



CAPITULO 4 BASE DE DATOS

| | | |
|--------------|---|---------|
| 4.1 | INTRODUCCIÓN | pág. 69 |
| 4.2 | DEFINICIÓN Y ADECUACIÓN DEL CONTENIDO DE LA BASE DE DATOS | pág. 69 |
| 4.2.1 | Datos diarios y mensuales observados | pág. 69 |
| 4.2.2 | Índices de extremos de precipitación y temperatura basados en datos observados | pág. 72 |
| 4.2.3 | Datos simulados de modelos climáticos | pág. 72 |
| 4.3 | DEFINICIÓN Y ADECUACIÓN DEL FORMATO DE LOS DATOS | pág. 72 |
| 4.4 | DESARROLLO DE LA ARQUITECTURA DE LA BASE DE DATOS | pág. 73 |
| 4.4.1 | Especificación detallada | pág. 73 |
| 4.4.2 | Arquitectura de la Base de Datos | pág. 74 |
| 4.5 | INTERFAZ DE VISUALIZACIÓN WEB DE LA BASE DE DATOS | pág. 78 |
| 4.6 | DESCRIPCIÓN DE LOS ARCHIVOS DE METADATOS | pág. 82 |
| 4.7 | DOCUMENTOS EXPLICATIVOS DE LA BASE DE DATOS | pág. 83 |
| 4.8 | IMPLEMENTACIÓN Y PUESTA A DISPOSICIÓN DE LA BASE DE DATOS | pág. 83 |



BASE DE DATOS

4.1 INTRODUCCIÓN

Este capítulo describe las características de la Base de Datos climáticos elaborada. La Base de Datos constituye un soporte informático que permite albergar y poner a disposición abierta y pública, información climática relevante para los estudios de impacto de cambio climático que se realizarán a futuro en el marco de la 3ra. Comunicación Nacional sobre Cambio Climático.

La Base de Datos incluye un conjunto de datos cuyos archivos correspondientes están descritos en la sección 4.2 así como la adecuación realizada a los mismos. La sección 4.3 define los datos y describe el proceso de adecuación de los formatos correspondientes mientras en el Anexo III se detallan los archivos recibidos u obtenidos de sus fuentes originales, que han ingresado a la Base de Datos. En la sección 4.4 presenta la arquitectura sobre la cual está siendo generada la Base de Datos, cuyo detalle se incluyen además en el Anexo IV. La interfaz de visualización web que permite a los usuarios consultar los archivos almacenados y sus metadatos, como así también visualizar mapas de localización de los datos, se presenta en la sección 4.5. La descripción general del contenido de los archivos de metadatos se incluye en la sección 4.6 y su detalle en el Anexo V mientras que la descripción de los documentos explicativos sobre la Base de Datos se presenta en la sección 4.7 y el detalle de los mismos en el Anexo VI. La información sobre la implementación y puesta a disposición de la Base de Datos mediante la inclusión de los conjuntos de datos se discute en la sección 4.8.

4.2 DEFINICIÓN Y ADECUACIÓN DEL CONTENIDO DE LA BASE DE DATOS

4.2.1 Datos diarios y mensuales observados

La Base de Datos contiene información de precipitación y temperatura provenientes de observaciones del clima reciente (1961-2010). Dichos datos están representados en un reticulado geo-referenciado de al menos 1 grado de latitud y longitud de resolución. Por las características de las regiones geográficas a cubrir por la base, los datos se agrupan teniendo en cuenta dos áreas principales: a) Argentina Continental, b) Antártida e Islas del Atlántico Sur.

a) Argentina Continental

Datos Mensuales

Se proveen datos de temperatura mínima, temperatura máxima, temperatura media y precipitación. Asimismo se proveen datos de temperatura de la superficie del mar que pueden ser de utilidad para diversos estudios de impacto. Las características de los conjuntos de datos seleccionados para ser incluidos en la Base de Datos son los siguientes:

CRU_TS 3.21:

VARIABLES: temperatura media, temperatura mínima, temperatura máxima, precipitación
Resolución: 0.5x0.5,



Frecuencia: Mensual,
Área: Global, Período: 1901-2012
URL: http://badc.nerc.ac.uk/browse/badc/cru/data/cru_ts/cru_ts_3.21/data

GPCC Versión 6.
Variables: Precipitación
Resolución: 1.0x1.0,
Frecuencia: Mensual,
Área: Global,
Período: 1901-2010
URL: ftp://ftp.dwd.de/pub/data/gpcc/html/fulldata_v6_doi_download.html

NOAA Optimum Interpolation Sea Surface Temperature (OI-SST V2).
Variable: Temperatura de la superficie del mar.
Resolución: 1.0x1.0,
Frecuencia: Mensual,
Área: Global,
Período 1981-2014
URL: <http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html>

Datos diarios:

Se proveen datos de temperatura mínima, temperatura máxima, temperatura media y precipitación. Asimismo se proveen datos de temperatura de la superficie del mar que pueden ser de utilidad para diversos estudios de impacto. Las características de los conjuntos de datos seleccionados para ser incluidos en la Base de Datos son los siguientes:

