

# ICOS Carbon Data Portal: Adaptación de un catálogo de metadatos para actuar como repositorio integrado de mediciones sobre gases de efecto invernadero a disposición de la comunidad científica

O. Fonts, M. García, F. González (1), J. Piera, J. Sorribas, J. Olivé (2)

<sup>(1)</sup> Red de desarrolladores SIG independientes geomati.co. <http://www.geomati.co> {oscar.fonts, micho.garcia, fernando.gonzalez} @geomati.co.

<sup>(2)</sup> Unidad de Tecnología Marina, Departamento de Telemática. Centro Superior de Investigaciones Científicas. <http://www.utm.csic.es> [utmtel@utm.csic.es](mailto:utmtel@utm.csic.es)

## RESUMEN

La infraestructura europea [ICOS \(Integrated Carbon Observation System\)](#), tiene como misión proveer de mediciones de gases de efecto invernadero a largo plazo, lo que ha de permitir estudiar el estado actual y comportamiento futuro del ciclo global del carbono.

En este contexto, [geomati.co](#) ha desarrollado un portal de búsqueda y descarga de datos que integra las mediciones realizadas en los ámbitos terrestre, marítimo y atmosférico, disciplinas que hasta ahora habían gestionado los datos de forma separada.

El portal permite hacer búsquedas por múltiples ámbitos geográficos, por rango temporal, por texto libre o por un subconjunto de magnitudes, realizar vistas previas de los datos, y añadir los conjuntos de datos que se crean interesantes a un “carrito” de descargas.

En el momento de realizar la descarga de una colección de datos, se le asignará un identificador universal que permitirá referenciarla en eventuales publicaciones, y repetir su descarga en el futuro (de modo que los experimentos publicados sean reproducibles).

*El portal se apoya en formatos abiertos de uso común en la comunidad científica, como el formato [NetCDF](#) para los datos, y en el [perfil ISO de CSW](#), estándar de catalogación y búsqueda propio del ámbito geoespacial. El portal se ha desarrollado partiendo de componentes de software libre existentes, como [Thredds Data Server](#), [GeoNetwork Open Source](#) y [GeoExt](#), y su código y documentación quedarán publicados bajo una licencia libre para hacer posible su reutilización en otros proyectos.*

**Palabras clave:** *ICOS, Carbon, data portal, Thredds, Geonetwork.*

## 1 INTRODUCCIÓN

### 1.1 ICOS

Los objetivos de la infraestructura europea ICOS <sup>1</sup> son:

- Realizar las observaciones, a largo plazo, necesarias para comprender el estado actual y predecir el comportamiento futuro del ciclo global del carbono y las emisiones de gases de efecto invernadero.
- Observar y evaluar la efectividad de la fijación de carbono y/o las actividades de reducción de las emisiones de gases de efecto invernadero en los niveles de composición global atmosférica, incluyendo la atribución de fuentes y sumideros por región y por sector.

Para ello se establecerá una red de observación sistemática a largo plazo, diseñada alrededor de unas instalaciones centrales, constituida por estaciones de medidas de alta precisión dedicadas a la monitorización de los flujos de los GEIs <sup>2</sup> de los ecosistemas terrestres y su concentración en la atmósfera, así como de los intercambios de CO<sub>2</sub> entre la atmósfera y los océanos. Las observaciones proporcionadas permitirán a los investigadores conseguir una plena comprensión de los intercambios de los GEIs sobre el continente europeo y de sus causas.

---

1 Integrated Carbon Observation System

2 Gases de Efecto Invernadero

## 1.2 Portal de datos ICOS-Spain

En el contexto de ICOS, **geomati.co** ha desarrollado un portal de datos que sirva como punto de acceso unificado a los datos proporcionados por la parte española de su infraestructura. Dicho portal es el que se describe en este artículo.

