



Centro Regional del  
Clima para el Sur de  
América del Sur

Centro Regional do  
Clima para o Sul da  
América do Sul



Serie Reportes Técnicos – Reporte Técnico CRC-SAS-2013-001

## Diseño del proceso de control de calidad de datos climáticos diarios en el Centro Regional del Clima para el Sur de América del Sur

Guillermo Podestá

*University of Miami, Rosenstiel School of Marine and Atmospheric Science, Miami, USA*

María de los Milagros Skansi

Natalia Herrera

Hernán Veiga

*Servicio Meteorológico Nacional, Buenos Aires, Argentina*

Santiago Rovere

*Facultad de Ingeniería, Universidad de Buenos Aires, Buenos Aires, Argentina*

15 Noviembre 2013

---

# 1 Introducción

Este documento describe el proceso de control de calidad de datos diarios de variables meteorológicas almacenadas en la base de datos del Centro Regional del Clima para el Sur de Sudamérica. Este documento se enfoca en el flujo de información; los detalles de los controles se incluyen en un documento separado (Reporte Técnico CRC-SAS 2013-002).

## 1.1 La base de datos del Centro Regional del Clima para el sur de América del Sur

El Centro Regional del Clima para el sur de América del Sur (en adelante, CRC-SAS) es un esfuerzo liderado por los servicios meteorológicos de Brasil y Argentina como parte del Marco Mundial para los Servicios Climáticos de la Organización Meteorológica Mundial (OMM) de la Organización de Naciones Unidas ([www.wmo.int/pages/publications/bulletin\\_es/documents/bulletin\\_es.pdf](http://www.wmo.int/pages/publications/bulletin_es/documents/bulletin_es.pdf)). El CRC-SAS involucra además la participación de Paraguay y Uruguay, así como Bolivia y Chile en carácter de miembros asociados. El objetivo principal del CRC-SAS es la producción y disseminación de datos, información y conocimiento climático que sea útil para apoyar la toma de decisiones en sectores de la sociedad sensibles a la variabilidad y cambio climático.

Dada la creciente comprensión de la influencia del clima en diferentes actividades humanas, existe una demanda creciente de datos y productos climáticos confiables. La disponibilidad de observaciones largas, continuas, bien calibradas y controladas es crítica para una caracterización adecuada de la evolución del sistema climático. Por este motivo, una de las actividades iniciales del CRC-SAS es la recopilación y control de calidad de datos climáticos históricos.

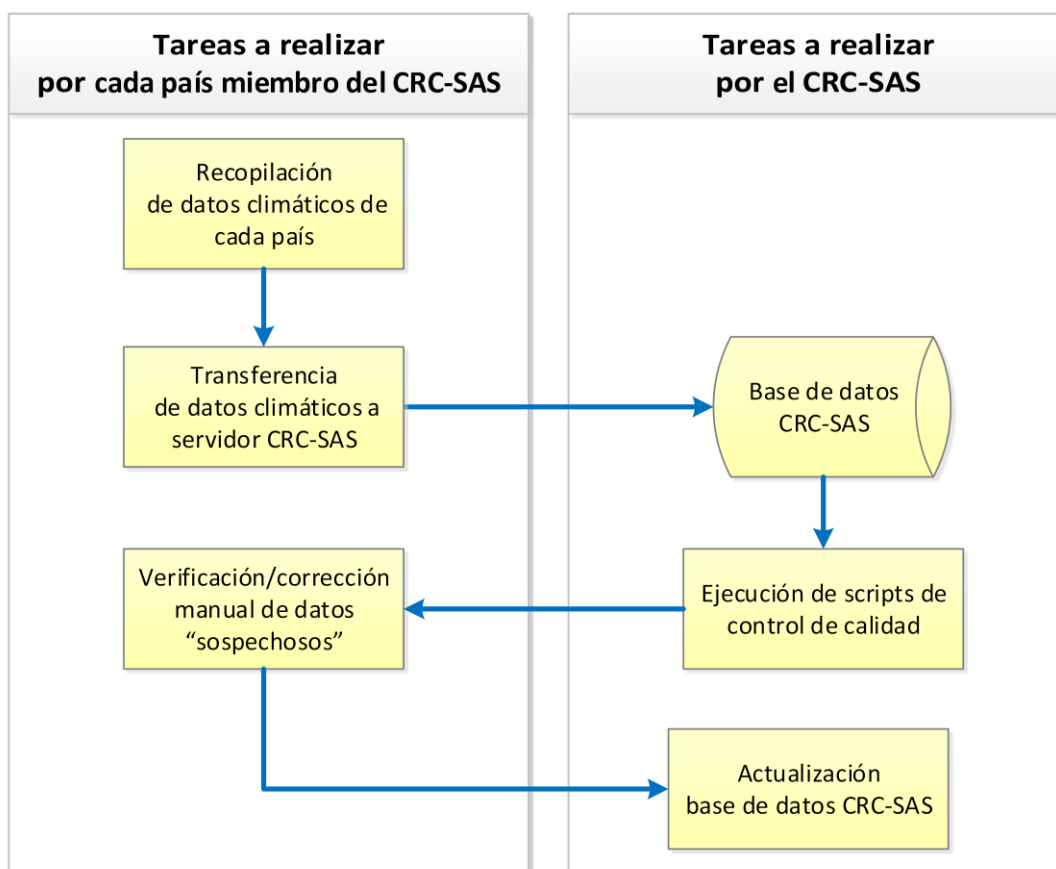
**Variables meteorológicas.** Se están recopilando datos diarios para una serie de variables meteorológicas (ver Apéndice A) para el período desde el 1 de enero de 1961 hasta el presente. Como un paso inicial, se están compilando datos provenientes de estaciones meteorológicas convencionales operadas por los países miembros del CRC-SAS. Las estaciones convencionales cumplen con ciertos requisitos: deben estar equipadas con instrumentos manuales operados por un observador meteorológico, el cual se debe encargar de efectuar mediciones meteorológicas, y puede también ser el encargado del funcionamiento y mantenimiento del instrumental si se le proporciona la formación adecuada ([ftp://ftp.wmo.int/Documents/MediaPublic/Publications/WMO488\\_GOSguide/488\\_2012\\_es.pdf](ftp://ftp.wmo.int/Documents/MediaPublic/Publications/WMO488_GOSguide/488_2012_es.pdf)). En una etapa siguiente se planea incorporar datos de estaciones meteorológicas automáticas, en los cuales los datos se registran electrónicamente sin un operador presente. Los datos históricos se están almacenando en una base de datos relacional PostgreSQL ([www.postgresql.org](http://www.postgresql.org)) administrada por el CRC-SAS. Esta base de datos se irá actualizando periódicamente (por ejemplo, cada 7-10 días) de modo de permitir la producción de diagnósticos que reflejen condiciones climáticas recientes en la zona abarcada por el CRC-SAS.

