19<sup>th</sup> November 2015: Building up a picture



### <u>Aims</u>

This was the second in a series of three meetings, and the aim of it was to use examples provided by the witnesses to explore ways in which overlaying big data sets and remote sensing can assess and communicate risk and resilience in food supplies and changes in biodiversity.

## <u>Witnesses</u>

**Dr Francois Kayitakire**, a senior scientist at the Joint Research Center (JRC) in the Institute of Environment and Sustainability (IES) from Ispra, Italy flew over to join **Dr Matthew Smith**, an ecologist working in the Computational Science Lab at Microsoft Research and **Craig Mills**, the CEO of Vizzuality.

## Research gaps

Francois outlined the food security programme at JRC and argued that resilience thinking calls for multidisciplinary approach, not only in the qualitative analyses, but also in quantitative models. Food security is an area that needs absolutely such an approach. An emerging focus of development for both them and the EC is to **think about food security in terms of resilience by building safety nets, putting in place response mechanisms and managing risk**. He argued that putting food security in a broader context of the conflicts surrounding resources and the vulnerability of resources is essential for policy and decision makers.

Matthew's group in Microsoft Research was originally founded to apply their expertise in predictive modelling to real world problems such as water security and food security and climate change. He is now concentrating on agriculture and 'agricultural intelligence' where he is working with companies to create geotemporal models that are being used to predict food supplies, represent demand and match the two together to minimise waste in supply chains. He argued that **some of the data challenges we face are fairly basic, simple and mundane** rather than being highly intellectual. Finding ways to better train the next generation to be able to bridge disciplines will be essential. Are there opportunities for better engagement between, for example, biologists and mathematicians, if the goal was problem based rather than discipline based?

Craig introduced Vizzuality and their aim of visualising complex datasets in a way that people can understand and interact with. Recently, they have been working with the World Resources Institute (WRI) to create <u>Global</u> <u>Forest Watch</u>, based on Matt Hansen's Global Forest Change datasets. Typically, **their challenge is to take data out and to strip the information back in order to be able to communicate clear messages**. For example, the aim of a <u>new project with the Zietz Foundation</u> is to take live data feeds from satellites and send information to poor famers in Kenya via their mobile phones to help them to improve their crop productivity. Most of them text only, so key questions centre on ways to present big data on a small screen: How does satellite data relate to what is happening on the ground? How can that information be communicated in a way that helps people to make practical decisions?

### Wicked problems and questions generated by the discussion included:

- How do we build resilient food systems in both developing and developed countries?
- The politics of data are very complex and can be politically changed and politically sensitive which will influence both how the data are collected and how it is used.
- Bringing remote sensing data down to a human scale: There is a disconnect between environmental information and people's understanding and use of that information. There are many new opportunities for open data and services, such as Copernicus, but as yet, there has been relatively little investment in how to communicate the information in a way that people can use to make decisions in the real world.
- Remotely sensed data does not replace the need for on the ground sensors and information, but instead compliments it. Finding ways to be able to support long term, ground and air based datasets will be an essential part of answering the questions we need to ask about food security in the future.
- Decisions are taken at multiple scales from local to international. What place does satellite data have in decision-making at all of these scales and is it feasible to use it to make local scale decisions?
- Is there scope for a growing role for citizen science in this 'new world' of open, big data?
- Although boring, data collection and storage standards are going to become increasingly important if we are going to be able to be able to cross-analyse and layer different datasets. Could lessons to be learnt from the experience of genetic open data be applied to environmental datasets?



Witnesses	
Dr Francois Kayitakire	Food Security Assessment Team in the Monitoring Agricultural Resources Unit (MARS), Institute for the Environment and Sustainability (IES), Joint Research Centre (JRC) of the European Commission (based in Ispra, Italy)
Dr Matthew Smith	Scientist in the Computational Science Lab at Microsoft Research
Craig Mills	CEO of Vizzuality

## **Biographies**

### **Dr Francois Kavitakire**

Francois leads a team working on resilience and on food and nutrition security assessment within the Food Security (FOODSEC) Group. His current activities focus on resilience for food and nutrition security, in particular the resilience measurement issues, food security assessment and classification methods and on agricultural risk management in developing countries. His team provides early warning on food security crisis using various data types and in particular satellite imagery and meteorological data, and they conduct research on modeling food security indicators. His area of interest is mainly Africa but also other developing countries.



