International Conference on Experimentation in Ecosystem Research in a Changing World: Challenges and Opportunities

ExpeER is constituted of four components (experimental, observational, modeling, analytical platform), each of which is substantial in itself. It will be a serious challenge to make them work together in a sustainable way. It requires vision on multiple levels for this large organization. The costs are high and funding agencies need to be convinced to support cooperation. We need to demonstrate the value of integration in order to show the net benefit for the sites of joining the network rather than working separately. This will also be the mission of the AnaEE preparatory phase.

Data sharing is important for integration. Data openness is the trend.

There is a general consensus that data in the long run should be openly accessible. Though licensing may be a barrier, sharing data should be a direction towards which our community could move.

We can learn from the experience of other research communities, e.g. the molecular biology and genomics community. In the domain of next generation sequencing, data are uploaded and shared in the ANGUS portal. A Next Generation Sequencer is a common equipment generating data in the molecular biology community. We may ask ourselves what is the common thing that we can all use in our community.

We can promote data sharing via funding: in order to get funds, research data should meet a certain quality standard and should be available through a portal.

Another way to promote data sharing is via publication: if all journals make deposited and publicly available dataset an obligation for publication, we can change the mindset around data sharing. Journals need to do it together. Quality control of published data is important. To get around the open data policy, people can publish their data by hiding some figures in their data and providing insufficient metadata.

Some ecological data is basic data in ecosystem research, and is needed in most experimentation (as is the case for meteorological data). They should be provided by a centralized organization. Some data collected by NEON and ICOS have this characteristic. ICOS will completely open its data. Data coming from ICOS towers go to the central repository and can be available the next day after production.

Documentation of data from many EU funded projects is insufficient. Better data sharing allows a better yield for the investment injected into these projects.

Incentives for data openness should be considered.

In the model community, a lot of effort has been made in model development. If modelers cannot benefit from sharing their work, they may lose incentives to support open access and transparency. A market approach may be better than imposing openness as an obligation. The economic value of the data cannot be neglected.

An index for the impact of the published data of a scientist may be a good way to proceed. Acknowledgement of data sources is important for data sharing.

Legacy data need to be accessible. Data interaction and comparison is important.

It was suggested that legacy data produced a hundred years ago be transformed into formats that can be searchable and accessible nowadays. Data format for ecological experiments is more diverse than sequencing or meteorology. A real infrastructure should be a data provider. A good description will give a clear overview of the data and allow users to better assess its quality.

Another opinion is that experiments carried out long time ago may not be correct, due to the insufficient technical level. Creating new data is more important than collecting legacy data. We have more powerful tools and more standardized protocols today. Data interaction and comparison is more important. The infrastructures have an important role in comparing measured data with the model predictions, and comparing data from scattered fields.

Traceability of data is important (data sources and raw data).

The source of data is essential: where they come from, for what purpose they are produced, how they link with each other. Integrating data in a way to predict the whole biosphere will be a major challenge.

Traceability of the data used by models is important. It's often very difficult to trace back the original data used in a model. A lot of product data are freely available today, but the background algorithm and the raw data are not free. If the raw data are wrong in the beginning, all downstream products will be wrong.

How can we deal with data?

We do need tools to manipulate the large amount of data sets.

Most modelers would like to make new models with new data. Applying existing models to existing data becomes quite tricky.

It's not clear enough who the targeted data end-users are, and what their needs are.

We should be more question-driven by asking what needs to be measured rather than driven by what the technology can measure. Otherwise we will end up measuring everything. But the impact of technology on science should not be neglected. New technology can provide a new wave of answers.

Since we don't know what the scientific question of tomorrow is, in some infrastructures, we are generating data whose actual use will only be determined in the future.

Computing power and data repository are essential.

Discovering a new way of analyzing existing data is as important as obtaining data. A repository (data storage) is urgently needed. We need seamless data pipelines that can talk to each other, mass computing power and good databases. It's more complicated for our community but that is what can move us forward. The limiting factor of our research is shifting from measurement techniques to data handling and the ability to ask the right questions.

Gap analysis can be carried out with a matrix approach.

In order to identify the gap, AnaEE uses a matrix approach. 3 major variables are analyzed: primary ecosystems in Europe, climate types, pressures and drivers. By limiting subcategories of variables, a matrix of a few hundred cells was produced. In the next step, cells in the matrix will be prioritized and filled with existing infrastructures. Infrastructures in different countries may cover the same cells in the matrix, so gap and overlap coexist.

In order to identify technological gaps, we can test different models from a simplified system to *in natura* system in various ecosystems, and improve the models with test output.

The approach of TERN (Terrestrial Ecosystem Research Network, Australia) is different. The goal was to cover the entire Australian continent based on 8 to 9 different parameters (vegetation types, landform, soil, etc.). Similar ecosystems were grouped. For example, if there are several functionally similar vegetation types, efforts are concentrated on one of them because of limited funding. Outcome will be extrapolated to the rest. The ecologists and the modeling community are collaborating to refine the gaps step by step.

We should raise the value of ecological data.

Weather data is well valued because it is needed by governments and it has commercial value for industry (e.g. for the aviation industry). Raising the value of ecological data is important. Generating data is the core of our talk. We are going to first make sure that the daily data are useful to ourselves, and hopefully useful for the users in the future. Means of communication will change for the next generation: maybe journals will be replaced by cloud sourcing, and the data value will change.

What can we learn from NEON?

