

Az adatpublikálás kulcsfontosságú elemei: A jelenlegi bevált gyakorlatok felhasználása az adatpublikálásra szolgáló referenciamodell kidolgozására

Szerzők: : Claire C. Austin^{1,2,3}, Theodora Bloom^{*4}, Sünje Dallmeier-Tiessen^{8,5}, Varsha K. Khodiyar⁶, Fiona Murphy^{8,7}, Amy Nurnberger⁸, Lisa Raymond⁹, Martina Stockhause¹⁰, Jonathan Tedds¹¹, Mary Vardigan¹², Angus Whyte¹³

Levelező szerzők: Sünje Dallmeier-Tiessen, Fiona Murphy

Közreműködők: Timothy Clark¹⁴, Eleni Castro¹⁵, Elizabeth Newbold¹⁶, Samuel Moore¹⁷, Brian Hole¹⁸

Szerzői affiliáció: ¹Environment Canada, ²Research Data Canada, ³Carleton University, ⁴BMJ, ⁵CERN, ⁶Nature Publishing Group, ⁷University of Reading, ⁸Columbia University, ⁹Woods Hole Oceanographic Institution, ¹⁰German Climate Computing Centre (DKRZ), ¹¹University of Leicester, ¹²University of Michigan/ICPSR, ¹³Digital Curation Centre - Edinburgh.

Szerzői nyilatkozat: Minden szerző kijelenti, hogy nincsenek be nem jelentett összeférhetetlenségeik.

A cikkben megfogalmazott vélemények a szerzők véleményei, és nem feltétlenül tükrözik azoknak a szervezeteknek az elveit, amelyekhez tartoznak.

A szerzők hozzájárultak a cikk megírásához és jelentősen az elemzéshez. A közreműködők megosztották a munkafolyamatokat a csoporttal (az elemzéshez). A szerzők ábécé sorrendben vannak felsorolva.

Absztrakt

Cél:

Az adatpublikálási munkafolyamatok óriási hatással lehetnek a kutatókra, a kutatási gyakorlatokra és a publikálási paradigmákra, valamint a finanszírozási stratégiákra, a karrier és a kutatás értékelésére.

Bemutatjuk az ilyen munkafolyamatok általános alkotóelemeit, hogy referenciamodellt nyújtsunk az említett érintetteknek.

Módszerek:

Az RDA-WDS Data Publishing Workflow csoportjának célja, hogy tanulmányozza az adatok közzétételével kapcsolatos munkafolyamatokat a tudományágak és intézmények között.

Munkafolyamatok széles skáláját vizsgálták, közös elemeket és bevált gyakorlatokat keresve, például alapvető önkiadói szolgáltatásokat, intézményi adatrepozitóriumokat, hosszú távú projekteket, hitelesített adatrepozitóriumokat, valamint a közös adatfolyóirat- és repozitórium megállapodásokat.

Eredmények:

E vizsgálat eredményeit felhasználták az általános összetevőkből álló adatpublikálási referenciamodell alkotására.

A jelenlegi adattárolási környezet értékelése alapján kiemeljük a figyelembe veendő fontos hiányosságokat és kihívásokat, különösen amikor összetettebb munkafolyamatokról van szó és azok szélesebb közösségi keretrendszerbe integrálódnak.

Következtetések:

Nyilvánvaló, hogy az adatpublikálás területén adott lehetőségek változatosak és dinamikusak, és vannak jelentős hiányosságok és kihívások is.

Az adatpublikálási rendszer különféle alkotóelemeinek a lehető legnagyobb mértékben zökkenőmentes és integrált módon kell működniük.

Ezért támogatjuk az adatrepozitóriumokra vonatkozó meglévő szabványok megvalósítását és az adatpublikálási folyamatok minden részét, valamint szükség esetén új szabványok kidolgozását.

A hatékony és megbízható adatpublikálást be kell ágyazni a dokumentált munkafolyamatokba. Mivel egyre több kutatóközösség törekszik a kutatással kapcsolatos adatok közzétételére, építhetnek az ebben a referenciamodellben megnevezett egy vagy több elemre.

TARTALOM

[Data availability](#)

[Introduction](#)

[Methods and materials](#)

[Results and analysis](#)

[Towards a reference model in data publishing](#)

[Definitions for data publishing workflows and outputs](#)

[Key components of data publishing](#)

[Detailed workflows and dependencies](#)

[Data deposit](#)

[Ingest](#)

[Quality assurance \(QA\) and quality control \(QC\)](#)

[Data administration and long-term archiving](#)

[Dissemination, access and citation](#)

[Other potential value-added services, and metrics](#)

[Diversity in workflows](#)

[Discussion and conclusions](#)

[Gaps and challenges](#)

[Best practice recommendations and conclusions](#)

[REFERENCES](#)

Adatok elérhetősége

A cikkben közölt elemzésben bemutatott adatok a következő helyen érhetőek el:

Bloom, T., Dallmeier-Tiessen, S., Murphy, F., Khodiyar, V.K., Austin, C.C., Whyte, A., Tedds, J., Nurnberger, A., Raymond, L., Stockhause, M., Vardigan, M. *Zenodo* doi: [10.5281/zenodo.33899](https://doi.org/10.5281/zenodo.33899) (2015)

Bevezetés

Az utóbbi években különféle adatpublikálási munkafolyamatok jelentek meg, amelyek lehetővé teszik a kutatók számára, hogy adatokat repozitóriumokon és kijelölt folyóiratokban publikálják.

Egyes tudományágak, mint például a társadalomtudományok, a genomika, a csillagászat, a földtudomány és a multidiszciplináris területek, mint például a poláris tudomány, kultúrákat alakítottak ki a kutatási adatok repozitóriumokon keresztüli megosztására, de hagyományosan nem mindenütt volt arra gyakorlat, hogy a kutatók adataikat hogyan osszák meg annak érdekében, hogy azok kereshetőek és újrafelhasználhatóak legyenek mások számára is.

Az adatmegosztásra általában vagy csak akkor kerül sor, ha egy közösség elkötelezte magát a nyílt megosztás mellett (pl. A Bermuda-alapelvek és a Fort Lauderdale találkozói megállapodása a genomi

adatokról), vagy akkor, ha erre törvényi követelmény vonatkozik, vagy ha a nagy kutatóközösségek hozzáférhetnek tudomány-specifikus létesítményekhez, műszerekhez vagy archívumokhoz.

Az előrelépés jelentős akadálya a bevált gyakorlatok és szabványok széles körű eltérése a tudományágak között és azokon belül. A bevált gyakorlatok példái közé tartozik a szabványosított adat archiválás a földtudományban, a csillagászatban és a genomikában. Sok más típusú adat számára az archiválás csak most kezd megjelenni, vagy egyáltalán nem is létezik.

A repozitóriumokon keresztül történő adatmegosztás egyik fő visszatartó ereje az, hogy ott az adatok publikálása időigényes, amely időtartamot úgy ítélnék meg a kutatók, hogy jobban is eltölthetnének olyan tevékenységekkel, amelyekre kreditet kapnak (például hagyományos kutatási publikációk, finanszírozásszerzés stb.). Sajnos, ha az adatokat a kutatók és intézményeik szétválasztják, az idő múlásával a visszakeresés valószínűsége gyorsan csökken.

A kiadók és a finanszírozó ügynökségek megbízatásának megjelenése a publikációk alapjául szolgáló adatokat hozzáférhetővé tétele érdekében a beszélgetést változtatni kellene: ahelyett, hogy azt kérdeznénk „A kutatók közzéteszik adataikat?“, az lenne a kérdés, hogy „Hogyan tudjuk megbízható módon közzétenni az adatokat?“ Így látjuk a nyitottság és az átláthatóság követelményeit, és arra törekszünk, hogy az adatokat első osztályú kutatási outputnak tekintsük. Az adatpublikálás jelentősen ösztönözheti a kutatókat adataik megosztására azáltal, hogy mérhető és méltányos eredményeket szolgáltatnak, ezáltal felgyorsítva a kialakulóban lévő paradigmaváltást.

Az adatok közzétételét még nem vizsgálják átfogó módon a kutatási értékelések és promóciók során, de fejlesztések és kezdeményezések zajlanak a különféle finanszírozási és más kutatási területeken, hogy ezeket az értékeléseket átfogóbbá tegyék. Míg továbbra is fennáll az az érzés, hogy az adatokra kisebb hangsúly kerül, mint a közzétett folyóiratcikkekre a foglalkoztatási és előléptetési döntésekkel kapcsolatban, a legfrissebb tanulmányok azt mutatják, hogy ha az adatok nyilvánosan hozzáférhetőek, az nagyobb számú publikációt eredményez.

Az adatmegosztás indokai az újrafelhasználás feltételezésein alapulnak - ha megosztják az adatokat, az odacsábítja a felhasználókat. Az adatok megosztásának, újra felhasználásának és új céljának megalkotásának képessége azonban a megfelelő tudás-infrastruktúra rendelkezésre állásától függ. Sajnos az infrastruktúra kiépítésének sok kísérlete kudarcot vallott, mert túl körülményes az alkalmazkodás. A megoldás lehet az infrastruktúra fejlesztésének a tudósok és akadémikus tényleges munkájához igazítása, ahelyett, hogy elvárnák tőlük, hogy az adatközpont, a szervezeti vezetők, a kiadók vagy a finanszírozók által kívánatos módon működjenek. Néhány felmérés azt mutatta, hogy a kutatók a repozitóriumok használatát a távoli harmadra helyezik - az egyedi kérések megválaszolása és a helyi weboldalakon való adatelhelyezés után.

Hagyományosan a közzétett kutatási eredmények független többszörözése volt a tudományos érvényesítés sarokköve. Azonban egyre növekvő aggodalomra ad okot a közzétett kutatás reprodukálhatósága, azaz pl. hogy a kutató közzétett eredményei reprodukálhatók-e a kutató által alkalmazott adatok, kód és módszerek felhasználásával. Itt is alapos kultúra változásra van szükség, ha a reprodukálhatóságot be kell építeni a kutatási folyamatba. Az adatok elérhetősége kulcsfontosságú a megismételhető kutatás szempontjából, és elengedhetetlen a tudományba vetett bizalom megőrzéséhez.

