

## Kulturális intézmények a szemantikus webben – az ALIADA szoftverben rejlő lehetőségek

### Egy az egyben

Cikkemet Kevin Kelly elmefuttatásának<sup>1</sup> összegzésével kezdem. Gondolatmenetét egy 2007-es TEDX előadáson ismertette. Meglátásai szerintem ma is helytállóak. Kelly előadásában kifejti, hogy a hálózat fejlődése – egy bizonyos szemszögből nézve – a következő szakaszokon ment keresztül: kezdetben a gépeket kapcsolták össze. Ez volt a Telnet és a Gopher korszaka, amikor a Telnet program segítségével egy nagyszámítógép termináljáról lehetett bejelentkezni egy másik, távoli nagyszámítógépre, és ott műveleteket lehetett végrehajtani. A Gopher pedig – leegyszerűsítve – a távoli számítógépen található információkhoz való hozzáférést könnyítette meg. Az egész technológia azon alapult, hogy a számítógépek csomagokat továbbítottak és osztottak meg. Minden számítógép részt vett ebben, függetlenül attól, hogy neki szólt-e a csomag vagy sem.

A fejlődés következő fázisában már a oldalakat kapcsoltak össze. A számítógépet mint egységet felváltotta az oldal mint egység. Ez a World Wide Web korszaka. Ezt mindenki ismeri, hiszen nagyrészt ma, 2015-ben is ebben a korszakban élünk. Ebben a világban mindenkinek meg kellett, meg kell osztania a neten azt, amije van honlapok és webes szolgáltatások (például online olvasói katalógus) formájában azért, hogy

---

<sup>1</sup> Kevin Kelly: The next 5,000 days of the web, TEDX, 2007 december:  
[http://www.ted.com/talks/kevin\\_kelly\\_on\\_the\\_next\\_5\\_000\\_days\\_of\\_the\\_web](http://www.ted.com/talks/kevin_kelly_on_the_next_5_000_days_of_the_web)

mások is csatlakozhassanak hozzá, és linket helyezhessenek el az oldalára. Ezáltal válunk, váltunk igazán a World Wide Web részévé.

A következő korszak Kelly 2007-es előadása szerint az adatok összekapcsolásának kora lesz. 2015-ben már azt mondhatjuk, hogy részben ebben a korban élünk. Ezt az időszakot szokás a szemantikus web korának is nevezni, melyben az adatokat **RDF állítások formájában** lehet közzétenni. Ebben a periódusban már adatokat kell közzélnünk magunkról, hogy más adatokkal interakcióba léphessenek. Minél több adatot közzlünk, annál többet vehetünk ki a rendszerből. Egy példával megvilágítva: minél több adatot közzlünk magunkról, annál több és testre szabottabb szolgáltatásban lehet részünk a szolgáltatók, a keresőmotorok, így a Google, a közösségi portálok, például a Facebook, vagy újabban a Windows 10 által. Előfordulhat, hogy valakinek nem éri meg a szolgáltatásokért fizetendő ár, azaz személyes adatainak, szokásainak a kiadása, de nyilvános közgyűjtemények esetében semmilyen tényező nem tarthatja vissza az intézményeket attól, hogy köztegyék a gyűjteményük egyes darabjait leíró metaadatokat.

Fontos megjegyeznünk, hogy az adatok összekapcsolása továbbra is (mint lentebb látni fogjuk az azonosítókkal kapcsolatban) internet technológián alapul, tehát nincs szükség új infrastruktúra kialakítására sem.

