

Transformación digital de un archivo histórico mediante el uso de un software de código abierto: el caso del Archivo Histórico de la Real Caja de Zacatecas

SARA MARÍA DEL PATROCINIO RODRÍGUEZ PALACIOS

Tecnológico de Monterrey (México)

sara.rodriquez@tec.mx

RESUMEN

Las entidades y agentes culturales tienen como misión administrar y gestionar sus bienes y servicios, estableciendo para ellos los procedimientos, normas y herramientas que garanticen el logro de sus objetivos. En su compromiso por preservar, dar visibilidad y acceso a sus contenidos, la Dirección Nacional de Patrimonio Cultural del Tecnológico de Monterrey gestionó la digitalización del Archivo Histórico de la Real Caja de Zacatecas, 1576-1936, mismo que custodia, conserva, difunde y tiene en comodato su Campus Zacatecas. En esta ponencia se compartirán algunos aspectos que se contemplaron para garantizar el éxito de este proceso: organización de los contenidos para reflejarlos en la herramienta tecnológica de gestión del Repositorio Institucional del Tecnológico de Monterrey (RITEC) para reflejar el cuadro de clasificación archivística; ajustes a la política de desarrollo de colecciones, definición de niveles de acceso, aprovisionar las capacidades de almacenamiento de información o de transferencia de datos; apego a un estándar de metadatos que permitiera registrar las características particulares; los procesos de preservación, la curaduría de metadatos y la importante colaboración de un gran equipo

interdepartamental e interdisciplinario. Se comparten los resultados y se propone una innovación tecnológica como estrategia de mejora para la gestión de los recursos digitalizados.

ABSTRACT

The mission of cultural entities and agents is to administer and manage their goods and services by establishing procedures, norms and tools that guarantee the achievement of their objectives. In its commitment to preserve, give visibility and access to its contents, the National Direction of Cultural Heritage of the Tecnológico de Monterrey managed the digitization of the Historical Archive of the Real Caja de Zacatecas, 1576-1936, which is kept, conserved, disseminated and on loan at its Zacatecas Campus. In this paper we will share some aspects that were contemplated to guarantee the success of this process: organization of the contents to reflect them in the technological management tool of the Institutional Repository of the Tecnológico de Monterrey (RITEC) in order to reflect the archival classification chart; adjustments to the collection development policy, definition of access levels, provisioning of information storage or data transfer capacities; adherence to a metadata standard that allowed the recording of particular characteristics; preservation processes, metadata curation and the important collaboration of a large interdepartmental and interdisciplinary team. The results are shared and a technological innovation is proposed as an improvement strategy for the management of digitized resources.

PALABRAS CLAVE

Acceso abierto; patrimonio cultural; transformación digital; repositorios de acceso abierto.

KEYWORDS

Open access; cultural heritage; digital transformation; open access repositories.

1. Introducción

El Archivo Histórico de la Real Caja de Zacatecas, 1576-1936 forma parte de la Colección Sociedad de Amigos de Zacatecas, A. C., es uno de los principales archivos históricos zacatecanos. Como menciona MAURICIO-ESCALANTE (2021), durante décadas este acervo documental estuvo en la Biblioteca Clements de la Universidad de Michigan en Ann Arbor y regresó a Zacatecas en 1992, gracias a las gestiones del Lic. Manuel Fernando Sescosse Varela. En este acervo documental se encuentra la administración y recaudación de impuestos que se realizaron desde el último cuarto del siglo XVI hasta principios del siglo XX, en la ciudad de Zacatecas. La minería, el comercio, las misiones religiosas al Norte de México, la milicia que se estableció en la ciudad durante el siglo XIX, el reparto de azogue y sal de las Salinas de Santa María y Del Peñol Blanco, el Colegio Seminario de San Luis Gonzaga, entre otros.

2. Desarrollo

2.1. Marco teórico

Los archivos históricos resguardan información social, política y económica de las sociedades. BABINES LÓPEZ hace hincapié en los valores secundarios: evidenciales, testimoniales e informativos.

