HADMÉRNÖK

IX. Évfolyam 2. szám - 2014. június

FLEINER Rita
fleiner.rita@nik.uni-obuda.hu

KAPCSOLT NYÍLT KORMÁNYZATI ADATOK BIZTONSÁGA

Absztrakt

Az utóbbi években világszerte jellemző közigazgatási elv lett a Nyílt Kormányzati Adatok publikálásának és felhasználásának folyamata. A Nyílt Adatok egy speciális típusa a Kapcsolt Nyílt Adat, mely mögött álló technológia a különböző adatforrások összekapcsolását és hatékony integrációját segíti elő. Számos ország elektronikus közigazgatásában találunk használatára példákat. A Kapcsolt Nyílt Adat fogalma és az ehhez kötődő technológia jelenleg is folyamatos fejlődés alatt áll. A Kapcsolt Nyílt Adatok kezelése számos különböző technológia egymásra épülő alkalmazását jelenti, ezért a biztonsági rések, hiányosságok is változó helyeken lehetnek jelen és használhatóak ki rosszindulatú célból. A publikáció bemutatja a Kapcsolt Nyílt Kormányzati Adatok használatával kapcsolatos fogalmakat és technológiákat; feltárja a Kapcsolt Nyílt Adatok kezelésének különböző architektúráit és elemzi a Kapcsolt Nyílt Kormányzati Adatok használatának biztonságát és ennek kormányzati vetületét.

In recent years it has become a recommended principle worldwide to publish, integrate and reuse Open Government Data. Linked Open Data is a special type of Open Data, which uses a technology that enables the connections of different data sources and facilitates their effective integration. In many countries there are various examples for the use of Linked Open Data in e-government. The concept of Linked Open Data and its underlying technology is under constant development. The use of Linked Open Data is composed of various interrelated technologies; therefore security gaps and vulnerabilities may be present in various locations in the architecture and can be used for malicious purposes. The aim of the publication is to describe the concepts and technologies related to the use of Linked Open Government Data, to explore the different architectures of Linked Open Data management and to analyze the security of the use of Linked Open Data in governmental processes.

Kulcsszavak: *nyílt adat, kapcsolt nyílt adat, kapcsolt nyílt kormányzati adat, biztonság ~ open data, linked open data, linked open government data, security*

BEVEZETÉS

Az utóbbi évtizedben világszerte találkozhatunk kormányzatok és közigazgatási szervezetek által indított Nyílt Kormányzati Kezdeményezésekkel. Az Európai Unióban 2003-ban adták ki először a Nyílt Kormányzati Adatok támogatását szolgáló PSI (Public Sector Information) irányelvet [1], amit 2013 júniusában módosítottak [2]. A módosított irányelv már nemcsak támogatja a közszféra adatainak újrahasznosítását, hanem kimondja, hogy a tagállamok számára kötelező a nyilvános adatokat újrahasznosítható formában közzétenni.

A nyílt kormányzati adatok olyan nyilvános adatok, amelyet a kormányzat állít elő vagy gyűjt, gépileg olvashatóak, feldolgozatlanok, bárki számára ingyenesen hozzáférhetőek, lehetőleg nyílt formátumúak és nincs rajtuk szerzői jogi korlátozás. A nyílt kormányzati adatok újrahasznosításának lehetővé tételekor a kormányzati szervezeteknek továbbra is biztosítaniuk kell a magánadatok védelmét, a nem publikus adatok bizalmasságát és a nemzetbiztonságot.

Az utóbbi években számos ország kormányzata Nyílt Adatait Kapcsolt Adatok formájában kezdte el közzétenni annak érdekében, hogy elősegítse a heterogén és komplex struktúrájú közigazgatási adatok újrahasznosítását, összekötését és integrációját. Jelen publikáció alapvető célja a Kapcsolt Nyílt Kormányzati Adatok témakörének bemutatása és biztonsági aspektusainak feltárása. Ennek érdekében a publikáció:

- bemutatja a Kapcsolt Nyílt Kormányzati Adatok használatával kapcsolatos fogalmakat és technológiákat;
- feltárja a Kapcsolt Nyílt Adatok kezelésének különböző architektúráit;
- elemzi a Kapcsolt Nyílt Adatok használatának biztonságát és ennek kormányzati vetületét.

KAPCSOLT NYÍLT KORMÁNYZATI ADATOK

Nyílt Kormányzati Adatok

Az állam által előállított adatok, adatbázisok egy része komoly piaci értékkel is bír. Az adatok az újrafeldolgozás és többletfunkciók hozzáadása hatására új szolgáltatásként adhatóak el a piacon, ezáltal érték teremthető elő. A nyílt kormányzati adatok, azáltal, hogy a könnyen megtalálhatóak, hozzáférhetőek és felhasználhatóak, elősegítik a vállalkozásokat, fejlesztéseket és tudományos felfedezéseket, és új állások létrehozását támogatják.