Clarís-LPB SAT 1.0 :
Variables: temperatura mínima, temperatura máxima
Resolución: 0.5x0.5,
Frecuencia: Diaria,
Área: Sudeste de Sudamérica (SESA, 20S-40S; 45W-75W)
Período: 1961-2000
URL: <http://wp32.at.fcen.uba.ar/gridded>

Clarís-LPB-rr-05.1:
Variables: precipitación
Resolución: 0.5x0.5,
Frecuencia: Diaria,
Área: Sudeste de Sudamérica (SESA, 20S-40S; 45W-75W)
Período: 1961-2000
URL: <http://wp32.at.fcen.uba.ar/gridded/prec>

NOAA/ESRL PSD SA
Variables: Precipitación,
Resolución: 1.0x1.0,
Frecuencia: Diaria,



Área: Sudamérica,
Período: 1940-2012

URL: ftp://ftp.cdc.noaa.gov/Datasets.other/south_america/sa24.daily.1.1940-2012.nc
http://www.esrl.noaa.gov/psd/data/gridded/data.south_america_precip.html

NCEP REANALYSIS I

Variables: temperatura mínima, temperatura máxima, temperatura media

Resolución: 1.875x1.915

Frecuencia: Diaria

Área: Sudamérica (20.5°S – 59.5°S, 79.5°W – 50.5°W)

Período: 1960-2010

URL: <http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.surfaceflux.html>

NOAA OI SST V2 HR

Variables: temperatura de superficie del mar

Resolución: 0.25x0.25,

Frecuencia: Diaria,

Área: Global,

Período: 1981-Presente

URL: <ftp://ftp.cdc.noaa.gov/Datasets/noaa.oisst.v2.highres/sst.day.mean.yyyy.v2.nc>

b) Antártida e Islas del Atlántico Sur:

Datos mensuales

Se proveen datos de temperatura mínima, temperatura máxima, temperatura media y precipitación. Las características de los conjuntos de datos seleccionados para ser incluidos en la Base de Datos son los siguientes:

NCEP REANALYSIS I

Variables: temperatura mínima, temperatura máxima, temperatura media, precipitación

Resolución: 1.875x1.915

Frecuencia: Mensual

Área: (60°S – 89°S, 25°W – 74°W)

Período: 1960-2010

URL: <http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.surfaceflux.html>

Datos diarios

Se proveen datos de temperatura mínima, temperatura máxima, temperatura media y precipitación. Las características de los conjuntos de datos seleccionados para ser incluidos en la Base de Datos son los siguientes:

NCEP REANALYSIS I

Variables: temperatura mínima, temperatura máxima, temperatura media, precipitación

Resolución: 1.875x1.915

Frecuencia: Diaria



Área: (60°S – 89°S, 25°W – 74°W)

Período: 1960-2010

URL: <http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.surfaceflux.html>

4.2.2 Índices de extremos de precipitación y temperatura basados en datos observados

Se proveen datos grillados de los 11 Índices de extremos de precipitación y temperatura descriptos en la sección 2.1.2 del Capítulo 2. Las características de los conjuntos de datos seleccionados para ser incluidos en la Base de Datos son los siguientes:

CLIMDEX HadEX2

Variables: FD, TR, TX_x, TN_n, TX90p, WSDI, Rx1, Rx5, CDD, R95pT

Resolución: 3.75x2.5

Frecuencia: Dependiente del índice

Área: Global

Período: 1960-2010

URL: <http://www.metoffice.gov.uk/hadobs/hadex2/download.html>

4.2.3 Datos simulados de modelos climáticos

La base incluye también los datos de precipitación y temperatura del clima del pasado reciente (1961-2010) y del clima futuro (2015-2039, 2075-2099) simulados por los modelos climáticos. Están a disposición los datos corregidos provenientes de todos los modelos climáticos que hayan sido seleccionados en el proceso de validación. Los datos simulados están disponibles en un reticulado geo-referenciado de 0.5 grados de latitud y longitud de resolución. Se proveerán los datos promedio del ensamble multi-modelo como de cada uno de los modelos individuales. Los datos climáticos simulados se proveen en resoluciones temporales diarias y mensuales. Los mismos cubren espacialmente la totalidad de Argentina subdividida en las 4 regiones de estudio. El listado de los modelos a incluir se encuentra en la sección 3.3.6 del Capítulo 3.

4.3 DEFINICIÓN Y ADECUACIÓN DEL FORMATO DE LOS DATOS

Los archivos de datos dentro del sistema están almacenados en formato NetCDF. Este formato facilita que los archivos grillados estén organizados espacial y temporalmente según las dimensiones fijadas en la metadata. Los archivos de datos están organizados según:

-tipo de dato: observados grillados; simulaciones; índices.

-período: presente; futuro cercano; futuro lejano.

-variable: temperatura media, temperatura máxima, temperatura mínima, precipitación, temperatura de superficie del mar.

-frecuencia: mensual, diaria.

Tratamiento previo:

La estructura interna que presentan los diferentes archivos que conforman la base no es homogénea. En consecuencia para poder procesar los archivos mediante comandos de



extracción automatizada se elabora una tabla con los parámetros necesarios, cuya descripción se encuentra en el Anexo IV.

