**Capture and Preservation of Born Digital News**

*March 5-6*

Universidad Nacional Autónoma de México, México City

El papel del archivista y bibliotecario en la documentación de eventos políticos y sociales: el caso del movimiento #RickyRenuncia en Puerto Rico

Joel Antonio Blanco-Rivera

Escuela Nacional de Conservación, Restauración y Museografía, Ciudad de México, México.

E-mail address: joel\_blanco\_r@encrym.edu.mx

Copyright © 2020 by Joel Antonio Blanco-Rivera. This work is made available under the terms of the Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0>

Abstract:

El 24 de julio de 2019, y luego de dos semanas de protestas masivas, el gobernador de Puerto Rico Ricardo Rosselló Nevares anunció su renuncia efectiva el 2 de agosto. Ante la importancia histórica de este evento para Puerto Rico, se comenzó en agosto de 2019 el Proyecto de Documentación #RickyRenuncia, una iniciativa colaborativa que busca recopilar, preservar, y dar acceso a documentos, principalmente de origen digital, sobre las protestas del verano del 2019. Un componente del proyecto fue la creación de una colección web sobre las protestas. Esta ponencia se enfoca en el trabajo realizado para el desarrollo de la Colección Web #RickyRenuncia, enfatizando en el componente colaborativo para lograr la selección y catalogación de los recursos. Se discutirán los desafíos relacionados a la captura de noticias digitales que forman parte de la colección, particularmente en formato multimedia, así como el papel de archivistas y bibliotecarios en proyectos de documentación de eventos políticos y sociales. Sobre esto último, se enfatiza en el papel activo que archivistas y bibliotecarios deben jugar para la preservación de contenidos digitales transmitidos tanto en medios de comunicación como en redes sociales.

**Keywords:** web archiving; proyectos colaborativos.

# Introducción

El 13 de julio de 2019, el Centro de Periodismo Investigativo (CPI) publicó 889 páginas de un chat de la aplicación Telegram donde participaron el entonces gobernador Ricardo Rosselló Nevares y sus más cercanos colaboradores (Valentín Ortíz y Minet, 2019). El chat incluye numerosos comentarios sexistas y homofóbicos. Además, se compartía información privilegiada de política pública aún cuando el chat incluía cabilderos, y se comentaba sobre posible acciones contra adversarios políticos. Días antes de la publicación del CPI ya se habían filtrado algunas de esas conversaciones, lo cual causó el inicio de protestas colectivas. Particularmente, el 11 de julio la Colectiva Feminista en Construcción convocó a una protesta en el aeropuerto internacional Luis Muñoz Marín, donde Rosselló llegaría en vuelo desde España. Las protestas se intensificaron con la publicación del chat por el CPI, y se movieron a la Calle Fortaleza del Viejo San Juan, donde ubica la mansión ejecutiva. Estas protestas diarias incluyeron dos marchas masivas, el 17 y 22 de julio, así como diversas formas de protesta. Además, se llevaron a cabo protestas en ciudades de Estados Unidos y otros países. En redes sociales, la etiqueta #RickyRenuncia se convirtió en la expresión principal para exigir la renuncia del gobernador y convocar a protestas. El miércoles 24 de julio, casi a la medianoche, Ricardo Rosselló anunció a través de un video transmitido por Facebook que renunciaba a su puesto de gobernador efectivo el viernes 2 de agosto (Mazzei y Robles, 2019).

Lo que se ha dado a conocer como el "Verano Puertorriqueño" fue un evento único en la historia de Puerto Rico. Por lo tanto, era importante identificar estrategias de documentación para seleccionar, organizar y preservar información valiosa tanto para propósitos de investigación como para construcción de memorias. Debido a que la gran mayoría de la información generada y transmitida era nativa digital, este proceso de documentación requería de una intervención prácticamente inmediata. Por lo tanto, archivistas y bibliotecarios debían asumir un rol activo en la documentación de los eventos alrededor de la renuncia de Rosselló Nevares.

