

Script of Presentation for WMO Data Conference, prep workshop 1 – The changing landscape of weather and climate data.

“The evolving meteorological data landscape”

[intro]



Good afternoon, my name is Charlie Ewen and I am the Director of Technology at the Met Office, the UK's National Met Service, and I would like to thank the Secretariat for being asked to speak at today's workshop to talk about the wider data landscape. I do have a tendency to talk quickly for which I apologise however a transcript will be available after this presentation for those that are not English-as-a-first language or, should there be any IT issues given that I am, as you can see, presenting from my workshop at home!

I have been at the UK Met Office for something over a decade and prior to that worked in the Private Sector in the Defence, Finance and Retail & Distribution industries working on what we would now call Digital Transformation.

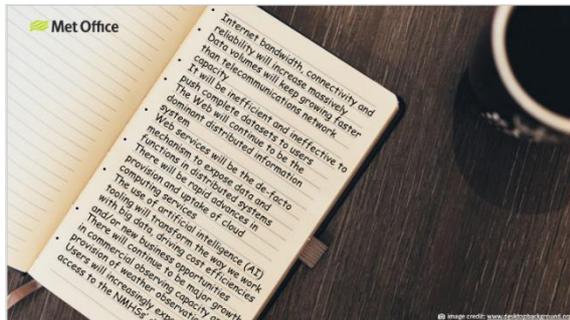
My main objective for today is to encourage our community to recognise that there has been a change in the way that we need to think about ourselves as National Hydrological and Meteorological Services and a need to re-think our approach to ensuring that by 2030, 'all nations are more resilient to impactful environmental events in the context of our changing climate.'

Firstly, let me elaborate on how the Met Office thinks the world has changed.

The Information Economy, Big Data and related themes and concepts are words that have existed for some time. It is easy to think 'yes I know – we are ready', and not respond. Back in the 1990's, the same things were being said about the Internet and its effect on retail. I was working in that industry at the time and for 4 or 5 years there was conference after conference discussing how the internet was going to change that world however despite this, most brands did not respond quickly enough. It has been noted by many that in technologically-driven change, recognition is early and often over-hyped, but that longer-term impact is almost universally underestimated – especially by leading sectors and companies. Just as was the case in my old industry in the late 1990's, I believe that we are now past the inflection point of exponential growth, the time at which change is doubling from a substantial number or base. Forgive me for UK centric numbers but the UK Data Economy increased from €43.8bn in 2013 to €78bn in 2019 and will top €89.7bn in 2020 and for the EU27 plus the United Kingdom exceeded 400 Billion Euro in 2019, with a growth of 7.6% over the previous year according to analyst firm, DataLandscape. According to Gartner, by 2022, 35% of large organizations will be either sellers or buyers of data via formal online data marketplaces, up from 25% in 2020. These numbers bear a striking resemblance to those that emerged in the late 90's when many brands that we have now forgotten didn't act quickly enough – the internet did change everything as will the arise of the information economy.

NEXT SLIDE PLEASE

[acknowledging that there's a huge amount of change in the data landscape]



[Source: EUMETNET Federated Data Coordination Mechanism requirements, Aug 2020]

Moving beyond statements about the ‘information economy’, I just want to acknowledge that there is a huge amount of change around the data landscape and the underpinning technology. Perhaps even exponential change – a concept that we’re all familiar with since the emergence of COVID-19. I could fill my allocated time today with just one of any of these topics – but I won’t. Instead, I want to share with you four things that are driving our thinking at the Met Office.

