

MEJORAS EN EL REPOSITORIO CIC DIGITAL. GESTIÓN DE DATOS PRIMARIOS Y APOYO A LOS SISTEMAS DE EVALUACIÓN



Centro de Servicios en Gestión de Información (CESGI)

Expositor: Dr. VILLARREAL, Gonzalo L.

Grupo de Trabajo: M. R. De Giusti, A. J. Lira, C. J. Nusch, M. M. Vila

<http://cesgi.cic.gba.gov.ar>

gonzalo@prebi.unlp.edu.ar

RESUMEN

Este trabajo presenta dos líneas de trabajo actuales del Centro de Servicios en Gestión de Información (CESGI) sobre el Repositorio Institucional CIC Digital. Una de las líneas de trabajo se relaciona con el tratamiento y visualización de datos primarios de investigación dentro del Repositorio Institucional. La segunda línea de trabajo se vincula con la incorporación de facilidades a dicho Repositorio para servir de apoyo a los sistemas de evaluación de las distintas convocatorias realizadas desde la CIC, así como también, para las presentaciones de informes de becas y de carrera. Para ambos proyectos, se presentan sus requerimientos mínimos, los desafíos que presentan y las ventajas que aportarán a la institución y a la comunidad de usuarios del Repositorio, tanto del ámbito académico científico como del público en general.

PRESENTACIÓN

El Repositorio CIC Digital tiene como objetivo organizar, preservar y difundir la producción científica realizada desde centros propios y asociados, como también por parte de los investigadores y becarios de la CIC que realizan sus actividades en cualquier ámbito de la comunidad académica y científica. Para realizar estas tareas, el CIC Digital brinda una amplia gama de servicios digitales para toda la comunidad de usuarios, como ser la integración de la CIC al Sistema Nacional de Repositorios Digitales (SNRD), el seguimiento del impacto y la visibilidad web de los recursos que aloja, el monitoreo de formatos y migración de objetos digitales a fin de asegurar su acceso a largo plazo, y el establecimiento de servicios de interoperabilidad con portales web institucionales que, junto al proyecto de desarrollo de portales web para Centros CIC, permite dar mayor difusión a toda la producción intelectual de los mismos y ahorrar costos asociados, tanto a la implementación y mantenimiento de estos portales, como a la actualización de los contenidos que se exponen en línea.

Al día de hoy, el repositorio aloja más de 6.500 objetos digitales, entre los que se incluyen: artículos en revistas, trabajos en eventos científicos y académicos, tesis, informes, libros y partes de libros. Este tipo de recursos pueden ser descargados y utilizados por cualquier usuario, ya que se trata esencialmente de documentos de texto: ocupan poco espacio de almacenamiento, pueden descargarse

en cuestión de segundos y no requieren el uso de herramientas de software avanzadas para su visualización. Sin embargo, este repositorio permite también el depósito de datos primarios, lo que requiere algunas consideraciones especiales para su exposición y tratamiento. Una de las características que incorporará el Repositorio de la CIC será la de permitir la visualización de datos primarios, tanto crudos como procesados, sin necesidad de descargarlos ni de contar con herramientas específicas para utilizarlos. Estos datos pueden ser mediciones, datos georeferenciados, encuestas, bases de datos, imágenes entre otros. El origen de los datos puede ser muy diverso, generado tanto automáticamente por dispositivos electrónicos (por ejemplo sensores dispuestos en determinados puntos geográficos) como también compilados manualmente en el marco de proyectos de estudio e investigación, y puede tratarse de datos crudos, o sea que son almacenados tal y como fueron obtenidos, como también de datos procesados, datos que recibieron algún tipo de tratamiento: normalización, limpieza, filtrado, agrupamiento, etcétera.