Az adatok jobb elérhetősége felé történő elmozdulás eredményeként megkezdődött egy közösségi beszélgetés a repositóriumok és adatfolyóiratok által használt szabványokról, munkafolyamatokról és minőségbiztosítási gyakorlatokról. A beszélgetések és a lehetséges megoldások elsősorban arra vonatkoznak, hogy a hatalmas mennyiségű adatot és a hozzájuk kapcsolódó metaadatokat miként lehet kezelni a különféle formátumokban. Az érdekelt csoportok különböző szintű szabványokat dolgoznak ki és fogadnak el olyan nemzetközi szervezetek révén, mint például a Research Data Alliance (RDA), the International Council for Science (ICSU-WDS), és tudományágak közösségein belül is. Például, a csillagászatban hosszú ideje folyik a metaadat-szabványok kidolgozása a Virtual Observatory Alliance (Nemzetközi Virtuális Megfigyelő Szövetség) (IVOA) révén, míg az éghajlattudományban a netCDF / CF egyezményt szabványos formátumként fejlesztették ki, amely metaadatokat tartalmaz a rács struktúrájú adatokhoz. Még a nagyon sokszínű területeken, mint például az élettudományok, a BioSharing kezdeményezés megkísérli koordinálni a szabványok közösségi felhasználását. Egyre nagyobb teret hódít az a szemlélet, hogy az adatok közzététele biztosítja az adatok hosszú távú megőrzését, és így megbízható ösztöndíjat eredményez, reprodukálható kutatásokat mutat be, elősegíti az új eredményeket, lehetővé teszi az újratelepítést, és így felismeri az előnyöket és maximalizálja a kutatási beruházások megtérülését.

De mi pontosan az adatpublikálás? Parsons és Fox megkérdőjelezi, hogy a publikálás-e a helyes kifejezés, amikor digitális információkkal van dolgunk. Azt sugallják, hogy az adatpublikálás fogalma lehet korlátozó és leegyszerűsített, és azt javasolják, hogy alternatív paradigmákat vizsgáljanak meg, például a szoftver kiadásának és finomításának modelleit, az egyszeri közzététel helyett. Természetesen a verziószabályozásnak az adatpublikálás szerves részévé kell válnia, és ez megkülönböztetheti a hagyományos folyóiratcikkétől.

A dinamikus adathivatkozás a kutatási adatkészlet egyik fontos jellemzője, ami majd fejlődni fog pl. adatok és longitudinális tanulmányok ellenőrzése. Az Earth System Science Data adatfolyóirat az „élő adatok” megközelítésével néz szembe ezzel a kihívással. Az RDA dinamikus hivatkozási munkacsoport átfogó specifikációt dolgozott ki az adatkészlet egy részhalmazának citálásától a dinamikusan előállított adatokig, „on-the-fly” technikával. A nemzetközi tudományos létesítmények általában tervezik a tanulmány teljes időtartama alatt időszakonként tudományosan feldolgozott adatközvetítéseket (például az

XMM-Newton röntgen-távcső forráskatalógus), az alapul szolgáló adatkészletek archívumokon keresztül történő hozzáférhetővé tételével az embargószabályoknak megfelelően.

2011-ben Lawrence et al. az „adatpublikálás” fogalmát úgy határozta meg: hogy az adatok a lehető legteljesebben legyenek elérhetőek az interneten. A közzétett adatok olyan folyamaton mennek keresztül, amely garantálja az egyszerűen emészthető információk megbízhatóságát, formátumát és tartalmát. Callaghan et al. kiegészíti ezt az elgondolást, azzal érvelve, hogy az adatok hivatalos publikálása túlmutató szolgáltatást is nyújt az adatkészlet weboldalon történő megjelentetésének egyszerű lépésénél, abban az értelemben, hogy magában foglalja az adatkészlet technikai (formátum, metaadat) vagy tartalmi jellegű (például az adatok pontosak?) ellenőrzését az adatkészletben.

A hivatalos adat publikálás az adatfelhasználó számára a kapcsolódó metaadatokat, az adatok megőrzésével kapcsolatos biztosítékokat és platformot kínál az adatkészlet megtalálására és értékelésére - ezek mindegyike elengedhetetlen az adatok újrahaználatához. Fontos szempont tanulmányunkban az, hogy a „normális” adatkezelés támogatása nem felel meg a bevált gyakorlatokra vonatkozó előírásoknak. Például, 32 nemzetközi online adatplatform felmérésének elvégzése után, a Standards & Interoperability Committee of Research Data Canada (RDC) arra a következtetésre jutott, hogy még sok tennivaló van annak biztosítása érdekében, hogy az online adatplatformok megfeleljenek a minimum követelményeknek: az adatok megbízható kezelésére és megosztására vonatkozó szabványokban, valamint iránymutatásokat dolgozott ki a kutatási adatok közzétételének és megőrzésének szempontjaira vonatkozóan.

A jelen tanulmánnyal az első lépés megtörtént egy referenciamodell felé, amely az adatpublikálásra szolgáló általános alkotóelemeket tartalmazza - amelynek elő kell segítenie a tudományterületek közötti szabványok meghatározását. Leírjuk a kiválasztott adat közzétételi megoldásokat, az adattárak és az adatfolyóiratok szerepét, és jellemezzük a jelenleg használt munkafolyamatokat. Elemzésünk a változatos munkafolyamatok azonosítását és leírását foglalta magában, beleértve az alapvető önkiadói szolgáltatásokat, a hosszú távú projekteket, a hitelesített repozitóriumokat, valamint a közös adatfolyóirat- és repozitórium megállapodásokat.

A kulcsfontosságú közös elemek és a bevett gyakorlatok az adatpublikálásra szolgáló referenciamodell részeként lettek azonosítva. Ezek hozzájárulhatnak a jövőben az adatpublikálási tevékenységek szabványosításához (miközben elegendő teret hagynak a tudományági vagy intézményi gyakorlatok számára). Érdeemes megjegyezni, hogy a kulcsfontosságú meghatározásokról továbbra is folytatódik a vita. A munkacsoport az elemzés alapján bemutatja az alapvető adat publikálás feltételeit.

Módszerek és anyagok

Az RDA-WDS adat-közzétételi munkafolyamatok munkacsoportját (WG) / (RDA-WDS Publishing Data Workflows Working Group) azért hozták létre, hogy elemezze a meglévő és a kialakuló munkafolyamatok és az adat publikálásra vonatkozó szabványok megfelelő módon reprezentatív körét, ideértve az adattárolást és a hivatkozást, valamint referenciamodellek és megvalósítások összetevőinek biztosítására az alkalmazáshoz új munkafolyamatokban. Jelen tanulmány kifejezetten egy olyan referenciamodell tervezetének megfogalmazására összpontosult, amely általános komponenseket tartalmaz az adatpublikálási munkafolyamatokhoz, amelyekre mások építhetnek. Azt is felismerjük, hogy a referenciamodellre azért is van szükség, hogy a kutatók körében népszerűsítse azokat a folyamatokat, amelyeket használhatónak és vonzóknak találnak.

Ennek elérése érdekében a munkacsoport az OASIS referenciamodell meghatározását követte:

„...Ez egy absztrakt keretrendszer a környezeti entitások közötti lényeges kapcsolatok megértéséhez és az ezt a környezetet támogató következetes szabványok vagy előírások kidolgozásához. A referenciamodell kevés egységes fogalomra épül, és alapul szolgálhat az oktatáshoz és a szabványok magyarázatához egy nem szakember számára. A referenciamodell nem kapcsolódik közvetlenül semmilyen szabványhoz, technológiához vagy más konkrét megvalósítási részletekhez, hanem célja egy olyan közös szemantika biztosítása, amelyet egyértelműen lehet használni a különböző megvalósítások között.”

Különösen releváns példa az OAIS referenciamodell egy nyílt archív információs rendszerre. Ez a modell alakította ki a Trusted Digital Repository (TDR) szabványokat, amelyek keretbe helyezték a repozitórium bevált gyakorlatait az archivált digitális objektumok befogadására, kezelésére és elérésére. Ezeket a közelmúltban példázza a DSA-WDS követelménykatalógusa, és különösen relevánsak a munkafolyamatok nyilvánvalóvá tétele szempontjából.

A munkacsoportban felmerült különös aggályaink az ilyen szabványokon alapulnak, hogy irányítsák a kutatási adatobjektumok minőségbiztosítását és szakértői értékelését, hivatkozásukat, valamint a kutatási és tudományos kommunikációs környezetben lévő más digitális objektumokkal való összekapcsolást.

Az esettanulmányi megközelítés összhangban áll a célkitűzéssel. Az esettanulmányok a jelenségeket kontextusukban vizsgálják. Hasonlóképpen, a fogalmi modell kidolgozása nem követeli meg, hogy általánosított igényeket gyártsunk az adattár populációjának egészére nézve, ugyanakkor arra kötelez minket, hogy a közösségi áttekintés és módosítás révén teszteljük annak relevanciáját az adattárak és más érdekelt felek számára.

Mivel az RDA-WDS Publishing Data Workflows WG tagsága megfelelő módon változatos volt a tudományágak felől és az érdekelt felek részvétele szempontjából, a csoport tudását és kapcsolatait használtuk, és felhívásokat tettünk az RDA és a WDS égisze alatt, a Force11 Implementation Group csoporttal együtt az adatpublikálási munkafolyamatok bevált gyakorlatainak és esettanulmányainak azonosítására. Az RDA plenáris ülésein bemutatott előadásokat és műhelymunkákat használták a megközelítés és az előrehaladás érvényesítésére. Ezzel az iteratív megközelítéssel azonosítottuk a kezdeti repozitóriumokat, projekteket és kiadói platformokat, amelyekről azt gondoltuk, hogy megfelelő módon reprezentatívak az intézményi hovatartozáshoz és a domain-specifikus vagy multidiszciplináris fókuszhoz. Ezek a munkafolyamatok esettanulmányként szolgáltak az elemzéshez, hogy azonosítsák az adattárolók, projektek és közzétételi platformok „adatok közzétételének” valószínű példáit, legyenek azok intézményi, domain-specifikusak vagy multidiszciplinárisak.

