



BIBLIOTHECA NATIONALIS HUNGARIAE

**Drótos László – Németh Márton**  
(Országos Széchényi Könyvtár)

---

# **A webarchiválás oktatása**

NETWORKSHOP 2018  
Eszterházy Károly Egyetem, Eger  
2018. április 4-6.

## Miről lesz szó?

---

- Meg kell tanulnunk megőrizni az internetet
- Webarchiválás-oktatási tervek az OSZK-ban
- A dán NetLab online kurzusának tapasztalatai
- Az IIPC Training Working Group tevékenysége
- Ajánlott források



## Meg kell tanulnunk megőrizni az internetet

---

- ❑ A digitálisan születő kultúránk megőrzése is közgyűjteményi feladat
- ❑ Webkönyvtárak, weblevéltárak és webmúzeumok kellene
- ❑ Ehhez speciális technikai, gyűjteményszervezési, metaadatolási, szolgáltatási, jogi stb. ismeretekre van szükség
- ❑ Ezeknek az ismereteknek be kell épülniük a képzésbe, a továbbképzésbe és az önképzésbe



## Webarchiválás-oktatási tervek az OSZK-ban

---

- 2017 tavaszától kísérleti webarchiváló projekt
- Honlap, wiki, demó archívum
- „404 Not Found” workshop
- Előadások, bemutatók, publikációk
- Kapcsolatfelvétel tanszékekkel
- Továbbképző tanfolyam a Könyvtári Intézetben
- „Blended learning” típusú tananyag terve





## OSZK WEBARATÁS – TESZT FÁZIS

(Legutóbbi módosítás: 2018. március 9.)

DEMO | WIKI | IRODALOM | LISTA | WORKSHOP

### A PROJEKT



Az Országos Széchényi Könyvtár 2017 áprilistól az OSZK projekt keretében elkezdett kísérletezni a webarchiválás technológiájával; az internetes források hosszú távú megőrzésének érdekében. Ez az alprojekt 2018 végéig tart és az informatikai hámeret a KIEÜ biztosítja.

Ennek a kutatási és fejlesztési munkának az a célja, hogy megalapozza egy leendő magyar internet archívum feltételeit:

- álljon rendelkezésre egy olyan műszaki infrastruktúra, amely képes a nyilvános internetről nagy tömegű, sokféle formátumú digitális tartalmat begyűjteni, feldolgozni, biztonságosan megőrizni és - a jogi státusz függvényében - szolgáltatni;
- legyenek a magyar közgyűjteményekben dolgozó könyvtárosok, levéltárosok és informatikusok között olyan szakemberek, akik értenek ehhez a tevékenységhez;
- készüljenek el olyan dokumentumok, amelyek alapján szabályozott módon folyhat ez a munka (pl. gyűjtőkör leírás és válogatási szempontok; a magyar webtér lehatárolása, metadátum struktúra, szerződésmenték a tartalomgazdák számára; az archiválási tevékenység és az archívumhoz való hozzáférés jogszabályi előírásai).

A teszt fázisban néhány száz kulturális és tudományos webhely kerül kiválasztásra (pl. könyvtári, levéltári, múzeumi honlapok, egyetemi és kutatóintézetek oldalai, elektronikus folyóiratok, szakmai blogok), melyeknek a tulajdonosait elektronikus levélben értesítjük erről és engedélyt kérünk az archiválásra ill. esetleg a lementett változat szolgáltatására is egy demonstrációs célra létrehozott gyűjteményben.

A megőrzésre érdemes magyar webhelyek válogatásához, illetve az archívum megtervezéséhez várjuk a téma iránt érdeklődő kollégák jelentkezését a [moz@mek.oszk.hu](mailto:moz@mek.oszk.hu) címre. (A gyűjtőkör részletesebb lehatárolása itt olvasható.)

### HÍREK

**2018. 03. 09.** A *demo archiválás* kísérleti jelleggel beüzemeltük a *WOLB* Wayback Search nevű felületet, amellyel a mentett webhelyek teljes szövegében lehet keresni és a találatok szűrhetőek: doménnevek, fájl típusok és a mentés éve szerint. A találati listában a weboldal vagy fájl (nagy betűvel kiemelt) címere kattintva jutunk el az archivált verzióra, az *URL* sorban lévő link pedig az eredeti honlapra/fájltra visz. A *Show full post* felirat alatt megnézhetők az adott találat részletes adatai, és ezen belül a *text* rovatban az a szövegtörzs is, ahol a keresett szó előfordul.

**2018. 03. 02.** 2019-től az Országos Széchényi Könyvtár tagja lett a webarchiválással foglalkozó intézményeket tömörítő nemzetközi

### SZAKIRODALOM MAGYARUL

- *Andriev, Alina: Web archiving made in Slovakia: Ezerfős projekt az elektronikus információtudomány gyűjtésén és archiválásán.*  
In: *Tudományos és Műszaki Tájékoztatás*, 2007, (34. évf.), 10. sz.
- *Balog, János - Thompson, David: Az első magyar webarchívum az Egyetem Székelyházban.*  
In: *Tudományos és Műszaki Tájékoztatás* 2006, (53. évf.), 10. sz.
- *Cserbóci, László: A web, web és a lementéspolitika-vezetés.*  
In: *Könyvtári Figyelő*, 2009, (55. évf.) 3. sz. p. 318-320.
- *Császár, Edit (ed. Delella László): Webarchiválás a világhon.*  
In: *Tudományos és Műszaki Tájékoztatás*, 2010, (57. évf.), 2. sz.
- *Dancs, Szabolcs: Webarchiválás politika.*  
In: *Könyv, könyvtár, könyvtáros*, 2011, (20. évf.), 10. sz.
- *Durgó Péter: A hazatérítés nemzeti bibliográfiája az Internet - Válogatás a hazatérítés az új kultúrákban.*  
Budapest: ELTE BTK, 2005.
- *Dénes László: Mű a Műt? Javaslat egy Magyar Internet Archívum létrehozására.*  
In: *Tudományos és Műszaki Tájékoztatás*, 2006, (53. évf.), 6. sz.
- *Dénes László: Az internet archiválás politikai feladatja.*  
In: *Tudományos és Műszaki Tájékoztatás*, 2017, (64. évf.), 7-8. sz.
- *Hegyházi, Érika: Hol tart ma a webarchiválás?*  
In: *Könyvtári Figyelő*, 2014, 4. sz.
- *Iller, Gábor: Webarchiválás a francia gyakorlatban.*  
In: *Könyvtári Figyelő*, 2009, (55. évf.) 3. sz. p. 353-354.
- *Jedlik, Esztergón: Elektronikus források begyűjtése az archívumokban.*  
Lemondás: [www.moz.oszk.hu/iranyelvek](http://www.moz.oszk.hu/iranyelvek)