La mayoría de los archivos, provistos por diversas fuentes, son de cobertura global. Por esta razón los archivos originales han sido recortados en espacio y tiempo para adecuar al área y tiempo de estudio. De esa forma, se evita un mal uso de espacio e incrementos de procesamiento innecesarios.

Los datos grillados observados se almacenaron en sus grillas espaciales originales. Mientras que los datos grillados simulados provenientes de los modelos climáticos se grillaron en una resolución de medio grado (0.5x0.5).

Armado de archivos de datos grillados:

Para cada archivo de datos (de ahora en más llamado “recurso”) que se almacenó en la Base de Datos, se requirió la definición de una tabla conteniendo las coordenadas geográficas de los puntos (latitud-longitud) contenidos en los archivos, asociando país, región, provincia. Para este proceso se diseñó un script, que recupera las áreas definidas en shapes de país-provincia, donde a su vez se integró las regiones definidas en el proyecto. Para cada punto en un archivo dado, el script realiza la búsqueda dentro de las shapes obteniendo la provincia, región, y país de cada punto. Esta información se almacena en la Tabla Provincias-Grilla, que se utiliza en la visualización de los mapas y en la exportación de los Datos.

4.4 DESARROLLO DE LA ARQUITECTURA DE LA BASE DE DATOS

4.4.1 Especificación detallada

La extensión espacial de los datos se circunscribe a aquella de la República Argentina incluyendo la Antártida e Islas del Atlántico Sur. Debido a su extensión territorial, la visualización de los mapas, se divide en “Continental” (20S-60S 80W-50W) y “Antártida e Islas del Atlántico Sur” (sur de 60°S entre 74W-25W). Los archivos de datos simulados validados y corregidos para el proyecto corresponden a la Argentina en su zona continental. Asimismo Argentina Continental es dividida en 4 regiones:

- Húmeda,
- Centro,
- Andes
- Patagonia.

Cada región está conformada por un conjunto de provincias.

-Húmeda: Buenos Aires, Santa Fe, Entre Ríos, Corrientes, Misiones

-Centro: Formosa, Chaco, Santiago Del Estero, Córdoba, San Luis, La Pampa

-Andes: Jujuy, Salta, Catamarca, La Rioja, San Juan, Mendoza

-Patagonia: Neuquén, Río Negro, Chubut, Santa Cruz, Tierra del Fuego e Islas Malvinas.

(La zona Antártida e Islas del Atlántico Sur no tiene sub regiones).

La Base de Datos cubre los siguientes períodos:

- Pasado reciente (1961-2010), datos grillados observados y simulados.



- Futuro cercano (2015-2039), solo simulaciones de los modelos seleccionados.
- Futuro lejano (2075-2099), solo simulaciones de los modelos seleccionados.

Asimismo se proveen algunos conjuntos de datos observados que cubren períodos temporales más cortos que el definido. Se decidió su provisión de manera de aumentar la oferta de datos.

Como se mencionó en la sección 4.3, los conjuntos de datos observados seleccionados para almacenar en la Base de Datos, están generados en grillas de distinta resolución que no coinciden entre sí. Los datos son almacenados en sus grillas originales y se provee para cada conjunto, la descripción de la grilla en la cual son representados. Cuando más de una resolución esté disponible, se elige para su almacenamiento aquella más cercana a un grado de resolución. Las simulaciones corregidas para cada región, se proveen en una grilla de medio grado de resolución (0.5x0.5).

Para la zona Antártida e islas del Atlántico Sur se hacen disponibles datos diarios y mensuales de provenientes de 9 modelos globales para todos los escenarios y variables solicitadas. Ninguno de los modelos regionales disponibles tiene un dominio que abarque esta zona. Con respecto a los índices, no se incluyen para esta zona dado que los índices calculados desde observaciones no incluyen a la Antártida. Por otra, porque los modelos si bien son globales la calidad de su información a esas latitudes tan altas es limitada como para justificar la calidad de los índices, y además porque la mayoría de los índices no tienen sentido práctico sobre la Antártida.

4.4.2 Arquitectura de la Base de Datos.

Esta sección describe en detalle las facilidades que la Base de Datos brinda y la arquitectura que se diseñó para su generación.

a) Diseño Conceptual:

El sistema maneja una cantidad de archivos, definida por las siguientes especificaciones:

- Datos Observados: 2 zonas, 1 período, 5 variables, 2 frecuencias.
- Datos simulados: 3 períodos, 9 modelos, 3 escenarios, 4 variables, 2 frecuencias.
- Datos de Índices: 1 zona, 3 períodos, 11 índices, 1 frecuencia.

Las simulaciones de tiempo presente cubren 50 años, se almacena en 2 archivos de 25 años cada una, esto reduce el tamaño de los archivos, que en la frecuencia diaria es crítico. Todos los archivos son indexados por la Base de Datos, que genera una tabla relacional, que contiene todas las propiedades de los archivos de manera de poder vincularlos y recuperarlos, según las consultas realizadas.

Los parámetros para los criterios de búsquedas que deben ser definidos para indexar las tablas son: zona, período, variable, frecuencia.

