

# A USER-CENTRIC APPROACH TO DATA CUBES

A RECENT OGC WORKSHOP LED TO THE DEVELOPMENT OF A NEW DEFINITION OF DATA CUBE AND UNDERSCORED THE NEED FOR A USER-CENTRIC APPROACH. **INGO SIMONIS** REPORTS



Geospatial data cubes are used frequently these days for their enabling of performant, cloud-compatible geospatial data access and analysis. But differences in their design, interfaces and handling of temporal characteristics are causing interoperability challenges for anyone interacting with more than one solution. Such challenges are unnecessarily wasting time and money, and – from a science perspective – affecting reproducibility.

To address these challenges, the Open Geospatial Consortium (OGC) and the Group on Earth Observation (GEO) invited global data cube experts to discuss the “state of the art” and find a way forward at the ‘Towards Data Cube Interoperability’ workshop. This two-day workshop in late April not only produced a new definition of the term ‘data cube’ but also underscored the need for a user-centric API-based approach that exposes not only the data available to the user, but also the processing algorithms that can be run on it – and enable users to add their own.

### **Data cubes from the users’ perspective**

Existing definitions of data cubes often focus on the data structure aspect as used in computer science. In contrast to this, the workshop emphasised the need to leave these definitions behind and focus on the user’s perspective. Users don’t care if the data is stored in a relational database, a cloud-based object store or a file server. What users are interested in is how they can access the data and the processing algorithms that they can apply to it. Any such standard for access should reflect this.

This led to an interesting rethinking of just what a data cube is and can be. Although it wasn’t agreed to on any formal consensus-basis, the workshop participants generally took a user-centric definition of a geo data cube to be “a discretised model of the earth that offers the estimated values of certain variables for each cell. Ideally, a data cube is dense (that is, it does not include empty cells) with constant cell distance for its spatial and temporal dimensions. A data cube describes its basic structure – its spatial and temporal characteristics and its supported variables (aka properties) – as metadata. It is further

defined by a set of functions. These functions describe the available discovery, access, view, analysis and processing methods by which the user can interact with the data cube.”

As we see, the data cube is described for the user, not the data. It does not matter if the data cube contains one, two or three spatial dimensions, or if time is given its own dimension(s) or is just part of the metadata of an observation — or isn’t relevant to the data at all. Similarly, it doesn’t matter how the data is stored. What will unify these heterogeneous data cubes is their use of a standardised HTTP-based API as their method of access and interaction.

The main concern of the user is what functions the data cube instance offers to apply to the data. These functions are what primarily differentiate the user-centric data cube definition over other definitions. A user needs to understand what questions can be asked to access data that fulfils specific filter criteria, how to visualise specific (sub-) sets of data, or how to execute analytical functions and other processes on the data cube. If supported, the user also needs to understand how to add their own processes to the data cube so that these can be executed directly on the data cube without the need to transfer vast amounts of data out of the cloud.

This isn’t to say that all other characteristics are of no concern to the user — they still need to be known. As such, they will be provided via the data cube API as metadata, so that the user can take them into account when assessing how best to process the data.

### **Interoperability through a data cube API**

Where does this leave OGC? We think an API-based, flexible approach to standards will provide end users, software developers and data cube operators with the best experience.

### **For end users**

A single, simple, standardised HTTP API to learn and/or code for, no matter where the data resides, will mean an increased selection of available software will support an increased choice of data cube providers and an increased number of processing algorithms. From a scientific perspective, this means that the atmospheric scientist doesn’t

additionally have to also be a Python expert, potentially using a low- or no-code platform GUI to create an algorithm that processes the data for their heatwave study across Germany. Another atmospheric scientist could then take that same processing algorithm and apply it to the UK with minimal changes – even if the required data is held by a different standards-compliant data provider – increasing the transparency and repeatability of scientific studies and other valuable analysis tasks.

### **For software developers**

A single, simple, standardised HTTP API means software developers don’t have to design their own vendor-specific methods for providing access to data cubes in their software. Instead, they interact with data cubes via HTTP calls, thus benefiting from simple standard Web communication, rather than interactions on the programmatic level. By coding to an agreed-upon standard, developers can work with any compliant data cube while minimising cube-specific adaptations. This increases the usability of the software, while decreasing the development and maintenance costs.

### **For data cube operators**

Using a single, simple, standardised HTTP API reduces development and maintenance costs while broadening the customer base. Being standards-compliant enables providers to access customers that are using any compliant software package, rather than just those using a select list of software coded to work with your specific instance. This means that more people will be coding for your data cube, even if they don’t know your service exists.

### **What’s next for OGC?**

It’s early days yet, but you can expect to see a data cube-related API become part of our suite of OGC API standards. Work towards such a data cube API builds upon the work of our Earth Observation Exploitation Platform (see ‘An App Store for Big Data’, *Geoconnexion International*, July/August 2020) and is under way as part of OGC Testbed-17.

**Ingo Simonis is chief technology innovation officer at OGC ([www.ogc.org](http://www.ogc.org))**