KÖNYVTÁR | LEVELEZŐ | MŰZEUUM | MŰVESZET | KÜLVETÉS | OKTATÁS | TÖRTÉNELEM | E-PERIODIKA

Ez a kis gyűjtemény az OSZK-ban zajló kísérleti webaratók projektjének keretében készült azokról a mentésekből, amelyeknél az eredeti honlap tulajdonos engedélyt adott a Nemzeti Könyvtár számára az archivált példány(ok) nyilvános szolgáltatására – egyelőre 2018 végéig. (További felajánlásokat örömmel veszünk a [mentes@oszk.hu](mailto:mentes@oszk.hu) e-mail címen.) A célja az, hogy demonstráljuk vele 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 nyílakra linkelt mentések a *Heritrix* szoftverrel készültek 2017 decemberével kezdődően, nem teljes mélységben, a videók és más nagy méretű állományok letöltésének kizárásával, és az eredeti webhelyen lévő [robots.txt](http://robots.txt) fájlban lévő tiltások tiszteletben tartásával. A megjelenítés az *Open Wayback* szoftverrel történt, amelyben a mentés dátumára kell kattintani az archív példány megnézéséhez. A mentett változatra és az első mentés időpontjában készült képernyőfotóra mutató nyíl mellett van egy-egy link az amerikai *Internet Archive*-ban található mentésekre, valamint az eredeti "élő" honlapra is (amennyiben az még létezik). A nyílakra kattintva a linkek új böngészőablakon nyílnak meg, így könnyen összehasonlíthatók a mentett verziók és az eredeti oldalak.

Kísérleti jelleggel a *SOJ.R. Wayback Search* nevű felület is kipróbálható, amellyel a mentett webhelyek teljes szövegében lehet keresni és a találatok szűkíthetők doménnevek, fájl típusok 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 lévő link pedig az eredeti honlapra/fájltra visz. A *Show full post* felirat alatt megnézhetők az adott találat részletes adatai, és ezen belül a *text* rovatban az a szövegek környezet is, ahol a keresett szó előfordul. A keresés magyar nyelvű automatikus szótárral történik, ami azt jelenti, hogy a ragozott és képzett alakokat is megtalálja a program. A leggyakoribb köztisztsági és névmások viszont ki lettek zárva az indexelésből, így ezeket figyelmen kívül hagyja a kereső. Az *Image search* és azon belül a *Geo search* opciókkal képekre és földrajzi nevekre lehet rákérdezni, de ez a keresés csak az alkönyvtár- és fájlnevekben történik, így kevés találatot ad.

## KÖNYVTÁRI HONLAPOK

A webhely neve	URL cím	OSZK mentés(ek)	Képernyőfotó	Internet Archive	Eredeti oldal
Berzsenyi Dániel Városi Könyvtár, Mórahalom	<a href="http://www.moraaloknyvtar.hu">www.moraaloknyvtar.hu</a>				
Csúcs Zoltán Városi Könyvtár, Érd	<a href="http://www.csucskalib.hu">www.csucskalib.hu</a>				
Evangelikus Országos Könyvtár	<a href="http://konyvtar.lutheran.hu">konyvtar.lutheran.hu</a>				
Extergomi Főiskolai Könyvtár	<a href="http://www.bibliotheca.hu">www.bibliotheca.hu</a>				
EX LIBRIS Könyvtár	<a href="http://exlibriskonyvtar.hu">exlibriskonyvtar.hu</a>				
Fejér György Városi Könyvtár, Keszthely	<a href="http://www.fgyvk.hu">www.fgyvk.hu</a>				
Hellischer József Városi Könyvtár, Extergom	<a href="http://www.vkextergom.hu">www.vkextergom.hu</a>				



KÖNYV

Ez a k...  
engede...  
örönm...  
korláto...  
hibákA piro...  
állomá...  
Wayba...  
időpon...  
"elő" f...  
verzióKísérle...  
szűkít...  
az arca...  
adatai...  
történe...  
az ind...  
rákeres

KÖNYV

A webb...

Berzseni

Cseke Z

Evangél

Eszterg

EX LIB

Fejér Gy

Helischi



MIA WIKI

Kezdőlap

Friss változtatások

Lapjaink

Segítség

Eszközök

Mi hivatkozik erre

Kapcsolódó változtatások

Speciális lapok

Nyomtatható változat

Hivatkozás erre a változatra

Lapinformációk

Szócikk

Vitalap

Olvasás

Lapforrás

Laptörténet

Keresés

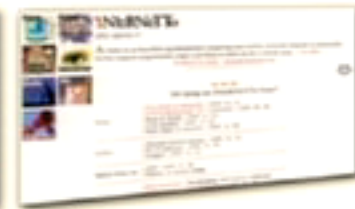
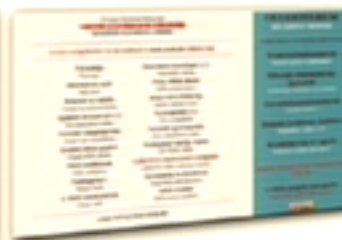
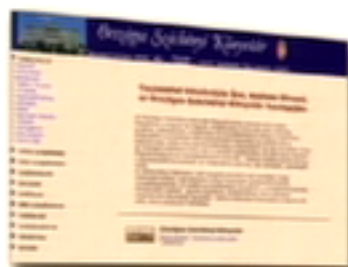


