

---

# Estrategia para la Integración de Catálogos de Fichas Impresas de Tres Bibliotecas Cubanas al Espacio de Datos Enlazados Abiertos

Pedro Urra González

---

## RESUMEN

En este trabajo se resumen los principales aspectos de la estrategia que se construyó como resultado del proyecto para la integración de catálogos impresos al espacio de la biblioteca digital cubana y de la WEB en general, la situación en la que se encuentra la iniciativa y se reflexiona sobre la relación con procesos paralelos de desarrollo de catálogos en línea, bibliotecas digitales y repositorios digitales del patrimonio cultural destacando las oportunidades del uso de técnicas de datos enlazados para estos fines.

## INTRODUCCIÓN

Desde finales de los años 1990 se generalizó el uso de catálogos automatizados en muchas bibliotecas cubanas con la extensión del uso del MicroI-sis, no obstante, a partir de la información recabada sobre los catálogos de tres importantes bibliotecas del país en el año 2014, se constató que un volumen considerable de los registros bibliográficos de sus colecciones se encontraban solamente en forma de catálogos de fichas impresas. La versión para microcomputadora del CDS/ISIS fue introducida y promovida en Cuba y el resto de América Latina desde finales de los años 1980 con el apoyo del programa de Información de la UNESCO (Ugobono 2011). En el caso de la Biblioteca Nacional de Cuba José Martí (BNJM), los documentos que ingresaron a sus fondos antes de 1996, salvo excepciones, no contaban con registros bibliográficos legibles por computadoras y por tanto no era posible su consulta a través de Internet. En una situación similar estaba la Biblioteca Central de la Universidad de la Habana (BCUH) y en el caso de la Biblioteca del Instituto de Historia de Cuba (BIHC) to-

dos los registros bibliográficos estaban en formato de fichas catalográficas impresas. Estas tres instituciones fueron seleccionadas para desarrollar un proyecto para la digitalización de catálogos impresos como parte de un proyecto de la Dirección de Informatización de la Universidad de la Habana bajo la supervisión del autor del presente trabajo.

Las tarjetas físicas de la Biblioteca Nacional cubrían más del 50% de su colección. Las tarjetas físicas de la Biblioteca Central de la Universidad cubrían más del 60% de su colección. No fue posible procesar todos estos artículos de nuevo para crear registros en un catálogo en línea. En el caso de la Biblioteca Nacional y la Universidad, en el propio año 2014, esta situación convivía con la explotación de sistemas automatizados de gestión bibliotecaria que solo contaban con las referencias bibliográficas de una parte minoritaria de sus acervos bibliográficos. Por tanto, la tarea de la conversión retrospectiva de los catálogos impresos en esas instituciones seguía siendo una necesidad no resuelta con las consiguientes limitaciones de la gestión de sus colecciones y la duplicidad de funciones en la gestión de los documentos y la prestación de servicios a usuarios. Al tratarse de instituciones con importantes colecciones nacionales y con catálogos impresos caracterizados por una alta calidad de las descripciones bibliográficas acumuladas históricamente, estaba descartada la opción de llevar adelante un proceso de recatalogación que desechara los registros bibliográficos impresos y no era viable por su volumen procesar nuevamente el conjunto de las colecciones. Al propio tiempo el volúmen y la complejidad de esta tarea, así como limitaciones de recursos, había aplazado por muchos años su solución. Todo esto justificó la propuesta de llevar adelante un proyecto para desarrollar una prueba de concepto, que ayudara a solucionar este problema con el apoyo de procesos de digitalización y de búsqueda de nuevas estrategias que priorizaran la incorporación de esa información, una vez digitalizada, al espacio de la WEB.

La visión de contar con una Biblioteca Digital Nacional de Cuba estaba condicionada en parte por esa realidad. Ello justificó la decisión de las instituciones seleccionadas, en colaboración con la Dirección de Informatización de la Universidad de la Habana, de desarrollar con recursos propios de las instituciones participantes, un proyecto que permitiera diseñar una estrategia para completar la conversión de catálogos impresos de estas bibliotecas. El proyecto partió de una revisión del estado del arte de ese tema, el estudio de las experiencias de otras bibliotecas que desarrollaron este tipo de procesos en décadas anteriores, el reconocimiento de las particularidades del contexto cubano y las nuevas oportunidades derivadas del avance tecnológico y metodológico en este campo. Desde el punto de vista de los propósitos, la estrategia se alineó con los principios para la construcción de una Biblioteca Digital Nacional de Cuba (Biblioteca Nacional de Cuba "José Martí" 2013).

## EL NUEVO CONTEXTO PARA LA CONVERSIÓN DE LOS CATÁLOGOS IMPRESOS

La conversión retrospectiva de catálogos impresos fue una tarea a la que las bibliotecas dedicaron muchos esfuerzos en décadas pasadas. Esto generó un grupo de experiencias, herramientas y conocimientos que constituyen un antecedente de cualquier esfuerzo en esta dirección (Chapman 1996; Harrison 1985; Hsueh 1992). No obstante a lo anterior, el escenario y las estrategias de estos procesos debe ajustarse a las nuevas realidades del contexto contemporáneo entre las que se destacan la generalización del uso de Internet y en particular del Web como plataforma para el acceso a la información, y el crecimiento de la disponibilidad de datos bibliográficos normalizados en catálogos de bibliotecas nacionales e internacionales como el Worldcat de OCLC y los ficheros de autoridades asociados a los mismos (Loesch 2011; Yang y Hofmann 2010).

