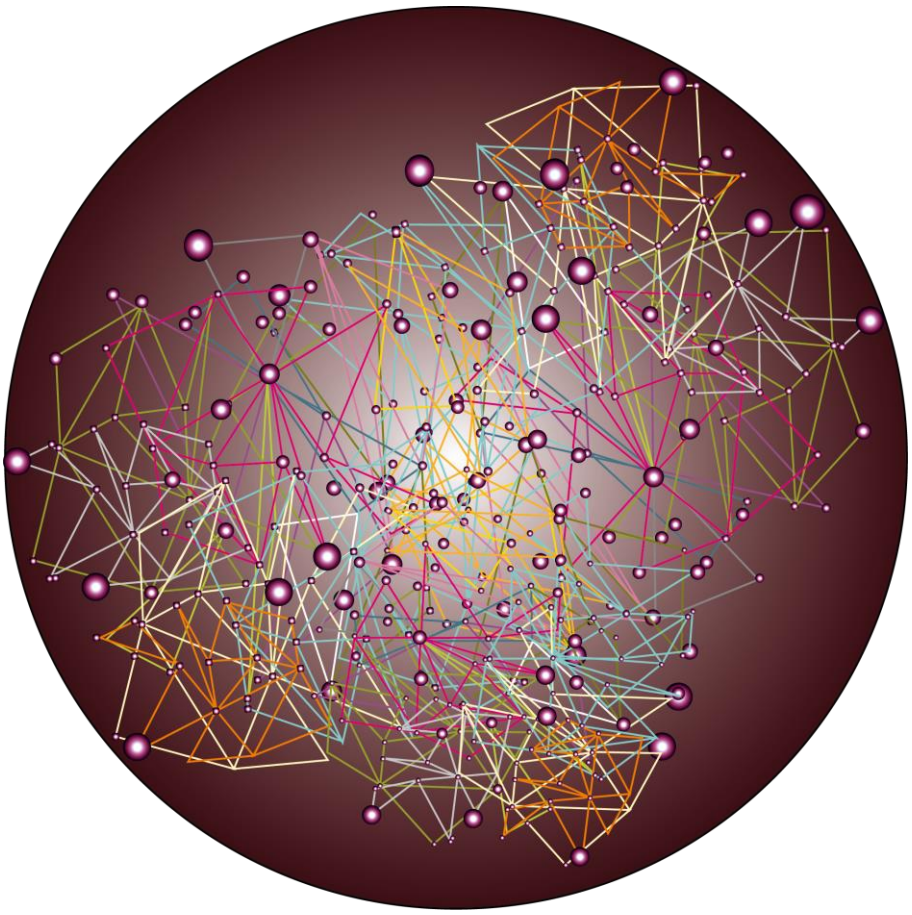
European Environment Agency

Big data and environmental knowledge

Summary report of the EEA Scientific Committee Seminar

Copenhagen, 10 October 2017



Acknowledgements

This report is based upon presentations and discussions that took place at a seminar of the European Environment Agency's (EEA) Scientific Committee, held on 10 October 2017 in Copenhagen. Seminar participants included:

Members of the EEA Scientific Committee

Per MICKWITZ, Finnish Environmental Institute, Finland, EEA Scientific Committee Chair
Mikael Skou ANDERSEN, Århus University, Denmark, Scientific Commitee Vice-Chair
Ana IGLESIAS, Technical University Madrid (UPM), Spain
Andrew James JORDAN, University of East Anglia, UK
Angel BORJA, Marine Research Division, AZTI-Technalia, Spain
Claire DUPONT, Vrije University, Belgium
Frank GEELS, University of Manchester, UK
Fred STEWARD, University of Westminister, UK
Greet SCHOETERS, VITO, Belgium
Jean-Louis WEBER, University of Nottingham (Hon. Prof.), UK
Joachim H SPANGENBERG, Helmholtz Center for Environmental Research (UFZ), Germany
Leonor TERRASON, Norwegian Institute for Air Research, Norway
Małgorzata GRODZIŃSKA-JURCZAK, Jagiellonian University, Poland
Philippe GRANDJEAN, University of Southern Denmark, Denmark
Tom H OLIVER, University of Reading, UK
Theodoros ZACHARIADIS, Cyprus University of Technology, Cyprus

Guests

Ahmet Mete SAATCI, Turkish Water Institute, Turkey
Caroline POULSEN, DEFRA
Evelin DILS, ETC WMGE/VITO
Kelly EDWARD, Amazon Web Services
Kitty van der HEIJDEN, NL
Maarten HAJER, Urban Futures Studio
Malgorzata BEDNAREK, NFP Poland
Mariana APOSTOL, NEPA Romania
Markus MEINERT, Federal Ministry for the Environment, Nature Conservation and Nuclear Safety
Nico PIETERSE, PBL
Peeter PÄRT, DG JRC, EC
Sander MUCHER, ETC BD/WEnR
Thomas BRUNHES, EC
Ulrika DOMELLÖF MATTSSON, EPA Sweden

Speakers in order of appearance

Chris STEENMANS, EEA
Hannes REUTER, DG Eurostat
Simon BINGHAM, Scottish EPA, IMPEL
Sven SCHADE, DG JRC
Vincent-Henri PEUCH, ECWMF, FR
Johan BROMAN, Amazon Web Services, SE
Janet RANGANATHAN, World Resource Institute, US

