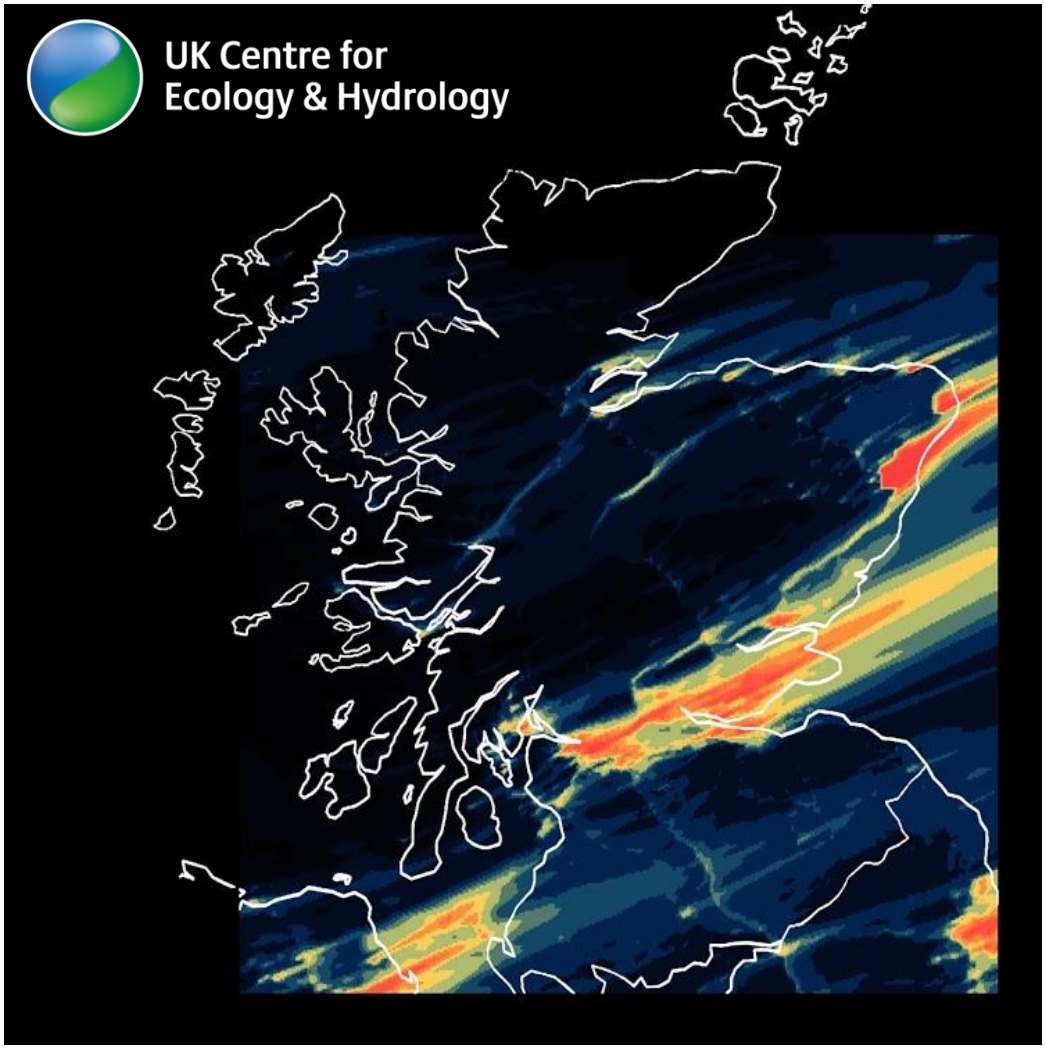

Stakeholder evaluation of modelled EMEP4UK gridded UK atmospheric composition data

UK-SCAPE
UK Status, Change and Projections of the Environment

Jan Dick and Massimo Vieno

March 2021

UK-SCAPE
UK Status, Change and Projections of the Environment

Title Stakeholder evaluation of modelled EMEP4UK gridded UK atmospheric composition data

Client UK Research and Innovation (UKRI)

Client reference NERC National Capability LTS-S: UK-SCAPE; NE/R016429/1

UKCEH reference UKCEH Project: 06948

UKCEH contact details Jan Dick
UKCEH Bush Estate, Penicuik, Midlothian, EH26 0QB

t: 0131 445 8578
e: jand@ceh.ac.uk

Author Jan Dick and Massimo Vieno

How to cite Dick, J. Vieno, M. (2021) Stakeholder evaluation of modelled EMEP4UK gridded UK atmospheric composition data. UK Centre for Ecology & Hydrology (UKCEH Project: 06948; NERC National Capability LTS-S: UK-SCAPE; NE/R016429/1). 21pp

Contents

Summary

Nine academic and government agency staff where interviewed on their opinion of the EMEP4UK gridded UK atmospheric composition data supplied by Massimo Vieno in connection with WP2.1 of the UK-SCAPE program.

The aim of the interviews was to understand from a user's perspective the utility of the UKSCAPE modelled EMEP4UK gridded UK atmospheric composition data and enable co-production of improvements. Specifically the interviewees opinion of four aspects were determined

- Co-producing impact from this science through Art/Science collaboration
- Ease of discovering EMEP4UK gridded UK atmospheric composition data
- Data delivery and user co-design improvements and finally
- How the data was used and acknowledged.

The interviewees priorities improvements including

- Providing a snippet of R and/or python code to accompany the data
- Providing a video of the data showing the spatial and temporal variability of a specific pollutant
- Providing a persistent identifier for the data e.g. DOI
- Providing an application programming interface enabling connection between computers or between computer programs i.e. API

Snippets of R and python are now under development within the WP2 and implemented in the DataLab (part of the WP2 Data Science Framework) and the other three suggestions will be included in future funding applications.

1 Aims

The aim of this study was to understand from a user's perspective the value of the UKSCAPE modelled EMEP4UK gridded UK atmospheric composition data and co-design improvements. The UKSCAPE Program Executive Board on 29 June 2020, during the pre-mid-term review meeting, agreed that the 12 individuals from the eight academic groups that had requested data from Massimo Vieno should be contacted and invited to offer their opinions to fulfil these aims. Essentially three aspects of the EMEP4UK gridded UK atmospheric composition data were investigated:

- Stakeholders' opinion of discovering the data,
- Stakeholders' opinion on how the data was delivered to them and their views on future data delivery options, and
- Stakeholders' feedback on how the data was used and acknowledged.

In addition, stakeholders view on the use of the EMEP4UK gridded UK atmospheric composition data in an arts project was sought (https://inspace.ed.ac.uk/everyone/).

2 Partipcants

Twelve individuals from eight institutions who had requested data from Massimo Vieno prior to July 2020 were contacted via email on 8th February 2021 (Appendix 1). Reminder emails were sent on 15th February 2021 to the individuals representing institutions who had not responded; University of Lancaster, Marine Scotland, and University of York (Appendix 1). The reminder emails resulted in no response so they were contacted by telephone when possible and Massimo Vieno sent personal emails. The representative from York University responded and an interview arranged. The representative from Marine Scotland was on maternity leave so Massimo Vieno offered a second contact – he was on paternity leave but agreed to be interviewed on his return to work (29/04/2021).

The representatives from University of Lancaster both replied that they had not used the data. One commented, "*I haven't actually gotten around to looking at this data yet*" and the other replied "*Sorry for the delay in responding. I too haven't got around to using this data yet".*

Partipcants in this study were offered anonymity (Appendix 2) so only the institution and number of staff interviewed are reported here (Table 1).

