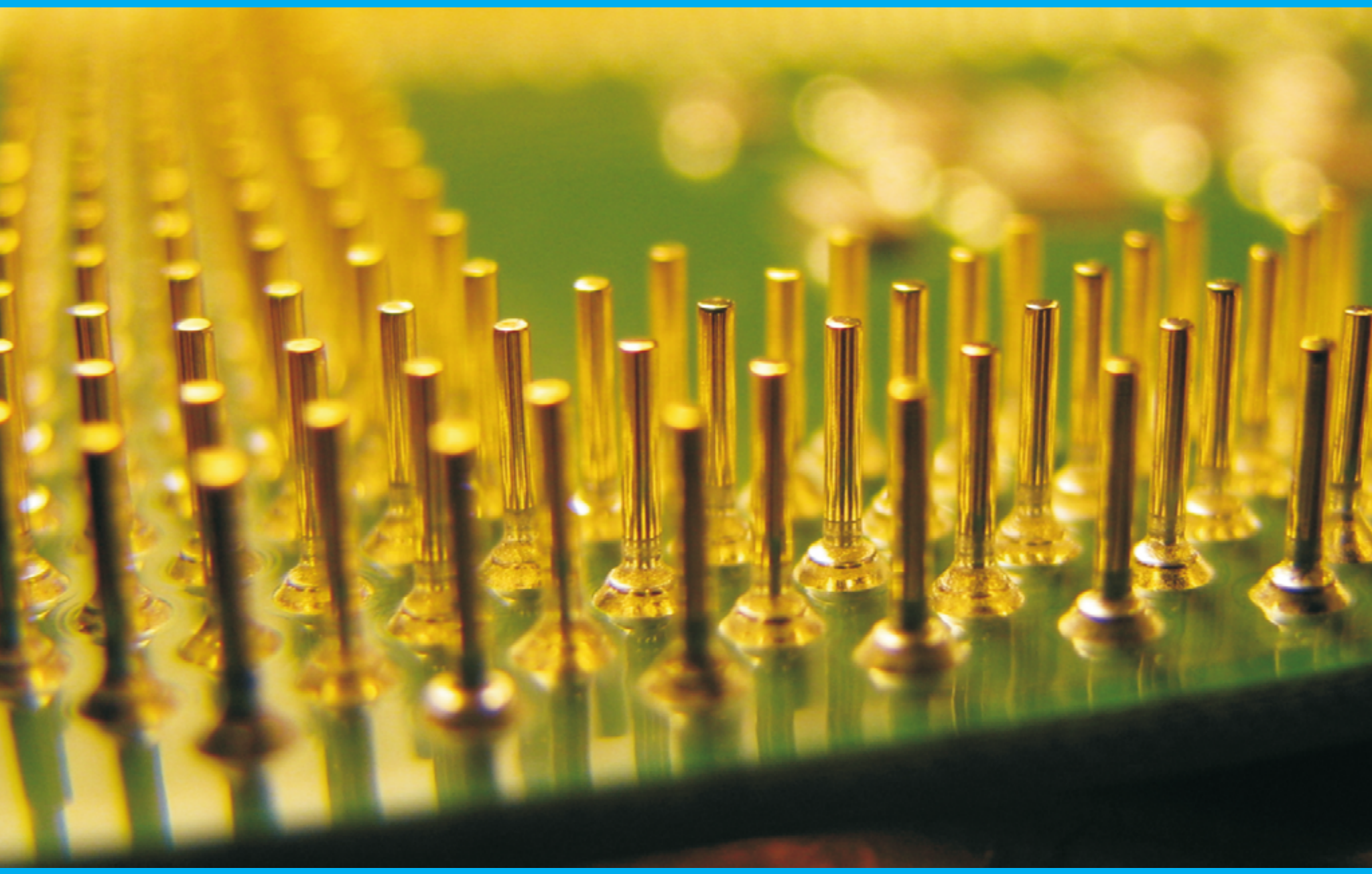
Mini/Manual



ARCHIVO
GENERAL
DE LA NACIÓN
COLOMBIA

Archivamiento Web

conceptos básicos, estrategias y mejores prácticas

ARCHIVO
GENERAL
DE LA NACIÓN
COLOMBIA

Mini/Manual

Archivamiento Web

conceptos básicos, estrategias y mejores prácticas

Subdirección de Tecnologías de la
Información Archivística y Documento Electrónico

Créditos

ARCHIVO GENERAL DE LA NACIÓN
JORGE PALACIOS PRECIADO -COLOMBIA
Establecimiento público
adscrito al Ministerio de Cultura

Consejo Directivo
Ministerio de Cultura
Ministra: Mariana Garcés Córdoba
Viceministra: María Claudia López Sorzano
Presidenta del Consejo

Representante de los Archivos del País
Margarita Monsalve Salas
Alcaldía Distrital de Barranquilla

Academia Colombiana de Historia
Juan Camilo Rodríguez Gómez
Presidente

Colciencias
Juanita León Peñarenas
Delegada de la Sra. Directora

Archivo General de la Nación
Carlos Alberto Zapata Cárdenas
Director General

Comité Editorial
Carlos Alberto Zapata Cárdenas
Claudia Ivonne Fáctor Lugo
Mauricio Tovar González
Jhon Alexander González Flórez
John Francisco Cuervo Alonso
Natacha Eslava Vélez
Dania Paola Asprilla Yurgaqui

Coordinación Editorial y Diagramación
Dania Paola Asprilla Yurgaqui
Sandra Cardona Carvajal
Catalina Lozano Ortega

Fotografía de Carátula
Atribución-NoComercial-SinDerivadas 2.0 Genérica (CC BY-NC-ND 2.0) -jadjadjad https://www.flickr.com/photos/jadjadjad/3116787127

Autor
Jhon Alexander González Flórez

Preparado por:
Iván Eduardo Triana Bohórquez

Gráficas
Ivan Triana Bohorquez

ISBN
978-958-8242-35-4

Archivo General de la Nación de Colombia
Carrera 6 No. 6-91
Teléfono: 328 2888 Fax: 337 2019
E-mail: contacto@archivogeneral.gov.co
Página web: www.archivogeneral.gov.co
Bogotá D.C., Colombia - 2015

Contenido

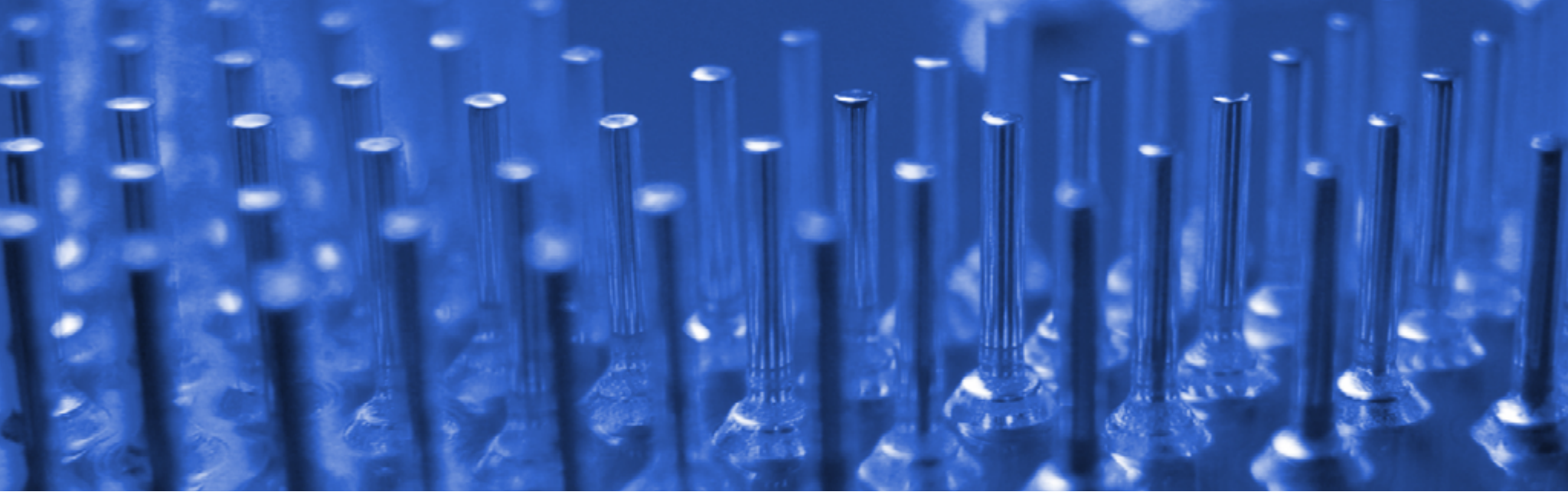
Introducción

Este Minimanual pretende ser un referente conceptual y de buenas prácticas para aquellas entidades públicas y privadas u otros, interesados en estructurar y desarrollar proyectos o iniciativas de archivamiento web, de cara al importante reto que asume la gestión documental en el país, con la penetración y uso de las nuevas tecnologías de la información y comunicación.

Está dirigido a la Administración Pública en sus diferentes niveles: nacional, departamental, distrital y municipal; a las entidades territoriales indígenas y demás entidades territoriales que se creen por Ley; a las divisiones administrativas; a las entidades privadas que cumplen funciones públicas, a las entidades públicas en las distintas ramas del poder; a las instituciones culturales y educativas, empresas del sector privado, autores y personas naturales interesadas en gestionar y preservar el patrimonio web.

El contexto normativo se enmarca en la **Ley 594 de 2000**, "Por medio de la cual se dicta la Ley General de Archivos y se dictan otras disposiciones" - Título XI, Conservación de Documentos, **el Decreto 2609 del 14 de diciembre de 2012**, "Por el cual se reglamenta el Título V de la Ley 594 de 2000, parcialmente los artículos 58 y 59 de la Ley 1437 de 2011 y se dictan otras disposiciones en materia de Gestión Documental para todas las Entidades del Estado" y el **Decreto 2693 21 de diciembre de 2012**, de Gobierno en Línea "Por el cual se establecen los lineamientos generales de la estrategia de Gobierno en Línea de la República de Colombia, se reglamentan parcialmente las Leyes 1341 de 2009 y 1450 de 2011, y se dictan otras disposiciones". Así mismo, en estándares tales como la Norma ISO 28500: *Information and documentation. The WARC File Format.*

Es así como para facilitar el entendimiento del lector, esta publicación se desarrolla en dos partes: la primera, aborda y define el concepto de archivamiento web, sus tipos, clases y principales retos. Igualmente, se hace referencia a las principales herramientas tecnológicas utilizadas y los casos de éxito más representativos a nivel mundial. La segunda, resume en cinco pasos, las mejores prácticas y estrategias para estructurar un proyecto de archivamiento web que permita garantizar la captura, organización, preservación, continuidad y consulta del patrimonio registrado en la web, a las generaciones actuales y futuras.