A lo anterior se suma el desarrollo de nuevos enfoques y técnicas para la gestión de los datos en las que tiene particular importancia la maduración y el creciente uso de las técnicas de datos enlazados por la comunidad bibliotecaria y de otras instituciones del patrimonio cultural (Moulaison y Million 2014). La más reciente guía de IFLA para las bibliografías nacionales dedica un apartado especial al uso de las técnicas de datos enlazados (IFLA 2015). Igualmente, OCLC y un grupo importante de bibliotecas nacionales han comenzado a exponer sus datos bibliográficos siguiendo estas pautas y la Biblioteca del Congreso de los Estados Unidos lanzó la iniciativa BIBFRAME, que supone un cambio fundamental en cuanto a las estrategias para la catalogación de las colecciones bibliotecarias, en las que reconoce a los datos enlazados como una oportunidad para revitalizar las funciones de las Bibliotecas en la era de Internet (Kroeger 2013, 2015, 2016; Zapounidou, Sfakakis, y Papatheodorou 2013, 2017).

Otro elemento a considerar como parte del contexto de los procesos de conversión de catálogos de bibliotecas es el surgimiento y desarrollo de las iniciativas para la construcción de Bibliotecas Digitales y Repositorios Digitales. Hoy prácticamente todos los países de la región Latinoamericana y caribeña han desarrollado proyectos de este tipo (Gómez-Acebo y Sánchez Nogales 2013). Sin embargo, salvo excepciones, coexiste un número considerable de colecciones parciales e incompletas, se duplican contenidos, y los datos de los acervos bibliográficos siguen estando desconectados entre sí y entre colecciones distintas. La disponibilidad de catálogos de calidad sigue siendo una carencia. Este proceso se multiplica y extiende y aunque crece aceleradamente la disponibilidad de copias digitales de documentos producidos en países latinoamericanos y caribeños integrados a repositorios globalizados, quedan ocultas importantes colecciones atesoradas localmente y se pierde la riqueza de los contextos culturales e históricos y las trazas asociadas a la vida social de dichos documentos en

cada uno de los países y regiones (ver por ejemplo Gregory y Williams 2014).

Esto convive con una ofensiva por parte de sectores empresariales por apropiarse de forma privada de los beneficios que genera el uso de la información acumulada en los fondos documentales de las bibliotecas. Lo anterior pone de relieve la importancia que tienen estos fondos no sólo en términos científicos y culturales sino también como activos económicos y plataforma para el desarrollo de numerosos servicios (Rieger 2008). La renovada Estrategia Europea para el desarrollo de la Biblioteca Europea es un ejemplo de enfoque en la dirección de rescatar y reforzar el papel proactivo de las bibliotecas y otras instituciones patrimoniales en la protección de los bienes públicos de carácter cultural y en su puesta al servicio de la comunidad (Poole, Racine, y Cousins 2014.)

Las iniciativas de bibliotecas y repositorios digitales, más que acciones desconectadas deben formar parte de un esfuerzo articulador e integrador vinculado a la defensa de la identidad cultural como base de una mejor convivencia entre los pueblos y las naciones. En este contexto también se pueden citar los esfuerzos para el rescate, la preservación y diseminación del patrimonio documental representado en la iniciativa Memoria del Mundo de la UNESCO y sus instancias regionales, subregionales y nacionales (Abid 1995; Sloggett 2005).

### BASES TEÓRICO-PRÁCTICAS DE LA ESTRATEGIA

Para orientar las acciones del proyecto se tuvo en cuenta la acumulación de conocimientos en procesos de conversión retrospectiva de catálogos impresos, se combinó la revisión profunda de la bibliografía disponible con la realización de entrevistas a expertos y grupos de discusión para identificar particularidades del escenario cubano así como prácticas y experiencias concretas. Se concluyó básicamente que en las condiciones de Cuba era necesaria una adecuación de la estrategia que tomara en cuenta las particularidades del entorno nacional.

Un estudio de fortalezas, oportunidades, amenazas y debilidades identificó una situación muy particular en la que junto a las limitaciones del acceso a Internet, la falta de equipamiento y recursos financieros, y las restricciones para el acceso a servicios y bases de datos internacionales convivían el legado de catálogos desarrollados de manera sistemática y con calidad, la existencia en las colecciones, de documentos valiosos únicos o escasos, la disponibilidad de fuerza de trabajo calificada, experiencias anteriores de procesos de digitalización y desarrollo de sistemas y la emergencia de nuevos paradigmas y posibilidades técnicas que ofrecen un espacio para la innovación y aprendizaje.

Como se señaló, el pilotaje con los Catálogos se insertó en el marco de una estrategia más global para el desarrollo de una Biblioteca Nacional Digital de Cuba, que se alinea al propio tiempo con estrategias regionales

y globales en esta dirección. También es parte de un proceso de aprender haciendo o construyendo.

