



BIBLIOTECA
NACIONAL
DE MÉXICO



Grupo de
Preservación
Digital



Preservación de sitios web: avances y retos

abril 2024



¿Qué es la preservación de sitios web?

La **preservación web** es el proceso de **recopilar** partes de la World Wide Web, **preservar** las colecciones en un formato específico y luego poner a **disposición** los archivos para su **acceso y uso**.



¿Qué es?

El International Internet Preservation Consortium (IIPC) define al archivado web (web archiving) a partir de cuatro procesos interdependientes:

- 1) Selección,
- 2) Captura,
- 3) **Preservación y**
- 4) **Accesibilidad futura.**



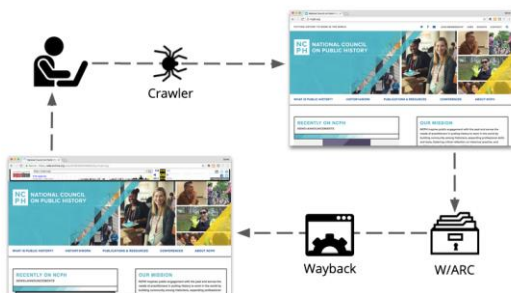
INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM

archivo web

PROCESO

- Selección,
- Captura,
- Preservación y
- Accesibilidad futura.

Preservación web
archivado web
web archiving



CONJUNTO DE INFORMACIÓN

Colecciones recopilada de la web por periodos definidos, los cuales pueden ser organizados de diferentes formas

Archivo web nacional
Archivo web temático

Australian Web Archive



ARCHIVO ESTÁNDAR

Archivo con formato estándar. Contiene todo lo que conforma el sitio web: CSS, htmls, JS, mp4...

archivo.warc
archivo.wacz

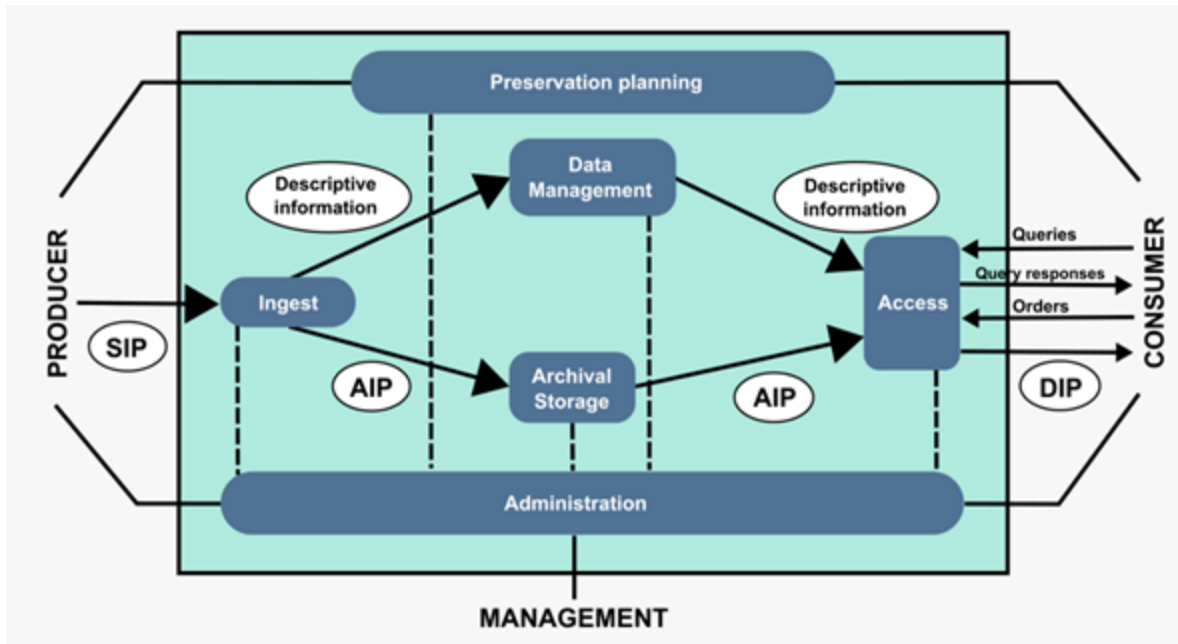


WARC
warc.gz
Antecesor: ARC



WACZ
Web Archive Collection Zipped
estándar de empaquetado para
archivos web en la web

Nuestras bases



OAIS



Ciclo de vida de archivo web



¿Por qué preservar sitios web?