A nyilvánosan hozzáférhető információkat a munkafolyamatok általános fogalmakkal való leírására használták. Ezenkívül a repozitórium képviselőit meghívták, hogy videokonferencia és szemináriumi találkozók útján mutassák be munkafolyamataikat. Hangsúlyt fektettek az adatok hivatkozását megkönnyítő munkafolyamatokra, és figyelembe vették az adatok „metrikájának” biztosítását. Az információkat egy összehasonlító mátrixba szerveztük és a csoporthoz továbbítottuk felülvizsgálat céljából, majd számos megjegyzés és javítás történt. Az üres mezőket ahol lehetséges volt, kitöltötték, és a kifejezéseket pedig átellenőrizték és harmonizálták az egész mátrixban. Huszonhat példát segítségével történt a jellemzők és a munkafolyamatok összehasonlítására. Azonban egy munkafolyamatot (Arkivum) úgy ítélték meg, hogy nem felel meg az „adatpublikálás” meghatározásának, ahogyan ez a kutatás során felmerült, tehát a záró táblázat huszonöt entitásból áll (1. táblázat).

1. Táblázat: A munkafolyamatok és egyéb tulajdonságok elemzésére kiválasztott repozitóriumok, projektek és publikáló platformok

Munkafolyamat-szolgáltató neve	Munkafolyamat-szolgáltató típus	Munkafolyamat-szolgáltató speciális kutatási terület, ha van környezeti tudományok földtudományok	Letét kezdeményező
ENVRI reference model PREPARDE	irányelvek irányelvek		Projekt-vezette kutató vezeti (<i>Geoscience Data Journal</i>) kutató vezeti
Ocean Data Publication Cookbook	irányelvek	Tengeri tudományok	kutató (szerző) vezeti
<i>Scientific Data</i> , Nature Publishing Group	Folyóirat		
<i>F1000Research</i>	Folyóirat	Élettudományok	Kutató vezette; A szerkesztőség ellenőrzi
Ubiquity Press OHDJ	Folyóirat	élet, egészség és társadalomtudományok	kutató vezeti

<i>GigaScience</i>	Folyóirat	élet-és orvosi biológiai tanulmányok	kutató (szerző) vezeti
<i>Data in Brief</i>	Folyóirat		szerző vezeti
<i>Earth System Science Data Journal</i> , Copernicus Publications	Folyóirat	földtudományok	Kutató által vezetett adatk. A kutatók által irányított adatoknak a repozitóriumba történő benyújtására
Science and Technology Facilities Council Data Centre	Repozitórium	fizika és űrtudományok	Kutató-vezette projekt részeként megvalósítandó
National Snow and Ice Data Center	Repozitórium	Poláris tudományok	projekt vagy kutató által vezetett
INSPIRE Digital library	Repozitórium	nagy energiájú fizika	kutató vezeti
UK Data Archive (ODIN)	Repozitórium	Társadalomtudományok	kutató vezeti
PURR Institutional Repository	Repozitórium		kutató/könyvtáros vezeti
ICPSR	Repozitórium	társadalom és viselkedéstudományok	Kutató, beszerzési tisztviselő és finanszírozó által vezetett
Edinburgh Datashare	Repozitórium		kutató végzi, könyvtáros segít
PANGAEA	Repozitórium	földtudományok	kutató vezeti
WDC Climate	Repozitórium	földtudományok	kutató vagy projekt vezette
CMIP/IPCC-DDC	Repozitórium	Éghajlat tudományok	projekt vezette
Dryad Digital Repository	Repozitórium	földtudományok	kutató vezeti
Stanford Digital Repository	Repozitórium		kutató vezeti
Academic Commons Columbia	Repozitórium		kutatók és repozitóriumi munkatársak
Data Repository for the University of Minnesota (DRUM)	Repozitórium		kutatók az intézetből
ARKIVUM and Figshare	Repozitórium		kutató vezeti
OJS/ Dataverse data repository, all disciplines	Repozitórium		kutató vezeti, a folyóirat cikk publikálás folyamat részeként

A munkafolyamatok az érintett tudományág, funkció, adatformátumok és szerepek alapján voltak leírva. Leírták azt is, hogy ezeknél milyen mértékben van jelen az adatpublikálással kapcsolatos következő 10 jellemző:

- Az állandó azonosítók (PID) hozzárendelése az adatkészletekhez és a használt PID típus - pl. DOI, ARK stb.
- Az adatok szakmai bírálata (pl. Kutató és szerkesztői bírálat)

- A metaadatok kurátori bírálata (pl. Intézményi vagy a tárgyi repozitóriumban)
- Technikai bírálat és ellenőrzések (pl. Az adatok integritása a repozitóriumban/ adatközpontban a befogadáskor)
- Kereshetőség: megtörtént-e az adatok indexelése, és ha igen, hol?
- Linkek további adattermékekhez (adat lap; bírálat; egyéb folyóiratcikkek) vagy „különálló” termékekhez
- Linkek, hogy biztosítsuk az információt, ahol van, és a szerző PID-k használata
- Az adatok idézésének megkönnyítése
- Hivatkozás az adatok életciklus-modelljére
- A szabványok betartása

A részletes információk és a kategorizálás megtalálhatók az összehasonlító mátrixot tartalmazó elemzési adatkészletben.

Eredmények és elemzés

Referenciamodell építése az adatpublikálásban

Az adatpublikálási munkafolyamatok és -kimenetek meghatározása

Az RDA-WDS Publishing Data Workflows WG által készített, az adatpublikációs munkafolyamatok összehasonlító mátrixának áttekintése feltárta, hogy a terminológia szabványosítására van szükség. Ezért hat kulcsfogalom meghatározását javasoljuk: kutatási adatok publikálása, kutatási adatok publikálásának munkafolyamatai, adatfolyóirat, adatscikk, adatáttekintés és repozitórium -bejegyzés.

Kutatási adatok publikálása

„A kutatási adatok publikálása nem más, mint a kutatási adatok, a kapcsolódó metaadatok, a kísérő dokumentáció és a szoftver kód (azokban az esetekben, amikor a nyers adatokat feldolgozták vagy manipulálták) kiadása újrafelhasználáshoz és elemzéshez, oly módon, hogy azok felfedezhetőek legyenek a weben és utalni lehessen rá következetes és egyedülálló módon. Az adatpublikálás dedikált adatrepozitóriumban és / vagy (adat) folyóiratokban zajlik, amelyek biztosítják, hogy a közzétett kutatási objektumok hosszú távon legyenek jól dokumentálva, kezelve, archiválva, kölcsönösen átjárhatók, idézhetőek legyenek, a minőségük legyen biztosítva és könnyen kereshető legyen- az adatok közzétételének minden szempontja, amely fontos az adatok harmadik fél általi végfelhasználók általi későbbi újrafelhasználására.”

Ez a meghatározás a bizalmas és érzékeny adatok publikálására is vonatkozik, a megfelelő védelemmel és a hozzáférhető metaadatokkal együtt. Az ilyen munkafolyamat konkrét példája lehet egy publikált folyóiratcikk, amely tartalmazza az adatkészlet fellelhetőségét és idézését az újrafelhasználáshoz való hozzáférési kritériumok miatt. A Harvard University jelenleg fejleszt egy olyan eszközt, amely végül a

Dataverse-hez lesz integrálva, azért, hogy a bizalmas és érzékeny adatokat felelősségteljes módon ossza meg és használja.

Kutatási adatok publikálásának munkafolyamatai

A kutatási adatok publikálásának munkafolyamatai olyan tevékenységek és folyamatok, amelyek a kutatási adatok, a kapcsolódó metaadatok és a kísérő dokumentumok és szoftver kód közzétételéhez vezetnek a weben. Az időközi vagy a véglegesen közzétett termékekkel ellentétben a munkafolyamatok arra szolgálnak, hogy menedzseljék, dokumentálják és áttekintsék, és így biztosítsák és növeljék a közzétett termék értékét. A munkafolyamatok tartalmazhatnak embereket, és gépeket, és gyakran az embereket a technológia támogatja, mivel azok a munkafolyamat egyes lépéseit hajtják végre. A hasonló munkafolyamatok részleteikben változhatnak, a kutatás tudományterülettől, az adatpublikálási terméktől függően és / vagy a munkafolyamat fogadó intézményétől függően (pl. Egyedi kiadó / folyóirat, intézményi repozitórium, tudományterület-specifikus repozitórium).

Adatfolyóirat

Az adatfolyóirat egy folyóirat (mindig Open Access), amely adatközzételt tesz közzé. Az adatfolyóirat általában sablonokat tartalmaz az adatok leírására, és útmutatást nyújt a kutatóknak arra vonatkozóan, hogy hol kell feltölteni, és hogyan kell leírni és bemutatni az adataikat. A folyóirattól függően az ilyen sablonok lehetnek általánosak vagy lehetnek tudományterülethez kötődőek. Egyes folyóiratok vagy kiadók saját adattárakat tartanak fenn. Az adatközzételt és a hozzá tartozó adatkészlet(ek) közötti kétirányú összeköttetés támogatása és a tartós azonosítási gyakorlatok megkönnyítése mellett az adatfolyóiratok munkafolyamatokat biztosítanak a minőségbiztosításhoz (azaz az adatok szakmai bírálata), és szerkesztési iránymutatásokat kell tartalmazniuk az adatminőség értékelésére.

Adatközzételt

Az adatközzételt egy „publikálási” termék, más néven „adatleíró”, amely megjelenhet egy adatfolyóiratban vagy más folyóiratban. Amikor a kiadók az „adatpublikálásra” hivatkoznak, általában az adatközzételtre gondolnak, nem pedig az alapul szolgáló adatkészletre. Az adatközzételt arra összpontosítanak, hogy az adatok kereshetőek, kölcsönösen átjárhatóak és újrafelhasználhatók legyenek, ahelyett, hogy hipotéziseket tesztelnének vagy új értelmezéseket mutatnának be (szemben a hagyományos folyóiratcikkekkel). Függetlenül attól, hogy egy külön repozitóriumban lévő adatkészlethez kapcsolódik, vagy az adatokkal együtt nyújtják be, az adatközzételt célja, hogy hivatalos utat biztosítson az adatmegosztáshoz. A szülő folyóirat dönthet úgy, hogy leírja-e az adatkészlet kezelésének, formázásának, elérhetőségének, megőrzésének vagy szakmai áttekintésének szabványait. Meghatározás szerint az adatközzételt ezen eszközök

tulajdonságainak leírására szolgál, valamint ösztönöz is ezekre. Az ilyen cikkek hossza változhat a mikro-tanulmányoktól (egy táblára vagy parcellára összpontosítva) a komplex adatkészletek nagyon részletes bemutatásáig.