El trabajo con los catálogos se abordó como un problema de integración de estos recursos al espacio abierto de los datos bibliográficos y no como un clásico problema de conversión retrospectiva que es lo que prevalece en la literatura y la práctica acumulada sobre el tema (Malmsten 2008). En lugar de hacer un proceso de conversión de los registros bibliográficos para obtener registros MARC como meta directa, se adoptó un proceso de iteraciones que permitiera incorporar progresivamente los datos de esos catálogos al espacio de Internet y lograr su interrelación con los datos existentes en la red. Aunque no se renunció a la posibilidad de obtener las representaciones normalizadas de las descripciones bibliográficas, en formato MARC, y poder incorporarlas a los Sistemas de Gestión Bibliotecaria de cada institución, se priorizó la digitalización de las fichas de los catálogos y su integración en un repositorio para una vez disponibles en Internet en formato digital, realizar procesos continuos de anotación con técnicas de datos enlazados

Existían las experiencias de desarrollo de los denominados “Card Images Public Access Catalogs (CIPACs)” estudiadas de manera detallada por Oberhauser (2002, 2003, y 2007). Se trata de la creación de representaciones “virtuales” de los catálogos impresos que de manera pragmática, ofrecen una representación digital en línea de los mismos mediante imágenes digitales de las fichas (ver fig. 1) ordenadas de manera similar o muy parecida a la forma en que existen físicamente. Así, sin alterar el orden original de los catálogos, se capturan las fichas bibliográficas y se disponen en un espacio Web de manera que el usuario pueda navegarlas en una lógica similar o bastante parecida a los catálogos en línea con la ventaja de estar en Internet, globalizando el acceso a los datos de las colecciones y potenciando la consulta a las mismas.

Estos catálogos, cuya primera experiencia documentada se remonta a los años 1990, aparecieron como una alternativa a los costosos procesos de conversión retrospectiva total de grandes bibliotecas y en algunos casos sirvieron como pasos intermedios de apoyo al completamiento de procesos de este tipo. En importantes bibliotecas siguen siendo la única alternativa para poder consultar desde Internet importantes colecciones. Existe un directorio de este tipo de catálogos disponibles en <http://cipacs.vfi-online.org>. Los mismos funcionan además como repositorios de imágenes de fichas que preservan la memoria documental y conservan trazas valiosas del funcionamiento de las instituciones bibliotecarias.

Algunos de los catálogos, por sus características particulares, su antigüedad y grado de preservación constituyen patrimonio del hacer bibliotecario del país y testimonio de la evolución histórica de los procesos técnicos, la sistematicidad del trabajo institucional y hasta testigos de etapas de abundancia y escasez reflejada en las características de las fichas (im-



Figura 1. CIPAC de uno de los catálogos de la Biblioteca Nacional Checa disponible en <https://retris.nkp.cz>.

portadas o hechas localmente), calidad de la mecanografía, nitidez de la impresión, alcance de las anotaciones y otros aspectos que dan para investigaciones disciplinarias muy interesantes.

La caracterización de los catálogos fue entonces uno de los pasos iniciales del proceso. Esta caracterización se realizó mediante revisiones in situ de los catálogos, la entrevista con responsables y personas vinculadas con su custodia y su historia. Se vieron los aspectos técnicos, materiales y socio culturales vinculados a los mismos así como su importancia dentro de la colección de las Bibliotecas. En el caso de la Biblioteca Nacional, se decidió por la dirección y el personal especializado de la institución hacer el pilotaje con la Sala Cubana Siglo XIX, considerando el valor de sus colecciones y también el tamaño relativamente menor (aproximadamente 130 gavetas) con respecto a la colección general (más de 700 gavetas). En el caso de la Biblioteca Central de la Universidad de la Habana, se decidió comenzar por el catálogo oficial de la Colección general por tener un tamaño manejable (92 gavetas) y estar en condiciones de conservación relativamente buenas. En el caso del Instituto de Historia, se seleccionó el único catálogo disponible el cual estaba en buen estado de conservación y tenía una cifra manejable (50 gavetas).

La experiencia de construcción de CIPAC, combinada con la aplicación de técnicas de datos enlazados apareció entonces como una alternativa interesante y viable en las condiciones cubanas lo cual fue presentado y discutido con expertos y en talleres en las instituciones involucradas. Se decidió entonces desarrollar un pilotaje para evaluar su viabilidad y aprender del proceso.

La solución propuso combinar técnicas de digitalización masiva, procesos de reconocimiento óptico de caracteres (OCR), anotación en red y

uso de las técnicas de datos enlazados. Los catálogos impresos se convirtieron entonces en objetos de digitalización dado su valor como testimonio de los procesos de construcción histórica de colecciones bibliográficas y como instrumentos para avanzar en la vinculación de los datos bibliográficos de las colecciones al espacio de datos enlazados.

## RESULTADOS

En consecuencia con los principios del proyecto de la Biblioteca Nacional Digital de Cuba, las intervenciones que se realizaron se orientaron a desarrollar capacidades locales y hacer un uso innovador y racional de los recursos disponibles. Estos principios guiaron la realización del pilotaje. A continuación se resumen un grupo de ejemplos de las acciones realizadas que ilustran los resultados en este sentido:

### *1. Hacer el Trabajo con el Personal Propio de las Bibliotecas y que Estuviera Directamente Vinculado con las Colecciones que Se Seleccionaron y con Acompañamiento de Expertos.*

Se decidió emprender el pilotaje con personal directamente relacionado con los procesos técnicos de la biblioteca; en el caso de la Biblioteca Nacional las tareas fueron completado por personal integrado a la propia Sala Cubana. El personal de otras áreas como la de informática, áreas de digitalización y otras, apoyaría el proceso, capacitaría, pero el trabajo se controlaría y lideraría desde el área responsable de la colección y sus catálogos. Se realizaron acciones de capacitación en la práctica y se vincularon a las personas de las tres instituciones estimulándose el apoyo mutuo y el entrenamiento de pares.

