

Recuperación, archivo y administración de registros y datos meteorológicos

por Jason Cooper, archivero del Centro Nacional de Datos Climáticos de la Administración Nacional del Océano y de la Atmósfera (NOAA), Estados Unidos de América



Los repositorios y archivos de datos de observación desempeñan un papel decisivo en el estudio del tiempo y del clima. Después de más de doscientos años de registros de observaciones en papel, las de los últimos 20 han sido grabadas en soportes digitales. Actualmente los repositorios mundiales disponen de datos en soportes físicos, como el papel y el microfilme, y en soportes digitales, como los discos de ordenador, las cintas magnéticas y los DVD; pero existe un riesgo de pérdida de datos independientemente del soporte en que se almacenen. Los repositorios deben identificar, evaluar, adquirir, archivar y migrar los datos a las nuevas tecnologías para que la comunidad científica pueda disponer de ellos. La administración y conversión de datos a formatos electrónicos convencionales es de vital importancia a fin de preservar y mejorar los registros climáticos mundiales.

Transformar colecciones de datos dispares en conjuntos completos y homogéneos es un reto fundamental de cara a la comunidad de investigadores climáticos. La mejor manera de lograrlo es mediante la cooperación al compartir recursos y áreas de conocimiento. El objetivo es poner una mayor cantidad de información en manos de los científicos que buscan comprender el sistema climático mundial y anticipar los cambios del clima.

Los datos observacionales en la actualidad

En el siglo XXI las observaciones meteorológicas son mayoritariamente recogidas en forma digital. La base de datos digital del Centro Nacional de Datos Climáticos de Estados Unidos de América (NCDC o el Centro) comenzó a funcionar formalmente a finales de la década de 1970, cuando la información satelital pasó a formatos digitales. En la década de 1990 el Servicio Meteorológico Nacional desplegó nuevos sistemas para obtener datos *in situ* y de radar, y reemplazó la antigua red de radares meteorológicos por la nueva red de radares meteorológicos de próxima generación (NEXRAD), con lo que las imágenes digitales comenzaban a llegar directamente al NCDC. Hoy en día, el Centro posee veinte años de datos NEXRAD en su

repositorio digital, al mismo tiempo que guarda más de 20 millones de imágenes anteriores a la red NEXRAD en microfilme.

Desde el comienzo de las observaciones meteorológicas horarias en los aeropuertos, a finales de la década de 1920, el NCDC ha almacenado decenas de millones de impresos de observación por todo el país, que se digitalizaron allá por la década de 1990 cuando el Servicio Meteorológico Nacional puso en marcha el sistema automático de observación en superficie. Este sistema transmite datos en formato digital directamente al NCDC, dando así por finalizado en gran medida el uso del papel en las observaciones. Así mismo, ante la utilización masiva de Internet por parte de la red de observadores voluntarios, se han establecido nuevos procedimientos de presentación de datos digitales.

Actualmente queda poco en formato de papel o de mapa analógico. Los únicos datos nuevos que llegan al NCDC en formato impreso en papel provienen de barcos en el mar —estaciones con problemas de transmisión intermitente— y de un número muy reducido de estaciones que aún no han efectuado la conversión al formato digital. La gran colección de registros en soporte físico del NCDC se incrementa lentamente. Ocasionalmente aún recibe pequeñas cantidades de datos en papel y en soportes digitales (disquetes, CD, etc.), encontrados en oficinas de predicción del Servicio Meteorológico Nacional o en oficinas climáticas estatales, proporcionados por personal saliente. Algunos datos han estado olvidados o abandonados durante décadas.

Rescate de datos

Con la transición masiva de las observaciones meteorológicas a los formatos digitales, los registros en papel o en microfilme envejecen en los repositorios de todo el mundo. El objetivo de las iniciativas de rescate de datos es el de aumentar la disponibilidad de estos datos históricos. En Estados Unidos, el enorme esfuerzo de digitalización conocido como Programa de modernización de las bases de datos climáticas ha convertido con

éxito más de cincuenta millones de registros de observación en formatos de imagen digital (PDF, TIFF, JPEG, etc.). El citado programa también contribuyó a informatizar cientos de millones de elementos meteorológicos para convertirlos en conjuntos de datos digitales. El éxito del programa ha dado como resultado la disponibilidad en formatos de imagen de los datos horarios de estaciones estadounidenses con un largo periodo de observación así como la informatización de los archivos de datos, y que Estados Unidos cuente con una densa cobertura de datos que se inicia en el siglo XIX.

