



SEMINAR PEMINDAHAN REKOD NEGARA : MELESTARIKAN PERPADUAN DAN WARISAN BANGSA

MALAYSIAN GOVERNMENT WEB ARCHIVE (MyGWA)

16 JUN 2023

HOTEL TENERA, BANGI



Muhammad Idzwan Bin Ramli
Pegawai Teknologi Maklumat
Seksyen Pengurusan Rekod Elektronik
Arkib Negara Malaysia

ISI KANDUNGAN

1. **PENGENALAN**
2. **OBJEKTIF**
3. **PELAKSANAAN PENGARKIBAN LAMAN WEB**
4. **ISU DAN CABARAN**
5. **KESIMPULAN**

PENGENALAN

- Pengarkiban laman web dilaksanakan kerana kandungan laman web adalah rekod awam yang mempunyai nilai kebangsaan dan sejarah yang perlu dipindahkan ke Arkib Negara Malaysia sebagai warisan khazanah negara seperti termaktub di dalam **Akta Arkib Negara Malaysia 2003, Akta 629.**
- Sistem MyGWA yang dibangunkan ini adalah gabungan beberapa perisian sumber terbuka 'open source' yang lain seperti *web crawler* dan *wayback machine* yang kemudiannya membentuk satu sistem pengarkiban laman web. *Web crawler* berfungsi menyalin keseluruhan kandungan laman web berkenaan secara sistematik dan berterusan sehingga selesai.
- Laman web yang telah disalin akan disimpan di dalam aplikasi MyGWA menggunakan format standard yang digunapakai oleh kebanyakan sistem pengarkiban laman web luar negara iaitu format WARC. Format WARC yang digunakan adalah mengikut standard ISO 28500:2009 seperti yang dinyatakan oleh International Internet Preservation Consortium (IIPC).



OBJEKTIF

1

Menyimpan dan memelihara versi kandungan laman web asal yang diarkibkan selaras dengan Akta Arkib Negara 2003 (Akta 629).

2

Memastikan kandungan laman web yang diarkibkan dapat diakses semula dan dirujuk oleh orang awam.

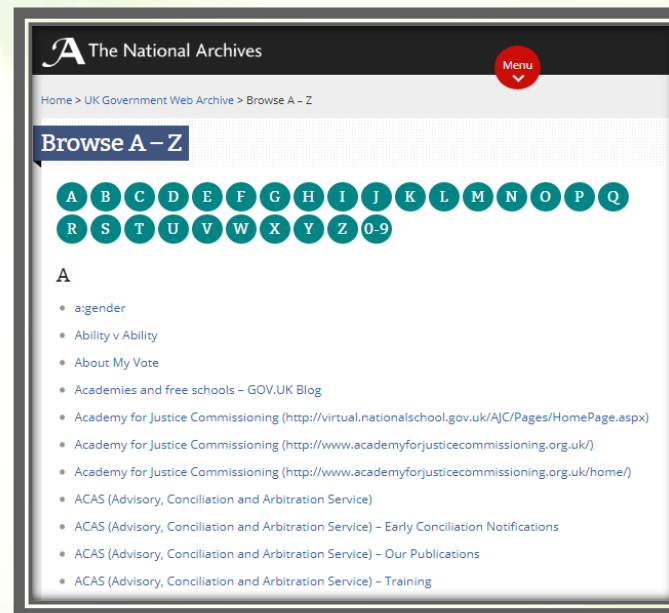
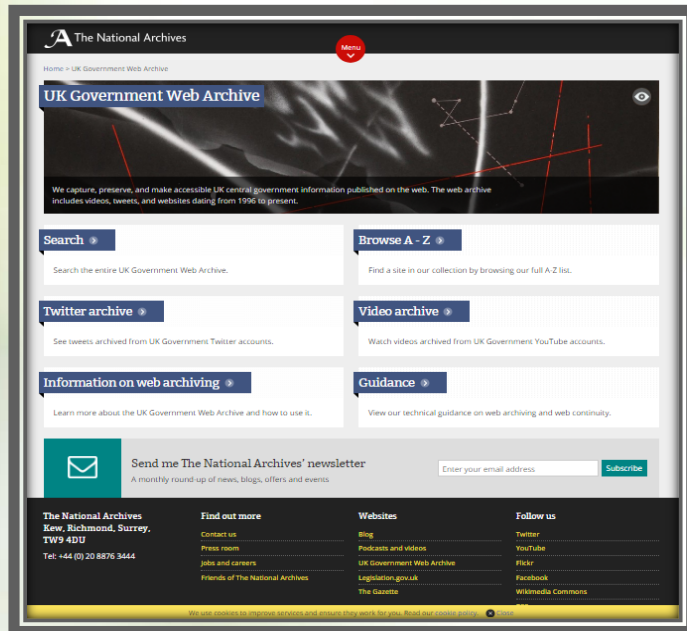
3

Disimpan sebagai koleksi atau dokumentasi yang boleh dijadikan warisan budaya dan dirujuk oleh generasi pada masa akan datang.

