

Masarykova univerzita

Fakulta informatiky



# SAMVERA

*Open Source Repository Framework*

PV070 - Závěrečná esej

**Nicole Benčíková**

Brno, 29.11.2018

<b>1. O projekte</b>	<b>2</b>
<b>2. Ciele</b>	<b>3</b>
<b>3. História</b>	<b>3</b>
<b>4. Samvera Community</b>	<b>4</b>
<b>5. Komponenty</b>	<b>6</b>
<b>6. Balíky</b>	<b>7</b>
6.1. Avalon	7
6.2. Hyku	7
6.3. Hyrax	8
<b>7. Zhodnotenie projektu</b>	<b>8</b>

## 1. O projekte

Samvera je open source framework pre tvorbu digitálnych repozitárov poskytujúci flexibilné a bohaté užívateľské rozhranie, ktoré je jednoducho prispôsobiteľné. Pozostáva z niekoľkých stavebných blokov, ktoré sú zdarma dostupné a užívatelia z nich môžu v rôznych kombináciách vytvárať repozitáre na mieru. Tento prístup je postavený na hlavnej filozofii projektu, a síce, že nie je možné vytvoriť jeden systém, ktorý by poskytoval kompletný rozsah riešení pre repozitáre inštitúcií a zároveň nie je možné, aby jedna inštitúcia samostatne spravovala kompletný rozsah riešení pre svoje repozitáre<sup>1</sup>. Samvera je projekt postavený na skupinovej práci tvorený komunitou, ktorá je základným stavebným kameňom pri vývoji a správe.

## 2. Ciele

Jeden z hlavných cieľov Samvery a Samvera Community je skvelo vystihnutý už v samotnom názve, Samvera, čo v preklade znamená súdržnosť. Samvera sa snaží vybudovať nielen framework pre repozitáre, ale aj celú komunitu, ktorá tento systém udržiava a vďaka nej môže aj naďalej napredovať. Zároveň sa snaží vybudovať svoj systém dostatočne flexibilný a robustný na to, aby mohol slúžiť všetkým častiam svojej komunity, vrátane veľkých aj malých inštitúcií a ponúka univerzálne riešenie repozitárov na mieru, ktoré sa dokážu prispôbiť potrebám užívateľov. V Samvera Community sa usiluje o maximálnu otvorenosť a transparentnosť vo všetkých ohľadoch. Svoj systém chcú sprístupniť taktiež aj čo najväčšiemu množstvu nových užívateľov, vďaka čomu je Samvera aj všetky jej ďalšie plugin-y či balíčky open source.

## 3. História

Na konferencii Open Repositories v roku 2008 bolo sformované partnerstvo medzi univerzitami University of Hull, University of Virginia, Stanford University a Fedora Commons, dnes známe ako DuraSpace. Toto partnerstvo za svoj cieľ považovalo priniesť komplexné riešenie pre digitálne repozitáre, ktoré by bolo dostatočne flexibilné, aby

---

<sup>1</sup> <https://samvera.org/samvera-community-sourced-software/>

mohlo adresovať konkrétne potreby týchto univerzít, a zároveň dobre škálovateľné a udržiavateľné.<sup>2</sup> Neskôr sa k tomuto partnerstvu pripojilo aj MediaShelf.

Ako jednou z prvých možností sa javili repozitáre Fedora, ktoré poskytovali flexibilný framework a svojim užívateľom poskytovalo možnosť výstavby repozitárov na mieru na základe ich lokálnych požiadaviek. Táto výstavba však bola veľmi komplexná a náročná, nakoľko užívatelia museli repozitáre zostavovať prakticky od základov, a preto vznikla potreba tento proces zjednodušiť. Bola vytvorená iniciatíva Hydra, ktorej telo tvorili práve repozitáre Fedora, nad ktorými stáli jedna alebo viacero hydra heads, ktoré poskytovali workflow a boli určené na zjednodušenie údržby týchto repozitárov. Tieto “hlavy” potom mohli byť jednoducho oddelené a znovu používané viacerými inštitúciami. V roku 2011 University Hull spustila svoju prvú implementáciu, ktorá je prístupná verejnosti a funguje dodnes.<sup>3</sup>

Hydra svoj názov bola nútená zmeniť v roku 2017 a to kvôli problémom pri registrácii ochrannej známky, keďže existuje rovnomenná nemecká firma, ktorá si svoju ochrannú známku zaregistrovala už skôr. Názov Samvera bol zvolený komunitou, nakoľko najlepšie vystihuje filozofiu tejto komunity.