It is relevant to note that Massimo Vieno is a participating researcher and co-author with three of the nine interviewees.

EMEP4UK gridded UK atmospheric composition data is useful to many communities resulting in UK-SCAPE data being used in a range of disciplines including monitoring flows of pollutants, hydrodynamic models and in connection to human health (Table 1).

Table 1. Details of institutions and number of individuals who have requested and been provided data by Massimo Vieno.

Institute	Type of data	Data size	Time covered	Usage	Number of contacts
London School of hygiene and Tropical medicine	Daily PM2.5	~40 Gb	2015-2018 2003-2014	Human health impact of air pollution assessment	2
University of Exeter	NO2, NO,PM2.5, PM10 and Ozone	~300Gb	2001, 2004, 2006, 2008, 2012 and 2015	To cross-link with the Millennium Cohort Study to examine the association between higher chronic exposure to air pollutants, namely O3, NO2, NO, PM2.5 and PM10 from birth and the rates of asthma development.	1
Environment Agency	Animated hourly SO2 and PM2.5	~100 Mb	2014 and 2015	Identifying SO2 plumes that can then be detected by satellite	2
University of Edinburgh	Global atmospheric composition	~1 Gb	2010	PhD student – Analysis of nitrogen fluxes across the globe	1
Marine Scotland	WRF hourly met data	~3Tb	2011, 2012, 2013	Scottish Shelf Model (http://www.marine.gov.scot/themes/scottish-shelf-model) running a hydrodynamic model for a specific time period	1
University of York	Annual AOT40	~1Gb	2001-2018	Identifying interesting years for a proposal	1
University of Leeds	Hourly PM2.5	~50Gb	2010-2016	Impact of weather types on UK ambient particulate matter concentrations	2
University of Lancaster	Hourly Ozone	~50Gb	2001-2015	They are exploring the inter annual and weather-dependent variability of surface ozone over the UK – but had not used the data and declined to be interviewed.	2

3 Interview protocol

A structured interview protocol, co-designed by Jan Dick (UKSCAPE Lead WP5) with Massimo Vieno (UKSCAPE Task lead WP2.1) was employed. The first author conducted all interviews collating both qualitative and quantitative data. The interviewees were asked to score each of 18 questions grouped into five themes (Fig 1) and asked to provide their thoughts and reasoning for their score. The questions were formulated as either scaled Likert type questions (scale 1-6) or categorical response classes. In addition, a final open question asked for any further comments.

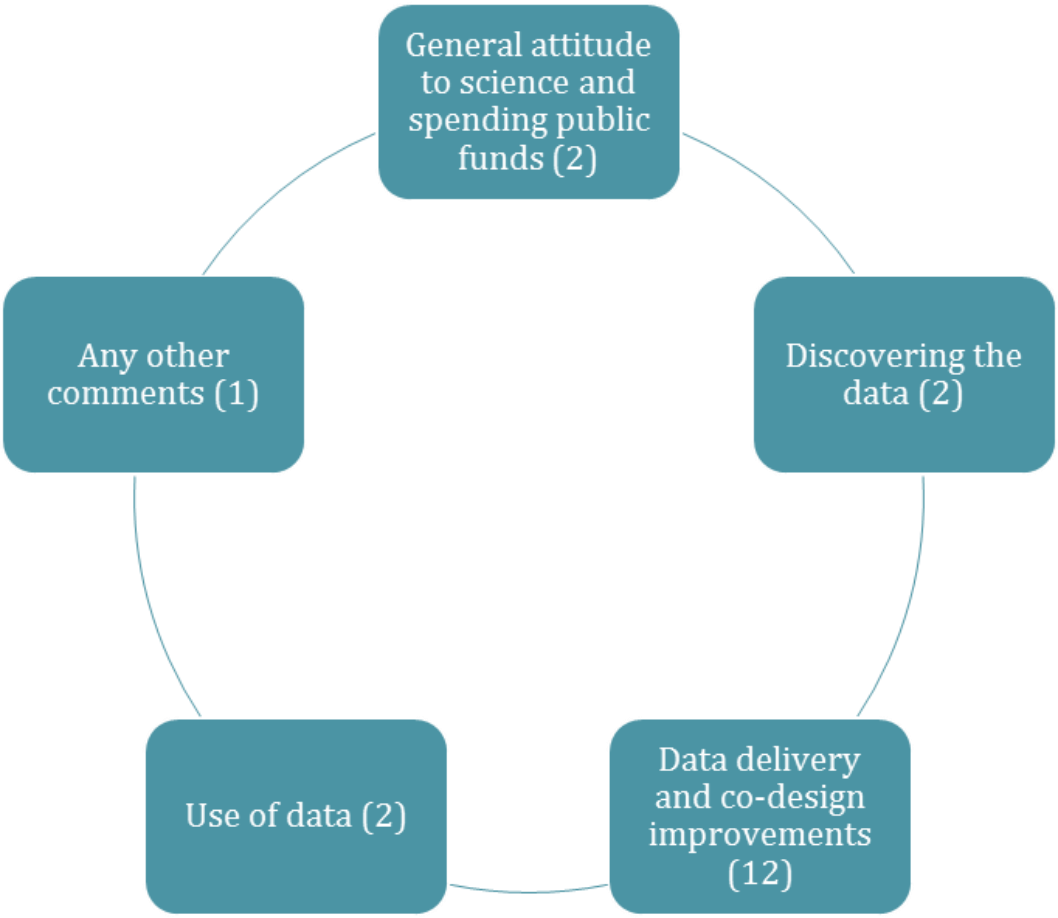

Figure 1. Five themes and number of questions in parenthesis forming the structure of the interview protocol.

A random order was not considered necessary rather the questions were asked in the same order for each interview (Table 2). The interviews were conducted either by telephone (1) or via zoom (8). The data was collected in a spreadsheet (excel) and returned to the interviewee for verification. This latter step was not consider important for the interviewees who were interviewed via zoom as the recording and transcription services of Zoom were employed. It was however considered important for the interview conducted via telephone when no recording was possible and therefore for consistency all interviewees were send a transcript of their interview.

The interview commenced with the interview giving verbal agreement that they had read the 'Participant information and consent sheet' (Appendix 2) and were happy for the interview to be recorded if conducted via Zoom (all agreed).

Two initial questions in the theme 'General attitude to science and spending public funds' were asked focused on the video animation,' Everyone', which is one of five art-science collaborations in an online exhibition organised by Inspace, the project space of the Design Informatics research institute in Edinburgh (https://inspace.ed.ac.uk/everyone/). The next two questions considered how the interviewee had found the EMEP4UK gridded UK atmospheric composition data. Twelve questions were devoted to the theme of data delivery as this theme was consider most important to foster co-design of future developments. Two questions focused on the theme of data use and acknowledgement and a final question invited the interviewee to add any other comments related to evaluating or improving the UK-SCAPE service.

Table 2. Questions and numerical scale utilised in interviews.

