
DESARROLLO DE LA RED BIOCLIMÁTICA DEL ESTADO MÉRIDA, VENEZUELA: ESTRATEGIAS DE CAPTURA, MANEJO Y PRESERVACIÓN DE DATOS AMBIENTALES

HILDA YELITZA CONTRERAS, ZULAY MÉNDEZ,
RODRIGO TORRENS y LUIS A. NÚÑEZ

RESUMEN

Se presenta la experiencia en la construcción y desarrollo de la Red Bioclimática del Estado Mérida (Venezuela) como prueba de concepto en la utilización de las técnicas y estrategias para la captura, manejo y preservación de datos ambientales provenientes de la e-investigación. Se describen los elementos éticos, conceptuales y técnicos para la preservación del patrimonio de datos, haciendo énfasis en los conceptos datos libres (open data),

metadatos y sus estándares. Se presenta el testimonio de la construcción del repositorio de datos bioclimáticos www.cecalc.ula.ve/redbc/, sus productos de información asociados y algunos resultados e impactos preliminares de su utilización. Se finaliza con algunas reflexiones y recomendaciones las cuales, quizá, permitan hacer sustentable y replicable esta iniciativa.

Una Nueva Manera de Producir, Diseminar y Preservar Datos

Cada vez con mayor frecuencia y efectividad los investigadores intercambian datos, ideas, publicaciones, referencias y artículos. Si bien una serie de emergentes y novedosos ambientes de colaboración electrónica no terminan de ser utilizados como herramientas cotidianas para estas interacciones (Borgman, 2006; Coles *et al.*, 2006; Collins *et al.*, 2007; De Roure y Frey, 2007) el correo electrónico, la mensajería instantánea y, sobre todo, la red, se erigen como los apoyos y motores a éstas nuevas formas de colaboración ubicua. Pero más allá de este intercambio entre pa-

res, se está transitando la era postgutemberg, en la cual los productores de información (investigadores, centros de investigación y/o instituciones académicas) tienen la capacidad de publicar y difundir directamente su producción intelectual, sin intermediarios editoriales y a costos cada vez menores.

Los términos ciberinfraestructura, e-ciencia y e-investigación, han sido acuñados para describir las nuevas formas de producción y diseminación del conocimiento, donde el uso intensivo de las tecnologías de información y comunicación (TIC), la distribución geográfica de los recursos de medición, procesamiento y análisis, pero sobre todo su acceso ubicuo, son sus características más resaltantes y descriptivas (ver Hey y Trefethen, 2003b, 2005;

Foster, 2005; y referencias allí citadas). Uno de los retos que habrá de enfrentarse en esta nueva manera de hacer investigación es manejar, administrar, analizar y preservar un “diluvio de datos” (Hey y Trefethen, 2003a) generado por una red de sensores a escala mundial y experimentos de grandes dimensiones (aceleradores de partículas, red de observatorios terrestres y satelitales, enormes bases de datos genéticas, por mencionar las más impactantes). Este alud de mediciones, desbordando toda capacidad para su manejo que no sea mediante las TIC, convierte a estos instrumentos en herramientas informáticas y a la experimentación en minería de datos.

Los grandes productores de datos son colaboraciones mundiales,

PALABRAS CLAVE / e-ciencia / e-investigación / Movimiento de Datos Libres / Repositorio de Datos /

Recibido: 16/05/2008. Modificado: 06/09/2008. Aceptado: 12/09/2008.

Hilda Yelitza Contreras. Licenciada en Computación, Universidad Central de Venezuela, (UCV), Venezuela. Maestría en Ciencias de la Computación, Universidad de Los Andes (ULA), Venezuela. Profesora, Facultad de Ingeniería y Centro de Simulación y Modelos (CESIMO), ULA, Venezuela. e-mail: hyelitza@ula.ve

Zulay Méndez. Geógrafa, ULA, Venezuela. Personal Asociado I, Corporación Parque Tecnológico de Mérida (CPTM), Venezuela. e-mail: zulay@ula.ve

Rodrigo Torrens. Ingeniero de Sistemas y Maestría en Ciencias de la Computación, ULA, Venezuela. Coordinador y Arquitecto de información de proyectos TIC del Centro de Teleinformación CTI-ULA, CPTM, Venezuela. e-mail: torrens@ula.ve

Luis A. Núñez. Licenciado en Física, Universidad Simón Bolívar, Venezuela. Doctor en Ciencias, UCV, Venezuela. Profesor. Facultad de Ciencias y Centro Nacional de Cálculo Científico (CeCALCULA-CPTM), ULA, Venezuela. Dirección: Corporación Parque Tecnológico de Mérida, Mérida 5101, Venezuela. e-mail: nunez@ula.ve

industriales y multinacionales que generan ingentes volúmenes de datos que deben ser distribuidos y mantenidos por los proyectos mientras duren. Muchos de esos datos nunca aparecerán publicados y cuando finalice la colaboración, muchas de las medidas experimentales se perderán o serán enviadas a reservorios (nacionales o internacionales) que nada tuvieron que ver con su producción. Gran parte de las decisiones y criterios para generarlos quedarán sepultados en una inmensa correspondencia electrónica de la que nadie dispondrá (Gray y Szalay, 2002). Igual suerte correrán los datos producidos por multitud de pequeños grupos de investigación quienes, atacando problemas similares, se encuentran distribuidos por todo el mundo. Todos ellos, grandes y pequeños productores de datos, enfrentarán problemas similares de catalogación, preservación y diseminación de datos y del conocimiento que de éstos surge. Es imperioso planificar y construir repositorios de datos que los almacenen mientras se producen y que conserven la traza testimonial de las decisiones y criterios que los generaron (Gray y Szalay, 2002; Karasti *et al.*, 2006; Borgman *et al.*, 2007; Murray-Rust, 2008).