Az 1. számú táblázat azt mutatja, hogy 17 EU tagállamnak van mára Nyílt Kormányzati Adatok portálja, aminek meglétet a PSI irányelv egyik alapvető következményének lehet tekinteni.

Mivel az irányelvet 24 hónapon belül minden tagállamnak be kell ültetnie a nemzeti jogrendjébe, várható, hogy a maradék 11 tagállam is a közeljövőben elindítja a Nyílt Kormányzati Adat portáljait. Ahogy a táblázat is sugallja, a Nyílt Kormányzati Adatok hasznosításának magyarországi, a gazdaságot és a hazai vállalkozásokat érintő hatása egyelőre nem vagy csak csekély mértékben kimutatható. Valószínűsíthető, hogy a közeljövőben ez a helyzet változni fog.

Az USA-ban 2009-ben indult el a Data.gov webhely, amely folyamatosan növekvő mértékben teszi hozzáférhetővé a kormányzati adatokat. Évtizedekkel ezelőtt az amerikai kormányzat szabadon elérhetővé tette az időjárási és a GPS adatokat. Az amerikai vállalkozók ezeknek az erőforrásoknak a segítségével navigációs rendszereket, időjárási előrejelző és figyelmeztető rendszereket, precíziós mezőgazdasági eszközöket és még nagyon sok más terméket és szolgáltatást hoztak létre, ezáltal, hozzájárultak a gazdasági növekedéshez és új munkahelyek teremtéséhez. [3]

Európai Unió tagállama	Nemzeti Kormányzati Nyílt Adat portál
Austria	data.gv.at
Belgium	data.belgium.be
Bulgaria	—
Croatia	—
Cyprus	—
Czech Republic	—
Denmark	digitaliser.dk
Estonia	www.opendata.ee
Finland	suomi.fi/suomifi/tyohuone/yhteiset_palvelut/avoin_data/
France	data.gouv.fr
Germany	govdata.de
Greece	geodata.gov.gr
Hungary	—
Ireland	www.opendata.ie
Italy	dati.gov.it
Latvia	—
Lithuania	—
Luxembourg	—
Malta	data.gov.mt
Netherlands	data.overheid.nl
Poland	—
Portugal	dados.gov.pt
Romania	—
Slovakia	data.gov.sk
Slovenia	—
Spain	datos.gob.es
Sweden	opengov.se
United Kingdom	data.gov.uk

1. táblázat. Nyílt Kormányzati Adatok portáljai az Európai Unióban

Kapcsolt Nyílt Adatok

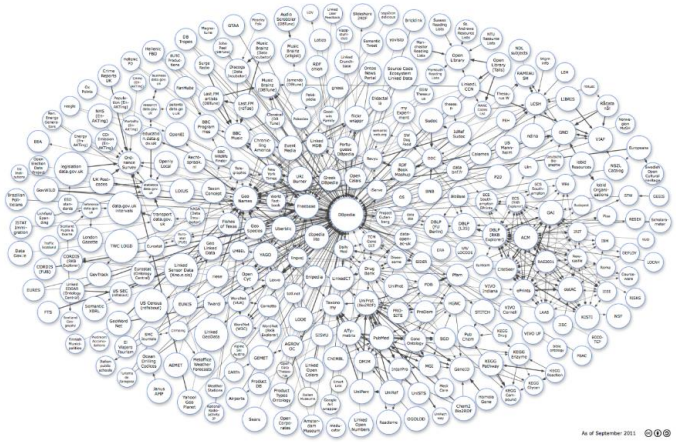
A Kapcsolt Nyílt Adatok (angolul Linked Open Data, rövidítve LOD) alapjait 2006-ban Tim Berners-Lee, a világháló feltalálója fektette le. A szemantikus web fogalma is az ő nevéhez köthető, mely a jövő világhálójának elképzelését jelenti. A szemantikus web célja a világhálón jelenlévő temérdek adat számítógép számára értelmezhető módon való közzététele. A szemantikus web víziójának megfelelően mára kezd gyakorlattá válni, hogy nagyszámú téma, szerteágazó tartalom esetében az adatokat összekapcsolt adatként teszik közzé. Az összekapcsolt adatok megkönnyítik a fejlesztők számára, hogy különböző forrásokból származó információkat összekapcsoljanak, ezáltal új és innovatív alkalmazásokat hozzanak létre.

