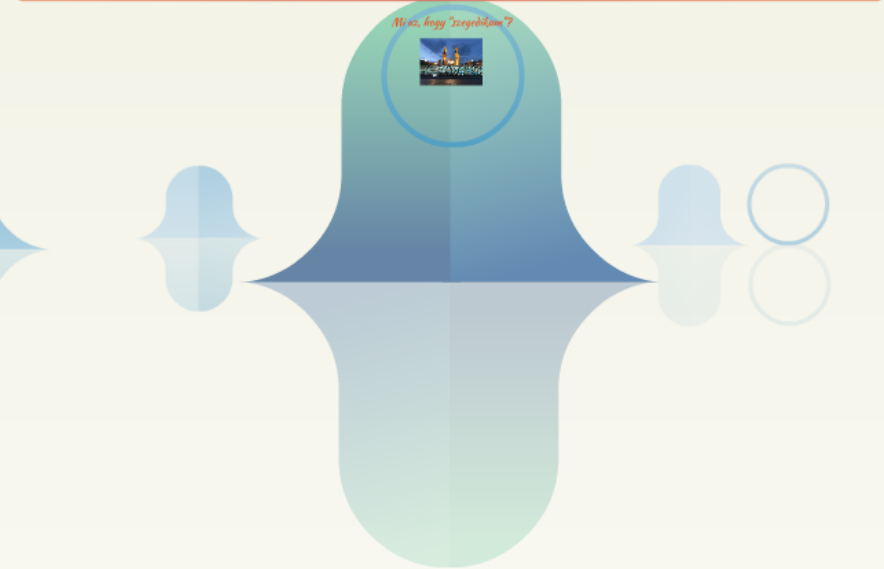
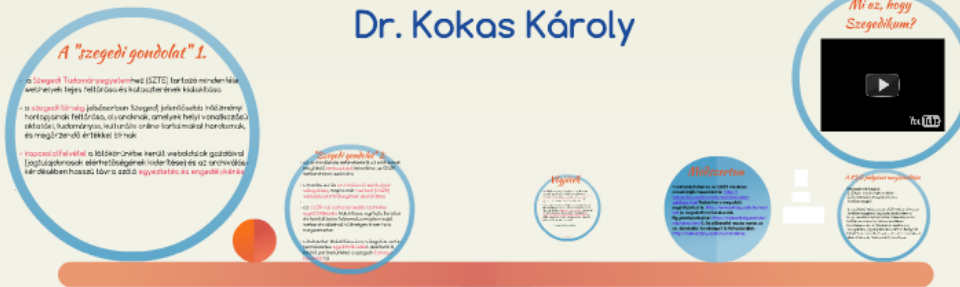


# Szegedikum a webarchivumban



## A helyi érdekű webarchiválás lehetőségei az OSZK webarchiválási programja keretében

Dr. Kokas Károly



# Szegedikum a webarchivumban



## A helyi érdeklő webarchiválás lehetőségei az OSZK webarchiválási programja keretében

Dr. Kokas Károly

### A "szegedi gondolat" 1.

- a Szegedi Tudományegyetemhez (SZTE) tartozó mindenféle webhelyek teljes feltárása és kataszterének kialakítása
- a szegedi létezés (jelősen Szeged) jelentősebb intézményi hierarchiájának feltárása, olyanoknak, amelyek helyi vonatkozású oktatási, tudományos, kulturális online tartalmakat hordoznak, és megőrzendő értékkel bírnak
- kapcsolatfelvétel a látókorunkbe került webladlak gazdáival (jogtulajdonosok elérhetőségének kiderítése) és az archiválás-kérdésében hosszú távra szóló egyeztetés és engedélykérés

### "Szegedi gondolat" 2.

- a már online webhelyekre az online megőrzés technológiák segítségével az OSZK webarchivum számára
- a már online webhelyekre az online megőrzés technológiák segítségével az OSZK webarchivum számára
- az OSZK web oldala adatainak a webarchivumban megőrzésének biztosítása, a webarchivumban megőrzésének biztosítása, a webarchivumban megőrzésének biztosítása
- a kataszter kidolgozása és a világhálóra való bemutatók megvalósítása, a kataszter kidolgozása és a világhálóra való bemutatók megvalósítása

