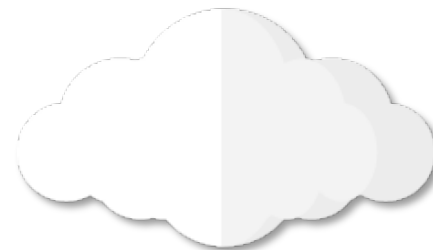




ELKH Cloud



Adatrepozitóriumok Bevezetés



Dr. Kovács László

SZTAKI DSD

laszlo.kovacs@sztaki.hu

DSD:



The screenshot shows the DSD website homepage. On the left is a dark blue vertical sidebar with the word "CONTENTS" written vertically. Below it are several small square icons. At the bottom of the sidebar is the text "MTA SZTAKI" and language selection options "en" and "hu". The main content area is a grid of six tiles: "Projects" (yellow cylinders), "Services" (metal grate), "Products" (blue spheres), "References" (white pyramid), "Publications" (ancient stone carvings), and "People" (white busts). At the bottom of the grid is a large dark blue tile with "DSD" and "DEPARTMENT OF DISTRIBUTED SYSTEMS" text, and a white tile with the number "25". At the very bottom of the page is the copyright notice "Copyright © 1994-2013 MTA SZTAKI - DSD."

- ▶ Reprodukciós krízis - világméretekben
- ▶ Mo: Évtizedes lemaradás a repozitóriumok területén
 - ▶ publikációs repozitóriumok (full text)
 - ▶ adatrepozitóriumok (kutatási adatok)
- ▶ Eötvös Loránd Kutatási Hálózat (ELKH) Repozitórium Munkabizottság
- ▶ Adatelvű tudomány (tézisek, vers. 1.3 (2020.05.29))
 - ▶ ELKH vízió 14 tézisben összefoglalva

Áttekintés



- ▶ Repozitórium fogalma
- ▶ Elvárások a modern repozitóriumokkal szemben (repozitórium funkciók)

Digitális repozitórium

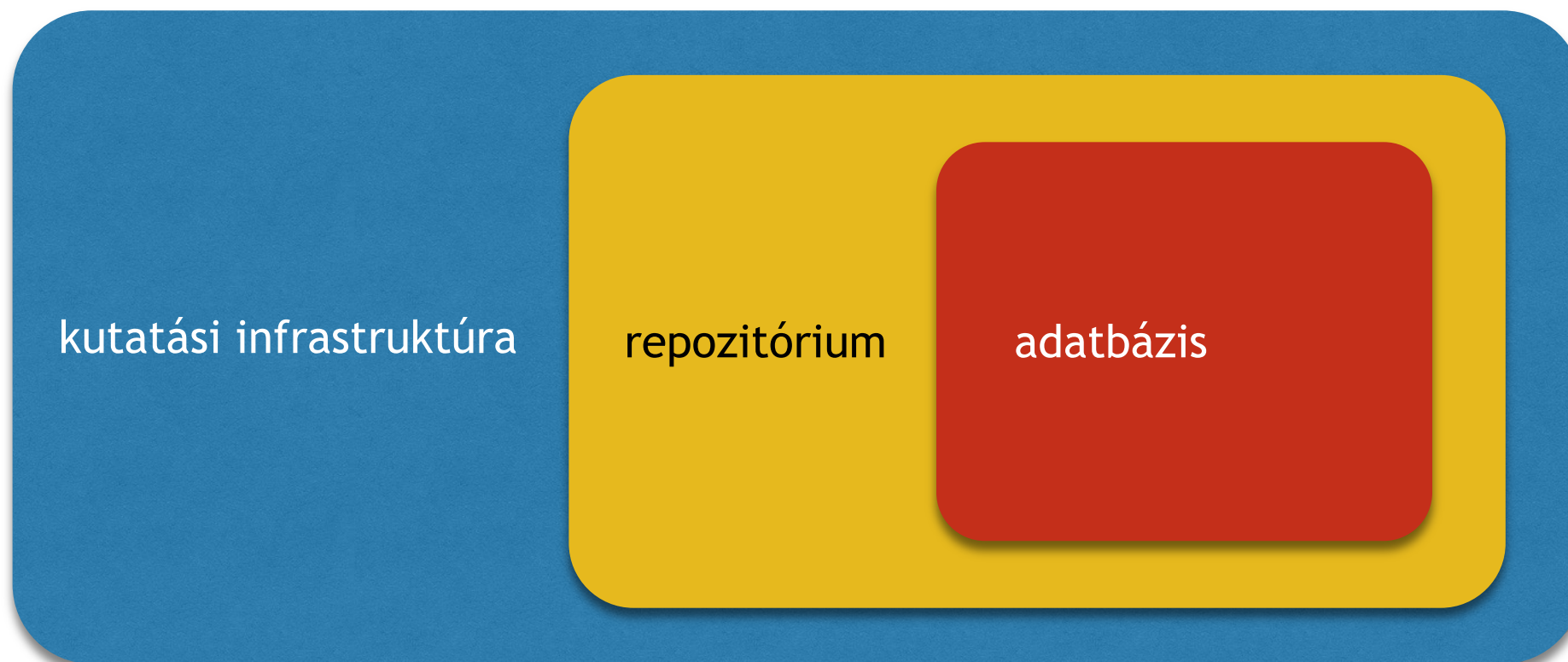
- ▶ A fogalom meghatározása (defined by the RDA Data Foundation and Terminology Working Group):
- ▶ “A digital repository is an infrastructure component that is able to store, manage and curate Digital Objects and return their bitstreams when a request is being issued.”
- ▶ A digital object (DO) is represented by a bitstream, is referenced and identified by a persistent identifier and has properties that are described by metadata.
- ▶ Metadata contains descriptive, contextual and provenance assertions about the properties of a DO.