## MIA WIKI

*A Magyar Internet Archivumhoz készülő tudásbázis*

A JELENLEGI KATEGÓRIÁK ÉS SZÓCIKKEIK SZÁMA:

- FOGALMAK / 90
- FORMÁTUMOK / 21
- FÓRUMOK / 9
- HASZNOSÍTÁS / 4
- IRODALOM / 2
- PROJEKTEK / 116
- RENDEZVÉNYEK / 31
- SZABVÁNYOK / 3
- SZERVEZETEK / 20
- SZOFTVEREK / 204
- SZOLGÁLTATÁSOK / 48
- Az összes szócikk listája



Bejelentkezés

emeltük a  
webhelyeklistában a  
jutunk el  
leti  
elők az  
zagja lett a  
etközi

ekt az

a az Egyesült

lágban

et :

tam

e

sa

Ez a k  
engede  
öröme  
korláto  
hibák

MIA

A piros  
állomá  
Wayba  
időpon  
"elő" f  
verzió

Kísérle  
szűkít  
az arca  
adatai,  
történe  
az ind  
rákere

Közölköz  
Friss változ  
Lapj felül  
Segítség  
Eszközök  
Mi hivatko  
Kezdeti változati  
Speciális  
Nyomtat  
Hivatkoz  
Lapfelirat

# 404 Not Found

## Ki őrzi meg az internetet?



### Az előadások prezentációi:

- Lendvai Miklós: A web, mint a kulturális gyűjtőkör része (7 MB)
- Kees Teszelszky: The harvest of the Dutch digital fields: the landscape of webarchiving in The Netherlands (15 MB)
- Peternák Miklós: 404 és akiknek nem kell - *Médiamúzeum Magyarországon: A "Magyar Tartalom" húsz éves jubileuma alkalmából* (55 MB)
- Kampis György: A MIA pilot rövid bemutatása (2013-2016) (3 MB)
- Drótos László - Németh Márton: Kísérleti webaratás projekt az OSZK-ban (10 MB)

### Az előadások videofelvételei a Videotortumban:

- "404 Not Found - Ki őrzi meg az internetet?" workshop

### Beszámolók:

- Ambrus Attila József tudósítása



KÖ

Ez a  
enge  
öröm  
korlá  
hibák

A pi  
állom  
Way  
időpe  
"elő"  
verzi

Kisé  
szűk  
az ar  
adata  
törté  
az in  
raker

KÖN

A web

Berze

Ceula

Evang

Exter

EX LI

Fejér

Helisc

# Kommunikáció és informatika alapjai

## 6. téma: az internet archiválása

Összeállította: Dr. Kosztyánné Dr. Mátrai Rita

(B)

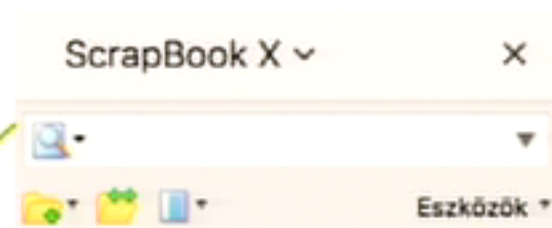
## Gyakorlat: archiváljunk ScrapBook X-szel!

- ▶ Telepítés után a Firefox menüsorában megjelenik a ScrapBook X
- ▶ Oldalsávon megjeleníthető a tartalom: ScrapBook X → Oldalsáv megjelenítése

Keresés a ScrapBook X-ben lévő oldalaink között.

- ◆ Új mappa...
- ◆ New elválasztó
- ◆ Új jegyzet
- ◆ New Note Page

Összes mappa kinyitása/bezárása



## 2 A SZAKMAI TOVÁBBKÉPZÉSI PROGRAMRA VONATKOZÓ ADATOK

### 2.1 A szakmai továbbképzési program megnevezése

Az internet archiválása mint közgyűjteményi feladat

### 2.2 A szakmai továbbképzés célja (minimum 400, maximum 600 karakter)

A résztvevők ismerjék meg az internetes tartalmak hosszú távú megőrzésére alkalmas technikákat, az ezzel foglalkozó külföldi projekteket és az Országos Széchényi Könyvtárban zajló webarchiválási tevékenységet. A kurzus elvégzése után legyenek képesek akár magán célra, akár intézményük számára weboldalakat és webhelyeket tartalmazó gyűjteményeket létrehozni, illetve bekapcsolódni a magyar internet archívummal kapcsolatos munkákba.

### 2.3 A szakmai továbbképzés elsődleges célcsoportja

Közgyűjtemények munkatársai: könyvtárosok, múzeumi és levéltári dolgozók, archivátorok, informatikusok. (A digitális dokumentumok kezelése, feldolgozása terén szerzett tapasztalat előny.)

### 2.4 Azoknak a munkaköröknek a megnevezése, amelyekben foglalkoztatottak számára javasolják a részvételt

#### Közművelődési szakmai munkakörök

- közművelődési szakember I.
- közművelődési szakember II.
- közművelődési szakember III.
- tudományos kutató
- informatikus
- közszolgálati és közösségi média szerkesztő és szervező

#### Muzeális intézményi szakmai munkakörök

- muzeológus asszisztens
- segédmuzeológus (vagy segédkurátor)
- muzeológus (vagy kurátor)
- főmuzeológus (vagy főkurátor)
- műtárgykolcsonzási szakember (vagy múzeumi regisztrátor)
- műtárgyvédelmi munkatárs: műtárgyvédelmi asszisztens

2

## „Az internet archiválása mint közgyűteményi feladat” c. továbbképzési program oktatói beosztása:

2.1

### Oktatók:

Drótos László (DL) <drotos.laszlo@oszk.hu> (E-könyvtári Szolgáltatások Osztály)

Kovács Péter (KP) <kpeter@oszk.hu> (Infrastruktúra Szolgáltatások Osztály)

Moldován István (MI) <moldovan@oszk.hu> (E-könyvtári Szolgáltatások Osztály)