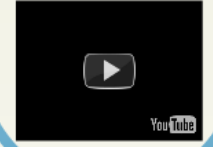
### Végül

...és a webarchivumban megőrzésének biztosítása, a webarchivumban megőrzésének biztosítása

### Módszertan

A webarchiválás az OSZK részben automatizált folyamat, amely során a webarchivumban megőrzésének biztosítása, a webarchivumban megőrzésének biztosítása

### Mi az, hogy Szegedikum?



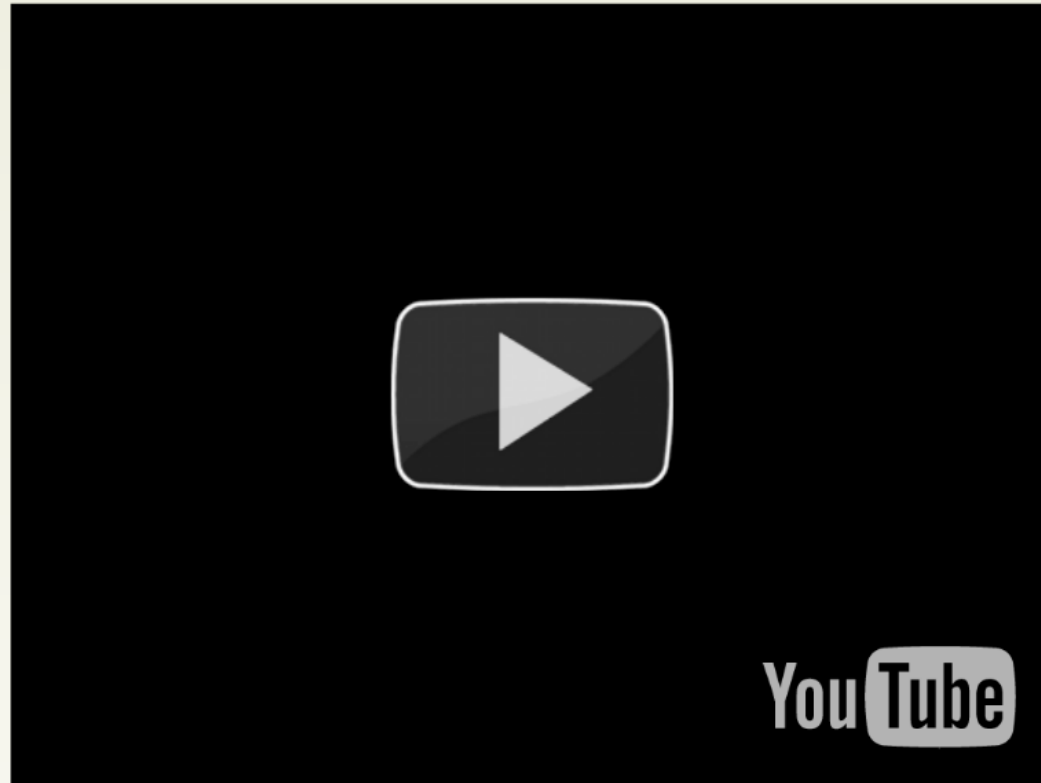
### A SZTE teljesítet megőrzéséről

A webarchivumban megőrzésének biztosítása, a webarchivumban megőrzésének biztosítása

### Mi az, hogy "szegedikum"?



# *Mi az, hogy Szegeedikum?*



# Mi az, hogy "szegeedikum"?







# A KDS pályázat megközelítése

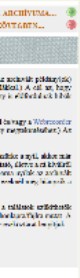
Pályázati felhívásban:

