

UNIVERSIDAD DE SALAMANCA
FACULTAD DE TRADUCCIÓN Y DOCUMENTACIÓN
MÁSTER EN SISTEMAS DE INFORMACIÓN DIGITAL

Trabajo de Fin de Máster



Archivabilidad web de las bibliotecas
universitarias: estudio comparado entre
China y España

AUTOR:

Shuqi SHI

DIRECTORA:

Dr.^a María Manuela Moro Cabero



**MÁSTER EN SISTEMAS DE
INFORMACIÓN DIGITAL**

UNIVERSIDAD DE SALAMANCA
FACULTAD DE TRADUCCIÓN Y DOCUMENTACIÓN
MÁSTER EN SISTEMAS DE INFORMACIÓN
DIGITAL

Trabajo de Fin de Máster

Archivabilidad web de las bibliotecas
universitarias: estudio comparado entre
China y España

AUTOR: Shuqi SHI

DIRECTORA: Dr.^a María Manuela Moro Cabero

Asiento catalográfico con formato del repositorio institucional GREDOS

Título:	Estudio de la archivabilidad de los sitios web de bibliotecas universitarias: análisis comparado entre China y España
Autor:	SHUQI SHI
Directoras:	Moro Cabero, María Manuela
Palabras clave:	[ES] Archivabilidad web; preservación digital; bibliotecas universitarias; ArchiveReady; comparación internacional [EN] Web archivability; digital preservation; academic libraries; ArchiveReady; international comparison
Clasificación UNESCO:	5701.06 Documentación 5701.02 Documentación Automatizada
Fecha:	2025-07-04
Resumen:	[ES] Este estudio analiza la archivabilidad de los sitios web de las bibliotecas universitarias en China y España, con el objetivo de identificar las dificultades técnicas y normativas que enfrentan estos entornos en la preservación digital. A través de un enfoque comparativo y empírico, se seleccionaron 60 universidades según rankings internacionales (ARWU y Wu Shulian) y se evaluaron sus sitios web utilizando la herramienta ArchiveReady, centrando el análisis en cinco dimensiones clave: accesibilidad, cohesión, metadatos, normalización y archivabilidad general. Los resultados muestran diferencias estructurales: mientras que las universidades españolas destacan en cumplimiento de normas y accesibilidad, las chinas presentan una alta cohesión y metadatos completos. Sin embargo, ambos países enfrentan desafíos como la captura de contenido dinámico y la fragmentación de estándares. El estudio concluye con recomendaciones orientadas a mejorar las estrategias institucionales de preservación digital, destacando la necesidad de políticas nacionales sólidas y tecnologías interoperables para garantizar el acceso duradero a los recursos académicos en la era digital. [EN] This study analyzes the web

archivability of university library websites in China and Spain, aiming to identify the technical and regulatory challenges faced in the context of digital preservation. Through a comparative and empirical approach, 60 universities were selected based on international rankings (ARWU and Wu Shulian), and their websites were evaluated using the ArchiveReady tool. The analysis focused on five key dimensions: accessibility, cohesion, metadata, standards compliance, and overall archivability. The results reveal structural differences: while Spanish universities stand out in standards compliance and accessibility, Chinese universities demonstrate strong cohesion and complete metadata. However, both countries face common challenges such as capturing dynamic content and fragmented metadata standards. The study concludes with recommendations to improve institutional digital preservation strategies, emphasizing the need for robust national policies and interoperable technologies to ensure long-term access to academic resources in the digital age.

Índice de contenido

Trabajo de Fin de Máster	1
Índice de Tablas	7
Índice de Figuras	8
Índice de Gráficas	9
0. Introducción	10
0.1 Antecedentes de la investigación	1
0.2 Objeto, pregunta y objetivos de investigación	2
0.3 Metodología de investigación	4
0.4 Dificultades observadas durante la investigación	5
0.5 Estructura del ensayo	6
Capítulo 1 Contextualización de la página Web.	9
1.1 ¿Qué es una página web?	9
1.2 Datos sobre el desarrollo de sitios Web, su vigencia y principales riesgos a los que están expuestos.	11
1.3 El Consorcio de Preservación Internacional de Internet-IIPC	14
Capítulo 2 Proceso y estado actual del archivado web	18
2.1 Proceso de archivado de los sitios web y análisis de su complejidad.	18
2.2 <i>Internet Archive</i> y su <i>wayback machine</i>	21
2.3 Una mirada a la legislación reguladora: China vs España.	23
2.4 Las herramientas principales del proceso de archivado empleadas en China vs España.	24
2.5.1 China	25
2.5.2 España	26
Capítulo 3 Evaluación del grado de archivabilidad de las páginas Web.	30
En este capítulo se detalla el concepto de archivabilidad, la herramienta empleada de uso gratuito personal “archiveready” y otras herramientas de utilidad para valorar problemas.	30
3.1 concepto de archivabilidad	30
3.2 La herramienta ArchiveReady	31
3.3 Otras herramientas complementarias	33
Capítulo 4 Selección de muestras y recolección de datos	35
4.1 Selección de muestras	35
4.2 Recopilación de datos	37
4.3 Datos obtenidos	37
Capítulo 5 Presentación de los resultados del estudio	43

5.1 Análisis de las tendencias generales	43
5.2 Diferencias entre los grupos	46
5.4 Evaluación global y tendencias generales	52
Capítulo 6 Análisis de casos. Enfoque particularizado	54
6.1 Caso 1: Universidad de Tsinghua (China) vs Universidad de Barcelona (España) – Grupo de élite	54
6.2 Caso 2: Universidad Normal Central de China vs Universidad de Salamanca (España) – Grupo intermedio	58
6.3 Caso 3: Universidad de Jimei (China) vs Universidad Carlos III de Madrid (España) – Grupo inferior	61
6.4 Consideraciones conclusivas	64
Capítulo 7. Conclusiones, análisis de causas y recomendaciones sobre la investigación	65
7.1 Conclusiones del estudio	65
7.2 Análisis de causas	66
7.3 Recomendaciones de política	68
7.4 Perspectivas de investigación	69
Referencias bibliográficas empleadas	70

Índice de Tablas

Tabla 1. Relación anual de sitios Web y de sitios web activos. (Fuente: Siteefy, https://siteefy.com/how-many-websites-are-there/	11
Tabla 2. Comparación de la legislación sobre el archivo web entre China y España	16
Tabla 3. Contenidos archivados en la Web española. Fuente: Wikipedia. Disponible en: https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives	27
Tabla 4. Comparativa del acceso entre archivos web españoles (Fuente: Wikipedia en español, https://es.wikipedia.org/wiki/Archivo_de_la_Web_Espa%C3%B1ola)	27
Tabla 5. Selección de la muestra.	35
Tabla 6. Datos obtenidos agrupados por categorías.	36
Tabla 7. Promedio de archivabilidad de los sitios web de bibliotecas universitarias	44
Tabla 8. Relación de datos atendiendo a los criterios.	46
Tabla 9. Comparativa entre dos universidades de élite de China y España	55
Tabla 10- Comparativa de dos universidades intermedias China y España.	58
Tabla 11. Comparativa de dos universidades de grupo inferior de China y España.	61

Índice de Figuras

Figura 1. Miembros del IIPC. Fuente. Sitio Web: https://www.google.com/maps/d/viewer?mid=14Pe_dyH97jAKh1kBJZ0j5mfzGVY&femb=1&ll=-3.81666561775622e-14%2C-113.11824539999998&z=1 ...	15
Figura 2. Fases del proceso. Elaboración propia	18
Figura 3. Mapa de iniciativas de archivado Web. Fuente: Wikipedia. Disponible en: https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives#/media/File:Map_of_Web_archiving_initiatives_58632FRT.jpg	24
Figura 4. Dimensiones de la archivabilidad. Fuente: sitio de N. Taylor. https://nullhandle.org/web-archivability/index.html	30
Figura 5. Interfaz principal de la herramienta ArchiveReady (Fuente: Sitio web oficial de ArchiveReady, https://www.archiveready.com/)	31
Figura 6. Resultado de evaluación de archivabilidad de https://bib.us.es/ con la herramienta ArchiveReady (Fuente: ArchiveReady, https://www.archiveready.com/)	38
Figura 7. Evaluación de archivabilidad web de sitios de bibliotecas universitarias en España (Fuente: Elaboración propia con datos obtenidos mediante ArchiveReady, junio de 2025)	39
Figura 8. Evaluación de archivabilidad web de sitios de bibliotecas universitarias en China (Fuente: Elaboración propia con datos obtenidos mediante ArchiveReady, junio de 2025)	41
Figura 9. Mapa de distribución geográfica de las universidades chinas según grupo de estudio (Fuente: Mapa generado con apoyo de ChatGPT a partir de los datos proporcionados por el autor.)	42
Figura 10. Resultados de la evaluación de estándares web del sitio de la Biblioteca de Tsinghua	56
Figura 11. Informe de archivabilidad de la web de la Biblioteca de la UB	57
Figura 12. Detalle técnico de la evaluación de archivabilidad web del sitio de la Biblioteca de la Universidad Normal Central de China	59
Figura 13. Detalle técnico de la evaluación de archivabilidad web del sitio de la Biblioteca de la Universidad de Salamanca	60
Figura 14. Detalle técnico de la evaluación de archivabilidad web del sitio de la Biblioteca de La Universidad de Jimei	62
Figura 15. Detalle técnico de la evaluación de archivabilidad web del sitio de la Biblioteca de La Universidad Carlos III	63

Índice de Gráficas

Gráfica 1. Comparación de los indicadores de archivabilidad web en bibliotecas universitarias de España y China	44
Gráfica 2. Promedio general de archivabilidad web en universidades de España y China	45
Gráfica 3. Comparación de los indicadores de archivabilidad web en universidades del grupo de élite	47
Gráfica 4. Comparación de los indicadores de archivabilidad web en universidades del grupo intermedio	48
Gráfica 5. Comparación de los indicadores de archivabilidad web en universidades del grupo básico	49
Gráfica 6. Comparación de la archivabilidad general de los sitios web entre España y China	49
Gráfica 7. Comparación del nivel de accesibilidad web en bibliotecas universitarias de España y China	50
Gráfica 8. Comparación del nivel de cohesión estructural en los sitios web de bibliotecas universitarias	50
Gráfica 9. Comparación de la integridad de los metadatos en los sitios web de bibliotecas universitarias	51
Gráfica 10. Comparación del nivel de conformidad con los estándares en los sitios web de bibliotecas universitarias	51
Gráfica 11. Comparación de los indicadores de archivabilidad entre la Universidad de Barcelona (España) y la Universidad de Tsinghua (China)	55
Gráfica 12. Comparación de los indicadores de archivabilidad entre la Huazhong Normal University (China) y la Universidad de Salamanca (España)	58
Gráfica 13. Comparación de los indicadores de archivabilidad entre Jiangxi Normal University (China) y la Universidad Carlos III de Madrid (España)	61

Agradecimientos

Quiero expresar mi más sincero agradecimiento a la profesora María Manuela Moro Cabero, por su guía paciente, su apoyo constante y su confianza en mí desde el inicio de este trabajo. Sus palabras de ánimo y su orientación académica han sido fundamentales en todo este proceso.

A mi familia, gracias por estar siempre a mi lado, incluso en la distancia. Su comprensión, cariño y fe incondicional en mí me han dado la fuerza para seguir adelante en los momentos más difíciles.

También agradezco a mis amigos, por cada mensaje de ánimo, cada conversación que me devolvió la calma, y por recordarme que no estoy sola en este camino.

Este trabajo no solo es el cierre de una etapa académica, sino también el reflejo de todo el apoyo y amor que he recibido. A todos ustedes, gracias de corazón.

Salamanca, julio de 2025

0. Introducción

En este apartado se describen los antecedentes, el objeto, problema y objetivos de investigación, así como las dificultades encontradas, la metodología empleada y la estructura que presenta dicha memoria.

0.1 Antecedentes de la investigación

Con el desarrollo de la sociedad de la información, las páginas web se han convertido en un canal central para difundir información institucional, gestionar servicios y trámites, así como para obtener información de todo tipo. Su función ya no es solo mostrar información básica, sino que se ha convertido en un sistema para registrar conocimientos y desarrollar y posibilitar gestiones. Su uso se ha generalizado en áreas gubernamentales y no gubernamentales, alcanzando a todos los campos, tales como la educación, la ciencia y la cultura, guardando contenidos de gran valor patrimonial histórico-cultural.

Las bibliotecas universitarias son centros de servicio a sus usuarios y ciudadanos que conforman un ejemplo típico. Sus sitios web integran recursos académicos clave como materiales de cursos, logros de investigación, archivos históricos institucionales y anuncios oficiales. A través de ellas, se realizan variados trámites, de los cuales quedan registros en sus logs.

Sin embargo, a diferencia de la estabilidad física de los documentos impresos, las actualizaciones frecuentes de las páginas web pueden hacer que las versiones antiguas se pierdan o sean eliminadas. Si hay cierre de servidores, rediseño de sitios web o cambios de dominio, muchas páginas desaparecerán para siempre. Esto no solo crea vacíos en la historia digital, sino que también dificulta el uso sostenible y el rastreo de los datos. Por ello, el archivado de la web (capturar, guardar y conservar páginas web mediante tecnología para asegurar su acceso futuro) se ha convertido en un tema urgente en la era digital. Fundamentalmente, para dejar rastro de las actividades e informaciones, dada la complejidad de las páginas y su volatilidad.

Actualmente, algunos países ya han establecido sistemas completos: Estados Unidos fue pionero en 1996 al crear el proyecto global "Internet Archive", mientras que la Unión Europea impulsa programas colaborativos entre países para archivar contenido web; España, por su parte, mediante el *Real Decreto 635/2015*¹, autorizó

¹ Boletín Oficial del Estado. (2015). Real Decreto 635/2015, de 10 de julio, por el que se regula el depósito legal de publicaciones en línea [Royal Decree 635/2015 of July 10th, regulating the legal deposit of online publications]. BOE-A-2015-8008. <https://www.boe.es/eli/es/rd/2015/07/10/635>

a la Biblioteca Nacional a capturar páginas web sin restricciones y obligó a los editores a colaborar, incluyendo el acceso con contraseña cuando sea necesario. No siempre, son las bibliotecas sedes del depósito legal, las guardianas de estos archivos. También los Archivos pueden almacenar las memorias institucionales web, tal y como es el caso de Colombia. En España, igualmente, existen otras iniciativas vinculadas a comunidades autónomas, como son las de la C.A. de Euskadi y la de Cataluña. China aún está desarrollando su sistema de archivo web, estudiando cómo regularlo por ley.

Archivar webs es difícil por problemas técnicos como código cambiante, lo que a veces da resultados incompletos. Por eso, evaluar la "capacidad de archivado" (es decir, la viabilidad técnica de guardar una página web de forma completa y precisa) se ha convertido en un paso clave. Entre los principales obstáculos observados ante el archivado se enumeran los siguientes: rotura de enlaces, ausencia de un mapa de sitio, empleo de programas que impiden el rastreo de las páginas, ausencia de normalización, volatilidad de contenidos, etc.

Los estudios existentes se centran, principalmente, en gobiernos, instituciones nacionales y grandes plataformas de noticias, prestando escasa atención a los sitios web de bibliotecas universitarias. Por regla general, se orientan hacia sus contenidos, su accesibilidad, sus prestaciones, su selección, sin tratar su grado de archivabilidad. Este tipo de sitios web en realidad tiene un valor único para la investigación; En primer lugar, presentan una alta densidad de información y un valor académico notable. En segundo término, se conforman de una arquitectura técnica diversa, que puede obstaculizar parte de su archivado, como el uso de interacciones con JavaScript o la carga dinámica de contenido. En tercer lugar, se actualizan con frecuencia y tienen estructuras de página muy variadas. Sin una evaluación específica, será difícil diagnosticar sus riesgos de preservación. Los estudios sobre capacidad de archivado no solo permiten identificar fallos técnicos, sino que también posibilitan comparar diferencias en estrategias de diseño web y políticas de preservación entre países, aportando bases para reducir la brecha digital.

0.2 Objeto, pregunta y objetivos de investigación

Este estudio se centra en la capacidad de archivado de los sitios web de los servicios bibliotecarios, tomando como estudio de caso las bibliotecas universitarias de China y España. Es decir, se trata de analizar si, en su caso, pueden preservarse de manera completa y fiable.

Se pretende responder a la siguiente pregunta de investigación: ¿Qué grado de dificultad presentan las páginas web de los servicios bibliotecarios de universidades destacadas en rankings para su archivado?

La relevancia de la respuesta radica en aspectos clave: busca identificar las dificultades concretas en el proceso de preservación de estas páginas, entendidas como conjunto de contenidos con tecnología compleja que son difíciles de capturar, o mediatizadas por una ausencia de información descriptiva esencial y de normalización.

En este sentido, se formula el siguiente objetivo general:

Objetivo General: Analizar la problemática que presentan las páginas web en los servicios bibliotecarios universitarios.

Para su logro, se formularon los siguientes objetivos específicos:

Objetivo específico 1- Contextualizar y conceptualizar la página web para entender su complejidad y posible problemática.

Objetivo específico 2- Analizar el proceso de archivado de páginas web y las herramientas empleadas.

Objetivo específico 3- Analizar en un entorno dado la problemática que se presenta, como es el caso de los servicios bibliotecarios universitarios, en China y España.

Objetivo específico 4-Explorar soluciones prácticas, a partir del estudio de caso comparado.

El estudio pretende aportar soluciones de mejora, a partir de los resultados obtenidos, ayudando a universidades con distintos recursos a proteger su patrimonio académico de manera más efectiva, al desvelar la problemática que presentan.

Al comparar la situación real de preservación web en diferentes tipos de universidades de España y China, este estudio busca destacar la influencia decisiva de las políticas y leyes en los resultados de conservación, así como los desafíos prácticos causados por diferencias en recursos, tales como pueden ser las limitaciones económicas y tecnológicas en universidades públicas, o la falta de atención por parte de las administraciones.

El estudio pretende proporcionar una base empírica valiosa para que China desarrolle o perfeccione futuras regulaciones, estándares técnicos y guías de trabajo

relacionadas. Resulta crucial que este estudio atienda las necesidades de preservación del conocimiento en universidades de todos los niveles, buscando reducir la 'brecha digital' causada por la desigualdad de recursos. Esto permite que no sólo las universidades élite en rankings, sino también, otras instituciones públicas conserven adecuadamente sus recursos académicos, garantizando que estos frutos del conocimiento - acumulados con sabiduría - permanezcan accesibles de manera duradera, segura y equitativa para profesores, estudiantes e investigadores. Así se convertirán en una 'memoria colectiva' que resista el paso del tiempo y no sea arrasada por las olas digitales.

0.3 Metodología de investigación

La investigación es de naturaleza descriptiva y empírica comparativa. El marco teórico, desde un enfoque descriptivo, se establece fundamentado en la literatura el contexto, concepto, proceso y técnica del archivado de la Web para exponer la problemática y complejidad que representa la preservación de los sitios web. Se parte de los conocimientos adquiridos en la asignatura Preservación Digital y el tema concreto otros modelos de almacenamiento, donde se introduce al estudiante en el archivado de la web y se proporciona literatura científica sobre esta temática. Al respecto, se han tenido en consideración, además de publicaciones periódicas, la formación en línea que aporta el Consorcio Internacional de Preservación de Internet, en el que se aborda en diferentes temas el proceso de archivado de la web (IIPC; 2025)² así como la monografía sobre preservación digital de la Web (Gomes et al.; 2021)³.

Desde un enfoque empírico, a su vez, se evalúan determinados sitios web del ámbito bibliotecario universitario para comprobar el grado de archivado que presentan y sus dificultades.

A su vez, se aprovecha el estudio de caso para realizar un estudio comparado entre los sitios de universidades españolas y chinas.

Por consiguiente, se establece una muestra para la selección siguiendo dos índices que nos proporcionarán universidades de élite y universidades en estados

² *IIPC Beginner's Training: How to Use*. International Internet Preservation Consortium. Recuperado de <https://netpreserve.org/web-archiving/training-materials/>

³ Gomes, D., Demidova, E., Winters, J., & Risse, T. (Eds.). (2021). *The Past Web: Exploring Web Archives*. Springer Nature Switzerland. <https://doi.org/10.1007/978-3-030-63291-5>

intermedios dentro de los propios índices. Estos son: el índice ARWU y el índice de Wushulian.

Para el análisis de los sitios web, se planea utilizar la herramienta existente ArchiveReady para evaluar por separado la calidad de archivado de las páginas web de los servicios de bibliotecas universitarias en China y en España. Esta herramienta se presenta con posterioridad y ella permite disponer de una evaluación total a partir de la media de cuatro criterios: accesibilidad, normalización, cohesión, metadata.

Se realizará a partir de una muestra establecida para comparar de manera más completa la situación de ambos países. Se tomará como referencia el ranking académico mundial de universidades ARWU y el ranking de Wushulian. Se seleccionarán las 10 universidades mejor clasificadas de China y de España (grupo de élite), así como 10 universidades de cada país que estén en el rango 500-700 a nivel mundial (grupo intermedio), y otras 10 universidades que se sitúen entre los puestos 700 y 900 (grupo inferior), para realizar una comparación entre los dos grupos.

El análisis se centrará en comparar las principales diferencias en la capacidad de archivado de sitios web de bibliotecas universitarias entre China y España, evaluando dimensiones clave como la integridad en la captura de contenido dinámico y el cumplimiento de estándares en metadatos de archivo. Posteriormente, se interpretarán estas diferencias a la luz de los contextos legales y políticos de ambos países en materia de preservación digital de información y protección del patrimonio cultural.

Atendiendo a los resultados del análisis y la comprensión de las políticas, se proponen recomendaciones prácticas para mejorar la preservación de sitios web de bibliotecas en ambos países. Estas sugerencias buscan ayudar a las universidades chinas y españolas a elevar sus estándares de conservación digital, garantizando que los valiosos recursos académicos permanezcan accesibles y utilizables a largo plazo.

0.4 Dificultades observadas durante la investigación

La investigación se ha visto limitada por los plazos, debido a que la asignatura de Preservación Digital se imparte en el segundo semestre, aspecto importante para obtener una visión general de la misma y disponer de conocimientos básicos para abordar la preservación como tal en un recurso determinado.

