Návrh a realizace distribuované personální databáze pro pro podporu systému evidence

Ing. Zbyněk Hubínka

Stávající situace

Cca 10 samostatných databází orientovaných na osobní data

Jedna osoba evidována několika způsoby

Změna příjmení znamená vznik nového záznamu

Osobní informace roztroušeny po různých tabulkách

Existující rozsáhlé, ale vzájemně málo kompatibilní databáze uchovávající data týchž osob, přičemž čerpání dat mimo individuální ruční vyhledávání je obtížné a zdlouhavé

Nemožnost propojit data do jednotného profilu

Nemožnost jednoznačné a úplné identifikace aktivit

Očekávaná konečná situace

Databáze unikátně klíčovaná

Databáze jednoduše, definovaným způsobem dostupná bez nutnosti paralelní úřední cesty (žádanky, potvrzení…)

Databáze s definovanými oprávněními k přístupu

Databáze otevřená (umožňuje-li to povaha poskytovaných dat)

Databáze integrovatelná do nových i stávajících systémů pracujících s osobními informacemi

Databáze splňující podmínky GDPR

Realizace

Jádrem jsou referenční data uchovávaná v LIANE (předpoklad: každý pracovník TUL má přinejmenším právo využívat služeb LIANE) – jméno, příjmení, e-mail

Další prvky základní struktury: národnost, jazyk – víc není třeba

Specifikující data: příznaky 1) aktivity (osoba je stále ve vztahu s TUL), 2) zařazení (akademik, THP, oboje); vlastník záznamu, tj. Osoba, která záznam pořídila

Unikátní klíč: celé kladné číslo rostoucí s každým dalším záznamem o 1 – vhodné jako prvek pseudonymizace podle GDPR; návazné systémy neoperují s osobou pod jménem a příjmením, ale pod číselným klíčem.

Pomocné záznamy

1) Záznam příslušníka TUL – definuje podrobnosti (telefon, kancelář, identifikátor ve STAGu… vesměs nepovinné, ale musí se vyskytovat právě jeden pro každého aktivního příslušníka

2) Záznam cizí osoby – definuje domovskou instituci, případně vazbu na další databáze. Záznamů může být více, podle počtu domovských institucí.

Během realizace se ukázalo jako výhodné připravit obecné schéma pro adresářovou databázi (LDAP), která by se mohla distribuovat mimo TUL, ale jednou klíčovou položkou by byla s popisovanou personální databází propojena – REALIZOVÁNO

Maticové záznamy

Propojení osoby a pracoviště – jednoznačně relace M:N, jeden člověk může být zaměstnán na více pracovištích a jedno pracoviště má více pracovníků

Definována nejnižší organizační jednotka – katedra, ústav, laboratoř, oddělení, a vyšší organizační jednotka – fakulta. Rektorát a CXI mají povahu fakulty, aniž by to bylo dále specifikováno. Pro studenty nezařazené je vytvořen speciální modul.

Každá aktivní osoba (pracovníci i studenti) je spojena s alespoň jednou nejnižší OJ, toto spojení má význam rozlišovací.

Autentizační model

Primárním autentizátorem je LIANE, sekundárním LDAP databáze podle definované struktury (může být i dislokovaná). Shibboleth v mnoha ohledech nevyhovuje – pouze pro webové aplikace, nemožnost obnovení sezení jinak než opakovaným přihlášením nebo reloadem okna prohlížeče.

Pro fakulty je k dispozici stínová databáze LDAP předávající autentizační informace LIANE – není nutné zpřístupňovat LIANE pro více než jedno zařízení na fakultu (technicky by mohlo být i jedno na univerzitu), přičemž lze autentizovat i osoby mimo TUL – realizováno v podobě informačního systému EF

Autentizace osob mimo TUL pomocí dislokovaných částí databáze (realizováno pro KNL)

Interakce

Data jsou primárně k dispozici prostřednictvím SQL dotazů, využívají je fakultní webové stránky (EF, FM, pravděpodobně FT – lze po dohodě s webmasterem TUL rozšířit i pro ostatní fakulty nebo katedry). Přímé propojení existuje s portálem Publikace a s právě spouštěným portálem Mobility, tyto systémy jsou dokonce oprávněny některá data přepisovat.

API – JSON pole pro webové vývojáře, přístup na vyžádání

Data z ostatních informačních systémů – nebylo pro ověření nutné, lze realizovat propojení se STAGem, pravděpodobmě i VEMA (bude nutné pro portál Mobility)

Externí identifikace, GDPR

Jako ideální vnější identifikátor byl vyhodnocen ORCID – může si jej zřídit každý (i neakademik), je přenosný, doživotní a obecně akceptovatelný. Bylo by vhodné v rámci TUL uvažovat o povinnosti tento identifikátor mít. Databáze se bez něj prozatím obejde.

GDPR – data propojující osoby s pracovištěm jsou uložena separátně a pseudonymizovaně (osoby jako čísla, pracoviště jako zkratky). Citlivá data jsou pouze rodná čísla, uchovávají se jen kvůli odesílání dat pro RIV v samostatné databázi, rovněž pseudonymizována.

Technické parametry

Samostatné zařízení povahy kancelářského PC s výhradou zajištění stálého provozu, databáze MariaDB a OpenLDAP, základní ovládací webové rozhraní (pro implementaci není nutné), API (realizováno vlastními silami). Původně navrhovaný minibox nezaručuje stálý provoz, nicméně jeden testovaný stále běží (majetek realizátora). Je třeba očekávat nárazově zvýšený počet přístupů v době odevzdávání dat do RIVu (případně jiných národních databází).

Závěr

Databáze běží, je využívána pro podporu evidence publikací a některých fakultních webů, identifikuje i osoby bez poměru ke škole, je propojena s evidencí dislokovanou pro KNL. Kromě nutnosti rozšíření o LDAP autentizaci k dalším změnám v projektu nedošlo.