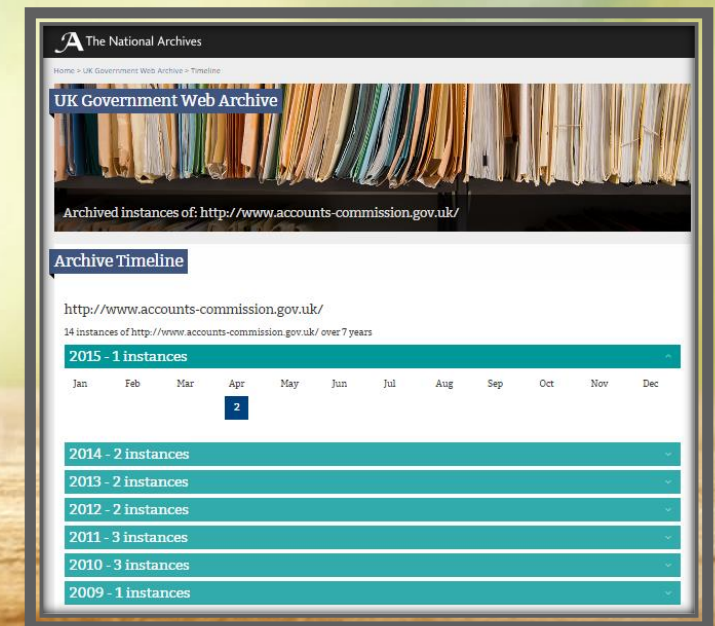


PENANDAARASAN

PENANDAARASAN PORTAL PENGARKIBAN LAMAN WEB



(UK WEB ARCHIVE)



PENANDAARASAN PORTAL PENGARKIBAN LAMAN WEB



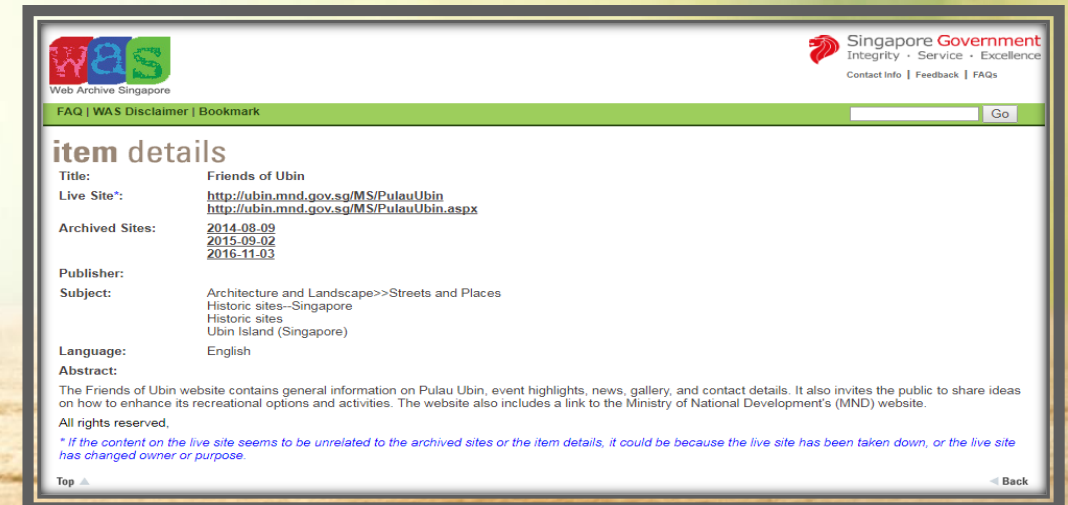
(AUSTRALIA WEB ARCHIVE)



PENANDAARASAN PORTAL PENGARKIBAN LAMAN WEB



(SINGAPORE WEB ARCHIVE)



Malaysian Government Web Archive (**MyGWA**)



Malaysian Government Web Archive (MyGWA)

Carian Terperinci

arkib negara

Q

Sumber: SEMUA

HASIL CARIAN

SENARAI HASIL CARIAN BERDASARKAN KATA KUNCI



Sumber

:

JABATAN ARKIB NEGARA MALAYSIA

URL

:

<http://www.arkib.gov.my>

Tajuk

:

PORTAL RASMI ARKIB NEGARA MALAYSIA

Subjek

:

SOSIAL

Subkategori

:

PERSEKUTUAN - JABATAN

Kategori

:

ARKIB AWAM

Jumlah Tika

:

13

Jumlah Capaian

:

356



Sumber

:

JABATAN ARKIB NEGARA MALAYSIA

URL

:

https://twitter.com/my_arkib_negara

Tajuk

:

LAMAN TWITTER - MY ARKIB NEGARA

Subjek

:

SOSIAL

Subkategori

:

PERSEKUTUAN - JABATAN

Kategori

:

ARKIB AWAM

Jumlah Tika

:

4

Jumlah Capaian

:

27

Senarai Subjek

Bilangan arkib yang telah dibahagikan di dalam kumpulan **SUBJEK** dan nombor di dalam kotak merah ini mewakili jumlah **LAMAN WEB**.

PENTADBIRAN	0
EKONOMI	0
KESELAMATAN	0
SOSIAL	25
TOTAL	25

Malaysian Government Web Archive (MyGWA)

Carian Terperinci

Sumber

kesihatan

Subjek

URL

Subkategori

Tajuk

Kategori

*Carian terhadap kepada data di dalam Bahasa Melayu

Cari

Q

Sumber: SEMUA

0

HASIL CARIAN

SENARAI HASIL CARIAN BERDASARKAN KATA KUNCI

Sumber

:

KESIHATAN, KEMENTERIAN

URL

:

http://www.moh.gov.my

Tajuk

:

PORTAL RASMI KEMENTERIAN KESIHATAN MALAYSIA

Subjek

:

SOSIAL

Subkategori

:

PERSEKUTUAN - JABATAN

Kategori

:

ARKIB AWAM

Jumlah Tika

:

19

Jumlah Capsian

:

9

Sumber

:

LEMBAGA PROMOSI KESIHATAN MALAYSIA

URL

:

https://mysihat.gov.my/

Tajuk

:

LAMAN RASMI LEMBAGA PROMOSI KESIHATAN MALAYSIA

Subjek

:

SOSIAL

Subkategori

:

PERSEKUTUAN - JABATAN

Kategori

:

ARKIB AWAM

Jumlah Tika

:

7

Jumlah Capsian

:

3

Sumber

:

