

MASARYKOVA UNIVERZITA

Fakulta informatiky



FEDORA

The Flexible Modular Open-Source Repository Platform

PV070

Alžbeta Strakošová

Obsah

Úvod	3
1. Flexible Extensible Digital Object Repository Architecture	3
2. Vznik Fedory	4
3. Fedora 4	4
4. Príklady použitia Fedory v praxi	6
Záver	6
Obrázky	7
Metadata v Dublin Core (DC)	8
Použitá literatúra	8

Úvod

Fedora je robustný modulárny repozitárový systém navrhnutý pre uchovávanie a správu digitálnych objektov. Je využívaná inštitúciami po celom svete pre správu rozsiahlych digitálnych kolekcii s najrôznejším obsahom. Jej vývoj je zameraný predovšetkým na perzistenciu dát a schopnosť interakcie systému s inými aplikáciami a službami.

1. Flexible Extensible Digital Object Repository Architecture

Systém Fedora, vyvinutý v roku 1998 na Cornellovej univerzite, bol založený na vedeckom článku od autorov Sandy Payette a Carla Lagoze s názvom Flexible Extensible Digital Object Repository Architecture (Fedora). V tejto práci predstavili architektúru repozitára digitálnych objektov, ktorá mala vyhovieť základnej požiadavke pre digitálne knižnice – poskytnúť spoľahlivé a zabezpečené metódy na ukladanie digitálneho obsahu a prístup k nemu. Okrem toho kládla veľký dôraz na rozšíriteľnosť a interoperabilitu.

Za kľúčové vlastnosti tejto architektúry označili okrem iného podporu heterogénnych typov dát, adaptáciu systému na nové dátové typy a schopnosť agregácie distribuovaných dát do komplexných objektov. Digitálnym objektom mal byť pre užívateľa zapuzdrený bitový prúd, ktorého význam bol definovaný jeho funkčnou vrstvou. Tá reprezentovala akési rozhranie vymedzujúce správanie objektu a takto transformovala čisté dáta do informácie čitateľnej pre človeka – či už išlo o články, multimediálne súbory, alebo čokoľvek iné.

Tento princíp bol kľúčovým pre zachovanie podpory neobmedzených dátových typov aj s výhľadom do budúcnosti. Payette a Lagoze v článku zdôrazňujú, že vlastnosťou dobrej architektúry je schopnosť bezproblémovo integrovať nové formy obsahu a mechanizmy pre prístup k nemu.

V čase publikovania článku bol už vytvorený prototyp popísanej architektúry s cieľom dôkladného otestovania jeho rozšíriteľnosti, interoperability, spoľahlivosti a zabezpečenia.

2. Vznik Fedory

V roku 2000 bola Fedora úspešne implementovaná na University of Virginia ako prototyp systému pre manažment a prístup k digitálnym kolekciam s rôznorodým obsahom. Projekt ukázal, že Fedora dokáže bežať ako webová aplikácia postavená na relačnej databáze. Na druhej strane ale nedosahoval úroveň interoperability, ktorý si dávali za cieľ Payette a Lagoze. Univerzita následne dostala grant od Andrew W. Mellon Foundation, aby v spolupráci s Cornellovou univerzitou vybudovali sofistikovaný a robustný repozitárový systém založený na architektúre Fedora, ktorý bude využívať najnovšie webové technológie a štandardy ako napríklad XML a Web Services. Cieľom bolo poskytnúť základ, na ktorom možno budovať webové digitálne knižnice schopné vysokej úrovne interoperability a vyhovieť tak potrebám vedcov, knihovníkov a iných informačných komúnit.

V roku 2003 bola predstavená Fedora open-source software. Dnes je zastrešená neziskovou organizáciou DuraSpace, ktorá podporuje projekty poskytujúce riešenia v oblasti manažmentu a dlhodobého ukladania dát. Okrem repozitárových systémov poskytuje napríklad službu DuraCloud, komerčnú cloudovú infraštruktúru určenú pre akademické knižnice a výskumné centrá, alebo linked data aplikáciu VIVO, ktorá poskytuje vedeckým inštitúciám platformu pre zdieľanie informácií o svojich výskumníkoch a ich práci.

