

# Az ELKH Adatrepozitórium Projekt (ARP) „motiváció”

Szilágyi Edit  
Wigner FK



# Tartalom

- Mi a kutatási adat? Kié a kutatási adat?
- Adatvesztések? Elkerülhetők? (adathordozó-, ill. programváltás, sérülés, személyi változás)
- Mi a repozitórium?
- Miért és milyen kutatási adatokat érdemes repozitóriumban elhelyezni?

# Mi a kutatási adat?

- „kutatási adat: digitális formátumú – a tudományos publikációtól eltérő – a kutatást végző vagy a kutatást finanszírozó szervezet által intézményi vagy tematikus adattárakon keresztül nyilvánosságra hozott közadat, amelyeket közfinanszírozású tudományos kutatási tevékenység keretében gyűjtenek vagy állítanak elő, és amelyeket a kutatási folyamat során bizonyítékként használnak, vagy amelyeket a kutatói közösség általában a kutatási megállapítások és eredmények megerősítéséhez szükségesként fogad el”
- Törvényjavaslat címe: **Egyes törvények közadatokkal összefüggő módosításáról**. Iromány száma: **T/1786**. Benyújtva: 2022-10-28 20:57

# Mi a kutatási adat?

- Kutatási adatokon lényegében mindent értünk a számítógépen futó rendszer- és kereskedelmi programokon kívül:
  - kísérletek tervezése, minták leírása, ...
  - mérési adatok, laborjegyzőkönyvek, ...
  - a mérések értelmezéséhez felhasznált adatok, táblázatok, ábrák, dokumentumok, ...
  - számítások (az ehhez szükséges saját fejlesztésű programok, makrók, stb.), ...
  - dokumentációk, kapcsolattartás, kommunikáció más kutatókkal, ...
  - pályázatok írásakor / projektek végrehajtása során keletkező anyagok ...
  - cikkek/előadások írásakor keletkező anyagok ...
  - ...

# Kié a kutatási adat?

- A kutatóké, akik előállították? (És ha igen, akkor melyik kutatóhoz tartozik leginkább?)
- A kutatási infrastruktúráé?
- Ahhoz a projekthez tartozik, amelyik a kutatást finanszírozta? A témavezetőé?
- Az intézeté?
- Ki gondoskodik az előállításáról, archiválásáról, felhasználásáról?

# Adatvesztés

- Ha az adat elvész (pl. adathordozó-sérülés miatt): megfelelő archiválás hiányában.
- Ha ugyan az adatok megvannak, de már valamiért használhatatlanná váltak. Pl.:
  - adathordozó-váltás,
  - programváltás, vagy
  - személyi változás miatt.

# Számítástechnika-történelem

- Abakusz (~5000 éve)
- Mechanikus számológépek: 1623. Wilhelm Schickard az első automatikus számológép (4 alpművelet)
- Elektromechanikus számítógépek: 1941. A **Z3** 22 bites szavakkal dolgozott, memóriájában 64 adatszót tudott tárolni, mint elődei, a Z1 és Z2. Felépítése a mai gépekhez hasonlít: processzort (aritmetikai-logikai egységet (ALU) és vezérlőegységet (Control Unit, CU)), memóriát, bemeneti egységet (szalag) és kimeneti egységet tartalmaz.
- Első generációs számítógépek: Mark I (1939 – 1944) - teljesen elektronikus számítógép (elektroncsöves)
- Második generációs számítógépek (1958-1964): már tranzistorokat tartalmaztak.
- 1958: Az első integrált áramkör (Jack Kilby, a Texas Instruments).
- Harmadik generációs számítógépek (1964-1971) már integrált áramköröket használnak.
- **1971 hozza a nagy fordulatot:** John Blankenbaker megépíti az első személyi számítógépet, a Kenbak I-t.
- Negyedik generációs számítógépek: mikroprocesszorok megjelenése (1974-1993)
- Ötödik generáció 1991-től napjainkig: párhuzamos és asszociatív működésű mikroprocesszorok
- Jövő: optikai számítógép? kvantumtechnológia?

# Adatvesztés, adathordozó-váltás

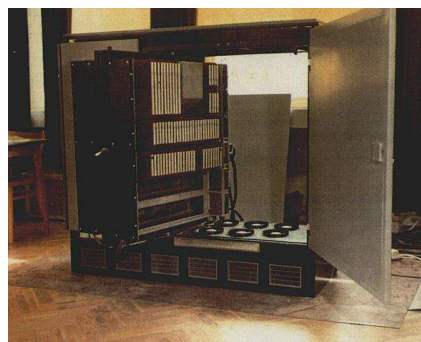
Sikerült transzformálnunk az összes adatunkat?