NEXT SLIDE PLEASE

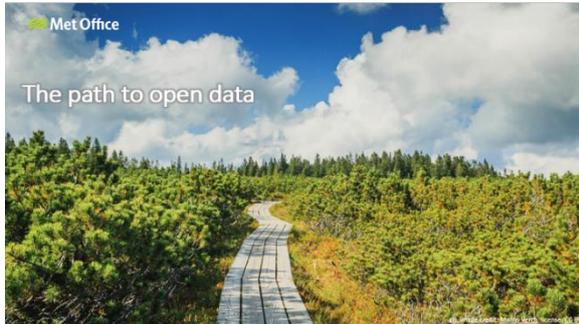
“We are a data company”



So how must we think differently? At the Met Office we often refer to a model called the four lenses – firstly we are a part of UK Government and therefore can easily see ourselves through this lens. Secondly, we are a research science organisation and as such, work within national and international academia in Environmental Science. We are part of an international community orchestrated by the WMO and again, a familiar lens. The Met Office is also a Trading Fund and as such gets involved in the direct delivery of services to industry as well as the public via our Public Weather Service. For those of you from other Meteorological or Hydrological organisations, the lenses may look a little different however I am sure that all of us have more than one, even if they are the need to work as an international community as well as a part of domestic government. I propose that there is a new and less familiar lens that we need to be prepared to look through, plan and act upon – this is the lens of a ‘Big Data’ company operating within a global Information Economy. Of course, this does not diminish the validity or importance of the others – it is just a new way to think, plan and act.

NEXT SLIDE PLEASE

“The path to open data”



Increasingly, we are seeing domestic governments, certainly in the UK and in the European Commission, develop national ‘data strategies’ and in doing so, tending towards the object of those data strategies and policies to simply be ‘data’. Understandably, when bureaucrats think about data, they tend towards the dictionary definition of something that describes fact such as “factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation”, taken from Webster. This approach immediately presents challenges for us where a significant fraction is anything but fact – just ask a member of the public about forecast accuracy!

The European Union seem to be dealing with this issue by introducing an ontological approach and using ‘acts, facts and information’ as a means to further break down the word ‘data’. At the Met Office we use a more straightforward approach with a concept that revolves around simply, ‘facts and opinions’ to distinguish between Observations and Models, Simulations, Predictions and Forecasts. As anyone that works with Observations will tell you, even they have a touch of ‘opinion’ about them especially remote sensing where the variables that we use are often derived from the raw sensor data and even with something as straightforward as rain gauge data, there are always some error bars. Essentially and especially as we transition to Ensemble Prediction Systems, models are the same with the ability to numerate the confidences and probability around a given variable however this is a developing area and the distinction of ‘facts and opinions’ is useful. So, what does this mean in practice? This was put very well by a colleague of mine when she said, “with a model, forecast or prediction, the metadata is probably as important as the data themselves’. We were discussing how useful models are at the time and just like any opinion, it is important to know the provenance and uncertainties around any opinion before we can make an important decision based upon it – at least we all should! We are not alone in recognising this concern. If you look at the FAIR data principles developed in the scientific community, metadata is considered equally as important as data. Data usage licenses, detailed provenance, and ‘domain-relevant community standards’ are specifically cited cases. I use this simple example to illustrate the issues associated with a blanket approach to data strategies and policies – it is simply too big a word to be effective without a common approach to classification and this is an area in which we, as a community, have not done enough. Each Met Service has different approaches to describing data – those ‘domain-relevant community standards’ mentioned in the FAIR data principles. Just within the Met Office we have many ways to describe what to non-expert users, would appear be the same thing.

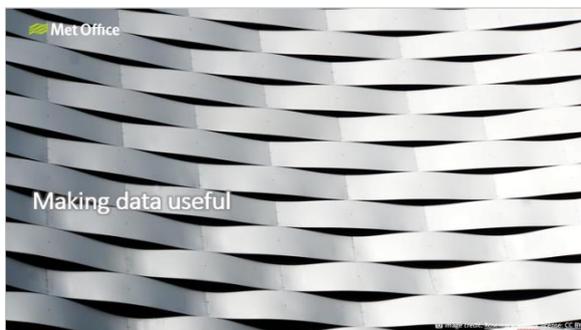
This lack of coordination, especially as it relates to prediction, model and forecast data makes us impossible to deal with as a community. We live in a global world and it is not surprising that industrial-scale users such as Google, Facebook, Amazon, Microsoft and so on have all turned to the private sector to obtain data – not necessarily because they believe the data to be ‘better’ in some

