



## Kutatási adatok infrastruktúrája a Berli Műszaki Egyetemen

A kutatási adatok hosszú távú és jövőbemutató kezelésének biztosításához a *Berli Műszaki Egyetem* (TU Berlin) 2012 szeptemberében kezdte kiépíteni a kutatási adatok infrastruktúráját. A cél az, hogy létrehozzák az egyetemen keletkezett kutatási adatok, s az ezekre épülő publikációk hosszú távú megőrzését és rendelkezésre bocsátását. A szervezeti és technikai komponensek megvalósítása egy kétéves, a teljes Műszaki Egyetemre kiterjedő projektben valósul meg, amelyben az *Egyetemi Könyvtár* (UB), a *Számítóközpont* és a *Kutatási Részleg* dolgoznak együtt. A tanulmány bemutatja a koncepciót és áttekintést nyújt a projekt jelenlegi állapotáról.

### 1. Bevezetés

2012 szeptemberében kétéves saját projekt keretében a TU Berlin kiépítette a kutatási adatok infrastruktúráját, ezáltal biztosítani tudja az egyetemen keletkezett kutatási adatok és az ezekre épülő publikációk hosszú távú tárolását és hozzáférhetőségét.

A Berli Műszaki Egyetem nagy hagyománnyal rendelkező, elismert kutatóegyetem, amelyen folyamatosan számos kutatási projekt zajlik, akár harmadik féllel együttműködve, esetenként a TU Berlin finanszírozásában. Mindeztidáig nem létezett az egyetemen a kutatási adatok tárolására infrastruktúra, s fennállt a veszély, hogy a kutatási adatok elvesznek, amint véget érnek a projektek, illetve a kutatók más intézményekbe kerülnek át. A különböző tudományterületeken az intézményi kutatási adatok infrastruktúrájának igénye nagyon eltérő: az elmúlt években – általában adott tudományos szakmai társaságok kezdeményezésére – világszerte jelentek meg szaktudományi repozitóriumok és archívumok, amelyekben a TU Berlin kutatói is jelen vannak. Azonban nem minden tudományterületen léteznek repozitóriumok, pedig igény lenne rájuk a többi tudományterületet átfedő kutatások és a kisebb projektek számára is.

A kutatási aktivitás megnövekedése miatt 2011 őszén az Egyetemi Könyvtár (UB), a tubIT IT-Service GmbH (tubIT), valamint a Kutatási Részleg kidolgozták a kutatási adatok kezelésével kapcsolatos koncepciót, amit 2012 márciusában mutattak be. A szervezeti és technikai koncepció a TU Berlin számára egy terhelhető, megbízható és tartós struktúrát céloz meg, továbbá a kutatók számára a kutatási eredmények kezeléséhez szükséges hosszú távú, jövőbemutató szolgáltatásokat is biztosít. Az infrastruktúra a technikai komponens mellett a tartalom menedzsmentjét is tartalmazza, amely a kutatási-adatokat, leírásukat, megjelenítésüket és tárolásukat szervezi meg és szabályozza.

### 2. Szervezeti és technikai koncepció

#### 2.1 Alapvető irányelvek és stratégiák

A koncepció a kutatási adatok már meglévő és működő infrastruktúráját követi, mely leginkább Nagy-Britanniára jellemző. Technikai szempontból szabványprotokollok és nyílt interfészeket biztosítják a más rendszerekkel történő átjárhatóságot.

A TU Berlin három részlege (UB, tubIT és Kutatási Részleg) rendelkezik azzal a specifikus tudással és lényeges képességekkel, amelyek alkalmazhatók a kutatási adatok infrastruktúrájának kiépítéséhez és üzemeltetéséhez. Ezért nincs szükség egy újabb egység létrehozására, hanem egy olyan hatékony szervezeti struktúrát kell megalkotni, amely a három egység képességeit egy „virtuális” szolgáltatási centrumban fogja össze.