A Kapcsolt Nyílt Adatok használatával strukturált adatokat a világhálón számítógép számára értelmezhető módon jelenítünk meg, az adatokat más adatcsoportokkal kapcsoljuk össze és lehetővé tesszük, hogy külső adatforrásokból az adatainkhoz kapcsolat legyen létesíthető. A Kapcsolt Nyílt Adatok négy alaptulajdonságát Tim Berners-Lee, a Web szülőatyja 2006-ban a következőképpen fogalmazta meg [4]:

1. Adatok elnevezésére és azonosítására URI-ket (Universal Resource Identifier) használjunk.
2. Feloldható HTTP URI-ket használjunk az adatok elnevezésére, így az adatokhoz tartalom társítható.
3. A URI mögötti tartalom szabványokon (RDF, SPARQL) alapuló hasznos információ legyen.

4. Az adatokhoz tartozó tartalom rendelkezzen más URI-khez mutató kapcsolatokkal, ezáltal más adatforrások információi is elérhetőek lesznek.

Az utolsó pont biztosítja azt, hogy a különböző forrású adathalmazok egymással összeköttetésben állhatnak és egymáshoz kapcsolódhatnak. Gyakorlatilag egy összefüggő hálózat alakul így ki, ez a Kapcsolt Nyílt Adatok Felhője, melyet a következő ábra szemléltet:



2. ábra: Kapcsolt Nyílt Adatok felhője [5]

A Kapcsolt Nyílt Adatok felhőjének középpontjában a DBpedia[1] adathalmaz áll. A DBpedia project célja a Wikipédia adataiból strukturált információkat kiemelve biztosítani az adatok hatékony visszakeresését és jobb felhasználhatóságát. A DBpedia lehetővé teszi strukturált lekérdezések végrehajtását a Wikipédia adatai felett és más, a weben elérhető adatkészletek összekapcsolását ezekkel az adatokkal.

Egy másik említésre méltó fejlesztés a Nyílt hozzáférésű adatok európai uniós portálja[2], mely egyablakos hozzáférést biztosít az Európai Unió intézményei és szervei által létrehozott és folyamatosan bővített adatbázisokhoz. A portál a különböző EU tagállami adatforrásokról és tartalmukról egy RDF adatbázist működtet, amiben keresni lehet az adatokra, illetve visszaadja a linket, ahonnan a keresett adat letölthető. A portál a tagállami adatkatalógusok metaadatait összekapcsolt adatként bocsájtja rendelkezésre. A portál üzemeltetői kinyilvánítják törekvésüket egyre több adatkészletet összekapcsolt adatként közzé tenni. Ez a példa is azt mutatja, hogy a Kapcsolt Nyílt Adatok jól használhatóak metaadatok közzétételére és az adatok közötti hatékony keresésre.

Kapcsolt Nyílt Adat szabványok

A Kapcsolt Nyílt Adatok egyik legnagyobb előnye, hogy azok a számítógép számára is értelmezhetőek. Az adatokat szabványosított technológiákon keresztül lehet elérni és módosítani. A már meglévő adathalmazok viszonylag könnyen bővíthetők és egy új adathalmaz könnyen összekapcsolható a már meglévőkkel.

A Kapcsolt Nyílt Adatok szabványosított adatmodellje az RDF (Resource Description Framework, Erőforrás Leíró Keretrendszer), ez egy gráf alapú adatmodell, mely lehetővé teszi az adatok strukturált ábrázolását és összekötését. Az RDF egy erőforrásokat leíró keretrendszer, mely alkalmas arra, hogy tetszőleges erőforrást metaadatokkal írjunk le, ahol az erőforrásokat http URI-kkel adjuk meg.

Az RDF segítségével kijelentéseket lehet tenni az erőforrásokról. Minden kijelentés három részből épül fel, ami <alany-állítmány/predikátum-tárgy> (angolul subject-predicate-object)

[1] http://dbpedia.org
[2] http:// http://open-data.europa.eu/hu

hármas (angolul triple). Az alany a leírandó erőforrást adja meg, az állítmány a leírandó erőforrás egy tulajdonságát, a tárgy pedig ennek a tulajdonságnak az értékét, ami vagy egy erőforrás vagy egy szöveges leírás.

