





Institut für Geschichte

Data-Driven Approaches to Studying the History of Museums on the Web: Challenges and Opportunities for New Discoveries

Dr. Nadezhda Povroznik

Wissenschaftliche Mitarbeiterin FG - Humanities Data Science and Methodology Institut für Geschichte Technische Universität Darmstadt



DigiHistCH24 conference theme:

Historical Research, Digital Literacy and Algorithmic Criticism



DigiHistCH24 conference theme:

Historical Research, Digital Literacy and Algorithmic Criticism



Digital History of Virtual Museums https://virtualmuseumhistory.com/

VM History

MAIN ABOUT THE PROJECT HISTORY VIRTUAL MUSEUMS INFOGRAPHICS USEFUL RESOURCES Q

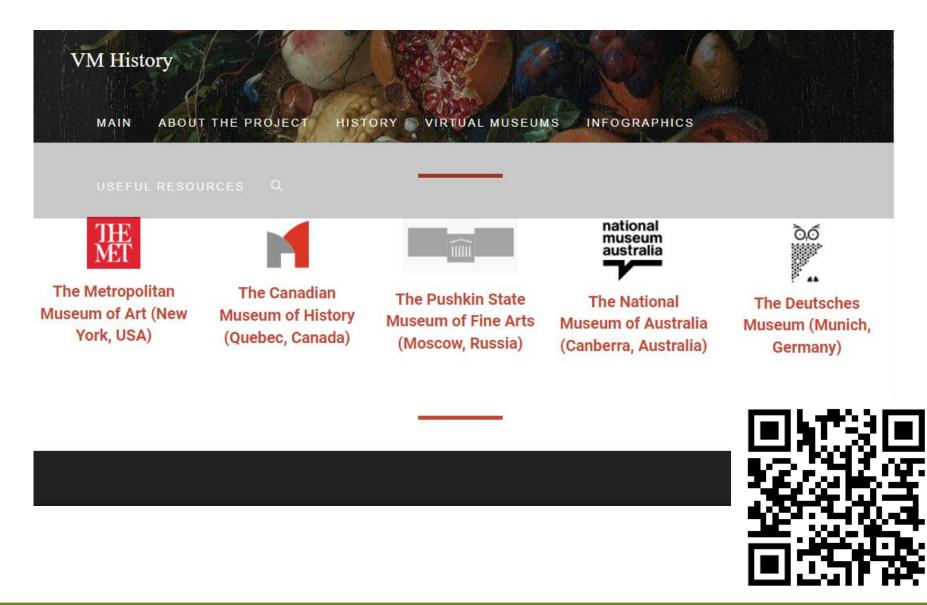
DIGITAL HISTORY OF VIRTUAL MUSEUMS

"Day by day, what you choose, what you think and what you do is who you become."

HERACLITUS



Digital History of Virtual Museums: Case studies https://virtualmuseumhistory.com/virtual-museums/

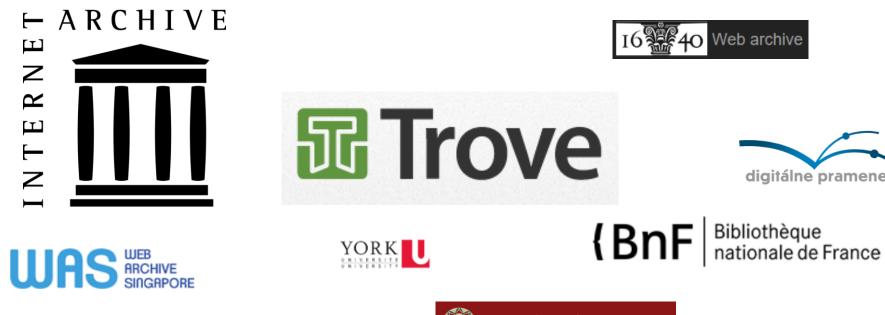




The image is generated by ChatGPT 40.

The prompt: illustrate data-driven approach to the history of the museums on the web

Digital History 2024 Switzerland 12-13 September University of Basel











Bibliothèque nationale du Luxembourg Nationalbibliothéik









Wayback Machine Growth

Wayback Machine	Pages Archived		
by Year	_		
2004	30,000,000,000 ((0-100B : Light blue)	
2005	40,000,000,000		
2008	85,000,000,000		
2012	150,000,000,000	(100B-450B : Yellow)	
2013	373,000,000,000		
2014	400,000,000,000		
2015	452,000,000,000	(450B-600B : Orange)	
2016	459,000,000,000		
2017	279,000,000,000		
2018	310,000,000,000		
2019	345,000,000,000		
2020	405,000,000,000		
2021	514,000,000,000		
2022	640,000,000,000	(600B-:Red)	
2024	866,000,000,000		