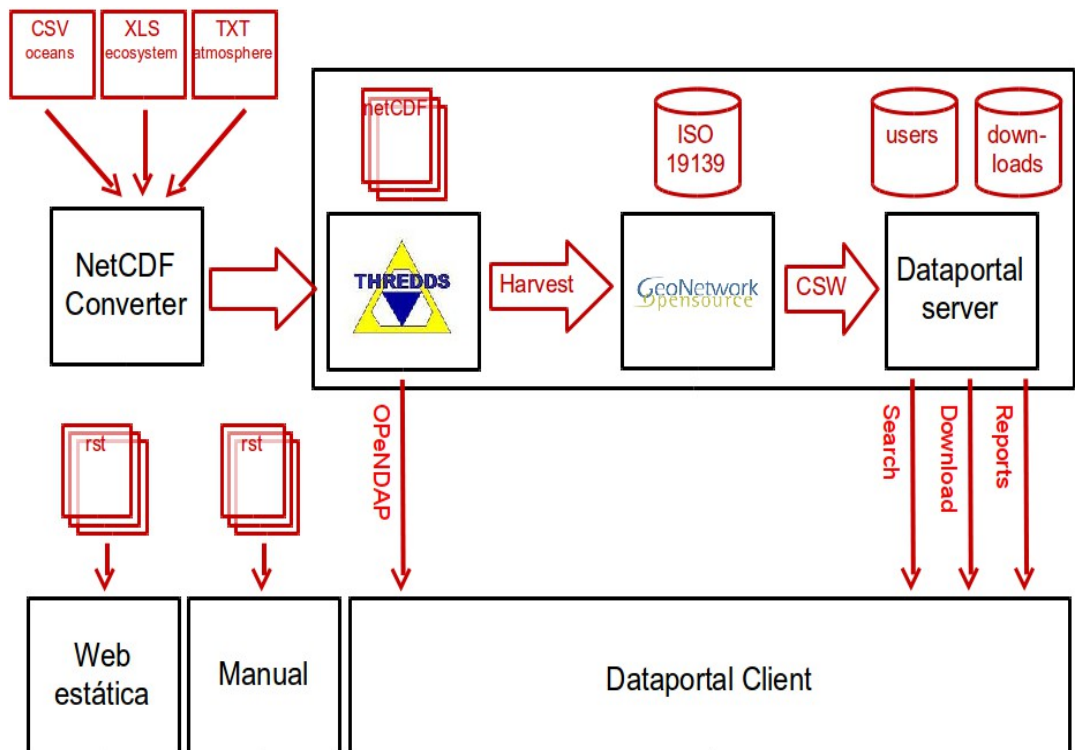
DataPortal permite:

- La agregación de datos procedentes de distintas áreas temáticas, convirtiéndolos a un formato común NetCDF, y su publicación en línea en uno o más repositorios mediante el uso del servidor Thredds Data Server.
- La extracción de metadatos e indexación en un catálogo ISO 19119 mediante el uso de GeoNetwork.
- El acceso a dichos datos mediante una aplicación web. La aplicación permite:
  - Búsqueda en el repositorio de datos, combinando los siguientes criterios:
    - Texto libre.
    - Uno o más ámbitos geográficos definidos como cajas contenedoras (*bounding boxes*).
    - Rango temporal en que se capturaron los datos.
    - Un conjunto de variables procedentes de un vocabulario controlado.
  - Previsualización de los resultados, entre los que destaca la posibilidad de inspeccionar los propios datos de forma gráfica antes de su descarga.
  - Selección de los conjuntos de datos que se desea descargar (al modo de un *carrito de la compra*).
  - Descarga de los datos, para lo que se requiere un registro previo de los usuarios.
  - Reproductibilidad de las descargas, a las que se asigna un identificador universal para ser referenciadas en posibles publicaciones.

- Generación de informes estadísticos sobre el uso del propio portal de datos.

### 1.3 Componentes del portal de datos

El desarrollo de DataPortal se ha basado en la reutilización de componentes de software libre y estándares abiertos de amplia implantación entre la comunidad científica y la industria geoespacial, y es a su vez un proyecto de software libre que esperamos poder poner a disposición de la comunidad a la mayor brevedad posible.