Összetartozó RDF hármasok gyűjteménye egyértelműen ábrázolható irányított címkézett gráfként. Az alanyok és a tárgyak alkotják a gráf csúcsait. Két csúcs között címkével ellátott irányított él szerepel, ha van egy kijelentés, aminek egyik csúcs az alanya, a címke az állítmánya, a másik csúcs pedig a tárgya. RDF segítségével két különböző weben lévő adatforrás könnyen összeköthető egymással. Az RDF hármasokat három attribútumú relációs adattáblában tárolják, amelyben a három attribútum rendre alany–állítmány-tárgy. Az RDF adatbázisokat Triplestore-nak is nevezik.

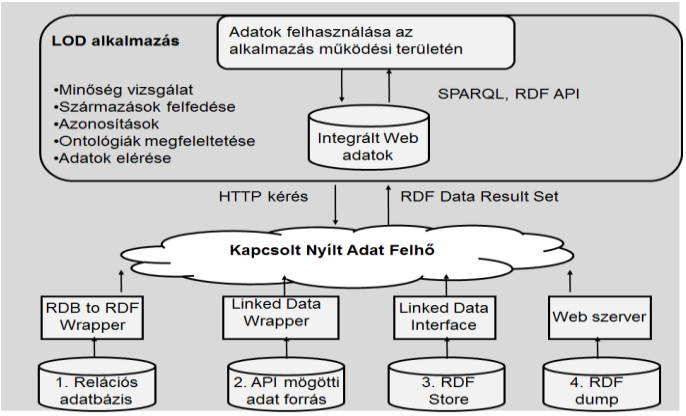
Annak megadására, hogy egy adathalmazban milyen típusú adatok lehetnek és köztük milyen kapcsolatok állhatnak fenn, szükséges egy absztrakt szint, ami minden esetben érvényes és általánosságban jellemzi adatainkat. Erre szolgál az ontológia, ami az adott területen (domain) értelmezett fogalmak osztályait, azok tulajdonságait és az osztályok közötti relációk formális reprezentációját írja le. Az ontológiával explicit módon megadjuk, hogy az adott területen belül milyen egyedeket, tulajdonságokat és kapcsolatokat használunk. Ontológia nyelv (például OWL vagy RDFs) az ontológiák kódolását lehetővé tevő formális nyelv.

A SPARQL az RDF lekérdező nyelve, mely RDF formátumú adatok elérését és manipulálását biztosítja. A SPARQL lekérdezések legtöbbször hármasok mintáit tartalmazzák, melyeket alap gráfmintának hívunk. Egy hármas minta az RDF hármashoz hasonlóan épül fel azzal a különbséggel, hogy az alany-állítmány-tárgy hármas közül bármelyik helyén állhat változó. A SPARQL lekérdezés eredménye vagy egy SQL eredményhalmazhoz hasonló táblázat vagy pedig egy újabb RDF gráf. SPARQL végpontnak nevezünk egy HTTP alapú lekérdezés szolgáltatást, mely SPARQL lekérdezéseket hajt végre Kapcsolt Adatok halmazán.

KAPCSOLT NYÍLT KORMÁNYZATI ADATOK ARCHITEKTÚRÁJA

Ahhoz, hogy a Kapcsolt Nyílt Adatok kormányzati alkalmazásának biztonsági aspektusait hatékonyan megvizsgálhassuk, első lépésben a biztonság alanyának elemzését végezzük el. Ennek érdekében áttekintjük a Kapcsolt Nyílt Adatok kezelésének lehetséges architektúráit, a szükséges rendszer elemeket és ezek viszonyát.

A következő ábra azt mutatja be, hogy a tetszőleges formátumban tárolt strukturált adatoktól hogyan jutunk el a Kapcsolt Nyílt Adatok felhőjéhez, aminek adattartalmát a megfelelő alkalmazások fel tudják dolgozni.



2. ábra: Kapcsolt Nyílt Adatok létrehozása és használata

A 2. ábrán legalul láthatóak az eredeti adatforrások, melyek relációs adatbázisban, tetszőleges API mögött lévő adattárolókban, illetve RDF formátumban lehetnek. Ezek az adatok megfelelő átalakítás után Kapcsolt Nyílt Adatként lesznek elérhetőek. Ha az adatokat feldolgozó alkalmazások felől nézzük a folyamatot, akkor ott az alkalmazásnak, esetleg egy előfeldolgozó egységnek a szükséges adatokat el kell érnie, a használatban lévő ontológiákat meg kell feleltetnie, a szükséges adat azonosításokat el kell végeznie, az adatok származását és egyéb minőségi vizsgálatokat el kell végeznie. Végül pedig SPARQL interfészen keresztül ki kell nyernie a célzottan szükséges adatokat, amiket már az alkalmazás a saját céljaira felhasználhat.