Wayback Machine. Wikipedia,

https://en.wikipedia.org/wiki/Wayback_Machine

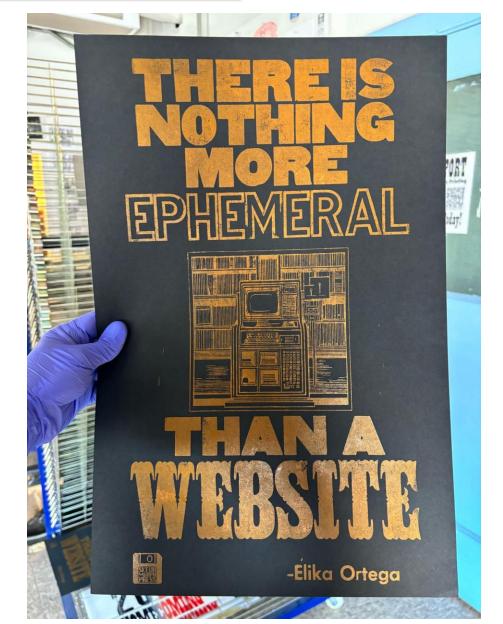
Big-Data Issues

- huge in volume, consisting of terabytes or petabytes of data;
- **high in velocity**, being created in or near real-time;
- diverse in variety, being structured and unstructured in nature;
- exhaustive in scope, striving to capture entire populations or systems (n = all);
- fine-grained in resolution and uniquely indexical in identification;
- relational in nature, containing common fields that enable the conjoining of different data sets;
- **flexible**, holding the traits of extensionality (can add new fields easily) and **scaleability** (can expand in size rapidly).

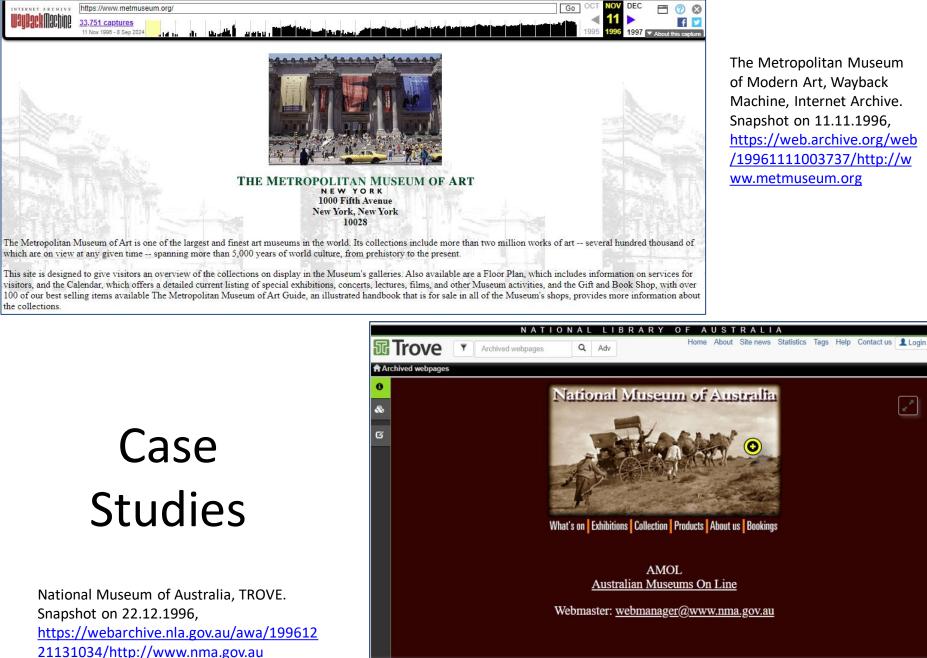
Kitchin, Rob. 2014. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data & Society* 1 (1). <u>https://doi.org/10.1177/2053951714528481</u>

What do we study?





Skeuomorph Press & BookLab at UIUC https://bsky.app/profile/ryancordell.bsky.social/post/3l3esvn ozbh26



This webpage snapshot was captured on the 22 Dec 1996 00:10

2.00

URL: http://www.nma.gov.au/

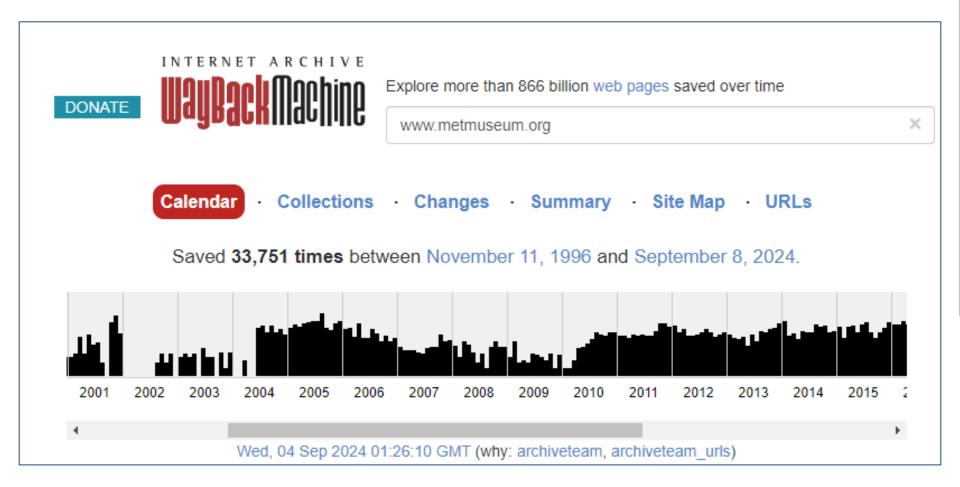
Digital History 2024

Switzerland

12-13 September

University of Base

Obtaining Data: The MET



metmuseum.org, Calendar, Wayback Machine, Screenshot on 08.09.2024, <u>https://web.archive.org/web/2024000000000*/metmuseum.org</u>