3.2 Elvárt eredménytermékek

3.2.1 Kötelezően megvalósítandó  
tevékenységek:

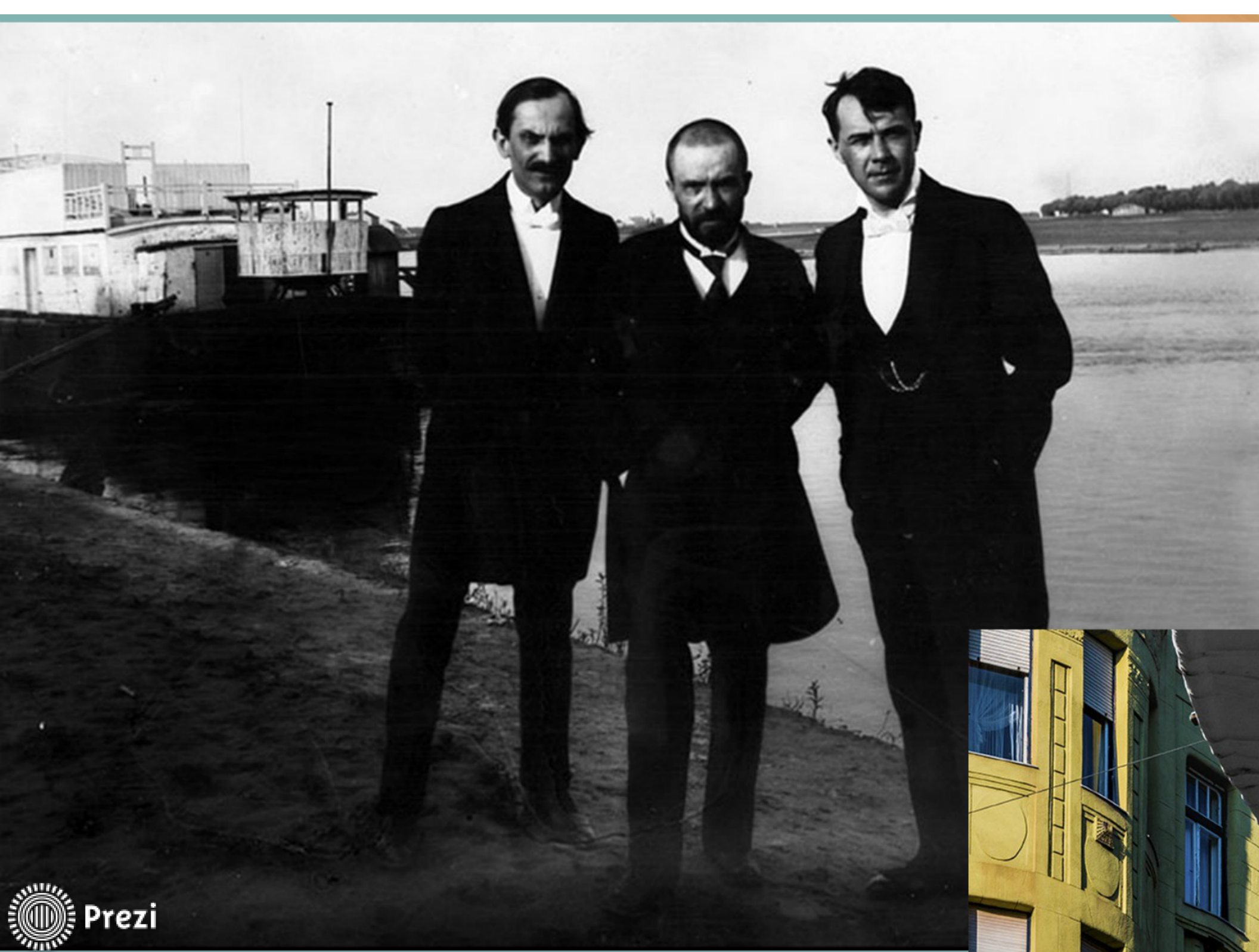
6. a pályázó feladata az OSZK webarchiválási  
tevékenységének segítése, elsősorban a  
helyi vonatkozású oktatási, tudományos és  
kulturális online tartalmak esetében.

Részfeladatok: archiválandó webhelyek  
válogatása, jogtulajdonosok elérhetőségének  
kiderítése, a mentett weboldalak minőségének  
ellenőrzése és metaadatok készítése.