Los evidenciales son los que muestran derechos y obligaciones en una sociedad; en este rubro podemos englobar a las leyes, reglamentos, decretos, etc., considerados base de la historia institucional. Los testimoniales nos muestran la evolución de las instituciones, se asigna este valor a los documentos generados por funciones sustantivas de la institución. Y por último, los informativos, documentos que sirven de instrumento de información para el usuario, la entidad productora y como fuente para la investigación política, social, demográfica, cultural, religiosa, etc. (BABINES LÓPEZ, 2022)

Los contenidos de los archivos históricos son, o deberían ser, para la sociedad. Sin embargo, temas como la preservación física (son documentos que requieren ciertas condiciones de luz, humedad, etc.), la ubicación geográfica (un investigador puede tener que hacer largos viajes para poder analizar los contenidos de los archivos), las gestiones requeridas para consultas (cartas, justificaciones, cuadrar agendas) y otros que se puedan acumular se convierten en una barrera difícil de sortear.

La transformación digital parece ser la forma como se puede garantizar el acceso y visibilidad más inclusivos. En palabras de TORRES (2020), esta transformación digital requiere tres procesos sucesivos:

- digitización (conversión de la información analógica a un formato digital) de los recursos documentales;
- digitalización (proceso de cambio) que implica garantizar el establecimiento de procesos, normas, técnicas y herramientas que garanticen la gestión, acceso y visibilidad de los recursos, y
- transformación digital (efecto o consecuencia de la digitalización de los procesos de negocios) que implica un cambio cultural, el cual involucra la participación de los contenidos en la vida digital de las organizaciones y sociedades.

2.2. Descripción y proceso de implementación de la innovación

La transformación digital del Archivo Histórico de la Real Caja de Zacatecas, 1576-1936 debió tener en cuenta muchos aspectos; se hará mención de aquellos más relacionados a la gestión digital.

Se seleccionó un software de código abierto que provee herramientas para la administración de colecciones digitales, DSpace en su versión 6.3, ya que es el medio utilizado por el Repositorio Institucional del Tecnológico de Monterrey (RITEC), sitio donde se decidió agregar estos recursos por ya existir una Comunidad de Patrimonio Cultural. El RITEC tiene como propósito ser la ventana única donde el Tecnológico de Monterrey comparte todos los recursos científicos, académicos y administrativos que genera, así como el patrimonio

documental y artístico que se encuentra bajo su resguardo y que está disponible en Acceso Abierto.

El RITEC se organiza utilizando la estructura ofrecida por DSpace, donde las comunidades son el nivel más alto de la jerarquía de contenidos. Pueden estar formadas por subcomunidades y por colecciones, estas últimas son las que contienen los ítems directamente, tal como se muestra en el GRÁFICO 1.

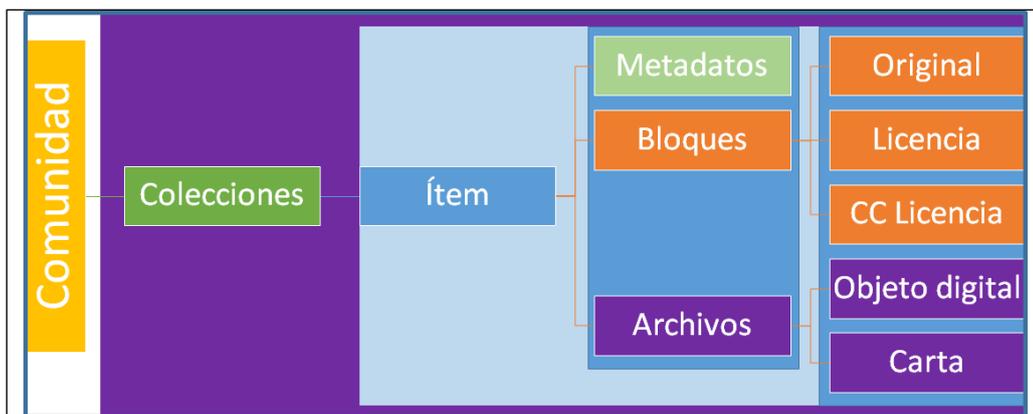


GRÁFICO 1. Estructura de DSpace 6.x

Como puede observarse, una Comunidad es el nivel más alto de la jerarquía de contenidos de DSpace. Pueden estar formadas por Subcomunidades y por Colecciones, estas últimas son las que contienen los ítems directamente.