Theme	Question	Scale
General attitude to science and spending public funds	Do you think the video is a useful way to communicate science?	1= total waste of time and money 6 =great visually and thought provoking
	What do you think of mixing art and science generally?	1= not useful it degrades the science 6=excellent all ways of engaging people in science is to be encouraged
Discovering the data	How easy did you find the data?	1= it was really hard to find - I just tripped over it; 2= Cold search on a bibliography; 3= a colleague/supervisor told me to contact him; 4= I know Massimo's work already
	How easy did you find contact details for Massimo?	1= very difficult to 6= very easy
Data delivery	Was the format of the data delivered appropriate?	1= yes exactly what I wanted 6=no I had to do a lot of work to make the data useful
	Was the NetCDF format useful ?	1= yes exactly what I wanted 6=no I had to do a lot of work to make the data useful
	Would a more window friendly format be useful?	1= yes exactly what I want 6=no I would still have to do a lot of work to make the data useful
	How useful would a website to download .cvs files be?	1=no value 6= brilliant exactly what I would want
	Would you like a website which would do simple calculations (for example adding two gridded layers, calculate UK or sub area average, max/min, etc. ?)	1=no value 6= brilliant exactly what I would want

	Would you like a machine readable format to obtain the data?	1=no value 6= brilliant exactly what I would want
	Would you prefer a snippet of R code ?	1=no value 6= brilliant exactly what I would want
	Would you prefer a snippet of Python code ?	1=no value 6= brilliant exactly what I would want
	Would you prefer a API ?	1=no value 6= brilliant exactly what I would want
	Would a DOI associated with the data be useful?	1=no value 6= brilliant exactly what I would want
	What would be a useful way for Massimo to make the data visual for you - a video showing the spatial and temporal variability of a specific pollutant?	1=no value 6= brilliant exactly what I would want
	What would be a useful way for Massimo to make the data visual for you - A processed UK map	1=no value 6= brilliant exactly what I would want
Use of data	How did you use the data?	1= did not use the data; 2= small contribution to a larger study - was not a critical dataset in the work; 3=critical for the work; 4= critical for a publication
	How were the results disseminated?	1= not disseminated
		2= to our local group
		3= to a national/international meeting
		4= in a report/thesis
		5= in a non-ISI paper
		6=in a ISI paper
Other	Themes	

4 Results

4.1 Number of responses and interview duration

In total, 11 individuals responded (92%) when contacted and nine individuals (75%) agreed to be interviewed from seven institutions. One invitee did not respond but his assistant represented the institution. The interviews were conducted primarily between 08/02/2021 and 16/3/2021 with a final interview on 29/4/2021. Only one interview was conducted via telephone and all the others were conducted via Zoom. The interviews lasted between 15 and 40 min with an average across all nine interviews of 30 min (Fig 2).

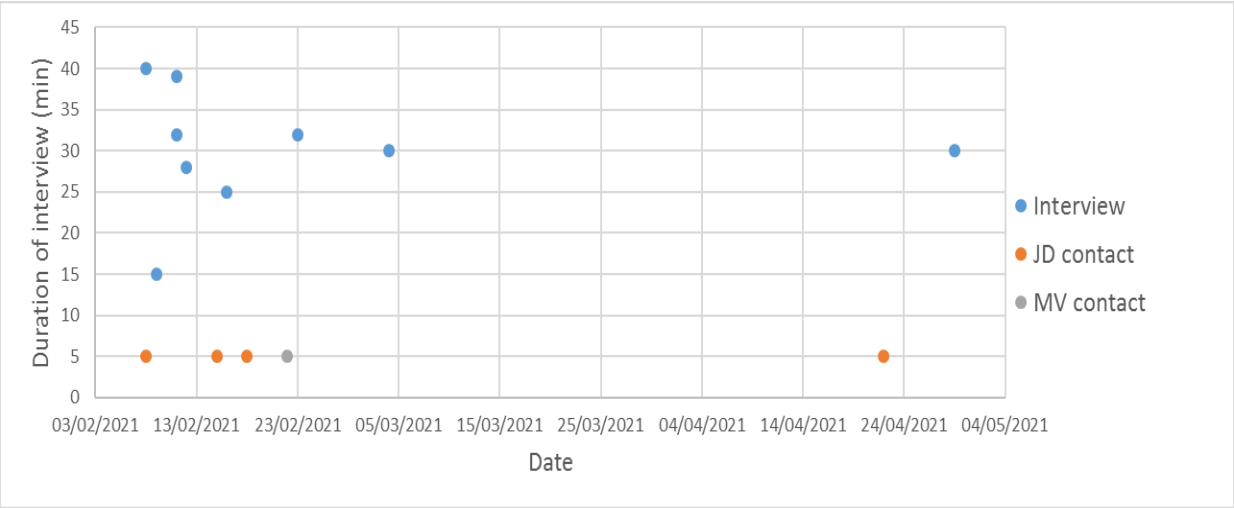

Figure 2. Date and duration of nine interviews with the dates (marked at 5 min point) when Jan Dick (JD) or Massimo Vieno (MV) made email or telephone contact.

Some individuals declined to answer some questions as they were outside their areas of expertise.

4.2 Themes

4.2.1 General attitude to science and spending public funds

One respondent interviewed on the telephone had not watched the video and so skipped the first two questions. The majority of the interviewees (88%) considered the 'Everyone' art installation that depicts reduction in air pollution in Edinburgh during Covid 19 lockdown was a useful way to communicate science (i.e. scored the question 4, 5 or 6 on the 1-6 Likert scale). Only two individuals however scored this question as a 6 i.e. great visually and thought provoking (Fig 3). The others thought it good but each qualified their answer to some extent (Table 3). Some found the art work unclear for example requesting a date stamp when the lock-down actually happened. Another commented on the lack of units and suggested that a voiceover might have been useful.

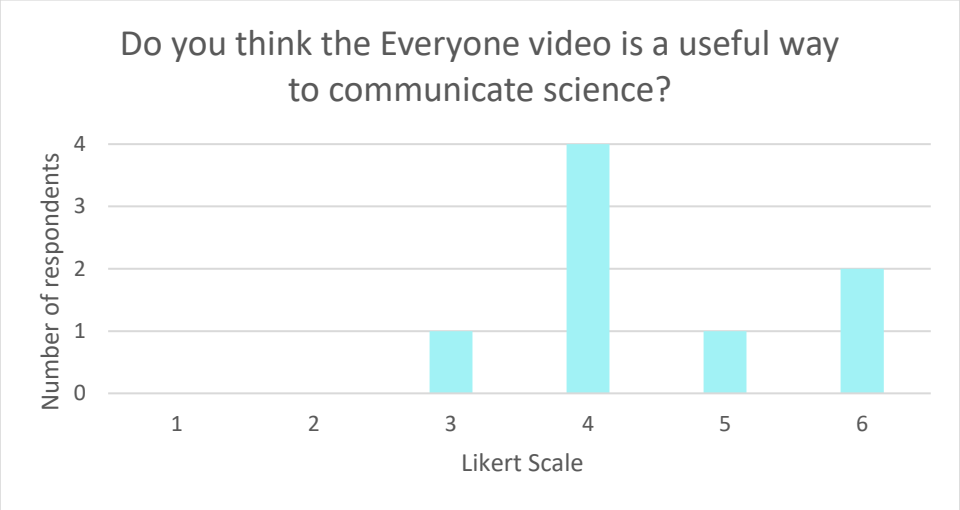

Figure 3. Response of interviewees (n=8) to the question Do you think the video is a useful way to communicate science? Likert Scale 1= total waste of time and money 6 =great visually and thought provoking.

Table 3 Narrative responses to the questions related to General attitude to science and spending public funds; Do you think the video is a useful way to communicate science?