Acompañamiento especializado del proceso por personal de más experiencia. Realización de acciones de control de calidad durante todo el proceso. La mayor parte de las anotaciones, transcripciones y revisiones del texto recuperado a partir del OCR de las fichas ha sido realizado por personal de las bibliotecas participantes destacándose la participación de los bibliotecarios de la Universidad. En este caso, dada la relativamente mala calidad de las imágenes de las fichas el OCR arrojó resultados de baja calidad y ello se compensó con la transcripción masiva de las fichas por parte de estos especialistas.

### *2. Tratar de Utilizar al Máximo los Recursos Técnicos y Materiales Disponibles y Fomentar la Cooperación con Otras Áreas.*

Se identificaron recursos locales que servirían para realizar tareas de digitalización. Se combinó el uso de computadoras disponibles y solo en el caso de la Biblioteca Nacional se dedicó una computadora nueva completamente al proceso de procesamiento del material digitalizado, en las otras instituciones se compartió tiempo de máquina. Se realizaron innovaciones como el diseño y construcción de “estaciones de captura digital” muy sencillas que facilitarían la digitalización de las fichas y que considera-



Figura 2. Bibliotecaria del Proyecto Operando Estación de Captura Digital de Fichas. (Photo cortesía del autor.)

ban este formato (ver fig. 2). Estas estaciones se diseñaron de manera que pudieran utilizar una cámara digital comercial estándar con determinadas prestaciones y con el único criterio de que capturaran imágenes que posteriormente pudieran visualizarse correctamente y hacerle procesos de OCR. Dadas las características de las fichas, su tamaño y relativa uniformidad se experimentaron y evaluaron diferentes alternativas hasta llegar a un prototipo con el que se capturaron más de 100 mil fichas. Utilizar la infraestructura de las instituciones participantes incluyó la infraestructura informática y de comunicaciones y en caso de no contar con ella buscar el apoyo de otras instituciones del proyecto global de la Biblioteca Nacional Digital. Los catálogos que ya están en línea están hospedados en una de las instituciones que colabora con el proyecto, dadas las facilidades para este fin. (El hospedaje del Sistema en línea que soporta los CIPACs se realiza con el apoyo de Infomed-Centro Nacional de Información de Ciencias Médicas de Cuba [<http://www.sld.cu/>], institución que ha sido un socio fundamental para el desarrollo del proyecto.) Adicionalmente cada institución posee copias de trabajo en sus redes locales que se sintonizan regularmente con el servidor principal.

Una vez que se había roto la inercia para la captura de fichas catalográficas con el apoyo de la solución antes mencionada fue posible contar con un escaner con alimentador de fichas y con una resolución adecuada para los procesos de OCR. Esto adicionó un elemento de infraestructura técnica que elevó la eficiencia del proceso de captura de las fichas y permitió que el reconocimiento de caracteres fuera mucho mejor.

The screenshot displays the CIPAC (Cuban Integrated Bibliographic Catalog) interface. At the top, there is a search bar with the text "Palabras en el texto de las fichas" and a "Buscar" button. Below the search bar is a navigation menu with letters A through W, and a "X" button. The main content area shows a book record for "C. V. Villaverde Vill 1882" with the title "Cecilia Valdés o, La Loma del Ángel, novela de costumbres cubanas. Nueva York, Impr. de El Espejo, 1882, xi, 590 p., front., 23 cm." The record also includes the author "Cecilia Valdés o, La Loma del Ángel, novela de costumbres cubanas. Nueva York, Impr. de El Espejo, 1882, xi, 590 p., front., 23 cm." and the call number "PQ7389.V55C4 1941". To the right of the main record is a sidebar with the heading "Biblioteca Nacional José Martí" and "Catálogo de LIBROS Y FOLLETOS-Oficial. Fichas:18817". The sidebar contains search options for "Aleatoria", "Gaveta: 22", "Ficha: 578", and "Ir a". Below this, there are search engines "Google", "Worldcat", and "DPLA". A "Texto de la ficha 22 / 678" section shows a snippet of the book record. At the bottom of the main record, there are navigation buttons "Anterior" and "Siguiente", and a pagination control showing "1/8" and "14" with a "Ir" button. The footer of the page indicates "Biblioteca Nacional José Martí" and "Última Actualización: 07/12/2017 16:48:14".

Figura 3. Pantalla que Muestra el CIPAC en Línea Desarrollado por el Autor.

### 3. Combinar Técnicas y Metodologías Avanzadas con Métodos Tradicionales en Función del Objetivo.

Se modeló y diseñó un sistema accesible en el WEB para la gestión de CIPACs que se fue desarrollando progresivamente con la filosofía de beta permanente para la exposición de los catálogos y para la incorporación de anotaciones progresivas. Este sistema fue programado en Python por el autor (disponible en <http://bnjm.sld.cu>) (ver fig. 3).