Vďaka možnosti poskytovať špecializovaný prístup k mohutným a komplexným digitálnym kolekciam je dnes systém Fedora široko využívaný akademickými, kultúrnymi, vzdelávacími a vedeckými organizáciami po celom svete. Medzi nich patria napríklad spomínaná Cornellova univerzita, University of Michigan alebo Dánska národná knižnica The Royal Library.

3. Fedora 4

Fedora je aktuálne dostupná vo verzii 4.x. Od jej začiatku sa udialo veľa zmien v štruktúre uložených dát, základná myšlienka ale ostala – poskytnutie stabilného základu v podobe repozitárových služieb založených na všeobecne prijímaných štandardoch. Tieto služby poskytujú cez aplikačné rozhrania RESTful APIs.

Fedora 4.x priniesla oproti jej predošlým verziám nový objektový model, ktorý podporuje hierarchiu. Obsah repozitára znázorňuje acyklický orientovaný graf (strom), v ktorom hrany reprezentujú vzťah medzi predkom a potomkom. Základným štruktúrnym prvkom je zdroj (resource) – akákoľvek webovo-adresovateľná entita s vlastným názvom a unikátnym identifikátorom. Zdroj môže byť identifikovaný tiež cestou (path) vedúcou od koreňa grafu, podobne ako v súborovom systéme. Jeho názov sa skladá z vopred registrovaného namespace prefixu a lokálneho názvu. Fedora obsahuje niekoľko preddefinovaných namespaces vrátane DC, FOAF, Fedora a LDP.

Príkladom zdroja je container, ktorý reprezentuje digitálny objekt, môže mať vlastnosti (properties) a potomkov. Jeho potomkom môže byť iný container, alebo binary. Binary, nazývaný tiež Non-RDF Source, môžeme chápať ako bitstream alebo súbor (file) a v repozitári je uložený vždy aj so svojim nonRdfSourceDescription zdrojom, ktorý ho popisuje.

Ďalšou oblasťou, na ktorú sa Fedora 4.x zamerala je podpora Linked Data. Okrem predošlého XML formátu umožňuje ukladať metadáta vo formáte RDF. Metadáta vo Fedore sú páry názov-hodnota, ktoré sa cez REST-API response preložia do RDF trojíc (triples). Tieto trojice sú vo forme podmet-predikát-predmet, kde podmetom je samotný zdroj, predikát je názov vlastnosti a predmet je hodnota vlastnosti. Hodnotou vlastnosti môže byť literál, alebo URI, ktorý môže patriť aj objektu mimo repozitára.

Jedným z mechanizmov pre zachovanie integrity dát vo Fedore je fixity checking – vypočítanie kontrolného súčtu a jeho porovnanie s kontrolným súčtom uloženým v repozitári spolu s dátami. Tento check môže byť zavolaný už pri nahrávaní dát do Fedory cez REST-API request, alebo kedykoľvek neskôr.

Umožňuje tiež export a import dát v štandardizovanom JCR/XML formáte. S novou stromovou štruktúrou je možnosť exportovať napríklad vybraný zdroj spolu so všetkými jeho potomkami – čiže nejaký podstrom daného repozitára.

Verzovanie vo Fedore 4.x so sebou prinieslo novú flexibilitu – okrem možnosti verzovania celého repozitára umožňuje zahrnúť a vylúčiť konkrétne datastreamy alebo typy obsahu, prípadne vytvoriť verziu objektu iba na REST-API request.

Fedora má veľmi aktívnu medzinárodnú komunitu developerov, prispievajúcu či už k samotnej Fedore – jadrú, alebo k vyvíjaniu kompletných aplikácií postavených na nej.

4. Príklady použitia Fedory v praxi

Ako už bolo spomenuté vyššie, Fedora je celosvetovo využívaná organizáciami predovšetkým z oblasti vzdelávania, vedy a výskumu.