Németh Márton (NM) <nemeth.marton@oszk.hu> (E-könyvtári Szolgáltatások Osztály)

Visky Ákos László (VÁL) <visky.akos.laszlo@oszk.hu> (E-könyvtári Szolgáltatások Osztály)

2.2

### Időbeosztás: *(az óraszámok 50 perces időtartamokat jelentenek!)*

#### 1. nap

1.1. Bevezető: Miért fontos a digitálisan születő, az interneten terjedő kultúra megőrzése, mi a közgyűjtemények felelőssége és mit tesz a nemzeti könyvtár? A kötelempéldány szabályozás ezen a téren hazánkban és külföldön. (MI 3 óra)

1.2. Áttekintés: Archiválási módszerek és archívumfajták. (NM 1 óra)

2.3

1.3. Külföldi projektek: Az Internet Archive, néhány nemzeti webarchívum, valamint egyéb típusú archívum ismertetése és kipróbálása. Az IIPC bemutatása (NM 4 óra)

#### 2. nap

2.1. Internetes tartalmak mentésére és megőrzésére használható ingyenes Windows szoftverek bemutatása és kipróbálása (DL 5 óra)

2.2. Weboldal- illetve webhely-archiváló online szolgáltatások bemutatása és az ingyenesek kipróbálása (DL 3 óra)

2.4

#### 3. nap

3.1. Webarchívum kialakítása Linux szerveren: a Heritrix arató-, az Open Wayback megjelenítő- és a NutchWAX kereső-rendszerek ismertetése és működésük demonstrálása, a WARC/ARC tároló- és a CDX indexformátum rövid bemutatása (KP 2 óra)

3.2 A Web Curator Tool keretrendszer bemutatása és kipróbálása, az archiválásra kiválasztott webhelyek metaadatolása (VÁL 3 óra)

3.3 A Netarchive Suite keretrendszer bemutatása (VÁL 1 óra)

3.4 Válogatási szempontok és utólagos minőségellenőrzés, archiválhatóság, jó és rossz példák bemutatása (VÁL 2 óra)

#### 4. nap

## A dán NetLab online kurzusa

---

- 2017 őszén megtartott ingyenes, angol nyelvű szeminárium az Aarhus Egyetem DIGHUMLAB kutatócsoportjával közösen
- Moodle-alapú távoktatási felület
- Öt fő feladat (assignment), majd az eredmények közös megvitatása
  - Your Areas of Interest
  - Reassessing Your Goals
  - Hands-on with Web Archives and Evaluating Your Finds
  - Hands-on with Doing Your Own Archiving
  - Web Archiving Course Evaluation
- Kézikönyv: Jan Nielsen: Using the Web Archives in Research)



[NETLAB ▾](#)[RESEARCH ▾](#)[FOR RESEARCHERS ▾](#)[SERVICES ▾](#)[RESOURCES ▾](#)[EVENTS ▾](#)

## Online Courses

*"Thanks for an interesting and practical course. It has certainly given me a lot to think about, and many useful ways of approaching my data gathering for my project!"*

*- Ally McCrow-Young, PhD Fellow, University of Copenhagen*

*"I finish the course with a much greater understanding of the challenges of modern web archiving, that I had not considered before. And with a bag full of new tricks on how to preserve things online and how to search in web archives. I think that a course like this should be taken by all historians and humanists."*

*- Barbara Ana Revuelta-Eugercios, Visiting researcher, Saxo Institute, University of Copenhagen*

### Online Course in Web Archives and Web Archiving

More than 98% of the world's data are now digital. A large part of the data are represented on the web, where content is added, deleted and edited constantly and worldwide. Web archiving is the preservation of content, and may be done individually, as well as by large-scale web archiving initiatives. Contemporary studies in the humanities and social sciences need to make use of web archives where earlier documents or versions may be retrieved, and to preserve findings on the live web for future analysis and reference.

The Online Course in Web Archives and Web Archiving will provide an opportunity to build an understanding of how and why archived web content differs from analogous material (such as printed books), and to get significant hands-on experience with web archives and web archiving, based on a focus in your own specific research interests.

The course is offered by NetLab, a national digital research infrastructure, and a community under [DIGHUMLAB](#).

## Course in Web Archives and Web Archiving

NetLab offers an online course in web archives and web archiving for research

The course is offered as online e-learning with a standard duration of five weeks, unless otherwise agreed. The workload during a course will be approximately 5-7 hours of studying per week, plus reading a fifty-page book. The course is problem-based, and for each assignment participants will get a feedback, which they are expected to react to before moving on to the next assignment.

Based in the participant's personal needs, interests or research topics, participants taking the full course will learn about the technical characteristics of archived web, existing web archives, archiving services, and how to do their own web archiving.

The course will teach participants about web archiving, web archives, and characteristics that set archived web content apart from printed material or other analogous types of data. (With archived web specific data may be incomplete or include other limitations that researchers need to be aware of, but archived web also offers large amounts of meta data, and a significant potential for big data analyses).

The course builds upon the source material and software presented in [NetLab's workshops](#). It is an opportunity to refresh knowledge from our workshops, or to acquire the same knowledge without attending a NetLab workshop. In both cases the course offers you the additional opportunity of building personal experience hands-on, based on your specific interests and needs.

Flexible versions of the course are offered.

A course certificate will be issued upon completion of the course.

For Ph.D. students, the course will count as a 3 ECTS Ph.D. seminar.

### Flexible Course

You can sign up as an individual or as a group (eight participants maximum). We have limited capacity (2-3 courses per semester), and will arrange courses as our capacity allows, and on a first-come, first-served basis.

If you are part of a research group or project with similar work areas, signing up as a group will allow you to share experiences, make choices from common interests, and to collaborate on task. Depending on your needs, the basic course plan of five weeks may be compressed down to three weeks, or extended up to eight weeks.

If you are a researcher with a specific interest in only one or two of the topics (web archives, doing your own archiving, characteristics of archived web), an abbreviated course may be arranged. Course certificates are only issued for the full course.