**Metadatos.** La base de datos también incluye metadatos, o sea datos sobre las estaciones meteorológicas. La OMM está actualmente desarrollando estándares para la recopilación de metadatos. Por este motivo, los metadatos a ser incluidos en la base de datos del CRC-SAS por el momento son muy simples. El contenido actual de los metadatos se lista en el Apéndice B.

## 2 Flujo del procesamiento de datos meteorológicos en el CRC-SAS

El uso de los datos climáticos en la base de datos del CRC-SAS para la generación de información climática útil y relevante requiere que los datos hayan sido sometidos a un proceso de control de calidad. Este proceso debe identificar valores sospechosos que podrían ser incorrectos y, en consecuencia, afectar indebidamente los productos o estadísticas derivados a partir de los datos originales. Por esta razón, el CRC-SAS– en colaboración con un proyecto de investigación financiado por el Instituto Interamericano para la Investigación del Cambio Global y por el Banco Interamericano de Desarrollo – ha recopilado e implementado una serie de procedimientos publicados en la literatura científica para el control de calidad de datos meteorológicos. Estos procedimientos se describen en detalle en otro documento (Reporte Técnico CRC-SAS-2013-002) Los planes actuales prevén que el control de calidad de los datos sea realizado en forma centralizada por el CRC-SAS, siguiendo el flujo que se ilustra en la Figura 1. En las siguientes secciones se describe muy sucintamente cada una de las etapas principales en el proceso.

Figura 1. Organización general del proceso de compilación y control de calidad de datos meteorológicos diarios en la base de datos del CRC-SAS.



## 2.1 Recopilación de datos climáticos de cada país miembro del CRC-SAS

Cada país miembro definirá las estaciones meteorológicas convencionales cuyos datos se contribuirán a la base de datos del CRC-SAS. Los criterios que se han consensado inicialmente sugieren que se contribuyan datos para:

- estaciones meteorológicas convencionales que estén actualmente activas; y
- estaciones meteorológicas convencionales que hayan operado por al menos 10 años consecutivos en el período desde el 1 de enero de 1961 al presente.

Debe enfatizarse que la acción de contribuir datos nacionales al CRC-SAS no implica que esos datos sean distribuidos a cualquier entidad o persona que los solicite. La diseminación de datos meteorológicos diarios será definida por cada país miembro. Un miembro puede estipular que sus datos puedan ser usados para la generación de productos por parte del CRC-SAS, pero que no sean publicados o transferidos a terceros.

Cada país se ha comprometido a proveer los datos en el formato consensado por los miembros. El formato de los archivos de texto que se usarán para la transferencia de los datos se ilustra en la Figura 2. Idealmente, debería desarrollarse en cada institución miembro un procedimiento automatizado para la exportación de los datos (sobre todo teniendo en cuenta que las actualizaciones futuras tendrán una frecuencia de 1-2 veces por mes).

Figura 2. Ejemplo de formato de archivo de texto con datos meteorológicos diarios para ser incluido en la base de datos del CRC-SAS. Las variables cuyo nombre se abrevia en el encabezamiento del archivo se describen en el

Apéndice A; el texto se escribe en forma vertical por razones de espacio.

omm_id	fecha	tmax	tmin	tmed	td	pres_est	Pres_nm	prcp	hr	helo	nub	vmax.d	vmax.f	vmed	Num_observaciones
87444	2002-07-31	11.4	2.2	6.1	-5.1	953.1	1027.8	0.001	48	8.7	1	\N	\N	7.4	4
87444	2002-08-01	12.6	-3.2	3.9	-8.1	952.8	1027.9	0	48	7.8	3	36	5.1	0.9	4
87444	2002-08-02	15.5	-8.8	4.3	-5.7	949.8	1025.2	23.2	56	8.9	1	36	4.6	0.9	4

Dado el largo de los datos requeridos para la creación de la primera versión de la base de datos, se sugiere que los datos históricos (1961 al presente) se organicen en un archivo de transferencia separado para cada estación meteorológica. Las actualizaciones (o sea, datos posteriores a la carga inicial de

---

datos), sin embargo, pueden incluir los datos para todas las estaciones dentro de un único archivo de transferencia – pero con el mismo formato.