Una segunda dificultad en la investigación es el número ingente de páginas Web, limitando el estudio de caso a un número exploratorio que posibilite determinar

algunos comportamientos y comparar resultados en dos territorios bien diferenciados, aunque con idéntica preocupación: mantener el conocimiento cultural de los entornos académicos en el tiempo y conservarlos para posibles investigaciones futuras.

Una tercera dificultad ha sido la derivada de la herramienta empleada, la cual aporta 5 criterios informativos. Así mismo, facilita el análisis con detalle de algunos aspectos a partir de la información que incluye. Por los objetivos fijados simplemente se aportan algunos datos sobre dos sitios web, uno de China y otro de España, seleccionados aleatoriamente para verificar el grado de información facilitada.

Una cuarta dificultad es disponer del tiempo necesario para abordar esta temática desde un enfoque empírico más integrador, lo cual no siempre ha sido posible.

La quinta dificultad se supedita a la existencia de web disponible y accesible para la herramienta empleada de análisis.

Todo ello, delimita la investigación, lógicamente, aunque no resta importancia a los resultados como estudio exploratorio.

0.5 Estructura del ensayo

Esta tesis se estructura en introducción y conclusiones a la investigación, marco teórico y marco empírico. Incluye, además, referencias y anexos. La introducción presenta el contexto, el problema y los objetivos de la investigación, así como se introduce en la metodología y en las dificultades observadas. Se enuncia la estructura del estudio.

El marco teórico, incluye dos capítulos, uno primero, destinado a contextualizar y conceptualizar un sitio web, su evolución aportando algunos datos sobre su complejidad. El capítulo segundo detalla el proceso de su archivado y la legislación existente al respecto, así como las herramientas más comunes empleadas.

El marco empírico, se estructura en tres capítulos. Un capítulo tercero en el que, se resume el estado actual de la investigación sobre el archivado web, incluyendo las herramientas para su evaluación. El capítulo cuarto informa sobre la selección de la muestra, la recogida de datos y los datos obtenidos. El capítulo cinco muestra los resultados del estudio de modo general, atendiendo a cada criterio para España, China, así como de modo comparado. El capítulo seis analiza varios casos representativos de universidades, con detalle para disponer de una triangulación sobre la investigación.

Finalmente, en conclusiones se responde a los objetivos fijados y se proponen recomendaciones.

Se establecen las referencias bibliográficas y en Anexos, el modelo de representación de las dificultades que aportan las herramientas empleadas.

Marco teórico: Páginas y archivado Web

Capítulo 1 Contextualización de la página Web.

En este capítulo se define, contextualiza y aportan datos sobre la evolución de las páginas web con el objetivo de que el lector pueda comprender el contexto de desarrollo de una página web.

1.1 ¿Qué es una página web?

En la norma UNE-ISO 2789: 2024 (p. 16) ⁴se define página Web como:

“3.2.27 sitio web de la biblioteca: Dominio único en Internet que consiste en una colección de páginas web publicadas por una biblioteca para proporcionar acceso a sus servicios y recursos”.

Mediante nota se especifica la interconexión de páginas, mediante enlaces hipertextuales, tal y como se indica: “NOTA 1 Las páginas del sitio web de una biblioteca generalmente están interconectadas mediante enlaces hipertextuales”.

A su vez, en notas 2 y 3 se establecen las exclusiones para colección digital, recursos gratuitos de internet, servicios web ofrecidos bajo el dominio de la biblioteca, pero cuyos beneficiarios serán otras organizaciones.

En España, la Ley 8 de 4 de mayo de 2022 retoma la definición aportada en el art. 2 del RD 635/ 2015, donde se regulan las publicaciones en línea objeto de Depósito Legal y actualiza el art. 3 de la Ley 23/2011 de Depósito Legal. Se observa que los sitios web son considerados publicaciones en línea, siendo estas definidas como:

“Información o contenido de cualquier naturaleza difundido en un soporte electrónico no tangible, archivado en un formato determinado y susceptible de identificación y tratamiento diferenciado, que sea objeto de difusión. Los sitios web se consideran publicaciones en línea”.

El art. 2 del RD 635 incluye la definición de sitio web como: “Punto de acceso electrónico formado por una o varias páginas agrupadas en un dominio de Internet”, siendo definido en dicho artículo un sitio Web accesible como: “aquel que tiene una URL pública y, por tanto, es visible para todos los usuarios de Internet”.

⁴ Asociación Española de Normalización. (2024). *UNE-ISO 2789:2024: Información y documentación – Estadísticas internacionales de bibliotecas*. Madrid: UNE. <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma/?c=N0072856>

El marco legal en España y China establece una sustancial diferencia en su conceptualización. La definición legal de las páginas web como publicaciones digitales constituye la base de la práctica archivística. España, mediante la Ley de Depósito Legal 23/2011, su RD 635/2015 y la actualización de la Ley como Ley 8/2022, ha incluido expresamente los 'contenidos en línea públicamente accesibles con URL fija' dentro del ámbito de preservación obligatoria de la Biblioteca Nacional (BNE), otorgándoles estatus de patrimonio cultural equivalente al de las publicaciones impresas.

Lourdes Santos de Paz, edita en 2018 ⁵en la Revista General de la Información un artículo sobre el Depósito Legal de las publicaciones en línea, señalando los desafíos para su desarrollo en España, entre los que destaca:

En contraste, En China la norma vigente GB/T 5271.20-1994 «Terminología de tecnologías de la información», solo se ofrece una descripción conceptual del «sistema de hipertexto». Desde el punto de vista jurídico, el artículo 47 de la Ley de Seguridad de la Red exige expresamente que los operadores de redes deben actuar de inmediato y notificar cuando detecten información ilegal, lo que refleja una orientación hacia la supervisión administrativa de los contenidos en línea (Comité Permanente del Congreso Nacional del Pueblo, 2016). Esta divergencia en el marco legal ha influido profundamente en el diseño posterior de los mecanismos de archivado en ambos países.

La norma UNE- ISO 2789:2024 define la voz 3.1 “biblioteca académica” como: “Biblioteca cuya función principal es satisfacer las necesidades de información en el ámbito del aprendizaje y la investigación. (p.8)

En nota (NOTA 1) se incluyen dos tipos de bibliotecas, aquellas de instituciones de educación superior y bibliotecas generales de investigación.” (2024, p.8)⁶.

Para los objetivos de este estudio se tienen en cuenta las definiciones de la norma, tanto de página web como de biblioteca académica. No obstante, todas ellas responden al dominio universitario para la biblioteca universitaria.

⁵ Santos de Paz, L. (2018). *Depósito legal en línea: Desafíos para su desarrollo en España*. **Revista General de Información y Documentación**, 28(1), 11–22. <https://doi.org/10.5209/RGID.60814>

⁶ Asociación Española de Normalización. (2024). *UNE-ISO 2789:2024 Información y documentación. Estadísticas internacionales de bibliotecas* (8ª ed.). UNE. <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma/?c=N0072521>

1.2 Datos sobre el desarrollo de sitios Web, su vigencia y principales riesgos a los que están expuestos.

El crecimiento exponencial del tamaño de Internet ha intensificado los desafíos para la preservación del patrimonio digital.

Aunque la cantidad exacta de sitios web cambia constantemente, existen más de 1.200 millones en la red mundial (1.180.650.484 según la Encuesta de Servidores Web de Netcraft en febrero de 2025,⁷ frente a 1.086.916.398 en febrero de 2024). Esto representa una expansión explosiva comparado no sólo anualmente, sino si fijamos la comparación con el primer sitio web creado en 1991.

Total Number of Websites by Year			
Year	All Websites	Active Websites	Active Websites %
January 2025	1,161,445,625	194,973,562	16.79%
January 2024	1,079,154,539	192,375,760	17.83%
January 2023	1,132,268,801	202,900,724	17.92%
January 2022	1,167,715,133	198,988,100	17.04%
January 2021	1,197,982,359	199,533,484	16.66%
January 2020	1,295,973,827	189,000,000	14.58%
January 2019	1,518,207,412	182,185,876	12.00%
January 2018	1,805,260,010	171,648,771	9.51%
January 2017	1,800,047,111	172,353,235	9.57%
January 2016	906,616,188	170,258,872	18.78%
January 2015	876,812,666	177,127,427	20.20%
January 2014	861,379,152	180,067,270	20.90%
January 2013	629,939,191	186,821,503	29.66%
January 2012	582,716,657	182,441,983	31.31%
January 2011	273,301,445	101,838,083	37.26%
January 2010	206,741,990	83,456,669	40.37%
January 2009	185,497,213	71,647,887	38.62%

Tabla 1. Relación anual de sitios Web y de sitios web activos. (Fuente: Siteefy, <https://siteefy.com/how-many-websites-are-there/>)

El hecho de registrar esta explosión, tal y como se observa en la tabla 1, no implica su vigencia en el tiempo. De facto, la vida útil promedio de un sitio web es de menos de tres años, y los recursos del sector educativo representan hasta el 18 % (Internet Live Stats, 2024). En la tabla 1 se incluyen los sitios Web activos, dando fe en la relación cronológica al rápido crecimiento de Web inactivas. Los datos se muestran inversamente proporcionales, mientras que en 2009 se contabilizan más de 185

⁷ Siteefy. (2025, March 2025). *How many websites are there?* Retrieved June 28, 2025, from <https://siteefy.com/how-many-websites-are-there/>

millones de sitios web, de los cuales sólo un 38.62% permanecen activos, en el 2025, son 1.165 millones de sitios web, representando un 16,79% los sitios activos.⁸

Además, el contenido de un sitio web puede variar en cuestión de segundos si nos detenemos en canales de comunicación y redes sociales incorporados en su sitio, o incluso en las tramitaciones que pueden ser realizadas, como es el caso de consultas de catálogos, de informaciones, etc.

Diversos estudios han evidenciado la rápida caducidad de los recursos web tras su publicación. Por ejemplo, según Archivo.pt, aproximadamente el 80 % de las páginas web dejan de estar disponibles en su forma original un año después de su publicación⁹. SalahEldeen y Nelson (2012) ¹⁰estiman que el 11 % de los recursos compartidos en redes sociales desaparecen tras el primer año, y este porcentaje se eleva al 27 % al cabo de dos años y medio. Asimismo, Klein et al. (2014)¹¹ señalan que cerca del 13 % de las referencias web incluidas en artículos científicos dejan de estar accesibles después de 27 meses.

Esta contradicción entre la gran escala y la fragilidad indica que: sin un archivado sistemático, los contenidos de alto valor en sitios académicos como bibliotecas universitarias (bases de datos de cursos, repositorios interactivos de investigaciones) enfrentarán pérdidas irreversibles, creando vacíos en la memoria digital.

Las primeras instantáneas de la Web datan de 1996, sin embargo, hemos señalado como la primera página web se crea en 1991. El periodo de un lustro entre 1991 y 1996 no se conserva desafortunadamente por lo que los primeros 5 años de dominio web ha desaparecido.

En el contexto universitario se observa una dificultad destacable: la vida útil de las páginas web como soportes de conocimiento se está acortando drásticamente. El

⁸ Siteefy. (2025, March 2025). *How many websites are there?* Retrieved June 28, 2025, from <https://siteefy.com/how-many-websites-are-there/>

⁹ Archivo.pt. (s.f.). *Preserve su sitio: Panfleto informativo*. Fundação para a Ciência e a Tecnologia. Recuperado de <https://sobre.arquivo.pt/wp-content/uploads/Panfleto-ENG.pdf>

¹⁰ SalahEldeen, H. M., & Nelson, M. L. (2012). *Losing My Revolution: How Many Resources Shared on Social Media Have Been Lost?* arXiv. <https://doi.org/10.48550/arXiv.1209.3026>

¹¹ Klein, M., Van de Sompel, H., Sanderson, R., Shankar, H., Balakireva, L., Zhou, K., & Tobin, R. (2014). *Scholarly context not found: One in five articles suffers from reference rot*. PLOS ONE, 9(12), e115253. <https://doi.org/10.1371/journal.pone.0115253>

análisis de datos de Internet Archive revela que más del 40% 12de las solicitudes de acceso a URI dependen de cookies o JavaScript para renderizar correctamente el contenido, lo que está directamente relacionado con los desafíos de captura de contenido dinámico.

Esta crisis es especialmente grave en el ámbito universitario, caracterizado por una alta heterogeneidad tecnológica. Los informes de Archive-It (que incluye numerosas bibliotecas universitarias) señalan constantemente que los "fallos en la captura de contenido dinámico" representan una de las categorías más frecuentes en las solicitudes de soporte técnico (generalmente entre el 30%-50% o más)¹³, aunque los porcentajes exactos varían según el año. Las herramientas tradicionales de rastreo no pueden capturar completamente este contenido (como bases de datos de cursos o gráficos interactivos).

Como demostraron Park y Tosaka (2010)¹⁴ en su estudio sobre repositorios digitales estadounidenses, la creación de metadatos presenta una amplia fragmentación en el uso de campos, la coherencia semántica y el control de calidad. Muchas instituciones, debido a la falta de estándares claros, recursos técnicos y personal adecuado, generan estructuras de metadatos muy dispares, lo que limita seriamente la trazabilidad a largo plazo y la interoperabilidad de los objetos digitales.

Otro riesgo específico es su accesibilidad a largo plazo: El análisis de referencias web en publicaciones académicas revela que aproximadamente el 20% 15de las citas a páginas web (incluyendo dominios .edu) sufren "descomposición de citas" (desaparición o alteración significativa del contenido) dentro de los 2 años posteriores a su publicación. Los dominios .edu muestran una tasa de descomposición ligeramente más lenta que el promedio, pero aún así presentan un problema significativo.

¹² Internet Archive. (2022, November 2). **Archiving and Data Services Update at the 2022 Archive-It Virtual Partner Meeting** [Video]. Internet Archive. <https://archive.org/details/archiving-and-data-services-update-at-the-2022-virtual-partner-meeting>

¹³ AlNoamany, Y., Weigle, M. C., & Nelson, M. L. (2013). *Access patterns for robots and humans in web archives*. En *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pp. 339–348. <https://doi.org/10.1145/2467696.2467722>

¹⁴ Park, J.-R., & Tosaka, Y. (2010). *Metadata creation practices in digital repositories: A survey*. *Cataloging & Classification Quarterly*, 48(8), 716–739. <https://doi.org/10.1080/01639374.2010.508711>

¹⁵ Jones, S. M., Klein, M., & Van de Sompel, H. (2021). *Robustifying links to combat reference rot*. *Code4Lib Journal*, 50. <https://journal.code4lib.org/articles/15509>

También es importante destacar la falta de normativas y la desigualdad en los recursos. Un vacío legal en cuanto a permisos y derechos. En España, el *Real Decreto 635/2015* estableció un mecanismo nacional obligatorio de recolección web que permite capturar incluso contenido encriptado sin restricciones de acceso, potestad mantenida en la Ley 8/2022. 16A nivel global, especialmente en países en desarrollo e instituciones con recursos limitados, desarrollar capacidades sistemáticas de preservación digital (incluyendo archivo web) sigue siendo un gran desafío. Muchas instituciones permanecen en fases "piloto" o "por proyectos", careciendo de marcos obligatorios a nivel nacional e inversión sostenible.

Se observa una marcada desigualdad en la asignación de recursos: evidencia empírica muestra que las universidades de élite chinas han mitigado parcialmente las deficiencias institucionales mediante un modelo híbrido (envío activo de metadatos + captura pasiva de archivos adjuntos). En un estudio sobre 42 universidades chinas de "clase mundial" (proyecto Double First-Class), solo el 23,8% (10 instituciones) contaban con proyectos de archivo web que incluían normativas institucionales, personal dedicado y mecanismos periódicos de rastreo. Entre universidades no pertenecientes a este grupo (muestra de 100), la tasa de implementación era inferior al 8% (Liu & Zhang, 2021, pp. 840-841)¹⁷. La escasez de recursos y la falta de concienciación han creado un círculo vicioso.

1.3 El Consorcio de Preservación Internacional de Internet-IIPC

Este Consorcio destinado a Preservar la Web se forma en julio de 2003 promovido por la Biblioteca Nacional de Francia, con 12 miembros participantes con el objetivo de desarrollar proyectos mediante grupos de trabajo. 18

El International Internet Preservation Consortium (IIPC), como actor clave en la gobernanza, impulsa prácticas de archivado estandarizadas mediante la colaboración de instituciones de patrimonio cultural de 35 países. Sus funciones principales incluyen:

¹⁶ UNESCO. (2021). PERSIST Guidelines for the selection of digital heritage for long-term preservation (2nd ed.). UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000244280>

¹⁷ Liu, Y., & Zhang, T. (2021). Web archiving practices in Chinese academic libraries: An empirical study based on the "Double First-Class" initiative. *Library Hi Tech*, 39(3), 832-848. <https://doi.org/10.1108/LHT-01-2020-0011>

¹⁸ International Internet Preservation Consortium. (s.f.). *About Us*. Recuperado de <https://netpreserve.org/about-us/>

establecer normas técnicas para la recolección, construir archivos colaborativos internacionales, promover el respaldo legal y fomentar el uso académico.

La relación de miembros se registra en la figura siguiente:



Figura 1. Miembros del IIPC. Fuente. Sitio Web:

https://www.google.com/maps/d/viewer?mid=14Pe_dyH97jAKh1kBJZ0j5mfzGVY&femb=1&ll=-3.81666561775622e-14%2C-113.11824539999998&z=1

Se destaca entre sus miembros la Biblioteca Nacional de China, así como la de España.

Cabe destacar que, entre los países miembros de dicho consorcio IIPC, solo 20 han completado una legislación específica sobre el tema. España, a través del Real Decreto 635/2015, ha establecido un sistema de recolección obligatoria que permite superar restricciones y capturar contenidos cifrados; mientras que China aún se encuentra en proceso de legislación (IIPC). Esta diferencia en las etapas legislativas convierte a China y España en un grupo de comparación ideal en los ámbitos legal y político.

En la tabla siguiente se incluye la información que aporta el Consorcio IIPC, sobre países con legislación, con directrices específicas o carentes de ella. En este caso, para España y China.

País	Legislación	Aspectos regulados
China	No dispone de información.	
España	Real Decreto por el que se regula el Depósito Legal de las publicaciones en línea de 2015 Ley de Depósito Legal de 2011 Modificación de la Ley de Depósito Legal de 2022	Fecha de inicio de recopilación desde 2009 , de acuerdo con la Ley General de Depósito Legal. 2011 y 2022. Amparo legal: desde 2015, de acuerdo con el Real Decreto 635 que regula el depósito legal de publicaciones en línea. Competencias: Las competencias en materia de depósito legal son compartidas entre la Biblioteca Nacional de España (BNE) y las bibliotecas autonómicas. Autorizaciones: permite la recopilación, sin necesidad de permiso. Conservación: La legislación sobre derechos de autor permite la copia para su conservación. Restricciones acceso: En caso de restricciones de la editorial, la Biblioteca puede exigir contraseñas o acceso a contenido protegido por contraseña. El depósito también es posible en estos casos. El acceso presencial debe estar disponible en la BNE y las bibliotecas regionales.

Tabla 2. Comparación de la legislación sobre el archivo web entre China y España

El sitio Web del Consorcio IIPC incluye rica información para sensibilizar, formar, informar y trabajar.

En su página Home, (<https://netpreserve.org/>) se informa sobre los países e instituciones miembros, sus grupos de trabajo pasados y activos para conformar

materiales formativos y promover proyectos de colaboración, así como reuniones internacionales y asambleas donde se exponen los avances.¹⁹

Muy interesante a su vez, es la información sobre eventos y sobre formación. Así, mediante el enlace a sus materiales (<https://netpreserve.org/web-archiving/training-materials/>) se dispone de un auténtico curso formativo introductorio con el desarrollo siguiente de 8 lecciones centradas en diapositivas que responden a interrogantes tales como ¿Por qué archivar sitios web?, ¿cuáles son las principales tecnologías, riesgos y beneficios en cada fase del proceso de archivar, así como redactar una política web.

Se incluyen además otros cursos más avanzados y enlaces a experiencias.²⁰

¹⁹ IIPC. (s.f.). *International Internet Preservation Consortium*. Recuperado de <https://netpreserve.org/>

²⁰ IIPC. (s.f.). *Training Materials*. International Internet Preservation Consortium. Recuperado de <https://netpreserve.org/web-archiving/training-materials/>

Capítulo 2 Proceso y estado actual del archivado web

En este capítulo se describe el proceso de archivado de un sitio web, su complejidad, el archivo de la web internacional denominado Internet Archive, así como la legislación básica empleada en China y España. Finalmente se presentan algunas de las herramientas empleadas en ambos países para el proceso de archivado.

2.1 Proceso de archivado de los sitios web y análisis de su complejidad.

El archivado web es un proceso sistemático diseñado para capturar, preservar y permitir el acceso a largo plazo a versiones históricas de páginas web. Este flujo generalmente consta de cuatro fases interrelacionadas que son representadas como sigue:



Figura 2. Fases del proceso. Elaboración propia

La fase de selección supone identificar sitios web, secciones o páginas específicas con valor de preservación a largo plazo. La estrategia debe considerar el valor del contenido (como valor académico, cultural o histórico), la frecuencia de actualización y el riesgo de desaparición. Los sitios web de bibliotecas universitarias, al albergar recursos docentes, logros de investigación y conocimiento institucional, son considerados de máxima prioridad para archivado.

La fase de recolección implica el uso de herramientas profesionales de rastreo para capturar páginas web objetivo y sus recursos asociados (HTML, CSS, JS, imágenes, videos, etc.). Esta es la fase con mayores desafíos técnicos.

La tercera fase es la de preservación, lo cual implica almacenar los contenidos capturados (normalmente encapsulados en formatos estándar como archivos WARC

[Web ARChive], La información sobre la capacidad de almacenamiento de este formato está disponible en la iniciativa de estudio de formatos impulsada por la Library of Congress y la IIPC, la cual aporta contenidos sobre los factores de sostenibilidad, calidad del formato, etc.)²¹ en sistemas confiables de repositorios digitales, implementando estrategias de preservación a largo plazo que incluyen: migración de formatos, verificación de integridad y recuperación ante desastres, entre otros.

