Topic 5a - Part 1: Future Innovations - Open data and emerging data services

[MUSIC PLAYING] One, I would say, winning point of Copernicus is the full open and free data policy. Data are available to everyone. It's not kept just for us. It's for the benefit of everyone. And I think that this is making our data widely used.

The previous bigger example was the Landsat in the land area. But I think that in some aspects, we are even going beyond what Landsat was doing because we are generating data in many other areas.

And as I said, sometimes we are slave of our success, because the data are so huge that sometimes the bottleneck is really to make these data available. But I think we are almost there. I think also this part will be solved. But I think in the end it is better to have this kind of challenge than not having any data at all.

We know about the several applications for smartphones that have been built on the progress from Copernicus in order to provide really some information to the citizens, in particular, related to their health. For example, to understand at city level what is the level of pollution. In a given day, if you can go out with your children or not, this is kind of example.

A challenge that we have is the huge amount of data that we are generating, and it is beyond what we were expecting. And for this reason we were, in some way, obliged to put a solution to that. Because it's good to generate data, but it's also more important to make these data used. And so we got in two new solutions. Now we are speaking about new platforms in order to give these data available to everyone.

The acronym for these platforms is DIAS. We have five of them on the pipeline. And in particular, there is one of these that is put in place by a team ECMWF, EUMETSAT, and Mercator. Together they are putting that in place, and this will be probably one of those that there will be more targeted on atmospheric composition.

In November 2017, the European Environment Agency and the European Commission launched the new European Air Quality Index. This is a new product for us. It combines what I would call traditional sources of air quality data with new information sources such as that available under the Copernicus services.

What the Index does is to convert what is often quite complex air quality data into a simple metric designed for the public and other key stakeholders to understand. So the measurements taken in the countries-- the satellite data-- is fed through to us. And within an hour of the measurement taking place in cities across Europe, the information is made available for the public in an easy, online tool that allows them to go and see the quality of the air where they live.

I think one of the key purposes and functions of the Air Quality Index is to improve the communication around air quality. What we have seen is there's a huge public appetite for reliable and timely information on air quality. The day of the launch, we had more than 240,000 individual visits to the website of the Index showing, indeed, the interest of air quality and its importance as an issue that affects us all.

What we have also seen in the Index is that originally when we launched not all countries were reporting the type of data that we needed to include in it. So since launch we've been very pleased to work with other countries across Europe to start reporting this information and make it available for the first time. So, for example, countries such as Bulgaria, Cyprus, and Estonia have all started reporting since the publishing of the Index online.

So when people think about the atmosphere, subconsciously they already have a perception from weather-- from clouds, from rainfall-- that it's what is happening around them that they want to know about straightaway. Same with other parts of the atmosphere now. We want to know locally for us what's our air quality. What's the change in our air quality? How is it improving?

You can imagine a situation-- and people are already working towards it-- where people want to take that satellite observations, mix it with observations at the ground and models, to give people a much clearer idea of the air quality where they live, the locality.

So I in this lovely garden, I get nice emissions of organic compounds from the trees. I get some air quality influences from the city around. Is this a good place to be? Is this a good place to live? I certainly think so. But I'd like to know so. I think that's that information.

We could do this interview somewhere else because we don't like the air quality here.

We could. We could.

The atmosphere is changing, and so it's not just about us right here. Also there are bigger changes in the atmosphere, and we're watching those as they happen. How is that changing things?

People, I think, want to know much moreso about their contributions to it, and how their own personal behavior-- and indeed, the behavior of the organizations they work for-- can help in that circumstance. How do they really choose the lifestyle, the transport systems, the technologies that they use?

I would like to think that in the future people are going to be much more positive about it. We tend to be a bit negative. It's about how we present a problem. But it could be about a choice. It could be about positive choice. Well, actually, we know if we do this, we get the right technology, we get the right atmosphere, we get the right environment. It's part of our life.

So we're starting to have the information to make those choices conscious. We can see various versions of the future.

Yeah.

And we might be able to say, well, actually, we choose this one.

Yes. And I think if you look at companies, they are, after all, composed of people, and they're managed by people. Those people think more and more, I think, about corporate responsibility. They can think about how does what we do here and where I live and my headquarters is, how is it related to the company? But also how about the factories or the systems I run somewhere else? It's a globalization in the industry, but it's also globalization of environmental sustainability of living.

What we've seen since the launch of the Air Quality Index is that a number of third parties have either adopted the Index in its entirety or have used its methodology for their own applications. So, for example, we've seen the Index is been used on the official ministry pages in a number of countries. Otherwise, we have instances where third party developers are actually using the data. So it's taking the Institute data from us coupled with the Copernicus data and using it in a range of different applications.