- Contenido web: frágil y efímera
- Importancia del Archivo web

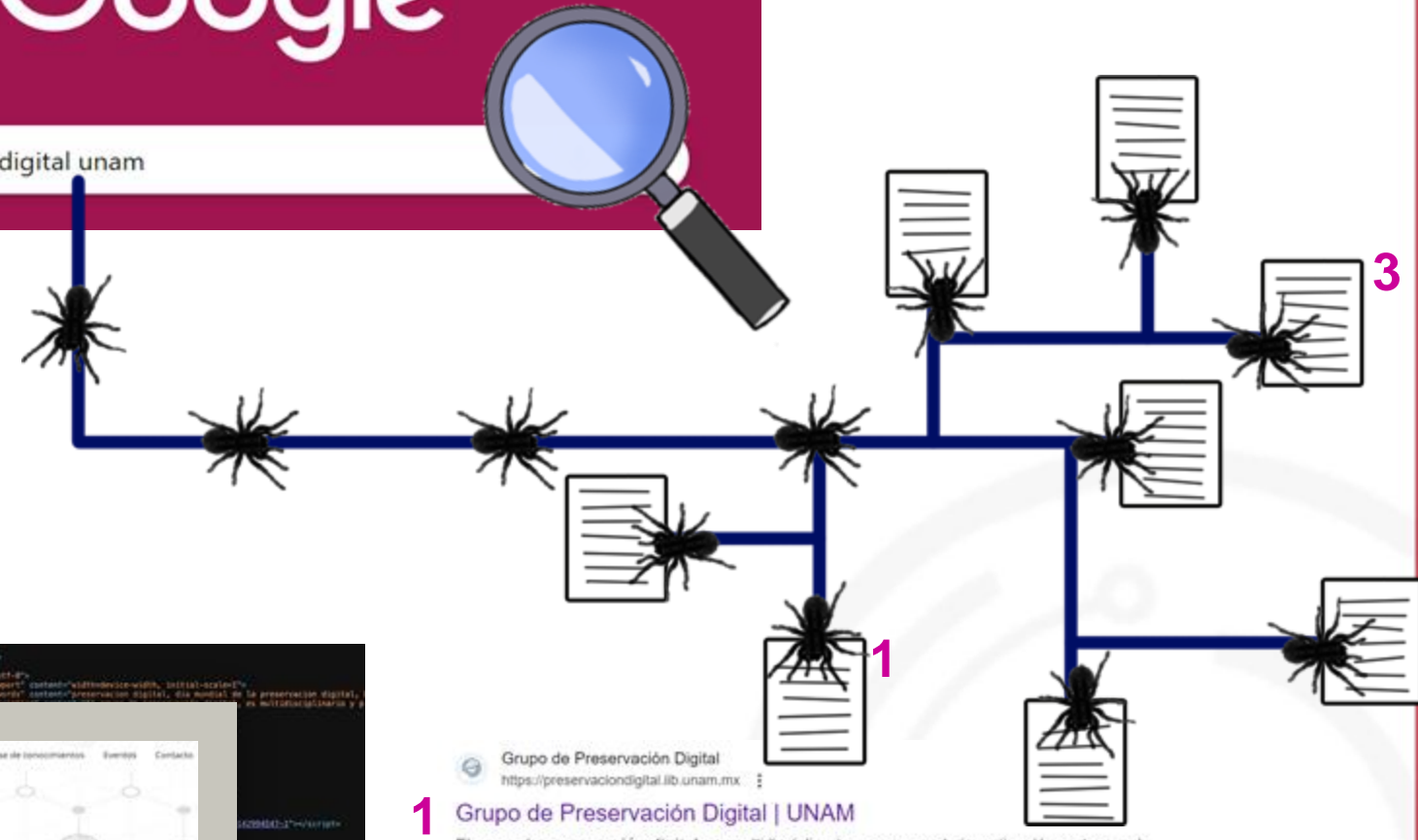


Video original:

<https://bit.ly/3xH1ZCW>

¿Cómo funciona?

