

Egy elképzelt, a bibliográfiai adattudományról szóló képzés vázlata¹

Király Péter

A bibliográfiai adattudomány egy viszonylag friss interdiszciplináris kutatási terület, amely a könyvtártudomány vagy tágasabban a közgyűjteményi tudományterületek, a történet- és társadalomtudomány, valamint az informatika bizonyos összetevőinek metszéspontján található. A bibliográfiai adattudomány célkitűzése, hogy (tipikusan, de nem kizárólagosan) nagyszámú, szerencsés esetben az adott témát érintő összes bibliográfiai rekordot tartalmazó adatforrásokra (pl. nemzeti bibliográfiákra) és az adattudomány módszertanára támaszkodva korábban rejtett vagy csak elszórt adatokkal igazolható, esetleg csak sejtett történeti, illetve gyűjteményi trendeket állapítson meg. Néhány példa a szakterület kutatási kérdéseire:

- Milyen volt a 17. századi német jogi disszertációk térbeli eloszlása és prozopográfiaja?²
- Milyen fokú interdiszciplinaritás figyelhető meg filozófiai disszertációk metaadatai alapján?³

¹ A vázlat a TOLONEN et al. 2023-ban kifejtett munkafolyamat továbbgondolása. Köszönettel tartozom a témáról való együttgondolkodásért Maróthy Szilviának (független kutató), Kokas Károlynak (Szegedi Tudományegyetem), Káldos Jánosnak (Országos Széchényi Könyvtár) és Thomas Wallningnak (Universität Wien), valamint Julia Damerow-nak (Arizona State University) a tanácsaiért és a DAMEROW, NELSON, HERNANDEZ 2025-ért.

² HESSBRÜGGEN-WALTER 2025.

³ HESSBRÜGGEN-WALTER 2024.

- Különböző régiókban hogyan változott a könyv formátuma és nyelve az időben?⁴
- Milyen az adott nyelvről készült fordítások mintázata, hogyan változott, melyek voltak adott korban a szupercentrális, a centrális és a periférikus nyelvek?⁵
- Milyen hatása van a kiadóknak a szépirodalom egészére?⁶
- Milyen profilja volt különféle könyvgyűjteményeknek?
- Van-e a könyv tartalma és formátuma között összefüggés?⁷
- Hogyan változtak a műfaji arányok?⁸
- Mennyi nyom nélkül elpusztult régi kiadvány lehetett?⁹
- Hogyan vizsgálható a művek befogadása bibliográfiai adatok segítségével?¹⁰
- Milyen a közgyűjteményi adatok minősége, és milyen javítási stratégiákat lehet kidolgozni?¹¹
- Hogyan segítik (vagy nehezítik) a közgyűjteményi adatok, adatszerkezetek és szabványok a fenti kérdések megválaszolását, milyen fejlesztési lehetőségeket javasol a kutatás a közgyűjteményei szabványosítás során?¹²

Bár a digitális bölcsészet oktatása az utóbbi években dinamikusan fejlődött, ezen belül a bibliográfiai források számítógépes elemzése sajnos alig jelenik meg, ahogyan ehhez hasonlóan hiányzik a könyvtártudományi és az informatikai oktatásból is. Ezt a hiányt javaslatom szerint egy új informális szakképzéssel lehetne orvosolni, amely azokat szólítja meg, akik érdeklődnek a fenti kérdések némelyike iránt, és akiknek már van az érintett tudományterületek valamelyikén (pl. könyvtártudomány, művelődéstörténet, irodalomszociológia, informatika) megszerzett ismerete. A könyvtári bibliográfiai szabványokra épülő rekordok elemzése valószínűleg a könyvtári képzésben sem lenne érdektelen. A képzés formája lehet akár nyári egyetem vagy több egyetemi szak által közösen indított szeminárium/tanfolyam. A képzés résztvevői lehetnének egyetemisták, illetve praktizáló szakemberek is.

⁴ LAHTI et al. 2019.

⁵ HEILBRON 1999.

⁶ BOURDIEU 2008.

⁷ LAHTI et al. 2019.

⁸ KIRÁLY, KISÉRY 2025.

⁹ FARKAS, KÁLDOS, KIRÁLY 2025.

¹⁰ SZEMES, DOBÁS 2025.

¹¹ KIRÁLY 2019.

¹² KIRÁLY et al. 2025.