### Conditions

The course is offered free of charge to researchers or Ph.D. students at Aarhus University, Aalborg University, University of Copenhagen, University of Southern Denmark, or The Royal Danish Library. When signing up for a course, you agree to do a course evaluation as your final task.

If you apply for a course as an individual, you may encounter other participants in the online work area for the course, with other or similar interests to your own. You may choose to work with them, or to work individually. If your project has a high confidentiality level, a closed individual course may be arranged, depending on our capacity.

Besides from researchers and Ph.D. students the course can also be made accessible as a resource for researchers who want to share it with their graduate students. We recommend that researchers start with taking the course themselves, and then discuss with us how to loan and best make use of the course.

Participants must have sufficient computer skills to download, install and use English language applications. NetLab will try to advise if problems are encountered, but we cannot offer individual support for third party applications.

The course materials and most of the software is in English. Communication during the course will be in Danish or English, depending on the participants.

### Sign up

To sign up for a course, please contact us at [info@netlab.dk](mailto:info@netlab.dk), and tell us who you are, where you are from, whether you are a group, and when you would like to get a course if possible. (The more flexible your wishes, the easier it will be for us to meet your wish).



NetLab, Høvelshøjsgade 14, DK-8200, Aarhus N, Denmark  
[www.netlab.dk](http://www.netlab.dk)

D I G H U M L A B

NetLab is a theme in DIGHUMLAB, with a special focus in web studies and digital research infrastructures. For more information, please visit [www.netlab.dk](http://www.netlab.dk), or [www.dighumlab.dk](http://www.dighumlab.dk).



After accessing the course the first time, you can bookmark it, so you don't have to access it from the portal every time. The direct link will lead you to a login screen first.

Direct link to the course page: <https://dighumlab.moodle.aau.dk/course/view.php?id=10>

This is the course page with "mapping" remarks. Functions without explanations should be ignored.

**Set menu language** (points to the language dropdown menu)

**Practical content:**  
 1) Important announcements if any.  
 2) Meta conference; your remarks, general questions, messages of being hindered etc. should go here.  
 3) Practical information documents (also sent by email).

**Practical information**  
 In this section you can find general information about the course.  
 Questions, suggestions, comments regarding the general course plan and design may be posted in the conference.  
 Questions or suggestions regarding the actual course design and content plan.  
 Course information  
 These are the documents with practical information on the course that you have also received by email.  
 Please read the whole course overview carefully and use it for planning your individual course participation.

**Course Conference**  
 In feedback, comments and general discussion about the assignments in the perspectives along in respect of your research interests with other course participants.  
 We aim at creating a continuous thread where it will be easy to make your thoughts to other users, and to get a feel for other participants.

**Course Materials**

**SEARCH FORUMS**  
 Advanced search

**UPCOMING ANNOUNCEMENTS**  
 Course closing soon after ending week 8 (weekend) 20th Sep 17:00 High Holiday  
 Welcome - please read this 28 Aug 17:00 High Holiday  
 Clear topics

**UPCOMING EVENTS**  
 There are no upcoming events.

**RECENT ACTIVITY**  
 Activity: Week 8 ending 19 September 2017, 8:22 PM  
 Full report of recent activity.

**COURSE UPDATES**  
 Updated: 1 week ago  
 Course information  
 Updated: 1 week ago  
 Updated: 1 week ago  
 Course Forum  
 Updated: 1 week ago  
 Updated: 1 week ago  
 Assignments 1-8

**This is a display for important messages if there are changes you should be aware of.** (points to the announcements section)

**This is a highlighting of upcoming events and deadlines from the course calendar.** (points to the upcoming events section)

**This is where you find the course materials.** (points to the course materials section)

**This is the course calendar with highlighted event dates. Hover cursor over date to see the event.** (points to the calendar)

**These are our main conferences for submitting and discussing assignments.** (points to the course forum and assignments)

**"Course Forum" is a supplementary meta conference, where you can ask for clarifications.** (points to the course forum)



## Task 4

Eveline Vlassenroot & Friedel Geeraert

### Table of Contents

<b>1 Data gathering strategy</b>	<b>2</b>
<b>2 Evaluation of results and challenges</b>	<b>2</b>
2.1 Manual software	2
2.1.1 Webrecorder	2
2.1.2 Print to .pdf	4
2.1.3 Digital footprints	5
2.1.4 Archive.is	8
2.1.5 Freezepage	8
2.1.6 Perma.cc	10
2.1.7 Video Download Helper	11
2.1.8 Firefox UnMHT-plugin	11
2.2 Automated software	14
2.2.1 Stillio	14
2.2.2 Pagefreezer	14
2.2.3 HTTrack website copier	15
<b>3 Conclusion</b>	<b>18</b>

Set menu  
language

A  
th  
D  
T



## Course in Web Archives and Web Archiving

### Course Evaluation

Please fill out, save and upload this form. We need your evaluation in order to maintain and offer the course, and it is also your final assignment to reflect upon what you learned, both in respect of what you expected, and in respect of the goals and interests you defined in assignment 1.

Where did you hear about the course?

Source	Mark or write field
NetLab website	<input checked="" type="checkbox"/>
National PhD Course page	
Aarhus University ARTS course page	
Other, please specify:	

Was the course workload as estimated?

Much less than estimated	A little less than estimated	As estimated	A little more than estimated	Much more than estimated
		<input checked="" type="checkbox"/>		
Elaborate if you like:				

Your overall satisfaction with the course – did we meet your expectations:

Very dissatisfied	Dissatisfied	Neutral	Satisfied	Very satisfied
				<input checked="" type="checkbox"/>
Elaborate if you like:				

NetLab, Helsingørsgade 14, DK-8200, Aarhus N, Denmark  
[www.netlab.dk](http://www.netlab.dk)

This course material is developed by NetLab may be re-used or distributed only by express permission.