NEON (National Ecological Observatory Network, US) has the advantage that it raises funds in a single country. In Europe, both funding and technical design are fragmented. We should overcome the difficulties through our ambition to catch up or lead research. Each side of the Atlantic has its own advantages. The American budgetary process is quite complicated, research funding in EU in some way is more stable than in the US. Mass investment in research infrastructure will pay off in the next generation.

NEON is spatially very extensive, from Puerto Rico to Hawaii, and from Texas to Alaska. It covers 20 domains, with each domain a set of standardized infrastructures. It was a big investment to build NEON and the operational costs are very high. As the available funding for AnaEE is limited, its boundaries should be clearly defined. Performance tests can be carried out for particular scientific questions before replication to other fields. This will also be the mission of a future European I3 project.

We should link models and experiments. Models should be involved in the early stage of research design and can help us better design follow-up experiments.

Finding the right model that fits the purpose and improving it is more important than building new models. Certain data sets are not compliable for a model and we need more experiments to generate data. In some other cases, we have enough data sets and we need to constrain models.

Broad ranges of observations can generate more data of high quality which allow better evaluation of models. Easier access should be provided. Infrastructures should provide easy tools to help researchers implementing models with data.

Models should be involved in the early stage of research design. They should be then adapted to each particular experiment. Models of a long term experiment can be improved progressively when more data are available. Historical data can be compared with model hindcast. Maybe we don't need to include everything we know in a model. We can bring the model to the scale where it's needed.

The values of models depend on how they are used. With existing data resulting from a terminated experiment, models can help us better design a follow-up experiment.

There are various types of models. A research infrastructure is not built to answer a single question with a single model. The needs of models for a short-term experiment (from a few months to a few years) are diverse and driven by various research questions. It's difficult for a long-term experimental infrastructure to foresee future technique development and adapt to it. We have to be flexible and pragmatic to cope with various experimental designs.

Getting modelers and experimentalists to work together is not an easy job. Each community has its own way of thinking. We agree that new science comes from better integration of different research communities. Each should move out of its comfort zone. For example, modelers have to commit some responsibility for field measurements. Laboratories should make changes of their measurement if the data output is not compliable with the models.

Some constants in models can be updated to variables.

Some constants created in formulas for simplification in early years should be updated to variables now. With bigger computing power, we can find out how the "constants" change.

Cooperation across disciplines is important. Multidisciplinary wprl requires common questions.

It's very important to have data and technique sharing across countries and continents because the challenge is common to the whole world. Predicting the biosphere is a great challenge. We need a well-structured network of networks organized behind the scientific questions.

Talking about a large infrastructure with multiple aims is difficult. We can take a step back to a smaller scale, ask what we want to achieve, identify the missing skills, and find people from different disciplines who have the missing skills. It takes time for a multidisciplinary team to find out a way of cooperating and interacting. A shared goal is essential. In TERN, business analysts were invited to identify the core questions and make sure the whole team understood the common goal.

Difficulties in multidisciplinary cooperation depend on the differences of discipline involved. When natural science community dialogues with social and economic science, it's more difficult. Boundary of discipline evolves too. For example, soil science and crop science is now merged into agro-system science.

Multidisciplinarity requires sharing common questions and balancing different interests. The scope can be extended progressively. For example, the boundaries of the critical zone in ecosystems extend deeper to the soil and groundwater and higher into the atmosphere now, and new disciplines are involved such as hydrology science and soil science.

Involving policy makers, industry, the general public and other stakeholders.

Other stakeholders such as NGOs should be involved in the discussion. In TERN, workshops involving scientists, policy makers and industry were organized. We will have more opportunities to get funding if we better understand what questions the policy makers want answered. The stakeholders and the society should be involved in identifying the common questions around which multidisciplinary groups will work.

This dialogue has its drawbacks as well. The demands of society change fast. For example in hydrology, the demand now is shifting to urban hydrology for green cities. We need some independence to make progress in natural science, and not be only driven by society's changing interests. Society's big questions change, but the subjacent scientific issues don't. For example, acidification was a hot topic in public domain and is not mentioned now. Scientific studies behind the acidification topic continue today under the questions of sustainability and climate change.

The question we want to solve is "what happens to the ecosystem if we manipulate something", "how ecosystems function" and "how they behave under certain scenarios". But for society, the question is "what we have to do", which is rather a social economic question. We can only answer part of their question. And this results in difficulties when we talk with stakeholders, especially policy makers. We need to involve specialists from other domains of society. We have to be aware of the limit of our (research) questions even if these questions are relevant and fundamental.

Scenario approach to dialogue with the general public.

We have multiple questions merging into ecosystem science but no buzzword to tag onto our science.

One way to dialogue with society is to make projections into the future and take some scenarios of "what if" (not predictions) with a particular climate condition. We can find out how the ecosystems behave and what kind of future the society would like. Then we might be able to provide advice on how we can work towards this future from now, what we need to know from our ecosystems now, what skills are needed, how experts should interact, what kind of information should be shared. This exercise can be carried out with broader stakeholders.

In conclusion: much done, much to do

Either we can continue business as usual, or we can operate differently. In order to build infrastructures, we need cooperation and integration, not only pure competition.

Society does want to react to the global challenge. There is progress, but due to inertia, it's slow. The world is evolving quickly and we need to react quickly.

Long term research infrastructure is a good tool to face the challenges, to provide answers. It also provides jobs to society. Infrastructures need to work hand-in-hand with stakeholders.

What is the next step? What is our horizon? How can we assimilate all the information exchanged during the conference? It's always a challenge to continue, to change and to adapt. This conference shows the great interest of the scientific community in infrastructure as a tool to provide new sciences. We hope participants found the conference fruitful.