Hoy, el rescate y la digitalización de datos muestran un aspecto diferente. El citado programa de modernización finalizó en 2011, pero aún quedan registros en soporte físico por digitalizar. El NCDC busca activamente colaboraciones para continuar con el rescate y con la conversión de datos a formatos digitales que puedan utilizarse en la investigación climática. En este sentido, el NCDC se ha involucrado en proyectos de colaboración para escanear antiguos registros satelitales que muestran la extensión de los hielos marinos al final de la década de 1960 y principios de la de 1970. El Centro Nacional de Huracanes ha comenzado a utilizar las primeras imágenes de satélites geoestacionarios y polares —que antes solo estaban disponibles en formato analógico— para llevar a cabo reanálisis de las temporadas de huracanes en el Atlántico entre 1966 y 1975. Posiblemente existan oportunidades de extender el reanálisis a las cuencas de los océanos Índico y Pacífico a través de la iniciativa del Archivo internacional de trayectorias más verosímiles para la administración climática.



NOAA

Archivos de la NOAA.

Como Centro mundial de datos de la OMM, el NCDC tiene en sus propios archivos unas 2 000 cajas de registros en papel provenientes del extranjero, y aproximadamente 5 000 carretes de microfilme. Así mismo posee registros de observaciones realizadas por las fuerzas armadas en campañas de guerra, desde la Segunda Guerra Mundial hasta los conflictos de los Balcanes y de Kuwait a principios de la década de 1990. Por primera vez los usuarios pueden encontrar estos y todos los registros en soporte físico en el NCDC, dado que actualmente puede accederse al banco de datos en soporte físico a través de la interfaz web del Sistema

de archivo para el seguimiento y el inventario de registros (WebARTIS). También existe la posibilidad de capturar y usar observaciones globales *in situ*. Las peticiones que requieran la conversión de formato físico en digital deben sufragar parte del coste que conlleva. Recientemente el Gobierno de Nueva Caledonia ha financiado el escaneo de las observaciones realizadas en la Segunda Guerra Mundial en sus islas y así rellenar un gran vacío existente en sus archivos climáticos. El Observatorio de Hong Kong descubrió que el NCDC posee registros originales de su observatorio realizados en la década de 1930; el Centro está creando copias digitales de los mismos.

El NCDC no es el único que posee registros meteorológicos históricos. Por todo el mundo existen valiosos registros nacionales que, si se digitalizan, reforzarían el registro climático y ayudarían a comprender el clima. El NCDC mantiene una estrecha relación con la Organización Internacional de Rescate de Datos Ambientales (IEDRO), que trabaja en la identificación y el rescate de registros meteorológicos en aquellos países que poseen gran cantidad de información del tiempo en formato analógico y que necesitan ayuda para conservarla y digitalizarla. El NCDC trabaja con la IEDRO en la definición de formatos comunes y en la recomendación de formatos de imagen para aquellos registros que son identificados y rescatados por organizaciones sin ánimo de lucro. El ámbito del rescate de datos incluye a quienes poseen y conservan registros en formato físico, a quienes buscan registros para escanear, a quienes digitalizan datos, y a quienes conservan y ofrecen acceso a colecciones digitales incluidas en los actuales registros climáticos.

Obsolescencia

El NCDC habitualmente acepta y se preocupa de aquellos datos que se encuentran en papel, en microfilme o en diferentes formatos digitales. Sin embargo, hay algunos formatos que presentan mayor dificultad en su mantenimiento y acceso. Aunque tiene capacidad para leer películas de 16 y de 35 mm, o incluso microtarjetas, el Centro necesita equipos especiales para leer microfilmes de 5 pulgadas, que son formatos muy poco comunes. Los repositorios deben ocuparse de un problema mucho más habitual: los datos contenidos en dispositivos portátiles desfasados o en formatos de archivo anticuados. El NCDC tiene la fortuna de disponer de una máquina con la que leer y manejar disquetes de 3,5 y 5,25 pulgadas pero, de nuevo, ha de buscar ayuda en la lectura de los “discos zip”, muy populares durante un corto periodo de la década de 1990. Tanto los CD-ROM como los DVD están perdiendo popularidad. Debido a la rápida obsolescencia y con objeto de continuar con la labor de preservación, el NCDC mantiene la política de pasar todos los datos que se encuentren en medios portátiles a un sistema digital de archivos gestionado en red.

Otro asunto relativo a la preservación digital es el uso de formatos que están pasados de moda o que son exclusivos. Muchos de ellos pertenecen a compañías

informáticas y, si estas dejan de existir o cambian a un nuevo programa informático, disminuye la asistencia técnica a los formatos antiguos. Cuando sucede esto, los repositorios deben migrar los ficheros a unos mejores formatos o bien conservar documentación para ayudar a los usuarios a hacerlo. Los repositorios digitales han de supervisar los formatos en sus fondos de datos y abandonar aquellos que pierden soporte técnico. La política del NCDC es la de usar el menor número posible de formatos, de manera que se minimice la cantidad de ellos a supervisar. La comunidad de archivos en general ofrece ideas acerca de los formatos más duraderos y la comunidad científica recomienda formatos autodescriptivos normalizados como el NetCDF (formulario de datos comunes en red).