way but often simply because it is useful, consistent and global. We DO have the agreed mechanism to respond to this in WIS2.0 however as with many things, we are being too slow. The link between openly available data and social and economic benefits is well established. Within EU countries, this frustration of data not being available will be addressed by a blanket acceleration of ‘Meteorological Data’ being identified as ‘High-Value Data’ and a directive that it shall be made open and freely available being stated in the recently updated Directive on Open Data which is scheduled for implementation next year.

This push toward open data is just one example of how we’re seen as data companies – in this case by governments. For many of us, this change in perception is driving a change in our purpose and defining a new role in the information economy that our governments expect us to adopt.

NEXT SLIDE PLEASE

“Making data useful”



I am sure that by now, most NMHSs understand the economics of Open Data. At the Met Office we are committed to providing easy discovery and open access to as much of our public task data as possible and in the main, already do. As I previously commented, the benefits case in doing this is clear. Where this does not go far enough is in making it useful. If you look at our community through the others’ perspectives, we are idiosyncratic and hard to deal with. We might be experts in our domain, but there is plenty we can learn from the broader ‘data community’. I believe that there are 3 important areas in which simply making observations and model data free and available does not go far enough when we start to think about real users.

1. Data Gravity – At the Met Office we use the phrase ‘data gravity’ to articulate the enormous scale of data produced by modern EPS model simulations as well as a growing issue from emerging observations platforms and the internet of things. Individual Met Services simply blasting petabytes of data into the internet from an FTP site does not help any but expert users and increasingly not even them, even if they can afford to ingest it.
2. Our Data will be consumed by Machines and Developers, not Experts – For the last 70 years or so we have made steady progress in being able to produce data that represents the future state of the atmosphere and its impacts at all timescales ranging from now to decades into the future.
3. We are a global community – We have always recognised we are a single global community and recent events have brought that into sharp focus for many more. Like the virus, weather and climate do not respect geo-political boundaries. At the same time, our users are global. Today, we as a community do not present seamless and global data but rather isolated islands of information.

These 3 issues are simply not addressed by any National Open Data Strategy that I have seen – indeed I am not sure that at least 2 of the 3 can be addressed at national level. I will briefly comment upon all 3.

1. Data Gravity, engaging with the standards body, Open Geospatial Consortium – standards such as Web Map and Web Coverage service allow for dynamic sub-setting of data cubes – on demand and by variable, time and location. Combined with effective API's such as the developing OGC standard for Environmental Data Retrieval this problem is tractable and makes data both available and useful for the majority. But there are also many scenarios that require the analysis of entire datasets. The solution here is to provide a big data 'platform' with massive data storage and co-located computing resource for analysing data in-situ. This is where cloud technology is particularly relevant. We are in the process of implementation of this stack at the Met Office in Public Cloud and this platform will allow for the re-use of Environmental Data at a scale not seen before by 2022.
2. Machines and Developers – not experts. In the information economy, whilst the ultimate consumers of data could still be anybody, the intermediate users will be developers, data scientists and businesspeople. They will NOT typically be weather and climate domain experts and have little if any appetite to unpick the intricacies of absent or poor metadata and information bound up in complex multi-dimensional arrays, understand the statistics associated with effective use of an EPS, or decode our specialist data formats such as GRIB and BUFR. This problem will be hard to solve however common approaches to non-expert friendly meta-data and common approaches to adoption of the standards I have commented on are the route to success. It is already clear that there is also a transition from users being people to users becoming machines and 'other people's algorithms. The Met Office is very active in working with the autonomous vehicle industry and the tech' giants to understand this growing challenge.
3. We are a global community, but we need to act like one, and quickly, working towards clear and common actionable outcomes. If we do not quickly progress as a community to serve the clear need for consistent, useful and global information – others understandably will. It is simply not realistic to expect the NMHSs in 193 countries to each individually develop the capabilities and platforms required – we need a community answer. In Europe we are making progress on the federated approach however I fear that it is too slow. It is possible to create an ecosystem of capability and as a community we might not individually 'own' all the components, but we should at least 'own' the overall design. For each in our own nations we are obliged to ensure that our citizens, industry and governments have access to the best possible information and that will not be delivered by a market-driven and fragmented approach – remember our data are often opinions and not facts!