A tananyag rövid vázlata:

- 1) A bibliográfiai adattudomány célkitűzése
- 2) Elméleti modellek
 - a) Tudományos elméletek és modellek (Darnton és követői, Bourdieu, a kvantitatív történetírás, a történeti események informatikai modelljei – Thibodeau és Thaller)
 - b) Közgyűteményi adatmodellek (a mű-kifejezési mód-manifesztáció-példány modell és elágazásai, ontológiák, levéltári adatmodellek)
- 3) Az adatelemzés munkafolyamata
 - a) Az adatok beszerzése. A bibliográfiai adatok és adatforrások főbb típusai (könyvtári katalógusok, tudományos publikációk adatbázisai, kutatásiadat-repozitóriumok, történeti források). Az adatok beszerzésének módszerei (szabványok, könyvtárakban használatos programozói interfészek és eszközök), tudnivalók az adatszerkezetről (metaadatsémák és szerializációs formátumok), a felhasználás feltételei.
 - b) Az adatok érvényességének vizsgálata. Műszaki érvényesség (XML-, JSON-érvényességvizsgálat) és az adattartalom minőségellenőrzése.
 - c) Előfeldolgozás. Formátumok, adatstruktúrák, konvertálás és az adatvesztés ellenőrzése.
 - d) Az adatok harmonizálása (normalizálás és adatgazdagítás). Az adatok reprodukálható módon történő átalakítása kvantitatív bölcsészeti elemzésre alkalmas adathalmazzá.
 - e) Adatelemzés és adatvizualizáció programozás (Python, R) és céleszközök segítségével.
 - f) Az eredmények közreadása. Szoftver és kutatási adat publikálása az újra-felhasználhatóságra is figyelemmel.
- 4) A kutatás után: a tágabb kontextus. Szakmai közösségek, konferenciák, folyóiratok, továbbképzési lehetőségek.

Jelen tanulmány mindebből pusztán az adatelemzés munkafolyamatának vázlatára szorítkozik.

A tananyag súlyát az adatelemzési munkafolyamatról szóló rész alkotja. A munkafolyamatban hat egymásra épülő szakaszt különböztetünk meg: az adatok beszerzését, előfeldolgozását, az adatharmonizációt, az elemzést, a validációt és végül a disszeminációt. A fenti szakaszok nem feltétlenül jelennek meg minden kutatás során. Az észet nemzeti könyvtár leírta azon adatpublikációs munkafolyamatát,¹³ amelynek a célja, hogy a kutatás számára elérhetővé te-

¹³ KRUSMAA, TINITS, NEMVALTS 2025.

gyék az észti nemzeti bibliográfiát. Az adatokat a katalógus OAI-PMH-felületén keresztül szerzik be, majd a MARCXML-formátumot táblázatos formátummá alakítják, amelyet külső adatforrások segítségével harmonizálnak és adatgazdagítanak, majd a folyamatot dokumentálva publikálják. Ebből a folyamatból maga az elemzés hiányzik, ezt – reményeik szerint – majd a kutatók fogják elvégezni, és nem a könyvtár. A rendelkezésre álló adatok és módszerek azonban visszahatnak, behatárolják a kutatási kérdéseket. Például ha a fordításokat szeretnénk tanulmányozni, de az adatbázisunkból hiányoznak a fordítókra vonatkozó adatok vagy a fordítás meglétére utaló tulajdonságok, akkor más adatforrásokhoz kell fordulni – feltéve, hogy adott időszakra és nyelvpárra ez rendelkezésre áll. Hasonlóképpen: a tipográfiai sajátosságok (szedéstükör, hasábok száma és mérete, betűtípusok) elemzéséhez rendelkezniünk kell az oldalak digitális fényképével, de ha csak a teljes szövegű állományhoz férünk hozzá, módosítanunk kell a kérdésfelvetést. Bizonyos kérdések nagyon komoly vagy specializált számítógépes háttérrel igényelnek. Ezek a feltételek nem mindig derülnek ki a kutatás elején. Fontos, hogy a kutató hívja fel az adatgazda közgyűjtemény figyelmét az általa tapasztalt – az adatokkal és az adatszerkezetekkel kapcsolatos – problémákra:¹⁴ a párbeszéd segít megérteni az adatokkal kapcsolatos elvárások különbözőségét, és hosszabb távon elvezethet a közgyűjteményi adatok szélesebb körű használatához.