Adatvizsgálat

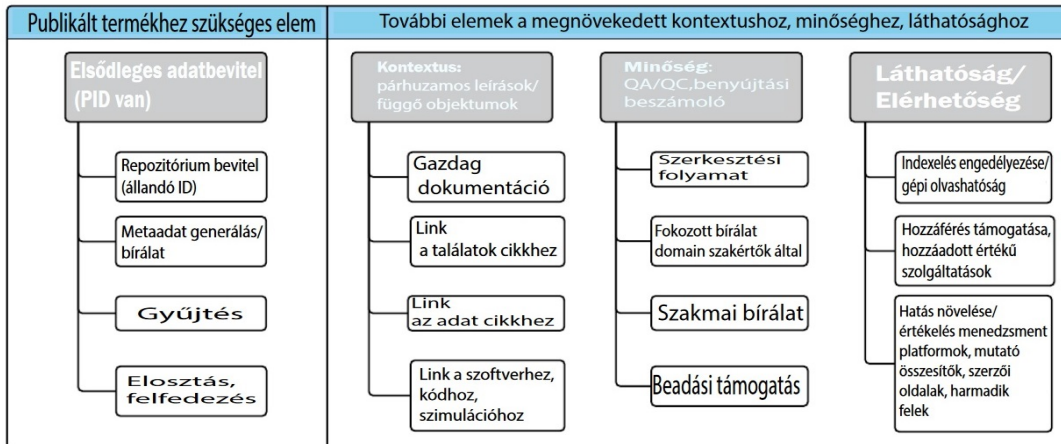
Az adatvizsgálat a minőség-értékelési munkafolyamatok széles skáláját foglalja magában, amely a metaadatok pontosságának technikai felülvizsgálatától az adatállományok és a dokumentáció megfelelőségének, valamint a számítások pontosságának kettős vak bírálattig terjedhet. A felülvizsgálati folyamatok többféle változata létezik, és olyan tényezőktől függ, mint például a kiadói követelmények, a kutatók elvárásai vagy az adatok érzékenysége. Egyes munkafolyamatok hasonlóak lehetnek a hagyományos folyóirat-munkafolyamatokhoz, amelyekben speciális szerepeket és felelőségeket ruháznak fel a szerkesztőkre és bírálókra az adatpublikálás minőségének biztosítása érdekében. Az adatbíráló folyamatok ezért magában foglalhatják a szakmai bírálatot, amelyet az adatfolyóiraton vagy az adattáron kívüli meghívott domain-szakértők végeznek, a repozitórium kezelés szakértők által végzett technikai adat-felülvizsgálatot annak biztosítása érdekében, hogy az adatok megőrizhetők-e, és / vagy a repozitórium tématerületi domain szakértői által végzett tartalmi bírálót.

Adattár-bejegyzés

Az adattár bejegyzés az adatpublikálás alapvető alkotóeleme, amely egy olyan állandó, egyedi azonosítót tartalmaz, amely egy olyan oldalra mutat, ahol adatleírást és részleteket találhatunk az adatok elérhetőségéről és a tényleges adatok elérésének eszközeiről.

Az adatpublikálás kulcsfontosságú elemei

A munkafolyamatok elemzése az RDA-WDS data publishing WG segítségével meghatározta azokat az összetevőket, amelyek hozzájárulnak az adatpublikáláshoz szükséges általános referenciamodellhez. Megkülönböztetjük az alapvető és a kiegészítő szolgáltatásokat. Az alapvető szolgáltatások egy megbízható adatrepozitóriumban lévő bejegyzéseket tartalmaznak, beleértve egy állandó azonosítót, szabványos metaadatokat és alapvető kezelést (1. ábra).



1. ábra: Az adatpublikálás kulcsfontosságú elemei. Az adatpublikáláshoz szükséges elemek a bal oldali panelen, a választható szolgáltatások és funkciók pedig a jobb oldali panelen láthatók.

Az opcionális kiegészítő szolgáltatások tartalmazhatnak olyan összetevőket, mint például a kontextusba helyezés az adatokat közlő tanulmányokba történő beágyazással vagy a hagyományos tudományos cikkekre mutató linkekkel. Egyes szerzők és megoldások különbséget tesznek a metaadatok publikálása és az adatok publikálása között. Úgy gondolnánk, hogy az adatokat és a hozzájuk kapcsolódó metaadatokat legalább kétirányban össze kell kapcsolni tartós módon, és közzé kell tenni, és csomagként kell megvizsgálni, mivel a metaadatok nélkülözhetetlenek az adat felhasználásához, megértéséhez és értelmezéséhez.

Fontos kiegészítő szolgáltatások a minőségbiztosítás / minőség-ellenőrzés (QA / QC) és a szakmai bírálat. Az ilyen szolgáltatásoknak különböző változatai léteznek, kezdve a szerző által irányított, a szerkesztő által vezérelt, a könyvtáros által támogatott megoldásoktól a (nyitott) szakmai bírálatig. Ezek az elemek elengedhetetlenek a jövőbeni újrafelhasználás és reprodukálható kutatás számára. Elemzésünk során azt találtuk, hogy sok szolgáltató nyújt, vagy fontolgatja ilyen szolgáltatások bevezetését. A kiegészítő szolgáltatások harmadik csoportja a láthatóság javítását célozza, amint az az 1. ábra jobb oldali paneljén látható. Ez a szolgáltatáskészlet jelenleg nincs jól kialakítva, és ez akadályozza az adatok újrafelhasználását. Egyéb feltörekvő szolgáltatások közé tartozik az adatközlési munkafolyamat összekapcsolása az indexelő szolgáltatásokkal, a kutatási információs szolgáltatásokkal (CRIS) vagy a mérőszám összesítővel.

Az adatok újbóli felhasználásának lehetősége érdekében az adatok közzétételének tartalmaznia kell legalább a kezelés alapvető elemeit, a QA/QC-t, és a hivatkozást, valamint a felhasználási esetnek

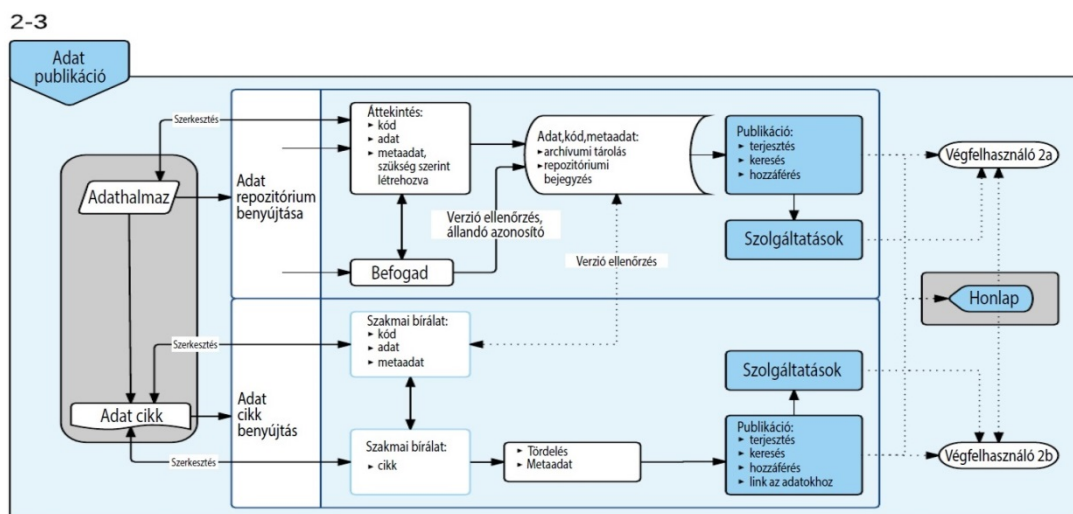
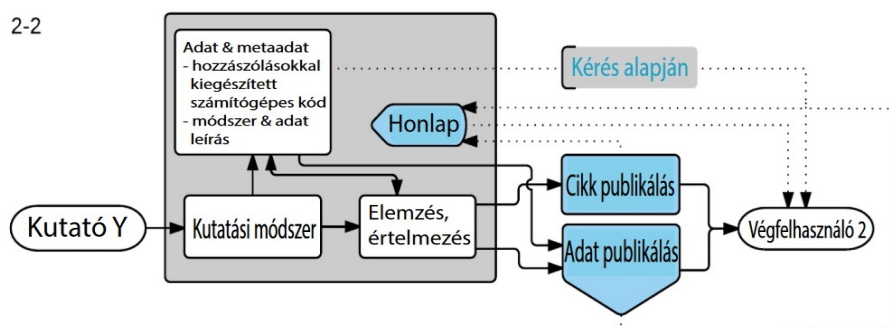
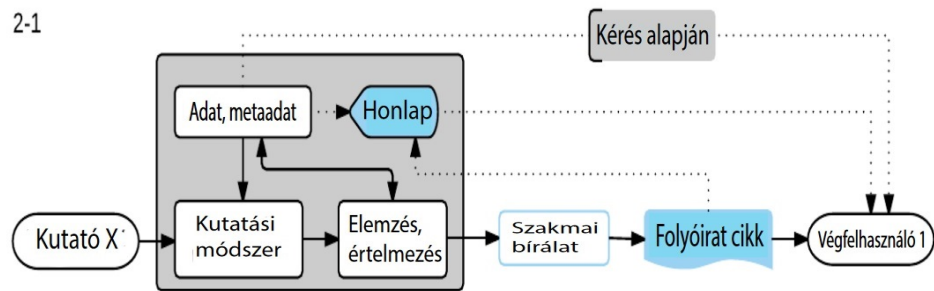
megfelelő kiegészítő elemeket (1. ábra). A felhasználástól függően azonban helyénvaló lehet egy adott elemkészletet kiválasztani a kulcskomponensek közül (néhány bevált gyakorlatot követve). A jövőbeli újrafelhasználás fényében azt állítanánk, hogy a kezelés alapvető elemeit, a QA/QC-t, valamint a hivatkozásokat mindig bele kell foglalni.

Részletes munkafolyamatok és függőségek

Bemutatunk egy hagyományos cikk publikációs munkafolyamatot (2-1. Ábra), egy reprodukálható kutatási munkafolyamatot (2-2. Ábra) és egy adat-publikációs munkafolyamatot (2-3. Ábra).