Your satisfaction with course materials:

Very dissatisfied	Dissatisfied	Neutral	Satisfied	Very satisfied
				<input checked="" type="checkbox"/>
Elaborate if you like:				

Your satisfaction with your communication with the instructor:

Very dissatisfied	Dissatisfied	Neutral	Satisfied	Very satisfied
				<input checked="" type="checkbox"/>
Elaborate if you like:				

Your satisfaction with your personal effort during the course:

Very dissatisfied	Dissatisfied	Neutral	Satisfied	Very satisfied
		<input checked="" type="checkbox"/>		
Elaborate if you like: Due to high workload I could not afford as much effort on the course as I expected before.				

NetLab, Helsingørsgade 14, DK-8200, Aarhus N, Denmark  
[www.netlab.dk](http://www.netlab.dk)

This course material is developed by NetLab may be re-used or distributed only by express permission.

## IIPC Training Working Group

---

- Az International Internet Preservation Consortium 2017. végén alakult munkacsoportja
- Megbeszélések Zoom videokonferencia formájában
- Kérdőíves felmérés az oktatási és továbbképzési igényekről 2018. januárjában (224 válasz érkezett)
- Legnagyobb igény a webarchiválási technológiák, a szoftverek és a szabványok oktatására lenne
- A legnépszerűbb képzési formák a webináriumok, illetve a személyes jelenlétet igénylő tanfolyamok
- Meglévő tananyagok nyilvántartása





## Training Working Group

Training is a top need and opportunity for IIPC. Exchange of "how to" information is a core purpose of IIPC, and is one of the primary reasons new members join IIPC. Current members have varying approaches to archiving reflecting different institutional mandates, legal contexts, technical infrastructure, etc.), but share a need for various types of training: technical, curatorial, for practitioners and for researchers.

The Training Working Group, chartered in late 2017, is working to fulfill the vision of making IIPC the world leader for training on web archiving to its members, web archivists and technologists engaged in web archiving.

- The IIPC training curriculum will be crowdsourced by its members, reflecting real life needs from practitioners, and serving as a form of member engagement and contribution.
- The IIPC training curriculum will be available online under an open (CC) license for asynchronous delivery; it will also be available as a kit for delivery in hands on workshops by "certified" (literally, or figuratively) web archiving instructors for distributed delivery.
- The curriculum will be continuously updated based on feedback and assessment of the effectiveness of the online and in person events and monitoring of new developments in web archiving.
- "Train the trainer" sessions will be delivered at IIPC events, and qualified trainers may then offer in person workshops at related events, such as the Society of American Archivists, IFLA, IPRES, etc.

### > Events

-  [Web Archiving Conference & General Assembly 2018](#)
-  [Web Archiving Week 2017](#)

## IIPC Training Survey

THU, 14 DECEMBER 2017 – IIPC PROGRAMME AND COMMUNICATIONS OFFICER

Recognizing the global need for practical training in web archiving, the IIPC chartered a new **working group dedicated to training** on October 31, 2017. While vital to preserving our common cultural heritage, web archiving remains a niche area, requiring specialized skills and knowledge to practice effectively.

The goal of this working group is to produce a quality curriculum – that can be delivered online or in person – to train the current and next generation of web archiving practitioners. By giving people the hands on learning they need to preserve the Web, we will empower IIPC members and the field at large to capture more and better archives, and help elevate web archiving worldwide.

One of the first actions for our Training Working Group is to survey the web archiving world to assess the current level of training needs. How do web archivists currently get training, and how good is it? What gaps are there, and where should we prioritize our efforts?

Netpreserve.org

IIPC YouTube Channel

@NetPreserve

Search ...

### Recent Posts

[Announcing the IIPC Technical Speaker Series](#)

[How Can We Use Web Archive?: A Brief Overview of WARP and How It Is Used](#)

[The Study of punk culture through the Portuguese](#)

**Inventory of Current WA Training Resources**  
(Please mark resources that cannot be shared externally)

1. [BoF Workshop 2012 \(How to fit in? Integrate a web archiving program in your organization\)](#)
2. [BI Workshop 2015 \(Heritrix, OpenWayback and Solr\)](#)
3. [Using Web Archives in Research - An Introduction](#), Janne Nielsen, NetLab.dk
4. [Researcher workshop](#), Emily Maemura, Christoph Becker, Ian Milligan, WAC&RESAW 2017
5. ['Awesome Web Archiving'](#) list
6. [NYARC's web archiving documentation](#)
7. [Getting Started in Web Archiving](#), Abbie Grotke, IFLA 2017
8. ['Getting Started with Web Archiving'](#), by Karl Blumenthal + Sumitra Duncan
9. [Capture All the URLs: First Steps in Web Archiving](#), Kristen Yarmey
10. [Digital Preservation at Oxford and Cambridge](#), DPC (you have to be a member)
11. [Investiga XXI](#): the Investiga XXI group promotes the usage Arquivo.pt as a tool and resource for scientific research.
12. ["Web Archiving Fundamentals"](#), Anna Perricci, fee-based Society of American Archivists webinar - Open [Zotero library](#) associated with this course (not all about training per se and some duplication with others on this list)
13. [Quality Assurance guide for Stanford Libraries web archiving](#), Nicholas Taylor
14. [Research Resources tutorials and examples](#), Nicholas Taylor
15. [Training courses provided by Arquivo.pt in the areas of web preservation and research](#)
16. [The Archives Unleashed Toolkit](#), AU Team (walkthrough to analyze W/ARC content)
17. [Archives Research Services Workshop](#), - Internet Archive, a self-guided workshop to provide researchers, developers, and general users an introduction to data mining and computational tools and methods for working with web archives.
18. [Educational Partnerships \(IA pulled together coursework from universities re web archiving\)](#) - various (I'm most familiar with Bert Lyon's class at the bottom of the page - AG)
19. [Library of Congress staff training](#) - some newer videos and a PDF of the high-level view of our confluence training guide for internal staff use, and mostly about our tool Digiboard, but might



## Review of Existing Training Materials - Claire N.xlsx

Fajl Szerkesztés Nézet Beállítás Formázás Adatok Eközök Bővítmények Súgó

MEGOSZTÁS

fx Up to date. Very comprehensive outline of web archiving practices, strategies and pros and cons. Great for a complete beginner. I'd give this to a new member of my team to read regardless of job role, grade etc. Case study on Danish Web Archive. Advice about how to create a 'micro web archive' using tools such as HTTrack and also do demand capture through Archive-IT etc.