Finalmente, la fase de acceso trata de proporcionar a los usuarios métodos para buscar, navegar y utilizar páginas web archivadas. Los permisos y modalidades de acceso están sujetos a restricciones de derechos de autor, privacidad, limitaciones técnicas y regulaciones legales.

Se especifican seguidamente algunos de los aspectos que nos ayudan a comprender la complejidad del proceso.

1. La complejidad en el diseño de estrategias y el dilema en la valoración de contenidos.

La formulación de estrategias efectivas de archivado enfrenta inicialmente complejos juicios de valor. Debemos evaluar integralmente múltiples dimensiones del contenido: valor académico, escasez cultural, relevancia histórica y su riesgo de desaparición (por ejemplo, los recursos docentes universitarios, con sus frecuentes actualizaciones y fuerte temporalidad, son extremadamente vulnerables a la pérdida). Sin embargo, estos criterios suelen ser difíciles de cuantificar y sus ponderaciones requieren ajustes dinámicos según cambios temporales y demandas sociales. Más complejo aún es que, en la práctica, los recursos institucionales (especialmente en bibliotecas universitarias) -como personal, tecnología y presupuesto- son siempre limitados. Este dilema de 'qué seleccionar y qué descartar' constituye el núcleo crítico del diseño estratégico.

2. Cuellos de botella y desafíos en la implementación técnica.

En el aspecto técnico, la naturaleza dinámica de las páginas web modernas plantea grandes obstáculos. Numerosos sitios dependen de JavaScript y AJAX para funciones interactivas (como scroll infinito o actualizaciones en tiempo real), lo que impide a los rastreadores tradicionales renderizar completamente el contenido. Para capturar estos elementos dinámicos, se requiere emplear

²¹ Library of Congress. (s.f.). *Sustainability of Digital Formats*: WARC. Recuperado de <https://www.loc.gov/preservation/digital/formats/fdd/fdd000236.shtml>

tecnologías como navegadores headless, pero esto incrementa significativamente el consumo de recursos computacionales, el tiempo de procesamiento y la tasa de fallos. Otro problema clave son las dependencias asociativas: la representación íntegra de una página exige todos sus recursos vinculados (HTML, CSS, JavaScript, imágenes, videos, etc.). No obstante, numerosos recursos externos pueden volverse inaccesibles por múltiples causas: servidores de origen inactivos, restricciones de acceso (como protección anti-hotlinking de CDN) o prohibiciones en robots.txt, lo que genera archivos incompletos. Adicionalmente, la preservación a largo plazo de recursos multimedia plantea desafíos excepcionales. Formatos complejos (como modelos 3D interactivos o streaming) dependen de códecs y entornos de reproducción específicos. Con el tiempo, estos formatos pueden quedar obsoletos, obligando a las instituciones archivísticas a invertir continuamente en migraciones periódicas de formato, incluso cuando los recursos ya están encapsulados en archivos estándar como WARC.

3. La fiabilidad de la preservación a largo plazo y la presión de los costos.

Garantizar la accesibilidad a largo plazo del contenido archivado es en sí mismo un sistema de ingeniería complejo. Esto requiere una infraestructura de almacenamiento altamente confiable que incluya almacenamiento redundante, copias de seguridad distribuidas y mecanismos completos de recuperación ante desastres, para resistir riesgos como fallas de hardware o desastres naturales. La gestión de la integridad de los datos es igualmente crucial, necesitando verificaciones periódicas de checksum en los archivos WARC para prevenir corrupción silenciosa de datos (Bit Rot), además de mantener una cadena de metadatos rigurosa que garantice la trazabilidad del origen y los cambios del contenido. Sin embargo, estas medidas de protección conllevan costos continuos elevados. Particularmente para archivos universitarios que contienen abundantes recursos multimedia como videos educativos y grabaciones de conferencias, los costos de almacenamiento, gestión y mantenimiento a largo plazo de grandes volúmenes de datos representan una pesada carga económica. Como señala el 'Informe de Desarrollo de Bibliotecas Universitarias de China 2022'²², el crecimiento de los costos de almacenamiento se ha convertido en uno de los

²² Comité Directivo del Trabajo de Información y Bibliotecas Universitarias del Ministerio de Educación de China. (2023). *Informe sobre el desarrollo de las bibliotecas universitarias en China en 2022*. Comité Directivo del Trabajo de Información y Bibliotecas Universitarias del Ministerio de Educación de China. Recuperado de <http://www.scal.edu.cn/sites/default/files/attachment/tjpg/2022%E5%B9%B4%E9%AB%98%E6%A0%A1%E5%9B%BE%E4%B9%A6%E9%A6%86%E5%8F%91%E5%B1%95%E6%8A%A5%E5%91%8A.pdf>

principales factores que limitan la sostenibilidad de los proyectos de preservación digital en muchas bibliotecas.

4. La contradicción profunda entre el acceso público y el cumplimiento legal.

Finalmente, permitir el acceso público o autorizado a los contenidos archivados también está plagado de complejidades. Los problemas legales y de derechos de autor son el primer obstáculo: materiales con copyright (como artículos académicos no autorizados), páginas con información personal privada, y las diferencias regulatorias entre países (por ejemplo, la Ley de Depósito Legal española vs las regulaciones chinas sobre contenido en internet) pueden impedir la apertura legal al público o restringir severamente los permisos de acceso. Técnicamente, recrear la funcionalidad y experiencia original de las páginas archivadas a menudo requiere emular su entorno de ejecución original (incluyendo bibliotecas JavaScript o *frameworks* específicos), pero estas tecnologías pueden quedar obsoletas e incompatibles con navegadores modernos, resultando en una experiencia de acceso deficiente o falta de funcionalidades. Para recursos universitarios, surge una paradoja destacable en el control de accesos: muchos recursos académicos valiosos (como bases de datos o contenidos de repositorios institucionales) suelen estar protegidos por autenticación IP o sistemas federados como Shibboleth. Durante el archivado, es necesario superar estos 'muros de permisos' para capturar el contenido; pero al proporcionar acceso, se debe recrear lógica de control similar para cumplir con licencias y derechos de autor (como se documenta en informes de CALIS), creando una contradicción operativa.

2.2 Internet Archive y su *wayback machine*

En la lucha contra la fugacidad de la información en Internet, la organización sin fines de lucro Internet Archive (IA) y su emblemático servicio Wayback Machine (WM) desempeñan un papel crucial. Fundada en 1996, IA tiene la ambiciosa misión de construir una "Biblioteca de Alejandría digital" que preserve el patrimonio cultural humano y ofrezca acceso universal. Según datos públicos recientes, su vasto repositorio no solo alberga cientos de miles de millones de capturas de páginas web (disponibles a través de WM), sino que también incluye millones de libros, audios, videos y otros recursos multimedia, con una capacidad de almacenamiento en constante crecimiento basada en una arquitectura distribuida personalizada.

Como servicio más conocido de IA, el valor central de Wayback Machine radica en su capacidad de recuperación histórica de páginas web. Los usuarios pueden introducir una URL y rastrear las capturas históricas de esa página desde 1996, con precisión hasta la fecha exacta de archivado. Esta función es insustituible para: recuperar contenido eliminado accidentalmente, rastrear cambios informativos (como modificaciones en páginas de políticas gubernamentales) o realizar investigaciones históricas. Además, WM ofrece la función 'Save Page Now' que permite guardar instantáneamente una URL, y proporciona una API para uso académico, reflejando su filosofía de acceso abierto. A pesar de los notables logros de WM, su implementación técnica y modelo operativo presentan limitaciones al atender demandas específicas, particularmente los requisitos de preservación profunda requeridos por instituciones académicas y bibliotecas:

1. Cuello de botella en la captura de contenido dinámico: La tecnología de rastreo tradicional de WM muestra limitaciones significativas al procesar páginas web modernas que dependen en gran medida del renderizado dinámico mediante JavaScript. Elementos interactivos (como gráficos de visualización de datos, flujos de contenido en tiempo real y aplicaciones de página única complejas) frecuentemente no pueden ser guardados de manera completa o precisa. Esto resulta en que numerosos recursos interactivos con importante valor académico o cultural (por ejemplo, herramientas educativas desarrolladas por universidades o exhibiciones de resultados de investigación) aparezcan en los archivos de WM en estado fragmentario o incluso ausentes por completo.
2. Deficiencia en la integridad de metadatos: El valor a largo plazo y la interpretabilidad de las páginas web dependen en gran medida de su información contextual (metadatos), como autor, institución editora, fecha de creación, historial de modificaciones, etc. Estudios han demostrado que las capturas de WM presentan altas tasas de pérdida de estos metadatos clave. Por ejemplo, investigaciones de Kelly et al. (TPDL 2013)²³ confirman que los contenidos interactivos generados dinámicamente y sus metadatos asociados frecuentemente no son capturados por Wayback Machine, lo que limita severamente la futura citación académica, el análisis de procedencia y la evaluación de confiabilidad de los materiales archivados.

²³ Kelly, M., Brunelle, J. F., Weigle, M. C., & Nelson, M. L. (2013). On the change in archivability of websites over time. TPDL 2013, Lecture Notes in Computer Science, 8092, 35–47.
https://doi.org/10.1007/978-3-642-40501-3_5

3. Limitaciones estructurales en la cobertura y diferencias regionales: la recolección de páginas web por parte de WM depende principalmente de su rastreador automático y de las contribuciones voluntarias de los usuarios (como la función "Save Page Now"). Falta un mecanismo sólido e institucionalizado de presentación obligatoria o de acuerdos de colaboración (como un sistema de depósito legal nacional). Esto provoca diferencias significativas en la cobertura según la región y el tipo de institución. En comparación con países como España, que tienen sistemas de archivo legal más desarrollados, las páginas web de universidades chinas en Wayback Machine (WM) suelen tener menor cobertura, ya que dependen principalmente de contribuciones de usuarios y rastreos automáticos (aunque hay debate sobre cifras exactas, la diferencia es evidente). Esto muestra que WM depende en gran medida de la iniciativa institucional y la cooperación internacional para obtener recursos.

2.3 Una mirada a la legislación reguladora: China vs España.

En cuanto a la protección legal del archivo web, España y China han adoptado enfoques muy diferentes. España estableció un marco legal sólido con el *Real Decreto 635/2015*. La clave de esta ley es su carácter obligatorio: el Artículo 9 autoriza explícitamente a la Biblioteca Nacional a superar barreras técnicas (incluyendo contenido con contraseña) para recopilar datos, mientras que el Artículo 12 impone multas de hasta 200.000 euros a editores que se nieguen a colaborar. Estas multas altas tienen un efecto disuasorio claro, garantizando la cooperación de instituciones (incluyendo universidades destacadas) con el sistema de recopilación de la Biblioteca Nacional, sentando una base legal firme para el archivo web nacional.

En cambio, en China el archivo web actual se basa principalmente en el Artículo 24²⁴ de la Ley de Bibliotecas Públicas, que menciona de forma general que "el Estado incentiva... la recopilación y preservación de recursos digitales". Aunque este artículo ofrece un fundamento legal para la conservación digital, carece de normas detalladas, autorizaciones operativas o mecanismos de ejecución sólidos (como sanciones claras o permisos legales para superar barreras técnicas). Esta ambigüedad legal limita en la práctica a instituciones como la Biblioteca Nacional cuando recopilan

²⁴ Comité Permanente de la Asamblea Popular Nacional de China. (4 de noviembre de 2017). *Ley de Bibliotecas Públicas de la República Popular China* (art. 24). Portal de la Asamblea Popular Nacional. <https://www.npc.gov.cn/zfxgk/zfxgkml/fl/202012/P020220223516904968244.pdf>

páginas web, especialmente frente a recursos con autenticación u obstáculos técnicos. Sus facultades y capacidad de ejecución dependen más de negociaciones institucionales y la voluntad de cooperación de los sitios web, dificultando una cobertura sistemática y obligatoria como la del modelo español.

2.4 Las herramientas principales del proceso de archivado empleadas en China vs España.

En la Wikipedia se recoge el mapa de iniciativas de archivado web, datado en 2021 que la propia IIPC recomienda para agrandar la consulta de herramientas. En la figura siguiente se aporta dicho mapa:

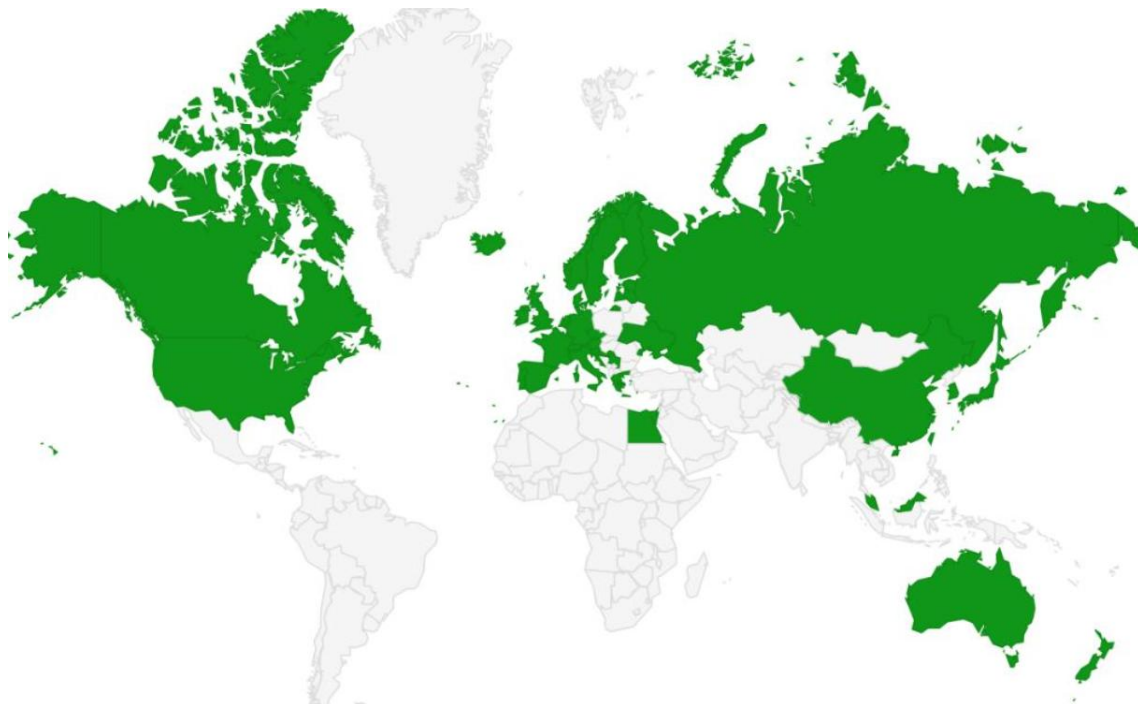


Figura 3. Mapa de iniciativas de archivado Web. Fuente: Wikipedia. Disponible en: https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives#/media/File:Map_of_Web_archiving_initiatives_58632FRT.jpg

Como es observable, tanto China como España, aparecen representadas en dicho mapa mundial. Para cada país se ofrece información sobre el nombre de la iniciativa de archivado, año, tecnologías, número de empleados a tiempo completo o parcial y observaciones. Igualmente, se aporta información sobre los datos archivado (contenidos, volumen, ,modalidad de rastreo y técnica empleada) y sobre el acceso (métodos de acceso mediante URL, mediante metadatos, en texto completo o mediante Memento) con detalles para cada país.

Detallamos para cada país algunos datos.

2.5.1 China

En China, el archivado web se lleva a cabo desde 2003.. Conocido como WIPC Se emplean herramientas tales como: Heritrix (herramienta de código abierto desarrollada por *Internet Archive* para rastrear y recolectar sitios web)²⁵, Wayback (plataforma del *Internet Archive* que permite consultar versiones archivadas de páginas web a lo largo del tiempo)²⁶y NutchWAX Archivado (una adaptación del motor de búsqueda Apache Nutch, diseñada para indexar y buscar contenidos archivados en formato ARC)²⁷el 26 de junio de 2015 en Wayback Machine .

Los datos sobre contenidos no son muy numerosos para China. Unicamente se señala la existencia de rastreos selectivos para algunos eventos esenciales o de gran influencia para la sociedad, economía, entre otros y generales para los sitios de dominio gov.cn.

Asimismo, el acceso en el archivo chino es posible mediante el empleo de metadatos y no cumple con las estipulaciones de Memento. El contenido del archivo solo está disponible en la intranet de la Biblioteca Nacional de China. Algunas colecciones son públicas y permiten la búsqueda de metadatos y la exploración por colección.

La herramienta de archivo web de la Biblioteca Nacional de China destaca en el procesamiento de codificación china (incluyendo el estándar GB18030-2022)²⁸, lo que garantiza una captura precisa de páginas en chino. Sin embargo, tiene limitaciones en contenido dinámico al no usar motores como Puppeteer, afectando su soporte para JavaScript. Además, existe un problema grave de metadatos: según una evaluación con ArchiveReady, 7 de las 15 mejores bibliotecas universitarias tenían

²⁵ Wikipedia contributors. (s.f.). *Heritrix*. Wikipedia, La enciclopedia libre.
<https://es.wikipedia.org/wiki/Heritrix>

²⁶ Wikipedia contributors. (s.f.). *Wayback Machine*. Wikipedia, La enciclopedia libre.
https://es.wikipedia.org/wiki/Wayback_Machine

²⁷ COPTR. (s.f.). *NutchWAX*. Community Owned digital Preservation Tool Registry. Recuperado de
<https://coptr.digipres.org/index.php/NutchWAX>

²⁸ Administración de Normalización de China. (2022). *GB 18030–2022: Requisitos técnicos para el procesamiento de información en chino: Juego de caracteres codificados*. Administración Estatal de Regulación del Mercado y Comité Nacional de Normalización de China.

metadatos incompletos (solo 75%-80% completos), debido a la falta de estándares unificados.

2.5.2 España

La información que se registra en la Wikipedia sobre España, atendiendo a la recomendación del IIPC es la siguiente: el Archivo de la Web Española es responsabilidad compartida entre la Biblioteca Nacional y las autonómicas con depósito legal. Las herramientas que emplea son: NetarchiveSuite , OpenWayback , Solr. Adopta un enfoque mixto de recolección selectiva y amplia. Desde 2009 a 2013, se han realizado anualmente recolectas de dominios .es completos en colaboración con Internet Archive mediante Heritrix y Wayback. Desde 2014, la Biblioteca Nacional de España realiza recolectas selectivas mediante NetarchiveSuite.

La Biblioteca Nacional dispone de 3 bibliotecarios a tiempo completo y 2 ingenieros de rastreo a tiempo parcial. Las bibliotecas regionales cuentan con varios bibliotecarios a tiempo parcial.

Como es observable, la Biblioteca Nacional de España ha establecido un proceso integral de archivo web: comienza con el renderizado de contenido dinámico para garantizar su captura completa, utiliza archivos en formato WARC para asegurar compatibilidad y gestión a largo plazo, emplea infraestructura nacional de almacenamiento en la nube para protección y persistencia de datos, y finalmente permite acceso público para cumplir su valor de preservación.

Un aspecto clave es que el marco legal español exige a los editores proporcionar copias sin DRM (protección digital de derechos), resolviendo directamente los obstáculos de archivo causados por protecciones técnicas y mejorando significativamente la capacidad de la BNE para preservar publicaciones digitales clave (como revistas electrónicas y tesis doctorales), lo que constituye un pilar fundamental de su sistema.

La información que se ofrece sobre la BNE es abundante, tal y como se expresa en la tabla siguiente:

Nombre	Contenido archivado (millones)	Espacio en disco ocupado (TB)	Formato de archivo	TLD/Rastros amplios	Rastros selectivos (Sí/No)	Comentarios
Archivo de la Web Española ^[56]	2539	117	WARC	.ES	Y	Rastros del dominio .ES (2009-2013): 2,421 millones de archivos (111 TB) en colaboración con Internet Archive. Rastros selectivos (2014-2015): 119 millones de archivos (6 TB). Se rastrearon unos 30 sitios web de medios de comunicación diariamente. Aún no se ha publicado.

Tabla 3. Contenidos archivados en la Web española. Fuente: Wikipedia. Disponible en:

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

Dispone de 2.539 millones de páginas con un volumen de 117 terabites y en formato Warc. Los rastros se enmarcan en el dominio “.es” y además realiza rastros selectivos. De la información registrada en la Wikipedia se deriva que se registran más sitios web de cuyos datos no se dispone información, como es el caso de sitios web derivados de medios de comunicación.

En la tabla siguiente se aporta información sobre el modelo de acceso, tanto para la BNE como para las autonómicas de Cataluña y la iniciativa vasca denominada Ondanaret.

Nombre	Historial de URL (Sí/No)	Búsqueda de metadatos (catálogo/avanzada) (Sí/No)	Búsqueda de texto completo (Sí/No)	Cumplimiento de Memento (No/Nativo/Proxy)	Comentarios
Archivo de la Web Española ^[56]	Y (Futuro)	Y (Futuro)	Y (Futuro)	No	Prever acceso en sitio en el corto-mediano plazo.
PADICAT: EI Archivo Web de Cataluña ^[57]	Y	Y	Y	No	Acceso abierto completo.
Archivo del Patrimonio Digital Vasco ^[58]	Y	Y	Y	No	

Tabla 4. Comparativa del acceso entre archivos web españoles (Fuente: Wikipedia en español, https://es.wikipedia.org/wiki/Archivo_de_la_Web_Espa%C3%B1ola)

Mientras que en el plano autonómico se aportan diferentes opciones de acceso, en el nacional, la biblioteca lo considera futurable, esto es, se encuentra en desarrollo,

en el momento de alimentar esta página de Wikipedia. En la actualidad el acceso en sitio es posible.

Estudio empírico: Estudio de caso

Capítulo 3 Evaluación del grado de archivabilidad de las páginas Web.

En este capítulo se detalla el concepto de archivabilidad, la herramienta empleada de uso gratuito personal “archiveready” y otras herramientas de utilidad para valorar problemas.

3.1 concepto de archivabilidad

La “archivabilidad web”, concepto central de este estudio, se refiere a la viabilidad técnica que una página o sitio web presenta para que pueda ser capturado, empaquetado y preservado de manera completa, precisa y confiable con las tecnologías actuales, garantizando su acceso y utilidad futura (Kelly et al., 2013).