Repozitórium

- ▶ repozitórium = tároló
- ▶ tartalom fajtája alapján:
 - ▶ full text publikáció, preprint, tézis, disszertáció, technical report, whitepaper, learning object, **research object**, adatrekord, **nyers kutatási adat**, **származtatott kutatási adat**,metaadat rekord
- ▶ a tartalom származási helye alapján:
 - ▶ személyes adat, folyóirat, intézet, részleg, projekt, ágazat, nemzeti, nemzetközi,...
- ▶ a tartalom felhasználási célja alapján:
 - ▶ hozzáférés biztosítása, keresés megkönnyítése, hosszú távú digitális tárolás, disszemináció, intézeti/céges tulajdonlás, megosztás, újrafelhasználás, ...

Repozitórium

- ▶ adatbázis - repozitórium - kutatási infrastruktúra



Korszerű repozitórium funkciók

- ▶ a tárolt objektumok egyértelmű és perzisztens azonosítása (weben), HTTP URI, DOI, handle,...
- ▶ az objektumok lehetséges felhasználásainak egyértelmű szabályozása (licenszek)
- ▶ objektumokhoz extra tartalom társítások (overlay content) belső funkcióként avagy third party szolgáltatások által (annotáció, review, comment,...)
- ▶ hálózatos objektum modellel rendelkező objektumok (LD, LOD) kezelése

Korszerű repozitórium funkciók 2

- ▶ az objektumok individuális exportja, importja (interfészek és protokollok)
- ▶ objektumok nagy tömegű exportja, importja (pl. harvest data mining célra)
- ▶ a repozitórium aktivitások loggolása, monitorozása (belső, külső célra)
- ▶ az objektum transzformációk és aktivitások külső transzparenciájának biztosítása (pl. real time notification, ajánlórendszer stb. szolgáltatások)
- ▶ felhasználó azonosítás (standard user ID, egyéniesített szolgáltatások)

Korszerű repozitórium funkciók 3

- ▶ a repozitórium működésének leírására és a felhasználás mikéntjére szabványosított metrikák (leíró metaadatok) bevezetése (működés, repozitórium kezelés optimalizáció, hálózati stb. külső szolgáltatások összehangolása, szervezetek számára metaadatszolgáltatások)
- ▶ hosszú távú digitális tárolás
- ▶ új tartalomfajták/formák kezelése (PDF, TEI XML, ...)
- ▶ összetett (kompozit, csomag) objektumok kezelése (**research object (RO)**, tárolási objektum, ...)