KESIHATAN, KEMENTERIAN

URL

:

https://www.infoshat.gov.my

Tajuk

:

LAMAN RASMI INFOSHAT

Subjek

:

SOSIAL

Subkategori

:

PERSEKUTUAN - JABATAN

Kategori

:

ARKIB AWAM

Jumlah Tika

:

48

Jumlah Capsian

:

38

Sumber

:

KESIHATAN, KEMENTERIAN

URL

:

http://covid-19.moh.gov.my/

Tajuk

:

LAMAN WEB COVID-19 MALAYSIA

Subjek

:

SOSIAL

Subkategori

:

PERSEKUTUAN - JABATAN

Kategori

:

ARKIB AWAM

Jumlah Tika

:

58

Jumlah Capsian

:

36

Senarai Subjek

Bilangan arkib yang telah dibahagikan di dalam kumpulan

SUBJEK dan nombor di dalam kotak merah ini mewakili

Jumlah LAMAN WEB.

PENTABESIAN

7

EKONOMI

0

KESELAMATAN

0

SOSIAL

175

TOTAL

186

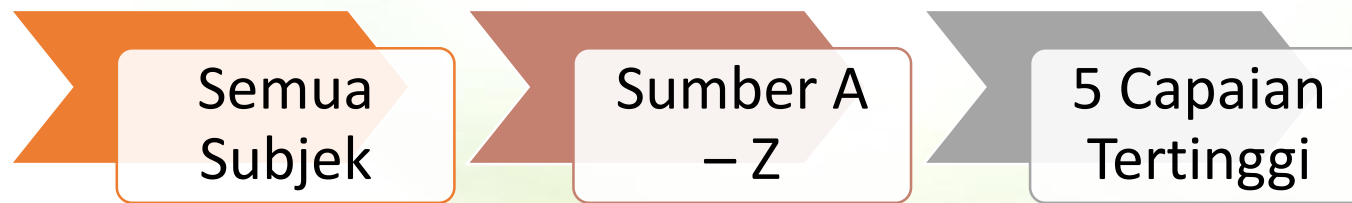
<<< Carian Terperinci >>>

Malaysian Government Web Archive (**MyGWA**)

Sumber: SEMUA	
Senarai Subjek	
Bilangan arkib yang telah dibahagikan di dalam kumpulan SUBJEK dan nombor di dalam kotak merah ini mewakili jumlah LAMAN WEB .	
PENTADBIRAN	175
EKONOMI	126
KESELAMATAN	86
SOSIAL	340
TOTAL	727

(Bilangan laman web yang dibahagikan mengikut 'Subjek')

Terdapat juga 3 jenis carian tambahan yang disediakan pada menu sistem MyGWA iaitu:



Malaysian Government Web Archive (**MyGWA**)

**MyGWA**
Malaysian Government
Web Archive
ARKIB: INTIPATI KEPADA MASA

LoginBahasa

Laman UtamaSenaraiBerkenaan MyGWAManual PenggunaSoalan LazimHubungi Kami

LANAM UTAMA ▸ SUMBER A - Z

Disember 12, 2017

SEMAK LAMAN WEB DIARKIBKAN DALAM SUMBER A KE Z

Pilih **Abjad** untuk melihat senarai

SEMUA A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

 – ARKIB NEGARA MALAYSIA, JABATAN (1)

 – BELIA DAN SUKAN MALAYSIA, KEMENTERIAN (4)

 – KEWANGAN MALAYSIA, KEMENTERIAN (1)

 – PFI ANCONGAN DAN KFRIDAYAAN MAI AYSIA KEMENTERIAN (1)

Dasar Privasi | Dasar Keselamatan | Terma dan Syarat |

Pihak Kerajaan dan Arkib Negara Malaysia tidak bertanggungjawab terhadap sebarang kehilangan atau kerosakan yang dialami kerana menggunakan maklumat dalam laman ini

Sesuai dipapar menggunakan versi terkini Chrome atau Mozilla Firefox dengan resolusi 1200 X 1000

Hakcipta Terpelihara 2017 © Arkib Negara Malaysia

Malaysian Government Web Archive (**MyGWA**)

**MyGWA**
Malaysian Government
Web Archive
ARKIB: INTIPATI KEPADA MASA

LoginBahasa

Laman UtamaSenaraiBerknaan MyGWAManual PenggunaSoalan LazimHubungi Kami

LAMAN UTAMA ▸ 5 CAPAIAN TERTINGGI

Jun 14, 2023

5 CAPAIAN TERTINGGI

-  — PELANCONGAN DAN KEBUDAYAAN MALAYSIA, KEMENTERIAN(24616)
-  — BELIA DAN SUKAN MALAYSIA, KEMENTERIAN(14444)
-  — KEWANGAN MALAYSIA, KEMENTERIAN (11960)
-  — JABATAN ARKIB NEGARA MALAYSIA(11343)
-  — PERDANA MENTERI, JABATAN (JPM)(6912)

Malaysian Government Web Archive (**MyGWA**)