A munkafolyamat-összehasonlítás során kiderült, hogy általában a kutató kezdeményezi a publikációs folyamatot, miután az adatok összegyűltek és közzétételre megfelelő állapotban vannak, vagy megfelelnek a benyújtásra vonatkozó repozitóriumi követelményeknek. Az adatkészletek közzétehetőek a repozitóriumban, társított adatkikkel vagy anélkül. Előfordul azonban, hogy az adatokat előállító eszköz és az adatrepoizitórium között közvetlen kapcsolat, úgynevezett csatorna van (az egyik ilyen példa a genom szekvenálás). A repozitóriumtól függően, mind az adatok hosszú távú archiválása előtt, mind tudományos, mind technikai minőségbiztosítási tevékenységek vonatkoznak az adatkészlet tartalmára, leírására, formátumára és a metaadat minőségére. A tipikus repozitórium létrehoz egy bejegyzést egy adott adatkészlethez vagy annak gyűjteményéhez.

A legtöbb adatrepoizitórium az adatkészletek szabványosított terjesztésébe, azaz az egyes közzétett elemek nyitóoldalába fektet be, ahogy azt a Force11 Data Citation implementációs csoport ajánlja. Egyes adattárak megkönnyítik a harmadik felek hozzáférését a kereshetőséghez vagy a metrikus szolgáltatásokhoz.



2. ábra. A kutatási adatpublikálás munkafolyamatai. Bemutatunk egy hagyományos cikk publikációs munkafolyamatot (2-1. ábra), egy reprodukálható kutatási munkafolyamatot (2-2. ábra) és egy adatkiadási munkafolyamatot (2-3. ábra).

Amint az a 2. ábrán látható, a kutatók számos különféle utat követhetnek és követhetnek is az adataikról való kommunikáció érdekében. Hagyományosan a kutatási eredményeket folyóiratokban teszik közzé, és

az adatok iránt érdeklődő olvasóknak (1. végfelhasználó) fel kell venniük a kapcsolatot a szerzőkkel ahhoz hogy elérjék a mögöttes adatokat, vagy meg kell próbálniuk hozzáférni azokhoz egy kutató által támogatott weboldalról (2-1. ábra). A kutatás nagyobb reprodukálhatóságát támogató, kialakulóban lévő folyamatok között szerepel az adatpublikálás valamilyen formája (2-2. ábra). Ez magában foglalja a különálló adatpublikáció speciális esetét, amelyek nem kapcsolódnak közvetlenül a dokumentumhoz. Ezek gyakoriak több domain területben (például a nagy éghajlati adatokat összehasonlító tanulmány, CMIP). A 2-3. ábra szemlélteti az elemzésünk során felmerülő két túlnyomórészt kialakulóban lévő adatpublikációs munkafolyamatot: (a) adatkészlet benyújtása a repozitóriumba; és b) adate cikk benyújtása adatfolyóiratba. Mindkét munkafolyamat megköveteli az adatkészletek elhelyezését a repozitóriumba.

A 2-3. ábrán bemutatott adat publikációs folyamat bármikor megindítható a kutatás során, amint az adatok kellően hiánytalanok és dokumentáltak, és különféle utakon haladhatnak. A repozitórium általában meghatározott sablonokat biztosít a metaadatoknak és további dokumentációknak (pl. módszertan vagy kódspecifikus metaadatok). A benyújtást ezután különféle szempontokból lehet felülvizsgálni, a repozitórium eljárásrendjétől és gyakorlatától függően. Ezek a felülvizsgálati folyamatok tartalmazhatnak formázási kérdéseket, tartalmi, metaadat vagy egyéb technikai részleteket. Egyes repozitóriumok is megkövetelhetik az adatkészlet verziószabályozását. Nagy a változatosság a repozitóriumoknál az elfogadott adatok típusa, a rendelkezésre álló források, a kínált szolgáltatások mértéke és a munkafolyamatok tekintetében. A 2-3. ábra a jelen tanulmányhoz kiválasztott repozitóriumok munkafolyamatainak közös elemeit szemlélteti (2-3. ábra), amelyek megegyeznek az 1. ábrán bemutatottakkal.

A kutató úgy is megkezdheti az adatpublikálás folyamatát, hogy benyújt publikálásra egy adate cikket egy adatfolyóiratba. Ezt a munkafolyamatot a 2-3. ábra is szemlélteti, és bár ez részben a repozitóriumoktól függ (az adatfolyóiratok általában azonosítják a jóváhagyott repozitóriumokat), az adate cikk publikációs folyamatának lehetősége van konzisztensebben biztosítani az adat publikálás előnyeit, ahogy azt az 1. ábra “További elemek” részén szemléltetjük. Az adatfolyóiratok hasonlítanak a hagyományos tudományos kutatási eredményekkel foglalkozó folyóirathoz (2-1. ábra), alapvető folyamataik közé tartozik a szakmai bírálat és az adate kszletek terjesztése. A bírálóknak természetesen publikáció előtti hozzáféréssel kell rendelkezniük a repozitóriumban lévő adate kszlethez, és verziószabályozó megoldásokra van szükség az adate kszletek és az adatpapírok számára. Függetlenül attól, hogy az adatokat folyóirate cikk formájában vagy repozitóriumon keresztül publikálják, mindkét munkafolyamat beépülhet a jelenlegi akadémiai értékelési rendszerbe és egy korszakalkotó folyamatot eredményezhet, nem pedig a korábbi rendszerektől való zavaró eltávolodást.

A reprodukálható kutatást támogató adat publikációs munkafolyamatok a végfelhasználók számára

hozzáférést biztosítanak a kezelt adatokhoz, kódokhoz és alátámasztó metaadatokhoz, amelyeket már elbíráltak és feltöltöttek egy megbízható repozitóriumba (2. ábra, 2a végfelhasználó). Ha egy kapcsolódó adatkikk kerül közzétételre, akkor a végfelhasználók további kontextuális információkkal is rendelkeznek majd (2. ábra, 2b végfelhasználó). A hagyományos folyóiratcikk a szokásos módon is megjelenhet, és összekapcsolható a publikált adatokkal és / vagy az adatkikkel is. Van néhány bevett automatizált munkafolyamat az adatok közzétételére (például az Open Journal Systems-Dataverse integrációval), vagy lehetnek alternatív automatizált vagy kézi munkafolyamatok a kutató támogatására (például Dryad).

Adattárolás

Megállapítottuk, hogy az adat publikációs munkafolyamatok alapjául szolgáló adattárolási mechanizmusok többségét a kutatók kezdeményezik, ám a tárolásra való benyújtás kezdeti lépését meghaladó részvételük a repozitóriumokban és a folyóiratokban eltérő volt. A platform célja (például az adatfolyóirat vagy a repozitórium célja), valamint az adatok letétbe helyezőjének végső észlelt célja és motivációja mind befolyásolja a folyamatot. Például egy tárgy-specialista repozitórium, mint amilyen a Science and Technology Facilities Council (STFC) vagy a National Snow and Ice Data Center (NSIDC) repozitóriumában található, átvizsgálja a benyújtásokat és értékeli a metaadatokat és a szükséges támogatás szintjét. Az adatfolyóiratok azonban rendszerint nem közvetlen megközelítést alkalmaznak: a folyóirat a közlemények specifikus helye, de az adatokat máshol tárolják. Ezért a folyóiratkiadó csoport gyakran külső felekre - a repozitóriumkezelőkre és általánosságban a kutatóközösségre - támaszkodik az adattárolás kezelésében és annak felmérésében, hogy az adattárolásra vonatkozó alapvető előírások betartásra kerülnek-e, vagy a minőségi előírások teljesülnek-e (lásd alább).

Befogadás

Megállapítottuk, hogy a tudományterület-specifikus repozitóriumoknak vannak a legszigorúbb befogadási és bírálati folyamatai, és hogy az általánosabb repozitóriumok, pl. az intézményi repozitóriumok (IR-k) vagy a Dryad elnézőbbek, tekintettel a különféle tudományágakból származó adatokat érintő használati esetek és gyakorlat nagyobb változatosságára. Néhány tudományág-specifikus repozitórium többlépcsős folyamatokkal rendelkezik, beleértve több QA / QC folyamatot és az OAIS-en alapuló munkafolyamatokat. Sok intézményi repozitórium szélesebb körű megközelítést vett át a misszióikhoz szükséges befogadáshoz, amely magában foglalja a campusokon készült kutatási eredmények archiválását, különösen azokat, amelyekhez kutatási anyagok hosszú sora tartozik, ideértve a történelmi adatokat is, amelyeket különféle módon lehet kezelni. Az adatstandardok kidolgozása és végrehajtása során, valamint a kutatók számára a modern adatkezelési gyakorlatokhoz szükséges eszközök, képzés és ösztönzők biztosítása mellett, a befogadási gyakorlatok kétségkívül javulnak.

Amikor az adat folyóiratok külső repozitóriumokra támaszkodnak a tényleges adatkezelés kezelése érdekében, akkor a folyóirat és a repozitórium kezelői között szoros együttműködésre van szükség, így bízhatnak benne, hogy a repozitórium az adatkezelést és a befogadást az elfogadható szabványos eljárások szerint végzi. Az adat folyóiratokat és a repozitóriumokat arra biztatják, hogy hozzanak nyilvánosságra és tegyenek átláthatóvá minden ilyen megállapodást (például szolgáltatási szintű megállapodásokat). Végül azonban az egyéni interakció ilyen szintje nem mérhető, ezért szükségesek lesznek automatizált eljárások és repozitóriumi szabványok.

Minőségbiztosítás (QA) és minőségellenőrzés (QC)

Megállapítottuk, hogy a minőségbiztosítás / minőségellenőrzés általában az adatpublikálási munkafolyamat során három ponton fordul elő: (1) adatgyűjtés és adatfeldolgozás során, az adatoknak repozitóriumba történő benyújtása előtt; (2) az adatok benyújtása és archiválása során; és (3) bírálati vagy szerkesztési eljárás során. Különbséget teszünk a hagyományosan értett szakmai bírálat és a belső bírálatok között, amelyeket az repozitóriumok és a folyóiratok általában végeznek (2. ábra), amelyek érinthetik a tartalmat, a formátumot, a leírást, a dokumentációt, a metaadatokat vagy egyéb technikai részletet. A minőségbiztosítási / minőségellenőrzési eljárások nagyon eltérőek, magába foglalják a tartalom és a dokumentáció minőségbiztosítási szerzőit / áttekintőit, a technikai minőségbiztosítási adatok kezelőit, könyvtárosokat és szerkesztőket. A minőségi kritériumok magukban foglalhatják az adatok, a metaadatok és a dokumentáció ellenőrzését a repozitóriumban, a tudományág és a projekt szabványai alapján.