Az adatszintű összekapcsolásnak (több más dologgal együtt) van egy általános, messzire mutató következménye. Ahhoz, hogy az adatokat össze tudjuk kapcsolni, azonosítókra van szükség. Ez evidencia már a relációs adatbázisok korától, sőt még korábbról is. Egy ügyes és valójában nagyon egyszerű technikai fogás, az internetcím megadása (<http://horvadam.hu/horvadam#me>) és/vagy egy, a feloldási mechanizmusába bevezetett új konvenció révén – melynek a részleteibe most nem megyek bele – megeremtődött a lehetősége annak, hogy ne csak az interneten létező dolgoknak, hanem a való világ dolgainak is internetes azonosítót adjunk. Azaz az URL-hez hasonló, azzal szinte teljesen megegyező azonosítóval, az URI-val azonosítsuk a weben valójában kívül eső dolgokat, így például egy embert. Ez azzal a következménnyel jár, hogy a web ettől kezdve már nemcsak az információk hálózata, hanem a dolgok hálózata is, ez az az EGY valami, amiben minden feloldódik.

A folyamat kétirányú. Nemcsak a dolgok költöznek a webre, hanem a web is beköltözik a dolgokba: például a cipőbe és az autóba is, különféle chippek és szenzorok formájában. Így már a cipőre gondolhatunk úgy is (újra Kevin Kellytől véve a hasonlatot), mint egy chipre, aminek talpa van, az autóra pedig úgy, mint egy chipre, aminek kereke van, és mindkettő rajta van a weben. A web ma már elválaszthatatlan része életünknek. Még általánosabban fogalmazva, a digitális világ és a materiális világ egyesülni látszik EGY valamiben, és egyre inkább kézzelfoghatóvá válik, hogy világunk alapvető építőköve az információ.

Az előző bekezdésekben utaltam rá, de most külön is kiemelem, hogy Kelly szerint a hálózat fejlődésére az a jellemző, hogy egyre többet osztunk meg magunkról. Kezdetben a hálózatra kapcsolt számítógépeknek kellett csomagokat megosztaniuk más számítógépekkel, később már weblapokat osztottunk meg, ma pedig már adatokat. Kelly szerint minél többet osztunk meg magunkról, annál többet tudunk kivenni a rendszerből. Szerinte tehát ez egy pozitív folyamat, amelyből nyertesként jövünk ki a személyre szabott szolgáltatások révén. Ebből nem lehet kimaradni, mert egész egyszerűen a web lesz a világunk. Már nem csupán a fiatalok és a technika iránt érdeklődők „játsszótere”, hanem a valóság mindenki számára megkerülhetetlen szelete, amelyből a kulturális intézmények sem maradhatnak ki.

Az adatokra épülő szemantikus web építésében a múzeumoknak és könyvtáraknak az a feladatuk, hogy az rekordjaikat RDF állítások sorozatává (azaz adatokká) konvertálják, állításaikat összekössék a mások által közölt állításokkal, és elérhetővé tegyék bárki számára. Emellett fontos, hogy tudassák a világgal, hogy az adataik hozzáférhetővé váltak a szemantikus weben.

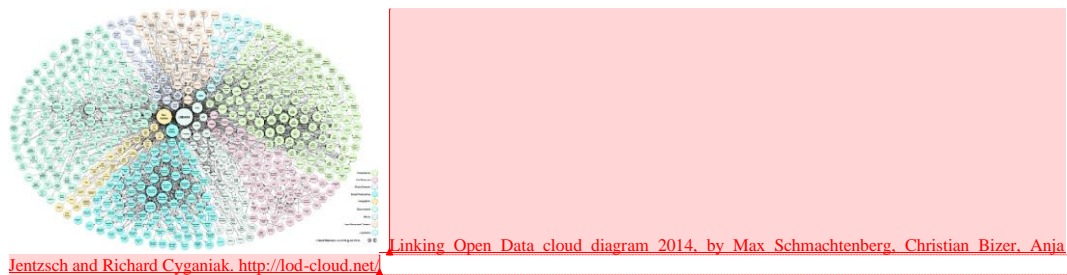
A címben említett ALIADA projekt célja egy olyan ingyenes szoftver fejlesztése, mely felgyorsítja a könyvtárak, múzeumok megjelenését a szemantikus weben, azért, hogy ezek az intézmények is részeivé válhassanak a fentebb – Kelly után szabadon és vázlatosan – felvázolt EGY-nek, és élvezhessék ennek előnyeit.