Prvý príklad nájdeme už na samotnej Cornellovej univerzite – jej projekt AgriKnowledge, zameraný na rozvoj v poľnohospodárstve. Jej obsah tvoria predovšetkým vedecké články, reporty, prezentácie a technické dokumenty. Užívateľ má možnosť vyhľadávať a filtrovať zdroje podľa oblasti poľnohospodárstva, jazyka, komodity alebo geografickej lokácie. Všetky zdroje sú pritom voľne dostupné v rámci Fair use. (Obr. 1)

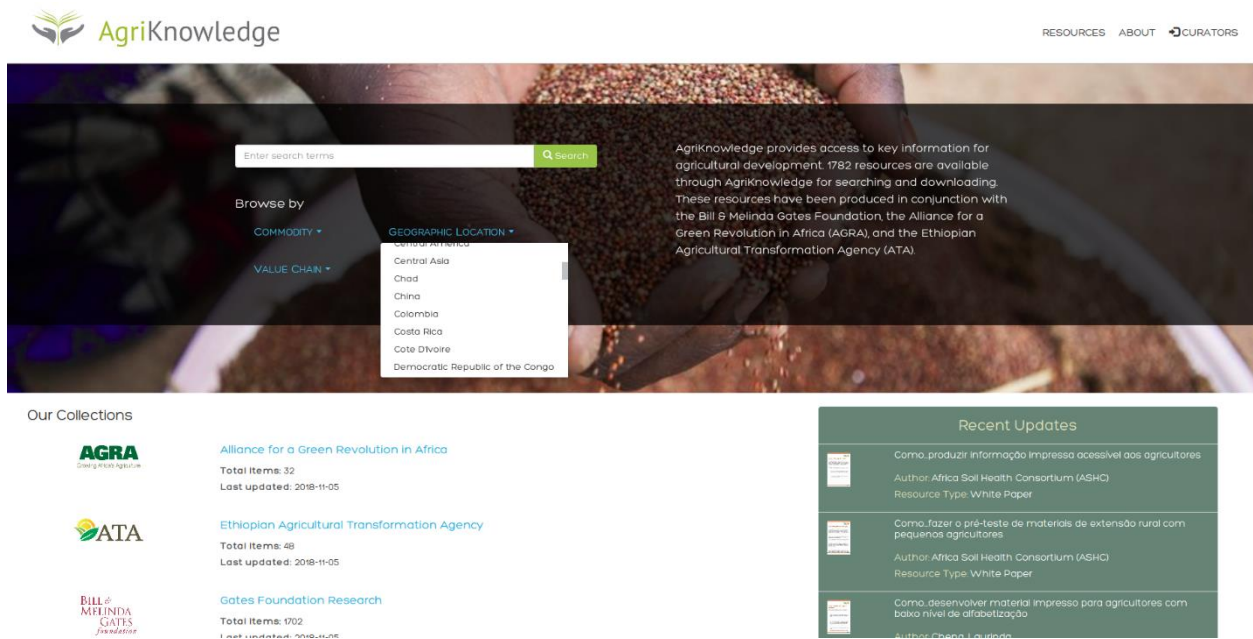
Ďalším príkladom je repozitár Deep Blue Data patraci University of Michigan Library, ktorý slúži na zdieľanie a archiváciu výskumných dát univerzity. Ponúka zdroje z oblasti vedy, techniky, sociálnych a humanitných vied, práva a mnoho ďalších. (Obr. 2)