Respondent score	Rationale for score
3	*I don't think I understood .... I didn't realize it was showing a trend I couldn't figure out when I was watching the video, and I think it was overlaying actual map and I didn't notice the date, in the top left hand corner, when I first watched it. I had to watch it a few times to realize that there was actually a timestamp at the top. So I thought the time series of the EMAP model clear, but the other stuff wasn't clear to me.*
4	*I think that the video is a potentially extremely useful way to communicate science to an audience who might not normally find out about these things. I think that there are concerns about presenting air quality data to an audience that may not have actually seen that data before cause things are very easy to misinterpret, in terms of what is a safe level and actually how the images on the video might effect an individual. I think it's extremely good idea , but I think care has to be taken when communicating data such as that to an external audience.*
4	*It was definitely usable yeah and especially for a lay audience, it can help get them interested in the topic, which is probably one of the issues with many scientific studies. They can be very interesting from a research perspective but it is difficult to get the public to understand them, and be interested in them, so I think it's definitely a positive interesting way to communicate science.*
4	*It wasn't clear what it was, maybe I missed a bit of text, but there was no units there's no description of what the field was that you were showing... I don't know daily precipitation or something. It looked pretty but it was some kind of atmospheric data, but it was*

	pretty meaningless to me, as a scientist.... You almost need a bit of a voiceover to explain what it is.
5	*Well, it is a good visualization and it's nice to look at but one issue that I would have with the video is because of the tapestry, or whatever running around the borders of it. The date was quite small on the top left hand corner, and so I find I had to keep going searching for the date finder. I just thought it could have been made more clear like this is the start of the clock-down. Just so you could keep better track of it and then maybe also they could have put a bit of text saying that lock-down begins on the day then you could have seen ..look at the immediate change in the air pollution.*
6	*I don't hear any music … my qualified statement is that I sort of feel like the more different ways that we can be presenting science the more likely it is that those different ways they're going to get picked up by a variety of different people who will, be more amenable to interpreting information in different ways. Probably for too long we have relied on papers and text and reports and that kind of thing, and I think this kind of thing is just great it just allows, especially in today's digital age, it allows greater visibility and I think it's all about accessibility.*
6	*I would make a point that it might be too long, but yeah it was useful definitely for communicating the science and how much of a reduction there was during the COVID*

Several interviewees, although they were critical scored this question above half way because they thought mixing art and science was in general a good thing.

This view was echoed in interviewees' responses to the second question "What do you think of mixing art and science generally?" All responds score this question as either a 5 (38%) or 6 (62%) on the 1-6 Likert scale. The visual aspect of the artwork was valued by the interviewees with one commenting *Much easier rather than reading a publication… better than wall to wall text.* Another responded suggested that perhaps more thought was required to *making it accessible for everyone …other ways of explaining data without just using visual and audio cues.* Another considered that it was important to communicate especially if the science impacted on people commenting *I think it* [air pollution] *affects all of us, regardless of our class or cast or anything so it's good if it's being used to communicate to the public.*

While positive one responded did remark on the need for funding such collaborations and another noted the disconnect between science and the public commenting: *I think we can be doing as much science as we'd like, but if it's not being picked up and accepted and understood by people, then it really isn't going to make much difference.*

Another interviewee was inspired after seeing the video and commented *It also made me think about if there are any ways I could interpret some of the results of the research I'm doing at the moment and attempt to communicate those to a wider audience.*

4.2.2 Discovering the data

All respondents answered both questions related to the theme of data discovery. The majority (89%) were recommended to contact Massimo by a colleague or knew him personally before requesting the data (Fig 4).

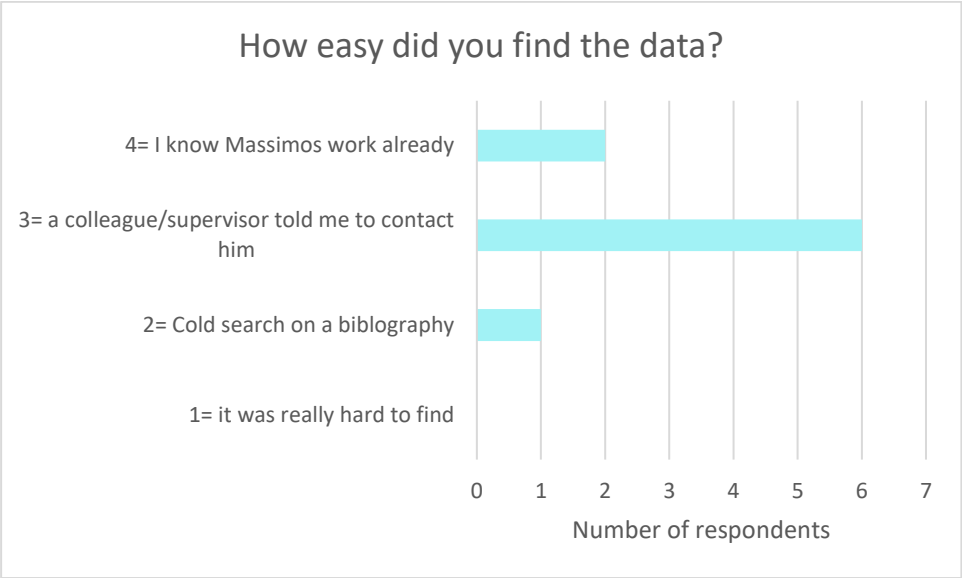

Figure 4. Response of interviewees (n=9) to the question “How easy did you find the data?” Likert Scale 1 to 4.

One respondent commented that they had found the data via the website (Table 4) and then contacted Massimo.

Table 4 Narrative responses and Likert score to the question “How easy did you find the data?” Likert score 1= it was really hard to find - I just tripped over it; 2= Cold search on a bibliography; 3= a colleague/supervisor told me to contact him; 4= I know Massimos work already .

Respondent score	Rationale for score
2	*I found the website online, so it will be 2, but it didn't have the data that I needed so then one of my supervisors had Massimo as a contact and put me in touch with him to get the data that I used.*
3	*actually when I started working this project with the ... professor that you contacted...They were already in conversations about using this data so basically because i'm expert a specialist dealing with this kind of data [it] was just a matter of getting in touch with Massimo and introducing myself in such a chat with him about this data, so I actually didn't have to search for it, this was already handled by Professor*
3	*my supervisors told me about Massimo and his work and that's how I contacted him*
3	*So in this particular case, it was through a colleague number three.*
3	*some colleagues suggested the use of this data, and it was very easy to find, especially because of the website... And easy to find*

	documentation on that and what it was about etc in a second, period I think we started digging into it, especially because it was fundamental, I will say for the.development of our project … we got in contact with the Massimo. directly , we were able, also to get more details and also of course he's a participation and contribution to the project and the way that we're using in our project, which I think it was critical.
3	*one of my supervisors is [UKCEH] colleague of Massimo*
3	*through contacts, colleagues, so it was actually a colleague who originally got in touch with is it Massimo ...And I, so I must admit I don't know how she made contact with him. Properly just through contacts professional contacts and yeah.*
4	*I mean, to be honest, I didn't even know that this project existed, but I was just asking Massimo for some at AOT40 data and then he sent it almost the same day. So it was it was definitely because I know that's what he does*
4	*I know Massimo by reputation; Massimo presented at the uk-usa air quality workshop about 4 years ago - an annual DEFRA, EA and US Environmental protection agency workshop*

All respondents reported that they found the contact details of Massimo easily (i.e. 100% scored this question as a 6 on the 1-6 Likert scale (1= very difficult 6= very easy).