EEA Staff

Hans BRYUNINCKX, EEA Executive Director
Chris STEENMANS, Head of Programme, ICT and data management
David STANNERS, Head of Programme, Partnerships and networks
Jock MARTIN, Head of Programme, Integrated Environmental Assessments
Paul MCALEAVEY, Head of Programme, Air and climate change
Ronan UHEL, Head of Programme, Natural systems and sustainability
Andrus MEINER, Natural systems and sustainability
Caroline WHALLEY, Natural Systems and Sustainability
Elena OSTÁRIZ, Integrated Environmental Assessments
Lars MORTENSEN, Partnerships and networks
Lorenzo BENINI, Integrated Environmental Assessments
Malene BRUUN, Partnerships and networks
Pernille FOLKMANN, Integrated Environmental Assessments Assistant
Stefan JENSEN, ICT and data management
Teresa RIBEIRO, Strategic coordination
Xenia TRIER, Integrated Environmental Assessments

This report was drafted by Stefan Jensen and Chris Steenmans, EEA.

Contents

Executive Summary

From an institutional perspective, big data has already started influencing some of the data and information processes used by the EEA, Eionet and partners in the European Commission. Classical data sources, such as data from reporting, are increasingly complemented by machine-generated data, often from *in situ* sensors or satellite monitoring (e.g. Copernicus), and by human-sourced data from social networks, internet searches and mobile content. Data at the EEA are growing and becoming more diverse, raising new issues and challenges related to processing, quality assurance and methodology.

The use of big data by the EEA could potentially lead to new ways of working through different partnerships with citizens, business and public authorities. This leads to new challenges related to the environment, climate change and sustainability. One of the areas revolves around new markets for data analytics.

Another growing discussion is that big data is triggering a debate around empowering citizens. Commonly agreed benefits from big data potentially lie in taking a more integrated approach leading to a better understanding of systemic issues.

Issues are emerging around relevance of data, securing privacy and confidentiality as well as data access and licencing. Views are diverging on the how relevant big data is to policy and the question if more or better data results in improved decision-making. Understanding and implementing big data related processes also require new skill sets.

The business and science related interventions and presentations showed the many opportunities that big data could provide to the EEA for creating new products with improved quality. New data processing and information platforms were introduced during these sessions, such as those currently being developed for Copernicus services, often based on cloud computing and machine learning. From a science and communication perspective, the revolution around big data relies on open data and open governance. When building applications based on big data, we need to look carefully at current and future user requirements and develop practical examples related to monitoring and reporting.

In the round-table discussion, several participants called for a focus on potential solutions rather than on the problems themselves. The EEA should be open for innovative ideas, participants felt. Nevertheless, a methodology shift from inductive insights, as applied in our current data analysis processes related to reporting data, to deductive approaches, needs to be evaluated. Therefore, pilot projects and collaboration with partners involved in big data analytics can be a way for the EEA to incorporate big data into future work programmes.

The seminar concluded by noting that the use of big data needs to be applied with caution, since the EEA has established a reputation over decades as a trustworthy information provider based on the Driving forces, Pressures, State, Impact and Responses (DPSIR) model. This is underpinned by robust and reliable data sources, which need to be sustained and shall only, after careful evaluation, be complemented by the new opportunities offered by big data.

Introduction

This report provides a summary of a European Environment Agency (EEA) Scientific Committee Seminar on big data[1] held at the EEA in Copenhagen, Denmark on 10 October 2017. Participants at the seminar considered the concept of big data, its current influence on the global data landscape and the impact it may have for future EEA work programmes.

Structure of the seminar

The seminar was organised around two sessions, entitled:

- Morning session: Big data – Institutional perspective
- Afternoon session: Big data – Science, business and policy perspective

Under each session, a number of experts provided presentations. A round table followed the afternoon session.

This meeting report provides a short summary of the presentations and discussions structured according to the programme of the workshop (see Annex 1), including opening remarks, the two sessions, the round table, summary and concluding remarks.

Opening remarks

Per Mickwitz, Chair of the EEA Scientific Committee, opened the seminar, welcomed participants, and introduced the meeting objectives.

The scientific committee seminar series has one principal focus: future knowledge. Big data is an important issue yet there is much to be defined as regards its role, identity and future planning. This seminar aims to form a view on what we should do in this area. There are a number of opportunities as technologies seem to drive the production, collection and storage of this data. Many new opportunities to search, integrate and analyse data are coming up. In the environmental area, interesting studies exist related to big data and social network data; innovation in renewable energies; or based on sensor data and text mining of administrative data. On the other hand, there are many ethical issues around privacy and security.

Hans Bruyninckx, EEA Executive Director, mentioned the frequent presence of big data in strategic discussions. It promises new products and services and faster and better decision-making – all aims of the EEA – these being reasons why it should be examined. Big data is a hot topic in several environmental domain discussions and is also viewed as a driver of Citizen Science. With all this in mind, a number of questions should be addressed: Will we be swamped by it? Do we have the right means, people and technology to correctly deal with big data? Will data in the future be treated differently in the way they relate to assessments and knowledge?

[1] The definition in the IMF reference document (see references section) provides a chapter entitled "What does big data mean?"

Morning session: Big data – Institutional perspectives

Chair: David Stanners (EEA)

Setting the scene

Presentation by Chris Steenmans (EEA)

Chris set the scene by reminding the audience about the various elements (the many "Vs") of the big data discussion.

