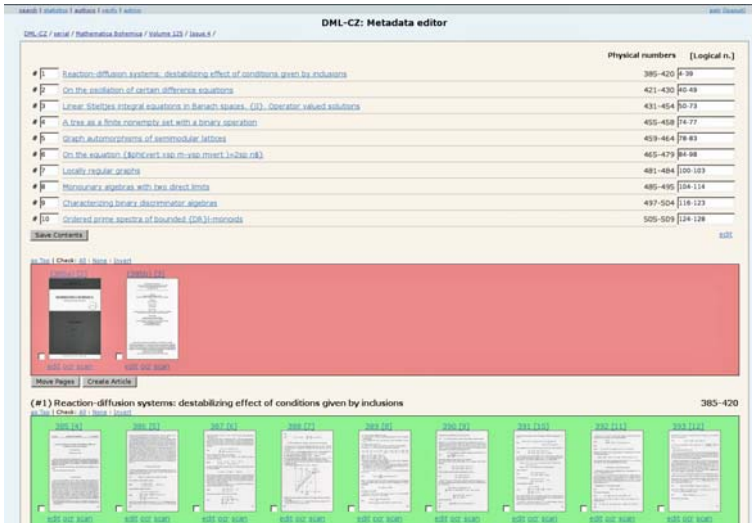
1. Vložení ZIP souboru s několika TIFF stránkami.
2. Nasdílení síťové složky na server a import stran v ní obsažených.
3. Zadáním rozsahu URL adres, odkud má ME stránky stáhnout.
4. Ruční vkládání jednotlivých stránek (při drobných korekcích).
5. Importem titulu či čísla ze systému Kramerius.

Po vložení obrazových souborů stran je možné zobrazit náhled všech stran čísla („číslem“ budeme v dalším rozumět i jednotlivý sborník nebo monografii), odkud je možné stránky jednoduše přesouvat, odstraňovat vakáty či stránky nahrazovat. Také je možné stránky hromadně číslovat (zvolením rozsahu stran a počátečního čísla strany), a to i pro vícestloupcové publikace (čísla sloupců oddělená lomítkem). Pro přehlednost se během číslování stránky zobrazuje výřez horního a spodního pruhu stránky, ve kterém bude číslo stránky pravděpodobně zobrazeno.



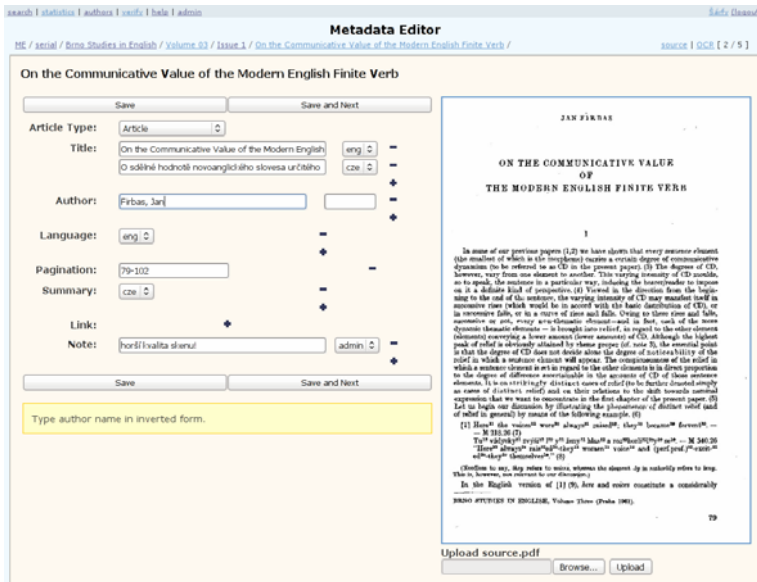
Obr. 2: Nástroj na hromadné přečíslování stran.

V dalším kroku je následně možné skupiny stránek sdružovat do článků. K dispozici je nástroj pro jejich přehledné zobrazení s možností přesouvání stránek mezi články, vytváření nových článků apod. (viz. obrázek č. 3).



Obr. 3: Vytváření struktury článků v čísle.