Si bien hace más de una década hubo algunos esfuerzos pioneros por generar un marco de recomendaciones que guiaran la preservación y diseminación de bases de datos científicos (Dozier *et al.*, 1995), es recientemente cuando organizaciones multilaterales y organismos planificadores de Europa y EEUU comenzaron a generar reportes técnicos, cada vez más detallados y con lineamientos respecto a como fomentar la preservación de colecciones de datos científicos (Lord y Macdonald, 2003; Arzberger *et al.*, 2004b; Simberloff *et al.*, 2005; Pilat y Fukasaku, 2007; Lyon, 2007; Clark, 2007). Incluso, algunos países como Canadá y China comienzan a fijar posición a ese respecto (Sabourin y Dumouchel, 2007; Xu, 2007). Sin embargo, muchas de las recomendaciones no han permeado hacia las comunidades productoras y/o custodios de colecciones de datos en esos países. Aún peor es la situación en Suramérica, donde todavía no hay convicción del cambio que experimentó la producción y diseminación del conocimiento, y tampoco existe la apropiación de un uso mínimo de las TIC en la docencia e investigación.

La creciente consciencia por preservar y compartir datos y, a partir de su uso y reutilización, generar nuevos conocimientos, ha impulsado una nueva faceta del Movimiento de Acceso

Libre al Conocimiento (OA, de *Open Access*). La preservación y diseminación libre a los datos científicos comienza a ser considerada, debatida y normada en distintos escenarios (Arzberger *et al.*, 2004a; Clark, 2007; Lyon, 2007; Pilat y Fukasaku, 2007; Sabourin y Dumouchel, 2007; Xu, 2007). Heredera de la reflexión y fundamento conceptual del OA, el Movimiento Datos Libres (OD, de *Open Data*) es un término reciente acuñado para indicar los estándares y procesos para que los datos de investigaciones puedan ser utilizados y reutilizados sin barreras ni costo. La comunidad académica percibe que los datos publicados le pertenecen como patrimonio. Sin embargo, muchos editores comienzan a demandar derechos de reproducción (*copyright*) lo cual, sin duda, es uno de los mayores impedimentos que habrá de enfrentar la e-investigación científica (Murray-Rust, 2008).

Este artículo pretende, por un lado, divulgar la reflexión que se viene dando en distintos foros internacionales referente al OD con el objeto de sensibilizar a los productores de datos para que asumamos la responsabilidad de difundir/preservar libremente nuestras mediciones y datos y, con ello, poder impulsar la creación de nuevos conocimientos. Por otro lado, se quiere mostrar un ejemplo de repositorio de datos libres que opera desde hace más de un lustro. Es una prueba de concepto de un repositorio de datos libres ambientales, el cual no solo colectiviza su recolección sino que comparte su custodia y curaduría. Se describen algunos conceptos y estrategias para la construcción de ese repositorio y se presenta una evaluación preliminar de su uso e impacto. Además, este artículo representa la continuación de un esfuerzo por garantizar y difundir libremente el conocimiento producido por los grupos de investigación de la Universidad de Los Andes, Venezuela (Núñez, 2002; Dávila *et al.*, 2006a, b). Este esfuerzo institucional por la preservación y diseminación del patrimonio intelectual de la ULA ha contribuido a la destacada posición de esta institución ocupa en los registros de actividad web de la academia en América Latina (Aguillo, 2005).

En la próxima sección se expondrán elementos de la reflexión general sobre OD, su relevancia y relación con el contexto del OA. Seguidamente, se presentan referencias técnicas relacionadas a los metadatos y su importancia para la difusión/preservación del patrimonio de mediciones, datos sintéticos y registros históricos. Luego se muestran algunos ejemplos de coleccio-

nes de datos, seguido de la experiencia en la creación y puesta en operación de la Red Bioclimática del Sur de Lago de Maracaibo y algunos de sus primeros resultados. Finalmente, se presentan algunas conclusiones y recomendaciones.

Los Datos: Un Patrimonio Intelectual Colectivo

La mayor parte de las investigaciones en Astronomía, Física de Altas Energía, Ecología y Medio Ambiente, Geología, Genética y Biología Molecular, por citar las áreas más relevantes productoras de datos para la e-investigación, están financiadas con fondos públicos. Por ello es de intuir que los datos provenientes de simulaciones y mediciones, y no solo las publicaciones producidas a partir de ellos, pertenecen a todos los ciudadanos. Los datos pueden ser el comienzo (o corroboración) de las ideas y, consecuentemente, deberían ser de libre acceso para que, con su uso y reutilización se pueda seguir la cadena de producción del conocimiento: Datos - Información - Conocimiento - Información - Datos. La idea central es que los datos generados por financiamientos públicos son patrimonio de la humanidad y deben estar accesibles y disponibles tan amplia y directamente como se pueda (Arzberger *et al.*, 2004a; Lessig, 2004; Alonso y Valladares, 2006). Esta visión contrasta con la actitud de investigadores y grupos de investigación que consideran los datos como su patrimonio y, sobre todo, se enfrenta a la reciente posición de muchos editores, quienes comienzan a exigir los datos que respaldan las publicaciones, haciéndoles extensivo el derecho de reproducción (*copyright*), con la consecuente restricción para su reutilización. Quizá el acceso a los datos pueda ser limitado si su utilización arriesga la seguridad de individuos o especies, compromete derechos de confidencialidad, o viola prerrogativas para su explotación temporal por quienes los recolectaron o generaron (Murray-Rust, 2008).