Ante la importancia histórica de este evento para Puerto Rico, se comenzó en agosto de 2019 el Proyecto de Documentación #RickyRenuncia, una iniciativa colaborativa que busca recopilar, preservar, y dar acceso a documentos, principalmente de origen digital, sobre las protestas del verano de 2019. Un componente del proyecto fue la creación de una colección web sobre las protestas, la cual incluye la selección, preservación y catalogación de noticias digitales. Esta ponencia se enfoca en el trabajo realizado para el desarrollo de la Colección Web #RickyRenuncia, enfatizando en el componente colaborativo para lograr la selección y catalogación de los recursos. Se discutirá el papel de archivistas y bibliotecarios en proyectos de documentación de eventos políticos y sociales. Se enfatiza en el papel activo que archivistas y bibliotecarios deben jugar para la preservación de contenidos digitales transmitidos tanto en medios de comunicación como en redes sociales.

**El Proyecto de Documentación #RickyRenuncia**

En los pasados años archivistas y bibliotecarios han asumido roles más activos en la documentación de eventos políticos y sociales significativos. Este rol activo incorpora la colaboración no solo multidisciplinaria, pero también a nivel de las comunidades impactadas por los sucesos documentados (Peaker and DeCaro, 2014; Atiso and Freeland, 2016). Entre las razones para desarrollar este tipo de proyectos está el reconocimiento de que archivos tradicionales no necesariamente proveen una mayor representación de la sociedad, y que las creación de archivos especializados en ciertas comunidades es parte de los intentos políticos para escuchar sus voces (ver Flinn, Stevens, and Shepherd 2009; Caswell, Cifor, and Ramirez 2016). Además, el hecho de que la mayor parte de la información sobre estos tipos de eventos es generada y transmitida a través de redes de comunicación digital requiere una intervención más inmediata de parte de archivistas y bibliotecarios para lograr la selección, organización y conservación de estos contenidos. En el caso del proyecto #RickyRenuncia, las conversaciones y las intervenciones comenzaron durante el periodo que sucedían las protestas.

El objetivo principal del Proyecto de Documentación #RickyRenuncia es adquirir, preservar y dar acceso a contenidos digitales relacionados a las protestas exigiendo la renuncia del gobernador de Puerto Rico Ricardo Rosselló Nevares, cubriendo principalmente el periodo entre el 13 de julio, día que se publica el chat del gobernador y sus ayudantes cercanos, y el 24 de julio, día en que Rosselló anuncia su renuncia. El proyecto tiene tres componentes principales: un dataset de sobre un millón de tuits recopilados bajo las etiquetas #RickyRenuncia y #RickyVeteYa, la Colección Web #RickyRenuncia, y la planificación de un repositorio digital conformado por materiales digitales donados por personas que deseen compartir sus experiencias y testimonios sobre las protestas. Nuestro plan es utilizar la plataforma Scalar como el espacio virtual central para proveer acceso a los materiales. Scalar es una plataforma de código abierto desarrollada por el *Alliance for Networking Visual Culture* que permite publicar proyectos interactivos que utilizan recursos de múltiples fuentes y en múltiples formatos. De los tres componentes enfocaré el resto de la ponencia en la colección web. Pero antes algunas palabras sobre los otros dos componentes. El dataset de tuits fue recopilado por este servidor utilizando la herramienta Twarc, diseñada por Ed Summers, miembro de Documenting the Now.[[1]](#footnote-1) Cubre el periodo del 10 al 25 de julio de 2019, y contiene poco más de 1.1 millones de tuits. En este caso el dataset de los tuits no estará accesible por dos razones. Primero, las políticas de Twitter no permiten el compartir los datos. Segundo, y más importante, existen consideraciones éticas sobre la publicación de una colección creada con tuits capturados sin solicitar permiso. Hemos incorporado el modelo de Documenting the Now, donde se publican los identificadores de los tuits, el cual un usuario puede convertir en el dataset, pero solo se recuperan los tuits que permanezcan públicos (https://www.docnow.io/catalog). El repositorio digital continúa en proceso de planificación y es la prioridad en el plan de trabajo del proyecto para el año 2020.