PORTAL RASMI ARKIB NEGARA MALAYSIA

INFORMASI RINCI LAMAN WEB



Tajuk	:	PORTAL RASMI ARKIB NEGARA MALAYSIA
Sumber	:	JABATAN ARKIB NEGARA MALAYSIA
Kata Kunci	:	
Diskripsi	:	ARKIB NEGARA MALAYSIA MERUPAKAN AGENSI YANG BERPERANAN MENYIMPAN REKOD – REKOD YANG BERNILAI SEJARAH UNTUK RUJUKAN. PORTAL RASMI ANM MENYEDIAKAN PERKHIDMATAN SEPERTI KEMUDAHAN PENYELIDIK, KONSULTANSI PENGURUSAN REKOD KERAJAAN, PUSAT PENDOKUMENTASIAN, PEMELIHARAAN AUDIOVISUAL NEGARA, PERKHIDMATAN DALAM TALIAN, PEROLEHAN BAHAN ARKIB, PUSAT SUMBER DAN LAIN – LAIN. PORTAL RASMI ANM JUGA MENYEDIAKAN KEMUDAHAN LAIN SEPERTI BERITA DAN PERISTIWA, PENGURUSAN MAKLUMAT, GALERI, MUAT TURUN BORANG PELUPUSAN, PANDUAN PELUPUSAN REKOD, BORANG LATIHAN DAN KERTAS PEMBENTANGAN SEMINAR.
URL	:	http://www.arkib.gov.my
Kategori	:	ARKIB AWAM
Subkategori	:	PERSEKUTUAN - JABATAN
Subjek	:	SOSIAL
Jumlah Capaian	:	356
Pautan Lain	:	

+	2021	(3 Tika)
+	2020	(3 Tika)
+	2019	(1 Tika)
+	2018	(4 Tika)
-	2017	(2 Tika)

Tika

Bilangan arkib yang telah dibahagikan di dalam kumpulan **TAHUNAN** dan nombor di dalam kotak merah ini mewakili **TARIKH HARI** diarkibkan.

DIS	
NOV	
OKT	30
SEP	26

Malaysian Government Web Archive (**MyGWA**)

TETAPAN PENJADUALAN
(CRAWLING)

Metadata Schedule Harvest

Schedule (JABATAN AP

No	Frequency
1	SEKARANG
2	SEKARANG
3	SEKARANG
4	SEKARANG
5	SEKARANG
6	SEKARANG
7	SEKARANG
8	SEKARANG

Form Crawl

Frequency *

SEKARANG

Start Date *

22-09-2021

08:00 AM

Max *

10 HARI

Attempt *

10 PERCUBAAN

Inside Hop *

10 HOP/KEDALAMAN PAUTAN

Engine *

WPULL

Inside Exclusion

Save

Outside Domain

19-06-2019 08-00-AM

APPROVED

COMPLETED

04-09-2019 04-00-PM

APPROVED



LAMAN WEB YANG DIARKIBKAN

CONTOH – JABATAN ARKIB NEGARA



CONTOH – JABATAN LATIHAN KHIDMAT NEGARA



CONTOH – SPAD



CONTOH – NADMA

https://mygwa.arkib.gov.my/wayback/20200417040217/http://www.nadma.gov.my/

PORTAL RASMI
AGENCI PENGURUSAN BENCANA NEGARA
JABATAN PERDANA MENTERI

 Mon - Fri
08:00-17:00 +603 8870 4800 +603 8870 4848 admin@nadma.gov.my

PORTAL BENCANA UTAMA KORPORAT **INFO** HUBUNGI KAMI E-PERKHIDMATAN Search ...

BANTUAN KHAS COVID-19

Garis Panduan Pemberian Bantuan Khas COVID-19

Syarat-Syarat Kelayakan

1. Warganegara Malaysia yang bekerja tetapi telah kehilangan sumber pendapatan atau tidak dibayar gaji sepanjang tempoh dikenakan Perintah Pengawasan Dan Pemerhatian Bagi Kontak Covid-19 Di Bawah Seksyen 15(1) Akta Pencegahan Dan Pengawalan Penyakit Berjangkit 1988 (Akta 342).
2. Warganegara Malaysia yang bekerja tetapi telah kehilangan sumber pendapatan atau tidak dibayar gaji sepanjang tempoh menjalani rawatan di wad hospital

CONTOH – INFOSIHAT

← → ↺ https://mygwa.arkib.gov.my/wayback/20210819024047/https://www.infosihat.gov.my/ 🔑 🔍 ☆ ⚙️ ⌛ Paused

Utama KCHS Projek Khas Multimedia Direktori Pautan Awam Warga Log masuk 🔍

Kita Mesti Menang

INGAT

Kita masih perlu patuhi SOP COVID-19 walaupun telah mendapat suntikan vaksin COVID-19



Duduk di rumah dan putuskan rantai COVID-19

- Pastikan pengudaraan yang baik di rumah

Sentiasa pakai pelitup muka, basuh tangan dan kekalkan jarak 1 meter.

Pengumuman 📢 ▶ Video Kesihatan Terkini

Media Terkini	Media Popular	Isu Semasa
---------------	---------------	------------

CONTOH – COVID 19


Kementerian Kesihatan Malaysia
R0 18/08/2021 1.04 (0.0 ==)
CARI / SEARCH

Utama
Terkini Harian
Terkini Negeri
Vaksin Covid-19
Garis Panduan
Penyelidikan
Infografik
Video
Hotline
Faq & Sop

COVID-19 lebih mudah tersebar di kawasan ini: 3S



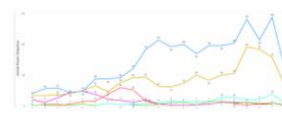
1 Kawasan sesak yang dipenuhi ramai orang

2 Kawasan di mana sembang/perbualan jarak dekat berlaku

3 Kawasan sempit dan tertutup dengan pengudaraan yang kurang baik

Risiko penularan lebih tinggi di kawasan yang mempunyai ketiga-tiga faktor ini. Walaupun sekatan telah dilonggarkan, sentiasa menilai kawasan yang dikunjungi dan kekal selamat dengan elakkan 3S.

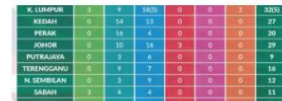
Sumber: <https://www.arkib.newsroom/qa-detail/coronavirus-disease-covid-19-how-is-it-transmitted>



TINDAKAN YANG DIPERLUKAN BAGI MENGURANGKAN RISIKO DAN KESAN PEMBUKAAN SEMULA LEBIH BANYAK SEKTOR EKONOMI DAN SOSIAL

Written on Wed, Aug 18, 2021.