En términos simples, mide la capacidad del contenido web para resistir la pérdida de información, ingresar exitosamente a archivos digitales y mantenerse usable en el futuro. Es un indicador multidimensional que evalúa cómo el contenido, estructura, funciones y recursos dependientes de una página mantienen su integridad y usabilidad durante todo el proceso de archivo.

N. Taylor define la archivabilidad como sigue:

“se refiere a la facilidad con la que el contenido, la estructura, la funcionalidad y la presentación de un sitio web se pueden conservar y volver a presentar posteriormente mediante herramientas modernas de archivado web.”²⁹

La archivabilidad incluye varias dimensiones, vinculadas a la mejora del acceso, la captura, el descubrimiento y otros recursos.



Figura 4. Dimensiones de la archivabilidad. Fuente: sitio de N. Taylor.
<https://nullhandle.org/web-archivability/index.html>

²⁹ Taylor, N. (n.d.). *Web Archivability Index*. <https://nullhandle.org/web-archivability/index.html>

Logicamente, ante el diseño de un sitio web se consideran dimensiones vinculadas con la accesibilidad, el rendimiento, la optimación de los motores de búsqueda, el cumplimiento de normas y la usabilidad, si bien, no sólo los desarrolladores de sitios piensan en las dimensiones de diseño, sino que también, se deben considerar las dimensiones de archivabilidad. El modo de captura y de descubrimiento del sitio y la facilidad en el acceso son importantes para hacer posible el archivado y la consulta con posterioridad.

Cabe destacar que la archivabilidad web no es una propiedad fija, sino que está significativamente condicionada por múltiples factores dinámicos.

3.2 La herramienta ArchiveReady

ArchiveReady fue una herramienta en línea cuyo objetivo principal era evaluar la viabilidad técnica de que una página web determinada pudiera ser capturada y preservada con éxito por servicios de archivo web como Internet Archive, es decir, su “archivabilidad”. Esta herramienta fue desarrollada por investigadores con el fin de ofrecer a los creadores de sitios web o profesionales de archivos un medio de diagnóstico preliminar.

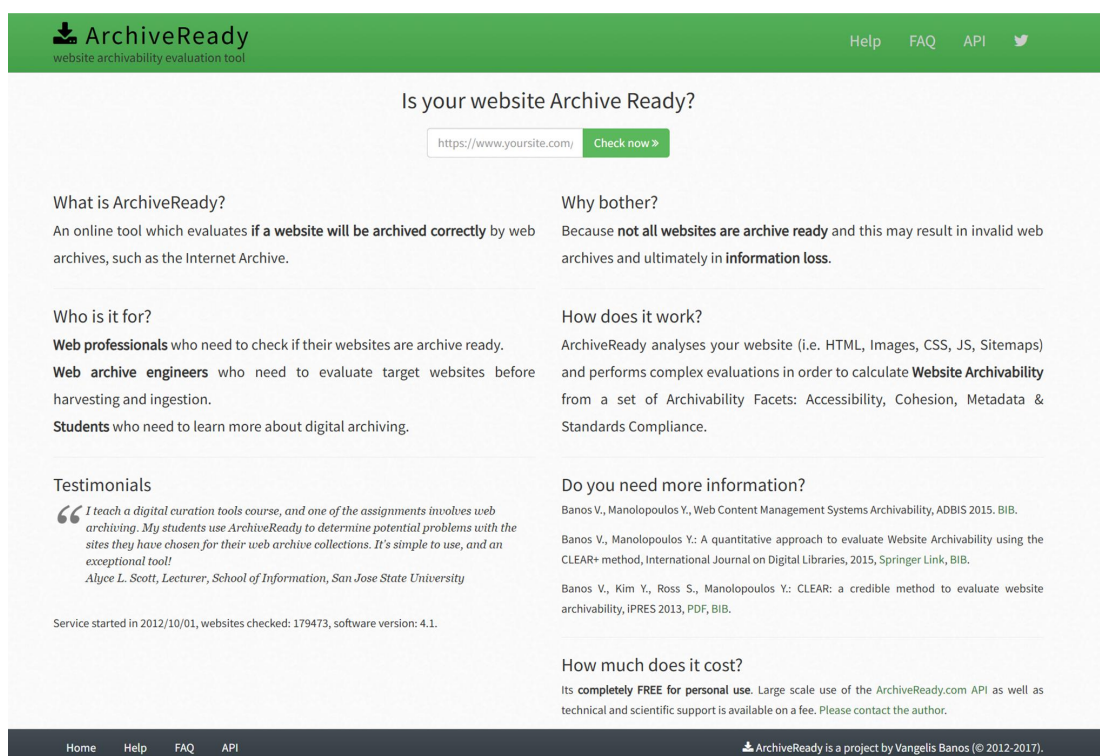


Figura 5. Interfaz principal de la herramienta ArchiveReady (Fuente: Sitio web oficial de ArchiveReady, <https://www.archiveready.com/>)

El usuario introduce la URL de la página web objetivo a través de la interfaz, y ArchiveReady procede a realizar un escaneo técnico automatizado del sitio. Una vez finalizado el análisis, la herramienta genera un informe estructurado. Este informe utiliza un sistema de codificación por colores (verde/naranja/rojo/azul) para indicar de forma visual el nivel de los problemas detectados en varios aspectos clave relacionados con el archivo web, lo que facilita la identificación rápida de posibles puntos de riesgo. Los datos de prueba temprana utilizados en este estudio (véase el Capítulo 4) fueron obtenidos a partir de dicha herramienta.

Se ofrece en dicha herramienta un análisis pormenorizado por categorías, las cuales son:

- **Accesibilidad:** Esta dimensión evalúa si el sitio web puede ser navegado y comprendido por diferentes perfiles de usuario, incluidos aquellos con discapacidades. Se consideran aspectos como la presencia de texto alternativo en imágenes, el uso de etiquetas semánticas y la estructura del contenido compatible con lectores de pantalla. Se evalúa con base en criterios establecidos por las Pautas de Accesibilidad al Contenido en la Web (WCAG³⁰).
- **Cohesión:** Se refiere al grado de integridad interna del sitio web. Analiza si todas las páginas están correctamente enlazadas, si las rutas de navegación son coherentes y si existen redirecciones innecesarias o errores 404 que rompen la estructura del sitio. Una cohesión alta implica una arquitectura web bien organizada, con enlaces funcionales y jerarquías claras.
- **Metadatos:** Esta categoría examina la inclusión de información estructurada y descriptiva que facilite la recuperación, preservación e interpretación del contenido archivado. Se consideran tanto los metadatos en cabeceras HTML (como <meta> tags), como el uso de esquemas normalizados como Dublin Core o Schema.org. La presencia de metadatos completos mejora significativamente la archivabilidad.
- **Cumplimiento de normas:** Evalúa hasta qué punto el sitio web sigue estándares técnicos establecidos, tales como el uso correcto del protocolo HTTP, la validez del código HTML y CSS según los validadores oficiales del W3C, y la adecuación a formatos interoperables como WARC u OAI-PMH. Un alto cumplimiento

³⁰ W3C. (2018). *Web Content Accessibility Guidelines (WCAG) 2.1*. World Wide Web Consortium (W3C). <https://www.w3.org/TR/WCAG21/>

normativo asegura mayor compatibilidad con herramientas de archivado automatizado.³¹

3.3 Otras herramientas complementarias

Además de herramientas de evaluación como ArchiveReady, existen muchas otras que son fundamentales para el proceso práctico de archivado web. Algunas categorías representativas y sus ejemplos incluyen:

1. **Webrecorder:** Permite capturar y reproducir páginas web de forma interactiva, siendo especialmente adecuado para sitios que contienen JavaScript y contenidos dinámicos. Su enfoque se centra en una captura de alta calidad y una reproducción auténtica del sitio original. <https://replayweb.page/>
2. **Brozzler:** Utiliza el motor de un navegador para renderizar y archivar páginas web, lo que permite capturar contenidos interactivos basados en JavaScript. Es compatible con el formato WARC, ampliamente utilizado en archivado web. <https://github.com/internetarchive/brozzler>
3. **webarchive-discovery:** Desarrollada por la Biblioteca Británica, esta herramienta permite indexar y analizar archivos WARC, lo que facilita tareas posteriores como el análisis de los contenidos archivados y la evaluación de su capacidad de recuperación. <https://github.com/ukwa/webarchive-discovery>
4. **Sitemap:** Permite a los motores de búsqueda y herramientas de archivado conocer la estructura del sitio web a través de un archivo XML. Su correcta implementación mejora la cobertura y organización de la captura. <https://www.sitemaps.org/>
5. **Robots.txt:** Archivo que indica qué partes del sitio web pueden ser rastreadas o excluidas por los bots. Una configuración restrictiva puede bloquear procesos de archivado automatizado. <http://www.robotstxt.org/>
6. **RSS:** Los canales RSS facilitan la identificación de contenidos dinámicos o actualizados, permitiendo una monitorización constante de nuevas publicaciones en sitios web. <https://en.wikipedia.org/wiki/RSS>

³¹ Banos, V., Manolopoulos, Y. A quantitative approach to evaluate Website Archivability using the CLEAR+ method. *Int J Digit Libr* 17, 119–141 (2016). <https://doi.org/10.1007/s00799-015-0144-4>

7. **CSS Validator:** La exactitud en los archivos CSS garantiza que la presentación y contenido de las páginas se mantenga coherente tras la captura. Esta herramienta permite validar la corrección del código CSS.
<http://jigsaw.w3.org/css-validator/>
8. **Protocolo HTTP:** El uso correcto del protocolo HTTP y su configuración (como el estado de redirecciones o errores 3xx/4xx) afecta directamente la accesibilidad del sitio para herramientas de preservación.
9. **Metadatos:** La inclusión adecuada de metadatos estructurados (por ejemplo, etiquetas <meta>, Dublin Core, Schema.org) mejora la interpretación del contenido archivado y su recuperación futura.

Capítulo 4 Selección de muestras y recolección de datos

Este capítulo tiene como objetivo explicar los principios de selección de muestras, los métodos de recolección de datos y el contenido real de los datos obtenidos en este estudio, proporcionando una base sólida para los análisis posteriores. La investigación se centra en las diferencias en las prácticas de archivo web entre bibliotecas universitarias de China y España, recopilando datos que abarcan dimensiones como: indicadores técnicos de archivabilidad web, estructura y características técnicas de los sitios web, y disponibilidad de políticas públicas. Para garantizar la representatividad y comparabilidad de los resultados, la selección de muestras y la obtención de datos siguieron estrictamente criterios predefinidos.

4.1 Selección de muestras

En el estudio se empleó muestreo estratificado para comparar bibliotecas universitarias chinas y españolas, utilizando como criterio principal el Ranking ARWU (por su neutralidad cultural e indicadores transparentes como publicaciones y premios), complementado para China con el Ranking Wu Shulian (que evalúa factores locales como reputación y enseñanza), garantizando así representatividad y evitando sesgos hacia países angloparlantes o potencias culturales.

En la tabla siguiente se recoge la relación de la población, la muestra y el porcentaje que representa sobre la población.

Índices empleados	Población total	% muestra	China	España
ARWU	China: 225	8.9%	20	30
	España: 36	83.3%		
Wu shulian	812	1.2%	10	
Total	1.073	60=100	30	30

Tabla 5. Selección de la muestra.

El marco de muestreo exploratorio se divide en tres niveles según la clasificación académica global, con 10 universidades de China y 10 de España en cada grupo:

1. Grupo élite: Se seleccionaron las 100 mejores universidades del ranking ARWU, representando las instituciones académicas más destacadas de ambos países. Para China, la muestra se basó en las top 10 nacionales del ARWU (ej. Universidad de Tsinghua, Universidad de Pekín), validadas por su estabilidad integral mediante el ranking Wu Shulian; para España, se tomaron directamente las 10 mejores españolas del ARWU (ej. Universidad Autónoma de Madrid, Universidad de Barcelona).
2. Grupo intermedio: Universidades en el rango 500-700 del ARWU, reflejando características típicas de instituciones con nivel académico medio.
3. Grupo básico: Universidades en el rango 700-900 del ARWU, representando instituciones de educación superior básica.

Se siguió estrictamente el principio de consistencia de datos: todas las muestras se seleccionaron según la misma edición anual del ranking ARWU (ej. versión 2024). Para los grupos intermedio y básico, se extrajeron universidades según su posición global directa; si había insuficientes instituciones dentro del rango, se completó con aquellas más cercanas a los límites establecidos.

En la tabla siguiente se detallan agrupadas por categorías los datos atendiendo a los índices empleados.

Índices empleados	Categorías establecidas	España	China	Total datos/% muestra
ARWU	Élite.	10	10	20 (33,3%)
Wu shulian			<i>Validación cruzada</i>	
ARWU	Intermedia – 500/700	10	10	20 (33,3%)
ARWU	Inferior: 700/900	10	10	20 (33,3%)
total	3 categorías	30	40	60 (100%)

Tabla 6. Datos obtenidos agrupados por categorías.

Cabe destacar que el ranking Wu Shulian solo se utilizó para la validación cruzada localizada del grupo élite chino (ej. verificando la coherencia entre las posiciones de Tsinghua en ARWU y Wu Shulian), sin aplicarse a universidades españolas ni comparaciones transnacionales, evitando así conflictos metodológicos.

4.2 Recopilación de datos

Se utilizó la herramienta ArchiveReady (versión 2.1) para realizar un escaneo estandarizado de las páginas principales de las bibliotecas de las universidades seleccionadas, obteniendo puntuaciones en cuatro indicadores clave:

- Accesibilidad: se analiza el código de estado HTTP, la proporción de enlaces rotos (como el error 404 Not Found) y la compatibilidad con el archivo robots.txt.
- Normalización: se verifica el cumplimiento de los estándares HTML/CSS (W3C) y el nivel de normalización en las rutas de los recursos.
- Cohesión: se evalúa la densidad de enlaces internos y la integridad del archivo sitemap.xml.
- Metadatos: se examina la cobertura de etiquetas <meta> y el grado de implementación de anotaciones semánticas con Schema.org.
- Así como la media de estos como medida general.

Mediante esta herramienta se trabajaron igualmente, una vez hallados los totales, 3 universidades chinas y 3 españolas de las diferentes categorías con mayor detalle. Aportando información sobre el tipo de errores y aciertos que la herramienta facilita.

Por tanto, la recopilación de datos se determinó a 3 niveles.

- 1) Nivel general, utilizando el criterio de medida general que se conforma a partir de la media de 4 criterios, para las 60 universidades.
- 2) Nivel especificado para los 4 criterios, anteriormente citados, para las 60 universidades, avanzando estudios específicos y comparados para cada criterio.
- 3) Análisis al detalle para 6 casos de los errores y aciertos. Razón por la que, en este caso, se confrontan 6 universidades, siendo la mitad de ellas de cada país.

4.3 Datos obtenidos

Debido a que las páginas web de algunas bibliotecas universitarias no pueden ser evaluadas con la herramienta ArchiveReady (como se muestra en la imagen), para asegurar el desarrollo fluido del estudio, se procederá a evaluar las siguientes páginas web de bibliotecas en la lista.

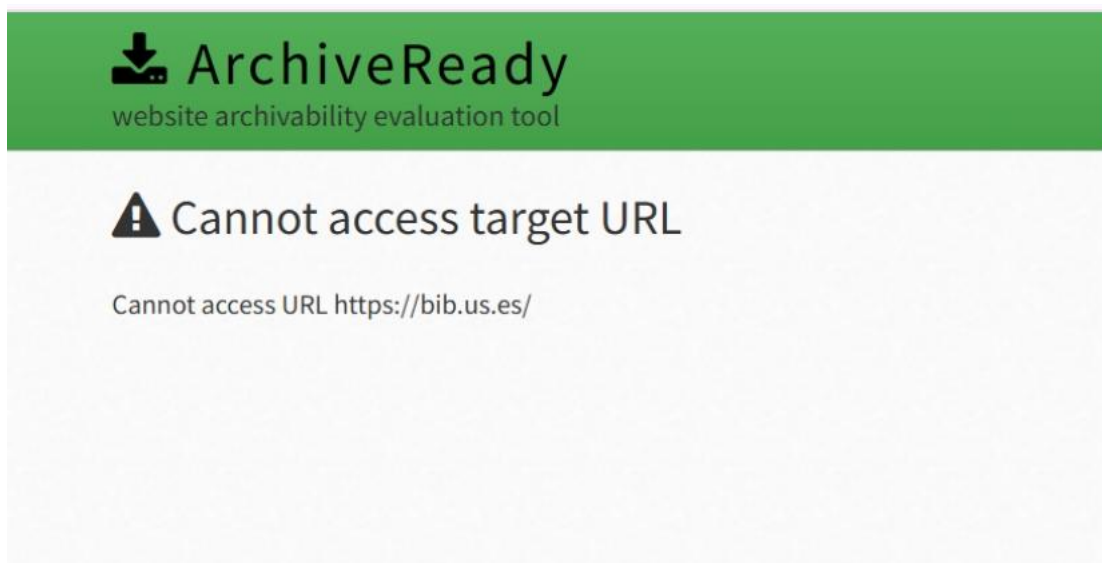


Figura 6. Resultado de evaluación de archivabilidad de <https://bib.us.es/> con la herramienta ArchiveReady (Fuente: ArchiveReady, <https://www.archiveready.com/>)

Ante este obstáculo, se dio continuidad a la lista. Las Universidades que no pudieron ser consultadas mediante la herramienta fueron:

- La universidad de Sevilla
- La universidad de Las Islas Baleares
- La universidad de Oviedo

Debido al número relativamente reducido de universidades en España, y especialmente al bajo número de instituciones que aparecen en los rankings internacionales más reconocidos como el ARWU (solo 36 universidades españolas frente a 225 universidades chinas), este estudio ha adoptado una estrategia de cobertura casi total para la selección de la muestra española.

En definitiva:

- Se han incluido un total de 30 universidades españolas, divididas en tres grupos comparables: élite, intermedio (rango 500–700) e inferior (rango 700–900), con 10 instituciones en cada grupo;
- Esta muestra representa aproximadamente el 83,3 % del total de universidades españolas clasificadas en los rankings seleccionados, lo que garantiza una amplia representatividad;
- Aunque esta estrategia sacrifica cierta aleatoriedad estadística, refuerza la validez empírica y comparativa del estudio, especialmente cuando el tamaño poblacional del país es limitado;

- Además, dado que las universidades españolas presentan diferencias regionales y tecnológicas más moderadas en comparación con China, esta muestra amplia resulta suficiente para reflejar la diversidad y el nivel general del sistema universitario español.

Por lo tanto, la decisión metodológica de utilizar una muestra altamente representativa en el caso español responde a una lógica de **equilibrio entre factibilidad, exhaustividad y comparabilidad internacional**.

La relación de páginas web de España se registra en la figura 7, como sigue.

Las universidades españolas						
universidad	tipo	URL	Website Archivability	Accessibility	Cohesion	Metadata Standards Compliance
grupo de elite						
1Universidad de Barcelona	público	https://www.ub.edu/biblioteca	79%	56%	100%	59%
2Universidad de Valencia	público	https://www.uv.es/urweb/servei-biblioteques-documentacio/c	69%	50%	100%	65%
3Universidad Autónoma de Barcelona	público	https://www.uab.cat/web/biblioteques-1345668785260.html	86%	77%	100%	94%
4Universidad Autónoma de Madrid	público	https://www.uam.es/uam/biblioteca	84%	68%	100%	94%
5Universidad Complutense de Madrid	público	https://biblioteca.ucm.es	79%	50%	87%	80%
6Universidad de Pompeu Fabra	público	https://www.upf.edu/web/biblioteca	84%	36%	100%	100%
7Universidad de Granada	público	https://biblioteca.ugr.es	80%	42%	89%	80%
8Universidad del País Vasco	público	https://www.ehu.eus/es/web/biblioteca	86%	65%	93%	84%
9Universidad Politécnica de València	público	https://biblioteca.uoc.edu	79%	50%	100%	65%
11Universidad de La Laguna	público	https://www.ull.es/servicios/biblioteca/	76%	57%	82%	63%
media			0.802	0.551	0.951	0.793
grupo intermedio						
12Universidad de Murcia	público	https://www.um.es/web/buc	71%	54%	89%	60%
13Universidad de Navarra	privado	https://www.unav.edu/bibliotecas	76%	68%	98%	40%
14Universidad de Salamanca	público	https://bibliotecas.usal.es/	69%	36%	96%	65%
15Universidad de Santiago de Compostela	público	https://www.usc.gal/gl/servizos/biblioteca	80%	65%	100%	78%
16Universidad de Zaragoza	público	https://biblioteca.unizar.es	62%	43%	92%	15%
17Universidad Politécnica de Madrid	público	https://www.upm.es/Estudiantes/Biblioteca	82%	60%	92%	78%
18Universidad Jaume I	público	https://www.uji.es/biblioteca	78%	41%	100%	94%
19Universidad de Alicante	público	https://biblioteca.ua.es	72%	36%	95%	79%
21Universidad de Lérida	público	https://bid.udl.cat/	81%	65%	84%	75%
22Universidad de Alcalá	público	https://www.uah.es/biblioteca	71%	47%	100%	56%
media			0.742	0.515	0.946	0.64
grupo inferior						
23Universidad de Castilla-La Mancha	público	https://www.biblioteca.uclm.es	75%	49%	100%	51%
24Universidad de Extremadura	público	https://biblioteca.unex.es	62%	42%	82%	48%
25Universidad de Málaga	público	https://www.biblioteca.uma.es	84%	50%	100%	88%
26Universidad Rovira i Virgili	público	https://www.urv.cat/ca/vida-campus/biblioteques/	90%	66%	100%	93%
27Universidad Politécnica de Cataluña	público	https://biblioteca.upc.edu	73%	38%	59%	94%
28Universidad de Girona	público	https://biblioteca.udg.edu	84%	43%	98%	94%
30Universidad de Vigo	público	https://cud.uvigo.es/biblioteca/	75%	58%	99%	44%
31Universidad Carlos III de Madrid	público	https://www.uc3m.es/biblioteca	73%	38%	59%	94%
32IE University	privado	https://www.ie.edu/university/library/	86%	60%	100%	85%
33Universidad Ramon Llull		https://www.url.edu/es/servicios/bibliotecas/bibliotecas	84%	66%	87%	81%
media			0.786	0.51	0.884	0.772
mediatotal			0.776667	0.525333	0.927	0.735

Figura 7. Evaluación de archivabilidad web de sitios de bibliotecas universitarias en España (Fuente: Elaboración propia con datos obtenidos mediante ArchiveReady, junio de 2025)

Seguidamente, pasamos a justificar la selección de universidades chinas. Con el objetivo de garantizar la representatividad y la solidez científica en la comparación de la capacidad de archivado web entre universidades chinas y españolas, la selección de las instituciones chinas se ha basado en una combinación de criterios clave: distribución geográfica, diversidad institucional, nivel de recursos académicos, sistemas de clasificación y equilibrio cuantitativo de la muestra.

