

Arassunk! Vagy ne arassunk?

Knapp Gábor

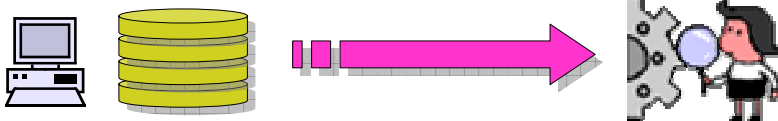
BME MOKK – knapp@mokk.bme.hu
NAVA projekt – knapp.gabor@nava.hu

Cél

- **Általános cél:**
 - Több forrásból származó, összetartozó adathalmaz (nem feltétlenül adatbázis!) „egykapus” elérhetőségének biztosítása
- **NDA cél (jelenleg):**
 - Kulturális gyűjtemények adatbázisainak együttes kereshetősége

Kellékek

- (Információ rögzítés)
- Adatbázis(ok), vagy valamilyen szinten strukturált adatok, adattárak (DP)
- Keresőalkalmazás, amely a felhasználók számára biztosítja az elérhetőséget (SP)
- Kommunikációs forma (protokoll) az adatok és a keresőalkalmazás között



NDA Café
2004. november 22



Lehetőségek

- Adattárolás
 - DP oldalon
 - SP oldalon
 - mindkét oldalon (DP+SP)
- Keresőalkalmazás
 - az SP oldalon („egykapus”)
- Kommunikáció
 - on-line, automatikus

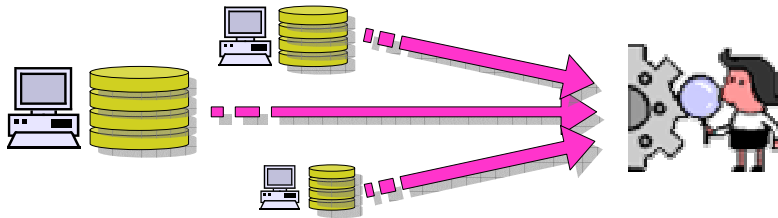
A különbség tehát az adattárolás helyében lehet
(de a választásnak a keresőalkalmazásra, kommunikációra is vannak hatásai)



NDA Café
2004. november 22



Adattárolás csak a DP oldalon



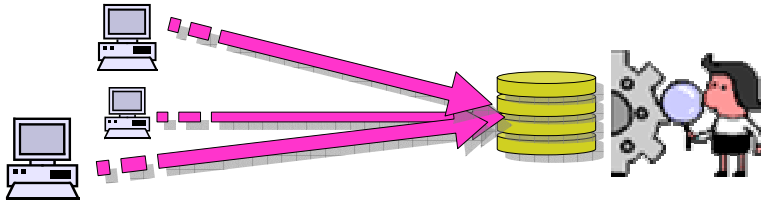
- A keresőalkalmazás oldalán nincs tárolt adat, minden kérdésnél végiglátogatja a kapcsolt adatbázisokat
- Kommunikáció: Z39.50, SRU/SRW, ODBC

Adattárolás a DP oldalon - Jellemzők



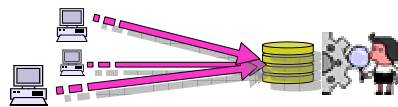
- Nincs szinkronizációs probléma, a mindenkori keresés pillanatában érvényes adatot kapjuk (valós idejű)
- Az SP csak a szabványos felületet látja, nem fér hozzá a DP belső adataihoz
- A DP monitorozhatja a használatot (kereséseket)
- DP oldalon is követelmény a nagy rendelkezésre állás (távoli elérhetőség)
- A DP szoftverkörnyezet összetett
- Egy helyen kell a szótárakat karbantartani
- Nem szabványos esetben kétoldalú megállapodások szükségesek

Adattárolás az SP oldalon



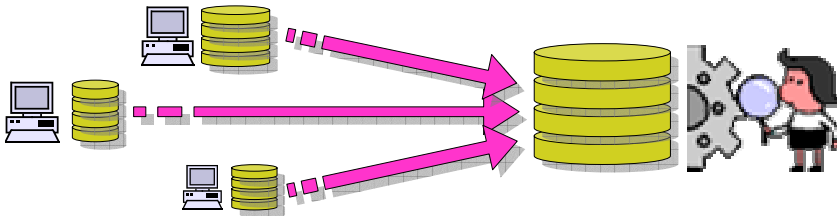
- A forrásoldalon (a fejekben) keletkező adatok közvetlenül az SP oldal adattárába kerülnek
- Kommunikáció: pl. HTTP

Adattárolás az SP oldalon - Jellemzők



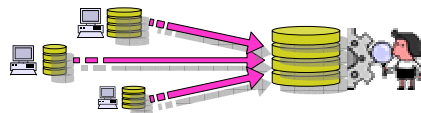
- Nincs szinkronizációs probléma, minden adat egy helyen van
- Biztonsági kérdések (authenticáció, az adatgazda közvetlenül érintkezik a „fő” adatbázissal)
- Csak az SP oldalon kell nagy rendelkezésre állás, minden ezen a szerveren múlik
- Egyszerű DP (vékony kliens)
- Az adatbevitel pillanatában lehet ellenőrizni, könnyen javítható
- Egy helyen kell a szótárakat karbantartani

Adattárolás mindkét oldalon



- A forrásoldal adatbázisainak tartalma alapján időről időre frissül a kereső oldal adatbázisa
- Kommunikáció: OAI-PMH (e-mail, pendrive stb.)

Adattárolás mindkét oldalon - Jellemzők



- Szinkronizációs probléma (kezelni kell az adatváltozásokat)
- Szabványos esetben nem kell előzetes megállapodás
- A lekérdezés nem valósidejű, így nem kell nagy rendelkezésre állás DP oldalon
- Egyszerű, olcsó szoftver a DP oldalon (OAI szerver)
- A hibák csak a begyűjtés után derülnek ki
- Az adatgazda nem naplózhatja a használatot, elveszti a felügyelet lehetőségét

Összefoglalás (DP szemszögből)

Jellemző	DP	SP	DP/SP
Példa protokoll	Z39.50	HTTP	OAI-PMH
DP rendszer bonyolultsága	***	*	**
DP rendelkezésre állás	***	*	**
Szinkronizáció, hibajavítás	**	*	***
DP kontroll lehetősége	***	*	*
Előzetes megállapodás	*	***	*

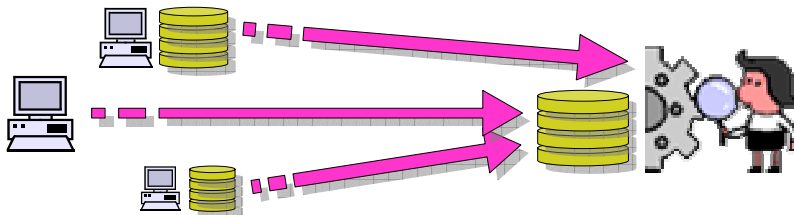
Akkor hát arassunk, vagy ne arassunk?



NDA Café
2004. november 22



Tanulság - Hibrid megoldás



- Valós idejű elérés (DP – Z39.50):
 - nagy gyűjtemény, komplex, megbízható informatikai rendszer
- Közvetlen bevitel (SP - HTTP):
 - kicsi gyűjtemény, informatikai rendszer nélkül
- Aratás (DP+SP, OAI-PMH):
 - közepes gyűjtemény, olcsó informatikai rendszer



NDA Café
2004. november 22