La Base de Datos incluye:

- los metadatos de los archivos,
- los descriptores necesarios para resolver las consultas impuestas,



- los descriptores de los archivos NetCDF, necesarios para extraer los datos con comandos automatizados,
- las grillas de latitud-longitud de cada archivo,
- las shapes de la Argentina

b) Diseño Físico:

La implementación física del modelo lógico arriba descrito se realiza centralizando toda la información de cada archivo en una tabla, denominada “Recurso”, esta tabla contiene todas las propiedades de los recursos. Cada recurso tiene un conjunto de posiciones latitud-longitud que georreferencian la variable que contiene.

Además se define una tabla “Grilla” que almacena las coordenadas, pares latitud-longitud, asociados a cada recurso, que incluye la provincia y país. Se define una tabla “NetCDFFile” que almacena los todos descriptores de los archivos NetCDF necesarios para la extracción de datos. Se define una tabla “Argentina” que contiene las shapes de la Argentina, provincias, y continente Antártico.

Se normaliza la tabla Recurso, se generan entonces las tablas asociadas: zona, región, variable, frecuencia, institución, autor. Se definen las relaciones entre las tablas que se muestra en el diagrama Entidad-Relación. Se definen los índices en las tablas para acelerar las búsquedas.

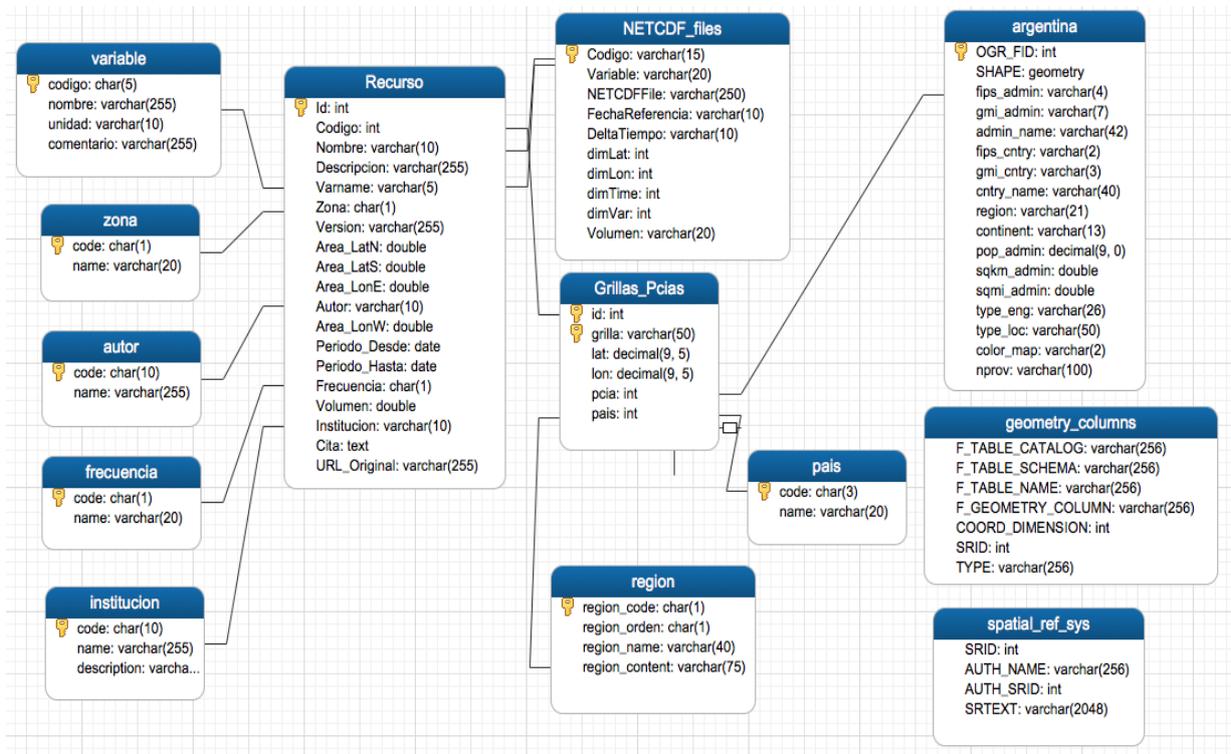
En el Anexo V se detalla la descripción de la tabla “Recurso”, las tablas asociadas resultado de la normalización de la Base de Datos, y las tablas de trabajo resultado del pre-procesamiento de datos para armar las formas, definir las regiones incluyendo las provincias contenidas y definir las grillas de los recursos.

c). Modelo Entidad-Relación:

El grafico que se incluye a continuación representa las entidades relevantes del sistema de Base de Datos así como sus interrelaciones y propiedades.



Figura 4.4.2.1 DIAGRAMA ENTIDAD-RELACION



d) Tecnologías utilizadas.