## Az ALIADA funkcionális leírása

A szemantikus weben való megjelenés egyik legnehezebb kérdése a konverzió, amit eddig mindenki, aki publikált a szemantikus weben, maga oldott meg valami módon, direkt erre a célra készített, saját

**Comentario [FL1]:** egy nagyon rövid leírás, h mi ez, 1-2 mondat lábjegyzetben?

fejlesztésű, (vagy megrendelt) speciális és egyedi szoftverek segítségével. Az ALIADA a szemantikus webben való megjelenés minden mozzanatát automatizálja. Inputként LIDO (múzeumi csereformátum), MARC (bibliográfiai és besorolási), valamint Dublin Core rekordokat fogad. Az inputot FRBRoo (és más, például WGS84, SKOS, SKOSXL, FOAF, DCTERMS, OWL-TIME) állításokká konvertálja, és elhelyezi egy mindenki által hozzáférhető adatbázisban (az úgynevezett „SPARQL végponton” keresztül). A SPARQL végponton rendelkező adatbázisba való betöltés jelenti azt, hogy az intézmény adatai megjelentek a szemantikus webben. A folyamat befejező mozzanata az, hogy az ALIADA szoftver automatikusan regisztrálja az így létrejött adathalmazt az Open Knowledge Foundation által létrehozott Datahub (<http://datahub.io>) adatbázisban. (Az ismert „LOD cloud” diagram ebből az adatbázisból készül.)



**Con formato:** Fuente: 8 pto

**Comentario [PK2]:** Laura felhívta a figyelmet a diagram megfelelő hivatkozására. A <http://lod-cloud.net/> honlapon ez szerepel.

**Con formato:** Fuente: 8 pto

Valójában ekkor tudatjuk a világgal, hogy megjelentünk a szemantikus webben. A konverzió fontos része, hogy a keletkezett állításokat az ALIADA szoftver – szintén automatikusan – összeköti más állításokkal: például a DBpedia (a Wikipedia szemantikus webes párja) és a VIAF (Virtual International Authority File) állításaival, ezáltal beleszöve a publikáló intézmény adatait a világméretű szemantikus adathálóba.

## A szemantikus webben való megjelenés tervezése

Az intézményben kell legalább egy ember, a rendszer-muzeológus/könyvtáros, aki tisztában van a szemantikus web alapjaival, és meg tudja tervezni az intézmény szemantikus webben való megjelenését. Először is ki kell találni, hogy az intézmény milyen adatokat akar közzé tenni a szemantikus webben most és a jövőben. Például egy múzeum, amelynek könyvtára is van, elhatározhatja, hogy mind a könyvtári, mind pedig a múzeumi leíró adatokat publikálni fogja. Azaz két adathalmazt (VOID nyelven [lásd alább] „dataset”-et) szeretne létrehozni.

A következő lépés az, hogy ki kell találni, hogy milyen URI-val akarja publikálni az adatokat. Ahhoz, hogy a rendszerkönyvtáros/muzeológus a tervezés eme szakaszát el tudja végezni, három dokumentumot kell ismernie: az egyik a Cool URIs for the Semantic Web (<http://www.w3.org/TR/cooluris/>), a másik a Describing Linked Datasets with the VoID Vocabulary (<http://www.w3.org/TR/void/>), a harmadik pedig a Designing URI Sets for the UK Public Sector (<http://www.cabinetoffice.gov.uk/sites/default/files/resources/designing-URI-sets-uk-public-sector.pdf>).