Se realizó un análisis profundo del cuadro de clasificación archivística para ajustar sus necesidades a la estructura que ofrece DSpace, por lo cual se pudo realizar el ajuste de manera exitosa, quedando como se muestra en el ejemplo del GRÁFICO 2:

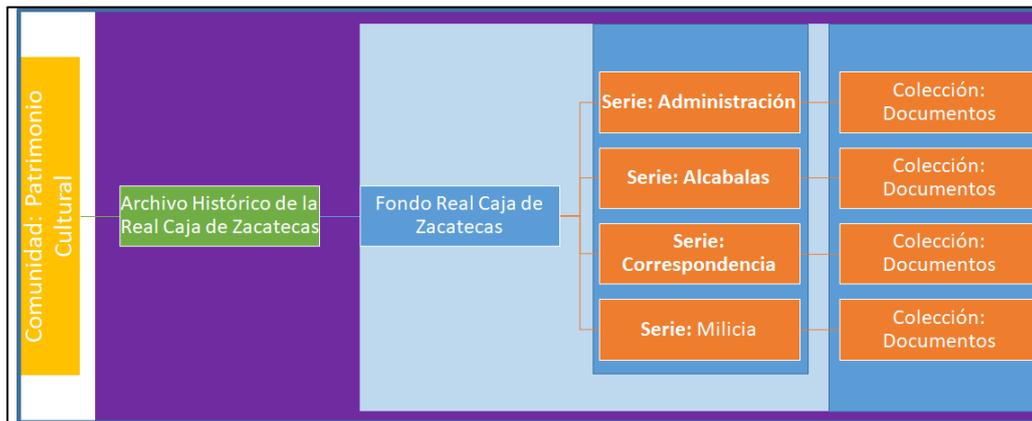


GRÁFICO 2. Estructura del Archivo Histórico de la Real Caja de Zacatecas, Fondo, Series y Colecciones

La política del desarrollo de colecciones del RITEC contemplaba inicialmente a todas las comunidades en general; sin embargo, hubo que hacer una actualización para ajustar las condiciones particulares que aplican a los materiales del archivo histórico, entre otros:

- formatos de ficheros
 - a) divulgación en PDF/A con reconocimiento óptico de caracteres (OCR)²
 - b) de preservación en TIFF
- nivel de acceso
 - a) formatos de divulgación en acceso abierto con Licencia Creative Commons específica por colección
 - b) formatos de preservación con acceso restringido, solo se permite su lectura a los administradores de las colecciones

Una vez digitalizados los recursos, se realizaron cargas masivas al RITEC mediante la herramienta Batch Metadata Editing, que permite realizar importaciones de ítems completos, que quedan depositados en el Repositorio en acceso abierto. Se siguió el proceso señalado en el GRÁFICO 3.

² El estándar PDF/A: PDF para la preservación a largo plazo. <https://es.scribd.com/document/240311146/m2u2-anexo09-estandar-pdf-a-pdf>