#### 4. Samvera Community

Pôvodné partnerstvo medzi University of Hull, University of Virginia, Stanford University a DuraSpace sa rozšírilo a v dnešnej dobe pozostáva z rozsiahlej komunity, ktorej organizácia v posledných mesiacoch tohto roku prešla niekoľkými zmenami a na ktorej sa momentálne stále pracuje. Aktuálna hierarchia tejto komunity je zobrazená na diagrame na nasledujúcej strane.

**Partneri** sú formálne zodpovední za vývoj Samvery a jej komunity a finančne tento projekt aj podporujú. Mimo samotného kódu a dizajnu, medzi ich povinnosti patrí napríklad organizácia rôznych konferencií a stretnutí, písanie dokumentácie, či tréning nových členov.

**Riadenie** pozostáva z pôvodných zakladateľov Hydry, ku ktorým sa neskôr pridali University of London, Northwest University, Indiana University, Emory University a University of Michigan. Riadenie formálne reprezentuje Samveru a je zodpovedné za právne a finančné záležitosti a komunikáciu medzi všetkými vrstvami komunity.

---

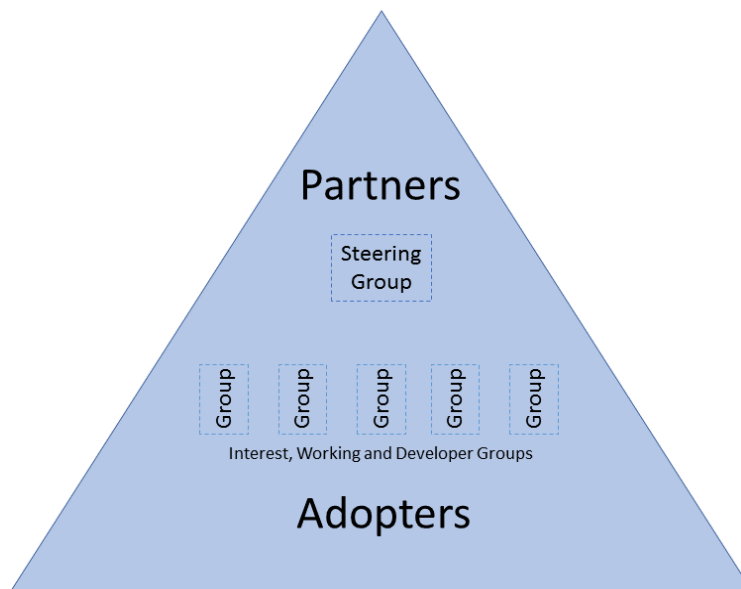
<sup>2</sup> <https://insights.uksg.org/articles/10.1629/uksg.383/>

<sup>3</sup> Hydra na University of Hull: <https://hydra.hull.ac.uk/>

V komunite ďalej existujú dva druhy skupín: **záujmové skupiny** a **pracovné skupiny** (skupiny vývojárov). Tieto skupiny sú ďalej vytvárané a sústreďujú sa na špecifické otázky a oblasti vývoju. Tieto skupiny medzi sebou úzko spolupracujú. Záujmové skupiny sa primárne sústreďujú na komunikáciu a diskusie. Pracovné skupiny zodpovedajú za vykonávanie úloh a kód. Vývojári musia povinne vlastniť Licenčnú Zmluvu Prispievateľa.

Kód vytvorený týmito skupinami ďalej prechádza peer review, než sa stane pevnou súčasťou systému. Je však vhodné spomenúť, že podľa skúseností používateľov táto peer review nie je dokonalá a v kóde sa nachádza veľa množstvo bug-ov, ktoré prácu komplikuje a je kameňom úrazu pre mnohých programátorov pracujúcich s týmto systémom.<sup>4 5</sup>

Mimo aktívnych súčastí táto komunita pozostáva aj z **pasívnych užívateľov**, ktorí Samveru a jej balíčky využívajú. V súčasnosti existuje vyše 70 inštitúcií, ktoré pre svoje repozitáre zvolili práve Samveru.