Alternativně lze místo stránkového zpracování vkládat na úrovni čísla rovnou hotová PDF článků, máme-li je k dispozici (např. z redakčního systému).



Obr. 4: Formulář pro vyplnění článkových metadat.

### 3.2. Vytvoření metadatového popisu

V tomto kroku se vytváří metadatový bibliografický popis článku, monografie, monografické série, sborníku, sborníkové řady či časopisu.

Pro vytváření metadatového popisu byla vyvinuta technologie *SForms*. Technicky je v ní formát metadatového popisu definován pomocí XML Schema šablony, a to včetně značek, které upravují vzhled zobrazovaného formuláře. U pole lze zvolit, zda se skládá z jiných polí, zda je povinné, opakovatelné, jaký formát má mít zapisovaná hodnota apod. Definici lze dále obohatit o uživatelskou dokumentaci (např. o katalogizační pravidla), která se pak zobrazuje přímo u vyplňovaného formuláře.

Tímto způsobem lze snadno nadefinovat množinu polí, které chceme na dané úrovni metadatového popisu evidovat. Vytvářená metadata jsou uložena v XML souborech, při každém uložení je navíc XML soubor validován oproti jeho XML Schema definici. Rovněž lze definovat počáteční XML obsahující implicitní hodnoty.

Ve standardní instalaci je schéma předpřipraveno pro systém Kramerius.

Ačkoliv je formát zápisu zcela volný, je vhodné, aby sémantiku některých základních polí (např. autor, název apod.) znal i Metadatový Editor. Obohacením XML Schema definice lze Metadatovému Editoru vysvětlit, které pole odpovídá kterému *internímu poli* Metadatového Editoru. ME pak umí podle těchto polí vyhledávat nebo je správně zobrazovat v popisích.

```
<xs:element name="Periodical">
  <xs:complexType form:render="Listing">
    <xs:sequence>
      <xs:element ref="SerialTitle" minOccurs="1"
        maxOccurs="1"/>
      <xs:element name="ISSN" minOccurs="0" maxOccurs="1">
        <xs:simpleType form:render="TextInput" form:size="12"
          me:meta="issn" form:hint="ISSN tištěného
            periodika">
          <xs:restriction base="xs:string">
            <xs:pattern value="(\d\d\d\d-\d\d\d[0-9X])" />
          </xs:restriction>
        </xs:simpleType>
      </xs:element>
      <!-- další prvky... -->
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

Obr. 5: SForms – XML Schema definice obohacená o vizuální (form:) a sémantické značky (me:).

Mechanismus SForms je použit pro vytváření metadatového popisu na všech úrovních vytvářené hierarchie. Na úrovni článku je během editování k dispozici náhled titulní stránky článku. To napomáhá k snadnějšímu vyplňování polí.

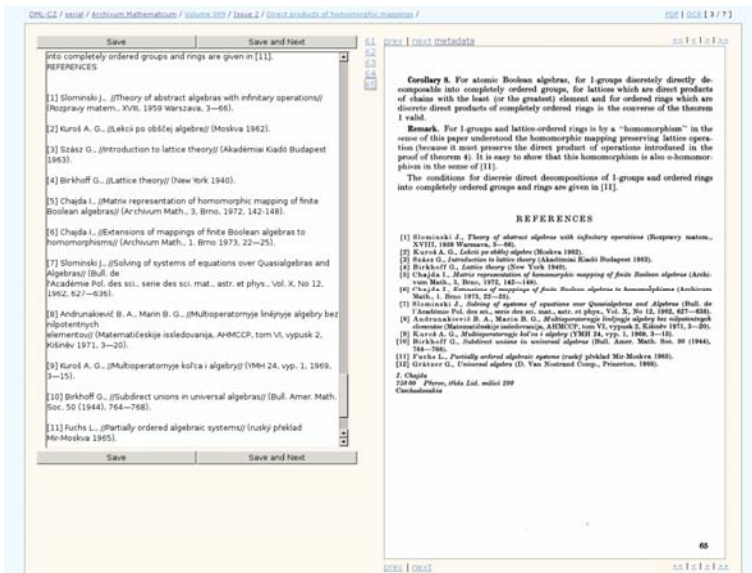
### 3.3. Pokročilé možnosti metadatového popisu

Metadatový popis lze dále kromě základního metadatového popisu obohatit o následující:

#### 3.3.1 Bibliografický záznam ve formátu MARC21

Metadatový popis lze rozšířit také o MARC21 záznam. ME obsahuje vestavěného klienta Z39.50 protokolu, pomocí kterého lze prohledat vzdálenou bibliografickou bázi a MARC21 záznam z ní stáhnout. Stažený záznam je možno jednoduchým rozhraním dále editovat. Seznam Z39.50 serverů je konfigurovatelný, včetně definice polí, přes které lze dohledávat.