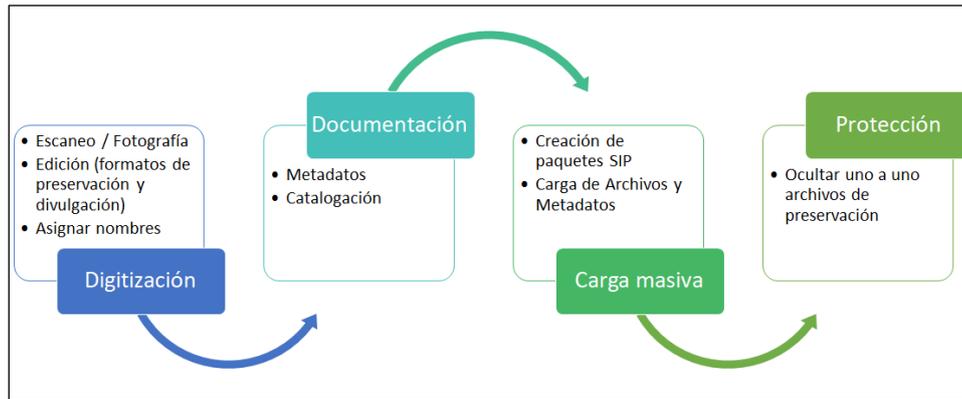


GRÁFICO 3. Flujo de depósito en el RITEC de recursos digitalizados

El sistema DSpace permite regular el nivel de acceso desde cualquiera de sus niveles, y siempre las modificaciones de un nivel afectan a todos los subniveles que contiene; por lo cual nos encontramos con un reto muy importante, ya que se tuvo que comenzar a hacer la protección de acceso solamente a los objetos digitales de preservación, esto implicó una mayor inversión económica y de tiempo; en el caso de ítems con archivos muy numerosos se tuvo que programar la restricción de acceso uno a uno. Como se muestra en el GRÁFICO 4, para el caso de una foja el trabajo que esto implica puede parecer rápido, sin embargo, cuando hablamos de cargas masivas y de expedientes numerosos surgen al menos tres áreas de oportunidad:

1. Seguridad (ventana de tiempo) los archivos de preservación que deben estar protegidos se encuentran expuestos en tanto se termina de realizar el cambio de nivel de acceso.
2. Tiempo invertido en protección. Es muy significativo.
3. Operación (no depender de horarios de carga). Como debe asegurarse la protección de todos los archivos de preservación tan rápido como sea posible, se tuvo que restringir el horario en que se realizaban las cargas masivas, para que el equipo colaborador tuviera el tiempo suficiente dentro de sus horas laborales para realizar la protección uno a uno.



GRÁFICO 4. Áreas de oportunidad para protección de archivos de preservación

A raíz de esta situación, se gestionó ante el equipo de administración tecnológica el desarrollo de un módulo que permitiera realizar estas cargas masivas con ficheros protegidos donde se pudieran asignar los privilegios de lectura de cada fichero a grupos de usuarios definidos. Este desarrollo permitió la eficiencia en tres rubros, como se muestra en el GRÁFICO 5, un menor costo ya que se utilizaron menos horas de trabajo del equipo especializado para realizar esta tarea; ahorro en tiempo ya que el añadir las características de nivel de acceso en el repositorio es inmediato en cuanto se realiza la carga masiva y, finalmente, una mejora muy significativa en cuanto al nivel de seguridad, pues ningún recurso queda expuesto desde el momento en que es depositado.

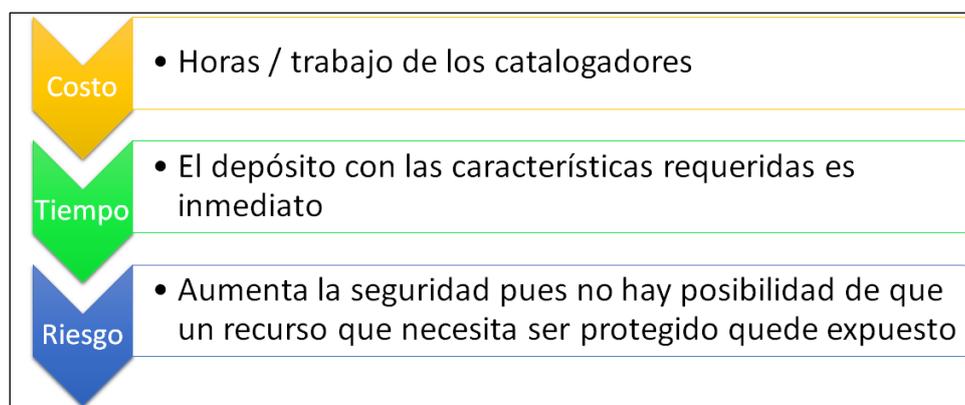


GRÁFICO 5. Eficiencia del desarrollo de cargas masivas protegidas en el repositorio