A legtöbb repozitórium és az összes adatfolyóirat, amelyet áttekintettünk, tartalmazott bizonyos minőség-ellenőrzési / minőségbiztosítási munkafolyamatokat, de a szolgáltatások szintje és típusa változott. A kialakított adat repozitóriumok (például ICPSR vagy Dataverse) általában rendelkeznek adatkezelő személyzettel, akiknek az a feladatuk, hogy segítsék az adatok szabványosítását és felülvizsgálatát a benyújtáskor és befogadáskor, különösen a metaadatok területén. Néhány domain-repozitórium (például ICPSR) tovább megy és alapos minőség-ellenőrzést végez az adatokkal, szükség esetén az eredeti vizsgálóval konzultálva felülvizsgálja az adatokat. A projekt adatok hosszú távú archiválásáért felelős egyéb adatrepozitóriumok (például az IPCC-DDC) dokumentálják minőségbiztosítási eredményeiket. Egyes adattárak a QA / QC munkafolyamatok során kutatókra támaszkodnak az adatok, metaadatok és dokumentáció tudományos szempontjainak validálásához. Technikai támogatást, adatok érvényesítését vagy minőségbiztosítási / minőségellenőrzési szolgáltatásokat néhány adatrepozitórium is végez, de az elköteleződés szintje a szolgáltatástól és az egyes intézményektől függően változott: egyesek ellenőrizték a fájl integritását, mások pedig komplexebb megőrzési műveleteket ajánlottak, például az on-the-fly adatformátum konvertálást. Néhány többcélú repozitórium támogatást nyújtott a kutatóknak a minőségbiztosítási / minőségellenőrzési munkafolyamatokhoz, de ez nem volt általános gyakorlat. Összességében az adat publikálás minőségbiztosítása / minőségellenőrzése „felkapott” téma, és a közösségen belül erősen és folyamatosan vitatják azt. Mayernik et al. leírja az adatok publikálásával kapcsolatos technikai és tudományos szakmai bírálat gyakorlatát.

A vizsgált folyóirat-munkafolyamatok általában az adatkészlet és az adatlap bírálatának kettős folyamatait hajtották végre, amelyeket külön végeztek el, majd ellenőriztek annak biztosítása érdekében, hogy a kettő közötti kapcsolat érvényes legyen. Az adatfolyóiratok ilyen minőségbiztosítási / minőségellenőrzési munkafolyamatai szoros együttműködést igényelnek a kutatói közösséggel és a szakmai bírálókkal, valamint a kiadók és az adatrepozitóriumok között a munkafolyamatok koordinálásában, verziókövetésében és konzisztenciájában.

Tekintettel a jelenleg kínált minőségbiztosítási és minőségellenőrzési szolgáltatások széles skálájára, a jövőbeli ajánlásoknak a következőket kell figyelembe venniük:

- Az olyan repozitóriumok, amelyek jelentős erőfeszítéseket tesznek a magas szintű

minőségbiztosításra / minőségellenőrzésre hasznosak azon kutatók számára, akiknek anyagai megegyeznek a repozitórium portfóliójával azáltal, hogy anyagaikat újrafelhasználhatóvá teszik. Ez leegyszerűsíti a kapcsolódó szakmai bírálat folyamatát és csökkenti a kutatók elé gördülő akadályok számát.

- Az általános kutatási repozitóriumok, amelyeknek sokféle adatot kell tartalmazniuk, rendelkezhetnek bizonyos korlátozásokkal a QA / QC munkafolyamatok tekintetében, és ezeket egyértelművé kell tenni.

- A minőség-szint-meghatározásoknak, a minőség-értékelési eljárásoknak és az eredményekkel kapcsolatos információknak egyértelműnek és könnyen hozzáférhetőnek kell lenniük felhasználók számára (és valószínűleg harmadik felek számára is, például összesítő vagy metrikus szolgáltatások felé).

Úgy tűnik, hogy tendencia mutatkozik a kutatási munkafolyamatban az adatok korai megosztására, olyan szakaszban, ahol az adatok még dinamikusak (lásd például Meehl et al.). Ezért szükség van olyan minőségbiztosítási / minőségellenőrzési eljárásokra, amelyek képesek dinamikus adatokat kezelni.

Adatkezelés és hosszú távú archiválás

Az adatnyilvántartási és -adatkezelési tevékenységek magukban foglalhatják a különféle fájlformátumok és formátumok kezelését, hozzáférési szintű korlátozások létrehozását, embargó eljárások létrehozását és végrehajtását, valamint az azonosítók kiosztását. Gyakorlatok széles választékát találtuk mindezen területeken. Ezek a különféle fájlformátumokkal kapcsolatos irányelvek megadásától az aktív fájlkonvertálásig terjednek; a hozzáférési korlátozások támogatásától a kizárólag a nyílt hozzáférés támogatásáig; rugalmas vagy szabványos embargóperiódusok adminisztrációján; és különféle típusú azonosítókat alkalmazását is beleértve. Több tudományterület-specifikus repozitórium már régóta megőrzi az adatokat, és részletes munkafolyamatokkal rendelkezik az archívumi megőrzésére. Más repozitóriumok meglehetősen újak ebben a párbeszédben, és továbbra is keresik a lehetséges megoldásokat.

A mintáink többségében található repozitóriumok elkötelezettséget mutattak a szabványok alkalmazása mellett. A bevált gyakorlatok és szabványok alkalmazása növeli annak valószínűségét, hogy a közzétett adatokat a későbbiekben is megőrzi, és ez kölcsönösen átjárható és fenntartható adatok publikálásához vezet. A repozitórium-hitelesítési rendszerek az

utóbbi években lendületet kaptak, és az adatpublikálási partnerekkel, mint a finanszírozókkal, kiadókkal és adatrezitóriumokkal való együttműködés révén elősegíthetik az adatok publikálását. A hitelesítési sémák közé tartoznak azok, amelyeket olyan szervezetek hajtanak végre, mint például a Data Seal of Approval (DSA) és a World Data System (ICSU-WDS). Az ilyen szabványok fokozott átvétele nagy hatással lenne az kölcsönösen átjárható és fenntartható adat publikációra.

Terjesztés, hozzáférés és idézés

Az általunk elemzett legtöbb rezitóriumi adatkészlet egyetlen oldalon voltak összefoglalva, amely általában néhány alapvető vagy bővített (ha nem minőség biztosított) metaadatot kínált. Ez általában tartalmazott egy DOI-t és néha ehelyett egy másik egyedi azonosítót is. Megállapítottuk, hogy az állandó azonosítók széles körben használtak, és felismertük, hogy az adatoknak citálhatóknak kell lenniük, ha azt akarjuk, hogy hasznosak legyenek. Meg kell jegyezni, hogy az adatpublikációs termékek terjesztését egyes esetekben a hagyományos folyóiratokkal való összekapcsolás és közzététel (például beágyazott megjelenítés) révén javították. Ez fontos, különös tekintettel arra, hogy szükségessé vált a kutatási közösségek kultúrájának változása, melyben az adatok publikálása normává kell váljon.

A terjesztési gyakorlatok széles körben változtak. Számos rezitórium támogatta a nyilvánosan hozzáférhető adatokat, de különbözött abban, mennyire optimálisan indexelik őket a megtalálhatósághoz. Ahogy az várható lenne, az adatfolyóiratok általában kapcsolódtak a keresőmotorokhoz, valamint az absztrakt készítő és indexelési szolgáltatásokhoz. Ezek azonban gyakran (ha nem mindig) az adatkikre vonatkoztak, nem pedig az adatkészletre mint olyan. A Thomson Reuters által elindított Data Citation Index és az olyan projektek, mint például az Data Discovery Index, felhívják a figyelmet az adatkeresés kihívásának fontosságára, és felgyorsíthatják a paradigmaváltást, hogy bevezessék az adatpublikálást a kutatói közösségekben.

Egy ilyen paradigmaváltás történt 2014-ben, amikor a Resource Identifier Initiative (RII) új nyilvántartást indított az orvosbiológiai szakirodalomban. A projekt fedezte az antitesteket, modell organizmusokat (egerek, zebrahal, legyek) és eszközöket (azaz szoftvereket és adatbázisokat), amelyek egészen átfogó kombinációt biztosítanak az adatokhoz, metaadatokhoz és platformokhoz, amellyel dolgozhatnak. Tizennyolc hónappal később a projekt képes volt beszámolni mind a kutatók viselkedésének kulturális változásáról, mind a vonatkozó kutatások

lehetséges reprodukálhatóságának jelentős növekedéséről. Amint azt Bandrowski és munkatársai tárgyalják, a kezdeményezés sikere az elfogadás és megértés megszerzése szempontjából kritikus tényező, az integrált megvalósítás volt. Az érdekelt felek csoportja, beleértve a kutatókat, a folyóirat szerkesztőket, az adott téma közösségének vezetőit és a kiadókat - egy meghatározott tudományágon belül, az idegtudományban - együtt dolgoztak a következetes üzenet biztosítása érdekében. Ez látványos indoklást, koherens folyóiratpolitikát (amely szükségessé tette az összeegyeztetést a leendő szerzők számára, hogy publikálhassanak), valamint speciális munkafolyamatot biztosított a regisztrációs folyamathoz (szükség esetén szakképzett, emberi támogatással kiegészítve). További munkára van szükség annak meghatározásához, hogy ezt a felhasználási esetet pontosan hogyan lehet kihasználni az alanyok, közösségek és más szereplők körében.

A FAIR alapelvei és egyéb szabályzatok határozottan említik, hogy az adatoknak hozzáférhetőnek kell lenniük. Az adat publikálási megoldások biztosítják ezt, de egyes munkafolyamatok csak bizonyos felhasználók számára teszik lehetővé az érzékeny adatok elérését. Példa erre azon kutatási adatok, amelyek olyan információkat tartalmaznak, melyek a válaszadók azonosításához vezethetnek. Ilyen esetekben egy leendő adatfelhasználó hozzáférhet a részletes kutatási metaadatokhoz annak meghatározása érdekében, hogy megállapítsa megfelelő kutatási igényeinek, de az adatkészlethez való hozzáférés megadása előtt adathasználati megállapodást kell aláírni. A metaadatokat, az adatsíkkéket vagy a leírókat nyílt hozzáféréssel lehet publikálni, esetleg Creative Commons licenccel, de az alapul szolgáló adatkészlet nem érhető el, kivéve regisztráció vagy más engedélyezési folyamat révén. Ebben az esetben az adatlap lehetővé tenné a közreműködő kutatók számára, hogy megfelelő kreditet szerezzenek, és egyszerre megkönnyítené az adatkeresést és újrahazsnálatot.