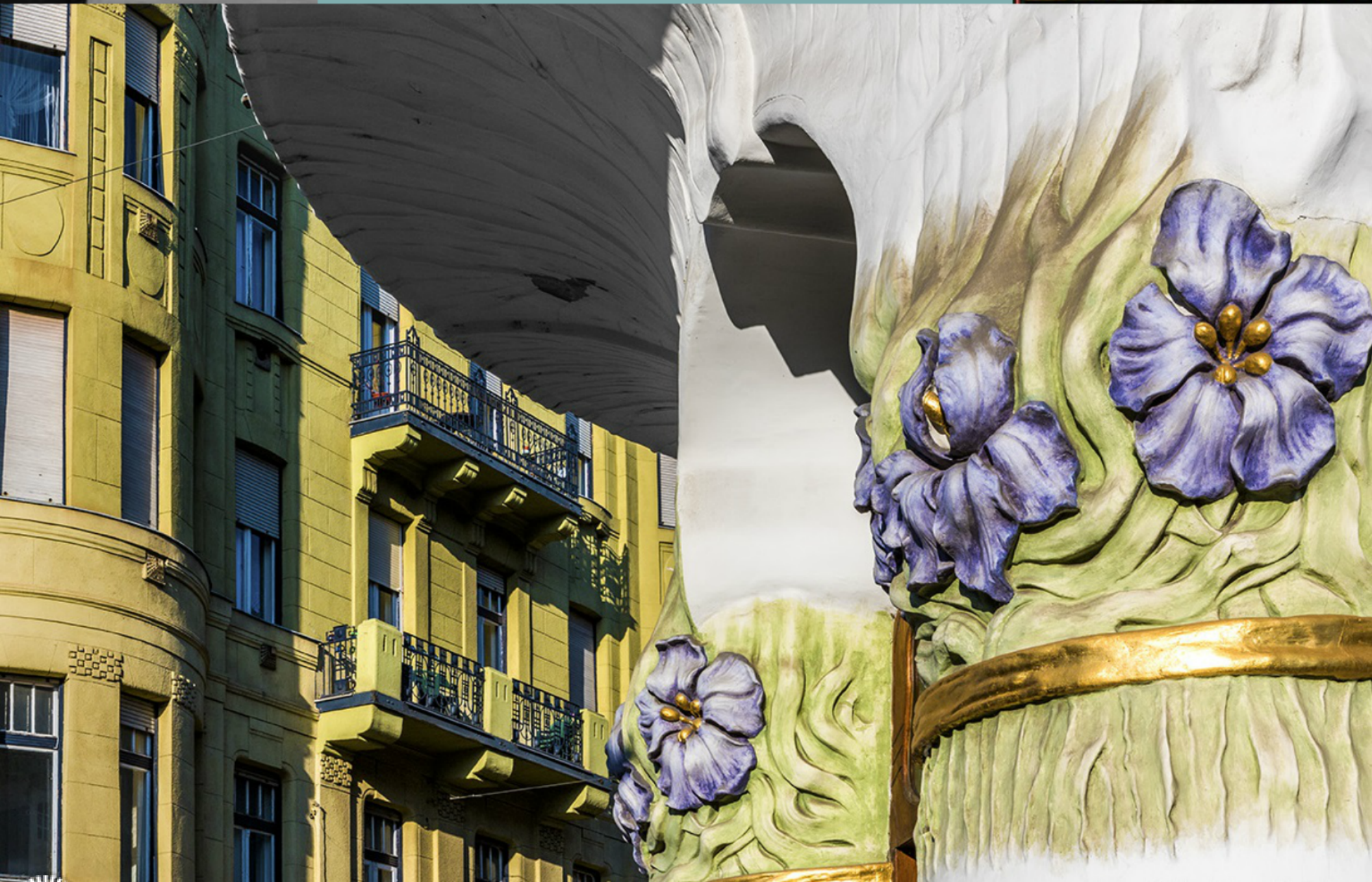




# A "szegedi gondolat" 1.

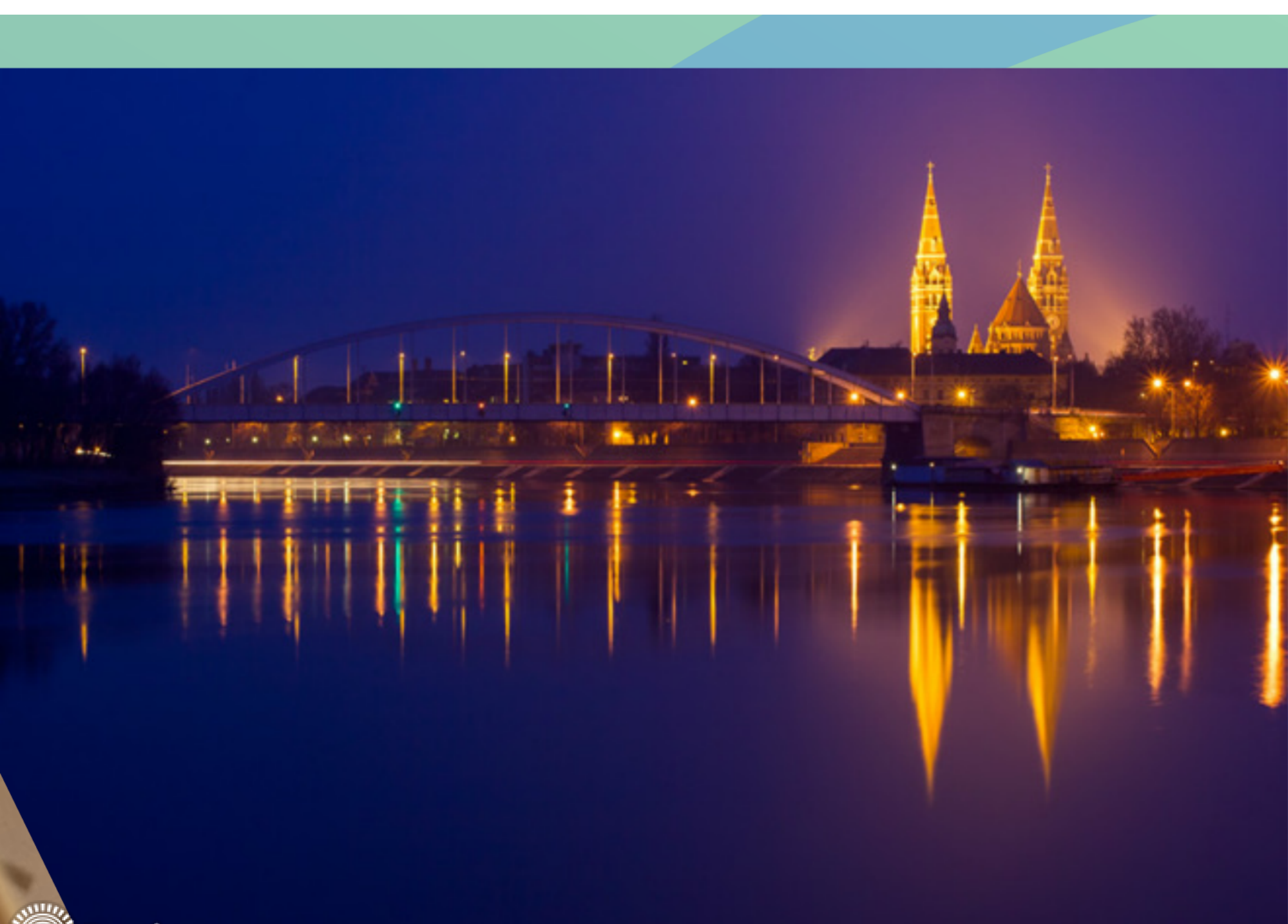
- a **Szegedi Tudományegyetem**hez (SZTE) tartozó mindenféle webhelyek teljes feltárása és kataszterének kialakítása
- a **szegedi térség** (elsősorban Szeged) jelentősebb intézményi honlapjainak feltárása, olyanoknak, amelyek helyi vonatkozású oktatási, tudományos, kulturális online tartalmakat hordoznak, és megőrzendő értékkel bírnak
- **kapcsolatfelvétel** a látókörünkbe került weboldalak gazdáival (jogtulajdonosok elérhetőségének kiderítése) és az archiválás-kérdésében hosszú távra szóló **egyeztetés és engedélykérés**