A koncepció a meglévő adatok használatát is támogatja, ezzel megkönnyítve a kutatók munkáját, egyúttal olyan többlétszolgáltatásokat hozva létre, amelyek ösztönzik a kifejlesztett eszközök használatát. Erre egy példa a projekt általános adatait (pl. projekt címe, absztrakt, részt vevő személyek stb.) tartalmazó adatmenedzsmentterv, melyhez az adatok a Kutatási Részleg elektronikus projektnyilvántartásából automatikusan átvehetők.

A kutatási tanácsadás a jövőben nagy hangsúlyt helyez a repozitóriumokról való tájékoztatásra és ajánlja használatukat. A koncepció figyelembe veszi, hogy a kutatók szorosan kapcsolódnak saját tudományterületükhöz, ezért a TU Berlin tagok kutatási eredményeinek nyilvántartása megosztottan történik: a kutatási eredmények egy része a TU Berlin intézményi repozitóriumába kerül, míg a másik része a világszerte fellelhető tudományterületi repozitóriumokba. A TU Berlin kutatóinak kutatási eredményei linkekkel és teljes szöveggel együtt szintén elérhetők: a metaadatok és a „Persistent Identifier” azonosítók automatizált workflow-n keresztül szabványprotokollokkal begyűjthetők, és a TU Berlin által üzemeltetett kutatási adatok és publikációk repozitóriumában tárolhatók, valamint az UB Primo tudományos portálján keresztül kutathatók.

## 2.2 A koncepció elfogadtatása a TU Berlinen

A TU Berlinen a kutatási adatok infrastruktúrájának elfogadtatása érdekében lényeges a kutatók és a kutatásért felelős részlegek körében a széles körű információmegosztás és tájékoztatás. A kutatásért felelős elnök, a karok kutatási dékánjai és a kutatási tanácsadás elfogadták a koncepciót. Ahhoz, hogy a kezdeményezést a TU Berlin megismerje, a „TU Intern” egyetemi lapban megjelent a „A kutatási adatok hosszabb életéért” című cikk. A kutatási adatok projekt egyetemen kívüli megismertetése 2013-ban a lipcsei könyvtári kongresszuson két előadással történt meg, melyek során a koncepciót, a projektet valamint a repozitóriumot mutatták be.

## 2.3 A „Kutatási adatok és publikációk szolgáltatócentruma”

A kutatási adatok infrastruktúrájának rögzítéséhez a TU Berlinen, az UB, tubIT és Kutatási Részleg egyesítették képességeiket a „Kutatási adatok és publikációk szolgáltatócentrumában” (SZF), és elkészült egy webes információs platform.

Az SZF a három részleg munkáját központosítja, s egy a TU-ra kiterjedő szolgáltatási ajánlatot kínál a kutatók számára. A Kutatási Részleg tájékoztatást nyújt a kutatási támogatások, projektkezdeményezések és egyéb szervezeti projektkérdések terén. A tubIT a technikai infrastruktúra rendelkezésre bocsátásáért és üzemeltetéséért felelős. Az UB a komplex metaadatok, a kutatási eredmények igazolásáért és a kutatási adatok menedzsmentjéért felel. Az SZF vezetését az UB látja el.

A TU Berlin újítása abból állt, hogy a kutatási adatok infrastruktúrájába bevonták a Kutatási Részleget. Míg a meglévő koncepciók a kutatási adatokra fókuszálnak és megkövetelik a könyvtár és a számítóközpont szoros együttműködését, addig a TU Berlin koncepciója már az adatok előállítására elvált feltűnik, vagyis már a projekt ötleténél, és a kutatók számára olyan „one-Stop-Service” valósul meg, amely a kezdetektől a kutatás teljes tartama alatt nyújt tanácsot és támogatást.

## 2.4 A kutatási adatok életrajza

A TU Berlin kutatási adat infrastruktúrája a kutatási adatok teljes életrajzát tartalmazza.