A hivatkozási szabályok, valamint a gyakorlat közösségenként és kultúránként is eltérő. A folyóiratok és kiadók egyre inkább belefoglalják az adatok idézésére vonatkozó irányelveket a szerzői támogató szolgáltatásaikba. A bevált gyakorlat vagy szabvány szempontjából a *Joint Declaration of Data Citation Principles* (Adatmegjelölési Alapelvek Közös Nyilatkozata) gyűjti a kritikus tömeget, és kezd általánosan elismertté válni. Párbeszéd alakul ki a részletesebb közösségi gyakorlatokról; például arról, hogy az adatkészletek és az adatpapírok publikálása - amelyet azután külön lehet hivatkozni a kapcsolódó elsődleges kutatási dokumentumoktól -

tisztességes gyakorlat-e egy olyan rendszerben, amely jutalmazza a magasabb hivatkozási arányokat. Az ésszerű gyakorlatok azonban megfogalmazhatók.

Egyéb potenciális hozzáadott értéket képviselő szolgáltatások és mutatók

Számos repozitórium- vagy folyóirat-szolgáltató túltekint azon munkafolyamatokon, amelyek információkat gyűjtenek a kutatási adatokról, és szeretnék ezeket az információkat láthatóvá tenni a terület más információszolgáltatói számára is. Ez hozzáadott értéket jelenthet a közzeendő adatokhoz. Ha az információ standardizált módon kerül közzétételre, akkor az adatok indexelhetők és harmadik fél által működtetett szolgáltatók számára is felderíthetők lehetnek pl. Adat összesítők (1. ábra). Tekintettel arra, hogy az ilyen adatösszesítők gyakran az eredeti adatszolgáltató hatáskörén vagy intézményi fókuszán túl dolgoznak, egyes adatszolgáltatók gazdagítják metaadataikat (például az adatpublikálás hivatkozásaival, kulcsszavakkal vagy részletesebb tárgykörrel), hogy lehetővé tegyék a jobb interdiszciplináris visszakeresést. Ideális esetben a kutatók kapnak visszajelzést arról, hogy mások hogyan töltik le vagy használják az adatokat. Ezen felül az olyan szolgáltatások, mint az ORCID, integrálva vannak annak érdekében, hogy a kutatók az anyagaikat különböző platformokon keresztül összekapcsolhassák. Ez nagyobb láthatóságot biztosít az adatoknak a különböző nyilvántartásokon keresztül, és lehetővé teszi a szerzők globális egyértelműsítését. Ez utóbbi különösen fontos a szerzői mutatók kialakításához. Vizsgálatunk során számos adatrepozitórium és adat folyóirat-szolgáltató érdeklődést mutatott az adatkészletek és a kapcsolódó objektumok új mutatói iránt. A megosztott anyagok felhasználásának, hatásainak és újrafelhasználásának nyomon követése gazdagíthatja az eredeti platformok tartalmát, és arra ösztönözheti a felhasználókat, hogy vegyenek részt további adatmegosztási vagy kezelési tevékenységekben. Ez az információ minden bizonnyal érdekes az infrastruktúra és a kutatás finanszírozói számára.

A munkafolyamatok sokfélesége

Noha a munkafolyamatok meglehetősen egyértelműnek tűnnek és valamelyest hasonlóak a hagyományos statikus közzétételi eljárásokhoz, az alapjául szolgáló folyamatok valójában meglehetősen összetettek és változatosak. A sokszínűség a kezelés területén volt a legszembetűnőbb. Azok a repozitóriumok, amelyek önálló publikálási lehetőségeket kínáltak kezelés nélkül, korlátozott eljárásokkal rendelkeztek, kevesebb erőforrást igényeltek, de potenciálisan kevesebb kontextuális információt és kevesebb minőségbiztosítást is nyújtottak. Azoknak a tudományági repozitóriumoknak, amelyek kiterjedt kezelést és minőségbiztosítást

végeztek, összetettebb munkafolyamatai voltak, esetenként egymást követő lépésekkel. Ezek elősegíthetnek több együttműködést a folyamat elején, vagy tartalmazhatnak szabványosított megőrzési lépéseket.

A metaadat heterogenitás volt megfigyelhető a tudományág-specifikus repositóriumok között. A magasan specializált repositóriumok gyakran a meghatározott metaadat-sémákra összpontosítottak, és ennek megfelelően folytattak kezelést. Egyes tudományágak metaadat-szabványokat alakítottak ki, hasonlóan a társadalomtudományok által alkalmazott Data Documentation Initiative szabványhoz. Ezzel szemben az általánosabb repositóriumok hajlamosak voltak konvergálni a domain-agnosztikai metaadat-sémákat, a tudományágakban közös mezőkkel, pl. a kötelező DataCite mezők.

Az adat folyóiratok az általános munkafolyamatokban hasonlóak, de a támogatás, az bíráló és a kezelés szintjében különböznek. Mint a repositóriumoknál, minél speciálisabb a folyóirat (például a földtudomány tudományága, előre meghatározott adatmegosztási gyakorlatokkal rendelkezik), annál előiróbb a szerzői útmutatás, és annál specializáltabbak a bírálati és minőségbiztosítási folyamatok. A nyilvános vagy a publikálást követő szakmai bíráló terjedésével egyes adat folyóiratok a szélesebb közösséget is a publikálási folyamatban való részvételre buzdítják.

A szélesebb kutatói közösség és néhány tudományág-alapú közösség jelenleg a kutatási adatok szabványos kiadásának kritériumait és gyakorlatait dolgozza ki. Az ezeket az erőfeszítéseket támogató szolgáltatások is - akár repositóriumok, akár folyóiratok - általában a késztermékek helyett a folyamatban lévő munkákat vagy a koncepciót bizonyító gyakorlatok jeleit mutatják. Ezt tükrözi az elemzési adatkészletünk. Részben a bírálati időszak (2015. február 1. - 2015. június 30.) előrehaladásának állapotától, valamint a tárgyterület sajátosságától függően, néhány munkafolyamat-bejegyzés meglehetősen homályos volt.

Összegzés és **következtetések**
Noha elemzésünk eredményei nagy változatosságot mutatnak az adat közzététel munkafolyamataiban, a kulcskomponensek meglehetősen hasonlóak voltak a szolgáltatók között. A közös alkotóelemeket az adatpublikálásra szolgáló referenciamodellbe csoportosítottuk és ábrázoltuk. Tekintettel az ezen a területen zajló gyors fejlődésre és a tudományági különbségekre, a munkafolyamatok különbségében

további növekedés várható. Az RDA munkacsoportján keresztül törekszünk további közösségi bírálatra és az általános referenciamodell alkotóelemeinek jóváhagyására, és hasonló tudományági variációk elemzésének kivitelezésére. Tanulmányunk eredményei azonban azt sugallják, hogy új megoldások (például az alulreprezentált tudományágak esetében) építhetnének az azonosított kulcskomponensekre, amelyek a legjobban megfelelnek a felhasználásuk esetének. Néhány nyilvánvaló hiányosság és kihívás (melyeket alább leírtak) akadályozzák a globális kölcsönös átjárhatóságot és a közös modell átvételét.

Hiányosságok

és

kihívások

Bár továbbra is sok olyan tudományág létezik, amelyekre vonatkozóan nem létezik konkrét domain-repozitórium, egyre több különféle típusú repozitóriumot látunk (a re3data.org több mint 1200 repozitóriumot indexel). A tudományágak repozitóriumai mellett számos új repozitóriumot is létrehoztak, amelyeket szélesebb gyűjtemények tárolására terveztek, pl. Zenodo, Figshare, Dryad, Dataverse, valamint a főiskolák és egyetemek intézményi repozitóriumai. Szintén létrejönnek „átmeneti” repozitóriumok, amelyek kiterjesztik a hagyományos munkafolyamatokat az együttműködési munkatérre - pl. Open Science Framework, amelynek kiadási munkafolyamata van az Dataverse-vel. Egy másik példa a SEAD (Sustainable Environment Actionable Data) projekt, amely olyan projekttereteket biztosít, amelyekben a tudósok adatokat kezelnek, találnak és megosztanak, és amely összekapcsolja a kutatókat a repozitóriumokkal, amelyek hosszú távon biztosítják az adatok elérését és megőrzését.

A sok új adatpublikálási tevékenység ellenére is, az esettanulmányok elemzése azt állapította meg, hogy a kihívások továbbra is fennállnak, különös tekintettel a bonyolultabb munkafolyamatokra. Ezek tartalmazzák:

- Kétirányú összeköttetés. Hogyan lehet az adatokat és a kiadványokat folyamatosan, automatikusan összekapcsolni? Több szervezet, köztük az RDA és a WDS, most dolgozik ezen a problémán. Ehhez kapcsolódó kérdés maguk a kapcsolatok megmaradása.
- Szoftver menedzsment. Megoldásokra van szükség a szoftverek kezeléséhez, megőrzéséhez, publikálásához és idézéséhez. Alapvető munkafolyamatok léteznek (a kódmegosztó platformok, a repozitóriumok és összesítők), de sokkal több munkára van szükség egy szélesebb keret létrehozásához, beleértve a közösség frissítését és a kapcsolódó adatokhoz való kapcsolódást magában foglaló kezdeményezéseket.
- Verziószabályozás. Általánosságban azt tapasztaltuk, hogy a repozitóriumok különféle módon kezelik a verziószabályozást, ami potenciálisan zavaró lehet. Noha néhány verziószabályozó megoldást a tudományterület-specifikus kihívásokhoz lehet igazítani, szükség van egységesítésre. Ez a kérdés az eredeti információkra is vonatkozik.