## 2.2 Transferencia de datos climáticos a servidor FTP en CRC-SAS

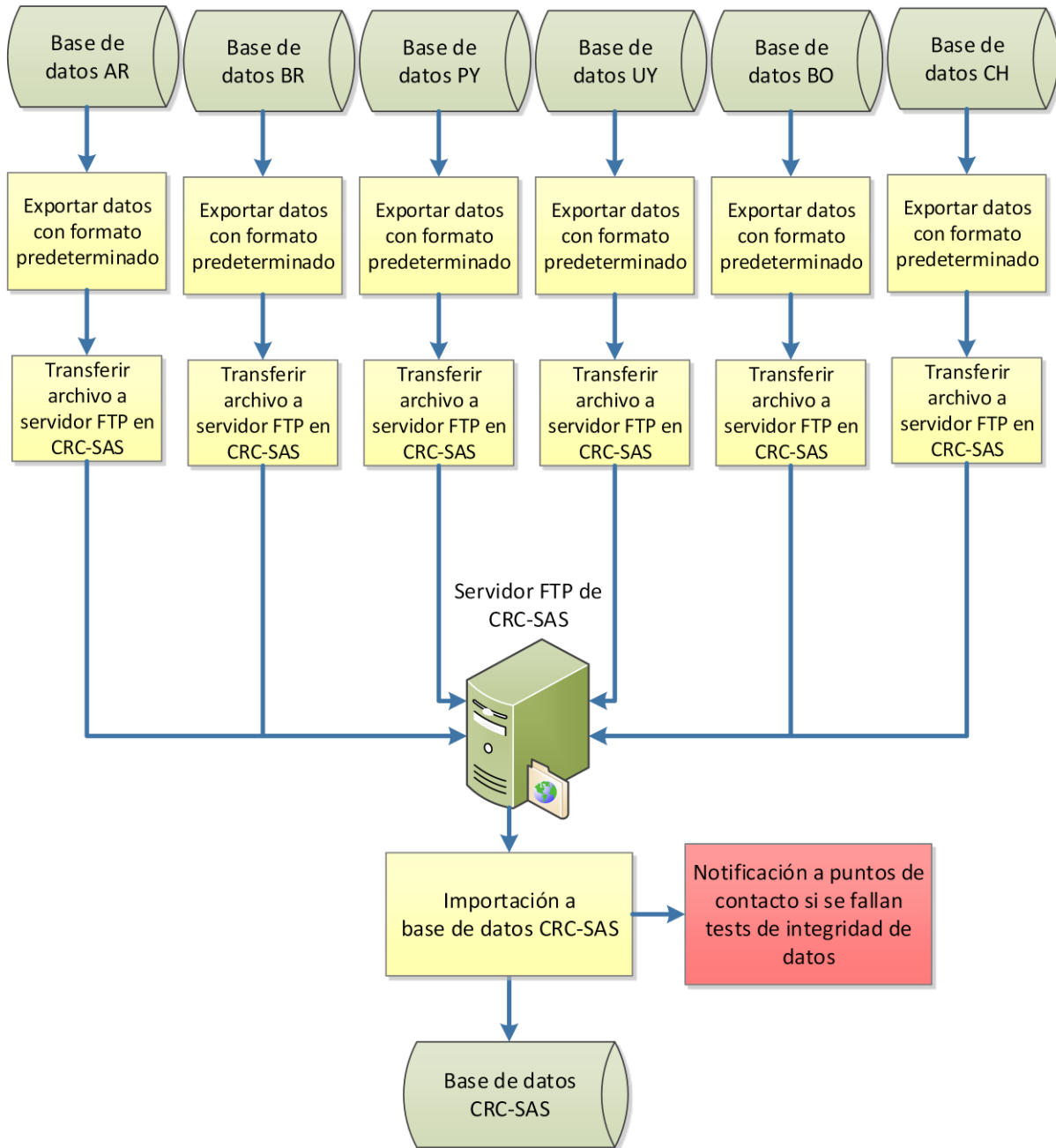
Los archivos de transferencia de datos climáticos preparados por cada país (ya sea con los datos históricos de 1961 al presente o actualizaciones posteriores) deberán ser copiados a un sitio ftp del CRC-SAS. La dirección del sitio FTP es [incluir en documento final]. Para realizar la transferencia cada miembro del Centro utilizará un nombre de usuario y contraseña a ser provisto por el CRC-SAS.

Un script de importación se activará periódicamente y listará archivos de transferencia en el sitio FTP que no se hayan procesado anteriormente. Este script realizará algunos controles de integridad de los datos – como verificar que no haya fechas duplicadas o que todos los datos importados para una estación tengan el mismo código de identificación – e importará los datos a la base de datos relacional del CRC-SAS. En caso de que haya problemas en el formato o contenido de los archivos de transferencia, se enviará en forma automática un mensaje a la casilla de correo electrónica del contacto designado por cada país para asuntos relacionados con la transferencia de datos.

Cada registro en la base de datos corresponde a una combinación única de una fecha y un ID de estación meteorológica. En algunos casos, los datos meteorológicos de una estación solamente incluyen aquellas fechas para las cuales haya datos. Por ejemplo, si la estación no ha observado datos entre el 25 y el 31 de agosto de 1962, los datos incluirán una fila para el 24 de agosto y, a continuación, una fila para el 1 de septiembre. Durante el proceso de importación, se completarán las series de modo de que haya una secuencia temporal completa. Para los registros que se agreguen, los valores de todas las variables (excepto fecha e ID de estación) se llenarán con datos nulos o faltantes. Siguiendo con el ejemplo anterior, en la base de datos se incluirán registros para las fechas del 25 al 31 de agosto de 1962. Todos los datos contendrán el valor “\N”.

Como parte del proceso de importación, se asigna una etiqueta “pendiente” a cada registro recientemente agregado a la base. De esta manera, la base de datos identifica los registros que deben pasar por los controles de calidad, que se realizan sólo una vez para cada registro. El proceso de transferencia de los datos se ilustra en la Figura 3.

Figura 3. Transferencia de datos meteorológicos diarios de los países miembros a la base de datos del CRC-SAS. El formato del archivo (o archivos) de transferencia se describe en el texto. Cada miembro del CRC transferirá esos archivos a un sitio ftp, de donde serán importados a la base de datos del Centro.