Az UB, a tubIT és a Kutatási Részleg a kutatási adatok életrajzából indulnak ki a munkamegosztás során. A Kutatási Részleg a projektekkel kapcsolatban ad tájékoztatást, az UB a kutatási adatok menedzsmentjéért, míg a tubIT az IT-infrastruktúráért felel. A három részleg mellett a kutatóknak is vannak feladataik, hiszen elő kell állítani a kutatási adatokat, leírják a projektet, megírják az annotációt, valamint kiválasztják a megőrzendő és törlendő kutatási adatokat.

## 2.5 A „DepositOnce” repozitórium

A TU Berlin kutatási adatok infrastruktúrája kezdetektől fogva olyan szolgáltatásokat és eszközöket bocsát rendelkezésre, amelyeket az egyetem összes tudományterületén lehet használni. Ezek egészülnek ki a tudományterületekre jellemző további (workflow, formátumok, metaadatsémák ...) szolgáltatásokkal.

A kutatásnyilvántartás technikai pillére az összes tudományterületre kiterjedő intézményi repozitórium, amelyben a kutatási eredmények, vagyis a kutatási adatok és a hozzájuk tartozó publikációk hosszú távon tárolhatók. A DepositOnce szolgáltatási platformként és többletszolgáltatásként működik, mely lehetővé teszi, hogy a kutatási adatok és publikációk a metaadatokkal együtt központosítva egyszerűen és tartósan legyenek tárolhatók, ezen kívül kapcsolódó szolgáltatásokat (honlap, Google Scholar, szakterületi repozitóriumok, az UB Primo Discovery & Delivery-rendszer stb.) is biztosít. Ebből adódóan a repozitórium nyílt platformként működik, amely támogatja a szabványokat és a szabványos hozzáférést más elismert repozitóriumokhoz, adatbázisokhoz és keresőrendszerekhez. Mind a publikációk, mind a kutatási adatok „Persistent Identifier” (DOI) azonosítóval vannak

ellátva, ezért állandóan elérhetők és hivatkozási lehetőségként szolgálnak.

A DepositOnce a TU identitásmenedzsmentjébe – s ezáltal a TU Berlin SSO-infrastruktúrájába került integrálásra. A hozzáférés- és jogosultságmenedzsment biztosítja a hozzáférési védelmet: míg a metaadatok az interneten keresztül nyíltan hozzáférhetők és ezzel teljesítik a nyílt hozzáférés kezdeményezés alapkövetelményeit, addig a kutatási adatok és a publikációk teljes szövege – a kutatók kérésének megfelelően – hozzáférési tilással ellátható vagy nyíltan hozzáférhető.

Az SZF munkamegosztásának megfelelően a repozitórium tartalmi koncepcióját az UB határozza meg és fejleszti. A TU Berlin IT-infrastruktúrájába történő integrálást és a technikai üzemeltetést, valamint a rendszeres biztonsági szolgáltatásokat a tubIT látja el.

### 3. Kutatási adatok projekt

A kutatási adatok projektet az UB, a tubIT és a Kutatási Részleg valósítja meg. Az UB a projekt indításakor egy informatikus állást hirdetett meg. A projekt az UB keretein belül a könyvtár vezetésével és az „AG Kutatási Adatok” projekt támogató-sával zajlik.

#### 3.1 Partnerekkel való együttműködés

A projekt a kutatók igényeinek figyelembevételéhez és a kutatási adatok infrastruktúrájának elismeréséhez a TU Berlinen partnerek együttműködésével valósult meg. Az első partnerek a 2012-ben engedélyezett 1029 „turbin” kutatás (képviselő: Prof. Dr.-Ing. R. King) és az audiokommunikációs kutatás (Prof. Dr. S. Weinzierl). A repozitórium struktúrájával, az adatmenedzsment tervének bevezetésével, a metaadatséma kérdésével és a „Persistent Identifier” azonosítóval kapcsolatban történt egyeztetés a partnerekkel, valamint prototípusokkal kerültek tesztelésre.