Pro metadatový formát systému Kramerius je dále k dispozici převodník, kdy při uložení MARC21 záznamu dojde ke konverzi bibliografických údajů do formátu systému Kramerius.



Obr. 6: Poloautomatizované označení referencí v plném textu článku z OCR.

### **3.3.2 Bibliografické citace**

Je-li to v lidských silách, lze u článku evidovat i jejich bibliografické citace. V ME je tento proces podpořen automatizovaným systémem, vychází se z OCR textu článku, ve kterém se systém pokusí najít blok referencí, v něm znakem prázdného řádku oddělí jednotlivé citace a v citacích znaky „/“ oddělí autora a název. Lidská obsluha však v dalším kroku musí takto označené reference projít a značky zkontrolovat, úspěšnost automatického procesu není stoprocentní a je silně závislá na kvalitě OCR.

Díky těmto značkám je následně vytvořeno XML se strukturovaným zápisem reference. Dále je možné referencí odkazovaný článek automatizovaně dohledat v citační databázi a čtenáři pak v publikačním systému u reference nabídnout klikatelný odkaz.

### **3.3.3. Linkování článků, článkové sekce**

Mechanismus linkování článků lze použít pro propojování článků, které spolu souvisejí, např. pokračování článku, odvozené práce, recenze článku, doporučené související články apod. Tato informace pak může být zohledněna v publikačním systému.

V případě, že jsou v originálním dokumentu články či kapitoly zanořené do více samostatných sekcí, je možné toto rozdělení reflektovat i v Metadataovém Editoru a následně i v publikačním systému.

### **3.3.4. Autoritní báze**

Pro zvýšení kvality autoritních údajů lze využít mechanismus autoritní báze. Ta umožňuje evidovat jednotlivé autory, různé způsoby zápisu jejich jména, jejich transliterace, případně lze jejich popis rozšířit o základní osobní údaje nutné pro jednoznačnou identifikaci autora. Propojením článků s autoritní bází pak umožňuje např. zobrazit všechny články zadaného autora, a to i v případě shody jména s jiným autorem nebo při publikaci článků pod pseudonymem.

Autoritní báze je inspirována autoritními bázemi používanými v knihovních systémech. V bibliografickém popisu či referenci se pole se jménem vyplní stejně, jak je zapsáno v originálním dokumentu, nicméně k poli se může přidat identifikátor osoby v autoritní bází. Tento model dovoluje rozlišit více autorů stejného jména, více jmen stejného autora nebo pokrýt případ, kdy si osobou autora nejsme jisti.

DML-CZ / Authors / B / BolzaB /

**BolzaB** « | « | « | « | » | » | » | »

**Id:**

**Description:**

**Profession:**

**Origin:**

**Date of Birth:**

**Date of Death:**

**Status:**

**Forms:**

	Surname	Name	Display	Transliterated	Attribute	
(3) <input type="checkbox"/>	Bolzano	Bernard	Bolzano, Bernard	Bolzano, Bernard	preferred	<a href="#">more</a>
(4) <input type="checkbox"/>	Bernard Bolzano	Bernard Bolzano	Bernard Bolzano	Bernard Bolzano	other	<a href="#">more</a>
(0) <input type="checkbox"/>	Bolzano	Bernard	Bolzano, Bernard	Bolzano, Bernard	other	<a href="#">more</a>

**Articles:**

<a href="#">Beiträge zu einer begründeteren Darstellung der Mathematik</a>	Bolzano, Bernard
<a href="#">Spisy Bernarda Bolzana. Svazek 2. Zahlentheorie</a>	Bolzano, Bernard
<a href="#">Betrachtungen über einige Gegenstände der Elementargeometrie</a>	Bolzano, Bernard
<a href="#">On the best state</a>	Bolzano, Bernard
<a href="#">Works of Bernard Bolzano: On the best state</a>	Bolzano, Bernard
<a href="#">Spisy Bernarda Bolzana. Svazek 5. Geometrische Arbeiten</a>	Bolzano, Bernard
<a href="#">Spisy Bernarda Bolzana. Svazek 1. Functionenlehre</a>	Bolzano, Bernard
<a href="#">The correspondence of B. Bolzano and F. Exner</a>	Eduard Winter: Bolzano, Bernard; Exner, František

Obr. 7: Editace autoritního záznamu.