*Componentes del portal de datos.*

Como se observa en la figura, los datos de partida están contenidos en ficheros de texto plano, en formato CSV o Excel, cuyo formato y especificación dependen del área temática de los que proceden.

La primera etapa consiste en la conversión de los distintos tipos de datos a uno común mediante un importador extensible a nuevos formatos. Una vez convertidos a un formato común, los datos se publican en la red mediante Thredds. La siguiente etapa (GeoNetwork) se encarga de extraer e indexar los metadatos necesarios, y de servirlos mediante estándares CSW e ISO.

DataPortal es la aplicación desarrollada ad-hoc para proporcionar una interfaz de usuario suficientemente rica y simple, y se divide en la parte servidora, que proporciona una serie de servicios web (búsqueda, descarga, autenticación, generación de estadísticas), y la parte cliente, que se ejecuta en el navegador del cliente, y que proporciona la interactividad necesaria para la selección, evaluación y descarga reproducible de los datos.

Para la documentación de la aplicación se ha utilizado el formato ReStructuredText.

## 2 FORMATO DE LOS DATOS

Se ha optado por el uso de NetCDF como formato común de los datos por tratarse de un formato:

- Abierto, cuya especificación es pública.
- Ampliamente utilizado por la comunidad científica, para el que existen decenas de aplicaciones de visualización y manipulación ya disponibles.
- Simple, que se basa en unos pocos conceptos muy básicos, y para cuya utilización no hace falta definir de antemano modelos de datos complejos.
- Compacto, puesto que la codificación de los datos es binaria, y el propio formato introduce muy poco *overhead*.
- Versátil, capaz de contener datos puntuales, series temporales, trayectorias, imágenes ráster (grid), todo ello con capacidad multidimensional, y pudiendo definir relaciones entre las variables.
- Potente, puesto que cada variable tiene asociada una serie de metadatos según convenciones desarrolladas en base a buenas prácticas, y que proporcionan información sobre el ámbito, granularidad, o unidades de cada una de las variables contenidas en los datos.

## 2.1 Metadatos

El *Data Portal* puede realizar búsquedas por ámbito geográfico, período temporal, conjunto de variables, o por texto libre, dentro de un conjunto de ficheros netCDF. Para que los datos originales sean recuperables mediante estos criterios, es necesario que éstos estén convenientemente descritos y puedan extraerse las variables que permitirán su búsqueda al final del flujo.

A tal efecto, se han seguido un subconjunto de las convenciones ya existentes para datos en el ámbito de la observación climática y predicción meteorológica<sup>3</sup>, y para el descubrimiento de los datos<sup>4</sup>.

En concreto, los datos NetCDF incluidos en la aplicación netCDF deberán contener los siguientes atributos globales:

Propuesta de atributos globales NetCDF mínimos para Data Portal

<i>Atributo</i>	<i>Descripción</i>
Metadata_Conventions	Literal: "Unidata Dataset Discovery v1.0".
id	Identificador de este dataset. Un UUID.
title	Descripción concisa del conjunto de datos.
summary	Un párrafo describiendo los datos con mayor detalle.
standard_name_vocabulary	Idealmente se usaría un vocabulario estándar como <sup>5</sup> . En caso de utilizar un vocabulario propio, indicar la URL donde pueda descargarse el vocabulario.
geospatial_lat_min	Límites geográficos en los que están contenidas las medidas, coordenadas geográficas en grados decimales, sistema

---

3 Climate & Forecast 1.5:

4 NetCDF Attribute Convention for Dataset Discovery:

5 CF Standard names:

<i>Atributo</i>	<i>Descripción</i>
	de referencia WGS84 (EPSG:4326).
geospatial_lat_max	<i>idem</i>
geospatial_lon_min	<i>idem</i>
geospatial_lon_max	<i>idem</i>
time_coverage_start	Instante del dato más remoto, en formato <a href="#">ISO8601</a> .
time_coverage_end	Instante del dato más reciente, en formato <a href="#">ISO8601</a> .
institution	Institución que publica los datos.

cdm\_data\_type

[Tipo de datos Thredd](#)