<sup>4</sup> On the present and future of samvera technical architectures. <https://bibwild.wordpress.com/2018/08/28/on-the-present-and-future-of-samvera-technical-architectures/>

<sup>5</sup> Are we still working on this? A meta-retrospective of a digital repository migration in the form of a classic Greek Tragedy. <https://journal.code4lib.org/articles/13581>

Obr. 1: Hierarchia Samvera Community

Nakoľko Samvera Community ale aj celá filozofia Samvery zakladajú na skupinovej práci a maximálnej transparentnosti, existuje hneď niekoľko spôsobov komunikácie medzi členmi. Medzi spôsoby online komunikácie patrí e-mailová komunikácia. Samvera udržuje mailing list pre svoju komunitu a zároveň aj separátne listy pre vývojárov. Ďalej využívajú aj Slack, LinkedIn či Twitter. Mimo textovej komunikácie sa bohato využívajú aj hovory na pravidelnú komunikáciu medzi jednotlivými skupinami. Všetky spôsoby komunikácie sú podrobne opísané na Samvera wiki<sup>6</sup>.

Samvera Community taktiež organizuje rôzne stretnutia a konferencie. Asi najvýznamnejšou konferenciou organizovanou partnermi Samvery je Samvera Connect, ktorá sa koná každoročne a trvá tri dni. Mimo zasadnutí a prezentácií s ňou sú súčasťou tejto konferencie aj rôzne workshopy. Samvera Connect sa snaží spojiť už existujúcu komunitu s jej novými členmi ale aj záujemcami, ktorí ešte nie sú jej súčasťou ale chcú sa dozvedieť viac. Ďalej Samvera ponúka napríklad regionálne stretnutia či 4-denný Samvera Camp, ktorý bol nedávno rozšírený o pokročilý Samvera Camp pre skúsenejších užívateľov.



Obr. 2: Samvera Connect 2018

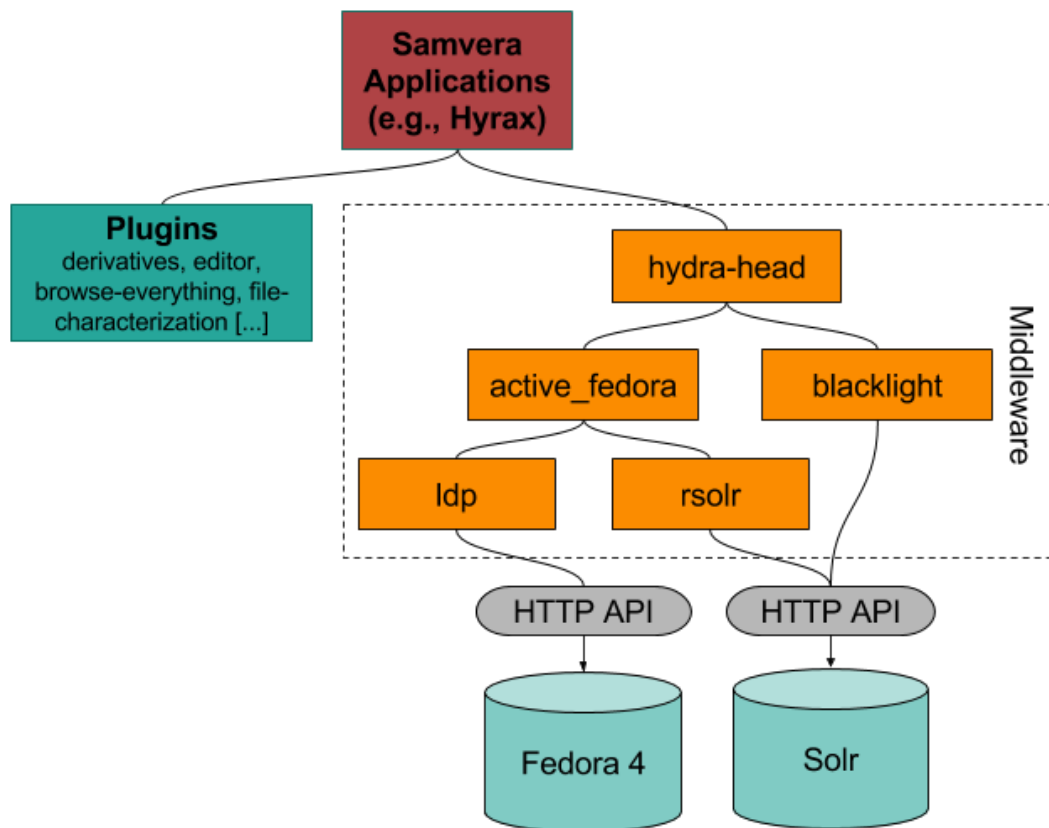
---

<sup>6</sup>Get in touch! - Samvera - DuraSpace Wiki.  
<https://wiki.duraspace.org/pages/viewpage.action?pageId=87460391>