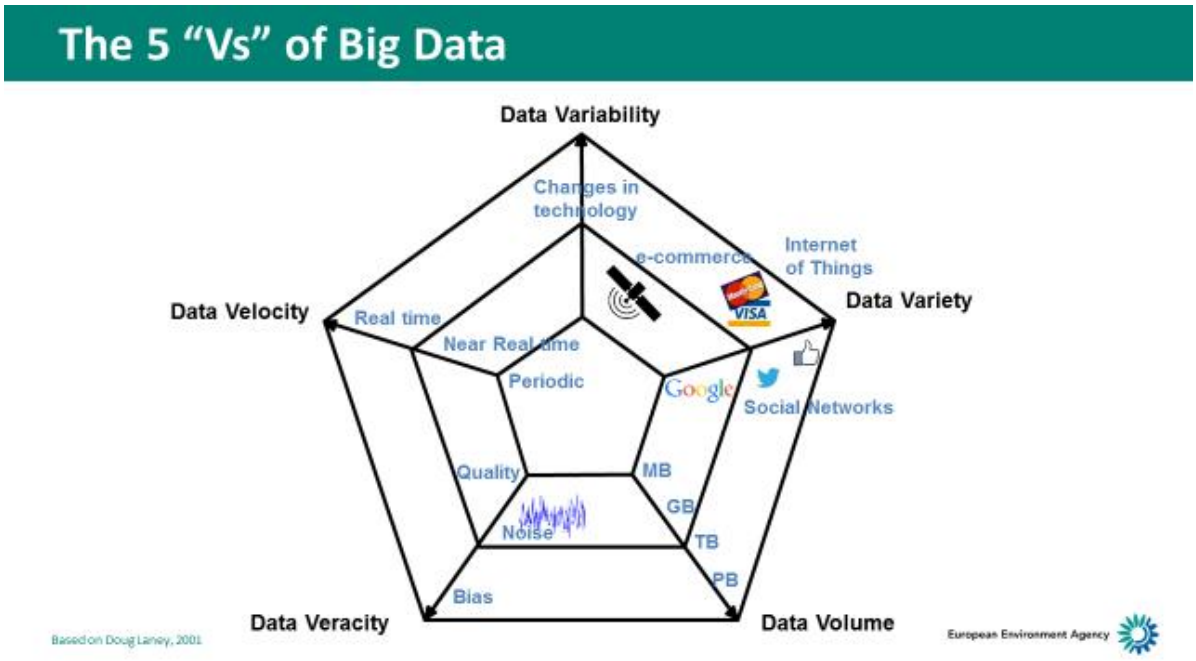


Figure 1: Key characteristics of big data

The EEA's newly launched air quality index can be seen as an example illustrating how big data can bring a diverse and dynamic range of source material. Over the years, the EEA has handled an ever-increasing amount of data from more and more sources. Our knowledge base is growing continuously. Copernicus is a key example of how big data works for the EEA. Co-creation of knowledge necessitates a long-term vision. New areas are emerging especially drivers from important international policies.

We have to reflect carefully on the role of the EEA and Eionet and how potential impacts, such as the need for staff with new skills using new technologies, will determine the ability of the organisation to deliver measurable and high quality results.

Discussion
There were questions on the use and relevance of big data for decision-making. Although not commonplace examples do exist, in particular from the climate change field.

Big data for European policy support

Presentation by Hannes Reuter (Eurostat)

Hannes explained Eurostat's approach to big data and how this will affect their production processes. There are increasing possibilities to integrate geospatial and statistical data within the relevant policy areas e.g. to determine the proper use of regional funds. He presented the topics behind the EU's big

data roadmap including involving the national statistical offices through a set of pilot projects under a specific action plan.

