

General requirements of ANAEE

Context of general requirements in ANAEE

Complete report on ANAEE general requirements is available at: <https://envriplus.manageprojects.com/projects/requirements/notebooks/470/pages/37>

Summary of ANAEE general requirements

Detailed requirements

AnaEE (analysis and experimentation on ecosystems) is a research infrastructure that focuses on providing innovative and integrated experimentation services for ecosystem research. From [1]:

AnaEE will be a research infrastructure for experimental manipulation of managed and unmanaged terrestrial and aquatic ecosystems. It will strongly support scientists in their analysis, assessment and forecasting of the impact of climate and other global changes on the services that ecosystems provide to society. AnaEE will support European scientists and policymakers to develop solutions to the challenges of food security and environmental sustainability, with the aim of stimulating the growth of a vibrant bioeconomy. AnaEE will accomplish this mission by building permanent and substantial links among researchers, science managers, policy makers, public and private sector innovators, and citizens.

Services will be provided to provide access to research facilities and their data, to promote exploitation and reuse of that data, and to provide high value analytical and modelling services. An emphasis is placed on world-class capabilities, providing the best facilities in order to better forecast the impact of global changes and feed into public policy.

Operation

AnaEE is being built upon existing infrastructure, but seeks to upgrade that infrastructure to handle the demands of innovative ecosystem research which the existing structures and resources cannot fulfil. AnaEE is currently in its preparatory phase. Current effort is being placed in the construction of infrastructure 'bricks' comprising vertical chunks of the infrastructure by locality (e.g. AnaEE in France)—policies, services and technologies are deployed in one brick in order to determine how best to proceed for the greater infrastructure.

As a distributed infrastructure, AnaEE will be built on experimental platforms of four types: *in natura*, *in vitro*, analytical and modelling, that can be linked together to provide a single data workflow through all the stages of experimentation.

Data is distributed, with different facilities belonging to different institutions. A portal with the ability to identify all data held in AnaEE is planned, based on the use of unified metadata to characterise data sets and direct users to the right facilities.

Data centres are provided by different institutions at a national level to secure data in the longer-term, using national data infrastructures or other dedicated service units. A European data modelling centre is foreseen to provide (among other services) some backup if individual institutions fail (but this has not been decided).

Data and computation

There are two basic classes of experiment that AnaEE needs to support: short-term and long-term. Short-term experiments are implemented on AnaEE experimental platforms with flexible design in relation with the scientific objectives and the platform capabilities. Each experiment lead to a specific data base gathering both core measurement delivered by the platform and specific data collected by the scientific user. Long-term experiments are based on a core experiment producing a continuous flow of data hosted on local data base for processing and quality check and then transferred to a central database hosted within the AnaEE infrastructure for backup and dissemination purposes. There exist a wide variety of experimental models, and AnaEE intends to be flexible regarding how researcher interact with the infrastructure.

As a general policy, data is open, a requirement of participation in AnaEE. There are no particularly sensitive data products that demand restricted access. The producers of data require that licenses for data use are adhered to, and proper attribution is given. Some constraints on data, defined in AnaEE data license, would then be required to ensure that the producer is properly acknowledged and that the data is not 'stolen' or otherwise abused. Private companies may access to platforms at a full cost rate, which leave them the possibility to control the data dissemination as they would own the data. Academic users are charged at marginal cost and then have to disseminate the data according to the AnaEE dissemination rules as mentioned above.

The use of academic embargos is already well established as a procedure to secure the cooperation of data producers, and as part of the quality assurance process. Within AnaEE-France, it is possible to define embargos of varying periods on individual data sets—a similar set up is envisaged at the European level.

As AnaEE is only in its preparatory phase, there are no specific technological choices in play at a European level yet; however the emphasis is on achieving interoperability between data centres and institutions via a semantics-driven, web services approach, rather than enforcing specific software choices.