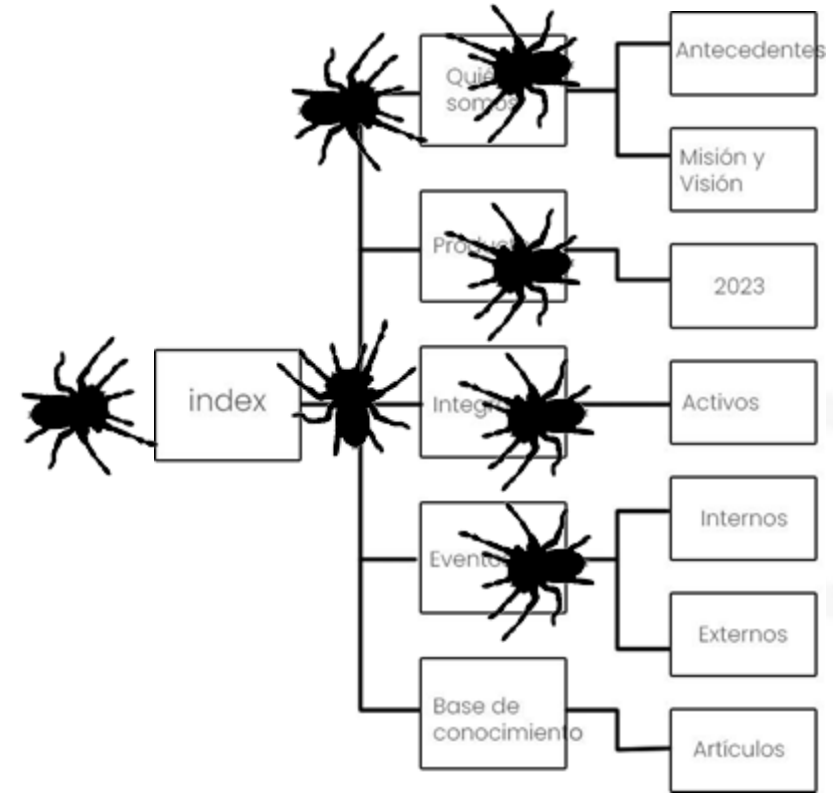
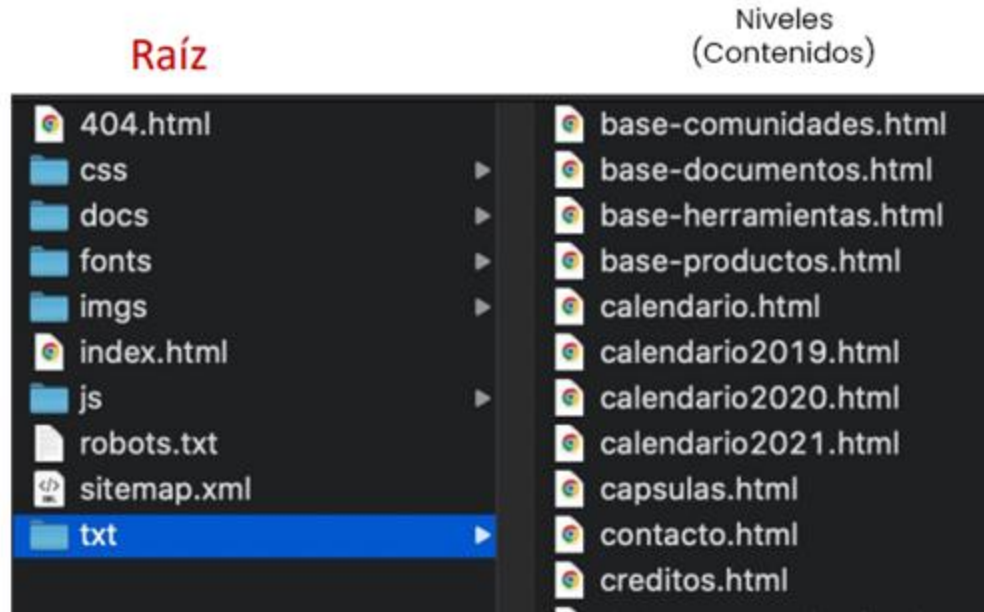
Crawler



- 1 Grupo de Preservación Digital | UNAM
El grupo de preservación digital, es multidisciplinario y promueve la investigación en torno a la problemática de la preservación de materiales digitales ...
 - 2 Preservación Digital – Dirección General de Repositorios ...
Aviso de Privacidad Simplificado e Integral de la DGRU-UNAM para las Personas Usuarias de las Plataformas Digitales Administradas por la DGRU; Aviso de ...
 - 3 Grupo de Preservación Digital
https://preservaciondigital.lib.unam.mx > integrantes
- Acerca del Grupo de Preservación Digital - UNAM

SEO

¿Cómo funciona?