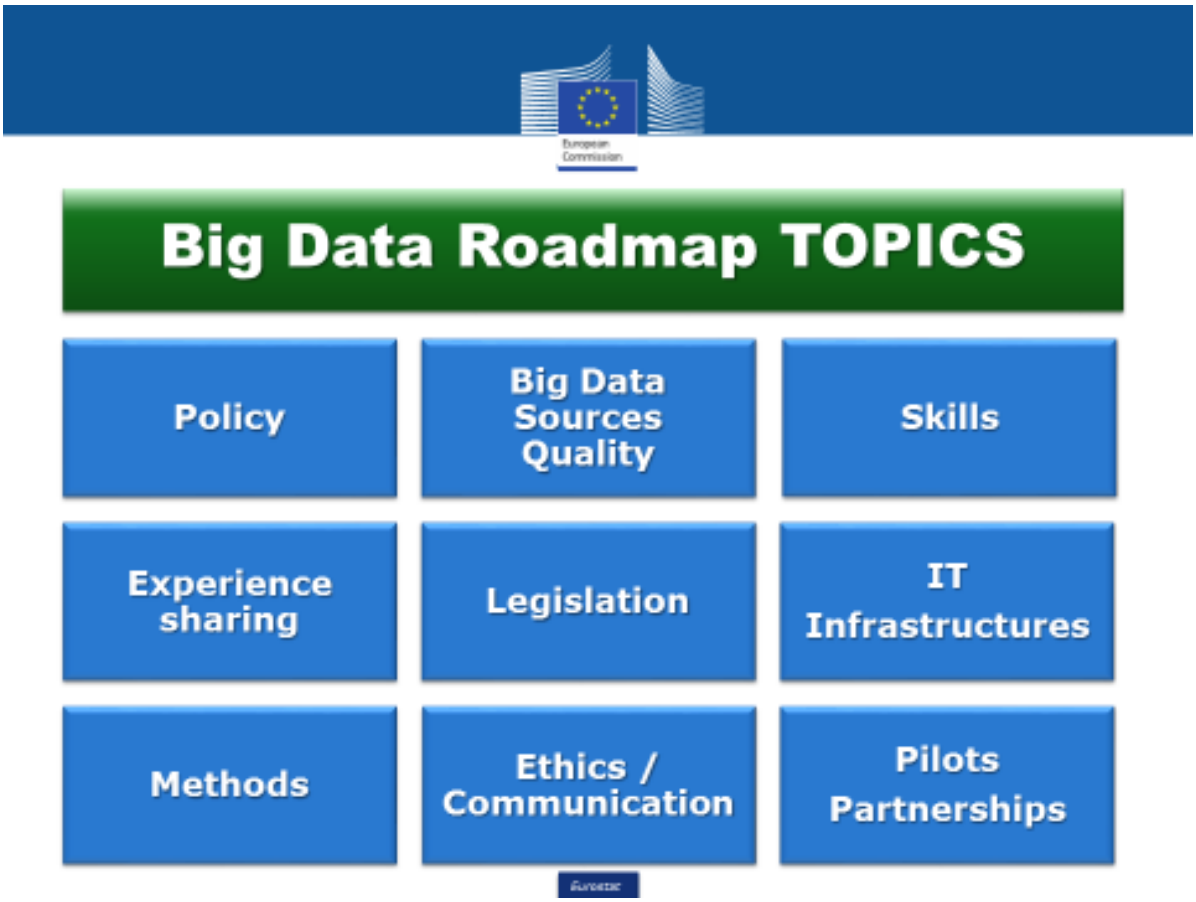


Figure 2: Illustrating the many areas of big data still to be addressed

Examples from these pilot projects were provided, which include smart electric meters and energy grids, shipping tracking, mobile phone data for population distribution. Knowledge gaps do exist which are addressed by a new "data for policy" initiative. Further activities include awareness raising on new analytical possibilities; the promotion of teamwork to avoid silo building; the harnessing of synergies between portfolios; and the re-use of data.

Discussion

Comments acknowledged the new technology used but criticised that it did not address sufficiently new ideas. One point raised took the form of a plea to use big data to modernise and move ahead of traditional statistics. A stronger role of social science data was also advocated. This type of data appears to be largely under-represented in the big data discussion.

Sweating the data (making our investments work)

Presentation by Simon Bingham (Scottish EPA)

Simon presented another institutional perspective from the point of view of environmental regulators. A complex, site-specific risk assessment system has been designed making use of own institutional data and bringing in data provided by partners.

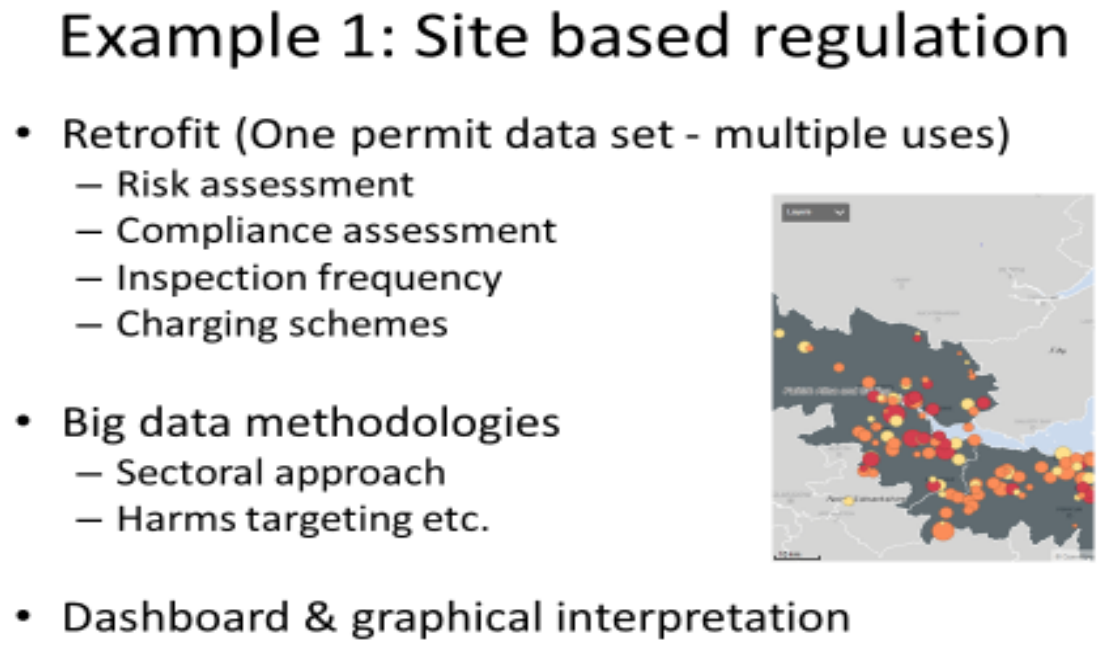


Figure 3: Risk assessment in a sectoral approach using big data technologies

Under the European Union Network for the Implementation and Enforcement of Environmental Law (IMPEL), many different best practise examples are being shared. The big data user community within this network is growing.

