

Intézményi interface technikai dokumentáció

verzió:1.2 (2011-09-30)

Definíciók

- AVIR – Adattár alapú Vezető Információs Rendszer
- Központi Adattárház – Központi és intézményi rendszerek történeti adatainak tárhelye
- Forrásrendszerek – Azon rendszerek, amelyekből a Központi Adattárház az adatait veszi
- Intézményi Adatok – a Felsőoktatási Intézmények kezelésében álló adatok halmaza
- Központi Adatok – központi adatkezelés alatt álló adatok halmaza

Bevezetés

Az AVIR projekt keretében egy Központi Adattárház kerül kialakításra, amely egyrészt a központi adatkezelés alatt álló adatokat, másrészt a különböző Felsőoktatási Intézmények rendszereiben található adatokat fogja tartalmazni.

Az adattárak az általános, napi működést támogató informatikai rendszerekhez képest nem arra vannak optimalizálva, hogy egyes tranzakciókat hajtsanak végre (mint például adatbevitel és adatmódosítások), hanem arra hogy sok adatot összevontan tudjanak kezelni, s azokból különböző döntés támogató és elemzői kimutatásokat lehessen készíteni.

Az adattárházak alapvetően a forrásrendszerekben létrehozott és megváltozott adatokat megadott gyakorisággal (például naponta vagy hetente) egyszer veszik át. Az átvett adatok különböző szűréseken, javításokon mehetnek keresztül, míg végül a központi adattárba kerülnek. A központi adattár a benne tárolt adatoknak a történetiségét is tárolja, így a forrásrendszerekből átvett adatoknak csak a változásokat kell tartalmaznia (ha egy adat nem változott, akkor az már biztosan ott van az adattárban).

Jelen dokumentum specifikálja, hogy milyen formában kell megvalósulnia az adatátvitelnek az intézményi forrásrendszerekből, és hogy az intézmények milyen formában kapják meg a rájuk vonatkozó adatokat a Központi Adattárházból.

Az AVIR projekt keretében az adatforrásokat két kategóriába lehet sorolni, az adatkezelő alapján:

- Intézményi Adatok,
- Központi Adatok.

Jelen dokumentum csak az Intézményi Adatok feldolgozásával foglalkozik. A következőkben a Betöltési Folyamat lesz bemutatva, azaz az, hogy az AVIR projektben a felsőoktatási intézmények által kezelt Intézményi Adatok milyen formában jutnak a Központi Adattárba. Ezt követi a Publikációs Folyamat ismertetése, azaz a különböző Kliens Adatfelhasználók – kiemelten a Felsőoktatási Intézmények – milyen formában kapnak adatokat a Központi Adattárból (Központi- és Intézményi Adatokat).

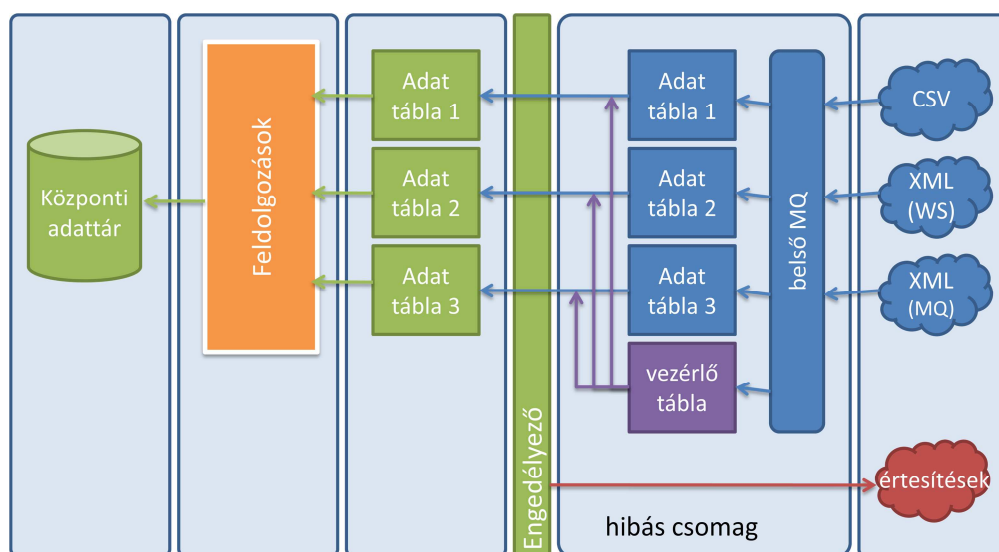
Betöltési folyamat

Architektúra

Az adatokat a Központi Adattár három lehetséges úton képes fogadni:

- Web szolgáltatás (WS) alapú felületen (XML),
- IBM WebSphere Message Queue-n (MQ) keresztül (XML),
- Webes felületen feltöltött CSV fájlok segítségével.

A beérkező adatok az érkeztetés után egy közös csatornában folytatják az útjukat. A feldolgozás csak a feltöltés befejezése után indul el, részben feltöltött csomagokat nem dolgoz fel a központi rendszer.



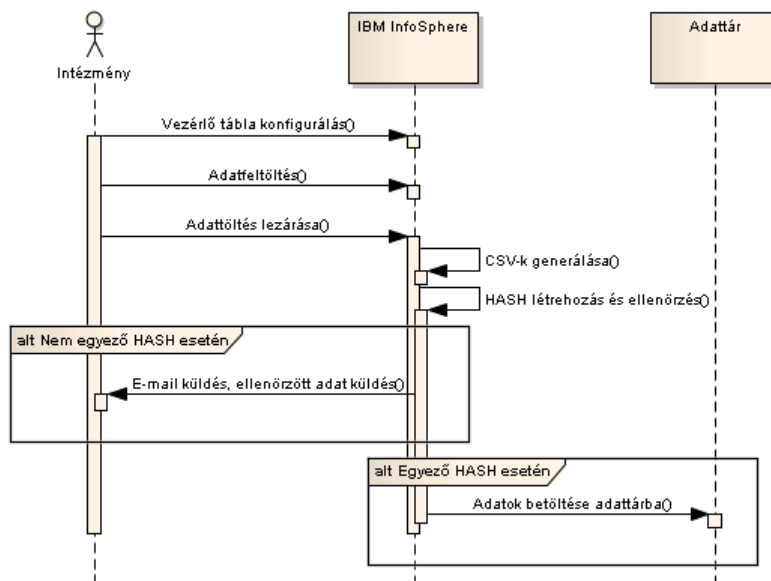
1. ábra: Intézményi adatok fogadása

Az intézmény táblánként egyszerre csak egy időszak adatait töltheti és azt teljes egészében kell töltenie. A teljesség ellenőrzéséhez két információt tartalmaz a vezérlő tábla:

- a beküldött sorok számát
- az adattartalom MD5 algoritmussal hash-elt kódját

Hogy melyik forrásnál melyik időszakot tölti, a vezérlőtábla megfelelő mezőjében adja meg. Ez a tábla tartalmazza azt is, hogy betöltési-, illetve feldolgozási hiba vagy sikeres adatbetöltés esetén milyen e-mailen kér az adatküldő visszaigazolást. Egy adott adatcsomag akkor kerül feldolgozásra, amikor az intézmény megadta az adott adatcsomag ellenőrző kódját, valamint lezárta az adatfeltöltést.

Adatfeltöltés lezárását követően a Központi Rendszer is elvégzi a beküldött adatokon ugyanazon algoritmus alapján az ellenőrző összeg számítását, majd összeveti azt a beküldöttel. Amennyiben eltérés van a két ellenőrző összeg között, hibajelentést küld a megadott e-mail címre, majd a beküldött adatokból generált CSV-t az adott intézmény felé publikálja.



2. ábra Adatfeltöltés folyamata

Adatfeldolgozás alatt nem tölthető újabb adatcsomag ugyanazon intézményi táblába.

A küldendő állományok felépítése

A küldendő XML állományok felépítését a mellékelt XSD határozza meg (intezmenyi_interface_minta adat.zip, XSD gyökere: Message.xsd).

A küldendő csv állományok felépítése augusztus folyamán kerül publikálásra.

Az XML üzenetek azonosítása

Az xml üzeneteket tranzakciós azonosítóval (/Kontener/KontenerFej/TranId) szükséges ellátni, amely felépítésében megegyezik a FIR-ben használt azonosító felépítésével. A különbséget a leírásban félkövér kiemeléssel jeleztük.

Az azonosító felépítése **AxxxxyÉÉHHNN**egyediazonosító, ahol:

- 'A' – kötött szám karakter (**értéke 8**)
- 'xxxx' – az intézmény négyjegyű azonosító kódja (képzése: szer_id-ból, 0 karakterekkel prefixelve, jobbra rendezve)
- 'y' – intézményen belüli küldő rendszer azonosító kódja (egyetlen ilyen alkalmazás esetén a kód 1) – több rendszer esetén a központi oldal osztja; használata főként azokban az intézményekben fontos, ahol a bejelentés több rendszerből fog történni
- 'ÉÉHHNN' – az azonosító létrehozásának dátuma
- egyediazonosító – az XML küldője által generált fixen 7 jegyű számsorozat, amely képzésénél a küldő garantálja, hogy a teljes azonosító egyedi lesz
- pl: 8012311007200000001

Vezérlőtábla feltöltése

A Vezérlőtábla egy olyan tábla, amely az adatküldésben érintett adattáblánként egy-egy rekordot tartalmaz. Az adott rekord azt adja meg, hogy milyen időszakra vonatkozik az adott csomag, hány rekordot küldtek, stb.

A vezérlő tábla kitöltése nélkül az adattáblák feldolgozása nem történik meg.

A vezérlőtábla szerkezete:

Szervezet: az adatokat küldő szervezet (felsőoktatási intézmény) FIR azonosítója.

Adattábla: az adattábla neve.

Időszak: azon időszak megjelölése, amelyre az adatok vonatkoznak. Ennek pontos módja a kiadott tartalmi dokumentációban van megadva.

Hash: az adattábla teljes tartalmának MD5 algoritmus szerinti hash értéke. A hash képzéshez az adattábla id szerint rendezve UTF8 kódolással előállított csv állományát kell MD5 algoritmus szerint hashelni.

RekordSzám: az adott időszakra vonatkozó rekordszám.

Megjegyzes: az elküldött adatok jellemzése. Értékkészlete:

- R: nem releváns az intézményre nézve;
- N: nincs információ (kötelező lenne küldeni, de nincs adat);
- E: ellenőrzött adatok;
- T: teszt adatok (ezt csak a teszt rendszer fogadja el)

Email: a feldolgozás eredményéről erre az email címre küld a rendszer üzenetet. Megadása opcionális

A fentiekből látszik, hogy vezérlő rekord akkor is szükséges, ha az intézménynek nincs egy adott témában elküldhető adata. Ilyenkor a vezérlő rekordban a nulla rekordszám és a „R” vagy „N” megjegyzés jelzi a központ felé, hogy nem érkezik adat az intézménytől.

Adatok feltöltése

Az adattáblák feltöltése a vezérlő táblával összhangban kell, hogy történjen. Az XML üzenetek pontos szerkezetét és egy minta üzenetet a melléklet tartalmazza. Az adatfájlok feldolgozását a rendszer úgy végzi, hogy az vagy teljes egészében beépül a Központi Adattárba, vagy egyáltalán nem. Amennyiben hibát észlel a betöltés során, akkor a teljes adattábla feldolgozása visszavonásra kerül.

Az adatküldések azonosítása

Az egy adatküldésbe tartozó XML üzeneteket az adatküldést egybekapcsoló azonosítóval (/Kontener/KontenerFej/AdatkuldesId) szükséges ellátni. Az AdatkuldesId tag-be az adatküldés

minden csomagja esetében a vezérlő tábla /Kontener/KontenerFej/TranId helyén szereplő azonosítóját szükséges írni.

Feldolgozás visszajelzései

A feldolgozás állapotáról az alábbi visszajelzések történnek:

- MQ-n érkező adatcsomagra MQ üzenet
- WS-en lekérdezhető státusz információ
- email értesítés a vezérlő táblában megadott email címre

Amennyiben az intézmény MQ-n keresztül küldi az adatcsomagot, annak státuszát is megkapja egy queue-ban. A státusz XML tervezett tartalma egy-egy vezérlőtábla által összefogott adatszolgáltatásra vonatkozóan a feldolgozás eredménye lesz (ami lehet sikeres; sikertelen, hiba okok felsorolásával). A státusz XML formátuma augusztus közepéig kerül véglegesítésre.

Bármilyen adatküldés történik, a feldolgozás státusza webservice-en keresztül is lekérhető, amely szintén augusztus során kerül kialakításra.

A feldolgozás hibája esetén a vezérlő táblában megadott email címre a rendszer értesítést küld. Fontos, hogy ez a visszajelzés csak hiba esetén működik, tehát a sikeresen beépült csomagokról nem kap semmiféle információt a megjelölt postaláda. Az email visszajelzés az üzemeltetést végző személyek számára készül, automatikus feldolgozásra nem ajánljuk.

Visszajelzések a WS interfészen

A WS interfészen a megfelelő szolgáltatás¹ meghívásával a feldolgozás állapotáról a következő státuszüzenetek érkehetnek:

WS	1000	XML üzenet átvéve
	1101	Nincs joga a szolgáltatás igénybevételéhez
	1201	A SOAPHeaderben megadott szervezet azonosító eltér a KontenerFejben megadott értéktől
	1301	A megadott adatküldés azonosító ismeretlen
	1302	Az üzenet tranzakciós azonosítója nem egyedi
	9999	Ismeretlen hiba lépett fel a program működése során
Csv2Xml	2000	CSV üzenet átvéve
	2101	Hiba az üzenet feldolgozása közben!
Validator	3000	Az üzenet megfelel az XSD-nek
	3101	Az üzenet nem felel meg az XSD-nek
	1101	Nincs joga a szolgáltatás igénybevételéhez (ezen a ponton MQ csatornán keresztül érkező üzenetekre készülhet ilyen válasz)
DATASTAGE	5000	Hash ellenőrzés sikeres
	5101	Hash ellenőrzés sikertelen!
	6000	Adattárház betöltés sikeres, hibátlan módon
	6010	Adattárház betöltés sikeres, a betöltés során nem kritikus hiba merült fel
	6101	Adattárház betöltés sikertelen!
	6102	

¹ Kontener operation meghívása a KontenerRequest paraméterrel (kitöltendő: KontenerTorzs elem AdatbekuldesStatuszKeres része)