Díky budování autoritní báze se výrazně zjednodušuje dohledávání a opravování překlepů jmen autorů. V plánech je také napojení autoritní báze na externí databáze, např. autoritní bázi Národní knihovny ČR.

### 3.4. Generování PDF a export dat

Při exportu dat je možno z naskenovaných obrázků stran generovat článkové PDF soubory. Pokud nad obrázky stran proběhlo OCR rozeznávání, lze generovat tzv. dvouvrstvé PDF, kdy horní (viditelnou) vrstvu tvoří obrázek stránky, zatímco spodní (neviditelnou, ale prohledávatelnou) tvoří textová reprezentace. Dle zadané šablony lze k vygenerovanému PDF vytvořit titulní stránku obsahující metadata článku, logo instituce, copyright a podobně. PDF lze dále podepsat digitálním podpisem zajišťujícím pravost dokumentu. Příklad takového PDF lze najít na stránkách digitální knihovny DML-CZ (viz <http://dml.cz>).

Metadatový Editor primárně slouží pro vytváření obsahu. Pro jeho zpřístupnění koncovým uživatelům je nutné použít některý z publikačních systémů. ME standardně obsahuje mechanismus pro export dat do publikačních systému DSpace a Kramerius, pro jiné systémy a pro jiné metadatové formáty ([5, 7]) je však nutné napsat si vlastní konverzní program.



## 4. Další funkce

### 4.1. Vyhledávání

Možnost procházení jednotlivých titulů, ročníků, čísel či výtisků je za účelem kontrol úplnosti a konzistence dat doplněna o možnost jednoduchého vyhledávání dle zadaných kritérií. Pole pro vyhledávání jsou v tuto chvíli limitována na prvky, které ME nativně zná, t.j. ty, které byly označeny jako příslušející interní pole v XML Schema definici používaného metadatového formátu (standardně autor, název, rok, ...).

Hledání lze omezit na vybraný titul, jazyk či kategorii článku. Režimů vyhledávání je několik:

- *exact* – hodnota pole je totožná s vyhledávaným výrazem
- *equal to* – záznamy obsahující zadaný výraz ve vybraných polích
- *none equal to* – žádný z vybraných (i více) polí neobsahuje vyhledávaný výraz
- *empty* – záznamy mající zadané pole prázdné
- *not empty* – záznamy mající zadané pole vyplněné

Pro přehlednost lze rovněž omezit seznam polí, které se ve výsledku budou zobrazovat.

The screenshot shows the ME search interface. At the top, there is a list of journals including 'General Topology and its Relations to Modern Analysis and Algebra (eng)', 'Applications of Mathematics (cze)', 'Archivum Mathematicum (eng)', 'Archivum Mathematicum (czech) (cze)', 'Časopis pro pěstování matematiky (cze)', 'Časopis pro pěstování matematiky a fysiky (cze)', 'Pismo o matematice a fysice pro pěstování matematiky a fysiky - JENSTEIN (cze)', 'Pismo o matematice a fysice - JENSTEIN (cze)', and 'Commentationes Mathematicae Universitatis Carolinae - JENSTEIN (cze)'. Below this is a 'Query' section with a table of element, property, category, relation, and display options. The 'Property' column has a dropdown menu set to 'Any'. Below the query table are search buttons and input fields for 'MSC' and 'Title'. At the bottom, there is a table of search results for '159 articles matches'.