3. a) Digitális adatokat három fő módon lehet beszerezni. A legkényelmesebb az, ha ezek letölthető fájlként (például újrafelhasználható kutatási adatként) rendelkezésre állnak. A QA Catalogue fejlesztése közben gyűjtöm és megosztom az ilyen forrásokról szóló információkat,¹⁵ de ez a fajta adatmegosztás viszonylag ritka. Gyakoribb, hogy az adatforrásokat valamilyen programozói interfészen keresztül lehet elérni. A leggyakoribb interfészekhez (OAI-PMH, Z39.50, SRW/SRU, SPARQL) különféle alkalmazások is elérhetőek, nem feltétlenül kell programozni, ugyanakkor időt kell szánni az interfészek intézményspecifikus beállításainak és paramétereinek tanulmányozására. Végül gyakori eset, hogy egyetlen korábbi lehetőség nem adott. Ekkor a weboldal HTML-forrásából nyerjük ki az adatokat, feltételezve, hogy a tipográfiai formázások következesen jelölnek bizonyos szemantikus elemeket¹⁶ – de ez esetben érdemes konzultálni a weboldal üzemeltetőjével, hogy nincs-e más,

¹⁴ Emellett az ad hoc problémabejelentés mellett a bibliográfiai adattudományban jelen vannak az adatok minőségére és az adatszerkezetekre összpontosító kutatások is.

¹⁵ <https://pkiraly.github.io/qa-catalogue/where-can-i-get-MARC-records.html> (Megtekintés: 2026. március 26.)

¹⁶ KIRÁLY 2025.

az oldalon nem dokumentált lehetőség. Bármelyik megoldást választjuk, győződjünk meg, hogy az adatok licensze megengedi-e az újrafelhasználást. Az adatok letöltése után az első lépés azok beolvasása. Különböző bibliográfiai formátumok esetében elérhetőek ezeket támogató programkönyvtárak, a MARC21 esetében például Java, Python, Go, JavaScript, R, PHP és más programozási nyelvekre. A könyvtörténeti kutatások esetében négy nagy adatforrástípussal számolhatunk: könyvtári katalógusokkal (például nemzeti könyvtárak vagy általános közös katalógusok mellett olyan, kimondottan régi könyves katalógusokkal is, mint amilyen a XVI., XVII. és XVIII. századi német nyelvű könyveket regisztráló VD16-, VD17- és VD18-sorozat, ennek olasz párja az EDIT16-, a Heritage of the Printed Book adatbázis), digitális könyvtárak katalógusaival (Europeana, Gallica, Német Digitális Könyvtár, Magyar Elektronikus Könyvtár, HathiTrust), szakirodalmi adatbázisokkal és kutatási adatrepozitóriumokkal (DataCite, Zenodo, OpenAlex, Open Citation), végül digitalizált (könyv)történeti források adatbázisaival (a Soci t  Typographique de Neuch tel adatb zisa, a XVIII. századi holland aukci s könyvkatalógusok adatb zisa, a MEDIANE, vagy az Eruditio).

3. b) Az adatok szab lyosság t  s minőség t k t m don szok s ellen rizni: l teznek ugyan az adatok strukt r j t le r  s m k, illetve olyan programok, amelyekkel ellen rizni lehet, hogy a dokumentum megfelel-e ezen s m knak, ugyanakkor a legt bbsz r ezek a s m k kiz r lag egy  ltal nos strukt ra le r s ra korl toz dnak (p ld ul egy MARC-rekordban kontroll-  s adatmez k található, az ut bbiban indik torok  s almez k lehetnek), vagyis csak az adatok legk ls bb r teg t  rintik.  rdemes tehát tov bbi ellen r z st v gezni – vagy adott form tum esetében rendelkez sre  ll  c lszoftverrel, vagy az  gynevezett felt r  adatelemz s (Exploratory Data Analysis) m dszertan val. A legegyszer bb az  gynevezett teljess gvizsg lat, mellyel azt n zz k meg, hogy az adatb zisban milyen adatelemek  s milyen ar nyban találhatóak.¹⁷ Az elemz sek t rj t k pez  fontosabb adatelemek tartalm t is  rdemes megvizsg lni, formai  s tartalmi szempontb l mennyire egys gesek: ugyanaz a szem ly vagy f ldrajzi n v vajon h ny alakban fordul el , vagy hogyan r gzített k a d tumokat. Adott adatelemben el fordul   rt kekr l k szített gyakoris gi list t b ng szve k pet kapunk az adatok természet r l  s az feldolgoz s tov bbi l p seiben k vetend  harmoniz ci s feladatokr l. Az ilyen lista lehet k dolt, p ld ul a Harald Klinka a Museum of Modern Art adatb zisában található d tumokat form tummint k szerint csoportosította (n gy s m, n gy s m-n gy s m, n gy s m-k t