4.2.3 Delivery of the data

Several interviewees declined to answer some of the questions in this theme, as they were not the person actually working with the data and so preferred not to provide input to the co-design process. The first questions asked about the current delivery of the data "Was the format of the data delivered appropriate?" Analysis revealed that 50% of the eight respondents consider that the data was exactly what they wanted (Likert scale = 1) and another three scored this question as a 2 almost exactly what they wanted (Table 5). One respondent reported they had quite a bit of work to do before they could use the data (Likert scale =5).

Table 5 Narrative responses and Likert scores to the question "Was the format of the data delivered appropriate?" (Likert Scale 1= yes exactly what I wanted 6=no I had to do a lot of work to make the data useful)

Respondent score	Rationale for score
1	*Again, that was easy and great procedure with Massimo. He was always very helpful ,*
1	*That's a one*
1	*Oh well, I didn't do the processing myself. Member of my team did it but from the feedback I have from her the data were in perfect shape and form of communication was very precise and comprehensive.*
1	*I was sort of a bit of a broker in some ways I passed the data on to a project that we were working on… i'm not quite sure how far they went with actually using that data.*

2	*2 it was very nearly the information that we needed to go.*
2	*I was learning about how to do all the analysis and how to look at the data, so I think that will be really subjective, for me, but I think that once I got the hang of it was fairly easy to look at the data. Also Massimo shared his Python script as well, so yeah it was not very hard to take the data from what he shared.*
2	*What you gave us was really useful. But then the nature of what we do is we're always going to want to reformat it and work on the data to get it to do what exactly what we want. And I would never I wouldn't really expected Massimo to have provided it in the format that we want, so why should I expect him to do my job.*
5	*so I got NetCDF file format which I wasn't really ..I have never come across that before, so I had to do quite a bit of work in ARC MAP to convert it, and the only issue then with converting it was how slow and unresponsive Arc MAP as a program I don't know if it was my laptop, but it was just tedious ...But then I had to change it on the raster layer and then it was simple enough that I just attached it to the LSOA area codes - lower super output area*

The following question asked more precisely “Was the NetCDF format useful ?”. Only 6 interviewees answered this question and again four scored one on the Likert Scale i.e. yes exactly what I wanted. However even although they scored this question highly some recognised that the format was specific and perhaps other formats could be useful for example one respondent commented *I believe that it is the smartest but is of course not, for all the communities in academia.* This sentiment was echoed by two other respondents who scored this question as a 3 and 5 i.e. they had to do a lot of work to make the data useful (Table 6).

Table 6 Narrative responses and Likert scores to the question “Was the NetCDF format useful ?” (Likert Scale 1= yes exactly what I wanted 6=no I had to do a lot of work to make the data useful)

Respondent score	Rationale for score
1	*one*
1	*one again*
1	*NetCDF is the format that Massimo uses ... I believe that it is the smartest but is of course not, for all the communities in academia, but is the easiest way to handle this data.*
1	*It was what I needed. I was just looking for a bit of a higher resolution, but I think that's something Massimo who had already provided the information about that… is it going to be one degree. I wanted a little higher than that but yeah I think what the model gives and what Massimo initially told it was , it was in the right format and I could look into it..*
3	*it's useful because I have access to people that can very easily use that NetCDF files.. If I didn't that I probably perhaps found it more useful as a comma delimited file, for example, but then I accept that probably takes a lot more space.*

5	*obviously contained everything I needed but yeah I needed to convert to make it usable format for myself.*

The questions asked about different formats and other potential deliverables revealed a range of responses depending the skills of the individual (Fig 5).

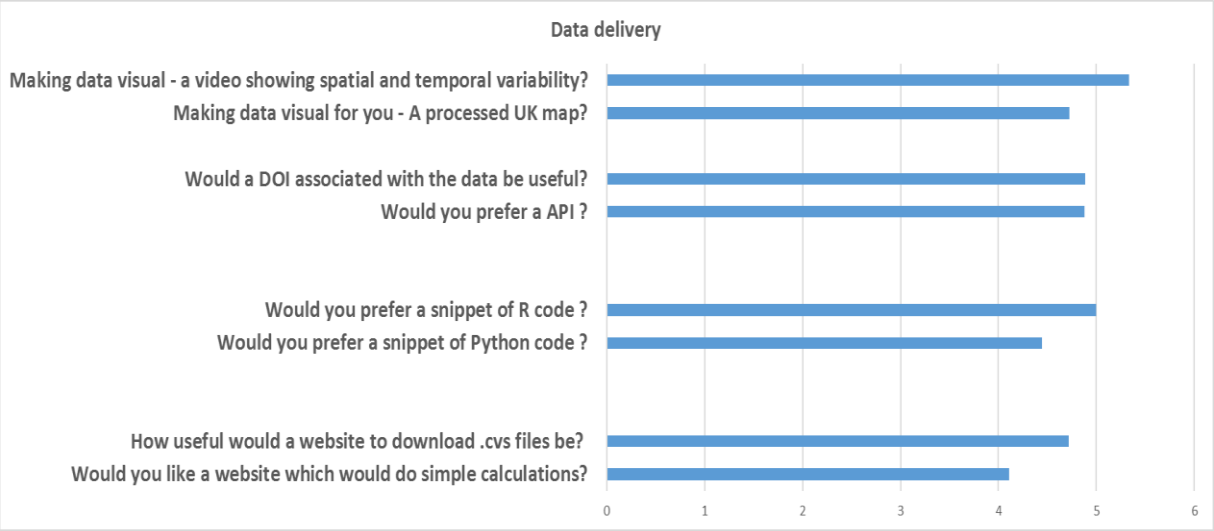

Figure 5 Response of interviewees to options for improved data delivery. Likert Scale 1= yes exactly what I wanted 6=no I had to do a lot of work to make the data useful

Overall respondents considered that "a video showing the spatial and temporal variability of a specific pollutant "would be the most useful addition (mean 5.3 on a likert scale 1=no value 6= brilliant exactly what I would want). The accompanying narrative responses showed that respondents were thinking generally about the community for example *people that don't work with spatial data normally they have some difficulties to understand how the data changes in space and in time.* This respondent considered a base map would be useful commenting:- *if you do a interactive platform where you place online and then you give the chance for people to zoom in and out and then these layers can appear at the scale, where you're dealing with, for example, if you're looking to the UK as a whole, you don't show the maps with streets. But if there's someone zoom in to Edinburgh, then the layer of the streets can appears, so these kind of things would be useful.* Another respondent commented on the multi-functional utility of such a video *i'm also thinking you know that's quite a useful thing to use at lectures and talks that you're giving people, so yeah that would be useful.*

Most of the respondents considered a DOI associated with the data would be useful (mean 4.9 on a 1=no value 6= brilliant exactly what I would want Likert scale). Five of the 9 respondents (55%) scored this question as a 6 on the Likert scale highlighting the longevity of the data link i.e. *it would be good...because sometimes when you direct your reference in a paper to a website, and this is the link where you can download this data, but sometimes the government change the path of this link and when people try it two year's time trying to access again is not working anymore so once you have a DOI it is there forever.* Even those that gave lower scores consider it useful but questioned the level of resources needed; one respondent who scored a 4 commented *I think it's easier to cite it and to refer to it, compared to using website which can go at some point so and I think it's recommended by many funders. Also to*

refer to different versions or database So yes. I would say 4 I mean it's not the fundamental but it can help.