Discussion
Question and answers highlighted that big data cannot necessarily be considered as open data, which raised data privacy issues.

Big data after the hype

Presentation by Sven Schade (JRC)

Sven Schade from the Joint Research Centre felt that the discussion on big data had evolved "beyond the hype". He reminded us that a key challenge is to filter the right data out of the growing amounts of potentially useful data. He presented the many challenges related to big data that the JRC has learned over time. Cost is becoming less of an issue. Mobile phone data can now be provided in a way that assures data privacy. Data re-use has been hampered by restrictive licences and non-transparent data handling practises. While data processing generally is becoming less of an issue it is still often done in a way that does not assure reproduction.

Figure 4: Experiences around data handling

He concluded by stating that finding the right data and creating the right evidence is the main challenge. Consciously using data from different sources should be valued against overly discussing the various "Vs" and their potential developments. Developments should also assure a cognitive balance between humans and machines.

Discussion
The discussion referred to a perceived lack of trust regarding the quality of citizen science data. However, in reality the quality of data from volunteers is not necessarily worse than official data.

One intervention suggested finding out how the method known as "mining the internet" is contributing to better data analytics.

Summarising the discussion on the institutions perspective

Stefan Jensen from the EEA provided a summary of the morning's presentations and discussions identifying several clusters.

- The presentations underlined that data is growing in volumes and variety; new ways to improve our knowledge base on the environment and sustainability are emerging.
- We can identify new ways to analyse and acknowledge the important role that new partnerships will play.
- Data quality needs to be better addressed. Finding and using the right data will be key when using big data for providing trustworthy information.
- Main benefits to date seem to lie in innovation, increasing re-use and cost reduction.

- Key issues focus on securing privacy and the lack of obvious solutions.
- It is unclear if better decision-making and higher policy relevance are sufficiently supported.
- There are many technologies involved in big data handling, most likely beyond the skills and competencies of the EEA and Eionet. Engagement with a wider network is needed.

Evelien Dils from VITO (BE) represents the EEA topic centre on waste and material flows and summarised her observations around the impact on research. She mentioned impacts on work related to human biomonitoring and to the circular economy. A new value chain is emerging which needs to link across existing value chains. A more systemic perspective needs to be taken; social data has to be better addressed.

Scientific Committee members questioned if we really know what we want to use big data for. In looking at the interplay between environmental, economic and social phenomena, data is currently available albeit with huge time differences. There is an expectation that big data may improve and ways will be found to overcome reporting time lags.

Afternoon Session: Big data – Science, business and policy perspectives

Chairs: Hans Bruyninckx (EEA) and Jean-Louis Weber (EEA Scientific Committee)

Using and sharing big data for weather, atmospheric composition and climate

Presentation by Vincent-Henri Peuch (ECMWF)

Vincent-Henri Peuch, from the European Centre for Medium-range Weather Forecasts (ECMWF), spoke from the perspective of the weather and atmospheric modelling community. The community is faced with a very significant growth in data in a domain which is very much aware of weather related disasters, and the resulting deaths and economic losses. Vincent referred to developments in the areas related to the aforementioned five "Vs". He illustrated a best practice example of handling complex data through a case study focusing on Hurricane Irma.