Pre jednoduchý a plynulý prechod na Samveru a jej rozličné balíky ponúka komunita prechodové plány z iných repozitárov na Samveru. Zároveň však myslia aj na to, že táto komunita nemusí fungovať večne a taktiež aj niektorí členovia môžu prejsť na iné systémy a z tohto dôvodu ponúkajú rovnako aj plán pre prechod na iný systém.

## 5. Komponenty

Samvera framework je zložený zo 4 hlavných komponent: Fedora 4, Solr, Blacklight a Ruby gems. Tieto komponenty sú doplnené plugin-mi, ktoré pridávajú dodatočnú, užšie špecializovanú funkcionálnosť.



Obr. 3: Komponenty Samvery

**Fedora 4** tvorí robustnú vrstvu pre ukladanie digitálnych objektov a ich metadát. Fedora taktiež poskytuje aj určitú mieru abstrakciu, ktorá nám umožňuje repozitáre chrániť

pred prípadnými budúcimi zmenami štruktúry.<sup>7</sup> Aktuálna verzia využívaná v Samvera Framework je Fedora 4.7.0. a Samvera úzko spolupracuje s DuraSpace pri aktualizovaní svojho softvéru a príslušných gem-ov na túto verziu.<sup>8</sup> Prístup k Fedore je zaistený pomocou HTTP API.

Obsah z týchto repozitárov je indexovaný pomocou **Solr**, čo zabezpečuje rýchly prístup k informáciám. Indexácia obsahu z Fedora repozitárov sa rieši pomocou **Sorizer**. Interakcia so Solr je zaistená pomocou HTTP API.

Na funkcionality ORM slúži **ActiveFedora**, ktorej základ je postavený na ActiveRecord a prispôsobený na používanie v Samvere. V budúcnosti sa plánuje tento gem nahradiť novým gem-om **Valkyrie**, ktorého vývoj začal v roku 2017 a umožňuje mimo Fedory využívať napríklad aj Postgres.

**Blacklight** je Ruby on Rails knižnica poskytujúca rozhrania pre vyhľadávanie a discovery. Blacklight je možné zakomponovať priamo do projektu, alebo využiť samostatné Blacklight aplikácie.

Medzi ďalšie zaujímavé gem-y patrí **Opinionated Metadata**, ktorý slúži na prevod XML schém (napr. MODS, ktoré Samvera pôvodne plánovala využívať) na doménovo špecifický jazyk pre Samveru. Metadata je v rôznych implementáciách možné ukladať pomocou RDF. Ďalšími gem-ami používanými v Samvere sú **Rsolr**, Ruby klient pre Solr, či **Linked Data Platform (LPD)** implementujúci interakciu s Fedora 4.

## 6. Balíky

Nakoľko sa v minulosti výstavba repozitárov zo Samvery a jej spravovanie preukázali byť náročné a drahé, Samvera Community sa rozhodla zamerať na produkciu balíčkov, ktoré stavajú na tomto framework-u a poskytujú základnú funkcionality.

### 6.1. Avalon

Avalon je balíček postavený na Samvera Framework, pomocou ktorého je možné spravovať audiovizuálne kolekcie. Tento projekt zastrešujú knižnice Indiana University Bloomington a Northwest University a o jeho správu sa stará Avalon Community.

Avalon sa sústreďuje primárne na pomoc pri vytváraní repozitárov pre univerzity. Jeho prvá verzia vyšla už v roku 2013.

---

<sup>7</sup> <https://wiki.duraspace.org/display/samvera/Hydra+Stack+-+The+Hierarchy+of+Promises>

<sup>8</sup> Samvera 2018 Annual Report



## 6.2. Hyku

Hyku je výsledkom projektu Hydra-In-a-Box pod vedením The Digital Public Library of America, Stanford University a DuraSpace.

Jedným z hlavných cieľov Hyku je interoperabilita a zjednodušenie inštalácie, konfigurácie a správy repozitárov.<sup>9</sup> Poskytuje robustné a flexibilné jadro. Veľkou výhodou Hyku je možnosť repozitáre spúšťať nielen lokálne, ale aj na cloude.