Tegyük fel, hogy a könyvtár – nagyon helyesen –, nemcsak leíró adatokat, hanem besorolási rekordokat is közzé szeretné tenni. Ezek lesznek a könyvtári adathalmaz alhalmazai („subsetjei”). A alhalmaz URI-jainak kialakításához tudnunk kell, hogy az ALIADA a „Designing URI Sets for the UK Public Sector” (<http://www.cabinetoffice.gov.uk/sites/default/files/resources/designing-URI-sets-uk-public-sector.pdf>) című dokumentumban felvázolt névkonvenciót követi. Az ALIADA felhasználók az ebben a dokumentumban lefektetett keretek között tudnak URI-kat kialakítani. Az idézett dokumentum alapján például ilyeneket lehet létrehozni a két „subset” részére: [data.library.szm.hu/id/bib/...](http://data.library.szm.hu/id/bib/...), illetve [data.library.szm.hu/id/auth/...](http://data.library.szm.hu/id/auth/...) (Az „id” szócska jelentését is az idézett dokumentumban találjuk meg.)

A tervezőnek természetesen ismernie kell a szemantikus web alapidokumentumát is. A legjobb, ha ennek az oldalnak az iránymutatásai alapján ássa bele magát a szabványba és a kísérő dokumentumokba: <http://www.w3.org/RDF/>.

## Mi végre?

A bevezetőből talán kiviláglott, hogy a szemantikus webben való megjelenés ma már egyfajta kötelezettsége a múzeumoknak és a könyvtáraknak. Nem elhanyagolható szempont, hogy részt vehetnek valami újnak, a

*Tim Berners Lee* által megálmodott szemantikus webnek a létrehozásában. (Tim Berners Lee álmolta meg a hagyományos webet is.)

A távoli és elvont célokon túl konkrét eredményeket is várhatunk a szemantikus webben való publikálástól. Az ALIADA nagyon hasznos lehet azon múzeumok számára, amelyek rendelkeznek könyvtárral is, mert használata révén létrejön annak az alapja, hogy mindkét gyűjtemény kereshetővé váljon azáltal, hogy állományuk adatait átkonvertálják FRBRoo állításokká, majd ezeket az állításokat kereshetővé teszik például a SolrRDF segítségével (<http://andregazzarini.blogspot.com.br/2014/12/a-solr-rdf-store-and-sparql-endpoint-in.html?m=1>).

További lehetőség, hogy az online katalógusban a felhasználók számára extra információkat lehessen nyújtani az ALIADA szoftver azon funkciója révén, hogy az a publikált adatokat összeköti más adathalmazok – például a már említett DBpedia – adataival. Ezeket az adatokat felhasználva kis fejlesztéssel arra is lehetőség nyílik majd például, hogy a hagyományos weben megjelenő katalógusban az alkotó nevével egyetlen kattintással el lehessen jutni az alkotó életrajzáig, vagy egy térképen meg lehessen mutatni a mű keletkezési helyét és egyéb földrajzi vonatkozásait.

Az ALIADA segítségével bárki, aki képes LIDO, MARC vagy DC metaadatokat előállítani az általa használt rendszerből, megjelenítheti az adatait a szemantikus webben. Ez azt jelenti, hogy a múzeumi HUNTEKA-t, a The Museum Systemet vagy a MuseumPlus használó múzeumok gond nélkül tudják alkalmazni az ALIADA-t.

## Végezetül

Az ALIADA nemcsak egy nyílt forráskódú szoftver, hanem egy, az Európai Unió által finanszírozott projekt neve is, melyhez a Szépművészeti Múzeum adatszolgáltatóként, valamint a szoftverrel szembeni elvárások kidolgozásával, teszteléssel és az eredmények széles körű megismertetésével járult hozzá. A projekt által fejlesztett ALIADA elérhető és telepíthető innen: <https://github.com/ALIADA/aliada-tool/>. A Szépművészeti Múzeum által publikált adatokat pedig innen kiindulva lehet elérni: <http://datahub.io/dataset/data-szepmuveszeti-hu>.