Francois' first assignment at the European Commission was within the Unit for Global Security and Crisis Management at the JRC. His work focused on building pieces of an armed conflict early-warning system and understanding their root causes. He worked also on monitoring natural resources that are susceptible to fuel armed conflicts, and on the use of satellite imagery to support post-disaster needs assessments.

Francois holds a Ph.D. degree in Agricultural Sciences received in 2006 from the Université catholigue de Louvain (UCL), Belgium. While working as researcher at the UCL, from 1998 to 2004, he focused on forest mapping and urban green area management using satellite imagery.

## **Dr Matthew Smith**

Matthew works in the Computational Science Lab at Microsoft Research, and is committed to improving societies (people, businesses, governments) abilities to predict geotemporal phenomena (properties and processes that can be associated with geographical space and time). He has worked in both theoretical and applied ecological science since he left highschool and has come to realise the enormous untapped value in predictive models of

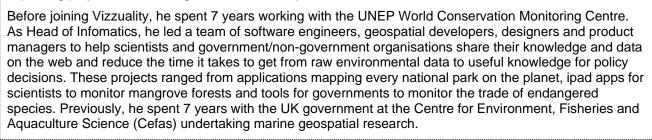


ecological and environmental systems and aims to unleash that potential on the world. In recent years he has also discovered so many other geotemporal phenomena that we can predict, anticipate and make decisions about much better than we have done to date, especially in the domains of agriculture, utilities and energy, to name some major business sectors.

He is currently working on some research projects with UK companies to investigate the value of predictive models of geotemporal phenomena to their businesses. While doing that, he maintains research interests in predicting crop dynamics, carbon and vegetation, human responses to climate change, and ecosystem structure and function.

#### Craig Mills

Vizzuality is a science and technology company focused on data visualization, web-GIS and tool development and committed to working on projects related to conservation, the environment and sustainable development. As CEO, Craig is responsible for figuring out what problems they should be trying to solve, guiding the company towards important world improving projects and working with NGOs to help them tell their stories.





### Setting the scene

This term, we are focusing on food and water supply resilience. We are jointly hosting these three meetings with the Joint Research Centre (JRC) of the European Commission and we are hoping to co-produce a brief white paper for EC policy makers identifying future research gaps and opportunities in this area.

This was the second meeting in the series, and we usedexamples provided by the witnesses to explore ways in which overlaying big data sets and remote sensing can assess and communicate risk and resilience in food supplies and changes in biodiversity.

Our first witness was **Dr Francois Kayitakire**, a senior scientist at the Joint Research Center (JRC) in the Institute of Environment and Sustainability (IES) based in Ispra, Italy. He leads a team working on resilience and on food and nutrition security assessment within the Food Security (FOODSEC) Group which focuses on resilience measurement issues, food security assessment and classification methods, and on agricultural risk management in developing countries.

He joined **Dr Matthew Smith**, an ecologist working in the Computational Science Lab at Microsoft Research whose research focuses on demonstrating how the deployment of valuable new environmental information can benefit businesses and society. He also maintains research interests in predicting crop dynamics, carbon and vegetation, human responses to climate change, and ecosystem structure and function.

Our third witness was **Craig Mills**, the CEO of Vizzuality, a science and technology company focused on data visualization, web-GIS and tool development and committed to working on projects related to conservation, the environment and sustainable development. Current projects related to our topic include developing a new way to visualise forest data for the World Resources Institute and a joint research project with the Zeitz Foundation to explore how to use satellite data and mobile phones to help small scale farmers in Kenya improve their food productivity.

### The questions we asked

They were all asked to choose a particular example from a project or a challenge that sparks their interest and then to give a bird's eye view of:

- 1) The problem
- 2) What we know already
- 3) What we need to know

These ideas then acted as a springboard for a round table discussion until 7pm which continued over dinner.