## 6.3. Hyrax

Hyrax je Ruby on Rails gem poskytujúci plnú funkcionality Samvery s moderným užívateľským prostredím a dodatočnými vylepšeniami. Vznikol spojením svojich dvoch predchodcov Sufia a Curation Commons.

Mimo bežnej funkcionality Samvery poskytuje Hyrax napríklad aj správu užívateľov, batch-editing nástroje pre metadáta, Google Analytics, full-textové indexy, verzovanie a iné.<sup>10</sup>

## 7. Zhodnotenie projektu

Samvera prichádza s unikátnym riešením pre digitálne repozitáre a poskytuje svojim užívateľom flexibilné, škálovateľné a robustné jadro v podobe Fedora repozitárov na ktorých stavia a zároveň ponúka obsluhu týchto repozitárov spolu s ďalšou pridanou funkcionality a užívateľským rozhraním. Ďalej ponúka balíky pre jednoduchšiu inštaláciu a správu. Okolo tohto projektu stojí priateľská komunita so silným záujmom Samveru udržať a ďalej ju rozvíjať.

Avšak po podrobnom prehľadávaní zdrojov som sa stretla aj s množstvom negatívnych ohlasov, väčšina z ktorých sa týkala bug-ov a prehnane komplexného a komplikovaného zavádzania plugin-ov či zdĺhavého prechodu na samotnú Samveru alebo ňou ponúkané balíčky. Taktiež bola zmienená aj zlá organizácia komunity, ktorá aj na základe týchto podnetov prešla za posledný rok mnohými zmenami.

Verím, že Samvera má potenciál napredovať a vyriešiť v nasledujúcich rokoch tieto problémy. V súčasnosti je využívaná mnohými inštitúciami primárne v USA a UK, ktoré podporujú jej ďalší vývoj a systematicky pracujú na jej zdokonalení.

---

<sup>9</sup> <https://wiki.duraspace.org/pages/viewpage.action?pageId=85530575>

<sup>10</sup> <http://hyr.ax/about/#q3>

## Metadata

<dc:title>Samvera - Open Source Repository Framework</dc:title>

<dc:creator>Nicole Benčíková</dc:creator>

<dc:description>Samvera je open source framework na tvorbu digitálnych repozitárov. Svojim užívateľom ponúka tvorbu repozitárov na mieru a to prostredníctvom flexibilného jadra kombinovaného s veľkým množstvom verejne dostupných plugin-ov. Samvera je tvorená Samvera Community, ktorá sa spoločne stará o vývoj a správu Samvery a ďalších produktov z jej rady.</dc:description>

<dc:date>2018-11-30</dc:date>

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

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

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

## Zdroje

- [1] Samvera.org [online]. Dostupné z: <https://samvera.org/>, 2018. Navštívené 22.11.2018.
- [2] The Project | Avalon [online]. Dostupné z: <http://www.avalonmediasystem.org/project>, 2018. Navštívené 22.11.2018.
- [3] Hydra-In-a-Box [online]. Dostupné z: <http://hydrainabox.samvera.org/>, 2017. Navštívené 22.11.2018.
- [4] Hyrax: a community-supported repository front-end [online]. Dostupné z: <http://hyr.ax/about/>, 2018. Navštívené 24.11.2018.
- [5] Samvera Wiki [online]. Dostupné z: <https://wiki.duraspace.org/display/samvera/>, 2018. Navštívené 22.11.2018.
- [6] From Hydra to Samvera: an open source community journey [online]. Dostupné z: <https://insights.uksg.org/articles/10.1629/uksg.383/>, 2017. Navštívené 22.11.2018.
- [7] On the present and future of samvera technical architectures [online]. Dostupné z: <https://bibwild.wordpress.com/2018/08/28/on-the-present-and-future-of-samvera-technical-architectures/>, 2018. Navštívené 27.11.2018.
- [8] Are we still working on this? A meta-retrospective of a digital repository migration in the form of a classic Greek Tragedy (in extreme violation of Aristotelian Unity of Time) [online]. Dostupné z: <https://journal.code4lib.org/articles/13581>, 2018. Navštívené 28.11.2018.
- [9] Samvera 2017 Annual Report
- [10] Hydra 2016 Annual Report