Se seleccionó una muestra probabilística de las obras representadas en el catálogo oficial de la Colección Cubana usando un muestreo por conglomerados (gavetas) bietápico con selección sistemática en la segunda etapa (las fichas) para estudiar el catálogo y estimar la cantidad de fichas que contaban con registros bibliográficos en forma de datos enlazados en la base de datos WorldCat, arrojando un 48 por ciento. Esto justificó la estrategia de enlazar las anotaciones de las fichas con esta base de datos para el posterior completamiento del proceso.<sup>1</sup>

Se estudiaron, e incorporaron técnicas de datos enlazados para diseñar la estrategia de exposición de datos en la WEB y desarrollar una WEB-API para acceder a los datos. Como resultado de la etapa de integración de los datos de los catálogos al espacio de datos enlazados abiertos está disponible una WEB-API Restful con datos RDF serializados en JSON-LD (Lanthaler y Gütl 2012). Se seleccionó el uso del vocabulario BIBFRAME 1.0 propuesto por la Biblioteca del Congreso de los Estados Unidos como vocabulario base para la representación de los datos ("Description of the Category View" n.d.) (ver figs. 4 y 5).

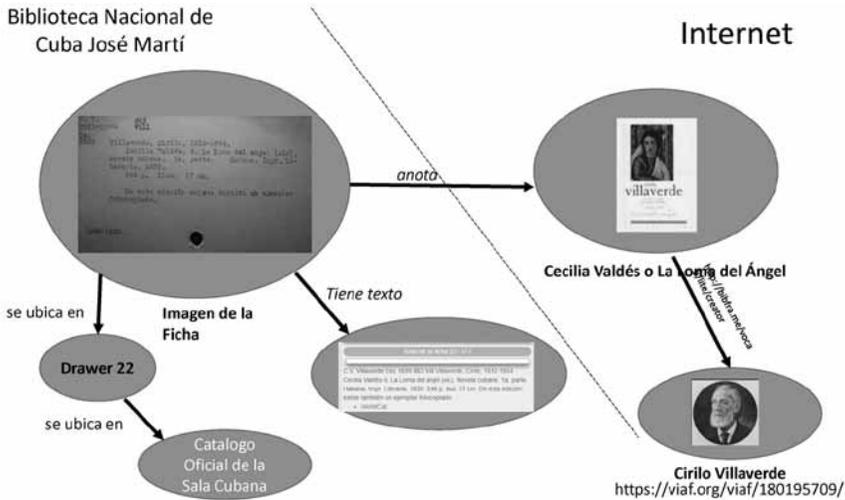


Figura 4. Grafo Informal de los Triples Básicos del Modelo.

Se diseñó un flujo de trabajo a partir de mejores prácticas ajustadas a los propósitos del proyecto, para la digitalización y procesamiento de las fichas y la progresiva incorporación de los textos de las fichas como anotaciones combinando técnicas de OCR (en lotes e interactivo), anotación en red y revisión descentralizada (ver fig. 6). Se integró la interfaz WEB con el uso de aplicaciones de escritorio para realizar el OCR y anotar cada ficha.

Se puso a funcionar un cosechador de datos y motor de búsqueda usando Lucene/SOLR para el procesamiento de los textos de la ficha que se fueran incorporando al espacio de datos del sistema y que permitiera la búsqueda a texto completo en el cuerpo del texto de las fichas procesadas (Grainger y Potter 2014; Smiley et al. 2015). La interfaz de búsqueda y navegación puede accederse en <http://bnjm.sld.cu/>.

#### 4. Ofrecer Productos y Servicios de Beneficio Directo de Usuarios, Bibliotecarios y Otros Especialistas desde Etapas Tempranas del Proceso.

La versión «virtual» de los catálogos están en línea y se exponen en dependencia del avance del proceso de captura. Las anotaciones de los textos se están realizando usando esta plataforma en red. De hecho los especialistas que realizan los procesos de control de calidad trabajan desde sus conexiones a la red. Las personas que realizan la captura, optimización, OCR y anotación de las tres instituciones tienen acceso a los catálogos y pueden visualizar su avance y colaborar técnicamente. Paralelamente al proceso de captura de las imágenes de las fichas se comenzó a exponer en línea, primero en la Intranet y luego en Internet, un portal para acceder a los resultados de la construcción de los catálogos y para su propia cons-

```

1 {
2   "@context": {
3     "@base": "http://bnjm.sld.cu/bnjmsculyfof/",
4     "@vocab": "http://bibframe.org/vocab/"
5   },
6   "@id": "./bnjmsculyfof001/bnjmsculyfof0010001.jsonld",
7   "@type": "Annotation",
8   "annotates": {
9     "@id": "http://worldcat.org/entity/work/id/320445285",
10    "@type": "Work"
11  },
12  "annotationAssertedBy": {
13    "@id": "http://viaf.org/viaf/169775455",
14    "@type": "Agent"
15  },
16  "annotationBody": {
17    "@id": "./bnjmsculyfof001/bnjmsculyfof0010001.txt"
18  },
19  "catalogCardImage": {
20    "@id": "./bnjmsculyfof001/images/bnjmsculyfof0010001.jpg"
21  }
22 }
23

```

Figura 5. Ejemplo de un registro JSON-LD con los datos enlazados de una ficha.