¹⁷ KRUSMAA, TINITS, NEMVALTS 2025.

szám stb.), így a sok egyéni dátum helyett egy jobban áttekinthető mintalistát kapott.¹⁸

3. c) Az előfeldolgozás során a beolvasott fájlokat olyan adatstruktúrává alakítjuk, melyek bemenetként alkalmasabban a szokásos adatelemzési módszerekkel való feldolgozásra (a Python esetében a legelterjedtebb a Pandas, az R esetében a Tibble „adatkeret”). Előfordulhat, hogy nem minden adatot transzformálunk, hanem csak bizonyos rekordokat (például egy nemzeti bibliográfiából csak a XVII. századi könyveket) vagy bizonyos adatelemeket (például kihagyjuk a könyvtári azonosítókat és más adminisztratív adatelemeket).

3. d) A mások által gondozott adatok ritkán felelnek meg minden szempontból pont annak az elemzési célnak, amelyre készülünk, az adatok számunkra fontos körét harmonizálni kell, azaz normalizálni (egységesíteni, az ellentmondásokat feloldani, bizonyos adattípusokat – például szöveges változókat numerikussá – konvertálni) és adatgazdagítani (derivált adatokat, például oldalszámokat számítani, külső adatforrásokból adatokat importálni). Az alábbiakban négy ilyen harmonizálási lépést vizsgálunk meg: a dátumok, helynevek, személyek és fogalmak harmonizálását. A dátumok nem csak a MoMe-gyűjteményben mutatnak nagyfokú változatosságot, szinte bármelyik könyvtári katalógusban találunk olyan dátumokat, amelyek eltérnek a programok számára könnyen kezelhető formátumtól. Például római számokkal („MDCCLXXX. [1780]”) vagy szöveges formában („druk janvier 2016.”), uralkodói évek szerint („Meiji 40”) vagy más időszámítás szerint megadott dátumokat tartalmaznak. További problémát jelent a bizonytalan dátumok kezelése (pl. „18--” és „18uu” a könyvtári katalógusokban egyaránt azt jelenti, hogy XIX. századi kiadásról van szó). A változatosság miatt az átalakítás nem triviális, de az sem, hogy végül is mire alakítsuk a dátumokat. Különbőféle megközelítések léteznek egyes részterületeken (lásd például a levéltári szabványokat vagy az Europeana gyakorlatát). Legfrissebb javaslatként a DHTech-közösség által létrehozott *undate* Python programkönyvtár¹⁹ a következő adatelemeket tárolja: a dátum forrásban szereplő változatlan formáját, az időszámítási rendszert, a dátum pontosságát, a legkorábbi és a legkésőbbi normalizált dátumot és az időtartamot – vagyis itt a dátum az egységes visszakeresés érdekében minden esetben egy időtartomány. A helynevek esetén külső helynévtárak állnak rendelkezésre az azonosításhoz és esetlegesen a térképi ábrázoláshoz vagy a nyelvi változatok megjelenítéséhez

¹⁸ <https://x.com/HxxxKxxx/status/1066805548866289664> (Megtekintés: 2026. március 12.)

¹⁹ KOESER et al. 2025.