A continuación, se detallan los motivos principales:

1. Amplia distribución geográfica y representatividad espacial

China es un país extenso con grandes disparidades regionales en términos de desarrollo económico, inversión educativa e infraestructura digital. Para evitar sesgos de concentración territorial, se ha optado por incluir universidades distribuidas por todo el país:

- Región norte: como Tsinghua, Peking, BUPT y UIBE, representando el nivel más alto de educación en la capital;
- Región este (delta del Yangtsé): con universidades como Fudan, Shanghai Jiao Tong, Nanjing, SEU, SDU, SHUFE, entre otras;
- Región sur (delta del río Perla): con universidades como Sun Yat-sen, SCUT y Jinan;
- Regiones centro-occidentales: como Central South University, Hunan University of Finance and Economics, Southwest University of Science and Technology, Sichuan Agricultural University;
- Regiones noreste y noroeste: como Harbin Medical University, Jilin University, Xidian University, Shaanxi University of Science and Technology.

Esta diversidad geográfica permite captar diferencias relevantes entre las prácticas institucionales de distintas zonas y asegura que los resultados tengan aplicabilidad nacional.

2. Diversidad institucional y cobertura de distintos niveles de recursos

Se ha considerado la heterogeneidad del sistema universitario chino, incluyendo:

- Universidades de élite nacional (miembros de los proyectos “985” y “211”, como Tsinghua, Peking, Fudan, SJTU, Sun Yat-sen), con altos niveles de recursos y gestión avanzada;
- Universidades regionales clave, como SDU, Harbin Medical, SEU o Central South, que muestran un nivel intermedio con fuerte influencia local;
- Universidades aplicadas o de nivel local, como HFUE o SWUST, que reflejan la realidad de instituciones con menor dotación tecnológica y organizativa;
- Universidades especializadas en economía, telecomunicaciones o tecnología (como SHUFE, BUPT, UIBE, Xidian), lo que enriquece la diversidad funcional de los sitios web institucionales.

Esta variedad institucional permite analizar cómo el nivel de recursos y el tipo de universidad influyen en las estrategias y resultados del archivado web.

3. Diseño basado en rankings académicos con validación cruzada

Las universidades fueron seleccionadas en base a dos sistemas de clasificación:

- ARWU (Academic Ranking of World Universities): internacionalmente reconocido, centrado en la excelencia investigadora y la visibilidad global;

- Ranking de Wu Shulian: de ámbito nacional, abarca más de 800 instituciones chinas y evalúa de forma holística los recursos, la docencia, el personal y el empleo.

La combinación de ambos rankings permite una validación cruzada rigurosa, asegurando que las universidades elegidas sean reconocidas tanto internacional como nacionalmente. Esta estrategia fortalece la legitimidad metodológica y proporciona una base cuantitativa para futuros análisis correlativos.

4. Equilibrio cuantitativo con la muestra española

Para facilitar una comparación directa y estadísticamente sólida, se ha adoptado una estructura simétrica de muestra:

- 10 universidades de élite,
- 10 universidades intermedias (rango 500–700),
- 10 universidades de nivel inferior (rango 700–900).

Este diseño refleja una división en tres grupos equivalentes tanto para China como para España (30 instituciones por país), lo que mejora la comparabilidad, la consistencia del análisis y la replicabilidad del estudio.

Igualmente, se muestra en la Figura 8 la relación de páginas web agrupadas por categorías para China.

las universidades chinas								
universidad	tipo	URL	Website Archivability	Accessibility	Cohesion	Metadata	Standards Compliance	
grupo de elite								
1 Tsinghua universidad 清华	publica	https://lib.tsinghua.edu.cn/	79%	44%	100%	100%		72%
2 Pekin universidad 北大	publica	https://www.lib.pku.edu.cn/	75%	44%	100%	100%		55%
3 Zhejiang universidad 浙大	publica	https://libweb.zju.edu.cn/	63%	40%	100%	80%		32%
4 Shanghai Jiao Tong universidad	publica	https://www.lib.sjtu.edu.cn/f/main/index.shtml	69%	40%	100%	75%		61%
5 USTC 中国科学技术大学	publica	https://lib.ustc.edu.cn/	75%	49%	100%	75%		77%
6 Fudan universidad 复旦	publica	https://library.fudan.edu.cn/	73%	40%	100%	80%		72%
7 Hong Kong Universidad 香港大学	publica	https://lib.hku.hk/	79%	38%	88%	100%		89%
8 Sun Yat-sen universidad 中山	publica	https://library-443.webvpn.sysu.edu.cn/	78%	41%	100%	100%		72%
9 Nanjing universidad 南京大学	publica	https://lib.nju.edu.cn/	84%	43%	100%	100%		91%
10 Wuhan Universidad 武汉大学	publica	https://www.lib.whu.edu.cn/	71%	28%	100%	80%		76%
media			0.746	0.407	0.988	0.89		0.697
grupo intermedio								
11 HIT 哈工大	publica	http://www.lib.hit.edu.cn/	71%	39%	100%	80%		66%
12 Yunnan Universidad 云南大学	publica	https://lib.ynu.edu.cn/	69%	38%	100%	100%		37%
13 Correos y Telecomunicaciones de Pekín	publica	https://lib.bupt.edu.cn/	84%	43%	100%	100%		93%
14 Normal del Centro de China Universidad 华	publica	https://lib.gzu.edu.cn/	76%	39%	100%	80%		84%
15 Quzhou Universidad 贵州大学	publica	https://cscv.gzu.edu.cn/	69%	45%	100%	100%		32%
16 Normal de Shaanxi Universidad 陕西师范大学	publica	https://lib.scnu.edu.cn/	82%	44%	100%	100%		83%
17 Normal del Sur de China Universidad 华南	publica	https://lib.snu.edu.cn/	76%	28%	100%	75%		100%
18 Suroeste Universidad 西南大学	publica	https://www.library.nenu.edu.cn/	75%	42%	100%	80%		77%
19 Normal del Noreste Universidad 东北师范大	publica	https://lib.sxu.edu.cn/	70%	46%	100%	100%		36%
20 Shanxi Universidad 山西大学	publica	https://lib.wzu.edu.cn/	80%	44%	100%	100%		77%
media			0.752	0.408	1.00	0.915		0.685
grupo inferior								
21 Universidad de Geociencias de Chengdu 成都	publica	https://lib.cdut.edu.cn/	76%	28%	100%	75%		100%
22 Universidad de Ciencias Médicas de Harbin	publica	https://lib.hrbmu.edu.cn/	80%	45%	100%	100%		76%
23 Universidad Politécnica de Lanzhou 兰州理	publica	https://lib.lut.edu.cn/index	65%	36%	100%	100%		24%
24 Universidad de Shantou 汕头大学	publica	https://www.lib.stu.edu.cn/	78%	37%	100%	100%		74%
25 Universidad de Jinbei 集美大学	publica	https://lib.jmu.edu.cn/	83%	45%	100%	100%		88%
26 Universidad de Ciencia y Tecnología de S	publica	https://library.sust.edu.cn/	66%	52%	98%	80%		36%
27 Universidad de Ciencia y Tecnología de W	publica	https://tsg.wust.edu.cn/	82%	40%	100%	100%		86%
28 Universidad de Xiangtan 湘潭大学	publica	https://www.xtu.edu.cn/xvsh/gzfw/tsg.htm	82%	45%	100%	100%		83%
29 Universidad de Finanzas, Economía y Dere	publica	https://library.zuel.edu.cn/	78%	54%	100%	100%		57%
30 Universidad Politécnica de Changsha 长沙理	publica	https://www.csust.edu.cn/tsg/	81%	45%	100%	100%		80%
media			0.771	0.427	0.998	0.955		0.704
mediatotal			0.756333	0.414	0.995333	0.92		0.695333

Figura 8. Evaluación de archivabilidad web de sitios de bibliotecas universitarias en China (Fuente: Elaboración propia con datos obtenidos mediante ArchiveReady, junio de 2025)

En la figura 9 se muestra la distribución geográfica de las universidades chinas



objeto de estudio.

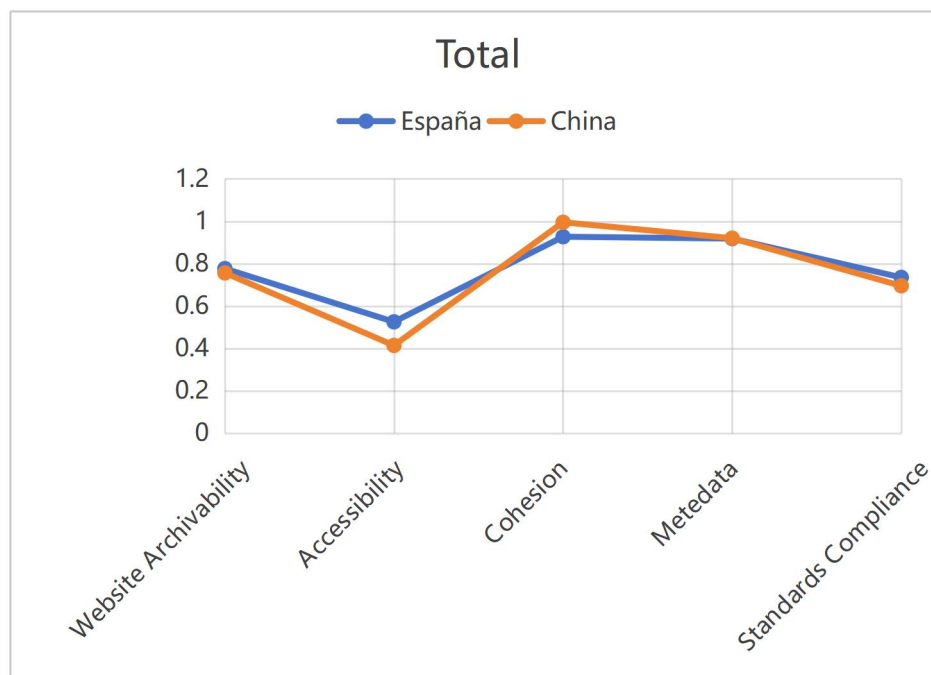
Figura 9. Mapa de distribución geográfica de las universidades chinas según grupo de estudio
(Fuente: Mapa generado con apoyo de ChatGPT a partir de los datos proporcionados por el autor.)

Capítulo 5 Presentación de los resultados del estudio

Este capítulo analiza sistemáticamente la archivabilidad web de las páginas principales de bibliotecas universitarias en China y España (grupos élite, intermedio y básico), basado en resultados de la herramienta ArchiveReady. La evaluación cubre cinco indicadores clave: archivabilidad general (Website Archivability), accesibilidad (Accessibility), cohesión estructural (Cohesion), integridad de metadatos (Metadata) y cumplimiento de estándares (Standards Compliance). Mediante análisis longitudinal de promedios grupales y comparaciones transversales China-España, se revelan diferencias estructurales en capacidades de preservación web.

5.1 Análisis de las tendencias generales

Este estudio recopiló los resultados de la evaluación de los sitios web de 30 bibliotecas universitarias de España y China. Las estadísticas muestran que la diferencia en la archivabilidad general entre los sitios web de ambos países es mínima: la puntuación media de España es de 0,7767, mientras que la de China es de 0,7563, con una diferencia de solo 0,0204. Sin embargo, los indicadores que componen esta puntuación general no son uniformes. España destaca en accesibilidad y conformidad con los estándares, con una accesibilidad media de 0,5253 y una conformidad con los estándares de 0,7350; China, por su parte, lidera en cohesión estructural e integridad de los metadatos, con una cohesión de 0,9953 y una integridad de los metadatos de 0,9200. Las ventajas complementarias de ambos países hacen que la puntuación total final tienda a converger.



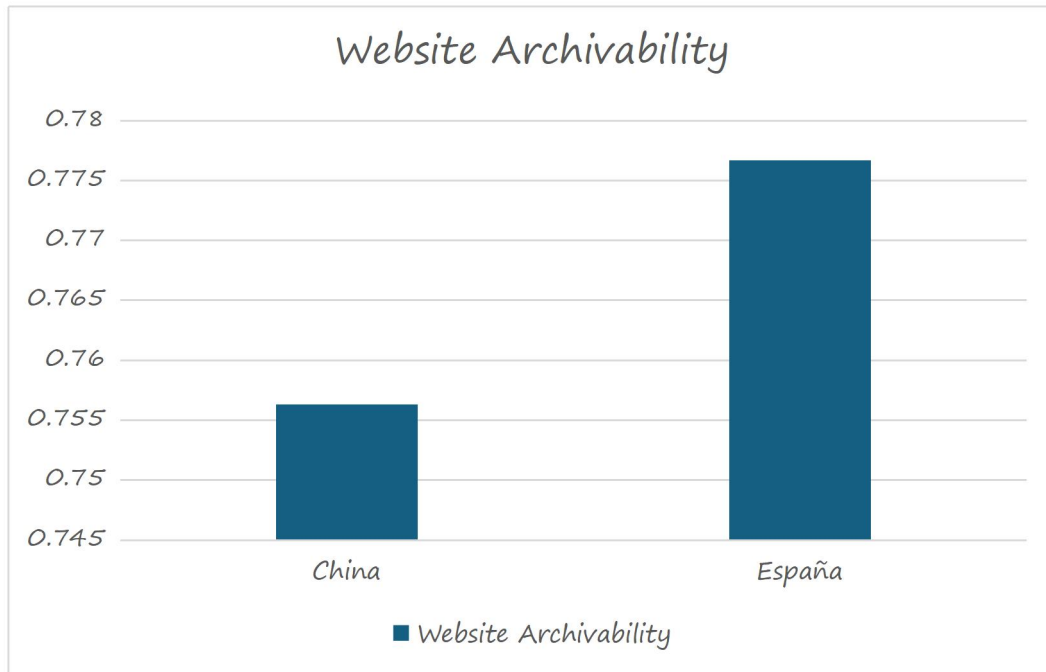
Gráfica 1. Comparación de los indicadores de archivabilidad web en bibliotecas universitarias de España y China

Los resultados mostrados en la tabla sobre el grado de archivabilidad general son como sigue y se muestra en la tabla siguiente.

País	Promedio de archivabilidad %
España	77,66
China	75,6

Tabla 7. Promedio de archivabilidad de los sitios web de bibliotecas universitarias

La capacidad general de archivado muestra una diferencia muy pequeña entre ambos países (0,776 frente a 0,756). Si bien, el valor más alto se registra para los servicios bibliotecarios de España. Los indicadores que conducen a estos porcentajes son variables en un país y otro, como veremos, reforzándose en unos criterios frente a otros, lo cual marca ciertas diferencias, aunque estas se reducen en el cómputo general.



Gráfica 2. Promedio general de archivabilidad web en universidades de España y China

5.2 Diferencias entre los grupos

Grupo	país	Website Archivability	Accessibility	Cohesion	Metadata	Standards Compliance
Elite	España	0.802	0.551	0.951	0.91	0.793
	China	0.746	0.407	0.988	0.89	0.697
	promedio	0.774	0.479	0.9695	0.9	0.745
Intermedio	España	0.742	0.515	0.946	0.87	0.64
	China	0.752	0.408	1.00	0.915	0.685
	promedio	0.747	0.4615	0.973	0.8925	0.6625
Inferior	España	0.786	0.51	0.884	0.975	0.772
	China	0.771	0.427	0.998	0.955	0.704
	promedio	0.7785	0.4685	0.941	0.965	0.738

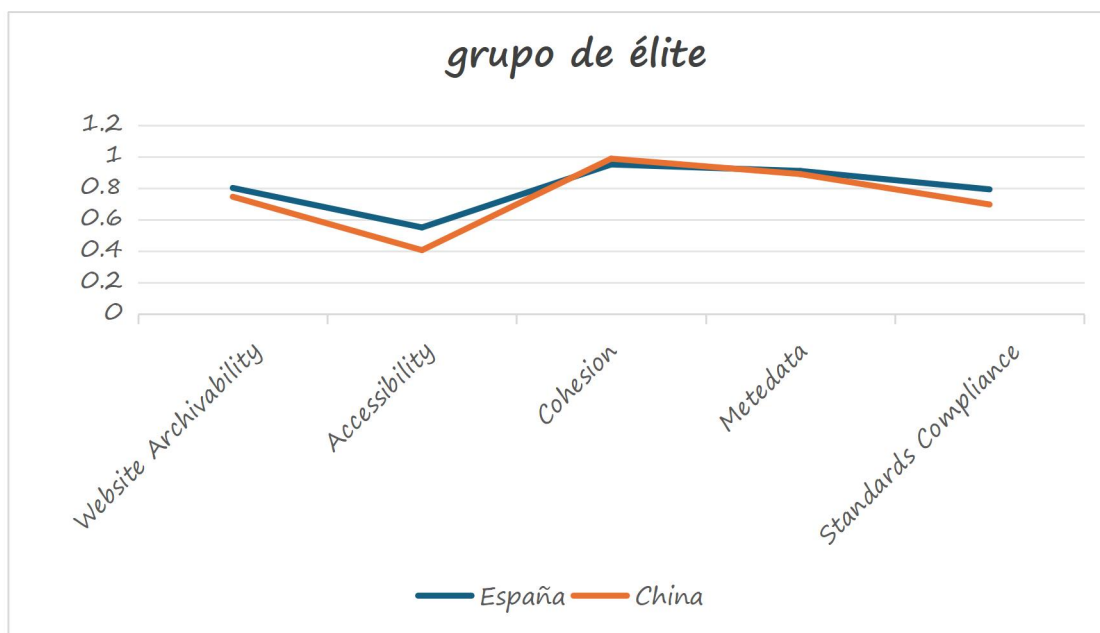
Tabla 8. Relación de datos atendiendo a los criterios.

Seguidamente se comenta la tabla atendiendo a los grupos establecidos.

Comparativa del grupo élite:

En el grupo de élite, los sitios web de las bibliotecas universitarias españolas alcanzaron una puntuación promedio de archivabilidad de 0,802, superior al 0,746 obtenido por las universidades chinas. España se destacó particularmente en los indicadores de Accesibilidad (0,551) y Cumplimiento de Estándares (0,793), lo que refleja un mayor grado de adecuación a los principios de diseño inclusivo y a las normativas técnicas como W3C. También obtuvo resultados sólidos en Cohesión (0,95) y Metadatos (0,91), aunque estas dos dimensiones también constituyen fortalezas del lado chino. De hecho, China alcanzó una cohesión de 0,988 y metadatos de 0,890, lo que muestra una integración estructural muy avanzada y descripciones bibliográficas completas. Cabe señalar que la cohesión casi perfecta de las universidades chinas indica una madurez notable en la organización interna de sus sitios web, mientras que las universidades españolas también se aproximan a ese nivel. Sin embargo, la

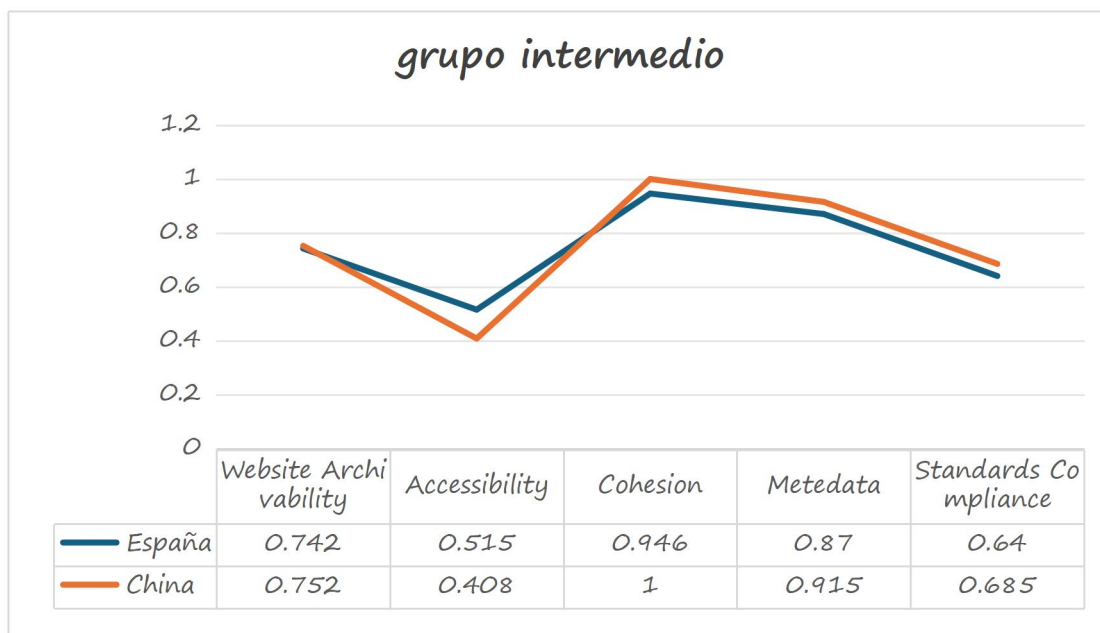
ventaja española en cumplimiento normativo podría estar asociada a la existencia de un marco legislativo consolidado sobre preservación digital, como el RD 635/2015.