## "Szegei gondolat" 2.

- az archiválándó webhelyekről az előírásnak megfelelő **metaadatok** készítése az OSZK webarchívum számára
- a munka során **archiválándó webhelyek válogatása**, majd a már **mentett (OSZK) weboldalak minőségének ellenőrzése**
- az **OSZK-val való szorosabb technikai együttműködés** kialakítása, egyfajta tanulási és konzultációs folyamat, a majdani saját webarchiváláshoz szükséges know-how megszerzése
- a kataszter kialakítása és a válogatás során természetes **együttműködést** alakítunk ki, kitűnő partnerünkkel, a szegei **Somogyi-könyvtárral**



Prezi

# Módszertan

A metaadatoláshoz az OSZK részletes útmutatóját használják: <http://mekosztaly.oszk.hu/mia/xml/utmutato-webhely.html> (beleértve a megadott segédfájlokat is: <http://mekosztaly.oszk.hu/mia/xml> (a megadott mintarekordok figyelembevételével: <http://mekosztaly.oszk.hu/mia/demo/xml/>)). Az előkészítő munka során az ún. demóoldal tanulságait is felhasználjuk: <http://mekosztaly.oszk.hu/mia/demo>
























Közi  
Közi terkep/m...



## Index of /mia/xml

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
 <a href="#">Parent Directory</a>		-	
 <a href="#">mezonevek-sugo.doc</a>	2019-08-10 13:30	394K	
 <a href="#">mezonevek-sugo.pdf</a>	2019-08-10 13:30	82K	
 <a href="#">mezonevek.doc</a>	2019-07-31 23:09	282K	
 <a href="#">mezonevek.html</a>	2019-07-31 23:09	60K	
 <a href="#">mezonevek.pdf</a>	2019-07-31 23:09	57K	
 <a href="#">mia-regi.xsd</a>	2019-07-30 12:13	84K	
 <a href="#">mia-regi.xsl</a>	2019-07-30 12:13	24K	
 <a href="#">mia.xsd</a>	2019-10-01 14:52	86K	
 <a href="#">mia.xsl</a>	2019-09-04 09:38	25K	
 <a href="#">mia_set.xsd</a>	2019-07-05 10:55	55K	
 <a href="#">mia_set.xsl</a>	2019-03-07 17:39	14K	
 <a href="#">ures-reszgyujtemeny-teljes.xml</a>	2018-09-23 14:10	4.0K	
 <a href="#">ures-reszgyujtemeny.xml</a>	2018-09-23 14:10	3.1K	
 <a href="#">ures-webhely-teljes.xml</a>	2019-08-05 17:39	6.5K	
 <a href="#">ures-webhely.xml</a>	2019-08-05 17:39	4.9K	
 <a href="#">utmutato-reszgyujtemeny.html</a>	2018-08-25 21:42	398	
 <a href="#">utmutato-webhely.html</a>	2018-08-25 21:42	391	
 <a href="#">xml_notepad.png</a>	2018-07-06 20:04	140K	





# Útmutató a szelektíven archivált egyedi webhelyek XML-alapú metaadatolásához

Szerkeszti: Drótos László és Visky Ákos László

*Utoljára módosítva: 2019-08-12*

## **Tartalom:**

[Az adatséma és az XML fájlok](#)

[XML Notepad](#)

[A metaadatolás előkészítése](#)

[A metaadatolás befejezése](#)

[BIBLIOGRÁFIAI \(LEÍRÓ\) METAADATOK](#)

[dc\\_identifier](#)

[dc\\_title](#)

[source\\_of\\_description](#)

[collector](#)

[dc\\_creator](#)

[dc\\_contributor](#)

# Mi az, hogy "szegedikum"?







## OSZK WEBARATÁS – DEMÓ ARCHÍVUM

OSZK-S WEBHELYEK ARCHÍVUMA...

KERESÉS A TELJES SZÖVEGBEN...