Záver

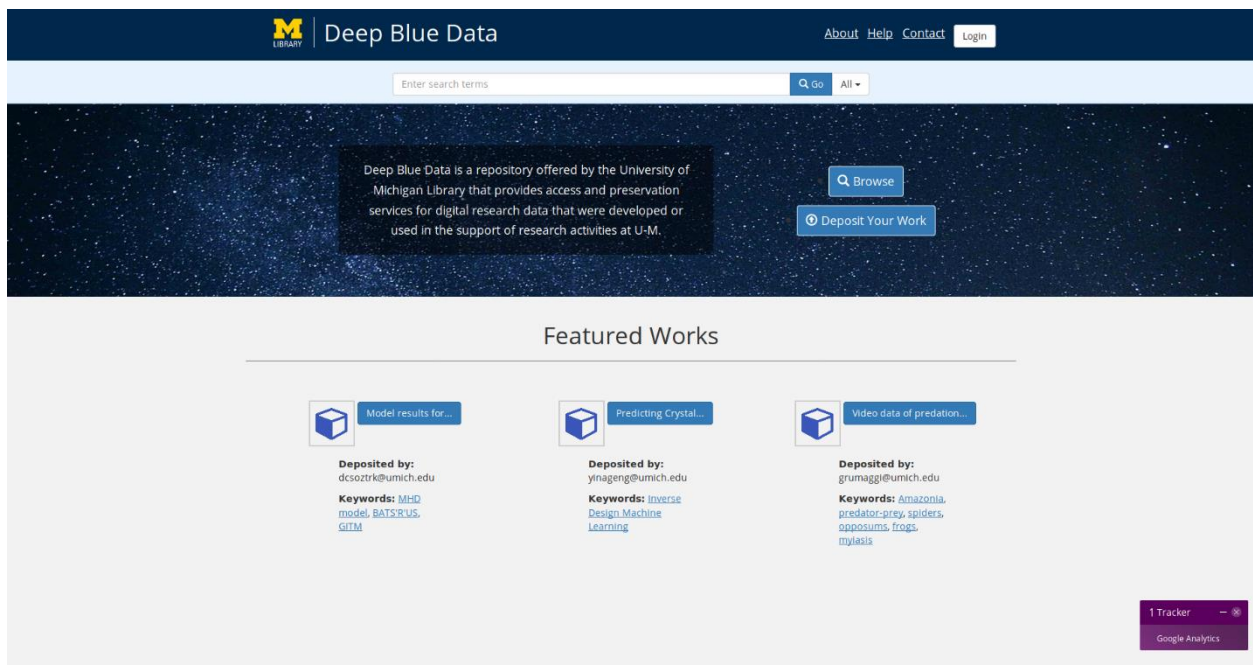
Fedora je skvelým riešením pre organizácie, ktoré potrebujú robustný, flexibilný a škálovateľný repozitárový systém pre spravovanie svojich digitálnych kolekcíí. Fedora aj s výhľadom do budúcnosti kladie dôraz na interoperabilitu a presadzovanie všeobecne prijímaných webových štandardov.

Je však dôležité zdôrazniť, že systém Fedora neponúka tzv. out of the box funkcionality, nakoľko neposkytuje potrebné užívateľské rozhranie – toto už je starosťou programátorov a správcov. Menšie organizácie možno preto radšej siahnu po jeho sesterskom projekte DSpace, aj za cenu nižšej flexibility, prípadne využijú front-end riešenia postavené na Fedore – ako sú napríklad Islandora alebo Samvera, donedávna nazývaná Hydra.

Obrázky



Obr. 1: AgriKnowledge (agriknowledge.org) – projekt Cornellovej univerzity



Obr. 2: Deep Blue Data (deepblue.lib.umich.edu/data) – projekt University of Michigan

Metadata v Dublin Core (DC)

<dc:title>FEDORA The Flexible, Modular, Open-Source Repository Platform</dc:title>

<dc:creator>Alžbeta Strakošová</dc:creator>

<dc:date>2018-11-29</dc:date>

<dc:type>Text</dc:type>

<dc:format>public</dc:format>

<dc:language>sk</dc:language>

Použitá literatúra

[1] Payette S., Lagoze C. (1998) Flexible and Extensible Digital Object and Repository Architecture (FEDORA). In: Nikolaou C., Stephanidis C. (eds) Research and Advanced Technology for Digital Libraries. ECDL 1998. Lecture Notes in Computer Science, vol 1513. Springer, Berlin, Heidelberg

[2] Payette S., Staples T. (2002) The Mellon Fedora Project Digital Library Architecture Meets XML and Web Services. In: Agosti M., Thanos C. (eds) Research and Advanced Technology for Digital Libraries. ECDL 2002. Lecture Notes in Computer Science, vol 2458. Springer, Berlin, Heidelberg

[3] Fedora – The Flexible, Modular, Open-Source Repository Platform. *Home – Duraspace.org* [online]. Copyright © 2018 Fedora [cit. 29.11.2018]. Dostupné z: <https://duraspace.org/fedora/>

[4] Home – Duraspace.org. *Home – Duraspace.org* [online]. Dostupné z: <https://duraspace.org/>

[5] Fedora 4.7.5 Documentation – Fedora 4.7.5 Documentation – DuraSpace Wiki. *Dashboard – DuraSpace Wiki* [online]. Dostupné z: <https://wiki.duraspace.org/display/FEDORA475/Fedora+4.7.5+Documentation>

[6] Fedora Repository Home – Fedora Repository – DuraSpace Wiki. *Dashboard – DuraSpace Wiki* [online]. Dostupné z: <https://wiki.duraspace.org/display/FF>