szükséges adatok letöltéséhez. A legfontosabbak között említhetjük a CERL Thesaurust,²⁰ a Getty Thesaurus of Geographic Names²¹ és a Geonamest,²² melyeket API-kon keresztül lehet lekérdezni. Bár ezek gazdag, sok forrásból és alapos ellenőrzéssel épített adatbázisok, a gyakorlat azt mutatja, hogy szinte minden bibliográfiai forrásban fogunk találni olyan névalakot, melyeket ezen szolgáltatások nem ismernek fel, ezért ezeket némi nem automatizált kézi adatfinomítással lehet beépíteni a saját adatbázisunkba. A személyek esetében ugyanezt az eljárást lehet követni, de természetesen más szolgáltatásokat igénybe véve: VIAF (Virtual International Authority File),²³ a CERL Thesaurus²⁴ személynévanyaga, ISNI (International Standard Name Identifier),²⁵ Wikidata.²⁶ Fontos megjegyezni, hogy egy tetszőleges adatbázisban természetesen jóval több személynév lehet, mint földrajzi név, ezért a találati arány valószínűleg kisebb is lesz. A fogalmak világa sokkal változatosabb, mint a földrajzi és személyneveké. Bár vannak univerzális fogalmi szótárak (tudásszervező rendszerek), szinte nincs olyan könyvtári katalógus, amelynek a rekordjaiban kizárólag egyetlen szótár fogalmait találnánk. Konkrét szótárak helyett a BARTOC nevű szolgáltatás (Basic Register of Thesauri, Ontologies and Classification)²⁷ használatát javasoljuk a kutatási kérdéseknek leginkább megfelelő szolgáltatás megtalálására. A harmonizálással kapcsolatban elengedhetetlen a pontatlan, hiányos, szubjektív és bizonytalan adatok kategóriáit megemlíteni.²⁸ A pontatlan adatra és kezelésére a dátum esetében láttunk példát. Hiányos adat az, amikor nem ismerjük az összes részletet, például nincs feltüntetve egy mű összes szerzője, vagy egy tárgy birtoklástörténetében lyukak vannak. Néhány adatot ki tudunk ugyan következtetni, ám nagyon nehéz a nem létezőket leírni. A szubjektív adat a provenienciára utal, vagyis hogy kié a szóban forgó állítás. Az ilyen állítások sokszor hipotetikusak és akár ellentmondók is lehetnek. Végül bizonytalan adat az, amikor az állítás igazságtartalma kétséges. A korábban idézett elméletek egyik fontos része, hogy a múlt konstruált, a források értelmezése az értelmező korábbi ismereteitől is függ, ebből következően a történelmi információs rendszerekben szükségszerűen meg kell engedni az egymásnak ellentmondó interpretációk egymás melletti szereplését, továbbá a

²⁰ <https://data.cerl.org/thesaurus/>

²¹ <https://www.getty.edu/research/tools/vocabularies/tgn/index.html>

²² <https://geonames.org>

²³ <https://viaf.org/>

²⁴ <https://data.cerl.org/thesaurus/>

²⁵ <https://isni.org/>

²⁶ <https://wikidata.org>

²⁷ <https://bartoc.org>

²⁸ MARIANI 2023.

kétértékű (igaz-hamis) logika helyett a bizonytalanságokat valószínűségi értékekkel lehetne leírni.²⁹ Például „Alexandre Dumas” (ha más információ nem áll rendelkezésre) feloldható az idősebb és az ifjabb íróra is – az előbbire nagyobb valószínűséggel, aminek az értékét az adatbázisban rögzíthetjük, és például a keresési találatok sorba rendezése során használhatjuk.

3. e) A munkafolyamat leglátványosabb fázisa az adatelemzés.³⁰ Itt a kutató a saját kutatási kérdéseit fordítja le a felhasznált eszközök kínálta műveletekre. Az elemzésekben kétfajta adatot használnak, egyfelől az adatharmonizációs lépések után rendelkezésre álló eredeti vagy gazdagított adatokat (cím, kiadási hely, nyelv, oldalszám), másfelől olyan számított adatokat, melyek ezek elemzésére épülnek (többek között betűtípus és -méret, a nyomtatvány készítéséhez felhasznált papír mennyisége, szavak száma, egy lakosra jutó nyomtatott szavak száma). Az adatelemzési módszerek a bibliográfiai adattudományra nézve nem egyediek, ezek olyan általános módszerek, melyekre rendelkezésre állnak általános és (kisebb mértékben) speciálisan bölcsészeti adattudományi tananyagok,³¹ valamint a kvantitatív történettudomány tankönyvei.³² A tananyag természetesen nem ölelhet fel minden szóba jöhető adattudományi fogást, de egy-egy példán keresztül ismertetnie kell a szakág eddigi története során leggyakrabban előforduló eljárásokat (a statisztika alapjai, idősoros elemzések, adatvizualizáció beleértve a térképi ábrázolást, szövegelemzés, hálózatelemzés).

3. f) A munkafolyamat utolsó lépése az eredmények terjesztése, amelybe beleértendők a hagyományos publikálási módok (tanulmány, könyv, konferencia-előadás) mellett újabb megközelítések, úgymint a folyamatban használt szoftverek, illetve generált adatok, valamint a kimondottan ezekre fókuszáló adat- és szoftvertanulmányok publikálása, blogolás és mikroblogolás, a prezentációs diák és felvételek közreadása, valamint a szakmai szervezetekben való részvétel. Külön ki szeretném hangsúlyozni az adatok speciális megosztásának a jelentőségét. Képzeljük el a következő esetet: egy kutató sok munkával adatgazdagított egy népszerű, sok kutatási potenciállal rendelkező, közgyűjtemény által gondozott adatforrást. Később egy másik kutató is szeretné bevonni a ku-

²⁹ THALLER 2020.