#### 3.2 Milyen kutatási adatokat tároljunk?

Nagyon sok hozadéka volt a partnerekkel történő egyeztetésnek a tárolandó kutatási adatok vonatkozásában. Az SFB mérnöktudományi projektek során előreláthatólag nagy adatmennyiség keletkezik, ezért az „adat” mint téma kiemelt szerepet kapott. Ahhoz, hogy meg lehessen becsülni az elvárt adatmennyiséget, a partnerekkel együtt kifejlesztettek egy kérdőívet a várható adatmennyiség-

gel és az adattípusokkal kapcsolatban. A válaszok a tubIT számára is segítséget nyújtanak.

A jövőben a kutatók számára, akik kutatási adataikat a repozitóriumba akarják feltölteni, felteszik a partnerekkel kifejlesztett kérdéseket, amelyeket a tudományterületek alapján kiegészítenek. A válaszból információt lehet szerezni a repozitórium tárolási igényéről és a szükséges bővítésekről.

Tervezik, hogy a kérdőívet webes alkalmazással fejlesszék, amit idővel minden egyes részt vevő tudományterület specifikusságával egészítenek ki.

Az alábbi felismerés vonható le a partnerekkel történt együttműködésből.

*A TU Berlin kutatási adatok és publikációk repozitóriuma a kutatási eredményeket tárolja, vagyis a kutatási adatokat és a hozzá tartozó publikációkat.*

Azokat a kutatási adatokat, amelyek a projekt mindennapjain nagyon nagy mennyiségben keletkeznek (pl. gépszimulációk), nem tárolják. A TU Berlinen a kutatási adatok kezelése érdekében sok tudományterület – legalábbis a természettudományok technikai területén és a mérnöktudományi szakterületeken – rendelkezik megoldásokkal és saját szerverrel, ahol a kísérleti adatokat tárolják, törlik, s újra létrehozzák.

A partnerekkel való együttműködés további fontos felismerése, hogy nem kell megőrizni azokat a kutatási adatokat, amelyek nem felelnek meg a DFG-definíciónak, amelyet a TUB-koncepció követ. A természettudományok technikai területén és a mérnöktudományi szakterületeken a kutatási adatoktól sokkal fontosabb a kísérlet felépítése és leírása (vázlatok, időterv, számítások, ...), amelyek alapján a tényleges kutatási adatok sokszor egyszerűen és gyorsan újra előállíthatók. E felismerés alapján a kutatási adatok projektben a megőrzésre vonatkozó definíciót megváltoztatták.

*A tudományterülettől függően a kutatási adatokat (a DFG-definíció alapján) és/vagy azokat az adatokat, amelyek az eredmény újbóli előállításához szükségesek, meg kell őrizni.*

Ezeket a felismeréseket, amelyek a szakterületek kis részére vonatkoznak, továbbiakban meg kell vitatni más szakterületek képviselőivel.

### **3.3 A technikailag önálló komponensek megvalósítása**

#### **3.3.1 Online információs platform (SZF-honlap)**

Az SZF-honlap teljes áttekintést nyújt a szolgáltatási centrumról, és a kutatási adatok infrastruktúrájával kapcsolatos szándékairól. A honlap 2013 májusától működik és a TU-honlapok legismertebb részéről is elérhető. A honlapok tartalmát az UB gondozza, s a Kutatási Részleg projektmunkatársai az egyes oldalak kutatási információiért felelősek.