---

## 2.3 Procedimientos para identificación de datos climáticos sospechosos

El siguiente paso es la realización de un conjunto de controles de calidad que se aplican a los datos diarios de cada estación meteorológica. Esta etapa será realizada en forma centralizada por el CRC-SAS. Sin embargo, si los miembros del Centro y otros investigadores o institución de la región tienen interés en aplicar estos controles a sus propios datos, más adelante en el documento se dan instrucciones para realizar los controles fuera del ambiente computacional del CRC-SAS.

### 2.3.1 Familias de controles de calidad

Aunque los detalles de los diferentes controles se discuten en otro documento, se presenta aquí una breve descripción. Los diferentes controles se extrajeron de varias publicaciones en la literatura científica; los detalles están en el documento citado arriba. Los controles de calidad de datos meteorológicos diarios se organizaron en seis grandes grupos o “familias”:

- a. Controles generales. Estos controles verifican la integridad general de los datos. Por ejemplo, se controla que no haya fechas duplicadas o fuera de secuencia en las observaciones diarias. Otra verificación que se realiza es la frecuencia con la cual se registran los valores decimales para cada variable. Desvíos muy marcados con respecto a una distribución aproximadamente uniforme de valores decimales de 0 a 9 (en el caso de las temperaturas, que se registran con un solo decimal) pueden alertar sobre la existencia de problemas potenciales en los datos.
- b. Controles de rango fijo. Estos controles aseguran que no existan valores físicamente no realistas. Por ejemplo, una temperatura máxima diaria de 99°C está por encima del record mundial. Es posible que un valor así corresponda a un código de valor faltante que no se ha definido apropiadamente como tal.
- c. Controles de rango variable. En esta familia de controles, los rangos o umbrales usados para “marcar” valores sospechosos varían a lo largo del año, por lo que son más finos que en los tests de rango fijo. Por ejemplo, se puede ajustar un ciclo anual a los valores de temperatura mínima diaria y los valores extremos se evalúan con respecto al valor esperado del ciclo anual para una fecha determinada.
- d. Controles de continuidad temporal. Estos controles detectan la presencia de saltos o picos inusuales en las series de datos. Por ejemplo, un valor de temperatura media muy bajo en relación a los valores de los días adyacentes (el anterior y el siguiente) puede ser marcado como “sospechoso” por este test.
- e. Controles de consistencia entre variables. Una serie de controles en esta familia o grupo evalúan la consistencia entre valores de variables que deben guardar cierta consistencia. Un ejemplo obvio es la verificación de que la temperatura mínima diaria sea menor o igual que la temperatura máxima diaria.
- f. Controles de consistencia espacial. Todos los controles descritos anteriormente se realizan sobre los datos de una única estación meteorológica (aunque en algunos controles se use más de una variable). En este grupo de tests, sin embargo, los valores de una variable para una estación

---

determinada (que generalmente se denomina la “estación central”) se comparan con los valores de esa variable para estaciones geográficamente cercanas (o “estaciones vecinas”).

### 2.3.2 Resultados de los controles de calidad

El resultado de esta etapa es una serie de datos marcados como “sospechosos”, que a continuación deben verificarse en forma manual usando como fuente los registros oficiales. Es decir, los controles de calidad no intentan “corregir” los datos, sino que se limitan a identificar valores que pueden ser errores verdaderos o no.

Etiquetas de registro. Antes del control de calidad, los registros recientemente importados tienen una etiqueta que indica su estado “pendiente” (o sea, que debe pasar por los controles). Una vez que un registro pasa por el control de calidad, se le asigna una nueva etiqueta cuyo valor depende del resultado de los controles de calidad. Las etiquetas asignadas a cada registro después de los controles de calidad pueden tomar los valores:

- “Dudoso” en el caso de que al menos una variable en el registro falle algún control de calidad; o
- “Validado” si todas las variables en un registro pasan todos los controles de calidad.

Etiquetas de variables. Además de las etiquetas que indican el estado de todo un registro completo, se asignan etiquetas a cada variable en un registro dentro de la base de datos. Después del control de calidad (pero antes de la verificación manual de los datos dudosos, descrita en la sección siguiente), una variable en un registro puede tomar los valores:

- “Faltante” si el valor de la variable no está disponible;
- “Sospechoso” si la variable falla al menos uno de los controles de calidad; o
- “Aprobado” si la variable pasa exitosamente todos los controles.