RO - Research Object

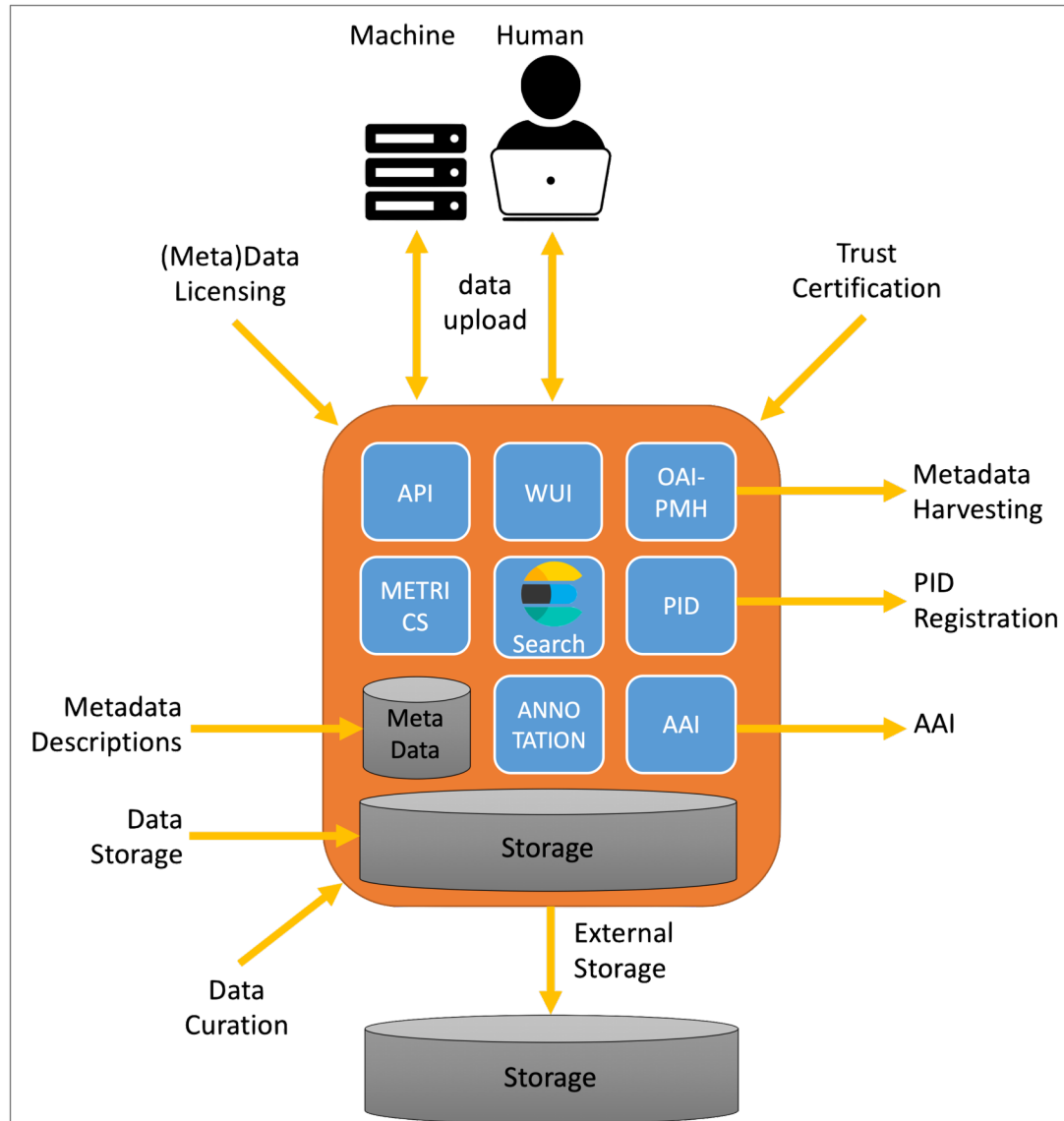
- ▶ Tudományos kutatások/kísérletek erőforrásainak és a kutatási körülmények (context) egy összecsomagolt, koncentrált ábrázolása
- ▶ Csomag tartalma:
 - ▶ **Adat** - a kutatás során felhasznált avagy létrehozott digitális adat-és tartalom- állományok
 - ▶ **Módszer** - az adatok előállításának, feldolgozásának, analizálásának módszere, a módszertan és/vagy kutatási/adatfeldolgozási workflow-k leírása
 - ▶ **Eredet információk** (provenance data) - az adatkeletkezés körülményeinek, forrásainak metaadatai, az adatfeldolgozás lépéseinek leíró metaadatai, a kísérleti elrendezések és paraméterbeállítások adatai,...
 - ▶ **Kutatók** - a kutatás résztvevői
 - ▶ **Annotációk** - az erőforrásokhoz társított egyéb overlay (meta)adatok, melyek segítik az interpretációk elkészítését, a megértést stb.
 - ▶ **Publikációk, interpretációk, disszeminációk** - cikk, whitepaper, prezentáció, ...

Korszerű repozitórium funkciók 4.