KKM mengambil maklum dan memahami kepentingan pengumuman Kerajaan mengenai pembukaan semula lebih banyak sektor ekonomi ...



KENYATAN MEDIA BERSAMA - KKM & MOSTI : PELAKSANAAN KAWALAN PANDEMIK DAN VAKSINASI TERUS BERJALAN LANCAR

Written on Wed, Aug 18, 2021.



STATUS TERKINI VARIANT OF INTEREST (VOI) DAN VARIANT OF CONCERN (VOC) DI MALAYSIA SEHINGGA 17/08/2021

Written on Wed, Aug 18, 2021.

Bilangan kes jangkitan COVID-19 dengan VOC dan VOI di seluruh negara setakat 17 Ogos 2021. Terdapat pertambahan Varian D...



BAGAIMANA MENGURANGKAN RISIKO COVID-19 APABILA ANDA BERKONGSI KERETA

Written on Wed, Aug 18, 2021.

Anda mempunyai risiko untuk mendapat Covid-19 di dalam kereta jika ada penumpang lain yang membawa virus. Virus boleh had...



SITUASI DAN MAKLUMAT TERKINI MENGENAI MUTASI SPIKE PROTEIN COVID-19

Written on Tue, Aug 17, 2021.

KKM ingin memaklumkan bahawa, kajian berterusan oleh Institut Penyelidikan Perubatan (IMR) telah mengenalpasti sebanyak ...

Sorotan

Kenyatan Media Bersama - KKM & MOSTI : PELAKSANAAN KAWALAN PANDEMIK DAN VAKSINASI TERUS BERJALAN LANCAR

WED, AUG 18, 2021

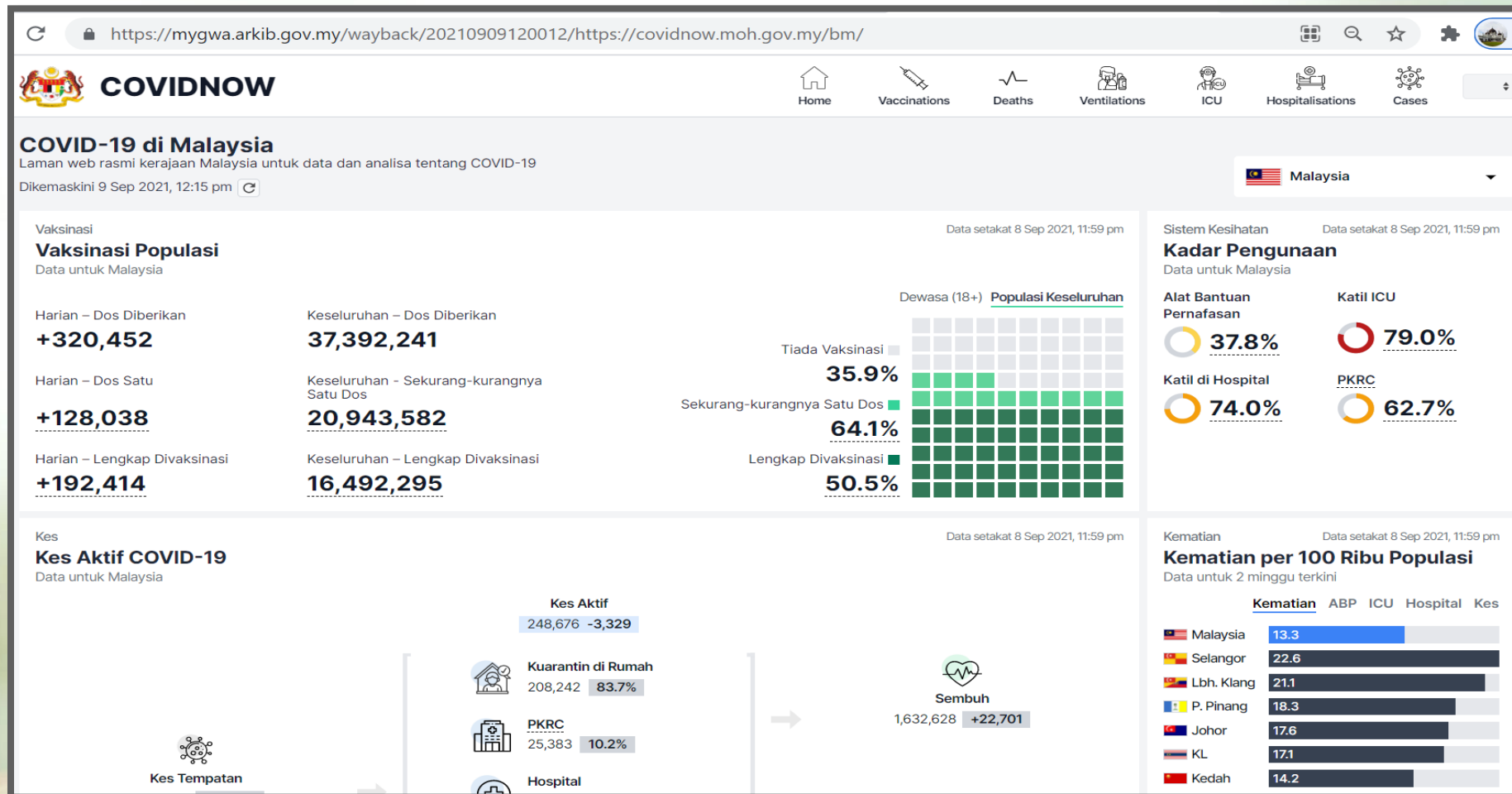
Capaian Popular

Informasi Vaksinasi

Senarai Fasilitas Kesihatan

Informasi Umum

CONTOH – COVIDNOW





1. Taklimat

- ☐ Pengurusan rekod elektronik sektor awam.

2. Persidangan

- ☐ Persidangan SARBICA di Singapura pada tahun 2019 dimana pembentangan mengenai MyGWA mendapat sambutan yang baik dan antara Institusi di Asia Tenggara yang mempunyai program Web Archiving

3. Seminar / Majlis

- ☐ Hari ICT ANM



4. Facebook Live ANM

☐ 25 Mac 2021

5. WIKI

☐ Kemaskini bahagian '*Web Archiving Initiatives*' di dalam WIKI dimana Malaysia juga merupakan salah satu jabatan yang mengimplementasi pengarkiban laman web.

6. Sesi Pengajaran Kolaborasi Bersama Industri Bagi Kursus Pengarkiban Web (ICM561) – UiTM Puncak Perdana

7. Kerjasama antara UiTM Puncak Perdana bagi persembahan di Persidangan Tahunan **Dublin Core Metadata Initiative (DCMI)**
- Best Practice: Metadata uses in web archiving

Arkib juga telah mengemaskini maklumat pada bahagian ‘Web Archiving Initiatives’ yang terdapat di laman web WIKI

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



https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives							
	Ivy Plus Libraries Confederation	United States	2013	Archive-It, Conifer	1	1	The Ivy Plus Libraries Confederation's Web Resources Collection Program is a collaborative collection development effort to build curated, thematic collections of freely available, but at-risk, web content in order to support research at participating Libraries and beyond. Participating Libraries are: Brown , Chicago , Columbia , Cornell , Dartmouth , Duke , Harvard , Johns Hopkins , MIT , Penn , Princeton , Stanford , and Yale . ^[104] Collections are accessible via Archive-It .
	Malaysian Government Web Archive (MyGWA)	Malaysia	2017	Wayback, WGET , WPULL	>1	0	National Archive of Malaysia started to archive websites of public sector in Malaysia since 2017.
	HTTP Archive						Crawls popular websites for Data analysis ^[105]
	National Library of Medicine (U.S.)	United States	2009	Archive-It, Conifer		~8	NLM web collecting is guided by the Collection Development Guidelines of the National Library of Medicine and other strategic collecting efforts. Collections include Global Health Events, the Opioid Epidemic, HIV/AIDS, Health and Medicine Blogs, and NLM's own web presence. ^[106]
	Smithsonian Libraries and Archives	United States	2000	Heritrix , Archive-It , Webrecorder , Conifer		5	The Smithsonian Libraries and Archives collects websites and social media accounts that document the history of the

“Facebook Live Pengarkiban Laman Web - Inisiatif Baru dan Inovatif” (25 Mac 2021)

Browser address bar: <https://www.facebook.com/myarkibnegara/videos/965113510961536>

Browser tabs: MyGWA - Malaysia..., Email MyGOVUC, ::Penyata Baki SSPN..., SSPN-i Plus ONLINE, Ddms 2.0 Login, ezHASIL, Download the DbVL..., Identity Self Service, Reading list

Facebook video player interface:

- Facebook logo and close button (X)
- Video thumbnail: 
- Video title: **Malaysian Government Web Archiving (MyGWA) Inisiatif Bar...**
- Video description: Inisiatif terkini Arkib Negara Malaysia.
- Engagement: 118 likes, 46 Komen, 1.1K Lihat
- Buttons: Suka, Komen
- Section: Komen (Sembunyikan)
- Sort: Paling Relevan
- Comments:  Pengongsi Laman FB Rasmi Arkib Negara Malaysia · 1:31 Selamat pagi semua



UNIVERSITI TEKNOLOGI MARA | Fakulti Pengurusan Maklumat

ICONS | **iSchools**

WEBINAR : PENGARKIBAN LAMAM SESAWANG: ISU DAN CABARAN

SESI PENGAJARAN KOLABORASI BERSAMA INDUSTRI
BAGI KURSUS PENGARKIBAN WEB (ICM561)

PENCERAMAH :
ENCIK MUHAMMAD IDZWAN BIN RAMLI
PEGAWAI TEKNOLOGI MAKLUMAT
KANAN, SEKSYEN PENGURUSAN
REKOD ELEKTRONIK, ARKIB NEGARA
MALAYSIA

**MODERATOR : DR. FARRAH DIANA
BINTI SAIFUL BAHRY**
PENYARAH
FAKULTI PENGURUSAN MAKLUMAT
UNIVERSITI TEKNOLOGI MARA

TARIKH: 29 JUN 2021 (SELASA)
MASA: 11:00 PAGI – 1:00 PETANG
**PLATFORM: GOOGLE MEET DAN FACEBOOK
LIVE "ICONS FPM"**

 **ICONS FPM**

“Sesi Pengajaran Kolaborasi Bersama Industri
Bagi Kursus Pengarkiban Web (ICM561) UiTM
Puncak Perdana”

(29 Jun 2021)

“Kerjasama dengan UiTM Puncak Perdana di Persidangan Tahunan Dublin Core Metadata Initiative (DCMI) - Best Practice: Metadata uses in web archiving” (5 Oktober 2021)

innovation in metadata design, implementation & best practice

Dublin Core™ Metadata Initiative

Home Specifications ▾ News Community ▾ Learning ▾ About ▾ Contact

[Home](#) / [DCMI Annual Conferences](#) / [DCMI Virtual 2021](#) / [DCMI Virtual: Presentations](#) / Metadata uses in web archiving initiatives in Malaysia

Best Practice: Metadata uses in web archiving initiatives in Malaysia

Title: Metadata uses in web archiving initiatives in Malaysia

Moderated by: Joseph Yi

Date: 2021-10-05 07:00

Resources:

Speaking:



Farrah Diana Saiful Bahry

Faculty of Information Management, Universiti Teknologi MARA

Farrah Diana Saiful Bahry is a senior lecturer and researcher in the past 10 years. Currently, she beholds a post as Head centre of Studies in Informatics at the Faculty of Information Management, Universiti Teknologi MARA. Her research interests are related to web design and user experiences, web evaluations and user behaviours, web archiving, and web content retrieval. With 7 years of experience in government bodies, she also put an interest in government websites, government digital initiatives, and government information technology adoption. She teaches in the areas of information systems, web design and content management, and technical organization.