Element	Property	Category	Relation	Display
<input type="checkbox"/> Title	<input type="checkbox"/> Any	<input type="checkbox"/> math	<input type="checkbox"/> equal to	<input type="checkbox"/> Title
<input type="checkbox"/> Author		<input type="checkbox"/> news	<input type="checkbox"/> none equal to	<input type="checkbox"/> Author
<input type="checkbox"/> MSC		<input type="checkbox"/> history	<input type="checkbox"/> empty	<input type="checkbox"/> MSC
<input type="checkbox"/> idMR		<input type="checkbox"/> politics	<input type="checkbox"/> not empty	<input type="checkbox"/> idMR
<input type="checkbox"/> idZBL		<input type="checkbox"/> editorial	<input type="checkbox"/> exact	<input type="checkbox"/> idZBL
<input type="checkbox"/> idJFM		<input type="checkbox"/> contents		<input type="checkbox"/> idJFM
<input type="checkbox"/> Note		<input type="checkbox"/> other		<input type="checkbox"/> Note
<input type="checkbox"/> Note: private		<input type="checkbox"/> review		<input type="checkbox"/> Note: private
<input type="checkbox"/> Error		<input type="checkbox"/> physics		<input type="checkbox"/> Error
<input type="checkbox"/> Language				<input type="checkbox"/> Language

Article ID	Type	Title	Author	MSC	idMR	idZBL	idJFM	Note	Note: private	Error	Language
math/CzechMath/23-1979-3/3	math	On the differentiation of complex functions in finite and infinite dimensional spaces (eng)	Zajíc, Luděk	52A05 46G05 28A27	MR536060	0429.46007					eng

Obr. 8: Vyhledávací mechanismus ME.

Výhledově počítáme s mechanismem obecného vyhledávání ve strukturovaných XML metadatech a také s funkcí vyhledávání v plných textech článků. Důraz je však spíše než na běžné „pravděpodobnostní“ vyhledávání, vhodné pro koncové uživatele, kladen na úplné a přesné výsledky, ulehčující hledání chybných či nesprávných údajů.

## 4.2. Automatizovaná verifikace

Pro udržení vysoké kvality digitální knihovny je vhodné data průběžně kontrolovat. Tuto činnost lze do jisté míry automatizovat. Metadatový Editor disponuje silným a snadno rozšiřitelným mechanismem kontrol správnosti a úplnosti dat, mezi které patří zejména:

- kontrola vyplnění povinných metadatových polí
- kontrola integrity datového úložiště (validace XML, kompletnost odkazovaných souborů, ...)
- kontrola správného pořadí stran
- porovnání údaje o jazyku článku oproti odhadu detekce jazyka z plného textu článku
- kontrola velikosti a rozlišení TIFF souborů skenovaných stránek
- kontrola syntaxe značek v referencích

Každý verifikační test je psán jako samostatný modul obsahující spustitelný program spolu s formálním popisem vstupních a výstupních parametrů. Formální popis je použit pro vytvoření vhodného uživatelského rozhraní v Metadatovém Editoru. Kontrolní test je možné spustit buď na všech datech nebo lze vybrat jen určitou hierarchickou podmnožinu.

Kontroly je možno také pouštět periodicky na pozadí, přičemž při chybě dojde systémovému administrátorovi varovný e-mail. Tento mechanismus lze použít pro pravidelnou (např. denní) kontrolu dat.

## 4.3. Správa uživatelů, uživatelské role

Systém disponuje rozsáhlým mechanismem autorizace uživatelů. Pro provedení jakékoliv operace v Metadatovém Editoru je potřeba mít v systému uživatelský účet, který má oprávnění jak pro danou operaci, tak pro daná data, kterých se operace týká.

Vytvářet uživatelské účty a nastavovat přístupová práva smí pouze osoba mající k tomu oprávnění, podle nastavení ME lze hesla k účtům mít uložena přímo v databázi ME nebo se heslo může ověřovat oproti autentizačnímu systému Kerberos.