masiva



sitemap



robot

selectiva

Formato estándar

.WARC

- Revisado
- Incluye metadatos
- Estándar

Web Archive



WARC

warc.gz
Antecesor: ARC

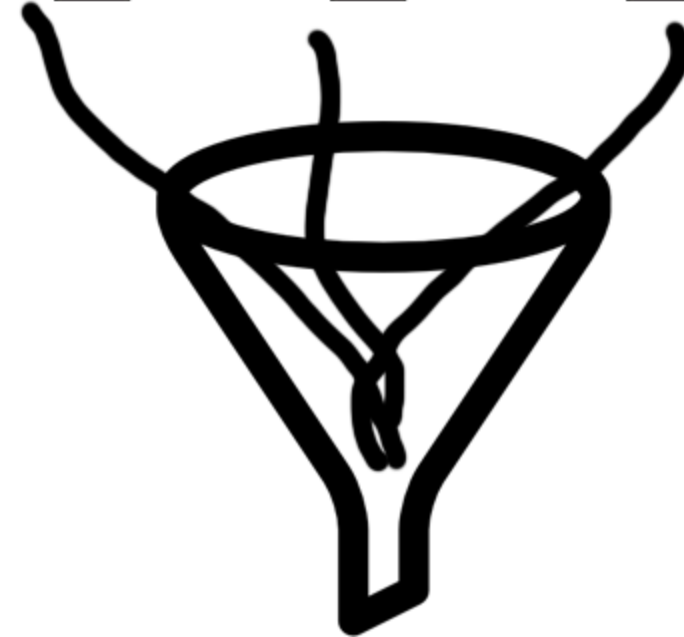


WACZ

Web Archive Collection Zipped
estándar de empaquetado para
archivos web en la web

Almacenar rastreos web como
serie de bloques de contenido
Con metadatos asignados

css • js • fonts



WARC

Objetivo general de la temática

Analizar a los elementos que conforman la **materialidad de los sitios web para destacar la importancia** de guardarlos con fines de preservación y exhortar a las **dependencias de la UNAM** e instituciones a **nivel nacional**, en la **realización** de esta buena práctica, trabajando de manera colaborativa, para recopilar, preservar y difundir el patrimonio documental digital nacional, garantizando su acceso universal a través de la Biblioteca Nacional.





Objetivo general de la temática





2024

SESIONES

- Modalidad virtual
- Dos veces al mes:
 - El 2º y 4º viernes de mes, a partir de febrero a noviembre de 2024
- Duración 1.5h por sesión
- Horario: 10-11:30h

FEB	MAR	ABR	MAY	JUN	AGS	SEP	OCT	NOV
9	8	12	24	14	9	13	11	22
23	22	26		28	23	27	25	



Integrantes

Carolina Silva Bretón
coordinadora

Equipo de trabajo:

- Alejandro Juárez Arriaga
- Ma Isabel Hernández Hernández
- Judith Alejandra Pérez
- Lizbeth Luna González
- Miguel Ángel G. Mejía Argueta
- Ricardo Montes Pérez
- Valeria Palafox Larios

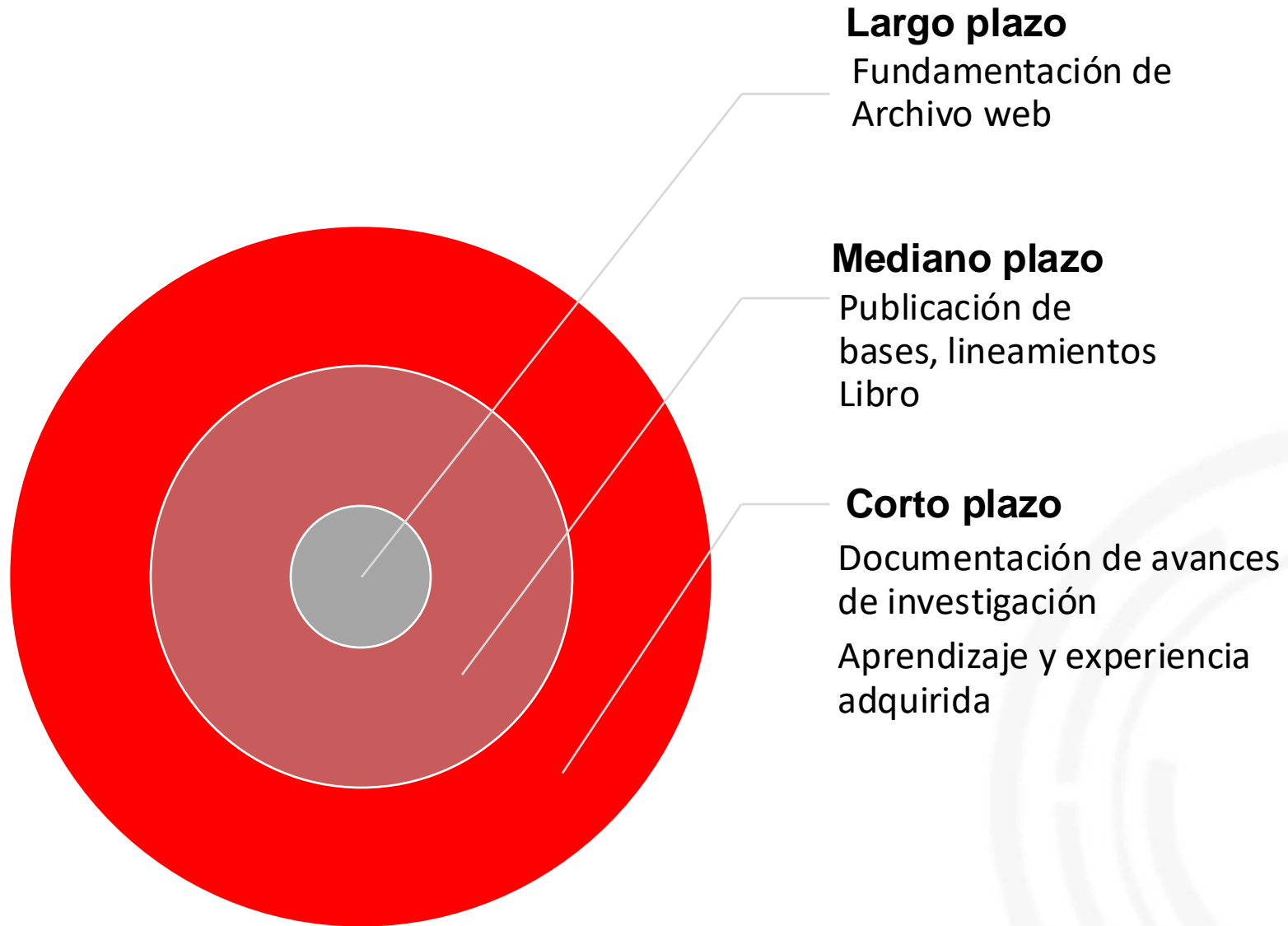
Agradecimientos:

- Jo Ana Morfin (2023)
- Esmeralda Martínez Montes (2023)

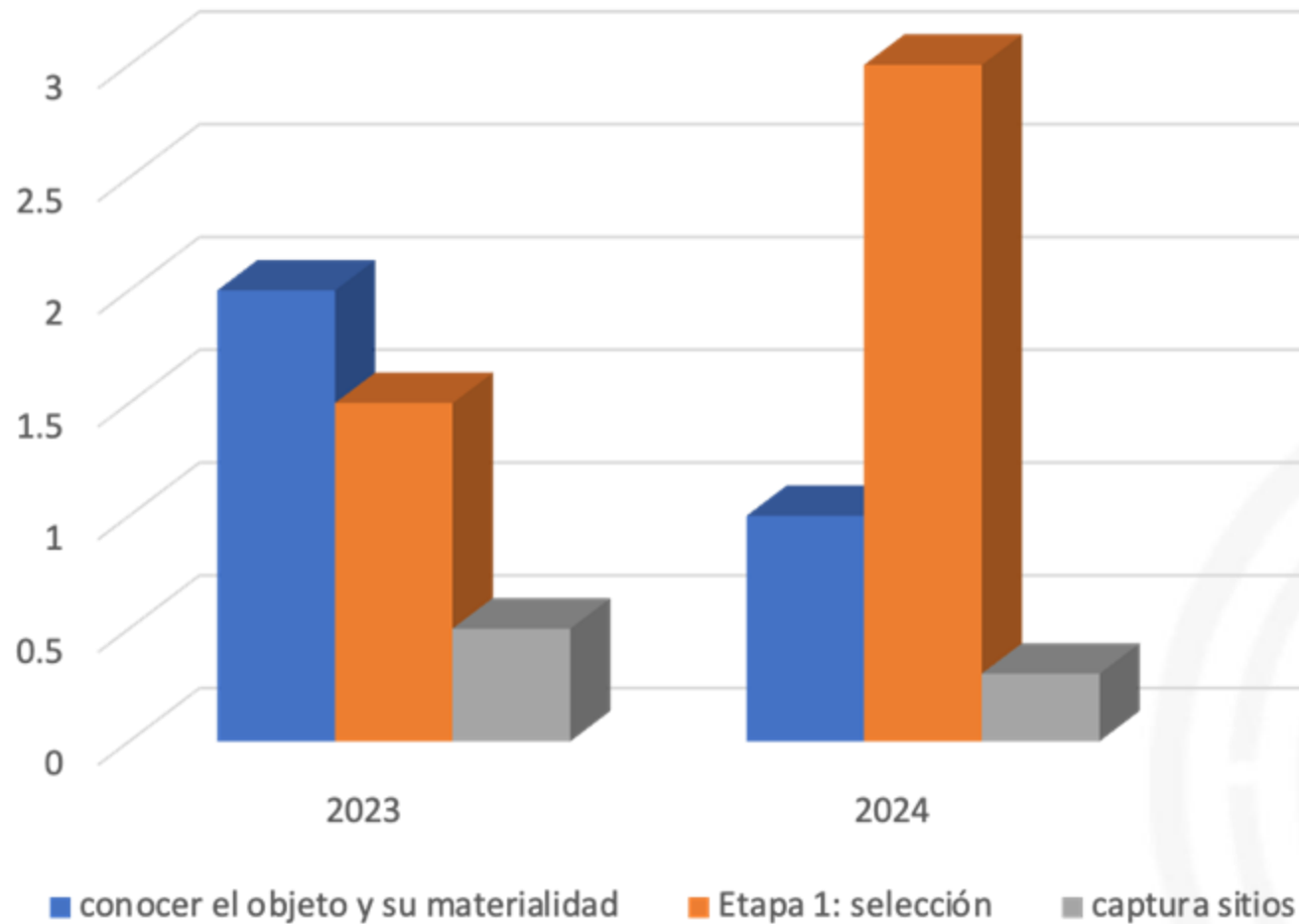




Metas



Avances de trabajo



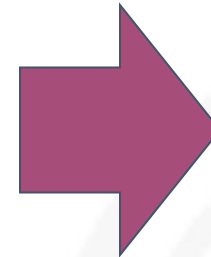
sesiones teórico-prácticas



2023



materialidad del objeto



G
L
O
S
A
R
I
O

Lenguaje en común

Frutos 2023



Glosario

(lenguaje en común)

Específico de archivo web



Concientización



Problemática Casos reales

Posible solución con esta buena práctica