1.

Importancia
del Archivamiento Web

El vertiginoso uso de la web como canal de comunicación y publicación de información en todos sus niveles, desde el gubernamental hasta el individual, demanda la necesidad de desarrollar estrategias e iniciativas que garanticen la disponibilidad de estos registros como evidencias de la gestión y la historia actual para las presentes y futuras generaciones.

Como respuesta a esta necesidad, el archivamiento web es el "proceso de recolección de fracciones o partes de la *World Wide Web* y la garantía de que la colección se conserva en un archivo o sistema de información para futuros investigadores, historiadores y público en general"[1].

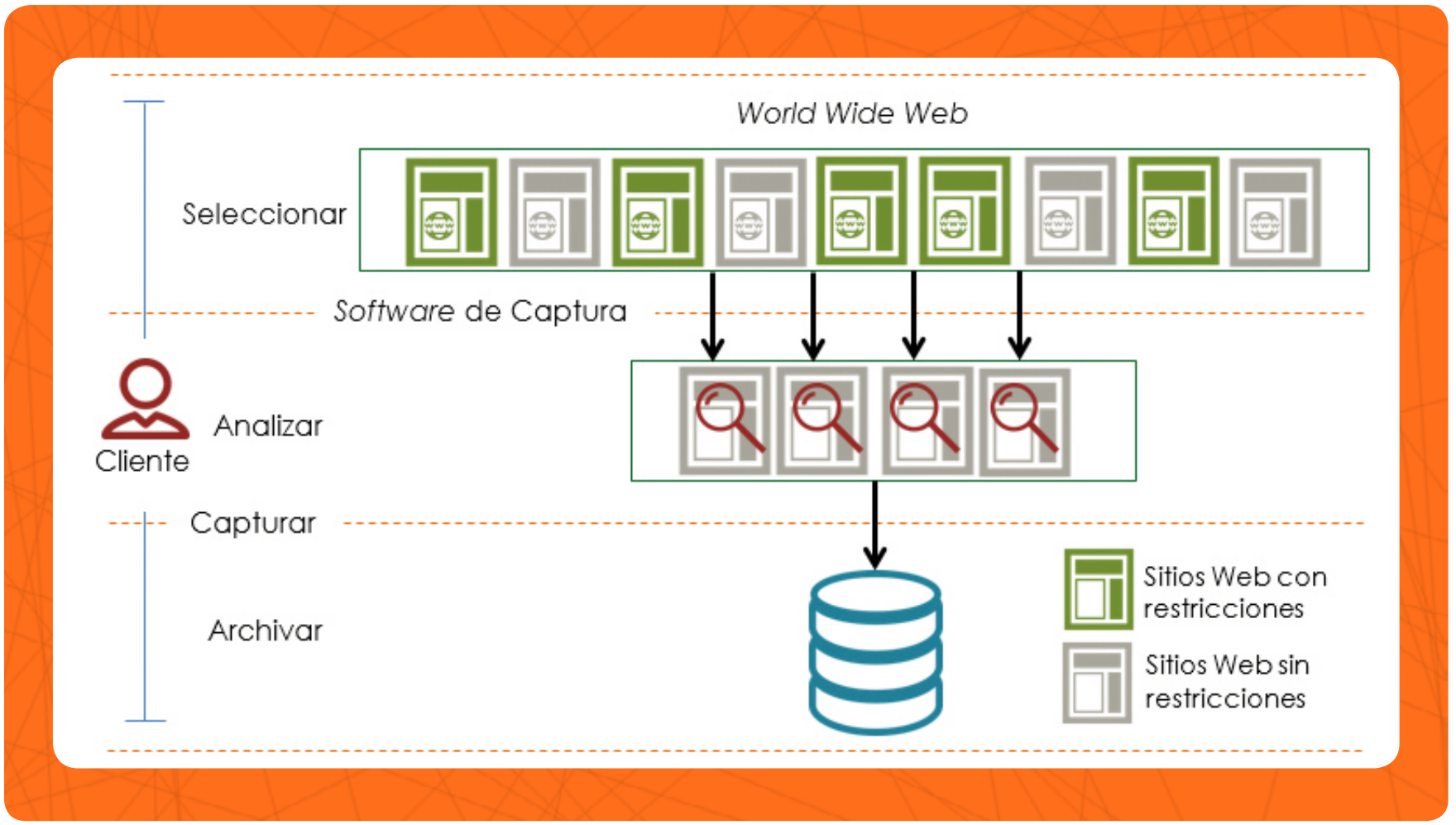
El proceso del archivamiento web es liderado por archivistas y desarrollado con las actividades tradicionales del archivo físico: seleccionar, almacenar, preservar y consultar. Sin embargo, por la cantidad de información contenida en la web, estas actividades son automatizadas con herramientas de *software* especialmente diseñadas para la recolección de los registros objeto de preservación.

1.1 Tipos de Archivamiento Web

Existen tres tipos para archivar contenidos web[2]. Su elección depende de la afinidad y concordancia con los objetivos y reque-

1. COLOMBIA. MINISTERIO DE TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES. Estrategia de Gobierno en Línea [En línea]. <http://programa.gobiernoenlinea.gov.co/apc-aa-files/e5203d1f18ecfc98d25cb0816b455615/minticmanual3.0.pdf> [citado el 2 de octubre de 2013]

2. UNITED KINGDOM. THE NATIONAL ARCHIVES. Web Archiving Guidance [En línea]. <http://www.nationalarchives.gov.uk/documents/information-management/web-archiving-guidance.pdf> [citado el 3 de octubre de 2013]



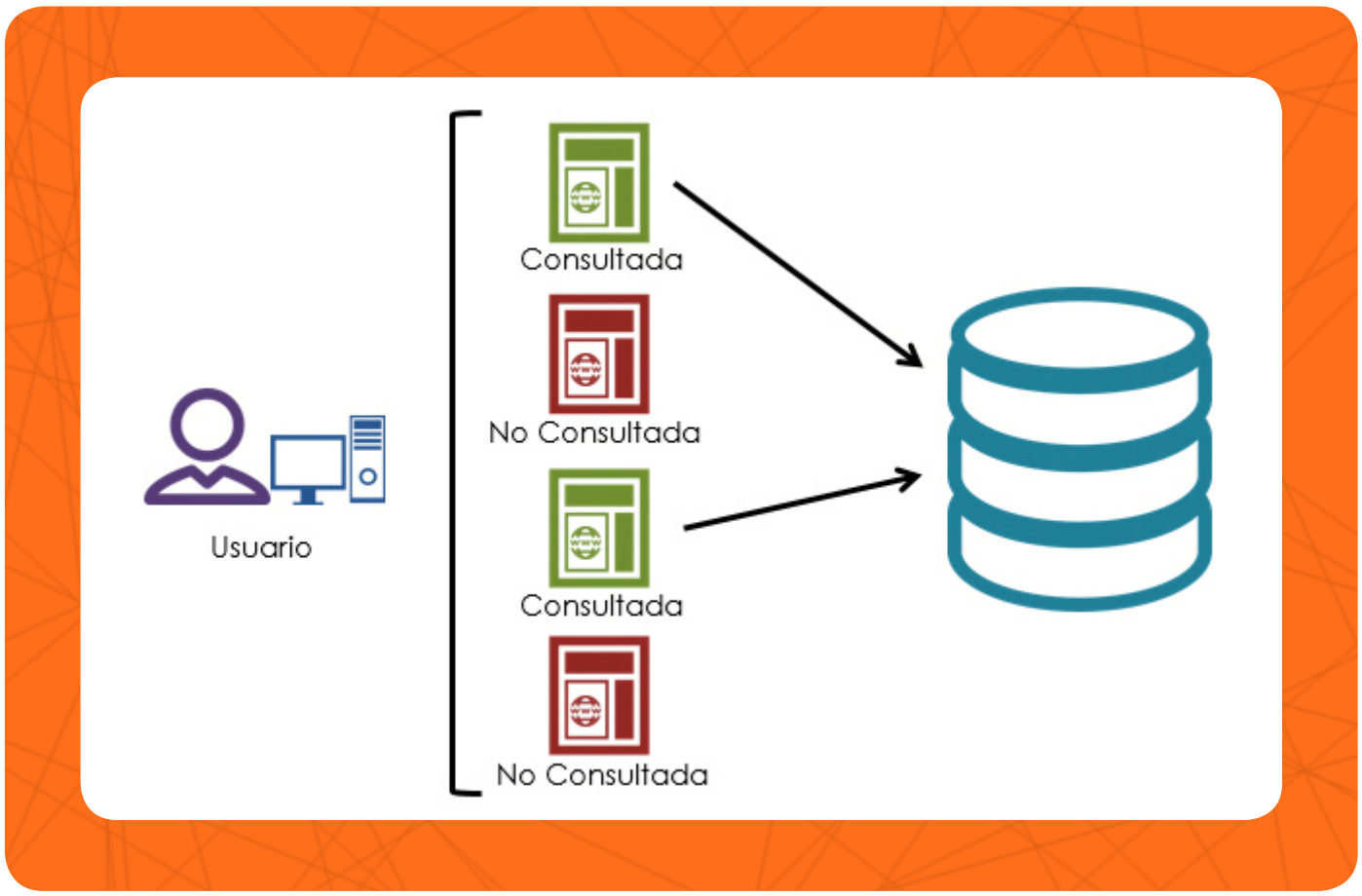
rimientos planteados en el proyecto de archivamiento web.

Archivamiento web de lado del cliente

Es el tipo archivamiento web más popular y empleado por instituciones interesadas en preservar la web, debido a su simplicidad y escalabilidad. Permite capturar cualquier sitio disponible abiertamente en la web, sin restricciones técnicas ni de derechos de autor. El *software* empleado navega por todo el sitio web y extrae los contenidos disponibles en cada enlace. El éxito de la captura de contenidos dependerá del nivel de optimización y accesibilidad del sitio web.