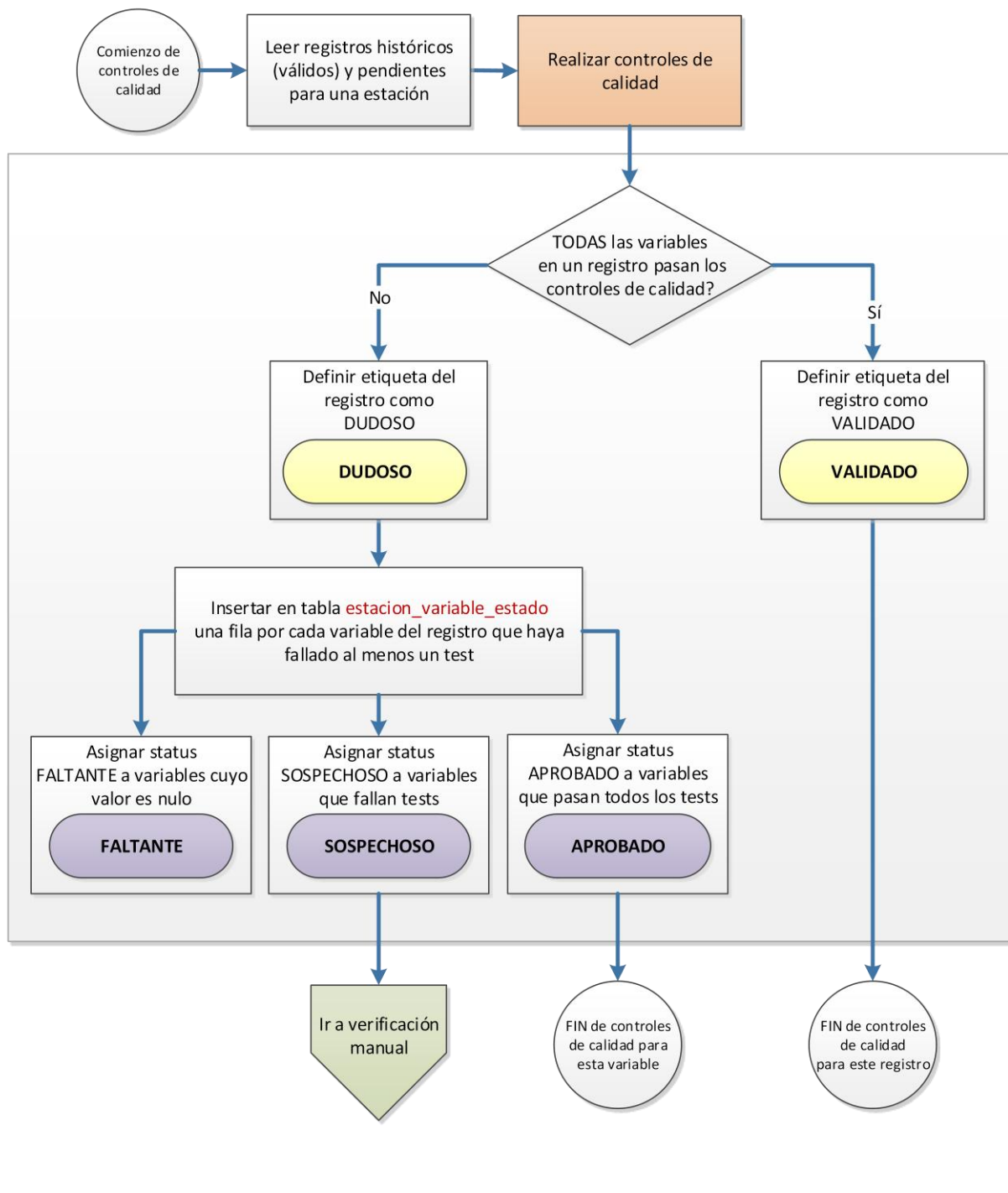
El flujo de asignación de etiquetas tanto a registros como a variables individuales se ilustra en la Figura 4.

### 2.3.3 Implementación de los controles de calidad

Todos los controles de calidad se implementaron en el lenguaje estadístico R (R Core Team, 2013). R es un software abierto y sin costo, lo que facilita la diseminación y utilización sin limitaciones de todos los controles implementados. El lenguaje R es sumamente flexible y potente; además, existen más de 5000 paquetes que expanden la funcionalidad del lenguaje. Estos paquetes, contribuidos por la comunidad global de usuarios de R, ahorran tener que programar todos los cálculos deseados. Por ejemplo, existen al menos dos paquetes que permiten calcular el largo del día (o sea, el número máximo de horas de sol) en función de la latitud y día del año.



Figura 4. Flujo de asignación de etiquetas de estado a registros y variables individuales durante la etapa de controles de calidad.



---

Aunque la ejecución de los controles de calidad destinados a identificar datos climáticos sospechosos se realizará en servidores del CRC-SAS (ver Figura 1), los procedimientos de control implementados se distribuirán a todos los miembros del CRC-SAS y a la comunidad de las ciencias climáticas en general.

## 2.4 Verificación manual de datos sospechosos

Los controles de calidad descritos en la sección anterior no intentan corregir automáticamente datos marcados como sospechosos. Los controles simplemente generan listas de posibles errores que en esta etapa deben ser verificadas manualmente. La verificación involucra la comparación de los valores sospechosos con la versión “oficial” de estos valores, por ejemplo las libretas donde los observadores de cada estación meteorológica registran los datos. Esta verificación será realizada por personal de los servicios meteorológicos e hidrológicos (SMHs) de cada país miembro del CRC-SAS, ya que las fuentes oficiales de datos se encuentran en cada país.

Para facilitar la verificación de datos sospechosos, el CRC-SAS (en colaboración con un proyecto financiado por el Banco Interamericano de Desarrollo) está desarrollando un procedimiento para que el personal de los SMHs miembros pueda acceder remotamente (vía un *browser*) a la base de datos del CRC-SAS. Este procedimiento todavía está en desarrollo, pero esencialmente se espera que se muestre una pantalla con cada registro “dudoso” (o sea, aquel para el cual por lo menos una variable haya fallado en algún test) y los registros para 5-6 días antes y después del registro examinado (para tener un contexto temporal). Para cada variable con valores sospechosos en un registro, el procedimiento también mostrará los valores de esa variable en las estaciones geográficamente adyacentes (lo que proporciona un contexto espacial para la interpretación de valores sospechosos). Se espera también que durante este procedimiento, el operador de un SMH tenga acceso a los registros meteorológicos oficiales. Un ejemplo preliminar del aspecto que tendrá una de las pantallas del sistema de verificación se muestra en la Figura 5.