KÖNYVTÁR | LEVÉLTÁR | MÚZEUM | MŰVÉSZET | TUDOMÁNY | OKTATÁS | ÖNKORMÁNYZAT |  
TURIZMUS | TÖRTÉNELEM | KÖNYV | E-PERIODIKA | HIRDETÉS | BLOG | SZEMÉLYES

Ez a kis gyűjtemény az OSZK-ban zajló **kísérleti webaratás projekt** keretében készül azokból a mentésekből, amelyeknél az eredeti honlaptulajdonos engedélyt adott a Nemzeti Könyvtár számára az archivált példány(ok) nyilvános szolgáltatására – egyelőre 2018 végéig. (További felajánlásokat, javaslatokat örömmel veszünk **ezen az úrlapon**. A nyilvános szolgáltatáshoz pedig **ezt a szerződést** kötjük a tartalomgazdákkal.) A cél az, hogy demonstráljuk a jelenlegi webarchiválási technológia lehetőségeit és korlátait. Bár a demóhoz kiválasztott webhelyek az automatikus módszerekkel viszonylag jól archiválhatók közül kerültek ki, még így is előfordulnak hibák és hiányok a lementett példányokban. Ezen problémák egy része a site-ok **robot-barát** és **archívum-barát** kialakításával orvosolható.

A piros nyilakra linkelt mentések a **Heritrix**, a **Brozzler**, a **Webrecorder**, vagy a **HTTrack** szoftverrel készültek, általában nem teljes mélységben. A megjelenítés az **Open Wayback** és a **PyWb** szoftverekkel és/vagy a **Webrecorder Player** online verziójával történik, továbbá a fájlrendszerben tárolt HTTrack mentések a webszerveren át is megnézhetők. (Minden esetben a mentés dátumát jelző linkre kell kattintani az archiv példány megtekintéséhez.) Az egyes archiváló és megjelenítő eszközök képességei eltérőek, ezért ahol több piros nyíl is van, ott érdemes mindegyiket kipróbálni.

A mentett változatra és az eredeti kezdőlapról készül oldalképre mutató nyilak mellett van egy-egy link az amerikai **Internet Archive**-ban található mentésekre, valamint az eredeti "élő" honlapra is. (Ha szürke a nyíl, akkor már nem él az oldal.) A nyilakra kattintva a weblapok új böngészőlapra nyílnak meg, így könnyen összehasonlíthatók a mentett verziók és az eredeti oldalak. A sárga gomb pedig az adott doménről kifelé mutató, illetve a rá kívülről hivatkozó linkek gráfját rajzolja ki - az archívumban levő mentések alapján (mivel a demó archívum még kicsi, ezért ez utóbbi, *ingoing* típusú linkek száma nagyon kevés). Az utolsó oszlopban levő barna nyilak az archivált webhelyeket leíró metaadat rekordokra visznek (a böngészőben a Ctrl/U megnyomásával látszik az eredeti XML forráskódjuk). Az újonnan felvett tételknél a metaadatok később készülnek csak el, így ezeknél még hiányzik a barna nyíl.

A fejlesztés alatt levő **SolrMIA** keresőnk mellett kísérleti jelleggel a **SolrWayback Search** nevű felület is kipróbálható, amellyel szintén a mentett webhelyek teljes szövegében lehet keresni és a találatok szűkíthetők doménnevek, fájltypusok és a mentés éve szerint. A találati listában a weboldal vagy fájl (nagy betűvel kiemelt) címére kattintva jutunk el az archivált verzióra, az *Url*: sorban levő cím pedig az eredeti honlapra/fájltra mutat. A *Show full post* felirat alatt megnézhetők az adott találat részletes adatai. Egy találatra kattintva további információk is megjeleníthetők az adott oldalról vagy doménről, ha a bal felső sarokban levő *Toolbar* eszköztárat lenyitjuk.