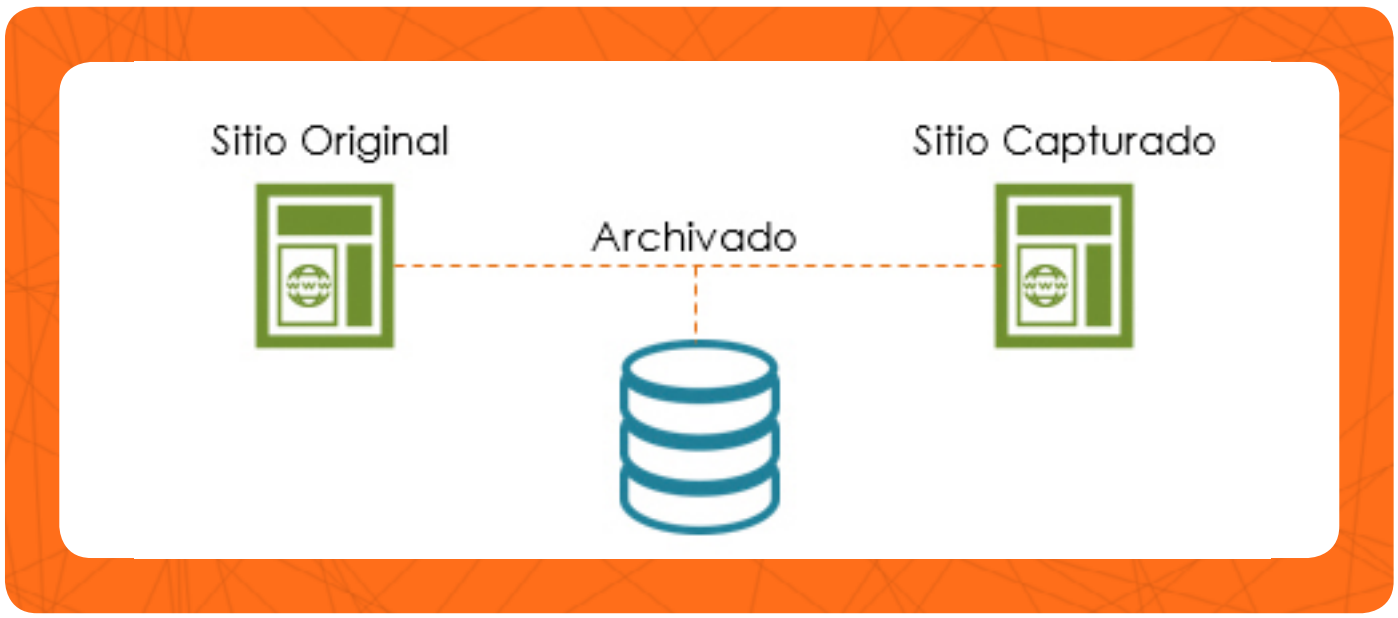
Archivamiento web basado en transacciones

Este modelo es operado desde el servidor que almacena el sitio web. Busca capturar sólo aque-



llos contenidos visualizados por los usuarios y evita los contenidos que nunca fueron visitados. Su principal ventaja es la asertividad en seleccionar para su preservación los contenidos que han sido de interés para la comunidad de usuarios.

Para emplear este tipo de archivamiento web, es necesario el trabajo en equipo con el administrador del servidor, para acceder a los informes de consulta y capturar los registros. Por sus condiciones técnicas, es un enfoque atractivo para proyectos internos de archivamiento web corporativo.



Archivamiento web del lado del servidor

El enfoque desde el lado del servidor, busca crear una copia del sitio web directamente del servidor que lo custodia. Al igual que el modelo anterior, requiere el consentimiento del administrador del mismo. Al crear una copia del sitio web, permite archivarlo conservando sus características de navegabilidad.

Los retos principales de este modelo, se centran en mantener la captura total y constante del sitio, más cuando los contenidos son dinámicos y generados a intervalos de tiempo cortos. Su principal beneficio, está en la capacidad de capturar contenidos inaccesibles por los *software* del archivamiento web del lado del cliente.

1.2 Retos para el Archivamiento Web

Para desarrollar un proyecto de archivamiento web exitoso, que cumpla con la totalidad de los requerimientos de calidad, captura y preservación, es necesario definir estrategias que superen los retos que se presentan en su implementación. Estos retos están clasificados en dos grupos: Técnicos y Administrativos[3].

[3]. BALL, Alex. Web Archiving [en línea]. <http://www.dcc.ac.uk/sites/default/files/documents/reports/sarwa-v1.1.pdf> [citado el 5 de octubre de 2013]

Retos del Archivamiento Web	
Administrativos	Técnicos
	
» Legal. » Selección y Alcance. » Asignación de responsabilidades.	» Coherencia Temporal. » Limitaciones de los rastreadores actuales. » Virus y Malware. » Duplicación. » Preservación a largo plazo.

1.2.1 Retos Administrativos

Son los relacionados con la planeación y dirección de quienes están gestionando el archivamiento web, incluyendo tanto a los líderes del proyecto como a los autores de los contenidos.

» **Legal:** Es el mayor reto no técnico al que se enfrenta un proyecto de archivamiento web, dado que un gran porcentaje de los sitios web y recursos publicados no especifican una licencia de uso de sus contenidos para ir acorde con las restricciones de derechos de autor y no capturar registros sin la autorización requerida.

» **Selección y alcance:** La falta de claridad en los objetivos y en el alcance del archivamiento web, son los principales causantes del fracaso del proyecto. Es indispensable definir con exactitud, los resultados esperados para de esta forma contar con el equipo de trabajo, la infraestructura tecnológica y el tipo de colección que se va a capturar, sea la colección completa de un dominio o un enfoque selectivo de recursos.

» **Asignación de responsabilidades:** Asumir una iniciativa que busque capturar y preservar el patrimonio web, exige que se compartan responsabilidades, procesos y recursos, de lo contrario, todo proyecto se asumirá como un esfuerzo aislado y de poca relevancia. El reto a superar es conformar un equipo de trabajo con responsabilidades definidas y capacidades claras.

1.2.2 Retos Técnicos

Los retos técnicos del archivamiento web están relacionados con los aspectos tecnológicos como el dinamismo de los contenidos, las limitaciones de los *software* de captura, los virus, la obsolescencia y la duplicidad de recursos.

» **Coherencia temporal:** Se refiere a la actualización constante de las páginas web. Un reto que es completo de abordar cuando el número de páginas a archivar incrementa por la falta de consistencia entre el recurso archivado y el sitio web disponible en línea. Cabe aclarar que este reto no se aborda cuando se archivan sitios web que ya no están en línea.

» **Limitaciones de los rastreadores actuales:** Para la automatización de las actividades del archivamiento web se utilizan *software* especializados. Para la selección y captura se utilizan *software* llamados rastreadores o *crawlers*. Por la complejidad de los contenidos disponibles en la web, y a pesar de los desarrollos y mejoras, aún existen limitaciones que evitan su selección y captura adecuada.

Los contenidos que evidencian las principales limitaciones de los rastreadores hacen parte de la web profunda. Por ejemplo:

* Contenidos dinámicos que se generan desde la base de datos del sitio en respuesta a la petición de un usuario.
* Archivos multimedia transmitidos por streaming.
* Contenidos protegidos con contraseña.
* Contenidos que sólo son accesibles con una búsqueda local dentro del sitio web.

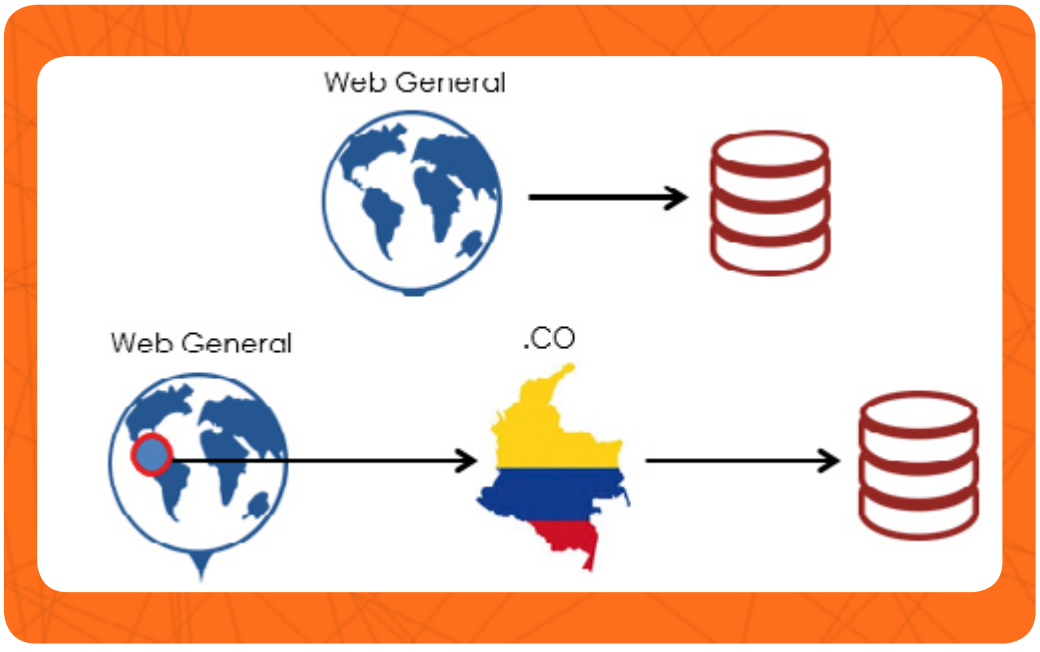
» **Virus y Malware:** Con el objetivo de mantener una captura integral de los contenidos web, el archivamiento web, de acuerdo con sus objetivos, políticas y alcances, debe convivir con los virus y el *malwere* en la captura de los sitios web, dado que pueden ser objeto de investigaciones para futuros usuarios. Es importante definir las herramientas y procedimientos necesarios para evitar alterar los contenidos a procesar y poner en riesgo la seguridad del repositorio de archivo.

» **Duplicación:** En los procesos de captura de recursos web, existen altas probabilidades de duplicar contenidos, que aunque sean extraídos de diferentes sitios, es el mismo. Esto entorpece la eficiencia del proyecto tanto en el acceso a la información como en el rendimiento del servidor destinado para el archivamiento web, siendo importante definir una estrategia que evite o elimine, con cierta frecuencia, los contenidos duplicados.