La razón por la que muchos grupos de investigación e investigadores individuales (financiados o no con fondos públicos) tienden a considerar los datos como su patrimonio se debe, principalmente, a la falta de claridad de los entes financieros y de las instituciones a las que pertenecen los investigadores y sus grupos de investigación. Las ventajas, y de allí la imperiosa necesidad, de preservar y difundir de forma libre los datos y mediciones no son (aún) apreciadas (Uhlir y Schröder, 2007). La disponibilidad de acceso

libre a los datos, además de incentivar la curiosidad y diversidad de análisis, promueve nuevas áreas de investigación transdisciplinaria al facilitar mecanismos automáticos de minería de datos.

Datos, Metadatos y la e-investigación

Se puede formular una idea simplificada de dato como aquella representación de información, capaz de ser procesada, interpretada, transmitida y preservada. Los datos surgen de mediciones de sensores (observacionales o experimentales), resultados de simulaciones (sintéticos o matemáticos, proveniente de modelos) o de registros en el tiempo (históricos).

La e-investigación impone un manejo automático de grandes volúmenes de datos. Para ser descubiertos, accedidos y analizados, los datos deben ser fácilmente identificables. En pocas palabras, para que los datos sean útiles deben poder ser descubiertos y para ello deben ser descritos apegados a estándares acordados por las comunidades productoras. A esa información básica utilizada para describir los datos: su contenido, formato, fechas importantes, condiciones de uso, fuente, propiedad y otras características, se conoce con el nombre de metadatos. Esta información permite al usuario evaluar si determinado conjunto de datos es adecuado para sus fines y facilitar el acceso a la información. Obviamente, los metadatos pueden ser o no digitales y los datos a los que están asociados pueden existir en ambas formas. La utilización de metadatos facilita (Michener *et al.*, 1997) la identificación y adquisición de datos para un tema determinado y para un período de tiempo o localización geográfica específica; el procesamiento, análisis y modelado automático de los datos; y la incorporación de elementos de conocimiento semántico asociado a los datos. Una adecuada documentación sobre el muestreo, procedimientos analíticos, anomalías y calidad de los datos, así como sobre la estructura de las colecciones de datos, ayudará a que esos datos puedan ser correctamente interpretados y reinterpretados en el futuro.

Conseguir la estandarización de los metadatos es muy importante porque permite definir una terminología común, y llevar a cabo la entrada, validación, acceso, integración y síntesis de los datos de manera automatizada, asegurando una documentación completa y precisa del contenido de los datos. Existen muchos y variados estándares de metadatos disponibles (*Dublin Core*, *Darwin Core*, *Content Standard*

for Digital Geospatial Metadata, ISO 19115, *Geographic Information Metadata*, *Ecological Metadata Language*, etc.). La razón de tantos estándares es las diversas aplicaciones de los metadatos (Bibliotecología, Biología, Geología, Ecología, Cartografía, etc.) y cada una de estas comunidades acuerda un conjunto de informaciones indispensables para identificar y manipular sus datos.

Las estructuras de metadatos están compuestas por elementos asociados a definiciones semánticas descriptivas de algunos de los posibles atributos del dato. Esas estructuras pueden ser arbitrariamente simples o complejas, y la información que contienen puede ser muy variada, dependiendo del tipo de dato y de las necesidades que para su uso impone la comunidad que lo generó. Un modelo de metadatos elaborado y auspiciado por la DCMI (*Dublin Core Metadata Initiative*) es el *Dublin Core*. Su flexibilidad y organización modular permiten describir, con el consenso internacional, cualquier tipo de recurso en sistemas de información. A partir del DCMI cada comunidad puede, dentro de su modelo de metadato, definir de manera diferente una propiedad o elemento, o extender el modelo básico. La iniciativa *Dublin Core* especifica un conjunto base de 15 elementos mientras que el modelo de metadato de *Learning Objects Metadata*, que está siendo desarrollado por la IEEE y otras organizaciones para describir recursos en ambientes de enseñanza-aprendizaje, tiene cerca de 100 elementos (Torrens, 2003).

Obviamente la incorporación de los metadatos demanda una inversión de tiempo y esfuerzo por parte de quienes generan y requieren preservar/compartir los datos. Es imperioso tomar en cuenta el tiempo que requiere la definición del modelo de metadatos, su aprendizaje y, posteriormente, su implantación. Luego vendrán los costos del mantenimiento a corto, mediano y largo plazo del sistema. Para que la implementación de un sistema basado en metadatos resulte exitosa debe existir un compromiso institucional del equipo de trabajo para la catalogación y preservación de los datos a largo plazo. En otras palabras, es imperioso involucrar en estas actividades al personal técnico de campo, los investigadores, estudiantes, personal técnico de informática y técnicos de laboratorio de la institución productora de datos.

Todo este esfuerzo se da de manera natural en grandes experi-

mentos, en la "gran ciencia". Sin embargo, es necesario que los grupos de investigación y los investigadores individuales tomen también conciencia de la importancia que tiene la catalogación y preservación de los datos. Solo así será posible garantizar su futuro uso y reutilización por parte de las generaciones futuras (Borgman *et al.*, 2007).

Ejemplos de Sistemas de Manejo de Colecciones de Datos

Una colección de datos es una serie de observaciones, simulaciones o registros, recolectados y/o generados con una misma metodología. En estas colecciones los metadatos definen los protocolos y mecanismos estandarizados para su identificación, localización, uso y reutilización. Las instituciones que coleccionan y distribuyen información (*clearinghouses* de datos) representan quizá la mayor experiencia en el uso de los repositorios de datos científicos dentro del ámbito geoespacial. Entre ellos cabe mencionar los repositorios del Comité Federal de Datos Geográficos (FGDC *Clearinghouse*) y de la Infraestructura Global de Datos Geoespaciales de EEUU (GSDI, *Global Spatial Data Infrastructure*).