Abstract:

Web archive initiatives in Malaysia have started and progressively take place. Web archive initiatives from Malaysia educational institution and national level will be exposed and explained in terms of implementation and metadata uses. Both education institution and national level initiatives (which is known as MyGWA) actively collect web archive content in specified web archive repositories. The web archive repository is either equipped with descriptive metadata to facilitate web users in accessing the collection publicly. However, there is a need to improve and extend the use of metadata to meet web archive extensive usage in the future especially using appropriate metadata standards. Thus, some comparative studies and benchmarking have been made and the findings will be visualized and presented. The result of this comparative study is useful for further enhancement of web archive repository towards web preservation.



Muhammad Idzwan bin Ramli

National Archives of Malaysia

Muhammad Idzwan bin Ramli Muhammad Idzwan is a Senior Information Technology Officer, Electronic Records Management Section, National Archives of Malaysia. He took a Bachelor of Computer Science (Software Engineering, Diploma in Information Technology at MSU. Mr. Muhammad Idzwan has a lot of experience in his own career for more than 15 years and has worked under large organizations such as Ambank, Permodalan Nasional Berhad (PNB) and Local Government Departments. Currently his description of work more focus on Web Archiving project for Malaysia government project.

 **DCMI** 2021
VIRTUAL

Online, 2021 October 4-15

https://www.dublincore.org/conferences/2021/presentations/metadata_uses_in_web_archiving_initiatives_in_malaysia/

ISU DAN CABARAN

- Kapasiti Storan
- Firewall - *blacklist*
- 3rd party tool;s – ie flippingbook
- Javascript
- Perubahan Teknologi - tools

Web Curator Tools

Web archiving software comparison

☆
🔍
☁

File Edit View Insert Format Data Tools Extensions Help

🖨

+

🔍

100%

🗨

Comment only

A1

fx

=HYPERLINK("https://github.com/archivers-space/research/tree/master/web_archiving", "For more information about this spreadsheet, please visit

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	For more information about this spreadsheet, please visit https://github.com/archivers-space/research/tree/master/web_archiving																	
2		General information					Users & developer features					"Out of the box"						
3	Name (link to home page)	Open source?	Source repo	Operating system(s)	Primary dev. language	Target audience	CLI	GUI	WUI	Library API	Network API	Extensibility framework	Parallel crawling	Scheduled crawling	Crawl storage format(s)	Capture raw responses	Follow links	URL
19	simplecrawler	BSD	GitHub	Lin, Win, macOS	Node.js	user	✓	✗	✗	✓	✗	Node modules, ES6+	✓	✗	files on disk	✓	✓	
20	Squidwarc	GPLv3	Github	Lin, macOS	Node.js	user, enterprise	✓	✗	✗	✓	✗	Node modules, ES6+	✓	✗	WARC	✓	✓	
21	StormCrawler	Apache	GitHub	Lin, Win, macOS	Java	user, enterprise	✓	✗	✗	✓	✗	Java classes	✓	limited	several db options	✗	✓	
22	WAIL (Electron)	GPLv3	GitHub	Lin, Win, macOS	Node.js (Electron)	user	✗	✓	✗	✓	✗	✗	✓	✓	WARC	✓	✓	
23	WAIL (py)	MIT	GitHub	Win, macOS	Python	user	✗	✓	*	✗	✗	✗	✓	✗	WARC	✓	✓	
24	WebMagic	Apache	GitHub	Lin, Win, macOS	Java	user	✗	✗	✗	✓	✗	Java		✗	files on disk		✓	
25	WebRecorder.io	Apache	GitHub	Lin, Win, macOS	Python	user	✓	✓	✓	✓	✗	✗	✗	✗	WARC	✓		
26	wget	GPL	Savannah	Lin, Win, macOS	C	user	✓	✗	✗	✗	✗	✗	✗	✗	WARC, files on disk	✓	✓	
27	wpull	GPL	GitHub	Lin, Win, macOS	Python	user	✓	✗	✗	✓	✓	plug-ins, scripts	✗	✗	WARC	✓	✓	

<https://docs.google.com/spreadsheets/d/1FqywaZnIhhQ7jDCC-W64NMRf5rDeh2Shx3u01MsBmTQ/edit#gid=0>

ISU DAN CABARAN

Patches

- ❖ Kemaskini naiktaraf *patches* yang digunakan oleh MyGWA

Tools

- ❖ Mengkaji tools - *open source* lain yang boleh digunakan bagi crawling laman web terutamanya laman web yang menggunakan javascript

Format

- ❖ Mengkaji format laman web terkini dan akan datang

Viewer

- ❖ Mengkaji penggunaan open wayback sebagai medium paparan laman web pada masa akan datang