» **Preservación a largo plazo:** La gran cantidad de formatos publicados en la web y los enlaces entre los diferentes recursos representan un reto importante para el archivamiento web. No sólo para garantizar una buena captura, sino por mantener el acceso futuro a los contenidos. La obsolescencia de formatos y los riesgos de romper el enlace entre los recursos, son variables que deben contemplarse al inicio del proyecto.

1.3 Archivamiento Web a Gran y Pequeña Escala

Para desarrollar un proyecto de archivamiento web existen dos grandes clases: gran y pequeña escala[4]. Dependiendo de los objetivos trazados, se hará la captura selectiva de recursos individuales o el archivamiento de dominios completos o de la web en general.



1.3.1 Gran Escala

Esta clase busca la captura de un gran volumen de recursos, como el archivamiento de la web en general o de un dominio completo, por ejemplo archivar todos los sitios web .co.

Para garantizar la calidad de las capturas, se hace la integración de dos condiciones: la elección del dominio y la definición de criterios de captura, que una vez definidas, se parametrizan en

4. BALL, Alex. Op. Cit.

las herramientas de *software* seleccionadas para esta labor.

Los criterios pueden incluir: la frecuencia con la cual se harán las capturas, el lenguaje de los contenidos, la relevancia de los recursos a capturar, los permisos de captura y reuso de contenidos, la captura de eventos y noticias importantes, entre otros, que se definan dentro del alcance. Es importante tener en cuenta, que elegir esta clase de archivamiento, exige analizar a profundidad estrategias que superen los retos mencionados en la sección anterior (técnicos y administrativos), dado que su complejidad da protagonismo a la mayoría.

1.3.2 Pequeña Escala

A diferencia de la clase anterior, la pequeña escala se enfoca en capturar recursos específicos de acuerdo con las necesidades o intereses de una comunidad reduci-da de usuarios (investigadores, académicos, usuarios individuales o autores). Sus principales ventajas se visualizan en procesos simples de captura, inversión reducida y enfoque en los contenidos puntuales de interés para los usuarios.

Para el proceso de archivamiento existen tres formas, cada una con su uso específico:

» **Archivado basado en la nube:** Consiste en que el propietario de la web, envía capturas de sus páginas a un tercero para su preservación.

» **Repositorio de citas: Captura todos los recursos citados en** publicaciones académicas digitales. Toma como punto de partida la bibliografía del documento e inicia con el proceso de archivamiento con el fin de mantener disponibles las fuentes utilizadas por los autores.

» **Archivo local:** El usuario tiene la posibilidad de realizar capturas directamente desde su equipo a los recursos web que considera importantes.

1.4 Herramientas de *Software*

Para llevar a cabo un proyecto de archivamiento web es indispensable analizar y elegir las herramientas de *software* más adecuadas para cumplir con los requerimientos y alcance deseado. En la siguiente tabla se mencionan las principales herramientas para la automatización de la selección, captura y visualización de recursos:

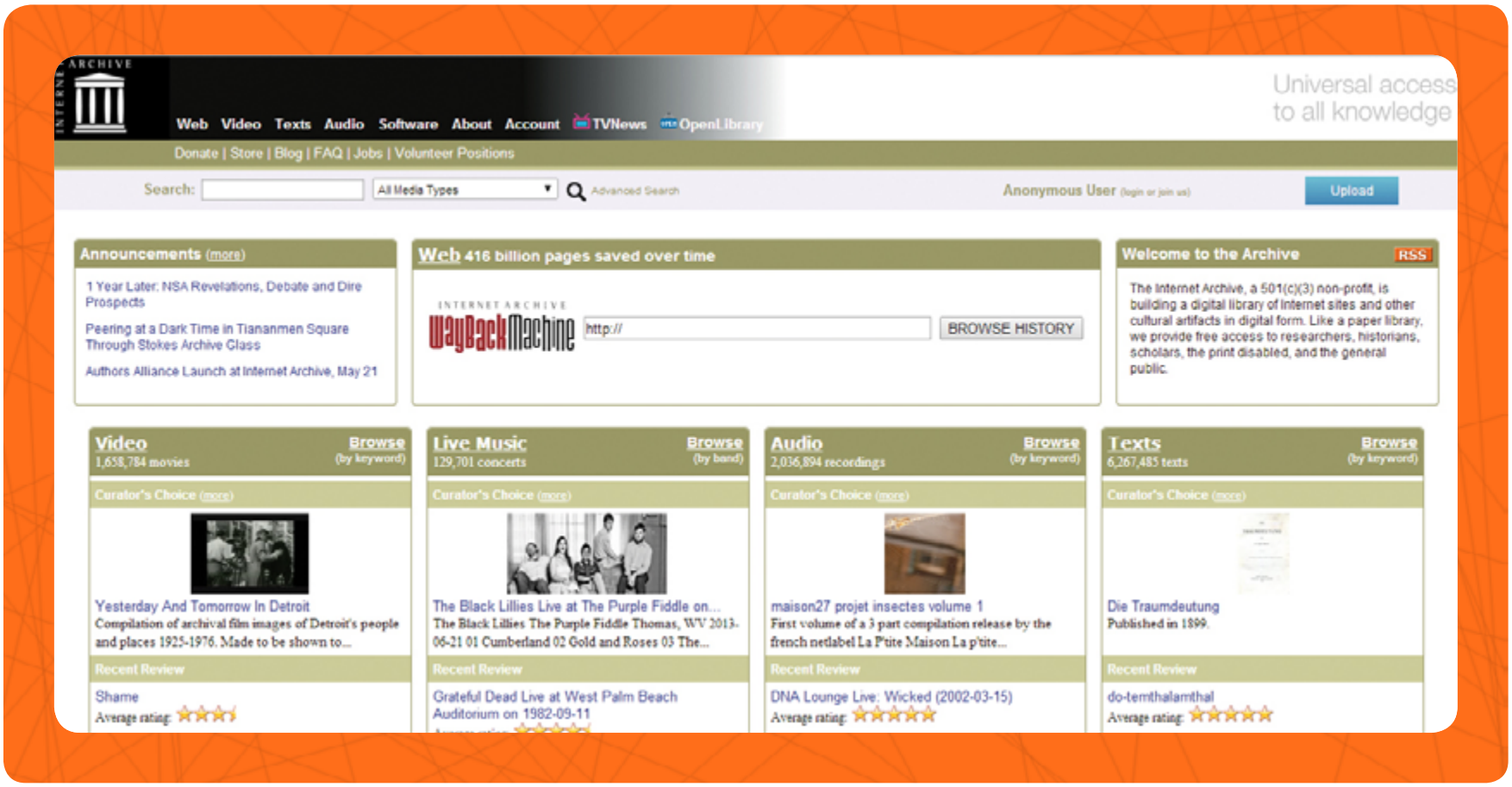
Software	Descripción
Heritrix http://webarchive.jira.com/wiki/display/Hiritrix/Hiritrix	Es un *software* rastreador desarrollado por la iniciativa Internet Archive en código abierto con licencia Apache 2.0. Esta aplicación sirve para identificar y capturar en la web los recursos seleccionados para su proceso de archivamiento. Respeta las restricciones de las etiquetas o ficheros robot.txt de cadapágina web a captura. Los resultados de rastreo los almacena en un fichero ARC.
HTTrack www.httrack.com	Es una aplicación de software libre que permite la descarga total o parcial de un sitio web a un equipo local, permitiendo su navegación sin conexión a Internet. Es ideal para el archivaminto local de pequeña escala.
Netarchivesuite http://sbforge.org/display/NASDOC42/NetarchiveSuite+Overview	Es una aplicación de código abierto desarrollada en el año 2007 y utilizada por el Archivo Digital de Dinamarca. Este *software* puede capturar la web de tres maneras: 1. Captura eventos específicos importantes como día de elecciones, movimientos sociales, catástrofes, entre otros; 2. Carptura selectiva de dominios específicos; 3. Captura a gran escala.

PANDAS pandora.nla.gov.au/pandas.htlm	(PANDORA *Digital Archiving System*) Es un *software* desarrollado por la Biblioteca Nacional de Australia. Facilita la automatización de los flojos de trabajo del archivamiento web como: la identificación, elección de los posibles recursos a archivar; la búsqueda y captura de los recursos con permisos concedidos; la gestión de metadatos; la configuración de restricciones de acceso; la programación automatizada de captura de recursos; y la visualización de contenidos.
Web Curator Tool (WCT) webcurator.sourceforge.net	Fue desarrollado en el año 2006 entre la Biblioteca Nacional de Nueva Zelanda y la Biblioteca Británica. Es una aplicación de código abierto disponible bajo licencia Apache. Esta aplicación facilita la gestión de flujos de trabajo para archivar selectivamente recuersos web. Automatiza la revisión de permisos concedidos en los recursos, la programación de rastreo, la captura de contenido y los metadatos descriptivos.
NutchWAX http://archiveaccess.sourceforge.net/projects/nutchwax/	Es una herramienta de indexación y búsqueda de colecciones web para archivo en formato ARC. Es patrocinado y utilizado por *Internet Archive*, *International Internet Preservation Consortium - IIPC* y el *Nordic Web Archive - NWA.*
WayBack Machine archive.org/web/web.php	Es una aplicación para la navegación de recursos archivados. Genera una base de datos con cada recurso capturado para facilitar su localización y visualización al usuario final, quien puede elegir le fecha de captura dek recurso que quiere consultar. Es un *software* de código abierto utilizado por el *Internet Archive*.
Memento www.mementoweb.org	Es una herramienta de navegación de las colecciones web archivadas. Permite al usuario final visualizar versiones anteriores de un sitio o página web a través de un menú de navegación por fechas de captura.