- Korlátozott felhasználású adatok megosztása. A tárolók és folyóiratok általában még nincsenek felszerelve a bizalmas adatok kezelésére. Fontos, hogy az adatmegosztási mechanizmus megfeleljen az adatok érzékenységi szintjének. Megérett az idő a szakértelem cseréjére ezen a területen.
- A szerep tisztasága. Az adatpublikálás az együttműködésre épül. A jobb felhasználói útmutatásért és a szolgáltatásokba vetett nagyobb bizalom érdekében a szerepkörök, a felelőségek és az együttműködés jobb megértése szükséges. A „ki mit csinál” dokumentálása a jelenlegi, közép és hosszú távon garantálja a zökkenőmentesebb szolgáltatást.
- Üzleti modellek. Nagy az érdeklődés a repozitóriumok értékének és fenntarthatóságának létrehozására. Beagrie és Houghton előállította az adatközponti tanulmányok szintézisét, amelyben mennyiségi és minőségi megközelítéseket kombináltak annak érdekében, hogy az értéket gazdasági szempontból számszerűsítsék, és bemutassák az egyéb, nem gazdasági hatásokat és előnyöket. A nemrégiben a Sloan által finanszírozott, 22 repozitóriumból álló találkozó egy tanulmány elkészítéséhez vezetett a Digitális Adatok Domain repozitóriumi Fenntartása területén. Sokkal több munkára van szükség az adatok közzétételéhez szükséges életképes pénzügyi modellek megértéséhez és a megbízható együttműködések megkülönböztetéséhez.
- Az adatok idézésének támogatása. Noha széles körű tudatosság látszik, az adatidézés végrehajtó csoportja (Data Citation Implementation Group) által ajánlott gyakorlatok és eljárások csak részben valósulnak meg. Számos PID jelenik meg, beleértve az ORCID, DOI, FunderRef, RRID, IGSN, ARK és még sok más. A tisztaságot és egyszerű használatot meg kell valósítani.
- Mutatók. Az adatot létrehozóknak, az intézményeiknek és finanszírozóiknak tudniuk kell, hogyan és milyen gyakran használják újra adataikat.
- Ösztönzők. Az adatpublikálás potenciális ösztönzőket kínál a kutatók számára, pl. Idézhető adatkészlet, állandó adatközlés és információk a kutatás hatásáról. Számos adattár is támogatást nyújt az adatok benyújtásához. Az adatpublikálás előnyeit jobban kell hangsúlyozni a kutatók felé. Ezenkívül az érdekelt feleknek terjeszteniük kell azt a tényt, hogy a formális adat-archiválás több iratot eredményez és így több tudományt, mint azt Piwowar és Vision, valamint Pienta et al. bizonyították. Nagyobb tisztaságot kell biztosítani a kutatási adatok hatásainak intézményes és finanszírozói felismerése tekintetében is.

A bonyolultabb adatok - különösen a nagy adatok és a dinamikus adatok - kihívásaival is foglalkozni kell. Míg az elmúlt 10 év folyamatainak középpontjában a visszavonhatatlan, teljesen dokumentált adatok állnak a korlátozás nélküli (kutatási) felhasználás céljából, az adatok közzétételének „jövőbiztosnak” kell lennie (Brase et al.). A kutatóközösségek elvárják az adatok idézését, még mielőtt azok elérnék az általános visszavonhatatlan állapotot, és mielőtt archiválnák azokat. Ez különösen igaz a nagy mennyiségű adatokkal rendelkező közösségekre (például nagy energiájú fizika; éghajlati tudományok), valamint az

adatcítáló entitásokra, amelyek több különálló adatkészletet tartalmaznak, amelyek esetében az általános stabil adatgyűjtés eléréséhez szükséges idő hosszú. Annak ellenére, hogy esettanulmányunk elemzése kimutatta, hogy az adatidézési munkafolyamatokat végrehajtják az adat publikálásban részt vevő számos csoport figyelembe veszi, a dinamikus adatok idézésével kapcsolatos kihívásokra nem kerül kellő figyelem. A verzió szabályozás és az adatkészletek eredet nyilvántartásának megfelelő vezetése szintén kritikus fontosságú az ilyen adatgyűjtések idézéséhez, és nélkülözhetetlen részei az adatpublikálási munkafolyamatnak.

A hiányosságok és a kihívások tekintetében felismerjük, hogy az elemzett esettanulmányok hatálya korlátozott. Ez kapcsolódik ahhoz az általános kihíváshoz-, amellyel a projekt során szembesültünk: nehéz a világos és következetes ember által olvasható munkafolyamat-ábrázolásokat találni a repozitóriumok számára. A megbízhatósági szabványok (például a Data Seal of Approval, a Nestor, az ISO 16363 és a World Data System) megkövetelik, hogy a repozitóriumok dokumentálják folyamataikat, így ez a jövőben változhat, de hozzátennék azt a javaslatot, hogy a repozitóriumok szabványos módon tegyék közzé munkafolyamataikat a nagyobb átláthatóság kedvéért. Ez erősítené a repozitóriumokba vetett bizalmat, és növeli a felhasználók részvételét. A sokféleség, amelyet találtunk, nem meglepő, és nem is szükségszerűen nemkívánatos. Az esettanulmányok és az adatgyakorlatok néprajzai azt mutatták, hogy a repozitóriumok „upstream” adatkezelésével kapcsolatos munkafolyamatok nagyon változatosak. Az adatmegosztási gyakorlatok sok esetben jelentősen eltérnek a szubdiszciplináris szinten (pl. Cragin et al.), így valószínűleg továbbra is fenn kell tartani a különféle megközelítések és a tájékozott választás támogatását, nem pedig az egységes vagy monolitikus modelleket (Pryor.). Elemzésünk azt mutatja, hogy munkafolyamatok sokfélesége fejlődött, és egyre több új jelenik meg, így a kutatók képesek lehetnek a legmegfelelőbbek kiválasztására olyan útmutatás alapján, amely megkülönbözteti a releváns jellemzőket, mint például a minőségbiztosítás / minőségbiztosítás és a különböző szolgáltatási vagy támogatási szintek.

A bevált gyakorlatokkal kapcsolatos ajánlások és következtetések

A kiválasztott esettanulmányok alapján meghatározásra kerültek az adatpublikálás kulcsfontosságú elemei, amelyek referenciamodellként eredményeznek az adatok publikálásában. Az elemzés és különösen az adatközlési munkafolyamatokban kulcsfontosságú résztvevőkkel folytatott beszélgetések rámutattak a bevált gyakorlatokra, amelyek hasznosak lehetnek az új munkafolyamatokat létrehozó szervezetek és a meglévő eljárások átalakítását vagy szabványosítását célzó szervezetek számára:

- € Kicsivel kell indítani és egyenként modulárisan építkezni, megértve, hogy az egyes építőelemek hogyan illeszkednek az általános munkafolyamathoz és a végső célhoz. Ezeknek az építőelemeknek nyílt forrású / megosztható alkotórészeknek kell lenniük.
- € Ha csak rendelkezésre áll, követni kell a szabványokat a kölcsönösen átjárhatóság megkönnyítése és a kiterjesztése érdekében, azok munkája alapján akik ugyanazokat a szabványokat alkalmazták. Például a Dublin Core egy széles körben használt metaadat-szabvány, amely megkönnyíti a metaadatok megosztását más rendszerekkel. Használjon fegyelmi előírásokat, ahol / amikor alkalmazhatók.
- € Különösen fontos az adatok idézésére vonatkozó szabványok végrehajtása és betartása, ideértve az állandó azonosítók (PID) használatát is. Az adatok és a publikációk közötti kapcsolatok automatikusan kinyerhetők, ha az adatokhoz használt DOI-kat rutinszerűen használják a papírokból. A kutatói PID-k, például az ORCID használata szintén kapcsolatot létesíthet az adatok és a dokumentumok, vagy más kutatási entitások, például a szoftver között. A PID-k lehetővé teszik a kapcsolt nyílt adatok funkcionalitását is.
- € Dokumentum szerepek, munkafolyamatok és szolgáltatások. A munkafolyamatok elemzésének egyik fő nehézsége a teljes, szabványosított és naprakész információ hiánya volt a maguk a platformok által nyújtott folyamatokról és szolgáltatásokról. Ez a szolgáltatások potenciális felhasználóit is érinti. A megbízható repozitóriumi hírnév fejlesztésének egy részének tartalmaznia kell egy rendszert, amely tisztázza a támogatási szinteket, a hosszú távú fenntarthatósági garanciákat, az alany szakértői erőforrásait stb.

Összegezve, a bemutatott referenciamodell és a bevált gyakorlatok ötletét követve egy olyan munkafolyamatot szeretnénk látni, amelynek eredményeként minden tudományos feladat összekapcsolódik, idézhető és lehetővé teszi, hogy a kutatók könnyedén tudjanak navigálni, mindemellett ismételt kutatásokat tesz lehetővé. Ez magában foglalja a dokumentációt, a kódot, az adatokat és a folyóirat cikkek közötti kapcsolatokat egy integrált környezetben. Ezenkívül az ideális munkafolyamatban ezeket a tárgyakat jól kell dokumentálni, hogy más kutatók (vagy amatőr tudósok stb.) újból felhasználhassák az adatokat új felfedezésekhez. Szeretnénk, ha az információkat API-k és más mechanizmusok révén szabványosítanánk és tennék közzé, hogy az adathasználat mutatói rögzíthetők legyenek. Megjegyezzük azonban, hogy a finanszírozás és az akadémiai jutalmazási rendszereknek első osztályú feladatként kell értékelniük az adatvezérelt másodlagos elemzést és a meglévő adatok újrafelhasználását, valamint az adatok közzétételét. A támogatók nagyobb figyelme (azaz jobban felismert érték) kulcsszerepet fog játszani a paradigma megváltoztatásában.

Az egyik nagy kihívás az, hogy intenzívebb együttműködésre van szükség az érdekelt csoportok között. Például a repositóriumok és a felsőoktatási intézmények (amelyek a kutatási adatok kritikus tömegét birtokolják) és a nagy folyóiratkiadók (amelyek a felfedezésre váró, közzétett kutatások kritikus tömegét tárolják) még nem kötelezték el magukat teljes mértékben egymás irányába. Noha új folyóiratformátumokat fejlesztenek ki, amelyek összekapcsolják az adatokat a lapokkal ezzel gazdagítva az olvasási élményt, a kulturális, műszaki és üzleti modell kérdései továbbra is akadályozzák az előrelépést.

Bemutattuk, hogy az adatközlési rendszer különféle alkotóelemeinek, ahol lehetséges, zökkenőmentesen és integrált környezetben kell működniük. Ezért támogatjuk a szabványok bevezetését és szükség esetén új szabványok kidolgozását a repositóriumok számára valamint az adatpublikálás folyamatának minden lépéséhez. Az adatpublikálást be kell ágyazni a dokumentált munkafolyamatokba, hogy elősegítsék az együttműködést a potenciális partnerekkel, és irányítsák a kutatókat, lehetővé téve és ösztönözve az újrafelhasználható kutatási adatok tartós tárolását, az eredet megőrzése közben.

Hivatkozások