³⁰ Lásd egyebek mellett a kutatási kérdések apropóján idézett szakirodalmat.

³¹ Néhány ezek közül: KARSDORP, KESTEMONT, RIDDELL 2021; KLINKE 2025; TAYLOR, TILTON 2024; BATORFY 2024; valamint a *Programming Historian* oldal: <https://programminghistorian.org>

³² LEMERCIER, ZALC 2019; FEINSTEIN, THOMAS 2002; HUDSON, ISHIZU 2016.

tatásába ugyanazt az adatbázist. Ha nem tud az előző kutató munkájáról, akkor az adatgazdagítást ő is kezdheti előlről. De még ha az első kutató publikálta is a maga adatgazdagítását, sokkal valószínűbb, hogy a következő kutatók az eredeti adatbázist fogják megtalálni. Ezt megelőzendő szükséges lenne, hogy a kutatók az általuk módosított adatokat visszajuttassák az eredeti adatforrásba. Szerencsére a MARC21-ben a 2022-es, 34. módosítással³³ bevezetett (és mostanra a mezők többségében elérhető) adatproveniencia almező segítségével megkülönböztethető a könyvtár, illetve a kutató által rögzített adat, ami elvi gyógyír lehet a könyvtár jogos igényére, hogy felelősséget vállaljon a saját adataiért. A kutató tehát az élettudományokban a laboratóriumi mérések publikálásának mintájára nanopublikációk segítségével egy-egy adatgazdagítási lépést meg tud osztani a könyvtárral, amely ezeket be tudja építeni a katalógusba anélkül, hogy saját felelősségvállalása és így hitelessége sérülne, a második kutató pedig már az adatgazdagított változaton dolgozhat. A vízió megvalósításához természetesen a felek közti kommunikáció szabványosítására is szükség van, amiben a kutatói közösség játszhatja a koordinátori szerepet.

Technikai módszertan. A tananyag metodológiailag elég széles területet ölel fel, amelynek nincs egyetlen szakértője, ezért a részterületek szakértőinek összefogására van szükség. Az írásához egy olyan platform használatát javaslom, amely lehetővé teszi a közös szerkesztést. Más, hasonló kezdeményezésekhez (DHTech, Programming Historian stb.) hasonlóan a végső fázisban az anyagokat célszerű Git-alapú platformra és ott kezelhető formátumba (pl. markdown) áthelyezni és automatizálni a publikációs formátumokba (HTML, PDF, epub) való átalakítás folyamatát. Az elkészült tananyagrészeket archiválni kell olyan repozitóriumban, ahol biztosított a változatok kezelése, és állandó azonosító (DOI) is kapnak. A programozási gyakorlatokhoz a böngészőben szerkeszthető és futtatható Jupyter notebookokat is létrehozunk, és azokat egy végrehajtást lehetővé tevő platformon helyezük el.

Célközönség. A tananyag olyan hallgatók és gyakorló szakemberek számára készül, akik szeretnék megismerkedni a könyvtörténet digitális módszereivel, illetve a bibliográfiai adattudománnyal, vagyis elsősorban a következő területeken alkalmazható: a) digitális bölcsészet, b) közgyűjteményi tanulmányok (könyvtár- és információs tudomány, levéltártan, muzeológia), c) informatika. Minden képzési alkalmat kissé hozzá kell igazítani a tényleges célközönség igényeire.

³³ <https://www.loc.gov/marc/up34bibliographic/bdapndxg.html> (Megtekintés: 2026. március 12.)

nyeihez, például a metaadatok szabványait eleve ismerő metaadat-szakértők képzésénél ehelyett több helyet kell szentelni a digitális módszereknek, míg az informatikusok képzésénél a bibliográfiai adatok programozásának sajátosságait kell bemutatni.

Minőségértékelés és fenntarthatóság. A digitális módszerek gyorsan változnak, és nagy az esélye annak, hogy a tananyag megírása idején releváns rész néhány évvel később már elavultnak számító eszközökön vagy módszereken alapul. A készítőknél el kell dönteniük, hogy mit kezdjenek az elavult részekkel. A Programming Historian például kidolgozott egy tananyag-kivonási elvet.³⁴ Fontos az is, hogy megakadályozzuk a gyors elavulást, ezért követni kell a terület legjobb gyakorlatait, például az ő útmutatóikat.³⁵ A tananyag hosszú távú karbantartásának kulcsa a rendszeres használata, a kidolgozóinak tehát elkötelezetten kell lenniük abban, hogy a tananyagot alkalmazzák a tanítás során, így garantálják annak rendszeres felülvizsgálatát.