El sistema está desarrollado utilizando herramientas de arquitectura abierta (Open Source). Funciona en plataforma Linux con un **servidor web Apache2**. Para alojar la Base de Datos relacional se utiliza **MySQL** y los lenguajes de programación y desarrollo utilizados son **PHP, SQL, JavaScript** (Ajax, **Jquery** y **Herramientas de desarrollo de Google Map** para mostrar los mapas). Los archivos en el sistema están almacenados internamente en formato **NetCDF**. El protocolo utilizado para la transferencia de hipertexto es **HTTP** (HiperText Transfer Protocol) que está basado en el envío de mensajes y establece el conjunto de normas mediante las cuales se envían peticiones de acceso y respuestas de una web, perteneciente a la capa de aplicación del modelo OSI. El **modelo de interconexión de sistemas abiertos OSI** (en inglés, *Open System Interconnection*) es el modelo de red descriptivo, que fue creado por la Organización Internacional para la Estandarización (ISO) en el año 1980 (ISO/IEC 7498-1). Es un marco de referencia para la definición de arquitecturas en la interconexión de los sistemas de comunicaciones.

Motivación en la selección de MySQL: El servidor de bases de datos MySQL es muy rápido, seguro, y fácil de usar. El servidor MySQL fue desarrollado originalmente para manejar grandes bases de datos mucho más rápido que las soluciones existentes y ha estado siendo usado exitosamente en ambientes de producción sumamente exigentes por varios años. Aunque se encuentra en desarrollo constante, el servidor MySQL ofrece hoy un conjunto rico y útil de funciones. Su conectividad, velocidad, y seguridad hacen de MySQL un servidor apropiado para acceder a bases de datos en Internet.



El almacenamiento de los datos en formato netCDF

La figura 4.3.2.1 muestra la estructura de un archivo netCDF mediante la anotación CDL (Idioma del formulario de datos comunes en red). CDL es el formato ASCII que se utiliza para describir el contenido de un archivo netCDF

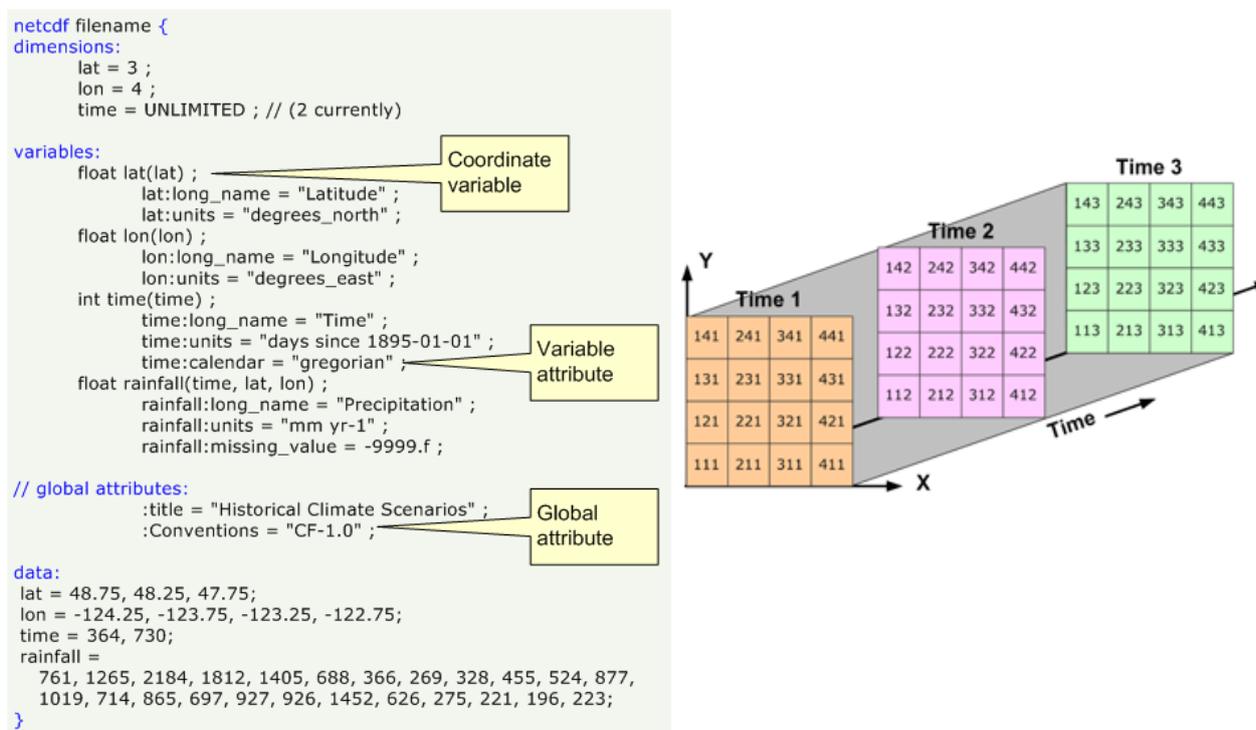


Figura 4.4.2.2: Componentes básicos de un archivo netCDF

Los datos de un archivo netCDF se almacenan en conjuntos. Por ejemplo: la temperatura que varía con el tiempo en una ubicación se almacena como un conjunto unidimensional. La temperatura de un área durante un tiempo determinado se almacena como un conjunto bidimensional.