La NASA también ha lanzado una iniciativa para reunir y distribuir gran cantidad de datos para el estudio de los efectos de cambios climáticos a escala global (gcmd.nasa.gov). A través de este sitio Web se puede obtener datos de diferentes tipos organizados de forma temática por su localización, por el tipo de instrumentación usada para obtenerlos, o por proyectos. Desafortunadamente, los cambios caprichosos de políticas que apuntan más a viajes interplanetarios que al deterioro del planeta, han afectado un número importante de proyectos de medición, catalogación y preservación de datos climáticos y las colecciones de datos, por ejemplo, sobre el calentamiento global.

Más recientemente las comunidades de ciencias ecológicas, ambientales y biológicas han desarrollado protocolos de preservación y la utilización de estándares de metadatos para el intercambio de información en colecciones de datos. Tal es el caso de la Red de Investigación Ecológica a Largo Plazo (LTER, de *Long Term Ecological Research*, www.lternet.edu), diseñada desde su concepción para considerar e incorporar el manejo y gestión de datos como una componente integral de la investigación. LTER dedica significativos esfuerzos al desarrollo de

métodos para manejar documentación, formatos, buenas prácticas para custodia y preservación de datos, así como también el desarrollo, y mantenimiento de códigos y aplicaciones informáticas (Torrens, 2003).

Conjuntamente con LTER, el proyecto Red del Conocimiento para la Biocomplejidad (KNB, de *Knowledge Network for Biocomplexity*) desarrolló el lenguaje EML para el intercambio de metadatos ecológicos y el software *Metacat* como un servidor de almacenamiento centralizado de colecciones de datos, que puede conectarse con otros servidores que usan los mismos mecanismos de comunicación e intercambio. Hoy día, KNB está desarrollando un sistema de almacenamiento y cómputo de datos a través de una *grid* (Ecogrid), partiendo de la infraestructura de información que tenían para el momento (para una descripción de KNB, Metacat y EML se puede consultar www.ecoinformatics.org).

En América Latina, el Programa Brasileño de Investigaciones Ecológicas de Larga Duración (PELD, de *Pesquisas Ecológicas de Longa Duração*) es un programa gubernamental que busca organizar y consolidar el conocimiento existente sobre los principales ecosistemas del Brasil. Comprende una docena de sitios que cubren selvas, lagos, ríos y manglares. Las mediciones incluyen datos físicos y químicos de aguas y suelos, densidad de especies y muestras de ADN, y caracterización temporal y espacial de mucha de esta información. La complejidad y diversidad de los datos, conjuntamente con la importancia que para Brasil tiene su medio ambiente, han permitido concretar una biblioteca digital de datos única en su tipo, con la cual los investigadores brasileños preservan y comparten estos datos (Barros *et al.*, 2007)

Red Bioclimática

En Venezuela, el uso de repositorios para el almacenamiento de datos ambientales, documentados por metadatos y accesible de forma libre a través de Internet es incipiente. Los pocos casos que existen prácticamente se limitan al *clearinghouse* de datos geospaciales del Instituto Geográfico de Venezuela Simón Bolívar (IGSM), al Sistema de Información Geográfico del Instituto Venezolano de Investigaciones Científicas (ECOSIG-IVIC, ecosig.ivic.ve) y a las iniciativas de la Universidad de Los Andes como son la Red Bioclimática de Mérida (RedBC) y la Red Venezolana de Estacio-

nes de Investigación Ecológica a Largo Plazo (EcoRed Venezuela).

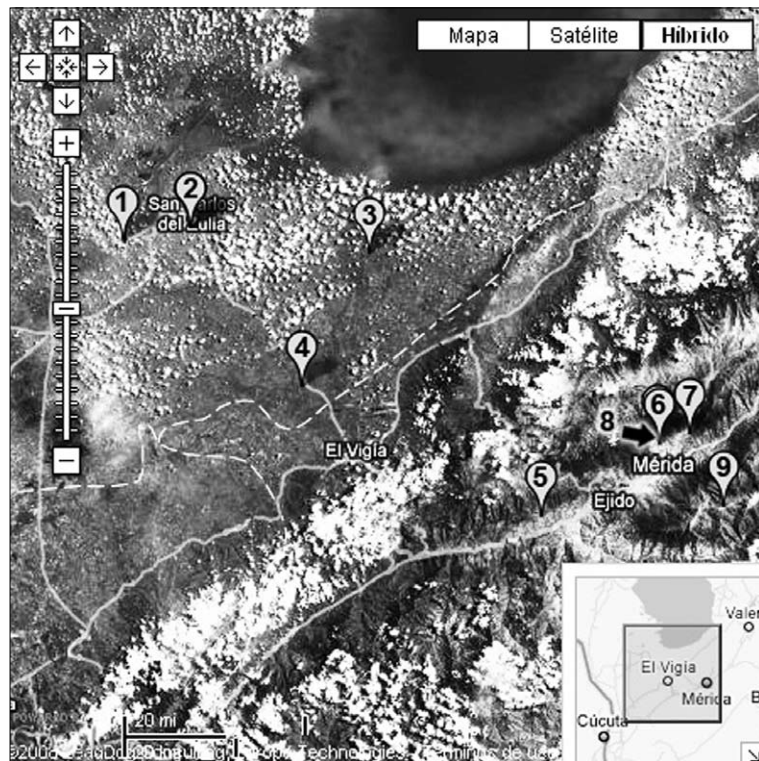
La EcoRed Venezuela fue establecida en 1997 con el apoyo del FONACIT, la Red LTER e instituciones científicas nacionales. Desde entonces, se han hecho esfuerzos para lograr que diferentes comunidades realicen la representación, manejo e intercambio eficiente de datos y colecciones de datos relacionados con las ciencias ecológicas, a través del uso de estándares internacionales y TIC. Con el apoyo del Centro Nacional de Cálculo Científico de la Universidad de Los Andes (CECALCULA) la EcoRed-Venezuela ha venido trabajando en el diseño y operación de una plataforma para el manejo de datos provenientes de proyectos de investigación ecológica a largo plazo.