Análisis de herramientas



BIBLIOTECA NACIONAL DE MEXICO



Grupo de Preservación Digital

Frutos 2023

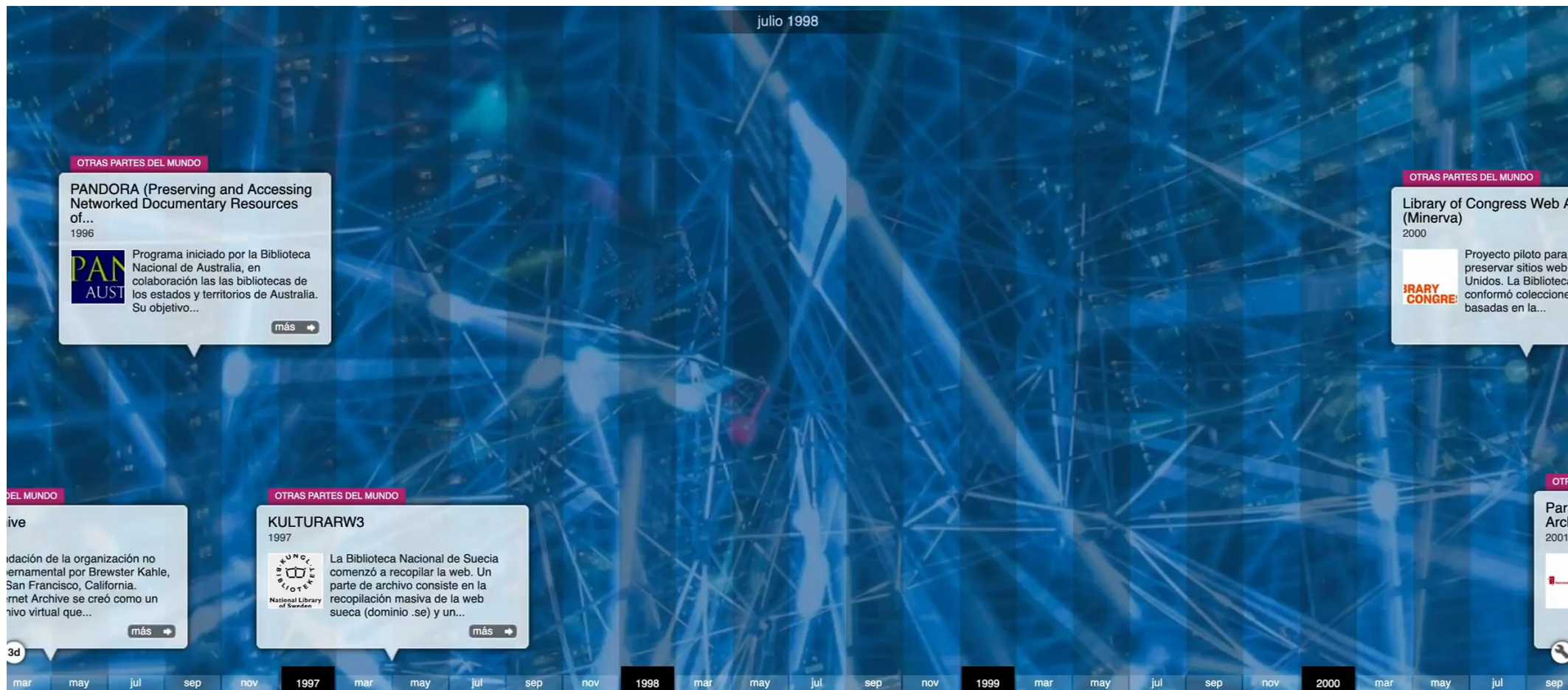
Estado del arte



Línea de tiempo interactiva

Revisión IIB-UNAM

Publicación y cómo citar





Crterios de seleccin

La preservacin web implica mantener y proteger el contenido en lnea para uso de futuros usuarios. Los criterios de seleccin en este contexto se basan en identificar qu contenido debe ser preservado.