Fueron requeridas cuatro semanas para el desarrollo de la herramienta que permitiera cubrir los siguientes requerimientos:

1. Generar estructuras de importación, que pueda proteger los ítems e indicar al sistema que suben en oculto.
2. El habilitador debe indicar, con una clave o con un metadato adicional en la hoja de cálculo donde se registran los metadatos, si un fichero asociado a un ítem tendrá permisos restringidos.
3. Que funcione con interfaz web preferentemente, o en su defecto, en modo consola (indicando los pasos a seguir para que una persona de la institución pueda hacer este trabajo en modo consola, e indicando también los comandos Linux de base).
4. Capacidad para seleccionar el grupo específico que tendrá acceso a los ficheros, básica y prioritariamente el Administrator y en segundo lugar algún otro que se pueda seleccionar dependiendo de las necesidades específicas de diseminación de las colecciones.

Tras esta actividad se implementó el flujo para cargas masivas protegidas, que tal como se muestra en el Gráfico 6 contempla los siguientes puntos:

Pasos previos

1. Se genera una hoja de cálculo en Excel donde cada columna se encabeza con la etiqueta Dublin Core de los metadatos a agregar. En una columna adicional se agregan los nombres de archivos y el nombre del Grupo que tiene privilegio de lectura.
2. Utilizando un habilitador desarrollado para esto, se generan carpetas SIP que relacionan los metadatos y sus respectivos archivos. Estos archivos se comprimen en un ZIP.

Carga masiva desde interfaz web

3. El usuario con privilegios de Administrator ingresa al Repositorio con su cuenta.
4. Selecciona la Herramienta Importación Batch (ZIP)
5. Elige la colección en donde se estarán depositando los recursos.
6. Selecciona el archivo ZIP generado.
7. Da clic en el botón [Subir un ZIP SimpleArchiveFormat]
8. DSpace genera un reporte con los recursos agregados.

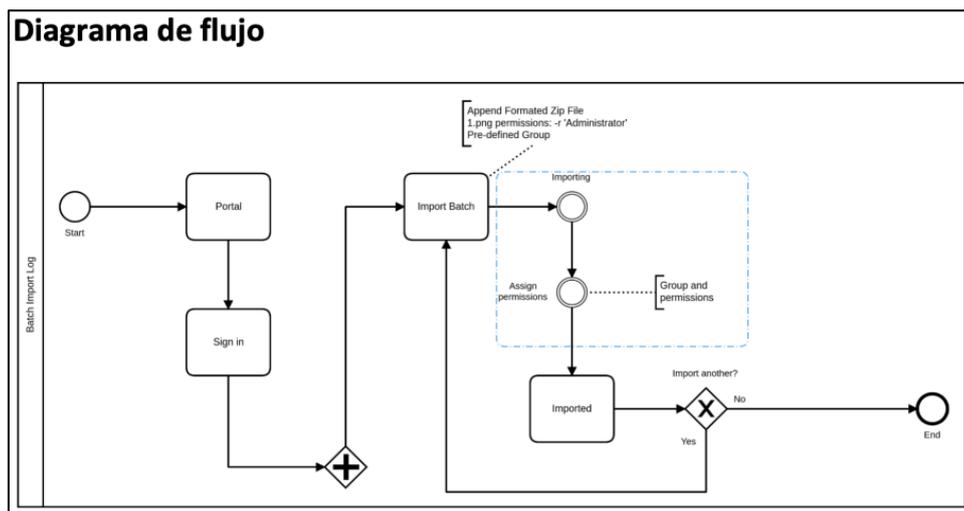


GRÁFICO 6. Diagrama de flujo de cargas masivas protegidas en el repositorio

Los metadatos son muy importantes para comprender qué contienen los recursos, además de aportar información indispensable sobre su estructura y son utilizados para su gestión administrativa. El cumplir con estándares internacionales garantiza la buena gestión, proporciona vínculos entre los recursos y su contexto de creación, permite la interoperabilidad para reconocer, procesar y usar en otros entornos, se cumple con requisitos legales de uso y distribución, reduce el riesgo de acceso no autorizado y facilitan los procesos de conversión, migración y conservación a largo plazo. Por ejemplo, en el caso de metadatos de fechas pues no aplica el estándar de uso exclusivo del formato ISO 8601 que dicta los cuatro dígitos del año, guion, dos dígitos para mes, guion, dos dígitos para día: AAAA-MM-DD. Las colecciones especiales y