	B	C	D	E	F	
	Title	Creator	Date	Audience	Format	Comments
2	How to fit in? Integrate a web IIPC (input from BNF, BL, IMR F		2012	Organisations new to web ar	Slides, Evaluation docu	Very high level overview show
3	IIPC Technical Training Workshop IIPC		2015	Technical	Videos	Videos of technical training ses
4	Living Web Archives in Resean Netlab (Jane Neilson)		2016	Researchers/Web Archive Us	Book (pdf)	Up to date. Very comprehensiv
5	WAC and RESAW User Workshop Various (for IIPC & Resaw)		2017	Researchers/Educators	Workshop	Overview of workshops only. D
6	Awesome Web Archiving Various (for IIPC)		Ongoing (last updat	Technical	GitHub page	Useful and up to date list of ve
7	NYARC Web Archiving Docs NYARC (Sumitra Durcan)		Ongoing	Web Archiving Team Membe	Website	Useful overview of all aspects
8	Getting Started in Web Archiv Library of Congress (Abbie Gro		2017	Organisations new to web ar	Paper	Up to date overview of things I
9	Getting Started in Web Archiv NYARC (Sumitra Durcan & Karl		2017	Organisations new to web ar	Workshop	Overview of workshops only. D
10	Capture All the URES: First Step University of Stirling (Kristen		2015	Organisations new to web ar	Slides	Case study of how a University
11	Digital Preservation at Oxford Bodleian Libraries, Cambridge		Ongoing	Digital Preservation Team M	Documents	Digital preservation focused to
12	Investiga XXI Portuguese Web Archive		Ongoing	Researchers/Web Archive Us	Slides, videos	Case studies only at the mom
13	Web Archiving Fundamentals Society of American Archivists		Ongoing (live cours	Web Archiving Team Membe	Videos, personal tutori	I can't view this because it is p
14	Johns Library for Web Archiv Various		Ongoing (live cours	Web Archiving Team Membe	Web pages, documents	Links to a variety of web archiv
16	Quality Assurance guide for St Stanford University (Nicholas T		Ongoing (last updat	Web Archiving Team Membe	Document	Good example of a QA strategy
18	Research Resources tutorials - Stanford University (Nicholas T		Ongoing	Researchers/Web Archive Us	Web page	Links to many tutorial resource
17	Training courses provided by Portuguese Web Archive		Ongoing	Researchers/Web Archive Us	Face to face training	Three training courses are curr
18	The Archives Unleashed Toolkit The Archives Unleashed Project		Ongoing	Researchers/Web Archive Us	Web page	Step by step technical instructi
19	Archives Research Services Wikimedia Archive		2015	Researchers/Web Archive Us	Web page	Originally created as an in pers
20	Educational Partnerships (IA p: Internet Archive		Ongoing (last updat	Students	Documents	Although aimed at university st
21	Library of Congress staff traini Library of Congress		Various	Curators, Web Archive Team	Documents, Videos	The older documents include a
22	Archive-IT Help Center Internet Archive		Ongoing	Web Archiving Team Membe	Web pages	Good clear guidance about sco
27	Archiv-it Video Trainings Internet Archive		Ongoing	Web Archiving team Membe	Videos	I haven't watched the videos, y
24	Metropolitan New York Librar MEERIC Web Archiving SIG		2016	Web Archiving Team Membe	Website	I haven't watched the videos, y
25	Web Curator Tool National Library of New Zealand		Ongoing (last updat	Web Archiving Team Membe	Web pages	Very tool specific. Contains a u
26	Heritrix Training Session Internet Archive (Gordon Moh		2011	Technical	Video	As a complete non-technical p
27	Introductions to Web Archivi British Library		2015	Researchers/Web Archive Us	Video	Three videos providing a basic

confluence training guide for internal staff use; and mostly about our tool Digiboard, but might

UDC Training M/C Inventory of Current LMA Training Resources

### Review of Existing Training Materials - Claire N.xlsx

Fájl Szerkesztés Nézet Beszúrás Formázás Adatok Eszközök Bővítmények Súgó A Drive mentet

100% - \$ % 0 00 123 - Calibri - 11 - B I A

Up to date. Very comprehensive outline of web archiving practices, strategies and pros and cons. Great for a complete beginner, I'd give this to a new member of my team to read regardless of job role, grade etc. Case study on Danish Web Archive. Advice about how to create a 'micro-web archive' using tools such as HTTrack and also on-demand capture through Archive IT etc.

Zoom

Marton Nemeth 33133791777

hbyrne

Participants (18)

Find a participant

- MN Marton Nemeth (ME)
- A abbiagrotta
- TC Tom Cramer
- 442074127000
- ES Ed Summers
- K Karol
- 33133791777
- 44209089213
- AP Anna Ferrici
- EW Eric Webharvesting
- H hbyrne
- LD Lori Donovan

Tom Cramer

28

confluence training guide for internal staff use, and mostly about our tool Digiboard, but might



## Getting Started in Web Archiving

### Tools

GROTKE, Abigail (2017) *Getting Started in Web Archiving*. Paper presented at: **IFLA WLIC 2017 - Wrocław, Poland - Libraries, Solidarity, Society**. in Session 186 - National Libraries with Information Technology.

Bookmark or cite this item: <http://library.ifla.org/id/eprint/1637>



PDF (413kB)

Language: **English (Original)**

Available under licence [Creative Commons Attribution](#).

Bookmark or cite this item: <http://library.ifla.org/1/186-grotke-en.pdf>

### Abstract

English

#### Getting Started in Web Archiving

This purpose of this paper is to provide general information about how organizations can get started in web archiving, for both those who are developing new web archiving programs and for libraries that are just beginning to explore the possibilities. The paper includes an overview of considerations when establishing a web archiving program, including typical approaches that national libraries take when preserving the web. These include: collection development, legal issues, tools and approaches, staffing, and whether to do work in-house or outsource some or most of the work. The paper will introduce the International Internet Preservation Consortium and the benefits of collaboration when building web archives.