**La Colección Web #RickyRenuncia**

Ed Summers (2019) explica que mientras la web está insertada en el día a día de los sistemas econónimos, de gobernanza, ciencia y cultura, su "precariedad arquitectónica" también es visible a diario. Esta pérdida de información genera, expone Summers, una "ansiedad archivística" que cuestiona "¿qué será recordado de nuestra época reciente?" (p. 2). En este contexto surge la práctica de *web archiving*, la cual establece procesos de gestión con intervenciones tanto de recursos humanos como de las tecnologías para la preservación y acceso de recursos de la web. El ciclo de vida del *web archiving*, elaborado por el Internet Archive, refleja estas intervenciones y el valor de la planificación y documentación de procesos (Bragg and Hanna, 2013).

En parte, esta ansiedad archivística ha sido una de las razones que nos a guiado a incorporar al proyecto de documentación la selección, catalogación y preservación de contenidos de la web sobre las protestas y la renuncia del gobernador Rosselló Nevares. Reconociendo la imposibilidad de preservarlo todo e incorporando la idea de que los archivos ofrecen solo una porción del universo de documentos generados (Harris 2002), buscamos capturar una representación del universo de contenidos web sobre estos eventos. El enfoque no eran solo las noticias, aunque sí son un componente significativo de la colección, sino además opiniones, cómo se estaba reportando sobre los eventos en el exterior, entradas de blogs y en redes sociales, entre otros. La idea de la Colección Web #RickyRenuncia se comenzó a trabajar luego de tener conocimiento de la iniciativa Spontaneous Events Collection del Internet Archive. Miembros del equipo de Archive-It, la plataforma que permite a instituciones crear colecciones web, colaboran con grupos para crear una colección web de un evento en particular. El 21 y 22 de julio de 2019 enviamos correos al Internet Archive proponiendo la creación de la Colección Web #RickyRenuncia. El 23 de julio recibimos la respuesta aceptando la propuesta.

En esta colaboración, el grupo de trabajo de Archive-It es responsable de todo lo relacionado a la administración de la plataforma: crear el espacio de la colección, realizar el "crawling" de los enlaces seleccionados, e importar los metadatos. Nuestra responsabilidad se enfoca en la selección de los recursos y la creación de metadatos. La selección final fue de 418 recursos, en español e inglés, disponibles en https://archive-it.org/collections/12491. La creación de los metadatos es un componente fundamental, ya que facilita navegar a través de la página y la búsqueda/recuperación de recursos en la plataforma. Tener solo una lista de enlaces en Archive-It no es la mejor estrategia, e inclusive se podría argumentar que, desde una perspectiva archivística y/o bibliotecológica, no sería una colección. La organización de la información es esencial, y por lo tanto los metadatos son indispensables.

Archive-It permite el importar los metadatos a través de una hoja de cálculo. Aún teniendo acceso a la plataforma, esta es la estrategia más adecuada en comparación con entrar los datos directamente a la forma de Archive-It. Una razón es que permite mantener copias de los metadatos sin depender totalmente de la plataforma; otra que facilita la colaboración sin que todos los participantes necesiten acceso a Archive-It. La estrategia de colaboración y "crowdsourcing" fue la que implementamos. El 30 de julio de 2019 publicamos en la página de Facebook del proyecto una convocatoria de voluntarios. Recibimos respuesta de trece personas interesadas en colaborar. A cada persona se le creó una hoja de cálculo con los quince elementos de Dublin Core, el esquema que utiliza Archive-It para los metadatos, y se le asignaron veinte enlaces. Una vez finalizados los metadatos, los mismos son revisados e incorporados a la hoja de cálculo que se envía al equipo Archive-It para importar. Hasta enero del 2020 hemos logrado catalogar sobre cien de los enlaces.