Organización y actores de la RedBC

La RedBC es una iniciativa para el diseño y construcción de ambientes de e-investigación y e-colaboración, desarrollada por CECALCULA. Consiste en un proyecto piloto, en base a una prueba de concepto de un repositorio de datos para la captura, procesamiento y difusión de informa-

ción bioclimática. Está concebido dentro de un esquema demostrativo y cooperativo donde instituciones o individuos que dispongan de datos pueden catalogarlos y enviarlos a CECALCULA para su preservación y difusión libre. Así se garantiza la preservación, catalogación, custodia y distribución libre de los datos generados por el proyecto y por distintas instituciones que se han sumado a la iniciativa.

La RedBC se inició a partir del proyecto Sistema de Información Bioclimático para el Sur de Lago de Maracaibo y la meseta de Mérida (SIBILA), el cual consistió en instalar una red de estaciones telemétricas meteorológicas, construir e instalar un conjunto de colectores de esporas y desarrollar un sistema de información a través de la web. La herramienta computacional permite el acceso de investigadores, productores y organismos relacionados con la actividad del agro a la información producida por esta red de estaciones. Se mostró la factibilidad de enviar datos, por telemetría, a centros de acopio en el Sur del Lago y desde allí a CECALCULA. Para una segunda fase, una vez acumulados un conjunto mínimo de datos que permitiera determinar correlaciones



Estaciones climatológicas

- | | | |
|-------------------|----------------|---------------|
| 1 La Chiquinquirá | 3 Chama-INIA | 6 Mucujún |
| 2 La Glorieta | 4 San Juan | 7 Santa Rosa |
| 3 CIPLAT | 5 La Hechicera | 8 Pico Espejo |

Figura 1. Ubicación geográfica de las estaciones climatológicas de la RedBC. Aclarar qué significan los signos rojos y los números.

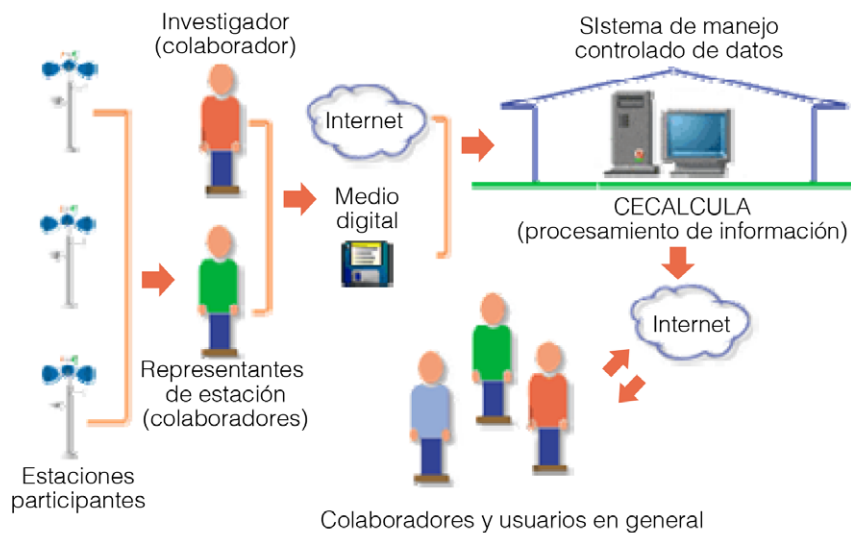


Figure 2. Esquema de procesos de envío y preservación de los datos bioclimáticos. El suministro de datos se realiza vía mensajería electrónica (correo electrónico o mensajería instantánea) o a través de envío en medios digitales (diskette, CD, DVD).

entre los datos climáticos y las incidencias de algunas afecciones que registran las plantaciones, se podrían generar, de forma automática, boletines informativos a los productores. Igualmente, a partir de los datos generados por las distintas estaciones, los investigadores podrían formular modelos microclimáticos y diseñar medidas fitosanitarias que permitirían el manejo efectivo de posibles desastres naturales. Recientemente comienzan a concretarse iniciativas de este tipo (Freitez, 2007) y este repositorio ya contiene los primeros datos de la incidencia de sigatoka negra en plantaciones de plátanos en el Sur del Lago de Maracaibo.

Estaciones climatológicas

En estos momentos son colaboradores de la RedBC el Centro Internacional del Plátano (CIPLAT), el Instituto Nacional de Investigaciones Agrícolas-Estación Chama (INIA-Chama), la Universidad del Sur de Lago Jesús María Semprúm (UNESUR), la Uni-

versidad de Los Andes (ULA) a través del Instituto de Investigaciones Agropecuarias (IIA-ULA) y el Laboratorio de Geofísica, el Parque Tecnológico de ULA a través de CECALCULA y el Instituto de Meteorología e Investigación Climática (*Institut für Meteorologie und Klimaforschung*, IMK) de la Universidad de Karlsruhe, Alemania.

Hasta ahora contribuyen con datos climáticos las siguientes estaciones (Figura 1):

- Est. Chama-INIA. Automatizada, localizada en Km. 41, vía Santa Bárbara. Sur del Lago de Maracaibo.
- Est. CIPLAT, Sur del Lago de Maracaibo. Automatizada, localizada en Pueblo Nuevo El Chivo. Sur del Lago de Maracaibo.
- Estación Mucujún-ULA, El Vallecito. Automatizada, pertenece al Centro de Investigaciones Atmosféricas y del Espacio (CIAE) de ULA.
- Est. La Hechicera-ULA, Mérida. Automática, perteneciente al Instituto de

Ciencias Ambientales y Ecológicas de ULA (ICAE).