Evaluación

Tanto si los datos están en formato físico como digital, siempre hay que considerar el procedimiento de archivo conocido como evaluación, es decir, el juicio subjetivo acerca de si un registro, un conjunto de datos o una colección pertenecen a un repositorio determinado. Las consideraciones respecto a la evaluación llevadas a cabo por el NCDC caen dentro de dos categorías básicas. La primera se refiere a la idoneidad del archivo, para la cual se pregunta lo siguiente (téngase en cuenta que la palabra "colección" puede sustituirse fácilmente por "datos" o "registros" según el soporte de que se trate):

- ¿La colección se encuentra ya archivada en algún lugar? Si es así, ¿se trata de archivos nacionales, universitarios o corporativos? En caso afirmativo, ¿deberíamos establecer acuerdos con finalidad de adquirirlos o tendría que ser el propietario quien disponga de ellos?
- ¿Es el NCDC el repositorio idóneo para la colección? Si no es así, ¿qué alternativas hay?
- ¿Es este el lugar donde los usuarios esperarían encontrar la colección? En caso negativo, ¿dónde buscarían? ¿Sería capaz ese repositorio de llevar a cabo su administración?
- ¿Estarían interesados en la colección los usuarios del NCDC? Se pide a los productores que identifiquen a la comunidad de usuarios de su colección.
- ¿La colección complementa nuestras existencias?
- ¿La colección sirve de apoyo al trabajo realizado en el NCDC, tal como desarrollo de productos, investigación o publicaciones?
- ¿La colección ha sido elaborada por la NOAA? La primera y más importante de las misiones del NCDC es la de servir como repositorio de su organismo matriz.

La situación ideal sería que el NCDC aceptara todos los registros que satisfagan el cuestionario anterior, pero los recursos disponibles y la capacidad para administrar los nuevos datos o registros en soporte físico juegan un papel fundamental y hacen necesario un segundo cuestionario:

- ¿Es capaz el NCDC de manejar el volumen de datos, tanto en transferencia como en almacenamiento?
- ¿El NCDC puede preparar los datos para su archivo? ¿Ayudará el productor en las tareas de transferencia, creación de metadatos y denominación de ficheros?
- ¿Los datos están en el formato preferido, o el NCDC puede acomodarlos a otro distinto?
- ¿El NCDC puede ofrecer acceso a los registros?
- ¿Es el NCDC competente en la materia como para responder a las preguntas de los usuarios?

Cuando se lleva a cabo una labor de evaluación de forma coherente, el repositorio desarrolla una identidad sólida y esta es percibida por la comunidad de usuarios que saben qué clase de datos y de registros físicos pueden encontrarse allí. Y a la inversa, una evaluación desordenada fomenta la fragmentación de las colecciones y la incertidumbre sobre el objetivo del repositorio, tanto desde la perspectiva de la comunidad a la que presta sus servicios como, incluso, desde la de sus propios empleados.

El repositorio y la investigación sobre el clima

El NCDC representa la vanguardia de los repositorios digitales para los registros que han sido convertidos a formatos de imagen y de datos informatizados. El papel del NCDC como Centro mundial de datos, así como su compromiso con la gestión digital, hacen del Centro el lugar idóneo donde enviar los datos para su archivo. El NCDC participa activamente en la Iniciativa internacional sobre las temperaturas en superficie, es decir, el repositorio mundial de las observaciones realizadas en la superficie terrestre desde principios del siglo XIX hasta la fecha. De igual forma participa en el Banco de datos internacional de presión en superficie, que contiene la mayor colección mundial de observaciones de presión atmosférica. Estas asociaciones de datos representan la instancia final del esfuerzo mundial de colaboración con respecto al rescate de datos: bases de datos globales tan completas y homogéneas como sea posible teniendo en cuenta las fuentes dispares de los registros originales de observación.

El rescate y la conservación de datos a lo largo del siglo XXI requieren cooperación para compartir recursos y experiencia. De esta forma, el rescate de datos se parece cada vez más a la investigación sobre el clima, con docenas de países, organizaciones sin ánimo de lucro y universidades contribuyendo a la mejora de los registros climáticos. Gracias a contar con unas bases de datos más ricas y completas, la comunidad científica climática de todo el mundo continuará avanzando en el conocimiento del sistema climático mundial y anticipando los futuros cambios en el clima.

Las referencias bibliográficas se encuentran en la versión en línea