Given previous comments, my personal view is that the longer-term effect well intended national data policy such as the EU Open Data Directive, has the potential to be catastrophic for some European Met Services without development and action by the community. This is NOT because the call for open and even potentially free is necessarily an issue but that it does not address the more fundamental problem of making an opinion 'useful'. As things stand, the directive is likely to request that all EU Met Services release low-level data and does not seem to have any focus on applying any kind of standards. This is likely to result in more data but that requires even more effort to make it consistent and 'useful' to non-experts. Again, we have the technical solution to this in WIS2.0 for consistency and the emerging Environmental Data Retrieval approach for API's to make the data useful.

NEXT SLIDE PLEASE

“A future data ecosystem?”



Just as ‘Data’ is too big a word, the same may be said of the phrase ‘National Met Service’. We are all similar at a functional level; collecting observations, generating forecasts and issuing warnings to save lives and property. But there is significant diversity within our global community in how we go about achieving this as a function of the national government’s perception of the role of a NMHS and the associated funding that brings. Whilst the technical solutions may be clear, the ability to deliver them is beyond reach for some and delivering them soon enough perhaps beyond the means of most. Building and operating cloud-based data platforms is a complicated and expensive challenge! In the EUMETNET community we are addressing this by developing the concept of ‘Federation’. Simply put, this is about solving the problems of open and available data as a community as opposed to expecting each Met Service to solve the problems independently. At the Met Office we are fortunate to have a ‘critical mass’ of capability that will allow us to implement the relevant technologies quickly and are in the process of doing so. The same is true at some other large Met Services and relevant member organisations such as ECMWF and EUMETSAT.

In closing may I say that I was asked to be provocative to stimulate some action-orientated and outcome focussed activity. In trying to do that in 15 minutes I have tried to generalise some complex issues and probably fallen foul of stating some opinions as facts (even if the metadata is OK!). The Met Office is completely committed to the ongoing work of the WMO and the brilliant community that it orchestrates and represents. We have huge respect for all the hard work that it leads. We are also very respectful of the fact that each Met Service, whatever size, plays a vital role for its government and citizens however at the same time we do see the variation in capability across the community. We recognise that, along with other major centres, we are privileged. With that privilege comes responsibility. The mantra of WMO is often to “leave no-one behind”, and certainly we must all be able to deliver high-quality services to our governments and citizens. Many of us are already content with the idea of handing off work to execute on public cloud platforms operated by others. Crucially we retain the ownership of those workloads: it’s still our work, we’re just running them remotely! Within WMO we see a future the echoes this model. Where we and other World Meteorological Centres provide capabilities such as cloud-based data platforms with access to high-fidelity forecast data upon which individual Met Services can build their own services. This vision of how we might operate in the future is an extension of how we already work today – with data from global modelling centres cascading to regional and national levels. As a community, and with the necessary training and capacity building, we can collectively provide the capability

needed to fully leverage advances in science and technology in delivering our sovereign Public Tasks and not be as constrained by the ability to build and maintain Big Data platforms locally.

NEXT SLIDE PLEASE

[thank you]

 Met Office



Thank you | Merci | Gracias | Спасибо | 谢谢 | شكرا
Charlie Ewen, CIO and IT Director, Met Office