– Est. Santa Rosa, Mérida. Convencional, localizada en el Instituto de Investigaciones Agrícolas de ULA, Santa Rosa, Mérida.

– Est. San Juan-ULA, San Juan de Lagunillas, Mérida. Convencional, pertenece al Instituto de Investigaciones Agrícolas de la ULA, localizada en San Juan de Lagunillas.

– Est. Pico Espejo-MARS, Mérida. Automatizada, forma parte de un convenio entre Venezuela (ULA) y el Gobierno de Alemania, localizada en la Estación Humboldt, Pico Espejo, Mérida, Parque Nacional Sierra Nevada.

– Est. La Glorieta-UNESUR. Automatizada, pertenece a UNESUR, localizada en Santa Bárbara del Zulia.

– Est. La Chiquinquirá-UNESUR. Automatizada, pertenece a UNESUR, localizada en Santa Bárbara del Zulia.

Procesos, datos y metadatos

La RedBC funciona a través de un Sistema de Manejo Centralizado de Datos en el que la unidad inicial de coordinación es CECALCULA. Tal y como puede apreciarse en la Figura 2, el suministro de datos se realiza vía mensajería electrónica (correo electrónico o mensajería instantánea) o a través de envío en medios digitales (diskette, CD, DVD). En CECALCULA, personal de soporte y apoyo a la RedBC revisa los archivos, organiza la información, genera los metadatos y verifica que éstos se ajusten a los estándares predefinidos. Los datos y metadatos son de libre acceso, se publican en el portal de la RedBC y pueden ser consultados vía Internet por los miembros de la red o por cualquier persona o institución que los necesite.

Colecciones de datos

Los datos se organizan en colecciones y se presentan en archivos clasificados por estación y por año. Sin ser una norma estricta, se recomienda que los archivos a ser enviados a la RedBC posean datos crudos, recabados en campo, en lugar de datos procesados.

Las colecciones de datos disponibles a través del portal de la RedBC son: Datos Climatológicos Est. Chama, 2001-2005; Datos Climatológicos CIPLAT, 2002-2003; Datos Climatológicos Est. Mucujún, 2000-2008; Da-

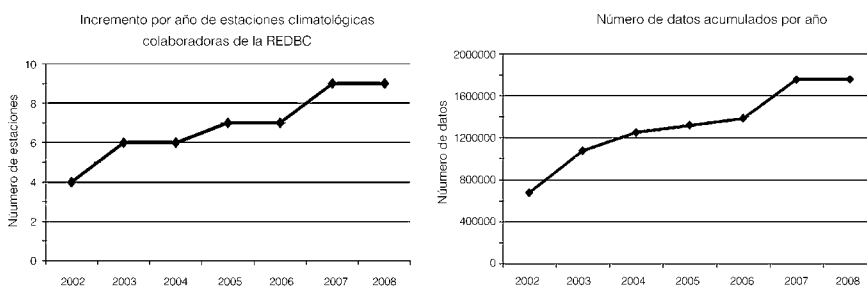


Figura 3. Estadísticas de crecimiento en la preservación y catalogación de datos. Es claro el incremento del número de estaciones y registros: cuatro estaciones y 679.094 registros en 2002 pasamos a nueve estaciones y 1.755.986 registros para finales del 2007.

tos Climatológicos Est. Santa Rosa, 1974-2001; Datos Climatológicos Est. San Juan, 1995-2001; Datos Climatológicos Est. Pico Espejo-MARS, 2004-2007; Datos Climatológicos Est. La Héchicera, 2000-2004; Datos Climatológicos Est. La Glorietal, 2005-2008; Datos Climatológicos Est. La Glorietal2, 2008; Datos Climatológicos Est. La Chiquinquirá, 2005-2008; Datos de evaluación de la enfermedad Sigatoka Negra en una plantación de plátano Hartón (*Musa AAB* cv. Hartón).

Datos libres de uso creciente e impacto fitosanitario

Actualmente, la RedBC proporciona al usuario, a través de Internet, una descripción completa de once colecciones de datos, de modo que puedan ser descubiertas, transferidas y usadas en el contexto apropiado. En la Figura 3a se puede apreciar el ritmo de crecimiento del número de estaciones que se han ido sumando a la RedBC. Se comenzó en 2002 con cuatro estaciones, dos en la zona Sur del Lago de Maracaibo y dos en la meseta de Mérida, y en el 2008 se dispone de nueve. Si bien el número de estaciones es limitado, el éxito de este enfoque voluntario se corrobora también con el incremento de la cantidad de datos almacenados, de 679094 registros en 2002 pasa a 1755986 para finales del 2007 (Figura 3b).

Recientemente (Freitez, 2007) un nuevo conjunto de datos se ha correlacionado con los datos climáticos. En Venezuela, el cultivo de musáceas (plátanos y bananos) es una de las actividades económicas más importantes en la región del Sur del Lago de Maracaibo. Más de la mitad del total de la superficie que en el país se destina a este tipo de cultivo se encuentra concentrada en esta región. En la última década, un hongo (la Sigatoka Negra o *Mycosphaerella Fijiensis* Morelet) ha impactado la producción del rubro disminuyendo su calidad y aumentando los costos producción por el uso periódico de agroquímicos, los cuales además contaminan significativamente aguas y suelos. Por ello, los investigadores del INIA, CIPLAT y ULA colocaron una serie de "caza esporas" que registran la densidad del inóculo en el ambiente. Con estos nuevos datos se ha comenzado a intentar correlacionar la densidad del inóculo en el ambiente con variables climáticas. Los resultados preliminares muestran que en algunos casos se logra predecir la intensidad de la enfermedad, con y sin discriminación por temporada climática (lluvia y sequía), con correlaciones has-

ta de 79,22%. En sequía (oct-mar) las variables correlacionadas con la densidad del inóculo en el ambiente son humedad relativa, precipitación, velocidad del viento y energía solar, mientras que durante las lluvias (abr-set), las variables mejor correlacionadas son temperatura del aire, evapotranspiración, humedad relativa y precipitación.