ISU DAN CABARAN

Storan

- ❖ Peningkatan storan kapasiti MyGWA - budget etc

Server

- ❖ Mengkaji keperluan spesifikasi server MyGWA

SSL

- ❖ Implementasi sijil SSL pada portal MyGWA

DRC

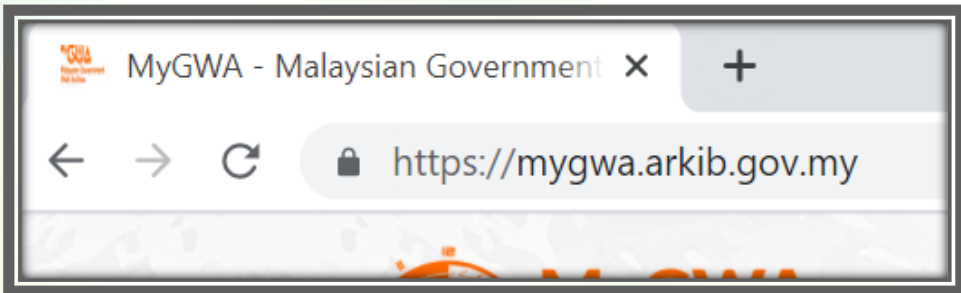
- ❖ Penyediaan DRC di Pusat data Sektor Awam (PDSA) bagi MyGWA sebagai sandaran (*backup*) sekiranya berlaku apa – apa masalah

ISU DAN CABARAN

```
[root@mygwa ~]# lastb -a | more
testuser ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
testuser ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
testuser ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
testuser ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
mysql ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
sirius ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
sirius ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
kk ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
kk ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
nicole ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
nicole ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
mysql ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
dev ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
dev ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
crms2 ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
crms2 ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
dadi ssh:notty Tue May 4 23:05 - 23:05 (00:00) localhost
dadi ssh:notty Tue May 4 23:04 - 23:04 (00:00) localhost
test ssh:notty Tue May 4 23:04 - 23:04 (00:00) localhost
test ssh:notty Tue May 4 23:04 - 23:04 (00:00) localhost
user ssh:notty Tue May 4 23:04 - 23:04 (00:00) localhost
```

```
[root@mygwa ~]# grep "Failed password" /var/log/secure
May 4 22:59:49 mygwa sshd[27129]: Failed password for invalid user
tom from ::1 port 52636 ssh2
May 4 22:59:51 mygwa sshd[27134]: Failed password for invalid user
doctor from ::1 port 52638 ssh2
May 4 22:59:54 mygwa sshd[27140]: Failed password for invalid user
node from ::1 port 52640 ssh2
May 4 22:59:56 mygwa sshd[27145]: Failed password for invalid user
```

4 Mei 2021 cubaan akses (*hacking*) ke server MyGWA melalui SSH. (SSH Brute Forcing)



SIJIL SSL

(<https://gpmi.mampu.gov.my>)

Certificate Viewer: mygwa.arkib.gov.my

General Details

Issued To

Common Name (CN)	mygwa.arkib.gov.my
Organisation (O)	ARKIB NEGARA MALAYSIA
Organisational Unit (OU)	<Not part of certificate>

Issued By

Common Name (CN)	GlobalSign Extended Validation CA - SHA256 - G3
Organisation (O)	GlobalSign nv-sa
Organisational Unit (OU)	<Not part of certificate>

Validity Period

Issued On	Tuesday, 7 February 2023 at 17:21:11
Expires On	Sunday, 10 March 2024 at 17:21:10

```
-----BEGIN PRIVATE KEY-----
IIEvgIBADANBgkqhkiG9w0BAQEFAASCBKgwggSkAgEAAoIBAQC9CVNsT6NnNF7
zfw592v4cnBjBwQc/60jgroih5lk8ME1b1UbKtsKrqjDXEE1zc4EiU840fr6z
HpazOknLUeijeZLH1X04DiSyag+A4dk3Lxa821NYq+bEPidZnCMcgA9Xz4Bn/r
dEbgJhrTuzhMqja6S2usrbzy2ZABuQkIPEIWGHQGYnrEIAfPdpxiTb01VjdC1
OkDjxGIY5hPzOVtKaFVgiG1zogCH7E6DG1kDNwZMXqvHMPwzVrvCw9ymkojeEL
fXwb69H5od6KJry/9d98yK+3QFy8F8AFacyoRcXqCfus6tfLuK06uda/khX4bc
zp2AnzRagMBAAECggEBAJnuHCZYeyCw/qGSCoV66YQtAdM/N2NfU/JUEDTfqMf
bjVjrvYuPzCnYq8fZ0aiMW7RJ814/qGXbyhhACoVzr1nh16IFYOWDnU03yEBSt
bgKHr6Y8yh2sbQ9j5TsM73JnvRat+0uznX5uEsC3hqTFESsd8AfZiCiORHh04
CAN9cw0kwaDs3MPytsSF83Ck/CPDhN88MamvzUVst1VR8Kaxdw4z++Qv+ivIif
8j/LEDtvv5BAJ9ESCTXDxaLFia0PeTYdxX7GgQ2s1c11Dy0LVssJSpWTQOXe4L
FrdvW+zCriHQop4vuYmJ3sgZIwrfxkySnxbDBHJFBKcGyEA4cCA6SG/ORJeh5l
JpNzdh+1ihj7z2qy2oe5Z9LKCIQ10UHoYrKjM35y5yetFr2Ra52EITNhnbo0Rv
W4KNfw8/lmGnidTLX2zF4bMWU4Ef1iR1IYvxf0iyXv3qft+Yaki3cjDDwog2B1
```

PENGURUSAN PERUBAHAN



KESIMPULAN

- Sistem Malaysian Government Web Archive (MyGWA) merupakan sistem pengarkiban laman web yang penting dan amat berguna dalam memudahkan proses pemindahan kandungan laman web secara dalam talian
- Sistem MyGWA memastikan laman web yang telah diarkibkan ini dapat diakses dan dipaparkan seperti laman web asal walaupun laman web berkenaan telah ditutup atau jumud
- Koleksi laman web yang tersimpan di dalam sistem MyGWA boleh dijadikan sebahagian koleksi warisan budaya negara yang boleh dirujuk oleh orang awam pada masa hadapan.

“Terima Kasih”