Los datos tridimensionales (3D), como p.ej., la temperatura de un área que varía con el tiempo, se almacenan como series de conjuntos bidimensionales. Datos tridimensionales: son los datos de un área que varían con el tiempo. Un archivo netCDF contiene dimensiones, variables y atributos. Estos componentes se utilizan conjuntamente para captar el significado de los datos y las relaciones entre los campos de datos en un archivo orientado a los conjuntos.

Dimensiones

La dimensión netCDF consta de nombre y tamaño. El tamaño de la dimensión es un número entero positivo de valor arbitrario. La dimensión se puede utilizar para representar una dimensión física real; por ejemplo, tiempo, latitud, longitud o altura. La dimensión también se puede utilizar como índice de otras cantidades, por ejemplo, número para ejecutar estación o modelo. Es posible utilizar la misma dimensión más de una vez para especificar una shape variable.



Variables

Una variable representa un conjunto de valores del mismo tipo. Las variables se utilizan para almacenar el volumen de los datos en un archivo netCDF. Cada variable tiene nombre, tipo de datos y forma descrita mediante la lista de sus dimensiones especificadas en el momento de crear la variable. El número de dimensiones es la clasificación (también denominado "dimensionalidad"). Una variable escalar tiene una clasificación de 0, un vector tiene una clasificación de 1 y una matriz tiene una clasificación de 2. Una variable también puede tener atributos asociados que se pueden agregar, eliminar o cambiar después de crear la variable.

Variables de coordenadas

Una variable unidimensional que tenga el mismo nombre que una dimensión es una variable de coordenadas. Se asocia a una dimensión constituida por una o más variables de datos y, normalmente, define la coordenada física que corresponde a esa dimensión.

Las variables de coordenadas no tienen ningún significado especial para la biblioteca netCDF. Sin embargo, el software que utilice esta biblioteca deberá manejar las variables de coordenadas de manera especializada.

Atributos

Los atributos netCDF se utilizan para almacenar datos secundarios o metadatos. La mayoría de los atributos proporcionan información sobre una variable determinada. Estos atributos se identifican por el nombre de la variable junto con el nombre del atributo. Los atributos que proporcionan información sobre todo el archivo netCDF se denominan atributos globales.

Convenciones

Las convenciones definen los metadatos que proporcionan una descripción definitiva de los datos de cada variable y sus propiedades espaciales y temporales. La convención ayuda a los usuarios de los datos procedentes de distintas fuentes decidir qué cantidades son comparables. El nombre de la convención se presenta como un atributo global en un archivo netCDF.

4.5 INTERFAZ DE VISUALIZACIÓN WEB DE LA BASE DE DATOS

Mediante esta interfaz el usuario deberá ser capaz de seleccionar las búsquedas previamente definidas en la Base de Datos, ingresando parámetros de selección de shape amistosa, de rápido accionar y que eviten errores. Para los archivos almacenados en la Base de Datos se muestra:

- La descripción general de cada archivo
- La información detallada de cada archivo
- Visualización previa de la información a recuperar mediante mapa grillado.

Para el archivo seleccionado se muestra, antes de descargar el archivo:

- Los metadatos asociados y valoración experta del archivo
- La gráfica donde se pueden obtener datos a descargar, identificando los dominios regionales de las 4 áreas de estudio, mediante selección de áreas.
- puntos/áreas seleccionadas con posibilidad de ser "des-seleccionadas" para reducir el tamaño del archivo a descargar, ajustándolo a las necesidades del usuario.



Se generan los archivos finales a descargar en formatos requeridos en los TDR, independientemente del formato de almacenamiento interno: ASCII y Shape

La interfaz proporciona un feedback del accionar del usuario, con ayudas y recuperación de errores de manera eficaz a través de una gráfica clara que facilita al usuario el reconocimiento visual de lo ofrecido en pantalla.

La interfaz incluye dos componentes principales:

- Una componente pública, a la que accede el público en general y comunica contenidos del proyecto al usuario común.
- Una componente a la que se accede por registración, destinada al usuario interesado en consultar y descargar los datos albergados en la Base de Datos

Diseño de la Interfaz Pública

Esta interfaz tiene como objetivo comunicar a los usuarios los objetivos del proyecto y de la Base de Datos, como así también provee información sobre quienes han colaborado en su construcción, contactos para obtener más información y links relacionados.

La página inicial tiene incluido un menú de ayudas y login. Ambos elementos están preparados para guiar al usuario interesado en usar la componente principal del sitio: “Base de Datos”. Se han agregado estructuras para que este sitio público pueda crecer, teniendo en cuenta que es un proyecto en desarrollo y a futuro puede ser necesario agregar más información.

Diseño de la Interfaz Base de Datos

Esta interfaz cuenta con una registración que inicialmente se usará para moderar los ingresos, y en su fase definitiva, maneja una registración de activación automática para minimizar la atención requerida para el sitio web por el administrador. Para ello cuenta con una pantalla de login, que incluye la registración de los usuarios, recuperación automática de contraseña por olvidos y el ingreso al sitio. La registración de los usuarios requiere que los mismos incluyan en el formulario en línea información de contacto y además las motivaciones y objetivos de los trabajos de investigación, desarrollo, asistencia técnica que requieren del uso de los datos.