Conclusiones y Recomendaciones

Sin duda, este lustro de operación del Repositorio de Datos Libres de la RedBC ha mostrado su viabilidad. En el éxito de este tipo de proyectos juegan un papel primordial los esquemas culturales y costumbres institucionales. Por lo tanto, es preciso tener claro que convencer a los investigadores e instituciones generadoras de datos para que los resguarden en un repositorio y los compartan es un proceso muy lento. El logro más importante de esta red es la cooperación y en ello ha sido clave el adiestramiento técnico y la visión cooperativa de los beneficios de compartir los datos. Varios talleres y cursillos de adiestramiento en el manejo de datos han surgido y se seguirán desarrollando dentro de la Red. Igualmente, la consultoría y asesoría en el mantenimiento de estaciones es casi permanente. Aunque existen no pocas dificultades para convencer a algunos investigadores y grupos de investigación de la importancia para compartir los datos y colocarlos con acceso libre, el fácil acceso de los datos en la web y su perpetuación, así como también el reconocimiento recibido de otras instituciones que los utilizan, ha estimulado el deseo de pequeños grupos a contribuir con sus datos. Es así como el éxito de la experiencia de la RedBC se puede medir por las estadísticas de incorporación de estaciones y de preservación de datos.

La RedBC proporciona al usuario, a través de Internet, una descripción completa de las colecciones de datos de tal modo que éstas puedan ser descubiertas, transferidas y usadas en el contexto apropiado por otros usuarios potenciales. La creación de este repositorio de datos científicos ha hecho posible que éste funcione como un proyecto que le permite a una parte de la comunidad científica venezolana empezar a catalogar y sistematizar sus colecciones de datos ambientales y, más importante aún, saber donde buscar al realizar investigaciones posteriores.

Si esta experiencia sirve de estímulo a otras instituciones y se logra dar un efecto multiplicador, nuestro

país saldrá beneficiado y fortalecido al disponer de redes de colaboración en todos los estados que inclusive pudieran interconectarse para ofrecer un mejor servicio. Un primer paso en esta replicación de esfuerzos en la preservación, lo ha constituido la iniciativa de la UNESUR en seguir el ejemplo de la RedBC y conformar también un repositorio para el almacenamiento, resguardo y difusión de sus datos bajo el asesoramiento de CECALCULA.

AGRADECIMIENTOS

Este proyecto ha sido financiado parcialmente por el Programa PIC Agenda Plátano del CDCHT ULA, bajo el número CVI PIC AGM01099. Igualmente los autores agradecen a Fundacite Mérida por el apoyo financiero brindado. Adicionalmente uno de nosotros (LAN) agradece también al FONACT (proyectos SI-2000000820 y F-2002000426) y la hospitalidad del Grupo de Relatividad y Gravitación de la Universidad de las Islas Baleares en España, que permitió la finalización de la redacción de este trabajo.

REFERENCIAS

- Aguillo IF, Granadino B, Llamas, G (2005) Posicionamiento en el web del sector académico iberoamericano. *Interciencia* 30: 735-738.
- Alonso B, Valladares F (2006) Bases de datos y metadatos en ecología: compartir para investigar en cambio global. *Ecosistemas* 6: 410.
- Arzberger P, Schroeder A, Beaulieu A, Bowker G, Casey K, Laaksonen L, Uhlir P, Wouters P (2004a) Promoting access to public research data for scientific, economic, and social development. *Data Sci. J.* 3: 135-152.
- Arzberger P, Schroeder P, Beaulieu A, Bowker G (2004b) Science and government: An international framework to promote access to data. *Science* 303: 1777-1778.
- Barros E, Laender A, Gonçalves M, Cota R (2007) Transitioning from the ecological fieldwork to an online repository: a digital library solution and evaluation. *Int J Dig. Libr.* 7: 109-112.
- Borgman C (2006) What can studies of e-learning teach us about collaboration in e-research? Some findings from digital library studies. *Comp. Supp. Coop. Work (CSCW)* 15: 359-383.
- Borgman C, Wallis J, Enyedy N (2007) Little science confronts the data deluge: habitat ecology, embedded sensor networks and digital libraries. *Int J Dig. Libr.* 7: 17-30.
- Clark R (2007) Database protection in Europe - recent developments and modest proposal. *Data Sci. J.* 6: OD12-OD20.
- Coles SJ, Frey JG, Hursthouse MB, Light ME, Milsted AJ, Carr LA, De Roure D, Gutteridge CJ, Mills HR, Meacham KE, Surrridge M, Lyon E, Heery R, Duke M, Day M