Gráfica 3. Comparación de los indicadores de archivabilidad web en universidades del grupo de élite

Comparativa del grupo intermedio:

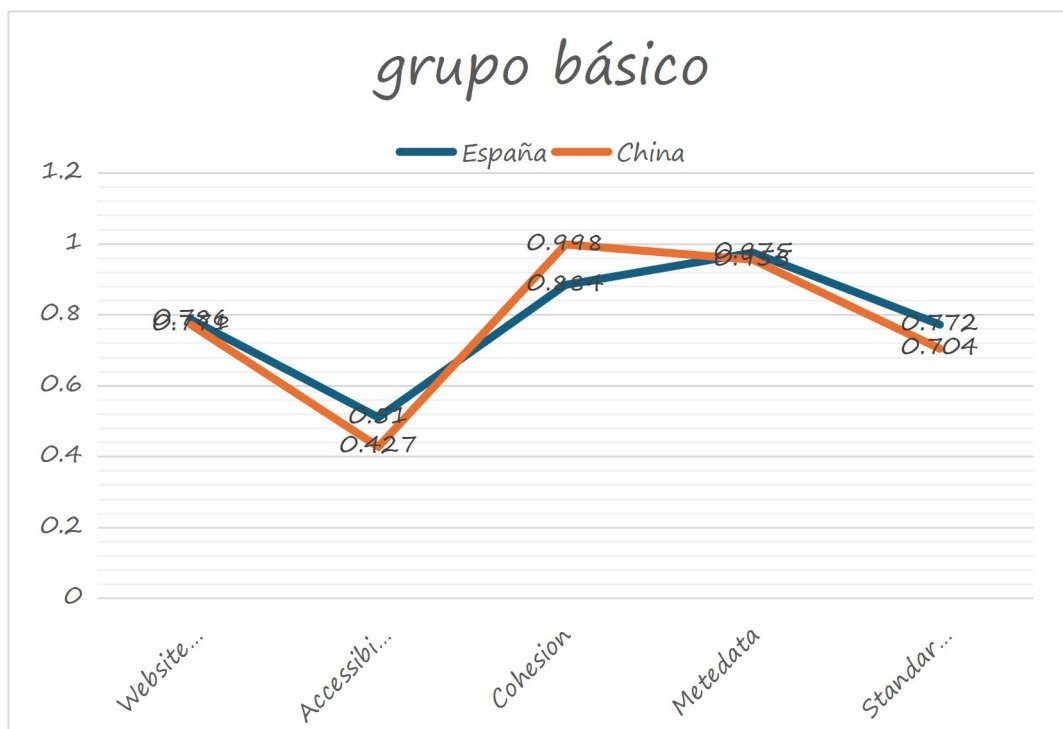
En el grupo intermedio, la diferencia en la archivabilidad web entre China (0,752) y España (0,742) se reduce, aunque los componentes internos divergen considerablemente. Las universidades chinas siguen liderando en metadatos (0,915) y alcanzan una cohesión estructural perfecta (1,000), lo cual subraya su énfasis en la uniformidad de diseño y contenido. Por otro lado, las instituciones españolas mantienen su liderazgo en Accesibilidad (0,515), aunque con menor margen que en el grupo de élite. Es importante destacar que China presenta una puntuación baja en Accesibilidad (0,408), lo que revela deficiencias comunes en aspectos como la navegación, el etiquetado semántico o las estrategias de presentación visual, lo cual exige reformas estructurales más profundas.



Gráfica 4. Comparación de los indicadores de archivabilidad web en universidades del grupo intermedio

Comparativa del grupo básico:

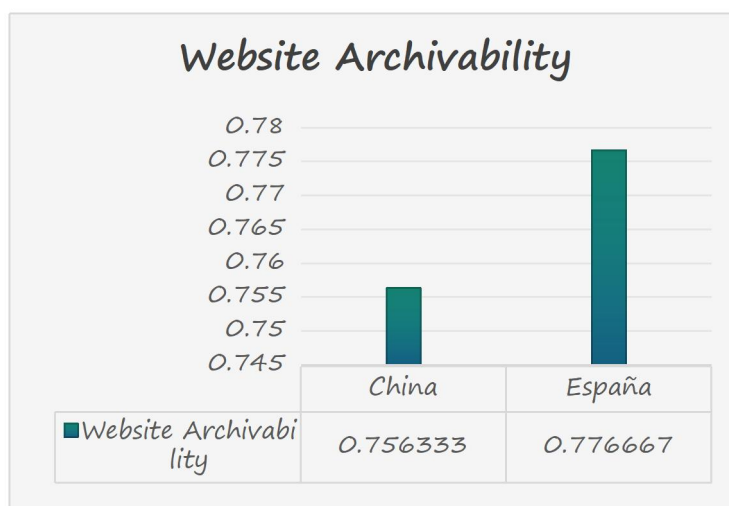
En el grupo inferior, China muestra una puntuación media de archivabilidad de 0,771, que aún se sitúa ligeramente por debajo del promedio español de 0,786. Aunque China mantiene su ventaja en Metadatos (0,955) y Cohesión (0,998), vuelve a quedar rezagada en el Cumplimiento de Estándares (0,704 frente a 0,772 en España). Esta brecha es más notable en Accesibilidad, donde las universidades españolas obtienen 0,510 frente a 0,427 en China. Esto sugiere que, aunque muchas universidades chinas de este grupo dependen de plataformas centralizadas para mantener la uniformidad estructural, todavía enfrentan limitaciones evidentes en cuanto a la implementación técnica y la experiencia del usuario, lo que limita su desempeño general en términos de archivabilidad.



Gráfica 5. Comparación de los indicadores de archivabilidad web en universidades del grupo básico

5.3 Análisis comparativo de indicadores clave

El análisis comparado de indicadores clave muestra lo siguiente:



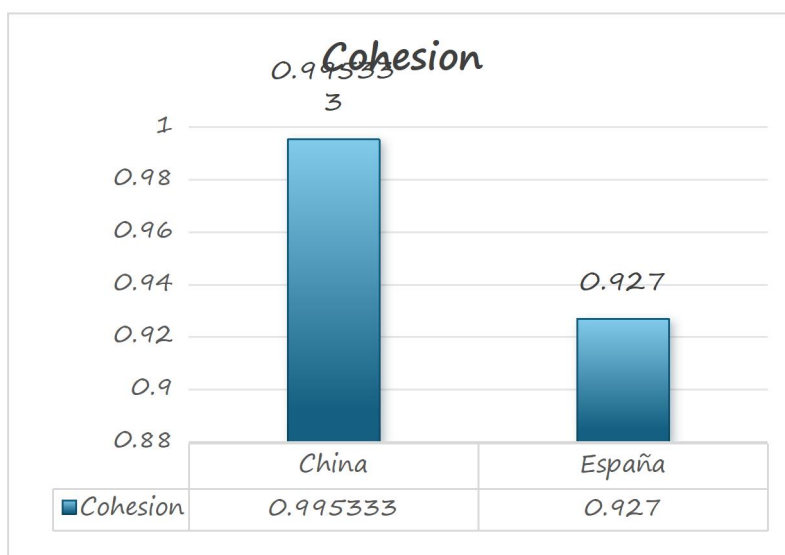
Gráfica 6. Comparación de la archivabilidad general de los sitios web entre España y China

En primer lugar, el indicador general de archivabilidad web (Website Archivability), en España se obtiene una puntuación ligeramente superior en general (0,776 vs. 0,756), lo cual se debe principalmente a una mejor normalización técnica y a medidas impulsadas por el marco legal.



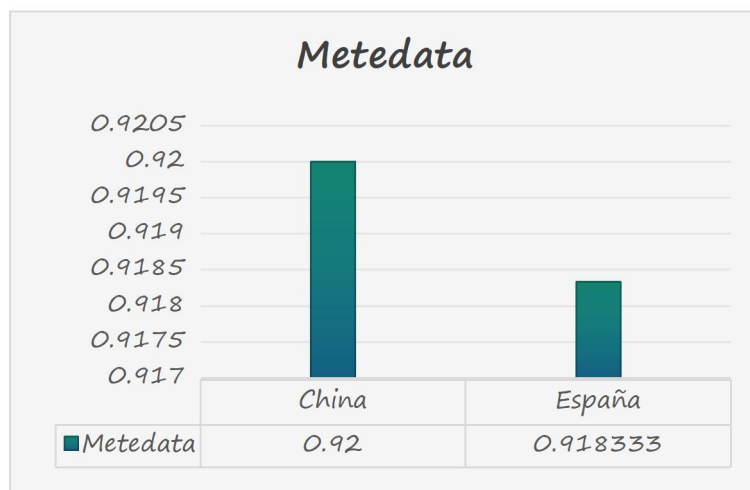
Gráfica 7. Comparación del nivel de accesibilidad web en bibliotecas universitarias de España y China

En relación con el indicador de accesibilidad (Accessibility), la diferencia es notable (España 0,525 vs. China 0,414). En los sitios chinos se detectan con frecuencia enlaces rotos, páginas sin respuesta o bloqueos por el archivo robots.txt.



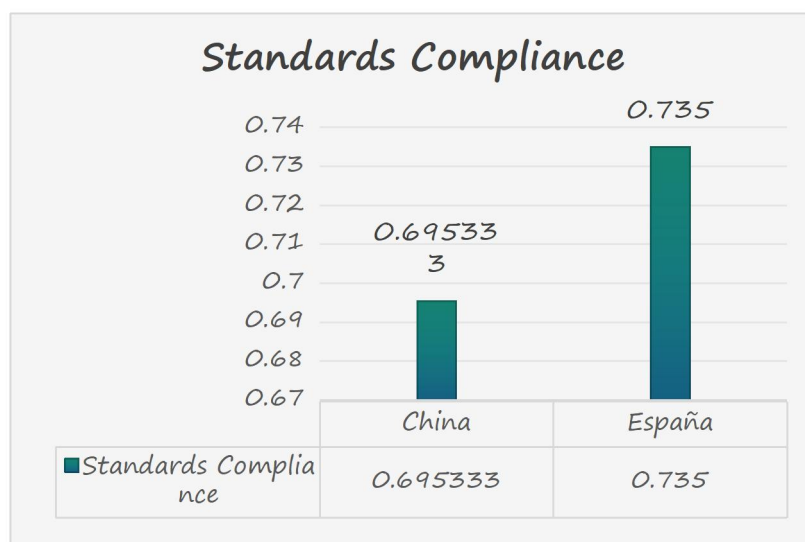
Gráfica 8. Comparación del nivel de cohesión estructural en los sitios web de bibliotecas universitarias

En cuanto a la cohesión estructural (Cohesion), las universidades chinas muestran un rendimiento sobresaliente (casi puntuación máxima), lo que refleja una fuerte conciencia de gestión estructural unificada.



Gráfica 9. Comparación de la integridad de los metadatos en los sitios web de bibliotecas universitarias

En lo referente al indicador de integridad de metadatos (Metadata), las universidades chinas logran puntuaciones estables y altas gracias al uso de diseños por plantillas y campos predefinidos. España obtiene puntuaciones algo más bajas, pero en general también presenta un nivel bueno.



Gráfica 10. Comparación del nivel de conformidad con los estándares en los sitios web de bibliotecas universitarias

Finalmente, en el indicador cumplimiento de estándares (Standards Compliance), España muestra una clara ventaja sistémica, con mayor coherencia y conformidad en validaciones W3C y el uso de etiquetas estructurales semánticas.

Desde un enfoque general, los sitios web de las bibliotecas universitarias españolas, se encuentran respaldadas por un marco legal claro y una implementación técnica de largo plazo, razón por la que presentan un sistema de preservación de sus sitios web más maduro y caracterizado por una “amplia cobertura y un alto nivel de normalización.

Por otro lado, las universidades chinas destacan en coherencia estructural y gestión de metadatos, pero afrontan limitaciones debido a la falta de apoyo legislativo y a una infraestructura técnica desigual, lo que genera cierta inestabilidad en la ejecución técnica.

Es importante señalar que las universidades chinas del grupo intermedio e inferior muestran un rendimiento relativamente equilibrado. Algunas han logrado mejorar su desempeño mediante el uso de plantillas estandarizadas, lo que refleja un potencial de mejora a través de la normalización. No obstante, si no se avanza simultáneamente en la optimización de los servidores y la compatibilidad con rastreadores web, estas ventajas técnicas podrían verse reducidas en el proceso real de archivado.

5.4 Evaluación global y tendencias generales

Desde una perspectiva general, las universidades españolas, respaldadas por un marco legal claro y una implementación técnica de largo plazo, presentan un sistema de preservación web más maduro, caracterizado por una “amplia cobertura y alto nivel de normalización”.

Por otro lado, las universidades chinas destacan en coherencia estructural y gestión de metadatos, pero enfrentan limitaciones debido a la falta de apoyo legislativo y a una infraestructura técnica desigual, lo que genera cierta inestabilidad en la ejecución técnica.

Es importante señalar que las universidades chinas del grupo intermedio e inferior muestran un rendimiento relativamente equilibrado. Algunas han logrado mejorar su desempeño mediante el uso de plantillas estandarizadas, lo que refleja un potencial de mejora a través de la normalización. No obstante, si no se avanza simultáneamente en la optimización de los servidores y la compatibilidad con rastreadores web, estas ventajas técnicas podrían verse reducidas en el proceso real de archivado.

Capítulo 6 Análisis de casos. Enfoque particularizado.

Con el objetivo de profundizar en la capacidad explicativa de los datos estadísticos presentados anteriormente, este capítulo realiza un análisis de casos sobre sitios web de bibliotecas universitarias representativas de China y España. A través del desglose de indicadores y la comparación estructural de las páginas, se examinan las manifestaciones concretas de los problemas relacionados con la archivabilidad. Los criterios de selección han sido los siguientes:

- **Una universidad del grupo de élite:** representa a las instituciones más destacadas del país, con los mayores niveles de recursos y capacidad tecnológica. Su análisis permite observar las mejores prácticas y estándares más avanzados en preservación web.
- **Una universidad del grupo intermedio:** representa a universidades de perfil medio, permitiendo examinar las estrategias adoptadas en contextos de recursos limitados pero con cierto grado de infraestructura digital consolidada.
- **Una universidad del grupo inferior:** representa a instituciones con menor dotación tecnológica y visibilidad internacional. Su inclusión permite identificar las principales dificultades y carencias en materia de preservación digital, sirviendo como muestra de los desafíos más comunes en contextos desfavorables.

6.1 Caso 1: Universidad de Tsinghua (China) vs Universidad de Barcelona (España) – Grupo de élite

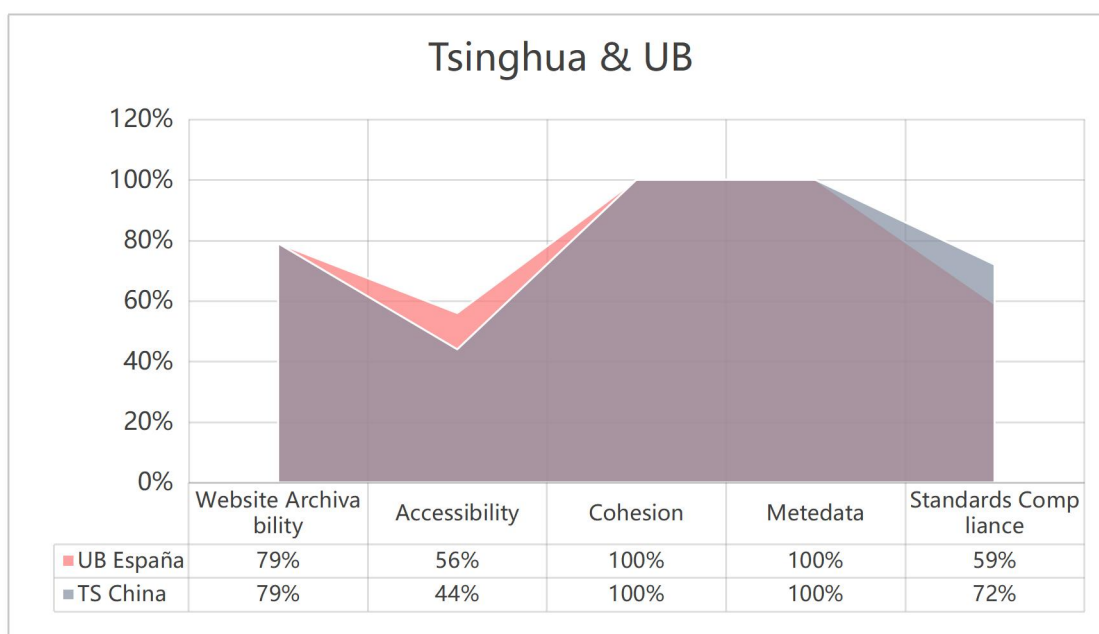
Este primer caso enfrenta a dos de las universidades más prestigiosas de sus respectivos países: la Universidad de Tsinghua, reconocida como una de las principales instituciones de educación superior en China, y la Universidad de Barcelona, considerada una de las universidades líderes en el ámbito académico español. Ambas pertenecen al grupo de élite según los rankings internacionales y cuentan con una infraestructura tecnológica altamente desarrollada.

En la tabla 9 se aportan los datos de la comparativa para los criterios analizados.

Nombre	País	Website Archivability	Accessibilidad	Cohesión	Metadatos	Standards Compliance
Tsinghua	China	79%	44%	100%	100%	72%
UB	España	79%	56%	100%	100%	59%

Tabla 9. Comparativa entre dos universidades de élite de China y España

Se observan diferencias en accesibilidad y cumplimiento de normas y coincidencias en el resto de criterios. Seguidamente se ofrece una representación y estudio de sus características del sitio web para cada una de las universidades.



Gráfica 11. Comparación de los indicadores de archivabilidad entre la Universidad de Barcelona (España) y la Universidad de Tsinghua (China)

Características específicas de la Universidad de Tsinghua:

El sitio web de la biblioteca de la Universidad de Tsinghua presenta una estructura general clara, con un uso coherente de plantillas y un alto grado de contenido estático, lo que favorece en principio su archivabilidad. Además, dispone de metadatos bien estructurados y completos, lo que facilita la descripción y recuperación posterior del contenido.

El enlace en el que se observan los principales fallos disponible en <https://archiveready.com/download-results/html/432381>

Sin embargo, parte de su contenido depende de la carga dinámica mediante JavaScript, lo que introduce barreras técnicas para herramientas de archivo automatizadas. A ello se suma la presencia de un archivo robots.txt con restricciones relativamente estrictas, que limita el acceso de los rastreadores web a ciertas secciones del sitio. Estas configuraciones afectan negativamente la accesibilidad del sitio para procesos de captura automatizada, reduciendo su archivabilidad efectiva.

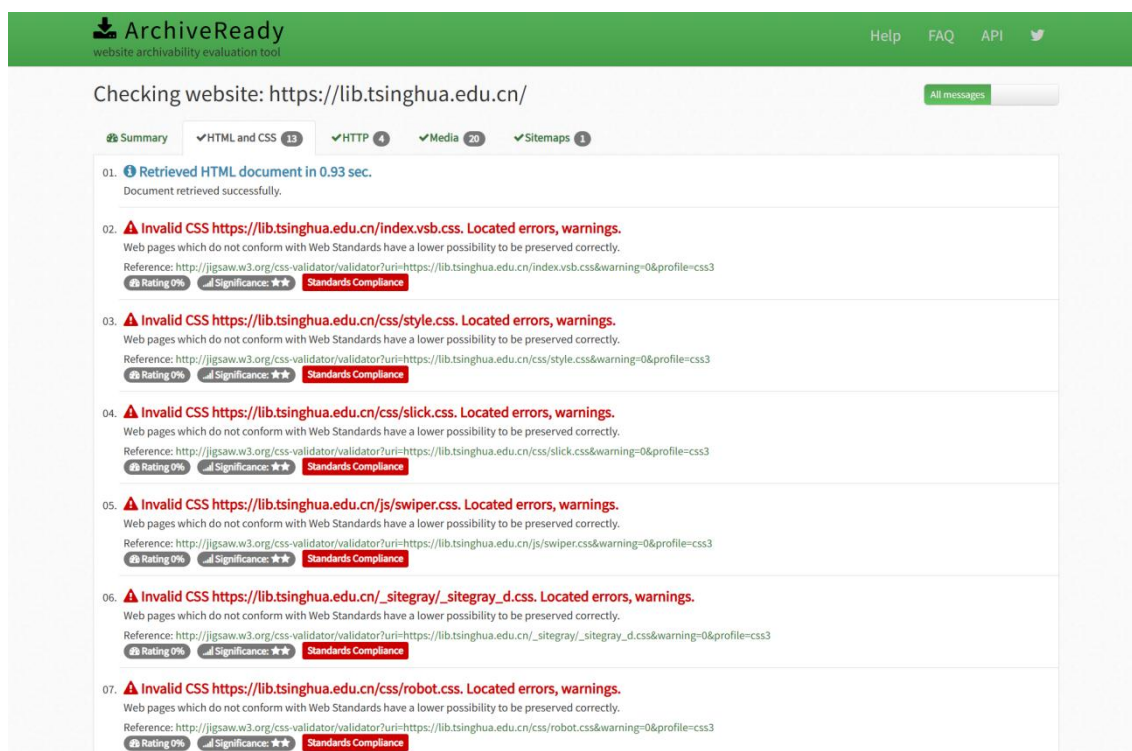


Figura 10. Resultados de la evaluación de estándares web del sitio de la Biblioteca de Tsinghua

Características específicas de la Universidad de Barcelona:

El sitio web de la Universidad de Barcelona presenta una alta legibilidad de las páginas y buena accesibilidad, con una navegación clara y estructurada, lo que facilita tanto la experiencia del usuario como la exploración automatizada por parte de herramientas de archivado.

No obstante, el portal muestra problemas de estandarización del código, como el uso inconsistente de etiquetas HTML y la presencia de páginas que no superan la validación W3C, lo cual afecta negativamente su puntuación en términos de conformidad técnica y puede generar dificultades durante la indexación o preservación estructural a largo plazo.

El enlace en el que se observan los principales fallos disponible en <https://archiveready.com/download-results/html/432384> muestra fallos vinculados para el CRAI en robot, mapa de sitios, CSS entre otros, considerando el validador CSS.