La identificación y actividad del usuario es mantenida por el sistema como un registro en el archivo de log incluyendo información como los accesos a la página y exportación de datos. Este permitirá realizar estadísticas confiables del uso de la Base de Datos.

Pasada la etapa de registración, se ingresa directamente a la búsqueda de recursos, según los parámetros definidos:

- Zona
- Período
- Variables
- Frecuencia



Se recuperan los registros que cumplen con la selección, visualizándose una tabla conteniendo:

- Nombre y descripción del Recurso
- Un botón que permite visualizar los metadatos.
- Un botón para visualizar gráficamente el área abarcada por el recurso, desplegando sobre un mapa los puntos que definen la grilla donde se seleccionará las series temporales de datos contenidos en la misma.

El usuario selecciona el “Mapa” del recurso elegido y en una ventana emergente para cada recurso seleccionado, se despliega:

- A izquierda un mapa con los puntos disponibles en color azul.
- A derecha aparece su identificación y un botón de “Info” para visualizar los metadatos. Si la variable cae en la zona Argentina Continental, se mostraran más ayudas de selección.

A medida que los puntos son seleccionados por los distintos métodos disponibles, se muestran en rojo. Los métodos disponibles de selección de puntos son:

- Selección por Región (solo para zona Argentina Continental). Se dispone de cajas de selección, para optar por las distintas regiones. En el mapa se visualizan las selecciones cambiando a color rojo.
- Selección por Provincia (solo para zona Argentina Continental). Las cajas de selección permitirán marcar la/s provincia/s, que en el mapa se visualizan cambiando a color rojo.
- Selección por Puntos individuales. Es posible seleccionar puntos en forma individual, para ello, parados sobre el mapa realizar un zoom de la zona con la que se va a trabajar y hacer click en los puntos elegidos. Los puntos seleccionados cambian a color rojo.
- Selección por Área rectangular, a elección del usuario definida sobre el mapa con el mouse.
- Selecciones mixtas, es posible combinar estas selecciones y así ir sumando áreas/puntos.
- Es posible deseleccionar puntos, provincias y/o regiones. En caso de realizar las deselecciones, los puntos vuelven a su color original azul. Es posible limpiar toda el área de trabajo, usando el botón “Limpiar Puntos”.

La interfaz permite realizar zooms sobre las áreas de interés del usuario identificando los nombres de las localidades hasta una resolución de metros. Es posible realizar “paneos” sobre el mapa hasta ubicar el o los puntos en la zona de interés. Es posible cambiar los fondos de los mapas entre: vista orográfica, vista satelital, o híbrida (con localidades y rutas). Es posible reconocer las coordenadas del punto grillado, señalando dicho punto con el mouse.

Una vez definida la selección final, usando el botón “Exportar Puntos” se dispara la consulta de exportación necesaria para los scripts que procesan los datos. La interfaz muestra un ícono en movimiento, para indicarle al usuario que la extracción está en curso. De esta manera se extraen de los archivos indexados por la Base de Datos, las Series de Tiempo (ST) de la variable seleccionada, en la frecuencia y período elegidos, para cada punto de grilla marcado. Se muestra un link al archivo resultado, y el tamaño del mismo.



Exportación de la Base de Datos

Las ST generadas, se exportan de la Base de Datos. Para cada punto seleccionado, se generará un archivo con la ST en formato texto “csv” (comma separated values: valores separados por coma), y para el conjunto de archivos obtenidos se construye:

- Un archivo GIS, conteniendo las posiciones de todos los puntos exportados en formato shape del área seleccionada.
- Un archivo conteniendo los metadatos asociados.

Se utiliza el formato “csv” porque es muy genérico y de fácil importación a distintos utilitarios (por ejemplo MS Excel, Access, o para ser usados por distintos lenguajes de programación: Matlab, R, etc).

La estructura del nombre de los archivos es:

- Lat{Latitud},_Lon{Longitud}.csv
- puntos.{shp|shx | dbf }
- detalle.cvs

Donde:

Latitud: es la latitud correspondiente a cada uno de los puntos exportados

Longitud: es la longitud correspondiente a cada uno de los puntos exportados.

Puntos: es el archivo shape conteniendo los puntos exportados.

Con el conjunto de archivos csv, el archivo shape y detalle, anteriormente descriptos, se construye un archivo comprimido (en formato zip). Se realiza esta compresión pues para el usuario es más eficiente la descarga de los archivos por internet, ya que reduce el tiempo de transferencia.

El formato del archivo Latx_Lony.cvs para las ST es:

Latitud (numérico -gg.mm)

Longitud (numérico -gg.mm)

Fecha (date dd/mm/aaaa)

Valor (numérico snnn.d)

El formato del archivo detalle.cvs es:

Registro 1:

Nombre de Variable; Unidad: texto libre

Registro 2:

Archivo original: texto libre

Registro 3:

Nombre Institución: texto libre

Registro 4:

Autor: texto libre

Registro 5:

Cita: texto libre

Registro 6:

Comentario Experto: Texto libre



El GIS, está compuesto por los archivos SHP, SHX y DBF, sistema WGS 84 en formato ESRI compatible.