#### **3.3.2 „DepositOnce” repozitórium**

A DepositOnce egy Dspace alapú szoftverrel működik, amely a legelterjedtebb nyílt forrású szoftverrendszerek egyike a kutatási adatok kezelésére. Az SZF munkamegosztásának megfelelően az UB munkája a repozitórium koncepciójának kidolgozása, valamint a kiegészítésekkel és a DSpace-nek való megfeleltetésekkel kapcsolatos programozói munkát jelenti. Ezzel párhuzamosan a tubIT a repozitóriumot beköti a TU Berlin infrastruktúrájába; a tubIT és az UB szorosan együttműködik. Ahhoz, hogy a fejlesztés zökkenőmentes legyen, az UB stratégiája az, hogy együttműködik a nemzetközi DSpace-fejlesztő közösséggel, miáltal a saját megfeleltetések és fejlesztések a hivatalos frissítésekben is megjelennek.

#### **3.3.3 Adatmenedzsmentterv**

Az adatmenedzsmentterv a kutatók számára eszközt nyújt a projektdokumentáció rögzítéséhez. A terv a WissGrid-projektben fejlesztett „kutatási adatok menedzsmentjének ellenőrző listáján” alapul, amely a TU Berlin igényeihez idomul. A programozási munkákat az UB végzi.

Az adatmenedzsmentterv olyan többletszolgáltatás, amely könnyen kezelhető webeszközként áll a kutatók rendelkezésére.

## **4 A fenntarthatóság lépései**

### **4.1 A repozitórium integrálása a TU Berlin IT-infrastruktúrájába**

A repozitórium integrálása a TU Berlin IT-infrastruktúrájába – szoros együttműködésben az UB-val és a tubIT-val – jelenleg is zajlik. 2013 júniusában a tubIT lezárta a DepositOnce szerverének beüzemelését és a TUB-Firewall által védett szer-

vert rendelkezésre bocsátotta a tubIT-környezetben. A szoftver integrálása majdnem lezárult, csak az adatbázis integrálása maradt hátra. 2013. augusztus vége óta a DepositOnce-ban tárolt kutatási adatok Backup-koncepciója zajlik. 2013 novemberében került beüzemeltetésre a repozitórium és az adatmenedzsmentterv.

Az integrálás fontos kritériuma, hogy a tubIT biztosítja a TU Berlin IT-infrastruktúrájáért biztonsági elvét, és az UB és a tubIT ennek megfelelő megoldásokat keres, és ezeket alkalmazza. Ezért az új szolgáltatás kezdetektől fogva minden szinten megfelel a TU Berlin magas minőségi követelményeinek.

## **5. A kutatási adatok infrastruktúrájának elfogadtatása a TU Berlinen**

Megállapítható, hogy a kutatási adat projektet és az SZF-t a TU Berlinen elfogadták. Ezt mutatja három kutató visszajelzése, akik több kutatási projektben is részt vettek. Az UB központi szerepet játszik az SZF-ben és a többi projektben. A részvétel célja a kutatási eredmények hosszú távú biztosítása, a keletkezett eszközök integrálása a TU Berlin kutatási adatok infrastruktúrájába.

A projektekben való részvétel stratégiai lépésnek is minősül, amely a jövőbe mutat. A projektkezdemenyezések meghatározása sok többletmunkával jár anélkül, hogy a kutatási adatok projektben azonnali haszon jelentkezne. A projektrészvételek „nyeresége”, amennyiben engedélyezik a projekteket, csak a következő években várható. Ez azért történik, mert a tudományos személyzetet és a szakmai tudást meg kell nyerni ahhoz, hogy a fejlesztések haladjanak.

## **6. Összefoglalás és előretekintés**

A két éves kutatási adatok projekt 2013. augusztus végére elérte az első félidőt, ezért ezt a tanulmányt köztes állapotfelmérésként kell tekinteni.

A projekt első feléből kiderül, hogy a kutatási adatok projektet a TU Berlin kutatói elfogadták. A TU-projektben való részvétellel rögzült a kutatási adatok infrastruktúrája és jelenléte a TU Berlinen.

Az egyetlen hiba a projektben ugyanaz, mint más projektek esetén is: a projektben részt vevő személyzet. A projektre fordított munkacapacitás hiányzik más területekről. Ezért a projekt főleg a

repozitórium esetén csúszott és az első verzió nem olyan működési terjedelemben valósul meg, ahogy eredetileg tervezték.