1.5 Casos de Éxito

Para ilustrar los resultados y los diferentes enfoques del archivamiento web, se describen los siguientes casos de éxito para que sirvan como referentes en la estructuración y diseño de futuras iniciativas:

Internet archive

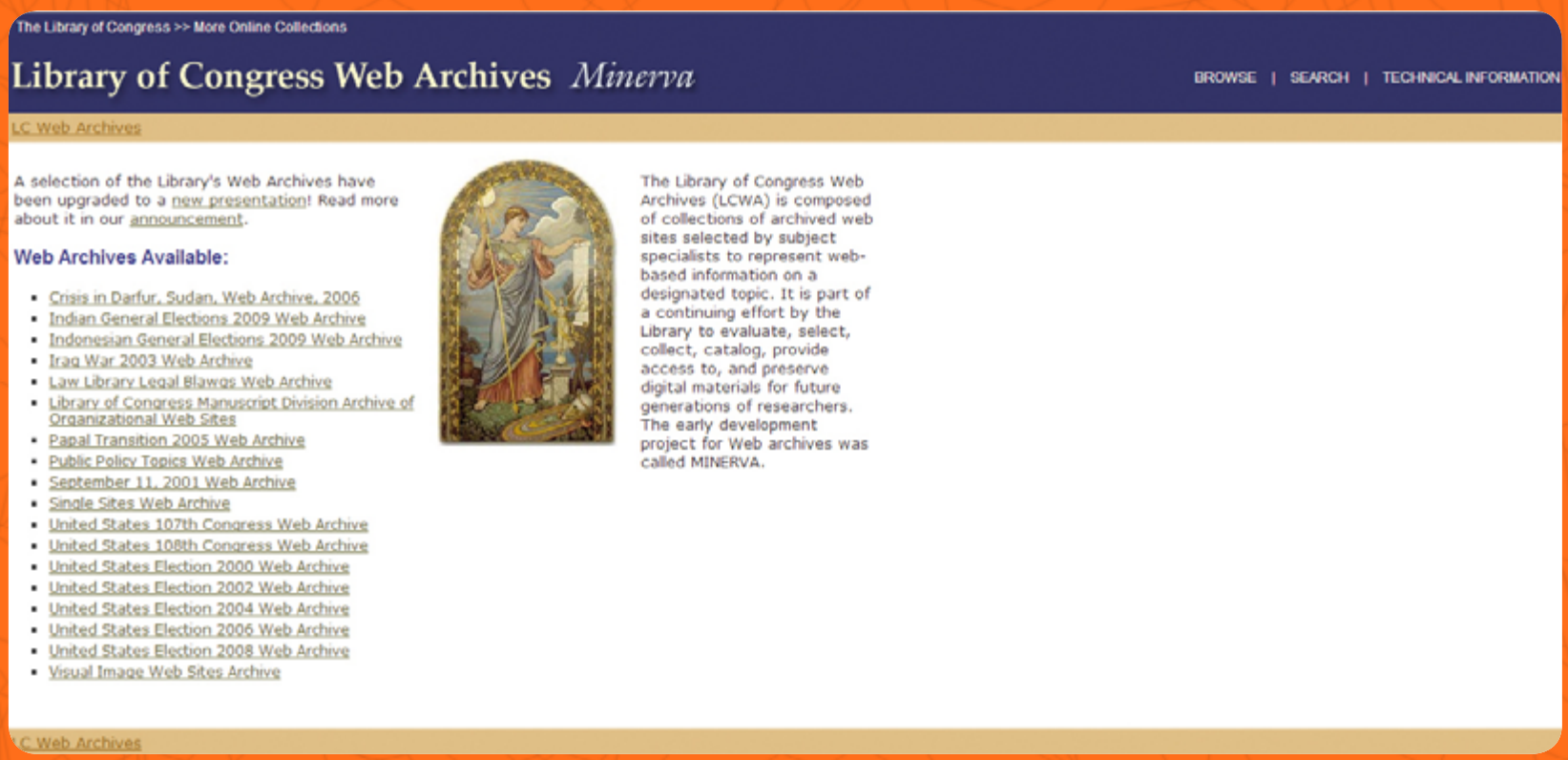


http://archive.org

Es una de las primeras iniciativas de archivamiento web a gran escala fundada en 1996, con el objetivo de construir una biblioteca de Internet que facilitara el acceso a investigadores, historiadores, académicos y al público en general, a sus colecciones web.

Esta iniciativa cuenta en este momento con una colección universal de más de 240 millones de páginas, que están disponibles en su portal para cualquier persona interesada. Dispone a su vez, de una interfaz muy intuitiva que permite hacer los filtros y búsquedas de manera fácil y rápida; con lo que el usuario tiene la posibilidad, por medio de un calendario que resalta las fechas de captura de cada sitio, de visualizar la evolución a través de la historia de su página web de interés.

Library of Congress Web Archives -LCWA

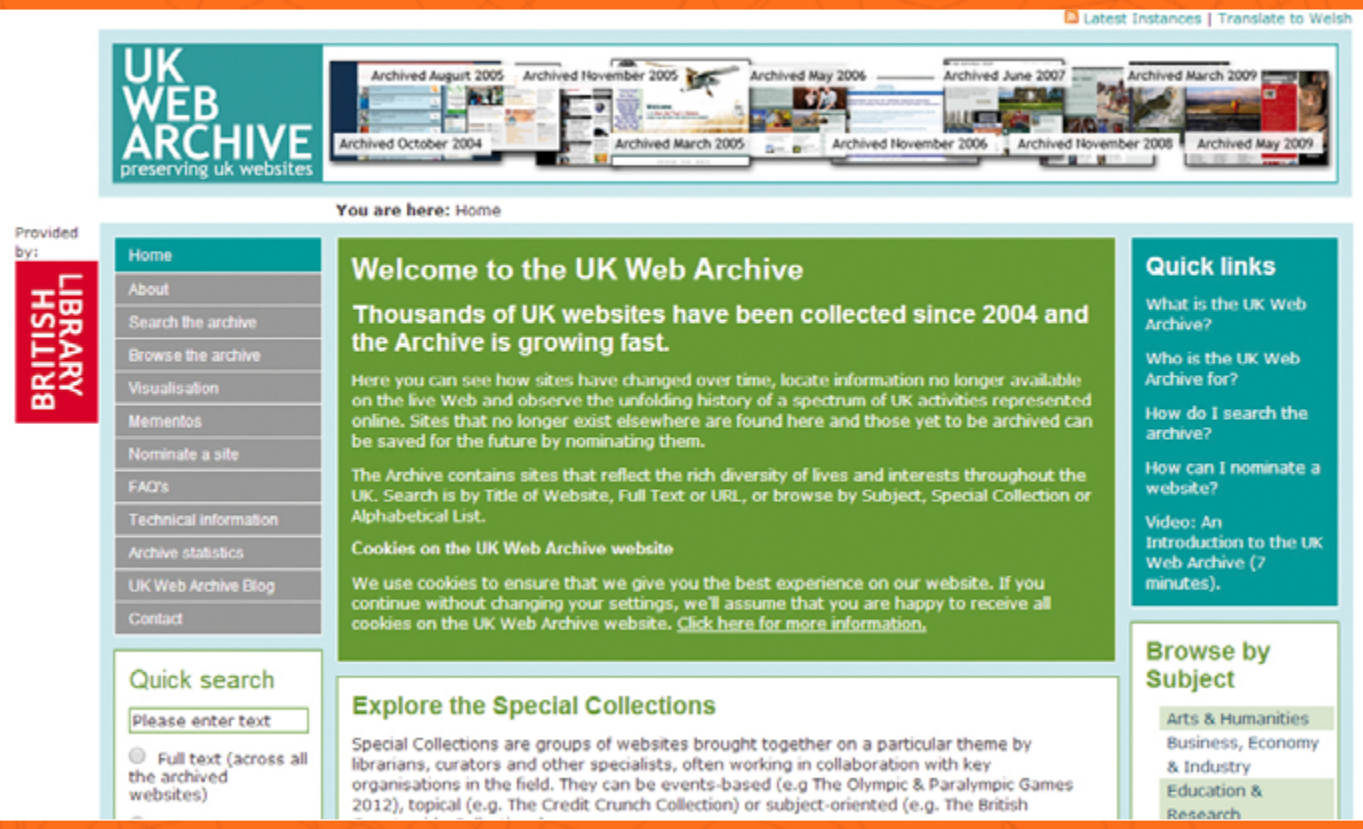


http://lcweb2.loc.gov/diglib/lcwa/html/lcwa-home.html

Inició en el año 2000 como un proyecto piloto para capturar y preservar los sitios web de Estados Unidos. Con este propósito conformó un equipo interdisciplinario para evaluar, seleccionar, recopilar, catalogar, preservar y proporcionar acceso a los recursos capturados.

La biblioteca ha conformado un archivamiento temático basado en eventos importantes de la nación estadounidense como las elecciones, la guerra en Irak y los sucesos del 11 de septiembre.