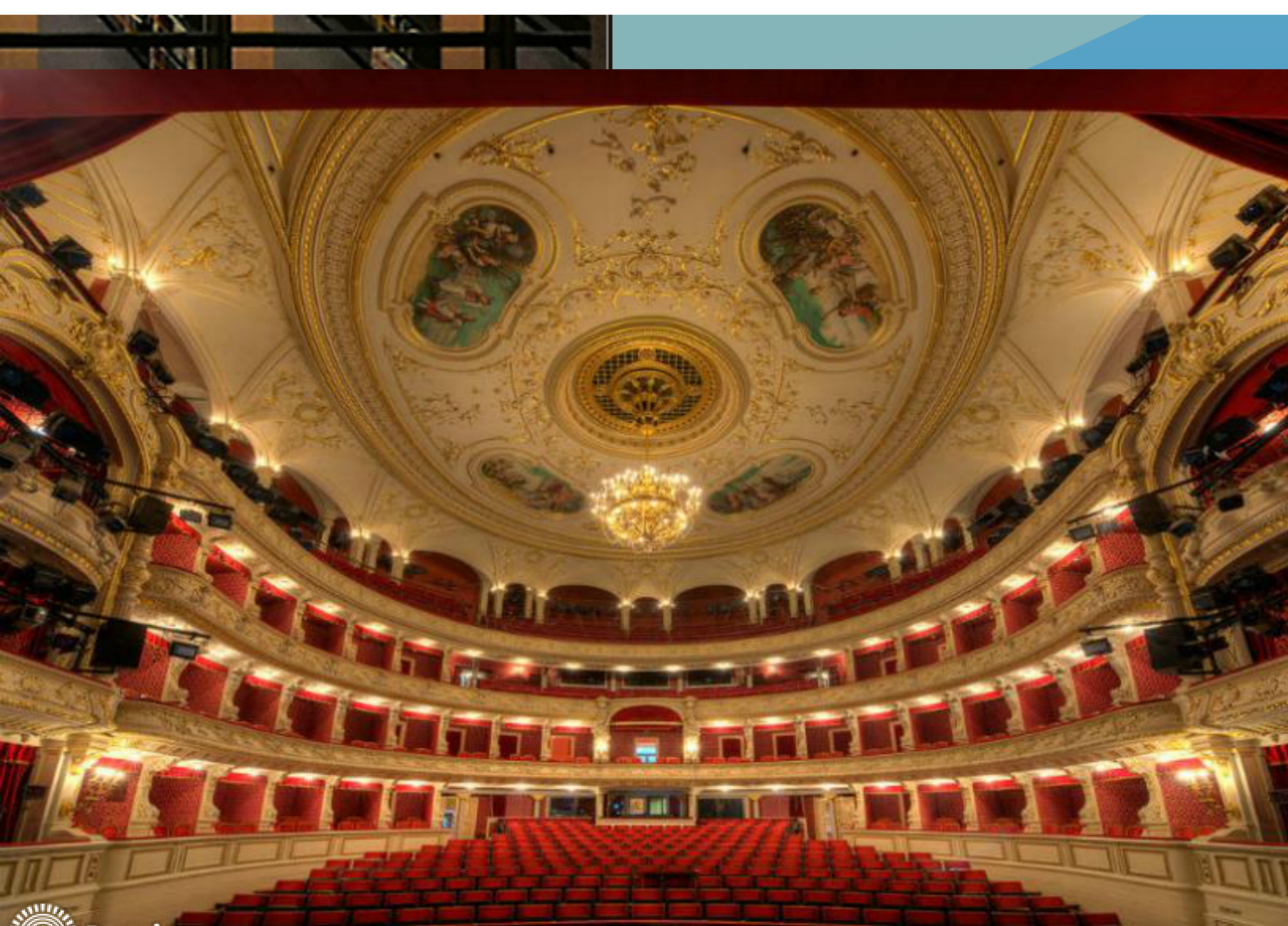
Jelmagyarázat a megjelenítő programokhoz:

Open Wayback

PyWayback

Webrecorder Player

HTTrack+webszerver





# Végjáték

- Az OSZK kérése szerint, addig is, amíg elkészül a tervezett adatkezelő rendszer, egy egyszerű szövegfájlban, vagy ha lehetséges, már **eleve XML-ben adjuk át a kiszemelt webhelyek metaadatait** (pl. ingyenes XML szerkesztő használatával).
- A feladatkör során tervezzük a későbbiekben, hogy bár kezdetben a javaslatunkra az archiválást az OSZK végzi, de később a **megtanult technológiákkal** hasonló, kompatibilis, de sokkal gyakoribb mentéseket **mi magunk is készítünk, saját hardver eszközökkel**, az OSZK-val azonos módszertannal, ez a két folyamat azután párhuzamosan is folyhat.

A megvalósítás korlátai...

*Köszönöm a figyelmet!*



*Köszönöm a figyelmet!*