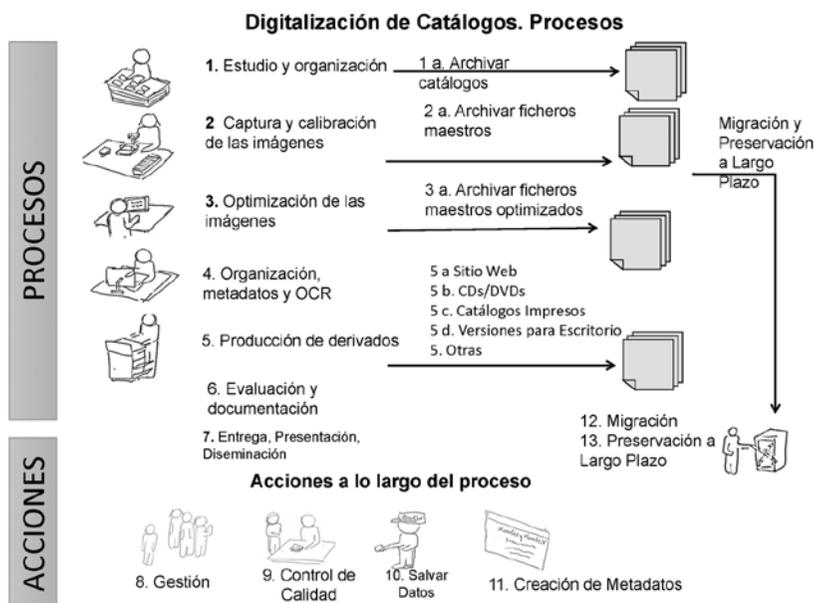


Figura 6. Diagrama del Flujo de Trabajo para la Digitalización y Procesamiento de las Fichas.



gentes para la interoperabilidad internacional de imágenes conocida por sus siglas IIF y el uso de la recomendación del Consorcio WWW de Anotaciones Web ha abierto nuevas posibilidades al proyecto. Se trabaja en una nueva versión que incorpora el conjunto de especificaciones técnicas IIF y que explota intensamente la recomendación para anotaciones citada (Crane 2017; Sanderson 2017; “International Image Interoperability Framework” 2018; Sironi 2017; McAulay 2017; Snyderman, Sanderson y Cramer 2015).

Lo anterior se adoptó como estrategia, dada la complejidad que implica la transformación de los textos libres de las fichas a una estructura del tipo MARC21 (Kroeger 2013). Nótese que para el caso del catálogo oficial de la Sala Cubana, la muestra permitió estimar que el 48 por ciento de las fichas cuenta con un registro en formato RDF en esta base de datos. Este abordaje reduce la cantidad de las fichas cuyos datos tienen que adicionarse manualmente al catálogo local. En el caso del catálogo oficial de la Sala Cubana de la Biblioteca Nacional la reducción es de casi la mitad del total de las fichas. Al propio tiempo se preservan las anotaciones locales para las fichas que se importaron y se abren posibilidades de vínculos con otros recursos siguiendo la ruta de los enlaces proporcionados por una base de datos de alta centralidad como WorldCat. Por esta vía se tiene acceso además a datos complementarios como las autoridades en bases de datos como el Virtual International Authorities File (<http://viaf.org>) y otras. La identificación de las fichas que cuentan con registros en el catálogo WorldCat se realiza combinando técnicas automatizadas con la revisión ha.

Finalmente, como parte de la visión progresiva de tratamiento de los datos, se han desarrollado herramientas complementarias de anotación. Se ha estudiado la experiencia de otros proyectos de datos enlazados y en particular de la plataforma de datos enlazados de Europea, la Digital Public Library of América y la BBC. Una de las herramientas busca «exprimir» datos del texto de las fichas para identificar entidades nombradas y poder representar estos datos como «datos enlazados.» La idea de un «exprimir de datos» en este contexto es una metáfora para referirse a procedimientos para extraer entidades nombradas a partir de textos para convertirlos en datos enlazados. La plataforma de datos enlazados de la BBC constituye un ejemplo muy innovador en esta dirección (ver Shearer 2013). Esto permitiría continuar adicionando valor a los registros y exponer los datos resultantes en la WEBAPI antes mencionada. Para resolver este problema se ha desarrollado una línea de investigación con estudiantes de la facultad de ciencias de computación de la Universidad de la Habana cuyos resultados serán probados y publicados próximamente.

La experiencia de incorporación de las técnicas de datos enlazados a los procesos de desarrollo de repositorios y bibliotecas digitales está funcionando como un proceso de aprendizaje y se presenta como una oportu-

nidad para dar respuesta al problema de la integración de catálogos impresos al espacio digital. La experiencia del proyecto cubano apunta en esa dirección y se necesitarán estudios para evaluar su verdadero impacto a mediano y largo plazo. A corto plazo se ha cambiado la situación de los catálogos impresos de tres importantes bibliotecas cubanas que comenzaron a exponer una parte importante de sus datos bibliográficos en el espacio abierto de la Web y a integrarse progresivamente a la red de datos enlazados abiertos.

## NOTAS

1. Ver, por ejemplo, Catálogo de Libros y Folletos—Oficial, Fichas: 15817, Biblioteca Nacional José Martí, n.d., <http://bnjm.sld.cu/cgi-bin/item.py?idficha=bnjmsculyfof0100175>; y su registro, <http://bnjm.sld.cu/cgi-bin/record.py?idficha=bnjmsculyfof0100175>.