****Una de las características que ha convertido a Dublin Core como uno de los esquemas de metadatos más utilizados en proyectos digitales es la flexibilidad que puede proveer los quince elementos principales. Sin embargo, esto se puede convertir en un problema si no se definen claramente los elementos en el contexto del proyecto. En nuestro caso, donde se lleva a cabo un proceso de crowdsourcing en la catalogación de los recursos, utilizamos una guía de metadatos, la cual especifica la definición de cada elemento, indicaciones sobre su uso (incluyendo si es obligatorio u opcional), y ejemplos (ver Figura 1). Por ejemplo, para el elemento Subject se tomó la decisión de utilizarlo para organizar los contenidos en categorías generales (ej. Gobierno, Movimientos de protesta), en vez de utilizar palabras claves o autoridades como el Library of Congress Subject Headings. Se tomó en consideración el volumen de la colección y las funcionalidades de Archive-It en relación a navegación, búsqueda y recuperación. Aunque la interface tiene un motor de búsqueda, no provee opciones de búsqueda por elemento. La interface permite navegar por elementos (ver Figura 2). Por lo tanto, puede ser útil para el usuario escoger de una lista concisa de categorías que agrupa los recursos según el tema principal. El proceso de crowdsourcing nos ayudó además a identificar aspectos de la guía que requiriesen revisión, y por lo tanto se invitó a los participantes a que compartieran sus observaciones y/o dudas que surgieran durante el proceso de catalogación. Recibimos observaciones, por ejemplo, sobre el elemento Subject, ya que se identificaron recursos que no tenían como tema principal una de las categorías establecidas.

Figura 1. Definición de elemento en la guía de metadatos.

En términos generales la estrategia de crowdsourcing ha sido positiva. Recibimos una buena respuesta de voluntarios que se integraron al proceso de catalogación, y la guía de metadatos ha funcionado, permitiendo además la participación de personas con distintos niveles de conocimiento en metadatos y Dublin Core. Sin embargo, aún no se ha logrado catalogar casi tres cuartas partes de los recursos. Es importante dar seguimiento a la catalogación de recursos asignados a los participantes, mientras se reconoce que es un trabajo que se va realizando de manera voluntaria dentro de los tiempos disponibles. Se podrían considerar otras estrategias, como por ejemplo sesiones virtuales para realizar catalogación en conjunto. Se puede además integrar la catalogación a espacios educativos, desarrollando actividades didácticas con estudiantes de programas de ciencias de la información.

 En términos técnicos uno de los desafíos principales de la gestión de colecciones web es las limitaciones que plataformas como Archive-It tienen para capturar recursos multimedios, y particularmente videos. Esto a su vez crea un limitante en el caso de recursos de medios noticiosos que publican video-reportajes. Esto se ve reflejado en la colección web #RickyRenuncia, donde se seleccionaron e incorporaron video-reportajes, pero la versión archivada no logró mantener la funcionalidad de la página y no se puede ver el video. Esto pone en perspectiva la importancia de considerar aspectos técnicos del *web archiving* al momento de establecer criterios de selección, y que estos criterios se documenten y se publiquen como parte de la colección web para que los usuarios estén informados sobre lo que no forma parte de la colección y por qué. Este ejercicio de transparencia es la propuesta presentada por Emily Maemura, Nicholas Worby, Ian Milligan y Christoph Becker (2018), la cual plantean bajo el concepto "web archives provenance". Los autores utilizan este concepto "para abarcar lo que usuarios deben conocer acerca de cómo una colección fue creada" (p. 1225). En el caso de la colección #RickyRenuncia esta información no está incluida en su espacio en Archive-It. Sin embargo, se está elaborando un manual de políticas y procedimientos que documentará estos procesos.

Figura 2. Página de la Colección Web #RickyRenuncia.

**Conclusión**

La literatura sobre planificación e implementación de proyectos de documentación colaborativos expone los desafíos y oportunidades relacionadas a la colaboración multidisciplinaria, y principalmente desde la perspectiva de proyectos con apoyos institucionales. En otras palabras, los proyectos se desarrollan dentro de un contexto donde el trabajo forma parte de las responsabilidades laborales de al menos los coordinadores. Sin embargo, ¿qué sucede cuando se busca desarrollar un proyecto de documentación que surge principalmente de un interés particular, independiente de apoyo institucional? El Proyecto de Documentación #RickyRenuncia es un ejemplo de este tipo de iniciativas. Aunque hemos logrado establecer estrategias de planificación y colaboración exitosas, la pregunta relacionada a la sustentabilidad se mantiene. Lo que hemos logrado es identificar oportunidades de colaboración para cada componente del proyecto, como es el caso de la colección web, y el uso de plataformas de código abierto como Scalar para crear el espacio virtual central para guiar al usuario a las diferentes partes del proyecto.