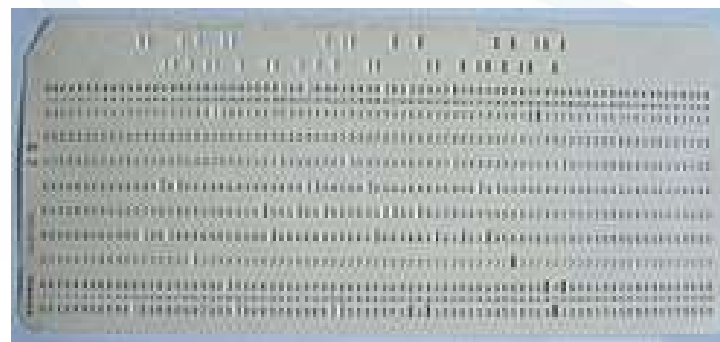
Lyukszalag

1901-1980 (1990)

R10, TPA gépek



TPA1001 (1968)



Lyukkártya (Hollerith-kártya)

18. századtól használják automatizálásra.

1960-as évektől kezdődően megjelennek az interaktív adatbeviteli lehetőségek.



# Adatvesztés, adathordozó-váltás

Sikerült transzformálnunk az összes adatunkat?

Magnókazetta.  
ZX Spectrum 16K/48K  
Commodore  
1982-1990



Speciális mágnesszalagokat  
háttértárolóként még akár ma is  
alkalmaznak.

Floppy-k  
(különböző méretben):  
1971-1987  
IBM Personal Computer XT  
(1981-87)



AT-k, Pentium-ok: CD, DVD, merevlemezek.  
Személyi számítógépek, laptopok, stb.: pendrive, (CD, DVD), merevlemezek

# Adatvesztés: programváltás

- Ha leszokunk egy korábbi szoftverről, egy idő után nem tudunk mit kezdeni a korábban keletkezett szövegekkel, ábrákkal, stb. (ChiWriter, Grapher ...)
- Visszafelé kompatibilisek a programok? És valóban azok, vagy csak ígérik? (MS Word, Origin, ....)

# Adatvesztés: sérülés, személyi változások

- Adathordozók sérülése: hogyan archiváljunk jól?
- Személyi változások: Átvehetők-e az adatok? Ki vegye át egy korábbi kolléga által kezelt adatokat?

# Mi a repozitórium?

- Röviden: egy mások által is kereshető adattár.
- Az itt elhelyezett adatcsomagok FAIR-elveket (FAIR = Findable, Accessible, Interoperable, Reusable – azaz az adatok megtalálhatóak, hozzáférhetőek, szabványosak és újrafelhasználhatók) kielégítik.

# Miért érdemes a kutatási adatokat repozitóriumban elhelyezni?

- Megszabadulhatunk bizonyos feladatoktól. A repozitóriumba letett adatokat már nem a kutatónak kell „gondoznia”.
- Publikálási „kényszer”. Több kiadó már elvárja, a jövőben várhatóan még több fogja elvárni.
  - Miért? Mert úgy tartják ellenőrizhetőnek a kutatásokat, ha a maguk a kutatási adatok is nyilvánosak.
- A kutatási adatok maguk is publikálhatók! (Scientific Data, ...)
- Kutatási pályázatoknál is elvárt lesz hamarosan a FAIR/CARE elveknek megfelelő adatkezelés, archiválás. (CARE = Collective benefit, Authority to control, Responsibility, Ethics; kollektív haszon, ellenőrzési jogkör, felelősség, etika).
- Adatmegosztás különböző intézetek, ill. projektrésztvevők között.

# Milyen kutatási adatokat érdemes elhelyezni a repozitóriumban?

- Feldolgozott adatokat:
  - Közleményekhez, projektekhez, mérésekhez tartozó összes adatot.
  - Programokat, makrókat.
- Megfelelően strukturáltan, csoportosítva, ellátva az összes tudnivalóval, hogy újrafelhasználható legyen. Nemcsak a repozitórium metaadatait kell megadni, hanem az újrafelhasználáshoz szükséges ún. belső (doménspecifikus) metaadatokat is.

Betartva a jogi előírásokat is: GDPR, copyright ...

# Összefoglalás

- Nagyon sok kérdést tettem fel; sokat meg se tudtam válaszolni.
- Sok kérdésre viszont jó válasz lehet az, ha megtanuljuk használni az adatrepozitóriumokat, és kihasználni az általuk nyújtott új lehetőségeket!

Köszönöm a figyelmet!