In AnaEE-France, pilot operation has been currently developed. A metadata standard for 1st level data products exists based on ISO 19115, compatible with the INSPIRE directive; management tools for this metadata are being developed. Common standards of vocabulary for environment and ecosystem data are desired in order to ensure that everything that might be described can be described. OBOE is used for ecology data, extended for the AnaEE context. The SSN ontology is used for describing sensors. Information about the nature and existence of data sets in the infrastructure (2nd level metadata) could be provided to a central portal to permit easier data discovery and access.

Within the context of the modelling platform, it is necessary to perform annotation of resources using the same ontology.

AnaEE does not directly provide access to services for computation, but rather a platform for use by researchers with tools that can be imported into the researchers' own environments.

AnaEE has a concept of 'modelling factories', which host models for use by researchers, and provide tools for producing new models or coupling models together into portable units and implementing facilities (link with data base, data assimilation ...). The idea is to provide advanced and flexible modelling tools to either experimentalist to improve the quality of the experiment (design, quality check, data interpretation, generalisation) or any modeller who wish to take profit of the high quality data produced by the infrastructure.

Dictating specific technologies for modelling factories is not a focus on AnaEE; maintaining support for factories and training researchers in their use is considered to be more important, with the overall strategy being one of interoperation.

The licenses for modelling factories depend on the factory—the core part is open source, but some parts may not be; there may be conditions on use of the factory, most commonly the need to (perhaps after a suitable embargo period) share scientific results.

Any software developed will be made available, open in use, but not always open source. The critical philosophy is to make all tools available to the community, but the specifics not yet fully defined in all cases—for example, it would not generally be possible to use tools or data for commercial services without some kind of special arrangement.

AnaEE expects to provide technical training for the platforms offered, and already provides some training for the models provided to researchers; training aimed at a broader section of society is also a possibility.

Interaction with ENVRI+ and other initiatives

It is the intention of AnaEE to provide excellent platforms with clear accessibility conditions and service descriptions, and a clear offering to researchers. The gathering of information in a common portal should help with this. Experiences gathered from the construction and operation of other platforms would be helpful to shape development.

AnaEE seeks to interact with all major ESFRI ecosystem/biodiversity research infrastructures, notably ICOS, LifeWatch and LTER. ENVRI+ is considered to be a good opportunity to enhance interaction and cooperation, both with the aforementioned infrastructures, and with infrastructures in other environmental science domains.

Operationally, there is already overlap between sites that contribute both to AnaEE and to ICOS. There is a need to share data management strategy and technical references. There is also a strong desire to enhance semantic interoperability, by the adoption of (for example) common metadata and service standards, and to develop a standard common view of data in general.

Within the context of ENVRI+, the AnaEE project is particularly interested in participating in the work on identification and citation, and on cataloguing as this is of fairly immediate concern to the infrastructure, and so it would be useful to synchronise approach with other research infrastructures. Processing is of some interest as well; particularly the interoperability between models and data, and the quality control of data produced by platforms.

Further notes

Being in the preparatory phase the data management plan for all partners of AnaEE is still under development; however an integrated procedure both for data access and modelling is in place in AnaEE-France, which may be scalable to the pan-European context. Adopting well-defined citation schemes, semantics and controlled vocabularies in common with other related research infrastructures is important.

Internally, AnaEE need to identify high-quality sites throughout Europe to integrate into its infrastructure, and has to measure (and acquire) the required investment to realise this. Strong governance processes need to be established to manage an infrastructure that crosses national boundaries.

It is not yet clear which organisations will be part of AnaEE after its preparatory phase completes, which limits the ability to make some decisions about (for example) data management. Funding for parts of an AnaEE ERIC may be provided by individual countries, but a European-level funding source for AnaEE activities is desirable.

Externally, linking both technically and organisationally with other infrastructure initiatives in adjacent domains is desirable to maximise research potential.

References

[1] AnaEE—analysis and experimentation on ecosystems, September 2015. <http://www.anaee.com/>, accessed 17th September 2015.

Formalities (who & when)

Go-between	Paul Martin
RI representative	Abad Chabbi, André Chanzy and Christian Pichot
Period of requirements collection	
Status	