Együttműködés. A tananyagfejlesztési időszak utolsó szakaszában érdemes felvenni a kapcsolatot potenciális oktatási partnerekkel: egyrészt egyetemekkel, másrészt informális oktatási kezdeményezésekkel, mint például a Programming Historian, a DHTech oktatási munkacsoportja,³⁶ a LIBER Adattudomány a könyvtárakban munkacsoportja,³⁷ a Library Carpentry,³⁸ a GLAM Labs³⁹ és a Kutatási Szoftverek Fejlesztése⁴⁰ közösségek. Fontolóra lehet venni az anyag adaptálását és beépítését néhány ilyen kezdeményezésbe.

³⁴ <https://programminghistorian.org/en/lesson-retirement-policy> (Megtekintés: 2026. március 9.)

³⁵ Sustainable Writing: <https://programminghistorian.org/en/author-guidelines#sustainable-writing>, Reviewer Guidelines for Assessing Lesson Sustainability: <https://programminghistorian.org/en/reviewer-guidelines#sustainability>, Editor Guidelines for Fostering Lesson Sustainability: <https://programminghistorian.org/en/editor-guidelines#c-sustainability--internationalization-review>.

³⁶ DHTech DHTech Training & Education Working Group, <https://dh-tech.github.io/wg-education-training/>

³⁷ LIBER Data Science in Libraries Working Group, <https://libereurope.eu/working-group/liber-data-science-in-libraries-working-group/>
<https://librarycarpentry.org/>

³⁸ <https://www.glamlabs.io/>

³⁹ Research Software Engineering, <https://society-rse.org/>, <https://us-rse.org/>, <https://de-rse.org/>

Hivatkozott irodalom

- BÁTORFY 2024 – BÁTORFY Attila, *Adatvizualizáció, Elmélet, rendszer, módszer, Bevezetés az adatok grafikus ábrázolásának elméletébe és gyakorlatába*, Bp., ELTE Eötvös Kiadó, 2024.
- BOURDIEU 2008 – BOURDIEU, Pierre, *A Conservative Revolution in Publishing = Translation Studies*, 1(2008/2), 123–153. <https://doi.org/10.1080/14781700802113465>
- DAMEROW, NELSON, HERNANDEZ 2025 – DAMEROW, Julia, NELSON, David Ragnar, HERNANDEZ, Jose, *Developing Training and Education Resources for Research Software Engineering in DH = Anthology of Computers and the Humanities*, 2(2025), 13–18. <https://doi.org/10.63744/X9dgVaDTvJHY>
- FARKAS, KÁLDOS, KIRÁLY 2025 – FARKAS Gábor Farkas, KÁLDOS János, KIRÁLY Péter, *A régi magyarországi kiadványok „sötét anyaga” = Magyar Könyvszemle*, 124(2025), 226–266. <https://doi.org/10.17167/MKSZ.2025.2.226-266>
- FEINSTEIN, THOMAS 2002 – FEINSTEIN, Charles, THOMAS, Mark, *Making History Count, A Primer in Quantitative Methods for Historians*, Cambridge, Cambridge University Press, 2002.
- HEILBRON 1999 – HEILBRON, Johan, *Towards a Sociology of Translation, Book Translations as a Cultural World-System = European Journal of Social Theory*, 2(1999/4), 429–444. <https://doi.org/10.1177/136843199002004002>
- HESSBRÜGGEN-WALTER 2024 – HESSBRÜGGEN-WALTER, Stefan, *Interdisciplinarity in the 17th Century?, A Co-Occurrence Analysis of Early Modern German Dissertation Titles = Synthese*, 203(2024/2), 1–19. <https://doi.org/10.1007/s11229-024-04494-2>
- HESSBRÜGGEN-WALTER 2025 – HESSBRÜGGEN-WALTER, Stefan, *Early Modern Dissertations in French Libraries, The EMDFL Dataset = Journal of Open Humanities Data*, 11(2025/36), 1–15. <https://doi.org/10.5334/johd.307>
- HUDSON, ISHIZU 2016 – HUDSON, Pat, ISHIZU, Mina, *History by Numbers, An introduction to quantitative approaches*, 2nd ed., Bloomsbury, 2016.
- KARSDORP, KESTEMONT, RIDDELL 2021 – KARSDORP, Folgert, KESTEMONT, Mike, RIDDELL, Allen, *Humanities Data Analysis, Case Studies with Python*, Princeton, Oxford, Princeton University Press, 2021.
- KIRÁLY 2019 – KIRÁLY, Péter, *Measuring metadata quality*, PhD dissertation, University of Göttingen, 2019. <https://doi.org/10.13140/RG.2.2.33177.77920>
- KIRÁLY 2025 – KIRÁLY, Péter, *Fetching Index Translationum = Bibliographic Data blog*, 2025. <https://bibliodata.substack.com/p/fetching-index-translationum>