Valor histrico,
cientfico y
cultural

Sitios
obligados

Temporalidad

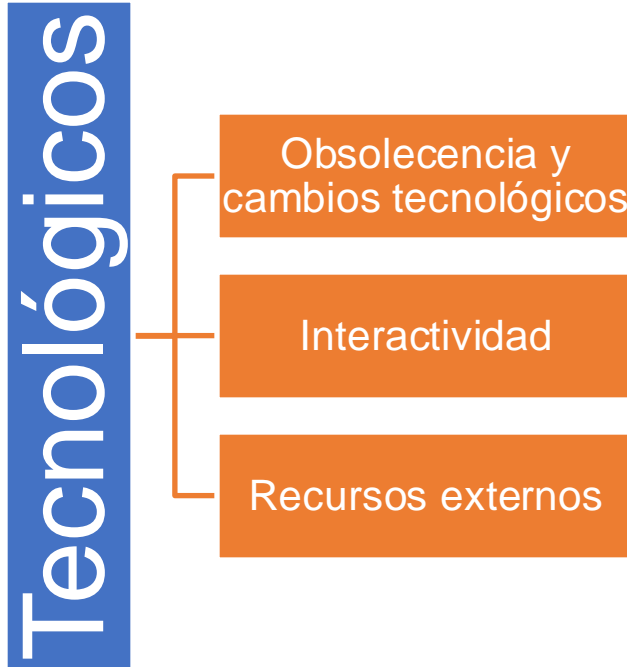
Actualidad

Diversidad

Popularidad

Relevancia

Aspectos a conservar



¿Todo se debe preservar?

- No todo es relevante
- Lo público y lo privado
- Derechos de autor
- Datos confidenciales





BIBLIOTECA NACIONAL DE MEXICO



Grupo de Preservación Digital

Frutos 2023



Colaboración en la publicación del GPD



Resumen...

Curso virtual



100 asistentes
Extranjeros y nacionales

Preservación de sitios web: concientización y etapa de selección

Alejandro Juárez Arriaga, Esmeralda Martínez Montes, Ma. Isabel Hernández Hernández, Jo Ana Morfín Guerrero, Judith Alejandra Pérez Sandoval, Lizbeth Luna González, Miguel Ángel G. Mejía Argueta y Valeria Palfox Laríos
Coordinadora: Carolina Silva Bretón

Resumen

Se entiende como preservación de sitios web a una serie de acciones que implican la selección, recopilación, almacenamiento, así como la implementación de herramientas, estándares, políticas y mejores prácticas que garanticen la gestión y acceso a los archivos web a través del tiempo. Dado que los sitios en la red reflejan la cultura, historia e información de nuestra sociedad, la preservación de éstos es importante para conservar el patrimonio documental de nuestra nación. Es importante resaltar que no todo lo que se produce en la web es objeto de preservación. Esta investigación con enfoque cualitativo de carácter exploratorio, abarca sólo la primera etapa de preservación, busca la concientización ante la problemática focalizada en el grupo interinstitucional de preservación de sitios web, del GPD, y valoriza el objeto de selección para su preservación, tomando en cuenta políticas institucionales y criterios técnicos, así como características únicas de su materialidad, con el fin de enmarcar lineamientos de acción para su permanencia y accesibilidad pública.

Palabras clave: preservación documental digital, selección

...ción y acceso al patrimonio documental a nivel y administración de los recursos'.

(IM), es la institución encargada de integrar, monio bibliográfico y documental mexicano, scrito: bibliografía, hemerografía, archivos, en el extranjero por mexicanos y todo ha institución, a través del Instituto de i de la Universidad Nacional Autónoma de inistra, desde el 2016 ha implementado servación bibliográfica mexicana, entre las servación Digital (GPD), en 2017.

...cidas digitales y alojadas en servidores, ón digital, en el 2020 nace la idea, en la y sitios web. Es hasta el 2023 cuando se grupo Preservación de sitios web, del de Investigaciones Bibliotecológicas y oyecto de archivo web, en donde la Además de ésta, la Universidad del on instituciones que han desarrollado s web en nuestro país. Sin embargo, esta perspectiva, son el Archivo web ngreso, por mencionar algunos.

