



A DPR KERETÉBEN ALKALMAZOTT KÉRDŐÍVES ÉS ADATINTEGRÁCIÓS TEVÉKENYSÉGEK RENDSZERÉNEK FELÜLVIZSGÁLATA



SZÉCHENYI 2020



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE

Készült az EFOP-3.4.5-VEKOP-17-2017-00001 azonosítószámú, *Rendszerszintű fejlesztések és hozzáférés bővítését szolgáló ágazati programok a felsőoktatásban* kiemelt projekt DPR Pillér megvalósítása során.

Szerző: Veres Edit, pillérvezető

Felelős kiadó:

Dr. Gloviczki Zoltán

© Oktatási Hivatal, 2020

Az Oktatási Hivatal kiadványa az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg. A kiadvány adatainak bármilyen formában történő nyilvános közzétételéhez vagy felhasználásához a kiadó előzetes írásbeli hozzájárulása szükséges.

TARTALOMJEGYZÉK

Bevezetés	4
1. Diplomás pályakövetés rendszere.....	5
A felülvizsgálat során áttevített kutatási modulok	5
Adminisztratív Adatbázisok Egyesítése (AAE)	5
Frissdiplomás kutatás	9
A diplomás pályakövetési rendszer egyéb moduljai	10

Bevezetés

Az EFOP 3.4.5 projekt keretei között a diplomás pályakövetési rendszer alapjaiban újul meg. A diplomás pályakövetés rendszerének működését a Nemzeti Felsőoktatásról szóló 2011 évi CCIV törvény, ennek végrehajtásáról szóló 87/2015 (IV.9) Korm. rendelet, valamint a döntés-előkészítéshez szükséges adatok hozzáférhetőségének biztosításáról szóló 2007. évi CI. törvény, és a döntéselőkészítéshez szükséges adatok hozzáférhetőségének biztosításáról szóló 2007. évi CI. törvény végrehajtásáról szóló 335/2007. (XII.13.) Korm. rendelet teszik lehetővé.

A Frissdiplomás kutatás, amely a pályakövetési rendszer online kérdőíves modulja a legrégebbi, az első adatfelvétel 2009-es. 2013-tól eleinte kísérleti jelleggel, mára már rendszerszerűen megjelent mellette az adatintegráción alapuló pályakövetés is. 2018-ra adottá vált annak a lehetősége is, hogy a két kutatási rendszer legjobb tulajdonságai egy új módszertanban, a regiszter alapú surveyben egyesüljenek.

A diplomás pályakövetés rendszere jelenleg egymástól elszigetelt és össze nem kapcsolódó kutatási modulok együttese, az újabb és újabb modulok azonban összemosták a határterületeket a különböző részkutatások között. Fontos kihívása volt a fejlesztésnek a megfelelő lehatárolások megtétele, illetve a lehetséges kapcsolódási pontok és szinergiák megtalálása, rájuk vonatkozó módszertan kidolgozása. A rendszer megújítása során törekedtünk a valóban rendszerszerű kutatási struktúra kialakítására: a párhuzamosságokat a kutatási modulokban letisztítottuk, a regiszter alapú survey pilottal az adatintegráció és survey módszereket egy új kutatási formába integráltuk.

1. Diplomás pályakövetés rendszere

A pályakövetés rendszerszintű kialakításához a kutatási modulok felülvizsgálata és újratervezése vált szükségessé.

A jelenlegi rendszer egyik fő problémája az, hogy a megerősödő adatintegrációs kutatási modulok mellett meggyengült a Frissdiplomás kutatás 'kutatási pozíciója', és hátrányává vált a kialakításakor bevezetett rendszerszintű előnye: sok kérdésre ad választ, azonban mindegyik kérdéskört (az adatfelvételi lehetőség korlátozottsága okán) felszínesen érinti. Adatintegráció hiányában nem volt máshonnan adat arról, hogy például mennyi a keresete egy frissen végzett diplomásnak, mennyi idő alatt talál állást. Az adatintegráció megerősödésével ezeket a tényadatokat jobban lehet vizsgálni a regiszter alapú kutatásokban, az *adatszolgáltatás* gyorsabb, pontosabb és megbízhatóbb ezekből az adatbázisokból, mint az önbevalláson alapuló, alacsonyabb elemszámú kérdőívekből. A hangsúly áttérődése azonban nem vonzotta magával automatikusan a Frissdiplomás kérdőív azirányú megújítását, hogy a párhuzamosságok megszüntetésével, elhagyva azon adatköröket, amelyek jobban vizsgálhatók a regiszterekben, lehetővé váljon bizonyos specializált témakörök valódi *mélyelemzése* (a kérdőíves keretek között). A Frissdiplomás kutatás kezdeti értéke, hogy mindenről szolgáltatott, bár felületes információt – a hátrányává vált.

A rendszerszintű megújítás koncepcionálisan azt jelentette a projekt keretei között, hogy felülvizsgáltuk az egyes kutatási modulok egymáshoz való viszonyát.

A felülvizsgálat során áttevezett kutatási modulok

Adminisztratív Adatbázisok Egyesítése (AAE)

A felülvizsgálat során a kutatási modul minden részterülete megújításra került, sorban:

- koncepcionális modell
- kutatásba bevont változók
- adatellenőrzés módszere
- adatfeldolgozás módszertana
- adatvizualizáció megújítása

Koncepcionális modell

Az AAE korábbi koncepcionális modelljében az adott évben legalább abszolutóriumot szerzett végzett hallgatók vizsgálata történt meg. Megtartva a korábbi módszertan azon jellegét, hogy a végzés éve

maradt az alapsokaság leválogatásának alapja, az abszolutórium mellett bővült a FIR rendszerből származtatható képzésről lemorzsolódók körével.

A felülvizsgálat során nyilvánvalóvá vált az is, hogy nem az alapsokaság száma határozza meg az erőforrást, melyet az adatintegráció előállítása igényel, hanem a lekért változók jellege, ezért egyrészt a korábbi 1 és 3 éves merítést 1-7 évesre bővítettük, másrészt jelentősen felülvizsgáltuk workshop keretei között a korábban lekért változókat és bővítettük azokat újabb adatkörökkel is.

A felülvizsgálat eredményeképpen egy sokkal áramvonalasabb, fókuszáltabb, könnyebben kezelhető adatállomány összekötése vált lehetségessé – nagyon alapsokaság, több információ és olcsóbb bekerülési költség mellett (30%-al csökkent az adatintegráció ára).

Kutatásba bevont változók

A felülvizsgálat során nyilvánvalóvá vált, hogy a megújítás előtti gyakorlatban rögzült az a megoldás, hogy egy-egy változó több adatszolgáltató partnertől került bekérésre, és esetlegesnek tűnt, hogy mikor melyik forrás alapján történik az adatszolgáltatás (az adatszolgáltató partnerek adatbázisaiban lehetnek jelentős eltérések, ezeknek módszertani okai vannak). A felesleges párhuzamosságokat megszüntettük, és egyértelműen meghatározásra került a módszertana annak, hogy melyik adatbázisból származó információt használjuk általánosan az adattáblákban.

Változók megújításának irányai a felülvizsgálat alapján:

NEAK:

- 2018-tól nem kérjük a családi állapotot és a személy állampolgárságát, mert ez utóbbi a FIR-ben is elérhető, és cél a párhuzamosságok elkerülése
- nem kérjük továbbá a munkaviszonyhoz kötődő adatokat tartalmazó táblát, ezeket az adatokat ugyanis megkapjuk a NAV-tól.
- kérjük viszont a külföldi tartózkodásra, kiküldetésre vonatkozó adatokat külön táblában

NGM

- nem kérjük a nem és a születési dátum adatot, mivel ezek pedig elérhetőek a FIR-ben
- járás szinten is megvalósítottuk az adatok elérhetőségét

MÁK

- a korábbi adatkéréseket kiegészítenénk a 2012-nél korábbi jövedelmi adatokkal. 2012 előtt éves összejövedelemre vannak elérhető adatok, ez elosztható azon napok számával, amikor a személynek volt munkaviszonya.
- érdemes megkülönböztetni 2018-tól a nyugdíjjárulék és a nyugdíjbiztosítási járulék köteles jövedelmet, mert az elsőnek van felső korlátja, és felette már nem kell járulékot fizetni, míg a

másodiknál ez nem merül fel, így azt érdemes használni. A különbség az a kettő esetében, hogy a nyugdíjjárulékot a munkavállaló a nyugdíjbiztosítási járulékot a munkáltató fizeti.

DHK

- a Diákhitel Központtól csak a Diákhitel 1 és Diákhitel 2 alapadatokat kérjük, a többi táblára, például részletes a törlesztési adatokra nincs szükség

FIR

Az egyik legfontosabb újítás, hogy a korábbi 14 tábla helyett csak 3 tábla kerül lekérésére. Ezzel a fentebb említett egyszerűbb kezelést és lekérdezést segítjük elő. Ez nem jelent adatkörök tekintetében csökkenést, sőt számos olyan, korábban nem szereplő adatkör is bekerült a lekérdezésbe, amely korábban nem szerepelt (pl. képzés nyelve). A következő három tábláról van szó:

- Személyi adatok tábla – ez tartalmazza a vizsgált személyek személyes adatait
- Képzési adatok tábla – ez tartalmazza a vizsgált alapsokaság vizsgált időszakba, és azon kívül eső képzési adatait
- Időszak tábla – ez tartalmazza a képzési adatokhoz kapcsolódó képzési időszakokat (félévek), így nyomon lehet félév szinten követni többek között a félévek aktív vagy passzív státuszát, finanszírozását

1. blokk személy adatok tábla

Főbb változások a korábbi adatkéréshez képest.

Az oszlopok nevét átneveztük, nem szükséges az oszlopok nevének DPR komincióval kezdődni, a hallgatóra vonatkozó adatok HLG betűsorról kezdődnek.

Ha a hallgatónak több intézményben is van élő (valid) rekordja, akkor a személyes adatblokk alapja az a rekord lesz, ahol az utolsó frissítés időpontja (HLG_LASTUPD) a legkésőbbi.

A hallgató célcsoportba tartozását (külföldi) az OSAP számítás alapján határozzuk meg, amely elérhető a FIR-ben, így további számítást nem igényel.

Bővítettük a személyes adatokat, bekerült a hallgató születési településének járása, megyéje és régiója. Ezek az adatok elérhetőek a FIR-ben, így ezek előállítására nem igényel további erőforrást, és jól használhatóak térbeli elemzésekre.

Bekerült az adatok közé, hogy van-e a hallgatónak Magyar Igazolványa. Ennek célja, hogy a külföldi állampolgárok közül láthatóvá váljon, kik azon a határon túli magyarok, akik bár nem magyar állampolgárok, és nem kettős állampolgárok, de az elemzés során magyarnak kell tekinteni őket.

Szintén új adatkörként szerepel a hallgató előnyben részesítése, amelynek a későbbiekben a szociális dimenzió mentén lehet hozzáadott értéke. Itt 5 új oszlop került be az adatok közé, minden esetben (1/0) értékekkel:

- o hallgató előnyben részesítendő-e (árva, félárva, családfenntartó, nagycsaládos)
- o hallgató fogyatékkal élő-e

- o hallgató gyermekgondozó-e
- o hallgató hátrányos helyzetű-e
- o hallgató halmozottan hátrányos helyzetű-e

2. blokk képzések tábla

A korábbi adatintegráció során több tábla szerepelt a hozott dokumentumokra és a nyelvvizsgákra vonatkozóan. Ezt a két táblát megszüntettük, ezzel jelentősen csökken az oszlopok száma, a korábban nem használt oszlopokat nem is kérjük le, továbbá mivel kevesebb tábla lesz az adatok között így az adatok könnyebben, kevesebb erőforrással kezelhetőek.

Alapvető változás, hogy nem tanévek szerint kérjük le az egyes képzések adatait, hanem egy táblában az összes releváns képzési adatot, így egy táblába lesznek rendezve az összes képzési adat.

Az adatkérésében nem a megadott tanéveknek megfelelő időszakokban lévő képzési elemek jelentik az alapot, hanem a gráfort kijelölő képzési elem az alapegység, illetve a meghirdethető képzés.

Nem az teljes időtartamot felölelő összes félév adatai kerülnek egy-egy külön táblába, hanem az összes képzési információ, kezdő és végdátummal.

Az adatok között szerepel a nyilvántartott szak és a meghirdetett képzés illetve ezek jellemzői: munkarend, képzési szint, illetve a képzés nyelve.

Az elemzéshez jelentősen hozzá tud járulni a képzés telephelye, ez bekerült a lekérő adatok közé, itt szerepel a gps koordináták, járás, kistérség, megye, régió.

Korábban nem szerepelt, de fontosnak látjuk a képzést lebonyolító intézmény fenntartójának típusát is, így ez is szerepel az adatok között.

Igyekeztünk az intézményi jogutódlás kérdését is kezelni, ennek megfelelően a képzés adatgazda intézménye és a jogutódlás esetén az utolsó ismert intézmény, kar szerepel.

A képzési terület mellett szerepel az adatok között a szakmacsoport és a tudományterület is, továbbá egy oszlop, amely az tartalmazza, hogy az adott képzés tanárképzés-e. Mindez azzal a céllal, hogy minél többféle bontást, elemzést el lehessen végezni.

Dátumok tekintetében szerepel a képzés megkezdésének, lezárásának dátuma, továbbá az abszolutórium, kimeneti vizsga, a nyelvvizsga követelmények teljesítésének és az oklevél szerzés dátuma.

A későbbi elemzéseknél fontos információ lehet a teljesített és elismert kreditek száma, illetve ezek nagysága.

A kreditekhez hasonlóan a félévek számát kérjük a FIR-ből: Képzésben teljesített félévek száma, aktív és passzív félévek száma, támogatott és nem támogatott félévek száma.

A képzés állami támogatottságára vonatkozóan a következő logikát alkalmazzuk: Ha a támogatott aktív félévek száma nagyobb vagy egyenlő, mint a nem támogatott félévek száma akkor támogatott, egyébként nem támogatott (1= támogatott, 0= nem támogatott).

Az első felsőfokú végzettség megszerzésére irányuló képzésre vonatkozóan az OSAP logikát átvettük. Az adatok közé bekerült a duális képzésekre vonatkozó információ is, amely fontos elemzési szempont lehet a szakpolitika számára.

Szintén fontos változás, hogy a képzési adatok között aggregált módon (a képzés során legalább egy félévig kapott-e...?) a hallgatói juttatásokra, ösztöndíjakra, illetve kormányzati ösztöndíjakra vonatkozó információ. Ezen információk segítségével nyomon lehet követni az egyes támogatási formák, kormányösztöndíjak hatását, adott esetben munkaerőpiaci jelentőségét.

3. blokk Időszak tábla

A képzések táblában tárolt képzési programok időszakaira, féléveire vonatkozó adatokat tartalmazza a képzési adatok tábla. Itt is szerepelnek a képzések főbb adatai, illetve féléves bontásban megtalálhatóak az adott félév jellemzői (státusza, finanszírozási módja). Ebből nyomon lehet követni a finanszírozott félévek számát, illetve az aktív, passzív félévek, illetve a képzési időszak finanszírozási módja egyéni, és képzés szinten.

Adatellenőrzés módszere

A korábbi, kézi tisztítási gyakorlatot elhagyva részletes adatprofilozást végzünk az adatminőség ellenőrzésére, visszajelezve a problémák az adatszolgáltató partnerek felé is, akik korábban ilyen jelzéseket nem kaptak

Adatfeldolgozás módszertana

Meghaladva a hagyományos szociológiai adatfeldolgozó programokat (pl. SPSS) az adatintegrációban tárolt adatmennyiség szükségessé tette új adatfeldolgozási módszertan kialakítását. A 2018-as adatintegráció több, mint kettőmilliárd egyedi adatot tartalmaz, ez a mennyiség a klasszikus szociológiai adatelemző módszerekkel nem feldolgozható – ezért szükségessé vált egységes módszertani sztenderd kialakítása az adatfeldolgozásra. Az SQL, az R és a PowerBI nyújtotta megoldásokat választottunk a feldolgozás új módszereinek a felülvizsgálat során.

Adatvizualizáció megújítása

Az adatvizualizációt teljesen új alapokra helyeztük: a vizualizálás PowerBI programmal történik.

Frissdiplomás kutatás

A Frissdiplomás kutatás végzett hallgatói moduljának felülvizsgálata az adatintegrációs kutatási modul megerősödésével és rendszeressé válásával elengedhetelenné vált. A felülvizsgálat eredménye az lett, hogy megállapításra került: a projekt keretén belül a kutatási modul teljes átdolgozása és megújítása szükséges.

A workshopsorozat nyilvánvalóvá tette azt is, hogy az aktív hallgatói modul sem hagyható változatlanul, hiszen nincsen hozzáadott értéke annak, hogy minden évben ugyanaz a kérdéssor kerül lekérdezésre az aktív hallgatói időszak alatt. A felülvizsgálat eredményeként az is megállapításra került, hogy a bevezetéshez képest jelentősen változott a felsőoktatási tér, és a kérdőívben nem szerepelnek jelenleg olyan kérdések, amelyek reflektálnának a meghatározó változásokra: pl. több típusú diákhitel konstrukció bevezetése, duális képzés, stb.

A diplomás pályakövetési rendszer egyéb moduljai

A diplomás pályakövetési rendszer a projekt során két új kutatási modullal bővült: egyrészt a felsőoktatási pályakövetés modullal (amely a projektben egy meghatározott képzési terület, az informatikai képzési területre fókuszált módszertan kialakítást és kutatás elvégzését jelenti, másrészt a FIR alapú kutatással, amelyet 'regiszter alapú survey'-nek nevezetünk el.

A két új kutatási modullal egyedülállóan komplex pályakövetési rendszer alakult ki. A strukturális jellegű megújítás további iránya lehet a képzési szint helyet egyéni szintű adatbázisok kialakítása, valamint a meglévő szakrendszerei adatbázisok mellett egyéb szakrendszerekben tárolt adatok hozzákötése az adatokhoz.