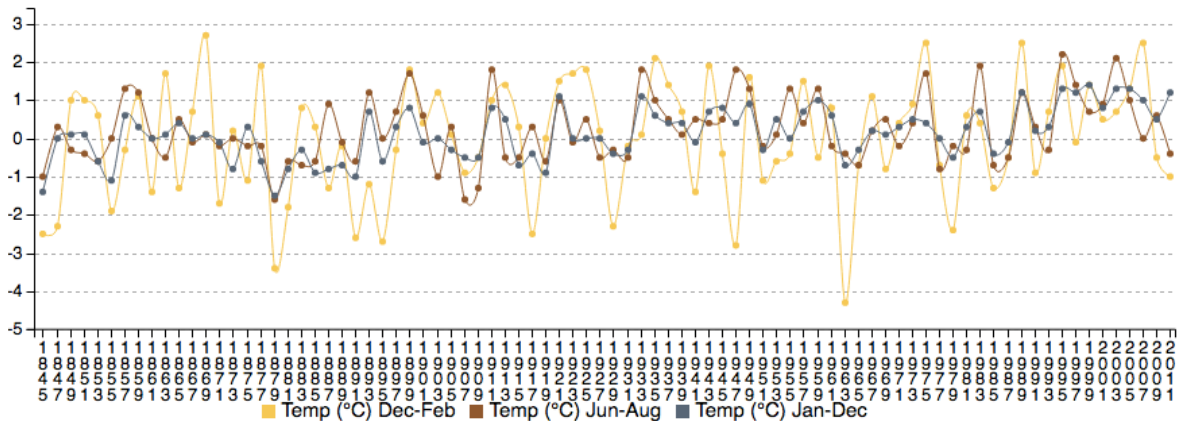
Además de constituir un servicio para preservar y difundir la producción intelectual de una organización, los repositorios institucionales pueden servir como sistema de apoyo a las actividades de gestión y evaluación que se realizan desde la Institución. Un caso concreto se da a la hora de reunir la producción de investigadores y centros en el marco de convocatorias y/o llamados a subsidios. Las distintas Comisiones Evaluadoras, así como también la Junta de Calificaciones y el Directorio de la CIC, requieren tener acceso a todas las publicaciones y trabajos generados por investigadores, profesionales de apoyo y becarios de la CIC. Sin embargo, algunos de estos materiales fueron publicados en determinados contextos o bajo licencias que no permiten su difusión abierta como el resto de los materiales, sino que requieren establecer nuevos circuitos para la recepción, organización y acceso controlado.

METODOLOGÍA Y RESULTADOS

La adición de los datos primarios como nuevo tipo de objeto digital en el repositorio implica el establecimiento de nuevas formas de organizarlos y preservarlos, lo que a su vez requiere estudiar e implementar esquemas de metadatos particulares para cada caso, nuevas formas de agrupar y explorar los recursos, sistemas de validación y migración de formatos específicos, entre otros. La recepción sistemática y el almacenamiento de estos datos también presenta un desafío adicional, dado el gran volumen de información que pueden representar. El repositorio debe ser capaz de lidiar con archivos de gran tamaño, almacenarlos y recuperarlos, y permitir su visualización y descarga en tiempos aceptables.