Durante el proceso de verificación manual, un operador (a) selecciona una estación meteorológica, (b) recibe una pantalla con registros sospechosos que todavía no han sido verificados, (c) controla cada valor sospechoso contra los registros oficiales), y (d) selecciona un estado final para cada variable sospechosa en un registro. Una vez que el operador selecciona un estado final para todas las variables sospechosas en un registro, la base de datos actualiza las etiquetas para ese registro, que pasa de “pendiente” a “revisado.”. También se actualizan las etiquetas correspondientes a cada variable en el registro. De acuerdo a la decisión del operador, la etiqueta de una variable puede tomar diferentes valores:

- “Ratificado” en el caso de que el valor sospechoso de la variable se haya confirmado usando los registros oficiales;
- “Corregido” si el valor sospechoso de la variable se corrigió en base a los registros oficiales (por ejemplo, una temperatura máxima de “41°C” en julio se cambia a “14°C”, sugiriendo una trasposición de dígitos);
- “Eliminado” en el caso de que un valor sospechoso no puede corregirse usando los registros oficiales, pero hay elementos para concluir que el valor es claramente erróneo;

- “No corregible” cuando (a) un valor sospechoso no puede corregirse usando los registros oficiales, pero (b) NO hay elementos suficientes para concluir que el valor es erróneo (o sea, el valor es plausible).

## 2.5 Actualización de la base de datos del CRC-SAS

El proceso de verificación manual de datos sospechosos actualiza automáticamente la base de datos del CRC-SAS cada vez que el operador de cada SMH pasa de un registro sospechoso al siguiente. Al terminar de verificar cada registro, las etiquetas de ese registro y de cada variable en el registro se actualizan automáticamente de acuerdo a las opciones elegidas por el operador. Una vez completada la verificación manual de un registro, la base de datos le asigna la etiqueta “Revisado.” El flujo de asignación de etiquetas de estado a registros y variables individuales durante la etapa de verificación manual de datos se ilustra en la Figura 6. Los valores posibles de las etiquetas para las variables en cada registro se discutieron en la sección anterior.

Figura 5. Ejemplo de una de las pantallas del sistema de verificación manual de datos sospechosos. Personal de cada país miembro accederá a la base de datos del CRC y controlará los valores considerados sospechosos utilizando registros. El registro siendo analizado se resalta, y las variables se colorean de acuerdo a los resultados de los controles de calidad (por ejemplo, en amarillo se indican los valores sospechosos). El mapa a la derecha de la pantalla proporciona contexto espacial, mostrando valores en estaciones vecinas (capacidad a implementar).

Centro Regional del Clima para el Sur de América del Sur  
Centro Regional do Clima para o Sul da América do Sul  
Fase demostración

Estaciones - [87544] Pehuajó Aero (Buenos Aires, Argentina) - Institución: SMN

Fecha	tmax	tmin	tmed	td	pres_est	pres_nm	prcp	hr	helio	nub	vmax_d	vmax_f	vmed
1961-01-01	37.5	12.3	24.3	9.8	991.2	NA	1.2	45	NA	3	NA	NA	5.4
1961-01-02	29.7	16.9	22.7	6.0	996.6	NA	0.0	42	NA	5	NA	NA	5.8
1961-01-03	30.6	17.3	23.2	4.5	998.6	NA	0.0	32	NA	3	NA	NA	2.2
1961-01-04	32.5	15.2	23.8	10.5	995.9	NA	0.0	47	NA	4	NA	NA	3.3
1961-01-05	26.4	18.4	21.6	13.3	992.0	NA	0.0	60	NA	6	36	10.3	4.1
1961-01-06	26.3	8.3	19.3	6.0	998.0	NA	0.0	47	NA	0	18	19.5	5.1
1961-01-07	34.5	11.9	24.5	8.3	994.5	NA	0.0	37	NA	3	36	13.4	5.7
1961-01-08	35.0	17.2	26.7	15.5	993.8	NA	0.0	53	NA	5	18	12.3	3.5

Referencia de colores:

- C Registro actual / Estación central (en mapa)
- Otro registro / Estación vecina (en mapa)
- Dato válido
- Dato sospechoso
- Dato no válido / no corregible

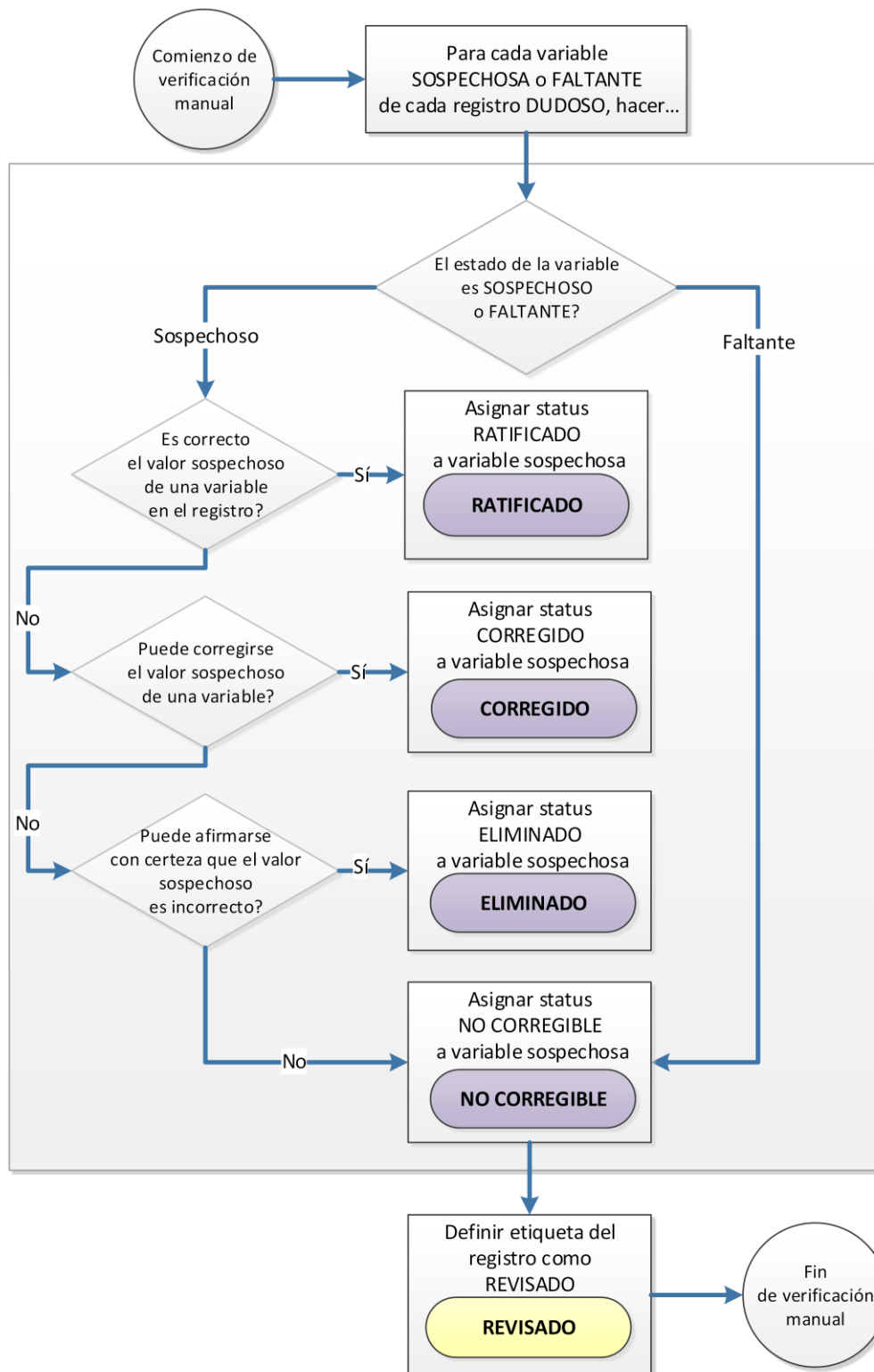
Este proyecto es financiado por las siguientes organizaciones