Obtaining Data: The MET



Artists (252) Interviews (244) Art History (241) Publication (238) Exhibitions (47) Publications (24) History (22) Programs (17) Archives (16) Libraries (16) Blogs (14) Fashion (8) Education (7) Politics and culture (4)

Subject

Art (348)



Details

Date archived

22 Dec 1996 00:10:34

http://www.nma.gov.au/

Related information View catalogue record in Trove for Organised crime in Australia

View catalogue record in Trove for 2020 Endeavour anniversary activities -

More snapshots...

National Museum of Australia

A Home / Websites

Title Front page

URL

A



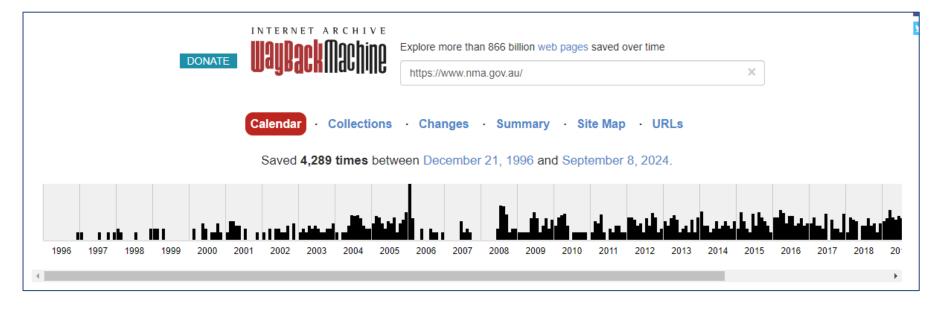


ABOUT	HELP	NEWS
-------	------	------

National Museum of Australia, TROVE. Snapshot on 22.12.1996, https://webarchive.nla.gov.au/awa/199612 21131034/http://www.nma.gov.au

NMA

Nma.gov.au, Calendar, Wayback Machine, Screenshot on 08.09.2024, https://web.archive.org/web/202400000000 0*/Nma.gov.au



National Museum of Australia

What's on Exhibitions Collection Products About us Bookings

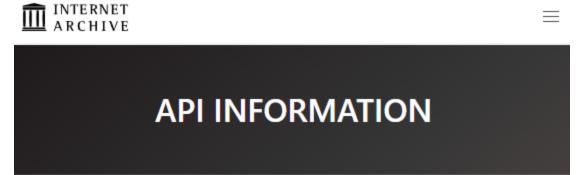
AMOL

Australian Museums On Line

Webmaster: webmanager@www.nma.gov.au

Digital History 2024

How do we obtain data?



Home / Archive.org / Technical Information / API information

Extensive and detailed information about Internet Archive APIs can be found at archive.org/services/docs/api/index.html



Digital History 2024 Switzerland 12-13 September University of Basel

UayBackMachine

host www.metmuseum.org

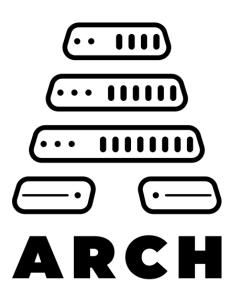
	Captures	Unique URLs
text/html	11,137,508	5,693,445
image/jpeg	4,431,051	305,778

Wayback Machine Downloader

https://github.com/hartator/wayback-machine-downloader

Downloading https://www.metmuseum.org/ to websites/www.metmuseum.org/ from Wayback Machine archives.
Getting snapshot pages
found 1372457 snaphots to consider.
file_list_curated: 1342067
1342067 files to download:
http://www.metmuseum.org:80/ -> websites/www.metmuseum.org/19961111003737/index.html (1/1342067)
http://www.metmuseum.org:80/ -> websites/www.metmuseum.org/19961219004113/index.html (2/1342067)
http://www.metmuseum.org:80/ -> websites/www.metmuseum.org/19970217203149/index.html (3/1342067)
http://www.metmuseum.org:80/ -> websites/www.metmuseum.org/19970416052424/index.html (4/1342067)
http://www.metmuseum.org:80/ -> websites/www.metmuseum.org/19970511181850/index.html (5/1342067)
http://www.metmuseum.org:80/ -> websites/www.metmuseum.org/19970619043010/index.html (6/1342067)
http://www.metmuseum.org:80/ -> websites/www.metmuseum.org/19971024144359/index.html (7/1342067)
http://www.metmuseum.org:80/ -> websites/www.metmuseum.org/19971210061150/index.html (8/1342067)
http://www.metmuseum.org:80/ -> websites/www.metmuseum.org/19980118222400/index.html (9/1342067)
http://www5.metmuseum.org:80/ -> websites/www.metmuseum.org/19981201090220/index.html (10/1342067)
http://metmuseum.org:80/ -> websites/www.metmuseum.org/19990117021812/index.html (11/1342067)
http://metmuseum.org:80/ -> websites/www.metmuseum.org/19990208005322/index.html (12/1342067)