Észre kell vennünk a következőket. Kapcsolt Nyílt Adatok kezelésekor sokszor több, egymástól független adatforrásból érjük el az adatokat, melyek általában egyenként kisebb mennyiségű, strukturált adatot szolgáltatnak. Lekérdezések végrehajtása szükséges az összekapcsolt adatforrások felett és az adatforrások tartalma kívülről HTTP GET paranccsal kérhető le, amire RDF tartalmat adnak vissza.

Az adatok szempontjából tehát elosztott rendszerrel van dolgunk, amit három fő szempont szerint osztályozunk. (1) Az adattárolás szerint megkülönböztetjük a központi elemben összegyűjtött adat kezelést, az adatok szempontjából kizárólagosan a forrásokban történő tárolástól. (2) Az adatok helyének megtalálását segítő index struktúrák szempontjából megkülönböztetjük az indexek központi kialakítását és tárolását, a kizárólag adatforrásokban meglévő indexektől, (3) végül pedig vizsgálhatjuk, hogy az adatokat tároló csomópontok tudnak-e egymással együttműködni és kommunikálni. Ezeket a szempontokat vizsgálva 3 különböző struktúrát különböztethetünk meg: (1) a központi adattárház kialakítását, (2) az elosztott adat tárolást úgynevezett mediátor segítségével és végül (3) a tisztán P2P rendszereket központi elem nélkül.

A központi adattárház architektúra esetében az összes adatforrás tartalmának tárolása egy központi RDF adatbázisban történik, központi index struktúrával, az ontológiák központi megfeleltetésével. A központi elem feladata az adatok begyűjtése, kapcsolt adattá alakítása, index struktúrák készítése, az adatok érvényességének feltárása. Ez a központi elem SPARQL interfészen keresztül érhető el és kérdezhető le az alkalmazások felől. A módszer hátránya a jelentős előfeldolgozás, az adatok és az index folyamatos frissítésének szükségessége, rossz skálázhatóság, felesleges adatok tárolása és a folyamatos frissítés szükségessége. Előnye viszont, hogy a központi elem a munka lényeges részét elvégzi, ezért az alkalmazások felöli adat lekérdezés megválaszolása hatékony, a lekérdezést a központi adattárház válaszolja meg.

Az adatok elosztott tárolását mediátor segítségével megvalósító architektúra esetében létezik egy központi elem, ahol az adatok speciális részhalmazának (azonosítók, meta adatok, statisztikák, indexek, adattartalmak) tárolása valósul meg a lekérdezések optimalizálása és az adatok forrásának megtalálása céljából. Az adatok lekérdezése valós időben történik, a központi elem szolgáltatásai segítségével. Hátránya ennek a megoldásnak a lekérdezések előzőnél hosszabb és bonyolultabb végrehajtása, illetve az indexek szükséges karbantartása. Előny viszont az előzőnél kevesebb tárolási igény és a lekérdezések optimalizálásának és helyes megválaszolásának az esélye a központi index struktúra miatt. Ha az adat források képesek kommunikálni, részben mentesíthetik a mediátort a különböző feladataitól [6], [7].

A tisztán Peer-to-Peer (P2P) rendszerek egyenrangú résztvevők együttműködésén alapszanak. A rendszer végpontjai közvetlenül egymással kommunikálnak, központi kitüntetett csomópont nélkül. A P2P hálózatok a kliens-szerver kapcsolathoz képest jelentősen eltérő módon működnek: a szerepek nincsenek előre meghatározva; többnyire követelmény is, hogy az összes resztvevő képes legyen valamilyen erőforrást a rendszer egésze számára elérhetővé tenni viszonzásképp az általa igénybevett szolgáltatásokért. A szerző véleménye szerint ez a

típusú struktúra nem lesz jellemző a kormányzati adatokra épülő alkalmazások esetén, ugyanis a hatékonyság érdekében célszerű bizonyos feladatokat központilag elvégezni.

KAPCSOLT NYÍLT KORMÁNYZATI ADATOK BIZTONSÁGI ASPEKTUSAI

Kapcsolt Nyílt Adat fogalma és az ehhez kötődő technológia az utóbbi néhány évben alakult ki és folyamatos fejlődés alatt áll. Nemzeti szinten folyamatosan jelennek meg a kormányzati Nyílt Adat portálok és az itt publikált adatokon működő alkalmazások. A jelenleg folyó kutatásokban a fő hangsúly a használatot és az elterjedést gátoló problémák leküzdésén van, mint például az adatok hatékony felhasználását jelentősen gátoló egységesen használt metaadatok hiánya, a több forrásban lévő adatok egyidejű lekérdezésének módja, a SPARQL lekérdezések írásának felhasználó szintű támogatása vagy a Kapcsolt Nyílt Adatok böngészésének vizuális támogatása.