en particular este archivo histórico requiere una entrada de texto abierto en fecha de creación para poder hacer anotaciones correspondientes a fecha cierta o aproximada, tales como: década probable, década segura, siglo probable, siglo seguro, fecha probable, posterior a, anterior a, cerca de o un periodo de tiempo comprendido entre dos fechas.

Aunado al esfuerzo de integrar estándares internacionales se realizó un trabajo colaborativo con el equipo especializado de catalogación para la asignación de metadatos y vocabularios controlados pertinentes, permitiendo que esta labor de curaduría refleje parte de la riqueza y belleza de las características particulares de cada tipo de recurso. Como resultado, se logró la creación de los siguientes documentos:

- Fuentes de consulta del vocabulario controlado para el Repositorio Institucional Tec (RITEC)³
- Glosario de términos bibliográficos, archivísticos, fotográficos y museográficos⁴

Se realizó, además, una proyección para determinar con anticipación las capacidades de almacenamiento de información o de transferencia de datos que serían requeridas, se muestra en la TABLA 1.

Tipo de recurso	Cantidad de ítems	Promedio de archivos por ítem		Peso unitario MB		Peso promedio Mb por ítem	Peso TOTAL Mb	Peso TOTAL Tb
		TIFF	PDF/A	TIFF	PDF/A			
Documentos y libros	8,710	20	1	25	150	650	5661500	5.3

TABLA 1. Cálculo de espacio para los recursos digitales del Archivo Histórico de la Real Caja de Zacatecas, 1576-1936 en el RITEC

³ Fuentes de consulta del vocabulario controlado para el Repositorio Institucional Tec (RITEC): <https://hdl.handle.net/11285/636607>

⁴ Glosario de términos bibliográficos, archivísticos, fotográficos y museográficos: <https://hdl.handle.net/11285/636686>

Dado este análisis, hubo que aprovisionar una capacidad de almacenamiento en el RITEC de 5,5 TB mediante Amazon S3 Glacier⁵, el cual ofrece el almacenamiento para los datos de larga duración que requieren una recuperación inmediata y copias de seguridad. La capacidad de almacenamiento requerida fue de 5,4 TB, muy similar a la proyectada. Adicionalmente se tiene una copia de todos los archivos en un disco externo.

2.3. Evaluación de resultados

Podemos comenzar a visualizar el gran impacto de la transformación digital de este acervo, simplemente en el periodo enero-abril 2021, el Archivo Histórico de la Real Caja de Zacatecas, 1576-1936 ha tenido 906 visitas y 858 descargas, se distribuyen como se muestra en el GRÁFICO 6.

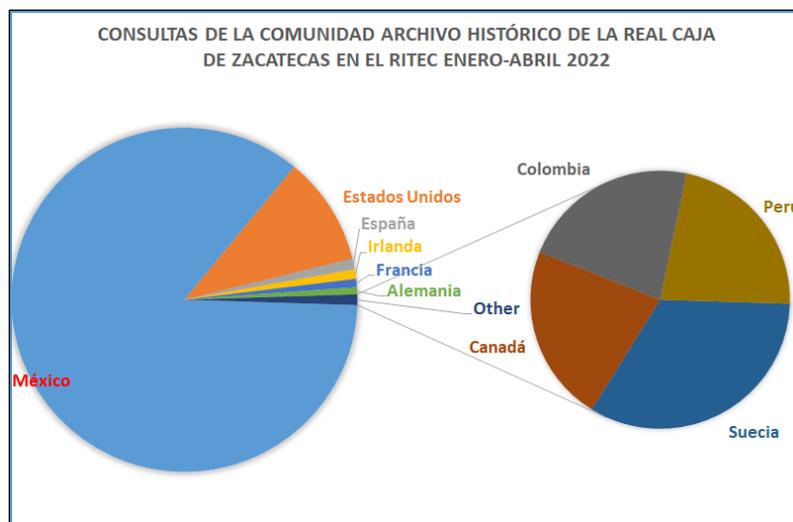


GRÁFICO 6. Visitas a ítems de las colecciones del Archivo Histórico de la Real Caja de Zacatecas en el RITEC en el periodo enero-abril 2022. Fuente:

<https://repositorio.tec.mx/handle/11285/636333>

La gran parte de visitas es de México; sin embargo, podemos observar que se extienden también hasta otros países y continentes. En la medida que se siga dando a conocer este acervo se tiene la expectativa que pueda tener un

⁵ Amazon S3 Glacier: <https://aws.amazon.com/es/s3/storage-classes/glacier/?refid=5970b1e9-218b-48cc-9862-f23c151d81b2>

crecimiento exponencial la visibilidad y consulta de los recursos que, en resumen; quedaron constituidos por un total de 8.718 ítems, divididos en 36 series (colecciones), tal como se muestra en el GRÁFICO 7.

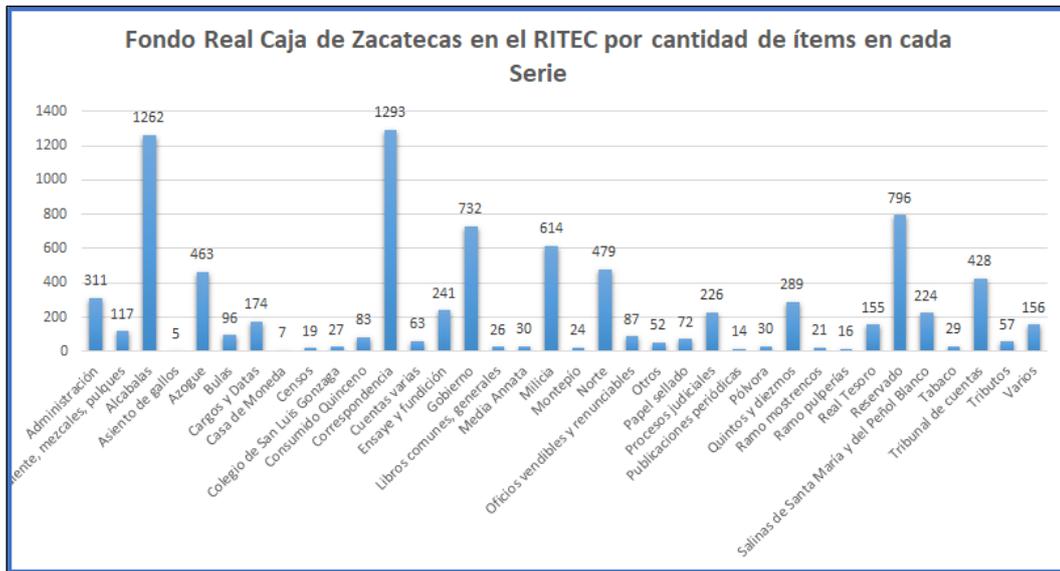


GRÁFICO 7. Fondo Real Caja de Zacatecas en el RITEC por cantidad de ítems en cada Serie.

Fuente: <https://repositorio.tec.mx/handle/11285/636333>

Cabe resaltar el trabajo previo de organización que permitió tener una estructura para el mejor aprovechamiento de sus contenidos. El proceso señalado es complejo e incluyó, como menciona MAURICIO ESCALANTE (2021), elaborar listas o inventarios, en los cuales se plasma una pre-catalogación que permite dar cuenta de la cantidad de material que contiene el acervo, definición de metros lineales y los volúmenes (expedientes o libros), temporalidad del archivo, autores, personajes y otros datos básicos que permiten enriquecer la catalogación de los recursos en el repositorio.