A projekt második felében be kell hozni az elmara-dást és új irányokat kell bevezetni. A repositórium technikai fejlesztése és a működésének kiterjesz-tése mellett ki kell dolgozni a „TU Berlin irányel-veit a kutatási adatok és publikációk terén”, majd bemutatni a TU Berlin grémiumának.

Összességében az első félidő eredményei sokat ígérnek. Még ha sok teendő is maradt, az eddig elérték pozitív irányba mutatnak és azt célozzák meg, hogy a projekt végére a TU Berlinen kiépül-jön a kutatási adatok infrastruktúrája a kutatók számára.

/KUBEREK, Monika: Die Forschungsdaten-Infra-struktur der TU Berlin. = Bibliotheksdienst, 47. köt. 11. sz. 2013. p. 833–846./

(Filó Krisztina)

## Obama információkat akar az amerikai IT-cégektől



Szorosabb együttműködést kért Barack Obama amerikai elnök a Szilíciumvölgyben működő cégektől a kiberbiztonsággal kapcsolatos információcserében.

Obama a *Stanford Egyetemen* vállalatvezetők előtt kijelentette, hogy hatékonyabb információcserére lenne szükség az olyan hackertámadások elleni védekezés-hez, mint amilyen a Sonyt érte a közelmúltban. „Ezt a kormány nem képes egyedül elvégezni. De tény az is,

hogy ezt a magánszektor sem képes egyedül elvégezni, mert az új fenyegetésekkel kapcsolatos legfrissebb információk a kormánynál vannak” – mondta. Az elnök szerint a kormány és a magánvállalatok együttműkö-dése és információcseréje az egyetlen módja annak, hogy Amerikát megvédjék a kibertámadásoktól.

Barack Obama külön is találkozott cégvezetők egy szűkebb csoportjával, hogy megvitassa velük, milyen üzleti veszteségeket okozott számukra a *Nemzetbiztonsági Ügynökség* (NSA) lelepleződött titkos adatgyű-jtési gyakorlata. Tiltakozásul amiatt, hogy az NSA-megfigyelés terén nem várható jelentősebb változás, a Google, a Facebook és a Yahoo vezérigazgatója távol maradt a pénteki tanácskozástól, amelyen viszont *Tim Cook* Apple-vezér és más vállalatok csúcsvezetői felszólaltak.

Az elnök pénteken rendeletet írt alá annak érdekében, hogy a cégeket arra bátorítsa: végezzenek fokozot-tabb adatcserét egymás között és a belbiztonsági minisztériummal a külön e célra létrehozott információ-csere- és elemzőközpontokon keresztül. Az egyre gyakoribb hackertámadások közepette az Obama-kormány fokozott jelentőséget tulajdonít a kiberbiztonság kérdésének. Lisa Monaco, Obama elnök terrorelhárítási és belbiztonsági tanácsadója kedden jelentette be, hogy egy új hírszerzési intézményt hoz-nak létre a kiberfenyegetésekre vonatkozó adatok rendszerezése és a kormányzaton belüli megosztására.

A *Kiberfenyegetettség Hírszerzés-integráló Központ* (Cyber Threat Intelligence Integration Center, CTIIC) feladata az lesz, hogy összesítse a *Központi Hírszerző Ügynökségtől* (CIA), a *Nemzetbiztonsági Ügynök-ségtől* (NSA) és más hasonló szervektől származó, az online fenyegetettségekre vonatkozó információkat, és azokat közölje más szövetségi intézményekkel. A CTIIC szervezetileg *James Clapper*, a nemzeti hír-szerzés igazgatójának (DNI) hivatala alá tartozik majd, amely már most is több titkosszolgálat munkáját hangolja össze.

/Forrás: <http://sg.hu/cikkek/110661/obama-informaciokat-akar-az-amerikai-it-cegektol/>

(B.Bné)