Archivo Web del Reino Unido

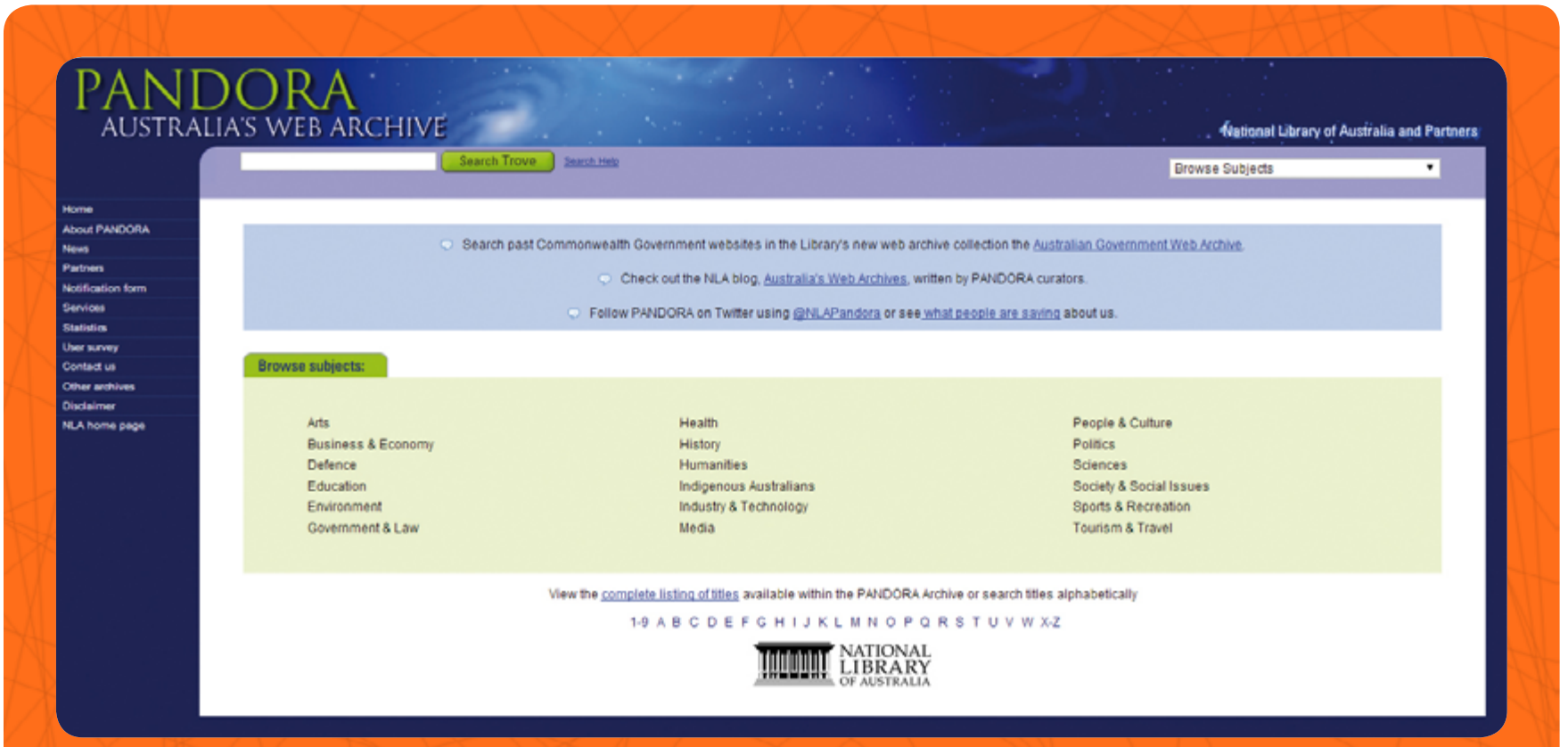


http://www.webarchive.org.uk

Este proyecto inició en el año 2004 por el Archivo Nacional del Reino Unido como estrategia para capturar y preservar la memoria web de la nación. Sus colecciones están compuestas por páginas web que reflejan la diversidad, intereses y actividades del Reino Unido. Igualmente archiva los sitios que registran los acontecimientos políticos, culturales, sociales y económicos de la nación.

Los usuarios tienen acceso público a sus colecciones desde su portal, el cual cuenta con varias operaciones de filtro, búsqueda y navegación.

Pandora



http://pandora.nla.gov.au/

Preserving and Accessing Networked Documentary Resources of Australia, es un proyecto liderado por la Biblioteca Nacional de Australia desde el año 1996. Su objetivo se centra en la captura selectiva de publicaciones y sitios relacionados con dicho país y los australianos. Sus colecciones contienen registros de la vida política, social, cultural e intelectual de la nación.

En este momento, debido a la complejidad del archivamiento, la Biblioteca ha buscado realizar un trabajo colectivo con las bibliotecas públicas y otras entidades relacionadas con la gestión del patrimonio cultural con el objetivo de compartir responsabilidades y recursos.

Netarkivet

http://netarkivet.dk/

Es una iniciativa que busca archivar todos los recursos web relacionados con los daneses, bajo el cumplimiento de la Ley Nacional de Depósito Legal.

Para la captura de los sitios web, combina tres estrategias:

1. Captura de todos los dominios daneses cuatro veces al año.

2. Captura selectiva diaria de recursos relacionados con los daneses.

3. Captura de eventos representativos del país cada dos o tres veces por año.

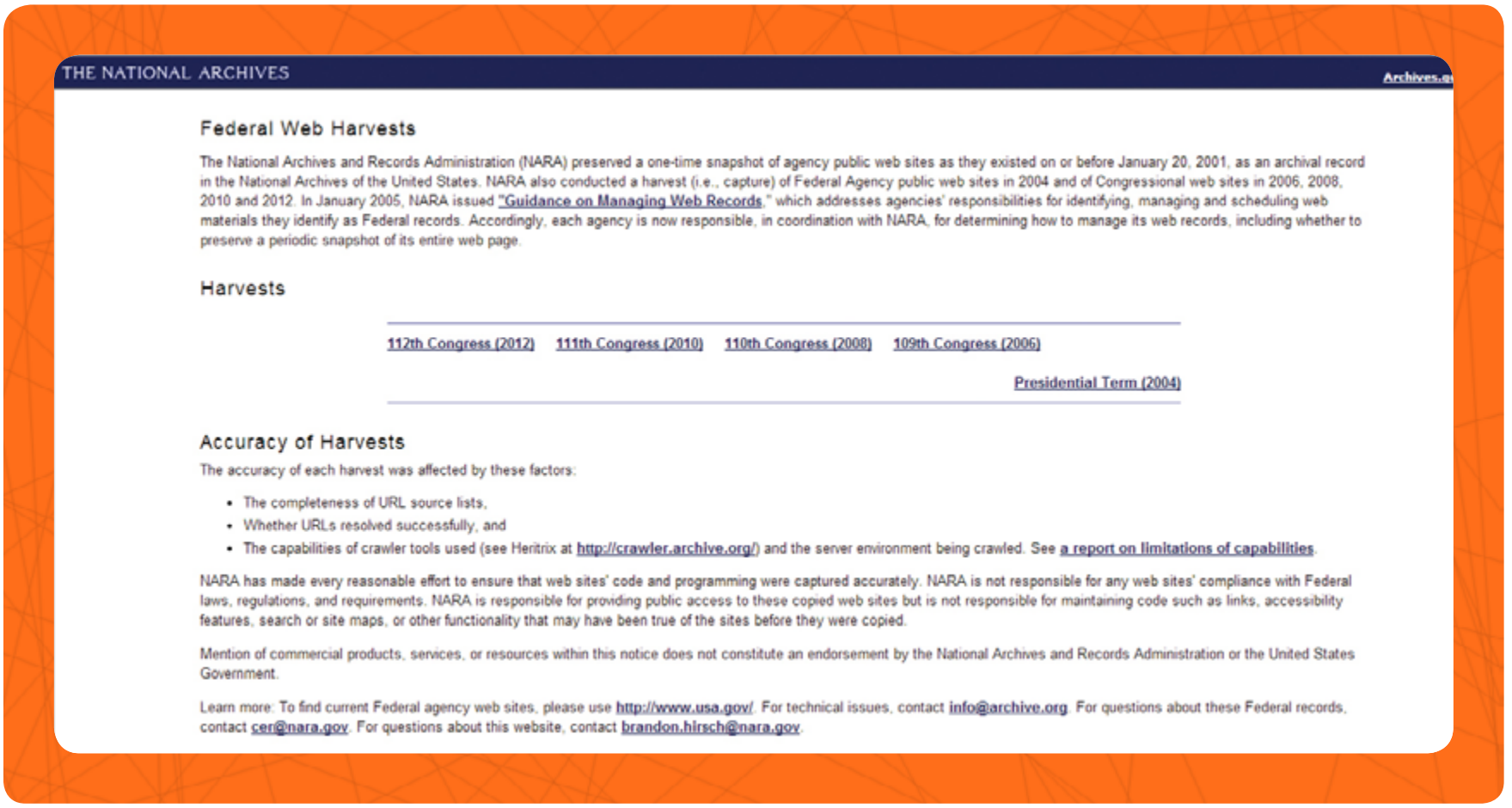
Padicat

http://www.padicat.cat/

Iniciativa liderada por la Biblioteca de Cataluña desde el año 2005, que busca la captura y preservación de los sitios web de Cataluña. Trabaja conjuntamente con el Centro de Servicios Científicos y Académicos de Cataluña, quien apoya los aspectos tecnológicos y técnicos.

A través de su portal, el usuario cuenta con varios filtros de búsqueda que facilitan la consulta y navegación de los recursos.

NARA



http://webharvest.gov/

El Archivo Nacional de Estados Unidos lidera el archivamiento web de todos los sitios de las entidades públicas del país.

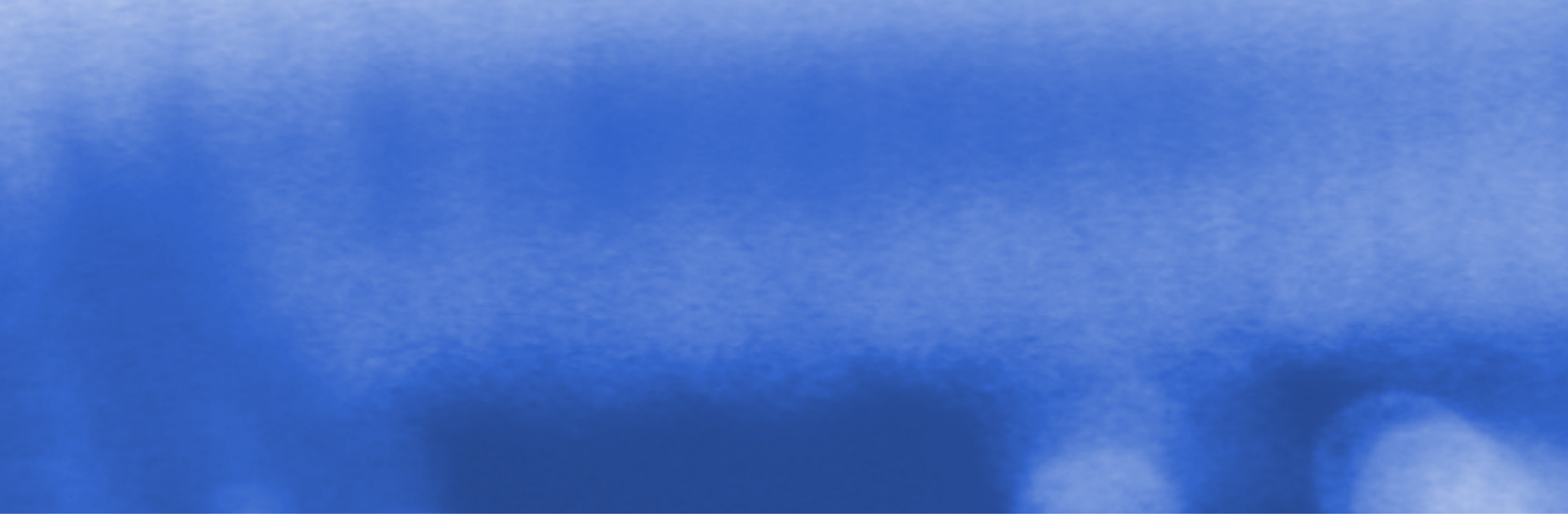
Su estrategia de archivamiento se basó en la definición de directrices para la optimización de sitios web, las cuales tuvieron que ser adoptadas por todas las entidades del estado. Esta estrategia ha facilitado la precisión y calidad en la captura de los recursos, que están disponibles para la ciudadanía de forma pública en su portal web.

