Accounting and Data Protection

Communities of researchers and students, as well as the research and computing and data infrastructures, need to be able to process data and meta-data about the users and their interaction with the systems. This is essential for accounting and for assigning use data to allocations, and to be able to follow up on incidents in the infrastructure. Before recommendation can be developed, it is necessary to make an inventory of the relevant use cases, identify the types of data generated within the infrastructure as a result of its use, and the respective roles of the participants in the infrastructure with respect to data protection. Using the identified roles and responsibilities (and taking into account the wide diversity in laws and regulations related to personal data protection throughout Europe!), we can then develop recommendations and template policies for the processing of personal data for each of the identified participants with the aim of providing recommendations that can be applied across the entire infrastructure.

This work in done is close collaboration with the e-Infrastructures (PRACE and EGI in particular), and with reference to organised user communities, since it is likely that policies and frameworks adopted at that level will gain sufficient traction in processing centres that it will have a harmonising effect across the European e-Infrastructure area.

- Requirements on data to protect from AAI, community, resource providers and e-infrastructure (MNA3.2)
- Recommendations and template policies for the processing of personal data by participants in the pan-European AAI (DNA3.5) - editable version available

With the new general data protection regulation (GDPR) - formally published on May 4th, 2016 - the context is shifting: Although on the one hand the regulation brings the advantage of better alignment in Europe, it also places some further limits on the ground for data processing, and for global collaboration there are many factors (including the Privacy Shield work) that drive change. In the 2nd year of AARC we investigate new approaches that align with the new environment - and still keep the collaboration from within Europe with the world at large - including approaches inspired by the 'corporate rules' mechanism and its potential applicability to coordinated e-Infrastructures and research infrastructures.

- Data Protection Impact Assessment - and its application to communities and proxies in the BPA ("AARC2/DNA3.1 - Report on the coordination of accounting data sharing amongst Infrastructures (initial phase)")

Under current legislation, only Model Contracts and Binding Corporate Rules appear to offer the framework required to transfer personal data within trans-national science e-Infrastructures. With hundreds of resource providers and user communities potentially exchanging data, it is impossible to conceive of each party executing a separate, legal agreement with all others as might be required by the standard use of Model Contracts. One possible solution is where each party would sign an adherence form acknowledging compliance with a Code of Conduct (as referred in GDPR Article 46.2(e)) . The principle proposal here is the GEANT Data Protection Code fo Conduct. It in itself already provides best practices and excellent guidance on how a service provider (and infrastructure) should conduct itself - e.g. including verbatim the Sirtfi requirements. If and when this DPCoCo can actually be an article 46.2(e) basis still remains open: the EDbP still has to be established before the process can even start.

So for now, we propose the BCR-inspired model as presented above as a suitable basis for distributed collaborative infrastructures where many independent organizations (with the user communities and their members represented in their professional capacity by their home organizations) collaborate within a well-controlled policy framework - which is a characteristic of most of the cross-national Infrastructures and the AARC selected use cases. For reference, the policy template *Policy on the Processing of Personal Data* developed jointly with EGI, WLCG, and GridPP, has been appended to this *Recommendation*.

- Recommendations and template policies for the processing of personal data (DNA3.5)

We would like to point out that this effort is focused on the protection of personal data that is generated as a result of participants working within the federated infrastructure. If you are about to exchange (research) data that in and of itself contains (sensitive) personal data, or where the combination of your data set with other data sets can result in the inference of personal or sensitive data, including other research data, please consult with your user community. E.g. for biomedical research data, look at ELIXIR and national initiatives.