BID Banco Interamericano de Desarrollo

IAI Inter-American Institute for Global Change Research

NSF United States National Science Foundation

Figura 6. Flujo de asignación de etiquetas de estado a registros y variables individuales durante la etapa de verificación manual de datos.



---

Un aspecto importante del proceso de control y actualización de datos es que los valores sospechosos que se confirmen como erróneos se guardan en la base de datos en tablas separadas. Es decir, si los registros oficiales indican que el valor registrado para una variable es incorrecto y debe reemplazarse, el valor originalmente almacenado en los datos se preserva, de modo de poder reconstruir los datos originales si fuera necesario.

### 3 Referencias

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, <http://www.R-project.org>.

### 4 Agradecimientos

Las actividades detalladas en este reporte fueron financiadas por el Banco Interamericano de Desarrollo (BID) a través del contrato C0121-13 con la Universidad de Miami como parte del proyecto *“Hydro-climate Services in La Plata River Basin.”* Otros fondos fueron provistos por el Instituto Inter-Americano para el Estudio del Cambio Global (IAI) mediante el proyecto CRN-035, *“Towards usable climate science: informing sustainable decisions and provision of climate services to the agriculture and water sectors of southeastern South America.”* El IAI está financiado por la Fundación Nacional de Ciencias de los Estados Unidos de Norteamérica (NSF) a través del grant GEO-1128040. Finalmente, uno de los autores (GP) fue apoyado económicamente por la Fundación Nacional de Ciencias de los Estados Unidos de Norteamérica a través del grant 1049109 del programa *“Decadal and Regional Climate Prediction using Earth System Models (EaSM)”*.

---

## Apéndice A

### A.1 Variables meteorológicas incluidas en la base de datos del Centro Regional del Clima para el Sur de América del Sur

Tabla A-1. Listado de variables meteorológicas a incluir en la base de datos del CRC-SAS. El número de observaciones diarias utilizado para calcular cantidades agregadas para un día (por ejemplo, temperatura media diaria o velocidad media del viento) se indica en la última columna.

<i>Columna</i>	<i>Variable</i>	<i>Nombre Abreviado</i>	<i>Unidades</i>
1	Temperatura máxima diaria	tmax	grados Celsius (°C)
2	Temperatura mínima diaria	tmin	grados Celsius (°C)
3	Temperatura media diaria	tmed	grados Celsius (°C)
4	Temperatura de rocío	td	grados Celsius (°C)
5	Presión atmosférica al nivel de la estación	pres_est	hectopascales (hPa)
6	Presión atmosférica reducida al nivel medio del mar	pres_nm	hectopascales (hPa)
7	Precipitación acumulada	prcp	milímetros (mm)
8	Humedad relativa	hr	porcentaje (%)
9	Horas diarias de sol (heliofanía)	helio	horas
10	Cobertura nubosa	nub	octavos
11	Dirección del viento máximo diario	vmax_d	decenas de grado (1,...,36)
12	Velocidad del viento máximo diario	vmax_f	metros por segundo ( $m s^{-1}$ )
13	Velocidad media del viento	vmed	metros por segundo ( $m s^{-1}$ )
14	Número de observaciones diarias usadas para calcular cantidades promediadas	num_observaciones	sin unidades



---

## A.2 Formato de los archivos de transferencia de datos climáticos

Como se ha discutido en el texto, los miembros del CRC-SAS contribuirán datos de variables meteorológicas archivadas a nivel diario a la base de datos del CRC. Estos datos se importarán a través de (a) archivos de texto organizados por estación meteorológica o, alternativamente, (b) un archivo único con datos de varias estaciones meteorológicas. Esta última opción quizás sea más recomendable para actualizaciones periódicas de datos, mientras que la importación de registros históricos largos (durante la carga inicial de la base de datos del CRC) será más fácil si los archivos están separados por estación. En cualquier caso, el formato de los archivos de texto debe ser el mismo.