So Mark, we've heard a lot in this course about data. And it's a very difficult thing to imagine because it's sort of somewhere. There's all these numbers somewhere. But the numbers are getting greater. But the important thing is that it is accessible. So there's a huge amount of it, but people do have access to it. But how is that link made?

So our data is stored in a big building just behind us. And it's an enormous building. All the EUMETSAT data is stored there. All the CAMS data will be stored over at ECMWF in Reading.

And one of the challenges that users face is well, that's nice. All the satellite data is here. All the models data there. Do I really have to download data from both places? So there's a big project that we're working with ECMWF, who host CAMS, and also an Ocean equivalent and ourselves to see if we can put all of the data in the cloud to make it much easier for the users to use.

So all of this data is theoretically accessible. There are no barriers. If you can use the software, everyone's allowed to get at the data.

Absolutely. So often the only barrier is a registration one. But then it's all to do with do you know how to use the software? Can you work with the data formats? So putting data in easier-to-use spaces is becoming important. But for the users, what they want to do is go to one place-- maybe in the cloud-- and just access the data.

Other barriers to use are once you've got the data, do you know what the format is? Do you have software to read it? So what we're trying to do is put in place easier software to use so that people can choose the data set they want for their problem and actually work on the science problem and exploit the data, rather than spending so long just wrestling with the formats.

Then there is all the technical side behind this. We can't take it for granted. There's a lot of different types of data and they're labeled in different ways and they have timestamps. But actually, the user doesn't want to know about any of that. They just want one thing.

It's a huge effort to put stuff in place, and it's a huge team effort on our side. But it's kind of a privilege-- when you see all the scientific papers written and you see the science output that gets done that utilizes all this data-- just to make it easier. Because we're the data experts in our data, but the actual application-- the solving and finding out of new things-- we want to get as easy as possible. So placing great quality data sets that are validated, verified, with easy-to-use tools in front of people, is really important.

And let's just talk for a second about the volume of data. Because we all get frustrated when it takes half an hour to download a movie online. But we're talking about much, much more data than that potentially.

We're into the world of petabytes of data. It's utterly huge. And this will only improve. So behind us we have pictures of the next generation of satellites getting launched at the end of this decade, the beginning of next. And there we're going to see 50 times more data flowing down towards Earth than we're seeing now. And so making that data accessible and usable is a real challenge.

And so there's a project in place to combine all this together. What's it called? And tell me a little bit about how it will work in the future.

So this is part of a European Commission organized thing. So there's a thing called DIAS. And so our one that we're working on with ECMWF and CMEM in Toulouse is a thing called Earth Hub. And so the idea is to put in front of the users all of the atmosphere and modeling data that ECMWF has, all of the marine data that Toulouse have, all the satellite data we have in one environment.

So a user will come to this place. They won't know where the data is, but it will all be very easy to use. And the real advantage there and the wonderful thing there is people being able to do interdisciplinary search across all of these data sets.

So for example, we've had some people who are looking at the way the ice when it melts flowing into the water just off of Greenland. So you can look at, well, what does the ECMWF

model say about that? What do some of the satellite data say about what's going on with the biology, with the temperature of the oceans? And to study the environment as a whole using all of the data and all of the sensors that are available.

And how far in the future is that? It sounds like a very sensible system. But how long is it until that will become accessible?

So we're building it this year. We're going to be piloting it at the end of next year. So in the next two or three years, this will become a reality for research users. And we should hopefully see some small and medium sized enterprises across Europe also starting to use it in the next, I would say, two years.

And the important thing here is making it accessible to those people who are not necessarily scientists, who don't have degrees in computer coding, who have lots of other interests but they're not specialist data management people.

Yeah. So if you don't know what netCDF is, if you don't know what a Python script looks like, that shouldn't be a barrier to you using our data. And so we really want to work hard to get the data to you in a way that you can use it, rather than you having to travel towards us and learn a whole bunch of new stuff. You've got a science problem to solve, and our data's useful. So it's just going that little bit of extra mile to make it useful.

So we know there's a lot of satellites up there now. We know they're already sending a lot of data. How is this going to change in the future?

So in the future, we've got two new programs launching at the end of this decade, beginning of next. So these two satellites, they get launched-- one in 2021, one in 2022-- and they're going to be the next generation of geostationary sensors looking over the earth. So one of them is an imager taking pictures of the earth all the time. The other one is a new instrument, what we call a sounder. So taking three-dimensional view of the earth looking at the temperature, the humidity, and the chemistry all the time.

These are going to produce a lot more data than the current satellites.

Huge. We're looking at another 50 times more data when these two sets of systems come online. So it's a huge data problem for the users, which we are working with the users to see how we can make it easier for them to get the data.

[MUSIC PLAYING]