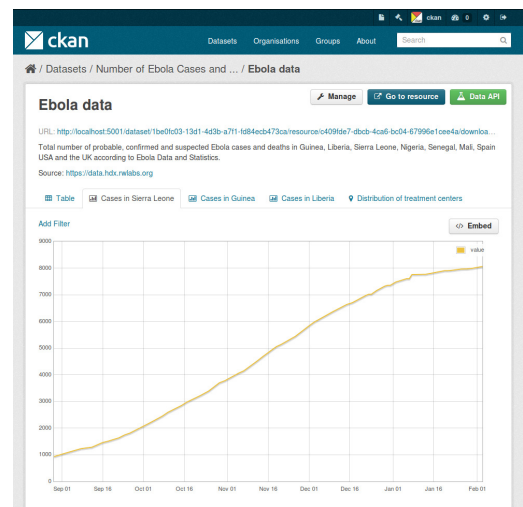
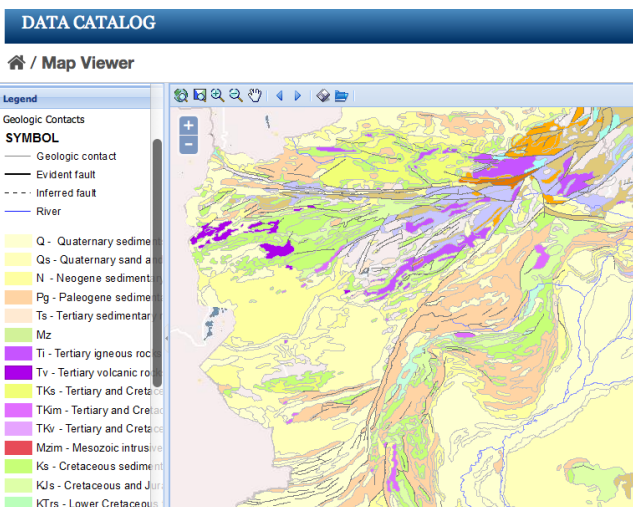
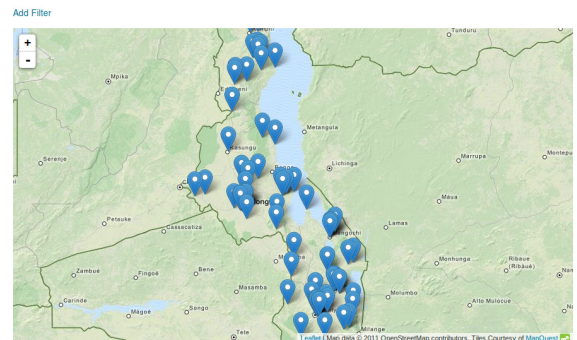
Por otro lado, se deben incorporar nuevos mecanismos para acceder a esta información, que también se ajustará al tipo de recursos en cuestión: visualizaciones interactivas, generación de mapas, galerías de imágenes, datos tabulares, etcétera. Esto requerirá también la implementación de sistemas de búsqueda de registros, selección de subconjuntos de datos, filtrado y descarga de paquetes de datos completos o parciales, entre otros requerimientos propios de un sistema de preservación de datos primarios de investigación. Si bien existen muchas herramientas informáticas y servicios de terceros que permiten visualizar y explorar determinados tipos de datos, el desafío aquí es integrar este amplio abanico de opciones bajo un mismo paraguas capaz de decidir, para cada conjunto de datos, cuál es la forma más apropiada para exponerlo a los usuarios y qué herramientas brindará para explorarlos, filtrarlos y descargarlos. Por ejemplo, si se trata de datos georeferenciados (como puede ser un archivo KML), se deberán seleccionar servicios de mapas interactivos, y aportar funciones para mostrar u ocultar capas (*layers*), mostrar información en determinados puntos, dibujar polígonos, etc; si se trata de datos

tabulares (por ejemplo un archivo CSV), la visualización deberá hacerse en forma de tabla, mostrando probablemente una parte de los datos e incorporando funciones para ordenar a partir de las columnas de la tabla, paginar resultados, realizar búsquedas y aplicar filtros. La capacidad de optar entre los distintos sistemas de visualización y acceso a cada conjunto de datos estará fuertemente vinculada con el nivel de detalle con que se describen los conjuntos de datos, lo que a su vez dependerá de las posibilidades que brindan los sistemas de catalogación y esquemas de metadatos a incorporar.



117 records

Ward	Alder...	Name	Address	Comments	Handicap ...	Latitu...	Longi...
106	1	Blackhawk Church	9620 Brader Way Madi...	Meeting Room on...	TRUE	43.05...	-89.55
42	6	Wil-Mar Neighborhood Center	953 Jenifer St Madison...	Yahara Room	TRUE	43.07...	-89.36
111	9	The Jefferson	9401 Old Sauk Rd Mad...	Activity Room	TRUE	43.07...	-89.55
9	3	East Police District	809 Thompson Dr Madi...	Community room	TRUE	43.10...	-89.29
59	8	Sept - May: Holt Commons, ...	800 Langdon St Madiso...		TRUE	43.07...	-89.39
58	8	UW Memorial Union	800 Langdon St Madiso...	Paul Bunyan Roo...	TRUE	43.07...	-89.39
107	9	Coventry Village	7707 Brookline Dr Mad...	Community Room	TRUE	43.04...	-89.51
105	1	Coventry Village	7707 Brookline Dr Mad...	Community Room	TRUE	43.04...	-89.51
109	9	High Point Church	7702 Old Sauk Rd Mad...	Micah Center - Pa...	TRUE	43.07...	-89.51
67	13	Brittingham Apartments	755 Braxton Pl Madiso...	Library	TRUE	43.06...	-89.39
88	19	Alicia Ashman Branch Library	733 High Point Rd Mad...	Meeting Room	TRUE	43.07...	-89.51
57	8	UW Memorial Library	728 State St Madison, ...	Room 116	TRUE	43.07...	-89.39
56	8	Sept - May: Gordon Dining, ...	728 State St Madison, ...		TRUE	43.07...	-89.39
117	19	Madison Ice Arena	725 Forward Drive Mad...		TRUE	43.04...	-89.48
104	1	Madison Ice Arena	725 Forward Dr Madiso...	Conference Room	TRUE	43.04...	-89.48
90	19	Madison Ice Arena	725 Forward Dr Madiso...	Conference Room	TRUE	43.04...	-89.48
65	13	Wingra School	718 Gilmore St Madiso...	First Floor Comm...	TRUE	43.05...	-89.43
49	4	Madison Fresh Market	703 University Avenue ...	Event Center on s...	TRUE	43.07...	-89.39
15	15	LaFollette High School	700 Pflaum Rd Madison...	Gym	TRUE	43.06...	-89.31
14	15	LaFollette High School	700 Pflaum Rd Madison...	Gym	TRUE	43.06...	-89.31