- KIRÁLY, KISÉRY 2025 – KIRÁLY, Péter, KISÉRY, András, *‘Mór Jókai, alas,’ The most successful Hungarian writer; A quantitative analysis = Patterns of Translation blog*, 2025. <https://translationpatterns.substack.com/p/mor-jokai-alas-the-most-successful>
- KIRÁLY et al. 2025 – KIRÁLY, Péter, UMERLE, Tomasz, MALÍNEK, Vojtěch, HERDEN, Elzbieta, KOPER, Beata, COLAVIZZA, Giovanni, JAGERSMA, Rindert, LAHTI, Leo, LINDEMANN, David, ŁUBOCKI, Jakub Maciej, MILANOVA, Alexandra, PÉTER, Róbert, RISSLER-PIPKA, Nanette, SIWECKA, Dorota, ROMANELLO, Matteo, ROSZKOWSKI, Marcin, TOLONEN, Mikko, VIMR, Ondřej, *Effects of Open Science and the Digital Transformation on the Bibliographical Data Landscape = Library Catalogues as Data, Research, Practice and Usage*, eds. Paul Gooding, Melissa Terras, Sarah Ames, Facet Publishing, 2025, 19–44.
- KLINKE 2025 – KLINKE, Harald, *Cultural Data Science, An Introduction to R*, Cham, Springer, 2025.
- KOESER et al. 2025 – KOESER, Rebecca Sutton, DAMEROW, Julia, CASTIES, Robert, CRAWFORD, Cole, *Undate, Humanistic Dates for Computation. Because Reality Is Frequently Inaccurate = Computational Humanities Research*, 1(2025), 1–9. <https://doi.org/10.1017/chr.2025.10006>
- KRUUSMAA, TINITS, NEMVALTS 2025 – KRUUSMAA, Krister, TINITS, Peeter, NEMVALTS, Laura, *Curated Bibliographic Data, The Case of the Estonian National Bibliography = Journal of Open Humanities Data*, 11(2025), 1–15. <https://doi.org/10.5334/johd.280>
- LAHTI et al. 2019 – LAHTI, Leo, MARJANEN, Jani, ROIVAINEN, Hege, TOLONEN, Mikko, *Bibliographic Data Science and the History of the Book (c. 1500–1800) = Cataloging & Classification Quarterly*, 57(2019/1), 5–23. <https://doi.org/10.1080/01639374.2018.1543747>
- LEMERCIER, ZALC 2019 – LEMERCIER, Claire, ZALC, Claire, *Quantitative Methods in the Humanities. An Introduction*, Charlottesville, London, University of Virginia Press, 2019.
- MARIANI 2023 – MARIANI, Fabio, *Introducing VISU, Vagueness, Incompleteness, Subjectivity, and Uncertainty in Art Provenance Data = Proceedings of the Workshop on Computational Methods in the Humanities*, (2022), Vol-3602. Workshop Proceedings, 2023. 63–84. <https://ceur-ws.org/Vol-3602/paper5.pdf>
- SZEMES, DOBÁS 2025 – SZEMES Botond, DOBÁS Kata, *A visegrádi országok digitális irodalmi emlékezete, Wikipedia, Wikidata – a regionális irodalomtörténet új alakzatai = Irodalomtörténeti Közlemények*, 129(2025/2), 191–212. <https://doi.org/10.56232/itk.2025.2.04>

- TAYLOR, TILTON 2024 – TAYLOR, Arnold, TILTON, Lauren, *Humanities Data in R. Exploring Networks, Geospatial Data, Images, and Text*, 2nd edition, Cham, Springer, 2024.
- THALLER 2020 – THALLER, Manfred, *On vagueness and uncertainty in historical data = Ivory Tower blog*, 2020. <https://ivorytower.hypotheses.org/88>
- TOLONEN et al. 2023 – TOLONEN, Mikko, VIMR, Ondřej, KIRÁLY, Péter, PANUŠKOVÁ, Charlotte, *Bibliographical Data Science, From Catalogues to Research Data = Social Sciences & Humanities Open Marketplace*, 2023. <https://marketplace.sshopencloud.eu/workflow/tE2HiC>