3. Conclusiones

El RITEC tiene un gran compromiso con la accesibilidad e inclusión, aunque en este aspecto aún queda mucho camino por recorrer, se propone implementar un software de reconocimiento óptico de caracteres (OCR) en el

Repositorio Institucional que, en conjunto con un visor de imágenes IIIF, brinde soporte para permitir la activación de la funcionalidad de búsqueda dentro del visor de IIIF, proporcionando búsqueda dentro de las imágenes, navegación a través de los resultados y resaltado en la imagen del texto OCR correspondiente a los términos de búsqueda ingresados. Se está ponderando el depósito de formatos de archivos de preservación que permitirán, además de garantizar la accesibilidad permanente de los objetos digitales, más adelante implementar estas iniciativas.

Adicional a ello, se sugiere fortalecer la preservación de los recursos digitalizados implementando un servicio de alta capacidad de almacenamiento y duplicación de la información que interactúe con el RITEC mediante la sincronización de metadatos, de tal manera que se eficienten los costos de almacenamiento y transferencia de datos.

Es de gran importancia dar visibilidad y acceso a estas colecciones patrimoniales como una gran estrategia para corresponder a los ideales de los Objetivos de Desarrollo Sostenible (ODS)⁶. Los contenidos de un archivo histórico que se pueden volver inaccesibles por temas de preservación física, podrán estar disponibles al mundo entero gracias al uso de la tecnología de manera equitativa e inclusiva, mediante procesos de transformación digital que promuevan oportunidades de aprendizaje permanente para todos.

Cerramos agradeciendo a todo el equipo que ha participado en el flujo de trabajo completo: la organización previa del archivo, la definición del cuadro de clasificación archivística, la revisión y documentación de plantillas de metadatos, la digitización, envío de archivos, recepción, catalogación y documentación, así como la gestión general de las colecciones patrimoniales en el Repositorio Institucional del Tecnológico de Monterrey.

⁶ Objetivos de Desarrollo Sostenible: <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>

Bibliografía

- AVILA BANDA, H. (2020, August 15). Glosario de Términos bibliográficos, Archivísticos, fotográficos y museográficos. Repositorio Institucional del Tecnológico de Monterrey. <https://hdl.handle.net/11285/636686>
- AVILA BANDA, H. (2020, July 23). Fuentes de Consulta del Vocabulario Controlado para el Repositorio institucional TEC (RITEC). Repositorio Institucional del Tecnológico de Monterrey. <https://hdl.handle.net/11285/636607>
- BABINES LÓPEZ, J. (n.d.). La importancia de los documentos históricos. – Archivo General del Estado de Oaxaca. <https://www.oaxaca.gob.mx/ageo/la-importancia-de-los-documentos-historicos/>
- EMAZ, A., & PÉREZ-RUIZ U. (2021). Amazon S3 Glacier. Amazon. <https://aws.amazon.com/es/s3/storage-classes/glacier/?refid=5970b1e9-218b-48cc-9862-f23c151d81b2>
- MAURICIO ESCALANTE, Z. M. (n.d.). Archivo Histórico de la Real Caja de Zacatecas custodiado por el Tecnológico de Monterrey en Campus Zacatecas. AMABPAC. <https://www.amabpac.org.mx/wp/archivo-historico-de-la-real-caja-de-zacatecas-custodiado-por-el-tecnologico-de-monterrey-en-campus-zacatecas/>
- MAURICIO ESCALANTE, Z. M., & MACÍAS CHIU, A. L. (2021, November 23). El problema de la Memoria en colecciones digitales. Preservación digital del patrimonio cultural. Retos, aprendizajes y reflexiones desde el Tecnológico de Monterrey. Patrimonio en un bit. <https://muac.unam.mx/patrimoniobit/en-vivo/el-problema-de-la-memoria-en-colecciones-digitales.html>
- PDF ASSOCIATION. (2022). PDF/a. <https://www.pdfa.org/wp-content/uploads/2011/07/pdfa-flyer-esp.pdf>
- TORRES, J. (2020, July 30). Los retos de la transformación digital para los repositorios y las bibliotecas digitales. figshare. https://figshare.com/articles/presentation/Dia_2_Joel_Torres_Transformacio_n_digital_pptx/12736124
- UNITED NATIONS. (2020). Objetivos y Metas de Desarrollo Sostenible - Desarrollo Sostenible. United Nations. <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>