Archivo Web de Coca Cola

http://www.coca-colacompany.com/stories/1s-and-0s-the-history-of-the-coca-cola-companys-website

Es un proyecto privado, cuyo objetivo es capturar y preservar los sitios web de las empresas locales de Coca Cola. Inició en el año 2009, utilizando un servicio comercial de archivado en el que se ha capturado y recuperado el patrimonio web de la empresa.

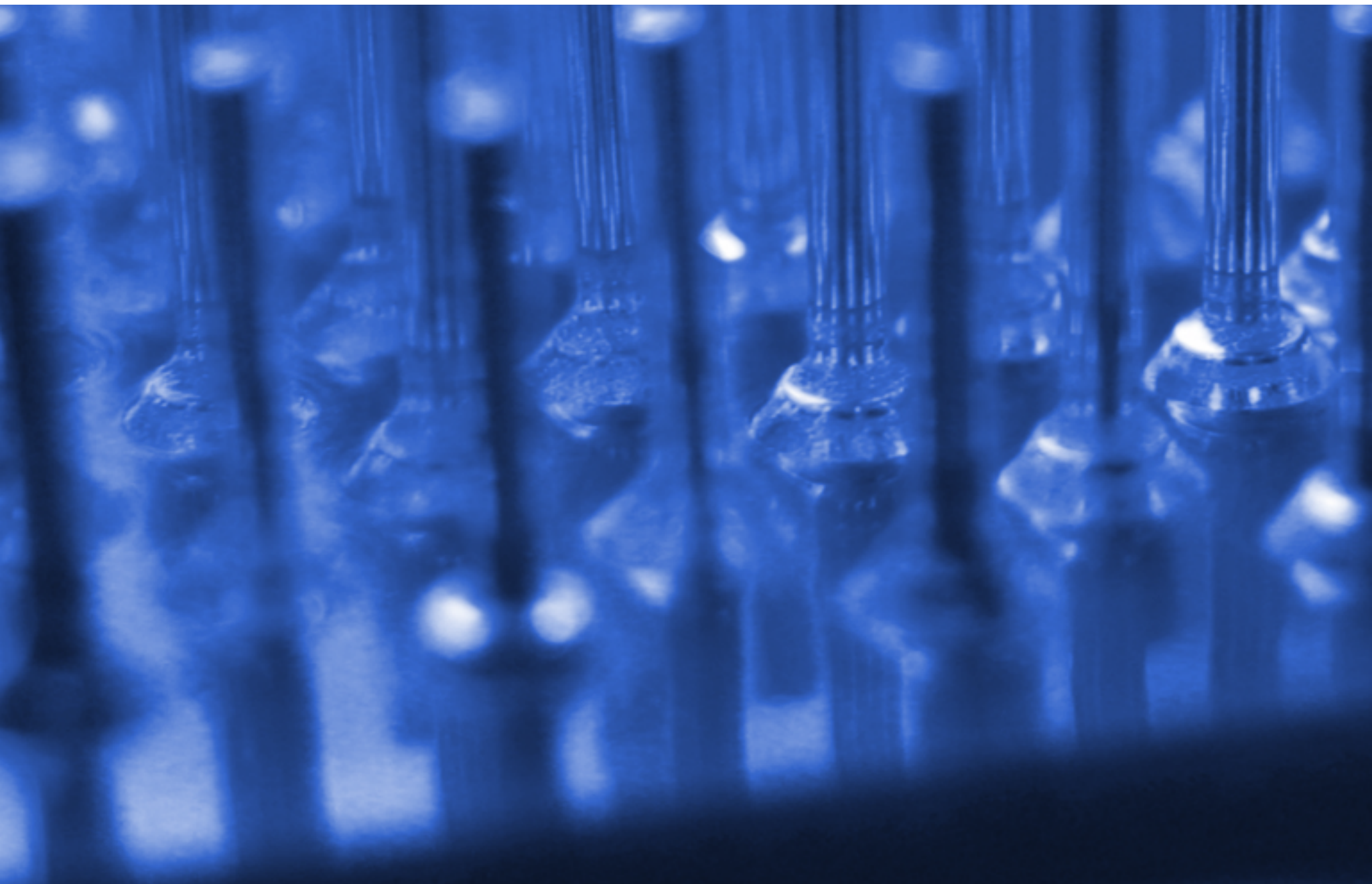
Adicionalmente, el proyecto ha facilitado el acceso a sus registros históricos y la captura de sus comunicaciones web han servido como evidencia ante instancias judiciales. El acceso es limitado y únicamente está disponible para los empleados de Coca Cola, a través de la herramienta de navegación de su proveedor. Su colección cuenta con más de seis millones de páginas web corporativas.



2.

Estrategias

y mejores prácticas: 5 pasos para el Archivamiento Web



Los 5 pasos para estructurar un proyecto de archivamiento web son formulados como punto de partida para facilitar la selección, captura, preservación y acceso de los recursos web conforme con los objetivos planteados por la organización interesada en proteger y mantener el patrimonio web.

Estos 5 pasos son planteados con la recopilación de buenas prácticas del Modelo del Ciclo de Vida del Archivamiento Web propuesto por el equipo de trabajo de *Archive-it* y la Guía de Archivamiento de Recursos Web del Archivo Nacional de Australia.

Paso 1: Definir objetivos

Toda organización interesada en emprender un proyecto de archivamiento web, debe evaluar y analizar sus funciones, plan estratégico, misión y visión, que le permita delimitar el alcance y la precisión de los objetivos del proyecto.

La definición adecuada de los objetivos, garantizará el éxito y sostenimiento del archivado, dado que se enmarca dentro del propósito de la organización, selecciona específicamente qué sitios web va a capturar, dimensiona la complejidad del proceso de archivado, identifica si es a gran o pequeña escala y elige el tipo de archivado y las estrategias adecuadas para superar los retos y riesgos asociados al proyecto.

Paso 2: Identificar aliados

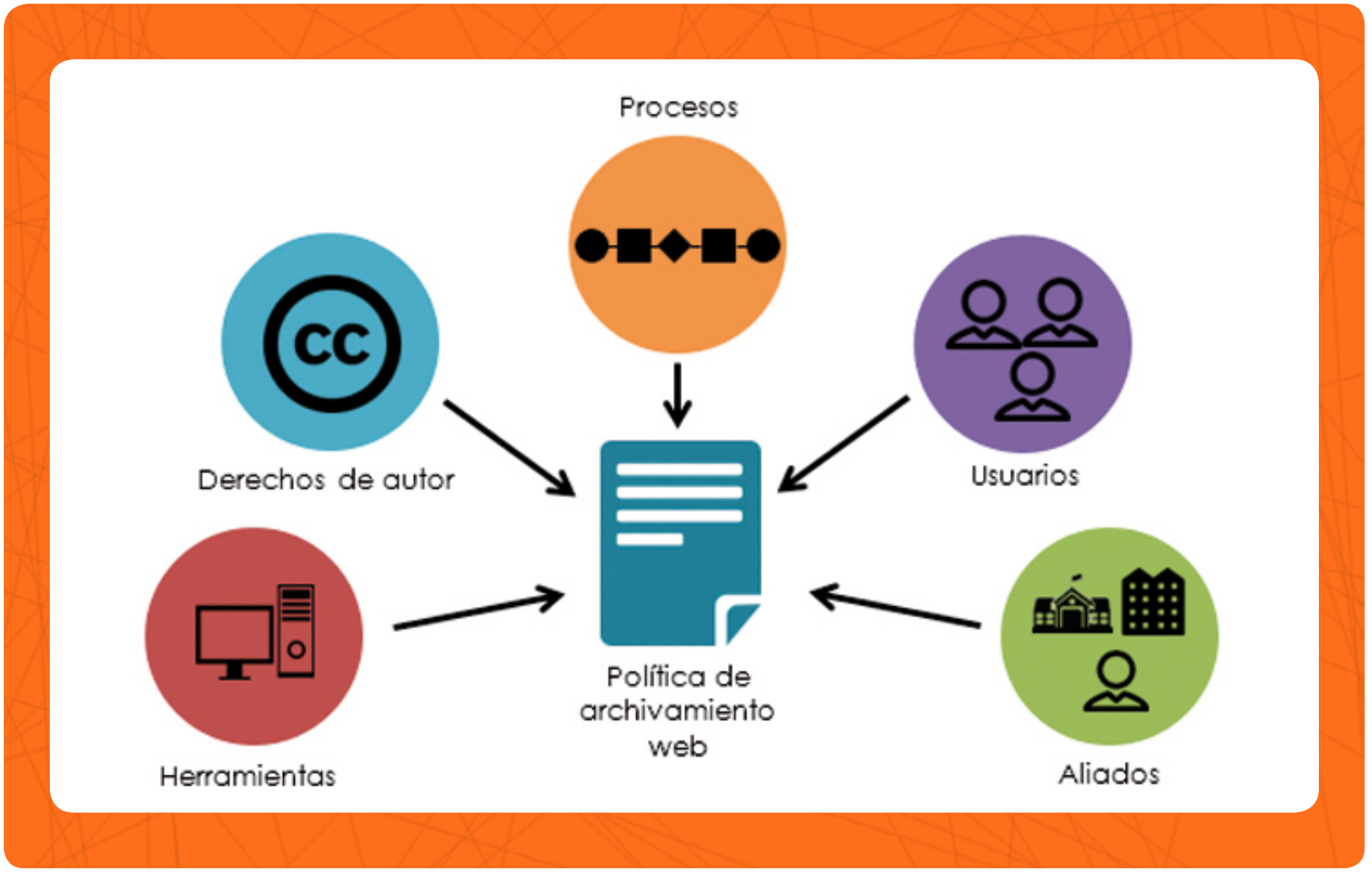
Abordar un proyecto de archivamiento web puede ser desgastante y muy costoso, dependiendo del alcance de los objetivos planteados. Para su-

perar este reto administrativo, la organización debe identificar aliados que se articulen con la iniciativa y estén interesados en integrarse al proyecto.

Es importante analizar las fortalezas y recursos disponibles de cada aliado para garantizar la definición y estandarización de los procesos y los flujos de trabajo del archivamiento web, la asignación de responsabilidades y los niveles de participación de las partes. Este paso requiere un nivel elevado de liderazgo por parte de la organización para unir esfuerzos dirigidos hacia un propósito en común.

Paso 3: Crear una política

La política de archivamiento web orientará y facilitará la toma de decisiones en la ejecución del proyecto, la elección de las herramientas de *software*, la definición y estandarización de procesos y flujos de trabajo, la asignación de responsabilidades y la adminis-



tración, uso, reuso y acceso de sus colecciones a la comunidad de usuarios interesados.