## People who came to the meeting

Alan O'Neill	Cavendish Laboratory
Alison Smith	Department of Plant Sciences
Craig Mills	Vizzuality
David Coomes	Department of Plant Sciences
Emily Shuckburgh	British Antarctic Survey
Francois Kayitakire	JRC (Joint Research Centre, European Commission)
Graeme Buchanan	RSPB
Gregoire Dubois	JRC (Joint Research Centre, European Commission)
Helen Curry	Department of the History and Philosophy of Science
Hildegard Diemberger	Department of Social Anthroplogy
lan Leslie	University of Cambridge Computer Laboratory
Koen Steemers	Department of Architecture
Kristen MacAskill	Department of Engineering
Marla Fuchs	Research Strategy Office
Martin Rees	Department of Astronomy
Martin Roberts	Cambridge Institute for Sustainability Leadership (CISL)
Matthew Smith	Microsoft Research, Cambridge
Neil Burgess	UNEP-WCMC
Pamela Kennedy	JRC (Joint Research Centre, European Commission)
Paul Linden	Department of Applied Mathematics and Theoretical Physics (DAMTP)
Rosamunde Almond	Department of Applied Mathematics and Theoretical Physics (DAMTP)
Rosemary Ostfeld	Department of Land Economy
Shailaja Fennell	Centre for Development Studes
Stephen Peedell	JRC (Joint Research Centre, European Commission)
Susan Owens	Department of Geography
Therese Rudebeck	Department of Geography



## Word Cloud

Created by using Word It Out - <u>www.worditout.com</u> – based on the transcript of the meeting (edited to exclude non subject-specific words).

### Will be added once the transcript is finished

### Introductions by the witnesses

**Francois Kayitakire** outlined the food security programme at JRC and highlighted the multi-dimensional approach they are taking to look at food crises and food need, based on bringing together information on social, economic, behavioural and diets with remote sensing data. An emerging area of development for both them and the EC is to think about food security in terms of resilience and of building safety nets, putting in place response mechanisms and managing risk.

The areas that he is primarily interested in are the Sahel and the Horn of Africa, both of which are at risk of drought. JRC plays an important role in providing advice and information to EC policy makers and he argues it is important to put food security in a broader context by talking about the conflict for resources and the vulnerability of resources.

**Matthew Smith's** group in Microsoft Research was originally founded to apply their expertise in predictive modelling to real world problems such as water security and food security and climate, for example how vegetation may respond to climate change. He is now concentrating on agriculture and what he termed 'agricultural intelligence' where he is working with companies to create models are being used to predict supply, represent demand and match the two together to minimise waste in supply chains.

Finding ways to better train the next generation in quantitative ecology, environmental science and agriculture, for example, will be essential and he sets Universities the challenge of designing courses to prepare students for the kinds of questions we will need to answer in the future.

**Craig Mills** agreed that commoditising data is a very competitive market and at Vizzuality, their aim it is visualise complex datasets in a way that people can understand and interact with. Recently, they have been working with the World Resources Institute (WRI) to create <u>Global Forest Watch</u>, based on Matt Hansen's Global Forest Change datasets.

Matthew argued that some of the data challenges we face are fairly basic, simple and mundane rather than being highly intellectual. He finds it frustrating situation when we have the intellectual ability to implement much more intelligent solutions but don't have the ability to use them. Craig agreed and added that typically, their challenge is to take data out and to strip the information back in order to be able to communicate clear messages.

For example, the aim of a <u>new project with the Zietz Foundation</u> is to take live data feeds from satellites and send information to poor famers in Kenya via their mobile phones to help them to improve their crop productivity. Although mobile phones are very widespread, most of them are low tech, text only phones rather than smartphones. Key questions therefore centre on ways to present big data on a small screen: How does the satellite data relate to what is happening on the ground? How can that information be communicated in a way that helps people to make practical decisions?

### Wicked problems and questions generated by the discussion included:

How do we build resilient food systems in both developing and developed countries?

There is a huge demand for interpreting and effectively communicating large volumes of data

How can we test how willing people are to change their behaviour?

Is it the research community's job to work out how to communicate their information better or will it / should it fall to external agencies and companies?

Overall, there is a disconnect between the information contained in big datasets and people's understanding of that information. New remote sensing systems such as Copernicus do offer many new opportunities for open data and services but there has been relatively little investment in talking to people about what they want the data for, finding ways to provide it to them and then feeding that information back into the system to make it more effective.



One of the challenges highlighted during the discussion was bringing remote sensing data down to a human scale. There is also a disconnect between remote sensing data and on the ground information, such as household services, both in how it is analysed and how it is used in practice to make decisions in the real world.

Matthew argued that forging connections between the information being delivered and the decisions makers is a risky project for academics and for companies as it is not necessarily publishable or open to commercialisation. How do we find ways to support the blue sky, future facing work?