#### Articles in this section

[Frequently Asked Questions](#)[Archive-It Crawling Technology](#)[Archive-It workflow](#)[Web Archiving Lifecycle White Paper](#)[Live Archive-It Training Sessions and Webinars](#)[Archive-It Trial Basics](#)[Using the Help Center](#)[Glossary of](#)

## Live Archive-It Training Sessions and Webinars

[Maria Praetzelis](#)

Updated 3 days ago

[Follow](#)

We offer several different training opportunities to our partners. If you are a new partner, we ask and recommend that any members of your team who plan to use our web application attend both of the online **Archive-It Informational Webinar** and **Archive-It Application Training**. More advanced training on specialized topics is made available regularly as well. Details for each may be found below. All training is performed live and online via Zoom. Partners will receive login and call-in information through email shortly before their selected session.

- [Archive-It Informational Webinar](#)
- [Archive-It Application Training](#)
- [Advanced Training Opportunities](#)

### Archive-It Informational Webinar: Archiving and Preserving the Web

Participants in this training session get an overview of Archive-It and the Internet Archive and see a demonstration of our service. These webinars are held twice a



# Metropolitan New York Library Council

29 feliratkozó

FELIRATKOZÁS 29

KEZDŐLAP

VIDEÓK

LEJÁTSZÁSI LISTÁK

CSATORNÁK

NÉVJEGY



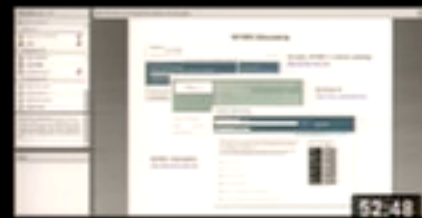
Feltöltések

ÖSSZES LEJÁTSZÁSA



## So You Want to Build A Web Archive

68 megtekintés · 1 éve



## Web Archiving Description and Access

49 megtekintés · 1 éve



## Web Archiving Building Collections

34 megtekintés · 1 éve



## Web Archiving: An Overview

117 megtekintés · 1 éve

A  
T  
P  
P  
T  
A  
13

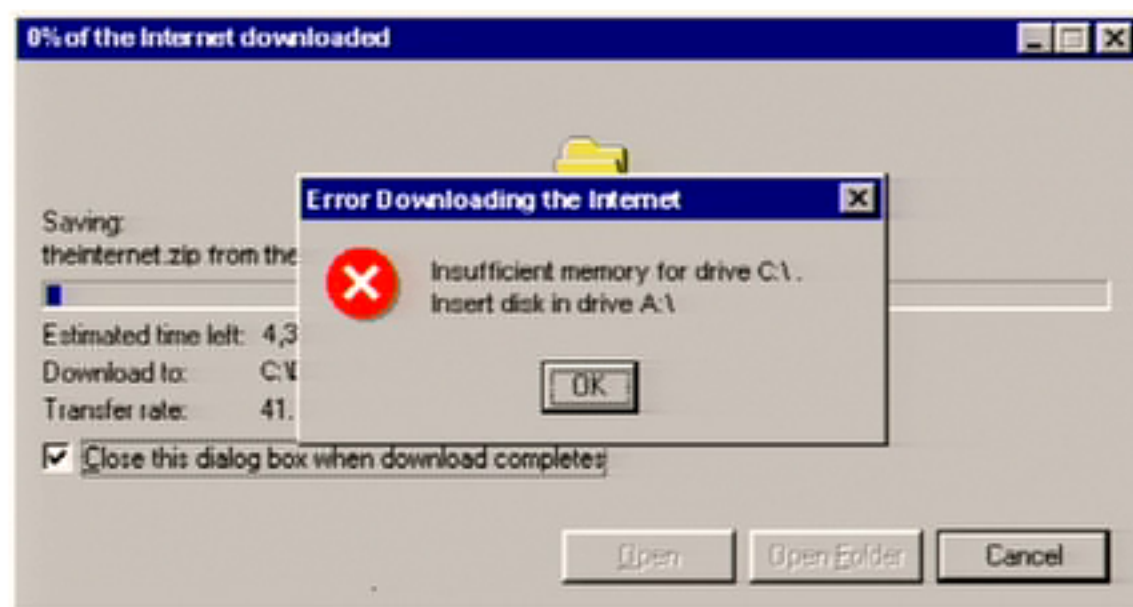
## Ajánlott források

---

- ❑ OSZK webaratás pilot projekt:  
<http://mekosztaly.oszk.hu/mia>
- ❑ NetLab Online Course in Web Archives and Web Archiving:  
<http://www.netlab.dk/services/courses/>
- ❑ Janne Nielsen: Using the web archives in research:  
[http://www.netlab.dk/wp-content/uploads/2016/10/Nielsen\\_Using\\_Web\\_Archives\\_in\\_Research.pdf](http://www.netlab.dk/wp-content/uploads/2016/10/Nielsen_Using_Web_Archives_in_Research.pdf)
- ❑ IIPC Training Working Group:  
<http://netpreserve.org/about-us/working-groups/training-working-group/>

# Köszönjük a figyelmet!

Kapcsolati e-mail cím: mia@mek.oszk.hu



Házi feladat: Töltsük le az Internetet, amíg még nem késő... ;-)

## 404 Not Found - Ki őrzi meg az internetet? (8)

404 Not Found - Ki őrzi meg az internetet? (8) **2017. október 13-án** szombati napot szentelt az **Országos Széchényi Könyvtár** a **Videotorium** online tartalmaként a **Központi Feladatnő: A weboldalak és a digitális kultúra** témakörében. A **videó** a **404 Not Found - Ki őrzi meg az internetet?** című előadásról készült felvétel.



Lendvai Miklós

5 perc - 17:25



Kacs-Teszelszky

5 perc - 17:25



Peternák Miklós

5 perc - 17:25



Dr. Prof. Kámpor György

5 perc - 17:25



Drótos László

5 perc - 17:25



Hozzászólások a workflop-on vitarendezőkhöz

5 perc - 17:25