4.6 DESCRIPCIÓN DE LOS ARCHIVOS DE METADATA

Como se describió en las secciones anteriores, las consultas y vistas que se pueden realizar en la Base de Datos incluyen la visualización de los archivos de metadatos asociado a cada recurso que se ofrece. El metadatos se organiza a través de una tabla que contiene lo siguiente:

- Código: uso interno de la Base de Datos
- Descripción: Se describe brevemente los datos y los métodos utilizados para generar el conjunto de datos grillados.
- Versión: se especifica en caso que el conjunto de datos cuente con más de una versión disponible.
- Zona: zona Continental o Antártica en la que se provee el recurso.
- Área Geográfica: límites en latitud y longitud de la zona en la que se provee el recurso.
- Período Desde/Hasta: período temporal en el que se provee el recurso.
- Frecuencia: Diaria o Mensual
- Institución: Nombre de la Institución o Instituciones de origen que proveen el recurso.
- Autor: Nombres del personal de esas Instituciones responsable de la generación del recurso.
- Cita: Referencia bibliográfica relacionada con el recurso.
- Tipo de Reticulado (grilla): espaciamiento en latitud y longitud del reticulado en el que se provee el recurso.
- Tipo de Datos: “Grillados Observados” en caso que el recurso se base en observaciones o “Grillados Simulados” en caso que el recurso provenga de simulaciones de modelos climáticos.
- Volumen: Tamaño en Megabites del recurso.
- Variables: Información sobre la variable, unidad, código de datos faltante, asociados con el recurso.
- URL: Dirección web donde el recurso es proporcionado por la Institución de origen.
- Nombre del Archivo: Nombre del archivo netcdf en el que se almacena el recurso en la Base de Datos.
- Comentario Experto: Espacio que en caso sea necesario se utiliza para proveer información adicional al usuario sobre el recurso. En particular se incluye información sobre la calidad del dato provisto por el recurso.

La descripción detallada del metadatos correspondiente a cada uno de los conjuntos de datos o recursos provistos en la Base de Datos y detallados en la sección 4.2 se incluye en el Anexo V.



4.7 DOCUMENTOS EXPLICATIVOS DE LA BASE DE DATOS

De manera facilitar a los usuarios el acceso y utilización de la Base de Datos a través de la interfaz web así como facilitar y recomendar una utilización adecuada y responsable de su contenido, se han elaborado un conjunto de documentos explicativos relacionados con:

- Acceso a la Base de Datos
- Uso de la Base de Datos
- Uso de los datos Grillados Observados
- Uso de los datos Grillados Simulados

El detalle de la primera versión de esos documentos se encuentra en el Anexo Cap4.IV. Los mismos estarán disponibles a través del link “Ayuda” que se encuentra en la parte superior derecha de la interfaz web.

Los documentos están organizados en el formato de Pregunta/Respuesta. Este formato se seleccionó dado que permite una identificación rápida por parte del usuario de la pregunta que resuelve su duda o su problema y a la vez este formato permite una fácil actualización y/o expansión de los temas de ayuda.

El contenido de los documentos proporcionados en este informe serían una primera versión que luego se actualizará incluyendo preguntas que surjan de los usuarios que utilicen la Base de Datos.

4.8 IMPLEMENTACIÓN Y PUESTA A DISPOSICIÓN DE LA BASE DE DATOS

La Base de Datos está implementada en los servidores del Centro de Investigaciones del Mar y la Atmósfera (CIMA), que cuenta con un servicio de acceso a internet de alta velocidad, provisto por la FCEN/UBA y acceso a Redes Avanzadas (Internet II) provistos por innova-red (innova-red.net). El host cuenta con servidor Web Apache2, Servidor de Base de Datos MySQL y motor PHP. Todos ellos protegidos por un firewall.

La figura 4.8.1 muestra en forma resumida la relación entre las consultas de los usuarios (Cliente Web), la interacción de los servidores y el resultado obtenido.

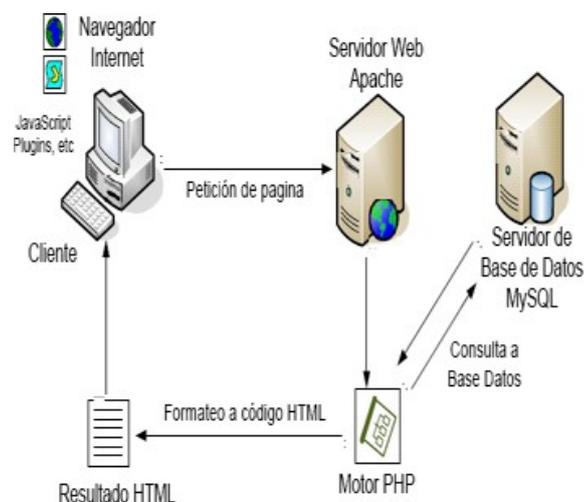


Figura 4.8.1: *Relación entre las consultas desde la Web e interacción de los servidores*

La Base de Datos se encuentra disponible y su consulta disponible a través de la interfaz web, en el sitio: <http://3cn.cima.fcen.uba.ar>.