Esta política debe crearse en conjunto con los aliados y en coherencia con los objetivos planteados en el proyecto.

Paso 4: Elegir estrategias de preservación

De acuerdo con la complejidad de las colecciones web, se deben elegir estrategias de preservación adecuadas al proyecto, que garanticen la disponibilidad y acceso a los recursos a largo plazo. La utilización de mejores prácticas y estándares internacionales es fundamental para afrontar los principales retos del archivamiento web.

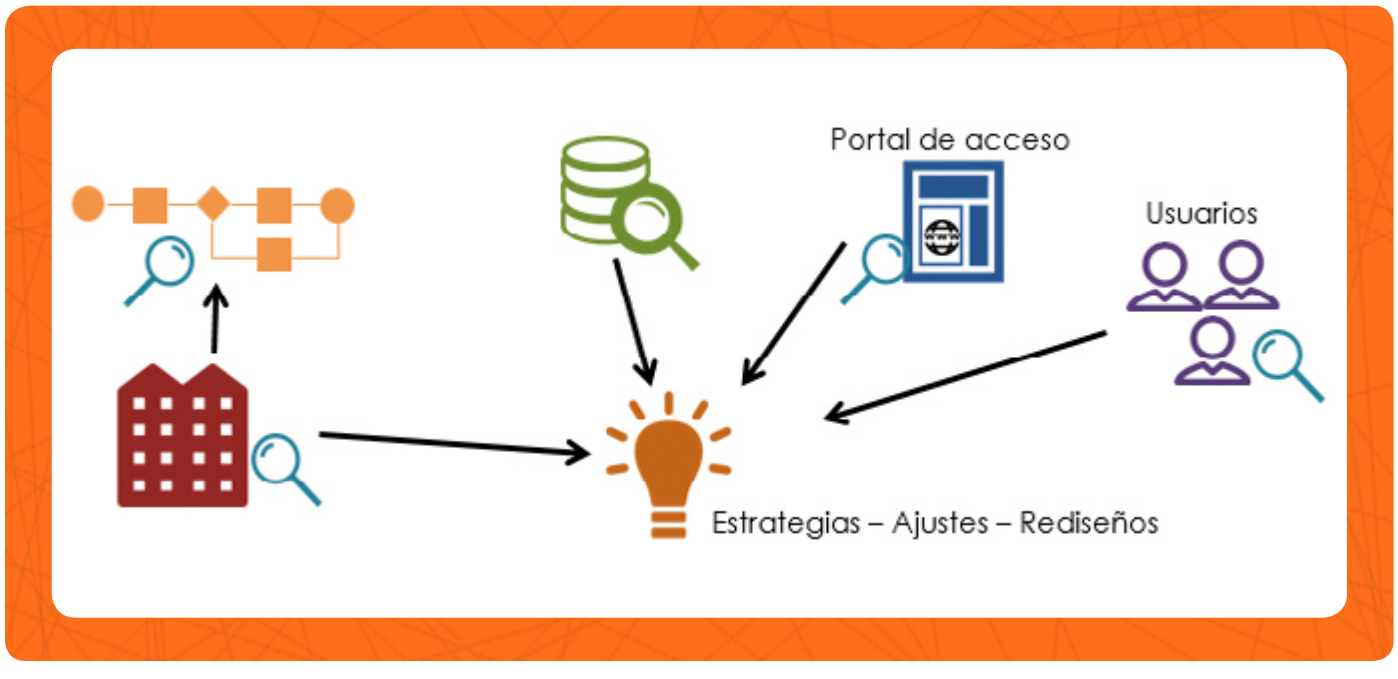
Sin embargo, la preservación digital es un tema en constan-

te evolución, lo cual exige una actualización y formación constante por parte de los líderes del proyecto.

Paso 5: Asegurar la calidad

El seguimiento y análisis en el cumplimiento de los procesos establecidos y las responsabilidades asignadas, de las herramientas tecnológicas, el desempeño, la asertividad de las estrategias elegidas para superar los retos y riesgos tanto técnicos como administrativos del archivamiento web, es una actividad que debe gestionarse de forma transversal y continua durante la ejecución del proyecto para identificar oportunidades de mejora y evitar desvíos en el enfoque de los métodos de trabajo.

El resultado de este paso debe generar estrategias o alternativas de solución para asegurar la calidad del archivamiento.



Glosario

ARC: Formato creado por Internet *Archive* para la captura y archivado de sitios web.

Crawler: *Software* que indexa o descarga contenido de la web de forma automática.

WARC: *Web Archive,* formato estándar por ISO 28500 para la captura y archivado de recursos web.

Bibliografía

AUSTRALIA. NATIONAL ARCHIVES OF AUSTRALIA. Archiving web resources: guidelines for keeping records of web-based activity in the commonwealth government [En línea]. <http://www.naa.gov.au/Images/archweb_guide_tcm16-47165.pdf> [citado el 12 de octubre de 2013]

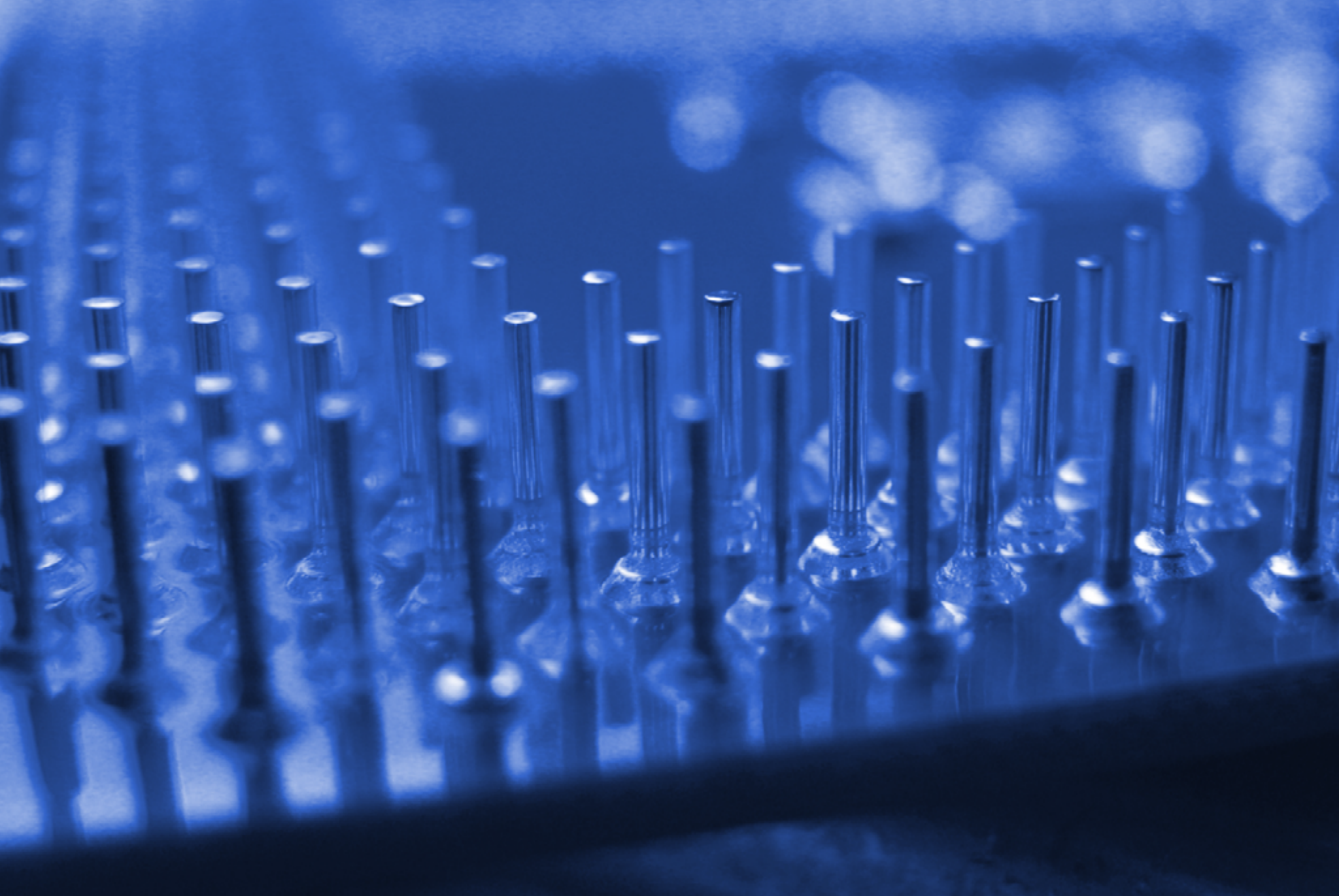
BALL, Alex. Web Archiving [en línea]. <http://www.dcc.ac.uk/sites/default/files/documents/reports/sarwa-v1.1.pdf> [citado el 5 de octubre de 2013]

BRAGG, Molly y HANNA, Kristine. The web archiving life cycle model [En línea]. <http://archive-it.org/static/files/archiveit_life_cycle_model.pdf> [citado el 11 de octubre de 2013]

COLOMBIA. MINISTERIO DE TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES. Estrategia de Gobierno en Línea [En línea]. <http://programa.gobiernoenlinea.gov.co/apc-aa-files/e5203d1f18ecfc98d25cb0816b455615/minticmanual3.0.pdf> [citado el 2 de octubre de 2013]

PENNOCK, Maureen. Web Archiving: DPC Technology Watch Report 13-01 March 2013[En línea]. <http://www.dpconline.org/component/docman/doc_download/865-dpctw13-01pdf > [citado el 2 de octubre de 2013]

UNITED KINGDOM. THE NATIONAL ARCHIVES. Web Archiving Guidance [En línea]. <http://www.nationalarchives.gov.uk/documents/information-management/web-archiving-guidance.pdf> [citado el 3 de octubre de 2013]



@ArchivoGeneral | Archivo General | CanalAGNColombia | AGN Colombia

Archivo General de la Nación - Colombia
Establecimiento público adscrito al Ministerio de Cultura
Carrera 6 No. 6-91 - Tel: 328 2888 - Fax: 337 2019
contacto@archivogeneral.gov.co - www.archivogeneral.gov.co
Bogotá D.C - Colombia

ARCHIVO GENERAL DE LA NACIÓN COLOMBIA