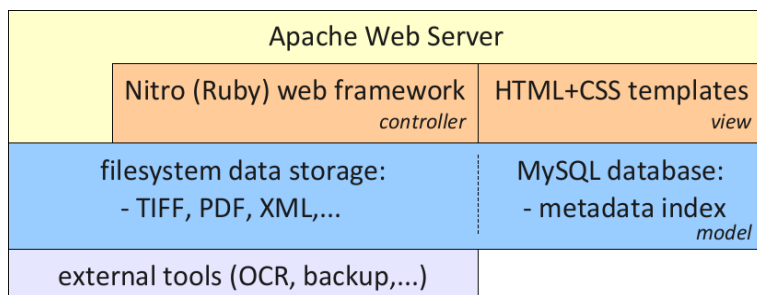
## 4.4. Systém nápovědy

Dokumentace k celému systému je přímo dostupná přes uživatelské rozhraní ME. Tato dokumentace je fyzicky uložena v *Google Docs* dokumentech, odkud se při zobrazování on-line stahují a zobrazují. V dokumentaci lze použít některé konstrukce, které vedou např. ke vložení „*screenshotu*“ obrazovky Metadatového Editoru (technicky je toto realizováno vložení „*inline*“ rámu s danou stránkou do HTML kódu). Tím je garantována aktuálnost snímků obrazovky (byť za cenu jistých technických komplikací).

Uživatelé mající k tomu oprávnění mohou dokumentaci přes aplikaci Google Docs rovnou upravovat. Na každé stránce s nápovědou zobrazuje ME tlačítko „upravit tuto stránku“ vedoucí na Google Docs, uložené změny se v *on-line* zobrazované nápovědě ME projeví okamžitě. Na vytváření dokumentace třeba i různých částí ME tak může spolupracovat více pracovníků.

## 5. Technické řešení

Metadatový Editor je, jak už bylo naznačeno, aplikace s architekturou klient-server, kde serverová část běží pod operačním systémem Linux a jako klientská část slouží běžný webový prohlížeč.



Obr. 9: Systémová architektura Metadatového Editoru.

### 5.1. Architektura serverové části aplikace

Serverová část má třívrstvou *Model-View-Controller* architekturu striktně oddělující vzhled aplikace, aplikační logiku a datovou reprezentaci. Jádro aplikace je napsáno použitím frameworku Ramaze v programovacím jazyku Ruby, pro svůj běh však využívá i celou řadu dalších technologií a knihoven:

- SForms – systém dynamických formulářů (perl a JavaScript)
- On-Line Image Server – zobrazování náhledů (výřezů) stránek (PHP)
- Kramerius IO – modul spolupracující se systémem Kramerius (Java)
- OCR – rozhraní k FineReader SDK pro automatizované OCR (C++)
- JQuery – knihovna pro pokročilé funkce uživatelského rozhraní (JavaScript)
- a další...

Výčet zdaleka není kompletní – jen pro zajímavost – při kompletní instalaci na systém Debian se celkově stahuje 157 MB softwarových závislostí.

### 5.2. Datové struktury aplikace

Jako primární datové úložiště je používána adresářová struktura s velice jednoduchou „samopopisující“ strukturou. Tato struktura, sloužící jako roz-

hraní mezi různými moduly ME, jeho rozšířeními či externími nástroji, se ukázala jako nanejvýš praktický koncept pro spolupráci nezávislých vývojářů. Importní a exportní skripty, dohledávání citací v citačních databázích, OCR, zálohování, kontrolní mechanizmy, to vše využívá jednotnou adresářovou strukturu nezávisle na zvoleném jazyku implementace nástroje.

Následuje příklad výpisu adresářové struktury. První číslo je stránkově orientované (pochází ze skenování), druhé je orientované článkově (PDF články dodané redakcí):

```
serial/  
  Lidovky/  
    detail.xml - popisná metadata seriálu, ve formátu dle XML  
    Schema definice  
    24-1974-4/ - ročník-rok-číslo, adresář reprezentující  
    časopisecké číslo  
      tiff/ - skenované číslo - originální TIFFy stránek  
        0001.tif - logický identifikátor stránky  
      pagemap.txt - názvy fyzických stran a jejich mapování na  
      logické  
      contents.txt - seznam článku v čísle (id článku a rozsah  
      stran)  
      #1/ - adresáře pro jednotlivé články  
        detail.xml - popisná metadata článku  
        references.xml - seznam referencí  
        item.pdf - PDF s článkem  
  
proceedings/  
  WSAA/ - akronym sborníkové řady  
    30-2002-1/ - ročník-rok-číslo, adresář reprezentující  
    konkrétní sborník  
      detail.xml - popisná metadata sborníku  
      #1/ - článkový adresář  
        detail.xml - popisná metadata článku  
        source.pdf - originální PDF článku
```

Na každé adresářové úrovni je soubor detail.xml popisující metadata objektu (odpovídající XML Schema definici konkrétní úrovně). Zda je číslo stránkově nebo článkově orientované závisí na existenci souboru pagemap.txt, obsahující seznam skenovaných stran. Adresářová struktura může být volně rozšířena o další soubory – např. MARC záznam, náhledy stran, zdrojové soubory článků apod.

Pro číslování stránek editor rozlišuje mezi interním jednoznačným identifikátorem stránky (*logickým číslem*) a popisným pojmenováním stránky (vytištěné, *fyzické číslo*). Zatímco logické číslo je přiřazováno systémem (vždy čtyřciferné číslo stránky) a nelze jej změnit, na fyzické číslo nejsou kladena žádná omezení a může obsahovat jakoukoliv textovou hodnotu (římské číslování, závorky, u vícsloupcových publikací lomítka apod.). Logické číslo

pochází z číslování během skenování a kvůli vakátům, duplicitám či dodatečně skenovaným stranám nemusí číselně odpovídat pořadí stran – posloupnost stran je definována pořadím ve výčtu logických stran v souboru pagemap.txt. Seznam článků čísla je v souboru contents.txt, u každého článku je uveden výčet (logických) čísel stran, které daný článek tvoří.

Jako identifikátor objektu se v rámci ME používá cesta k objektu v adresářové struktuře (např. proceedings/WSAA/30-2002-1/#1 pro první článek uvedeného sborníku). Kromě URL se tyto identifikátory používají např. při linkování článků apod.

ME si z adresářové struktury dále vytváří index v MySQL databázi pro urychlení přístupu a pro funkci vyhledávání. Tento přístup kombinuje výhody obou úložišť – pružnou administraci i rychlost přístupu.

### **5.3. Uživatelské rozhraní**

Uživatelské rozhraní aplikace je, díky webové klient-server architektuře, dostupné přes webový prohlížeč. Vzdálený autorizovaný přístup umožňuje distribuovat práci mezi pracovníky s různou mírou znalostí či zkušeností.

Síťová komunikace je šifrována pomocí HTTPS spojení a používaný model přístupových práv zohledňuje oprávnění uživatele vzhledem k datům, ke kterým přistupuje. Veškeré změny jsou logovány, statistiky jde použít například i pro odhad pracovní činnosti pracovníka.

Pro zobrazování obrazových souborů na straně klienta je použit pro tyto účely vyvinutý systém OLIS – On-Line Image Server – transparentní vrstva mezi webovým prohlížečem a Metadatovým Editorem, mající za úkol konvertovat obrazové soubory (TIFF, PDF, DjVu, ...) do formátu a velikosti vhodné pro klienta. Generované náhledy či výřezy stran systém ukládá do vyrovnávací paměti pro urychlení opakovaného přístupu. Další funkcí systému OLIS je zobrazování náhledu stránky ve vysokém rozlišení včetně možnosti plynulého zvětšování a posouvání po stránce.

### **5.4. Instalace aplikace**

Doporučený způsob instalace ME je na dedikovaný Linuxový server. Distribuční balík však kromě aplikačních kódů pro instalaci na Linuxovém serveru obsahuje také VMware obraz s předinstalovaným Metadatovým Editorem, který je možno pouštět na libovolném operačním systému, na kterém běží VMware, čili prakticky na všech verzích Microsoft Windows.

Aplikaci je tedy možno provozovat dvěma způsoby – nativně na Linux serveru nebo pomocí VMware obrazu i na jiných operačních systémech (např. na Microsoft Windows). Součástí dokumentace je popis instalace jak na Linux server, tak pro provozování ve VMware virtuálním stroji. Instalace

aktualizací zdrojových kódů ME je postavena na systému pro správu verzí Subversion.