A continuación se dan algunos detalles del formato que los países deberían observar para minimizar los errores en la importación de datos.

- g. Si se transfieren datos para una única estación meteorológica, el nombre del archivo debería ser el ID internacional de la estación y la extensión “.csv” (aunque las columnas del archivo estén separados por tabulaciones). Por ejemplo el archivo “87544.csv” contendrá datos para la estación cuyo identificador es 87544 (Pehuajó, Argentina).
- h. Si, en cambio, se transfieren datos para múltiples estaciones, por ejemplo como parte de una actualización periódica de datos, entonces el archivo debe tener el nombre abreviado de la institución y la extensión “.csv”. Por ejemplo, un archivo que contenga los datos transferidos por el Servicio Meteorológico Nacional de Argentina para los últimos 10 días debería llamarse “SMN.csv”. Los nombres abreviados para otras instituciones son “INMET” para Brasil, “DMH” para Paraguay, “INUMET” para Uruguay y “SENAMHI” para Bolivia.
- i. El archivo debe contener un encabezado (o sea, la primera línea en el archivo) con los nombres abreviados de las variables en cada columna en el orden citado en la Tabla A-1. El encabezado se usa para verificar la integridad de los datos durante la importación a la base de datos relacional, por lo que es necesario respetar estos nombres y su orden. Los nombres a utilizar en el encabezado se muestran en la columna “Nombre abreviado” de la Tabla A-1 (“omm\_id”, “fecha”, etc.). Es indistinto usar mayúsculas o minúsculas en el encabezado: “FECHA” y “fecha” son ambos válidos.
- j. Cada fila subsiguiente en el archivo representa las observaciones para un día y una estación meteorológica.
- k. Cada columna corresponde a una variable. Las columnas deben estar separadas por una tabulación (tab o “\t”). El orden de las columnas debe ser el indicado en la Tabla A-1. Aunque no se incluyan todas las variables listadas en la Tabla A-1 (por ejemplo, algunos miembros solamente transfieren datos de temperaturas máxima y mínima y precipitación), todas las columnas sin datos deben llenarse con valores faltantes (ver ítem siguiente).
- l. Los valores faltantes para una variable se deben indicar como “\N” (sin las comillas), que será interpretado por la base de datos como un valor NULL o faltante. También es válido dejar en blanco (vacío) el valor faltante de una variable. En el archivo de texto, un espacio vacío se indica como dos tabulaciones consecutivas “\t\t” (que separan la columna anterior y la posterior a la



---

columna en blanco. Sin embargo, NO debe incluirse un espacio en blanco como dato faltante – es decir, debe evitarse “\t \t” (notar el espacio entre las dos tabulaciones).

- m. Debe utilizarse el punto “.” (no la coma “,”) como separador de decimales en los valores de las variables. No debe usarse coma para separar miles (por ejemplo, NO usar “1,013.7” sino “1013.7”).
- n. Las fechas pueden listarse (a) en formato ISO 8601, o sea, el año, mes y día están separados por un guión medio (AAAA-MM-DD); o (b) como día, mes y año separados por una barra (DD/MM/AAAA).

---

## Apéndice B

### B1. Información incluida en la versión *preliminar* de los metadatos para cada estación meteorológica en la base de datos del Centro Regional Climático para el Sur de América del Sur

Tabla B-1. Listado de datos a incluir en el archivo de metadatos para cada estación meteorológica a incluirse en la base de datos del CRC-SAS.

<b>Nombre abreviado</b>	<b>Descripción</b>
omm_id	Número internacional de la estación (código OMM)
nombre	Nombre de la estación
lon_grad	Longitud (grados sexagesimales)
lon_min	Longitud (minutos sexagesimales)
lon_seg	Longitud (segundos sexagesimales)
lon_hem	Hemisferio longitudinal (E/O)
lat_grad	Latitud (grados sexagesimales)
lat_min	Latitud (minutos sexagesimales)
lat_seg	Latitud (segundos sexagesimales)
lat_hem	Hemisferio latitudinal (N/S)
lon_dec	Longitud (grados decimales)
lat_dec	Latitud (grados decimales)
elev	Elevación sobre el nivel del mar (m)
pais_iso2c	Código alfabético del país (estándar ISO 3166-1 alpha-2)
nivel_adm_1	Nivel administrativo 1 (estado, provincia)
nivel_adm_2	Nivel administrativo 2 (departamento, partido)
fecha_inicio	Fecha de inicio de operación de la estación
fecha_fin	Fecha de fin de operación de la estación
activa	La estación está activa en la actualidad? (sí/no)
operador	Código de institución que opera la estación (ej., "1" por INMET)
id_interno	Número interno de la estación dentro de la organización
tipo_estacion	Código de tipo de estación meteorológica (C: convencional, A; automática, etc.)

---

Algunas consideraciones sobre el archivo de metadatos:

- El archivo de metadatos debe tener codificación UTF-8 de modo de poder almacenar acentos y otros caracteres especiales;
- Valores negativos de latitud y longitud decimal indican, respectivamente hemisferios S y W;
- La fecha de inicio de las observaciones en una estación meteorológica no necesariamente debe coincidir con la primera fecha para la cual se transfieren datos. La estación puede haber comenzado a funcionar en 1947, por ejemplo, pero se contribuyen datos desde principios de 1961; y
- La fecha de fin de las observaciones sólo debe llenarse en el caso de estaciones meteorológicas que hayan dejado de tomar observaciones. Esta fecha indica cuándo ha dejado de operar una estación meteorológica. Por lo tanto, para estaciones que estén en actividad este campo debe estar vacío (o lleno con “\N”). Este campo NO debe llenarse con la fecha de la última observación transferida a la base de datos del CRC.