Respondents equally considered that an API (an application programming interface enabling connection between computers or between computer programs) a useful addition (mean 4.9 on a Likert scale 1=no value 6= brilliant exactly what I would want). Fifty percent of the eight respondents scored this question as a six on the Likert scale (Table 7) indicating a strong desire for this suggested addition. Only one respondent scored this question less than 4 on the Likert scale commenting *I have used those in the past and find them temperamental, so I think I would prefer to just manually download data from a website.*

Table 7 Narrative responses and Likert scores to the question "Would you prefer a API ?" (Likert Scale 1=no value 6= brilliant exactly what I would want).

Respondent score	Rationale for score
6	*Yes, so an API allowing me to pull down a subsection of data would be really useful.*
6	*I do know what an API is… I think if it's talking about API for downloading the data then that's very useful that will be six*
6	*sound extremely helpful, then, because I need to speak to Massimo against and put in a separate request, so I guess 5 … they might not deliver exactly what I want, so 5.*
6	*A six*
5	*sound extremely helpful, then, because I need to speak to Massimo against and put in a separate request, so I guess 5 … they might not deliver exactly what I want, so 5.*
4	*I tend to prefer a softer code that is just because i'm a lot more used to that as a statistician lets say compared to a data researcher or probably environmental modeller but I think it would be useful, anyway, mostly depends on how easy is to use API. Sometimes it depends really on the API some of them are very flexible and easy to use some of them are not so I will give it a 4*
4	*it's not my ability to use it and also, I think the R Code and the Python code would probably be more useful.*
2	*have used those in the past and find them temperamental, so I think I would prefer to just manually download data from a website.*

Opinions on the option of providing a snippet of R or Python code to accompany the data varied depending on how familiar the respondent was with these two programing languages (Fig 5.). For example one respondent who scored the desire for a snippet of python code as a 6 (i.e. brilliant exactly what I would want) commented *I would say a 6 because my code is in Python. W*hereas (s)he scored the same question related to R code as a 4 and commented *I don't think I would use it, but I think it would be very useful for people who are just starting out who might want to look at the data.* All respondents indicating that a snippet of R and python code to accompany the data would a useful addition with their preference dependant on their preferred coding language.

4.2.4 Use and acknowledgement of data

Two questions focused on (i) How did you use the data? and (ii) How were the results disseminated? Both were open questions accompanied by a categorical Likert scale. All nine interviewees provided answers to both questions.

Two-thirds of interviewees reported that they consider the data critical for their work. The scoring is interesting as respondents struggled sometimes to decide between the four options. Only two respondents reported the data had been critical for the resultant publication (Fig 6) but one responded that scored the question 3- critical for the work commented *It was critical for the work and yes for the publication - critical but might have found another way.* Another responded was not sure whether to score "2-small contribution to a larger study" or "3- critical to the work". They commented *either a 2 or 3 I will be generous 3 because the thing that happened is even though we ultimately didn't actually use the visuals they enabled us to focus our attention on certain pollutants in know certain regions. So I guess, then we couldn't deliver the project, ultimately, without the data... the data was very helpful to help us focus our reference and save a huge amount of time.*

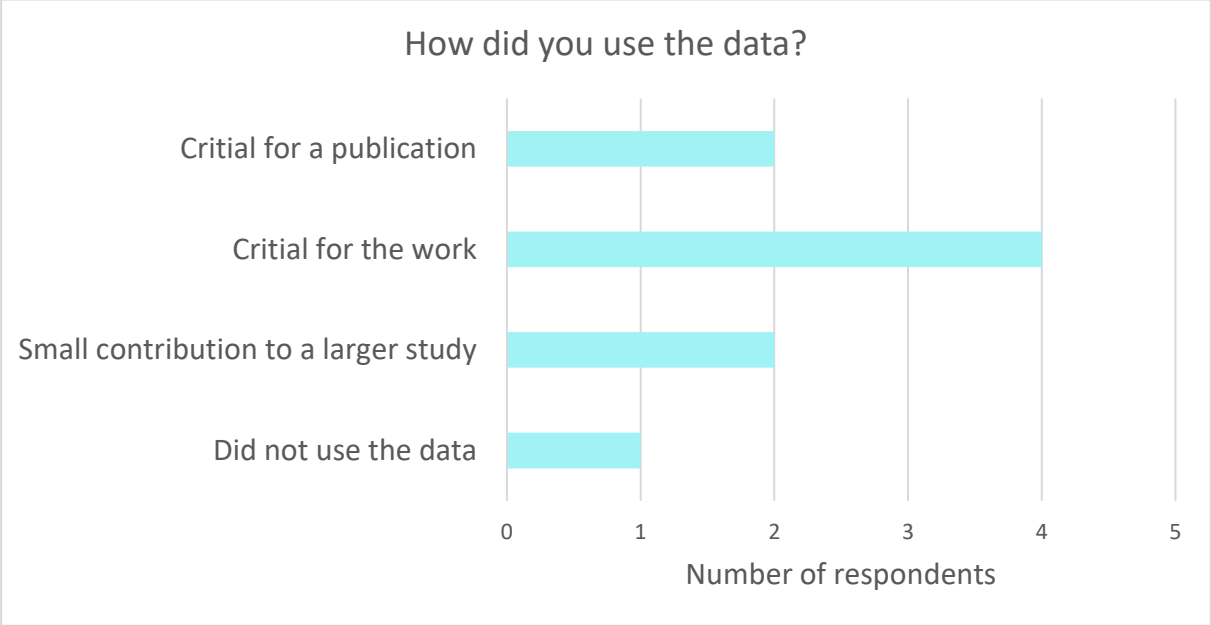

Figure 6. Response of interviewees (n=9) to the question "How did you use the data?" Likert Scale 1 to 4.

Similarly, one respondent who scored this question as a 2- small contribution to a larger study commented *That's hard ...Two or three so it wasn't completely critical, but it was very useful*. Only one respondent scored the question as a "1-did not use the data". They provided a detailed explanation:

I didn't end up using the data or even digging a lot of things out of it. I think one of the biggest reason for me was EMAP is an offline model, and I really needed the feedback between the meteorology and chemistry variables for my studies. So I wanted an online model which takes in meteorology and feedbacks like temperate ... second was resolution I wanted a regional study over India and not over in the UK ... I thought it was more developed for UK use rather than the Asia.

Responds were invited to select from a six point categorical scale when answering the question ‘How where the results disseminated?’ Several respondents selected more than one category (Fig 6).

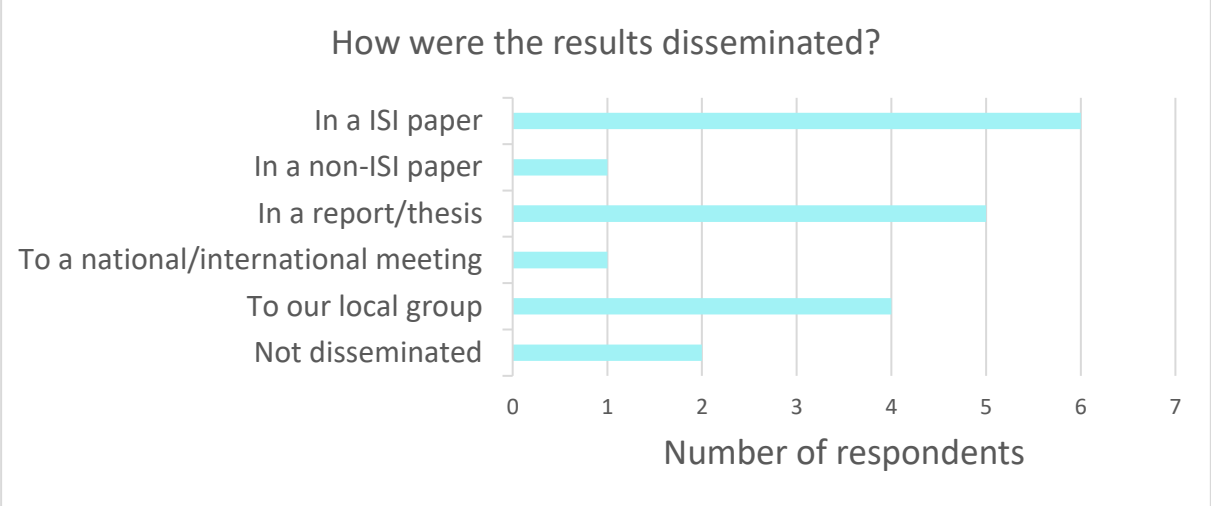

Figure 6. Response of interviewees (n=9) to the question “How were the results disseminated?”

Six respondents reported that the work had been published and when questions all said they had acknowledge the data source. Massimo was co-author with three of the respondents and they reported that he had provided the text that acknowledged the funding. One respondent reported that the work had not yet advanced to a published paper but stated this was their endpoint.

Two respondents reported not disseminating the data because they had not used the data. As described above one student had investigated and was pleased to have received the data but finally selected to focus their work on another model. The other had evaluated the data and passed on to colleagues who were bidding for a Science Technology and Facilities council - UKRI grant aiming to improve food technology.

4.2.5 Further comments

All except one interviewee provided thoughts when asked if there were any other comments (Table 8). These comments ranged from simple praise of Massimo and the data to suggestions of improvements as discussed above to positive feedback on the process which gave the users of the data a voice.

Table 8. Interviewees comments when asked if they had any other comments

EMEP4UK videos were very useful for reconnaissance and targeting of satellite-data for air-quality analysis at regulated industrial sites; and over the next few years this could be an important feature of EMEP4UK outputs for the new generation of geostationary and constellation satellites. appropriate
I think if the data is more easily available, online more people will use it. A lot of people will just look elsewhere if they can't, if they realize they have to contact someone because it's… well, people just want the easiest option don't they...And I think yeah the more like papers, the data is in and the more exposure it gets so the more people will download it

It was a really useful chunk of data and he actually gave me more than I needed , and I was only looking at five waves … of the Millennium cohort study ... I was looking at the first five waves so that's 2001 2004 2006 2008 and 2011. so I only needed those five years of air pollution data, but he even sent me data from 2001 up until 2012 which would have been useful if I wanted to look at maybe lag effect or something...
I think the one thing that would benefit the health applications is going even further back in time so, for example, Massimo provided this data from 2000 on.But if it would be possible to go back because, as I mentioned, there is this cohort.In the health applications where people are being followed in terms of their medical conditions even back in time, like 1940 1960 the UK is very good in geospatial data and also health data so to address the exposure of these people in that specific decade would be good to have the data
might be worth having a more user friendly interface for data if that makes sense.Because I only knew about it, because some one at the EA told me then it was all done by speaking to Masimo and he obviously was busy as it was just when COVID started ..yeah so there's actually two things really - maybe some kind of more formal process, for requesting the data via the website might be useful.... and in terms of a Service level agreement type thing .. [JD like a licennce] No ...see if there was a sort of guarantee of receiving the data within certain amount of time becasue it was uncertainty when we would get the data ... I am definitely definitely definitely definitely not saying anything bad about the kind of timescales that the data was actually send in.
I didn't find many people, except the people from ceh were using EMAP I didn't find anyone in my colleagues from my school to be knowing about EMAP at all,.so I think that also made me feel like I might not get as much help ...but that's not true, I think Massimo would have helped anyway, ...I didn't find many resources on the Internet, about EMAP, they were papers sure, but... less information than other models that I was looking at.
And the sort of processing work that had to do with the data that Massimo provided was to convert the data into the the right sort of files structure that we needed to input to the model. We have to do a little bit of regridding [JD should Massimo offer more] I think he's best delivering it in his way, but the fact that he uses universal format NetCDF makes our lives very easy and something like an API allowing us to pull data in down in chunks... the chunks of data that we need is also would be really good as well and that might have meant ...might have meant I could miss out a step, for example..
i'm sort of intrigued by the whole kind of user you know I think it's a really good idea to try to reach out to the users to understand what they actually need and we develop thing you know tools that are actually useful for People ... but it's quite a difficult thing to is quite time consuming isn't it and there's so many potential users out there ... there's all sort of an issue of sort of advertising, so that people know about it..

5 Conclusions

Overall the interviewees where positive concerning the EMEP4UK gridded UK atmospheric composition data provided via the UK-SCAPE funding. The respondents considered the data could be used more by the UK research community if it was more available e.g. accessible remotely. They welcomed the suggestion to visualise the data via a video and provide a persistent identifier for the data such as a Digital Object Identifier (DOI). In addition, a small snippet of R and Python code to accompany the data was considered desirable.

Although not specifically asked all the interviewees highly praised Massimo Vieno's work and personal commitment to servicing the research community commenting for example *great procedure with Massimo. He was always very helpful* and *What Massimo did .. in terms of the what we asked, for that was exactly what we wanted.*

Several interviewees appreciated that UK-SCAPE had funded the work to reach out and give the user community a voice and the opportunity to co-design improvements. They commented for example *so you know what you're doing here is fantastic really important work, I would say.* While another commented *it has been good to know that all these feedbacks are taken and my feedback has been useful in some way and I think it's a really good idea to try to reach out to the users to understand what they actually need …develop thing you know tools that are actually useful for people.*

6 Next Steps

One of the aim of this study was to consider future work and following this analysis, all future requests will be accompanied by a snippet of R or python code depending on the wishes of the recipients. The other suggested improvements are all in consideration but are dependent on funding.

The interviewees in this study recognised that setting up, running and creating large gridded dataset from atmospheric chemistry transport models (ACTMs) is a complex task but very valuable to the scientific community. In addition, each model, model input, included in the chain needs to be updated (usually annually) to keep the data product up to date with the ever-changing scientific knowledge.

It was clear from the responses included in this research that an up to date gridded UK dataset needs to be created, maintained and made available to the wider UK and global scientific community. Complex UK gridded atmospheric composition data is used in a wide range of scientific fields from meteorology e.g. impact of weather types on UK ambient particulate matter concentrations to human health impact of air pollution assessments.

7 Acknowledgements

The authors are grateful to all the interviewees for giving so freely of their time and engaging so fully in this endeavour.

Appendix 1 – Example email invitation and reminder

From: Dick, Jan
Sent: 08 February 2021 14:14
To:
Subject: EMEP4UK gridded UK atmospheric composition data

Dear

I understand that Massimo Vieno send you modelled EMEP4UK gridded UK atmospheric composition data last year (i.e. [insert 'Type of data' from Table 1]) which was part funded by the UK-SCAPE project. As part of the community evaluation of this project, I would really appreciate the opportunity to discover how you found the experience and how we may improve delivery of public funded research. I write you a joint email but separate interviews would be ideal if you can both spare the time.

Given the situation with COVID-19, I would very much like to discuss your views remotely by telephone or preferably via Zoom. I can guarantee that the information supplied will remain confidential and you opinions will remain anonymous (unless you desire otherwise).

Would you be willing to have a chat (10-20 min) about the UK-SCAPE data access and utility? If so can we arrange an appointment preferably this week or next?

jan

I work a flexible working pattern so may send emails out of 'typical' working hours.

Please be assured that I do not expect a response outside of your own working hours.

Dr Jan Dick

Senior Socio-ecologist,

Ecology, Evolution and Environmental Change Group,

UK Centre for Ecology & Hydrology

Bush Estate, Penicuik, Midlothian, EH26 0QB, UK

Direct Telephone ++ 44 (0)131 445 8578

Office Telephone ++ 44 (0)131 445 4343

skype ceh_jand

Orcid.org/0000-0002-4180-9338

ResearcherID I-6839-2012

Scopus Author ID:7201749564

Environmental Change Network www.ecn.ac.uk

EU FP7 INTERACT www.eu-interact.org

EU H2020 eLTER http://www.lter-europe.net/lter-europe/projects/eLTER

www.ceh.ac.uk | @UK_CEH

_____

From: Dick, Jan
Sent: 15 February 2021 10:38
To:
Subject: RESENT: Evaluation of UK-SCAPE data you received

Dear

As I wrote below I am conducting a survey of people who have received data from Massimo Vieno, part funded by the UK-SCAPE project.

I have to report to NERC the response of all those I have contacted so I would really appreciate it if you can let me know if you would be willing to be interviewed.

I fully appreciate that your time is limited but hope you both will take a moment to reply to this message.

Thanks in anticipation

Jan

<Included fully message above>

Appendix 2 - Participant Information and Consent Sheet

Value of the UKSCAPE modelled EMEP4UK gridded UK atmospheric composition data

Contents

Rationale for interview

You are invited to an interview because you requested data from Massimo Vieno, UKCEH, which was part funded by the UK-SCAPE project. As part of the community evaluation of UK-SCAPE Jan Dick has been invited to discover how you found the experience and how UKCEH may improve delivery of public funded research.

This *Participant Information and Consent Sheet* explains the procedure. Before you decide whether you wish to participate in the interviews, it is important that you read the information provided below. This will help you to understand why and how the research is being carried out and what participation will involve. Please contact Dr Jan Dick (jand@ceh.ac.uk), who will conduct the interview, if anything is unclear or you have any questions.

You can refuse and withdraw at any stage during the interviews. Please note that information from the interviews will be anonymised and incorporated into a single report of all participating interviewees. Consequently, your views cannot be withdrawn after the interview is complete and the data analysed.

Who is conducting this research?

The key contacts from the project team are Massimo Vieno (mvi@ceh.ac.uk), Jan Dick (jand@ceh.ac.uk) and the program lead is Eleanor Blyth, (emb@ceh.ac.uk) all are staff members of UK Centre for Ecology and Hydrology.

Who is funding the research?

The interviews and delivery of the air quality data are funded by UKRI/Natural Environment Research Council award number NE/R016429/1 as part of the UK-SCAPE programme delivering National Capability.

What is the purpose of the research?

Aims of interviews is to understand from a user's perspective the value of the UKSCAPE modelled EMEP4UK gridded UK atmospheric composition data and co-design improvements.

Do I have to take part?

No. Taking part in this knowledge sharing activity is completely voluntary and deciding not take part will not disadvantage you in anyway. You are free to withdraw from the interview at any time without explanation or penalty. The best way to withdraw from the interview is to alert Dr Jan Dick conducting the interview that you wish your contributions removed. Withdrawing your contributions once they have been anonymised and summarised will not be possible.

What will happen if I take part?

Participating will entail one interview of approximately 30 min to one hour (Feb-march 2021) scheduled at a time mutually agreed with Dr Jan Dick (weekend and evening interviews are possible if you desire as we recognise that many scientists are working non-standard hours during this period of lock-down).

The purpose of the interview is to understand your motivation for requesting data from Massimo Vieno, UKCEH, how easy and useful the requesting process was and your feedback will also be gathered on the utility or any constraints that may affect you requesting further data and your views to co-design better access.

The interview is designed as a semi-structured interview i.e. open-ended questions, allowing for a discussion with the interviewee rather than a straightforward question and answer format. Five themes will be explored (i) knowledge of the data prior to your request, (ii) accessibility of the data from your perspective (iii) how you used the data (iv) how can access be improved from your perspective and (v) any other relevant issue raised by the interviewee.

The interview will be recorded, with your permission, to enable Dr Jan Dick to check when she writes the report that she is correctly representing the views you express. If individual quotes are used, they will not be attributable to an individual researcher.

Are there any risks in taking part?

There are no risks to taking part in the interview, which the research team can foresee. The research team are not part of any UK regulatory agencies and access to the data in future will not be restricted in any way, it is in fact hoped to improve access to the data through analysis of these interviews.

What are the possible benefits of taking part?

There are no immediate direct benefits to taking part in this project; however, we hope that following the results of the interviews we will have a better understanding of data access and use from the researchers perspective and can co-design better access.

Will my taking part in this project be kept confidential?

Yes - All output will be anonymised to ensure no identifiable data is made public. Dr Jan Dick will be managing your contact details to ensure you are invited and will keep those contact lists secure. There will not be any records linking your contributions back to your name or contact details, unless you state you which to be identified for some reason.

What will happen to the information I provide?

The information you provide will be captured electronically and via a recording of the interview if you agree. The data will be stored to support analysis and a future report documenting this co-production process. We intend to archive the anonymised data for future research use; however, there will be no way for these data to be linked to project participants. If you wish to withdraw your contribution, this must occur during the interview, when UKCEH will still be able to identify your statements and remove them from the analysis. Once the report has been anonymised, it will not be possible to withdraw your contribution. If you are interested to access any of the results of the project, you can contact Dr Jan Dick (jand@ceh.ac.uk).

Data Protection

The personal data that will be collected and processed in this study are your name and contact details, solely for facilitating the arrangements for the interviews and will not be used for any analysis or reporting.

The UKCEH asserts that it is lawful for it to process your personal data in this project, as the processing is necessary for the performance of a task carried out in the public interest (contacting you to arrange interviews). Following the completion of the UK-SCAPE project (Dec 2024) Jan Dick will delete the file with your contact details.

The UKCEH respects your rights and preferences in relation to your data and if you wish to update, access, erase, or limit the use of your information, please let us know by emailing Dr Jan Dick (jand@ceh.ac.uk). Please note that some of your rights may be limited where personal data is processed for research, but these occasions do not relate to this project. If you wish to complain about the use of your information please contact the UKCEH's Data Protection Officer in the first instance (email: Quentin Tucker, Data Protection Officer quetuc@ceh.ac.uk). You may also wish to contact the Information Commissioner's Office (https://ico.org.uk/).

BANGOR
UK Centre for Ecology & Hydrology
Environment Centre Wales
Deiniol Road
Bangor
Gwynedd
LL57 2UW
United Kingdom
T: +44 (0)1248 374500
F: +44 (0)1248 362133

EDINBURGH
UK Centre for Ecology & Hydrology
Bush Estate
Penicuik
Midlothian
EH26 0QB
United Kingdom
T: +44 (0)131 4454343
F: +44 (0)131 4453943

LANCASTER
UK Centre for Ecology & Hydrology
Lancaster Environment Centre
Library Avenue
Bailrigg
Lancaster
LA1 4AP
United Kingdom
T: +44 (0)1524 595800
F: +44 (0)1524 61536

WALLINGFORD (Headquarters)
UK Centre for Ecology & Hydrology
Maclean Building
Benson Lane
Crowmarsh Gifford
Wallingford
Oxfordshire
OX10 8BB
United Kingdom
T: +44 (0)1491 838800
F: +44 (0)1491 692424

enquiries@ceh.ac.uk

www.ceh.ac.uk