The background features a stylized tree with a trunk and branches in shades of blue, green, and yellow. The tree is surrounded by several overlapping circles in various colors (purple, blue, green, yellow). In the center, there is a circular logo with a white dot in the middle, surrounded by concentric circles and dots, resembling a molecular or atomic structure.

**A tudományos
adatok újrafelhasznál-
hatóságának és
FAIR kezelésének
irányelvei**

PARTHENOS

Fellelhető

Hozzáférhető

Interoperábilis

Újrafelhasználható

Az irányelvekről

A jelen útmutató abban kíván tanácsokat adni, hogy hogyan hangoljuk össze a társadalom- és bölcsészettudományokban a kutatási adatok előállítóinak, archiválóinak és újrafelhasználóinak munkáját annak érdekében, hogy a kutatási adatok minél jelentősebb része újrafelhasználható legyen.

Az irányelveket összefoglaló útmutató a PARTHENOS projektben részt vevő több, mint 50 tag együttes munkájának eredménye. A munkacsoport arra kereste a választ, hogy milyen közös vonások és jó gyakorlatok lelhetők fel a tudományos adatok kezelésével foglalkozó vezérelvek alkalmazásában. Munkájuk során egyaránt támaszkodtak a fellelhető szakirodalomra, továbbá különböző szakemberek körében végzett kérdőíves és interjú vizsgálatokra. A munka során közel 100 adatkezelési irányelvet tekintettek át (beleértve formális és megegyezően alapuló irányelveket preferált formátumokra, adatellenőrzési irányelvekre, legjobb módszerekre stb.)

A (meta)adatok és a repozitóriumok minőségbeli kritériumaira összpontosítva a PARTHENOS csoport 20 olyan alapelvet azonosított, amelyek egyaránt érvényesek a társadalom- és bölcsészettudományok tudományterületeire.

A könnyen követhetőség és idézhetőség érdekében az egyes alapelveket egy általánosabb elméleti keretbe, a FAIR-elvekbe - FAIR principles: Fellelhető (Findable), Hozzáférhető (Accessible), Interoperábilis (Interoperable) és Újrafelhasználható (Reusable) - ágyazva vesszük sorra. A modern tudományosság alapértékeit magukba foglaló FAIR alapelveket 2016-ban publikálta a FORCE11 csoport elsősorban a tudományos adatok minőségét és újrafelhasználhatóságát elsősegítendő.

Valamennyi PARTHENOS alapelvet a tudományos adatok előállítóinak, (újra)felhasználóinak valamint az adatok archiválásáért, kezeléséért felelős szakembereknek szóló speciális ajánlásokkal egészítettünk ki. Az alapelvek alatti ikonok mutatják, hogy melyik célcsoportnak szólnak az egyes ajánlások.



A lámpa ikon az adatok előállítóinak, kurátorainak és (újra)felhasználóinak, például a történettudomány, régészet, nyelvészeti diszciplínák, vagy a társadalomtudományok területén dolgozó kutatói közösségeknek szóló ajánlásokat jelzi.



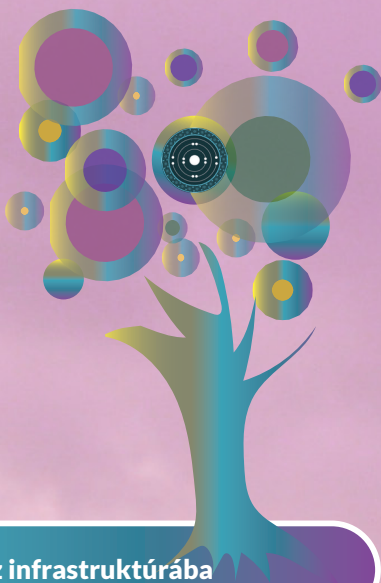
A kerék ikon a kutatási infrastruktúráknak, tudásközpontoknak, kutatóintézeteken belül működő archívumoknak vagy a közgyűjteményeknek (könyvtáraknak, levéltáraknak, múzeumoknak) szóló ajánlásokat jelzi.



A PARTHENOS egy tizenhat európai kutatóintézetet és tudásközpontot magába foglaló tudományos konzorcium. A PARTHENOS tagok célja, hogy a tudományos adatok újrafelhasználhatóságának növelése érdekében kapcsolatot képezzenek a kutatói közösségek, repozitóriumok, kutatási infrastruktúrák és örökségvédelmi intézmények adatainak életciklusai között a társadalom- és bölcsészettudományokban.

20 irányelv

a tudományos adatok újrafelhasználhatóságáért és FAIR kezeléséért



1

Invesztáljon szakemberekbe és az infrastruktúrába

A további irányelv tizenkilenc alkalmazásához szükséges fontos előfeltétel, hogy investáljunk a modern adatkezeléshez szükséges infrastruktúrába, továbbá, hogy adatszakértőket képezzünk és alkalmazzunk.



*Ismerje meg a tudományos adatkezelést érintő legjobb gyakorlatokat. Ehhez ajánljuk a **PARTHENOS** adatkezeléssel foglalkozó online tananyagát vagy a **CESSDA Data Management Expert Guide** -jét.*



Fektessenek be adatszakértők képzésébe és alkalmazásába és határozzanak meg a költségvetésében egy technikai infrastruktúrára és az azt üzemeltető munkaerőre fordítható összeget.

Fellelhetőség

A tudományos adatoknak könnyen fellelhetőnek és könnyen azonosíthatóknak kell lenniük mind az emberek, mint pedig számítógépek számára. Ehhez jó minőségű és kellően részletes metaadatok szükségesek, amelyek lehetővé teszik a releváns adatok felfedezését.

2

Használjunk állandó azonosítókat (Persistent Identifiers, PIDs)

Az adatok helyének meghatározása minden további adatkezeléssel kapcsolatos művelet alapfeltétele a hozzáféréstől az újrahaznosításig. Az egyes adatelemek akkor lesznek egyértelműen fellelhetőek és egyedien azonosíthatók hosszú távon is, ha állandó azonosítóval (PID) vannak ellátva. Ezek az azonosítók akkor is garantálják az adatelemek megtalálhatóságát, ha azok eredeti webcíme megváltozik. A PID-ek különböző formákat ölthetnek, lehetnek például DOI-k, PURL-ok, URL-ek vagy Handle-ök.



Tüntesse fel az adatelemeihez tartozó PID-eket a tudományos publikációiban.



Jelöljenek ki egy, az Önök gyűjteményeihez és infrastruktúrájához illő állandó azonosító sémát és rendeljenek azonosítókat a gyűjtemények/adatállományok egyes darabjaihoz. Az NCDD PID kalauza segít a megfelelő PID rendszer kiválasztásában.

3

Hivatkozás a kutatási adatokra

Könnyebben megtalálhatók a kutatási adatok, ha állandó azonosítóval vannak ellátva és ha a közösségi standardoknak megfelelően hivatkozunk rájuk.



Ismerkedjen meg az Ön tudományterületén használatos adathivatkozási irányelvekkel és ezeket követve hivatkozzon az Önnek releváns kutatási adatokra.



Tájékoztassa a kutatói közösségeket az adathivatkozás bevált gyakorlatairól és könnyítse meg a kutatási és forrásadatok hivatkozását például egy 'Hogyan idézzük ezt az adatsort' gomb alkalmazásával.

4

Használjunk állandó szerzői azonosítókat

Egy állandó szerzői azonosító (mint például a VIAF, ISNI vagy ORCID) segít összekapcsolni egymással a különböző adatsorokat, kutatási tevékenységeket, publikációkat a kutatóikkal, szerzőikkel és lehetővé teszi elismerésüket és fellelhetőségüket.



Különböztesse meg magát a többi, esetleg hasonló nevű kutatótól vagy kutatócsoporttól. Ha még nincs állandó szerzői azonosítója, kérvenyezzen egyet, és tüntesse fel az adatainál és eredményeinél.



Tüntesse fel a szerzői azonosítókat a metaadatokban.

5

Válasszunk megfelelő metaadat-sémákat az adatok leírásához

A metaadatoknak különösen fontos szerepük van az adatok fellelhetőségében. Különösen igaz ez a leíró és a hivatkozást segítő metaadat komponensekre. A metaadat séma egy tulajdonképpeni lista, amely az adatok, források leírásához szükséges elemeket tartalmazza standardizált formában, mint például cím, azonosító, készítő neve vagy a létrehozás dátuma. A már létező, szélesebb körben használatos metaadat sémák alkalmazása elősegíti a könnyebb adatcserét nemzetközi kontextusokban is.



A minél könnyebben fellelhetőség érdekében a lehető legteljesebben és legkövetkezetesebben dokumentálja a kutatási adatokat. Gondoljon arra, hogy később más kutatók vagy felhasználók ezek alapján az információk alapján fognak hozzáférni az adatsorhoz és értelmezni azt. Amennyiben lehetséges, használjon már létező, bevett, és az adat típusához illő metaadat-sémákat.



Tegyék egyértelművé, mely metaadat-sémát alkalmazzák és ajánlják a kutatói közösségnek. Az adatsorokat gazdagítandó, az adatok repozitóriumban való elhelyezésekor ajánljuk egy kiegészítő adatok (például keletkezéstörténeti információk) gyűjtésére szolgáló űrlap alkalmazását.

Hozzáférhetőség

A kutatási adatoknak könnyen hozzáférhetőnek és kinyerhetőnek kell lenniük pontosan meghatározott hozzáférési feltételekkel és standardizált elérési csatornákkal.

6

Válasszunk megbízható repozitóriumot a kutatási adatoknak

A hitelesített repozitóriumok megbízható otthont adnak az adatainknak. A hitelesítés garantálja az adatok biztonságos tárolását, elérhetőségét, fellelhetőségét hosszú távon is. Ilyen hitelesítési tanúsítványok például a CoreTrustSeal, a nestor seal vagy az ISO 16363-as tanúsítvány.



Tegye elérhetővé repozitóriumok egy megbízható repozitóriumon keresztül. A repozitórium által használt standardok követésével (preferált formátum, metaadat-sémák etc.) egyúttal a FAIR követelményeknek is megfelelhet.



Egyértelműen tüntessék fel a repozitórium minősítését igazoló tanúsítványt a honlapjukon. Ha a repozitórium még nem rendelkezik ilyen tanúsítvánnyal, írják le, hogyan kívánják biztosítani az adatok hosszú távú elérhetőségét, hozzáférhetőségét, fellelhetőségét, és újrahasznosíthatóságát.

7

Egyértelműen fogalmazzuk meg az adatokhoz való hozzáférés feltételeit

Amikor az adatok bekerülnek repozitóriumba, tisztázni kell, hogy milyen hozzáférési lehetőségeket adhat meg a szerző vagy kurátor az adatokhoz vagyis hogy mások milyen feltételek mellett férhetnek hozzájuk.



Amikor hozzáférési lehetőséget választ, vegye figyelembe a jogi előírásokat, a szakterület-specifikus eljárás módokat és adott esetben az etikai protokollt. Ha lehet, válassza az adatokhoz való nyílt hozzáférést (Open Access). Ha személyes adatokkal dolgozik, alaposan tervezze meg, hogy pontosan mihez kéri az adatszolgáltatók hozzájárulását. Győződjön meg róla a megosztás előtt, hogy az így gyűjtött adatok semmi olyan részletet nem tartalmaznak, amelyek alapján személyes adatok nyilvánossá válhatnak és egyértelműsítse, hogy milyen intézkedéseket tett a személyes adatok védelme érdekében.



Ösztönözzék, támogassák a (meta)adatok szabad hozzáférést (Open Access) publikálását. Közöljék egyértelműen, ha az érzékeny (meta)adatok csak korlátozással férhetők hozzá és jelezzék, ha valamely információ nem integrálható a nyilvánosan hozzáférhető metaadatok közé. Ebben az esetben törekedjenek arra, hogy ezek az érzékeny adatok egy alaposan dokumentált és kontrollált hozzáférési eljáráson keresztül válhassanak elérhetővé.

8

Ha szükséges, használjunk adatszárlatot

Az adatok megosztására, szabad hozzáférhetőségére vonatkozó zárlatot hatálya alatt csak az adatsorok jellemzése válik publikussá, maguk az adatok nem. A (meta)adatok egy meghatározott időperiódus leteltével válnak csak elérhetővé.



Ha az adatok megosztására vonatkozóan zárlatot alkalmaz, pontosan jelezze, miért van erre szükség és milyen időszakra szól. Tegye a (meta)adatokat elérhetővé milyen hamar csak lehet.



Határozza meg, hogy lehetséges-e adatebargó használata az Önök rendszerében és ha igen, milyen feltételek mellett.

9

Alkalmazzunk standardizált adatcsere-protokollokat

A standardizált adatcsere-protokollok használata lehetővé teszi a kutatóintézetek, infrastruktúrák számára, hogy szabadon hozzáférhetővé és olvashatóvá tehessek a(z) (meta)adataikat akár keresőmotorok számára is. Ezáltal az adatok sokkal könnyebben válnak kereshetővé és hozzáférhetővé.



Használjanak sztenderdizált adatcsere-protokollokat mint például a SWORD, az OAI-PMH, a ResourceSync vagy a SPARQL. Konvertálják a metaadat-sémákat XML-lé vagy RDF-fé. Tartsanak fenn egy jegyzéket (registry), amelyben számon tartják és közlik a kutatási adatok lekérdezésére szolgáló végpontokat és elérési útvonalakat.

A gyorsabb felfedezhetőséget és az új tudományos felismeréseket elősegítendő, a kutatási adatoknak mind az emberek, mind pedig a számítógépes rendszerek által könnyen kombinálhatónak kell lenniük.

Interoperabilitás

10

Létesítsünk jól dokumentált, gépileg működtethető API-kat.

A jól dokumentált és gépileg automatizált API-k – vagyis alprogramok, protokollok és szoftvereket fejlesztő eszközök együttese – lehetővé teszik a különböző repozitóriumból adattárakból származó (meta)adatok, források automatikus indexelését, lehívását és kombinálását.



Ügyeljenek rá, hogy az API-jaik alaposan dokumentáltak legyenek és tegyék lehetővé, hogy a (meta)adatmodellek sémáját is közvetítsék. Javasoljuk, hogy adjanak meg példákat, hogyan bányászhatók a leghatékonyabban adatok a különböző végpontokon keresztül és hogy az így kinyert adatok hogyan kombinálhatók új kutatási témákhoz használható újabb adatsorokká.

11

Használjunk nyílt és jól dokumentált szakszótárakat

A metaadat-elemek leírásának nyílt, jól dokumentált és széles körben ismert közösségi konvenciókon alapuló szakszótárakat (vocabularies) kell követniük. Az ilyen jegyzékek pontosan meghatározzák az adatokban kifejezésre jutó minőségeket és fogalmakat.



Használjon az Ön tudományterületén releváns szakszótárakat, jegyzékeket és a kutatási projekt kezdetétől fogva ezek szerint strukturálja és gazdagítsa kutatási adatait és eredményeit.



Az egyes kutatási területek sajátosságait figyelembe véve mutasson példákat közösségi standardokon nyugvó, széles körben használatos, de ugyanakkor diszciplínaszpecifikus szakszótárakra.

12

Dokumentáljuk a metaadat-modelleket

A metaadat-modellek világos és részletes dokumentálása segíti a fejlesztőket a különböző modellek és a metaadat komponensek közötti megfelelések (mappings) felállításában és átjárhatóvá tételükben.



Publikálják az Önök intézményében használatos metaadat-modelleket. Dokumentálják a technikai specifikációkat, csoportokat (közös tulajdonságokkal rendelkező dolgok csoportjai) és jellemzőket (a metaadatok egyes alrészeit jellemző tulajdonságokat és a metaadatok egyes részei közötti viszonyokat). Az egyes metaadat-komponensek közötti megfeleltetésekhez tüntessék fel a kötelező és az ajánlott jellemzőket.

13

Írjunk elő és használjunk átjárható adatszabványokat

Egy, a kutatói közösségek által erősen támogatott adatszabvány használata nagyban növeli az egyes gyűjtemények, adatsorok megoszthatóságának, újrafelhasználhatóságának és kombinálhatóságának esélyeit.



Nézzon utána és egyeztessen a repozitóriummal, ahol kutatási adatait tárolni kívánja az adattér által alkalmazott standardokról. Szerkessze és strukturálja kutatási adatait ezek szerint a standardok szerint már a projekt elejétől fogva.



Pontosan határozzák meg az Önök intézménye által alkalmazott standardokat és ösztönözzék a kutatói közösségeket, hogy vegyenek részt ezek megvitatásában, fenntartásában és alakításában, különös tekintettel az egyes standardok közötti átjárhatóságra. Jó példák a nyelvtudományokban használatos CMDI és a régészetben elterjedt SIKB0102.

14

Alkalmazzunk az adatok minőségét javító eljárásokat

Az adatok minőségének javítása érdekében hozunk létre (lehetőleg automatizált) (meta)adattisztító és gazdagító eljárásokat. Ezek az adatsorok közti átjárhatóságot is segítik.



Az adatgyűjtéskor bizonyos eljárásokkal minimalizálhatja a potenciális hibákat például ha kézi begépelés helyett naptárból választja ki a dátumokat.



Fektessenek be olyan eszközökbe, amelyek segítik a (meta)adatok tisztítását és konverzióját standardizált, interoperábilis formátumokba. Közös fejlesszenek ki munkafolyamatokat és szoftveres megoldásokat ezeknek a folyamatoknak az automatizálására például gépi tanuláson alapuló eszközök használatával.

15

Írjunk elő és használjunk „jövőálló” fájlformátumokat

Valamennyi repozitóriumban tárolt fájlnak nyílt, nemzetközileg standardizált formátumúnak kell lennie, így biztosítva hosszú távú elérhetőségüket, átjárhatóságukat, hozzáférhetőségüket és fenntarthatóságukat.



A kutatási projekt kezdeteitől használjon nyílt, és az Önnek releváns repozitórium által támogatott fájlformátumokat. Fontos, hogy ezek a formátumok ne függjenek se egy speciális szoftvertől, se pedig fejlesztőtől vagy más szolgáltatóktól.



Ösztönözzék a hosszú távú megőrzést elősegítő fájlformátumok például a PDF-A, CSV és MID/MIF használatát. Közzöljenek egy részletes, de könnyen átlátható áttekintést az Önök intézménye által támogatott fájlformátumokról.

A kutatási adatoknak jövőbeli projektekhez is elérhetőnek és felhasználhatónak kell lenniük. Ahhoz, hogy megbízhatóan építhessünk a korábbi kutatások eredményeire és hogy ezek az eredmények reprodukálhatók legyenek, már az adatok előállításának, gondozásának korai szakaszaiban szem előtt kell tartani a későbbi újrafelhasználhatóságukat.

Újrafelhasználhatóság

16

Szisztematikusan dokumentáljuk az adatokat

Ahhoz, hogy világossá tegyünk, milyen elvárásokkal közelíthetünk egy adott adatsorhoz vagy repoiztórumhoz, az adatok rendszeres és alapos dokumentálása szükséges. Ha pontosan és átláthatóan írjuk le, mi van az adatokban és mik az esetleges korlátai az anyagnak, a kutatói közösség nagyobb bizalommal fogja majd (újra)használni a munkánk eredményét.



Tegye elérhetővé a kódkönyvét a részletes módszertani leírással és a rövidítések jegyzékével együtt. Írja bele az adatgyűjtés esetleges hiányosságait is az adatbázis keletkezéstörténetével együtt.

17

Használjunk egyezményes megnevezéseket

Ha pontos és konzisztens megnevezési konvenciókat követünk (ezek az adatfájlok elnevezésére szolgáló, általánosan elfogadott sémák), sokkal könnyebbé tesszük a jövő generációinak, hogy hozzáférjenek az általunk generált vagy szerkesztett adatokhoz, vagy hogy hatékonyan tudják feldolgozni és megérteni azokat.



Nézzon utána, hogy melyek az Ön tudományterületéhez leginkább illő, azon belül a leggyakrabban használt egyezményes megnevezések.



Tegyük világossá a névadási konvenciókkal kapcsolatos legjobb gyakorlatokat. Készítsenek és/vagy kövessenek specifikus névadási konvenciókat.

18

Használjunk közismert fájlformátumokat

Az egyes kutatói csoportokon, tudományterületeken belül széles körben használt, standardizált fájlformátumok használata növeli az adtok újrafelhasználhatóságát.



A kutatási adatok hatékonyabb megosztása érdekében az archiváláshoz szükséges fájlformátumok mellett használjon aktuálisan népszerű formátumokat is, például Excel (xlsx), CSV vagy ESRI Shapefiles formátumokat a MID/MIF file-ok mellett.



Az archiváláshoz szükséges formátumok mellett publikálják az adatokat népszerű, az újrafelhasználást megkönnyítő formátumokban is, ha ez a kettő nem egyezik.

19

Őrizzük meg az adatok integritását

A fárasztó munkával összegyűjtött, megszerkesztett adatokat teljességükben kell megőrizni és továbbadni. Az adatok hitelességét biztosítandó, iktassunk közbe az adatok integritását ellenőrző lépéseket.



Alkalmazzunk hatékony verziókezelési módszereket. Az adatok hitelességének egyik legfontosabb feltétele, hogy valamennyi átdolgozás és új verzió alaposan dokumentált és jól követhető egyen.



A fájlokban bekövetkezett módosítások követéséhez különösen fontos, hogy rögzítsék a keletkezéstörténeti információkat (az adatok eredetét és az időközben bekövetkezett változtatásokat). Az is fontos továbbá, hogy az eredeti példány a másolatokkal összehasonlítható maradjon. Biztosítsák, hogy az adatsorok különböző verziói elérhetők és kereshetők maradjanak például az állandó azonosítók verziókövetésével.

20

Licencek az újrahasznosítás jegyében

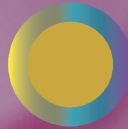
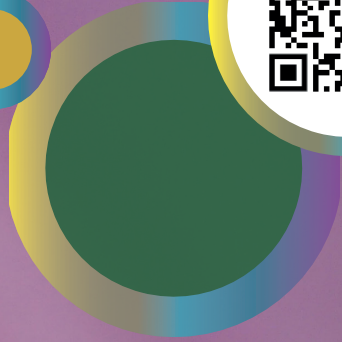
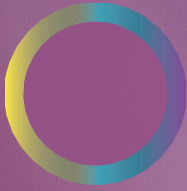
A lehető legszélesebb körű újrahasznosításhoz, tisztázni kell, kihez vagy kikhez tartoznak a(z) (meta)adatok szerzői jogai és milyen licenc érvényes rájuk.



Bizonyosodjon meg arról, hogy kihez vagy kikhez tartoznak a szerzői jogok mielőtt megosztja kutatásai adatait.



Kommunikálják a (meta)adatokhoz tartozó licenceket és az újrafelhasználási feltételeket világosan és gépileg is olvasható formátumban. A jobb átjárhatóság kedvéért érdemes széles körben ismert és használatos keretrendszert használni a licenszeléshez, mint amilyen például a Creative Commons.



A PARTHENOS az Európai Bizottság Horizon2020 programja által támogatott projekt. A jelen publikációban kifejtett nézetek és vélemények a szerző(k) álláspontját tükrözik, és nem szükségszerűen egyeznek az Európai Bizottság álláspontjával.

Ez a mű a Creative Commons CC-BY 4.0 licenc feltételeinek megfelelően használható fel. Design: Verbeeldingskr8. DOI: 10.5281/zenodo.3363355