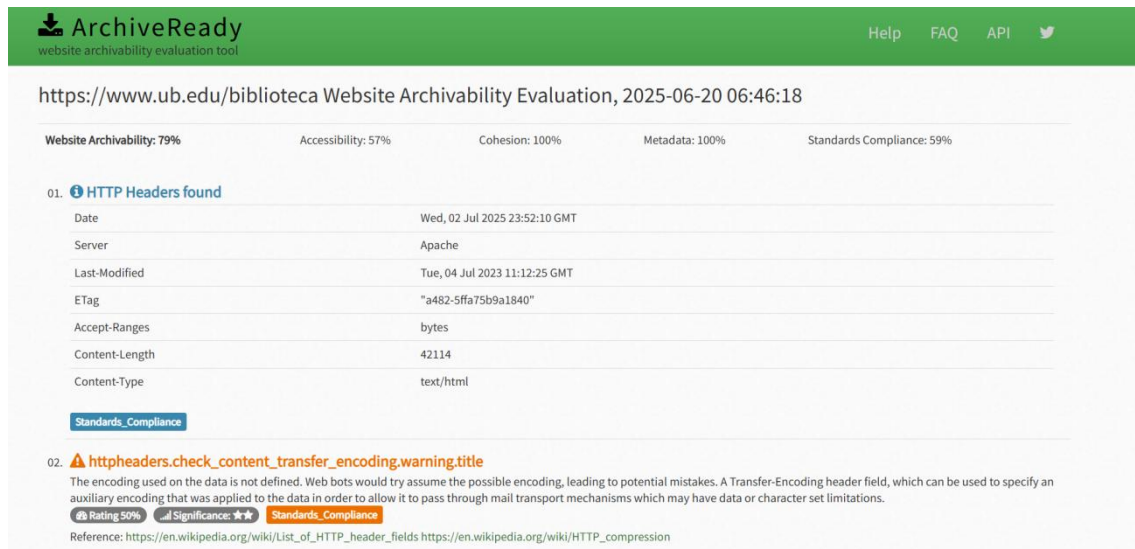


Figura 11. Informe de archivabilidad de la web de la Biblioteca de la UB

Conclusión comparativa:

Ambos sitios web presentan un desempeño similar en cuanto a estructura general y calidad de metadatos, reflejando niveles elevados de organización y documentación. Pero existen diferencias notables en la forma en que cada institución implementa su arquitectura técnica:

- La Universidad de Tsinghua, si bien demuestra una infraestructura avanzada, adopta una implementación tecnológica más compleja, con servicios interactivos como el sistema de reservas o el acceso a documentos que dependen del renderizado del lado del cliente. Esta dependencia introduce barreras para su captura archivística, reduciendo la calidad e integridad de las versiones archivadas.
- La Universidad de Barcelona, por su parte, muestra una estructura más convencional y técnicamente estable, sin grandes innovaciones en la presentación o la interacción, pero con una mayor compatibilidad con herramientas de rastreo automático y una arquitectura más amigable para los procesos de preservación web.

6.2 Caso 2: Universidad Normal Central de China vs Universidad de Salamanca (España) – Grupo intermedio

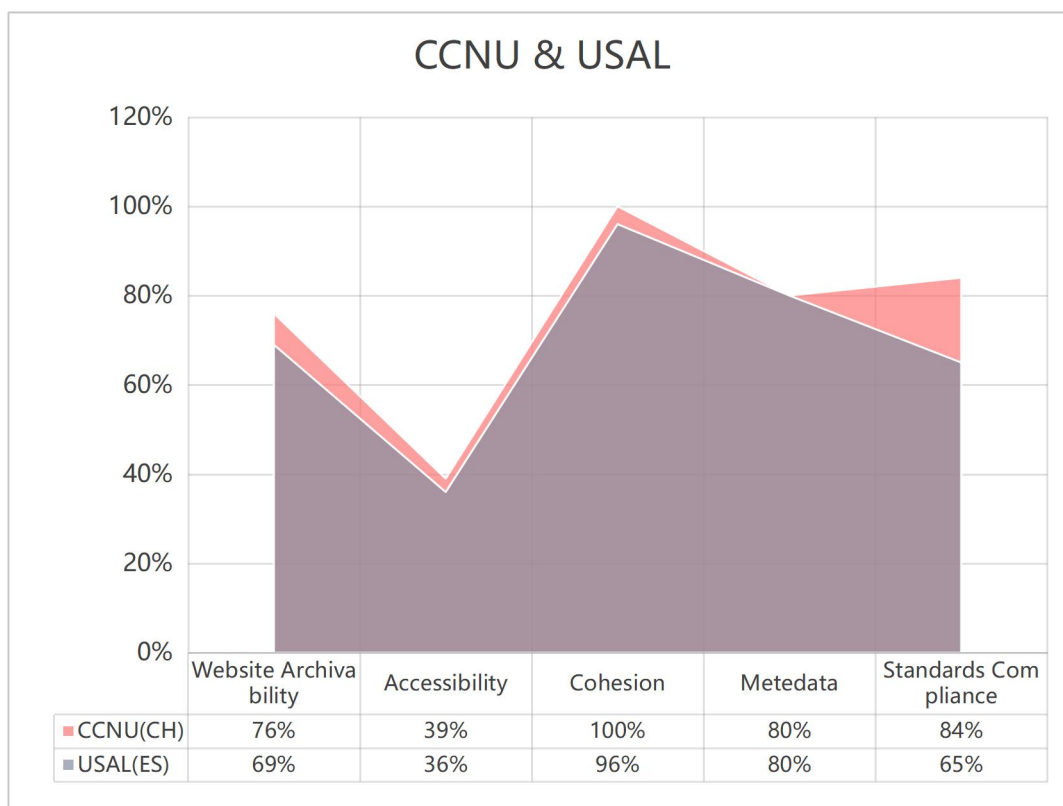
Ambas instituciones poseen una infraestructura digital funcional, aunque con recursos más limitados que las universidades del grupo de élite. El análisis permite observar cómo enfrentan los desafíos de archivabilidad en contextos de desarrollo tecnológico medio. Al igual que en el grupo de élite se muestran los porcentajes y la representación de cada una con su análisis de sus características.

La tabla 10 incluye el sitio de la biblioteca de la CCNU y el de la Universidad de Salamanca, en el que esta última aporta % menores.

Nombre	país	Website Archivability	Accessibility	Cohesion	Metadata	Standards Compliance
CCNU	China	76%	39%	100%	80%	84%
USAL	Espana	69%	36%	96%	80%	65%

Tabla 10- Comparativa de dos universidades intermedias China y España.

Si bien, tal y como se observa en la gráfica dichos porcentajes son relevantes especialmente para la USAL en el cumplimiento de estándares.



Gráfica 12. Comparación de los indicadores de archivabilidad entre la Huazhong Normal University (China) y la Universidad de Salamanca (España)

Características específicas de la Universidad Normal Central de China (CCNU):

En términos específicos, el sitio web de la biblioteca de CCNU utiliza un sistema de desarrollo propio con abundantes elementos incrustados mediante iframes. Se detecta una falta de estandarización en la marcación de metadatos, ya que muchas páginas carecen de etiquetas <title> o <meta>, lo cual complica la descripción y recuperación automatizada de la información. Además, se observan redirecciones frecuentes de URL (por ejemplo, de /library a /web/lib), lo que afecta la coherencia estructural del sitio. Durante el proceso de análisis se identificaron numerosos errores 403 y 404, lo que indica problemas significativos de estabilidad de acceso.

El enlace en el que se observan los principales fallos disponible en <https://archiveready.com/download-results/html/432385>

x-frame-options	SAMEORIGIN
x-frame-options	SAMEORIGIN
frame-options	SAMEORIGIN
accept-ranges	bytes
vary	Accept-Encoding
x-envoy-upstream-service-time	0

Standards Compliance

02. **Transfer encoding was clearly defined in HTTP Headers.**
 A Transfer-Encoding header field, which can be used to specify an auxiliary encoding that was applied to the data in order to allow it to pass through mail transport mechanisms which may have data or character set limitations.
 Rating 100% Significance: ★★ Standards Compliance
 Reference: https://en.wikipedia.org/wiki/List_of_HTTP_header_fields https://en.wikipedia.org/wiki/HTTP_compression
03. **Content type was clearly defined.**
 Correctly defined HTTP headers help web bots understand the content.
 Rating 100% Significance: ★★ Standards Compliance Metadata
 Reference: <http://en.wikipedia.org/wiki/MIME>
04. **HTTP caching headers are not available.**
 HTTP caching headers are critical to improve web archiving bots performance because they help avoid downloading unmodified content multiple times.
 Rating 50% Significance: ★★ Accessibility Metadata
 Reference: <https://developers.google.com/speed/docs/best-practices/caching>
05. **No sitemap.xml found.**
 Sitemaps are an easy way for webmasters to inform web bots about available web pages so that they can more intelligently crawl the site.
 Rating 0% Significance: ★★ Accessibility
 Reference: <http://www.sitemaps.org/>

Figura 12. Detalle técnico de la evaluación de archivabilidad web del sitio de la Biblioteca de la Universidad Normal Central de China

Características específicas de la Universidad de Salamanca:

Por su parte, la página web de la biblioteca de la Universidad de Salamanca presenta una estructura general clara y jerárquica, pero el uso de múltiples iframes anidados dificulta la captura de contenido a nivel de página única. Además, muchos recursos académicos o secciones específicas (como bases de datos o plataformas de investigación) están alojados en dominios externos, lo que impide una cobertura archivística centralizada. La funcionalidad de cambio de idioma modifica las rutas base de las URL, reduciendo la coherencia de los enlaces entre versiones del mismo contenido.

El enlace en el que se observan los principales fallos disponible en <https://archiveready.com/download-results/html/432390>

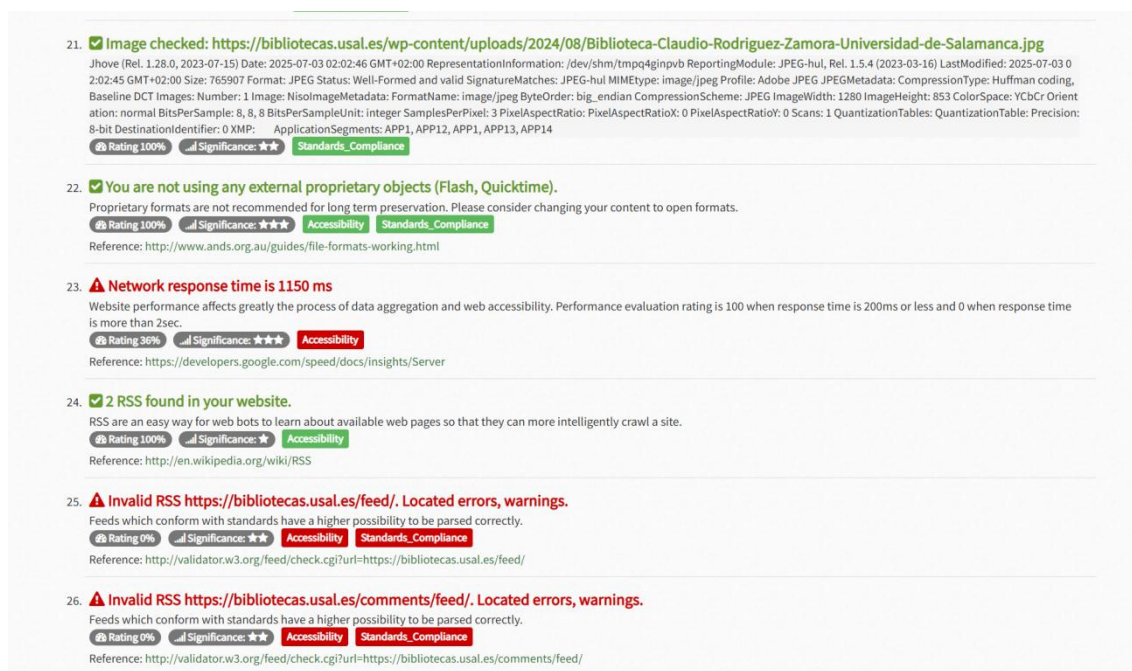


Figura 13. Detalle técnico de la evaluación de archivabilidad web del sitio de la Biblioteca de la Universidad de Salamanca

Conclusión comparativa:

Ambos sitios presentan un rendimiento técnico medio y muestran dificultades distintas en la archivabilidad:

- En términos comparativos, ambos sitios presentan limitaciones derivadas de sus decisiones técnicas. CCNU destaca por una mayor coherencia en la organización de enlaces internos, pero sufre por una implementación poco estandarizada, con un uso excesivo de iframes, redirecciones no documentadas y deficiencias en metadatos. Por otro lado, USAL cumple de manera más consistente con los estándares técnicos y semánticos, pero la fragmentación estructural, el alojamiento externo de contenidos y la inestabilidad en las rutas de enlace afectan negativamente su archivabilidad.
- En conclusión, CCNU ofrece mayor consistencia interna en los enlaces, pero requiere importantes mejoras en estandarización y accesibilidad. USAL, si bien técnicamente más conforme a los estándares, debe abordar los problemas estructurales que dificultan una captura archivística integral.

6.3 Caso 3: Universidad de Jimei (China) vs Universidad Carlos III de Madrid (España) – Grupo inferior

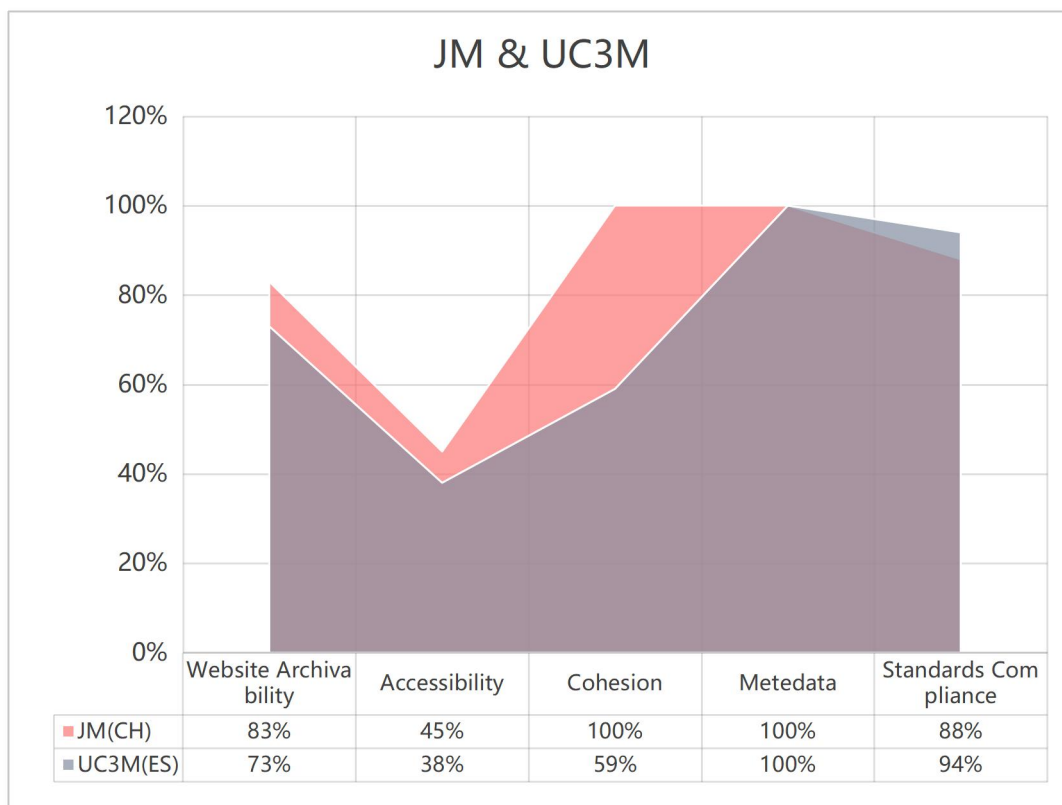
Este tercer caso se centra en dos universidades que pertenecen al grupo inferior en términos de posicionamiento académico global. Ambas instituciones representan contextos con recursos digitales limitados y niveles de visibilidad internacional relativamente bajos. El análisis de sus sitios web permite identificar los principales obstáculos que enfrentan en materia de archivabilidad.

La tabla 11 aporta los porcentajes de los sitios web de las bibliotecas de Jimei y de la Universidad Carlos III de Madrid.

Nombre	pais	Website Archivability	Accessibility	Cohesion	Metadata	Standards Compliance
Jimei	China	83%	45%	100%	100%	88%
UC3M	Espana	73%	38%	59%	100%	94%

Tabla 11. Comparativa de dos universidades de grupo inferior de China y España.

Tal y como se observa en la gráfica comparada, se muestran notables desequilibrios y coincidencia únicamente en metadatos.



Gráfica 13. Comparación de los indicadores de archivabilidad entre Jiangxi Normal University (China) y la Universidad Carlos III de Madrid (España)

Características específicas de la Universidad de Jimei:

El sitio web de la biblioteca utiliza la plantilla unificada del Ministerio de Educación de China (EduCMS), lo que garantiza una estructura coherente y estandarizada. Las etiquetas HTML están bien implementadas, con un alto grado de estandarización y ausencia de errores sintácticos. La mayoría de las páginas son archivos HTML estáticos, lo que facilita su captura por herramientas de archivado. El servidor es estable, pero el sitio carece de adaptación multilingüe y no implementa protocolos abiertos como OAI-PMH. Además, la frecuencia de actualización es baja, y algunos enlaces presentan problemas de caducidad.

El enlace en el que se observan los principales fallos disponible en <https://archiveready.com/download-results/html/432404>

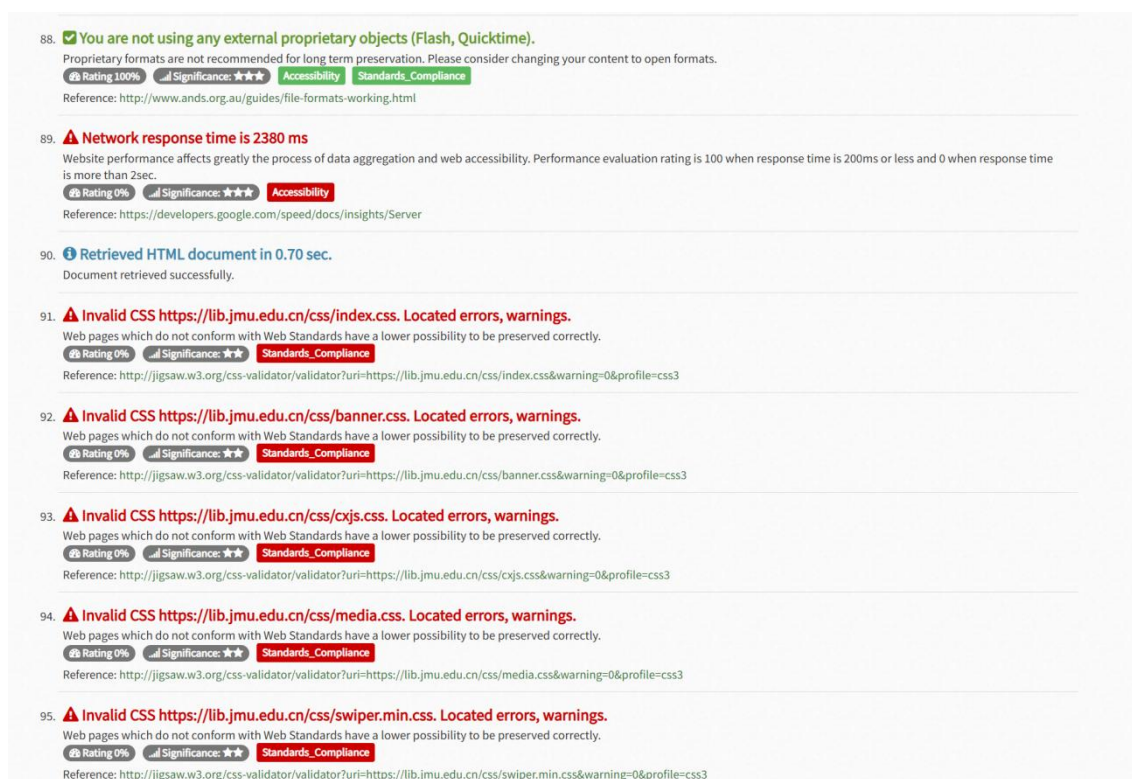


Figura 14. Detalle técnico de la evaluación de archivabilidad web del sitio de la Biblioteca de La Universidad de Jimei

Características específicas de la Universidad Carlos III de Madrid:

El sitio emplea un diseño responsivo, compatible con estándares de accesibilidad. Dispone de metadatos completos, compatibles tanto con Schema.org como con Dublin Core. No obstante, su estructura es compleja, con múltiples bloques independientes y fragmentación de enlaces, lo que resulta en una puntuación baja en cohesión (59%). En cuanto a accesibilidad, parte del contenido requiere redirecciones

a subdominios o plataformas externas, lo que complica la navegación para los rastreadores. A pesar de ello, su nivel de conformidad técnica es excelente, sirviendo como un modelo de referencia para la validación W3C.