Research Infrastructures



Archives Research Compute Hub

github.com/internetarchive/arch

🛠 GLAM Workbench

GLAM Workbench

Home	
About	>
Help	>
GLAM data	>
Trove	>
DigitalNZ	
Archives	>
Heritage	>
Libraries	>
Web archives	~
Notebooks	>
Data	>
Museums	>
Government	>

glam-workbench.net/web-archives/

Web archives

Notebooks

Timegates, Timemaps, and Mementos

Exploring the Internet Archive's CDX API

Comparing CDX APIs

Timemaps vs CDX APIs

Get a Memento

Find all versions of a web page

Harvest text from web pages

Harvest data about a domain

Explore Powerpoints from a domain

Explore all subdomains from gov.au

Compare versions of a page

Observe change in a page over time

Create full-page screenshots

Create a series of screenshots

Display changes in page text

Find when text appears

Data

GLAM Workbench

🛠 Web Archives

0

Q

GLAM Workbench

Timegates, Timemaps, and Mementos

web archives

%

API documentation

Works with AWA, IA, NLNZ, UKWA, & UKGWA

Systems supporting the Memento protocol provide machine-readable information about web archive captures, even if other APIs are not available. In this notebook we'll look at the way the Memento protocol is supported across five web archive repositories – the UK Web Archive, the National Library of Australia, the National Library of New Zealand, the UK Government Web Archive, and the Internet Archive.

Run live on Binder

GLAM Workbench, https://glam-workbench.net/

Digital History 2024

>

Switzerland 12-13 September University of Basel

What do we obtain? URL-analysis

https://web.archive.org/web,/20080515145825/http://www.metmuseum.org/explore/van_gogh/vangogh_alternate.html

https://web.archive.org/ web/ 20080515145825/ http://www.metmuseum.org/ explore/ van_gogh/ vangogh_alternate.html

What do we obtain? Subdomains issue

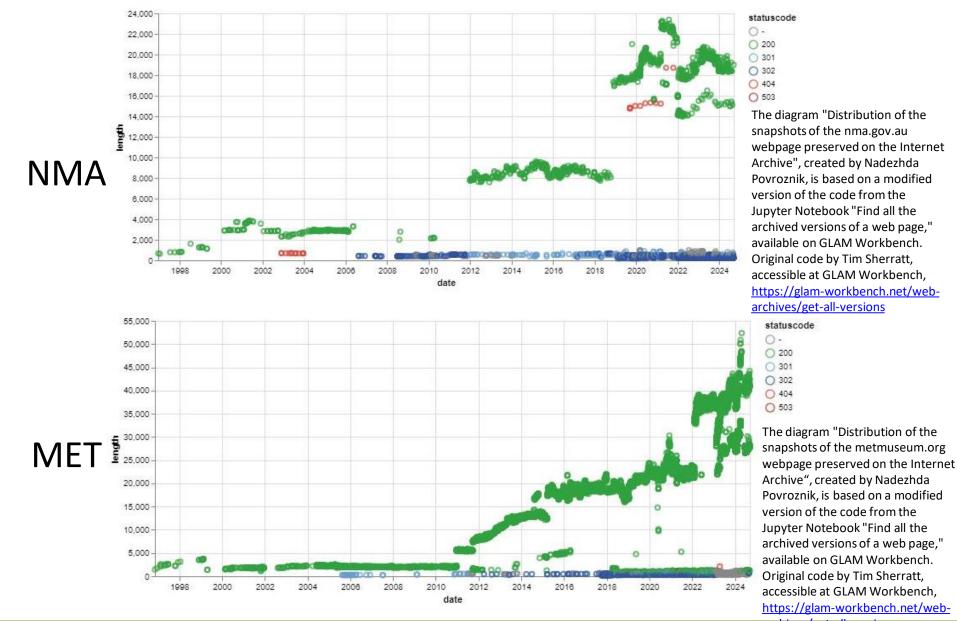
https://web.archive.org/ web/ 20080515145825/ http://www.metmuseum.org/ explore/ van_gogh/ vangogh_alternate.html



Title: The Artist Project | The Metropolitan Museum of Art URL: http://artistproject.metmuseum.org

The Metropolitan Museum of Art Web Archive, Archive-IT, <u>https://www.archive-it.org/collections/13135</u>