- (2006) An e-Science environment for service crystallography from submission to dissemination. *J. Chem. Inf. Model.* 46: 1006-1016.
- Collins L, Martínez M, Mane K, Powell J, Kieffer C, Simas T, Heckethorn S, Varjabedian K, Blake M, Luce R (2007) Collaborative e-Science libraries. *Int. J. Digit. Libr.* 7: 31-33.
- Dávila JA, Núñez L, Sandia B, Silva JG, Torrens R (2006a) www.saber.ula.ve: un ejemplo de repositorio institucional universitario. *Interciencia* 31: 29-37.
- Dávila JA, Núñez L, Sandia B, Torrens R (2006b) Los repositorios institucionales y la preservación del patrimonio intelectual académico. *Interciencia* 31: 22-29.
- De Roure D, Frey J (2007) Three perspectives on collaborative knowledge acquisition in e-science. En *Workshop on Semantic Web for Collaborative Knowledge Acquisition*. 20th Int. Joint Conf. on Artificial Intelligence. Hyderabad, India. Enero 2007.
- Dozier J, Alexander S, Courain M, Dutton JA, Emery W, Gritton B, Jenne R, Kurth W, Lide D, Richard BK, Warnow-Blewett J (1995) *Preserving Scientific Data on Our Physical Universe: A New Strategy for Archiving the Nation's Scientific Information Resources*. Reporte Técnico. National Research Council. Inglaterra. pp. 49-61.
- Foster I (2005) Service-oriented science. *Science* 308: 814-817.
- Freitez JA (2007) *Desarrollo de un Modelo Predictivo del Brote de Sigatoka Negra para las Plantaciones de Plátano al Sur del Lago de Maracaibo*. Tesis. Universidad de Los Andes. Mérida, Venezuela. pp. 61-86.
- Gray J, Szalay A (2002) The world-wide telescope. *Comm. ACM* 45: 50-55.
- Hey T, Trefethen A (2003a) The data deluge: An e-science perspective. En Berman F, Fox G, Hey T (Eds.) *Grid Computing: Making the Global Infrastructure a Reality*. Wiley. Nueva York, EEUU. pp. 809-824.
- Hey T, Trefethen AE (2003b) E-science and its implications. *Phil. Trans. R. Soc. Lond. A*, 361: 1809-1825.
- Hey T, Trefethen AE (2005) Cyberinfraestructura for e-science. *Science* 308: 817-821.
- Karasti H, Baker K, Halkola E (2006) Enriching the notion of data curation in e-science: Data managing and information infrastructuring in the long term ecological research (LTER) network. *Computer Supported Cooperative Work (CSCW) 15*: 321-358.
- Lessig L (2004) *Free Culture*. Penguin. Nueva York, EEUU. pp. 119-179.
- Lord P, Macdonald A (2003) *Data curation for e-science in the UK: an audit to establish requirements for future curation and provision*. Reporte Técnico. The JISC Committee for the Support of Research. The Digital Archiving Consultancy Limited. Twickenham, Reino Unido. pp. 28-81.
- Lyon L (2007) *Dealing with data: Roles, rights, responsibilities and relationships*. Reporte Técnico. UKOLN. Bath, Inglaterra. pp. 16-42.
- Michener W, Brunt J, Helly J, Kirchner T, Stafford S (1997) Nongeospatial metadata for the ecological sciences. *Ecol. Applic.* 7: 330-342.
- Murray-Rust P (2008) Open data in science. *Nature Precedings*. hdl:10101/npre.2008.1526.1. precedings.nature.com.
- Núñez L (2002) La reconquista digital de la biblioteca pública. *Interciencia* 27: 195-201.
- Pilat D, Fukasaku Y (2007) OECD principles and guidelines for access to research data from public funding. *Data Sci. J.* 6: OD4-OD11.
- Sabourin M, Dumouchel B (2007) Canadian national consultation on access to scientific research data. *Data Sci. J.* 6: OD26-OD35.
- Simberloff D, Barish BC, Droegemeier KK, Etter DM, Fedoroff NV, Ford KM, Lanzerotti LJ, Leshner A, Lubchenco J, Rossmann MG, White JAJr (2005) Long-lived digital data collections enabling research and education in the 21st century. Reporte Técnico NSB-05-40. National Science Foundation. Washington, DC; EEUU. pp. 23-40.
- Torrens R (2003) *Desarrollo de Sistemas de Información Bio-climática*. Tesis. Universidad de Los Andes, Mérida, Venezuela. pp. 29-50.
- Uhlir P, Schröder P (2007) Open data for global science. *Data Sci. J.* 6: OD36-OD53.
- Xu G (2007) Open access to scientific data: Promoting science and innovation. *Data Sci. J.* 6: OD21-OD25.

BIOCLIMATIC NETWORK OF MERIDA STATE, VENEZUELA: ENVIRONMENTAL DATA RETRIEVAL, HANDLING AND PRESERVATION STRATEGIES

Hilda Yelitz Contreras, Zulay Méndez, Rodrigo Torrens and Luis A. Núñez

SUMMARY

This paper presents the experience accumulated in the construction and development of the Bioclimatic Network of the Merida State (Venezuela) as a test of concepts in the use of techniques and strategies for the retrieval, handling and preservation data from e-research. The ethical, conceptual and technical aspects involved in the preservation of the data are described, with emphasis on the concepts of open data and metadata, and their

standards. The construction of the bioclimatic data repository at www.cecalc.ula.ve/redbc/ is reviewed, as well as its associated information products and some preliminary results and impacts of its utilization. Finally, some reflections and recommendations are made that will, perhaps, allow the maintenance and replication of this initiative.

DESENVOLVIMENTO DA REDE BIOCLIMÁTICA DO ESTADO MÉRIDA, VENEZUELA: ESTRATÉGIAS DE CAPTURA, MANEJO E PRESERVAÇÃO DE DADOS AMBIENTAIS

Hilda Yelitz Contreras, Zulay Méndez, Rodrigo Torrens e Luis A. Núñez

RESUMO

Apresenta-se a experiência na construção e desenvolvimento da Rede Bioclimática do Estado Mérida (Venezuela) como prova de conceito na utilização das técnicas e estratégias para a captura, manejo e preservação de dados ambientais provenientes da investigação eletrônica. Descrevem-se os elementos éticos, conceituais e técnicos para a preservação do patrimônio de dados, fazendo ênfase nos conceitos dados livres (open data), metada-

dos e seus estándares. Apresenta-se o testemunho da construção do repositório de dados bioclimáticos www.cecalc.ula.ve/redbc/, seus produtos de informação associados e alguns resultados e impactos preliminares de sua utilização. Finaliza-se com algumas reflexões e recomendações as quais, quem sabe, permitam fazer sustentável e replicável esta iniciativa.