Keith Richards argued that it is easy to make assumptions that big datasets, particularly open source datasets, are value- free. In fact, the politics of data are very complex and they and their content can be politically changed and politically sensitive. This will influence both how the data are collected and how it is used. Hildegard added that although she can see how useful it may be to make information more accessible, there are also challenges associated with the information itself and what people do with it.

Craig argued that the private sector are already providing information commercially for a price. The starting point is always how to display that information and that is when a technical problem becomes a human problem.

In general, a lot of information is held by individual institutions and companies although some, such as Google and Microsoft, are opening up, it is neither automatic or easy to do.

Trust in the data and its source are very important and it is very challenging to create 'filters' to show what information is important and can be trusted and what is not.

Free and open environmental data are great but it does pose problems for the IGO and NGO sector, for example biodiversity conservation, because they are often expected to provide information free as well. UNEP-WCMC and IUCN get around this by charging companies only for data access. In other words, the data are free but people are charged for the layers of analysis and information that has been put on top of it – the 'business intelligence'.

Matthew argues that there is both a need for, and a consumer demand for, Met Office-style predictive models for biological systems. Paul Linden agreed in principle, but questioned whether it was realistic, given our knowledge of the complexity of biological systems and the layers of uncertainty at each level within it. Matthew replied by saying that uncertainty lies at the heart of the predictive models that Microsoft is producing. He agreed that we cannot re-create biological systems but by incorporating risk and uncertainty and taking a 'bird's eye view' of the system as a whole, he believes we will be able to create effective predictive models.

Francois cautioned against assuming that remotely sensed data replaces the need for on the ground sensors and information, but instead compliments it. Finding ways to be able to support long term, ground and air based datasets will be an essential part of answering the questions we need to ask about food security in the future.

Decisions are taken at multiple scales from local to international. What places does satellite data play in decision-making at all of these scales? Hildegard then asked, to what extent can satellite information be used to make local decisions?

Craig stressed that although boring, data collection and storage standards are going to become increasingly important. Even basic levels of compatibility are not yet in place to be able to cross analyse and layer different datasets.

Putting data in the public domain is a risky business, both in terms of what will be done with it and whether funding will be there to maintain it.

If there is a hypothetical £10 million available, is it best to put it into satellites and research to support our understanding of the environment, or is it better to invest in green initiatives and projects?

Is there scope for a growing role for citizen science in this 'new world' of open 'big' data? Graeme argued that currently, it is very hard to use citizen science to collect long term, comprehensive datasets. There are initiatives, such as ispot (which identifies wildlife and has over half a million contributors). However, the experience of conservation NGOs such as RSPB and BTO is that although very useful, citizen science tends to yield data that is patchy in both space and time so more work is needed to overcome these challenges.

Getting hold of data is expensive but curating, storing and managing it is as well. How do you engineer consistency across datasets? How can you engineer a culture of open data and also put in place the infrastructure to be able to support it?



Alison argued that systems to manage large dataset and to standardise and collect consistent information was put in place 20 years ago for genetic datasets. Are there lessons to be learnt from that process for environmental datasets?

## Key points people took away from the meeting

### **Communication**

There is an urgent need for data and information to be 'processed' / 'translated' into a format which can be digested by the wider public and more specifically by decision-makers. There appears to be no shortage of funding to develop the technology to acquire data, but, too little effort put in to transforming data into relevant and accessible information in a timely manner. Are there ways in which this 'translation' service can be promoted and provided?

Modernising data collection methods and science communication to fit with the development of communication technologies (use of smartphones and crowd-sourcing). Analytical methods for data collected in such settings are not yet well developed.

The Vizzuality approach was the most striking to me - make something intuitive and easy and people are much more likely to use it, even if it isnt perfect. I have sent the link for survey monkey to Francois to show him how a simple template can be used for surveys - as well as making it easy to design and complete, the standard format means that the data from several different surveys can be combined more easily.

**Visualisation of information is crucial.** Not only is this valuable to communicate the key messages more effectively and easily, but our brain can often be better at identifying patterns in images and graphics which computers may struggle with.

It is amazing that so much satellite data is potentially available, but, it is either: (a) not captured from the satellite or (b) not made available in a form that people can readily access.

The need to develop advanced data visualisation techniques. (Craig) The amount and heterogeneity of data, as well as their complexity, demand imaginative approaches to data visualisation. Before that we need to find the information in the data that will prove to be important, and we often don't know what we are looking for. With its skills in machine learning, Cambridge could contribute significantly to this data exploratory activity with a view to bringing out the information content by tying into advanced visualisation tools, such as are being developed in industry (and possibly in this University).

#### Monitoring the environment

Wherever you go, you hear about the reduction of resources to maintain environmental monitoring stations. There is the danger that long-standing measurements/surveys are just stopped, without having an alternative system in place. Crowd sourcing was proposed as a substitute, but, we are far from being able to substitute long-term monitoring networks with data collated from crowd sourcing. Will there be data gaps in the future, before a solution is found?

### Taking risks?

Given the current status of the European economy, it was striking to hear how reserved the private sector is, when it comes to taking risks associated with development. It is well-known in advancing science and technology, that for every successful development there are always multiple developments which do not take off. Is the community thus being stifled by a weak economy (and limited resources for research and development), in terms of innovation? How can we tackle this? Given that the freedom we once had is much reduced, should efforts be placed/focused on improving Institutional partnerships (public-private), multi-disciplinary long-term programmes and finding other ways to reduce risks by creating buffers.

One area we explored was the **impact that the global recession was having on public tax revenues and public funding for the current generation of huge data gathering platforms**. It's likely this trend will continue and we discussed the pros and cons of private sector funding for their more specific needs and the sectors growing capacity to analyse this data.



### Training the next generation

**Finding ways to channel young researchers/scientists' ideas into the policy process**. It seems to me that there is such a rich resource of ideas and opinions that we do not tap into when listening to public opinions and stakeholders' views. It's probably something we need to take to Rob Doubleday for discussion, but, I really think, at least for what I saw in relation to Agriculture and the CAP, the need to find ways / routes to channel the young scientific communities' views into policy making. In the UK, this may have to be done through the existing channels of the Ministry, or the Farmers' Union, but I think there is a strong justification to find a voice and a conduit for the opinions and fresh ideas of young scientists. I don't have any solutions, but, would be happy to work on this to try to find how to move this idea forward, if it meets with interest in the Cambridge side.

My question about **how we better train the next generation in quantitative ecology / environmental science / agriculture** was not fully answered. For example, I have been struck in the past that while Cambridge University has one of the best ecology/biology/environmental science degrees in the world, the graduates from those degrees are still woefully under-prepared for the kinds of work and analysis that is needed. That's a major bottleneck.

I liked Matthew Smith's description of mathematically competent people as "Quants" having not heard the term before. On the other hand, I wonder if there is an opportunity for better engagement between biologists and mathematicians if the goal was problem based rather than discipline based.

The lack of trained people on the use of satellite and other data, including meteorological data. (Francois) This is big issue, of course, not just in the area of sustainability and environment. I think this University could have a concerted response to this shortfall. I'm personally looking with my colleagues in the Centre for Scientific Computing to develop capabilities and training in the use of relevant machine learning techniques and "big data" analysis.

### Taking a multi-disciplinary approach

**Big data**: **there's a need to consider multi-source and multi-scale big data**, i.e. combining data from different disciplines and scales in the same analytical framework. Resilience thinking calls for multi-disciplinary approach, not only in the qualitative analyses, but also in quantitative models. Food security is an area that needs absolutely such an approach.

**Connecting the tech folks with the human folks seems critical**, how to create the structures within the university to get good cross discipline research (I'd guess it's already happening..)

The broad range of end-users and communities is leading to the rapid multiplication of various information portals while we need **more integration to properly address the complexity of the world**. This is especially true for conservation issues which require much stronger links with the actors involved in the global competition for land.

### Food security and tracking progress towards the Sustainable Development Goals (SDGs)

The idea being developed by Nazia about **building scenarios and roadmaps to the SDG is extremely interesting**. We're in the process of defining the indicators and targets of the SDGs. In many cases, the scientific evidence of the targets is very weak. This can be a research topic for MSc and PhD theses.

I liked Francois Kayitakire's description of how **providing social protection rather than food aid is more appropriate for establishing resilience**. To those in the West it is often hard to consider what difficulties people in developing countries need under normal circumstances rather than just in times of crisis.