What do we obtain? A set of the webpages



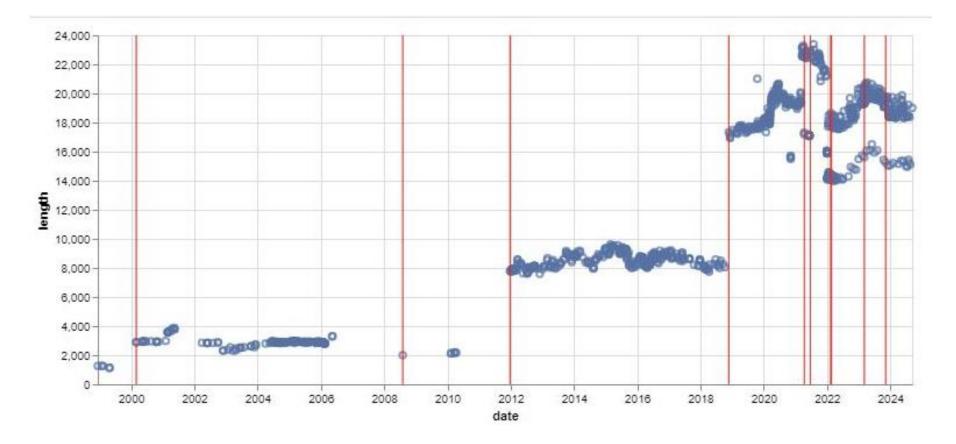
Digital History 2024

Switzerland 12-1

12-13 September University of Basel

archives/get-all-versions

What do we obtain? A set of the webpages



NMA: Periodisation

The diagram "Distribution of the snapshots of the nma.gov.au webpage preserved on the Internet Archive", created by Nadezhda Povroznik, is based on a modified version of the code from the Jupyter Notebook "Find all the archived versions of a web page," available on GLAM Workbench. Original code by Tim Sherratt, accessible at GLAM Workbench, <u>https://glam-workbench.net/web-archives/get-all-versions</u>

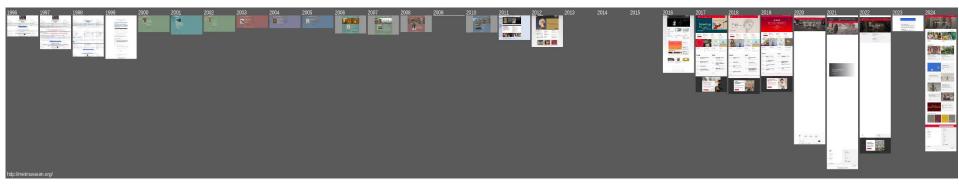


What do we obtain? A Complex of formats



The diagram "Screenshots 1996-2024 of the nma.gov.au webpage", created by Nadezhda Povroznik, is based on a modified version of the code from the Jupyter Notebook "Using screenshots to visualise change in a page over time," available on GLAM Workbench. Original code by Tim Sherratt, accessible at GLAM Workbench, <u>https://glam-workbench.net/web-archives/create-screenshots-overtime/</u>

NMA



The diagram "Screenshots 1996-2024 of the metmuseum.org webpage", created by Nadezhda Povroznik, is based on a modified version of the code from the Jupyter Notebook "Using screenshots to visualise change in a page over time," available on GLAM Workbench. Original code by Tim Sherratt, accessible at GLAM Workbench, <u>https://glam-workbench.net/web-archives/create-</u>screenshots-over-time/

MET

Digital History 2024 Switzerland 12-13 September University of Basel

What do we obtain? **Complex of formats**

19970411130920	19980613231552	19990427020716	20000226202431	20000507130209	20000619203655	20001207203000	20010707111546	20020620100529	20021227185340	20030423054343	20030620192416	20030823233712	20040627012301	20041023222134	20041216203209	20050901172546	20060221163448
19970411131008	19980613231558	19990427023816	20000302112114	20000507130647	20000805152906	20001207204000	20010707111810	20020620100552	20021227185958	20030423100357	20030620192754	20030823234640	20040627012522	20041024011152	20041216204319	20051001101739	20060221163502
19980131152021	19980613231605	19990428005554	20000304224728	20000507142111	20000806124302	20001207204400	20010707111821	20020620100610	20021227190340	20030423100445	20030620193534	20030823235247	20040627012809	20041024012715	20041216204836	20051023235759	20060221163718
19980131152211	19980613231612	19990428012040	20000305012307	20000507155347	20000806160303	20001207205400	20010707112331	20020620100635	20021227190438	20030423100736	20030620193804	20030919070908	20040627013100	20041024013325	20041216210241	20051025182247	20060221163732
19980131153842	19980613231619	19990428021051	20000305030650	20000507163641	20000806160304	20001207210300	20010707112339	20020620100921	20021227190443	20030423102141	20030620194219	20031003203826	20040627013200	20041024014126	20041216211116	20051101213333	20060221164101
19980131165438	19980613231626	19990428023823	20000305054516	20000507171121	20000806160306	20001207210900	20010707112409	20020620100944	20021227190644	20030424043824	20030621170304	20031004150907	20040627013426	20041024231438	20041216212437	20051107005846	20060221164121
19980131165917	19980613231632	19990428030610	20000305102541	20000507174845	20000815220319	20010201091000	20010707113102	20020620101133	20021227190852	20030427192739	20030622230115	20031007071814	20040627013442	20041025174044	20041216212721	20051220111551	20060221164350
19980131174107	19980613231639	19990508155152	20000305112804	20000507190957	20000824115256	20010210021858	20010707113418	20020620101351	20021227191050	20030427193022	20030718161510	20031010023753	20040627013712	20041027074230	20041216213710	20051227164331	20060221164408
19980613211403	19980613231646	19990922160649	20000305165737	20000507202741	20000827050833	20010210204443	20010818143149	20020620101449	20021227191319	20030427193031	20030802062319	20031018141413	20040627013834	20041117092558	20041216215546	20060105043056	20060221165541
19980613211410	19980613231653	19991010152549	20000305191557	20000509070304	20000827093034	20010210204457	20010818143233	20020620101551	20021227191359	20030427193313	20030819175745	20031018141506	20040627013949	20041124085406	20041216220630	20060113220638	20060221165548
19980613211417	19980613231700	19991010185306	20000305195817	20000509104322	20000828014957	20010210204510	20010827133339	20020620101641	20021227191507	20030427193325	20030819181034	20031018141754	20040627042043	20041204122444	20041216223816	20060202001948	20060221170207
19980613211423	19980613231707	19991010212842	20000306000804	20000509125519	20000829123309	20010210204959	20010827141812	20020620101805	20021227191515	20030427193411	20030819181231	20031018141852	20040627042358	20041204123655	20041224052018	20060206042836	20060221170254
19980613211511	19980613231715	19991010230505	20000306012614	20000509163355	20000831000752	20010210205012	20011108171803	20020620102056	20021227191714	20030427193534	20030819181418	20031018141900	20040627042402	20041204200429	20050129091057	20060207052021	20060323165242
19980613211519	19980613231722	19991011001849	20000306031203	20000510151158	20000902001541	20010210205159	20011118030046	20020620102154	20021227191716	20030427193557	20030822195053	20031018141932	20040627042422	20041204203100	20050204072644	20060207135212	20060421222604
19980613211526	19980613235059	19991011011817	20000306052210	20000510161106	20000905001150	20010210205200	20011118030211	20020620102217	20021227191749	20030427193615	20030822195345	20031018142003	20040627042540	20041204203855	20050204173322	20060221151700	20060503035844
19980613211532	19980613235125	19991011013906	20000306123500	20000520032055	20000905031838	20010210205340	20011124044944	20020620102245	20021227192300	20030427193631	20030822195415	20031018142005	20040627044230	20041204205303	20050207134158	20060221151710	20060516093214
19980613211539	19980613235333	19991011030737	20000306141820	20000527121114	20000905094826	20010214044819	20011125223324	20020620102657	20021227192317	20030427193906	20030822195820	20031018142036	20040627044235	20041204205654	20050209201904	20060221151716	20060516093756
19980613211546	19980613235339	19991011064813	20000306191321	20000527124055	20000905200152	20010216074457	20011125224330	20020621063845	20021227192349	20030427194038	20030822195927	20031018142617	20040627044501	20041204211108	20050212185355	20060221151722	20060516093810
19980613211559	19980613235346	19991011081114	20000309022045	20000601084924	20000905234923	20010224120509	20011125224909	20020621065304	20021227192457	20030427194153	20030822200405	20031018142934	20040627044519	20041204211920	20050212190326	20060221151727	20060516093829
19980613211606	19980613235434	19991011095802	20000309050343	20000605223321	20000914050321	20010224174545	20011125225220	20020621065450	20021227193237	20030427194216	20030822200413	20031019060445	20040627044906	20041204223938	20050212191247	20060221151734	20060516093908
19980613211612	19980613235442	19991011120706	20000309074044	20000605233307	20000928034849	20010301200742	20011125230515	20020621065814	20030118165553	20030427194324	20030822200414	20031028083518	20040627045008	20041204224712	20050212191657	20060221151740	20060516093938
19980613231154	19980613235449	19991011135252	20000309101042	20000606035718	20000928072903	20010308224210	20011125231132	20020726233922	20030118170533	20030427194345	20030822200605	20031118215030	20040627045231	20041204225556	20050212192026	20060221151750	20060516093945
19980613231202	19980613235455	19991011152123	20000309125054	20000606100313	20000928093411	20010402065336	20011125232007	20020802020126	20030118171148	20030427194621	20030822200749	20031129095205	20040627045245	20041204230405	20050213011921	20060221151808	20060516093952
19980613231208	19980613235502	19991011185836	20000309152613	20000606144626	20000928135820	20010417232843	20011125232452	20020804110057	20030121151831	20030427194640	20030822200822	20031204002128	20040627045749	20041204231757	20050213012101	20060221151821	20060516094007
19980613231216	19980613235509	19991011205941	20000309205044	20000606155523	20001001235649	20010418043712	20011125232458	20020816230629	20030130074959	20030427194651	20030822201456	20031212195613	20040627065406	20041204232350	20050213012138	20060221151826	20060516094017
19980613231223	19980613235516	19991011232952	20000309231625	20000606225414	20001001235708	20010418153701	20011125232940	20020816231349	20030201210652	20030427194719	20030822201530	20031220223441	20040627065613	20041205012341	20050213012144	20060221151845	20060516094027
19980613231230	19980613235524	19991013162725	20000416124732	20000607022911	20001001235727	20010618011424	20011125233514	20020816231934	20030210204915	20030427194929	20030822201809	20031230164111	20040627070728	20041205013039	20050213012208	20060221151852	20060516094034
19980613231237	19980613235531	19991013172704	20000416150643	20000607032708	20001012115234	20010624221633	20011125234431	20020816233913	20030218054049	20030427195003	20030822201810	20040205172033	20040627071733	20041216164617	20050213012218	20060221151912	20060516094111
19980613231244	19980613235538	19991013181642	20000416173342	20000607093422	20001012131106	20010629094640	20011125235015	20020816234022	20030219131945	20030427195028	20030822202019	20040205172551	20040627072155	20041216165448	20050213012358	20060221151915	20060516094147
19980613231251	19980613235545	19991013184821	20000428030618	20000607120853	20001014194912	20010629094656	20011217072653	20020926094119	20030219132623	20030427195143	20030822202240	20040205173715	20040627072340	20041216175340	20050213012414	20060221151925	20060526090507
19980613231258	19980613235552	19991013195750	20000428105641	20000607154744	20001018064620	20010629094814	20011218002304	20021002215734	20030302135304	20030427195215	20030822202743	20040206070101	20040627082847	20041216175814	20050213121255	20060221152006	20060610053256
19980613231305	19980613235601	19991014010349	20000430000259	20000607192738	20001120015800	20010629095804	20011221012359	20021008164349	20030302135528	20030427195219	20030822202758	20040206175518	20040627083617	20041216181933	20050213122143	20060221152016	20060613045304
19980613231311	19980613235608	19991014014022	20000504183438	20000608001931	20001120030500	20010629101034	20020112202237	20021017212519	20030302135808	20030427195244	20030822203132	20040210213446	20040627084410	20041216182248	20050213122551	20060221152021	20060615092455
19980613231317	19980614000514	19991103001843	20000504183806	20000608034400	20001120074500	20010629101124	20020124114143	20021101064900	20030302135904	20030427195256	20030822203406	20040220105349	20040627084743	20041216183028	20050213153654	20060221152046	20060617053546
19980613231323	19981202110219	19991103014207	20000504192148	20000608122147	20001120140500	20010629102145	20020209115725	20021101064911	20030302140010	20030427195329	20030822203558	20040220105533	20040627085450	20041216183644	20050216041042	20060221152058	20060617053600
19980613231331	19981203024854	19991103024307	20000505214946	20000608144056	20001206112300	20010629102146	20020213213928	20021101070532	20030302140038	20030427195332	20030822203730	20040220105941	20040627085703	20041216184536	20050303014130	20060221152107	20060617155612
19980613231337	19981203053628	19991106052406	20000505235709	20000608184017	20001206134400	20010629103323	20020403231550	20021101070824	20030302140130	20030427195441	20030822204115	20040220110151	20040627090229	20041216184809	20050308124823	20060221152119	20060617155907
19980613231345	19981206203208	19991109061323	20000506005530	20000609004838	20001206142100	20010629103439	20020421153416	20021101070942	20030302140530	20030427195458	20030822204134	20040220110336	20040627090306	20041216185427	20050308212706	20060221152125	20060617160415
19980613231352	19981212024332	19991109142648	20000506005545	20000609012506	20001206150200	20010629104721	20020421154024	20021101071011	20030302140644	20030427195701	20030822204335	20040220110739	20040627091251	20041216185523	20050310044803	20060221152134	20060618175627
19980613231358	19990208012642	19991109181357	20000506031549	20000609020318	20001206153800	20010629104907	20020421154525	20021101071934	20030302140659	20030427195930	20030822204340	20040220110834	20040627091307	20041216190318	20050320151336	20060221152143	20060618175923
19980613231405	19990219103542	19991110100915	20000506214118	20000609050525	20001206160800	20010629105350	20020421154712	20021101072034	20030302140707	20030428034521	20030822204357	20040220111034	20040627091724	20041216190337	20050405005359	20060221152218	20060621064636
19980613231411	19990219114346	19991110171328	20000506223218	20000609070725	20001206172400	20010629110018	20020421155010	20021101072237	20030302141334	20030428034556	20030822204645	20040329045945	20040627092128	20041216191554	20050421185300	20060221152226	20060621065000
19980613231418	19990219172736	19991111044430	20000506223657	20000613152532	20001206175900	20010629110613	20020421155032	20021101073627	20030302141609	20030428034957	20030822204832	20040405152735	20040627092500	20041216191717	20050507102247	20060221152232	20060621065004
19980613231424	19990220012331	19991111075423	20000507035600	20000613153331	20001206190700	20010629110922	20020421155512	20021101073910	20030302141715	20030526223408	20030822204844	20040405214737	20040627093049	20041216192911	20050517015451	20060221152240	20060621065045
19980613231431	19990220022235	20000121121252	20000507042216	20000613161329	20001206194100	20010629231909	20020421155717	20021101074219	20030302141740	20030608125233	20030822205118	20040411050841	20040627175835	20041216193459	20050706084609	20060221152245	20060621065128
19980613231438	19990220045337	20000124175547	20000507042512	20000613162742	20001206201500	20010706121847	20020421155815	20021123023422	20030302142109	20030608125639	20030822205604	20040413154404	20040627180201	20041216193518	20050725121625	20060221152251	20060621065139
19980613231444	19990220051821	20000129213303	20000507051355	20000613173459	20001206224500	20010707055916	20020421160312	20021125204304	20030302142153	20030608130140	20030822205705	20040430104646	20040627180657	20041216193741	20050725121631	20060221152258	20060621065154
19980613231451	19990220070325	20000201044222	20000507052155	20000613202109	20001206231900	20010707060108	20020421160551	20021202170352	20030302142759	20030608131256	20030822205706	20040515082802	20040627180659	20041216193909	20050725121634	20060221161047	20060621065218
19980613231459	19990220081353	20000226015041	20000507062221	20000613230911	20001206235500	20010707060527	20020527235444	20021202200356	20030302142957	20030608132851	20030822210024	20040519100842	20040803165407	20041216194423	20050725121637	20060221161214	20060621065306
19980613231505	19990220103246	20000226023256	20000507062255	20000614013833	20001207184200	20010707060857	20020601023739	20021219150222	20030302143249	20030611190225	20030822210145	20040603201212	20040907052528	20041216195414	20050725121641	20060221161400	20060621065410
19980613231512	19990302075402	20000226042615	20000507071113	20000614170250	20001207185000	20010707061354	20020612100902	20021219150336	20030302143445	20030618114849	20030822210303	20040609081412	20040911004956	20041216195737	20050725121644	20060221161412	20060621065422
19980613231519	19990422024100	20000226092237	20000507072037	20000614204019	20001207194200	20010707105914	20020613221932	20021227184537	20030320221212	20030618120229	20030822210306	20040620164103	20041016082714	20041216200052	20050725121648	20060221162308	20060621065434
19980613231526	19990422051415	20000226100302	20000507074324	20000615024209	20001207195300	20010707110049	20020616191021	20021227184653	20030325045202	20030620183130	20030822210359	20040621102706	20041018144123	20041216200324	20050725121650	20060221162314	20060621065504
19980613231532	19990422054306	20000226135406	20000507083954	20000615124644	20001207201000	20010707110613	20020620100017	20021227184656	20030403084851	20030620191159	20030822210628	20040622031331	20041018145257	20041216201044	20050812235759	20060221162320	20060621113719
19980613231538	19990422080100	20000226164135	20000507091703	20000615143352	20001207201600	20010707110647	20020620100154	20021227184724	20030414192138	20030620191523	20030822210653	20040622225142	20041018212045	20041216202404	20050830080829	20060221163220	20060621123838
19980613231545	19990422082309	20000226182053	20000507114806	20000615175045	20001207202400	20010707111312	20020620100245	20021227185315	20030419015404	20030620192049	20030822210953	20040627012201	20041022085112	20041216202502	20050831020943	20060221163238	20060621123900
									-								