## REFERENCIAS

- Abid, Abelaziz. 1995. "Memory of the World-Preserving the Documentary Heritage." *IFLA Journal* 21 (3): 169–74. <https://doi.org/10.1177/034003529502100302>.
- Alemu, Getaneh, Brett Stevens, Penny Ross, and Jane Chandler. 2012. "Linked Data for Libraries: Benefits of a Conceptual Shift from Library-Specific Record Structures to RDF-Based Data Models." *New Library World* 113 (11/12): 549–70. <https://doi.org/10.1108/03074801211282920>.
- Avram, Henriette D. 1975. *MARC: Its History and Implications*. Washington, DC: Library of Congress. <https://eric.ed.gov/?id=ED127954>.
- Biblioteca Nacional de Cuba "José Martí." 2013. "Proyecto Para la Construcción en Red de la Biblioteca Nacional Digital de Cuba José Martí." [http://www.bnjm.sld.cu/info/acerca/bndc/proyecto\\_bndc\\_version1.0.html](http://www.bnjm.sld.cu/info/acerca/bndc/proyecto_bndc_version1.0.html).
- Chapman, Ann. 1996. "Retrospective Catalogue Conversion: A National Study and a Discussion Based on Selected Literature." *Libri* 46 (1): 16–24.
- Crane, Tom. 2017. "An Introduction to IIIF." Digirati. March 2017. [http://resources.digirati.com/iiif/an-introduction-to-iiif/?utm\\_source=Cultural+Digital&utm\\_campaign=d5d79a3877-culturaldigital074&utm\\_medium=email&utm\\_term=0\\_f5c318bb03-d5d79a3877-140375265](http://resources.digirati.com/iiif/an-introduction-to-iiif/?utm_source=Cultural+Digital&utm_campaign=d5d79a3877-culturaldigital074&utm_medium=email&utm_term=0_f5c318bb03-d5d79a3877-140375265).
- "Description of the Category View of the BIBFRAME Vocabulary." n.d. Washington, D.C.: Library of Congress. Accessed October 12, 2018. <http://www.loc.gov/bibframe/docs/vocab-category.html>.
- Gómez-Acebo, José Luis Bueren, and Elena Sánchez Nogales. 2013. "Biblioteca Digital del Patrimonio Iberoamericano: Open Source Technology in the Service of a Major Cooperative Project." Paper presented at IFLA WLIC, Singapore, August 19, 2013. <http://library.ifla.org/78/>.
- Grainger, Trey, and Timothy Potter. 2014. *Solr in Action*. Shelter Island, NY: Manning Publications and Dream Tech Press.
- Gregory, Lisa, and Stephanie Williams. 2014. "On Being a Hub: Some Details Behind Providing Metadata for the Digital Public Library of America." *D-Lib Magazine* 20 (7/8). <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/july14/gregory/07gregory.html>.
- Harrison, Martin. 1985. "Retrospective Conversion of Card Catalogues into Full MARC Format Using Sophisticated Computer-Controlled Visual Imaging Techniques." *Program* 19 (3): 213–30.
- Hsueh, Daphne. 1992. "Recon Road Maps: Retrospective Conversion Literature, 1980–1990." *Cataloging & Classification Quarterly* 14 (3/4): 5–22. [https://doi.org/10.1300/J104v14n03\\_02](https://doi.org/10.1300/J104v14n03_02).
- International Federation of Library Associations and Institutions. 2015. "IFLA—Best Practice for National Bibliographic Agencies in a Digital Age." <https://www.ifla.org/node/7858>.
- "International Image Interoperability Framework." 2018. Wikipedia. Last modified October 1, 2018, 07:46. [https://en.wikipedia.org/wiki/International\\_Image\\_Interoperability\\_Framework](https://en.wikipedia.org/wiki/International_Image_Interoperability_Framework).