Propojení s instalací systému Kramerius spočívá v instalaci speciálního rozšíření do systému Kramerius a v nasdílení disku s repozitářem Krameria. Modul je používán při komunikaci se systémem Kramerius, sdílený repozitář slouží jako překladiště dat mezi ME a Krameriem.

Hardwarové nároky na aplikaci nejsou přehnané, na základní provoz postačí jakýkoliv modernější server. Výkon hardware je limitující pouze při provádění plošných údržbových operací.

## 6. Závěr, další vývoj

Metadatový Editor je v současné době produkčně používán v několika instalacích. V rámci projektu DML-CZ bylo jeho pomocí zpracováno více než 200 000 stran v 15 000 člancích [2]. ME je dále nasazen na Filozofické fakultě Masarykovy univerzity v Brně na digitalizaci sborníků vydávaných na MU. Další instalace, propojená se systémem Kramerius, je v Moravské zemské knihovně v Brně.

Další vývoj ME bude orientován na větší podporu administrativních metadat, generování dokumentů v archivním PDF/A formátu, v plánu je podpora formátu systému Manuscriptorium, výstup do podoby statického webu (včetně metadat pro *Google* a *Google Scholar* indexaci), integrace s Registrem Digitalizace [8] a další.

Vadou na kráse ME se však pomalu stává jeho rozsáhlost. Postupem času v ME vzniklo množství rozšíření kladoucí netriviální nároky na snadnost nastavování, správy a údržby. V blízkém budoucnu proto také plánujeme konsolidaci některých funkcí ME a dotažení systému webově orientované konfigurace. I přes tyto nedostatky však bylo během času prověřeno, že ME je nástroj schopný efektivně posloužit při jinak náročné tvorbě digitální knihovny.

## Literatura

[1] Bartošek, Lhoták, Rákosník, Sojka, Šárfy: DML-CZ: The Objectives and the First Steps. In *Communicating Mathematics in the Digital Era*.

London : A. K. Peters Ltd., Transatlantic Publishers Group, 2008. Sv. 1, ISBN 9781568814100, s. 69–79.

[2] Bartošek, Kovář, Šárfy: DML-CZ Metadata Editor. In *Towards Digital Mathematics Library*. Brno : Masaryk University, 2008. ISBN 978-80-210-4658-0, s. 139–151. 28.7.2008, Birmingham, UK.

- [3] Bartošek, Miroslav: Česká digitální matematická knihovna. *INFORUM 2008: 14. ročník konference o profesionálních informačních zdrojích*. Praha : Albertina icome Praha, 2008. 11 s. 2008, Praha.
- [4] Martin Duda: Kramerius v praxi. *Knihovny současnosti*. Sborník z 15. konference, konané ve dnech 11.–13. září 2007 v Seči u Chrudimi. Brno 2007. 515 s. ISBN 978-80-86249-41-7
- [5] Krejčíř, Vlastimil: Systémy pro tvorbu digitálních knihoven. *Ikaros* [online]. 2006, roč. 10, č. 5/2 [cit. 2009-04-15]. Dostupný na World Wide Web: <<http://www.ikaros.cz/node/3457>>. URN-NBN:cz-ik3457. ISSN 1212-5075.
- [6] Krejčíř, Vlastimil: DSpace 1.4. [online] Slajdy z přednášky pro předmět *Digitální knihovny* na Fakultě informatiky MU. Podzim 2006. [cit. 2009-04-15]. Dostupný na World Wide Web: <<http://www.dspace.cz/soubory/pv070slajdy.pdf>>.
- [7] Lhoták, Martin: Open source pro digitální knihovnu. In *Automatizace knihovnických procesů – 11* : sborník z 11. ročníku semináře pořádaného ve dnech 16.–17. května 2007 v Liberci. Praha : ČVUT, 2007. 136 s. ISBN 978-80-01-03691-4.
- [8] Pavlová, Lucie: RD.CZ: Evidence digitalizovaných dokumentů a sledování procesu zpracování. *Knihovny současnosti 2008*. Sborník z 16. konference, konané ve dnech 16.–18. září 2008 v Seči u Chrudimi. Brno 2008. 373 s. ISBN 978-80-86249-49-0