Ejemplos de visualizaciones de datos primarios dentro de repositorios de datos en forma de gráficos, tablas y mapas

En lo relativo al uso del Repositorio Institucional como sistema de apoyo a la evaluación, se están realizando modificaciones en CIC Digital, en particular en los módulos de gestión de usuarios y de control de acceso a los documentos. El objetivo aquí es permitir sólo a determinados usuarios especiales dentro del Repositorio, acceder a estos recursos, tanto a los metadatos como a los objetos digitales. Para ello, se ampliaron las capacidades del módulo de autorizaciones del software *Dspace* (encargado de la gestión de roles y permisos) utilizado por el Repositorio. Dado que esta herramienta no considera este tipo de usuarios, se implementaron las funciones que permiten a este módulo dar soporte al nuevo rol de usuarios dentro del repositorio, con permisos para visualizar y descargar los documentos que forman parte del proceso de evaluación, más allá de su estado abierto o embargado. Los cambios sobre el software *Dspace* involucraron también al módulo *Workflow*, responsable de la gestión y seguimiento de los distintos flujos de trabajo dentro del Repositorio, y se establecieron nuevos flujos de trabajo que se aplicarán sobre los documentos, tanto al momento del autoarchivo como durante la recepción por parte de los administradores del repositorio y el acceso para su evaluación por parte de los nuevos usuarios especiales. Estos cambios incluyen, por ejemplo, la generación de colecciones especiales para el depósito de estos materiales, la adición de metadatos descriptivos que permitan a las comisiones evaluadoras identificar unívocamente cada recurso, y la incorporación de mecanismos de búsqueda interna para agilizar el acceso a múltiples recursos relacionados (por ejemplo, publicaciones de un investigador en un rango de fechas).

CONCLUSIONES

Los repositorios institucionales brindan un conjunto de servicios para su comunidad de usuarios, y estos servicios se ajustan a las características y requerimientos que la comunidad de usuarios demanda. Un Repositorio Institucional debe ser capaz de evolucionar y adaptarse a los cambios en el texto, a medida que la institución a la que pertenece crece y cada vez que la comunidad se amplía o demanda nuevos servicios. Los proyectos presentados en este trabajo exponen algunas líneas de trabajo sobre del Repositorio Institucional CIC Digital que buscan adaptar el Repositorio a los tiempos que corren mediante la incorporación de nuevos servicios internos a la organización como para la comunidad de usuarios investigadores y el público en general. Las modificaciones que se han realizado o se realizarán en el corto plazo sobre el Repositorio CIC Digital, permitirán por un lado, asegurar la disponibilidad y el acceso a los recursos financiados con fondos públicos como se establece en la Ley Nacional de Acceso Abierto (Ley 26899), en particular a los datos obtenidos en proyectos de I+D, con el objetivo de transparentar el trabajo realizado por las instituciones que componen el Sistema Nacional de Ciencia, Tecnología e Innovación (SNCTI), así como también, para promover su reuso en diferentes proyectos y múltiples disciplinas. La disponibilidad de datos primarios en repositorio permitirá repetir experimentos, continuar y extender líneas de investigación, contrastar resultados y visualizar informes por parte de cualquier usuario que ingrese al portal web del repositorio, sin la necesidad de instalar ni conocer herramientas informáticas específicas para analizar y procesar estos datos.

Por otro lado, el depósito de publicaciones y todo tipo de materiales utilizados para las evaluaciones que se realizan desde la CIC, permiten no sólo centralizar y agilizar el acceso a estos documentos durante las distintas convocatorias de la CIC, sino también asegurar la preservación y el acceso a largo plazo de tales materiales. Esto servirá también para darle mayor visibilidad al repositorio en el ámbito académico y permitirá ampliar el alcance de los recursos generados por la comunidad académica.