El enlace en el que se observan los principales fallos disponible en <https://archiveready.com/download-results/html/432407>

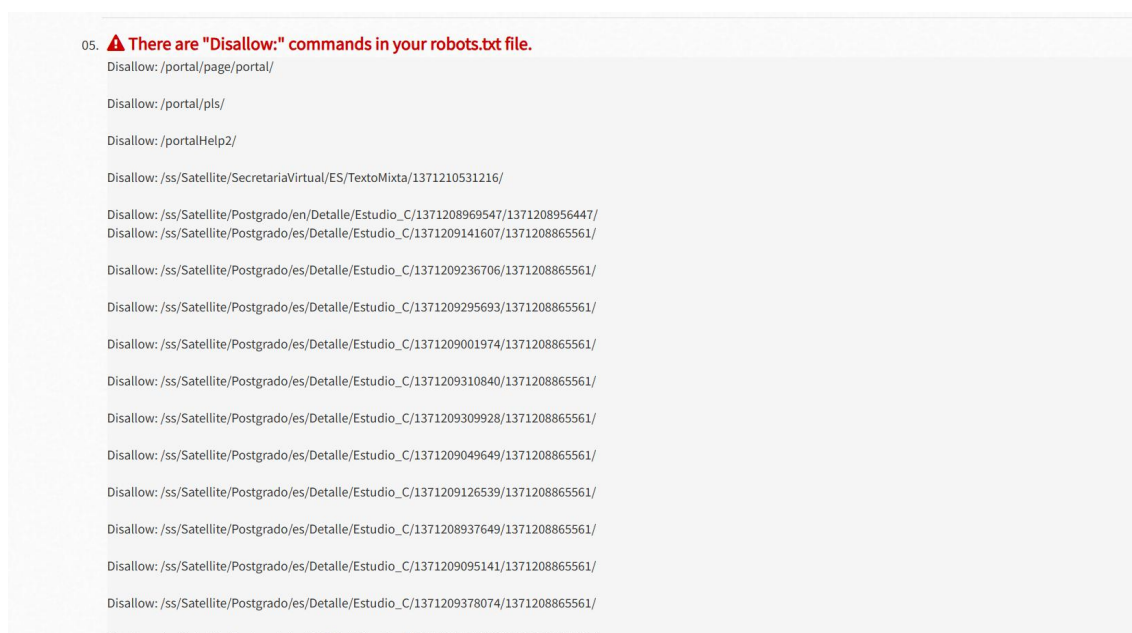


Figura 15. Detalle técnico de la evaluación de archivabilidad web del sitio de la Biblioteca de La Universidad Carlos III

Conclusión comparativa:

Ambas universidades presentan fortalezas distintas en archivabilidad dentro del grupo de instituciones con menor nivel de recursos.

- La Universidad de Jimei destaca por su diseño estructural claro y uniforme, lo que facilita considerablemente el proceso de archivo. Su uso de plantillas consistentes mejora significativamente los puntajes de metadatos y cohesión, convirtiéndola en un ejemplo típico de cómo las universidades chinas de menor rango pueden lograr una archivabilidad efectiva a bajo coste gracias a la estandarización y las políticas institucionales.
- La Universidad Carlos III, aunque muestra cierta debilidad estructural por la fragmentación de enlaces y saltos frecuentes entre módulos, presenta una alta calidad técnica, con un marcado cumplimiento de los estándares de accesibilidad, codificación semántica y validación. Este enfoque demuestra los avances de las universidades españolas en aspectos normativos y metadatos

semánticos, compensando las limitaciones estructurales con una fuerte base de interoperabilidad.

6.4 Consideraciones conclusivas

A través de la comparación entre los tres niveles institucionales analizados, se evidencian patrones diferenciados entre China y España en cuanto a la archivabilidad de los sitios web universitarios:

- Las universidades chinas muestran una ventaja sistémica en la consistencia estructural y la configuración de metadatos. Esta fortaleza se manifiesta especialmente en las instituciones de nivel medio e inferior, que han logrado mejorar significativamente su capacidad de ser archivadas mediante estrategias de estandarización y uso de plantillas unificadas.
- Las universidades españolas, por su parte, destacan en aspectos como la accesibilidad web y la conformidad con estándares internacionales, impulsadas por un entorno normativo más desarrollado y una mayor capacidad técnica para aplicar buenas prácticas en el diseño web.
- En última instancia, el análisis sugiere que el éxito en la preservación web no depende exclusivamente de los recursos tecnológicos disponibles, sino sobre todo del grado de concienciación institucional respecto a la archivabilidad y de la capacidad para implementar estrategias integrales y sostenidas en el tiempo.

Capítulo 7. Conclusiones, análisis de causas y recomendaciones sobre la investigación

Este estudio ha llevado a cabo una evaluación sistemática de la archivabilidad de los sitios web de bibliotecas universitarias en China y España, combinando análisis cuantitativo con estudios de caso representativos. A partir de los resultados obtenidos, se han identificado diferencias estructurales, técnicas y organizativas en las prácticas de preservación digital entre ambos países. A continuación, se desarrollan los hallazgos principales desde tres dimensiones: conclusiones generales, análisis de causas y propuestas de mejora.

7.1 Conclusiones del estudio

Este estudio se centra en la archivabilidad de los sitios web de las bibliotecas universitarias, tomando como muestra a universidades de China y España. A través de un análisis comparativo desde múltiples dimensiones —estructura general, implementación técnica, políticas legales y condiciones de recursos— se busca responder a la pregunta central: ¿qué dificultades enfrentan los sitios web de bibliotecas universitarias durante el proceso de archivado?, así como explorar los mecanismos profundos que influyen en la eficacia del mismo.

En términos generales, el estudio revela que las bibliotecas universitarias de ambos países cuentan con una base relativamente sólida para la preservación web, con puntuaciones de evaluación superiores al 0,75 en la mayoría de los casos. Esto indica que tanto en China como en España, la infraestructura digital cumple con condiciones básicas para permitir la captura y conservación. No obstante, al analizar en detalle, las universidades de ambos países muestran fortalezas destacadas y debilidades estructurales diferenciadas, reflejando contrastes tanto en la lógica técnica como en la institucional.

Las universidades chinas, en especial las de nivel medio e inferior, suelen adoptar plantillas unificadas y campos estandarizados, lo cual mejora notablemente la coherencia estructural de los sitios y la integridad de sus metadatos. Esta estrategia de estandarización impulsada desde niveles superiores ha demostrado ser eficaz en contextos con recursos limitados, facilitando la operación del archivado y reduciendo la complejidad de captura. Sin embargo, también se evidencian deficiencias como baja interactividad, escaso cumplimiento de estándares internacionales y configuraciones

inestables de los servidores, lo que restringe la fidelidad de la preservación en la práctica.

Las universidades españolas, por su parte, presentan un desempeño superior en términos de accesibilidad web y conformidad con estándares semánticos. Gracias a un mayor uso de etiquetas semánticas, codificación HTML estandarizada y diseño responsivo, sus sitios permiten una mejor detección por herramientas de archivado y una reconstrucción estructural más precisa. Además, desde 2009, España aplica una política de depósito legal para sitios web, que establece un marco normativo claro para su conservación y define responsabilidades institucionales específicas. En contraste, aunque China cuenta con leyes como la Ley de Bibliotecas o la Ley de Ciberseguridad, aún carece de un régimen legal específico dedicado al archivado web, lo que ha generado una falta de orientación institucional y estándares unificados en las prácticas de preservación.

Cabe destacar que el estudio también identifica una serie de obstáculos comunes en la práctica, tales como la configuración inadecuada de archivos robots.txt que limita el rastreo, la pérdida de contenido por dependencia de JavaScript para el renderizado, la fragmentación de enlaces por alojamiento externo de recursos, y las rutas inconsistentes entre versiones multilingües que dificultan el reconocimiento estructural. Estos problemas se presentan en ambos países, lo que demuestra que las dificultades del archivado web no solo dependen de la inversión en recursos, sino también del diseño conceptual y la normalización técnica del entorno digital.

7.2 Análisis de causas

Detrás de las diferencias anteriormente descritas, el diseño institucional y los mecanismos de ejecución son, sin duda, los factores más fundamentales. La razón por la cual las universidades españolas han logrado una mayor estandarización y conformidad en el archivado web es que la Biblioteca Nacional del país posee autoridad legal para realizar recolecciones obligatorias, acompañadas de mecanismos sancionadores. Este tipo de protección legal de arriba hacia abajo obliga a que, incluso en condiciones de recursos limitados, las universidades deban cumplir con estándares básicos de preservación. De hecho, entre sus dominios, se recopila periódicamente el rastreo de las universidades españolas.

Por el contrario, China aún no ha establecido un marco legal obligatorio similar, lo que hace que el archivo web dependa en gran medida de la iniciativa propia de las universidades o del impulso de proyectos de investigación, careciendo de un mecanismo de ejecución uniforme.

Desde la perspectiva del ecosistema técnico, el sistema educativo chino utiliza ampliamente plataformas de plantillas unificadas, como EduCMS, lo que ha permitido a muchas universidades construir rápidamente sitios web con estructuras uniformes y formatos estandarizados. Aunque este modelo centralizado presenta una ventaja significativa en términos de coherencia estructural, también limita la capacidad de los rastreadores web debido a problemas como la actualización poco frecuente y una baja accesibilidad.

En cambio, las universidades españolas suelen utilizar plataformas CMS de código abierto y estructuras de servicios distribuidos, lo que favorece la expresividad semántica y la compatibilidad con estándares de accesibilidad. Sin embargo, la inestabilidad de los enlaces entre sitios y la fragmentación de rutas afectan negativamente su archivabilidad global.

En cuanto al mecanismo organizativo, también existen diferencias en la asignación de responsabilidades dentro de las universidades. En la mayoría de los casos en China, los departamentos de gestión de redes son los encargados del diseño y mantenimiento de los sitios web, y las bibliotecas desempeñan un papel secundario, por lo que la preservación web rara vez se considera una prioridad. En cambio, en España, las bibliotecas asumen directamente funciones clave en la preservación de recursos informativos y el acceso abierto, por lo que la conformidad, accesibilidad y capacidad de preservación a largo plazo de los sitios web forman parte de sus objetivos institucionales. Esto refleja diferencias culturales e institucionales en la comprensión del rol de la biblioteca digital.

Además, en cuanto a la concepción del diseño web, las universidades chinas suelen centrarse más en la integración funcional y la presentación visual, dando prioridad a portales integrados de servicios, pero a menudo descuidan la legibilidad estructural y la compatibilidad con rastreadores. Por el contrario, las universidades españolas prestan más atención a la descripción semántica, las normas de codificación y la interoperabilidad de los datos, y, bajo el impulso de políticas de ciencia abierta, están empezando a considerar la archivabilidad web como parte integrante de la gestión a largo plazo de la información académica.

7.3 Recomendaciones de política

En respuesta a los problemas mencionados anteriormente, este estudio propone las siguientes recomendaciones, con el objetivo de promover un desarrollo de mayor nivel en la práctica del archivado web universitario, partiendo de las realidades específicas de China y España.

Para las universidades chinas, se debería promulgar lo antes posible una normativa con fuerza legal sobre el archivado web, que defina claramente las responsabilidades y obligaciones de los distintos tipos de universidades en la preservación de recursos digitales. Se recomienda establecer, a nivel de la Biblioteca Nacional o del Ministerio de Educación, una plataforma unificada de archivado, que combine un mecanismo de entrega obligatoria y evaluaciones anuales, formando así una red de preservación sistemática y sostenida. Sobre la base de los sistemas de plantillas existentes, se debería mejorar su compatibilidad con la preservación digital, incluyendo la apertura del archivo robots.txt, el aumento del cumplimiento de la sintaxis HTML y la incorporación de funciones que permitan el renderizado de contenido dinámico. Asimismo, se sugiere que las universidades integren la archivabilidad web en los sistemas de evaluación del nivel de informatización, y establezcan el puesto de “responsable del archivado web”, reforzando el papel funcional de las bibliotecas o departamentos de información en la preservación digital a largo plazo.

Para las universidades españolas, aunque ya cuentan con ventajas notables en mecanismos legales y cumplimiento normativo, todavía hay margen de mejora en cuanto a la uniformidad estructural y organización del contenido. Se recomienda que la Biblioteca Nacional lidere el desarrollo de componentes comunes de navegación del sitio, mapas del sitio y plantillas estructurales para las universidades de nivel medio e inferior, con el fin de reducir los saltos de página y la rotura de enlaces. Además, se podría promover más ampliamente el uso de tecnologías de archivado de contenido dinámico, como herramientas de captura basadas en renderizado (Puppeteer, etc.), para compensar las limitaciones actuales en la captura de contenidos generados con JavaScript. Frente al problema de la falta de equipos técnicos en universidades pequeñas y medianas, se puede considerar la creación de un Centro Nacional de Asistencia para el Archivado Web, que proporcione asesoría técnica, formación y servicios de alojamiento centralizados, a fin de mejorar la capacidad general de archivado del sistema.

En el ámbito de la cooperación internacional, se recomienda que las instituciones de preservación digital y los equipos técnicos de distintos países trabajen conjuntamente para desarrollar estándares de evaluación de archivabilidad compatibles con múltiples lenguas y culturas, y para promover la estandarización de las tecnologías de captura de páginas dinámicas. Al mismo tiempo, se debería establecer un repositorio de muestras de sitios web universitarios, que sirva como banco de pruebas para que las instituciones de todo el mundo realicen autoevaluaciones y mejoras en su capacidad de preservación web.

7.4 Perspectivas de investigación

Debido a limitaciones de tiempo y herramientas, el presente estudio aún presenta ciertas restricciones, como la cobertura limitada de la muestra y las diferencias de adaptación entre herramientas de evaluación. Las investigaciones futuras podrían ampliar el espectro de análisis a más países y tipos de instituciones universitarias, especialmente aquellas de carácter periférico, como los centros de formación profesional o institutos técnicos superiores.

Asimismo, sería valioso incorporar la perspectiva del usuario, evaluando el grado de reconstrucción funcional e inteligibilidad de la información de las páginas archivadas durante su uso real. También se podría desarrollar un análisis comparativo más profundo entre distintas herramientas de archivado web —como Webrecorder, Heritrix o Brozzler— en cuanto a su adaptación, estabilidad y rendimiento técnico en entornos universitarios, explorando cómo sus distintos mecanismos de captura afectan a la autenticidad del contenido preservado.

A través de este estudio, se espera fomentar una mayor cooperación entre China y España en la construcción de políticas y estándares técnicos relacionados con el archivado web universitario, y así contribuir al desarrollo coordinado de la preservación del patrimonio académico digital a nivel global, asegurando que los resultados del conocimiento puedan ser fiablemente registrados, transmitidos y reutilizados en entornos digitales.

Referencias bibliográficas empleadas

1. Administración de Normalización de China. (2022). *GB 18030–2022: Requisitos técnicos para el procesamiento de información en chino: Juego de caracteres codificados*. Administración Estatal de Regulación del Mercado y Comité Nacional de Normalización de China.
2. AlNoamany, Y., Weigle, M. C., & Nelson, M. L. (2013). Access patterns for robots and humans in web archives. En *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (pp. 339–348). <https://doi.org/10.1145/2467696.2467722>
3. Arquivo.pt. (s.f.). *Preserve su sitio: Panfleto informativo*. Fundação para a Ciência e a Tecnologia. <https://sobre.arquivo.pt/wp-content/uploads/Panfleto-ENG.pdf>
4. Asociación Española de Normalización. (2024). *UNE-ISO 2789:2024. Información y documentación – Estadísticas internacionales de bibliotecas* (8ª ed.). UNE. <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma/?c=N0072521>
5. Banos, V., Manolopoulos, Y. A quantitative approach to evaluate Website Archivability using the CLEAR+ method. *Int J Digit Libr* 17, 119–141 (2016). <https://doi.org/10.1007/s00799-015-0144-4>
6. Boletín Oficial del Estado. (2015). *Real Decreto 635/2015, de 10 de julio, por el que se regula el depósito legal de publicaciones en línea [Royal Decree 635/2015 of July 10th, regulating the legal deposit of online publications]*. BOE-A-2015-8008. <https://www.boe.es/eli/es/rd/2015/07/10/635>
7. Comité Directivo del Trabajo de Información y Bibliotecas Universitarias del Ministerio de Educación de China. (2023). *Informe sobre el desarrollo de las bibliotecas universitarias en China en 2022*. Recuperado de <http://www.scal.edu.cn/sites/default/files/attachment/tjpg/2022%E5%B9%B4%E9%AB%98%E6%A0%A1%E5%9B%BE%E4%B9%A6%E9%A6%86%E5%8F%91%E5%B1%95%E6%8A%A5%E5%91%8A.pdf>
8. Comité Permanente de la Asamblea Popular Nacional de China. (2017, 4 de noviembre). *Ley de Bibliotecas Públicas de la República Popular China* (art. 24). Portal de la Asamblea Popular Nacional. <https://www.npc.gov.cn/zfxgk/zfxgkml/fl/202012/P020220223516904968244.pdf>
9. COPTR. (s.f.). *NutchWAX*. Community Owned digital Preservation Tool Registry. Recuperado de <https://coptr.digipres.org/index.php/NutchWAX>

10. Gomes, D., Demidova, E., Winters, J., & Risse, T. (Eds.). (2021). *The Past Web: Exploring Web Archives*. Springer Nature Switzerland. <https://doi.org/10.1007/978-3-030-63291-5>
11. International Internet Preservation Consortium (IIPC). (s.f.). *Training Materials*. Recuperado de <https://netpreserve.org/web-archiving/training-materials/>
12. International Internet Preservation Consortium (IIPC). (s.f.). *About Us*. Recuperado de <https://netpreserve.org/about-us/>
13. Internet Archive. (2022, 2 de noviembre). *Archiving and Data Services Update at the 2022 Archive-It Virtual Partner Meeting* [Video]. Internet Archive. <https://archive.org/details/archiving-and-data-services-update-at-the-2022-virtual-partner-meeting>
14. Internet Archive. (s.f.). *Archive-It*. Recuperado de <https://archive.org/details/archiveit>
15. Jones, S. M., Klein, M., & Van de Sompel, H. (2021). Robustifying links to combat reference rot. *Code4Lib Journal*, 50. <https://journal.code4lib.org/articles/15509>
16. Klein, M., Van de Sompel, H., Sanderson, R., Shankar, H., Balakireva, L., Zhou, K., & Tobin, R. (2014). Scholarly context not found: One in five articles suffers from reference rot. *PLOS ONE*, 9(12), e115253. <https://doi.org/10.1371/journal.pone.0115253>
17. Library of Congress. (s.f.). *Sustainability of Digital Formats: WARC*. Recuperado de <https://www.loc.gov/preservation/digital/formats/fdd/fdd000236.shtml>
18. Liu, Y., & Zhang, T. (2021). Web archiving practices in Chinese academic libraries: An empirical study based on the “Double First-Class” initiative. *Library Hi Tech*, 39(3), 832–848. <https://doi.org/10.1108/LHT-01-2020-0011>
19. Park, J.-R., & Tosaka, Y. (2010). Metadata creation practices in digital repositories: A survey. *Cataloging & Classification Quarterly*, 48(8), 716–739. <https://doi.org/10.1080/01639374.2010.508711>
20. SalahEldeen, H. M., & Nelson, M. L. (2012). Losing My Revolution: How Many Resources Shared on Social Media Have Been Lost? *arXiv*. <https://doi.org/10.48550/arXiv.1209.3026>

21. Santos de Paz, L. (2018). Depósito legal en línea: Desafíos para su desarrollo en España. *Revista General de Información y Documentación*, 28(1), 11–22. <https://doi.org/10.5209/RGID.60814>
22. Siteefy. (2025, marzo). *How many websites are there?* Recuperado el 28 de junio de 2025, de <https://siteefy.com/how-many-websites-are-there/>
23. Taylor, N. (n.d.). *Web Archivability Index*. <https://nullhandle.org/web-archivability/index.html>
24. UNESCO. (2021). *PERSIST Guidelines for the selection of digital heritage for long-term preservation* (2^a ed.). UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000244280>
25. Wikipedia contributors. (s.f.). *Heritrix*. Wikipedia, La enciclopedia libre. <https://es.wikipedia.org/wiki/Heritrix>
26. Wikipedia contributors. (s.f.). *Wayback Machine*. Wikipedia, La enciclopedia libre. https://es.wikipedia.org/wiki/Wayback_Machine
27. W3C. (2018). *Web Content Accessibility Guidelines (WCAG) 2.1*. World Wide Web Consortium (W3C). <https://www.w3.org/TR/WCAG21/>

Referencias en idioma Chino

1. Liu, B. (2024). Retos y estrategias en la preservación de recursos de información en línea: Una perspectiva basada en el desorden de URLs y diferencias temporales. *Lantai Neiwai*, (16), 31–33. 刘冰. (2024). 网络信息资源归档挑战与对策——基于URL无序性和时效性差异的视角. *兰台内外*, (16), 31–33.
2. Liu, P. (2019). Revisión del contenido de la literatura sobre la preservación de información en línea en China. *Lantai Shijie*, (12), 36–39, 42. <https://doi.org/10.16565/j.cnki.1006-7744.2019.12.07> 刘佩. (2019). 我国网络信息归档文献内容研究综述. *兰台世界*, (12), 36–39+42. <https://doi.org/10.16565/j.cnki.1006-7744.2019.12.07>
3. Liu, Y. (2021). Exploración práctica de la recolección y archivo de información importante en sitios web universitarios. *Chengjian Dang'an*, (09), 120–121. 刘赟博. (2021). 高校重要网页信息采集归档实践. *城建档案*, (09), 120–121.

4. Pei, J. (2021). Prácticas de gestión de archivo web del IIPC y sus implicaciones para China. *Daxue Tushuguan Qingbao Xuekan*, 39(04), 87–91. 裴佳越. (2021). IIPC 网页归档管理工作实践及其对我国的启示. *大学图书情报学刊*, 39(04), 87–91.
5. Wu, Q. (2022). Investigación sobre el archivo web del Internet Archive de EE. UU. y sus implicaciones (Tesis de maestría, Universidad Normal de Fujian). CNKI. <https://doi.org/10.27019/d.cnki.gfjsu.2022.000568> 吴倩. (2022). 美国 Internet Archive 网页归档研究及启示 (硕士学位论文, 福建师范大学). 中国知网. <https://doi.org/10.27019/d.cnki.gfjsu.2022.000568>
6. Wu, X., Wang, J., & Chen, D. (2024). Análisis de casos sobre la preservación de sitios web en universidades estadounidenses y sus implicaciones: Los casos de Harvard, Michigan y Columbia. *Lantai Shijie*, (07), 75–78, 85. <https://doi.org/10.16565/j.cnki.1006-7744.2024.07.18> 吴晓茹, 王继武, & 陈丹. (2024). 美国高校网页归档的案例分析及启示——以哈佛大学、密歇根大学、哥伦比亚大学为例. *兰台世界*, (07), 75–78+85. <https://doi.org/10.16565/j.cnki.1006-7744.2024.07.18>
7. Xie, W. (2022). Investigación e implementación del archivo en línea de documentos electrónicos universitarios. *Dang'an Guanli*, (01), 75–76. <https://doi.org/10.15950/j.cnki.1005-9458.2022.01.028> 谢文杰. (2022). 高校电子文件在线归档的研究与实现. *档案管理*, (01), 75–76. <https://doi.org/10.15950/j.cnki.1005-9458.2022.01.028>