A Kapcsolt Nyílt Adatok használatának biztonsági kérdéseivel azonban a szerző véleménye szerint kevesen foglalkoznak. A Kapcsolt Nyílt Adatok technológiája az adatok kezelésének és tárolásának egy alapjaiban új módszerét jelenti. Fontos lenne a fejlődés elejétől fogva a figyelem középpontjába állítani a biztonsági kérdéseket is. A Kapcsolt Nyílt Adatok kezelése sokkal több különböző technológia egymásra épülő alkalmazását jelenti, összehasonlítva például a relációs adatbázisok esetével, ezért a biztonsági rések, hiányosságok is könnyebben és nagyobb valószínűséggel lesznek jelen és használhatóak ki rosszindulatú célból.

A következőkben a Kapcsolt Nyílt Adatok kezelésének biztonsági aspektusait tekintjük át. Ezen belül megvizsgáljuk a biztonság alanyát, ennek védendő tulajdonságait, illetve lehetséges fenyegetéseit. A biztonság alanyának kiindulásnak tekinthetjük a különböző adat forrásokat, illetve az adatok elérését biztosító informatikai rendszert. Előzőekben vizsgáltuk a Kapcsolt Nyílt Adatok kezelésének különböző architektúráit. Mindhárom típusra jellemző, hogy számtalan különféle formátumú adatforrás tartalmára épülhet. Az adatok hatékony integrációját és lekérdezését egyéb adatstruktúrák segítik elő, mint például cashe adatok, indexek, ontológiák, szótárak és névterek. Ezek védelme sokkal összetettebb és nehezebb feladat, mint pl. egy relációs adatbázis-kezelő rendszer védelme.

Hasonló mondható el az adatok kezelését megvalósító informatikai rendszerről is, mely számtalan részrendszerből áll össze. A teljesség igénye nélkül megemlítjük a különböző típusú forrásból származó adatokat RDF formátumra alakító wrappereket, a SPARQL lekérdezés feldolgozó rendszert, az adat indexelést létrehozó rendszert, stb.

A szemantikus web komponenseit [8] figyelembe véve a Kapcsolt Nyílt Adatok kezelésének biztonságát a következő ábrán szemléltetett rétegek szerint építhetjük fel:

3. ábra: Kapcsolt Nyílt Adatok biztonságának elemei

Ahhoz, hogy az Kapcsolt Nyílt Adatok kezelésének folyamata biztonságos legyen, a különböző technológia rétegek biztonságát garantálni kell. A 3. ábra szerint 5 szint különböztethető meg, az első a fizikai adatátvitel biztonságához kapcsolódik, a második az RDF adatok hordozó formátumának biztonságához, a harmadik magának az RDF adatmodellnek a biztonságához, a negyedik szinten található az ontológiák, különböző szótárak biztonsága, míg

legfelül a különböző adatforrásokból származó adatok összekapcsolásának, integrációjának biztonsága.

A Kapcsolt Nyílt Kormányzati Adatok biztonságának elemzésekor vizsgálni kell a három legfontosabb biztonsági tulajdonságot és ezek szerepét. Adatkezelési folyamatokban a bizalmasság annak biztosítását jelenti, hogy az adatok csak az arra jogosultak számára elérhetőek, a bizalmasság elvesztése az adatok illetéktelenek általi hozzáférését, megismerését jelenti. Mivel vizsgálatunk tárgyában nyílt, vagyis publikus adatok kezeléséről van szó, a bizalmasság kérdése legtöbbször nem játszik szerepet, tehát ennek biztosításáról nem kell gondoskodni.

A sértetlenség azt jelenti, hogy a tárolt adatot, illetve az azt kezelő rendszert csak az arra jogosultak változtathatják meg, azok észrevétlenül nem módosulhatnak és nem törölhetők. A Kapcsolt Nyílt Kormányzati Adatok szempontjából a sértetlenség biztosítása legtöbbször nem a kiemelt kategóriába tartozik, a kezelt adatok köre miatt. Ugyanakkor a sértetlenséget biztosítani kell a működőképesség, a rend, az államba vetett bizalom fenntartása érdekében. Felmerülhetnek azonban olyan felhasználási területek is (pl. vészhelyzeti kommunikáció kezelése), amikor kiemelt kategóriába eshet a szóban forgó biztonsági szint. Ha vannak alkalmazások, amelyek ezekre az adatforrásokra építenek, akkor magának az alkalmazásnak a biztonsági szintjét kell meghatározni és ez fog továbbgyűrűzni az adatok elvárt biztonsági szintjére. Érdemes szem előtt tartani, hogy azokat az adatokat is védeni kell, ezekre is értelmezni kell a védelmet, melyek segédeszközei a Kapcsolt Nyílt Adat kezelési technológiának (például ontológiák, névterek, indexek, cashe adatok).

A rendelkezésre állás annak biztosítása, hogy a felhatalmazott felhasználók hozzáférnek a szükséges adatokhoz. A Kapcsolt Nyílt Kormányzati Adatok szempontjából a rendelkezésre állás fontosságáról hasonlóan lehet gondolkodni, mint azt a sértetlenség esetében láttuk. A rendelkezésre állás megsértése azt jelenti, hogy az adatokhoz, illetve az azokat kezelő rendszerhez való hozzáférés egy adott időtartamra nézve megsérül, teljes mértékben megszűnik vagy tervezési, illetve megvalósítási hibákból kifolyólag létre sem jön a kívánt módon. A probléma létrejöhet támadások hatására, de lehet kivitelezési hiba is. A. Flemming és O. Hartig összegyűjtötték a Kapcsolt Adat források minőségi kritériumait [9], melynek elemzéséből kiderül, hogy az adatok tervezett elérésének is számtalan buktatója lehet a helytelen megvalósítási folyamat miatt. Fontos megvalósítási elemek a következők:

- Helyes URI tervezés és használat
- Belső és külső URI-k feloldhatóságának helyes megvalósítása
- Tartalom egyeztetés használata és ennek helyes szerver oldali beállítása
- SPARQL végpontok helyes formátumú feltüntetése
- SPARQL végpontok elérésének biztosítása
- RDF adat fájlok elérésének biztosítása

A Kapcsolt Nyílt Adatokra épülő rendszerek támadásainak elemzésekor a Kapcsolt Nyílt Adat speciális tulajdonságait kihasználó eseteket tekintjük át a [10] publikáció alapján. Természetesen az ismert, más informatikai rendszereknél is fennálló fenyegetéseket, mint például a hálózat támadásait, DoS támadásokat vizsgálatunk tárgyát képező rendszereknél is lehet alkalmazni.

A támadók célja lehet a Kapcsolt Nyílt Adatokat felhasználó alkalmazások megfertőzése, a világháló hamis és értéktelen adattal való elárasztása, illetve a legális tartalmak módosítása és eltüntetése. Céljuk lehet még a keresőmotorok által kiszámított rangsorok befolyásolása, illetve rosszindulatú adattartalmak észrevétlenségének elérése.

A támadás kivitelezése épülhet az adattartalom rosszindulatú módosítására, méghozzá a struktúra különböző pontjait érintve.

(1) RDF hármasokat lehet illegálisan törölni, illetve beszúrni vagy módosítani népszerű ontológiák vagy szótárak elemeit felhasználva (például rdfs:comment vagy dc:creator) az RDF predikátumok helyén. Ezáltal az adatok tartalma és megjelenítése változtatható meg. RDF hármasok manipulálásával el lehet érni különböző adatforrások közötti kapcsolatok módosítását. Az owl:sameAs predikátum használatával, illetve RDF hármasokban található URI-k módosításával téves entitás egyezéseket lehet elérni.

(2) Kapcsolt Nyílt Adat források meglévő szótárakat, ontológiákat használnak, melyeket a támadó meghamisíthat és megtévesztő, az eredetihez nagyon hasonló URI alatt közzétehet. Vannak olyan webhelyek (például prefix.cc vagy schemapedia.com), ahol különböző ontológiák leírása, jellemzése és elérhetősége található. Fejlesztők használják ezeket a helyeket a számukra ideális ontológia megtalálásához. Ha itt a támadó meghamisítja az ontológia URI azonosítóját, akkor a saját meghamisított ontológiájának használatát tudja elérni, amivel a működést jelentősen befolyásolni tudja.

(3) Kapcsolt Nyílt Adatok használatával különböző adatforrások integrációját lehet elérni. Egy adott adatforrásban más adatforrások elérését szabályozó kifejezések támadásával (például a VoID séma) a szükséges adatforrások felfedezése, elérése hiúsítható meg.

(4) Erőforrás hivatkozások feloldási mechanizmusának, illetve tartalom egyeztetési folyamat támadásával hamis tartalmak megjelenítését lehet elérni.

(5) Kapcsolt Nyílt Adatok lokális tárolása esetén az adatbegyűjtést végző motor (crawler) megfertőzésével el lehet érni hamisított adatforrások tartalmának a felhasználását.

(6) RDFa az a formátum, amivel HTML oldalakba lehet RDF tartalmat elhelyezni oly módon, hogy azt a kereső motorok felismerik. Weblapokba meghamisított RDFa metaadatok elhelyezésével a keresőmotorok által kialakított rangsort lehet illegálisan befolyásolni.

(7) Az SQL injekcióhoz hasonlóan a Kapcsolt Nyílt Adatok lekérdezése esetén is létezik kódszintű injektálás, amit SPARQL injekciónak nevezünk. Ez a támadás a lekérdezések machinálására épít, ami által adatmódosítást, beszúrást vagy törlést tud véghezvinni a támadó.

A fentiekben ismertetett a támadási vektorok is alátámasztják, hogy a Kapcsolt Nyílt Adatok kormányzati alkalmazása megkívánja, hogy a biztonsági követelményeket felmérjük, a veszélyekkel tisztában legyünk és a megfelelő védelmi intézkedéseket gyakoroljuk.

ÖSSZEGZÉS, KÖVETKEZTETÉSEK

Számos ország elektronikus közigazgatásában megtalálhatók kormányzati Nyílt Adat portálok és az itt publikált adatokon működő alkalmazások. A Nyílt Adatok egy speciális típusa a Kapcsolt Nyílt Adat, mely mögött álló technológia a különböző adatforrások összekapcsolását, hatékony integrációját és az adatok újrahasznosítását segíti elő. A Kapcsolt Nyílt Adat fogalma és az ehhez kötődő technológia jelenleg is folyamatos fejlődés alatt áll. A napjainkban folyó kutatásokban a fő hangsúly a használatot és az elterjedést nehezítő problémák leküzdésén van, a Kapcsolt Nyílt Adatok biztonsági kérdéseivel kevesen foglalkoznak. A Kapcsolt Nyílt Adatok kezelése számos különböző technológia egymásra épülő alkalmazását jelenti, ezért a biztonsági rések, hiányosságok is változó helyeken lehetnek jelen és használhatóak ki rosszindulatú célból.

A publikációban áttekintettük a Kapcsolt Nyílt Adatok kezelésének architektúráját, biztonsági aspektusait, meghatároztuk a biztonság alanyát, elemeztük a biztonsági tulajdonságok szerepét és bemutattuk a technológiára jellemző fenyegetéseket. Egy következő kutatás témája lehet a Kapcsolt Nyílt Kormányzati Adatok és az ún. Big Data jelenség kapcsolatának vizsgálata [11]. Ennek keretében a Big Data témakört érintő biztonsági kihívásokat [12] célszerű lenne összevetni a Kapcsolt Nyílt Adatok biztonsági kérdéseinek vizsgálatával.

Felhasznált irodalom

[1] Az Európai Parlament és a Tanács 2003/98/EK irányelve (2003. november 17.) a közszféra információinak további felhasználásáról.

[2] Az Európai Parlament 2013. június 13-i jogalkotási állásfoglalása a közszféra információinak további felhasználásáról szóló 2003/98/EK irányelv módosítására irányuló európai parlamenti és tanácsi irányelvről szóló javaslatról

[3] The White House Office of the Press Secretary: Executive Order -- Making Open and Machine Readable the New Default for Government Information, 2013. május 9. http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government (2014.02.10)

[4] T. Berners-Lee: Linked data-design issues (2006). http://www.w3.org/DesignIssues/LinkedData.html (2014.02.10)

[5] Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

[6] Androutsellis-Theotokis Stephanos, Diomidis Spinellis: A survey of peer-to-peer content distribution technologies. ACM Computing Surveys (CSUR) 36.4 (2004): 335-371.

[7] Görlitz Olaf, Steffen Staab: Federated data management and query optimization for linked open data. In: New Directions in Web Data Management 1. Springer Berlin Heidelberg, 2011. p. 109-137.

[8] Thuraisingham Bhavani: Security standards for the semantic web. Computer Standards & Interfaces, 2005, 27.3: 257-268.

[9] Hartig Olaf; Flemming A: Quality criteria for linked data sources. http://sourceforge.net/apps/mediawiki/trdf/index.php?title=Quality_Criteria_for_Linked_Data_sources (2014.02.10)

[10] Hasnain, A., Al-Bakri, M., Costabello, L., Cong, Z., Davis, I., & Heath, T. (2012, November): Spamming in Linked Data. In Third International Workshop on Consuming Linked Data (COLD2012).

[11] Joel Gurin: Big data and open data: what's what and why does it matter? Guardian Professional, 15 April 2014

[12] Top Ten Big Data Security and Privacy Challenges. Cloud Security Alliance, 2012: