



# A hivatalos statisztika modernizációja az adatforradalom korában

Vukovich Gabriella  
150 éves a Központi Statisztikai Hivatal  
Földi sokaságok – égi tünemények konferencia  
2017. október 18.

---

# A statisztika előtt álló kihívások

- Gyorsan változó felhasználói igények, változatos adat-felhasználási formák
  - Változó társadalom, változó gazdaság
  - Adatforradalom
  - Adatszolgáltatói terhek, adatszolgáltatói hajlandóság
  - Korlátozott erőforrások
-

# Változatos felhasználói igények és adat-felhasználási formák

- Professzionális felhasználók köre és száma bővül
- Mikroadatok iránti igény
- A statisztikai információ, mint bármely más hír pillanatok alatt szétterjed a közösségi médián és az egyéb internetes hírfelületeken ⇒ „amatőr” felhasználók sokasága
- A hivatalos statisztikai adat-előállító és az alternatív adat-előállítók keverednek a felhasználás során
- A statisztikai információ mint hírkategória szóljon a jelenről, esetleg a jövőről
- Korlátozott az idő és a tér a szakértői értelmezésre (valós idejű információ 140 karakterben)
- Sokféle felhasználó = sokféle értelmezés

*Korábban:* kevesebb adat, szakszerű értelmezés (statisztikusok, kutatók, elemzők)

*Ma:* adatözön, döntően a felhasználóra bízott értelmezés

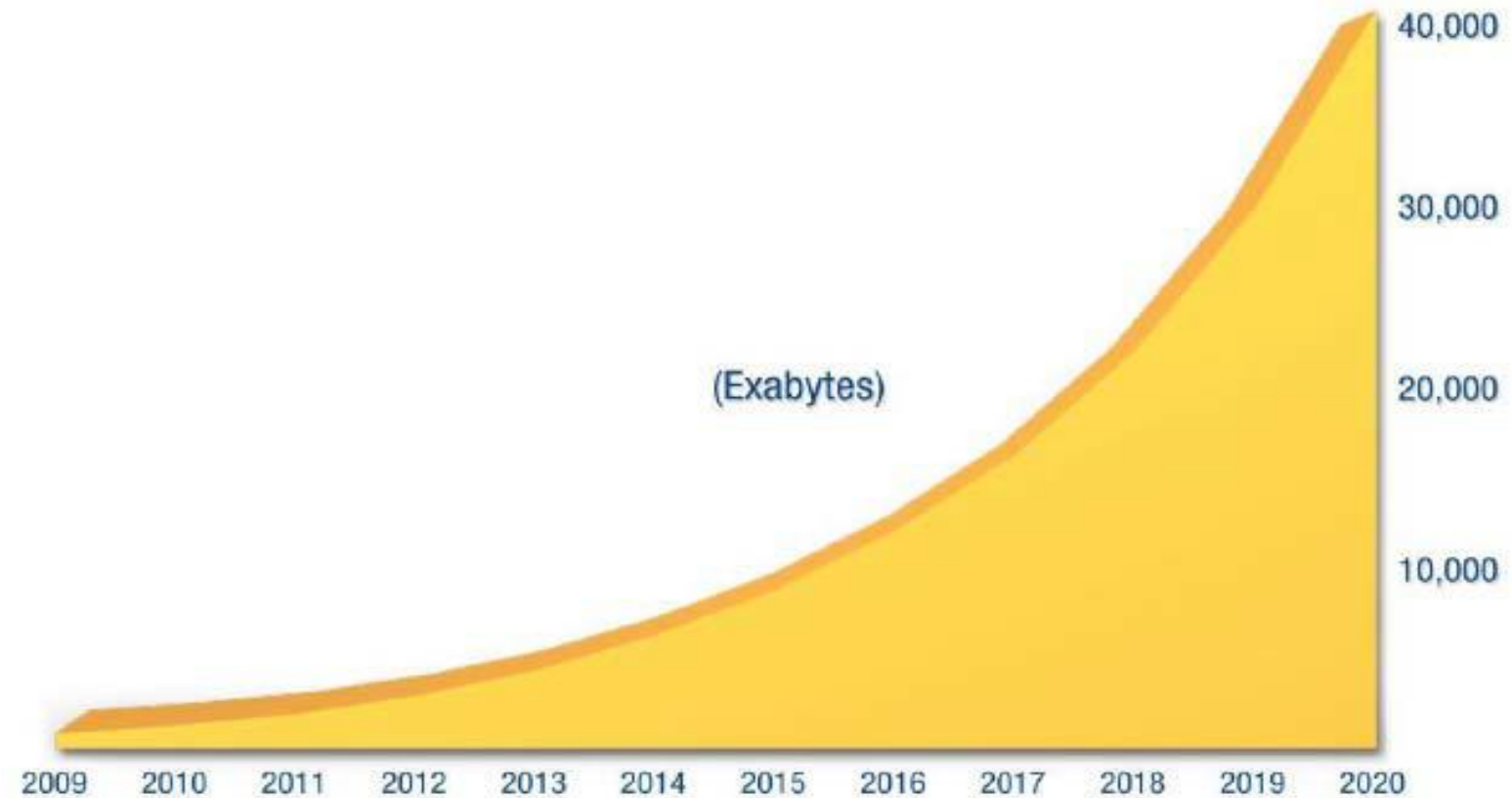
- A felhasználók többségétől távol áll a módszertani ismeret, a metaadatok használata stb.; közérthető statisztikai információkat kell közzétenni, lehetőleg látványos formában
- Az adat és az adat-előállító akkor is hiteltelennek tűnik, ha a felhasználó értelmezi rosszul az adatot

# Változó társadalom, változó gazdaság

- Társadalmi szerkezet
  - Háztartások összetétele, együttélési formák, korstruktúra,
  - Új és megszűnő foglalkozások, új foglalkoztatási formák
  - Közérdeklődés témái változnak (pl. fenntarthatósági kérdések előtérbe kerültek)
  - Globalizáció – gazdasági szereplők és kapcsolatok
  - Változó migrációs mintázatok
  - Határokat átlépő gazdasági tevékenységek körének és típusainak bővülése
  - Digitális technológiák ⇒ új vállalkozási modellek (pl. közösségi gazdaság - „sharing economy”)
  - Szolgáltató ágazatok térnyerése
-

- Nagyrészt technológia- és kínálatvezérelt spontán (?) folyamatok
  - A (domináns) kínálati oldal
    - az adatok mennyiségének robbanásszerű növekedése (kétévente megkétszereződik a keletkező adatok mennyisége) és keletkezésének gyorsulása
    - az adatforrások robbanásszerű bővülése (közösségi média, mobil alkalmazások, állami nyilvántartások és adatbázisok, üzleti adatbázisok, szenzorok, kamerák, térinformatikai eszközök, statisztikai összeírások stb.)
    - az adat-előállítók, adatfeldolgozók számának gyors növekedése – adat-előállítók versenye
  - ⇒ az adatok keletkezése, tárolása és elemzése jelenleg határtalannak tűnő expanzióban van
  - Keresleti oldal
    - az adatok iránti igény robbanásszerű növekedése a társadalom minden szegmensében
-

## The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020



Source: IDC's Digital Universe Study, sponsored by EMC, December 2012

1 EB = 1,13 x 10<sup>15</sup> kB

---

# Adatszolgáltatói terhek, adatszolgáltatói hajlandóság

- Az adatszolgáltatók (vállalkozások, intézmények) adminisztratív terhei rontják a versenyképességet
  - Az adatszolgáltatók jelentős része nem érzékeli a statisztika hasznosságát
  - Az adatszolgáltatói hajlandóság (vállalkozások, lakosság) romlik
  - Rengeteg állami adatbázis, nyilvántartás stb. létezik, ezek elvileg helyettesíthetik a statisztikai adatgyűjtéseket
  - A gyakorlatban is, de gyakran korlátokkal
  - Az adminisztratív és egyéb ú.n. másodlagos adatforrások statisztikai célú felhasználásának bővítése minden statisztikai hivatal célja
  - Előnyök/hátrányok
-

# A hivatalos statisztika hagyományos lehetőségeinek és korlátainak feszegetése

- Kis területi egységekre vonatkozó adatok iránti igény
  - Kevésbé vagy egyáltalán nem lefedett sokaságokra vonatkozó adatok (speciális helyzetű vagy igényű csoportok, kisebbségek)
  - Tömegjelenségek mérése vs rendkívüli részletezettség; mikroadat-állományok; adatvédelem
  - Másodlagos adatforrások szerepe megnő
-





# Másodlagos adatforrások (Big Data is) a hivatalos statisztikában

- Egyedüli adatforrásként
  - Adatgyűjtések kiegészítésére
  - Adatforrásként különböző másodlagos adatforrások összekapcsolásával
  - Modell alapú becslésekhez
  - Adatok validálására
-

# Korlátozott erőforrások

- Pénz
  - Emberi erőforrások
    - új adatbányászati, technikai ismeretek, új készségek
    - elemzői készségek
  - Ha az erőforrások nem bővülnek, a hatékonysági tartalékokat kell még jobban feltárni
-

# A hivatalos statisztika modernizációja

- Az elmúlt 150 évben a hivatalos statisztika folyamatosan megújult
  - DE: a XXI. század elején a hagyományos keretek között már nem tud megfelelni a sokasodó és egyre gyorsabban változó elvárásoknak
  - Az információs piacon részben ismeretlen (és ismeretlen minőségű adatokat előállító) versenytársakkal kell megküzdeni
  - A modernizáció célja, hogy küldetésünket a jelenben és a jövőben is teljesíteni tudjuk: ***hiteles és megbízható minőségű hivatalos statisztikai adatokkal és elemzésekkel támogassuk a tényeken alapuló döntéshozatalt, a tudományos kutatást és a társadalom szereplői közötti párbeszédet***
-

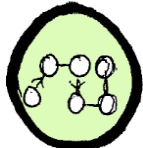
# A kihívásokra adott válasz: modernizáció



⇒ Emberek



⇒ Adatforrások, módszerek



⇒ Folyamatok



⇒ Technológia, informatikai eszközök, rendszerek



⇒ Keretrendszerek, standardok



⇒ Támogató és adminisztratív szolgáltatások

# Az adat-előállítás egyes lépéseinek jellemzői

Folyamatszakasz	Tegnap és ma	Holnap
Igényspecifikáció	Időkorlátok és korlátozott rugalmasság	A felhasználói igények feltárása és fokozott érvényesítése
Tervezés	Gyakran egyedi megoldások	Modulokból építkezés
Fejlesztés	Gyakran egyedi megoldások	A legtöbb szakstatisztika igényeit kielégítő egységes megoldások
Adatfelvétel megvalósítása	Döntően elsődleges statisztikai adatgyűjtés	Az adminisztratív és egyéb másodlagos források használatának előtérbe kerülése
Adatfeldolgozás	Egyedi és élők munkáigényes megoldások	Egységes adatfeldolgozási rendszerek, általánosan használható eszközök
Elemzés	Kapacitás-, eszköz- és időkorlátok	Az elemzések előtérbe kerülnek, fokozott szakértelem, hozzáadott érték és adatintegráció
Közreadás, tájékoztatás	Inkább statikus és kiadványcentrikus	Továbbfelhasználást lehetővé tevő adatközlések, interaktív és újszerű vizualizációs eszközök

## A modernizáció jellemzői

- Továbbra is fókuszban a minőség; a hivatalos statisztika legfontosabb eszköze az információs piaci versenyben a dokumentált minőség
  - Rugalmas alkalmazkodás a külvilág és a felhasználói igények változásához
  - Új termékek és szolgáltatások, magasabb hozzáadott érték
  - Szakstatisztikai, módszertani és technikai, technológiai kutatások, fejlesztések
  - Adatintegráció, modell-alapú becslések
  - Adatgyár helyett információ-menedzsment
  - Nemzetközi együttműködés az intézményi és folyamatmodernizáció terén is
-



# Intézményi értékek mentén megvalósuló modernizáció

- **A 150 éves intézmény alapvető értékei meglepően állandóak**
    - Szakmai függetlenség, pártatlan működés, adataink, elemzéseink objektivitása
    - Minőség iránti elkötelezettség
    - Felhasználó-központúság
    - A bizalmas egyedi adatok védelme
    - A munkatársak szakmai felkészültsége, elhivatottsága
    - Tiszteletre méltó hagyományunk a megújulás képessége
-



Elkezdődött a második 150 év

---