- Kroeger, Angela J. 2013. "The Road to BIBFRAME: The Evolution of the Idea of Bibliographic Transition into a Post-MARC Future." *Cataloging & Classification Quarterly* 51 (8): 873–90.
- . 2015. "A Brief Overview of BIBFRAME." *Criss Library Faculty Proceedings and Presentations* 66. <http://digitalcommons.unomaha.edu/cgi/viewcontent.cgi?article=1069&context=crisslibfacproc>.
- . 2016. "Hello BIBFRAME 2.0: Changes from 1.0 and Possible Directions for the Future." *Criss Library Faculty Proceedings and Presentations* 65. [http://digitalcommons.unomaha.edu/crisslibfacproc/65/?utm\\_source=digitalcommons.unomaha.edu%2Fcrisslibfacproc%2F65&utm\\_medium=PDF&utm\\_campaign=PDFCoverPages](http://digitalcommons.unomaha.edu/crisslibfacproc/65/?utm_source=digitalcommons.unomaha.edu%2Fcrisslibfacproc%2F65&utm_medium=PDF&utm_campaign=PDFCoverPages).
- Lanthaler, Markus, and Christian Gütl. 2012. "On Using JSON-LD to Create Evolvable RESTful Services." In *Proceedings of the Third International Workshop on RESTful Design*, edited by Rosa Alarcon, Cesare Pautasso, and Erik Wilde, 25–32. New York: ACM. <http://dl.acm.org/citation.cfm?id=2307827>.
- Loesch, Martha Fallajay. 2011. "VIAF (The Virtual International Authority File)—<http://viaf.org>." *Technical Services Quarterly* 28 (2): 255–56.
- Malmsten, Martin. 2008. "Making a Library Catalogue Part of the Semantic Web." In *Metadata for Semantic and Social Applications: Proceedings of the International Conference on Dublin Core and Metadata Applications, 22–26 September 2008*, edited by Jane Greenberg and Wolfgang Klas, 146–52. Singapore: Dublin Core Metadata Initiative and Universitätsverlag Göttingen. <http://www.oapen.org/download?type=document&docid=610315>.
- McAulay, Lisa. 2017. "Benefits of the International Image Interoperability Framework (IIIF) Featuring Medieval Palimpsest Manuscripts." Poster presented at the Digital Initiatives Symposium 34, San Diego, February 5, 2017. <http://digital.sandiego.edu/symposium/2017/2017/34>.
- Moulaion, Heather Lea, and Anthony J. Million. 2014. "The Disruptive Qualities of Linked Data in the Library Environment: Analysis and Recommendations." *Cataloging & Classification Quarterly* 52 (4): 367–87. <https://doi.org/10.1080/01639374.2014.880981>.
- Oberhauser, Otto C. 2002. "Card-Image Public Access Catalogues (CIPACs): A Critical Consideration of a Cost-Effective Alternative to Full Retrospective Catalogue Conversion." Master's thesis, University of Central England (Birmingham, UK). <http://eprints.rclis.org/8552/>.
- . 2003. "Card-Image Public Access Catalogues (CIPACs): An International Survey." *Program* 37 (2): 73–84.
- . 2007. "The International CIPAC List im neuen Gewand." *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 60 (1): 71.
- Poole, Nick, Bruno Racine, and Jill Cousins. 2014. "We Transform the World with Culture: Europeana Strategy 2015–2020" (policy report). The Hague: Europeana Foundation. [https://pro.europeana.eu/files/Europeana\\_Professional/Publications/Europeana\\_Strategy\\_2020.pdf](https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana_Strategy_2020.pdf).
- Rieger, Oya Y. 2008. *Preservation in the Age of Large-Scale Digitization: A White Paper*. Washington, DC: Council on Library and Information Resources. <https://www.clir.org/wp-content/uploads/sites/6/2016/09/pub141.pdf>.
- Sanderson, Robert. 2017. "Building Distributed Online Exhibitions with IIIF." Paper presented at MW17: Museums and the Web 2017, Cleveland, April 21, 2017. <http://mw17.mwconf.org/paper/building-distributed-online-exhibitions-with-iiif/>.
- Shearer, Matt. 2013. "BBC News Lab: Linked Data." *BBC News*, January 28, 2013. <http://www.bbc.co.uk/blogs/internet/entries/63841314-c3c6-33d2-a7b8-f58ca040a65b>.
- Sironi, Giorgio. 2017. "Dynamically Serving Scientific Images Using IIIF." Labs (blog). eLife (website). July 13, 2017. <https://elifesciences.org/labs/d6044799/dynamically-serving-scientific-images-using-iiif>.
- Sloggett, Robyn. 2005. "Valuing Significance or Signifying Value? Culture in a Global Context. [The UNESCO Memory of the World Programme]." *Archives and Manuscripts* 33 (2): 110. <https://search.informit.com.au/fullText;res=IELAPA;dn=200606881>.
- Smiley, David, Eric Pugh, Kranti Parisa, and Matt Mitchell. 2015. *Apache Solr Enterprise Search Server*. 3rd ed. Birmingham, UK: Packt.
- Snydman, Stuart, Robert Sanderson, and Tom Cramer. 2015. "The International Image Interoperability Framework (IIIF): A Community and Technology Approach for Web-Based Images." Paper presented at the Society for Imaging Science and Technology Archiving 2015 Conference, Los Angeles, May 20, 2015. <https://pdfs.semanticscholar.org/b27b/109dfa6dc983b3bd2836d206292a1ef8d363.pdf>.

- Ugobono, Mónica. 2011. "Cronología del CDS/Isis." *Boletín Electrónico ABGRA* 3 (3): 1–7. <http://www.abgra.org.ar/newsletter/Historia%20ISIS.pdf>.
- Yang, Sharon Q., and Melissa A. Hofmann. 2010. "The Next Generation Library Catalog: A Comparative Study of the OPACs of Koha, Evergreen, and Voyager." *Information Technology and Libraries* 29 (3): 141.
- Zapounidou, Sofia, Michalis Sfakakis, and Christos Papatheodorou. 2013. "Highlights of Library Data Models in the Era of Linked Open Data." In *Metadata and Semantic Research*, edited by Emmanouel Garoufallou and Jane Greenberg, 396–407. Basel, Switzerland: Springer. [http://link.springer.com/chapter/10.1007/978-3-319-03437-9\\_38](http://link.springer.com/chapter/10.1007/978-3-319-03437-9_38).
- . 2017. "Representing and Integrating Bibliographic Information into the Semantic Web: A Comparison of Four Conceptual Models." *Journal of Information Science* 43 (4): 525–53. <https://doi.org/10.1177/0165551516650410>.

---

Pedro Urra González, MS, es profesor titular de tecnologías de la información de la Universidad de la Habana. Ha dedicado su carrera profesional al uso y la promoción de las tecnologías de la información en bibliotecas y centros de información. Fundador de la Red Cubana de Información en Ciencias de la Salud, Infomed, fue director del Centro Nacional de Información de Ciencias Médicas de Cuba entre los años 2002 y el 2010. Formó parte del equipo de desarrollo de la Biblioteca Virtual de Salud de América Latina y el Caribe y el Campus Virtual de Salud Pública. [urrape@gmail.com](mailto:urrape@gmail.com).