Matthew from Microsoft made a reference to the **pressures in the local supermarket supply chain and how shipments are made at short notice just to meet requirements**. That really reinforced to me a huge problem that the UK could take action on.

#### Storing, managing and co-ordinating big datasets

Just how basic our data challenges are. These are not highly intellectual. They are simple and mundane. It's a very frustrating situation when we have the intellectual ability to do so much more intelligent solutions but don't have the ability to use them.



Solving the problem of patchy data in time and space seems important if we are to provide good information for sustainable planning.

**More attention needs to be put into the long term management of data** (all kinds, from Earth Observation to citizen science) to ensure open access and services matching the very different needs from different end-users. If we want to become more resilient to changes, we must ensure more solid foundations to be prepared for addressing a broad range of possible scenarios

Data are only as good as what users will make out of it. There is an underestimated responsibility in both communities (providers and end-users) in defining the way information has to be produced and used.

**Big data: questions about its resilience, consistency and value.** 'Resilience' in terms of how it is stored and curated, 'consistency' with regards to how gaps in data can be avoided, and 'value' in terms of whether the data actually address the questions we are looking to answer. The latter point is something that I have often wondered, particularly as 'big data' has become more fashionable: Just because we can and are collecting more and more data, that does not mean it is helpful and may in fact be leading us up blind alleys. Big data can be valuable but it is essentially only a 'dumb' resource and in some cases has become a discipline in itself!

The need for and opportunities afforded by cloud computing if that can be co-ordinated or consolidated appropriately. It would be massively helpful if we had a virtual computing space to access data, develop algorithms and apps, working with developers and customers in an open environment. (Matthew) A number of groups are developing platforms, but their user base is likely to be small, because they have not captured the market in this area. The sense I have is that a big player needs to step up to the plate, with Google and Microsoft being obvious contenders. It seems to me that Cambridge should develop a view about this, with the possibility of forging a major partnership with one or the other (if not both). My inclination would be Microsoft.

## The politics of data

I regretted that we didn't get into this more, or press the speakers to be more critical themselves about the values that are coded into the seemingly neutral numbers that they crunch. It seems to me especially that Microsoft, with its great ambitions for data provisions, should surely have to think about this aspect in greater detail than was evident in the discussion and responses.

### Joining the dots

More research is needed to find the **best ways of using data to support small scale localised decision making**.

How in practical terms can we **harness global citizens in a two way conversation** about landcover, health, farming and biodiversity change?

How can we make the **best use of new remote sensing and new on the ground technologies** to better understand change?

How can we turn remote sensed maps of changes in tree cover at 30m or 10m scale to maps of changes in a) biodiversity b) firewood c) medicines d) livelihoods?

Land cover mapping is getting more sophisticated so that specific elements can be targeted and different maps could then be layered on top of each other. Yet, there is an argument that satellites still can't really capture the true biodiversity of a location.

## Using machine learning to classify different types of agricultural crops using satellite imagery

Storing all agricultural concession data on the Cloud, then creating algorithms to determine if companies are deforesting illegally. This would be similar to the Global Forest Watch platform, but for all agricultural commodities, and storage on the Cloud could provide a layer of anonymity, unless a company is violating laws / planting outside of concession boundaries

**Creating a "universal" land classification system**, then developing a smartphone platform so the general public could participate in "citizen science" and take photos / videos of land with GPS locations. This data could then be used in tandem with satellite data to classify larger swaths of land.

**Top-down v bottom-up data**: The links between satellite and 'ground source' data can be very weak. Mismatches between simplified large scale categories (e.g. climate type, biodiversity, demographics, etc.) and



what is observed on the ground (local microclimates, species and cultures/traditions) are not uncommon. Techniques to match up these different types and sources of data are important, and possibly crowdsourcing and low-tech phone messaging can be helpful.

The most interesting elements were the **clear belief that predictive models of land cover change can be built, and the fact that these models are capable of taking more into account than remote sensing variables**. Integration of socio-economic data for example. And the positive of this is that these will be mechanistic models rather than necessarily mathematical models.

But the topic remains that these models will be running before we can walk - **we still do not know what is happening across much of the globe**, and that basic information is more useful than predictions of what will happen. In the analogy of the evening, we need to know if it is raining now, rather than whether it might rain at 11 or 11:30 next Thursday. We should be able to look out of the window and decide if we will get wet today.