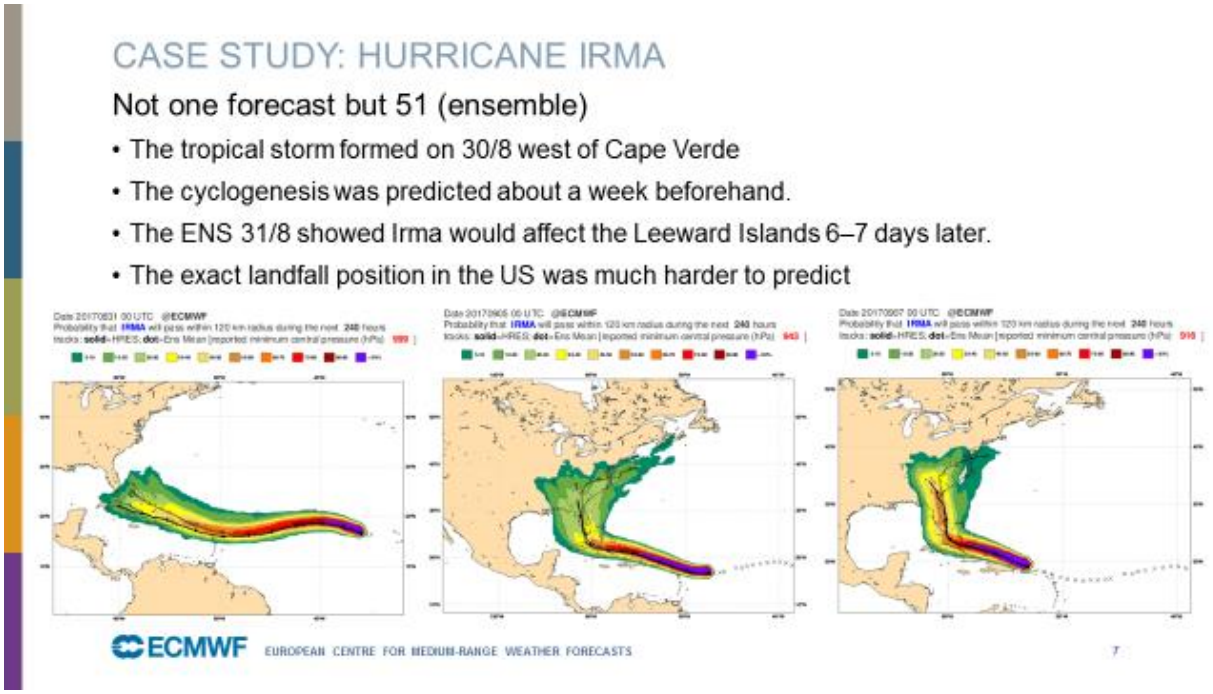


Figure 5: Example of a big data case study

ECMWF identified that for their purposes a powerful data store is needed to provide data to the users. However, high-performance computing and archiving technologies are no longer able to provide the efficiency gains of the past. A paradigm change has taken place: a climate data store is being developed where users can carry out the processing and then retrieve the results rather than only downloading the data. A "Data and Information and Access Service" (DIAS) is being developed, together with Copernicus partners, exploring the potential of cloud computing-based services.

Discussion
The discussion identified that big data has a relative definition – it is changing over time as data volume grows and technological solutions change continuously. Citizen science activities in the domain were mentioned and how they could be integrated. Examples of pilot projects were provided, which link to approaches and tools to deal with data of very different quality and which follow agreed criteria to document data quality accordingly. The current way of working cannot scale up, hence the need for cloud computing.

Amazon and big data

Presentation by Johan Broman (Amazon Web Services)

Johan Broman, the head of the solution architecture team in the Nordic countries at Amazon Web Services (AWS), provided an industry perspective. He underlined the role of cloud computing and stated that a defining element of big data is that current tools and processes are no longer suited to handle the data at stake. Amazon focusses on solutions addressing data volume, variety and velocity in its products offered.

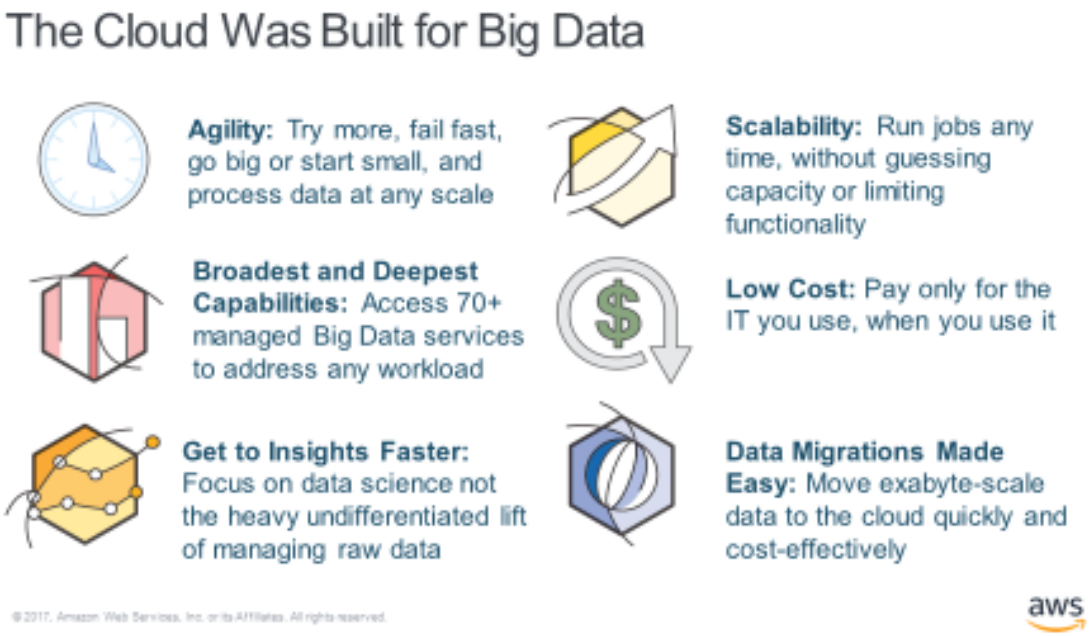


Figure 6: Cloud – big data interactions

The Amazon business model is based on the cloud being the sole platform. The solutions presented relate to the issues raised in the ECMWF presentation on how to address data handling and computing limitations. Up to 90 services and tools are offered by Amazon to handle data. Users can configure this flexibly and build incrementally different usages. They put data into a "data lake" and use the computational frameworks needed. Increasingly customers are using these solutions. AWS offers a set of public environmental and spatial datasets, which can be used on the platform.

Discussion
Questions raised focused on the standards used and on future technologies. Responses indicated that there are no specific requirements and an innovative process is ongoing focusing on updating the technologies.

Monitoring to Mobilize – Big data analytics in practise

Presentation by Janet Ranganathan (World Resource Institute)

Janet is the science director of the World Resource Institute (WRI) – a global research organisation operating as an NGO from Washington DC. In her presentation, she spoke about data management at a global level. For her, big data has to be seen in the context of the data revolution. Open data and open governance are key WRI principles. A big gap exists between data supply and its use to solve real world problems. The usage of big data has to be problem focussed. Janet provided examples from global projects on water risks and forests where users' needs and decision-making are in particular focus. For the success of these projects, other technologies apart from big data, such as machine learning, also play a vital role.

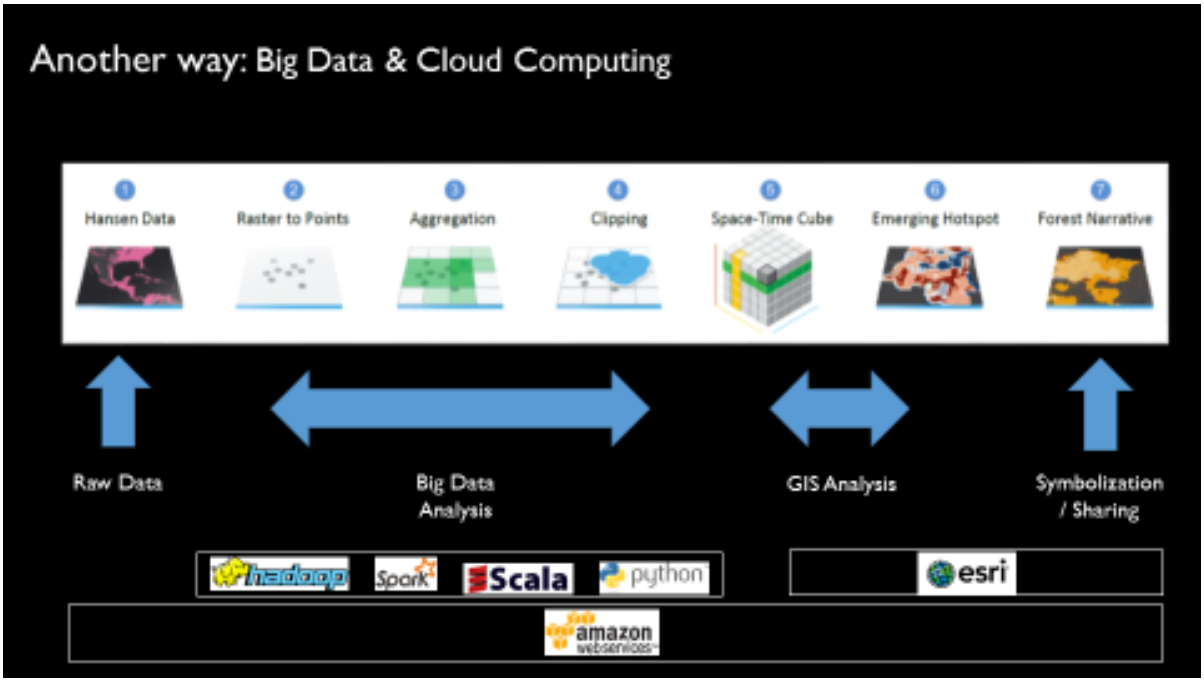


Figure 7: An infrastructure example to power WRI tools

It is very clear that for WRI, big data is only one of many tools to power platforms, apps and other user-oriented tools.

Discussion
Questions focused on the content of WRI's work highlighting the new opportunities for ecosystem analysis. WRI uses a workflow model to bring data, citizen engagement and other elements into successful citizen monitoring tools (thematic "watches").

Round table

Chaired by Leonor Tarrason (EEA Scientific Committee)

Scientific Committee members discussed a wider set of topics, which were geared to advise the EEA's work programme. A majority of Scientific Committee members advocated for the EEA to take a middle ground approach in the uptake of big data and related analytics. Partnerships are seen as key and the Agency should already be exploring with its various partners and networks what a middle ground approach actually means in practise. The need to exercise caution in terms of working in a transparent manner was raised. To identify the sources of these forms of new data; how they arrived and what assumptions are imbedded in them; were all viewed as important. The EEA needs to avoid getting this wrong because there is a risk in the organisation no longer being viewed as a reliable knowledge, information and data provider.

Summary and Outcomes

EEA director Hans Bruyninckx *summarised the outcomes of the seminar*

The seminar discussed the potential use of big data for improving our environmental knowledge in the context of the EEA's core task to provide relevant, reliable, timely and targeted information and knowledge to policy makers. A major challenge for the EEA is to integrate the use of big data in its key

business processes as an authoritative information provider. Trust has to be maintained irrespective of how the EEA makes use of innovative data flows. Satellite based monitoring is an example of a new way of producing reliable data based on big data.

Generally, it is difficult to define precisely what big data is. While the concept is still contested by many, the use of the term big data is gaining momentum based on the practices that have emerged around the general idea. An institutionalisation of the concept is taking place as nowadays conferences, journals, the roles of senior members of staff in large companies, and business models are all arranged around big data. A set of practices are emerging on which we need to have a discussion. A key assumption for the EEA's operational model is that we assume that better knowledge leads to better policies.

An important item to consider is exactly how big data relates to the EEA's analytical model - DPSIR. We need to identify how this new paradigm of input data can create new knowledge and contribute to strengthening parts of the chain that links integrated assessments.

The discussion identified six elements where big data would support better knowledge creation:

1) Faster access to data and knowledge – a key example is Copernicus where data availability and processing time has become much faster. The first cycle of Corine Land Cover took 10 years, now it takes 2 or 3 years. Annual cycles could soon become a reality.
2) Big data can allow organisations to be more precise such as through better emission projections contained in some of our regular reports. This is clearly contributing to improved knowledge creation, and also has the potential to be more precise. Urban air quality is just one example where improvements in data have taken place.
3) Big data supports more data integration. Some of the best examples in the presentations focused on integrating different types of knowledge. Through such integration, we may get a better understanding of what it means to think systemically and to build systemic knowledge.
   One of the bigger challenges for the EEA is to deal with the increased variety of data, sources, flows and types of data and to combine those elements in a qualitative and authoritative way.
4) Is big data creating more relevant data? In the context of the SOER we need to understand the fundamental unsustainability of systems of production and consumption and progress to systemic change. Can this be done by combining data on, for example, production, consumption and consumers? These are all elements that are multi-facetted and multi-causal. This systemic understanding needs to include financial data at the European level, which so far has rarely been included into environmental assessments. The possibilities for such new combinations will lead to new ways to analyse.
5) Big data has the potential to be more persuasive if supported by effective communication. The WRI presentation served as a very good example of this. The work by Swedish statistician Hans Rosling uses innovative visualisation to reflect forward-looking ways of dealing with big data. With the big data evolution, visualization gets highly dynamic.
6) Is big data more empowering? Because if our model of change is that better data and better knowledge leads to better policies, we need to provide knowledge that prompts people into action or empowers them. This can take place at the institutional level, but does it also contribute to stronger citizenship or to environmental democracy? International agreements exist, such as the Aarhus convention, which specifically asks for transparent knowledge that empowers citizens.

All these elements can really contribute to the EEA's core business. The EEA has always written a substantial amount of reports. The amount of data we are handling to underpin our reports has grown by roughly a factor of 1 000. This poses the question of what has essentially changed in those

reports, in the access to them and in their usability. The future role of big data has to be identified against this background.

Another key issue is how big data can contribute to improving policies. If the usage of big data leads to changes that bring better policies it can lead to faster reactions. Potential examples are early warning systems or precision farming in Africa, which is now supported by satellites to optimise the usage of seeds.
The systemic polices that we need to support, such as the one on land use, land-use change and forestry (LULUCF), have enormous scope for big data to provide better information.
For areas such as public health and socio-economic data, it is not so clear – there we could do a lot better.
Another model of change is through public pressure. There we need to reflect on what the public nature of big data is and what role the EEA has. We have a strong commitment to reaching out to the public in a transparent way and will have to define how we can utilise big data for this purpose.

Big data and related algorithms could potentially support future oriented work, such as for models and scenarios, but this would need to be further evaluated.

As an official institution, the EEA is obliged to work with legally determined data. In this context, the EEA has to evaluate the big data opportunities and consider potential changes to a data gathering system dating back 30-40 years.

Closing remarks

***Per Mickwitz, EEA Scientific Committee**, provided some closing remarks.*

The relevance of big data will be measured by how we produce better quality knowledge on key issues in the future. There has been a thorough discussion of most of those aspects during the day.

From a scientific perspective, it is important to have three elements that fit together for the production of new knowledge. A first key element, which was strongly addressed, is related to data and methods. Asking the right questions is the second crucial element. The third important element is the theoretical background and the concepts that are being used. These elements need to fit together and it will be a task in the coming years to address each of these in detail for a successful use of big data in future EEA products and services.

Annex 1: Seminar agenda

Joint EEA - Scientific Committee seminar *big data and environmental knowledge*
Tuesday 10 October 2017, 09.30 to 17:00
European Environment Agency, Kongens Nytorv 6, Copenhagen, Denmark

09.00-09.30	**Registration and Coffee**
09.30-09.45	**Welcome and meeting objectives** Hans Bruyninckx, EEA Executive Director Per Mickwitz, Chair, EEA Scientific Committee
Morning session	**Big data - Institutional perspectives – discussion included** Chair: David Stanners
09.45-10.45	• Setting the scene (Chris Steenmans, EEA) • Big data for European policy support (Hannes Reuter, Eurostat)
10.45-11.15	*Coffee break*
Continued	**Big data - Institutional perspectives – discussion included**
11.15 – 12.15	• Sweating the data (making our investments work) (Simon Bingham, Scottish EPA, IMPEL) • Big data after the hype (Sven Schade, JRC)
12.15-12.30	Summarising the discussion on the institutions perspective – what are the new opportunities? (Evelien Dils, ETC WMGE/VITO, Stefan Jensen, EEA)
12.30-13.30	*Lunch*
Afternoon session	**Big data – Science, business and policy perspectives – discussion included** Chairs: Hans Bruyninckx, EEA – Jean-Louis Weber, EEA Scientific Committee
13.30-15.00	• Using and sharing big data for weather, atmospheric composition and climate (Vincent-Henri Peuch, ECMWF) • Amazon and big data – status and outlook (Johan Broman, Amazon Web Services) • Monitoring to Mobilize – big data analytics in practise (Janet Ranganathan, World Resource Institute)
15.00-15.30	*Coffee break*
Round table:	Chair: Leonor Tarrason, EEA Scientific Committee
15.30 – 16.45	• How will big data impact the EEA and Eionet monitoring, reporting and future indicator developments? – strategic and

	practical issues (Round table members: EEA scientific Committee)
16.45-17.00	Closing remarks by Per Mickwitz, chair of the EEA Scientific Committee and Hans Bruyninckx, EEA Executive Director

References

Concept note on EEA Scientific Committee Seminar on "Big data and environmental knowledge", October 2017, 68th Scientific Committee, Doc. EEA/SC/68/Sem 2, European Environment Agency.

IMF, 2017, 'Big data: Potential, Challenges, and Statistical Implications' (https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2017/09/13/Big-Data-Potential-Challenges-and-Statistical-Implications-45106) accessed 5 February 2018.