En relación a la Colección Web #RickyRenuncia, el apoyo del Internet Archive y su iniciativa Spontaneous Event Collections fue fundamental debido a que nuestras instituciones no cuentan con el servicio de Archive-It. La estrategia de crowdsourcing también ha generado resultados positivos, y esperamos continuar el trabajo hasta finalizar la catalogación de los 418 recursos. Será importante estudiar los recursos que forman parte la colección para realizar una evaluación que nos ayude a identificar lagunas en la colección y replantear la estrategia de selección. Finalmente, será importante identificar estrategias de difusión para cumplir con uno de los objetivos del proyecto, que es proveer acceso a documentación sobre un evento que se ha insertado en la memoria colectiva de los puertorriqueños.

References

Atiso, K., & Freeland, C. (2016). Identifying the social and technical barriers affecting engagement in online community archives: A preliminary study of "Documenting Ferguson" archive. *Library Philosophy and Practice*, 1.

Bragg, M. & Hanna, K. (2013). The web archiving life cycle model. Recuperado de http://ait.blog.archive.org/files/2014/04/archiveit\_life\_cycle\_model.pdf.

Caswell, M., Cifor, M., & Ramirez, M. H. (2016). “To suddenly discover yourself existing”: uncovering the impact of community archives. *The American Archivist*, *79*(1), 56-81.

Flinn, A., Stevens, M., & Shepherd, E. (2009). Whose memories, whose archives? Independent community archives, autonomy and the mainstream. *Archival science*, *9*(1-2), 71.

Harris, V. (2002). The archival sliver: power, memory, and archives in South Africa. *Archival Science*, *2*(1-2), 63-86.

Maemura, E., Worby, N., Milligan, I., & Becker, C. (2018). If these crawls could talk: Studying and documenting web archives provenance. *Journal of the Association for Information Science and Technology*, *69*(10), 1223–1233. https://doi.org/10.1002/asi.24048

Mazzei, P. y Robles, F. (2019, 24 de julio). Ricardo Rosselló anunció su renuncia a la gobernación de Puerto Rico. *The New York Times*. Recuperado de https://www.nytimes.com/es/2019/07/24/espanol/america-latina/rossello-renuncia-puerto-rico.html

Peaker, A., & DeCaro, J. (2014). *Building communities of collaborators at Our Marathon: The Boston Bombing Digital Archive*. Bucknell University Digital Scholarship Conferenc, Lewisburg, Pennsylvania. Recuperado de https://digitalcommons.bucknell.edu/digital-scholarship-conference/budsc2014/presentations/2/

Summers, E. (2019, 15 de marzo). Appraisal practices in web archives. https://doi.org/10.31235/osf.io/75mjp

Valentín Ortíz, L. J. y Minet, C. (2019, 13 de julio). Las 889 páginas de Telegram entre Rosselló Nevares y sus allegados. *Centro de Periodismo Investigativo*. Recuperado de http://periodismoinvestigativo.com/2019/07/las-889-paginas-de-telegram-entre-rossello-nevares-y-sus-allegados/

1. Para un estudio de caso que explica el uso de Twarc ver Blanco-Rivera, Joel A. "Curaduría digital y la preservación de contenidos web: creando una colección de tuits sobre la huelga de la Universidad de Puerto Rico," Encuentro Latinoamericano de Bibliotecarios, Achivistas y Museólogos," 9 de octubre de 2017, Ciudad de México. https://www.institutomora.edu.mx/EBAM/2017/Ponencias/Curaduria%20digital%20y%20la%20preservacion%20de%20contenidos%20web%20creando%20una%20coleccion%20de%20tuits%20sobre.pdf. [↑](#footnote-ref-1)