- ▶ repozitórium beágyazhatósági, interoperabilitási feltételek (pl. beágyazás kutatási infrastruktúrába)
- ▶ kommunikációs interfészek, API, szabványos protokollok megléte (adattranzferszintű interoperabilitás)
- ▶ szemantikus szintű metaadateleírások (szemantikus interoperabilitás)
- ▶ repozitórium működési transzparencia (működési interoperabilitás)
- ▶ repozitórium stabil fenntartása hosszú távon (működési, üzleti, finanszírozási, ... modellek)
- ▶ repozitórium fenntartó - felhasználó kommunikáció, feedback (szolgáltatás minőségének fejlesztése, új igények kiszolgálása)

Korszerű repozitórium (példa)

EOSC hub konferencia:



Digital Repository

From: Presentation:

Mark van der Sanden & Heinrich Widmann

2020 EOSC-hub Week 20.05.2020

Szuprastruktúra



Adat- és tartalom-infrastruktúra

Kommunikációs- és számítási-
infrastruktúra

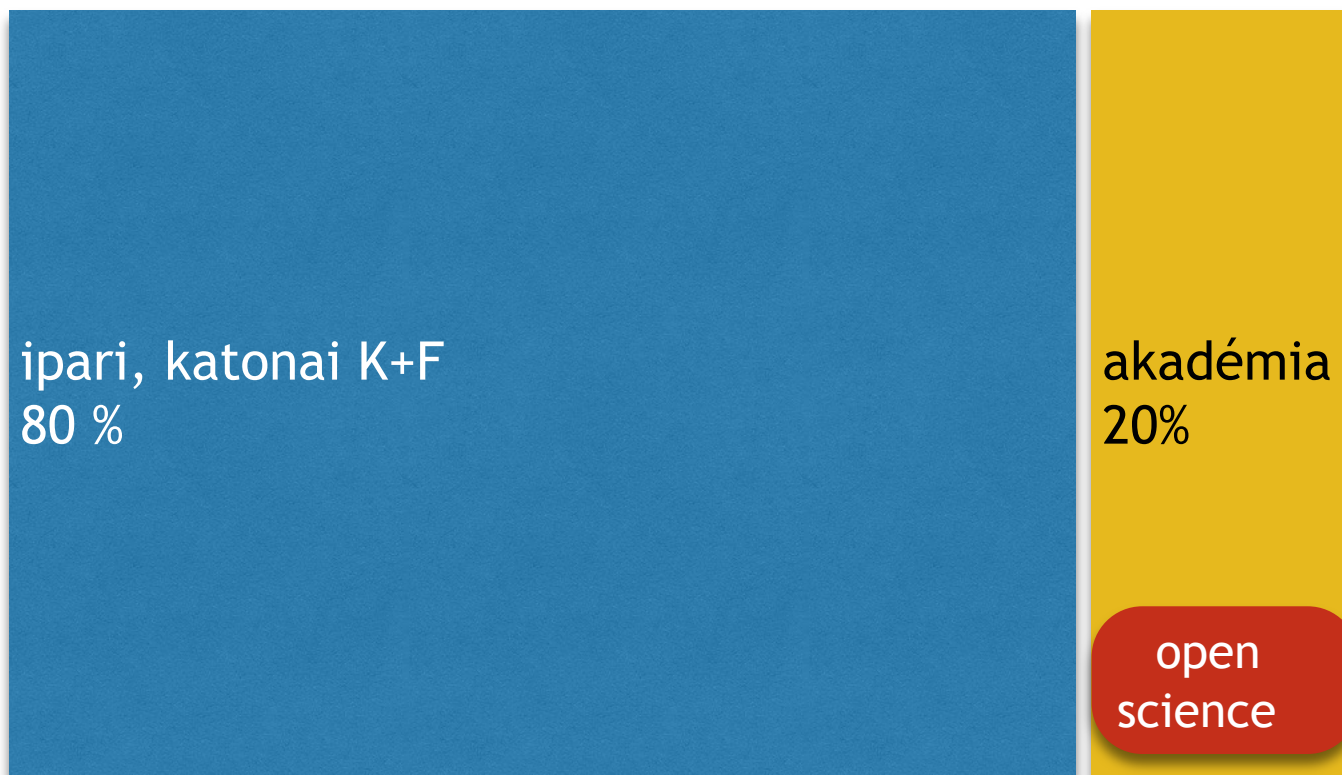


Adat- és tartalom-szuprastruktúra

- ▶ adat- és tartalom-infrastruktúra (szuprastruktúra)
 - ▶ adatszolgáltatások illeszkedő (interoperábilis) rendszere
 - ▶ alap-adatok és infrastrukturális alap-szolgáltatások (pl. névterek, ontológiák, tezauruszok,....)
 - ▶ interoperabilitás
 - ▶ szabványos interfészek és kommunikációs protokollok
 - ▶ szabványmegfelelési ellenőrzésére konformancia központok és konformancia szolgáltatások
 - ▶ szemantikus interoperabilitás
 - ▶ működési interoperabilitás
 - ▶ fenntarthatóság (működési, üzleti, finanszírozási modellek, intézményi működés)

Open Science ?

a változás iránya





Adatrepozitóriumok

Dr. Kovács László

SZTAKI DSD

Elosztott Rendszerek Osztály

laszlo.kovacs@sztaki.hu