Digital History 2024

Switzerland

12-13 September University of Basel

Building a Corpora

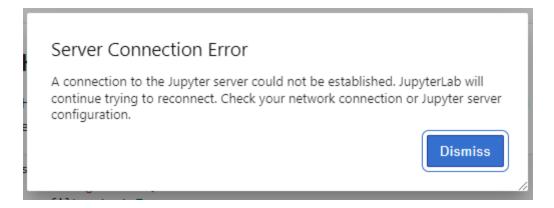
Harvesting a single url or list of urls

Get all human-visible text from all captures of a single url in the Australian Web Archive.

save_texts_from_url("ia", "http://nma.gov.au/", filter_text=False)

Captures: 0%

19/4299 [02:15<17:56:14, 15.09s/it]



The Metropolitan Museum of Art

Collection API

The Metropolitan Museum of Art Collection API, <u>https://metmuseum.github.io</u>

Museums' APIs: Object Images

OUR COLLECTION	- Collection API						
			API				

NMA, Collection API, <u>https://www.nma.gov.au/about/our-</u> collection/museum-api

Switzerland

Complexity

host www.metmuseum.org

Indexed on April 11, 2024. See also: · Parent Domain Summary: metmuseum.org · TLD Summary: org

Saved 33,798 times between November 11, 1996 and September 10, 2024.

MIME-types		Year Start	Year End
Any MIME-types		✔ 1996	2024
All text image	application	font audio video	1
Summary on M	IIME-types	Count	
Quick search on MIME-types.		< 1 2 > >>	Captures
	Captures	Unique URLs	
text/html	11,137,508	5,693,445	
image/jpeg	4,431,051	305,778	
text/css	869,592	50,560	
image/png	428,525	3,861	
image/svg+xml	424,359	492	
font/woff2	374,467	1,049	
image/gif	368,511	17,229	metmuseum.org, Summary, Wayback Machine,
application/json	225,383	16,729	Screenshot on 08.09.2024,
application/javascript	224,760	178,209	https://web.archive.org/web/20240000000000*/metmu
application/octet- stream	134,124	420	eum.org



DigiHistCH24 conference theme:

Historical Research, Digital Literacy and Algorithmic Criticism





DigiHistCH24 conference theme:

Historical Research, Digital Literacy and Algorithmic Criticism

- Valuable data
- New insights
- Source criticism
- Web preservation

- Complexity in obtaining data
- Lowering the entry barrier

- Completeness
- Reliability
- Biases
- Limits
- Challenges

Thank you!