...heredado, realizará acciones en pos de su

...eb, ha sido el escenario perfecto para esta ar sobre esta buena práctica a través de a rificar la importancia de las páginas que se 1. Los retos, oportunidades, fortalezas, i usuarios tengan herramientas suficientes captura de los sitios.

...ortium (IIPC)' define al archívado web se interdependientes: 1) Selección, 2) to de preservación, ni todo lo que es lo de ciertos dominios de

El Grupo de Preservación Digital organizado por la Biblioteca Nacional de México, convocan a participar en el curso virtual para celebrar el:

Día Mundial de la Preservación Digital 2023

Contenido

- I. Introducción
Jo Ana Morfín [BIBLIOTECA NACIONAL DE MEXICO] y Carolina Silva [BIBLIOTECA NACIONAL DE MEXICO]
- II. Problemáticas
Miguel Ángel G. Mejía [GPD, UNAM]
- III. Antecedentes
Ma. Isabel Hernández [CONACI, HUMANIDADES, UNAM] y Esmeralda Martínez [CONACI, UNAM]
- IV. Criterios de selección
Alejandro Juárez [BIBLIOTECA NACIONAL DE MEXICO] y Judith Alejandra Pérez [BIBLIOTECA NACIONAL DE MEXICO]
- V. Política de recopilación
Valeria Palfox [GPD, UNAM]

Coordinadora: Dra. Carolina Silva Bretón, BNA / IB, UNAM

viernes 17 de noviembre de 2023

De 10:00 a 13:00 horas / Duración: 3 horas
En línea / Gratuito / Cupo limitado
Se otorgará constancia de asistencia.



Curso
Preservación de sitios web.
Etapa de selección



2024



Tesouro

a partir de los criterios/alcance y objetivos



Política de selección



¿QUÉ?



¿QUIÉN?



¿PARA QUÉ?



¿CÓMO?



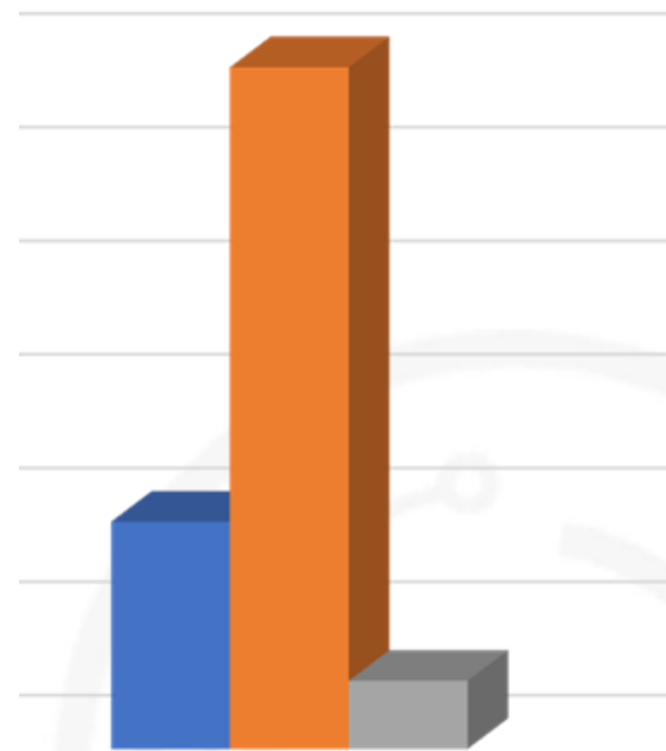
¿CUÁNDO?



¿DÓNDE?



Captura de 3 sitios web



2024

■ conocer el objeto y su materialidad

■ Etapa 1: selección

■ captura sitios

Retos identificados

- **Concientizar** a nuestros directivos institucionales
- **Colaboración** interinstitucional
- **Diversidad de contenidos** y toma de decisiones
- **Cambio tecnológico**
- **Derechos de autor**
- **Financiamiento y recursos**
- **Preservación a largo plazo**



Consideraciones futuras

- **Políticas y normativas**
- **Educación y conciencia**
- **Compartir conocimiento y experiencia internacional**
- **Desarrollo de tecnologías**
- **Acceso público**
- **Inclusión y diversidad**



Necesidades urgentes del GPD

Vinculación con otras áreas :

- Políticas
- Almacenamiento
- Planes de preservación





GRACIAS





Contacto

Dra. Carolina Silva Bretón
carolinasb@unam.mx

Correo con asunto: Interesado en preservación de sitios web

Nombre:

Profesión:

Dependencia:

¿Por qué estás interesado?: